USMAN ABIDEMI SARUMI

STUDENT ACADEMIC PERFORMANCE PREDICTION WITH WI-FI IoT ATTENDANCE SYSTEM

MASTER THESIS

2023

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF ARTIFICIAL INTELLIGENCE ENGINEERING

STUDENT ACADEMIC PERFORMANCE PREDICTION WITH WI-FI IoT ATTENDANCE SYSTEM

M.Sc. THESIS

Usman Abidemi SARUMI

Nicosia

June, 2023

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF ARTIFICIAL INTELLIGENCE ENGINEERING

STUDENT ACADEMIC PERFORMANCE PREDICTION WITH WI-FI IoT ATTENDANCE SYSTEM

M.Sc. THESIS

Usman Abidemi SARUMI

Supervisor

Prof. Dr. Fadi Al-Turjman

Nicosia

June, 2023

Approval

We certify that we read the thesis submitted by **Paul Biriko BEWA** titled "Analysis of International Conflict Resolution Mechanisms: Impact of the United Nations in the Aggression in Eastern DR Congo". In addition, we are of the view that it fulfils all the requirements, both in terms of its breadth and its level of quality, to be a thesis for the Master of Law degree.

Jury Members	Name-Surname	Signature
Head of the Committee:	Asst. Prof. Dr. Özde Dereboylular	.....
Committee Member:	Dr. Erkut Ziya Sivrikaya	.....
Supervisor:	Asst. Prof. Dr. Perçem Arman	.....

Approved by the Head of the Department

06/09/2023

.....

Asst. Prof. Dr. Nabi Berkut

Approved by the Institute of Graduate Studies

..../..../.....

.....

Prof. Dr. Kemal Hüsnü Can Başer

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.

Usman Abidemi Sarumi

…./…./2023.

Day/Month/Year

Sarumi Usman Abidemi: STUDENT ACADEMIC PERFORMANCE PREDICTION WITH WI-FI IoT ATTENDANCE SYSTEM
Approval of Directore of Institute of Graduate studied
Prof.Dr.K.Husnu Can Baseri

We certify this thesis is satisfactory for the Award of the Degree of Master In Artificial Intelligence

Examining Comimittee in charge:

Prof.Fadi Al-Turjman	**Supervisor, Department of Artificial Intelligence Engineering, NEU**
Assoc. Prof. Sertan Serte	**Committee member, Department of Electrical & Electronics Engineering, NEU**
Asst. Prof. Dr. Pwadubashiyi Coston PWAVODİ	**Committee member, Department of Biomedical Engineering, CIU**

Acknowledgement

Firstly, I want to give God the glory for guiding me through my research and making it successful. I would like to acknowledge my research supervisor Prof. Dr. Fadi Al-Turjman for all of his advice, wisdom, and expertise. An appreciative of his wonderful sense of humor, inspiration, and constructive feedback. I would like to express my gratitude to Prof. Ali serener and Dr. Mubarak Saleh for their support in making this effort feasible. Thanks to their guidance and advice, I was able to complete every writing aspect of my assignment. I appreciate you.

I would want to express my deep gratitude to my parents, Mr. and Mrs. Musifiu Sarumi, for their unwavering support, love, and sacrifices throughout my life. I would especially like to thank my family for their inspiration and support. I want to thank Mercel and Tommy, two of my co-workers and friends, for their compassion, sincere support, and willingness to be there for me.

Finally, I want to express my gratitude to everyone who helped me, directly or indirectly, finish my research.

Abstract

“STUDENT ACADEMIC PERFORMANCE PREDICTION WITH WI-FI IoT ATTENDANCE SYSTEM”

SARUMI USMAN ABIDEMI

M.sc, Department of Artificial Intelligence

May 2023,

In today's education landscape, student engagement and academic performance are crucial aspects of effective learning. One key factor that can motivate students to exert the necessary effort is the ability to monitor their academic progress in real time. Traditional methods of performance evaluation, and choosing features such as end-of-semester grades, parental influence, age, gender etc., often fall short in providing timely feedback for remediation. To address this challenge, researchers have explored the use of machine learning techniques to enable students to predict their performance at any given moment. However, existing approaches in the literature have a significant limitation: they are not trained with real-time collected data.

This Master's thesis presents a pioneering study that tackles the aforementioned limitation by designing an AI and IoT-based system capable of predicting student performance based on real-time attendance records and GPA. To obtain accurate and up-to-date attendance data, we have implemented a WiFi-based attendance taking system. This system leverages short-range Wi-Fi technology integrated into a mobile-friendly web application, ensuring that students can mark their class participation by sending a post request to a remote server. This approach effectively eliminates the potential for proxy-attendance, thus enhancing the reliability and authenticity of the attendance records.

The core focus of this research lies in developing a machine learning model that utilizes the collected attendance data to predict student performance in a timely manner. By employing advanced machine learning algorithms and techniques,

including but not limited to classification and regression models, feature selection, etc., our system offers an accurate and reliable performance prediction framework.

To validate the effectiveness of our proposed system, a comprehensive evaluation is conducted by comparing its performance against existing approaches reported in the literature. The evaluation metrics include accuracy, reliability, and predictive capability. The results of this study clearly demonstrate that our AI and IoT-based attendance monitoring system achieves higher accuracy and reliability in predicting student performance, outperforming the existing methods in the literature.

The contributions of this research extend beyond academic evaluation, as the developed system holds the potential to revolutionize educational practices. By providing students with real-time performance predictions, educators can identify struggling individuals early on and implement targeted interventions for improved learning outcomes. Furthermore, the integration of AI and IoT technologies in education fosters a data-driven and adaptive learning environment, enhancing student engagement and promoting personalized education.

Keywords; student academic performance, attendance record, machine learning, Internet of things, classification.

ÖZET

"STUDENT ACADEMIC PERFORMANCE PREDICTION WITH WI-FI IoT ATTENDANCE SYSTEM"

SARUMI USMAN ABIDEMI
M.sc, Department of Artificial Intelligence
May 2023,

Günümüzün eğitim ortamında öğrenci katılımı ve akademik performans, etkili öğrenmenin önemli unsurlarıdır. Öğrencileri gerekli çabayı göstermeye motive edebilecek temel faktörlerden biri, akademik ilerlemelerini gerçek zamanlı olarak izleyebilmektir. Geleneksel performans değerlendirme yöntemleri ve dönem sonu notları, ebeveyn etkisi, yaş, cinsiyet vb. gibi özelliklerin seçilmesi, iyileştirme için zamanında geri bildirim sağlamada genellikle yetersiz kalır. Bu zorluğun üstesinden gelmek için araştırmacılar, öğrencilerin herhangi bir andaki performanslarını tahmin etmelerini sağlayacak makine öğrenimi tekniklerinin kullanımını araştırdılar. Ancak literatürdeki mevcut yaklaşımların önemli bir sınırlaması vardır: gerçek zamanlı olarak toplanan verilerle eğitilmezler.

