Mustafa ABABNEH THE THESIS TITLE PhD THESIS 2023

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

DEVELOPMENT AND TESTING OF PERFORMANCE SCALE APPLICATION AS A LEARNING SUPPORT FOR STUDENTS IN SECONDARY SCHOOL

Ph.D. THESIS

MUSTAFA ABABNEH

Nicosia

February 2023

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

DEVELOPMENT OF PERFORMANCE AND TESTING SCALE APPLICATION AS A LEARNING SUPPORT FOR STUDENTS IN SECONDARY SCHOOL

PhD THESIS

MUSTAFA ABABNEH

Supervisor

Prof. Dr. Fezile Özdamli,

Co-Supervisor

Assoc. Prof. Dr. Damla Karagozlu

Nicosia

February 2023

Approval

The Thesis defence was held online. The Jury members declared their acceptance verbally, which is recorded.

Examining Committee	Name-Surname	Signature
Head of Committee:	Assist. Prof. Dr. Selin UZELALTİNBULAT	
Committee Member:	Assist. Prof. Dr. Mustafa BABAGIL	
Committee Member:	Assoc. Prof. Dr. Boran SEKEROGLU	
Committee Member:	Prof. Dr. Hüseyin BİCEN	
Supervisor:	Prof. Dr. Fezile ÖZDAMLI	
Co-Supervisor:	Assoc. Prof. Dr. Damla KARAGOZLU	

Approved by the Head of the Department

20/3/2023

Prof. Dr. Nadire Çavuş
Head of the Department

Approved by the Institute of Graduate Studies

/ /2023

Prof. Dr. Kemal Hüsnü Can Başer
Head of the Institute

Declaration

I hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Mustafa ABABNEH

9/2/2023

Acknowledgments

This research project would not have been possible without the help, support and patience of my supervisor Prof. Dr. Fezile Özdamli and my Co-Supervisor Prof. Dr. Damla Karagozlu. Her experience was invaluable in formulating the topics and methodology of the research. Her insightful remarks prompted me to refine my thinking and take my work to a higher level. I am thankful for her valuable guidance throughout my studies and her feedback that gave me the right experience that enabled me to choose the right direction and successfully complete my thesis.

I am also pleased to thank all the faculty members that have taught me during my coursework at the esteemed Department of Computer Information Systems, Near East University. Then I would like to express my appreciation to my colleagues and fellow researchers from my department for their wonderful cooperation and support.

In addition, my gratitude goes to my family, who supported me and motivated me, as well as finding factors and opportunities to make me confident in putting my mind at ease during the completion of my PhD study. I would also like to thank my friends for their wise advice and friendly support. They have always been my first support for me. I would especially like to thank my lovely wife and daughter for their personal support and great patience at all times.

Special thanks to the thesis defence Jury members for their understanding and insightful comments that improved the quality of the thesis document and shed more light on the research endeavour.

Mustafa ABABNEH

Abstract

Development and Testing of Performance Scale Application as a Learning Support for Students in Secondary School

Mustafa ABABNEH

PhD, Department of Computer Information Systems

February 2023, 259 pages

The academic achievement of secondary school students is impacted by factors such as a lack of self-learning, appraisal, fulfilment, and the inability to perform well in exams. The inability of Jordanian students to select the right specialty during their secondary education may also have an impact on their performance and, in turn, their academic achievement. These constitute the key problems constraining secondary schools in meeting students' needs and developing effective instructional tools, which encourage the students' poor achievement or failure. Mobile learning apps provide students with an opportunity to make the right choice, plan their time and effort in the learning process. Performance scale applications (PSAs) are still underutilized in the classroom, in spite of the rapid advancement of smartphone and e-learning technology. The main objective of this study is to develop and evaluate PSA as a skill support for students in secondary school. The PSA was developed on the Android smartphone operating system using Java to predict secondary school student achievement. Three specialties: scientific, literary, and industrial, were employed. The participants consisted of 60 students in one experimental study and 75 in another study at Irbid Secondary School in Jordan. In addition, 15 teachers were sampled for their opinions towards PSA. The variables measured include improving evaluation (IME), improving scientific (IMSC), improving communication (IMC), and satisfaction of learning (SOL). The results showed that PSA correctly indicated the students' preference for specialist, IMC, IMSC, SOL, personalized learning (L), distance learning (L), mobile learning (L), self-learning (L), and specialty L. The results revealed a significant and positive effect of PSA on students' learning achievement in secondary school. Among the tested ML algorithms, extra tree regression was found to be an effective tool and was used to predict GPA in PSA. The results showed that the means for the experimental and control groups were comparable in the pre-test; however, they were higher in the post-test for the experimental group. The averages of the experimental groups were superior to the control groups in each post-test. Moreover, the application was effective in the scientific and literary specializations but ineffective in Arabic. In conclusion, the performance scale application can efficiently predict the students' academic achievements and choice of specialty in secondary schools. Future study should consider additional functionalities of the PSA that are consistent with the findings of this study.

Key words: mobile learning, performance scale application, secondary school, students' academic achievement, ML algorithm.

Öz

ORTAOKUL ÖĞRENCİLERİNİN ÖĞRENME SÜRECİNE DESTEK OLARAK PERFORMANS ÖLÇEĞİ UYGULAMASININ GELİŞTİRİLMESİ VE TEST EDİLMESİ

Mustafa ABABNEH

Doktora, Bilgisayar Enformatik Anabilim Dalı

Şubat 2023, 259 pages

Ortaokul Ortaokul öğrencilerinin akademik başarıları, kendi kendine öğrenme eksikliği, değerlendirme, memnuniyet ve sınavlarda başarılı olamama gibi faktörlerden etkilenmektedir. Ürdünlü orta eğitim öğrencilerinin öğrenimleri sırasında doğru uzmanlık alanını seçememeleri de performanslarını ve dolayısıyla akademik başarılarını etkilemektedir. Mobil öğrenme uygulamaları, öğrencilere öğrenme sürecinde doğru seçimi yapma, zamanlarını ve çabalarını planlama fırsatı sunar. Cep telefonu ve e-öğrenme teknolojisindeki hızlı gelişmelere rağmen, sınıfta performans ölçeği uygulaması (PSA) kullanımı hala yeterince kullanılmamaktadır. Bu çalışmanın temel amacı, ortaokul öğrencilerine yönelik bir beceri desteği olarak PSA'yı geliştirmek ve test etmektir. PSA, ortaokuldaki öğrenci başarısını tahmin etmek için Java dili kullanılarak Android mobil işletim sisteminde geliştirilmiştir. Çalışmaya 11. sınıf temel uzmanlık öğrencileri katılmıştır. Katılımcılar, Ürdün'deki Irbid Ortaokulunda 60 ve 75 kişilik deney ve kontrol olmak üzere iki gruptan oluşmaktadır. Ayrıca, PSA'ya yönelik görüşleri için 15 öğretmenden görüş alınmıştır. İncelenen değişkenler, değerlendirmeyi geliştirmeyi (IME), iletişimi geliştirmeyi (IMC), bilimselliği geliştirmeyi (IMSC) ve öğrenme memnuniyetini (SOL) içerir. Bulgular, performans ölçeği uygulamasının öğrencilerin uzmanlık, IMC, IMSC, SOL, kişiselleştirilmiş öğrenme (L), mesafe L, mobil L, kendi kendine L ve uzmanlık L seçimini doğru bir şekilde öngördüğünü göstermiştir. Bulgular, performans ölçeğinin olumlu ve anlamlı bir etkisi olduğunu ortaya koymuştur. Ortaokul öğrencilerinin öğrenme başarısı üzerine uygulamadan elde edilen ML tutmaları arasında ekstra ağaç düzenlemesi etkili bir araç olarak PSA'da genel başarı ortalamsını tahmin etmek için etkili bir sonuç olarak belirlenmiştir. Sonuçlar, deney ve kontrol grupları için ortalamaların ön testte karşılaştırılabilir olduğunu gösterdi; ancak deney grubu için

son testte daha yüksektiler. Her son testte deney grubunun ortalaması kontrol gruplarından daha yüksek olarak belirlenmiştir. Ayrıca uygulama ilmî ve edebî ihtisaslarda etkili iken Arapçada etkisiz kalmıştır. Sonuç olarak, performans ölçeği uygulaması, ortaokullarda öğrencilerin akademik başarılarını ve uzmanlık seçimlerini etkili bir şekilde tahmin edebilir. Gelecekteki araştırmalar, bu çalışmanın bulgularıyla tutarlı olan PSA'nın ek işlevlerini dikkate almalıdır.

Anahtar kelimeler: mobil öğrenme, performans ölçeği uygulaması, ortaokul, öğrencilerin akademik başarısı, ML algoritması.

Table of Contents

List of Tables

List of Figure

List of Abbreviations

ANN:	Artificial Neural Network
BRR:	Bayesian Ridge Regression,
CFA:	Confirmatory Factor Analysis
GPA:	Cumulative Grade Point Average
DT:	Decision Tree
ET:	Extra Trees
GB:	Gradient Boost
IME:	Improving Evaluation
IMSC:	Improving Scientific
KNN	K Nearest Neighbour
L:	Learning
LR:	Linear Regression
PSA:	Performance Scale Application
RF:	Random Forest,
SOL:	Satisfaction of Learning
SVM:	Support Vector Machine
AVE:	Variance Extracted

CHAPTER I

INTRODUCTION

1.1 Introduction

This chapter introduces the study's subject. The study's background, problem statement, research questions, objectives of the investigation, study scope, and significance of the study are all included in this section. It also highlights about the study's contribution and, last but not least, its limitations.

1.2 Background of the study

In today's technological and data-driven world, virtual devices, such as smart devices include computers, pad, and smartphones, are frequently used. These tools gather enormous amounts of diverse data, which can help researchers identify previously unidentified patterns and orientations (Chevet et al., 2022). Such data propagate quickly, which prompts the development of new methods for enhancing and evaluating big data (Svennberg et al., 2022). The methods used for data analysis are divided into three categories: statistics methods (Almaiah et al., 2022), artificial intelligence methods (Berei & Pusztai, 2022), and machine learning methods (Rathod, Kulkarni, & Saha, 2022). All of these methods are used to extract patterns, predict behaviors, or describe directions based on desired outcomes (Ofori, Maina, & Gitonga, 2020).

Digital revolution cut across all sectors of life. Education is deemed the most conspicuous area influenced by this shift at all levels. Education establishments, including elementary schools, secondary schools, higher institutions, and online educational service sources, have recently improved their capacity to gather, use, and communicate data quickly and easily (Malkawi, 2020; Alhasani & Orji, 2022). Today, educational data mining and learning analysis play a significant function in coordinating the development of the educational practice, with the data focusing on describing the acquisition of valuable knowledge (Pallathadka et al., 2021). Make smart interventions for the process in order to apply the knowledge learned and utilize it to anticipate a variety of significant educational outcomes, such as performances (Amjad et al., 2022). Regarding learning analysis, it concentrates on using data analysis to implement the learning process with optimism (Qiu et al., 2022). It is really beneficial to have learning analysis data that is focused on developing and making some choices on both the teaching and learning processes (Du et al., 2019). As a result, data analysis can have a significant impact on everyone involved, starting with educators, students, leadership trainers, and school management.

The internet as a tool for information search, eLearning, and teaching platforms are only a few examples of recent technology advancements in education (Almaiah, Al-Khasawneh, & Althunibat, 2020; El Gourari et al., 2020). Mobile learning platforms, websites, and educational platforms have all been useful in advancing secondary education and enhancing student performance (Kew & Tasir, 2022). Mobile technology assists pupils in both interpreting difficult abstract concepts and supporting the construction of those conceptions in their minds by using its identifiable multidimensional tools and technical enablers (Siwawetkul & Koraneekij, 2020). The phrase "mobile learning" describes a method of instruction that makes use of mobile devices and the services they provide, including Wireless Application Protocol (WAP), application software, and other technologies (Astuti & Suratman, 2022). As a result, mobile learning apps give users the chance to schedule their time and effort for learning (Shams, Ajmal, & Jumani, 2021). Additionally, it promotes and offers the methods that students could apply while learning, which inspires them to be more imaginative. In order to enhance the conventional learning process, mobile technology should be employed in the classroom, which is more practical and engaging learning environment for the modern electronic lifestyle (Klimova, 2019). Studies have demonstrated the benefits of using mobile learning tools to instruct many ideas, including how they have enhanced student engagement, choice, and capacity for making important decisions while also enhancing the study of many courses in schools (Rohmah, Pramono, & Yusuf, 2020; Pilatso & Chukwuere, 2022). Teachers and students now have easy access to a variety of information and instructional resources because to the advancement of information technology (Bossman & Agyei, 2022). The technology assisting students with their learning process and utilizing various learning techniques, such as digital devices, to enhance student performance in distance learning, self-learning, mobile learning, and personalized learning (Nawar & Mouazen, 2017; De Backer et al., 2022).

Student academic accomplishment is a crucial component of education (Ozdemir & Ozturk, 2022). It is believed to be the focal point around which the whole learning system spins. Academic success determines whether students succeed or fail in school (Cachia, Lynam, & Stock, 2018). Along with the main purposes and objectives of the school and the most influential instructors in the classroom, the significance of students' academic advancement is also emphasized (Hines et al., 2020). Learning success can also be defined as the knowledge obtained during a predetermined period that can be assessed by a teacher using grades and school policies. Students receive useful tools through education in addition to the knowledge and skills they need to overcome obstacles in their academic careers. They have to be competent at applying what they have learnt in class to actual circumstances (Ztürk, 2017). It's important to teach students how to comprehend the course information, particularly how to translate a metaphorical idea into concrete information.

The conventional educational model, which emphasizes memorization, in-person instruction, putting the instructor at the heart of the educational process, and using books as the main source of knowledge, is not now the most successful model. It is rather, a new paradigm for distance learning tools, like performance scale applications or mobile applications, has emerged as a result of the advancement of technology (El Gourari et al., 2020). In order to effectively deliver distance and classroom learning to students, mobile education relies on the usage of eLearning technology and applications (Amjad et al., 2022). With the use of performance scales, students are better able to make wise decisions for themselves, study for themselves outside of the classroom, communicate clearly with their lecturers and other students, and apply what they are learning to practical circumstances.

1.3 Statement of the Problem

The use of performance scales in the classroom for secondary school students is a relatively new concept and approach in teaching and learning in Jordan. It is not a novel notion to teach students through the use of visual aids (Sharma et al., 2018; Al Zieni, 2019), but using mobile apps to aid students in self-learning and moral decision-making enhances their academic performance. Although mobile apps have been used in educational settings for decades (Kattayat, Josey, and Asha, 2017; Ugur-Erdogmus & Akr, 2022), Secondary school pupils' academic performance and self-learning are still issues. This is due to the fact that secondary school students in Jordan select their area of study or class, whether it be a scientific, literary, or industrial field (Malkawi, 2020). Their poor choices have an effect on their academic success. Secondary school students still struggle with self-learning, which has a big impact on how well they do in school. There was no research that seeks to help Jordanian secondary school learners select their academic specialty. Studies have generally revealed that university students choose their medical specializations (Goldenberg, Williams, & Spollen, 2017; Jordan & Duckett, 2018).

Even though literature has long recognized the value and importance of integrating technology in educational settings, particularly mobile and Android applications, to achieve a range of goals like adapting to change, boosting students' academic performance, motivating students to learn, etc. (Han, 2022; Nafasov et al., 2022; Ugur-Erdogmus & Çakır, 2022), however no study has focused on choice of specialty. The main debate focused on the utility of mobile applications in distance learning (Shams, Ajmal, & Jumani, 2021), self-learning (Sayibu et al., 2021), mobile learning (Chao, 2019; Onuba, 2022), and personalized learning (Xie et al., 2019; Lin, Zhao, Liu, & Pu, 2020).

In Jordan, many students continue to select majors that are not a good fit for their aptitudes, which has a hugely detrimental influence on their prospects in science and other fields, despite the diversity and depth of studies undertaken to improve learning management systems. Many students, particularly those that choose the wrong field in secondary school, end up with very poor grades or even flunk their major as a result of their poor decision (Ogowewo, 2010; Sattin-Bajaj et al., 2018). Middle Eastern nations, like Jordan, continue to encounter several difficulties with the current teaching and learning systems that help teachers place secondary school students in various departments, as well as a method to help schools assign students to each department in in line with their scientific aptitudes (Malkawi, 2020). Hence, this period marks the actual start of students' future educational and professional careers. Therefore, The aim of this project is to develop a new method for evaluating a student's aptitudes and preferences in the eleventh grade using information and grades from all subjects from the seventh to the eleventh grade.

1.4 Research Objective

The key objective of this work is to develop and test performance scale application as a skill support for students in secondary school.

The following are the specific objectives that support the key one above:

1. To apply machine learning algorithms with the highest predictive efficiency for performance-scale applications, such as artificial neural networks, extra-tree regression, logistic regression, nave Bayes, K-Means clustering, and decision trees.

2. To investigate the effect of the developed performance scale application on students' academic achievement.

3. To investigate the effectiveness of designed performance scale application in improving students' choice of specialties (including scientific, literary, and industry) and academic achievement.

4. To determine the performance scale application in improving students' scores in selected specialties.

5. To determine a correlation between the results of the pre-test and the post-test in all groups and subjects.

6. To examine teachers' perceptions towards performance scale application to effectively improve students' ability to make a choice and their academic achievement.

7. To determine the efficiency of performance scale application to effectively improve the evaluation (IME) ability of students in secondary school.

8. To determine the efficiency of performance scale application to effectively improve the communication (IMC) ability of students in secondary school.

9. To determine the efficiency of performance scale application to effectively improve the scientific (IMSC) ability of students in secondary school.

10. To determine the efficiency of performance scale application to effectively improve the satisfaction of learning (SOL) ability of students in secondary school.

1.5 Research questions

The following research questions guide the objective of this study:

1. Which machine learning algorithms have high predictive efficiency in building performance-scale applications?

2. What is the effect of the developed performance scale application on students' academic achievement?

3. Does the effectiveness of designed performance scale application enhance the students’ choice of specialties and academic achievement?

4. Does the performance scale application effectively improve students’ scores in selected specialties?

5. Does a correlation exist between the results of the pre-test and the post-test among the groups and subjects?

6. What are the perceptions of teachers towards performance scale application to effectively improve students’ choice and their academic achievement?

7. Does the efficiency of performance scale application effectively improve the evaluation ability of the students?

8. Does the efficiency of performance scale application effectively improve the communication ability of the students?

9. Does the efficiency of performance scale application effectively improve the scientific ability of the students?

10. Does the efficiency of performance scale application effectively improve the satisfaction ability of the students?

1.6 Significance of the Study

Developing mobile apps was previously thought to be a possibility that called for a particular amount of software development expertise. However, nowadays such perception has changed with progressive implementation of mobile apps in education and academic related usage. Hence, the development of performance scale application can help to the students to have better experience of academic and improve their learning and performance.

Additionally, one of this study's major contributions is its support for efficient instructional teaching via mobile studying application. There is no research on performance scale software concentrating on mobile learning system and secondary school students' academic achievement in Jordan. Consequently, this work will thereby contribute to the body of knowledge. The creation of PSA using user interface concepts and ML methods aims to assist secondary school students with their academic work, such as class assignments, self-learning, self-evaluation, and selection of specialty. By doing this, the student can make the right specialization decision, increase their interest in and participation in class, improve their exam results, and enhance their overall secondary school learning performance. The results of this study can help schools use the curriculum effectively and make the best use of their time to improve student learning. Instead of spending time ineffectively teaching students to make judgments or placing them in the wrong speciality or area of interest, which leads to failures and weak achievement.

The study's goal is to address a problem that pertains to secondary school pupils that has not receive enough attention in previous studies. Providing a mobile app will make it easier to guide students. The performance scale application aids in making forecasts for the future and investigating numerous patterns, enabling timely and accurate decision-making among students. Ensuring that the students are guided effectively by selecting the right specialties for them based on their scientific knowledge and expertise would help them avoid failing. The technique can be expanded upon and used in various disciplines. The suggested method is simple to integrate with the existing academic information systems using a student's record to evaluate their performance. The capacity to examine trends and zero down on the most crucial data in the student's file is something that five years intends to offer.

1.7 Contributions of the Study

The previous systematic analysis of the literature revealed that there are not many studies that focus on secondary school students and helping the students to decide on their academic specialties. In addition, the application of artificial intelligence techniques, particularly data mining and machine learning, is unquestionably a large area that is crucial to the area of education and is regarded as the newest technology. In this study, uses ML approach to help students choose appropriate scientific specializations in secondary school will add to the knowledge of the literature and solve the students' academic achievement.

E-learning systems have also been the focus of several studies to forecast student performance and student assessment (Almaiah, Al-Khasawneh, & Althunibat, 2020; El Gourari et al., 2020; Kew & Tasir, 2022). Utilizing machine learning algorithms within the context of e-learning to appraise students' performance and support them to become more successful was the subject of some studies (Ofori, Maina, & Gitonga, 2020; Yadav et al., 2020; Ahsan et al., 2021; Sandra, Lumbangaol, & Matsuo, 2021). Finding of this study will show the prevailing trends in students' academic data over the course of five years, as well as demonstrating the potential benefits of data mining algorithms for performance scale application.

There have been numerous studies conducted in the Middle East, including Jordan, that aim to predict success using past achievement rates, grades, or treatment skills. These studies more indicators of the students' performance to forecast their success at the university level, which make the outcomes to have poor precision and lack of effectiveness in implementations in schools and classrooms (Ariffin, Sidek, & Mutalib, 2018; Jordan & Duckett, 2018; Tawafak et al., 2020). We do not utilize additional information, such as age, gender, or homework, and only use the grades students achieve in past performance evaluations, such as the prior 5-year GPA with the level test score. Because of this, our algorithm may be used in any hands-on, in traditional classroom, or online learning environment. This serves as another contribution.

We found that the application of artificial intelligence and ML methods in the area of education is significant and pervasive after reviewing prior systematic reviews. This is because there has been a lot of study on applying artificial intelligence approaches to support the instructive system and boost academic achievement rates by creating tools that evaluate students and predict their academic achievement using ML algorithms (Halde, 2016; Sandra, Lumbangaol, & Matsuo, 2021; Afshar et al., 2022). The contribution of this study will be outcome from the test results of machine learning methods including linear/logistic regression, k-nearest neighbors, extra trees regression, ANN, and decision trees, among others, are used in data mining as standard algorithms, to test against secondary students' data. This information will be valuable for researchers and school management and regulations in Jordan and other parts of the world.

Previous studies demonstrated that strategies for determining precise steps to improve student performance are among the most crucial ways to attain academic achievement (Batool, 2020; Ozdemir & Ozturk, 2022; Ugur-Erdogmus & Çakır, 2022). Since this study is aiming at accuracy, data from five previous years in addition to current data against the algorithm are used to predict the students' choice, self-learning, specialty, and academic achievement more accurately.

1.8 Summary

This study has explained the performance scale application and its related concepts and methods in the background, with problems defined and objectives enumerated based on the problems identified. The research questions were also listed, which guided the objectives of the research. The significance, contributions, and limitations of this study have also been discussed.

This study has also explained that the use of PSA in the classroom for secondary school students is a relatively new concept and approach in teaching and learning in Jordan. Although mobile apps have been used in educational contexts for decades, there are still problems with self-learning and academic success. By assisting students in both decoding complex abstract notions and helping them organize those concepts in their minds, mobile technology plays a part in the learning process. It also offers students the methods that they could apply while learning, which inspires them to be more creative.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

This part presents and discusses the relevant literature in relation to the topic and objectives of this study. The literature covers the basic concepts, definitions, and descriptions of mobile applications in education, e-learning, and technology, and the methods involved in the selection, categorization, and analysis of digital devices, such as performance scale applications.

2.2 Mobile application in education

The development of information and communication technologies has led to development in many fields, including the educational field. It is no longer appropriate to use the conventional educational model, which relies on memory, indoctrination, reliance on the instructor as the center of the educational process, and the book as the main source of knowledge without the teacher. Instead, a new paradigm of distant learning was created, and this was attributed to the technology revolution and development (Safitri et al., 2021). The usage of computers and regional and international communication networks in education is attributed to the electronic revolution. Due to the introduction of the e-learning model, which made face-to-face distant learning possible, and the wireless revolution, a new model has now emerged. This is mobile learning, often known as mobile learning for training and distance learning that depends on wireless technologies. This featured cell phones, PDAs, and laptops, which caused a change from a wired learning environment to a wireless learning environment (Al-Rahmi et al., 2022).

The term "educational" is wide concept consist of many different types of software that fall under the category of education when it comes to mobile apps (Ali et al., 2022). Some educational apps are made for secondary school students, and student grades, while others are better suited for educators. Mobile technology is a great invention in 21st century with most ground-breaking productivity tools (Papadakis, 2022). In education, new knowledge and new abilities can be developed using smartphones and tablets. The quality of an educational app is a crucial factor in the constantly evolving world of teaching and learning, and it can significantly improve the educational experience of the student (Voshaar et al., 2022).

Due to their emphasis on modern communication techniques and technological advancements, many technologists and educators conflate the ideas of M-learning and eLearning. M-learning, also known as mobile learning, is defined as earning in numerous contexts, through interactions with people and content, using personal electronic devices (Hartley & Andújar, 2022). M-learning is a new method of leveraging mobile devices to access educational content. M-learners employ educational mobile device technology at their convenience, which is a sort of distant education (Matzavela & Alepis, 2021). While eLearning in the context of mobile technology is one of the simplest and quickest methods to obtain educational content is through mobile devices. In addition to the World Wide Web, mobile devices have a built-in application store where users can download original apps (Al-Rahmi ET AL., 2022). These apps were all created with a specific objective in mind. Apps for gaming with an amusing focus come to mind. However, the options are limitless, and many games are created with an instructional focus (Laurens Arredondo & Valdés Riquelme, 2022). The usage of mobile games in the classroom has been shown to improve civic engagement, attention retention, active teamwork during gameplay, and problem-solving and cooperation skills (Laranjeiro, 2021). Of course, in addition to gaming apps, there are a wide range of other applications and technology used in education (Qashou, 2021).

2.2.1 Mobile learning

The growth of mobile devices has resulted in the appearance of numerous new applications that provide rapid user communication as well as the birth of numerous other applications in academic learnings (Kumar & Moral, 2019). The development of wireless networks, the introduction of new mobile e-learning devices, and the appearance of fourth- and fifth-generation networks. Mobile learning is a type of distant learning that enables flexibility and engagement in the teaching and learning practices wherever and whenever they occur (Sevinç & Yavuz, 2021). Small and convenient wireless gadgets like mobile phones, smartphones, tablets, and PADs are utilized in mobile learning to achieve this.

Internet-based applications for managing mobile learning can be used at any time and from any location. The operating systems of smart and mobile devices are compatible with these instructional packages (Fuad & Al-Yahya, 2021). Mobile devices are being utilized for more than just messaging and playing educational games or apps because new ones have been developed that may be used in the classroom, the library, the schoolyard, and training halls. Using class applications, teachers can check attendance and absences, record events, take notes, and complete other activities to better manage their courses (Shah, 2021). Sharing screens between the instructor's device and the students' devices during class apps entails the teacher having control over the students' devices while they are in the classroom. File transfers from one device to another are made possible through this. The most popular of these applications is the Nearpod app, which works with both Apple and Android devices (Wang & Chia, 2022).

Some educators see mobile learning as an extension of e-learning, while others see it as a separate educational system with its own needs and obligations. Because smart gadgets are more affordable than computers, mobile learning can accomplish what e-learning has not been able to (Nafasov et al., 2022). Due to the convenience of portable devices, mobile learning can occur anytime, anyplace. Mobile learning enables use of programs, browsing of mail, text, audio, and video communications, as well as instant messaging with notification of message receipt to the sender (Babalola & Omolafe, 2021).

2.2.2 Advantage, usage, and challenges of mobile learning

The key advantage of mobile learning is that it may take place anywhere and whenever. With mobile learning, a learner can immediately connect to the information network (Ugur-Erdogmus & Çakır, 2022). The simplicity of messaging between students is a defining feature of mobile learning. This technology is affordable, readily available, and reasonably inexpensive (LeBeau, Huey, & Hart, 2019). It is simple to transport this technology because of its compact size and capability for high-speed access, as well as assist them in developing a fresh framework for schooling (Emedoli et al., 2022).

The advantages of mobile learning also enable students to communicate with one another rather than remain secluded behind large monitors. Smart mobile devices that are portable and easy to use in the classroom can replace bulky computers that take up more space (Mihail-Văduva, 2022). Numerous books and educational audio and video clips can be downloaded onto mobile devices, which are lighter and smaller. The potential for handwriting with electronic writing instruments on portable devices and with capacity to distribute and exchange files using SHAREit and Bluetooth technology (Tachinamutu et al., 2022). These gadgets are portable and available at all times. Engage students' enthusiasm in learning by allowing them to use gaming and mobile technology (al-Qallawi & Raghavan, 2022). These tools can benefit children who struggle with learning. Low-cost Internet bundles made it possible for mobile devices to connect to the network practically continuously (Huilcapi-Collantes et al., 2020). They may be able to find information more quickly and work together more successfully online as a result. Smart phones have evolved into a need of life because they not only serve as a means of communication but also come with social networking features, games, and e-books (Sharma et al., 2018). The availability of these gadgets, as well as the leisure and instructional applications they offer, must be utilized by those in charge of the educational process (Wisniewski & Torous, 2020).

Each step in the process of utilizing and implementing mobile learning applications contains fundamental skills that both students and teachers must do and master in order to turn on a successful mobile learning system that meets the required objectives. The following are some of the fundamental abilities needed for mobile learning: the process of creating the educational material's text, graphics, audio, and video content and making content ready for the authoring system with the aid of multimedia programs (Pambayun et al., 2019). Utilize the authoring system to create and publish content for various mobile platforms. making lesson plans and preparations on a mobile device. tethering several instructional tools to the tablet computer. using a tablet computer to prepare for electronic tests (Sulisworo, Yunita, & Komalasari, 2017).

Mobile learning faces a number of difficulties, including mobile devices, notably smartphones and personal digital assistants, have small screens, which limits the quantity of information that can be presented. Particularly in mobile phones and other portable digital devices, storage space is limited. Because batteries operate quickly, they need to be constantly charged (Al Zieni, 2019). Compared to desktop computers, it is more likely to be stolen or misplaced. When employing mobile devices to hack wireless networks, there are security risks that the user may encounter (Khery et al., 2019). For these gadgets to be used properly and efficiently, teachers and students must receive training. Wireless networks and contemporary gadgets must be established in order to implement the mobile learning concept (Onaolapo & Oyewole, 2018).

2.3 Comparisons of Mobile Learning and e-Learning

Digital culture, which is focused on digesting knowledge and assisting the student to be the center of the learning process rather than the teacher, is introduced by e-learning and mobile learning (Chao, 2019). Both e-learning and mobile learning (mLearning) are expensive, especially in the early stages of implementation (Kattayat, Josey, & Asha, 2017). As the e-learning model calls for desktop computers, the creation of educational software, and the construction of electronic curricula that are made available online, the infrastructure must be ready (Almaiah et al., 2022). Teachers and students are being trained on how to work with the contemporary technology through e-learning and electronic curriculum that are not supported by the Internet. Creating a wireless network, getting mobile wireless devices like smartphones, creating electronic curriculum, and training the human component are all necessary for the mobile learning concept (Mergany, Dafalla, & Awooda, 2021).

Given that it relies on self-learning, e-learning and mobile learning encourage student engagement and successful acquisition of the scientific information. Both ways, still and moving photos, video clips, and graphics are used to portray scientific content (Sayibu et al., 2021). Students can access the course material using the two models by going online and browsing. Both systems give students the ability to contact their teachers whenever they want and ask questions. The learner's critical and creative thinking skills are developed by the two models, which rely on the problem-solving approach (Kew & Tasir, 2022). Both options provide the admission of an infinite number of students from throughout the globe, as well as electronic and digital teaching resources are simple to update.

Mobile learning relies on the use of wireless technologies, whereas e-learning relies on the use of wired electronic devices like desktop computers and laptops. Personal digital assistants, tiny computers, and smart phones are examples of wireless devices that are used in m-learning (El Gourari et al., 2020). Electronic learning devices are wired to the Internet; thus, one must be in one of the locations where dial-up Internet access is offered. Mobile learning makes it easier to access and browse the Internet at any time and from any location because it connects wirelessly and can be done anywhere without requiring a specific location (Bossman & Agyei, 2022). In contrast to e-learning, which necessitates sitting in front of desktop or laptop computers in certain locations, mobile learning can be assisted at anytime and anywhere. aids in the file and e-book interchange amongst students using the mobile learning concept (Almaiah, Al-Khasawneh, & Althunibat, 2020). Mobile learning uses wireless technologies, which have lower storage capacities than e-learning does, which uses wired technology.

2.4 e-Learning Technology

E-learning is described as a method for delivering educational training programs to students or trainees whenever and whenever using ICT and interactive tools, including synchronous or asynchronous (Chau, Law, & Tang, 2021). E-learning can also be viewed as an educational approach that relies on imparting knowledge and teaching learners how to use information and communication technology (Ilgaz, 2021). Through the use of multiple media, e-learning enables the student to actively engage with the course material, the instructor, and fellow students at the same time or asynchronously. Using the electronic technologies intended for managing all educational scientific activities and their requirements online (Zaidi et al., 2021).

Although there is a lot of overlap between mobile learning and e-learning, there are also numerous differences between the two that many people are unaware of.

Understanding these differences may assist one to better grasp the methods each educational system uses and how its fields are defined (Marunevich et al., 2021). If the term "mobile learning" is defined as "the use of smart devices to facilitate learning through the applications they offer," then According to the definition of e-learning, it is "a way of learning employing modern communication systems such as computers, networks, and multimedia, in a synchronous or asynchronous manner, or both" (Munoto, Sumbawati, & Sari, 2021).

2.4.1 Concepts and Advantages of e-learning

The term "direct e-learning" refers to a technique and set of educational practices that use the Internet to connect students and teachers and exchange information about classes and research topics (Asghar, BarberÃ, & Younas, 2021). The growth of the e-learning offers the chance for engagement and the adoption of direct e-learning over the Internet in order to mimic the efficacy of real-world teaching techniques.

Computer-based e-learning is a technique that is comparable to conventional basic education and is seen as an addition to existing educational practices rather than a departure from them. a collection of strategies and tactics (Alsharida, Hammood, & Al-Emran, 2021). If it is difficult to broadcast an educational film over the Internet, for instance, it is OK to provide it on computer-based e-learning as long as this raises the caliber and degree of instruction and training.

E-learning and information technology are tools for spreading knowledge and fulfilling the known goals of education rather than being a goal or an end in themselves (Alyoussef, 2021). This includes preparing the learner to fulfill all of the demands of actual life. It became reliant on information technology and its dynamic character in one way or another. Virtual courses, educational seminars, video conferences, e-learning, instructional websites on the internet, and student self-evaluation are some of the ways that e-learning differs from traditional education (Daraghmi, 2022). Along with management, follow-up, and results preparation, interactive relationships between schools, students, and teachers, and conflating education with entertainment in online learning are further issues.

However, it may be claimed that the following are the most significant advantages, rationalizations, and benefits of e-learning: Improving the likelihood of interaction between students and the learning management. using online venues like discussion boards and chat rooms to express the various points of view of students (Rajeh et al., 2021). Since the communication tools give every student the chance to share their opinions whenever they choose and without feeling self-conscious, there is a sense of equality. Access to the teacher should be simple as soon as feasible, outside of regular school hours. The ability to adapt the teaching style: Students can obtain the technical material in a manner that suits them because e-learning and its resources enable them to apply the resources in a variety of ways in accordance with the style of instruction that works best for them (Alyoussef, 2021). This helps to account for the individual differences between students.

Appropriateness of various teaching techniques composing and putting together the lecture or lesson in his or her own unique fashion, e-learning enables the learner to concentrate on key concepts. Students can learn whenever they choose because curricula are available every day of the week (24 hours a day, 7 days a week). Utilize time to its fullest potential. easing the teacher's administrative responsibility (Qiu et al., 2022). With the use of electronic tools, e-learning offers the ability to send and receive with the knowledge of the student's acknowledgement of these files. Lessen the activities on the school by using the resources provided by e-learning, which can send student files and records to the college database as well as evaluate grades, test results, and statistics on them (Kew & Tasir, 2022). E-learning promotes group learning, teamwork, and improved inter-learner communication. offers education to those whose line of work or unique situations prevent them from enrolling in a teacher's course.

2.5. Types of e-learning

Mobile learning: Compact and portable wireless technologies, such as mobile phones, are being used to guarantee that students have access to educational content from any location at any time. There will be no set time or location where learners must travel in order to learn. Because they can learn whenever and wherever they desire, learners will feel more empowered with mobile learning (Almaiah et al., 2022). Additionally, the information that is provided to pupils is not necessary for them to learn it. For both official and casual learning, users can access additional, individualized learning resources via the Internet or the host ith the aid of wireless mobile technology. Mobile technology allows employees to access training resources and information whenever they need it for just-in-time training while they are on the job (Chao, 2019). Since knowledge is accessed and applied immediately rather than being learned and then applied later, just-in-time learning promotes high level learning. Educators and trainers are granted more authority since they can convey with learners through mobile technology at any time and from any place. In addition, educators and trainers have access to learning resources 24/7/365 for class planning and delivery (Mergany et al., 2021).

Synchronous Learning: Using learning tools like virtual classrooms, instant chat, or text chatting, the teaching and learning styles bring the two parties together simultaneously. When a group of students are all learning at the same time, the term "synchronous learning" is used (Chau, Law, & Tang, 2021). Prior to the advent of synchronous learning environments, the majority of online education was conducted through asynchronous learning techniques. When teachers and students come together at the same time and location (physical or virtual), they are said to be learning synchronously and interacting in "real-time." As the lesson is happening, students can ask questions and receive responses in real-time. The instructor can monitor the pupils' comprehension in the moment and modify the lesson as necessary (Weber & Ahn, 2021). There is a greater sense of the teacher "being there" for the students. Asynchronous methods execute all the way through before returning to the caller. Asynchronous methods begin a task in the background and immediately return to the caller. Synchronous Techniques. Typical synchronous methods finish their execution and immediately return their results to the caller. The establishment of learning communities among online learners may be facilitated through synchronous e-learning, which is often enhanced with chat and videoconferencing (Khan et al., 2022). Synchronous online learning is more social and less irritating for both students and educators because questions are asked, and responses are given in real time.

Asynchronous Learning: Asynchronous learning is a word that refers to educational processes that do not take place simultaneously or at the same location (Rehman & Fatima, 2021). Emails, blogs, social networks, digital instructional information, and private Encyclopaedia are a few examples of asynchronous learning methods. When learning occurs asynchronously, students can access resources at their own leisure and engage in longer-duration social interactions (Basri, Husain, & Modayama, 2021). It makes use of tools that allow a network of people to share knowledge without being constrained by space or time. Students can primarily complete their schoolwork on their schedules via asynchronous online learning. At their convenience, students can access lecture recordings and required readings (Amilyana & Noer, 2021). These online courses frequently rely on email correspondence, discussion boards, and audio and video recordings. Virtual libraries and lecture notes are two more typical asynchronous teaching aids. Under asynchronous online learning, students can finish their schoolwork while still maintaining their other professional and personal commitments (Muzaini et al., 2021). Although, there are still deadlines for schoolwork, students can look over their weekly lectures and assignments at their own pace.

Distance Education*:* commonly referred to as distance learning, is a method of studying without physically being present in a classroom. It is one of the teaching techniques that relies heavily on communication to get over the issue of the teacher and student being separated by great distances. Distance education is described as instruction provided by a teacher to students who are geographically separated from one other and who communicate with each other through one or more technical means. Flexibility for students in distance education has received great marks. In addition, remote access to distance learning courses enables learners globally to complete their coursework while at home or on the go. The ability to set up their own schedules gives students the freedom to remain autonomous. Students who are school now and want to pursue higher education along with their careers are drawn to this type of schooling or distance education.

2.6 Difference between distance learning, smart learning and e-learning

The features and modes of usage of mobile learning and e-learning systems differ greatly from one another. We go over them in the contrast between e-learning and mobile learning by highlighting the parallels and discrepancies between the two. Location is the primary difference between distance learning and e-learning. With e-learning, students can complete their online classes and exams while seated in a group in a classroom (Al Zieni, 2019). Electronic learning is another name for e-learning. E-learning programs are intended for those who have reliable internet access at their place of residence or employment. Alternatively, distance learning can be completed online and at home by the student. Students may access information quickly and easily with the use of mobile-friendly teaching strategies, and they can begin studying whenever it is convenient for them (Asghar, BarberÃ, & Younas, 2021). As a result, smartphone users finish their coursework 40% quicker than those who use a desktop computer. Based on administrative perspective, mobile learning saves time (Alowayr & Al-Azawei, 2021).

Moreover, mobile learning courses should not have intricate details or complicated graphics, in contrast to e-learning courses that might use more elaborate designs and modules. Additionally, mobile learning should have simple interfaces and navigation. Instructors distribute course materials using desktop and laptop computers in e-learning. This necessitates the optimization of course materials for huge screens. Because students can view more information on a single screen with larger monitors, productivity increases (Huang et al., 2021). Larger screens typically have higher resolutions, enabling live HD video streaming of classes. A computer's larger screen allows for the use of whiteboards and screen sharing by lecturers.

Since mobile devices are used to distribute information, mobile learning modules must be designed for small screens. High-resolution data, intricate graphics, or media that consume voluminous data or bandwidth are not permitted in mobile learning modules. The finest information design features include one idea per screen, big buttons, and straightforward navigation (Babalola & Omolafe, 2021). The developer has more time to apply and is able to use more of it for a bigger image while creating an online learning course. The reader's eyes must always be considered while designing for mobile learning. In a mobile experience, pupils will have to glance to read text and see small-sized images (Sánchez-Otálvaro et al., 2022). Any courses that fall into the electronic and mobile categories at the same time can automatically be solved with receptive website themes and software.

E-learning and mobile learning both have their own predetermined time limits for how long users should spend on each module within a course. Classes for online learning should last 20 to 30 minutes. Mobile classes are substantially shorter, lasting no more than 3–10 minutes (Marunevich et al., 2021). Although it may appear that these learning systems serve the same objective, this is not actually the case. E-learning is designed to help people gain specialized information and abilities. Mobile learning aims to assist ongoing learning processes while on the go or provide access to knowledge when it is needed. Mobile learning may undoubtedly be utilized to improve e-learning, notwithstanding the systems' differences. However, studies suggest that some elements might either hinder or increase a student's propensity to utilize mobile educational technology. The ability of students in this area was highly influenced by attitude, subjective norm, and perceived behavioral control (Sevinç & Yavuz, 2021).

2.7 Distance Learning

Distance learning is a type of education where students get teaching through online classrooms, or any other audio-visual means (Mastan et al., 2022). It makes it possible for individuals to get education without needing to be physically present in a classroom. Others also defined distance learning as any educational or learning process where there is no student-to-student interaction, and the teachers and students are physically separated (Blinova, 2022). The physical separation of teachers and students during instruction, often known as distant education or e-learning, as well as the use of a range of technologies to improve student-teacher and student-student communication, are the fundamental elements of this type of learning (De & Goswami, 2018). One of the benefits of remote learning is that it actually makes it easier to obtain course materials, provides greater convenience for both teachers and students, and allows for scheduling flexibility.

Distance learning is becoming more and more popular throughout the world and is expected to replace traditional classroom instruction as the preferred method of instruction in the future. It offers those with diverse demographic and cultural backgrounds who are unable to continue their formal education a way to do so. Because of the development of technology, distance learning is now capable of supporting the concept of lifetime learning (Shams, Ajmal, & Jumani, 2021).

The COVID-19 pandemic is among the main causes of the need for distance learning (Almaiah, Al-Khasawneh, & Althunibat, 2020). Geographical distance, population growth, growing social demand for education, knowledge and technology advancements, and economic affordability are all cited as some of the most significant factors supporting distant learning (De & Goswami, 2018)**.** Numerous circumstances force us to recognize the value and necessity of distance learning. Distance learning is crucial for keeping up with the scientific and technological development and for improving the educational system's tool by looking for contemporary patterns or systems (Safitri et al., 2021). Learning is appropriate for the time in which we live, which is marked by an explosion in population and knowledge, as well as a rise in interest in learning, the need for teachers, and the demand for it. New educational patterns and forms, like open education, e-learning, virtual universities, and others, emerged as a result of this (Bossman & Agyei, 2022). Through the advancement of teaching and learning techniques, it helps to accelerate the development cycle. It is also regarded as one of the educational models dedicated to assisting the learners in acquiring the knowledge, science, training, and experiences they need to be an active individual and a good citizen who contribute to the growth and development of their society with all the information and understanding they possess (El Gourari et al., 2020).

Its significance is demonstrated by the introduction of distance learning programs for the majority of students and by offering educational possibilities to anybody who so chooses, regardless of age, gender, or living circumstances. Students can get several degrees, assist in community education efforts, and prepare skilled laborers specialized in various industries (Safitri et al., 2022).

2.7.1 Advantages of Distance Learning in Education

Distance learning in education enhance the student's academic performance. Individual education is based on closely monitoring each student one-on-one and developing a lesson plan to monitor his progress across the board in academic disciplines (Avcı & Ergün, 2022). It strengthens the subjects in which the students are poor or sluggish to learn compared to their peers, which has been seen to have a favorable effect on students by schools who use distance learning education (Wang & Chia, 2022). The students’ motivation to put in more effort and achieve higher results is increased by building on strengths and improving flaws. bridging the academic gap between students and teachers.

When studying at the same educational level, there will always be individual variances and learning gaps amongst learners. This is a result of their mental abilities and skills, which vary from person to person. This is where the role of individual education comes into play. Teachers put forth a lot of effort to assist students who lag behind their peers in closing these gaps and attaining the objectives of the study materials (Jordan & Duckett, 2018). The regularity observed in conventional classrooms, which is very different from distance learning for individual schooling. Because individual education in distance learning enables teachers to get to know their students better, analyse their personalities and abilities, and create plans specifically for each student, as well as present information in various ways to accommodate individual differences and achieve the desired results (Shams, Ajmal, & Jumani, 2021). Distance learning helps to shape teachers' personalities and give them a distinctive professional experience by improving their communication between the teachers and the learners. The bond between the student and teacher is strengthened and their interaction is increased via distance learning for individual learning (Ferri, Grifoni, & Guzzo, 2020). The teacher can allot more time for a student who needs it rather than for another who only needs reinforcement and stimulation, which strengthens the bond between them and improves the effectiveness of the outcomes. The teacher can also allocate more time for a student who needs it than for another who only requires stimulation.

Given that the use of different apps and connection to mobile phone gadgets characterize the present era (such as performance scale application), it appears that distant education will be the predominant pattern of supported education in the future (Hartley & Andújar, 2022). As a result, incorporating technology into the classroom has gained popularity on a global scale, and engaging with instructional content on mobile devices has replaced conventional study methods as a driving force behind learning.

2.8 Performance Scale Application

A scale that is used to determine the scale of success in any given degree of performance is referred to as a "performance scale" (Terhorst et al., 2020). Specifically designed for mobile devices like Android phones and tablets, performance scale applications are a type of software (Meddar et al., 2022). Users frequently have access to services similar to those accessed on PCs through mobile applications (Nafasov, Akhtamova, & Sadullaeva, 2022).

Due to the quick growth in their use, these portable computers have received excellent press, and they are now a popular trend among students. With the aid of mobile technology that may be visualized, students have the chance to fully express their creativity in an organized setting (Sevinç & Yavuz, 2021). The information that is directly connected to their education can be used, uploaded, and shared with others by the students. Students' reactions to learning outcomes may be improved by the performance scale application.

The PSA is a component of an e-learning program intended to transform the global education system, including secondary level (Sayibu et al., 2021). The successful integration of e-learning platforms and distant learning in secondary school depends on the participation of all stakeholders, particularly students and teachers (El Gourari et al., 2020). Both students and teachers must be open to use e-learning resources in the classroom with the implementation of performance scales in secondary school. Teachers can virtually reach a broader audience of pupils with e-learning systems and distant learning platforms, such as mobile learning platforms (Almaiah, Al-Khasawneh, & Althunibat, 2020).

Schools can offer students learning resources in a range of methods as part of their learning process, including applications, games, movies, slides, audio, chat rooms, and computer testing platforms (Kew & Tasir, 2022). Contrarily, students can search, access, and exchange instructional materials to further their understanding of the performance scale application (Rasheed et al., 2022). Despite the facts, performance scale apps as an e-learning system are not now widely used, especially in developing countries. More study is still needed in these countries (such as Jordan) to promote their usage in high schools.

The conventional educational model, which emphasizes memorization, in-person instruction, placing the instructor at the center of the learning process, and using books as the main source of knowledge, no more represents the most successful model; rather, a new paradigm for distance learning tools, like performance scale applications or mobile applications, has emerged as a result of the advancement of technology (El Gourari et al., 2020). Mobile learning relies on the usage of radio receiver technologies and functions to effectively teach students outside of the classroom and via remote learning (Amjad et al., 2022). With the use of performance scales, students may connect their learning to real-world situations, make wise decisions for themselves, learn for themselves outside of the classroom, and engage with their instructors and school efficiently.

2.8.1 Personalized Learning

Learning that is specifically geared toward a student's skills and weaknesses is known as personalized learning. Personalizing education to each learner's unique needs, interests, and abilities is the aim of customized learning. According to this idea, one way to deliver teaching is through personalized learning. Based on their learning approach, background knowledge, abilities, and interests, each student receives a customized "learning strategy" (Lin et al., 2020). It goes against the "one-way fits all" methodology that is prevalent in most schools. Pupils learn in a variety of ways and at varying rates (Xie et al., 2019).

In collaboration with their teachers, students create both immediate and long-term goals. Students that use this technique become more responsible for their education. Teachers ensure that project-based learning or lesson plans adhere to academic requirements (Sereno, 2018). Additionally, they observe students to see if they are developing the abilities that are required of them as they advance in their schooling. Special education is still necessary, even with personalized learning. It is a general education strategy that can be used in conjunction with specialist intervention programs and individual education plans (Bourekkache et al., 2020). But for individualized learning to be effective, adjustments, supports, and accessible learning methodologies must be included. If done correctly, learning will be more engaging for all pupils. Additionally, pupils who are having difficulty will receive assistance sooner. Students with disabilities risk falling further behind if things are not done properly (Troussas, Krouska, & Sgouropoulou, 2020).

Personalized learning is a method that not only enables but also supports students to study in a way that best fits their individual learning styles and aptitudes for assimilation and application of the knowledge they acquire in class. Individual students' desires and interests are taken into consideration when teaching lessons, and extra help is given as required using the same method (Iyer et al., 2022). According to Thai, Bang, & Li (2022), with the aid of personalized instruction, students can advance at a pace better suited to their requirements. Personalizing information and delivery methods for each student based on their level of cognitive development in any subject is the foundation of personalized learning, a philosophical approach to teaching and learning. After finishing a task, students' success on the task decides whether they go on to more difficult projects or receive further training to hone the abilities they will need to accomplish more long-term goals. Their teacher continuously assesses their progress and takes action to close any gaps that may arise. This will increase each student's level of confidence and motivate him to work harder on his skills and close any educational gaps in order to achieve better results.

With individualized learning, no two schools will be precisely alike. The following are four prevalent models that schools utilize, though. Each of these models holds all students to high standards and binds their academic success to a rigid set of specifications.

1. Student profiles in schools. This category of school keeps a present record that enables a thorough comprehending of each student's different capabilities, needs, enthusiasms, advancement, and desires. These updates occur much more frequently than they would on a typical report card. Additionally, these thorough updates assist teachers in making choices that will enhance students' learning. Students can monitor their personal development with the use of a student profile. It enables the instructor, the learner, and, in various schools, the parent to determine whether they should alter their teaching strategies or their objectives prior to the student performs poorly or fails (LeGeros et al., 2022).

2. Personalized learning pathway in schools. With the assistance of this category of school, each student can create a learning pathway that responds to or adjusts based on their development, motivations, and desires. For example, depending on weekly updates regarding academic performance and hobbies, a school might plan a student's timetable. Each student has a different schedule. But a variety of teaching techniques are probably included. The combination may involve one-on-one teaching with a teacher, solo study on specific skills or challenging assignments, and task-based learning with a small set of peers. With a tailored learning path, a student can work on various abilities at various speeds. But that does not imply that the school will permit a student to lag significantly behind in any subject or specialty. Each student is attentively observed by teachers, who also offer additional help when required (Chang, Li, & Huang, 2022).

3. Competence-based learning in schools. This category of school regularly evaluates pupils to track their advancement toward predetermined objectives. Students can easily understand what they must grasp in this system. These competences encompass particular abilities, information, and attitudes like fostering resilience. Options for how and when to show their understanding are given to the students. For instance, a student and a teacher might work together to integrate specific mathematical concepts into an internship at a retail establishment. The learner may work on a number of competencies at once. They go on to the next after mastering one. Every student receives the assistance or services required to help them master the skills. The focus is not on taking a test and scoring well or poorly on it. Instead, it emphasizes lifelong learning and possessing (Zheng et al., 2022a).

4. Schools utilizing adaptable learning contexts. This category of school modifies the education context for each student according to how they best learn. That includes details like the classroom's physical arrangement, the day's schedule, and the distribution of teachers. Schools can, for instance, consider strategies to provide teachers extra time for small group instruction. Redesigning how teachers utilize the time, space, and resources in the classroom is not simple. However, using design idea in this way can help the classroom adapt to the requirements of students (Zheng et al., 2022b).

2.8.2 Self-learning

The development of self-education has its origins in the depths of human history, according to Funa & Talaue (2021), who attribute it to Socrates (399–496 BC) and his (Socratic) method, an instructional dialogic approach in which the teacher guides the student through dialogue and questions. One of the words that has gained popularity is "self-learning," which refers to the use of connected, scientifically sound resources and the relationship between teachers and students.

Self-learning relies on the application of the theoretical knowledge acquired during the educational stage. The ongoing that was made public for the job that was proposed in the initial stage of education (Sayibu et al., 2021). The educational process has undergone evident and noticeable modifications as a result of the growing load that knowledge in numerous scientific domains has placed on both teachers and students. Given the limited amount of time allowed to many disciplines, it has become challenging for the instructor to give students access to all the sciences and experiences. Since the curriculum in schools is no longer adequate, students must continue their education outside of the classroom in order to meet their needs (Güngörmez, 2021).

In order to maintain their independent self-learning, students develop the abilities and learning habits of continual improvement. teaching students problem-solving skills and cultivating a creative atmosphere. establishing an ongoing learning community. Obtaining everlasting, ongoing education reliance on oneself during the learning process by the learner. fostering a sense of responsibility. want to discover new things (Dorji, Tamang, & Tilak, 2021). It is crucial to give the student the skills he will need to teach himself, including the capacity to read. writing ability, information collection, documenting, and search expertise are other skills that should be provided to the student (Diesta & Ferolino, 2021).

2.9 Academic Achievement

Performance is measured by academic success. Academic achievement or academic performance is the level of accomplishment that a student, instructor, or institution has attained or their predetermined educational objectives within a predetermined timeframe (Bayar & Kurt, 2021). Academic success is primarily influenced by classroom performance, which demonstrates how effectively pupils have accomplished the learning objectives (Yan, 2020).

Studies have concentrated on creating a tool for predicting student performance, self-evaluation, self-learning, classroom achievement, failure rates, and lowering exam cheating proportions inside e-learning practices (Bernacki, Greene, & Crompton, 2020; Wahyu et al., 2021). Others have e-learning tools, like programs that anticipate student success and assess students, like PSA (Mergany, Dafalla, & Awooda, 2021; Ugur-Erdogmus & akr, 2022). Student information systems served as the primary source of the majority of the data used in these investigations. Some studies, in contrast to past research, concentrated on applying machine learning methods to evaluate performance and aid in student improvement (Sandra, Lumbangaol, & Matsuo, 2021; Hossain, Rahman, & Uddin, 2022).

The practice of anticipating academic success utilizing scale measures to improve student achievement has been shown in prior study to be one of the most important approaches for obtaining academic success (Halde, 2016; Ofori, Maina, & Gitonga, 2020; Sokkhey & Okazaki, 2020). In order to improve student performance in schools, new mobile software need be developed that will guide high school students toward the specialization that is best for their academic achievement, scientific abilities.

2.10 Choice of Specialization

The majority of students desire to choose a specialization they genuinely love and find it fun, but they often lack the background knowledge necessary to make an informed decision in a variety of fields (Alonso-Virgós et al., 2021). It's difficult for many students to accept this one. If a student's first option is in a difficult specialization or field, it is advantageous to build a substitute specialty that they would like to practice as well as gaining expertise (Ogowewo, 2010; Sattin-Bajaj et al., 2021). Application of performance scales may increase students' chances of choosing the appropriate speciality (Ogowewo, 2010; Stoyanov et al., 2016). Test results, enthusiasm in the topic, and sound suggestions are all essential considerations when choosing a candidate. Improved potential, curiosity, easy access and a well-known location, a reward, future chances, and appreciation are the main drivers of secondary school students' decision to specialize (Ogowewo, 2010; Chao, 2019). Schools and regulators take these issues into account. There is a dearth of research on secondary school pupils' specialization preferences. Therefore, additional attention is required to close this research gap.