Bu Yüksek Lisans tezi, gerçek zamanlı devam kayıtlarına ve genel not ortalamasına dayalı olarak öğrenci performansını tahmin edebilen yapay zeka ve IoT tabanlı bir sistem tasarlayarak yukarıda belirtilen sınırlamayı ortadan kaldıran öncü bir çalışma sunmaktadır. Doğru ve güncel katılım verilerini elde etmek için WiFi tabanlı bir yoklama alma sistemi uyguladık. Bu sistem, mobil uyumlu bir web uygulamasına entegre edilmiş kısa menzilli Wi-Fi teknolojisinden yararlanarak öğrencilerin uzak bir sunucuya bir gönderi isteği göndererek derse katılımlarını işaretleyebilmelerini sağlar. Bu yaklaşım, vekaleten katılım potansiyelini etkili bir şekilde ortadan kaldırır, böylece katılım kayıtlarının güvenilirliğini ve gerçekliğini artırır.

Bu araştırmanın temel odağı, öğrenci performansını zamanında tahmin etmek için toplanan devam verilerini kullanan bir makine öğrenimi modeli geliştirmektir. Sistemimiz, sınıflandırma ve regresyon modelleri, özellik seçimi vb. dahil ancak

bunlarla sınırlı olmamak üzere gelişmiş makine öğrenimi algoritmaları ve tekniklerini kullanarak doğru ve güvenilir bir performans tahmin çerçevesi sunar.

Önerilen sistemimizin etkinliğini doğrulamak için, performansını literatürde bildirilen mevcut yaklaşımlarla karşılaştırarak kapsamlı bir değerlendirme yapılmaktadır. Değerlendirme ölçütleri doğruluk, güvenilirlik ve tahmin yeteneğini içerir. Bu çalışmanın sonuçları, yapay zeka ve IoT tabanlı devam izleme sistemimizin, öğrenci performansını tahmin etmede literatürdeki mevcut yöntemlerden daha iyi performans göstererek daha yüksek doğruluk ve güvenilirliğe ulaştığını açıkça göstermektedir.

Geliştirilen sistem eğitim uygulamalarında devrim yaratma potansiyeline sahip olduğundan, bu araştırmanın katkıları akademik değerlendirmenin ötesine uzanmaktadır. Eğitimciler, öğrencilere gerçek zamanlı performans tahminleri sunarak, zorluk yaşayan bireyleri erkenden tespit edebilir ve daha iyi öğrenme sonuçları için hedeflenen müdahaleleri uygulayabilir. Ayrıca, AI ve IoT teknolojilerinin eğitime entegrasyonu, veri odaklı ve uyarlanabilir bir öğrenme ortamını teşvik ederek öğrenci katılımını artırır ve kişiselleştirilmiş eğitimi teşvik eder.

Anahtar Kelimeler; öğrenci akademik performansı, devam kaydı, makine öğrenimi, nesnelerin interneti, sınıflandırma

Table of Contents

CHAPTER IV

CHAPTER V

List of Table

Page

List of figures

Page

List of Abbreviations

IoT:	Internet of Things
ML:	Machine Learning
RF:	Random Forest
LR:	Logistic Regression
SVM:	Support Vector Machine
GPA:	Grade point assessment
WI-FI:	Wireless Fidelity
J48:	Homogenous Ensemble Decision Tree Classifier
DT:	Decision Tree
MLP:	Multilayer perceptron
OneR:	One Rule Algorithm
AI:	Artificial Intelligence
JRip:	Joint Reserve Intelligence Program

CHAPTER I

1. INTRODUCTION

1.1. Background of Study

The educational sector is one of the essential parts that contribute development to a country, due to this crucial impact, every developed country has evaluation criteria in their educational system. For any educational system to be successful, a standard and accurate evaluation method must be incorporated in the system. The quick digitization of educational institutions has a huge impact on how data collection is facilitated for educators and employees. Along with changes in learning styles and governance of educational institutions from previously offline-based to online-based, the abundance of data on educational institutions is also moving extremely quickly. Explicit programming, which is currently frequently employed as a technique to provide solutions to problems, is no longer sufficient in some situations, much like other industries affected by the pattern of modern Industry. In this situation, machine learning bridges the gap and offers a method for handling this enormous and dynamic data. ("Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review," 2021). Part of The main necessity for all stakeholders, including students, teachers, parents, administrators management, industry, and the environment, is that pupils succeed academically in any educational institution or higher education institutions can improve professionally and academically with regular input from all stakeholders, but they must also embrace cutting-edge technologies that can accelerate their growth. Identifying at-risk students early, projecting an appropriate branch or course, and early success prediction of students utilizing trending artificial intelligence technologies like machine learning can also assist management and students enhance their academic performance (Saluja et al., 2023). In modern educational systems there are a lot of methods used to measure student academic performance due to it crucial part it takes in the job market, this is one of the reasons why student, lecturers and parents put more attention to academic success achieved during study (Carvalho Brito et al., 2008). Several data are been generated from educational institution by applying different technique e.g. attendance records generated from Internet of things Wi-Fi attendance system (Abidemi et al., 2022) , data mining etc. Basically student academic performance is always measured by their individual CGPA after the

evaluation from examination result obtained; Therefore if any institution has the ability to forecast student performance in advance, it will be a very impressive move to improve the academic achievement of the students (Yadav et al., 2012b). The most thorough reflection of what student learned in a course is their grades, and as a result, they have a direct impact on their overall academic achievement, this is due to the rapid development of data and internet of things technologies. Tutors should maintain track of students' academic progress because they act as a means of guidance in the process of their development and successful outcome, they should think about how to turn their backlogs into progress, assisting pupils in making academic plans in preparation and thoroughly examining the situation. Student academic achievement, a crucial component of academic work, there have been claims that the grades of the students are strongly correlated with the submission status of assignments, project reports, laboratory reports, exams and attendance records (Tao et al., 2022). It is difficult to fully predict student achievement using different educational data today because of the dirty properties of the data, such as isolated data or low content information in these systems, and the lack of effective methods. Today, all these data, which are generated in large volumes every day, may not be used effectively due to these factors. The same data in multiple systems may have different names because there isn't a single data standard, and it may also be inconsistent since it isn't timely updated or synchronized. By fixing the aforementioned issues with campus information systems, teachers will receive useful feedback, be able to forecast student progress, and have useful assistance in making important decisions(Qu et al., 2018).