2.11 Data Mining Techniques

One of the often-employed techniques for information analysis and mining is an intelligent educational system. The eLearning data mining process is now simpler, more dependable, and more precise thanks to the introduction of artificial intelligence and machine learning algorithms (Baashar et al., 2022). Automated learning algorithms have been developed in numerous research to evaluate pupils and predict their academic success, and these systems have demonstrated great accuracy (El Gourari et al., 2020; Ahsan et al., 2021; Hossain, Rahman, & Uddin, 2022).

An algorithm for calculating the grade point average, or "GPA," has been developed through research (Sekeroglu, Dimililer, & Tuncal, 2019). This process involved looking at a range of factors, such as learning tendencies in academic achievement, the effects of group achievement on individuals, and instructional methods for students. Clustering analysis was used to validate the technique once the dataset was gathered from school learners. Despite the success of this technique, the data in this instance can only be collected from the pupils' point of view.

Instead of utilizing the clustering method to evaluate students' performance, a study reveals pupils who were unable to keep up with the programming classes (Shams, Ajmal, & Jumani, 2021). The researcher collected the data using the student profiles and found that the algorithm generated correct findings when K means were employed. Different research also created an algorithm to identify pupils with overdue assignments and boost their academic achievement. The students were enrolled in a mixed-model course, and the researcher used an online course from Internet Scholar to collect the data (Lim et al., 2022).

For use in other study report, the investigator designed a method of visual analysis known as "performance Vis" (Nawar & Mouazen, 2017). The study evaluates information on students' achievement over time using three crucial variables: the composition of the exam questions, student characters, and assignment completion. When a student is registered in a course, the method manages them. This technique's development incorporated the information system the students used to store their data. However, the results demonstrated how well the performance Vis predicted students' course performance. The performance Vis should have been created to coincide with the course when it was still being taught, if possible. Therefore, the advantages will be larger if the pupils can quickly pinpoint their regions of vulnerability. It might have helped teachers and school management spot pupils who are likely to drop out of school soon.

2.12 Previous Studies

The goal of many studies is to enhance the learning process by developing instructional strategies for learners. In one of these investigations, the researchers develop an algorithm that collects the grade point average, or "GPA." This process is carried out by looking at a range of factors, such as learning developments in academic achievement, the effects of groups' levels of academic achievement, and instructional methods for students. The data was gathered from Thai universities, specifically the University of Phayao, and clustering analysis was used to validate the algorithm. Despite the fact that this algorithm showed its achievement, the research in this case was limited to the data collected from the students' perspective, with all other relevant sides being disregarded (Nuankaew et al., 2019).

To help students build their educational status so that it could be quickly evaluated by their teachers and lecturers, the procedure of earlier academic level and performance prediction was carried out. However, teachers can make some decisions during this process, particularly regarding the best teaching method. As a result, the researcher created a sample that uses genetic algorithms and is referred to as "Fuzzy Logic Model" (Yldz et al., 2012). Along with tracking the number of hours a student logs into the system and the amount of time they spend online, data is also gathered here using the distant education system. It should also be noted that the investigator has been looking at the data for six weeks. At the end of the semester, a comparison to the actual findings was made, and the model's accuracy was estimated to be roughly 84.52 percent. The most accurate model is this one, however it lacks a wide range of data sources. In other words, the outcomes would be extremely accurate if the data resources utilized in this technique were greater.

The study paper "Self-Regulated Learning," which sought to assess and judge the pupils' abilities, also surfaced. Depending on the online courses that employ "Educational Process Mining" approaches (Cerezo et al., 2020). For Spanish college students, a course log file was examined online. The investigation included a pre-treatment to carry out the "mining" process, which was then followed by a discovery. Additionally, the "Inductive Miner" algorithm was employed. The data that was analyzed for this study only included third-year students, despite the fact that substantial results were obtained. As a result, compared to the freshmen's scores, which would likely be substantially different, this presented a hurdle.

In order to be used in another study report, the investigator developed a visual analysis technique dubbed "performance Vis." It is utilized to analyse data on students' success over time by taking into account three essential criteria, including student traits, homework success, and exam question scheme (Deng et al., 2019). This approach might aid students in previewing their prior performances as well as in pinpointing the precise areas of challenge in this course. Recognizing students who might fail in a short amount of time can also help teachers and administrators. This technique's development incorporated the information system the students used to store their data. Unluckily, it would have been better if it had been developed to go along with the course while it was still being taught. Thus, if students can quickly pinpoint their areas of weakness, the benefits will be larger. So, even if they are taking the course, they can manage them.

The study's methodology is based on identifying and forecasting the students who are most likely to fail or perform below expectations by enrolling in various mixed practical courses. In addition to gathering information from the "learning management system" and the outcomes of online quizzes or midterm exams, Kumamoto University in Japan uses the neural network algorithm. These two algorithms apparently didn't use enough data. The idea is that the more data that is used, the more precise the outcomes will be. Contrary to earlier studies, the data in this one was acquired from professors who disregarded the viewpoint of the students (Sukhbaatar et al., 2019).

In a different study, the investigator attempts to identify the students who cannot carry on with the programming courses rather than evaluating their performance through the clustering technique. The investigator gathered the information from the SIS or the profiles of the pupils. In this instance, the investigator proposed a strategy to anticipate learners leaving e-learning methods when it became clear that the algorithm's usage of K means doesn't produce thorough results and that the lack of data entry restricts the accuracy of the results (Oeda & Hashimoto, 2017).

Numerous research studies also make behavior predictions for students. This is being done in an effort to boost the poor students' performance before it's too late by implementing the right solutions and powerful strategies. In this situation, the researcher created a "Neural Fuzzy Inference System" to predict the students' success based on their past accomplishments. For example, a set of data was used in a public institution in India, and the outcomes were excellent. However, the researcher's analysis indicates that more data would produce more thorough and precise results (Moitra & Mandal, 2019).

An investigator has also developed a system to enhance pupils' academic achievement by reviewing their late homework. Using a scholar course on the internet, the investigator gathered the information from grouping the learners in a particular mixed-learning class. However, this method encountered numerous issues, particularly when the number of students in the classes increased. In other words, when dealing with enormous data, the algorithm is ineffective (Akram et al., 2019).

Additionally, in this review study, the researcher looked at previous work that has been done specifically over the past five years and is closely related to the issue of predicting academic performance. This study's goal was to enhance a prediction model that diligently seeks to enhance students' success-oriented steps by employing the "Educational Data Mining" technique. Only college students with bachelor’s degrees are eligible for this study. The SIS, which houses student information systems, served as the source of all of their data (Alyahyan & Düştegör, 2019).

The researcher in this case went to look for the information that directs students (Vialardi et al., 2011). To do this, a single algorithm was created. It was a "RISP DM" methodology that was used with SIS-derived data. The goal of this algorithm is to assist college students as they register for their courses and textbooks. It aids students in selecting the right course for their skill level. After testing this method, it became clear that it produced flawless results. However, in such a short amount of time, grades cannot be used to determine a student's success or failure.

Another significant study developed an algorithm that, when used in an online learning environment, highlights the students' expected academic achievement (Aydoğdu, 2020). This algorithm depends on the analysis of numerous factors, including the number of participants and the amount of time spent by students browsing. Although the results were excellent for such an algorithm, it would be better to include more variables in the analysis to aid in improving the research's accuracy.

This program relies on big data to forecast academic success. Data was extracted from the "SIS, LMS, Decision Tree, and Artificial Neural Network" systems at West Scotland University, which included information on roughly 141 students (Adejo & Connolly, 2018). With the help of an upgraded version of the "Multi-model Heterogeneous Ensemble Approach" model, this method intends to aid in the early identification of students who may be in danger of discontinuing their higher education. Like previous research, this one might have been improved if it had had a large number of variables to be examined in order to obtain more accurate and trustworthy results.

In addition to improving student performance and evaluating it, the research that examined learning management systems also worked to develop teachers (Ababneh et al., 2021). Both the teachers' performance and their capabilities need to be improved. Data for this study was gathered using MySQL. This type of study was conducted to determine the effects of each strategy on the performance of the third-year students at Bahçeşehir University. The primary criterion used to evaluate this study was the opinions of the students. Additionally, the R Tools were utilized to manage this data because this study is vital for assessing and improving the effectiveness of teachers. It does, however, have a lot of precise quantification. For instance, professors might not interact in the online classes, which is not always a weakness but could be due to a lack of materials that are tailored to them. This could diminish the research's findings because the variety and variations in perspectives could produce a lot of insightful findings.

2.13 Summary

In conclusion, machine learning and artificial intelligence techniques are widely used in the field of education. Numerous studies have been done to see if these techniques may help the educational system and improve academic accomplishment levels by creating automated learning methods to assess students and forecast their academic achievement. High academic success rates have been attained in many of these trials. E-learning systems can be used to forecast student performance and assessment. A model for forecasting student dropout rates in e-learning systems has been the focus of other investigations. Student information systems were the main source of data in all previous research. As opposed to past studies, several studies concentrated on utilizing ML algorithms to evaluate teachers' performance and assist them in improving it. The systematic analysis of the studies revealed a dearth of studies focusing on secondary school students. No studies were found that attempted to assist secondary school students in selecting their learning direction at this point, even though only two research were discovered that assisted university students. There is a huge gap in the improvement of the academic achievement process since this reason is among the most important ones in doing literary research on the subject. The use of AI systems, particularly ML, is another significant area that is essential to the education subject and is recognized as the most recent technology.

Way forward based on a review of literature. Previous research has demonstrated that strategies for identifying corrective steps to improve student performance are among the most crucial means to realize educational achievement. Based on the results of this systematic research, we will propose developing a new approach to guide secondary school students in selecting suitable scientific subspecialties based on their academic performance in all specialties throughout the previous school years. The semester and cumulative averages will not be taken into account in the prior investigations. Moreover, the estimates of the subjects the student took, the final cumulative GPA, the results for each subject taken separately, and the degree to which they are related to one another will be considered to be fundamental components in assessing outcomes, forecasting, and assessing the student's academic achievement. But in this study, we will focus on what's important for evaluating a student's academic achievement and setting goals for their outcomes. This technique is highly helpful in assisting the student in selecting a suitable major, preventing future student failure, and saving both the learners and the teachers' time and effort.

CHAPTER III

MATERIALS AND METHODS

3.1 Introduction

This section discusses the materials and methods used in this study. Along with a detailed description of the data, its methods of collection, and its ethical approval. Additionally, the chapter will discuss every stage of the suggested system in detail along with the algorithms utilized and the method used to create mobile applications.

3.2 Reasons for Selecting Performance Scale Application

Student achievement is very important since it reflects the effectiveness of secondary education. In order to solve the issues regarding students' poor performance and failure in their placed specialties, the performance scale application (PSA) may improve student achievement and early projections for the futuristic outcomes may be helpful. Another reason for choosing the PSA is to serve as a mobile platform that improves teaching and learning by delivering learning content and instructions from teachers and schools to students. In addition, it serves as a guide for students to select specialties at secondary levels. The PSA will help students, teachers, and educational institutions to expect an accurate student's achievement, GPA, and the best specialties for the 12th grade in order to avoid student failure and depression.

Another reason is that academic achievement, participation in learning-focused activities, satisfaction, gaining desired information, skills, and competencies, achieving educational outcomes, and post-school performance are all defined as components of student achievement by the guide of PSA. The student academic record can be preserved, tabulated, and structured through PSA. Using PSA, information about the student is useful for the evaluation and academic growth processes. In addition to being used as a means of documentation for schools using the student's academic record and data from the Jordanian Secondary School Leaving Certificate Board.

Also, PSA can aid in specifying performance benchmarks that stand in for the calibre of simulations that a student is able to locate and use in schools. This is because it provides assessment tools to gather data and information required to evaluate the performance of the learners. It provides a simple tool for completing tasks and assignments because it is quick and convenient. Similarly, because the digitization tsunami has swept into education, the trend in education is shifting. Due to the increasing prevalence of mobile apps, eLearning is now the most recent requirement for students. Its success is a result of its originality and ability to make learning enjoyable for students. Other reasons include modern learning techniques, easy access, timesaving, cost-effective, effective parent-student-teacher communication, and a comprehensive and systematic approach.

3.3 Research Design and Framework

This strategy seeks to predict students' secondary school curriculum before they join the 12th grade in order to prevent pushing students to choose a specialization that is not a good fit for their abilities based on their prior grades. This is based on a fundamental exam for each secondary school specialty. The study methodology consists of two main Phases:

Phase 1: System Development Process (prediction model and the mobile application development).

Phase 2: Experimental Study

3.3.1 Methodology Phase 1: System Development Process

The Phase 1: System Development Process consists of two main stages the first one is a prediction model and the second one is a mobile application development, thus the prediction model consists of three main components

The first portion is a clever model using ML algorithms to forecast the student's GPA in the event that they choose to specialize. Numerous ML methods were examined. This comprises the following techniques: Gradient Boosting, Model Stacking, Support Vector Machine, K-Nearest Neighbor, Linear Regression, Bayesian Ridge Regression, Artificial Neural Network, Extra Trees Regression, and Random Forest. Then, the system (performance scale application) was used to compare each model to determine which was the most accurate and efficient.

The second module is a mobile application (PSA) that is used to gather student grades from the essential courses for the specialization they passed in 11th grade first. Then, it establishes a test difficulty level for the student's selected specialty and expresses the exam's results as a percentage. The application's last output for 11th graders' performance in foundational courses and exam outcomes was an AVG output.

A genetic model, the third element, was employed to forecast student performance and direct them into the most appropriate speciality. The model selects the optimal outcome, which is displayed in the application for the performance scale (Figure 1).

Several other tasks were also included, such as:

Firstly, the typical student performance in the fundamental subjects for the eleventh grade:

Since students still have opportunity to alter their thoughts prior to entering the 12th grade, this step is taken since they have only partially decided on their specialization at this point. They would not be able to switch majors during this grade. Moreover, this plan is essential for improving the method so that the curriculum and learning setting in this class are equivalent to those in the 12th grade, such as the difficulty and course content. Due to this, it is simpler to assess whether or not a student possesses the abilities and degree of output required to successfully achieve this course. In this level, the result is average.

Secondly, establishing a level test for the primary courses in each specialty:

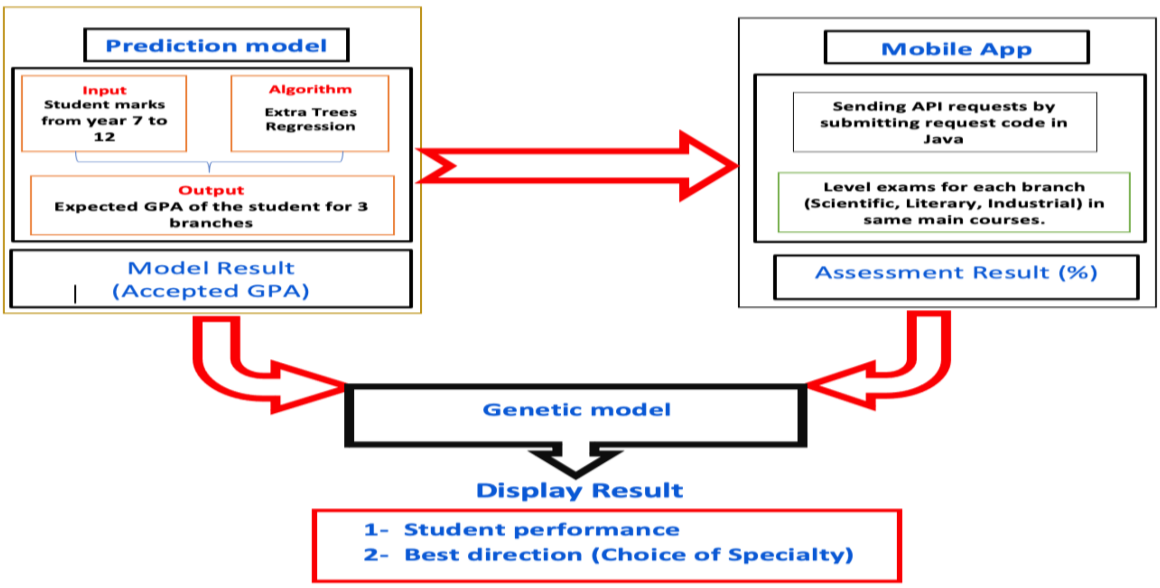
As the results of the exam would reflect the student's performance level, this phase is essential for increasing and improving the precision of the system functioning. The result of this stage is therefore given as a percentage, and the average was also utilized to determine the sum of the results of the first and second steps. In this instance, one of the inputs to the prediction system is the average.

Thirdly, the additional factor in the anticipated GPA is:

The findings from the mobile app, which comprise the expected GPA, average level exam scores, and basic course grades, are fed into the genetic algorithm in this last step. Moreover, the mobile app sends the average of the major course grades and the outcomes of the level exams to the genetic algorithm in ML via the input full API. Due to the fact that the genetic processes utilize this information as inputs, this is accomplished by creating a Java instruction. The ultimate output of this suggested system is given for exhibition on the mobile application when the genetic method's prediction of the student's achievement regarding their best speciality has been completed.

Figure 1

The Prediction Model Framework



3.3.1.1 Prediction Model Development

Data mining is used to draw knowledge from a large amount of data by identifying new and precise types of information that are useful to the researcher.

This study employs machine learning data mining strategies to develop samples that assess data and forecast future outcomes that may be advantageous to both students and teachers. We must first distinguish between regular programming, often known as conventional programming, and machine learning programming. Traditional programming relies on providing the machine with information about a problem by giving solutions through instructions on how the system will handle them. This process is referred to as explicit programming. Machine learning, on the other hand, relies on providing the computer with actual data that includes the imports on the output from which the computer infers the link between them. Then, even if it had never encountered or dealt with it before, it is prepared to receive new data.

3.3.1.1.1 The first component: prediction model using machine learning algorithms (expected GPA)

The tested and used ML algorithms in this study are discussed presented in this section. However, only the key ML algorithms were shown below. The extra trees regression approach was shown to be the most accurate prediction algorithm out of 10 tested algorithms. The extra trees regression has a precision of 98.92% and a 98.92% accuracy when compared to other methods with the three specialties investigated.

(A) ***Supervised learning***

A supervised learning algorithm was employed to spontaneously create an input–output function (a predictor) f (x): M Z to compute estimates of outputs as a function of inputs given a sample $(mi, zi)^N$ of input–output sets (where m i ∈ M and yi ∈ Z)[1]. The algorithm searches in the hypothesis space H, which is a suppositionally big but limited set of input–output functions (a subset of the space Z M). Extra trees regression hypothesis spaces are used.

The collection of all samples with finite sizes is represented by

$$(M \times Z)^* = (M \times Z)^N$$

$N = 1$

$A: (M \times Z)^* \rightarrow H$

from $(M \times Z)^*$ into the hypothesis space H. For a given H_1

sample *ls* ∫ $(M \times Z)^*$ represents *A*(*ls*) with the function replaced by the algorithm *D*.

We designate by *D*(*ls*) the function produced by the algorithm *D* for a particular H_1 sample, *ls*(M Z).

The probabilistic validation of the supervised algorithm takes into account m: $\ell \rightarrow M$ and z: $\ell \rightarrow Z$, which are two random variables specified by a certain probability space (ℓ, λ, Θ). Let *PM,Z* represent their combined probability distribution described by $M \times Z$ and let $\ell : Z \times Z \rightarrow R+$ be a non-negative loss function described by $Z \times Z$, while any $f\ \delta$ H is represented by

$$L(f) = X \qquad M \times Z\ \text{Ʀ}(f(m), z)dP\ M,Z$$

(B) ***Linear predictor***

The extra trees regression was used as a linear predicting function $f(k, i)$ to estimate the probability that observation *i* has output *k*, in the subsequent function:

$$f(k, i) = \beta_{0,k} + \beta_{2\mathrm{k}}\, x_{2,i} + \ldots + \beta_{\mathrm{M,k}}\, x_{M,i}$$

where $\beta_{\mathrm{M,k}}$ is a regression constant linked with the *M*Type equation here.th instructive variable and the *k*th output. The regression constant and instructive variables are typically classified into vectors of dimension *M* + 1, such that a more concise version of the predicting function is represented as:

$$f(k, i) = \beta_{\mathrm{k}} \cdot x_i$$

where β_{k} is the set of regression constants linked with output *k*, and a row vector is the group of instructive variables linked with observation *i*.

Extra Trees Algorithm

Extra trees: Function for training extra trees classifier or regression.

This function performs the extra trees developing algorithm (executed in Java).

Algorithm 1:

```
##
extraTrees(m, z,
        ntree=605,
        mtry = if (!is.null(z) && !is.factor(z))
            max(floor(ncol(x)/3),           1)           else
floor(sqrt(ncol(x))),
        nodesize = if (!is.null(z) && !is.factor(z)) 7 else 1,
        numRandomCuts = 1,
        evenCuts = FALSE,
        numThreads = 1,
        quantile = F,
        weights = NULL,
        subsetSizes = NULL,
        subsetGroups = NULL,
        tasks = NULL,
        probOfTaskCuts = mtry/ncol(x),
        numRandomTaskCuts = 1,
        na.action = "stop",
        ...)
```

(C) ***Set of independent binary extra tree regression***

The extra trees algorithms are obtained running $K - 1$ independent binary extra trees regressions for each of the K potential outputs. In these algorithms, few outputs are selected as the “swing,” and the remaining $K - 1$ outputs are then individually regressed on the swing output. If output K (the final output) is selected as the swing, then the following occurs:

In $\frac{\Pr(Y_i=1)}{\Pr(Y_i=K)}$

In $\frac{\Pr(Y_i=2)}{\Pr(Y_i=K)=\beta_1}$

$\ldots. \qquad . x_i = \beta_2$

In $\frac{\Pr(Y_i=K-1)}{\Pr(Y_i=K)=\beta_{K-1}}$

$. x_i$

The common Softmax transform, which is employed in compositional data analysis, is another name for this equational algorithm. For each potential result, we have established a distinct series of regression coefficients.

When both sides were exponentiated and the probabilities were calculated, we obtained:

$$\Pr(Y_i = K)\, e^{\beta_1 \cdot X_i} \quad = \Pr\,(Y_i 1)$$

$$\Pr\,(Y_i = \ \Pr(Y_i = K)\, e^{\beta_1 \cdot X_i}$$

$$\Pr(Y_i \quad = K \_ 1) = \Pr\,(\,Y_i = K)\, e^{\beta_{k-1} \cdot X_i}$$

Given that all K of the probabilities must add up to 1, we determine:

$$\Pr(Y_i \quad = K) = 1 \quad -\sum_{k=1}^{K-1} \Pr(Y_i \quad = k) = 1 \ - \sum_{k=1}^{K-1} \Pr(Y_i \quad = k)\, e^{\beta_1 \cdot X_i}$$

$$\Longrightarrow \Pr(Y_i \quad = K) = 1 + \frac{1}{\sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

This helped us determine the other probability:

$$\Pr(Y_i \ = 1) = 1 + \frac{e^{\beta_i \cdot X_i}}{\sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

$$\Pr(Y_i = 2) = 1 + \frac{e^{\beta_i \cdot X_i}}{\sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

$$\Pr(Y_i = K \_^1) = 1 + \frac{e^{\beta_i \cdot X_i}}{\sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

where, normally, the sum goes from to *K*:

$$\Pr(Y_i = k) = 1 + \frac{e^{\beta_i \cdot X_i}}{\sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

Where β_k is described as 0. The algorithm depends on the presumption of independence of unrelated alternatives as indicated above, which is why we conducted extra trees regressions.

(D) ***Extracting, transforming, and feature selection***

(i) TF-IDF Feature Extractor

To mine text in the system, the feature vectorization technique known as term frequency-inverse document frequency (TF-IDF) is commonly utilized to indicate the significance of a phrase to a particular document in the corpus. Put a term (t) before a document (d) and the corpus (C) before a document. The frequency of a word is expressed as T F (t, d), where t is the number of times the term occurs in document d, and DF (t, C), where t is the sum of documents that include that term. Basically, the sum of TF and IDF makes up the TF-IDF measured.

$$\mathbf{T\,F\,IDF}\ (t, d, C) = \mathbf{T\,F}\ (t, d) \cdot \mathbf{IDF}\ (t, C).$$

where |C| represents the corpus's overall document count. Because a logarithm is employed, a term's IDF value becomes 0 if it occurs in every document. In order to prevent division by zero for terms outside of the corpus, a smoothing term is used.

(ii) VectorSlicer

A transformer called VectorSlicer was used to accept a feature vector as input and produces a new feature vector containing a sub-collection of the prime features. A vector column's features were extracted using it. The vector column with the supplied index numbers was input into the VectorSlicer, which subsequently produces a new vector column with the values chosen using those index numbers. AttributeGroup implemented matches on an attribute's name region.

(iii) Locality-Sensitive Hashing

A significant class of hashing methods known as locality-sensitive hashing (LSH) is often employed in large datasets for outlier recognition, estimate closest neighbor search, and clustering. The basic idea behind LSH is to hash data points into buckets using a family of functions (referred to as "LSH families"), with the aim of making sure that data points that were close to one another are most probable in the similar bucket, whereas data points that were far apart are most probable in different buckets.

An LSH family is a family of functions L that satisfies the following criteria in the metric space (M, d), where M is a cluster and d is a distance function of M.

$\forall s, t \in M,$

$d(s, t) \leq r1 \Rightarrow P\ u(L(s) = L(t)) \geq s1$

$d(s, t) \geq r2 \Rightarrow P\ u(L(s) = L(t)) \leq s2$

It is known as the (u1, u2, s1, s2)-sensitive LSH family.

Working Steps

1. Reading files and extracting data using imported library (Figure 2).

Figure 2

Import Libraries and Load Datasets

Import Libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Load Datasets

In [2]:
```
# Year 7
year7_df = pd.read_excel('student data new.xlsx', sheet_name=0, index_col='Name ')
new_columns = [col.strip() for col in year7_df.columns]
year7_df.columns = new_columns
year7_df = year7_df.iloc[:, :9]
year7_df.head(5)
```

2. The exploration data analysis (EDA):

   At this stage, the data was checked carefully for errors and to discover the relationship between the features or its relationship with the final target. In addition, the visualization of data was made in order to have a full image of the data handled.

3. The cleaning:

   In this stage, the data errors were introduced before corrected in the previous stage (EDA) in order to guarantee getting clean data, which means better performance with the model (Figure 3).

Figure 3

Data Cleaning

Cleaning

```
In [23]: # Giving 9COMPUTER the minimum = 35
         year8_df.loc['KL', '9COMPUTER'] = 35
         year8_df.loc['ZWLL', '9COMPUTER'] = 35

In [24]: year8_df[year8_df['9COMPUTER'] < 35]

Out[24]:
        9MATH  9PHYSIC  9CHYMESTRY  9BIOLOGY  9COMPUTER  9ENGLISH  9ARABIC  9SOCIAL  9RELIGION
Name
```

4. Data emerging:

   Since each class belongs to the same student, but the used data is dispersed across the years of study, this is a unique situation that requires the data to be combined in order to be handled properly.

5. Feature engineering:

   Since the preceding stage's conceptualization is complete at this point, numerous operations are performed on the key characteristics to produce additional ones that are more valuable to the model, such as the normalization and mean procedures.

6. Preferred specialties:

   Depending on the model used, this is also a unique scenario. There are two data frames, one for each major, because researcher took the student data from the same class. Then, a model was created for each major so that each student's grades may be predicted accurately based on that major.

7. Preprocessing:

   At this stage, the data is split into two categories: 80 percent of the data are used for training, while 20 percent are used for testing. They are now prepared for modeling (figure 4).

Figure 4

Data Preprocessing

Pre-processing

```
In [32]: # spliting data
         from sklearn.model_selection import train_test_split

In [33]: # Scientific
         x = scientific_df.drop('average', axis=1)
         y = scientific_df['average']
         x_train_s, x_test_s, y_train_s, y_test_s = train_test_split(x, y, train_size=0.8, random_state=42)

In [34]: # Literary
         x = literary_df.drop('average', axis=1)
         y = literary_df['average']
         x_train_l, x_test_l, y_train_l, y_test_l = train_test_split(x, y, train_size=0.8, random_state=0)

In [35]: # Industrial
         x = industrial_df.drop('average', axis=1)
         y = industrial_df['average']
         x_train_i, x_test_i, y_train_i, y_test_i = train_test_split(x, y, train_size=0.8, random_state=42)
```

8. Modeling:

In order to extract the appropriate note that connects between the inputs and the outputs in accordance with the nature of the model itself, the models are now taken from this "Sklearn library" and put to the test on training data (figure 5).

Figure 5

The Modeling Stage

Modeling

```
In [4]: # Models
        from sklearn.linear_model import LinearRegression
        from sklearn.linear_model import BayesianRidge
        from sklearn.svm import SVR
        from sklearn.neighbors import KNeighborsRegressor
        from sklearn.tree import DecisionTreeRegressor
        from sklearn.ensemble import RandomForestRegressor, ExtraTreesRegressor, GradientBoostingRegressor, StackingRegressor
        from sklearn.metrics  import r2_score, mean_squared_error

        import tensorflow as tf
        from tensorflow.keras import datasets, layers, models, callbacks, optimizers
```

9. Evaluation:

    In addition to determining whether the model was a useful tool or not, whether it needs to be edited, and whether there are any issues like underfitting or not, the accuracy and performance of the model are assessed at this step. In addition to describing how each model functions, we'll also go over the outcomes for each major.

10. Saving:

    Here, the model was kept and was not used until it was when required. It is consistently recalled for its accuracy in expecting the outcomes of the expected GPA.

The Data Mining using ML Algorithms:

Data mining (DM) is defined as a computational tool for processing data utilized in many different fields with the goal of extracting knowledge from the data. In order to create a specific model that will be used to identify the new information, the DM approaches were used. Several different types were applied in this study, including Decision Trees, Artificial Neural Networks, K-Nearest Neighbor, Random Forest, Linear Regression, Support Vector Machines, Gradient Boosting, Extra Trees Regression, Model Stacking, and Bayesian Ridge Regression.

1. Linear regression:

With the use of this research, it was possible to forecast the value of a variable, like GPA, based on the value of a different variable. The independent variable was the one utilized to predict the value of the other variable. This type of analysis can more accurately anticipate the value of the dependent variable by estimating the coefficients of a linear equation including one or more independent variables. It is allowed for linear regression to use a straight line or surface that reduces the discrepancies between the projected values and the actual output values.

2. Decision Tree:

It is regarded as one of the most typical approaches to expectancy. The technique has been extensively adopted as the best method for researchers since it has many beneficial characteristics, including simplicity and understandability for identifying both small and large data structures and anticipating their values. In data mining, decision trees were used to generate trees after measuring the training group. It is employed to make predictions. The most popular and useful classification tools were thought to be decision tree classifiers.

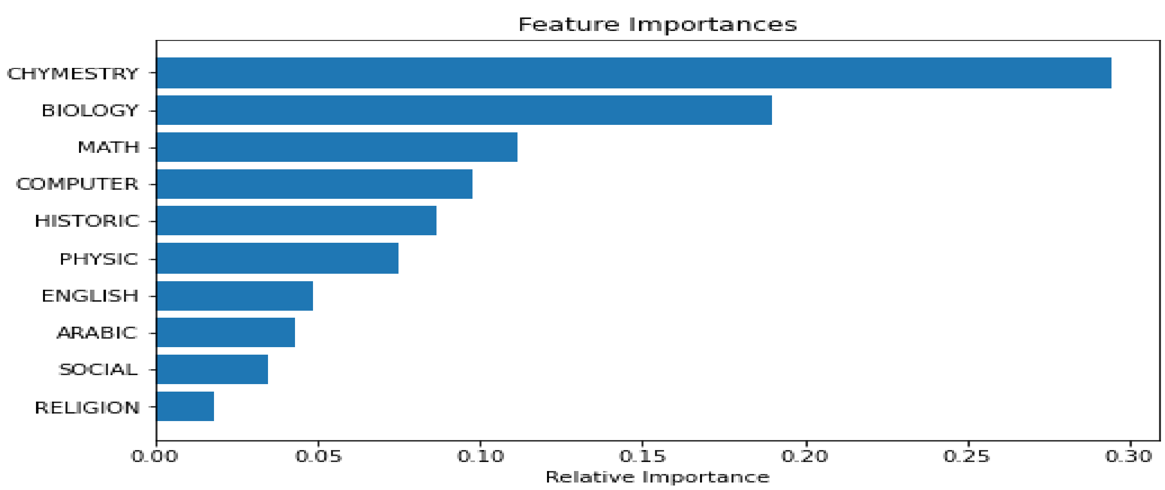
Typically, decision tree classifiers have a tree-like structure with leaf nodes at the end and root attributes at the beginning. It has several specializations with dissimilar features, and each specialty has a leaf node to display a class or kind of class distribution. In addition to determining the relative relevance of qualities, the decision tree algorithms also attempt to define the link between attributes. Decision trees' benefits are defined as characterizing rules that are simple to understand and interpret for users. Additionally, they perform admirably with both numerical and categorical variables and do not require laborious data preparations.

3. Extra Trees Regression:

This type of algorithm uses averaging to improve predictive accuracy in addition to minimizing over-fitting and is extremely ideal for several random decision trees (also known as extra-trees) utilized for many sub-samples of the dataset (figure 6).

Figure 6

The Feature Importance of Random Fore Extra Trees Regression



4. Neural Network:

As one of the often-employed methods in educational data mining, neural networks was used as one such method. The advantage of a neural network is that it can find all potential interactions between the predictor's variables. Even in intricate nonlinear relationships between both types of independent and dependent variables, neural networks were able to detect the object without a doubt. One of the top prediction tools was ultimately decided to be the neural network approach.

5. K-Nearest Neighbor:

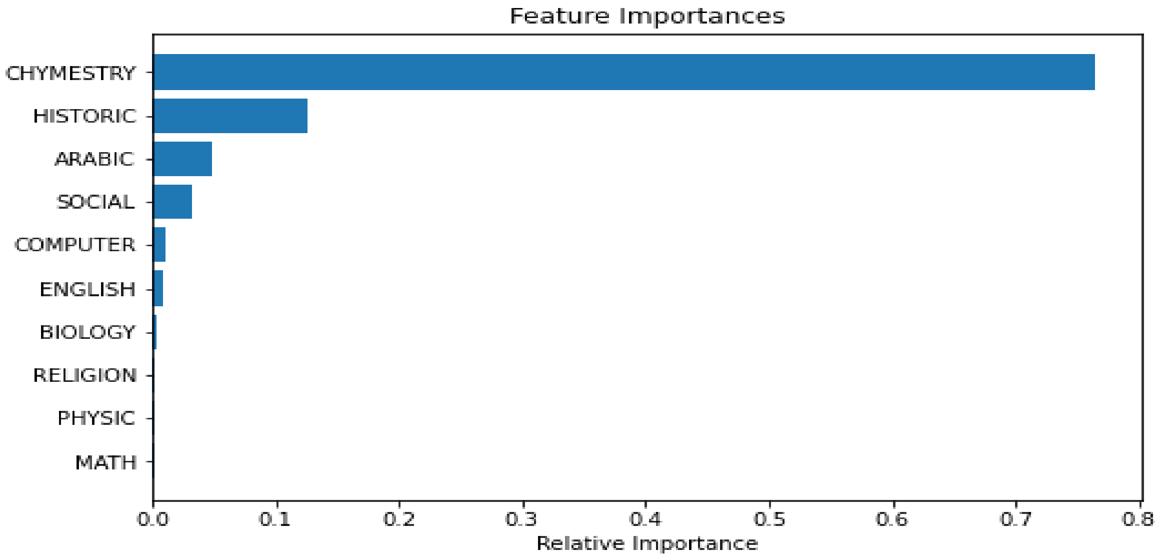
Based on the nearby training samples in the feature space, the k-Nearest Neighbor algorithms (k-NN) can create structured objects. K-NN is seen as an example of instance-based learning (or lazy learning). In this instance, a close approximation of the function was used, and the entire calculation was delayed in anticipation of categorization. The main problem with the k-NN method is that loud or inappropriate features have the potential to seriously affect the system's accuracy. Similar to this, if the feature balance is unreliable in relation to their relevance, its accuracy may become unfortunate.

6. Gradient boosting:

One of the most potent algorithms in the field of ML is the gradient boosting technique, which is used to measure bias errors and variance errors in ML algorithms. One boosting approach that is utilized to reduce the bias error of the models is known as gradient boosting (figure 7).

Figure 7

The Feature Importance of Gradient Boosting



7. Bayesian Ridge Regression:

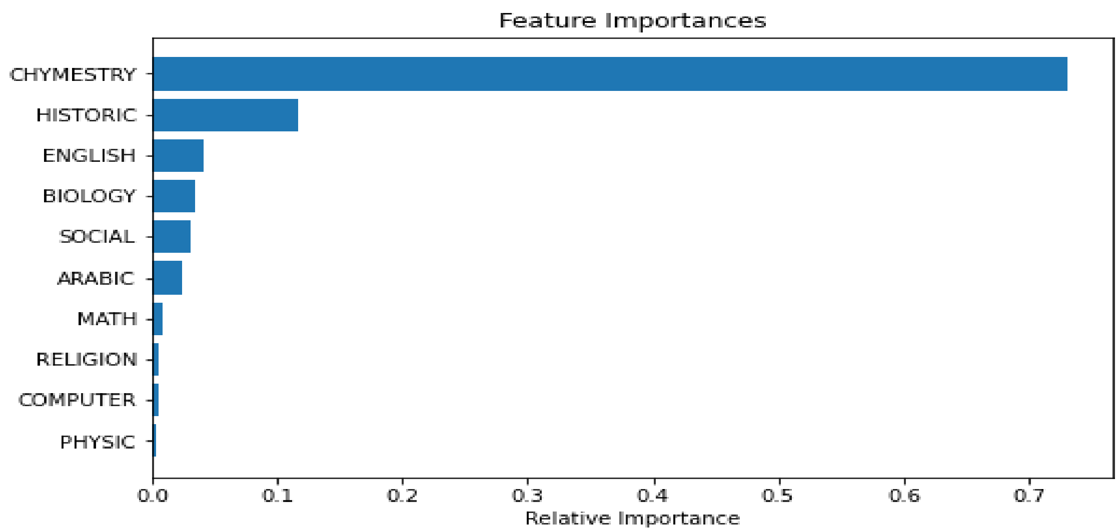
Bayesian regression can let insufficient data survive or disseminate data by expressing linear regression with probability distributors rather than point estimations. As a result, it is anticipated that the output "y" will be drawn from a probability distribution rather than being estimated as a single value. A helpful method for dealing with limited or even unevenly distributed data is provided by Bayesian linear regression. Additionally, it permits applying a prior to the noise and coefficients in the absence of data, where the priors will assume control. Additionally, the components of the Bayesian linear regression that are effective for the data were utilized. If it is confident in it or not, as well as which aspects are uncertain, was the question at hand (it might be based on the priors completely).

8. Random Forest:

A supervised ML algorithm known as random forest is used extensively in classification and regression problems. Its task is to build decision trees using a variety of samples, and it must decide on their generality for both classification and average accuracy when a regression is involved. Perhaps the most important characteristic of the Random Forest Algorithm is its capacity to manage data sets containing constant variables analogous to regression and particular variables similarly to classification (figure 8). Additionally, it does well in classification issues. In essence, random forests were constructed from subsets of data, with the final result relying on the ranking average and addressing the issues with overfitting. As a result, the Random Forest employs no formula groups and instead chooses observations at random, builds decision trees, and takes the average result.

Figure 8

The Feature Importance of Random Forest



9. Model Stacking

By collecting the outcomes of different models and running them via a meta-learner, another ML model, model stacking is a method for improving model predictions. In essence, a stacked model works by funneling the output of several separate models through a "meta-learner." Although a linear regressor or/and classifier is frequently used, other models, like decision trees, are also possible. The meta-learner seeks to optimize the positive aspects of each model while minimizing its negative ones. The outcome is typically a highly strong model that generalizes new data.

10. Support Vector Machine

The term "Support Vector Machine" (SVM) refers to a supervised ML process that can be utilized for problems involving both classification and regression. However, classification issues were utilized. The SVM method represented each data point as a point in an n-dimensional area, where n is the number of features, and each feature's value correlates to a certain position. Then, we carry out classifications by locating the hyper-plane that successfully distinguishes between the two courses.

11. Merging All Together and Testing Model:

The subjects and the years were merge together using the following ML algorithms

In [22]:

```
# Creating 1 dataframe for all data
df = year11_df.copy()
```

In [23]:

```
# All in One Dataframe
df['ARABIC'] = (year7_df['8ARABIC'] + year8_df['9ARABIC'] + year9_df['10ARABIC'] + year10_df['11ARABIC']) / 4
df['ENGLISH'] = (year7_df['8ENGLISH'] + year8_df['9ENGLISH'] + year9_df['10ENGLISH'] + year10_df['11ENGLISH']) / 4
df['RELIGION'] = (year7_df['8RELIGION'] + year8_df['9RELIGION'] + year9_df['10RELIGION'] + year10_df['11RELIGION']) / 4

df['MATH'] = (year7_df['8MATH'] + year8_df['9MATH'] + year9_df['10MATH'] + year10_df['11MATH']) / 4
df['PHYSIC'] = (year7_df['8PHYSIC'] + year8_df['9PHYSIC'] + year9_df['10PHYSIC'] + year10_df['11PHYSIC']) / 4
df['CHYMESTRY'] = (year7_df['8CHYMESTRY'] + year8_df['9CHYMESTRY'] + year9_df['10CHYMESTRY'] + year10_df['11CHYMESTRY']) / 4
df['BIOLOGY'] = (year7_df['8BIOLOGY'] + year8_df['9BIOLOGY'] + year9_df['10BIOLOGY'] + year10_df['11BIOLOGY']) / 4
df['COMPUTER'] = (year7_df['8COMPUTER'] + year8_df['9COMPUTER'] + year9_df['10COMPUTER'] + year10_df['11COMPUTER']) / 4
df['SOCIAL'] = (year7_df['8SOCIAL'] + year8_df['9SOCIAL'] + year9_df['10SOCIAL'] + year10_df['11SOCIAL']) / 4
```

In [24]:

```
df.head()
```

Out[24]:

All the three models of this study were tested together using the following ML algorithms.

In [89]:

```
science = pickle.load(open('scientific_model.sav', 'rb'))
literature = pickle.load(open('literary_model.sav', 'rb'))
industry = pickle.load(open('industrial_model.sav', 'rb'))
```

In [90]:

```
for i in [0, 5, 7, 17, 25, 27]:
    recored = np.array(x_train_i.iloc[i].tolist()).reshape(1,-1)
    s = science.predict(recored)[0]
```

```
    l = literature.predict(recored)[0]
    i = industy.predict(recored)[0]
    print(f'Scientific: {s:.2f}, Literary: {l:.2f}, Industrial: {i:.2f}')
print('- - - - - - - - - - - - - - - - - - -')

for i in [0, 5, 7, 17, 25, 27]:
    recored = np.array(x_train_l.iloc[i].tolist()).reshape(1,-1)
    s = science.predict(recored)[0]
    l = literature.predict(recored)[0]
    i = industy.predict(recored)[0]
    print(f'Scientific: {s:.2f}, Literary: {l:.2f}, Industrial: {i:.2f}')
print('- - - - - - - - - - - - - - - - - - -')

for i in [0, 5, 7, 17, 25, 27]:
    recored = np.array(x_train_s.iloc[i].tolist()).reshape(1,-1)
    s = science.predict(recored)[0]
    l = literature.predict(recored)[0]
    i = industy.predict(recored)[0]
    print(f'Scientific: {s:.2f}, Literary: {l:.2f}, Industrial: {i:.2f}')
Scientific: 49.73, Literary: 48.72, Industrial: 49.38
Scientific: 69.63, Literary: 72.78, Industrial: 79.50
Scientific: 60.62, Literary: 80.44, Industrial: 65.88
Scientific: 49.61, Literary: 49.36, Industrial: 47.62
Scientific: 60.62, Literary: 80.44, Industrial: 65.88
Scientific: 57.76, Literary: 72.50, Industrial: 54.00
- - - - - - - - - - - - - - - - - - -
Scientific: 58.27, Literary: 69.25, Industrial: 54.15
Scientific: 60.09, Literary: 50.00, Industrial: 62.68
Scientific: 67.61, Literary: 84.86, Industrial: 77.88
Scientific: 49.37, Literary: 48.50, Industrial: 49.26
Scientific: 50.88, Literary: 49.88, Industrial: 63.37
Scientific: 68.01, Literary: 49.00, Industrial: 73.18
- - - - - - - - - - - - - - - - - - -
Scientific: 88.73, Literary: 86.76, Industrial: 87.91
Scientific: 48.44, Literary: 58.86, Industrial: 50.82
Scientific: 68.00, Literary: 83.45, Industrial: 68.12
Scientific: 87.44, Literary: 82.02, Industrial: 71.76
Scientific: 86.01, Literary: 81.81, Industrial: 86.17
Scientific: 91.22, Literary: 88.58, Industrial: 90.99
```

Done!

3.3.1.1.2 The Second component: The Mobile Application (level examination)

At this time, Android was the mobile operating system used to construct the PSA (Figure 9). There are two user interfaces for the application: one is for students and the other is for teachers. According on their performance levels in class, teachers were employed to provide daily tasks and feedback to the students using teaching materials. The teaching academic materials are available to the pupils through the PSA. They can also utilize the program to turn in schoolwork or tasks and have private discussions with their teachers and peers about any issues they might be having. Moreover, it allows students to schedule appointments with advisors who specialize in their particular subject area in order to get assistance, direction, and solutions for any issues they might encounter while carrying out their other academic duties. Using the PSA, the teacher can also communicate with students through messages, respond to their inquiries, and discuss their ideas and concerns with them.

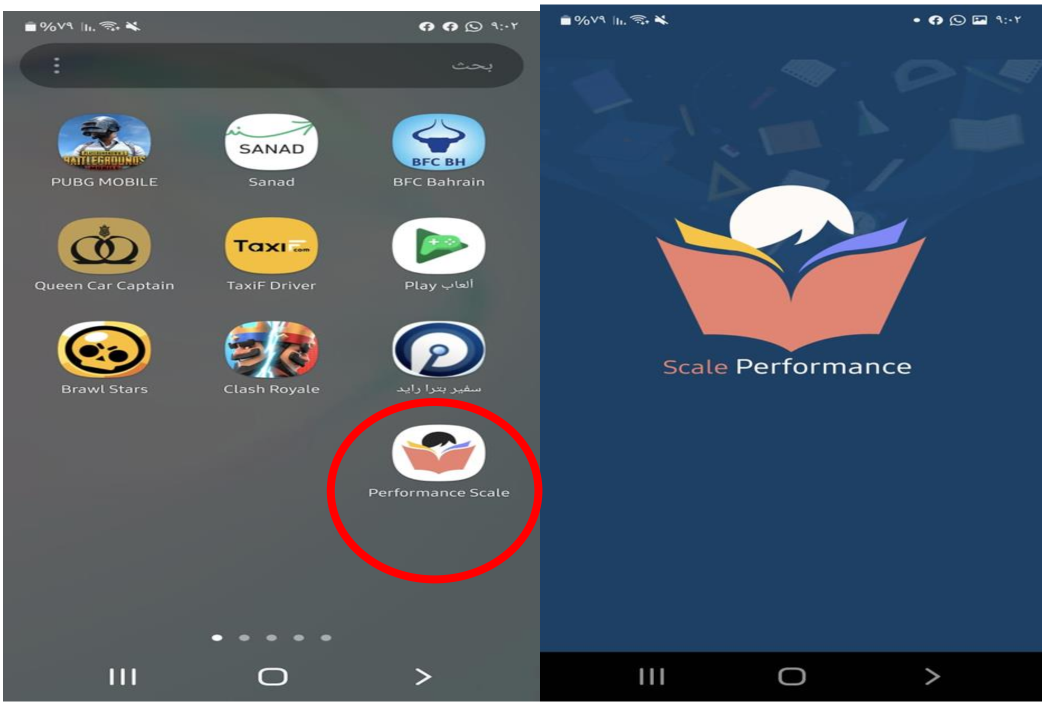
The key functions and tasks of the include:

First, the typical grade 11 student performance in the fundamental topic is: This step is taken because students still have time to alter their opinions before it's too late in the 12th grade even if they've already chosen a portion of their specialization (specialist of study). In this grade, they are unable to switch their major. The fact that this strategy will make the course and learning setting in this class more similar to that of the 12th grade, right down to the level of difficulty of the courses, makes it essential for progressing the system and achieving its goals. As a result, it is simpler to identify whether a student has the abilities and level of productivity required to effectively achieve this course with high performance.

Secondly, the development of a level test for the foundational courses in each specialization. The outcome of the exam exposes the student's performance level and scientific level in this field, making this stage crucial in raising and improving the correctness of the systems' work. Since the average is one of the inputs into the forecasting system, the result of this step is expressed as a percentage. The average will also be utilized to determine the sum of the results of the first and second steps. The additional input, however, is the projected GPA.

Figure 9

The Scale Performance Application



It should be noted that the results, including the projected GPA and the average exam level with basic course marks, were submitted from the mobile application to the genetic method. Moreover, using the rest full API, the mobile application communicated the genetic method in the ML the results of the level exams as well as the average of the major course grades. Because these data were regarded inputs for the genetic algorithms, this was done by constructing a Java code. After concluding the genetic method's forecast of the student's achievement concerning their best specialty, the proposed system was given the green light to produce the final output, which would then appear on the mobile app.

3.3.1.1.3 The linkage mechanism between the mobile app and the machine learning

The process of establishing a connection between the mobile application and ML was carried out by sending API (rest complete API) requests by submitting request code in Java. There was a database record for this request query. This assignment was completed when the user enters grades from 7th to 11th using a mobile app, or when they calculate the average level exam % using the average of their basic course grades.

Additionally, when the machine learning responds to a request from the mobile app, a response is also sent via the full API. The machine learning responds by obtaining the anticipated GPA and final result, which is the anticipated outcome of students' performance, and by assisting them in selecting the specialty that best matches their skills and abilities. To illustrate, sending an API request and responding to it with a full API allows ML and mobile applications to communicate with each other.

- The rest full API for mobile app:

1. Entering students' grades for class hours from grades 7^{th} to 11^{th}.
2. Obtaining the total average of the sum of both level exams and the average of marks for the basic subjects.

- The request full API response ML:

1. Obtaining the expected GPA
2. Obtaining the final result using a genetic algorithm

3.3.1.1.4 The Third component: Genetic Model

Global search heuristics include genetic algorithms. They are regarded as search techniques that are utilized in computers with the intention of finding precise and accurate answers to problems that need to be optimized and searched. It employs certain techniques like crossover, modification, selection, or even inheritance that are inspired by biological evolution.

This search-based algorithmic approach is used to address the precise optimization problem in machine learning. Since it solves challenging issues that require time and effort to solve, it is very important. Data centers, electronic circuit design, codebreaking, image processing, artificial creativity, and other real-world applications have all employed it. In order to handle search and optimization issues, the genetic algorithm is a heuristic search technique. These algorithms are a subset of evolutionary algorithms. They were mainly employed in computing. They offer answers based on the theories of genetics and natural selection. They are also distinguished because they are more intelligent than random ones because they employ prior data to focus the search on the area of the solution space with the highest performance.

i. Working Steps:

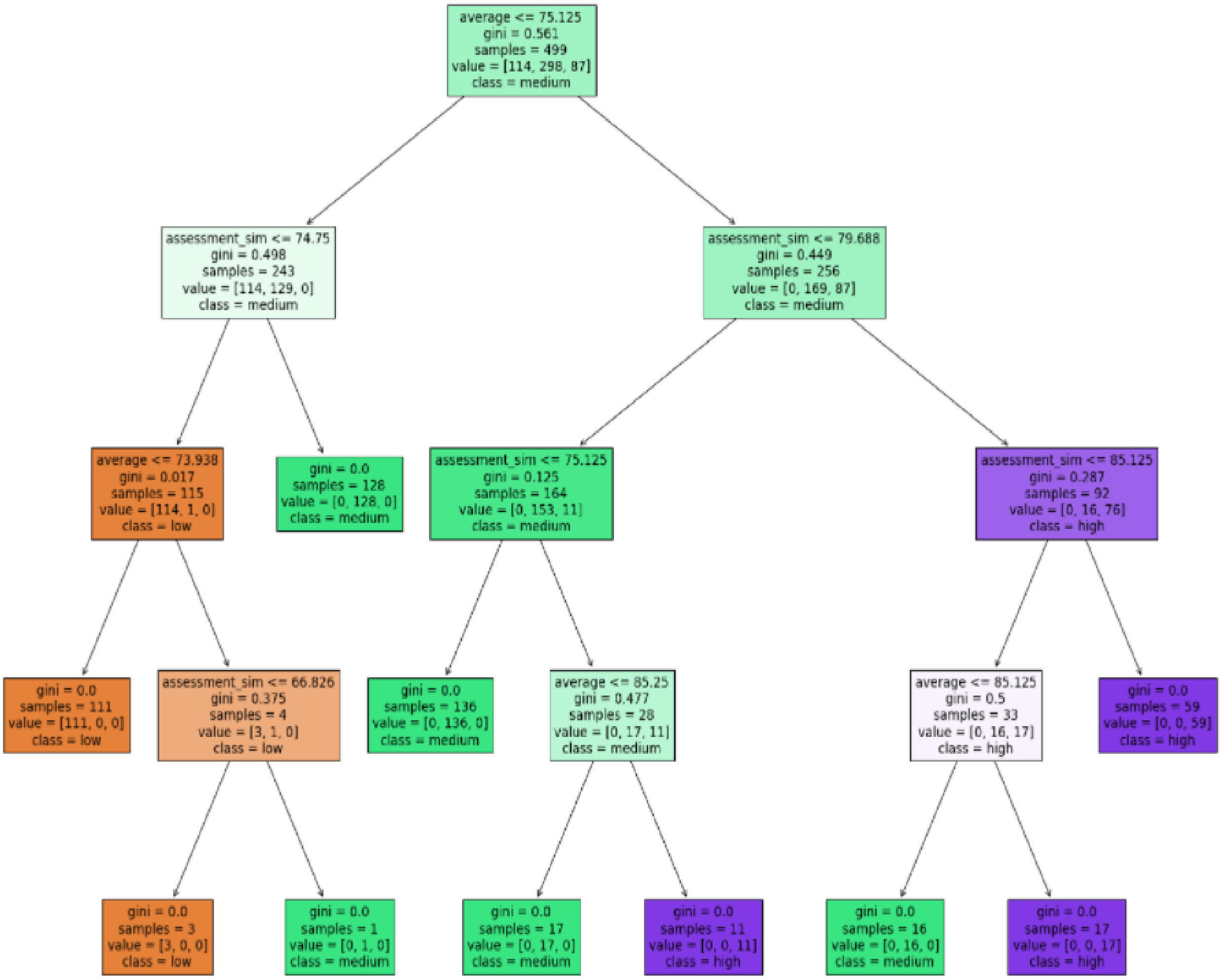
1. Reading data as mentioned previously, but here the GPA is read only.
2. Making a simulation for the assessment result through making a copy of the GPA and mixing it randomly.
3. Defining some rules.
4. Applying the rules and getting the final performance according to the rules defined previously.
5. Using the division tree model and exercising it where the inputs are the expected GPA and assessment results and the output is the students' performance,

The genetic algorithm way of working depends on two main inputs: the expected GPA and the average of the total percentage of level exams; and the average of the main subjects' marks. However, the inputs are either high, medium, or low for the expected students' performance (figure 10).

- In the high case, this means that students' decisions are suitable for their abilities and skills to be correct.
- In medium cases, a student's decision is good, and he can continue in this specialty, but sometimes there is a better choice.
- In this case, the student's decision is rather good to his abilities compared to other specializations, or it is not appropriate for him completely and he must change it.

Figure 10

The Structure of Genetic Model



The work mechanism was divided into four sections given numbers, as follows:

- 0_ 65 (Possible)
- 65_75 (Good)
- 75_85 (Very good)
- 85_100 (Excellent)

ii. **Advantages**:

- Accuracy about 99.2%.
- Better interpretation ability.
- Easy to change and configure.

3.3.1.2 Mobile Application Development

The PSA includes two user interfaces that users can select from when the system first launches: a instructor interface and a student interface. The system opens when the user inputs the code, provided it is available or has previously been registered in the database as a user. User interface (UI) functionality is the foundation for these operations.

The PSA's development started with the UI because a decent UI may help users, and since the objective and content of any electronic mobile app are what define it, that's where it started (students, instructors, and schools, with a principal emphasis on students) in maintaining continuous concentration on a chosen and targeted item or subject. What the PSA UI functions include is listed below.

Connectivity, or the ability to quickly and conveniently obtain information, is the first element. Additionally, it enables simple pausing, restarting, and pausing of the PSA. The second factor, which comprises of three characteristics, is usability. The initial characteristic of the application is to use a straightforward design plan to reduce memory usage. This makes the application simple to remember and motivates the pupils to concentrate more. The user can read it fast because to the second feature, which is information that is brief and to the point. The third characteristic is split attention, and because users of e-mobile tools frequently need to complete multiple tasks at once, the user interface (UI) shouldn't be unduly complicated.

The PSA's capacity to lead the user via a stepwise process, menu, or decision that requires obtaining related data from the system is referred to as the third component, or direction. The PSA's emphasis on the pupils' capacity for decision-making was its greatest asset. The PSA receives information from the fourth component. It is one of the most helpful UI features and a crucial requirement for communicating fundamental and important information. In fact, the knowledge can be broadly disseminated provided it is combined with the appropriate arrangement. The fifth element is interactivity, which is how students interact with the PSA. This element has a few important characteristics. The number of options on the landing page was considered in order to prevent overcrowding and losing the interactive element.

Sixth on the list is usability. To establish a positive and dependable student knowledge for efficient learning, a user-friendly UI was built. The eighth element is richness or comprehensiveness. A comprehensive collection of elements must be developed by the software in order to support content modification. This allows the program's users—learners, instructors, and schools—the opportunity to manage it in accordance with their individual talents and knowledge. This is the PSA's contribution to this development's second major strength. Consistency is the seventh element. A successful user interface (UI) will have consistent interactions. Users will be helped by consistency to form their ideas and interpretations of the software, which will create room for improvement. The ninth element is personalization. Experienced learners strongly desire to feel in charge of the UI and to have the UI respond to their actions. They will have control in the way of favourite over how their learning activities are planned and carried out. The PSA also features three (3) different kinds of screens.

In this stage, the researcher designed mobile app to helping the students' pass in the secondary stage on time by guide them to choosing the best branch in high school and by enhance their academic performance, to avoid failure and depression, as well as save their time etc (figure 11). To develop the system, several steps were applied like:

1. The First Step:

Predicting the academic performances and the best branch for each student, this step relied on two inputs see:

I. The expected cumulative average for the secondary stage in the three branches (scientific, industrial, literary), which is obtained by applying the extra tree regression algorithm on the grades of students from the seventh to the tenth grade.

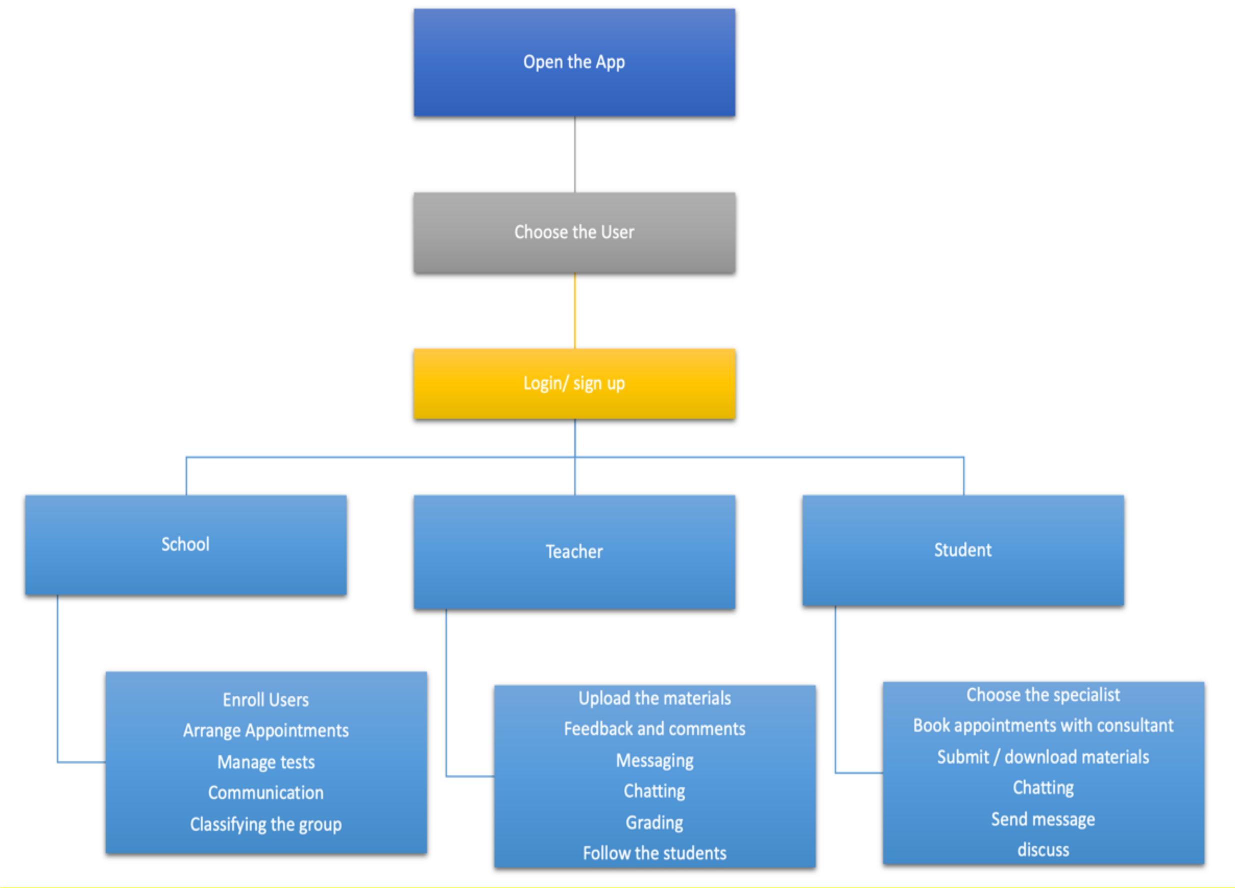
II. The result of the level exam in percentage value, this exam was hold for each student in the basic courses in each branch by using mobile app.

2. The Second Step:

Here the system begins the process of helping the student pass the stage by improving his performance to finish this stage on time. Whether he wants to stay in his branch or change to the branch suggested to him by the system. Students in each subject in all branches will be divided into three groups based on their level of academic performance (high, medium, low), in the four years preceding high school.

Figure 11

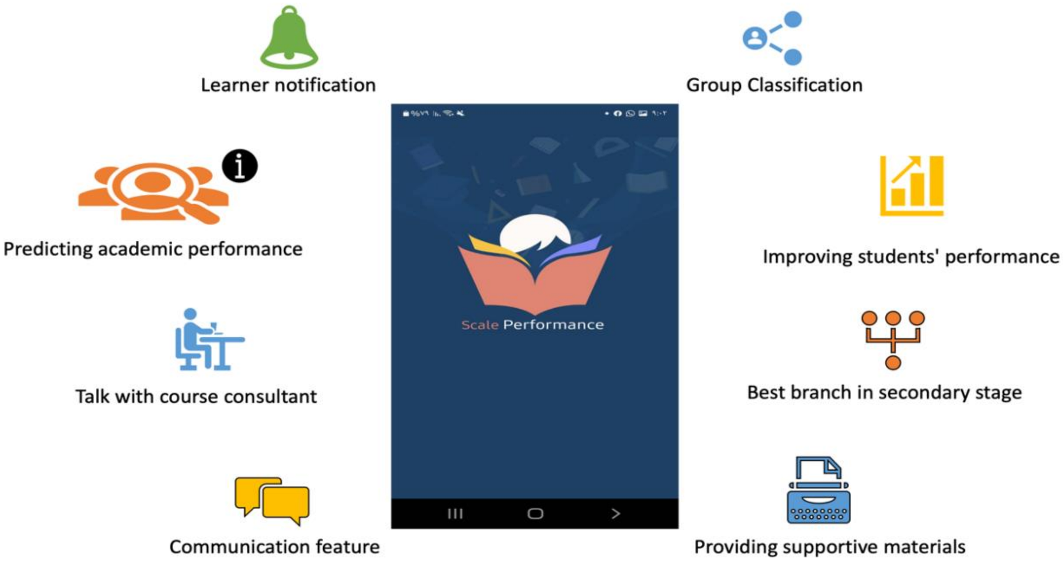
The Mobile Application Development Design



Here the system begins the process of helping the student pass the stage by improving his performance to finish this stage on time. Whether he wants to stay in his branch or change to the branch suggested to him by the system (figure 12).

Figure 12

The Mobile Application Development Design



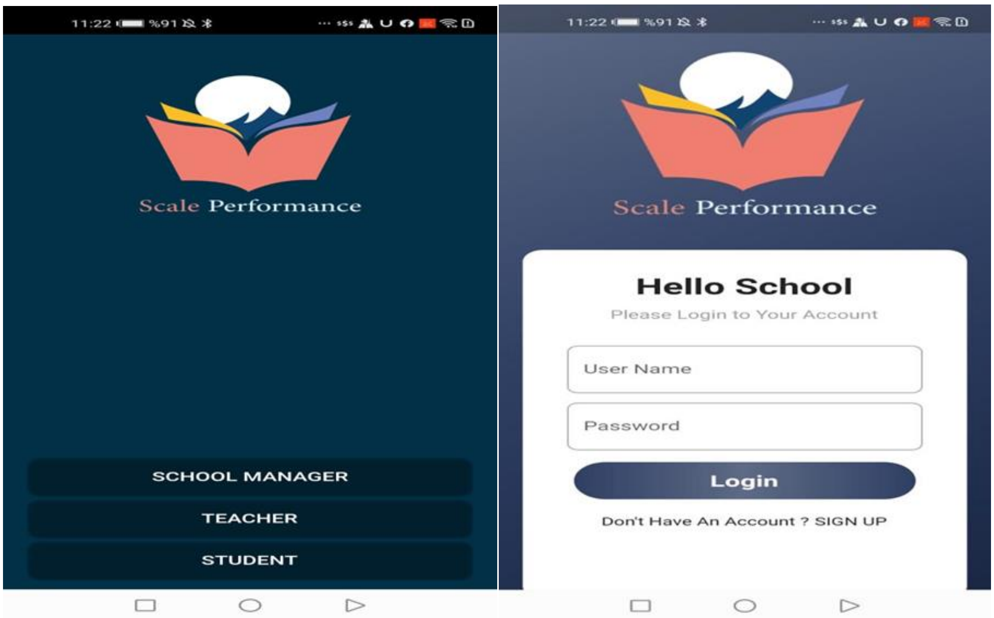
3.3.1.2.1 The Mobile Application Features

Firstly, students in each subject in all branches will be divided into three groups based on their level of academic performance (high, medium, low), in the four years preceding high school. Secondly, the school begins to help students in each subject according to the group to which the student belongs, by providing him with lessons and supportive exams, linking him to famous sites, and increasing focus on weak points in each group. Thirdly, the application also provides a feature comment between the student to discuss some topic, also send private message to the teacher. Fourthly, application also provides a feature allow teachers to send a message to one student or to all students such as announcements. fifth, application also provides the feature of alerting the student when he receives any message or when providing him with any academic content or any assignment. Finally, application also provides a feature allow students to speak with course consultant.

From the interfaces, the user chooses whether they are school or a teacher or a student at the start of the system and login (figure 13).

Figure 13

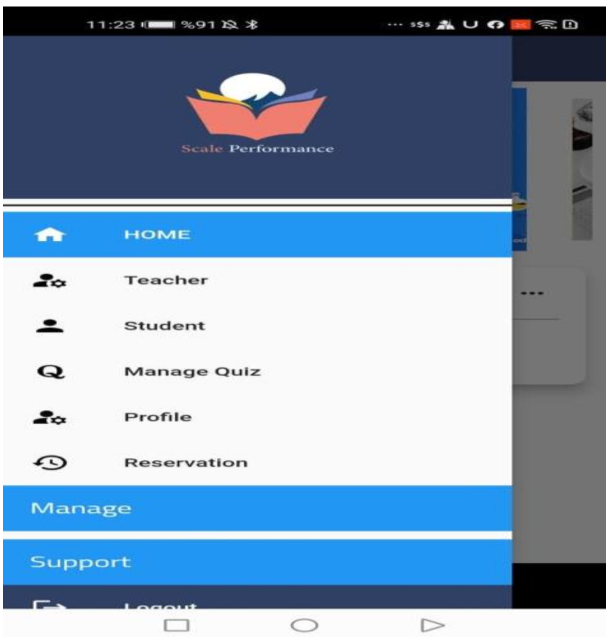
The Login And Sign-Up Screen In PSA



In addition to the school dashboard allow to us to enroll users (student, teachers), manage quiz and created it, update profile and arrange the appointment (figure 14).

Figure 14

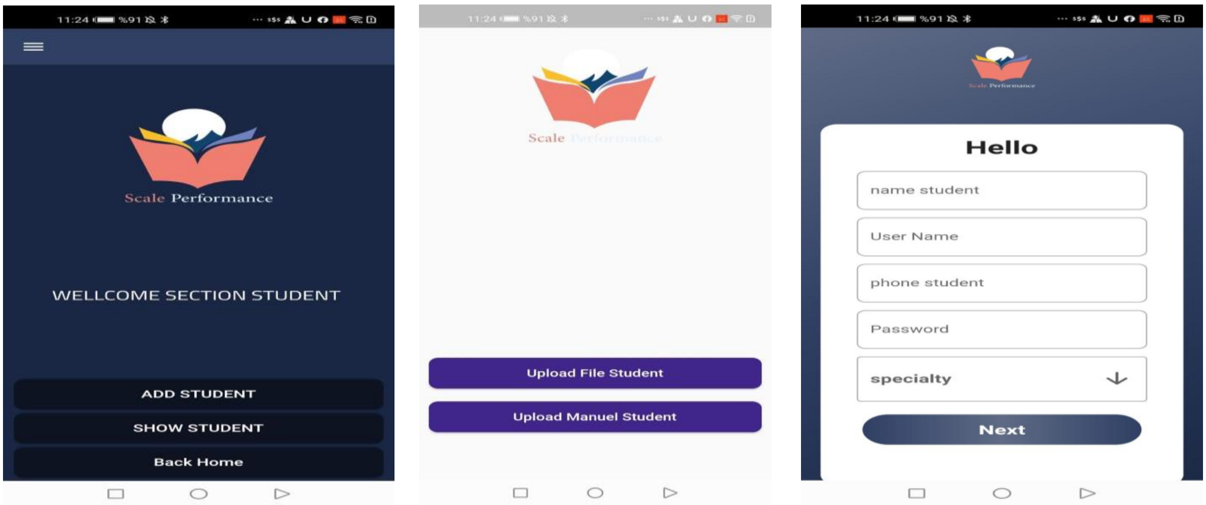
The Dashboard Screen In Performance Scale Application



The school can enroll student and enter his marks from grade 7th to grade 10th manual or upload csv file (figure 15).

Figure 15

The Enroll Students Screen In Performance Scale Application



The school can enroll teacher and his department and privileges (figure 16).

Figure 16

The Enroll Teachers Screen In Performance Scale Application

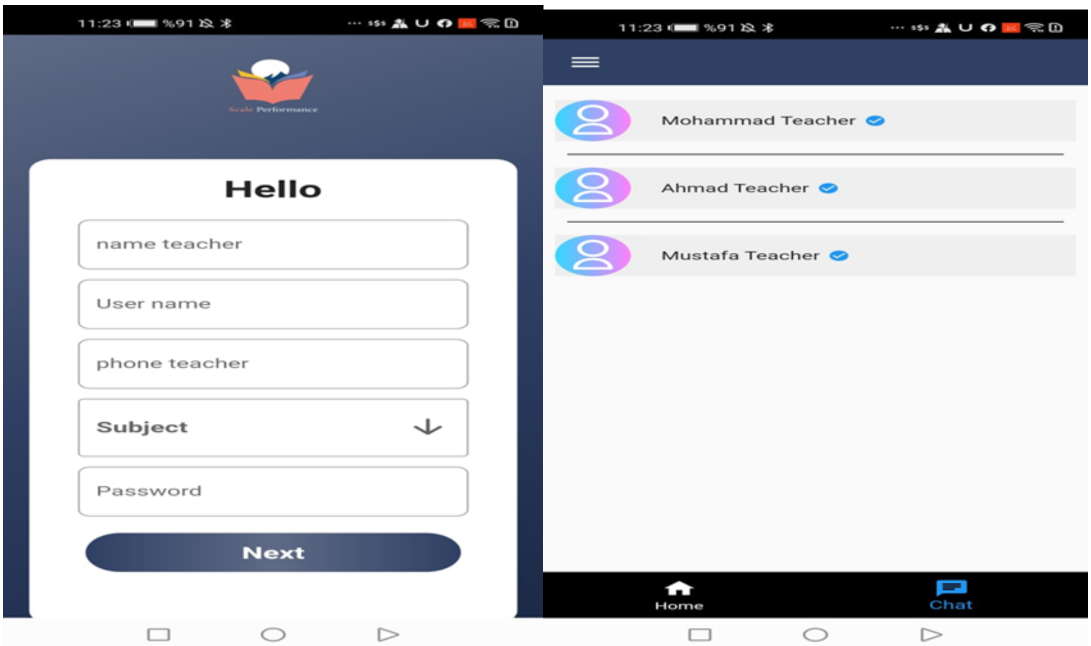


Figure depicts a panel that gives users complete control over level assessments, particularly by modifying their numbers to reflect changes in courses, educational curricula, the questions, and the amount of time provided for each exam (figure 17).

Figure 17

Screen for Exam Control

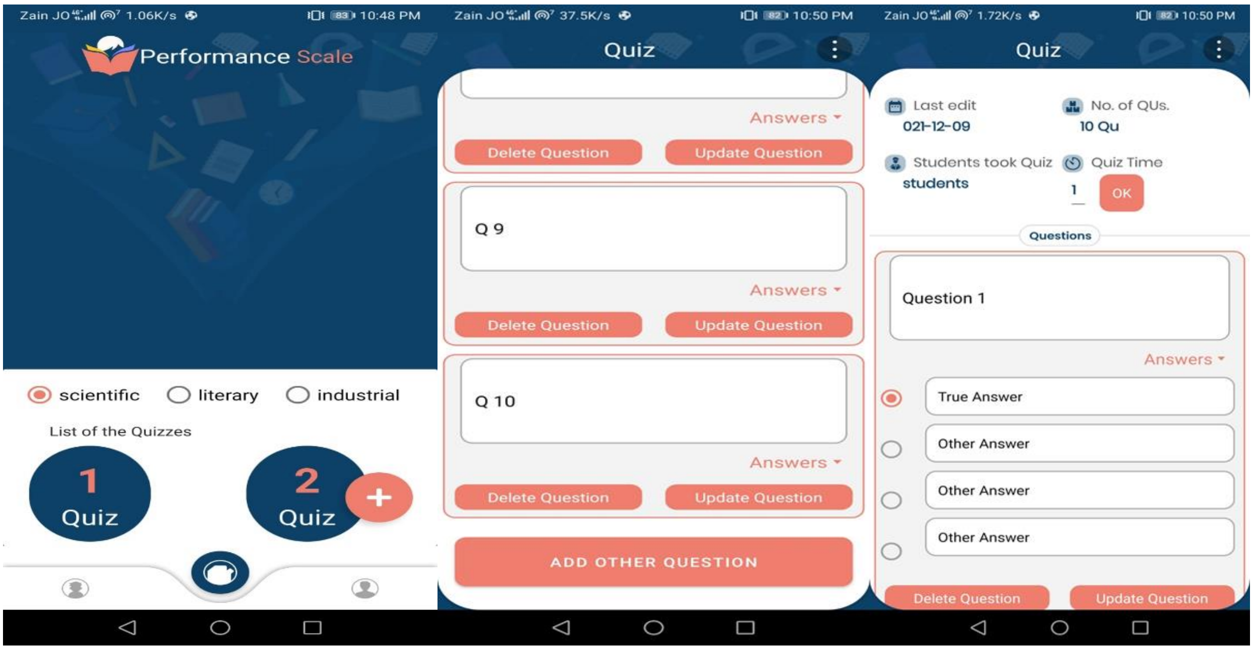


Figure shows a screen containing the names of the students, their specializations, and the application's outcomes (Figure 18). The school will be taken to a new screen after selecting the student's name, which will show the student's level exam scores, the average of their basic subjects, the genetic algorithm's findings (figure 19).

Figure 18

Student Names And Specializations On A Screen And The Application's Results

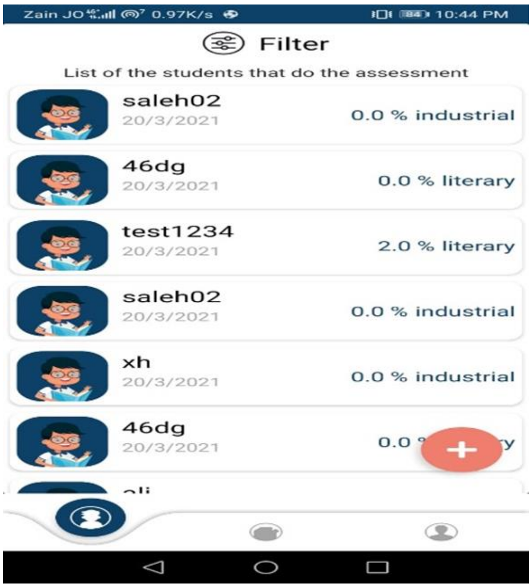
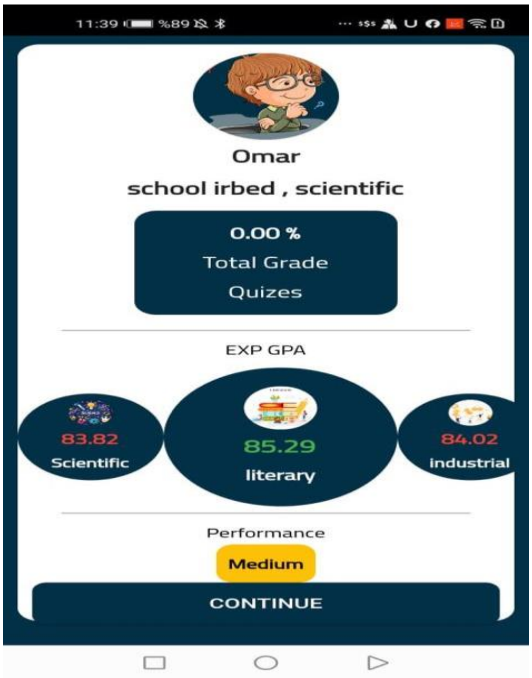


Figure 19

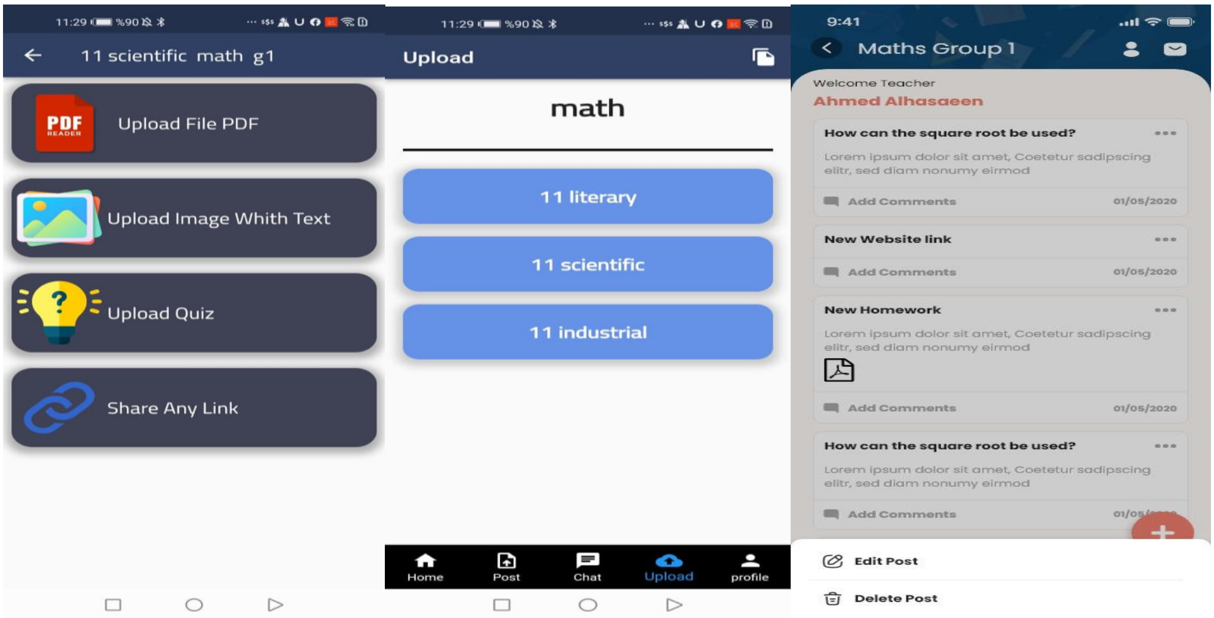
The average grade in the fundamental subjects, a student's level exam marks, and a prediction of their performance



PSA allows teacher to upload information for the students, as shown in (figure 20). They can share any instructional links or add any quiz content.

Figure 20

The Upload Screen



The Ministry of Education and subject-matter specialists collaborated to produce level exams for PSA.

The showing the result page give him choice if he wants to change his direction or continue on his choice whether the student's choice was excellent or terrible and whether there were any better options (figure 21).

Figure 21

The Choice of Branch in Performance Scale Application

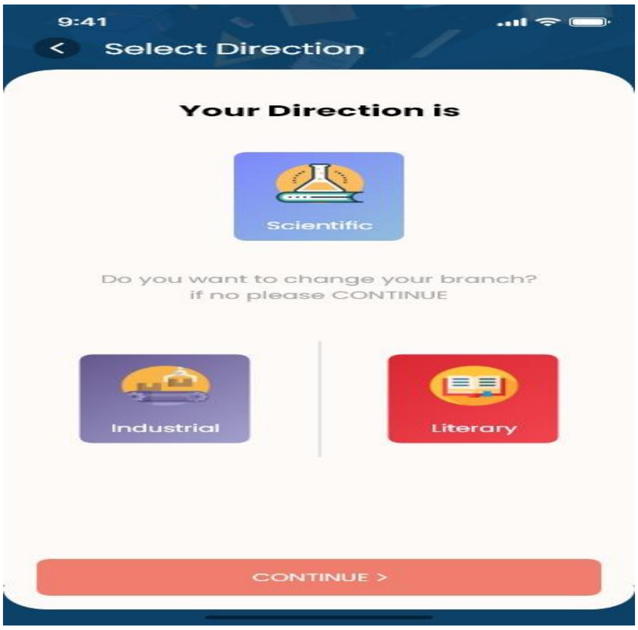


Figure shows a screen containing Then the student is classified in each subject according to his academic status and alerting the student when he receives any message or when providing him with any academic content or any assignment (Figure 22). the student can make reservation for appointment to skill support if he has any technical issue or management problem (figure 23).

Figure 22

The Classified in Each Subject in Performance Scale Application

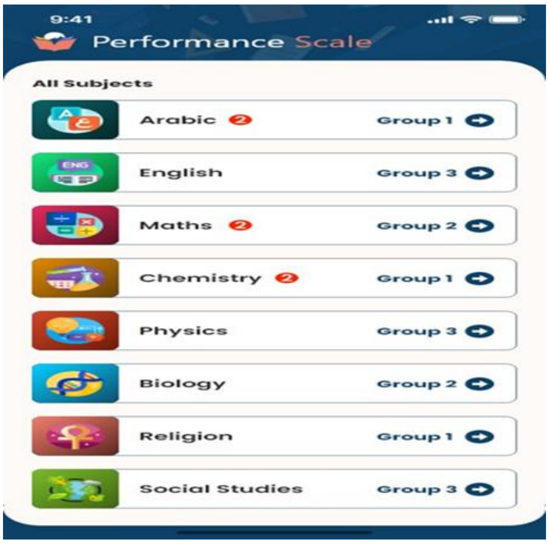
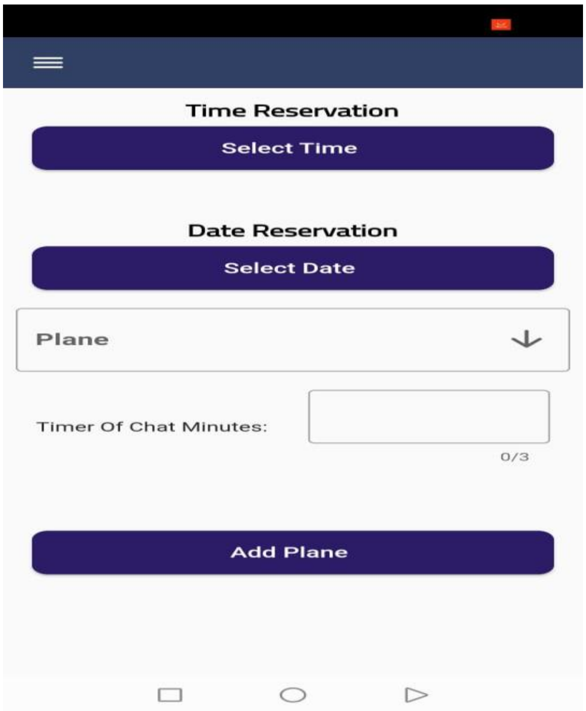


Figure 23

The Reservation Plane in Performance Scale Application

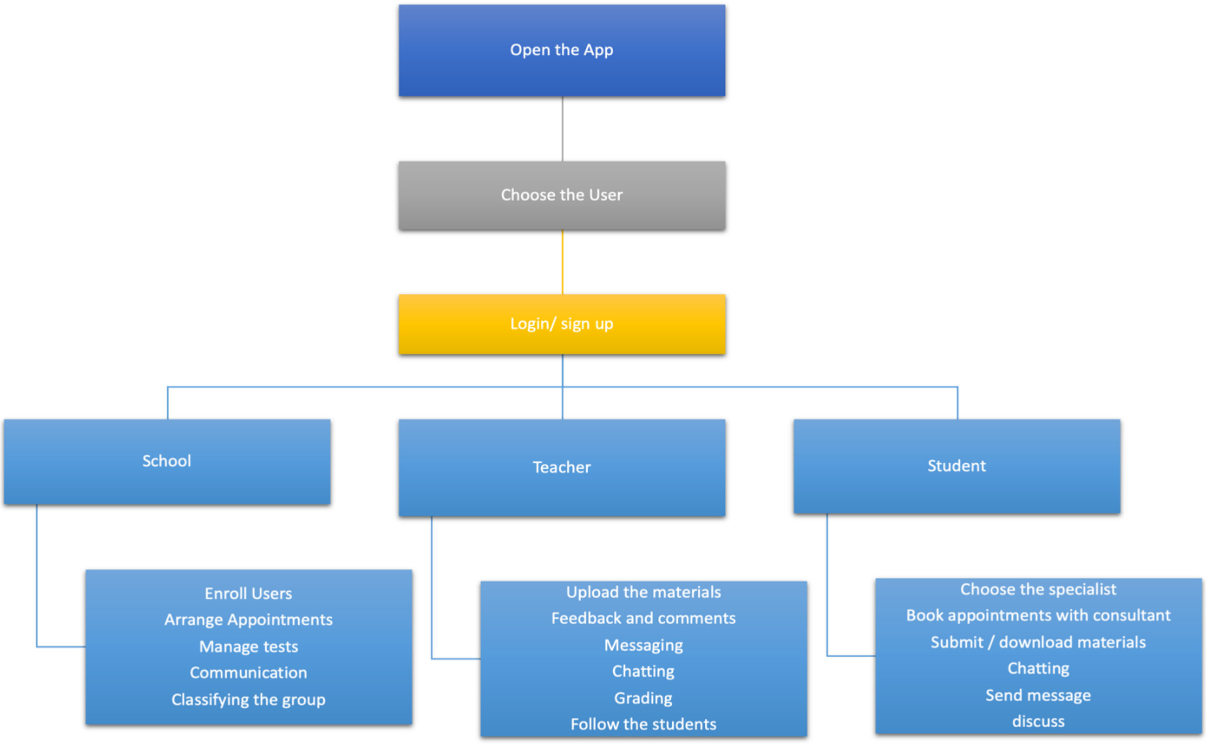


3.3.2 Methodology Phase 2: Experimental Study

This section deals with presenting the procedures followed in this research in terms of using the research methodology and adopting the appropriate experimental approach, as well as defining the research population, selecting the sample, and making an equivalence between groups (see Figure 24). This experimental research intends to examine the effectiveness of the developed PSA as a mobile application as an assistive tool to guiding the students in secondary school and help them to pass in this level on time. Data from both quantitative and qualitative sources are gathered and examined using a mixed-methods methodology.

Figure 24

The Mobile Application Development Design



3.3.2.1 Study Design

The present study used a classic controlled experimental design. In this approach, the research model consists experimental or control group when the students are randomly assigned to both groups (Melo, 2019). The experimental group was taught using the developed PSA mobile application, while the control group was taught according to the traditional method to measure the academic achievement of both groups.

3.3.2.2 Ethical Committee Approval

Most prevalent protocols that investigators must adhere to while collecting data is ethical permission; they must protect the privacy of the data and only use it for the intended purpose (Buchanan & Hvizdak, 2009; Brown, Spiro, & Quinton, 2020). To obtain approval, a request was made to the university's ethical committee. A formal endorsement was given on November 16, 2021. The copy of the ethical approval is in Appendix D.

3.3.2.3 Participants, Sampling, and Data Collection

The population of this study constitute 1248 secondary students at Irbid Secondary School in Jordan. In the Middle East, Jordan is among the nations with the best educational systems. Jordan applies the standard for beginning secondary education and choice of specialization. The participants' records were sampled from 2017 till 2022. These records include the students' grades in all subjects from seventh grade through twelfth grade, with male gender of the students, as shown in Table 1. These records were acquired from the Jordanian Ministry of Education, which keeps records of students' grades long after they graduate.

The samples of this study were collected in two stages with different numbers of participants as explained below:

Beginning with: There were 75 participants overall from the Jordanian Irbid Secondary School. The students were between the ages of 16 and 17. The quantitative approach to designing user interfaces for mobile applications put forward by Jiang et al. served as the foundation for the simple sampling of students (2019). The scientific, literary, and industrial specializations, each with twenty-five pupils, are divided into three groups with the remaining students.

The second stage: Overall, 60 students participated in the case study at Irbid Secondary School in Jordan. The students were between the ages of 16 and 17. Twenty of these students belong to the scientific specialty, 20 to the literary specialty, and 20 to the industrial specialty, respectively. The students in each specialty were split into two control and experimental groups (Patton, 1987), each group consisting of 10 students, for a total of 30 students in the control group and 30 students in the experience group. All of the 60 students were from the lowest academic performance based on their grades in the four years prior to the secondary stage, for more explanation look Table 2.

Table 1

Sample of School Data

Name	ARABIC	ENGLISH	RELIGION	HISTORIC	MATH	PHYSIC	CHYMESTRY	BIOLOGY	COMPUTER	SOCIAL	average
A	80.6	84.4	91.4	85	89.25	78.50	89.50	93.50	94.00	94.00	85.111111
B	88.6	94.0	96.4	91	94.50	95.75	97.25	96.75	96.75	98.75	94.200000
I	75.4	79.2	80.8	60	58.75	52.25	62.00	62.75	73.25	89.25	55.000000
J	60.8	64.6	71.4	55	85.50	82.50	58.50	53.25	70.00	62.75	49.666667
K	70.0	73.6	80.8	60	86.00	79.25	83.75	79.75	86.25	83.25	71.666667

Table 2

Participants, Sampling, and Data Collection

	Course content & activities	Learning Strategy	Evaluation of the study	Participants
Experimental Group	Course materials and learning activities in each subject on the developed mobile application.	Active learning strategy by using the mobile application & traditional learning.	• Academic achievement (pre-test) & (post-test).	30 students from weakest group/ all department - Male gender - Average age is 16-18
Control Group	Traditional course materials and traditional	Active learning strategy by using traditional learning.	Academic achievement (pre-test) & (post-test).	30 students from weakest group/ all department

	learning activities.			- Male gender - Average age is 16-18
Questionnaire	N/A	N/A	• Students' opinions.	75 students from all group/ all department - Male gender - Average age is 16-18
Teachers Interview	N/A	N/A	• Teachers Interview & opinions.	15 teacher - Male gender - Average age is 40

3.3.2.4 Data Collection Tools

The data collection tools which were implemented in this study were: the "Academic Achievement Test (Pre-test and Post-test)" , "Teachers Opinion Interview" , and "The students questionnaire". In inductive research, there is an urgent need to know that the study measures what it claims to (Ogunsanya et al., 2019). Thus, the content and face validity was conducted for data collection tools by presenting the content to a panel of five IT experts, and special education and field experts' opinions were collected to identify the validity of the data collection tools and the academic achievement test used in this study. This is in order to indicate the extent to which the test represents the content of the material to be measured through the judgment of experts. Content validity is concerned with how accurately a test measures the knowledge and skills identified in the educational objective.

Academic Achievement Test: There was an academic achievement test for measuring academic performance. The researcher with the help of the teachers, prepared the academic achievement test sheet based on the ministry of education standards. Additionally, pre-test and post-test cover the some subjects for each literary, scientific, and industrial discipline.The researcher prepared it according to the following steps. First, identify educational topics. Secondly, deriving and formulating objectives. The researcher, with the help of the teachers in the school, applied the academic achievement test to the two groups. The test was done at the same time for the two groups. The researcher formulated the academic achievement test content of 10 true or false type questions and 5 essay questions, which are: the level of remembering , the level of understanding, and the level of application of the students based on the cognitive domain levels of Bloom's classification. It has been determined that 50 minutes was an appropriate time for the answer.

Pre-Test: A pre-test was conducted a week before the remedial intervention in order to assess the students' previous performance in the two main subjects in each department before using PSA, also to ensure if the control an experminatal group in same academic level. It consists from 10 true or false type questions and 5 essay questions.

Post-Test: The post-test was conducted 8 weeks after the application of the therapeutic intervention. In this test, the same mechanism and procedures used in the pre-test were followed with different questions.Therefore, for measuring academic performance, , there was a post test at the end of the experiment dependent on standard and criteria (MOE). There were three main areas that the students were tested for:

Main topic in Math and Physics of Scientific branch.
Main topic in Arabic and Geography of Literary branch.
Main topic in Drawing and Engineering of Industry branch.

Interview Questions: In this study, the results of the quasi-experimental study were supported by a semi-structured interview (Gay et., Al 2011). A semi-structured interview was conducted between 15-20 minutes for 15 teachers after they used PSA. The time ranged differently depending on the teachers' responses. In general, the interview seeked to identify and collect teachers' perceptions and opinions about PSA.

Students' Opinions Questionnaire: The goal of the students' opinions is to test and verify research questions and hypothesis. The students' responses to a questionnaire (Appendix A) were utilized to compile the data. To gauge how the intervention affected the pupils, questionnaires with response options on a five-point Likert scale (1 = completely disagree, 5 = completely agree) were utilized. These items include improving evaluation (IME), improving communication (IMC), improving scientific (IMSC), and satisfaction of learning (SOL). Table 3 shows how the survey's variables are distributed.

Table 3

Questionnaire Variables Distribution

Axe No.	Variables	Symbol	Sub Axe		No. Items
1	Personal Data	PD	Specialty: Scientific, literary, Industrial		1
2	Performance scale (Independent variable)	IME	Improving evaluation, (X1, X2, …, X5)		5
3	Mobile learning skills (Dependent variables)	IMC	Improving communication , (A1, A2, …, A7)	Personalized learning A1, A2, A3	3
				Distance learning A4, A5, A6	3
				A7	1
		IMSC	Improving scientific content, (B1, B2,…, B5)	Self-learning B1, B2, B3	3
				B4, B5	2
		SOL	Satisfaction of learning, (C1, C2, …, C10)	Specialty learning D1, D2, D3, D4	4
				Mobile learning D5, D6	2
				D7, D8, D9, D10	4

3.3.2.5 Cronbach's reliability

The appropriate Cronbach's dependability coefficient lies between 0.977 and 0.978. If the composite reliability and Cronbach's alpha values were both equal to or greater than 0.70, it was approved (Peterson & Kim, 2013; Bonett & Wright, 2015). The variable's index value had to be less than 0.80 (Franke & Sarstedt, 2019), the average variance extracted rate had to be at least 0.5, and the AVE square had to be greater than the inter-construct correlations (IC) components. These three criteria were used in the study to assess the discriminant validity (Farrell & Rudd, 2009). Confirmatory factor loadings also reached 0.7 and above.

3.3.2.6 Data analysis

For the gathering and analysis of data for this study, both quantitative and qualitative methods were employed (Yin, 2009). Using SPSS version 26, the quantitative data for the research topics was analyzed. First, a paired sample t-test was used to determine the mean differences between the experimental and control groups. The independent sample t-test was then used to analyze the variance in averages between groups in each course. AMOS was used to evaluate hypotheses regarding complex variable correlations.

Second, Pearson's Correlation was used to analyse two tests from each specialty. Subsequently, the independent sample t-test was used to determine whether the mean scores of the exams for the experimental group and control group in each course were equal. Third, the performance of the same group in both pre-test and post-test was compared using the ANCOVA and ANOVA analysis. For these ANCOVA, effect sizes were presented as a square and interpreted using Cohen's (Cohen, 2013) guidelines: 0.01-0.05 = small effect; 0.06-0.13 = moderate effect, $\geq .14$ = large effect. ANCOVA was used to manage initial differences between the two groups (experimental and control) regarding using the proposed mobile application. We considered the post-test in each course as the continuous dependent variable, the pre-test in the same course as a covariate, and the method (with and without application or empirical group and control group) as the interaction effects of a categorical variable.

Fourth, repeated measures analysis, or one-way repeated measures An ANOVA was used to compare the performance of the same group of participants under different experimental conditions. The significance threshold was set at 0.05.

The Kaiser-Meyer-Olkin (KMO) and Bartlett's Test were applied to determine how well the components accounted for one another in terms of the partial correlation between the variables. If the statistical probability P-value for the Levene equality of variance test was more than 0.05, then the variances of the groups were equal. Following that, the "Equal variances assumed" row was used to display the t-test results. The "Equal variances not assumed" report, however, would have been used if Levene's equality of variance test result had been significant.

3.3.2.7 Experiment Procedure

The experiment was lasting for 8 weeks, starting from 15 February 2022 to 15 April 2022 in 2nd semester classes. Three days per week, each class lasts for 50 minutes. The experimental group had the chance to use the mobile application at home. Also, experimental group could use the trail quizzes as an extra resource. In other hand the control group only study from their course book for the homework and quizzes. The teacher taught them the same content material as a traditional method without using mobile application. These procedures were the same. They were not in the same classroom physically, because the researcher believes that the two groups being in the same classroom might affect the quality of academic achievement. The whole process was observed by the researcher (see Figure 25).

- **First week**
    - Responsibilities and expectation during study.
    - Installed PSA on their devices and experimental group.
    - Small training session for students/teachers: Participants received training about how to use the system.
    - Students have been divided into the two groups in each department.
    - Conduct test.
    - Pre-test: Made sure that students in same level dependent on a previous mark in each subject.

    - Started delivering the pre-test grades.
- **8 middle weeks**
    - During the experiment, the process was observed daily by the researcher for both groups.
    - Normal classes, to learn main topic in each subject.
- **Last week**
    - Taking the post-test grades.
    - Send the Questionnaire by the researcher for the students to find out their feeling about PSA.
    - Teachers interview to find out their opinions about PSA.

Figure 25

Experimental Group using the PSA Mobile Application

CHAPTER IV

FINDINGS

4.1 Introduction

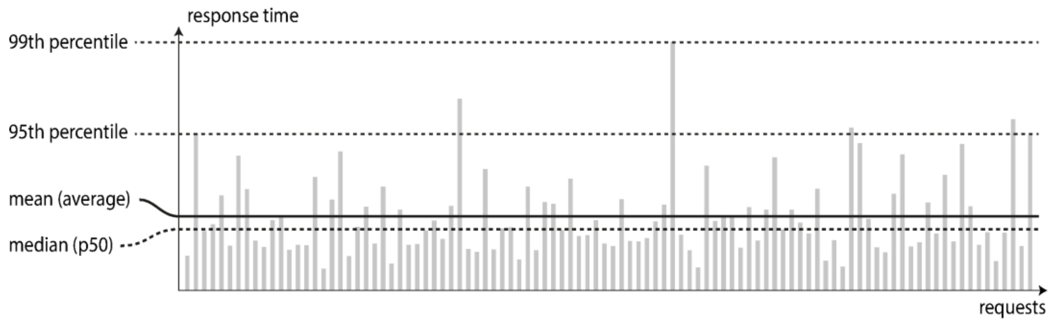
This section explains the results of this study based on the collected data. This includes the results of the performance of tested machine learning algorithms used, system operation and performance, and the accuracy of the performance scale in determining the students' choice, specialty, achievements, and learning outcomes. Also discusses the hypotheses of this study in terms of the results obtained.

4.1.2. *PSA Response Times*

Figure 26 displays the median and percentile response times for a request made in the PSA. The PSA function was used to gauge how well or poorly the outliers performed using the 95%, 99.0%, and 99.9% percentiles. These percentiles demonstrated that the PSA's response time thresholds were 99.9, 99, or 95% of requests were fulfilled faster than any other threshold. If the 95th percentile response time is 2 seconds, then 95 out of 100 requests will be fulfilled in under 2 seconds, while 9 out of 100 requests will require at least 2 seconds to be fulfilled. In Figure 10, each request is represented by a grey bar, and the height of the bar shows how long it took for the request to respond. The bulk of queries are handled rather rapidly, however occasionally some outliers need a much longer processing time. There are many other reasons why this could happen, including the loss of a network pack, a delay in the rubbish collection, a page mistake, and mechanical vibrations in the server store.

Figure 26

The PSA's Percentiles and Mean for Request Response Times

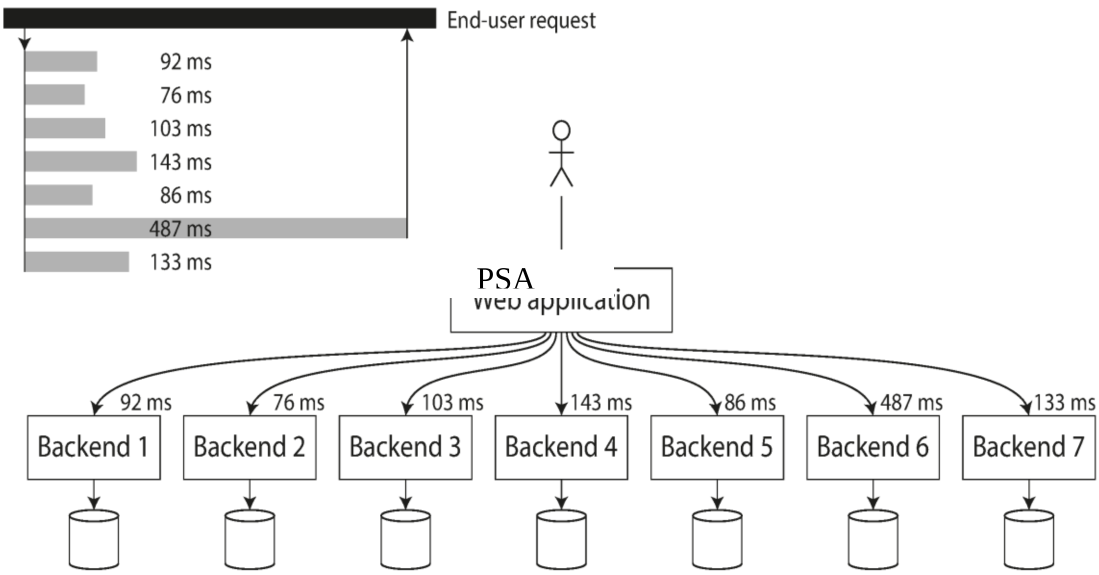


4.3. PSA Backend Calls

When one end-user request is fulfilled frequently by backend activity or response, high percentiles become more important. Even though the calls are made in parallel, this request still waits for the longest call to end. A single sluggish call can make all requests slow (Figure 27). If an end-user request requires numerous backend calls, the risk of receiving a sluggish call increases even when only a small percentage of backend calls are slow. Consequently, a higher percentage of user requests are unsuccessful.

Figure 27

PSA Backend Calls in End-User Request



4.2 Descriptive Statistics

4.2.1 Results of Tested Algorithms

Table 4 presents the results of tested algorithms according to each specialty, accuracy, and mean. These algorithms include Linear Regression (LR), Bayesian Ridge Regression, Support Vector Machine (SVM), K Nearest Neighbour (KNN), Decision Tree, Random Forest, Extra Trees, Gradient Boost, Stacking, and Artificial Neural Network (ANN). Three specialties (departments), namely scientific, literary, and industrial, were taken into consideration. The results showed that the extra trees had the highest accuracy (mean = 99.22) compared to stacking (mean = 98.73), gradient boost (mean = 97.47), random forest (mean = 97.57), and decision tree (mean = 97.17). The algorithms that showed less accuracies include Linear Regression (mean = 89.03), Bayesian Ridge Regression (mean = 88.90), Support Vector Machine (mean = 92.07), and K Nearest Neighbour (mean = 92.47).

Table 4

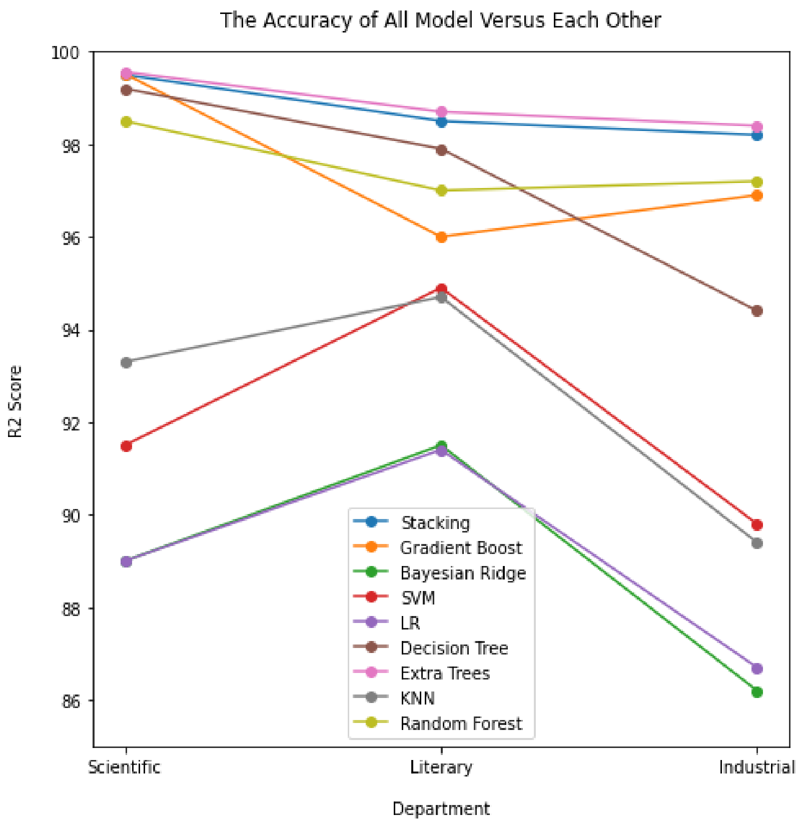
The Results of Tested Algorithms

	Model	Department	Accuracy	Mean
0	LR	Scientific	89	
1	LR	Literary	91.4	89.03
2	LR	Industrial	86.7	
3	Bayesian Ridge	Scientific	89	
4	Bayesian Ridge	Literary	91.5	88.90
5	Bayesian Ridge	Industrial	86.2	
6	SVM	Scientific	91.5	
7	SVM	Literary	94.9	92.07
8	SVM	Industrial	89.8	
9	KNN	Scientific	93.3	
10	KNN	Literary	94.7	92.47
11	KNN	Industrial	89.4	
12	Decision Tree	Scientific	99.2	
13	Decision Tree	Literary	97.9	97.17
14	Decision Tree	Industrial	94.4	
15	Random Forest	Scientific	98.5	
16	Random Forest	Literary	97	97.57
17	Random Forest	Industrial	97.2	
18	Extra Trees	Scientific	99.56	
19	Extra Trees	Literary	98.7	99.22
20	Extra Trees	Industrial	99.4	
21	Gradient Boost	Scientific	99.52	
22	Gradient Boost	Literary	96	97.47
23	Gradient Boost	Industrial	96.9	
24	Stacking	Scientific	99.5	
25	Stacking	Literary	98.5	98.73
26	Stacking	Industrial	98.2	
27	ANN	Scientific	68.9	
28	ANN	Literary	0	43.97
29	ANN	Industrial	63	

The outcome of all evaluated models' accuracy predictions is shown in Figure 28. The Extra Trees Regression algorithm, which had the greatest R2 score among the other nine tested algorithms for prediction accuracy, came out on top.

Figure 28

Accuracy of All Tested Models



A) Stacking

Table 5 displays the results of the tested stacking algorithm, while Table 6 shows the parameter description for this stacking algorithm. Overall, 20 folds of stacking were created using the create MultiFolds function. The results showed multiple three-level predictions that were above average. This further reduces the variance of the stacking predictions, which were computed to create repeated stacks (A, B, C, and D). The results of repeated stacks showed that stacking precision ranges from above average to high precision.

Depending on the value of k, the first three steps were carried out iteratively for the k-folds. The model was trained on the three folds (StackingA, StackingB, and StackingC) and makes predictions with precision using the recall control if k = 4 (the fourth fold is StackingD). The entire dataset's overall predictions were obtained by repeating this stage k times. The stacking appeared to perform better than StackingA, StackingB, and StackingD. These findings suggest that, while dimensionality cutback and the multi-class stacking model may play important roles, they may not be the primary causes of StackingC's enhancement.

Table 5

The Results of Tested Stacking Algorithm

Scientific		**Literacy**		**Industrial**		**Recall control**	
StackingA	**PrecisionA**	**StackingB**	**PrecisionB**	**StackingC**	**PrecisionC**	**StackingD**	**PrecisionD**
0.7460	74.6%	0.8820	88.2%	0.9950	99.5%	0.7380	73.8%
0.7760	77.6%	0.8400	84.0%	0.9960	99.6%	0.9920	99.2%
0.8840	88.4%	0.7940	79.4%	0.9970	99.7%	0.8880	88.8%
0.7480	74.8%	0.8760	87.6%	0.9970	99.7%	0.8570	85.7%
0.7660	76.6%	0.8000	80.0%	0.9990	99.9%	0.7040	70.4%
0.7390	73.9%	0.8000	80.0%	0.7700	77.0%	0.8210	82.1%
0.7970	79.7%	0.8000	80.0%	0.7270	72.7%	0.8040	80.4%
0.7450	74.5%	0.8000	80.0%	0.9000	90.0%	0.9970	99.7%
0.7530	75.3%	0.8000	80.0%	0.8650	86.5%	0.7070	70.7%
0.7040	70.4%	0.8070	80.7%	0.7280	72.8%	0.9940	99.4%
0.8040	80.4%	0.8070	80.7%	0.7660	76.6%	0.8170	81.7%
0.8090	80.9%	0.8450	84.5%	0.7060	70.6%	0.8620	86.2%
0.7510	75.1%	0.8000	80.0%	0.8330	83.3%	0.9130	91.3%
0.8380	83.8%	0.8090	80.9%	0.9960	99.6%	0.9060	90.6%
0.8640	86.4%	0.8110	81.1%	0.8420	84.2%	0.8010	80.1%
0.8980	89.8%	0.9270	92.7%	0.7190	71.9%	0.7940	79.4%
0.7380	73.8%	0.9910	99.1%	0.9310	93.1%	0.8320	83.2%
0.7560	75.6%	0.9930	99.3%	0.8310	83.1%	0.8190	81.9%
0.7560	75.6%	0.9950	99.5%	0.9270	92.7%	0.8280	82.8%

Table 6

Parameter Description for Stacking Algorithm in Table 5

Parameter	**Description**
StackingA	It is employed to categorize different forest type A, which is a multi-level classification algorithm.
PrecisionA	Precise predictor K* algorithm utilized for a single stacking base-classifier A
StackingB	It is employed to categorize different forest type B, which is a multi-level classification algorithm.