Predicting student academic achievement has long been a hot topic in educational research, spanning a wide range of academic fields and grade levels. It is crucial for educators to be capable of predicting students' academic performance because it enables early identification of those who need more assistance, aids in the design and implementation of timely interventions, and broadens our theoretical understanding of the variables that can affect academic outcomes. Although there were considerable variances in the variables that the various researchers selected as being the most crucial, the various models' predictive accuracy varied. Adopting ML techniques and methodologies can help educational research create models that can be used for making predictions on data as well as for elaborating on the causes of predictors of student performance in their academic activities (Puah, 2021).

Furthermore, one of the main resources of the university are there students, the quality of the students' work will determine which graduates and postgraduates are intelligent and capable enough to lead the country and make up its labour force, and be in charge of its social and economic growth. The performance of university students should concern employers in the labour market, school administrators and teachers. Academic achievement is one of the most crucial factors employers look at when hiring new hires, especially recent graduates. Therefore, students must put in the most effort in their academic work to satisfy the demands of the company. The Cumulative Grade Point Average (CGPA) is a metric used to assess students' academic performance. The CGPA which accounts for the average of all exam grades over all university semesters shows a student's gross academic success. Numerous factors may both hinder and help students achieve a high CGPA that accurately represents their overall academic performance. The introduction of information technology has produced the storage of enormous amounts of data in a variety of formats, including student, teacher, alumni, resource, and other data. To make better decisions using the data gathered from various applications, a suitable method of knowledge extraction from vast repositories is needed (Yadav et al., 2012a). As the lecturer is required by higher education institutions to track and record student attendance, attending class is very crucial. Unfortunately, it becomes a very important problem to immediately and accurately record attendance. There is a notion that student attendance might predict academic performance, the conventional method of keeping attendance records is obsolete, and modern technologies like the internet of things, image processing, and attendance records are more practical. Due to convenience and concealment, identity verification and recognition using student ID or number characteristics have grown more commonplace in the digital age (Ismail et al., 2022).

Moreover, Academic success is typically seen as a direct correlation with college attendance. Regular attendance often improves learning since it allows for more one-on-one time with teachers and classmates. Those who support making attendance in class a requirement believe that attendance in class positively corresponds with academic achievement. It was suggested that a student's academic performance, including their skills and aptitudes, attendance in class, and other factors, contribute to their ability to succeed academically. Studies have found a strong correlation between high school success and college success, with Regardless of a student's

personal situation, indicators like GPA and class attendances are critical to assessing academic performance in higher education. In essence, the most important indicator of college success, stating that "relationship between student's attendance and college GPA is so powerful that it would seem more important for colleges to fully consider this measure in deciding on placement. And also regardless of the quality or type of school attended, cumulative grade point average (GPA) in academic subjects in high school has proven to be the best overall predictor of student performance in college (Hazaa et al., 2021) .

This study (Gump, 2010) shows that going to class is essential for students who want to excel in their academic activities. Even though there are many other elements that affect academic achievement, going to class is one of the simplest ones that students can manage, thus it is important to emphasize its significance. However, it is crucial to emphasize and give students chances to practice the abilities that can be used to optimize the advantages of attending class, such as question-asking techniques and note-taking techniques. Additionally, these results are consistent with a study that shows a high GPA can be used to predict first-year undergraduate student performance and provide information for selecting the right courses, both of which could boost the likelihood that students will succeed in college courses.

So, this project aims to explore class attendance record of every students in correlation with the CGPA which may be used to forecast academic success by inserting this data into a machine learning algorithm such as logistic regression, support vector machine, linear regression etc., this predictive model used can provide instinctive and better study advice based on data generated. Figure 1 indicates how these two technologies integrate with each other to achieve a great task.

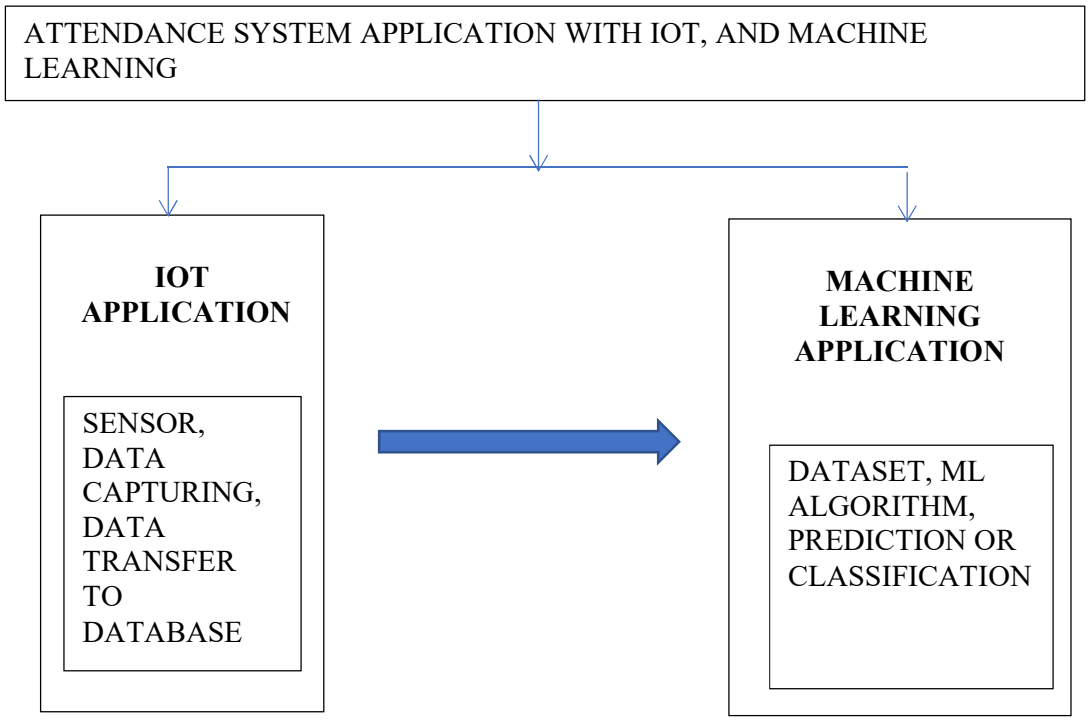


Figure 1: The connection of IoT attendance system and machine learning application

Description of Machine Learning:

Artificial intelligence (AI) research in the area of machine learning focuses on creating algorithms and models that automatically learn from experience and get better over time without explicit programming. As shown in Figure 2, the main objective of machine learning is to make it possible for computers to analyse and interpret complex data, spot patterns, and make predictions or judgments based on that data. When employing a dataset containing input data and associated output labels or target values, a model is trained as part of the learning process in machine learning. The model learns from this labelled data by identifying patterns and relationships that exist between the inputs and outputs, it then generalizes these patterns to make predictions or decisions on new, unseen data.