PrecisionB	Precise predictor K* algorithm utilized for a single stacking base-classifier B
StackingC	It is employed to categorize different forest type C, which is a multi-level classification algorithm.
PrecisionC	Precise predictor K* algorithm utilized for a single stacking base-classifier C
StackingD	It is employed to categorize different forest type D, which is a multi-level classification algorithm.
PrecisionD	Precise predictor K* algorithm utilized for a single stacking base-classifier D

B) Gradient Boost

Table 7 presents the results of the tested gradient boost algorithm, while Table 8 presents the parameter description for the gradient boost algorithm. The target variable is the test percent for each specialty, while the independent variables are scientific, literary, and industrial. A regression is utilized in this situation since the target variable is continuous. The root level (level 0) for the first estimator includes all of the data. The expected test percent speciality at this level is 61.58, which is within the same mean as the final prediction (64.04). The error is equivalent to the difference between the actual and expected test specialty percentages.

In order to compute the second estimator, the gradient boosting algorithm uses the first estimator's residues (test percent specialty) as root nodes (Table 7). For instance, specialties that have false literacy go into one child node, whereas specialties that have true literacy go into a different child node. In this case, scientific is false. Therefore, the first estimator's projected test percent speciality is 41.07. The value of literacy for the second predictor was computed, and we found that it is true. Therefore, -4.74 is the true literacy mean in the second estimator. This produces the first estimator's prediction precision of 25. Therefore, the algorithm's final prediction was 45.12 with a final residue of -1.41.

Table 7

The Results of Tested Gradient Boost Algorithm

Scientific	**Literacy**	**Industrial**	**Precision**	**Prediction from 1st estimator**	**Mean from 2nd estimator**	**Final Prediction**	**Final Residue**
False	True	True	25	41.07	-4.74	45.12	-1.41
True	True	False	30	41.13	-4.99	45.13	-1.66
True	True	True	32	42.67	-4.89	46.12	-1.58
False	True	True	35	46.68	-4.86	46.11	-1.51
True	True	True	37	48.80	-4.70	51.32	1.45
True	False	True	40	55.55	-4.82	58.11	-1.75
False	True	False	42	55.61	-4.80	58.12	-1.75
False	True	True	45	57.62	5.00	60.12	-1.75
True	True	False	47	57.66	4.71	60.13	1.75
False	True	True	50	57.74	4.99	60.13	1.75
True	True	False	55	57.83	6.82	60.32	1.75
True	False	True	60	57.86	6.86	60.12	-0.94
True	False	True	65	65.44	6.91	68.11	-1.63
True	True	True	67	65.21	6.91	68.12	-1.63
False	True	False	70	65.67	6.80	68.13	-1.63
False	True	True	72	75.68	6.79	77.11	-1.63
True	True	False	75	76.75	6.83	79.11	1.63
False	True	False	80	80.44	6.82	86.13	-0.89
True	True	True	85	87.31	6.83	88.12	1.53
True	False	True	94	94.89	6.83	95.13	1.53
Average			55.30	61.58	5.85	64.04	1.56

Table 8

Parameter Description for Gradient Boost Algorithm in Table 7

Parameter	**Description**
Test %	It is test percentage of each specialty
Prediction from 1st estimator	It is a base estimator. Decision Stump serves as the Gradient Boost algorithm's fixed base estimator.
Mean from 2nd estimator	It is mean estimate using n_estimator. It uses AdaBoost. In this, n estimator's default is 100.
Final Prediction	It is gradient boosting with two estimators.
Final Residue	It is root nodes determined by the first estimator.

C) Bayesian Ridge

Bayesian ridge regression, which computes a probabilistic model of the regression problem, is a key practical form of Bayesian regression. Table 9 displays the results of the tested Bayesian ridge algorithm, while Table 11 shows the parameter description for Bayesian ridge in Table 10. A distribution of potential model parameters was based on the data and the prior that produced the results of performing Bayesian ridge (the posterior). The results showed varying values for the posterior mean and posterior t-stat, with scientific having a higher mean followed by literacy and then industrial. However, all the specialties have comparable precision. Moreover, based on the results in Table 10, literacy produced better mean, P(incl) and P(incl|data) compared to scientific and industrial.

Table 10 also contains the posterior mean. There were no discrepancies between the results among specializations' Bayesian ridge based on the mean and 95% credible interval for both the lower and upper value. All coefficients are positive and significant, according to the Bayesian ridge regression.

Table 9

The Results of Tested Bayesian Ridge Algorithm

Variables	Posterior Mean	Posterior t-stat	Precision
Scientific_1	7.36	24.21	3.80
Scientific_2	1.36	1.83	2.00
Scientific_3	0.41	0.36	2.71
Literacy_1	5.45	1.83	3.99
Literacy_2	1.31	0.48	2.82
Literacy_3	1.73	1.2	2.86
Industrial_1	1.31	8.94	3.91
Industrial_2	0.37	0.79	2.79
Industrial_3	1.27	3.4	2.83

Table 10

The Results of Posterior Coefficients for the Tested Bayesian Ridge Algorithm

						95% Credible Interval	
Coefficients	**Mean**	**SD**	**P(incl)**	**P(incl|data)**	**$Bf_{inclusion}$**	**Lower**	**Upper**
Intercept	5.31	0.012	1	1	2.69	3.217	3.92
Scientific	0.18	0.045	0.04	0.03	1.55	0.72	0.86
Literacy	0.04	0.037	0.01	0.02	3.22	0.09	0.31
Industrial	0.29	0.062	0.05	0.04	6.71	0.18	0.08

Table 11

Parameter Description for Bayesian Ridge in Table 10

Parameter	**Description**
Posterior Mean	It expressed estimate of the regression coefficients. The weighted average is hence the posterior mean.
Posterior t-stat	It express the relationship between the predictor factors and the response.
P(incl)	The Pr(incl) represents the p-value associated with the value in the incl value column. If the P(incl) is less than a certain significance level (i.e., $\alpha<0.05$) then the predictor variable is said to have a statistically significant association with the response variable in the model.
P(incl|data)	Regression's p values used to measure if the associations in the data also existed in the entire datasets.
$BF_{Inclusion}$	The shift from before to posterior inclusion odds is represented by the inclusion Bayes factor, or "$BF_{Inclusion}$."

D) Support vector machine (SVM)

Table 12 shows the results of the tested SVM algorithm. The results showed that the accuracies and precisions of all variables (scientific, literacy, and industrial) were closely related and comparable for all parameters (i.e., linear, quadratic, MLP, RBF, RBF_Sigma, and Poly). The sensitivity appeared to be high across all the specialties, with a good recall. The Fp rate also produced a good response across specialties except for the linear parameter under literacy. The SVM specificity for linear, quadratic, MLP, RBF, RBF_Sigma, and poly appeared good, though the F measure values for literacy parameters were higher compared to scientific and industrial. Overall, the algorithms appeared excellent in terms of accuracy and precision.

Table 12

The Results of the Tested SVM Algorithm

			Scientific				
Kernels	Accuracy	Precision	Sensitivity	Recall	Fp rate	F measure	Specificity
Linear	0.74	1.36	0.36	0.62	1.75	2.13	0.41
Quadratic	0.73	1.06	0.27	0.49	1.75	2.61	0.44
MLP	0.68	1.27	0.32	0.52	1.94	2.61	0.63
RBF	0.76	1.23	0.41	0.27	1.63	2.23	0.31
RBF_Sigma	0.66	1.83	0.34	0.19	1.63	2.23	0.45
Poly	0.77	1.23	0.32	0.49	1.63	2.53	0.42
			Literacy				
Linear	0.78	1.11	0.42	0.69	2.05	3.13	0.35
Quadratic	0.76	1.20	0.39	0.74	1.8	2.61	0.39
MLP	0.71	1.25	0.47	0.76	1.77	2.88	0.42
RBF	0.66	1.73	0.41	0.68	1.69	3.23	0.37
RBF_Sigma	0.66	1.42	0.34	0.72	1.99	3.09	0.40
Poly	0.69	1.54	0.35	0.68	1.88	3.53	0.36
			Industrial				
Linear	0.80	1.62	0.39	0.55	1.88	1.75	0.48
Quadratic	0.81	1.71	0.44	0.60	1.96	1.93	0.47
MLP	0.84	1.33	0.46	0.63	1.78	1.87	0.66
RBF	0.86	1.22	0.41	0.70	1.85	1.91	0.39
RBF_Sigma	0.82	1.61	0.38	0.58	1.68	1.89	0.40
Poly	0.79	1.36	0.39	0.69	1.65	1.94	0.46

Table 13

Parameter Description for SVM in Table 12

Parameter	**Description**
Linear	It is used to describe SVM data linearity with dataset separated using a straight line.
Quadratic	It is used as a binary classifier.
MLP	MLPs is used for classification prediction with inputs arranged a class.
RBF	RBF is the default kernel used within the sklearn's SVM classification algorithm
RBF_Sigma	It is used as a fine-tune for bandwidth kernel function. The amount of nonlinearity incorporated into the model is controlled by this parameter. The decision border is particularly non-linear when the sigma is not big.
Poly	It illustrates how closely vectors (training data) resemble each other in a feature space over polynomials of the initial inputs, enabling the learning for non-linear model.

E) Linear Regression

Table 14 displays the results of the tested linear regression model. Based on the results, the tested models produced good R^2 of 96.12, 92.60, and 90.30 for scientific, literacy, and industrial, respectively. The scientific had a higher R^2 compared to literacy and industrial. All the models produced significant p-values. Similarly, the results of the regression coefficient from simple linear models produced a higher R^2 with better precision for all models.

Table 14

The Results of the Tested Linear Regression Algorithm

Model	Explanatory variable	F	df	p-value	R^2 adjusted
1	Scientific	264.2	1, 25	<0.001	96.12
2	Literacy	141.5	1, 25	<0.001	92.60
3	Industrial	87.6	1, 25	0.014	90.30

	Regression coefficient from Simple Linear models				Regression coefficient from Multiple Linear models			
Explanatory variable	Estimate	95% Cl	p-value	r^2	Estimate	95% Cl	p-value	Precision
Scientific	1.34	0.75, 0.94	0.011	0.9887	2.89	093, 2.34	0.024	1.99
Literacy	14.06	8.52, 16.45	<0.001	0.9798	5.73	12.66, 7.98	<0.001	1.41
Industrial	28.33	6.99, 47.82	<0.001	0.9875	1.87	4.42, 9.11	0.001	1.66

Cl=Conference Interval,

F) Decision Tree

Figure 29 displays the results of the decision tree test. The top table lists the features of a made-up data set with two variables, 807 data points, and one class. The parents of the class are both variables A and B. For instance, the table displays 101 and 105 data items with the variables A and class set to 0 and 1, respectively. The class value with the maximum probability was chosen by the algorithm for each branch. If multiple values from a class have the highest probability, the algorithm chose one at random. If, for instance, variable scientific (var_sci) was 0 and variable literacy (var_Lit) was also 0, the probability of having a class of either of the 0 was 0.50, then either. 0 was picked by the algorithm. Finally, values from neighboring branches that share variables, parents, and classes are concatenated, which was used by the decision tree regressor.

Figure 29

Decision Tree Tables

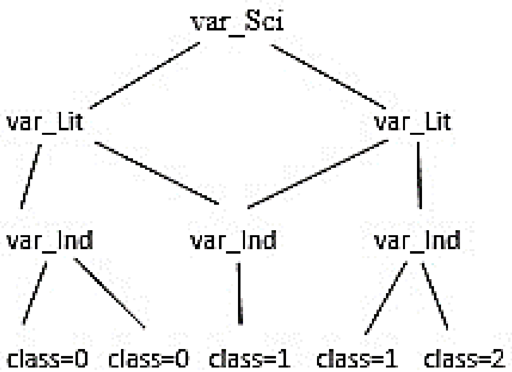
Class=0					Class=1				
var_A=0	var_A=1	var_B=0	var_B=1	var_B=2	var_A=0	var_A=1	var_B=0	var_B=1	var_B=2
101	105	72	60	75	110	85	80	65	54

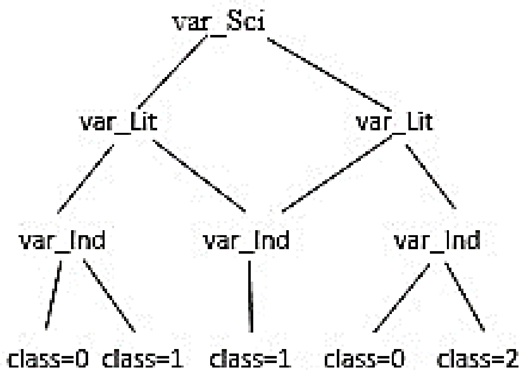
			Class		
Sci	Lit	Ind	0	1	2
0	0	0	0.50	0.51	0.52
0	1	1	0.46	0.50	0.52
1	1	0	0.47	0.49	0.51
1	0	1	0.48	0.50	0.45
1	0	1	0.50	0.50	0.50
2	2	2	0.53	0.47	0.47

Conditional Table

			Class		
Sci	Lit	Ind	0	1	2
(var_Sci)0	(var_Lit)0	(var_Ind)0	0.53	0.58	0.51
(var_Sci)0	(var_Lit)1	(var_Ind)1	0.49	0.49	0.52
(var_Sci)1	(var_Lit)1	(var_Ind)0	0.47	0.52	0.49
(var_Sci)1	(var_Lit)0	(var_Ind)1	0.53	0.50	5.51
(var_Sci)1	(var_Lit)0	(var_Ind)1	0.55	0.53	0.46
(var_Sci)2	(var_Lit)2	(var_Ind)2	0.53	0.53	0.47

Decision Tree Table





Decision Tree table

Sci = Scientific

Lit = Literacy

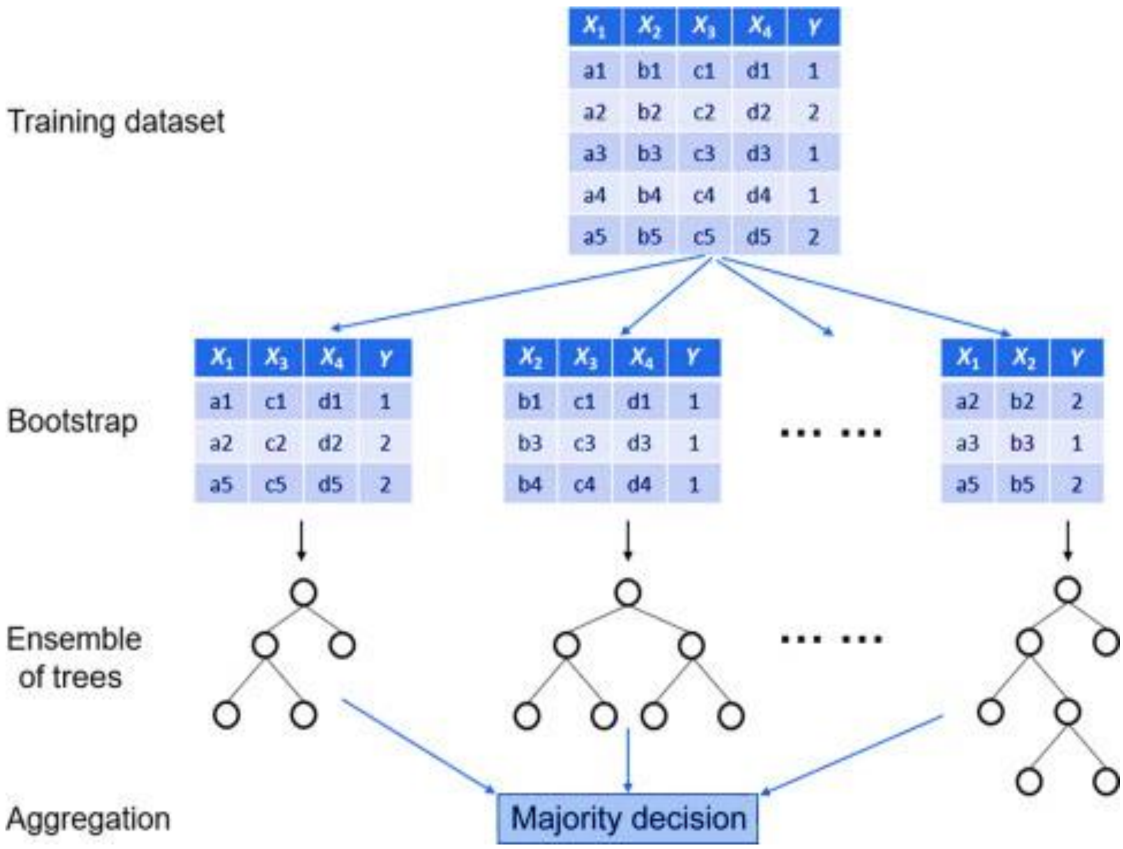
Ind = Industrial

G) Extra Trees

Figure 30 presents the results of the extra trees. The results showed that the extra tree classifier takes a random chunk of data without replacing it. As a result, nodes were separated according to random separations rather than the optimal separations. The randomness in the extra tree therefore derives from the data's random divisions rather than from the bootstrap cluster. Extra Trees is also a lot faster because it randomly splits data rather than searching for the best split at each node. Hence, it was observed that there was barely any difference among the datasets in this algorithm's outcomes, based on the table findings in Figure 30. The difference was negligible based on these results, which showed the better performance of the extra trees algorithm. The variance of the extra trees was lower, as already indicated, but the difference was once again essentially negligible.

Figure 30

Extra Trees Results



	Extra Trees	
	F-Score	Time(s)
Dataset 1	0.773 ± 0.009	0.213
Dataset 2	0.775 ± 0.027	0.206
Dataset 3	0.807 ± 0.015	0.206
Dataset 4	0.755 ± 0.016	0.238
Dataset 5	0.758 ± 0.017	0.182

H) k-nearest Neighbor Algorithm

The results of the tested k-nearest neighbor (KNN) algorithm are shown in Table 15. There were no significant differences between the values among the variables (scientific, literacy, and industrial). However, fuzzy KNN performed better than conventional KNN, with excellent precision across the models. Since the fuzzy was a c-means clustering algorithm with high precision, it was chosen over the conventional KNN.

Table 15

The Results of the Tested k-nearest Neighbor Algorithm

Variable	**Conventional KNN**	**Fuzzy KNN**	**Precision**
Scientific_k1	0.71	0.85	0.01
Scientific_k2	0.72	0.86	0.01
Scientific_k3	0.71	0.85	0.03
Literacy_k1	0.74	0.85	0.01
Literacy_k2	0.71	0.85	0.03
Literacy_k3	0.74	0.86	0.02
Industrial_k1	0.71	0.85	0.01
Industrial_k2	0.72	0.85	0.02
Industrial_k3	0.71	0.84	0.01

I) Random Forest

The results of the tested random forest algorithm are shown in Table 16, while Table 17 shows parameter descriptions for random forest in Table 16. The findings revealed that accuracy datasets produced the greatest results. The ideal situation was when they practically had TNR. Many true negatives were present in other data with low accuracy levels. The majority of the replicated data for each specialty had higher accuracy levels than could be shown (99.90 percent). The data with the highest imbalance levels showed the highest reliability rates, at 66.66 percent, whereas other data showed reliability levels of 33.33 percent.

Table 16

The Results of the Tested Random Forest Algorithm

Dataset	Accuracy %	Reliability	Kappa	True Negative Rate (TNR)	Negative Predictive value
Scientific_KDD1	99.90	33.33	0	0	0
Scientific_KDD2	99.90	33.33	0	0	0
Scientific_KDD3	99.90	33.33	0	0	0
Scientific_KDD4	99.90	66.67	0.0033	0.9967	0
Scientific_KDD5	100.00	33.33	0	0	0
Scientific_KDD6	99.90	33.33	0	0	0.01278
Literacy_KDD1	99.90	33.33	0	0	0
Literacy_KDD2	100.00	66.63	0.0037	-0.9963	0
Literacy_KDD3	99.90	33.33	0	0	0
Literacy_KDD4	99.90	33.33	0	0	0
Literacy_KDD5	100.00	33.33	0	0	0
Literacy_KDD6	99.90	65.99	0.0051	0.9949	0
Industrial_KDD1	99.90	33.33	0	0	0.00163
Industrial_KDD2	99.90	33.33	0	0	0
Industrial_KDD3	100.00	33.33	0	0	0
Industrial_KDD4	99.90	33.33	0	0	0
Industrial_KDD5	99.90	33.33	0	0	0
Industrial_KDD6	99.90	66.88	0.0042	0.9958	0

Table 17

Parameter Description for Random Forest in Table 16

Parameter	**Description**
Kappa	It functions as a statistic for contrasting observed accuracy and presumed accuracy (i.e., random likelihood).
True Negative Rate (TNR)	It is likelihood that a real negative will result in a test result of zero, which also known as specificity.
Negative Predictive value	It presents the percentage of data with a negative test result.

4.2.3 Descriptive statistics of groups for Pre-Test and Post-Test

Table 18 presents descriptive statistics of groups (pre-test and post-test). The mean scores (55.4) of the experimental group were greater than those of the control group (52.7) in the post-test, while their scores were similar in the pre-test for the experimental (51.4) and control group (52.2), respectively, for the math course. The same goes for the results of other courses. The empirical group achieved an average score higher than the average score of the control group in all post-tests and the exact opposite in the pre-test. In all courses, the passing score was 50 and the maximum score was 100.

Table 18

Groups' descriptive statistics for the pre- and post-test

		Group →	**Empirical**		**Control**	
Specialty	Test	Subjects	Mean	Std. Deviation	Mean	Std. Deviation
scientific	Post	Math	55.4	4.949	52.7	6.255
	Pre	Math	51.4	5.168	52.2	6.161
	Post	Physics	54.4	4.789	50.8	5.350
	Pre	Physics	50.6	5.641	50.9	5.195
literary	Post	Arabic	56.2	5.160	51.6	5.602
	Pre	Arabic	51.3	6.019	51.9	4.818
	Post	Geography	57.1	5.763	52.8	5.884
	Pre	Geography	51.8	4.733	52.4	5.400
Industrial	Post	Drawing	52.3	4.218	49.4	5.797
	Pre	Drawing	49.5	5.701	50	4.899
	Post	Engineering	55	5.011	50.9	4.175
	Pre	Engineering	50.7	4.138	51.4	5.400

4.3 Inference Statistics

Pearson correlation was utilized to analyze the results of the pre-test and the post-test. Table 19 displays the finding of Pearson correlation between the pre- and post-test. In the scientific specialty, the correlation between pre-math and post-math was estimated to be r = 0.871, with a significant 0.00. The correlation between the pre-physics and post-physics was estimated to be r = 0.867, with a significant 0.00. The literary specialty Pearson correlation between pre-Geography and post-Geography was estimated to be r = 0.869, with a significant 0.00. However, the Pearson correlation between pre-Arabic and post-Arabic was estimated to be r = 0.15, but not significant 0.527 > 0.05. Finally, the literary specialty Pearson correlation between the pre-industrial and post-industrial drawings was estimated to be r = 0.835, with a significant 0.00, as shown in Table 6. However, the correlation between pre-engineering and post-engineering was not significant because P = 0.208 > 0.05.

Table 19

Pearson Correlation Between the Pre- and Post-Test

Specialty	First Test	Second Test	Pearson Correlation	Sig. (2-tailed)
Scientific	Pre-Math	Post-Math	0.871	0.000
	Pre-Physics	Post-Physics	0.867	0.000
Literary	Pre-Arabic	Post-Arabic	0.150	0.527
	Pre-Geography	Post-Geography	0.869	0.000
Industry	Pre-Drawing	Post-Drawing	0.835	0.000
	Pre-Engineering	Post-Engineering	0.294	0.208

Table 20 displays the findings of the paired sample test. The findings revealed significant differences between post- and pre-math (p = 0.002) and Geography (p = 0.000). The rest of the subjects were statistically significant. Similarly, the results showed that Geography had the highest significant t (4.205) followed by Mathematics (t = 3.537), while Arabic (t = 1.425) and Engineering (1.481).

Table 20

Paired Samples Test

Course	Mean Difference	Std. Deviation	t	df	Sig. (2-tailed)
Post-Math – Pre-Math	2.25	0.845	3.537	19	0.002
Post-Physics – Pre-Physics	1.85	0.720	3.042	19	0.007
Post-Arabic – Pre-Arabic	2.3	0.219	1.425	19	0.170
Post-Geography – Pre-Geography	2.85	0.031	4.205	19	0.000
Post-Drawing – Pre-Drawing	1.1	0.972	1.655	19	0.114
Post-Engineering – Pre-Engineering	1.9	0.739	1.481	19	0.155

An independent t-test according to Pallant (2020) was used. Table 21 showed the findings of independent sample t-test for the groups in the different courses. The results revealed that all of Levene's tests were not significant across the specializations, including scientific, literary, and industrial. However, the results of the t-test were statistical significance for scientific, literary, and industrial. for each course except for the Arabic (0.109 > 0.05), which was not significant. This study observed that there was no mean difference within groups for all specializations. However, there was a mean difference across groups noted for Arabic (5.2 for both assumed and not assumed) and drawing (4.9 for both assumed and not assumed). This indicates that the performance scale application effectively improves the scores of the students in all studied courses except for Arabic.

Table 21

Independent Sample T-test for the Groups in the Different Courses

Specialty	Course	Equal variances	Levene's Test F	Levene's Test Sig.	t-test for Equality of Means t	df	Sig. (2-tailed)	Mean Difference
scientific	Math	Assumed	0.75	0.398	3.452	18	0.003	3.5
		Not assumed			3.452	15.224	0.003	3.5
	Physics	Assumed	3.394	0.082	4.607	18	0.000	3.9
		Not assumed			4.607	15.6	0.000	3.9
literary	Arabic	Assumed	0.857	0.367	1.687	18	0.109	5.2
		Not assumed			1.687	16.657	0.110	5.2
	Geography	Assumed	0.729	0.404	3.075	18	0.007	3.4
		Not assumed			3.075	15.786	0.007	3.4
Industrial	Drawing	Assumed	0.008	0.931	6.297	18	0.000	4.9
		Not assumed			6.297	17.447	0.000	4.9
	Engineering	Assumed	0.75	0.398	3.452	18	0.003	3.5
		Not assumed			3.452	15.224	0.003	3.5

The results of ANCOVA analysis-based assessments of type III between-subjects effects are shown in Table 22. The datasets were checked statistically for fulfilling normality, linearity, reliable measurement of covariates, homogeneity of variances, and a homogeneous regression slope. For this reason, we excluded Arabic and engineering from this analysis because they do not fulfil the condition. Math (p =.004), Physics (p =.004), Geography (p =.007), and Drawing (p =.007) had statistically significant differences in adjusted means. The calculated R^2 and adjusted R^2 were found across all specializations of subjects to be close to each other, which implies that there was no big disparity among results. This performance scale prediction was quite accurate for assessing the student's performance. Therefore, the ANCOVA results showed that the experimental group outperformed the control group in these specializations through programmed learning and not by chance.

Table 22

Tests of Between-Subjects Effects: Type III

Specialty	Source	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
	Dependent Variable: **Post -Math**						
	Model	520.919	2	260.46	50.298	0.000	0.855
	Intercept	9.89	1	9.89	1.91	0.185	0.101
	Pre-Math	484.469	1	484.469	93.558	0.000	0.846
	Method	58.499	1	58.499	11.297	0.004	0.399
	Error	88.031	17	5.178			
	Total	59037	20	$\mathbf{R^2}$ = 0.855 (Adjusted $\mathbf{R^2}$ = 0.838)			
scientific	Dependent Variable: **Post -Physics**						
	Model	472.258	2	236.129	70.994	0.000	0.893
	Intercept	13.256	1	13.256	3.986	0.062	0.19
	Pre-Physics	407.458	1	407.458	122.506	0.000	0.878
	Method	74.559	1	74.559	22.417	0.000	0.569
	Error	56.542	17	3.326			
	Total	55864	20	$\mathbf{R^2}$ =0.893 (Adjusted $\mathbf{R^2}$ = 0.880)			
	Dependent Variable: **Post-Geography**						
	Model	653.010	2	326.505	111.146	0.000	0.929
	Intercept	0.908	1	0.908	0.309	0.585	0.018
literary	Pre-Geography	560.56	1	560.56	190.821	0.000	0.918
	Method	122.507	1	122.507	41.703	0.000	0.71
	Error	49.94	17	2.938			
	Total	61093	20	$\mathbf{R^2}$ =0.929 (Adjusted $\mathbf{R^2}$ = 0.921)			
	Dependent Variable: **Post-Drawing.**						
	Model	406.515	2	203.257	35.246	0	0.806
	Intercept	15.503	1	15.503	2.688	0.119	0.137
Industrial	Pre-Drawing	364.465	1	364.465	63.201	0	0.788
	Method	55.086	1	55.086	9.552	0.007	0.36
	Error	98.035	17	5.767			
	Total	5219	20	$\mathbf{R^2}$ = 0.806 (Adjusted $\mathbf{R^2}$ = 0.783)			

The effect size (η^2) was used to answer the question 5 to determine practical performance and significance. Table 23 presents the results of one-way repeated measures (ANOVA). This analysis was conducted to evaluate the null hypothesis: there is no significant change in mean scores due to the use of the proposed application. The results indicated a significant method effect for the Math course, Wilks' Lambda = 0.477, F (1, 18) = 19.703, P < 0.01, η^2 = 0. 523. Based on these results, the null hypothesis was rejected for the Math course. Likewise, there was a significant method effect for the Physics course, Wilks' Lambda = 0.485, F (1, 18) =19.102, P<0.01, η^2 = 0.515. Based on these results, the null hypothesis was rejected for the Physics course.

Table 23

One-way Repeated Measure (ANOVA)

Effect	Multivariate Tests	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Math	Pillai's Trace	0.523	19.703	1	18	0.00	0.523
	Wilks' Lambda	0.477	19.703	1	18	0.00	0.523
	Hotelling's Trace	1.095	19.703	1	18	0.00	0.523
	Roy's Largest Root	1.095	19.703	1	18	0.00	0.523
Math * Method	Pillai's Trace	0.398	11.919	1	18	0.003	0.398
	Wilks' Lambda	0.602	11.919	1	18	0.003	0.398
	Hotelling's Trace	0.662	11.919	1	18	0.003	0.398
	Roy's Largest Root	0.662	11.919	1	18	0.003	0.398
Physics	Pillai's Trace	0.515	19.102	1	18	0.00	0.515
	Wilks' Lambda	0.485	19.102	1	18	0.00	0.515
	Hotelling's Trace	1.061	19.102	1	18	0.00	0.515
	Roy's Largest Root	1.061	19.102	1	18	0.00	0.515
Physics* Method	Pillai's Trace	0.541	21.223	1	18	0.00	0.541
	Wilks' Lambda	0.459	21.223	1	18	0.00	0.541
	Hotelling's Trace	1.179	21.223	1	18	0.00	0.541
	Roy's Largest Root	1.179	21.223	1	18	0.00	0.541

Table 24 presents the results of ANOVA for Geography and Drawing. The results showed a significant method effect for Geography course, Wilks' Lambda = 0.251, F (1, 18) = 53.653, P < 0.01, η^2 = 0.749. Thus, the null hypothesis was rejected because method effect for the Geography course. Similarly, there was no significant method effect for drawing course, Wilks' Lambda = 0.82, F (1, 18) = 3.960, P < 0.1. Thus, the null hypothesis was accepted because there was no method effect for the

Drawing course. This further support the fact that the application does not work for the industrial specialty Drawing subject.

Table 24

One-way Repeated Measure (ANOVA)

Effect	Multivariate Tests	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Geography	Pillai's Trace	0.749	53.653	1	18	0.00	0.749
	Wilks' Lambda	0.251	53.653	1	18	0.00	0.749
	Hotelling's Trace	2.981	53.653	1	18	0.00	0.749
	Roy's Largest Root	2.981	53.653	1	18	0.00	0.749
Geography * Method	Pillai's Trace	0.688	39.65	1	18	0.00	0.688
	Wilks' Lambda	0.312	39.65	1	18	0.00	0.688
	Hotelling's Trace	2.203	39.65	1	18	0.00	0.688
	Roy's Largest Root	2.203	39.65	1	18	0.00	0.688
Drawing	Pillai's Trace	0.18	3.960	1	18	0.062	0.18
	Wilks' Lambda	0.82	3.960	1	18	0.062	0.18
	Hotelling's Trace	0.22	3.960	1	18	0.062	0.18
	Roy's Largest Root	0.22	3.960	1	18	0.062	0.18
Drawing * Method	Pillai's Trace	0.344	9.458	1	18	0.007	0.344
	Wilks' Lambda	0.656	9.458	1	18	0.007	0.344
	Hotelling's Trace	0.525	9.458	1	18	0.007	0.344
	Roy's Largest Root	0.525	9.458	1	18	0.007	0.344

4.4 Teacher Opinions towards Performance Scale Application

The results of the teachers’ responses were presented based on the statistics displayed in (Figure 31) The findings were discussed according to the supporters, opponents, and advisers as follows:

a) Supporters

The supporters’ responses were summarized into nine positive points: performance scale application (PSA) improved students’ performance and helped them choose the suitable department for their abilities; it motivated students to learn; it made the educational process easier and more effective; and it increased student interaction. Some of the teachers were of the opinion that PSA increased the student's efficiency (point 1): “During using the PSA, the student's efficiency and their academic achievement increased, especially when they learned grammar and vocabulary” (Teacher 1).

Some teachers indicated that PSA made the teaching process easier (point 2): “When I used the PSA in the geography topics, I noticed it made the teaching process easier and increased students' interaction” (Teacher 3). The "Performance Scale depends on ease of use that attracts the student towards the study desire, which we sometimes suffer the opposite in the classroom” (Teacher 4). From the point of view that studying through the application is easy and encourages the student to self-learn, the students are always waiting for the tasks presented to them, and this is evident through their active participation (Teacher 6). “The PSA was easy to use and communicate with, which helped us achieve a big and noticeable improvement in student performance” (Teacher 2).

A teacher also indicated that the PSA guides the students to the appropriate educational level (point 3): “From my opinion, I think that the idea of helping the student put himself in a correct picture of his educational level and the best suitable section for him would enable him to pass the stage with ease and effectiveness, and it is better than leaving them without guidance” (Teacher 15). Another teacher also responded that PSA always made the educational process continuous (point 4): “In my view, the communication made by Performance Scale made the educational process continuous and wonderful at all times” (Teacher 11).

Moreover, yet another teacher revealed that increasing student motivation for self-learning (point 5): “Supporting the student and providing him with enrichment curriculums according to his academic level in the physics subject was one of the important points made by the Performance Scale to increase student motivation for self-learning” (Teacher 12). Teacher 7 expressed that the student has preferred distance learning using PSA (point 6): “I believe that Performance Scale motivated the student to learn in general and distance learning in particular, where the student has become more effective in any times, also in many cases, the student has preferred distance learning over being in the classroom” (Teacher 7).

According to several teachers, PSA makes the student more active in participation (point 7): “I think that the Performance Scale application in Industrial Engineering topics helped to achieve one of its desired goals, which is to make the student more active in participation and motivate him to pass this stage” (Teacher 9). “In my opinion, the Performance Scale contributed to effective communication between me and the students at any time and the activation of a mechanism for expressing opinions and discussion that improves the educational process” (Teacher 8). “When sharing a video or an assignment through the Performance Scale application in the industrial engineering topics, we noticed that student participation and interaction is much greater than inside the classroom, especially for shy and quiet students” (Teacher 10).

Other teachers showed that the PSA increases the student's self-confidence (point 8): “From my view, the ease of use and communication made possible by the Performance Scale made the student more involved and effective, reduced his weaknesses like a lack of self-confidence, and brought us to effectiveness in the class” (Teacher 14). For me, as a math teacher for ten years, dividing students into categories contributed significantly to increasing the student's self-confidence and the focus became greater than before. We succeeded in improving the weaknesses of students, motivating them to express their opinions and share them with their teacher and colleagues, improving their performance, and providing sustainable learning (Teacher 13). Lastly, another teacher revealed that PSA reduces the level of fear and anxiety in the student (point 9): “I believe that the student's level of fear and anxiety has become almost non-existent because the student found himself in a group compatible with his educational level” (Teacher 5).

These nine teachers’ opinions support the results of this study in the previous section.

b) Opponents

The supporters’ responses were summarized into nine negative points: time constraints, the intensity of the curriculum, and some problems with Internet communication. Some teachers showed that PSA did not consider time constraints: "From my teaching experience, I think that the PSA did not take into account the student's time constraint, which is divided between the student's private life and studies, as there is not enough time after school to learn on a daily basis through the application" (Teacher 15). Another teacher responded that the PSA did not consider the intensity of the curriculum and problems with Internet communication: "From my point of view, the researcher did not take into account the intensity of the curriculum in the secondary stage and the difficulty of this stage for the student and the parents" (Teacher 16). Also, another teacher mentioned the problems with internet communication: "When I looking at the group of students that I teach, where some students suffer from poverty and do not have access to the Internet, and also some students work at this age to help their families in the life and do not have time and we cannot ignore the psychological pressure that students face to finish this stage” (Teacher 17).

There are advantages and disadvantages to all teaching strategies, but the disadvantages mentioned here can be overcome with time, especially with the rapid technological development in the world of communications and the provision of new inexpensive devices, as well as the extension of Internet networks to cover most of the world at the lowest cost.

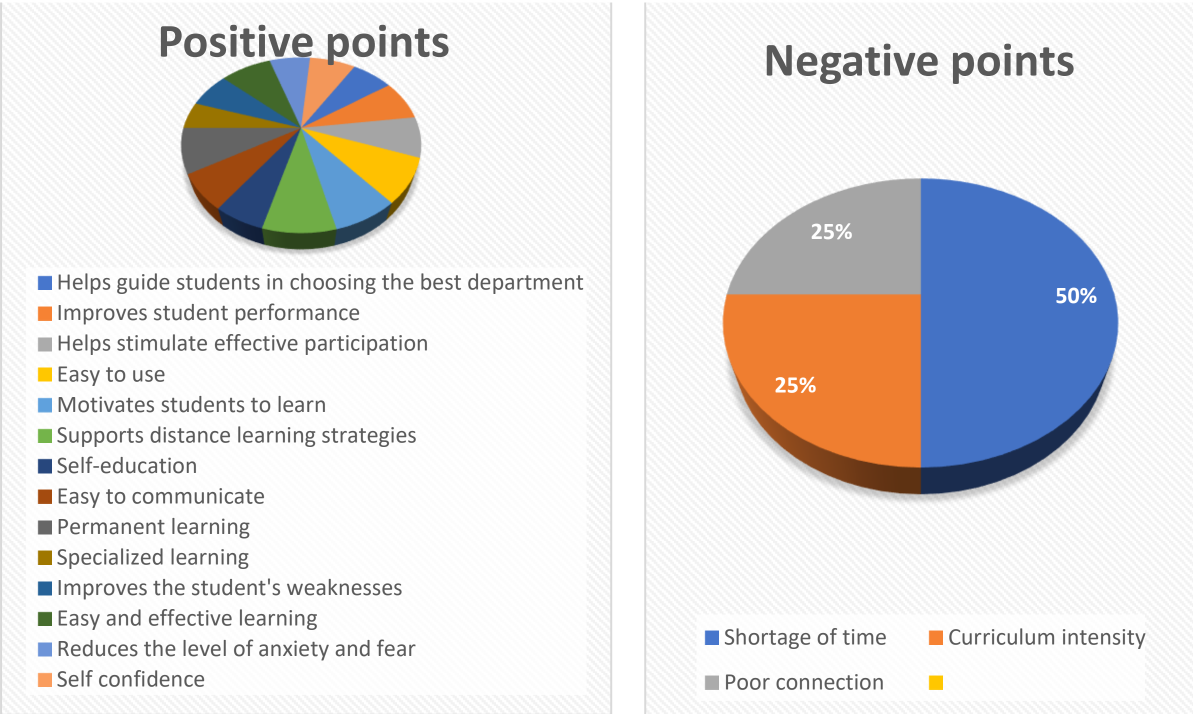
c) Advisers

Some teachers also recommended the implementation of PSA at various levels of study and activating distance teaching directly. A teacher has shown that combining PSA with traditional teaching yields positive results: "I believe that the PSA should be combined with traditional teaching to save students time and effort." (Teacher 9). Another teacher indicated that the PSA should be implemented at all educational levels:

"In my opinion, implementing the Performance Scale application in all educational levels that precede the secondary stage will give better results for future generations." (Teacher 1). Therefore, this study considers this advice to be important and worthy of attention and consideration.

Figure 31

Opinions of Teacher about Performance Scale Application



4.5 Results of Experimental Study

Table 3 displays descriptive data from the entire sample (N = 75). The intervention's impact on students' IME, IMC, IMSC, and SOL is evaluated using statements with multiple-choice answers on a five-point Likert scale (1 = completely disagree, 5 = completely agree). The author utilized these measurements to investigate how the PSA affected the motivation of students to improve their academic performance in Jordan. A total of 75 students participated in this study in all specializations (scientific, literary, and industry) from high school.

As stated in Table 25, a total of 75 students from all specialities participated in this study, with 49% coming from the scientific field, 41% coming from the literary field, and 10% coming from the industrial sector.

Table 25

Response Rate of Participants by Specialty

Specialty	Frequency	Percent
scientific	37	49%
literary	31	41%
industry	7	10%
Total	75	100

The findings of KMO and Bartlett's test are shown in Table 26. The KMO measure of sample adequacy for the results was excellent at 0.873. The study was able to move further with factor analysis because the significant value of Bartlett's test of sphericity was good and sufficient.

Table 26

The Results Of Kmo And Bartlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.873
Bartlett's Test of Sphericity	Approx. Chi-Square	2187.205
	Df	351
	Sig.	0.000

The variance inflation factor (VIF), tolerance (>0.1), and Pearson product-moment correlation coefficients (0.90) of the data were all extremely good (See Tables 27 and 28).

Table 27

Results Of Correlation Analysis (Dependent Variable: SOL)

Item	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	-0.455	0.286		-1.589	0.122		
IME	0.365	0.107	0.35	3.417	0.002	0.311	3.215
IMC	0.318	0.134	0.27	2.374	0.024	0.253	3.957
IMSC	0.416	0.139	0.388	2.998	0.005	0.195	5.139

Table 28

Correlation Analysis by AMOS

Variable	min	max	skew	c.r.	kurtosis	c.r.
IME	2.2	5	-0.597	-2.109	0.215	0.38
IMC	2	5	-0.422	-1.491	0.211	0.373
IMSC	2.14	5	-0.595	-2.103	0.381	0.674
SOL	1.8	5	-0.647	-2.287	1.175	2.077
Multivariate					19.294	12.059

4.5.1 Cronbach's reliability

The findings of the reliability analysis are shown in Table 29. Every variable satisfies the requirements of the Cronbach alpha coefficient, which ranges from 0.977 to 0.978. All 27 of the item-to-total correlation values were determined to have high values that were above the permitted threshold. The factor loadings of the 27 items were also enhanced (>0.3), and the coefficient alpha values also showed increased dependability. The results of this investigation provide strong evidence for the instruments' high levels of accuracy, applicability, and acceptability. The data were then analyzed using inferential statistics to assess the research hypotheses.

Table 29

Results Of Reliability Analysis

Cod	Item	Item-Total Correlation	Cronbach's Alpha if Item Deleted	Factor Loadings	Cronbach's Alpha Analysis
Improving evaluation					0.909
	X1	0.680	0.978	0.685	
	X2	0.810	0.977	0.869	
IME	X3	0.656	0.978	0.743	
	X4	0.660	0.978	0.789	
	X5	0.704	0.978	0.852	
Improving communication					0.929
	A1	0.693	0.978	0.797	
	A2	0.721	0.978	0.837	
	A3	0.744	0.978	0.824	
IMC	A4	0.733	0.978	0.795	
	A5	0.782	0.978	0.879	
	A6	0.737	0.978	0.780	
	A7	0.822	0.977	0.837	
Improving scientific content					0.941
	B1	0.778	0.978	0.858	
	B2	0.771	0.978	0.843	
IMSC	B3	0.677	0.978	0.791	
	B4	0.688	0.978	0.688	
	B5	0.768	0.978	0.826	
Satisfaction of learning					0.954
	C1	0.727	0.978	0.813	
	C2	0.765	0.978	0.840	
	C3	0.738	0.978	0.854	
	C4	0.837	0.977	0.798	
	C5	0.833	0.977	0.862	
SOL	C6	0.884	0.977	0.834	
	C7	0.731	0.978	0.851	
	C8	0.753	0.978	0.774	
	C9	0.725	0.978	0.866	
	C10	0.732	0.978	0.764	
Total					0.978

Partial least squares SmartPLS 3 software was employed as a crucial statistical technique to assess the results based on CFA for the remaining hypotheses. The outcomes of the dependent variables and evaluation of the effects depicted in Figure 6 were displayed in Table 30. Acceptable results were obtained for the constructs, items, and CFA generate factor loading of 0.5 or higher (Hilkenmeier et al. 2020). The average variance extracted (AVE) of the construct's elements was above 0.50, which is a highly acceptable number, demonstrating convergence validity (Anis et al., 2020; Shrestha, 2021). According to our investigation, the AVE ratings are 0.63 (IME), 0.68 (IMC), 0.65 (IMSC), and 0.68. (SOL). While Cronbach's alpha values ranged from 0.847 to 0.948, the appropriateness indices for CR and CA, which satisfy all requirements, corroborate the model specificity. All were above the permissible threshold of 0.70.

Table 30

Overall Validity And Reliability

	Fornell-Larcker Criterion				Construct Reliability and Validity			
Item	IME	IMC	IMSC	SOL	CA	RA	CR	AVE
IME	0.791				0.847	0.853	0.892	0.625
IMC	0.849	0.822			0.920	0.921	0.936	0.675
IMSC	0.830	0.846	0.804		0.861	0.865	0.901	0.646
SOL	0.834	0.836	0.897	0.826	0.948	0.955	0.956	0.683

4.5.2 The Hypothesis of the Study

H1: There are statistically differences between the sample members' use of the performance measure application by a specialty.

H2: The use of the Performance scale application in education improves communication positively.

H3: The use of the PSA in education improves scientific content, positively.

H4: The use of the PSA in education improves Satisfaction of learning, positively.

H5: The use of the PSA in education improves personalized learning, positively.

H6: The use of the PSA in education improves distance learning, positively.

H7: The use of the PSA in education improves Satisfaction of learning, positively.

H8: The use of the PSA in education improves self-learning, positively.

H9: The use of the PSA in education improves specialty learning and is positively and substantially.

4.5.2.1 Testing of Hypotheses

A technique for approximating a sole regression model with numerous effect variables is known as multivariate regression. A model is referred to be multivariate multiple regression when it has several predictor variables. We used a multivariate analysis of covariance to examine the impact of the speciality on the dependent variables (MANCOVA; Table 31). This approach was used to take into account covariance between the dependent variables. The variance homogeneity of the variables included in the model was not disproved, according to the Box M and Levene's tests (Table 32), which show that this assumption was correct ($p > 0.05$). According to MANCOVA's findings ($p = 0.005$), the specialization had an effect on the dependent variables. Due to this, we choose the alternative hypothesis (H1), which states that there are statistically significant variations in the way each specialization is represented in the sample of students.

Table 31

Testing the Overall Effect through Multivariate Analysis of Covariance

Effect	Wilks' λ	F	df1	df2	Sig.	η^2	***Result***
Intercept	0.798	5.636	3	67	0.002	0.202	Supported
IME	0.304	51.208	3	67	0.000	0.696	Supported
Specialty	0.760	3.279	6	134	0.005	0.128	Supported
IME* Specialty	0.803	2.585	6	134	0.021	0.104	Supported

Table 32

Levene's Test of Equality of Error Variances and Box's Test of Equality of Covariance Matrices

Name of Test	Item	F	df1	df2	Sig.
Levene's Test of Equality of Error Variancesa	IMC	1.847	2	72	0.165
	IMSC	0.456	2	72	0.636
	SOL	0.027	2	72	0.974
Box's M	15.127	1.102	12	1257.455	0.354

According to Figure 32 and Table 33 ($\beta = 0.849$, $t = 6.478$, $p < 0.001$), IME was substantially and favorably related to IMC. As a consequence, Hypothesis 2 is supported and shown to be true, suggesting that using the performance scale application has enhanced students' learning success by enhancing their communication. A substantial and positive association between IME and IMSC was also discovered ($\beta = 0.\ 834$, $t = 6.707$, $p < 0.001$). The implementation of the PSA has enhanced students' learning attainment by enhancing their scientific content, confirming Hypothesis 3 to be valid and accepted. The results also showed that IME was positively and substantially linked with SOL ($\beta = 0.830$, $t = 6.614$, $p < 0.001$). The utilization of the PSA has enhanced students' learning success by enhancing their pleasure with learning, which proves Hypothesis 4 to be correct and approved.

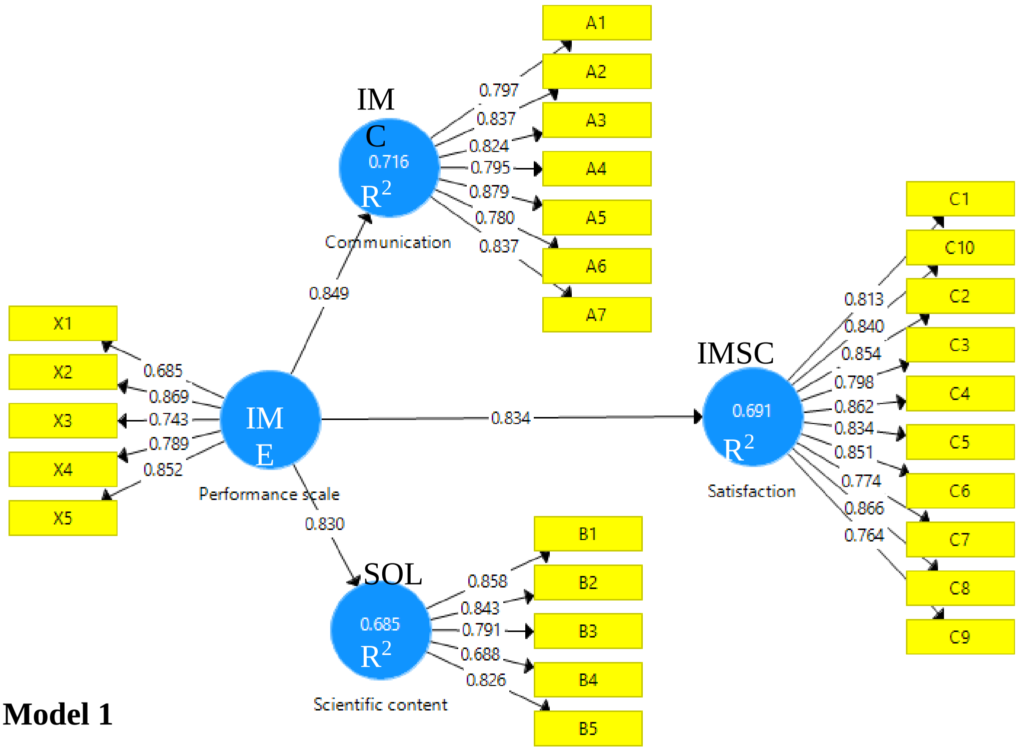
The results also showed that IME was significantly and favorably linked with individualized L ($\beta = 0.791$, $t = 6.491$, $p < 0.001$) (Figure 32 and Table 33) The utilization of the PSA has enhanced students' learning attainment by enhancing their tailored learning, which proves Hypothesis 5 is correct and acceptable. Regarding claim number 6, the results suggest that IME was significantly and favorably correlated with Distance L ($\beta = 0.803$, $t = 6.154$, $p < 0.001$). As a consequence, Hypothesis 6 is confirmed, demonstrating that the performance scale application has enhanced students' distant learning and improved learning attainment. Similar to this, the findings suggest that IME has a substantial and positive relationship with mobile L ($\beta = 0.721$, $t = 6.097$, $p <$

0.001). As a consequence, Hypothesis 7 is accepted, proving that using the performance scale application has enhanced mobile learning, which has enhanced student learning attainment. Additionally, the results indicate that IME was positively and strongly correlated with Self L ($\beta = 0.833$, $t = 6.547$, $p < 0.001$). The utilization of the PSA has increased students' learning success by enhancing their self-learning, which is why Hypothesis 8 is accepted. Finally, the findings suggest that there is a significant and positive relationship between IME and IME ($\beta = 0.765$, $t = 6.245$ $p < 0.001$). Therefore, Hypothesis 9 is confirmed and accepted, which states that through enhancing students' speciality learning, the performance scale application has increased students' learning attainment (Figure 32 and Table 33).

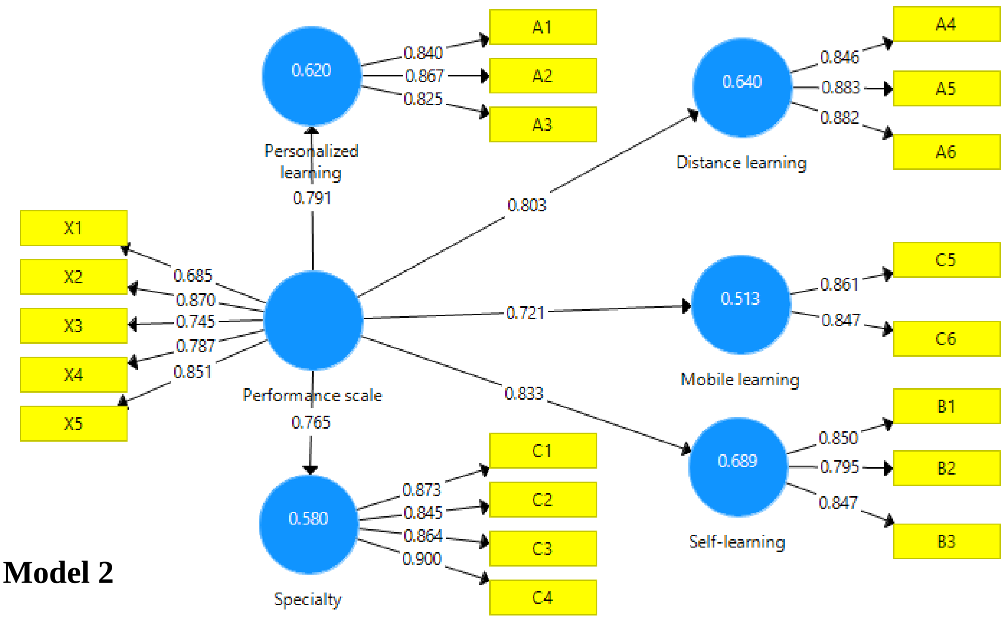
Furthermore, Figure 30 showed that the range of the R^2 value is from 0 to 1, with greater values suggesting a stronger explanatory power. R^2 values of 0.75, 0.50, and 0.25 are mostly considered as being significant, moderate, and weak, respectively (Franke & Sarstedt, 2019; Hilkenmeier et al., 2020). R^2 scores for the two models range from 0.515 to 0.716. The model has a medium level of predictive power as a consequence.

Figure 32

Measurement of the Dependent Variables for the Proposed Model 1 and 2.



Model 1



Model 2

Table 33

Structural Models For Hypothesis Testing Results

H	Independent	Relationship	Dependent V.	Estimate	S.E.	C.R.	P	***Result***
H2	IME	→	IMC	0.849	0.119	6.478	***	Supported
H3	IME	→	IMSC	0.834	0.138	6.707	***	Supported
H4	IME	→	SOL	0.830	0.149	6.614	***	Supported
H5	IME	→	Personalized L.	0.791	0.123	6.491	***	Supported
H6	IME	→	Distance L.	0.803	0.128	6.154	***	Supported
H7	IME	→	Mobile L.	0.721	0.123	6.097	***	Supported
H8	IME	→	Self L.	0.833	0.155	6.547	***	Supported
H9	IME	→	Specialty L.	0.765	0.123	6.245	***	Supported

CHAPTER V

DISCUSSION

This chapter presents the discussion for the findings of this study in comparison with the previous studies.

This study has developed a mobile application named the Performance Scale Application (PSA) as a guide and support for effective e-learning apps for students in secondary schools in Jordan. The app has effectively supported distance learning and self-learning, improves student efficiency, increases their effective participation, specialty choice ability, and guides them to the appropriate sections of their scientific abilities. In addition, it supports students in making an appropriate choice of specialty of study areas or departments and makes learning easier than traditional learning. The application was tested on a sample of students, and the results showed that the PSA was an effective educational tool because it supported learning strategies such as self-learning, personalized, and distance learning, which in turn enhanced the academic achievements or performances of the students, particularly in math, geography, and physics subjects, but it was not effective in the industrial specialty. It appeared that the application was effective for the scientific specialty and partially effective in the literary specialty, especially with the subject of geography, and ineffective for the Arabic course. The application is partially effective for the industrial specialty, especially with respect to the subject of the drawing course. The reason may be due to the lower average scores of the control group in the post-test than in the pre-test because of their lack of motivation. Another reason may be due to the practical nature of this specialty. The findings of this study corroborate and reinforce by earlier studies that highlight the value of employing mobile applications in secondary education settings, where Alanezi (2017) emphasized the value of using the Edmodo app in courses on mathematics and computer science. The importance of incorporating mobile learning has been stressed in numerous earlier studies (Kuimova et al., 2018; Bettayeb et al., 2020) because it creates a more engaging learning environment and encourages students to participate more and express their thoughts.