Machine learning algorithms can be categorized into different types, such as supervised learning, unsupervised learning, and reinforcement learning. Supervised learning involves utilizing labelled data to train a model., unsupervised learning focuses on discovering patterns or structures in unlabelled data, and reinforcement learning deals with training an agent to decide in accordance with interactions with an environment and receiving rewards or penalties.

The typical steps involved in machine learning include:

1. Data collection: Gathering relevant and representative data that contains the patterns and relationships to be learned. This data can originate from a variety of sources, such as database, sensors, or user interactions.

2. Data pre-processing: Cleaning and preparing the data for analysis. This step may involve removing outliers, handling missing values, normalizing or scaling features, and splitting the data into training and testing sets.

3. Model selection: Choosing an appropriate machine learning model or algorithm that is best suited for the problem at hand. The choice of model depends on variables like the type of data, the complexity of the problem, and the available computational resources.

4. Training: Using the training data to train the machine learning model. The model learns from the input data by adjusting its internal parameters or weights based on the patterns and relationships present in the training examples.

5. Evaluation: Assessing performance evaluation of the trained model using test data. The model's predictions or decisions are compared to the true values or labels in the testing set to measure its accuracy and generalization ability.

6. Model refinement: Fine-tuning the model based on the evaluation results. This step may involve adjusting hyper-parameters (settings that control the learning process) or exploring different model architectures to improve performance.

7. Prediction or decision-making: Once the model is trained and evaluated, On the basis of fresh, unforeseen facts, it can be used to create predictions or judgments. The model processes input data, employs previously discovered patterns, and generates an outcome or forecast.

Furthermore, Machine learning has several uses in a variety of fields, such as speech and image recognition, natural language processing, recommendation systems, fraud

detection, autonomous cars, and more. It enables computers that use data to learn and adapt their behaviour to perform tasks more accurately and efficiently than traditional rule-based programming approaches.

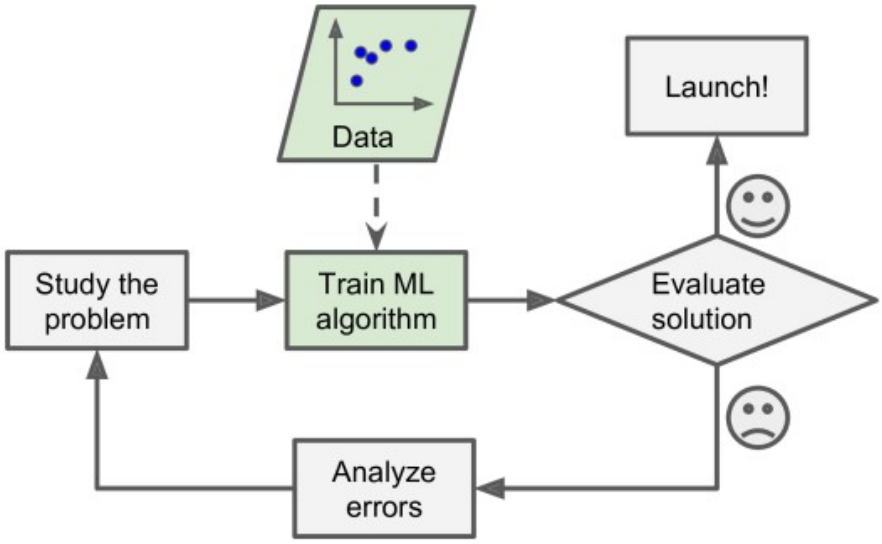


Figure 2: Machine learning approach

1.2. Statement of the problem:

After several reviews of related project to predict student academic performance using some factors for feature selection, it was noticed that most of the previous works have low accuracy, this project provides a high accuracy that tend to perform better for the academic prediction purpose using two very important features (student attendance record and GPA) that have significant role to play for student academic evaluation. These selected features on a dataset will be used to train a model for better performance in other to create the awareness to lecturer and student.

1.3. Aim of the research:

The aim of this project is to design a prediction system that will make use of students GPA and attendance data generated from IoT Wi-Fi attendance system to train a machine learning model to perform student academic performance prediction by investigate how machine learning methods can be utilized for forecasting student achievement and at-risk student in other to make improvement. Also, to examine the

connection between attendance and GPA to see if these factors can serve as trustworthy predictors of a student's success or failure.

1.4. Objective of the research:

The objective of the study is to develop a predictive model that can accurately forecast students' academic performance based on their attendance records and Grade Point Average (GPA).

The specific objectives of the study are as follows:

1. Analyse the correlation between attendance records and GPA: The study seeks to examine the extent to which attendance patterns are associated with a student's GPA. By analysing historical data, it aims to identify any patterns or trends that suggest a significant relationship between attendance and academic performance.
2. Develop a predictive model: Using machine learning or statistical techniques, the study aims to build a predictive model that can forecast student academic performance based on attendance records and GPA. The model should accurately predict whether a student is likely to excel or struggle academically, using the two variables as input features.
3. Evaluate the model's performance: The study will assess the predictive model's accuracy, precision, recall, and other relevant metrics to determine its effectiveness in predicting student academic performance. This evaluation will help determine the model's reliability and practical applicability.
4. Provide insights for intervention strategies: Based on the findings of the study, the objective is to provide insights and recommendations for educational institutions to develop intervention strategies. These strategies can be targeted towards students at risk of underperforming academically, identified using the predictive model, with the aim of improving their attendance and GPA.

1.4. Limitations of the Research

Previous research has examined several methods for prediction of student academic performance with many drawbacks; including using non numerical data input, low accuracy algorithm, consideration of too many factors for the algorithm inputs. Only two features are used in this research and one of the features i.e. the GPA is input manually in combination with generated attendance record to form our dataset while

other features like student number, course title and codes, class name, date and time are removed for effective dataset normalization.