According to the approach, there was a substantial rise in scores, indicating that those who participated in the experimental group had mean scores that were greater than those of the control group's non-participants. But the classes in Arabic, drawing, and engineering failed some statistical tests. The cause could be attributed to the control group's lower mean results in the post-test compared to the pre-test due to their lack of enthusiasm. despite the fact that the experimental group's average scores greatly improved across all courses. This can be because of how the course is structured or because of uncontrollable external events. For the courses in Math, Physics, and Geography, follow-up comparisons showed a significant effect between performance in these courses and possible lifelong learning. According to Hu et al. (2021), mobile learning is a crucial component in achieving "permanent learning," or lifelong learning.

Many teachers confirmed the results obtained from the students with supportive, opponent, and advice regarding the PSA, which proves its importance as an effective educational tool that supports learning strategies at the secondary stage. The teachers' responses strongly showed that the PSA can support students in secondary schools to engage in distance learning when it is needed, for example during COVID-19. Similarly, according to studies (Heggart & Yoo, 2018; Jordan & Duckett, 2018), the Google Classroom application efficiently supports remote learning practices and encourages students to engage in self-directed learning. According to Bai (2019), the use of mobile learning apps in high school has been shown to be effective in treating students' deficiencies, which supports the idea of a personalized learning strategy.

Findings also showed that there were a few teachers who had some negative opinions regarding the PSA in terms of the students' lack of time to cover all topics. Because the secondary school curriculum is so dense, the student is unable to complete and study all of the content curricula.Consistent with this, according to Darmaji et al. (2019), any instrument that enhances academic achievement also maintains the student's time management. Therefore, mobile learning helps students conserve time rather than waste it (Sun et al., 2017). It is important to note that one of the teachers' negative opinions was in line with many other studies, as many researchers stressed that the issue of bad network and Internet connection is a significant barrier to mobile learning, as was the case in the studies of Vishwakarma (2015) and Sharma et al. (2018). Jeong (2022) also indicated that having access to educational content via mobile applications enables students to continuously learn for themselves, supporting the sustainability of learning. According to Huilcapi-Collantes et al. (2020), some user-friendly mobile applications make it simple for teachers and students to communicate at all times and locations for the lowest possible cost. Sun et al. (2017) also stated that students' self-confidence increases, and their levels of worry and anxiety decrease while using mobile learning applications that enhance learning practices.

The research's findings have confirmed that the additional trees regression technique is a useful tool for determining students' secondary school achievement while developing PSA. This is because, after the usage of PSA ation, it was discovered that all tested hypotheses supported the students' secondary successes in terms of their choice of speciality, IMC, IMSC, SOL, customized L, distance L, mobile L, self L, and specialty L. The results demonstrated that using PSA on mobile apps is a successful technique to raise student accomplishment levels through the eLearning revolution in the classroom. It's never too late to learn. Due to the accessibility of self-learning through mobile apps, eLearning has taken center stage (Sayibu et al., 2021). Due to the fact that a PSA is feature-oriented and produced by the extra-tree regression technique, students can use it at their own pace and take their time learning.

Students' choice of specialization is influenced by the performance scale application, and this option was found to significantly affect the IMC, IMSC, SOL, personalized L, distance L, mobile L, and self-L. Results from the IME, which had a strong and favorable correlation with specialty learning, corroborated this. This demonstrated that the decision-making process for specialization begins with student preference. According to Alonso-Virgó et al. (2021; De Backer et al. 2022) and other researchers, specialty is the outcome of a complex interaction between student GPA expectations, specialty area expectation, and contest for unfilled employment. The present study, especially in personalized learning and distant learning, identified a PSA for following learners' specialized choices as a learning intervention. The approach promotes specialist choice while enhancing secondary school performance, especially for individuals who have just made or are about to make their specialty choice. The use of a performance scale as a specific guide can help students demonstrate their skills in proportion to their accomplishments.

IMC is benefited by IME. This demonstrated how the performance scale application helps students in secondary schools communicate better with their peers and teachers about the course topics, their concerns, and their areas of specialization. According to Bimayu and Nopriadi (2019), communication varies among students pursuing method-oriented specialization, which may be a response to the question of how crucial it is for a student's specialization to complement their personality, interests, and talents. Mobile apps can be used to encourage interaction amongst students via chat, forums, and other characteristics that facilitate student participation (Almaiah et al., 2020). The success of pupils depends on collaboration between teachers and students via eLearning platforms, such as performance scale applications (Meddar et al., 2022). Numerous aspects of education are using mobile applications, and their use has been expanding significantly (Liu & Li, 2022).

Furthermore, IME has a favorable impact on IMSC. This conclusion is highly supported by the great response percentage in science (49%) than literature (41%) and business (1%), as found in this research. Fostering a real interest in science among high school students is a fundamental goal of IMSC on mobile apps, as it is an essential element of scientific literacy. Recent research (Sari et al., 2019; Gültekin & Altun, 2022) have found that school-based scientific material has been helpful in raising students' IME, which in turn affects their interest in science. This research confirms what we found. Given the abundance of studies showing the benefits of evaluation skills on a range of learning findings, it is rational to assume that learners have strong evaluation skills that can help them choose the right specialization because they are less probable or even unable to involve with significant societal issues associated with science.

Furthermore, IME also has a favorable impact on SOL. The findings show that motivating students to evaluate their interests via mobile learning applications (also known as PSA) resulted in higher levels of satisfaction. Huang et al. (2022) assertion that application usage influences mobile user happiness provided additional support for the findings. The findings are consistent with earlier study (El Gourari et al., 2020; Qiu et al., 2022), which showed that learners are more expected to be inspired to utilize PSA as an eLearning assistance if they feel them appropriate to their learning effects and specialist support. If they have IME and feel it appropriate for their learning, learners are more probable to be pleased with the application of the performance scale.

IME has a favorable impact on customized L. This shows that performance scale applications that enable tailored learning can boost student success via the school's support system. Personalized learning may necessitate a number of structural and curriculum changes in schools. However, tailored learning's outcomes suggest that this method of instruction could significantly improve students' academic achievement (Sereno, 2018; Ugur-Erdogmus & Akr, 2022). Research suggests that using mobile apps can boost students' participation in their studies (Bourekkache et al., 2020). Personalized learning is impacted by student choice, which revealed a connection between SOL, IMSC, accomplishment, and specialist choice. According to researchers, having a choice increases pupils' sense of competence and autonomy, which raises their achievement (Alkhateeb & Al-Duwairi, 2019).

Distance L is benefited by IME. The effective use of eLearning may be credited with this important finding. Internet access and application-enabled devices are the main tools for implementing e-learning in the classroom. Students are taking more ownership of their information acquisition because to the growing popularity of online learning (Blinova, 2022). Students can learn and decide for themselves at the secondary level what their areas of interest and expertise are acknowledges to eLearning apps like the PSA. According to AlAlami, Adwan, & Alsous, higher GPA students have been demonstrated to do better academically when learning remotely and at their own pace (2022). Their impressive academic performance has mostly been the result of individualized instruction. They could feel less at ease in a virtual environment where it might be more difficult to have meaningful interactions without a mobile app.

The same is true for how IME benefits mobile L. This result suggests that if students utilize mobile devices for academic learning, their academic performance may be affected. Application of the performance scale revealed that mobile learning had a significant influence on students' learning performance and that their attitudes regarding utilizing their phones for learning were relatively positive in their selected field. In their study of how instructors viewed students' learning achievement when mobile devices were utilized in the classroom, Ugur-Erdogmus & Akr (2022) discovered that 91% of the instructors at different grade levels examined thought that mobile gadgets had a significant influence on students' learning in the classroom. This work agrees with what we discovered. More and more study are being done to figure out how to encourage students to use their portable devices in and beyond of the classroom. Nevertheless, the best approach has not yet been agreed upon. Based on the study's findings, performance sales applications that employ ML algorithms could be dependable mobile applications that aid secondary school learners in enhancing their academic performance.

IME also has a favorable impact on self-L. This shows that, in comparison to learners whose grades do not factor into the final grade, students who engage in self-learning can obtain grades that increase their GPA. Research have shown that students are rational, supporting the study's conclusion. If students are conscious of the fact that their activities can influence how they ultimately choose their speciality, mobile apps can help them foster better self-evaluation (De Backer et al., 2022; Fairlamb, 2022; Lei et al., 2022).

CHAPTER VI

CONCLUSION AND RECOMMENDATIONS

This section describes conclusions according to the research findings consistent with the objective of the research. The chapter also provides recommendations for future studies and outlines the limitations of this research.

6.1 Conclusion

This study has successfully developed, tested, and evaluated a performance scale application (PSA) for student usage at secondary school as a mobile and e-learning support to improve their academic learning and achievements. The learning achievement of pupils in secondary school in Jordan was found to be positively and significantly impacted by PSA in this study. The findings revealed that the PSA was both supportive and effective as an educational tool for secondary students. The PSA's two primary functions are: the first one involves calculating the average grade for each student in the 11th grade based on their performance in the core subjects and the specialization they chose in the prior year. The second involves taking a level test in the academic topics relevant to each specialty; the results are given as percentages based on how well the students performed in this specialty. In addition, it serves as a preparation year for the students from grade 11 and grade 12. This demonstrates that the ML technique, also known as Extra Trees Regression, is a useful tool for creating PSA. The algorithm has effectively predicted the Cumulative Grade Point Average (GPA) and guides the students to make the appropriate choice of an appropriate specialty or branch of area to avoid facing failure at secondary level in grade 12. The best model is therefore only guaranteed as a first entry passing through genetic and fuzzy modelling, resulting in students' academic achievement and the best specialization placement suitable for their ability, performance, and school's administration records.

The results of all investigated hypotheses were determined to be in favor of the students' academic success. Each of the assumptions is confirmed by the results. The system's performance allowed it to correctly forecast the students' speciality choice, SOL, mobile L, IMC, personalized L, IMSC, distance L, self-L, and specialty L. The results showed that students who utilized the PSA indicated positive self-learning as measured by the IME, IMC, IMSC, and SOL.

In addition, PSA was effective in the scientific specialty while partially effective in the literary specialty, especially in geography, and ineffective in Arabic. But the application was not effective for the industrial specialty. The results also confirmed that the usage of the developed PSA reduced students’ anxiety and student efficiency, increased their effective participation, increased self-confidence, improved choice ability, and improved weaknesses. It also made it easier for the student to request any advice without any hesitation, and also made it easier for the learner to access any educational content at any time and any place.

In conclusion, the PSA is effective at predicting secondary school students' academic success and choice of specialization. The PSA can provide academic assistance for administrators in arranging instructional materials, grouping students, keeping their academic records, and providing them with support according to their individual abilities and scientific skills. Thus, this can be used as an e-learning device in Jordan's secondary school.

6.2 Recommendations for Future Studies

As a prerequisite for any study to provide recommendations for future studies in order to pave the way for possible improvements on the outcomes of this research, this study provides the following suggestions:

- Based on the results of this investigation, it was suggested that further features for the PSA be included in future research. Response time, screen changes, demand per second, network utilization, memory use, task completion time, and throughput are some of these functionalities. Future research can be used to carry out these enhancements.

- Future research might also evaluate this PSA in a number of carefully chosen secondary schools with regard to a crucial factor and the availability of diverse specialties. This study's variables can also be taken into account.

- The recommendations of this study include that it may be possible in the future to include some educational activities and features to the PSA that assist students with special needs, which may also include certain instructional games.

- It is not advised to disseminate the results widely for undertaking additional research on other contexts and situations as well as using larger samples to further confirm the effectiveness and viability of the PSA. Due to time restrictions and COVID-19 pandemic impact on social isolations, this study use the relatively moderate sample size.

- We also advise undertaking further research to determine the real causes of the outcomes of the application's failure in the Arabic and industrial specializations.

6.3 Limitations of the Study

There is no study without limitations or challenges faced during the process of doing the study, particularly in data collection processes, analysis, and reporting. The limitations of this study are discussed below.

Data collection is seen as one of the most challenging aspects of the study, particularly as the research's timing corresponds with the COVID-19 outbreak and necessitates information from schools and other relevant regulators.

One of the most important constraints faced in this research is the existence of a group of learners who are not interested in studying generally and do not want to make any effort. Also, some topics in the industrial specialties need a practical explanation, such as engineering and drawing. Therefore, future studies need to design the PSA with motivating features and re-simplify the industrial courses to encourage these students who are not willing to be involved in the study to actively participate.

In addition, lack of full cooperation from majority of the students in participating in this study due to social distancing and isolation caused by COVID-19. This further limited the numbers of projected numbers of participants supposed to have involved.

Despite the fact that the bulk of the pertinent studies presume academic performance, a small number of them are reported about the secondary school students, which constituted the challenges of this study during information collection and literature review process.

Due to the fact that only Jordan's secondary schools and those with comparable settings and features were considered, the results cannot be generalized. One of the biggest problems is the utilization of a single school to derive a conclusion.

REFERENCES

Ababneh, M., Aljarrah, A., Karagozlu, D., & Ozdamli, F. (2021). Guiding the students in high school by using machine learning. *TEM Journal*, *10*(1), 384-391.

Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, *10*(1), 61-75.

Afshar, M., Sharma, B., Dligach, D., Oguss, M., Brown, R., Chhabra, N., ... & Karnik, N. S. (2022). Development and multimodal validation of a substance misuse algorithm for referral to treatment using artificial intelligence (SMART-AI): a retrospective deep learning study. *The Lancet Digital Health*, *4*(6), e426-e435.

Ahsan, M. M., Mahmud, M. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, *9*(3), 52.

Akram, A., Fu, C., Li, Y., Javed, M. Y., Lin, R., Jiang, Y., & Tang, Y. (2019). Predicting students' academic procrastination in blended learning course using homework submission data. *Ieee Access*, *7*, 102487-102498.

Al Zieni, H. (2019). The effect of mobile learning on learner autonomy: A suggested measurement tool to assess the development of learner autonomy. *Journal of Asia TEFL*, *16*(3), 1020.

Al-Alami, Z. M., Adwan, S. W., & Alsous, M. (2022). Remote learning during Covid-19 lockdown: A study on anatomy and histology education for pharmacy students in Jordan. *Anatomical sciences education*, *15*(2), 249.

Alanezi, Y. (2017, October). EffectEffective use of educational platform ((Edmodo)) for students of mathematics and Computer specialty in t. In *Proceedings of International Academic Conferences* (No. 5907221). International Institute of Social and Economic Sciences.

Alhasani, M., & Orji, R. (2022, March). SortOut: Persuasive Stress Management Mobile Application for Higher Education Students. In *International Conference on Persuasive Technology* (pp. 16-27). Springer, Cham.

Ali, A. D., & Hanna, W. K. (2022). Predicting students' achievement in a hybrid environment through self-regulated learning, log data, and course engagement: A data mining approach. *Journal of Educational Computing Research*, *60*(4), 960-985.

Ali, W., Riaz, O., Mumtaz, S., Khan, A. R., Saba, T., & Bahaj, S. A. (2022). Mobile Application Usability Evaluation: A Study Based on Demography. *IEEE Access*, *10*, 41512-41524.

Alkhateeb, M. A., & Al-Duwairi, A. M. (2019). The Effect of Using Mobile Applications (GeoGebra and Sketchpad) on the Students' Achievement. *International Electronic Journal of Mathematics Education*, *14*(3), 523-533.

Almaiah, M. A., Al-Khasawneh, A., & Althunibat, A. (2020). Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic. *Education and information technologies*, *25*(6), 5261-5280.

Almaiah, M. A., Hajjej, F., Lutfi, A., Al-Khasawneh, A., Alkhdour, T., Almomani, O., & Shehab, R. (2022). A Conceptual Framework for Determining Quality Requirements for Mobile Learning Applications Using Delphi Method. *Electronics*, *11*(5), 788.

Alonso-Virgós, L., Fondon, M. D., Espada, J. P., & Crespo, R. G. (2021, April). Women in science and technology studies. A study about the influence of parents on their children's choice of speciality. And about the trend of the different specialities in Spanish students. In *2021 IEEE Global Engineering Education Conference (EDUCON)* (pp. 122-130). IEEE.

Alowayr, A., & Al-Azawei, A. (2021). Predicting mobile learning acceptance: An integrated model and empirical study based on higher education students' perceptions. *Australasian Journal of Educational Technology*, *37*(3), 38-55.

al-Qallawi, S., & Raghavan, M. (2022). A review of online reactions to game-based assessment mobile applications. *International Journal of Selection and Assessment*, *30*(1), 14-26.

Al-Rahmi, A. M., Al-Rahmi, W. M., Alturki, U., Aldraiweesh, A., Almutairy, S., & Al-Adwan, A. S. (2022). Acceptance of mobile technologies and M-learning by university students: An empirical investigation in higher education. *Education and Information Technologies*, 1-22.

Alsharida, R., Hammood, M., & Al-Emran, M. (2021). Mobile Learning Adoption: A Systematic Review of the Technology Acceptance Model from 2017 to 2020. *International Journal of Emerging Technologies in Learning (IJET)*, *16*(5), 147-162.

Althunibat, A., Altarawneh, F., Dawood, R., & Almaiah, M. A. (2022). Propose a New Quality Model for M-Learning Application in Light of COVID-19. *Mobile Information Systems*, *2022*.

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1), 1-21.

Alyoussef, I. Y. (2021). E-Learning acceptance: The role of task–technology fit as sustainability in higher education. *Sustainability*, *13*(11), 6450.

Amilyana, A. S., & Noer, T. K. M. S. (2021). A Teaching Material Based on Science, Environment, Technology, and Society to Improve Student's Critical Thinking Skills: Synchronous and Asynchronous Learning During Covid-19 Pandemic. *IJORER: International Journal of Recent Educational Research*, *2*(4), 372-391.

Amjad, S., Younas, M., Anwar, M., Shaheen, Q., Shiraz, M., & Gani, A. (2022). Data Mining Techniques to Analyze the Impact of Social Media on Academic Performance of High School Students. *Wireless Communications and Mobile Computing*, *2022*.

Anis, L., Perez, G., Benzies, K. M., Ewashen, C., Hart, M., & Letourneau, N. (2020). Convergent validity of three measures of reflective function: Parent development interview, parental reflective function questionnaire, and reflective function questionnaire. *Frontiers in Psychology*, *11*, 574719.

Ariffin, S. A., Sidek, S. F., & Mutalib, M. F. H. (2018). A Preliminary Investigation of Malaysian Student's Daily Use of Mobile Devices as Potential Tools for STEM in a Local University Context. *International Journal of Interactive Mobile Technologies*, *12*(2), 80-91.

Asghar, M. Z., BarberÃ, E., & Younas, I. (2021). Mobile learning technology readiness and acceptance among pre-service teachers in Pakistan during the COVID-19 pandemic. *Knowledge Management & E-Learning: An International Journal*, *13*(1), 83-101.

Astuti, I., & Suratman, D. (2022). Development Of Mobile Learning Based On A Scientific Approach To Learning Elementary School Themes. *Jurnal Scientia*, *10*(2), 291-301.

Avcı, Ü., & Ergün, E. (2022). Online students' LMS activities and their effect on engagement, information literacy and academic performance. *Interactive Learning Environments*, *30*(1), 71-84.

Aydoğdu, Ş. (2020). Predicting student final performance using artificial neural networks in online learning environments. *Education and Information Technologies*, *25*(3), 1913-1927.

Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., Alsariera, Y. A., Ali, A. Q., ... & Tiong, S. K. (2022). Toward predicting student's academic performance using artificial neural networks (ANNs). *Applied Sciences*, *12*(3), 1289.

Babalola, E. O., & Omolafe, E. V. (2021). Effect of Developed Mobile Application on Undergraduates Academic Performance in Computer Science. *ASEAN Journal of Science and Engineering Education*, *2*(3), 215-222.

Bai, H. (2019). Pedagogical practices of mobile learning in K-12 and higher education settings. *TechTrends*, *63*(5), 611-620.

Basri, M., Husain, B., & Modayama, W. (2021). University students' perceptions in implementing asynchronous learning during Covid-19 era. *Metathesis: Journal of English Language, Literature, and Teaching*, *4*(3), 263-276.

Batool, S. S. (2020). Academic achievement: Interplay of positive parenting, self-esteem, and academic procrastination. *Australian Journal of Psychology*, *72*(2), 174-187.

Bayar, M. F., & Kurt, U. (2021). Effects of mobile learning science course on students' academic achievement and their opinions about the course. *Science Education International*, *32*(3), 254-263.

Berei, E. B., & Pusztai, G. (2022). Learning through Digital Devices—Academic Risks and Responsibilities. *Education Sciences*, *12*(7), 480.

Bernacki, M. L., Greene, J. A., & Crompton, H. (2020). Mobile technology, learning, and achievement: Advances in understanding and measuring the role of mobile technology in education. *Contemporary Educational Psychology*, *60*, 101827.

Bettayeb, H., Alshurideh, M. T., & Al Kurdi, B. (2020). The effectiveness of Mobile Learning in UAE Universities: A systematic review of Motivation, Self-efficacy, Usability and Usefulness. *International Journal of Control and Automation*, *13*(2), 1558-1579.

Bimayu, W., & Nopriadi, N. (2019). Learning Based on Information Technology and Learning Strategy Communication in Facing Industrial Revolution Era 4.0. *International Journal of Educational Review*, *1*(2), 77-86.

Blinova, O. (2022, March). What Covid Taught Us About Assessment: Students'perceptions Of Academic Integrity In Distance Learning. In *Proceedings of INTED2022 Conference* (Vol. 7, p. 8th).

Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of organizational behavior*, *36*(1), 3-15.

Bossman, A., & Agyei, S. K. (2022). Technology and instructor dimensions, e-learning satisfaction, and academic performance of distance students in Ghana. *Heliyon*, *8*(4), e09200.

Bourekkache, S., Tigane, S., Kazar, O., & Kahloul, L. (2020, December). Mobile and personalized learning system for computer science students. In *2020 Sixth International Conference on e-Learning (econf)* (pp. 189-193). IEEE.

Bradley-Dorsey, M., Beck, D., Maranto, R., Tran, B., Clark, T., & Liu, F. (2022). Is cyber like in-person? Relationships between student-student, student-teacher interaction and student achievement in cyber schools. *Computers and Education Open*, *3*, 100101.

Brown, C., Spiro, J., & Quinton, S. (2020). The role of research ethics committees: Friend or foe in educational research? An exploratory study. *British Educational Research Journal*, *46*(4), 747-769.

Buchanan, E. A., & Hvizdak, E. E. (2009). Online survey tools: Ethical and methodological concerns of human research ethics committees. *Journal of empirical research on human research ethics*, *4*(2), 37-48.

Cachia, M., Lynam, S., & Stock, R. (2018). Academic success: Is it just about the grades?. *Higher Education Pedagogies*, *3*(1), 434-439.

Cerezo, R., Bogarín, A., Esteban, M., & Romero, C. (2020). Process mining for self-regulated learning assessment in e-learning. *Journal of Computing in Higher Education*, *32*(1), 74-88.

Chang, Y. C., Li, J. W., & Huang, D. Y. (2022). A Personalized Learning Service Compatible with Moodle E-Learning Management System. *Applied Sciences*, *12*(7), 3562.

Chao, C. M. (2019). Factors determining the behavioral intention to use mobile learning: An application and extension of the UTAUT model. *Frontiers in psychology*, *10*, 1652.

Chau, K. Y., Law, K. M., & Tang, Y. M. (2021). Impact of self-directed learning and educational technology readiness on synchronous E-learning. *Journal of Organizational and End User Computing (JOEUC)*, *33*(6), 1-20.

Chau, K. Y., Law, K. M., & Tang, Y. M. (2021). Impact of self-directed learning and educational technology readiness on synchronous E-learning. *Journal of Organizational and End User Computing (JOEUC)*, *33*(6), 1-20.

Chevet, G., Baccino, T., Marlot, L., Vinter, A., & Drai-Zerbib, V. (2022). Effects of interruption on eye movements and comprehension during reading on digital devices. *Learning and Instruction*, *80*, 101565.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Daraghmi, E. (2022). The Adoption of Augmented Reality Technology in e-learning: Case of a Mobile Application During Covied-19. In *International Conference on Business and Technology* (pp. 915-936). Springer, Cham.

Darmaji, D., Kurniawan, D., Astalini, A., Lumbantoruan, A., & Samosir, S. (2019). Mobile learning in higher education for the industrial revolution 4.0: Perception and response of physics practicum. *Education for Primary Care*, *22*(5), 297-303.

De Backer, L., Van Keer, H., De Smedt, F., Merchie, E., & Valcke, M. (2022). Identifying regulation profiles during computer-supported collaborative learning and examining their relation with students' performance, motivation, and self-efficacy for learning. *Computers & Education*, *179*, 104421.

De, K., & Goswami, S. (2018). An Introspection of the Responses of the Ruralites and the Elderly to Information and Communication Technology in Open Distance Learning. In *Technology for Efficient Learner Support Services in Distance Education* (pp. 227-242). Springer, Singapore.

Deng, H., Wang, X., Guo, Z., Decker, A., Duan, X., Wang, C., ... & Abbott, K. (2019). Performancevis: Visual analytics of student performance data from an introductory chemistry course. *Visual Informatics*, *3*(4), 166-176.

Diesta, A. R. L., & Ferolino, C. H. (2021). Adaptive Coping Strategies in Learning General Mathematics Instructions Through Self-Learning Modules (SLMs). *United International Journal for Research & Technology*, *2*(11), 9-14.

Dorji, T., Tamang, S. T., & Tilak, T. V. S. V. G. K. (2021). Self-learning on COVID-19 among medical students in Bhutan: A cross-sectional study. *Heliyon*, *7*(7), e07533.

El Gourari, A., Skouri, M., Raoufi, M., & Ouatik, F. (2020, December). The future of the transition to E-learning and distance learning using artificial intelligence. In *2020 Sixth International Conference on e-Learning (econf)* (pp. 279-284). IEEE.

Emedoli, D., Alemanno, F., Houdayer, E., Brugliera, L., Iannaccone, S., & Tettamanti, A. (2022). Mobile application tool for remote rehabilitation after discharge from coronavirus disease-19 rehabilitation unit. *Healthcare Technology Letters*.

Fairlamb, S. (2022). We need to talk about self-esteem: The effect of contingent self-worth on student achievement and well-being. *Scholarship of Teaching and Learning in Psychology*, *8*(1), 45.

Farrell, A. M., & Rudd, J. M. (2009). Factor analysis and discriminant validity: A brief review of some practical issues. Anzmac.

Ferri, F., Grifoni, P., & Guzzo, T. (2020). Online learning and emergency remote teaching: Opportunities and challenges in emergency situations. *Societies*, *10*(4), 86.

Franke, G., & Sarstedt, M. (2019). Heuristics versus statistics in discriminant validity testing: a comparison of four procedures. *Internet Research*.

Fuad, A., & Al-Yahya, M. (2021). Analysis and classification of mobile apps using topic modeling: A case study on Google Play Arabic apps. *Complexity*, *2021*.

Funa, A., & Talaue, F. (2021). Constructivist learning amid the COVID-19 pandemic: Investigating students' perceptions of biology self-learning modules. *International Journal of Learning, Teaching and Educational Research*, *20*(3), 250-264.

Gay, L.R., Mills, G.E. and Airasian, P. (2009) Educational Research Competencies for Analysis and Applications. Pearson, Columbus.

Goldenberg, M. N., Williams, D. K., & Spollen, J. J. (2017). Stability of and factors related to medical student specialty choice of psychiatry. *American Journal of Psychiatry*, *174*(9), 859-866.

Gültekin, S. B., & Altun, T. (2022). Investigating the Impact of Activities Based on Scientific Process Skills on 4th Grade Students' Problem-Solving Skills. *International Electronic Journal of Elementary Education*, *14*(4), 491-500.

Güngörmez, H. G. (2021). An Investigation into the Relationship between Secondary School Students' Self-Learning Using Technology and Test Anxiety. *International Journal of Education and Literacy Studies*, *9*(4), 61-68.

Halde, R. R. (2016, September). Application of Machine Learning algorithms for betterment in education system. In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)* (pp. 1110-1114). IEEE.

Han, Y. (2022). Using mobile applications in the study of vocal skills. *Education and Information Technologies*, 1-21.

Hartley, K., & Andújar, A. (2022). Smartphones and Learning: An Extension of M-Learning or a Distinct Area of Inquiry. *Education Sciences*, *12*(1), 50.

Heggart, K., & Yoo, J. (2018). Getting the most from Google Classroom: A pedagogical framework for tertiary educators. *Australian Journal of Teacher Education*, *43*(3), 140-153.

Hilkenmeier, F., Bohndick, C., Bohndick, T., & Hilkenmeier, J. (2020). Assessing distinctiveness in multidimensional instruments without access to raw data–a manifest Fornell-Larcker criterion. *Frontiers in psychology*, *11*, 223.

Hines, E. M., Moore III, J. L., Mayes, R. D., Harris, P. C., Vega, D., Robinson, D. V., ... & Jackson, C. E. (2020). Making student achievement a priority: The role of school counselors in turnaround schools. *Urban Education*, *55*(2), 216-237.

Hossain, M. S., Rahman, M. F., & Uddin, M. K. (2022). Analyzing and Predicting Learner Sentiment Toward Specialty Schools Using Machine Learning Techniques. In *Applying Data Science and Learning Analytics Throughout a Learner's Lifespan* (pp. 133-158). IGI Global.

Hsiao, H. S., Lin, Y. W., Lin, K. Y., Lin, C. Y., Chen, J. H., & Chen, J. C. (2022). Using robot-based practices to develop an activity that incorporated the 6E model to improve elementary school students' learning performances. *Interactive Learning Environments*, *30*(1), 85-99.

Hu, H., Feng, F., Lai, Z., Yang, J., & Wang, T. (2021, October). Optimization of Self-Learning Speed-Tracking Control for Permanent Magnet Synchronous Motor. In *International Conference on Electrical and Information Technologies for Rail Transportation* (pp. 232-240). Springer, Singapore.

Huang, R. T., Jabor, M. K., Tang, T. W., & Chang, S. C. (2022). Examine the moderating role of mobile technology anxiety in mobile learning: a modified model of goal-directed behavior. *Asia Pacific Education Review*, *23*(1), 101-113.

Huang, R. T., Yu, C. L., Tang, T. W., & Chang, S. C. (2021). A study of the use of mobile learning technology in Taiwan for language learning. *Innovations in Education and Teaching International*, *58*(1), 59-71.

Huilcapi-Collantes, C., Hernández Martín, A., & Hernández-Ramos, J. P. (2020, October). Pedagogical and user interface usability evaluation of an educational mobile app that promotes visual literacy. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 315-321).

Huilcapi-Collantes, C., Hernández Martín, A., & Hernández-Ramos, J. P. (2020, October). Pedagogical and user interface usability evaluation of an educational mobile app that promotes visual literacy. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 315-321).

Ilgaz, H. (2021). The role of mobile technologies in shifting to digital. *Educational Technology Research and Development*, *69*(1), 189-190.

Iyer, L. S., Bharadwaj, S., Shetty, S. H., Verma, V., & Devanathan, M. (2022). Advancing Equity in Digital Classrooms: A Personalized Learning Framework for Higher Education Institutions. In *Socioeconomic Inclusion During an Era of Online Education* (pp. 225-245). IGI Global.

Jeong, K. O. (2022). Facilitating sustainable self-directed learning experience with the use of mobile-assisted language learning. *Sustainability*, *14*(5), 2894.

Jiang, Z., Yin, H., Luo, Y., Gong, J., Yang, Y., & Lin, M. (2019, October). Quantitative Analysis of Mobile Application User Interface Design. In *2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC)* (pp. 1-8). IEEE.

Jordan, M. M., & Duckett, N. D. (2018). Universities confront 'tech disruption': Perceptions of student engagement online using two learning management systems. *The journal of public and professional sociology*, *10*(1), 4.

Jordan, M. M., & Duckett, N. D. (2018). Universities confront 'tech disruption': Perceptions of student engagement online using two learning management systems. *The journal of public and professional sociology*, *10*(1), 4.

Kattayat, S., Josey, S., & Asha, J. V. (2017). Mobile Learning Apps in Instruction and Students Achievement. *Int. J. Interact. Mob. Technol.*, *11*(1), 143-147.

Kew, S. N., & Tasir, Z. (2022). Developing a learning analytics intervention in e-learning to enhance students' learning performance: A case study. *Education and Information Technologies*, 1-36.

Khan, R. A., Atta, K., Sajjad, M., & Jawaid, M. (2022). Twelve tips to enhance student engagement in synchronous online teaching and learning. *Medical Teacher*, *44*(6), 601-606.

Khery, Y., Nufida, B. A., Suryati, S., Rahayu, S., & Budiasih, E. (2019, April). Mobile Learning with Oriented Nature of Science (NOS): Does undergraduate school need it?. In *3rd Asian education symposium (AES 2018)* (pp. 227-232). Atlantis Press.

Klimova, B. (2019). Impact of mobile learning on students' achievement results. *Education Sciences*, *9*(2), 90.

Kuimova, M., Burleigh, D., Uzunboylu, H., & Bazhenov, R. (2018). Positive effects of mobile learning on foreign language learning. *TEM journal*, *7*(4), 837-841.

Kumar, S. R., & Moral, C. S. (2019). Attitude towards mobile learning in relation to academic performance of pre-teachers. *ZENITH International Journal of Multidisciplinary Research*, *9*(3), 293-303.

Laurens Arredondo, L. A., & Valdés Riquelme, H. (2022). M-learning adapted to the ARCS model of motivation and applied to a kinematics course. *Computer Applications in Engineering Education*, *30*(1), 77-92.

LeBeau, K., Huey, L. G., & Hart, M. (2019). Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the mobile application rating scale. *JMIR mHealth and uHealth*, *7*(5), e13019.

LeGeros, L., Bishop, P., Netcoh, S., & Downes, J. (2022). Informing the Implementation of Personalized Learning in the Middle Grades through a School-Wide Genius Hour. *RMLE Online*, *45*(1), 1-22.

Lei, H., Chiu, M. M., Wang, D., Wang, C., & Xie, T. (2022). Effects of game-based learning on students' achievement in science: a meta-analysis. *Journal of Educational Computing Research*, 07356331211064543.

Leite, W. L., Kuang, H., Jing, Z., Xing, W., Cavanaugh, C., & Huggins-Manley, A. C. (2022). The relationship between self-regulated student use of a virtual learning environment for algebra and student achievement: An examination of the role of teacher orchestration. *Computers & Education*, *191*, 104615.

Lim, H., Kim, S., Chung, K. M., Lee, K., Kim, T., & Heo, J. (2022). Is college students' trajectory associated with academic performance?. *Computers & Education*, *178*, 104397.

Lin, J., Zhao, Y., Liu, C., & Pu, H. (2020, June). Personalized learning service based on big data for education. In *2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI)* (pp. 235-238). IEEE.

Liu, P., & Li, Y. (2022). Response time evaluation of mobile applications combining network protocol analysis and information fusion. *Information and Software Technology*, *145*, 106838.

Lutfi, A., Saad, M., Almaiah, M. A., Alsaad, A., Al-Khasawneh, A., Alrawad, M., ... & Al-Khasawneh, A. L. (2022). Actual use of mobile learning technologies during social distancing circumstances: case study of King Faisal University students. *Sustainability*, *14*(12), 7323.

Malkawi, N. A. M. (2020). Attitudes of Secondary School Students in Jordan toward Contemporary Educational Concepts. *Journal of Education and Learning*, *9*(1), 229-236.

Marunevich, O., Kolmakova, V., Odaruyk, I., & Shalkov, D. (2021). E-learning and M-learning as tools for enhancing teaching and learning in higher education: a case study of Russia. In *SHS Web of Conferences* (Vol. 110). EDP Sciences.

Mastan, I. A., Sensuse, D. I., Suryono, R. R., & Kautsarina, K. (2022). Evaluation of distance learning system (e-learning): a systematic literature review. *Jurnal Teknoinfo*, *16*(1), 132-137.

Meddar, J. M., Ponnapalli, A., Azhar, R., Turchioe, M. R., Duran, A. T., & Creber, R. M. (2022). A Structured Review of Commercially Available Cardiac Rehabilitation mHealth Applications Using the Mobile Application Rating Scale. *Journal of Cardiopulmonary Rehabilitation and Prevention*, *42*(3), 141-147.

Mendoza, N. B., Yan, Z., & King, R. B. (2022). Domain-specific motivation and self-assessment practice as mechanisms linking perceived need-supportive teaching to student achievement. *European Journal of Psychology of Education*, 1-24.

Mergany, N. N., Dafalla, A. E., & Awooda, E. (2021). Effect of mobile learning on academic achievement and attitude of Sudanese dental students: a preliminary study. *BMC medical education*, *21*(1), 1-7.

Mihail-Văduva, D. (2022). The Effectiveness of a Multimedia Mobile Application. In *Education, Research and Business Technologies* (pp. 15-30). Springer, Singapore.

Moitra, D., & Mandal, R. K. (2019). Automated grading of non-small cell lung cancer by fuzzy rough nearest neighbour method. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *8*(1), 1-9.

Munoto, M., Sumbawati, M. S., & Sari, S. F. M. (2021). The use of mobile technology in learning with online and offline systems. *International Journal of Information and Communication Technology Education (IJICTE)*, *17*(2), 54-67.

Muzaini, M., Rahayuningsih, S., Nasrun, N., & Hasbi, M. (2021). Creativity in synchronous and asynchronous learning during the covid-19 pandemic: a case study. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *10*(3), 1722-1735.

Nafasov, M. M., Akhtamova, L. A. Q., & Sadullaeva, F. A. (2022). Technology of creating educational mobile applications. *Academic research in educational sciences*, *3*(4), 926-933.

Nawar, S., & Mouazen, A. M. (2017). Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. *Catena*, *151*, 118-129.

Nuankaew, W. S., Nuankaew, P., Teeraputon, D., Phanniphong, K., & Bussaman, S. (2019). Perception and attitude toward self-regulated learning of Thailand's students in educational data mining perspective. *International Journal of Emerging Technologies in Learning (Online)*, *14*(9), 34.

Oeda, S., & Hashimoto, G. (2017). Log-Data Clustering Analysis for Dropout Prediction in Beginner Programming Classes. *Procedia computer science*, *112*, 614-621.

Ofori, F., Maina, E., & Gitonga, R. (2020). Using machine learning algorithms to predict studentsâ€™ performance and improve learning outcome: a literature based review. *Journal of Information and Technology*, *4*(1).

Ogowewo, B. O. (2010). Factors Influencing Career Choice Among Secondary School Students: Implications for Career Guidance. *International Journal of Interdisciplinary social sciences*, *5*(2).

Ogunsanya, M., Cho, S., Hudson, A., & Chong, B. (2019). Validation and reliability of CLEQoL. British Journal of Dermatology, 180(6), e238-e238. https://doi.org/10.1111/bjd.17950

Onaolapo, S., & Oyewole, O. (2018). Performance expectancy, effort expectancy, and facilitating conditions as factors influencing smart phones use for mobile learning by postgraduate students of the University of Ibadan, Nigeria. *Interdisciplinary Journal of e-Skills and Lifelong Learning*, *14*(1), 95-115.

Onuba, S. O. (2022). Inluence Of Perception And Attitude Towards Mobile Learning Application (App) On Students'academic Performance In General Studies In Nsukka Local Government Area, Enugu State, Nigeria. *Irish Journal of Educational Practice*, *5*(2).

Ozdemir, D., & Ozturk, F. (2022). The Investigation of Mobile Virtual Reality Application Instructional Content in Geography Education: Academic Achievement, Presence, and Student Interaction. *International Journal of Human–Computer Interaction*, 1-17.

Öztürk, M. (2017). The effect of computer-supported teaching applications on 8th grade students' academic success and knowledge retention in science. *Hasan Ali Yücel Egitim Fakültesi Dergisi*, *14*(2), 59.

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.

Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2021). Classification and prediction of student performance data using various machine learning algorithms. *Materials today: proceedings*.

Pambayun, B., Wirjawan, J. V., Herwinarso, H., Wijaya, A., Untung, B., & Pratidhina, E. (2019). Designing Mobile Learning App to Help High School Students to Learn Simple Harmonic Motion. *International Journal on Social and Education Sciences*, *1*(1), 24-29.

Papadakis, S. (2022). Apps to Promote Computational Thinking and Coding Skills to Young Age Children: A Pedagogical Challenge for the 21st Century Learners. *Educational Process: International Journal (EDUPIJ)*, *11*(1), 7-13.

Patton, M. Q. (1987). *How to use qualitative methods in evaluation* (No. 4). Sage.

Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of applied psychology*, *98*(1), 194.

Pilatso, P., & Chukwuere, J. E. (2022). Investigating the Impact of Fourth-Generation (4G) Mobile Technologies on Student Academic Performance. *Nigerian Journal of Technology*, *41*(3), 547-559.

Qiu, F., Zhang, G., Sheng, X., Jiang, L., Zhu, L., Xiang, Q., ... & Chen, P. K. (2022). Predicting students' performance in e-learning using learning process and behaviour data. *Scientific Reports*, *12*(1), 1-15.

Rajeh, M. T., Abduljabbar, F. H., Alqahtani, S. M., Waly, F. J., Alnaami, I., Aljurayyan, A., & Alzaman, N. (2021). Students' satisfaction and continued intention toward e-learning: a theory-based study. *Medical education online*, *26*(1), 1961348.

Rasheed, H. M. W., He, Y., Khalid, J., Khizar, H. M. U., & Sharif, S. (2022). The relationship between e-learning and academic performance of students. *Journal of Public Affairs*, *22*(3), e2492.

Rathod, U. H., Kulkarni, V., & Saha, U. K. (2022). On the Application of Machine Learning in Savonius Wind Turbine Technology: An Estimation of Turbine Performance Using Artificial Neural Network and Genetic Expression Programming. *Journal of Energy Resources Technology*, *144*(6).

Rehman, R., & Fatima, S. S. (2021). An innovation in Flipped Class Room: A teaching model to facilitate synchronous and asynchronous learning during a pandemic. *Pakistan Journal of Medical Sciences*, *37*(1), 131.

Rohmah, F., Pramono, S. E., & Yusuf, A. (2020). Problem based learning assisted by mobile learning to improve conceptual understanding of primary school students. *Educational Management*, *9*(1), 51-58.

Safitri, H., Hamidah, I., Setiawan, W., & Kaniawati, I. (2021, October). Problem Solving Ability of Distance Education Student on Electrostatic Topic. In *Journal of Physics: Conference Series* (Vol. 2019, No. 1, p. 012048). IOP Publishing.

Sánchez-Otálvaro, L. M., Jiménez-Rivero, Y., Velasquez, R. A., & Botero, J. E. (2022). Development and testing of a mobile application for periodontal diagnosis. *Journal of Clinical and Experimental Dentistry*, *14*(3), e269.

Sandra, L., Lumbangaol, F., & Matsuo, T. (2021). Machine learning algorithm to predict student's performance: a systematic literature review. *TEM Journal*, *10*(4), 1919-1927.

Sari, R., Sumarmi, S., Astina, I., Utomo, D., & Ridhwan, R. (2019). Measuring students scientific learning perception and critical thinking skill using paper-based testing: school and gender differences. *International Journal of Emerging Technologies in Learning (IJET)*, *14*(19), 132-149.

Sattin-Bajaj, C., Jennings, J. L., Corcoran, S. P., Baker-Smith, E. C., & Hailey, C. (2018). Surviving at the street level: How counselors' implementation of school choice policy shapes students' high school destinations. *Sociology of Education*, *91*(1), 46-71.

Sayibu, M., Jianxun, C., Akintunde, T. Y., Hafeez, R. O., Koroma, J., Amosun, T. S., & Shahani, R. (2021). Nexus between students' attitude towards self-learning, Tencent APP usability, mobile-learning, and innovative performance. *Social Sciences & Humanities Open*, *4*(1), 100217.

Sayibu, M., Jianxun, C., Akintunde, T. Y., Hafeez, R. O., Koroma, J., Amosun, T. S., & Shahani, R. (2021). Nexus between students' attitude towards self-learning, Tencent APP usability, mobile-learning, and innovative performance. *Social Sciences & Humanities Open*, *4*(1), 100217.

Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019, March). Student performance prediction and classification using machine learning algorithms. In *Proceedings of the 2019 8th International Conference on Educational and Information Technology* (pp. 7-11).

Sereno, M. A. (2018). *The impact of a personalized learning framework on student achievement*. Edgewood College.

Sevinç, T. A. Ş., & YAVUZ, A. (2021). The Effect of Geometry Teaching Designed with the Developed Mobile Application on the Academic Achievement of the

Students. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, *16*(1), 227-245.

Shah, P. S. I. (2021). I-Taska Management Apps. *Innovation Insights*, 17.

Shams, A., Ajmal, F., & Jumani, N. B. (2021). Effectiveness of Mobile Learning for Academic Progress of MS Education Students: A Case Study of Distance Learning Program at IIUI. *Pakistan Journal of Distance and Online Learning*, *7*(1), 33-46.

Sharma, B. N., Nand, R., Naseem, M., Reddy, E., Narayan, S. S., & Reddy, K. (2018, December). Smart learning in the Pacific: Design of new pedagogical tools. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 573-580). IEEE.

Sharma, B. N., Nand, R., Naseem, M., Reddy, E., Narayan, S. S., & Reddy, K. (2018, December). Smart learning in the Pacific: Design of new pedagogical tools. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 573-580). IEEE.

Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), 4-11.

Siwawetkul, W., & Koraneekij, P. (2020). Effectof 5E instructional model on mobile technology to enhance reasoning ability of lower primary school students. *Kasetsart Journal of Social Sciences*, *41*(1), 40-45.

Sokkhey, P., & Okazaki, T. (2020). Hybrid machine learning algorithms for predicting academic performance. *Int. J. Adv. Comput. Sci. Appl*, *11*(1), 32-41.

Stoyanov, S. R., Hides, L., Kavanagh, D. J., & Wilson, H. (2016). Development and validation of the user version of the Mobile Application Rating Scale (uMARS). *JMIR mHealth and uHealth*, *4*(2), e5849.

Sukhbaatar, O., Usagawa, T., & Choimaa, L. (2019). An artificial neural network based early prediction of failure-prone students in blended learning course. *International Journal of Emerging Technologies in Learning (iJET)*, *14*(19), 77-92.

Sulisworo, D., Yunita, L., & Komalasari, A. (2017). Which mobile learning is more suitable on physics learning in indonesian high school?. *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, *5*(1), 97-104.

Sulla, F., Aquino, A., & Rollo, D. (2022). University Students' Online Learning During COVID-19: The Role of Grit in Academic Performance. *Frontiers in Psychology*, *13*, 825047-825047.

Sun, Z., Lin, C. H., You, J., Shen, H. J., Qi, S., & Luo, L. (2017). Improving the English-speaking skills of young learners through mobile social networking. *Computer assisted language learning*, *30*(3-4), 304-324.

Sun, Z., Lin, C. H., You, J., Shen, H. J., Qi, S., & Luo, L. (2017). Improving the English-speaking skills of young learners through mobile social networking. *Computer assisted language learning*, *30*(3-4), 304-324.

Svennberg, E., Tjong, F., Goette, A., Akoum, N., Di Biase, L., Bordachar, P., ... & Sinner, M. (2022). How to use digital devices to detect and manage arrhythmias: an EHRA practical guide. *Europace*, *24*(6), 979-1005.

Tachinamutu, M., Said, M. N. H. B. M., Abdullah, Z. B., Ali, M. F. B., & Tahir, L. B. M. (2022). The Effect of Using'Microb'Mobile Application in Science Learning in Primary School. *International Journal of Interactive Mobile Technologies*, *16*(2).

Tawafak, R. M., Romli, A. B., Arshah, R. B. A., & Malik, S. I. (2020). Framework design of university communication model (UCOM) to enhance continuous intentions in teaching and e-learning process. *Education and Information Technologies*, *25*(2), 817-843.

Terhorst, Y., Philippi, P., Sander, L. B., Schultchen, D., Paganini, S., Bardus, M., ... & Messner, E. M. (2020). Validation of the mobile application rating scale (MARS). *Plos one*, *15*(11), e0241480.

Thai, K. P., Bang, H. J., & Li, L. (2022). Accelerating early math learning with research-based personalized learning games: A cluster randomized controlled trial. *Journal of Research on Educational Effectiveness*, *15*(1), 28-51.

Troussas, C., Krouska, A., & Sgouropoulou, C. (2020). Collaboration and fuzzy-modeled personalization for mobile game-based learning in higher education. *Computers & Education*, *144*, 103698.

Ugur-Erdogmus, F., & Çakır, R. (2022). Effect of Gamified Mobile Applications and the Role of Player Types on the Achievement of Students. *Journal of Educational Computing Research*, 07356331211065679.

Ugur-Erdogmus, F., & Çakır, R. (2022). Effect of Gamified Mobile Applications and the Role of Player Types on the Achievement of Students. *Journal of Educational Computing Research*, 07356331211065679.

Ugur-Erdogmus, F., & Çakır, R. (2022). Effect of Gamified Mobile Applications and the Role of Player Types on the Achievement of Students. *Journal of Educational Computing Research*, 07356331211065679.

Vialardi, C., Chue, J., Peche, J. P., Alvarado, G., Vinatea, B., Estrella, J., & Ortigosa, Á. (2011). A data mining approach to guide students through the enrollment process based on academic performance. *User modeling and user-adapted interaction*, *21*(1), 217-248.

Vishwakarma, A. (2015). Benefits and challenges of mobile learning in education. In *Promoting active learning through the integration of mobile and ubiquitous technologies* (pp. 24-36). IGI Global.

Voshaar, J., Knipp, M., Loy, T., Zimmermann, J., & Johannsen, F. (2022). The impact of using a mobile app on learning success in accounting education. *Accounting Education*, 1-26.

Wahyu, Y., Supardi, K., Mulu, M., & Narut, Y. F. (2021). The Effect of Problem-Based Approach Assisted with Mobile Augmented Reality toward Students' Science Process Skills And Achievement. *Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram*, *9*(2), 420-427.

Wang, J., & Chia, I. (2022). Engaging students via Nearpod® in synchronous online teaching. *Management Teaching Review*, *7*(3), 245-253.

Weber, W., & Ahn, J. (2021). COVID-19 Conferences: resident perceptions of online synchronous learning environments. *Western Journal of Emergency Medicine*, *22*(1), 115.

Wisniewski, H., & Torous, J. (2020). Digital navigators to implement smartphone and digital tools in care. *Acta Psychiatrica Scandinavica*, *141*(4), 350-355.

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education*, *140*, 103599.

Yadav, B. P., Ghate, S., Harshavardhan, A., Jhansi, G., Kumar, K. S., & Sudarshan, E. (2020, December). Text categorization Performance examination Using Machine Learning Algorithms. In *IOP Conference Series: Materials Science and Engineering* (Vol. 981, No. 2, p. 022044). IOP Publishing.

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, *45*(2), 224-238.

Yıldız, O., Bal, A., GŸlseçen, S., & Kentli, F. D. (2012). A genetic-fuzzy based mathematical model to evaluate the distance education students’ academic performance. *Procedia-Social and Behavioral Sciences*, *55*, 409-418.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). sage.

Zaidi, S. F. H., Osmanaj, V., Ali, O., & Zaidi, S. A. H. (2021). Adoption of mobile technology for mobile learning by university students during COVID-19. *The International Journal of Information and Learning Technology*.

Zheng, L., Long, M., Zhong, L., & Gyasi, J. F. (2022a). The effectiveness of technology-facilitated personalized learning on learning achievements and learning perceptions: a meta-analysis. *Education and Information Technologies*, 1-24.

APPENDICES

Appendix A

.

Questionnaire

Dear student, we are pleased that you evaluated the application (performance scale) after you using so that we can be developed it.

1) What is your specialty?

A) Scientific	B) literary	C) Industrial

Table Axis (1): Role of the application (Performance scale) in facilitating the evaluation process.

NAme	**Sup Axis**	Code	**Statement**
Evaluation process	Performance scale	X1	The app helped me perform some tests and assignments remotely.
		X2	The app helped me identify my weaknesses and defects.
		X3	The app made it easy to know my grades in exams and assignments as soon as they were released.
		X4	The app made it easy to determine my academic level compared to my peers.
		X5	The app made it easy to search for previous exams and solve them myself.

Table Axis (2): Role of the application (performance scale) in improving communication.

Old No.	**Code.**	**Sub Axis**	**Statement**
2	A1	Personalized learning	The teacher focused on the weaknesses I encountered in the curriculum.
4	A2		The app provides instant and quick support to enhance learning.
6	A3		Communication with the advisor was very important in some subjects.
1	A4	Distance learning	The application made communication easy with teachers and asked them.
3	A5		The app made collaborating with my peers easy to understand scientific material.
5	A6		The app made talking with an article advisor easy.
7	A7		The app made it easy to participate actively in any dialogue that revolves around the scientific material.

Table Axis (3): Role of the application (performance scale) in providing scientific content.

Old No.	**Code.**	**Sub Axis**	**Statement**
1	B1	Self-learning	The app made it easy to access the scientific material required for self-study anytime and anywhere.
4	B2		The app provided me with additional books and references to expand my understanding.
5	B3		The app motivated me to search on the web for answers to teacher assignments.
2	B4		The app alerts me to all new duties and lessons downloaded.
3	B5		The app made it easy to submit assignments on time.

Table Axis (4): Satisfaction of learning with the app (performance scale).

Old No.	**Code.**	**Sub Axis**	**Statement**
1	C1	Specialty learning	The app made choosing the right major for me easy.
3	C2		The app made learning is exciting and engaging.
4	C3		The app improved my skills and abilities.
8	C4		The app educational content is structured, graded, and suits my level.
7	C5	Mobile learning	The app made me use my mobile for more useful work.
9	C6		The app motivated me to search for media related to the study subject.
2	C7		The app increased my self-confidence and I was able to present and discuss my ideas.
5	C8		The app organizes my time in navigating between scientific subjects.
6	C9		The app organizes my thoughts and set my priorities.
10	C10		The app motivated me to participate in class discussions.