1.5. Structure of the thesis

The thesis is split into six chapters, as well as a conclusion, an appendix, and references, make up the dissertation. An overview of the research and its setting is given in Chapter 1. In Chapter 2's literature review. The theoretical foundation that will be needed to understand the rest of this thesis is introduced in this chapter. Reference to relevant works and comparison an overview of research methods and a description of the study's methodology are provided in Chapter 3. Chapter 4 provides an overview of the Performance and Result Analysis and in Chapter 5 the conclusion.

CHAPTER II

2.0. LITERATURE REVIEW

In a research project (Hassan et al., 2021), the authors of this research work emphasis on students who run the danger caused by not completing high school at expected time, there research methodology serves as a beginning to venture for future research on other poor academic outcomes. They concentrate on a type of framework formulate the problem, extract the necessary feature, perform classification, evaluate the applies criteria and generalizes the unfavourable academic output, such as under matching or not applicable to college, they describe the important features to achieve this task, and also many classification algorithms are applied, which their performance is assessed using metrics that are significant to school administrators in this research using machine learning framework to described and find such students. In (Ofelia San Pedro et al., 2013) it was stated that a software model was used as types of detectors to predict college attendance for a good number of students who are using the system. These models may offer more useful information compare to the existing models for prediction if they can be used to enrolment prediction for the college students, there detection is effective and possibly more actionable than prior methods; the software is capable to tell the difference between a student who plans to go to college and one who won't 68.6% of the time. To demonstrate Decision Tree, which had the highest accuracy score of 66.9%, is the superior model over Random Forest. When an experiment was demonstrated that certain pertinent factors do have an impact on students' academic achievement with respect to some relevant features that can have influence on student performance in their academics (Kaunang & Rotikan, 2018) . Making use of data mining approach by considering Questionnaires containing the demographics, prior GPA, and family history of the students were used to gather the data, and train a model for forecasting students' academic achievement. K-mean clustering technique was also put into practice to carryout analysis of student test results in a project (Oyelade et al., 2010) . The model was used to determine and analyse the student outcomes of a private institution in Nigeria, which serves as a good benchmark for tracking development of students' academic achievement in higher education institutions in order to help academic planners make effective decisions. An approach to categorizing student scores based on performance levels utilizing cluster analysis and widely used statistical methods

was also described. When a Decision Tree-based supervised learning model for projecting students' academic achievement. A group of classifiers, J48 and MLP was used for evaluation of the performance of predictive model created by the student on the dataset. The effectiveness of these classifiers is further enhanced by the application of an ensemble technique; the outcome demonstrates the suggested ensemble model, which incorporates the Decision Tree (J48) algorithm, attained the highest accuracy. The proposed model will thereafter be evaluated on a sizable dataset with a greater number of features (Imran et al.,2019). The uses of Naïve Bayes machine learning techniques in article (Dinh et al., 2020) to forecast students' eventual Grade Point Averages based on personal variables (including gender and place of residence), university entrance test scores, gap years, and first year level and second year level academic success. For both students and educational institutions, predicting students' academic achievement is a major concern. With this institutional administrators and lecturers can support students' learning plans by modifying the academic curriculum and lectures module based on the prediction of learning outcomes, an academic progress, individual traits, and behaviours related to learning activities are only a few of the variables that can affect a student's performance in school. During the analysis of current academic analytics in a research projects (*A Review on Predictive Modeling Technique for Student Academic Performance Monitoring | MATEC Web of Conferences*, 2019) That is primarily concerned with forecasting student academic success. To achieve the optimal performance model employing a variety of student data, strategies, algorithms, and tools, numerous approaches have been put out by prior academics. Classification, regression, and grouping are just a few of the learning tasks that are related to the predictive modelling used to forecast student performance. Several variables have been picked and evaluated to determine the most important characteristics to do prediction for choosing the most perfect model for the prediction. In (For et al., 2021) the parameters examined were effective prospective features for the model to effectively identify at-risk pupils in order for educators to provide these struggling students with corrective measures by carrying out a quantitative and qualitative research considering some factors such as attendance record , number of workload etc., at the outcome of this project a software was designed that make use of data mining and machine learning algorithm to develop a model that can detect or forecast student academic performance in an effort to determine which pupils are most likely to fail

that particular course. Preventing the risk of students who run the danger of not completing their college education with the expected time was focused on by the authors of this paper (Lakkaraju et al., 2015) there research methodology serves as a springboard for upcoming studies on additional unsuccessful academic outcomes. They concentrate on methodology, which includes problem formulation, feature extraction, classifiers, and evaluation criteria, is generalizable to other undesirable academic outcomes, such as under matching or not registering for college, These attributes that are advantageous for this task are described, many classification algorithms are applied, and their performance is assessed using metrics that are significant to school administrators in this research using A machine learning framework to described and find such students. To show that building and training a prediction model with a reliable accuracy rate is feasible utilizing a small amount of data was the major aim of this research (Abu Zohair, 2019) and also to examines the potential applications of finding the key signs in the tiny dataset that will be incorporated into the prediction model using visualization and clustering techniques. The best indicators were assessed using a number of machine learning techniques, and they chose the model with highest accuracy for implementation. There results showed that clustering technique, out of the ones used, is capable of locating important indications in tiny datasets. During this study the authors utilize support vector machines and learning discriminant analysis techniques for training on tiny datasets was demonstrated and delivering of accuracy rates and reliability tests for a valid classification. In a particular research work (Hamsa et al., 2016) the author intends to create a prediction model for a student's academic performance undergraduate and post graduate students to achieve this, factors such as internal and sessional scores, and admission score were chosen. Internal grades are calculated by adding the assignment grade, the average of the two sessional exam scores, and the attendance grade was put into consideration. The author utilized The use of data mining methods in education is known as educational data mining with an educational background with the goal of revealing underlying patterns and knowledge pertaining to student performance. Utilizing two selected classification techniques —Decision Tree and Fuzzy Genetic Algorithm. In research article (Nabil et al., 2021) the authors investigate the effectiveness of deep learning in the area of EDM, particularly in forecasting students' academic success and identifying at-risk individuals. This study created forecasting models using the deep neural network

(DNN), decision tree, random forest, gradient boosting, logistic regression, support vector classifier, and K-nearest neighbour to predict students' academic performance in upcoming courses given their grades in the preceding courses of the first academic year. It was noticed that the With an accuracy of 89%, the suggested DNN model can predict students' performance in a data structure course and identify those who are most likely to fail before the semester is up. This is superior to models like the K-nearest neighbour, logistic regression, support vector classifier, and decision tree. In order to analyse current academic analytics research projects (Mat et al., 2019) that are primarily concerned with forecasting student academic success. To achieve the optimal performance model employing a variety of student data, strategies, algorithms, and tools, numerous approaches have been put out by prior academics. Classification, regression, and grouping are just a few of the learning tasks that are related to the predictive modelling used to forecast student performance. Several variables have been picked and evaluated to determine the most important characteristics to do prediction in order to obtain the best prediction model. Using different type of technology to measures performance of students in their academic task by implemented machine learning algorithm was used on student dataset to forecast student performance, after the experiment the analysis of the result shows that random forest classifier algorithm have more accuracy that the support vector algorithm (Rai et al., 2021) . In this research article (Song et al., 2021) the authors compare the suggested forecasting model along with seven other methods using the actual dataset in order to confirm its superiority by construct a Model for predicting academic performance based on sequential involvement based on their regular learning activities by assessing students' effectiveness in their online learning. In order to forecast how well students will learn given their retrieved engagement trends, prior results, and demographic information, the second component works in conjunction with the next step is to determine the trends in how students engage in their everyday learning activities. During the review done by the authors in this research paper (Abidemi et al., 2023.) the significance of attendance record was emphasized correlation with student grade point GPA, furthermore the authors review several types of attendance system with different technologies and mode of operation and data storage, ranging from attendance system that work without internet i.e. localhost, attendance system that work with internet i.e. internet of things IoT for operation, some are integrated with artificial intelligence application to

achieve this task. The goal of this study is to research (Saleem et al., 2021) is to comprehend and forecast student performance using data from electronic learning management systems. With the aid of electronic learning management systems, professors and students can interact live with their institutional online portals and carry out instructional tasks digitally. The researcher's main focus is on how to effectively assess student performance using various indicators gathered during online sessions, which would affect both parents and teachers. The learning management system used in this study's dataset offers a number of variables that can be used to predict students' performance, there the five components of the suggested model conventional machine learning algorithms that are further improved by applying four ensemble techniques: bagging, boosting, stacking, and voting. By proactively informing the kids, it can be helpful for predicting student achievement and supporting educators in making knowledgeable judgments. Ensemble techniques have significantly improved the model's performance the stacking model, which combines all five classifiers, beat previous ensemble approaches and achieved the highest accuracy score (0.800%), the prediction ratio was increased and the integration of the ML models outperformed all other ensemble techniques. The authors of this research (Kumar et al., 2021) used a data analysis tool to explore and utilize fresh data from a historical academic dataset. The academic dataset is transformed into an interpretable format by the data analysis tools, this could significantly affect the objective of the educational research. An analysis of several "Machine Learning" methodologies is done using the 10-fold cross-validation method. Accuracy, Mean Absolute Error, and Root Relative Squared Error are among the evaluation metrics that are taken into consideration when evaluating these procedures. Relative Absolute Error and Root Mean Squared Error. The findings indicate that J48 (75.83%), Multilayer Perceptron (78.33%), Naive Bayes (67.70%), Locally Weighted Learning (66.04%), Decision Table (72.70%), and Naive Bayes (67.70%) all provide the highest levels of accuracy, hence confirming the validity of the initial study topic. Furthermore, the application of the ensemble approaches (Bagging, AdaBoostM1, and Random SubSpace) tends to improve the effectiveness of the selected classifier in light of the implementation findings. The results show a significant improvement when using the suggested EMT model approach. It's been demonstrated that the suggested method can produce excellent accuracy (up to 80.33%). According to the study's done by (Walia et al., 2020) it was stated that

student performance data can be used to anticipate the factors that will have an impact on a student's academic schedule. The quality of education must also be improved in order to keep pupils interested in attending school. In order to forecast academic performance, researchers in educational data mining typically employ CGPA and internal performance marks. But according to the results of their implementation, school and study time both have an impact on a student's final grade. They discovered that classification algorithms like OneR, JRip, and Decision Tree are similarly effective in predicting student results and have more JRIP rules than 80.00% accuracy.

This paper work (Kotsiantis et al., 2010) aim to bridge the gap between existing ML techniques and empirical predictions of student performance with the use of six trained ML algorithms. The learning algorithm might then be able to anticipate how incoming students will perform; making it a useful tool for identifying expected underachievers. This study's objective is to contrast a few cutting-edge learning algorithms. Six algorithms that were trained using data sets given by the Hellenic Open University were used in two studies. The Naive Bayes algorithm was shown to be the most suitable to be utilized for the creation of a software assistance tool, has more than adequate accuracy, and its overall sensitivity is quite satisfactory, among other notable findings. The goal of this study (Aziz et al., 2014) is to create prediction models for student academic performance for the first semester of the Bachelor of Computer Science program at Universiti Sultan ZainalAbidin (UniSZA). Three classification techniques have been chosen for this study: Naive Bayes, Rule Based, and Decision Tree. To determine which classification model for prediction works best, a comparative study is also carried out. According to the experiment, models created using the Rule-Based and Decision Tree algorithms perform better than those created using the Naive Bayes algorithm. In order to conduct this study, five independent criteria (gender, race, hometown, family income, and way of entry into the institution) have been chosen. These factors were selected based on previous research, including studies from the social sciences. The findings show that race has the greatest impact on students' performance, followed by family income, gender, mode of entry into higher education, and hometown location. The professor might utilize the prediction model to categorize the pupils and take immediate measures to boost their performance.

The student performance dataset from UCI was examined in this study (Malini & Kalpana, 2021) to identify the numerous factors that influence student achievement. Only highly ranked attributes were employed to decrease the dimensionality and undesirable features of the dataset, which comprises many attributes. The impact of the student's economic background is examined and how it affects them in many ways. The findings thus demonstrate that solely economic factors have an impact on pupils' performance. With the economic data, MLP provides 72% accuracy, Bagging classifiers demonstrate 88% accuracy, and Multi Boost classifiers provide 86% accuracy.

CHAPTER III

3.0. RESEARCH METHODOLOGY

3.1.1 Study Design:

To achieve this study objectives, according to the framework phases depicted in Figure. 1, a study research approach was carried out from beginning to the end. From dataset generation, dataset preparation, model training and testing, model evaluation, model deployments as output /result visualization, and also an approach like heat maps and hierarchy to identify the top correlated indicators, are included in various phases of the experiment. The best model for forecasting student achievement will then be selected after the indicators have been applied to several algorithms.