(5) Do you have some ideas for app development? Record it here:

Appendix B

Import Libraries

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Load Datasets

In [2]:

```
# Year 7
year7_df = pd.read_excel('student data new.xlsx',
sheet_name=0, index_col='Name ')
new_columns = [col.strip() for col in year7_df.columns]
year7_df.columns = new_columns
year7_df = year7_df.iloc[:, :9]
year7_df.head(5)
```

Out[2]:

	8MATH	8PHYSIC	8CHYMESTRY	8BIOLOGY	8COMPUTER	8ENGLISH	8ARABIC	8SOCIAL	8RELIGION
Name									
A1	85.0	79.0	87.0	89.0	90.0	83.0	78.0	90.0	91.0
A2	98.0	96.0	99.0	97.0	98.0	97.0	90.0	98.0	100.0
A3	67.0	70.0	65.0	75.0	75.0	82.0	80.0	87.0	90.0
A4	85.0	87.0	66.0	69.0	90.0	70.0	67.0	73.0	92.0
A5	47.0	43.0	51.0	50.0	54.0	57.0	50.0	55.0	53.0

In [3]:

```
# Year 8
year8_df = pd.read_excel('student data new.xlsx',
sheet_name=1, index_col='Name ')
new_columns = [col.strip() for col in year8_df.columns]
year8_df.columns = new_columns
year8_df = year8_df.iloc[:, :9]
```

```
year8_df.head(5)
```

Out[3]:

Name	9MATH	9PHYSIC	9CHYMESTRY	9BIOLOGY	9COMPUTER	9ENGLISH	9ARABIC	9SOCIAL	9RELIGION
A1	89.0	76.0	90.0	94.0	92.0	87.0	80.0	93.0	96.0
A2	98.0	95.0	97.0	100.0	100.0	95.0	92.0	100.0	99.0
A3	69.0	71.0	68.0	78.0	79.0	87.0	84.0	93.0	90.0
A4	88.0	89.0	65.0	67.0	95.0	72.0	70.0	78.0	90.0
A5	51.0	50.0	68.0	57.0	53.0	50.0	58.0	51.0	57.0

In [4]:

```
# Year 9
year9_df = pd.read_excel('student data new.xlsx',
sheet_name=2, index_col='Name ')
new_columns = [col.strip() for col in year9_df.columns]
year9_df.columns = new_columns
year9_df = year9_df.iloc[:, :9]
year9_df.head(5)
```

Out[4]:

Name	10MATH	10PHYSIC	10CHYMESTRY	10BIOLOGY	10COMPUTER	10ENGLISH	10ARABIC	10SOCIAL	10RELIGION
A1	95.0	79.0	92.0	98.0	98.0	84.0	82.0	96.0	94.0
A2	94.0	100.0	99.0	100.0	100.0	98.0	96.0	100.0	100.0
A3	72.0	70.0	73.0	75.0	81.0	91.0	85.0	96.0	95.0
A4	92.0	90.0	68.0	70.0	93.0	79.0	78.0	83.0	95.0
A5	60.0	54.0	65.0	51.0	59.0	55.0	51.0	50.0	53.0

In [5]:

```
# Year 10
```

```
year10_df = pd.read_excel('student data new.xlsx',
sheet_name=3, index_col='Name ')
new_columns = [col.strip() for col in year10_df.columns]
year10_df.columns = new_columns
year10_df = year10_df.iloc[:, :9]
year10_df.head(5)
```

Out[5]:

	11MATH	11PHYSIC	11CHYMESTRY	11BIOLOGY	11COMPUTER	11ENGLISH	11ARABIC	11SOCIAL	11RELIGION
Name									
A1	88.0	80.0	89.0	93.0	96.0	83.0	83.0	97.0	92.0
A2	88.0	92.0	94.0	90.0	89.0	90.0	83.0	97.0	95.0
A3	66.0	69.0	62.0	61.0	70.0	83.0	80.0	82.0	87.0
A4	85.0	82.0	60.0	62.0	88.0	72.0	71.0	80.0	88.0
A5	51.0	50.0	51.0	53.0	55.0	57.0	53.0	51.0	54.0

In [6]:

```
# Year 11
year11_df = pd.read_excel('student data new.xlsx',
sheet_name=4, index_col='Name ')
new_columns = [col.strip() for col in year11_df.columns]
year11_df.columns = new_columns
year11_df = year11_df.iloc[:, :18]
year11_df.head(5)
```

Out[6]:

	ARABIC	ENGLISH	RELIGION	HISTORIC	عربي تخصص L	تاريخ العرب L	جغرافيا L	رياضيات L	حاسوب L	إدارة المشروعات I	العلوم الصناعية I	الرسم الصناعي I	رياضيات S/I	فيزياء S/I	علوم الأرض S	كيمياء S	أحياء S	حاسوب S
Name																		
A1	80.0	85.0	84.0	85.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	85.0	88.0	89.0	90.0	92.0	95.0
A2	82.0	90.0	88.0	91.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	90.0	92.0	89.0	93.0	87.0	85.0
A3	80.0	78.0	80.0	73.0	70.0	75.0	72.0	70.0	72.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
A4	81.0	80.0	75.0	78.0	NaN	NaN	NaN	NaN	NaN	80.0	82.0	77.0	70.0	65.0	NaN	NaN	NaN	NaN
A5	43.0	33.0	50.0	54.0	38.0	55.0	51.0	53.0	57.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [7]:

```
# Department and Avg
target_df = pd.read_excel('student data new.xlsx', sheet_name=5, index_col='Name ')
new_columns = [col.strip() for col in target_df.columns]
target_df.columns = new_columns
target_df = target_df[['department','average']]
target_df.head(5)
```

Out[7]:

	department	average
Name		
A1	S	85.111111
A2	S	94.200000

	department	average
Name		
A3	L	81.125000
A4	I	79.250000
A5	L	49.500000

EDA - Exploratory Data Analysis

- Let's see the statistics of each dataframe

In [8]:

```
year7_df.describe()
```

Out[8]:

	8MATH	8PHYSIC	8CHYMESTRY	8BIOLOGY	8COMPUTER	8ENGLISH	8ARABIC	8SOCIAL	8RELIGION
count	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000
mean	75.826923	74.362981	71.436699	72.154647	78.310096	76.233974	75.237981	78.104167	82.317308
std	14.178963	14.231688	12.460073	11.575461	11.314807	11.335083	11.049583	12.647540	12.907180
min	42.000000	40.000000	50.000000	50.000000	43.000000	45.000000	43.000000	45.000000	45.000000
25%	63.000000	60.000000	63.000000	65.000000	71.000000	68.000000	68.000000	70.000000	78.000000
50%	80.000000	79.500000	69.000000	70.000000	80.000000	77.000000	77.000000	79.000000	86.000000
75%	88.000000	87.000000	80.000000	81.000000	87.000000	84.250000	83.000000	89.000000	91.000000
max	99.000000	98.000000	99.000000	100.000000	99.000000	98.000000	100.000000	99.000000	100.000000

In [9]:

```
year7_df[(year7_df['8MATH']<50)]
```

Out[9]:

	8MATH	8PHYSIC	8CHYMESTRY	8BIOLOGY	8COMPUTER	8ENGLISH	8ARABIC	8SOCIAL	8RELIGION
Name									
A5	47.0	43.0	51.0	50.0	54.0	57.0	50.0	55.0	53.0
A20	42.0	40.0	55.0	55.0	51.0	53.0	61.0	54.0	59.0
A22	47.0	50.0	53.0	51.0	65.0	84.0	71.0	75.0	85.0
A629	49.0	45.0	51.0	55.0	54.0	53.0	50.0	55.0	57.0
A631	44.0	51.0	50.0	53.0	80.0	72.0	75.0	77.0	75.0
A646	45.0	53.0	52.0	51.0	60.0	84.0	71.0	69.0	79.0

In [10]:

```
year8_df.describe()
```

Out[10]:

	9MATH	9PHYSIC	9CHYMESTRY	9BIOLOGY	9COMPUTER	9ENGLISH	9ARABIC	9SOCIAL	9RELIGION
count	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000
mean	75.541667	73.828526	71.551282	72.206731	77.565705	76.004808	74.958333	77.826923	81.471154
std	14.309160	14.337923	12.718680	11.932067	12.314465	11.922429	11.597425	13.241287	13.548539
min	50.000000	45.000000	50.000000	50.000000	8.000000	48.000000	44.000000	49.000000	50.000000
25%	62.000000	59.000000	62.000000	64.000000	71.000000	68.000000	67.000000	70.000000	76.000000

	9MATH	9PHYSIC	9CHYMESTRY	9BIOLOGY	9COMPUTER	9ENGLISH	9ARABIC	9SOCIAL	9RELIGION
50%	78.000000	77.000000	70.000000	70.000000	80.000000	78.000000	77.000000	79.000000	85.000000
75%	88.000000	87.000000	81.000000	81.000000	87.000000	85.000000	83.000000	89.000000	91.000000
max	100.000000	99.000000	99.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000

- Min 8 ?!!!

In [11]:

```
year8_df[year8_df['9COMPUTER'] < 50]
```

Out[11]:

	9MATH	9PHYSIC	9CHYMESTRY	9BIOLOGY	9COMPUTER	9ENGLISH	9ARABIC	9SOCIAL	9RELIGION
Name									
A323	86.0	84.0	63.0	69.0	8.0	78.0	71.0	75.0	77.0

In [12]:

```
year8_df[(year8_df['9COMPUTER'] < 51)]
```

Out[12]:

	9MATH	9PHYSIC	9CHYMESTRY	9BIOLOGY	9COMPUTER	9ENGLISH	9ARABIC	9SOCIAL	9RELIGION
Name									
A7	52.0	51.0	64.0	60.0	50.0	55.0	68.0	73.0	67.0
A24	50.0	52.0	61.0	57.0	50.0	55.0	44.0	53.0	51.0
A28	57.0	55.0	62.0	54.0	50.0	51.0	52.0	59.0	50.0
A65	53.0	54.0	51.0	56.0	50.0	57.0	52.0	51.0	58.0
A71	56.0	52.0	55.0	57.0	50.0	54.0	51.0	55.0	53.0

Name	9MATH	9PHYSIC	9CHYMESTRY	9BIOLOGY	9COMPUTER	9ENGLISH	9ARABIC	9SOCIAL	9RELIGION
A218	58.0	55.0	57.0	54.0	50.0	53.0	51.0	50.0	56.0
A246	53.0	55.0	51.0	56.0	50.0	54.0	50.0	57.0	54.0
A266	57.0	53.0	59.0	51.0	50.0	54.0	57.0	53.0	59.0
A284	67.0	59.0	54.0	55.0	50.0	52.0	59.0	53.0	50.0
A323	86.0	84.0	63.0	69.0	8.0	78.0	71.0	75.0	77.0
A529	53.0	50.0	57.0	52.0	50.0	51.0	57.0	53.0	55.0
A542	50.0	52.0	53.0	51.0	50.0	56.0	54.0	59.0	57.0
A631	52.0	51.0	64.0	60.0	50.0	55.0	68.0	73.0	67.0
A648	50.0	52.0	61.0	57.0	50.0	55.0	44.0	53.0	51.0
A652	57.0	55.0	62.0	54.0	50.0	51.0	52.0	59.0	50.0
A689	53.0	54.0	51.0	56.0	50.0	57.0	52.0	51.0	58.0
A695	56.0	52.0	55.0	57.0	50.0	54.0	51.0	55.0	53.0
A842	58.0	55.0	57.0	54.0	50.0	53.0	51.0	50.0	56.0
A870	53.0	55.0	51.0	56.0	50.0	54.0	50.0	57.0	54.0

	9MATH	9PHYSIC	9CHYMESTRY	9BIOLOGY	9COMPUTER	9ENGLISH	9ARABIC	9SOCIAL	9RELIGION
Name									
A890	57.0	53.0	59.0	51.0	50.0	54.0	57.0	53.0	59.0
A908	67.0	59.0	54.0	55.0	50.0	52.0	59.0	53.0	50.0
A1153	53.0	50.0	57.0	52.0	50.0	51.0	57.0	53.0	55.0
A1166	50.0	52.0	53.0	51.0	50.0	56.0	54.0	59.0	57.0

- give it minimum = 50

In [13]:

```
year9_df.describe()
```

Out[13]:

	10MATH	10PHYSIC	10CHYMESTRY	10BIOLOGY	10COMPUTER	10ENGLISH	10ARABIC	10SOCIAL	10RELIGION
count	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000
mean	74.848558	72.957532	71.334936	71.825321	77.514423	75.721955	74.649840	77.613782	80.739583
std	14.522012	14.626292	12.696507	12.278116	11.880646	12.031585	11.920883	13.292823	14.052096
min	49.000000	45.000000	50.000000	50.000000	50.000000	43.000000	50.000000	49.000000	50.000000
25%	61.000000	59.000000	61.000000	63.000000	71.000000	67.000000	66.000000	69.750000	74.000000
50%	73.000000	75.000000	70.000000	70.000000	80.000000	78.000000	77.000000	80.000000	85.000000
75%	88.000000	86.000000	80.000000	81.000000	86.000000	85.000000	84.000000	89.000000	92.000000

	10MATH	10PHYSIC	10CHYMESTRY	10BIOLOGY	10COMPUTER	10ENGLISH	10ARABIC	10SOCIAL	10RELIGION
max	100.000000	100.000000	99.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000

In [14]:

```
year9_df[year9_df['10ENGLISH'] < 50]
```

Out[14]:

	10MATH	10PHYSIC	10CHYMESTRY	10BIOLOGY	10COMPUTER	10ENGLISH	10ARABIC	10SOCIAL	10RELIGION
Name									
A202	49.0	53.0	53.0	59.0	56.0	47.0	50.0	52.0	55.0
A382	54.0	60.0	53.0	59.0	53.0	43.0	50.0	60.0	52.0
A826	60.0	50.0	57.0	51.0	60.0	48.0	55.0	52.0	58.0
A1006	56.0	56.0	53.0	59.0	51.0	47.0	50.0	52.0	50.0

In [15]:

```
year10_df.describe()
```

Out[15]:

	11MATH	11PHYSIC	11CHYMESTRY	11BIOLOGY	11COMPUTER	11ENGLISH	11ARABIC	11SOCIAL	11RELIGION
count	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000	1248.000000
mean	72.451923	70.691506	69.141827	69.330128	74.741186	73.726763	72.495192	75.555288	78.240385
std	15.598283	15.340305	13.532932	13.134978	12.873993	12.894981	13.101723	14.096306	14.881405
min	44.000000	26.000000	49.000000	50.000000	43.000000	48.000000	46.000000	49.000000	48.000000

	11MATH	11PHYSIC	11CHYMESTRY	11BIOLOGY	11COMPUTER	11ENGLISH	11ARABIC	11SOCIAL	11RELIGION
25%	58.000000	57.000000	57.000000	60.000000	65.000000	62.000000	61.000000	62.000000	67.000000
50%	71.000000	70.000000	67.000000	67.000000	75.000000	76.000000	75.000000	78.000000	83.000000
75%	87.000000	85.000000	79.000000	79.000000	85.000000	84.000000	83.000000	87.000000	90.000000
max	100.000000	98.000000	99.000000	100.000000	99.000000	99.000000	100.000000	98.000000	100.000000

- 11PHYSIC 26?!!

In [16]:

```
year10_df[year10_df['11PHYSIC'] < 50]
```

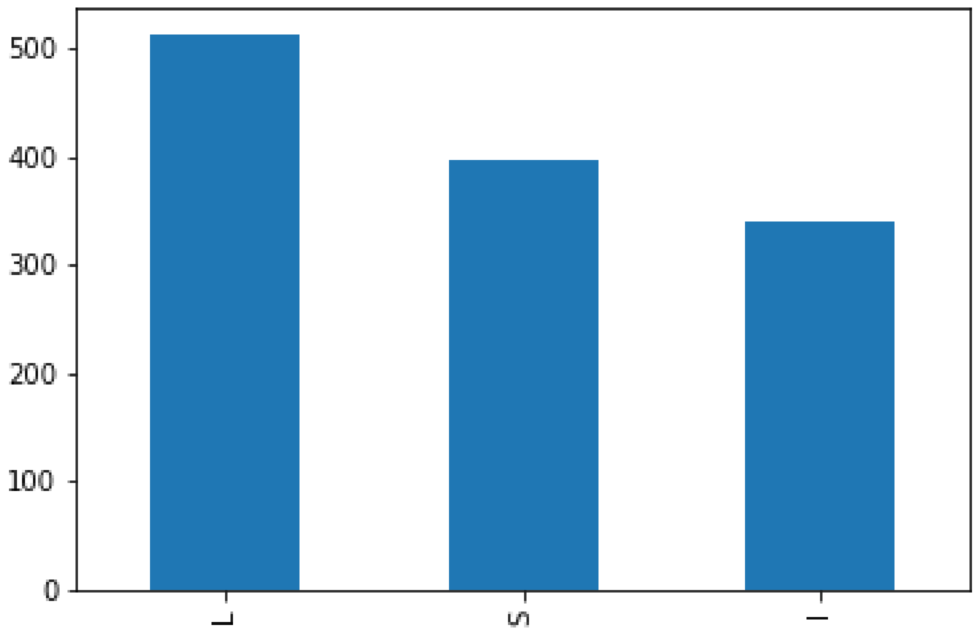
Out[16]:

	11MATH	11PHYSIC	11CHYMESTRY	11BIOLOGY	11COMPUTER	11ENGLISH	11ARABIC	11SOCIAL	11RELIGION
Name									
A249	60.0	47.0	59.0	54.0	55.0	59.0	50.0	60.0	53.0
A455	58.0	44.0	61.0	54.0	54.0	56.0	55.0	51.0	59.0
A529	56.0	39.0	51.0	55.0	47.0	51.0	52.0	54.0	48.0
A546	51.0	42.0	53.0	53.0	60.0	54.0	51.0	52.0	56.0
A584	44.0	41.0	52.0	50.0	53.0	59.0	60.0	52.0	55.0
A585	47.0	41.0	50.0	53.0	50.0	54.0	53.0	51.0	50.0
A589	46.0	35.0	57.0	52.0	60.0	59.0	53.0	61.0	50.0

	11MATH	11PHYSIC	11CHYMESTRY	11BIOLOGY	11COMPUTER	11ENGLISH	11ARABIC	11SOCIAL	11RELIGION
Name									
A605	57.0	26.0	50.0	56.0	52.0	50.0	46.0	55.0	56.0
A608	54.0	47.0	55.0	51.0	56.0	52.0	58.0	54.0	52.0
A873	51.0	45.0	53.0	54.0	58.0	56.0	50.0	51.0	53.0
A1079	51.0	45.0	53.0	54.0	58.0	56.0	50.0	51.0	53.0
A1153	51.0	34.0	55.0	51.0	47.0	51.0	52.0	50.0	49.0
A1170	57.0	49.0	50.0	53.0	57.0	54.0	51.0	55.0	58.0
A1208	44.0	41.0	52.0	50.0	53.0	59.0	60.0	52.0	55.0
A1209	47.0	41.0	50.0	53.0	50.0	54.0	53.0	51.0	50.0
A1213	47.0	31.0	55.0	54.0	51.0	59.0	50.0	52.0	51.0
A1229	57.0	26.0	50.0	56.0	52.0	50.0	46.0	55.0	56.0
A1232	54.0	47.0	55.0	51.0	56.0	52.0	58.0	54.0	52.0

In [17]:

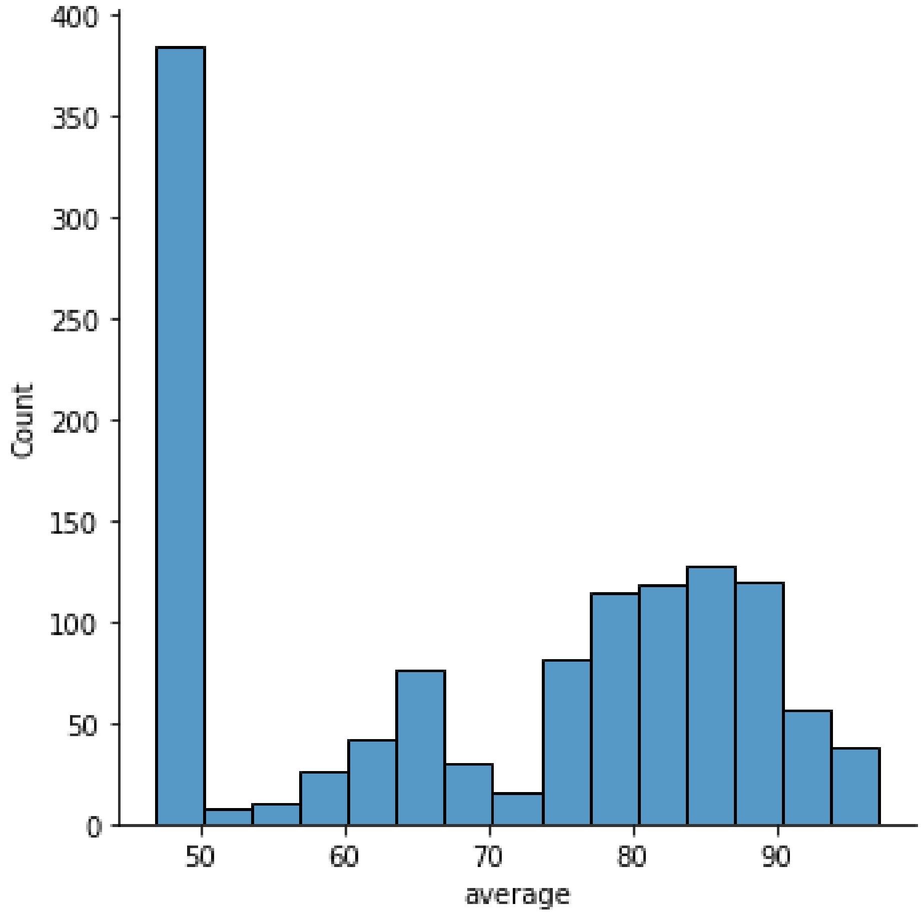
```
# number of records of each department in our data
target_df['department'].value_counts().plot(kind='bar');
```



In [18]:

```
# The distribution of our target "average (expected gpa)"

sns.displot(target_df['average'], bins=15);
```



In [19]:

```
target_df[target_df['average']<50]['average'].value_counts()
```

Out[19]:

```
49.875000    38
49.750000    32
49.625000    28
49.125000    24
49.000000    22
49.250000    22
49.375000    18
49.500000    16
48.625000    10
49.666667    10
48.500000    10
47.750000    10
48.875000     8
48.000000     8
48.750000     8
48.375000     6
49.333333     6
47.875000     6
47.125000     6
47.500000     6
47.625000     6
48.125000     4
47.375000     4
49.777778     4
48.333333     2
48.250000     2
46.875000     2
47.000000     2
48.444444     2
49.888889     2
47.777778     2
Name: average, dtype: int64
```

Cleaning

In [20]:

```
# Giving 9COMPUTER the minimum = 35
year8_df.loc['A323', '9COMPUTER'] = 35
```

In [21]:

```
year8_df[year8_df['9COMPUTER'] < 35]
```

Out[21]:

	9MATH	9PHYSIC	9CHYMESTRY	9BIOLOGY	9COMPUTER	9ENGLISH	9ARABIC	9SOCIAL	9RELIGION
Name									
A323	86.0	84.0	63.0	69.0	8.0	78.0	71.0	75.0	77.0

In [24]:

```
df.head()
```

Out[24]:

	ARABIC	ENGLISH	RELIGION	HISTORIC	عربي تخصص L	تاريخ العرب L	جغرافيا L	رياضيات L	حاسوب L	إدارة المشروعات I	...	علوم الأرض S	كيمياء S	أحياء S	حاسوب S	MATH	PHYSIC	CHYMESTRY	BIOLOGY	COMPUTER	SOCIAL
Name																					
A1	80.75	84.25	93.25	85.0	NaN	NaN	NaN	NaN	NaN	NaN	...	89.0	90.0	92.0	95.0	89.25	78.50	89.50	93.50	94.00	94.00
A2	90.25	95.00	98.50	91.0	NaN	NaN	NaN	NaN	NaN	NaN	...	89.0	93.0	87.0	85.0	94.50	95.75	97.25	96.75	96.75	98.75
A3	82.25	85.75	90.50	73.0	70.0	75.0	72.0	70.0	72.0	NaN	...	NaN	NaN	NaN	NaN	68.50	70.00	67.00	72.25	76.25	89.50
A4	71.50	73.25	91.25	78.0	NaN	NaN	NaN	NaN	NaN	80.0	...	NaN	NaN	NaN	NaN	87.50	87.00	64.75	67.00	91.50	78.50
A5	53.00	54.75	54.25	54.0	38.0	55.0	51.0	53.0	57.0	NaN	...	NaN	NaN	NaN	NaN	52.25	49.25	58.75	52.75	55.25	51.75

5 rows × 24 columns

In [25]:

```
# Adding target dataframe
df = df.merge(target_df, right_index=True, left_index=True)
```

In [26]:

```
df.info()
<class 'pandas.core.frame.DataFrame'>
Index: 1250 entries, A1     to A1250
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   ARABIC             1248 non-null   float64
 1   ENGLISH            1248 non-null   float64
 2   RELIGION           1248 non-null   float64
 3   HISTORIC           1248 non-null   float64
 4   عربي تخصص L        516 non-null    float64
 5   تاريخ العرب L      516 non-null    float64
 6   جغرافيا L          516 non-null    float64
 7   رياضيات L          516 non-null    float64
 8   حاسوب L            516 non-null    float64
 9   إدارة المشروعات I  336 non-null    float64
 10  العلوم الصناعية I  336 non-null    float64
 11  الرسم الصناعي I    336 non-null    float64
 12  رياضيات S/I        732 non-null    float64
 13  فيزياء S/I         732 non-null    float64
 14  علوم الأرض S      396 non-null    float64
 15  كيمياء S           396 non-null    float64
 16  أحياء S            394 non-null    float64
 17  حاسوب S            394 non-null    float64
 18  MATH               1248 non-null   float64
 19  PHYSIC             1248 non-null   float64
 20  CHYMESTRY          1248 non-null   float64
 21  BIOLOGY            1248 non-null   float64
 22  COMPUTER           1248 non-null   float64
 23  SOCIAL             1248 non-null   float64
 24  department         1248 non-null   object
 25  average            1248 non-null   float64
dtypes: float64(25), object(1)
memory usage: 296.0+ KB
```

In [27]:

```
df.columns
```

Out[27]:

```
Index(['ARABIC', 'ENGLISH', 'RELIGION', 'HISTORIC', 'عربي تخصص
L',
       'تاريخ العرب L', 'جغرافيا L', 'رياضيات L', 'حاسوب L',
       'إدارة المشروعات I', 'العلوم الصناعية I', 'الرسم الصناع
ي I',
       'رياضيات S/I', 'فيزياء S/I', 'علوم الأرض S', 'كيمياء S',
'أحياء S',
       'حاسوب S', 'MATH', 'PHYSIC', 'CHYMESTRY', 'BIOLOGY', 'C
OMPUTER',
       'SOCIAL', 'department', 'average'],
      dtype='object')
```

In [28]:

```
# Separate only important features
df = df[['ARABIC', 'ENGLISH', 'RELIGION',
```

```
        'MATH', 'PHYSIC', 'CHYMESTRY', 'BIOLOGY',
        'COMPUTER','SOCIAL', 'department', 'average']]
```

In [29]:

```
# Spliting Science Department
scientific_df = df[df['department']=='S']
scientific_df = scientific_df[['ARABIC', 'ENGLISH',
'RELIGION',
             'MATH', 'PHYSIC', 'CHYMESTRY', 'BIOLOGY',
'COMPUTER', 'SOCIAL',
             'average']]

scientific_df.head()
```

Out[29]:

	ARABIC	ENGLISH	RELIGION	MATH	PHYSIC	CHYMESTRY	BIOLOGY	COMPUTER	SOCIAL	average
Name										
A1	80.75	84.25	93.25	89.25	78.50	89.50	93.50	94.00	94.00	85.111111
A2	90.25	95.00	98.50	94.50	95.75	97.25	96.75	96.75	98.75	94.200000
A9	78.00	83.50	86.00	58.75	52.25	62.00	62.75	73.25	89.25	55.000000
A10	61.50	67.00	76.25	85.50	82.50	58.50	53.25	70.00	62.75	49.666667
A11	72.00	77.00	85.00	86.00	79.25	83.75	79.75	86.25	83.25	71.666667

In [30]:

```
# Spliting Literature Department
literary_df = df[df['department']=='L']
literary_df = literary_df[['ARABIC', 'ENGLISH', 'RELIGION',
             'MATH', 'PHYSIC', 'CHYMESTRY', 'BIOLOGY',
'COMPUTER', 'SOCIAL',
             'average']]

literary_df.head()
```

Out[30]:

	ARABIC	ENGLISH	RELIGION	MATH	PHYSIC	CHYMESTRY	BIOLOGY	COMPUTER	SOCIAL	average
Name										
A3	82.25	85.75	90.50	68.50	70.00	67.00	72.25	76.25	89.50	81.125
A5	53.00	54.75	54.25	52.25	49.25	58.75	52.75	55.25	51.75	49.500
A6	65.00	69.00	85.50	55.75	54.25	57.25	61.25	70.00	79.50	50.000
A8	59.75	60.00	71.25	84.00	81.75	56.25	55.00	71.00	62.50	49.750
A16	87.75	88.00	97.50	62.25	62.00	63.75	66.75	78.75	95.00	81.750

In [31]:

```
# Spliting Industrial Department
industrial_df = df[df['department']=='I']
industrial_df = industrial_df[['ARABIC', 'ENGLISH', 
'RELIGION',
             'MATH', 'PHYSIC', 'CHYMESTRY', 'BIOLOGY', 
'COMPUTER', 'SOCIAL',
             'average']]

industrial_df.head()
```

Out[31]:

	ARABIC	ENGLISH	RELIGION	MATH	PHYSIC	CHYMESTRY	BIOLOGY	COMPUTER	SOCIAL	average
Name										
A4	71.50	73.25	91.25	87.50	87.00	64.75	67.00	91.50	78.5	79.250
A7	77.00	69.25	77.25	54.00	53.50	57.75	55.50	70.50	76.0	62.125
A12	61.75	61.50	85.75	79.75	83.00	58.50	62.75	80.00	70.0	74.125

	ARABIC	ENGLISH	RELIGION	MATH	PHYSIC	CHYMESTRY	BIOLOGY	COMPUTER	SOCIAL	average
Name										
A13	63.25	67.50	84.50	89.50	87.00	69.00	65.25	80.25	70.0	79.875
A20	68.50	62.50	63.25	51.50	48.75	54.75	51.25	56.50	62.5	49.000

Pre-processing

In [32]:

```
# spliting data
from sklearn.model_selection import train_test_split
```

In [33]:

```
# Scientific
x = scientific_df.drop('average', axis=1)
y = scientific_df['average']
x_train_s, x_test_s, y_train_s, y_test_s = train_test_split(x,
y, train_size=0.8, random_state=42)
```

In [34]:

```
# Literary
x = literary_df.drop('average', axis=1)
y = literary_df['average']
x_train_l, x_test_l, y_train_l, y_test_l = train_test_split(x,
y, train_size=0.8, random_state=0)
```

In [35]:

```
# Industrial
x = industrial_df.drop('average', axis=1)
y = industrial_df['average']
x_train_i, x_test_i, y_train_i, y_test_i = train_test_split(x,
y, train_size=0.8, random_state=42)
```

Scientific df

In [38]:

```
# Linear Regression
lr = LinearRegression()
lr.fit(x_train_s, y_train_s)
calculate_accuracies(lr, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.8157778287049504
Testing    R2: 0.8681651966167288

Training   MSE: 39.45692320091661
Testing    MSE: 27.610978447817843
```

In [39]:

```
# Bayesian Ridge Regression
br = BayesianRidge(n_iter=300, normalize=True)
br.fit(x_train_s, y_train_s)
calculate_accuracies(br, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.8151558310624973
Testing    R2: 0.8682390141334553

Training   MSE: 39.59014339388738
Testing    MSE: 27.595518388629376
```

In [40]:

```
# Support Vector Machine Regression
svr = SVR(kernel='linear')
svr.fit(x_train_s, y_train_s)
calculate_accuracies(svr, x_train_s, x_test_s, y_train_s,
y_test_s)
print()

svr = SVR(kernel='rbf')
svr.fit(x_train_s, y_train_s)
calculate_accuracies(svr, x_train_s, x_test_s, y_train_s,
y_test_s)
print()

svr = SVR(kernel='poly')
svr.fit(x_train_s, y_train_s)
calculate_accuracies(svr, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.7959184689890086
Testing    R2: 0.8659682813268429

Training   MSE: 43.71042442513298
Testing    MSE: 28.07109200784928

Training   R2: 0.8420062591059158
Testing    R2: 0.8987999118646041

Training   MSE: 33.839286861401334
Testing    MSE: 21.194960516611577

Training   R2: 0.8626382018234304
Testing    R2: 0.9143982522677386

Training   MSE: 29.420312893350168
Testing    MSE: 17.928103589305486
```

In [41]:

```
# K Nearest Neighbour Regressor
knn = KNeighborsRegressor(n_neighbors=7)
knn.fit(x_train_s, y_train_s)
calculate_accuracies(knn, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.9277062584837957
Testing    R2: 0.9485373364596175

Training   MSE: 15.483959324001509
Testing    MSE: 10.778143990929705
```

In [42]:

```
# Decision Tree Regressor
dt = DecisionTreeRegressor(max_depth=6)
dt.fit(x_train_s, y_train_s)
calculate_accuracies(dt, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.9898226602567123
Testing    R2: 0.9901905230069249

Training   MSE: 2.179794700157963
Testing    MSE: 2.0544594514450534
```

In [43]:

```
# Random Forest Regressor
rf = RandomForestRegressor(n_estimators=120, max_depth=20,
random_state=42)
rf.fit(x_train_s, y_train_s)
calculate_accuracies(rf, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.9982154072034025
Testing    R2: 0.9924327423710692

Training   MSE: 0.38222620233631094
Testing    MSE: 1.5848575788751649
```

In [44]:

```
# Extra Trees Regressor
et = ExtraTreesRegressor(n_estimators=40, max_depth=16,
random_state=42)
et.fit(x_train_s, y_train_s)
calculate_accuracies(et, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.9999991774120106
Testing    R2: 0.9996604730159785

Training   MSE: 0.00017618287144027214
Testing    MSE: 0.07110923669387728
```

In [45]:

```
# Gradient Boosting Regressor
gbr = GradientBoostingRegressor(n_estimators=120,
max_depth=16, criterion='mse', random_state=42)
gbr.fit(x_train_s, y_train_s)
```

```
calculate_accuracies(gbr, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.999999999989572
Testing    R2: 0.9999465324019551

Training   MSE: 2.2334878036452576e-09
Testing    MSE: 0.011198049827423187
```

In [46]:

```
# Stacking Technique
et = ExtraTreesRegressor(n_estimators=40, max_depth=16,
random_state=42)
gbr = GradientBoostingRegressor(n_estimators=120,
max_depth=16, criterion='mse', random_state=42)
rf = RandomForestRegressor(n_estimators=120, max_depth=20,
random_state=42)
regressors = [('et', et), ('gbr', gbr), ('rf', rf)]

from sklearn.ensemble import StackingRegressor
sr = StackingRegressor(estimators=regressors,
final_estimator=LinearRegression())
sr.fit(x_train_s, y_train_s)
calculate_accuracies(sr, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.9998716510726553
Testing    R2: 0.999496842860537

Training   MSE: 0.027489925526119813
Testing    MSE: 0.10537931242018801
```

In [47]:

```
# ANN
model = models.Sequential([
    layers.Dense(10, activation='relu', input_dim = 9),
    layers.Dense(units = 8, activation='relu'),
    layers.Dense(units = 4, activation='relu'),
    layers.Dense(units = 1)
])
```

In [48]:

```
model.compile(optimizer = 'adam', loss = 'mean_squared_error')
```

In [49]:

```
early_stopp = callbacks.EarlyStopping(patience=5,
restore_best_weights=True)
```

In [50]:

```
history = model.fit(x_train_s, y_train_s,
validation_data=(x_test_s, y_test_s),
                    epochs=100, batch_size=125,
callbacks=[early_stopp])
```

```
Epoch 1/100
3/3 [==============================] - 3s 141ms/step - loss: 5
866.2129 - val_loss: 5913.6382
Epoch 2/100
```

```
3/3 [==============================] - 0s 25ms/step - loss: 58
65.7617 - val_loss: 5913.1851
Epoch 3/100
3/3 [==============================] - 0s 22ms/step - loss: 58
65.3125 - val_loss: 5912.7329
Epoch 4/100
3/3 [==============================] - 0s 21ms/step - loss: 58
64.8599 - val_loss: 5912.2798
Epoch 5/100
3/3 [==============================] - 0s 24ms/step - loss: 58
64.4106 - val_loss: 5911.8271
Epoch 6/100
3/3 [==============================] - 0s 20ms/step - loss: 58
63.9580 - val_loss: 5911.3750
Epoch 7/100
3/3 [==============================] - 0s 21ms/step - loss: 58
63.5068 - val_loss: 5910.9224
Epoch 8/100
3/3 [==============================] - 0s 21ms/step - loss: 58
63.0547 - val_loss: 5910.4697
Epoch 9/100
3/3 [==============================] - 0s 24ms/step - loss: 58
62.6050 - val_loss: 5910.0151
Epoch 10/100
3/3 [==============================] - 0s 22ms/step - loss: 58
62.1553 - val_loss: 5909.5615
Epoch 11/100
3/3 [==============================] - 0s 22ms/step - loss: 58
61.7026 - val_loss: 5909.1079
Epoch 12/100
3/3 [==============================] - 0s 19ms/step - loss: 58
61.2495 - val_loss: 5908.6548
Epoch 13/100
3/3 [==============================] - 0s 24ms/step - loss: 58
60.7998 - val_loss: 5908.2007
Epoch 14/100
3/3 [==============================] - 0s 20ms/step - loss: 58
60.3491 - val_loss: 5907.7476
Epoch 15/100
3/3 [==============================] - 0s 27ms/step - loss: 58
59.8965 - val_loss: 5907.2939
Epoch 16/100
3/3 [==============================] - 0s 20ms/step - loss: 58
59.4463 - val_loss: 5906.8408
Epoch 17/100
3/3 [==============================] - 0s 21ms/step - loss: 58
58.9946 - val_loss: 5906.3877
Epoch 18/100
3/3 [==============================] - 0s 22ms/step - loss: 58
58.5444 - val_loss: 5905.9346
Epoch 19/100
```

```
3/3 [==============================] - 0s 23ms/step - loss: 58
58.0908 - val_loss: 5905.4814
Epoch 20/100
3/3 [==============================] - 0s 22ms/step - loss: 58
57.6426 - val_loss: 5905.0283
Epoch 21/100
3/3 [==============================] - 0s 20ms/step - loss: 58
57.1904 - val_loss: 5904.5757
Epoch 22/100
3/3 [==============================] - 0s 23ms/step - loss: 58
56.7397 - val_loss: 5904.1230
Epoch 23/100
3/3 [==============================] - 0s 21ms/step - loss: 58
56.2925 - val_loss: 5903.6709
Epoch 24/100
3/3 [==============================] - 0s 37ms/step - loss: 58
55.8403 - val_loss: 5903.2178
Epoch 25/100
3/3 [==============================] - 0s 24ms/step - loss: 58
55.3867 - val_loss: 5902.7656
Epoch 26/100
3/3 [==============================] - 0s 21ms/step - loss: 58
54.9375 - val_loss: 5902.3135
Epoch 27/100
3/3 [==============================] - 0s 22ms/step - loss: 58
54.4863 - val_loss: 5901.8599
Epoch 28/100
3/3 [==============================] - 0s 20ms/step - loss: 58
54.0371 - val_loss: 5901.4058
Epoch 29/100
3/3 [==============================] - 0s 22ms/step - loss: 58
53.5854 - val_loss: 5900.9536
Epoch 30/100
3/3 [==============================] - 0s 21ms/step - loss: 58
53.1357 - val_loss: 5900.5010
Epoch 31/100
3/3 [==============================] - 0s 20ms/step - loss: 58
52.6846 - val_loss: 5900.0483
Epoch 32/100
3/3 [==============================] - 0s 25ms/step - loss: 58
52.2344 - val_loss: 5899.5967
Epoch 33/100
3/3 [==============================] - 0s 22ms/step - loss: 58
51.7856 - val_loss: 5899.1436
Epoch 34/100
3/3 [==============================] - 0s 24ms/step - loss: 58
51.3354 - val_loss: 5898.6929
Epoch 35/100
3/3 [==============================] - 0s 23ms/step - loss: 58
50.8823 - val_loss: 5898.2412
Epoch 36/100
```

```
3/3 [==============================] - 0s 21ms/step - loss: 58
50.4360 - val_loss: 5897.7881
Epoch 37/100
3/3 [==============================] - 0s 23ms/step - loss: 58
49.9858 - val_loss: 5897.3359
Epoch 38/100
3/3 [==============================] - 0s 22ms/step - loss: 58
49.5356 - val_loss: 5896.8838
Epoch 39/100
3/3 [==============================] - 0s 24ms/step - loss: 58
49.0845 - val_loss: 5896.4326
Epoch 40/100
3/3 [==============================] - 0s 23ms/step - loss: 58
48.6338 - val_loss: 5895.9805
Epoch 41/100
3/3 [==============================] - 0s 24ms/step - loss: 58
48.1851 - val_loss: 5895.5273
Epoch 42/100
3/3 [==============================] - 0s 25ms/step - loss: 58
47.7329 - val_loss: 5895.0757
Epoch 43/100
3/3 [==============================] - 0s 21ms/step - loss: 58
47.2847 - val_loss: 5894.6240
Epoch 44/100
3/3 [==============================] - 0s 26ms/step - loss: 58
46.8340 - val_loss: 5894.1719
Epoch 45/100
3/3 [==============================] - 0s 24ms/step - loss: 58
46.3823 - val_loss: 5893.7197
Epoch 46/100
3/3 [==============================] - 0s 21ms/step - loss: 58
45.9351 - val_loss: 5893.2671
Epoch 47/100
3/3 [==============================] - 0s 22ms/step - loss: 58
45.4839 - val_loss: 5892.8154
Epoch 48/100
3/3 [==============================] - 0s 21ms/step - loss: 58
45.0337 - val_loss: 5892.3643
Epoch 49/100
3/3 [==============================] - 0s 22ms/step - loss: 58
44.5859 - val_loss: 5891.9126
Epoch 50/100
3/3 [==============================] - 0s 23ms/step - loss: 58
44.1353 - val_loss: 5891.4604
Epoch 51/100
3/3 [==============================] - 0s 23ms/step - loss: 58
43.6860 - val_loss: 5891.0098
Epoch 52/100
3/3 [==============================] - 0s 22ms/step - loss: 58
43.2368 - val_loss: 5890.5576
Epoch 53/100
```

```
3/3 [==============================] - 0s 26ms/step - loss: 58
42.7871 - val_loss: 5890.1055
Epoch 54/100
3/3 [==============================] - 0s 26ms/step - loss: 58
42.3398 - val_loss: 5889.6533
Epoch 55/100
3/3 [==============================] - 0s 23ms/step - loss: 58
41.8862 - val_loss: 5889.2017
Epoch 56/100
3/3 [==============================] - 0s 21ms/step - loss: 58
41.4404 - val_loss: 5888.7495
Epoch 57/100
3/3 [==============================] - 0s 23ms/step - loss: 58
40.9868 - val_loss: 5888.2993
Epoch 58/100
3/3 [==============================] - 0s 23ms/step - loss: 58
40.5391 - val_loss: 5887.8472
Epoch 59/100
3/3 [==============================] - 0s 18ms/step - loss: 58
40.0884 - val_loss: 5887.3955
Epoch 60/100
3/3 [==============================] - 0s 24ms/step - loss: 58
39.6396 - val_loss: 5886.9434
Epoch 61/100
3/3 [==============================] - 0s 25ms/step - loss: 58
39.1909 - val_loss: 5886.4922
Epoch 62/100
3/3 [==============================] - 0s 23ms/step - loss: 58
38.7397 - val_loss: 5886.0405
Epoch 63/100
3/3 [==============================] - 0s 25ms/step - loss: 58
38.2910 - val_loss: 5885.5889
Epoch 64/100
3/3 [==============================] - 0s 22ms/step - loss: 58
37.8423 - val_loss: 5885.1382
Epoch 65/100
3/3 [==============================] - 0s 25ms/step - loss: 58
37.3940 - val_loss: 5884.6865
Epoch 66/100
3/3 [==============================] - 0s 22ms/step - loss: 58
36.9438 - val_loss: 5884.2354
Epoch 67/100
3/3 [==============================] - 0s 21ms/step - loss: 58
36.4937 - val_loss: 5883.7842
Epoch 68/100
3/3 [==============================] - 0s 23ms/step - loss: 58
36.0425 - val_loss: 5883.3330
Epoch 69/100
3/3 [==============================] - 0s 22ms/step - loss: 58
35.5986 - val_loss: 5882.8813
Epoch 70/100
```

```
3/3 [==============================] - 0s 22ms/step - loss: 58
35.1450 - val_loss: 5882.4307
Epoch 71/100
3/3 [==============================] - 0s 23ms/step - loss: 58
34.6987 - val_loss: 5881.9785
Epoch 72/100
3/3 [==============================] - 0s 24ms/step - loss: 58
34.2471 - val_loss: 5881.5273
Epoch 73/100
3/3 [==============================] - 0s 21ms/step - loss: 58
33.7993 - val_loss: 5881.0752
Epoch 74/100
3/3 [==============================] - 0s 22ms/step - loss: 58
33.3506 - val_loss: 5880.6250
Epoch 75/100
3/3 [==============================] - 0s 23ms/step - loss: 58
32.9019 - val_loss: 5880.1733
Epoch 76/100
3/3 [==============================] - 0s 26ms/step - loss: 58
32.4512 - val_loss: 5879.7241
Epoch 77/100
3/3 [==============================] - 0s 23ms/step - loss: 58
32.0054 - val_loss: 5879.2725
Epoch 78/100
3/3 [==============================] - 0s 23ms/step - loss: 58
31.5571 - val_loss: 5878.8228
Epoch 79/100
3/3 [==============================] - 0s 23ms/step - loss: 58
31.1064 - val_loss: 5878.3735
Epoch 80/100
3/3 [==============================] - 0s 27ms/step - loss: 58
30.6582 - val_loss: 5877.9229
Epoch 81/100
3/3 [==============================] - 0s 27ms/step - loss: 58
30.2109 - val_loss: 5877.4712
Epoch 82/100
3/3 [==============================] - 0s 23ms/step - loss: 58
29.7603 - val_loss: 5877.0205
Epoch 83/100
3/3 [==============================] - 0s 25ms/step - loss: 58
29.3125 - val_loss: 5876.5693
Epoch 84/100
3/3 [==============================] - 0s 22ms/step - loss: 58
28.8638 - val_loss: 5876.1182
Epoch 85/100
3/3 [==============================] - 0s 24ms/step - loss: 58
28.4141 - val_loss: 5875.6680
Epoch 86/100
3/3 [==============================] - 0s 23ms/step - loss: 58
27.9673 - val_loss: 5875.2163
Epoch 87/100
```

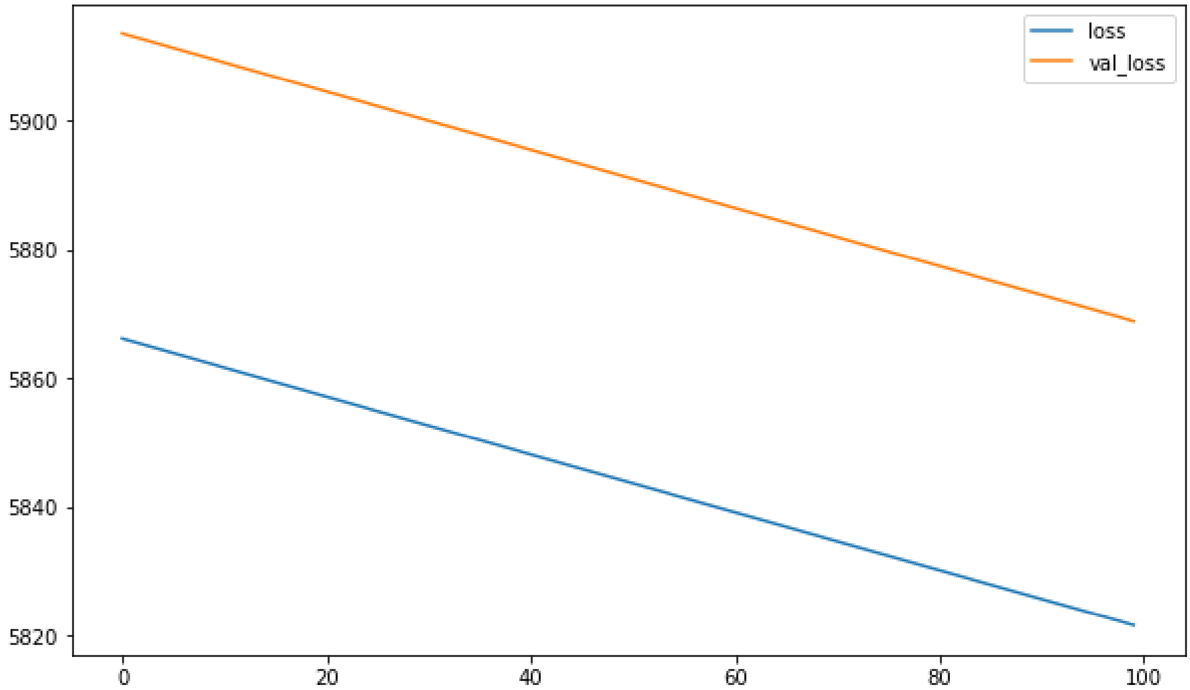
```
3/3 [==============================] - 0s 21ms/step - loss: 58
27.5176 - val_loss: 5874.7666
Epoch 88/100
3/3 [==============================] - 0s 22ms/step - loss: 58
27.0679 - val_loss: 5874.3149
Epoch 89/100
3/3 [==============================] - 0s 22ms/step - loss: 58
26.6201 - val_loss: 5873.8633
Epoch 90/100
3/3 [==============================] - 0s 22ms/step - loss: 58
26.1694 - val_loss: 5873.4131
Epoch 91/100
3/3 [==============================] - 0s 23ms/step - loss: 58
25.7227 - val_loss: 5872.9619
Epoch 92/100
3/3 [==============================] - 0s 21ms/step - loss: 58
25.2725 - val_loss: 5872.5107
Epoch 93/100
3/3 [==============================] - 0s 21ms/step - loss: 58
24.8262 - val_loss: 5872.0605
Epoch 94/100
3/3 [==============================] - 0s 21ms/step - loss: 58
24.3760 - val_loss: 5871.6094
Epoch 95/100
3/3 [==============================] - 0s 20ms/step - loss: 58
23.9272 - val_loss: 5871.1577
Epoch 96/100
3/3 [==============================] - 0s 21ms/step - loss: 58
23.4800 - val_loss: 5870.7080
Epoch 97/100
3/3 [==============================] - 0s 21ms/step - loss: 58
23.0303 - val_loss: 5870.2578
Epoch 98/100
3/3 [==============================] - 0s 25ms/step - loss: 58
22.5806 - val_loss: 5869.8091
Epoch 99/100
3/3 [==============================] - 0s 21ms/step - loss: 58
22.1367 - val_loss: 5869.3589
Epoch 100/100
3/3 [==============================] - 0s 20ms/step - loss: 58
21.6855 - val_loss: 5868.9092
```

In [51]:

```
pd.DataFrame(history.history).plot(figsize=(10, 6))
```

Out[51]:

```
<AxesSubplot:>
```



In [52]:

```
print('Training   R2: ', r2_score(y_train_s,
model.predict(x_train_s)))
print('Testing    R2: ', r2_score(y_test_s,
model.predict(x_test_s)))
print()
print('Training   MSE: ', mean_squared_error(y_train_s,
model.predict(x_train_s)))
print('Testing    MSE: ', mean_squared_error(y_test_s,
model.predict(x_test_s)))
```

```
Training   R2:  -26.17960918100228
Testing    R2:  -27.022422047506122

Training   MSE:  5821.360939059526
Testing    MSE:  5868.909205712267
```

In [53]:

```
# Best model for scientific department is Extra Trees
Regressor
scientific_model = ExtraTreesRegressor(n_estimators=40,
max_depth=16, random_state=42)
scientific_model.fit(x_train_s, y_train_s)
```

Out[53]:

```
ExtraTreesRegressor(max_depth=16, n_estimators=40, random_stat
e=42)
```

In [54]:

```
# Save Model
import pickle
pickle.dump(scientific_model, open('scientific_model.sav',
'wb'))
```

In [ ]:

Literary df

In [55]:

```
# Linear Regression
lr = LinearRegression()
lr.fit(x_train_l, y_train_l)
calculate_accuracies(lr, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.7478220115315022
Testing    R2: 0.7872905930355578

Training   MSE: 71.0140093453123
Testing    MSE: 58.33806736330305
```

In [56]:

```
# Bayesian Ridge Regression
br = BayesianRidge(n_iter=300, normalize=True)
br.fit(x_train_l, y_train_l)
calculate_accuracies(br, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.74746910249076
Testing    R2: 0.7888230670044328

Training   MSE: 71.11338949371267
Testing    MSE: 57.91776827590214
```

In [57]:

```
# Support Vector Machine Regression
svr = SVR(kernel='linear')
svr.fit(x_train_l, y_train_l)
calculate_accuracies(svr, x_train_l, x_test_l, y_train_l,
y_test_l)
print()

svr = SVR(kernel='rbf')
svr.fit(x_train_l, y_train_l)
calculate_accuracies(svr, x_train_l, x_test_l, y_train_l,
y_test_l)
print()

svr = SVR(kernel='poly')
svr.fit(x_train_l, y_train_l)
calculate_accuracies(svr, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.7277272465008384
Testing    R2: 0.8051452760033394

Training   MSE: 76.67274998459561
Testing    MSE: 53.44120966157014

Training   R2: 0.8253280039554438
Testing    R2: 0.8808873891379246
```

```
Training   MSE: 49.18811048816814
Testing    MSE: 32.66804047576635

Training   R2: 0.8274873320838395
Testing    R2: 0.874390825645711

Training   MSE: 48.580037797840355
Testing    MSE: 34.449799750296634
```

In [58]:

```
# K Nearest Neighbour Regressor
knn = KNeighborsRegressor(n_neighbors=7)
knn.fit(x_train_l, y_train_l)
calculate_accuracies(knn, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.92617283634484
Testing    R2: 0.8992809689796525

Training   MSE: 20.789930642183528
Testing    MSE: 27.623383940955026
```

In [59]:

```
# Decision Tree Regressor
dt = DecisionTreeRegressor(max_depth=6)
dt.fit(x_train_l, y_train_l)
calculate_accuracies(dt, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9881008414928604
Testing    R2: 0.9833833616186944

Training   MSE: 3.3508354894857306
Testing    MSE: 4.5573093502268085
```

In [60]:

```
# Random Forest Regressor
rf = RandomForestRegressor(n_estimators=120, max_depth=20,
random_state=42)
rf.fit(x_train_l, y_train_l)
calculate_accuracies(rf, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9980138337621363
Testing    R2: 0.9945352508810357

Training   MSE: 0.5593098296706511
Testing    MSE: 1.4987719949733245
```

In [61]:

```
# Extra Trees Regressor
et = ExtraTreesRegressor(n_estimators=35, max_depth=17,
random_state=42)
et.fit(x_train_l, y_train_l)
calculate_accuracies(et, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.999998612491685
Testing    R2: 0.9993824629101646
```

```
Training   MSE: 0.00039072612582569414
Testing    MSE: 0.1693668411767668
```

In [62]:

```
# Gradient Boosting Regressor
gbr = GradientBoostingRegressor(n_estimators=120,
max_depth=16, criterion='mse', random_state=42)
gbr.fit(x_train_l, y_train_l)
calculate_accuracies(gbr, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9999987790269449
Testing    R2: 0.9958204505146635

Training   MSE: 0.0003438293424260309
Testing    MSE: 1.1462908147949347
```

In [63]:

```
# Stacking Technique
et = ExtraTreesRegressor(n_estimators=35, max_depth=17,
random_state=42)
gbr = GradientBoostingRegressor(n_estimators=120,
max_depth=16, criterion='mse', random_state=42)
rf = RandomForestRegressor(n_estimators=120, max_depth=20,
random_state=42)
regressors = [('et', et), ('gbr', gbr), ('rf', rf)]

from sklearn.ensemble import StackingRegressor
sr = StackingRegressor(estimators=regressors,
final_estimator=LinearRegression())
sr.fit(x_train_l, y_train_l)
calculate_accuracies(sr, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9999076027959456
Testing    R2: 0.9991463572990018

Training   MSE: 0.026019304666720283
Testing    MSE: 0.23412159389522597
```

In [64]:

```
# ANN
model = models.Sequential([
    layers.Dense(10, activation='relu', input_dim = 9),
    layers.Dense(units = 8, activation='relu'),
    layers.Dense(units = 4, activation='relu'),
    layers.Dense(units = 1)
])
```

In [65]:

```
model.compile(optimizer = 'adam', loss = 'mean_squared_error')
```

In [66]:

```
early_stopp = callbacks.EarlyStopping(patience=5,
restore_best_weights=True)
```

In [67]:

```
history = model.fit(x_train_l, y_train_l,
validation_data=(x_test_l, y_test_l),
                    epochs=100, batch_size=125,
callbacks=[early_stopp])
```

```
Epoch 1/100
4/4 [==============================] - 1s 57ms/step - loss: 13
136.6611 - val_loss: 10814.4268
Epoch 2/100
4/4 [==============================] - 0s 14ms/step - loss: 10
892.4375 - val_loss: 9002.7031
Epoch 3/100
4/4 [==============================] - 0s 16ms/step - loss: 91
22.8359 - val_loss: 7682.1006
Epoch 4/100
4/4 [==============================] - 0s 16ms/step - loss: 78
48.2993 - val_loss: 6725.4805
Epoch 5/100
4/4 [==============================] - 0s 14ms/step - loss: 69
64.8687 - val_loss: 6165.8491
Epoch 6/100
4/4 [==============================] - 0s 14ms/step - loss: 64
22.4751 - val_loss: 5705.7002
Epoch 7/100
4/4 [==============================] - 0s 16ms/step - loss: 59
49.7241 - val_loss: 5290.7856
Epoch 8/100
4/4 [==============================] - 0s 14ms/step - loss: 55
19.9009 - val_loss: 4911.2378
Epoch 9/100
4/4 [==============================] - 0s 17ms/step - loss: 51
25.3794 - val_loss: 4559.8276
Epoch 10/100
4/4 [==============================] - 0s 14ms/step - loss: 47
64.3086 - val_loss: 4298.7329
Epoch 11/100
4/4 [==============================] - 0s 14ms/step - loss: 45
37.4526 - val_loss: 4203.8613
Epoch 12/100
4/4 [==============================] - 0s 14ms/step - loss: 44
73.4204 - val_loss: 4190.4404
Epoch 13/100
4/4 [==============================] - 0s 14ms/step - loss: 44
63.6118 - val_loss: 4186.7495
Epoch 14/100
4/4 [==============================] - 0s 14ms/step - loss: 44
61.2896 - val_loss: 4186.3340
Epoch 15/100
4/4 [==============================] - 0s 17ms/step - loss: 44
60.8589 - val_loss: 4185.9185
Epoch 16/100
4/4 [==============================] - 0s 15ms/step - loss: 44
60.4272 - val_loss: 4185.5010
```

```
Epoch 17/100
4/4 [==============================] - 0s 15ms/step - loss: 44
59.9956 - val_loss: 4185.0806
Epoch 18/100
4/4 [==============================] - 0s 15ms/step - loss: 44
59.5605 - val_loss: 4184.6572
Epoch 19/100
4/4 [==============================] - 0s 17ms/step - loss: 44
59.1221 - val_loss: 4184.2334
Epoch 20/100
4/4 [==============================] - 0s 15ms/step - loss: 44
58.6846 - val_loss: 4183.8052
Epoch 21/100
4/4 [==============================] - 0s 15ms/step - loss: 44
58.2393 - val_loss: 4183.3745
Epoch 22/100
4/4 [==============================] - 0s 15ms/step - loss: 44
57.7920 - val_loss: 4182.9414
Epoch 23/100
4/4 [==============================] - 0s 15ms/step - loss: 44
57.3472 - val_loss: 4182.5049
Epoch 24/100
4/4 [==============================] - 0s 15ms/step - loss: 44
56.8940 - val_loss: 4182.0659
Epoch 25/100
4/4 [==============================] - 0s 16ms/step - loss: 44
56.4399 - val_loss: 4181.6255
Epoch 26/100
4/4 [==============================] - 0s 13ms/step - loss: 44
55.9824 - val_loss: 4181.1826
Epoch 27/100
4/4 [==============================] - 0s 18ms/step - loss: 44
55.5254 - val_loss: 4180.7368
Epoch 28/100
4/4 [==============================] - 0s 14ms/step - loss: 44
55.0635 - val_loss: 4180.2896
Epoch 29/100
4/4 [==============================] - 0s 15ms/step - loss: 44
54.6001 - val_loss: 4179.8408
Epoch 30/100
4/4 [==============================] - 0s 14ms/step - loss: 44
54.1362 - val_loss: 4179.3911
Epoch 31/100
4/4 [==============================] - 0s 13ms/step - loss: 44
53.6714 - val_loss: 4178.9399
Epoch 32/100
4/4 [==============================] - 0s 16ms/step - loss: 44
53.2036 - val_loss: 4178.4878
Epoch 33/100
4/4 [==============================] - 0s 15ms/step - loss: 44
52.7363 - val_loss: 4178.0342
Epoch 34/100
```

```
4/4 [==============================] - 0s 15ms/step - loss: 44
52.2686 - val_loss: 4177.5786
Epoch 35/100
4/4 [==============================] - 0s 14ms/step - loss: 44
51.7979 - val_loss: 4177.1221
Epoch 36/100
4/4 [==============================] - 0s 13ms/step - loss: 44
51.3257 - val_loss: 4176.6655
Epoch 37/100
4/4 [==============================] - 0s 14ms/step - loss: 44
50.8535 - val_loss: 4176.2085
Epoch 38/100
4/4 [==============================] - 0s 14ms/step - loss: 44
50.3828 - val_loss: 4175.7510
Epoch 39/100
4/4 [==============================] - 0s 12ms/step - loss: 44
49.9058 - val_loss: 4175.2935
Epoch 40/100
4/4 [==============================] - 0s 15ms/step - loss: 44
49.4336 - val_loss: 4174.8345
Epoch 41/100
4/4 [==============================] - 0s 15ms/step - loss: 44
48.9565 - val_loss: 4174.3730
Epoch 42/100
4/4 [==============================] - 0s 13ms/step - loss: 44
48.4814 - val_loss: 4173.9102
Epoch 43/100
4/4 [==============================] - 0s 14ms/step - loss: 44
48.0029 - val_loss: 4173.4448
Epoch 44/100
4/4 [==============================] - 0s 15ms/step - loss: 44
47.5200 - val_loss: 4172.9795
Epoch 45/100
4/4 [==============================] - 0s 14ms/step - loss: 44
47.0415 - val_loss: 4172.5132
Epoch 46/100
4/4 [==============================] - 0s 15ms/step - loss: 44
46.5576 - val_loss: 4172.0483
Epoch 47/100
4/4 [==============================] - 0s 15ms/step - loss: 44
46.0776 - val_loss: 4171.5820
Epoch 48/100
4/4 [==============================] - 0s 16ms/step - loss: 44
45.5923 - val_loss: 4171.1157
Epoch 49/100
4/4 [==============================] - 0s 16ms/step - loss: 44
45.1099 - val_loss: 4170.6470
Epoch 50/100
4/4 [==============================] - 0s 16ms/step - loss: 44
44.6284 - val_loss: 4170.1772
Epoch 51/100
```

```
4/4 [==============================] - 0s 14ms/step - loss: 44
44.1421 - val_loss: 4169.7090
Epoch 52/100
4/4 [==============================] - 0s 15ms/step - loss: 44
43.6548 - val_loss: 4169.2397
Epoch 53/100
4/4 [==============================] - 0s 15ms/step - loss: 44
43.1689 - val_loss: 4168.7676
Epoch 54/100
4/4 [==============================] - 0s 16ms/step - loss: 44
42.6836 - val_loss: 4168.2930
Epoch 55/100
4/4 [==============================] - 0s 17ms/step - loss: 44
42.1924 - val_loss: 4167.8198
Epoch 56/100
4/4 [==============================] - 0s 16ms/step - loss: 44
41.7012 - val_loss: 4167.3467
Epoch 57/100
4/4 [==============================] - 0s 28ms/step - loss: 44
41.2158 - val_loss: 4166.8726
Epoch 58/100
4/4 [==============================] - 0s 15ms/step - loss: 44
40.7251 - val_loss: 4166.4009
Epoch 59/100
4/4 [==============================] - 0s 16ms/step - loss: 44
40.2358 - val_loss: 4165.9277
Epoch 60/100
4/4 [==============================] - 0s 16ms/step - loss: 44
39.7471 - val_loss: 4165.4551
Epoch 61/100
4/4 [==============================] - 0s 15ms/step - loss: 44
39.2588 - val_loss: 4164.9819
Epoch 62/100
4/4 [==============================] - 0s 16ms/step - loss: 44
38.7690 - val_loss: 4164.5078
Epoch 63/100
4/4 [==============================] - 0s 15ms/step - loss: 44
38.2808 - val_loss: 4164.0342
Epoch 64/100
4/4 [==============================] - 0s 17ms/step - loss: 44
37.7905 - val_loss: 4163.5601
Epoch 65/100
4/4 [==============================] - 0s 18ms/step - loss: 44
37.2998 - val_loss: 4163.0859
Epoch 66/100
4/4 [==============================] - 0s 17ms/step - loss: 44
36.8105 - val_loss: 4162.6123
Epoch 67/100
4/4 [==============================] - 0s 16ms/step - loss: 44
36.3208 - val_loss: 4162.1372
Epoch 68/100
```

```
4/4 [==============================] - 0s 15ms/step - loss: 44
35.8296 - val_loss: 4161.6606
Epoch 69/100
4/4 [==============================] - 0s 16ms/step - loss: 44
35.3354 - val_loss: 4161.1851
Epoch 70/100
4/4 [==============================] - 0s 15ms/step - loss: 44
34.8452 - val_loss: 4160.7075
Epoch 71/100
4/4 [==============================] - 0s 18ms/step - loss: 44
34.3511 - val_loss: 4160.2300
Epoch 72/100
4/4 [==============================] - 0s 17ms/step - loss: 44
33.8589 - val_loss: 4159.7520
Epoch 73/100
4/4 [==============================] - 0s 17ms/step - loss: 44
33.3608 - val_loss: 4159.2739
Epoch 74/100
4/4 [==============================] - 0s 16ms/step - loss: 44
32.8682 - val_loss: 4158.7939
Epoch 75/100
4/4 [==============================] - 0s 16ms/step - loss: 44
32.3730 - val_loss: 4158.3149
Epoch 76/100
4/4 [==============================] - 0s 16ms/step - loss: 44
31.8770 - val_loss: 4157.8369
Epoch 77/100
4/4 [==============================] - 0s 16ms/step - loss: 44
31.3843 - val_loss: 4157.3608
Epoch 78/100
4/4 [==============================] - 0s 17ms/step - loss: 44
30.8926 - val_loss: 4156.8848
Epoch 79/100
4/4 [==============================] - 0s 16ms/step - loss: 44
30.3984 - val_loss: 4156.4072
Epoch 80/100
4/4 [==============================] - 0s 14ms/step - loss: 44
29.9038 - val_loss: 4155.9302
Epoch 81/100
4/4 [==============================] - 0s 14ms/step - loss: 44
29.4097 - val_loss: 4155.4507
Epoch 82/100
4/4 [==============================] - 0s 15ms/step - loss: 44
28.9155 - val_loss: 4154.9692
Epoch 83/100
4/4 [==============================] - 0s 14ms/step - loss: 44
28.4160 - val_loss: 4154.4893
Epoch 84/100
4/4 [==============================] - 0s 14ms/step - loss: 44
27.9189 - val_loss: 4154.0073
Epoch 85/100
```

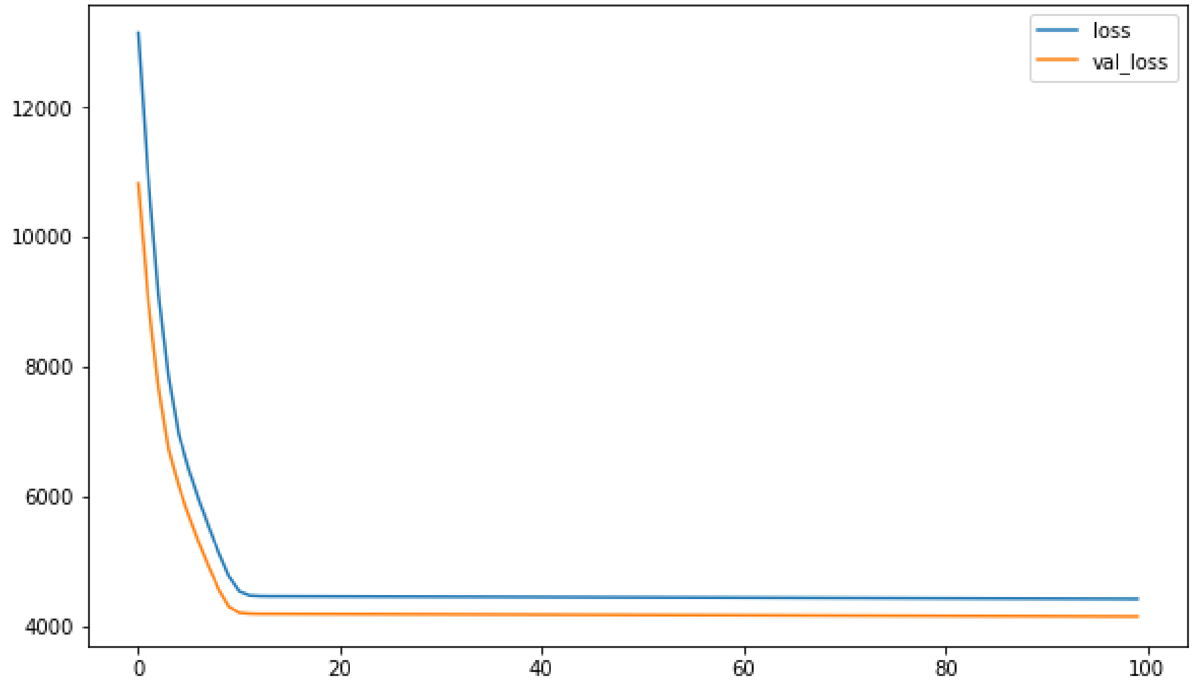
```
4/4 [==============================] - 0s 14ms/step - loss: 44
27.4219 - val_loss: 4153.5254
Epoch 86/100
4/4 [==============================] - 0s 15ms/step - loss: 44
26.9253 - val_loss: 4153.0415
Epoch 87/100
4/4 [==============================] - 0s 15ms/step - loss: 44
26.4229 - val_loss: 4152.5591
Epoch 88/100
4/4 [==============================] - 0s 14ms/step - loss: 44
25.9238 - val_loss: 4152.0747
Epoch 89/100
4/4 [==============================] - 0s 13ms/step - loss: 44
25.4253 - val_loss: 4151.5903
Epoch 90/100
4/4 [==============================] - 0s 16ms/step - loss: 44
24.9238 - val_loss: 4151.1060
Epoch 91/100
4/4 [==============================] - 0s 16ms/step - loss: 44
24.4243 - val_loss: 4150.6255
Epoch 92/100
4/4 [==============================] - 0s 14ms/step - loss: 44
23.9302 - val_loss: 4150.1436
Epoch 93/100
4/4 [==============================] - 0s 13ms/step - loss: 44
23.4297 - val_loss: 4149.6646
Epoch 94/100
4/4 [==============================] - 0s 14ms/step - loss: 44
22.9341 - val_loss: 4149.1860
Epoch 95/100
4/4 [==============================] - 0s 15ms/step - loss: 44
22.4390 - val_loss: 4148.7065
Epoch 96/100
4/4 [==============================] - 0s 16ms/step - loss: 44
21.9414 - val_loss: 4148.2241
Epoch 97/100
4/4 [==============================] - 0s 17ms/step - loss: 44
21.4438 - val_loss: 4147.7412
Epoch 98/100
4/4 [==============================] - 0s 15ms/step - loss: 44
20.9463 - val_loss: 4147.2593
Epoch 99/100
4/4 [==============================] - 0s 15ms/step - loss: 44
20.4482 - val_loss: 4146.7778
Epoch 100/100
4/4 [==============================] - 0s 14ms/step - loss: 44
19.9482 - val_loss: 4146.2954
```

In [68]:

```
pd.DataFrame(history.history).plot(figsize=(10, 6))
```

Out[68]:

```
<AxesSubplot:>
```



In [69]:

```
print('Training   R2: ', r2_score(y_train_l,
model.predict(x_train_l)))
print('Testing    R2: ', r2_score(y_test_l,
model.predict(x_test_l)))
print()
print('Training   MSE: ', mean_squared_error(y_train_l,
model.predict(x_train_l)))
print('Testing    MSE: ', mean_squared_error(y_test_l,
model.predict(x_test_l)))
```

```
Training   R2:  -14.694435218431137
Testing    R2:  -14.118019196847348

Training   MSE:  4419.595762658301
Testing    MSE:  4146.2953373417195
```

In [70]:

```
# Best model for scientific department is Extra Trees
Regressor
literary_model = ExtraTreesRegressor(n_estimators=35,
max_depth=17, random_state=42)
literary_model.fit(x_train_l, y_train_l)
```

Out[70]:

```
ExtraTreesRegressor(max_depth=17, n_estimators=35, random_stat
e=42)
```

In [71]:

```
# Save Model
import pickle
pickle.dump(literary_model, open('literary_model.sav', 'wb'))
```

In [ ]:

Industrial df

In [72]:

```
# Linear Regression
lr = LinearRegression()
lr.fit(x_train_i, y_train_i)
calculate_accuracies(lr, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.7639030638455095
Testing    R2: 0.7359899790513678

Training   MSE: 53.50180606704221
Testing    MSE: 41.059621095769124
```

In [73]:

```
# Bayesian Ridge Regression
br = BayesianRidge(n_iter=300, normalize=True)
br.fit(x_train_i, y_train_i)
calculate_accuracies(br, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.7625287123292441
Testing    R2: 0.725747287103448

Training   MSE: 53.813247161911235
Testing    MSE: 42.65259491119902
```

In [74]:

```
# Support Vector Machine Regression
svr = SVR(kernel='linear')
svr.fit(x_train_i, y_train_i)
calculate_accuracies(svr, x_train_i, x_test_i, y_train_i,
y_test_i)
print()

svr = SVR(kernel='rbf')
svr.fit(x_train_i, y_train_i)
calculate_accuracies(svr, x_train_i, x_test_i, y_train_i,
y_test_i)
print()

svr = SVR(kernel='poly')
svr.fit(x_train_i, y_train_i)
calculate_accuracies(svr, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.7343443832377701
Testing    R2: 0.7124225619232052

Training   MSE: 60.200083576404296
Testing    MSE: 44.72489567137537

Training   R2: 0.8848764544854175
Testing    R2: 0.7787330406994488
```

```
Training   MSE: 26.08808782610011
Testing    MSE: 34.412093439670116

Training   R2: 0.8167573895344026
Testing    R2: 0.7737800851904137

Training   MSE: 41.52451432886793
Testing    MSE: 35.182391763099105
```

In [75]:

```
# K Nearest Neighbour Regressor
knn = KNeighborsRegressor(n_neighbors=7)
knn.fit(x_train_i, y_train_i)
calculate_accuracies(knn, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.962618057478594
Testing    R2: 0.8365168594419065

Training   MSE: 8.471102894282712
Testing    MSE: 25.425382653061227
```

In [76]:

```
# Decision Tree Regressor
dt = DecisionTreeRegressor(max_depth=6)
dt.fit(x_train_i, y_train_i)
calculate_accuracies(dt, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.9891348240954182
Testing    R2: 0.9244548707237283

Training   MSE: 2.4621519601205537
Testing    MSE: 11.74900245289602
```

In [77]:

```
# Random Forest Regressor
rf = RandomForestRegressor(n_estimators=100, max_depth=20,
random_state=42)
rf.fit(x_train_i, y_train_i)
calculate_accuracies(rf, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.9975845958075834
Testing    R2: 0.9440684935640412

Training   MSE: 0.5473535098805145
Testing    MSE: 8.698633685661767
```

In [78]:

```
# Extra Trees Regressor
et = ExtraTreesRegressor(n_estimators=50, max_depth=17,
random_state=42)
et.fit(x_train_i, y_train_i)
calculate_accuracies(et, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.9999999955722503
Testing    R2: 0.9681369628899654
```

```
Training   MSE: 1.0033700980387839e-06
Testing    MSE: 4.955433986928101
```

In [79]:

```
# Gradient Boosting Regressor
gbr = GradientBoostingRegressor(n_estimators=25, max_depth=20, 
criterion='mse', random_state=42)
gbr.fit(x_train_i, y_train_i)
calculate_accuracies(gbr, x_train_i, x_test_i, y_train_i, 
y_test_i)
```

```
Training   R2: 0.9948462247926799
Testing    R2: 0.9334023880654079

Training   MSE: 1.1678943663832355
Testing    MSE: 10.35745803167624
```

In [80]:

```
# Stacking Technique
et = ExtraTreesRegressor(n_estimators=50, max_depth=17, 
random_state=42)
gbr = GradientBoostingRegressor(n_estimators=25, max_depth=20, 
criterion='mse', random_state=42)
rf = RandomForestRegressor(n_estimators=100, max_depth=20, 
random_state=42)
regressors = [('et', et), ('gbr', gbr), ('rf', rf)]

from sklearn.ensemble import StackingRegressor
sr = StackingRegressor(estimators=regressors, 
final_estimator=LinearRegression())
sr.fit(x_train_i, y_train_i)
calculate_accuracies(sr, x_train_i, x_test_i, y_train_i, 
y_test_i)
```

```
Training   R2: 0.9999593724151156
Testing    R2: 0.9670398021340392

Training   MSE: 0.009206596251780363
Testing    MSE: 5.1260676801402285
```

In [81]:

```
# ANN
model = models.Sequential([
    layers.Dense(10, activation='relu', input_dim = 9),
    layers.Dense(units = 8, activation='relu'),
    layers.Dense(units = 4, activation='relu'),
    layers.Dense(units = 1)
])
```

In [82]:

```
model.compile(optimizer = 'adam', loss = 'mean_squared_error')
```

In [83]:

```
early_stopp = callbacks.EarlyStopping(patience=5, 
restore_best_weights=True)
```

In [84]:

```
history = model.fit(x_train_i, y_train_i,
validation_data=(x_test_i, y_test_i),
                    epochs=100, batch_size=125,
callbacks=[early_stopp])
```

```
Epoch 1/100
3/3 [==============================] - 1s 86ms/step - loss: 64
84.8325 - val_loss: 6061.2671
Epoch 2/100
3/3 [==============================] - 0s 19ms/step - loss: 57
51.2432 - val_loss: 5344.2305
Epoch 3/100
3/3 [==============================] - 0s 22ms/step - loss: 50
60.2710 - val_loss: 4749.9375
Epoch 4/100
3/3 [==============================] - 0s 21ms/step - loss: 44
95.0962 - val_loss: 4281.9180
Epoch 5/100
3/3 [==============================] - 0s 21ms/step - loss: 40
40.9688 - val_loss: 3813.5276
Epoch 6/100
3/3 [==============================] - 0s 23ms/step - loss: 35
58.8098 - val_loss: 3309.5396
Epoch 7/100
3/3 [==============================] - 0s 21ms/step - loss: 30
66.8928 - val_loss: 2815.8735
Epoch 8/100
3/3 [==============================] - 0s 22ms/step - loss: 25
94.0930 - val_loss: 2353.5188
Epoch 9/100
3/3 [==============================] - 0s 22ms/step - loss: 21
52.5732 - val_loss: 1934.8745
Epoch 10/100
3/3 [==============================] - 0s 22ms/step - loss: 17
57.1669 - val_loss: 1571.4828
Epoch 11/100
3/3 [==============================] - 0s 23ms/step - loss: 14
25.0161 - val_loss: 1285.2365
Epoch 12/100
3/3 [==============================] - 0s 20ms/step - loss: 11
63.2186 - val_loss: 1048.7931
Epoch 13/100
3/3 [==============================] - 0s 22ms/step - loss: 94
4.4349 - val_loss: 847.2625
Epoch 14/100
3/3 [==============================] - 0s 21ms/step - loss: 75
7.9783 - val_loss: 672.5178
Epoch 15/100
3/3 [==============================] - 0s 23ms/step - loss: 59
5.1105 - val_loss: 524.4380
Epoch 16/100
3/3 [==============================] - 0s 22ms/step - loss: 46
0.1856 - val_loss: 402.5860
```

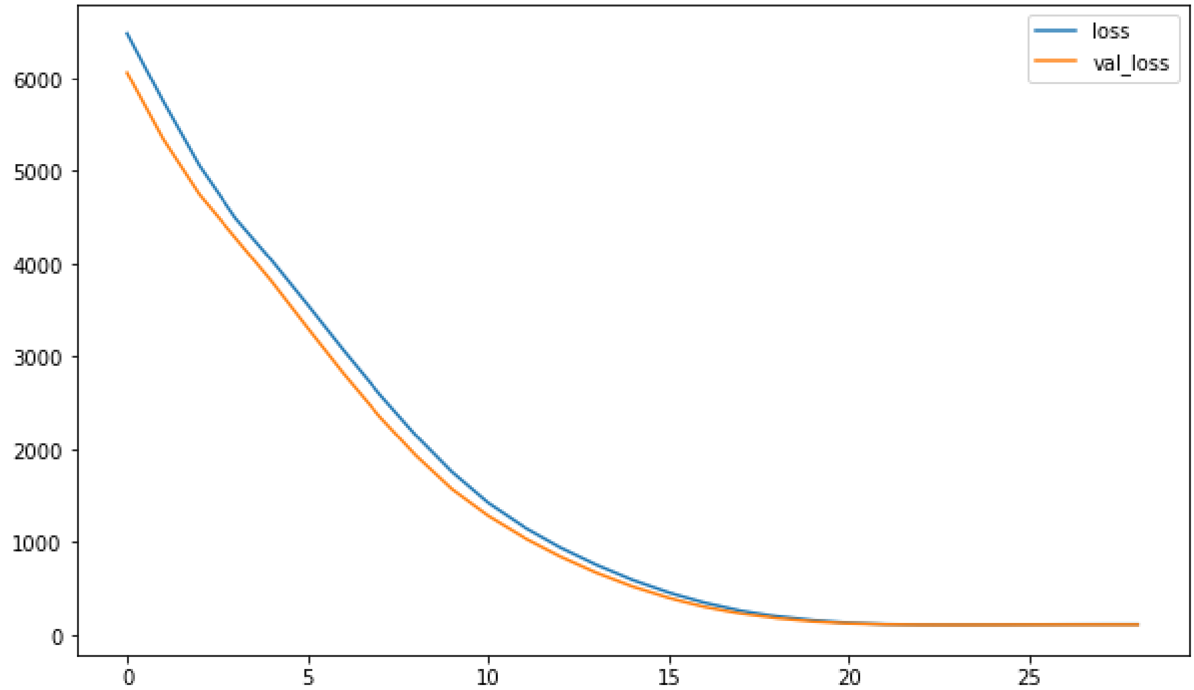
```
Epoch 17/100
3/3 [==============================] - 0s 22ms/step - loss: 35
1.0483 - val_loss: 305.7966
Epoch 18/100
3/3 [==============================] - 0s 22ms/step - loss: 26
3.9883 - val_loss: 232.2294
Epoch 19/100
3/3 [==============================] - 0s 22ms/step - loss: 20
1.2844 - val_loss: 179.1780
Epoch 20/100
3/3 [==============================] - 0s 21ms/step - loss: 15
7.2394 - val_loss: 143.7552
Epoch 21/100
3/3 [==============================] - 0s 29ms/step - loss: 12
9.0856 - val_loss: 122.6676
Epoch 22/100
3/3 [==============================] - 0s 24ms/step - loss: 11
3.9756 - val_loss: 111.8591
Epoch 23/100
3/3 [==============================] - 0s 23ms/step - loss: 10
8.1187 - val_loss: 107.5055
Epoch 24/100
3/3 [==============================] - 0s 22ms/step - loss: 10
7.6638 - val_loss: 106.8273
Epoch 25/100
3/3 [==============================] - 0s 20ms/step - loss: 10
8.5872 - val_loss: 107.7059
Epoch 26/100
3/3 [==============================] - 0s 21ms/step - loss: 11
0.9544 - val_loss: 108.8096
Epoch 27/100
3/3 [==============================] - 0s 21ms/step - loss: 11
2.5534 - val_loss: 109.3596
Epoch 28/100
3/3 [==============================] - 0s 21ms/step - loss: 11
3.3748 - val_loss: 109.4855
Epoch 29/100
3/3 [==============================] - 0s 22ms/step - loss: 11
3.4539 - val_loss: 109.1131
```

In [85]:

```
pd.DataFrame(history.history).plot(figsize=(10, 6))
```

Out[85]:

```
<AxesSubplot:>
```



In [86]:

```
print('Training   R2: ', r2_score(y_train_i,
model.predict(x_train_i)))
print('Testing    R2: ', r2_score(y_test_i,
model.predict(x_test_i)))
print()
print('Training   MSE: ', mean_squared_error(y_train_i,
model.predict(x_train_i)))
print('Testing    MSE: ', mean_squared_error(y_test_i,
model.predict(x_test_i)))
```

```
Training   R2:  0.5221801031623673
Testing    R2:  0.31310901222581167

Training   MSE:  108.27852267787645
Testing    MSE:  106.82732265527996
```

In [87]:

```
# Best model for scientific department is Extra Trees
Regressor
industrial_model = ExtraTreesRegressor(n_estimators=50,
max_depth=17, random_state=42)
industrial_model.fit(x_train_i, y_train_i)
```

Out[87]:

```
ExtraTreesRegressor(max_depth=17, n_estimators=50, random_stat
e=42)
```

In [88]:

```
# Save Model
import pickle
pickle.dump(industrial_model, open('industrial_model.sav',
'wb'))
```

In [ ]:

Appendix C

VISUALIZATION OF THREE (3) DEPARTMENT TOGETHER

Scientific df

Linear Regression

In [39]:

```
# Linear Regression
lr = LinearRegression()
lr.fit(x_train_s, y_train_s)
calculate_accuracies(lr, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.8606312894965049
Testing    R2: 0.8906620424513275

Training   MSE: 29.710203200075487
Testing    MSE: 23.331598920627588
```

In [62]:

```
# Linear Equation
equation = f'expected_gpa = {lr.intercept_:.2f}'
for i in range(10):
    equation += f' + ({lr.coef_[i]:.2f}
{x_train_s.columns[i]})'

print(equation)
```

```
expected_gpa = -27.31 + (-0.08 ARABIC) + (0.17 ENGLISH) + (0.0
6 RELIGION) + (0.33 HISTORIC) + (0.33 MATH) + (-0.11 PHYSIC) +
(0.30 CHYMESTRY) + (0.20 BIOLOGY) + (0.02 COMPUTER) + (0.07 SO
CIAL)
```

Bayesian Ridge Regression

In [63]:

```
# Bayesian Ridge Regression
br = BayesianRidge(n_iter=300, normalize=True)
br.fit(x_train_s, y_train_s)
```

```
calculate_accuracies(br, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.8595519265543166
Testing    R2: 0.8908809966129169

Training   MSE: 29.94029855091282
Testing    MSE: 23.284876347838264
```

In [64]:

```
# Bayesian Ridge Linear Equation
equation = f'expected_gpa = {br.intercept_:.2f}'
for i in range(10):
    equation += f' + ({br.coef_[i]:.2f}
{x_train_s.columns[i]})'

print(equation)
```

```
expected_gpa = -27.05 + (-0.05 ARABIC) + (0.15 ENGLISH) + (0.0
8 RELIGION) + (0.31 HISTORIC) + (0.23 MATH) + (-0.01 PHYSIC) +
(0.26 CHYMESTRY) + (0.22 BIOLOGY) + (0.04 COMPUTER) + (0.05 SO
CIAL)
```

Support Vector Machine

In [69]:

```
# Support Vector Machine Regression
svr = SVR(kernel='linear')
svr.fit(x_train_s, y_train_s)
calculate_accuracies(svr, x_train_s, x_test_s, y_train_s,
y_test_s)
print()

svr = SVR(kernel='rbf')
svr.fit(x_train_s, y_train_s)
calculate_accuracies(svr, x_train_s, x_test_s, y_train_s,
y_test_s)
print()

svr = SVR(kernel='poly')
svr.fit(x_train_s, y_train_s)
calculate_accuracies(svr, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.8533782336637841
Testing    R2: 0.8850552375935233

Training   MSE: 31.25638786256632
Testing    MSE: 24.528033581575773

Training   R2: 0.8052938548606032
Testing    R2: 0.8635968005904039

Training   MSE: 41.506871345055586
Testing    MSE: 29.107043989717567
```

```
Training   R2: 0.9027477545282534
Testing    R2: 0.9152088724983622

Training   MSE: 20.73194165455634
Testing    MSE: 18.09355710724109
```

K Nearest Neighbour

In [74]:

```
# K Nearest Neighbour Regressor
knn = KNeighborsRegressor(n_neighbors=7)
knn.fit(x_train_s, y_train_s)
calculate_accuracies(knn, x_train_s, x_test_s, y_train_s, 
y_test_s)
```

```
Training   R2: 0.9144482313007791
Testing    R2: 0.9330968815740753

Training   MSE: 18.237669151110982
Testing    MSE: 14.276439405391795
```

Decision Tree

In [83]:

```
# Decision Tree Regressor
dt = DecisionTreeRegressor(max_depth=6)
dt.fit(x_train_s, y_train_s)
calculate_accuracies(dt, x_train_s, x_test_s, y_train_s, 
y_test_s)
```

```
Training   R2: 0.9957284518078535
Testing    R2: 0.9872805355882599

Training   MSE: 0.9105958167303676
Testing    MSE: 2.7142032720687057
```

In [86]:

```
# Show tree structure as text
from sklearn.tree import export_text
print(export_text(dt))
```

```
|--- feature_6 <= 73.62
|   |--- feature_3 <= 71.50
|   |   |--- feature_0 <= 74.30
|   |   |   |--- feature_1 <= 74.10
|   |   |   |   |--- feature_6 <= 64.75
|   |   |   |   |   |--- feature_3 <= 56.00
|   |   |   |   |   |   |--- value: [49.52]
|   |   |   |   |   |--- feature_3 >  56.00
|   |   |   |   |   |   |--- value: [50.00]
|   |   |   |   |--- feature_6 >  64.75
|   |   |   |   |   |--- feature_0 <= 65.00
|   |   |   |   |   |   |--- value: [48.33]
|   |   |   |   |   |--- feature_0 >  65.00
```

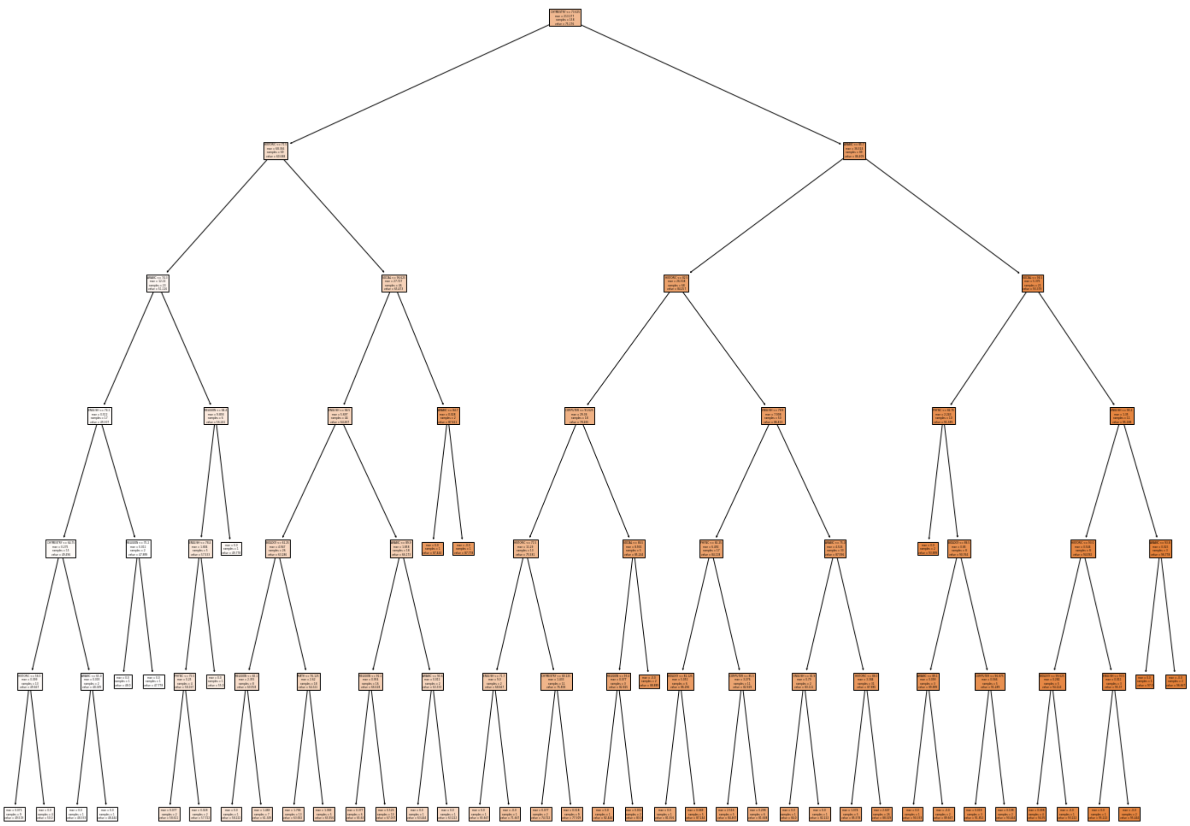
```
|   |   |   |   |   |   |--- value: [48.44]
|   |   |   |--- feature_1 >  74.10
|   |   |   |   |--- feature_2 <= 75.20
|   |   |   |   |   |--- value: [48.00]
|   |   |   |   |--- feature_2 >  75.20
|   |   |   |   |   |--- value: [47.78]
|   |   |--- feature_0 >  74.30
|   |   |   |--- feature_2 <= 84.40
|   |   |   |   |--- feature_1 <= 78.40
|   |   |   |   |   |--- feature_5 <= 75.50
|   |   |   |   |   |   |--- value: [58.61]
|   |   |   |   |   |--- feature_5 >  75.50
|   |   |   |   |   |   |--- value: [57.72]
|   |   |   |   |--- feature_1 >  78.40
|   |   |   |   |   |--- value: [55.00]
|   |   |   |--- feature_2 >  84.40
|   |   |   |   |--- value: [49.78]
|   |--- feature_3 >  71.50
|   |   |--- feature_9 <= 96.62
|   |   |   |--- feature_1 <= 84.50
|   |   |   |   |--- feature_7 <= 62.25
|   |   |   |   |   |--- feature_2 <= 85.10
|   |   |   |   |   |   |--- value: [58.22]
|   |   |   |   |   |--- feature_2 >  85.10
|   |   |   |   |   |   |--- value: [61.35]
|   |   |   |   |--- feature_7 >  62.25
|   |   |   |   |   |--- feature_4 <= 76.12
|   |   |   |   |   |   |--- value: [63.69]
|   |   |   |   |   |--- feature_4 >  76.12
|   |   |   |   |   |   |--- value: [65.96]
|   |   |   |--- feature_1 >  84.50
|   |   |   |   |--- feature_0 <= 89.90
|   |   |   |   |   |--- feature_2 <= 92.10
|   |   |   |   |   |   |--- value: [65.63]
|   |   |   |   |   |--- feature_2 >  92.10
|   |   |   |   |   |   |--- value: [67.07]
|   |   |   |   |--- feature_0 >  89.90
|   |   |   |   |   |--- feature_0 <= 90.60
|   |   |   |   |   |   |--- value: [63.44]
|   |   |   |   |   |--- feature_0 >  90.60
|   |   |   |   |   |   |--- value: [63.22]
|   |   |--- feature_9 >  96.62
|   |   |   |--- feature_0 <= 84.70
|   |   |   |   |--- value: [87.44]
|   |   |   |--- feature_0 >  84.70
|   |   |   |   |--- value: [87.78]
|--- feature_6 >  73.62
|   |--- feature_0 <= 85.70
|   |   |--- feature_3 <= 82.50
|   |   |   |--- feature_8 <= 91.62
|   |   |   |   |--- feature_3 <= 71.50
|   |   |   |   |   |--- feature_1 <= 71.70
```

```
|   |   |   |   |   |   |--- value: [65.67]
|   |   |   |   |   |--- feature_1 >  71.70
|   |   |   |   |   |   |--- value: [71.67]
|   |   |   |   |--- feature_3 >  71.50
|   |   |   |   |   |--- feature_6 <= 80.12
|   |   |   |   |   |   |--- value: [74.72]
|   |   |   |   |   |--- feature_6 >  80.12
|   |   |   |   |   |   |--- value: [77.31]
|   |   |   |--- feature_8 >  91.62
|   |   |   |   |--- feature_9 <= 86.50
|   |   |   |   |   |--- feature_2 <= 93.40
|   |   |   |   |   |   |--- value: [82.44]
|   |   |   |   |   |--- feature_2 >  93.40
|   |   |   |   |   |   |--- value: [83.00]
|   |   |   |   |--- feature_9 >  86.50
|   |   |   |   |   |--- value: [88.89]
|   |   |--- feature_3 >  82.50
|   |   |   |--- feature_1 <= 78.90
|   |   |   |   |--- feature_5 <= 85.25
|   |   |   |   |   |--- feature_7 <= 81.12
|   |   |   |   |   |   |--- value: [81.56]
|   |   |   |   |   |--- feature_7 >  81.12
|   |   |   |   |   |   |--- value: [87.24]
|   |   |   |   |--- feature_5 >  85.25
|   |   |   |   |   |--- feature_8 <= 85.50
|   |   |   |   |   |   |--- value: [84.47]
|   |   |   |   |   |--- feature_8 >  85.50
|   |   |   |   |   |   |--- value: [81.65]
|   |   |   |--- feature_1 >  78.90
|   |   |   |   |--- feature_0 <= 75.00
|   |   |   |   |   |--- feature_1 <= 81.90
|   |   |   |   |   |   |--- value: [84.00]
|   |   |   |   |   |--- feature_1 >  81.90
|   |   |   |   |   |   |--- value: [82.22]
|   |   |   |   |--- feature_0 >  75.00
|   |   |   |   |   |--- feature_3 <= 86.00
|   |   |   |   |   |   |--- value: [85.58]
|   |   |   |   |   |--- feature_3 >  86.00
|   |   |   |   |   |   |--- value: [88.33]
|   |--- feature_0 >  85.70
|   |   |--- feature_9 <= 94.50
|   |   |   |--- feature_5 <= 82.75
|   |   |   |   |--- value: [93.89]
|   |   |   |--- feature_5 >  82.75
|   |   |   |   |--- feature_7 <= 89.50
|   |   |   |   |   |--- feature_0 <= 89.10
|   |   |   |   |   |   |--- value: [90.33]
|   |   |   |   |   |--- feature_0 >  89.10
|   |   |   |   |   |   |--- value: [89.67]
|   |   |   |   |--- feature_7 >  89.50
|   |   |   |   |   |--- feature_8 <= 96.88
|   |   |   |   |   |   |--- value: [91.85]
```

```
|   |   |   |   |   |--- feature_8 >  96.88
|   |   |   |   |   |   |--- value: [90.44]
|   |   |--- feature_9 >  94.50
|   |   |   |--- feature_1 <= 95.20
|   |   |   |   |--- feature_3 <= 93.00
|   |   |   |   |   |--- feature_7 <= 99.62
|   |   |   |   |   |   |--- value: [94.35]
|   |   |   |   |   |--- feature_7 >  99.62
|   |   |   |   |   |   |--- value: [93.22]
|   |   |   |   |--- feature_3 >  93.00
|   |   |   |   |   |--- feature_1 <= 93.50
|   |   |   |   |   |   |--- value: [95.22]
|   |   |   |   |   |--- feature_1 >  93.50
|   |   |   |   |   |   |--- value: [95.44]
|   |   |   |--- feature_1 >  95.20
|   |   |   |   |--- feature_0 <= 93.80
|   |   |   |   |   |--- value: [97.00]
|   |   |   |   |--- feature_0 >  93.80
|   |   |   |   |   |--- value: [96.67]
```

In [84]:

```
# Show Tree Structure as plot
from sklearn.tree import plot_tree
plt.figure(figsize=(25,20))
plot_tree(dt,feature_names=x_train_s.columns, filled=True);
```



Random Forest

In [87]:

```
# Random Forest Regressor
rf = RandomForestRegressor(n_estimators=120, max_depth=20,
random_state=42)
rf.fit(x_train_s, y_train_s)
calculate_accuracies(rf, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.9934780489035981
Testing    R2: 0.9858591199072837

Training   MSE: 1.390329950209265
Testing    MSE: 3.0175188022976713
```

In [94]:

```
# Feature Importance
features = x_train_s.columns
importances = rf.feature_importances_
indices = np.argsort(importances)

plt.figure(figsize=(8,5))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices],
align='center')
plt.yticks(range(len(indices)), [features[i] for i in
indices])
plt.xlabel('Relative Importance')
```

Out[94]:

```
Text(0.5, 0, 'Relative Importance')
```

Extra Trees

In [95]:

```
# Extra Trees Regressor
et = ExtraTreesRegressor(n_estimators=40, max_depth=16,
random_state=42)
et.fit(x_train_s, y_train_s)
calculate_accuracies(et, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.9999999633463226
Testing    R2: 0.9956405796702124

Training   MSE: 7.813720893899859e-06
Testing    MSE: 0.9302555941358002
```

In [96]:

```
# Feature Importance
features = x_train_s.columns
importances = et.feature_importances_
indices = np.argsort(importances)
```

```
plt.figure(figsize=(8,5))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices],
align='center')
plt.yticks(range(len(indices)), [features[i] for i in
indices])
plt.xlabel('Relative Importance')
```

Out[96]:

```
Text(0.5, 0, 'Relative Importance')
```

Gradient Boosting

In [98]:

```
# Gradient Boosting Regressor
gbr = GradientBoostingRegressor(n_estimators=120,
max_depth=16, criterion='mse', random_state=42)
gbr.fit(x_train_s, y_train_s)
calculate_accuracies(gbr, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.999999999989572
Testing    R2: 0.9952760613932878

Training   MSE: 2.2230160883279274e-09
Testing    MSE: 1.0080400564315612
```

In [99]:

```
# Feature Importance
features = x_train_s.columns
importances = gbr.feature_importances_
indices = np.argsort(importances)

plt.figure(figsize=(8,5))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices],
align='center')
plt.yticks(range(len(indices)), [features[i] for i in
indices])
plt.xlabel('Relative Importance')
```

Out[99]:

```
Text(0.5, 0, 'Relative Importance')
```

Stacking

In [46]:

```
# Stacking Technique
et = ExtraTreesRegressor(n_estimators=40, max_depth=16,
random_state=42)
gbr = GradientBoostingRegressor(n_estimators=120,
max_depth=16, criterion='mse', random_state=42)
```

```
rf = RandomForestRegressor(n_estimators=120, max_depth=20,
random_state=42)
regressors = [('et', et), ('gbr', gbr), ('rf', rf)]

from sklearn.ensemble import StackingRegressor
sr = StackingRegressor(estimators=regressors,
final_estimator=LinearRegression())
sr.fit(x_train_s, y_train_s)
calculate_accuracies(sr, x_train_s, x_test_s, y_train_s,
y_test_s)
```

```
Training   R2: 0.9994
Testing    R2: 0.9955

Training   MSE: 0.1198
Testing    MSE: 0.9577
```

Artificial Neural Network

In [47]:

```
# ANN
model = models.Sequential([
    layers.Dense(10, activation='relu', input_dim = 10),
    layers.Dense(units = 8, activation='relu'),
    layers.Dense(units = 4, activation='relu'),
    layers.Dense(units = 1)
])
```

In [48]:

```
model.compile(optimizer = 'adam', loss = 'mean_squared_error')
```

In [49]:

```
early_stopp = callbacks.EarlyStopping(patience=5,
restore_best_weights=True)
```

In [50]:

```
history = model.fit(x_train_s, y_train_s,
validation_data=(x_test_s, y_test_s),
                    epochs=100, batch_size=125,
callbacks=[early_stopp])
```

```
Epoch 1/100
2/2 [==============================] - 1s 335ms/step - loss: 8
844.0732 - val_loss: 8369.5371
Epoch 2/100
2/2 [==============================] - 0s 53ms/step - loss: 82
54.7900 - val_loss: 7790.4673
Epoch 3/100
2/2 [==============================] - 0s 40ms/step - loss: 76
86.3418 - val_loss: 7235.6431
Epoch 4/100
2/2 [==============================] - 0s 52ms/step - loss: 71
41.4106 - val_loss: 6705.3110
```

```
Epoch 5/100
2/2 [==============================] - 0s 70ms/step - loss: 66
16.0752 - val_loss: 6198.5938
Epoch 6/100
2/2 [==============================] - 0s 59ms/step - loss: 61
18.6333 - val_loss: 5715.1860
Epoch 7/100
2/2 [==============================] - 0s 43ms/step - loss: 56
43.2324 - val_loss: 5255.5693
Epoch 8/100
2/2 [==============================] - 0s 52ms/step - loss: 51
92.0806 - val_loss: 4819.9302
Epoch 9/100
2/2 [==============================] - 0s 53ms/step - loss: 47
59.4238 - val_loss: 4407.0430
Epoch 10/100
2/2 [==============================] - 0s 58ms/step - loss: 43
51.2476 - val_loss: 4015.3523
Epoch 11/100
2/2 [==============================] - 0s 68ms/step - loss: 39
65.1995 - val_loss: 3644.5867
Epoch 12/100
2/2 [==============================] - 0s 68ms/step - loss: 36
02.3181 - val_loss: 3295.1680
Epoch 13/100
2/2 [==============================] - 0s 42ms/step - loss: 32
57.6787 - val_loss: 2967.2534
Epoch 14/100
2/2 [==============================] - 0s 44ms/step - loss: 29
34.9016 - val_loss: 2660.1753
Epoch 15/100
2/2 [==============================] - 0s 48ms/step - loss: 26
29.2935 - val_loss: 2372.5168
Epoch 16/100
2/2 [==============================] - 0s 53ms/step - loss: 23
42.7568 - val_loss: 2097.4155
Epoch 17/100
2/2 [==============================] - 0s 61ms/step - loss: 20
72.7646 - val_loss: 1836.3362
Epoch 18/100
2/2 [==============================] - 0s 57ms/step - loss: 18
14.6926 - val_loss: 1592.0374
Epoch 19/100
2/2 [==============================] - 0s 57ms/step - loss: 15
72.1643 - val_loss: 1365.7571
Epoch 20/100
2/2 [==============================] - 0s 50ms/step - loss: 13
46.0172 - val_loss: 1149.7445
Epoch 21/100
2/2 [==============================] - 0s 48ms/step - loss: 11
15.4851 - val_loss: 931.0494
Epoch 22/100
```

```
2/2 [==============================] - 0s 53ms/step - loss: 89
1.2341 - val_loss: 699.7899
Epoch 23/100
2/2 [==============================] - 0s 67ms/step - loss: 65
8.8511 - val_loss: 462.7523
Epoch 24/100
2/2 [==============================] - 0s 63ms/step - loss: 43
7.7462 - val_loss: 271.0106
Epoch 25/100
2/2 [==============================] - 0s 50ms/step - loss: 25
3.6341 - val_loss: 143.6966
Epoch 26/100
2/2 [==============================] - 0s 45ms/step - loss: 13
9.9116 - val_loss: 79.0259
Epoch 27/100
2/2 [==============================] - 0s 47ms/step - loss: 84
.8471 - val_loss: 66.2434
Epoch 28/100
2/2 [==============================] - 0s 59ms/step - loss: 74
.0447 - val_loss: 85.9329
Epoch 29/100
2/2 [==============================] - 0s 56ms/step - loss: 97
.4233 - val_loss: 117.5698
Epoch 30/100
2/2 [==============================] - 0s 61ms/step - loss: 12
8.8107 - val_loss: 143.6599
Epoch 31/100
2/2 [==============================] - 0s 53ms/step - loss: 15
3.4807 - val_loss: 154.4269
Epoch 32/100
2/2 [==============================] - 0s 52ms/step - loss: 16
0.7472 - val_loss: 148.7057
```

In [51]:

```
pd.DataFrame(history.history).plot(figsize=(10, 6))
```

Out[51]:

```
<AxesSubplot:>
```

In [52]:

```
print('Training   R2: ', r2_score(y_train_s,
model.predict(x_train_s)))
print('Testing    R2: ', r2_score(y_test_s,
model.predict(x_test_s)))
print()
print('Training   MSE: ', mean_squared_error(y_train_s,
model.predict(x_train_s)))
print('Testing    MSE: ', mean_squared_error(y_test_s,
model.predict(x_test_s)))
```

```
Training   R2:  0.6521266262998191
Testing    R2:  0.6895663333826705

Training   MSE:  74.158601189533
```

```
Testing    MSE:  66.24336107386237
```

Literary df

Linear Regression

In [101]:

```
# Linear Regression
lr = LinearRegression()
lr.fit(x_train_l, y_train_l)
calculate_accuracies(lr, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9184658235522126
Testing    R2: 0.9148877985809863

Training   MSE: 22.424321566988173
Testing    MSE: 25.144774841431797
```

In [102]:

```
# Linear Equation
equation = f'expected_gpa = {lr.intercept_:.2f}'
for i in range(10):
    equation += f' + ({lr.coef_[i]:.2f}
{x_train_s.columns[i]})'

print(equation)
```

```
expected_gpa = -10.70 + (0.16 ARABIC) + (0.33 ENGLISH) + (-0.1
1 RELIGION) + (0.78 HISTORIC) + (-0.01 MATH) + (-0.10 PHYSIC)
+ (-0.02 CHYMESTRY) + (0.07 BIOLOGY) + (-0.04 COMPUTER) + (0.0
5 SOCIAL)
```

Bayesian Ridge Regression

In [105]:

```
# Bayesian Ridge Regression
br = BayesianRidge(n_iter=300, normalize=True)
br.fit(x_train_l, y_train_l)
calculate_accuracies(br, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9183206666493151
Testing    R2: 0.9152737664392925

Training   MSE: 22.46424402907776
Testing    MSE: 25.030748007072777
```

In [106]:

```
# Bayesian Ridge Linear Equation
equation = f'expected_gpa = {br.intercept_:.2f}'
```

```
for i in range(10):
    equation += f' + ({br.coef_[i]:.2f}
{x_train_s.columns[i]})'

print(equation)
```

```
expected_gpa = -10.95 + (0.17 ARABIC) + (0.31 ENGLISH) + (-0.0
7 RELIGION) + (0.76 HISTORIC) + (-0.02 MATH) + (-0.10 PHYSIC)
+ (-0.00 CHYMESTRY) + (0.06 BIOLOGY) + (-0.03 COMPUTER) + (0.0
4 SOCIAL)
```

Support Vector Machine

In [57]:

```
# Support Vector Machine Regression
svr = SVR(kernel='linear')
svr.fit(x_train_l, y_train_l)
calculate_accuracies(svr, x_train_l, x_test_l, y_train_l,
y_test_l)
print()

svr = SVR(kernel='rbf')
svr.fit(x_train_l, y_train_l)
calculate_accuracies(svr, x_train_l, x_test_l, y_train_l,
y_test_l)
print()

svr = SVR(kernel='poly')
svr.fit(x_train_l, y_train_l)
calculate_accuracies(svr, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9144
Testing    R2: 0.9107

Training   MSE: 23.5464
Testing    MSE: 26.3889

Training   R2: 0.8974
Testing    R2: 0.8853

Training   MSE: 28.2233
Testing    MSE: 33.8919

Training   R2: 0.9434
Testing    R2: 0.9490

Training   MSE: 15.5722
Testing    MSE: 15.0619
```

K Nearest Neighbour

In [58]:

```
# K Nearest Neighbour Regressor
knn = KNeighborsRegressor(n_neighbors=7)
knn.fit(x_train_l, y_train_l)
calculate_accuracies(knn, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9316
Testing    R2: 0.9477

Training   MSE: 18.8048
Testing    MSE: 15.4522
```

Decision Tree

In [115]:

```
# Decision Tree Regressor
dt = DecisionTreeRegressor(max_depth=6)
dt.fit(x_train_l, y_train_l)
calculate_accuracies(dt, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9979985315710823
Testing    R2: 0.9796050278653815

Training   MSE: 0.5504632978658462
Testing    MSE: 6.025305111044806
```

In [116]:

```
# Show tree structure as text
from sklearn.tree import export_text
print(export_text(dt))
```

```
|--- feature_3 <= 72.50
|   |--- feature_1 <= 75.60
|   |   |--- feature_0 <= 67.50
|   |   |   |--- feature_8 <= 88.88
|   |   |   |   |--- feature_3 <= 65.00
|   |   |   |   |   |--- feature_8 <= 69.88
|   |   |   |   |   |   |--- value: [49.04]
|   |   |   |   |   |--- feature_8 >  69.88
|   |   |   |   |   |   |--- value: [49.59]
|   |   |   |   |--- feature_3 >  65.00
|   |   |   |   |   |--- feature_0 <= 59.00
|   |   |   |   |   |   |--- value: [54.88]
|   |   |   |   |   |--- feature_0 >  59.00
|   |   |   |   |   |   |--- value: [49.81]
|   |   |   |--- feature_8 >  88.88
|   |   |   |   |--- value: [57.50]
|   |   |--- feature_0 >  67.50
|   |   |   |--- feature_5 <= 69.00
```

```
|   |   |   |   |--- value: [50.00]
|   |   |   |--- feature_5 >  69.00
|   |   |   |   |--- value: [64.75]
|   |--- feature_1 >  75.60
|   |   |--- feature_1 <= 78.50
|   |   |   |--- feature_2 <= 84.30
|   |   |   |   |--- feature_4 <= 76.50
|   |   |   |   |   |--- feature_1 <= 75.90
|   |   |   |   |   |   |--- value: [69.50]
|   |   |   |   |   |--- feature_1 >  75.90
|   |   |   |   |   |   |--- value: [72.38]
|   |   |   |   |--- feature_4 >  76.50
|   |   |   |   |   |--- value: [75.62]
|   |   |   |--- feature_2 >  84.30
|   |   |   |   |--- feature_9 <= 91.50
|   |   |   |   |   |--- feature_8 <= 72.12
|   |   |   |   |   |   |--- value: [79.44]
|   |   |   |   |   |--- feature_8 >  72.12
|   |   |   |   |   |   |--- value: [80.12]
|   |   |   |   |--- feature_9 >  91.50
|   |   |   |   |   |--- value: [76.88]
|   |   |--- feature_1 >  78.50
|   |   |   |--- feature_3 <= 62.00
|   |   |   |   |--- value: [50.00]
|   |   |   |--- feature_3 >  62.00
|   |   |   |   |--- value: [49.75]
|--- feature_3 >  72.50
|   |--- feature_3 <= 83.50
|   |   |--- feature_2 <= 82.70
|   |   |   |--- feature_6 <= 91.62
|   |   |   |   |--- feature_8 <= 71.62
|   |   |   |   |   |--- value: [69.25]
|   |   |   |   |--- feature_8 >  71.62
|   |   |   |   |   |--- feature_6 <= 84.38
|   |   |   |   |   |   |--- value: [75.18]
|   |   |   |   |   |--- feature_6 >  84.38
|   |   |   |   |   |   |--- value: [78.62]
|   |   |   |--- feature_6 >  91.62
|   |   |   |   |--- feature_2 <= 77.10
|   |   |   |   |   |--- value: [85.25]
|   |   |   |   |--- feature_2 >  77.10
|   |   |   |   |   |--- value: [85.88]
|   |   |--- feature_2 >  82.70
|   |   |   |--- feature_1 <= 80.50
|   |   |   |   |--- feature_0 <= 77.50
|   |   |   |   |   |--- feature_9 <= 92.88
|   |   |   |   |   |   |--- value: [79.42]
|   |   |   |   |   |--- feature_9 >  92.88
|   |   |   |   |   |   |--- value: [79.88]
|   |   |   |   |--- feature_0 >  77.50
|   |   |   |   |   |--- feature_7 <= 62.25
|   |   |   |   |   |   |--- value: [77.62]
```