This study adopts and illustrates more details and steps entails in each phase of the experiment as indicated in Figure 3. For instance data generation phase explain the description of software and hardware configuration of the system with their application, while dataset properties and preparation entails , dataset preparations, features encoding, input of missing data, data normalization was done in data preparation phase, all these methods such as heat maps and hierarchical, are used to identify the best linked indicators. The best model for forecasting student achievement will then be selected after the indicators have been applied to several algorithms via there evaluation metrics, lastly the model was deployed using flask API. This will be covered in more detail in the sections that follow.

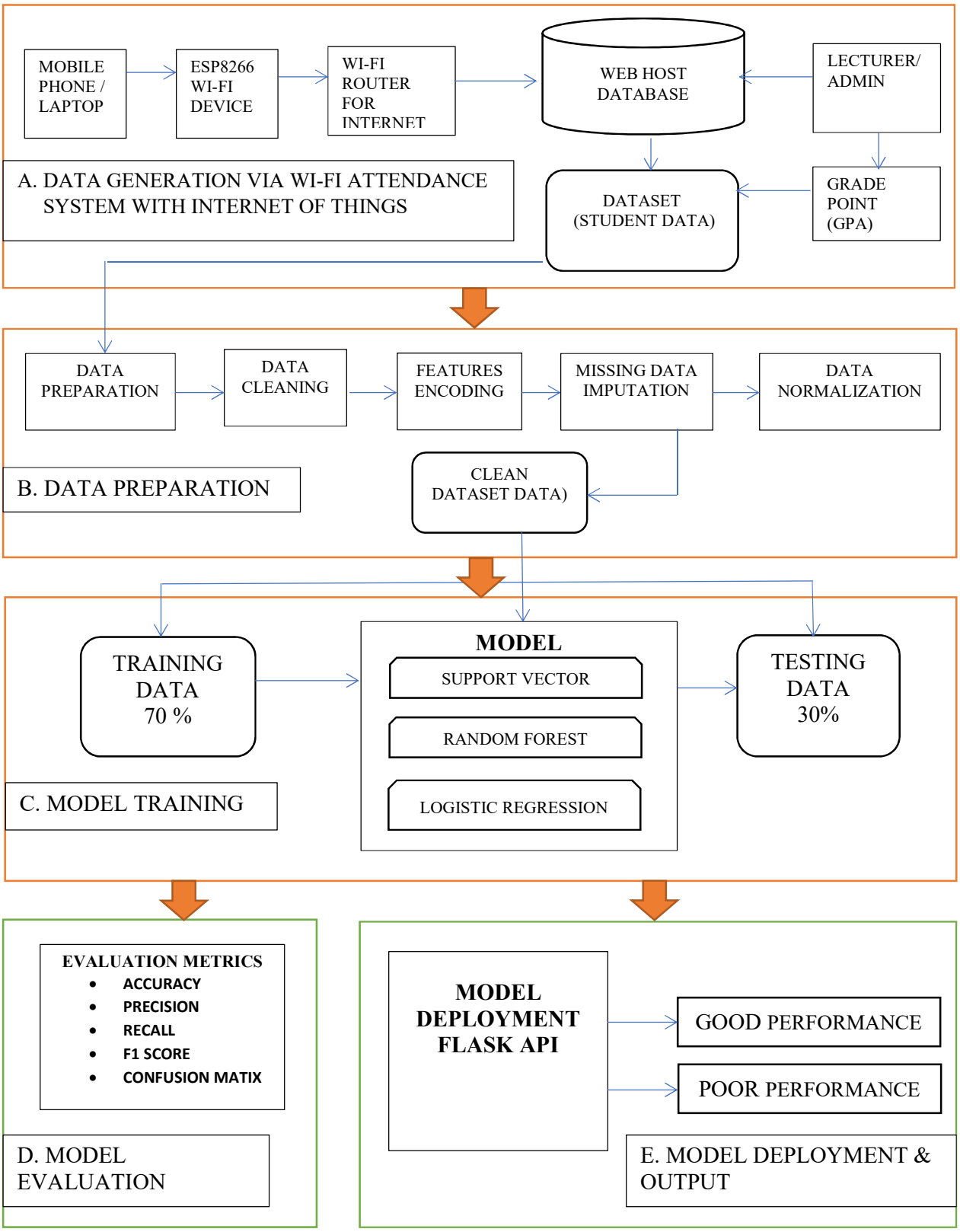


Figure 3: The block diagram of student academic performance system via IoT attendance system architecture

3.3 Materials and Method

The design methodology of this project is divided into different segment; these are the Hardware configuration part, Database configuration, Data Captive portal and the front end web design for Graphic users interface. Below is the list of hardware and software requirement.

- NodeMCU ESP 8266
- MySQL Database
- Webhost domain
- Arduino IDE with Libraries
- Embedded C++
- Python IDE with Libraries
- PHP
- HTML

3.2 Data generation via IoT Wi-Fi attendance system

This is a type of attendance system whereby a student initiate attendance by connecting to a Wi-Fi micro controller programmed to accept student number after it has been connected with via an IP address **http://192.168.4.1/,** The student will input there student number in a page displayed after connection to the micro controller and click submit button, this student ID number will be saved on a webhost database alongside with some vital information such as the student full name , the date and time utilizing the internet of things, after completion of the class the student will also repeat the same process for sign-out attendance. Tutor or Administrator will be able to access this attendance on a graphic user interface GUI webpage via internet and also can generate or download the class attendance in csv file to build a dataset that can be used to train artificial intelligence model as indicated in Figure 4.