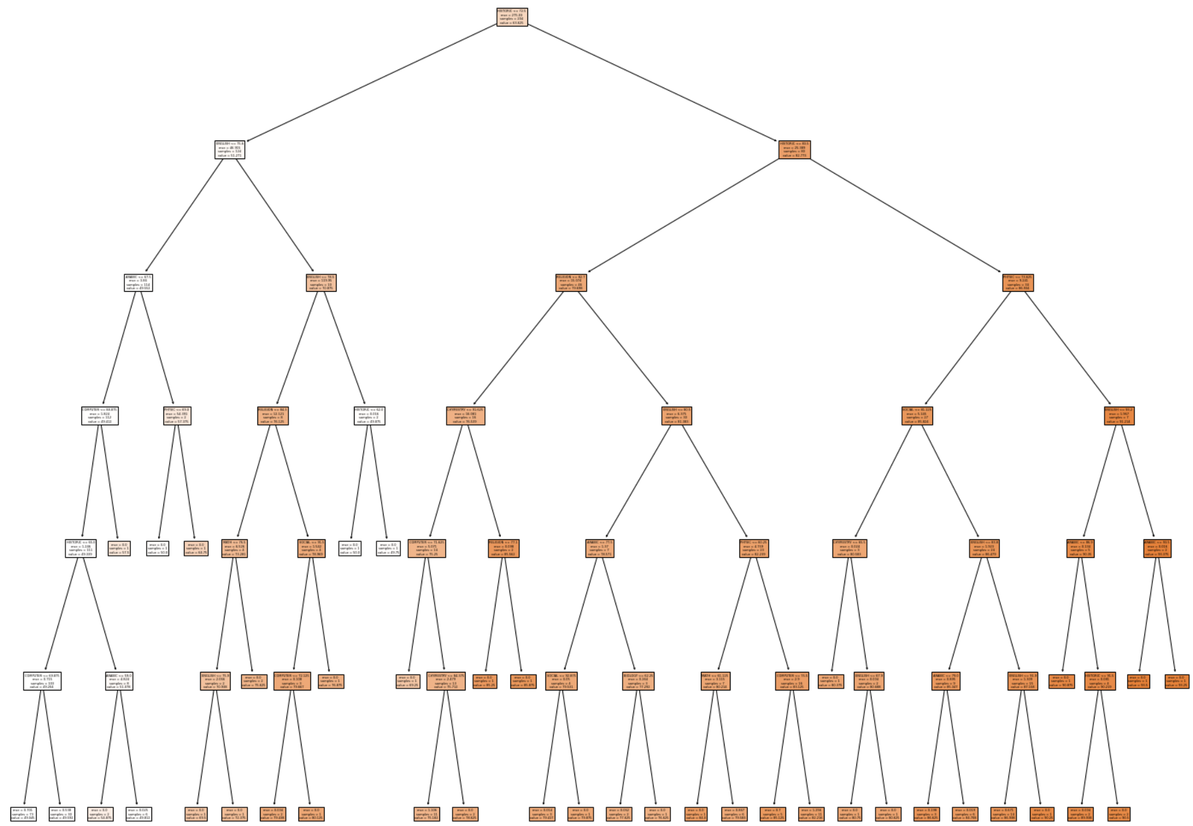
```
|   |   |   |   |   |--- feature_7 >  62.25
|   |   |   |   |   |   |--- value: [76.62]
|   |   |   |--- feature_1 >  80.50
|   |   |   |   |--- feature_5 <= 60.25
|   |   |   |   |   |--- feature_4 <= 61.12
|   |   |   |   |   |   |--- value: [84.00]
|   |   |   |   |   |--- feature_4 >  61.12
|   |   |   |   |   |   |--- value: [79.58]
|   |   |   |   |--- feature_5 >  60.25
|   |   |   |   |   |--- feature_8 <= 75.50
|   |   |   |   |   |   |--- value: [85.12]
|   |   |   |   |   |--- feature_8 >  75.50
|   |   |   |   |   |   |--- value: [82.22]
|   |--- feature_3 >  83.50
|   |   |--- feature_5 <= 73.62
|   |   |   |--- feature_9 <= 81.12
|   |   |   |   |--- feature_6 <= 65.50
|   |   |   |   |   |--- value: [80.38]
|   |   |   |   |--- feature_6 >  65.50
|   |   |   |   |   |--- feature_1 <= 67.90
|   |   |   |   |   |   |--- value: [80.75]
|   |   |   |   |   |--- feature_1 >  67.90
|   |   |   |   |   |   |--- value: [80.62]
|   |   |   |--- feature_9 >  81.12
|   |   |   |   |--- feature_1 <= 83.80
|   |   |   |   |   |--- feature_0 <= 79.00
|   |   |   |   |   |   |--- value: [86.62]
|   |   |   |   |   |--- feature_0 >  79.00
|   |   |   |   |   |   |--- value: [84.71]
|   |   |   |   |--- feature_1 >  83.80
|   |   |   |   |   |--- feature_1 <= 91.90
|   |   |   |   |   |   |--- value: [86.94]
|   |   |   |   |   |--- feature_1 >  91.90
|   |   |   |   |   |   |--- value: [90.25]
|   |   |--- feature_5 >  73.62
|   |   |   |--- feature_1 <= 93.20
|   |   |   |   |--- feature_0 <= 86.90
|   |   |   |   |   |--- value: [90.88]
|   |   |   |   |--- feature_0 >  86.90
|   |   |   |   |   |--- feature_3 <= 91.50
|   |   |   |   |   |   |--- value: [89.94]
|   |   |   |   |   |--- feature_3 >  91.50
|   |   |   |   |   |   |--- value: [90.50]
|   |   |   |--- feature_1 >  93.20
|   |   |   |   |--- feature_0 <= 93.50
|   |   |   |   |   |--- value: [93.50]
|   |   |   |   |--- feature_0 >  93.50
|   |   |   |   |   |--- value: [93.25]
```

In [117]:

```
# Show Tree Structure as plot
from sklearn.tree import plot_tree
```

```
plt.figure(figsize=(25,20))
plot_tree(dt,feature_names=x_train_s.columns, filled=True);
```

```
C:\Users\Wssam\anaconda3\lib\site-packages\sklearn\utils\depre
cation.py:143: FutureWarning: The sklearn.tree.export module i
s  deprecated in version 0.22 and will be removed in version 0
.24. The corresponding classes / functions should instead be i
mported from sklearn.tree. Anything that cannot be imported fr
om sklearn.tree is now part of the private API.
  warnings.warn(message, FutureWarning)
```



Random Forest

In [121]:

```
# Random Forest Regressor
rf = RandomForestRegressor(n_estimators=120, max_depth=20,
random_state=42)
rf.fit(x_train_l, y_train_l)
calculate_accuracies(rf, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.994622480497317
Testing    R2: 0.9704876121372158

Training   MSE: 1.4789776731004902
Testing    MSE: 8.718871506911059
```

In [122]:

```
# Feature Importance
features = x_train_s.columns
importances = rf.feature_importances_
indices = np.argsort(importances)

plt.figure(figsize=(8,5))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices],
align='center')
plt.yticks(range(len(indices)), [features[i] for i in
indices])
plt.xlabel('Relative Importance')
```

Out[122]:

```
Text(0.5, 0, 'Relative Importance')
```

Extra Trees

In [123]:

```
# Extra Trees Regressor
et = ExtraTreesRegressor(n_estimators=35, max_depth=17,
random_state=42)
et.fit(x_train_l, y_train_l)
calculate_accuracies(et, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9999989511892269
Testing    R2: 0.9872393779712936

Training   MSE: 0.00028845413129820885
Testing    MSE: 3.769882136743306
```

In [124]:

```
# Feature Importance
features = x_train_s.columns
importances = et.feature_importances_
indices = np.argsort(importances)

plt.figure(figsize=(8,5))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices],
align='center')
plt.yticks(range(len(indices)), [features[i] for i in
indices])
plt.xlabel('Relative Importance')
```

Out[124]:

```
Text(0.5, 0, 'Relative Importance')
```

Gradient Boosting

In [126]:

```
# Gradient Boosting Regressor
gbr = GradientBoostingRegressor(n_estimators=120,
max_depth=16, criterion='mse', random_state=42)
gbr.fit(x_train_l, y_train_l)
calculate_accuracies(gbr, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9999999999895703
Testing    R2: 0.9606685766028231

Training   MSE: 2.8684547807588635e-09
Testing    MSE: 11.619718078330688
```

In [127]:

```
# Feature Importance
features = x_train_s.columns
importances = gbr.feature_importances_
indices = np.argsort(importances)

plt.figure(figsize=(8,5))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices],
align='center')
plt.yticks(range(len(indices)), [features[i] for i in
indices])
plt.xlabel('Relative Importance')
```

Out[127]:

```
Text(0.5, 0, 'Relative Importance')
```

Stacking

In [63]:

```
# Stacking Technique
et = ExtraTreesRegressor(n_estimators=35, max_depth=17,
random_state=42)
gbr = GradientBoostingRegressor(n_estimators=120,
max_depth=16, criterion='mse', random_state=42)
rf = RandomForestRegressor(n_estimators=120, max_depth=20,
random_state=42)
regressors = [('et', et), ('gbr', gbr), ('rf', rf)]

from sklearn.ensemble import StackingRegressor
sr = StackingRegressor(estimators=regressors,
final_estimator=LinearRegression())
sr.fit(x_train_l, y_train_l)
calculate_accuracies(sr, x_train_l, x_test_l, y_train_l,
y_test_l)
```

```
Training   R2: 0.9995
Testing    R2: 0.9850

Training   MSE: 0.1314
```

```
Testing    MSE: 4.4203
```

Artificial Neural Network

In [64]:

```
# ANN
model = models.Sequential([
    layers.Dense(10, activation='relu', input_dim = 10),
    layers.Dense(units = 8, activation='relu'),
    layers.Dense(units = 4, activation='relu'),
    layers.Dense(units = 1)
])
```

In [65]:

```
model.compile(optimizer = 'adam', loss = 'mean_squared_error')
```

In [66]:

```
early_stopp = callbacks.EarlyStopping(patience=5,
restore_best_weights=True)
```

In [67]:

```
history = model.fit(x_train_l, y_train_l,
validation_data=(x_test_l, y_test_l),
                    epochs=100, batch_size=125,
callbacks=[early_stopp])
```

```
Epoch 1/100
2/2 [==============================] - 1s 294ms/step - loss: 4
603.8818 - val_loss: 5006.9297
Epoch 2/100
2/2 [==============================] - 0s 45ms/step - loss: 45
24.8877 - val_loss: 4924.6553
Epoch 3/100
2/2 [==============================] - 0s 71ms/step - loss: 44
56.6489 - val_loss: 4856.1997
Epoch 4/100
2/2 [==============================] - 0s 53ms/step - loss: 43
96.6528 - val_loss: 4802.6699
Epoch 5/100
2/2 [==============================] - 0s 42ms/step - loss: 43
49.4106 - val_loss: 4767.2046
Epoch 6/100
2/2 [==============================] - 0s 39ms/step - loss: 43
25.2891 - val_loss: 4760.1372
Epoch 7/100
2/2 [==============================] - 0s 44ms/step - loss: 43
21.6372 - val_loss: 4759.7227
Epoch 8/100
2/2 [==============================] - 0s 59ms/step - loss: 43
21.3442 - val_loss: 4759.4565
Epoch 9/100
```

```
2/2 [==============================] - 0s 58ms/step - loss: 43
21.0908 - val_loss: 4759.1909
Epoch 10/100
2/2 [==============================] - 0s 55ms/step - loss: 43
20.8364 - val_loss: 4758.9248
Epoch 11/100
2/2 [==============================] - 0s 61ms/step - loss: 43
20.5850 - val_loss: 4758.6587
Epoch 12/100
2/2 [==============================] - 0s 42ms/step - loss: 43
20.3301 - val_loss: 4758.3926
Epoch 13/100
2/2 [==============================] - 0s 41ms/step - loss: 43
20.0776 - val_loss: 4758.1265
Epoch 14/100
2/2 [==============================] - 0s 54ms/step - loss: 43
19.8237 - val_loss: 4757.8604
Epoch 15/100
2/2 [==============================] - 0s 55ms/step - loss: 43
19.5693 - val_loss: 4757.5942
Epoch 16/100
2/2 [==============================] - 0s 70ms/step - loss: 43
19.3154 - val_loss: 4757.3276
Epoch 17/100
2/2 [==============================] - 0s 53ms/step - loss: 43
19.0649 - val_loss: 4757.0615
Epoch 18/100
2/2 [==============================] - 0s 50ms/step - loss: 43
18.8086 - val_loss: 4756.7949
Epoch 19/100
2/2 [==============================] - 0s 36ms/step - loss: 43
18.5547 - val_loss: 4756.5283
Epoch 20/100
2/2 [==============================] - 0s 52ms/step - loss: 43
18.3027 - val_loss: 4756.2622
Epoch 21/100
2/2 [==============================] - 0s 43ms/step - loss: 43
18.0508 - val_loss: 4755.9951
Epoch 22/100
2/2 [==============================] - 0s 47ms/step - loss: 43
17.7964 - val_loss: 4755.7295
Epoch 23/100
2/2 [==============================] - 0s 48ms/step - loss: 43
17.5405 - val_loss: 4755.4634
Epoch 24/100
2/2 [==============================] - 0s 89ms/step - loss: 43
17.2866 - val_loss: 4755.1973
Epoch 25/100
2/2 [==============================] - 0s 47ms/step - loss: 43
17.0361 - val_loss: 4754.9307
Epoch 26/100
```

```
2/2 [==============================] - 0s 41ms/step - loss: 43
16.7803 - val_loss: 4754.6646
Epoch 27/100
2/2 [==============================] - 0s 46ms/step - loss: 43
16.5298 - val_loss: 4754.3975
Epoch 28/100
2/2 [==============================] - 0s 40ms/step - loss: 43
16.2734 - val_loss: 4754.1318
Epoch 29/100
2/2 [==============================] - 0s 47ms/step - loss: 43
16.0215 - val_loss: 4753.8657
Epoch 30/100
2/2 [==============================] - 0s 64ms/step - loss: 43
15.7690 - val_loss: 4753.5986
Epoch 31/100
2/2 [==============================] - 0s 67ms/step - loss: 43
15.5127 - val_loss: 4753.3330
Epoch 32/100
2/2 [==============================] - 0s 63ms/step - loss: 43
15.2593 - val_loss: 4753.0669
Epoch 33/100
2/2 [==============================] - 0s 68ms/step - loss: 43
15.0078 - val_loss: 4752.7998
Epoch 34/100
2/2 [==============================] - 0s 47ms/step - loss: 43
14.7539 - val_loss: 4752.5342
Epoch 35/100
2/2 [==============================] - 0s 46ms/step - loss: 43
14.5015 - val_loss: 4752.2681
Epoch 36/100
2/2 [==============================] - 0s 48ms/step - loss: 43
14.2456 - val_loss: 4752.0020
Epoch 37/100
2/2 [==============================] - 0s 58ms/step - loss: 43
13.9932 - val_loss: 4751.7354
Epoch 38/100
2/2 [==============================] - 0s 58ms/step - loss: 43
13.7383 - val_loss: 4751.4692
Epoch 39/100
2/2 [==============================] - 0s 76ms/step - loss: 43
13.4863 - val_loss: 4751.2026
Epoch 40/100
2/2 [==============================] - 0s 50ms/step - loss: 43
13.2324 - val_loss: 4750.9365
Epoch 41/100
2/2 [==============================] - 0s 54ms/step - loss: 43
12.9780 - val_loss: 4750.6699
Epoch 42/100
2/2 [==============================] - 0s 59ms/step - loss: 43
12.7251 - val_loss: 4750.4033
Epoch 43/100
```

```
2/2 [==============================] - 0s 55ms/step - loss: 43
12.4712 - val_loss: 4750.1372
Epoch 44/100
2/2 [==============================] - 0s 59ms/step - loss: 43
12.2183 - val_loss: 4749.8706
Epoch 45/100
2/2 [==============================] - 0s 69ms/step - loss: 43
11.9644 - val_loss: 4749.6035
Epoch 46/100
2/2 [==============================] - 0s 59ms/step - loss: 43
11.7109 - val_loss: 4749.3369
Epoch 47/100
2/2 [==============================] - 0s 63ms/step - loss: 43
11.4561 - val_loss: 4749.0708
Epoch 48/100
2/2 [==============================] - 0s 45ms/step - loss: 43
11.2036 - val_loss: 4748.8047
Epoch 49/100
2/2 [==============================] - 0s 46ms/step - loss: 43
10.9497 - val_loss: 4748.5386
Epoch 50/100
2/2 [==============================] - 0s 59ms/step - loss: 43
10.6973 - val_loss: 4748.2720
Epoch 51/100
2/2 [==============================] - 0s 48ms/step - loss: 43
10.4419 - val_loss: 4748.0059
Epoch 52/100
2/2 [==============================] - 0s 58ms/step - loss: 43
10.1904 - val_loss: 4747.7393
Epoch 53/100
2/2 [==============================] - 0s 49ms/step - loss: 43
09.9331 - val_loss: 4747.4736
Epoch 54/100
2/2 [==============================] - 0s 63ms/step - loss: 43
09.6821 - val_loss: 4747.2065
Epoch 55/100
2/2 [==============================] - 0s 74ms/step - loss: 43
09.4272 - val_loss: 4746.9409
Epoch 56/100
2/2 [==============================] - 0s 48ms/step - loss: 43
09.1748 - val_loss: 4746.6743
Epoch 57/100
2/2 [==============================] - 0s 56ms/step - loss: 43
08.9214 - val_loss: 4746.4072
Epoch 58/100
2/2 [==============================] - 0s 63ms/step - loss: 43
08.6665 - val_loss: 4746.1411
Epoch 59/100
2/2 [==============================] - 0s 55ms/step - loss: 43
08.4131 - val_loss: 4745.8745
Epoch 60/100
```

```
2/2 [==============================] - 0s 63ms/step - loss: 43
08.1582 - val_loss: 4745.6074
Epoch 61/100
2/2 [==============================] - 0s 77ms/step - loss: 43
07.9033 - val_loss: 4745.3408
Epoch 62/100
2/2 [==============================] - 0s 58ms/step - loss: 43
07.6533 - val_loss: 4745.0732
Epoch 63/100
2/2 [==============================] - 0s 50ms/step - loss: 43
07.3975 - val_loss: 4744.8066
Epoch 64/100
2/2 [==============================] - 0s 50ms/step - loss: 43
07.1436 - val_loss: 4744.5400
Epoch 65/100
2/2 [==============================] - 0s 45ms/step - loss: 43
06.8896 - val_loss: 4744.2734
Epoch 66/100
2/2 [==============================] - 0s 67ms/step - loss: 43
06.6372 - val_loss: 4744.0068
Epoch 67/100
2/2 [==============================] - 0s 51ms/step - loss: 43
06.3813 - val_loss: 4743.7412
Epoch 68/100
2/2 [==============================] - 0s 66ms/step - loss: 43
06.1284 - val_loss: 4743.4746
Epoch 69/100
2/2 [==============================] - 0s 60ms/step - loss: 43
05.8774 - val_loss: 4743.2075
Epoch 70/100
2/2 [==============================] - 0s 67ms/step - loss: 43
05.6201 - val_loss: 4742.9419
Epoch 71/100
2/2 [==============================] - 0s 46ms/step - loss: 43
05.3682 - val_loss: 4742.6753
Epoch 72/100
2/2 [==============================] - 0s 48ms/step - loss: 43
05.1128 - val_loss: 4742.4092
Epoch 73/100
2/2 [==============================] - 0s 52ms/step - loss: 43
04.8613 - val_loss: 4742.1426
Epoch 74/100
2/2 [==============================] - 0s 46ms/step - loss: 43
04.6084 - val_loss: 4741.8765
Epoch 75/100
2/2 [==============================] - 0s 56ms/step - loss: 43
04.3516 - val_loss: 4741.6099
Epoch 76/100
2/2 [==============================] - 0s 69ms/step - loss: 43
04.1011 - val_loss: 4741.3438
Epoch 77/100
```

```
2/2 [==============================] - 0s 49ms/step - loss: 43
03.8457 - val_loss: 4741.0771
Epoch 78/100
2/2 [==============================] - 0s 50ms/step - loss: 43
03.5933 - val_loss: 4740.8105
Epoch 79/100
2/2 [==============================] - 0s 50ms/step - loss: 43
03.3398 - val_loss: 4740.5439
Epoch 80/100
2/2 [==============================] - 0s 53ms/step - loss: 43
03.0859 - val_loss: 4740.2783
Epoch 81/100
2/2 [==============================] - 0s 55ms/step - loss: 43
02.8325 - val_loss: 4740.0117
Epoch 82/100
2/2 [==============================] - 0s 63ms/step - loss: 43
02.5767 - val_loss: 4739.7466
Epoch 83/100
2/2 [==============================] - 0s 63ms/step - loss: 43
02.3262 - val_loss: 4739.4790
Epoch 84/100
2/2 [==============================] - 0s 46ms/step - loss: 43
02.0723 - val_loss: 4739.2129
Epoch 85/100
2/2 [==============================] - 0s 50ms/step - loss: 43
01.8169 - val_loss: 4738.9468
Epoch 86/100
2/2 [==============================] - 0s 61ms/step - loss: 43
01.5659 - val_loss: 4738.6802
Epoch 87/100
2/2 [==============================] - 0s 58ms/step - loss: 43
01.3110 - val_loss: 4738.4136
Epoch 88/100
2/2 [==============================] - 0s 40ms/step - loss: 43
01.0586 - val_loss: 4738.1475
Epoch 89/100
2/2 [==============================] - 0s 70ms/step - loss: 43
00.8047 - val_loss: 4737.8813
Epoch 90/100
2/2 [==============================] - 0s 57ms/step - loss: 43
00.5498 - val_loss: 4737.6152
Epoch 91/100
2/2 [==============================] - 0s 48ms/step - loss: 43
00.2983 - val_loss: 4737.3491
Epoch 92/100
2/2 [==============================] - 0s 49ms/step - loss: 43
00.0444 - val_loss: 4737.0835
Epoch 93/100
2/2 [==============================] - 0s 44ms/step - loss: 42
99.7910 - val_loss: 4736.8174
Epoch 94/100
```

```
2/2 [==============================] - 0s 47ms/step - loss: 42
99.5391 - val_loss: 4736.5518
Epoch 95/100
2/2 [==============================] - 0s 71ms/step - loss: 42
99.2871 - val_loss: 4736.2856
Epoch 96/100
2/2 [==============================] - 0s 58ms/step - loss: 42
99.0342 - val_loss: 4736.0200
Epoch 97/100
2/2 [==============================] - 0s 44ms/step - loss: 42
98.7783 - val_loss: 4735.7554
Epoch 98/100
2/2 [==============================] - 0s 42ms/step - loss: 42
98.5273 - val_loss: 4735.4902
Epoch 99/100
2/2 [==============================] - 0s 41ms/step - loss: 42
98.2739 - val_loss: 4735.2236
Epoch 100/100
2/2 [==============================] - 0s 51ms/step - loss: 42
98.0210 - val_loss: 4734.9585
```

In [68]:

```
pd.DataFrame(history.history).plot(figsize=(10, 6))
```

Out[68]:

```
<AxesSubplot:>
```

In [69]:

```
print('Training   R2: ', r2_score(y_train_l,
model.predict(x_train_l)))
print('Testing    R2: ', r2_score(y_test_l,
model.predict(x_test_l)))
print()
print('Training   MSE: ', mean_squared_error(y_train_l,
model.predict(x_train_l)))
print('Testing    MSE: ', mean_squared_error(y_test_l,
model.predict(x_test_l)))
```

```
Training   R2:  -14.626740482678613
Testing    R2:  -15.027296421305657

Training   MSE:  4297.81803035513
Testing    MSE:  4734.958714633687
```

Industrial df

Linear Regression

In [103]:

```
# Linear Regression
lr = LinearRegression()
lr.fit(x_train_i, y_train_i)
```

```
calculate_accuracies(lr, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.8469248649947998
Testing    R2: 0.8675829126708228

Training   MSE: 33.1578143741665
Testing    MSE: 26.04989934289757
```

In [104]:

```
# Linear Equation
equation = f'expected_gpa = {lr.intercept_:.2f}'
for i in range(10):
    equation += f' + ({lr.coef_[i]:.2f}
{x_train_s.columns[i]})'

print(equation)
```

```
expected_gpa = -21.55 + (0.05 ARABIC) + (0.26 ENGLISH) + (0.04
RELIGION) + (0.50 HISTORIC) + (0.71 MATH) + (-0.08 PHYSIC) + (
-0.17 CHYMESTRY) + (0.03 BIOLOGY) + (-0.07 COMPUTER) + (-0.04
SOCIAL)
```

Bayesian Ridge Regression

In [107]:

```
# Bayesian Ridge Regression
br = BayesianRidge(n_iter=300, normalize=True)
br.fit(x_train_i, y_train_i)
calculate_accuracies(br, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.8450365963667343
Testing    R2: 0.8620096437480173

Training   MSE: 33.566834824521344
Testing    MSE: 27.146306894054987
```

In [108]:

```
# Bayesian Ridge Linear Equation
equation = f'expected_gpa = {br.intercept_:.2f}'
for i in range(10):
    equation += f' + ({br.coef_[i]:.2f}
{x_train_s.columns[i]})'

print(equation)
```

```
expected_gpa = -22.44 + (0.09 ARABIC) + (0.21 ENGLISH) + (0.06
RELIGION) + (0.46 HISTORIC) + (0.55 MATH) + (0.05 PHYSIC) + (-
0.12 CHYMESTRY) + (0.02 BIOLOGY) + (-0.03 COMPUTER) + (-0.06 S
OCIAL)
```

Support Vector Machine

In [74]:

```
# Support Vector Machine Regression
svr = SVR(kernel='linear')
svr.fit(x_train_i, y_train_i)
calculate_accuracies(svr, x_train_i, x_test_i, y_train_i, y_test_i)
print()

svr = SVR(kernel='rbf')
svr.fit(x_train_i, y_train_i)
calculate_accuracies(svr, x_train_i, x_test_i, y_train_i, y_test_i)
print()

svr = SVR(kernel='poly')
svr.fit(x_train_i, y_train_i)
calculate_accuracies(svr, x_train_i, x_test_i, y_train_i, y_test_i)
```

```
Training   R2: 0.8229
Testing    R2: 0.8205

Training   MSE: 38.3615
Testing    MSE: 35.3218

Training   R2: 0.7890
Testing    R2: 0.8035

Training   MSE: 45.6991
Testing    MSE: 38.6574

Training   R2: 0.8977
Testing    R2: 0.8984

Training   MSE: 22.1593
Testing    MSE: 19.9842
```

Nearest Neighbour

In [75]:

```
# K Nearest Neighbour Regressor
knn = KNeighborsRegressor(n_neighbors=7)
knn.fit(x_train_i, y_train_i)
calculate_accuracies(knn, x_train_i, x_test_i, y_train_i, y_test_i)
```

```
Training   R2: 0.9297
Testing    R2: 0.8940

Training   MSE: 15.2303
Testing    MSE: 20.8624
```

Decision Tree

In [118]:

```
# Decision Tree Regressor
dt = DecisionTreeRegressor(max_depth=6)
dt.fit(x_train_i, y_train_i)
calculate_accuracies(dt, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.994609567940544
Testing    R2: 0.94766305484669

Training   MSE: 1.1676288615909078
Testing    MSE: 10.296043967265735
```

In [119]:

```
# Show tree structure as text
from sklearn.tree import export_text
print(export_text(dt))
```

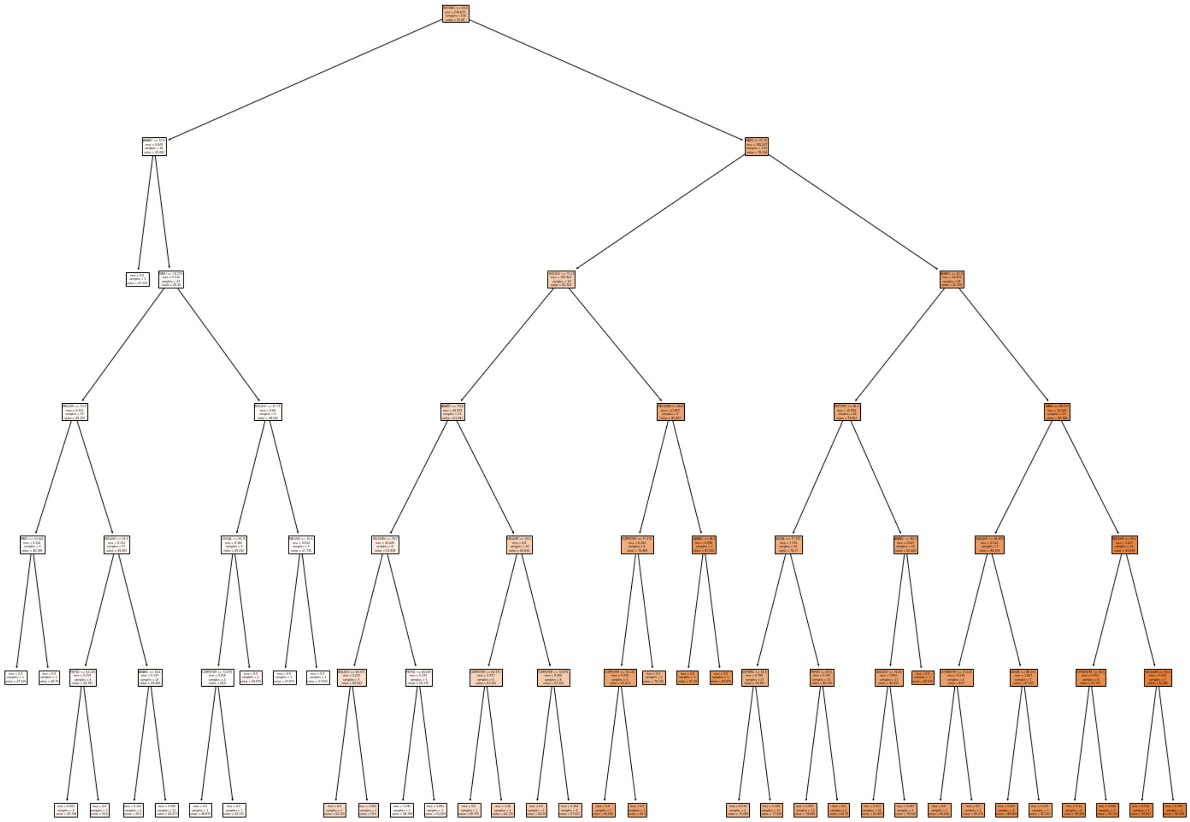
```
|--- feature_3 <= 60.00
|   |--- feature_0 <= 53.50
|   |   |--- value: [47.12]
|   |--- feature_0 >  53.50
|   |   |--- feature_4 <= 74.38
|   |   |   |--- feature_1 <= 53.30
|   |   |   |   |--- feature_4 <= 53.62
|   |   |   |   |   |--- value: [47.62]
|   |   |   |   |--- feature_4 >  53.62
|   |   |   |   |   |--- value: [48.75]
|   |   |   |--- feature_1 >  53.30
|   |   |   |   |--- feature_1 <= 55.50
|   |   |   |   |   |--- feature_5 <= 54.12
|   |   |   |   |   |   |--- value: [49.71]
|   |   |   |   |   |--- feature_5 >  54.12
|   |   |   |   |   |   |--- value: [50.00]
|   |   |   |   |--- feature_1 >  55.50
|   |   |   |   |   |--- feature_0 <= 58.40
|   |   |   |   |   |   |--- value: [49.00]
|   |   |   |   |   |--- feature_0 >  58.40
|   |   |   |   |   |   |--- value: [49.48]
|   |   |--- feature_4 >  74.38
|   |   |   |--- feature_7 <= 67.75
|   |   |   |   |--- feature_9 <= 69.75
|   |   |   |   |   |--- feature_8 <= 71.88
|   |   |   |   |   |   |--- value: [48.88]
|   |   |   |   |   |--- feature_8 >  71.88
|   |   |   |   |   |   |--- value: [49.12]
|   |   |   |   |--- feature_9 >  69.75
|   |   |   |   |   |--- value: [49.88]
|   |   |   |--- feature_7 >  67.75
|   |   |   |   |--- feature_1 <= 61.40
```

```
|   |   |   |   |   |--- value: [47.88]
|   |   |   |   |--- feature_1 >  61.40
|   |   |   |   |   |--- value: [47.62]
|--- feature_3 >  60.00
|   |--- feature_4 <= 71.75
|   |   |--- feature_7 <= 74.25
|   |   |   |--- feature_0 <= 79.40
|   |   |   |   |--- feature_2 <= 79.70
|   |   |   |   |   |--- feature_7 <= 63.38
|   |   |   |   |   |   |--- value: [62.12]
|   |   |   |   |   |--- feature_7 >  63.38
|   |   |   |   |   |   |--- value: [59.00]
|   |   |   |   |--- feature_2 >  79.70
|   |   |   |   |   |--- feature_5 <= 65.62
|   |   |   |   |   |   |--- value: [48.71]
|   |   |   |   |   |--- feature_5 >  65.62
|   |   |   |   |   |   |--- value: [52.62]
|   |   |   |--- feature_0 >  79.40
|   |   |   |   |--- feature_1 <= 83.50
|   |   |   |   |   |--- feature_8 <= 64.38
|   |   |   |   |   |   |--- value: [60.38]
|   |   |   |   |   |--- feature_8 >  64.38
|   |   |   |   |   |   |--- value: [64.17]
|   |   |   |   |--- feature_1 >  83.50
|   |   |   |   |   |--- feature_8 <= 70.88
|   |   |   |   |   |   |--- value: [66.25]
|   |   |   |   |   |--- feature_8 >  70.88
|   |   |   |   |   |   |--- value: [67.52]
|   |   |--- feature_7 >  74.25
|   |   |   |--- feature_2 <= 89.90
|   |   |   |   |--- feature_8 <= 77.38
|   |   |   |   |   |--- feature_8 <= 74.12
|   |   |   |   |   |   |--- value: [81.62]
|   |   |   |   |   |--- feature_8 >  74.12
|   |   |   |   |   |   |--- value: [82.00]
|   |   |   |   |--- feature_8 >  77.38
|   |   |   |   |   |--- value: [76.12]
|   |   |   |--- feature_2 >  89.90
|   |   |   |   |--- feature_0 <= 84.40
|   |   |   |   |   |--- value: [87.25]
|   |   |   |   |--- feature_0 >  84.40
|   |   |   |   |   |--- value: [87.88]
|   |--- feature_4 >  71.75
|   |   |--- feature_0 <= 82.80
|   |   |   |--- feature_3 <= 81.50
|   |   |   |   |--- feature_9 <= 77.38
|   |   |   |   |   |--- feature_3 <= 65.50
|   |   |   |   |   |   |--- value: [73.69]
|   |   |   |   |   |--- feature_3 >  65.50
|   |   |   |   |   |   |--- value: [77.57]
|   |   |   |   |--- feature_9 >  77.38
|   |   |   |   |   |--- feature_5 <= 87.50
```

```
|   |   |   |   |   |   |--- value: [79.61]
|   |   |   |   |   |--- feature_5 >  87.50
|   |   |   |   |   |   |--- value: [84.75]
|   |   |   |--- feature_3 >  81.50
|   |   |   |   |--- feature_0 <= 82.30
|   |   |   |   |   |--- feature_7 <= 76.25
|   |   |   |   |   |   |--- value: [82.66]
|   |   |   |   |   |--- feature_7 >  76.25
|   |   |   |   |   |   |--- value: [79.04]
|   |   |   |   |--- feature_0 >  82.30
|   |   |   |   |   |--- value: [89.62]
|   |   |--- feature_0 >  82.80
|   |   |   |--- feature_4 <= 89.38
|   |   |   |   |--- feature_7 <= 66.88
|   |   |   |   |   |--- feature_6 <= 73.12
|   |   |   |   |   |   |--- value: [83.62]
|   |   |   |   |   |--- feature_6 >  73.12
|   |   |   |   |   |   |--- value: [83.38]
|   |   |   |   |--- feature_7 >  66.88
|   |   |   |   |   |--- feature_9 <= 86.38
|   |   |   |   |   |   |--- value: [88.47]
|   |   |   |   |   |--- feature_9 >  86.38
|   |   |   |   |   |   |--- value: [85.33]
|   |   |   |--- feature_4 >  89.38
|   |   |   |   |--- feature_1 <= 93.20
|   |   |   |   |   |--- feature_6 <= 92.50
|   |   |   |   |   |   |--- value: [90.84]
|   |   |   |   |   |--- feature_6 >  92.50
|   |   |   |   |   |   |--- value: [92.31]
|   |   |   |   |--- feature_1 >  93.20
|   |   |   |   |   |--- feature_2 <= 93.60
|   |   |   |   |   |   |--- value: [93.67]
|   |   |   |   |   |--- feature_2 >  93.60
|   |   |   |   |   |   |--- value: [95.09]
```

In [120]:

```
# Show tree structure as plot
from sklearn.tree import plot_tree
plt.figure(figsize=(25,20))
plot_tree(dt,feature_names=x_train_s.columns, filled=True);
```



Random Forest

In [128]:

```
# Random Forest Regressor
rf = RandomForestRegressor(n_estimators=100, max_depth=20,
random_state=42)
rf.fit(x_train_i, y_train_i)
calculate_accuracies(rf, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.9941071294385533
Testing    R2: 0.9722373214488768

Training   MSE: 1.2764627527573529
Testing    MSE: 5.461643933823531
```

In [129]:

```
# Feature Importance
features = x_train_s.columns
importances = rf.feature_importances_
indices = np.argsort(importances)

plt.figure(figsize=(8,5))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices],
align='center')
```

```
plt.yticks(range(len(indices)), [features[i] for i in indices])
plt.xlabel('Relative Importance')
```

Out[129]:

```
Text(0.5, 0, 'Relative Importance')
```

Extra Trees

In [130]:

```
# Extra Trees Regressor
et = ExtraTreesRegressor(n_estimators=50, max_depth=17, random_state=42)
et.fit(x_train_i, y_train_i)
calculate_accuracies(et, x_train_i, x_test_i, y_train_i, y_test_i)
```

```
Training   R2: 0.999999960538571
Testing    R2: 0.984741183455248

Training   MSE: 8.54779411765032e-06
Testing    MSE: 3.001807720588238
```

In [131]:

```
# Feature Importance
features = x_train_s.columns
importances = et.feature_importances_
indices = np.argsort(importances)

plt.figure(figsize=(8,5))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices], align='center')
plt.yticks(range(len(indices)), [features[i] for i in indices])
plt.xlabel('Relative Importance')
```

Out[131]:

```
Text(0.5, 0, 'Relative Importance')
```

Gradient Boosting

In [132]:

```
# Gradient Boosting Regressor
gbr = GradientBoostingRegressor(n_estimators=25, max_depth=20, criterion='mse', random_state=42)
gbr.fit(x_train_i, y_train_i)
calculate_accuracies(gbr, x_train_i, x_test_i, y_train_i, y_test_i)
```

```
Training   R2: 0.9948462247926799
Testing    R2: 0.9690530708897035
```

```
Training   MSE: 1.11636629714353
Testing    MSE: 6.088069180157669
```

In [133]:

```
# Feature Importance
features = x_train_s.columns
importances = gbr.feature_importances_
indices = np.argsort(importances)

plt.figure(figsize=(8,5))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices],
align='center')
plt.yticks(range(len(indices)), [features[i] for i in
indices])
plt.xlabel('Relative Importance')
```

Out[133]:

```
Text(0.5, 0, 'Relative Importance')
```

Stacking

In [80]:

```
# Stacking Technique
et = ExtraTreesRegressor(n_estimators=50, max_depth=17,
random_state=42)
gbr = GradientBoostingRegressor(n_estimators=25, max_depth=20,
criterion='mse', random_state=42)
rf = RandomForestRegressor(n_estimators=100, max_depth=20,
random_state=42)
regressors = [('et', et), ('gbr', gbr), ('rf', rf)]

from sklearn.ensemble import StackingRegressor
sr = StackingRegressor(estimators=regressors,
final_estimator=LinearRegression())
sr.fit(x_train_i, y_train_i)
calculate_accuracies(sr, x_train_i, x_test_i, y_train_i,
y_test_i)
```

```
Training   R2: 0.9984
Testing    R2: 0.9821

Training   MSE: 0.3432
Testing    MSE: 3.5147
```

Artificial Neural Network

In [81]:

```
# ANN
model = models.Sequential([
    layers.Dense(10, activation='relu', input_dim = 10),
```

```
    layers.Dense(units = 8, activation='relu'),
    layers.Dense(units = 4, activation='relu'),
    layers.Dense(units = 1)
])
```

In [82]:

```
model.compile(optimizer = 'adam', loss = 'mean_squared_error')
```

In [83]:

```
early_stopp = callbacks.EarlyStopping(patience=5,
restore_best_weights=True)
```

In [84]:

```
history = model.fit(x_train_i, y_train_i,
validation_data=(x_test_i, y_test_i),
                    epochs=100, batch_size=125,
callbacks=[early_stopp])
```

```
Epoch 1/100
2/2 [==============================] - 1s 293ms/step - loss: 1
618.6349 - val_loss: 1327.4579
Epoch 2/100
2/2 [==============================] - 0s 54ms/step - loss: 13
80.5208 - val_loss: 1109.0869
Epoch 3/100
2/2 [==============================] - 0s 42ms/step - loss: 11
54.7411 - val_loss: 906.2011
Epoch 4/100
2/2 [==============================] - 0s 52ms/step - loss: 94
7.3502 - val_loss: 721.1635
Epoch 5/100
2/2 [==============================] - 0s 56ms/step - loss: 75
6.5266 - val_loss: 556.9769
Epoch 6/100
2/2 [==============================] - 0s 73ms/step - loss: 58
6.5121 - val_loss: 415.2522
Epoch 7/100
2/2 [==============================] - 0s 58ms/step - loss: 44
0.3659 - val_loss: 298.1283
Epoch 8/100
2/2 [==============================] - 0s 45ms/step - loss: 31
7.4913 - val_loss: 206.3595
Epoch 9/100
2/2 [==============================] - 0s 45ms/step - loss: 22
1.7663 - val_loss: 140.0979
Epoch 10/100
2/2 [==============================] - 0s 45ms/step - loss: 15
1.2513 - val_loss: 98.1662
Epoch 11/100
2/2 [==============================] - 0s 56ms/step - loss: 10
5.3379 - val_loss: 77.3191
Epoch 12/100
2/2 [==============================] - 0s 63ms/step - loss: 80
.1353 - val_loss: 72.7461
Epoch 13/100
```

```
2/2 [==============================] - 0s 50ms/step - loss: 72
.5620 - val_loss: 78.9059
Epoch 14/100
2/2 [==============================] - 0s 64ms/step - loss: 74
.9863 - val_loss: 90.2112
Epoch 15/100
2/2 [==============================] - 0s 60ms/step - loss: 84
.1944 - val_loss: 101.9969
Epoch 16/100
2/2 [==============================] - 0s 72ms/step - loss: 93
.9409 - val_loss: 110.4603
Epoch 17/100
2/2 [==============================] - 0s 59ms/step - loss: 10
1.2035 - val_loss: 113.2582
```

In [85]:

```
pd.DataFrame(history.history).plot(figsize=(10, 6))
```

Out[85]:

```
<AxesSubplot:>
```

In [86]:

```
print('Training   R2: ', r2_score(y_train_i,
model.predict(x_train_i)))
print('Testing    R2: ', r2_score(y_test_i,
model.predict(x_test_i)))
print()
print('Training   MSE: ', mean_squared_error(y_train_i,
model.predict(x_train_i)))
print('Testing    MSE: ', mean_squared_error(y_test_i,
model.predict(x_test_i)))
```

```
Training   R2:  0.6673721602932563
Testing    R2:  0.6302165682534584

Training   MSE:  72.05097133705964
Testing    MSE:  72.74605845786573
```

Plots

In [1]:

```
import matplotlib.pyplot as plt
import seaborn as sns
```

Linear Regression

In [82]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 89, 91.4, 86.7, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a',
'#000000']
```

In [83]:

```
plt.figure(figsize=(7,7))
```

```
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
plt.title('The Accuracy of Linear Regression Model\nfor Each
Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

Bayesian Ridge Regression

In [84]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 89, 91.5, 86.2, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a',
'#000000']
```

In [85]:

```
plt.figure(figsize=(7,7))
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
plt.title('The Accuracy of Bayesian Ridge Regression
Model\nfor Each Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

Support Vector Machine

In [86]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 91.5, 94.9, 89.8, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a',
'#000000']
```

In [87]:

```
plt.figure(figsize=(7,7))
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
plt.title('The Accuracy of Support Vector Machine Model\nfor
Each Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
```

```
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

K Nearest Neighbour

In [88]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 93.3, 94.7, 89.4, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a',
'#000000']
```

In [90]:

```
plt.figure(figsize=(7,7))
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
plt.title('The Accuracy of K Nearest Neighbour Model\nfor Each
Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

Decision Tree

In [91]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 99.2, 97.9, 94.4, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a',
'#000000']
```

In [92]:

```
plt.figure(figsize=(7,7))
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
plt.title('The Accuracy of Decision Tree Model\nfor Each
Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

Random Forest

In [93]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 98.5, 97, 97.2, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a',
'#000000']
```

In [94]:

```
plt.figure(figsize=(7,7))
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
plt.title('The Accuracy of Random Forest Model\nfor Each
Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

Extra Trees

In [95]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 99.56, 98.7, 98.4, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a',
'#000000']
```

In [96]:

```
plt.figure(figsize=(7,7))
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
plt.title('The Accuracy of Extra Trees Model\nfor Each
Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

Gradient Boosting

In [97]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 99.52, 96, 96.9, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a',
'#000000']
```

In [98]:

```
plt.figure(figsize=(7,7))
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
```

```
plt.title('The Accuracy of Gradient Boosting Model\nfor Each Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

Stacking

In [99]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 99.5, 98.5, 98.2, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a', '#000000']
```

In [100]:

```
plt.figure(figsize=(7,7))
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
plt.title('The Accuracy of Stacking Regressor Tree Model\nfor Each Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

Artificial Neural Network

In [101]:

```
departments = ['', 'Scientific', 'Literary', 'Industrial', '']
accuracies = [0, 68.9, 0, 63, 0]
depart_positions = [0,1,2,3,4]
colors = ['#000000', '#3274a1', '#e1812c', '#3a923a', '#000000']
```

In [102]:

```
plt.figure(figsize=(7,7))
plt.bar(depart_positions, accuracies, width=0.3, color=colors)
plt.title('The Accuracy of Artificial Neural Network\nfor Each Department', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.xticks(depart_positions, departments, rotation=45)
plt.ylim([0,100])
plt.xlim([0,4])
plt.tight_layout()
```

All Together

In [112]:

```
models = ['LR']*3 + ['Bayesian Ridge']*3 + ['SVM']*3 +
['KNN']*3\
         + ['Decision Tree']*3 + ['Random Forest']*3 + ['Extra
Trees']*3\
         + ['Gradient Boost']*3 + ['Stacking']*3 + ['ANN']*3
```

In [106]:

```
departments = ['Scientific', 'Literary', 'Industrial'] * 10
```

In [114]:

```
accuracy = [  89, 91.4, 86.7,
              89, 91.5, 86.2,
              91.5, 94.9, 89.8,
              93.3, 94.7, 89.4,
              99.2, 97.9, 94.4,
              98.5, 97, 97.2,
              99.56, 98.7, 98.4,
              99.52, 96, 96.9,
              99.5, 98.5, 98.2,
              68.9, 0, 63  ]
```

In [115]:

```
import pandas as pd

df = pd.DataFrame({
   'model' : models,
   'department':departments,
   'accuracy':accuracies
})
df
```

Out[115]:

In [121]:

```
list(set(models))
```

Out[121]:

```
['Stacking',
 'Gradient Boost',
 'Bayesian Ridge',
 'SVM',
 'LR',
 'Decision Tree',
 'Extra Trees',
 'ANN',
```

```
 'KNN',
 'Random Forest']
```

In [122]:

```
df[df['model'] == 'Stacking']
```

Out[122]:

	Model	Department	Accuracy
24	Stacking	Scientific	99.5
25	Stacking	Literary	98.5
26	Stacking	Industrial	98.2

In [136]:

```
plt.figure(figsize=(7,7))

for model in list(set(models)):
    if model != 'ANN':
        selected_model = df[df['model'] == model]
        plt.plot(selected_model['department'],
selected_model['accuracy'], 'o-' , label=model)

plt.title('The Accuracy of All Model Versus Each Other',
pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.legend()
plt.ylim(85, 100);
```

In [140]:

```
plt.figure(figsize=(7,7))

selected_model = df[df['model'] == 'ANN']
plt.plot(selected_model['department'],
selected_model['accuracy'], 'ko-' , label='ANN')

plt.title('The Accuracy of ANN', pad=15)
plt.xlabel('Department', labelpad=15)
plt.ylabel('R2 Score', labelpad=15)
plt.legend()
plt.ylim(0, 70);
```

Appendix D

ETHICAL APPROVAL LETTER

YAKIN DOĞU ÜNİVERSİTESİ

BİLİMSEL ARAŞTIRMALAR ETİK KURULU

16.11.2021

Dear Mustafa Ababneh

Your application titled **"Data Mining Mobile Application to Guiding the Students in High School by Using Machine Learning"** with the application number NEU/AS/2021/129 has been evaluated by the Scientific Research Ethics Committee and granted approval. You can start your research on the condition that you will abide by the information provided in your application form.

Assoc. Prof. Dr. Direnç Kanol

Rapporteur of theScientificResearchEthicsCommittee

Direnc Kanol

Note: If you need to provide an official letter to an institution with the signature of the Head of NEU Scientific Research Ethics Committee, please apply to the secretariat of the ethics committee by showing this document.

Appendix X

Turnitin Similarity Report

Thesis

by Mustafa Ababneh

Submission date: 15-Mar-2023 03:33PM (UTC+0200)
Submission ID: 2037768594
File name: Thesis__Mustafa_15_march2023.docx (6.33M)
Word count: 51573
Character count: 306179

Thesis

ORIGINALITY REPORT

15%	11%	7%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	rishit-dagli.github.io Internet Source	1%
2	Submitted to New College of the Humanities Student Paper	1%
3	www.researchgate.net Internet Source	1%
4	github.com Internet Source	1%
5	Submitted to Infile Student Paper	<1%
6	Submitted to University of Brighton Student Paper	<1%
7	Sayan Mukhopadhyay, Pratip Samanta. "Chapter 5 Deep Learning and Neural Networks", Springer Science and Business Media LLC, 2023 Publication	<1%
8	Submitted to University of Oxford Student Paper	<1%

CURRICULUM VITAE

PERSONAL DATA

Name: Mustafa Abdel-Karim Ahmad

Nationality: Jordanian

Date of Birth: April 26th, 1986

Marital Status: Married

Mobile: **+97333579370**
E- mail: Mustafa.ababneh86@gmail.com

Languages:

Arabic Mother Language, full command of speaking, reading and writing.

English Fluent in writing, speaking and reading.

Objective

Seeking a Trainer position where I can share my knowledge with passion to enlighten the trainees to excel is computer science field with innovation technologies.

Educational Qualifications

2023 PhD in computer information system (Artificial Intelligence) at Near East University

2019 Master degree in Computer Science; 3.93 Overall GPA at Amman Arab University in filed Information Retrieval and Database.

2010 Bachelor degree (BA computer information system) at Jordan University of Science and Technology

Work Experience (x Years)

2010 – 2011 was working as Math teacher in Directorate of Military Education and Culture

2011 - 2012 was working as an IT teacher in a private school “Al Arabia School Irbid”.

2012 - 2013 was working as an IT teacher in Ministry of Education in Alhushain Bin Talal School.

2013 - 2022 “IT teacher and General Manager” in the Cambridge Academy (My private academy it is specialist in languages teaching and university material teaching).

2017 - 2019 was working as a part time in laboratory instructor for (C++, SQL, HTML, Vb) in Amman Arab University.

2021 - 2023 was working as an AI Architect in creative cloud company, Sudia Arabia.

PUBLICATIONS

[1] Al-Jarrah, Ayat Amin, Ghassan Kanaan, and Mustafa Abdel-Kareem Ababneh. "Retrieving Arabic Textual Documents Based on Queries Written in Bahraini Slang Language." Modern Applied Science 13.6 (2019).

[2] Ababneh, Mustafa Abdel-Kareem, Ghassan Kanaan, and Ayat Amin Al-Jarrah. "Enhanced Arabic Information Retrieval by Using Arabic Slang." Modern Applied Science 13.6 (2019).

[3] M. Ababneh, A. Al-Jarrah, D. Karagozlu. "The Role of Big Data and Machine Learning in COVID-19". In BRAIN. Broad Research in Artificial Intelligence and Neuroscience on Web Science. 2020.

[4] Ababneh, M., Aljarrah, A., Karagozlu, D., & Ozdamli, F. (2021). Guiding the Students in High School by Using Machine Learning.

[5] Aljarrah, A., Ababneh, M., Karagozlu, D., & Ozdamli, F. (2021). Artificial Intelligence Techniques for Distance Education: A Systematic Literature Review. Tem Journal-Technology Education Management Informatics, 1621-1629.

[6] Ozdamli, F., Aljarrah, A., Karagozlu, D., & Ababneh, M. (2022). Facial Recognition System to Detect Student Emotions and Cheating in Distance Learning. Sustainability, 14(20), 13230.

[7] Ozdamli, F., Ababneh, M., Karagozlu, D., & Aljarrah, A. (2022). Development and Testing of Performance Scale Application as an Effective Electronic Tool to Enhance Students' Academic Achievements. Electronics, 11(23), 4023.

[8] M. Ababneh, A. Al-Jarrah, N. Cavus. Social media usage in education: Big Data Perspective." Proceedings of the 11th ACM Conference on Web Science. 2020.

[9] Aljarrah, A. A., Ababneh, M. A. K., & Cavus, N. (2020). The role of massive open online courses during the COVID-19 era: Challenges and perspective. New Trends and Issues Proceedings on Humanities and Social Sciences, 7(3), 142-152.

Courses Taught

[1] Programming Basics in JavaScript through Grasshopper from Google.

[2] Python Programming Course using PyCharm from Python Institute.

[3] Machine learning Course with Raspberry pi from IBM.

[4] Data base programming course from Amman Arab University.

[5] Multimedia Systems.

[6] Introduction to Information Technology from Pearson level 3&4.

[7] Mobile Application Development from Pearson level 3&4.

[8] Computer vision from Pearson level 3&4.
[9] Principles of Programming Languages from Pearson level 3&4.
[10] Big data management from Pearson level 3&4.
[11] Deep Learning from Pearson level 3&4.
[12] Data Mining from Pearson level 3&4.

Computer Skills

Operating Systems	:	Microsoft Windows, Mac, Linux
Languages	:	C, C++, C#, Java, JavaScript, HTML, python
Scripting	:	JavaScript
Databases	:	Oracle

REFEREES

[1] Assist. Professor Dr. Damla Karagozlu, Computer Education and Instructional Technology department, Cyprus International University, Mersin 10, Turkey

E-mail: dkaragozlu@ciu.edu.tr

[2] Assoc. Prof. Dr. Fezile Özdamli, Computer Education and Instructional Technology department, Near East University, Mersin 10, Turkey

E-mail: fezile.ozdamli@neu.edu.tr

[3] Prof. Ghassan Kanaan, Dean of IT College, Amman Arab University, Jordan Mobile: +962-795687155 E-mail: Ghassan.Kanaan@yahoo.com

[4] Assoc. Prof. Dr. Mohammed Shatnawi, Faculty of Computer and Information Technology, Jordan University of Science & Technology (JUST), Jordan Mobile: +962-797266261 E-mail: mshatnawi@just.edu.jo