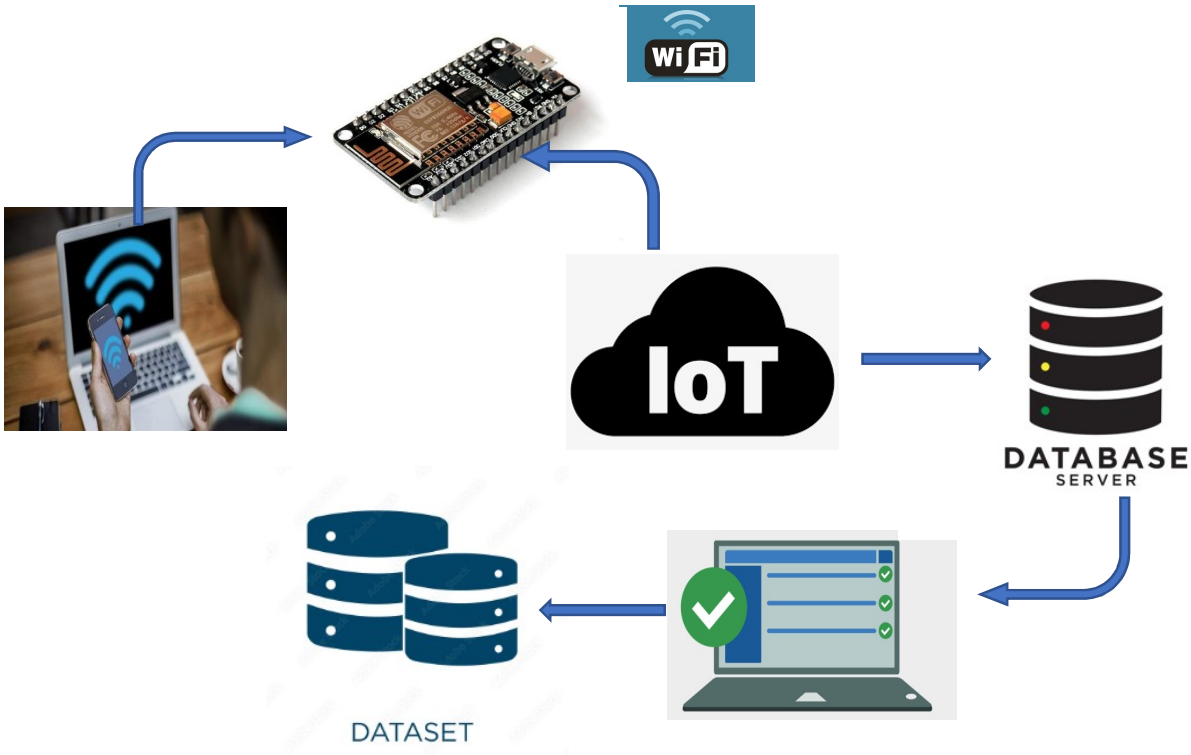


Figure 4: Data generation via IoT Wi-Fi attendance system

3.2.1 Description of IoT Wi-Fi attendance system design method for data generation

The design methodology used to achieve this, is broken down into various segments such as the hardware and software configuration part for database configuration, and web design for the graphic user interface's front end.

3.2.1.1 Hardware Configuration

The hardware configuration phase comprises programming step. Esp8266 Wi-Fi module is the microcontroller that was programmed using the Arduino programming IDE software, a free software package for creating code that can be uploaded to a microcontroller board, debugged, and compiled. The internet connectivity establish communication channel between the esp8266 Wi-Fi module and the database. There are several steps involved in achieving connectivity, the first is to link students to a local hotspot that has been set up for the project's hardware esp8266 Wi-Fi micro

controller to serve as both an access point (AP) and a station mode (STA) simultaneously.

The hardware serving as access point, other devices, such as a computer and a mobile phone, can find the network and connect to it to use it as a Wi-Fi hotspot. It is set up using the class name as the ESP8266 device name so that students can connect to it based on the class where the lecture will be held.

The ESP8266 device then operates in station mode will be able to connect to another Wi-Fi router in order to access the internet as illustrated in Figure 5 below.

```
// Set AP credentials
#define AP_SSID "VT-1-D03" // ESP name as Room Number
#define AP_PASS "12345678" // Password for connecting to the ESP
```

Figure 5: Arduino Code to Assign the room number as ESP name.

The ESP8266 Wi-Fi micro controller was programmed with embedded C++ code in combination with HTML codes, this programmed was successfully complied with the help of Wi-Fi manager library to facilitate the connection to internet through Wi-Fi router and setting up the operational mode which give the micro controller a default IP address of 192.168.4.1 for other device to connect to it. After the users (students) have connected to it, an HTML form will display for the student to enter their student ID.

3.2.1.2 Software configuration

The software configuration entails the database which was created using MySQL and paired with PHP code to keep a real-time record of attendance. It will include the student's ID, first and last name, and the class where the lectures are held, date and time. The student database is set up so that the attendance can be checked by looking up the student's name, student number, date, and time. This stage, which is also known as the back end stage doesn’t create duplicate information in the database.

The student will record their attendance by connecting to the ESP8266 Wi-Fi module micro controller access point deployed in the classroom to access the data captive portal, while the lecturer/admin must log in to the admin page to see the attendance record depending on what has been recorded for every student that have submit/mark attendance. A graphical user interface GUI will be used by the student/user, the

instructor/admin, to complete a one-time registration at the beginning of the semester. PHP, CSS, and HTML were used to for design of this GUI. MySQL is used for the backend while sublime text is used for the frontend web page that can be accessed using a web browser, such as Google Chrome can be used to get access to the GUI of the software implementation's front end and it's where the output can be checked. Faster data collection, aggregation, and processing are made possible by this, the administrator or lecturer can access the attendance record whenever they want over the internet.

3.3 Dataset properties and preparation

To prepare the datasets for the analysis phase, they will go through a pre-processing stage (cleaning, missing data imputation, etc.) as indicated in Figure 6 below.

Figure 6: Dataset Analysis Framework

3.3.1 Data preparation

In this study the data of 193 students was sourced for. These records include the student's ID, student grade point assessment GPA and number of attendance recorded, the list of the most popular properties, their datatypes, and other pertinent information are present in the dataset.

Multiple modifications have been made to the provided dataset in order to get the dataset ready for analysis that was done using Python Integrated Development Environment and Microsoft Excel. The datasets initially comprised valueless attributes include incomplete occurrences, inadequate data types for the characteristics, and other flaws, that necessitated its preparation prior to submission to the analysis step. The datasets went through the following stages of preparation as a consequence.

3.3.2 Data cleaning

To clean up the dataset, all information that wasn't essential to the research was removed, including course code, description of course, semester, session, and name of institution. Also the students with insufficient records, such as students without

grades, only the needed data are allowed to remain in the dataset as shown in the Figure 7.

```
    data.info()
[17] ✓ 0.2s
...  <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 193 entries, 0 to 192
     Data columns (total 4 columns):
      #   Column              Non-Null Count  Dtype
     ---  ------              --------------  -----
      0   s_no                193 non-null    int64
      1   attendance _record  193 non-null    float64
      2   gpa                 193 non-null    float64
      3   status              193 non-null    int64
     dtypes: float64(2), int64(2)
     memory usage: 6.2 KB
```

Figure 7: Description of the dataset from the data frame

3.3.3 Features encoding

For a variety of reasons, all of the properties of the datatypes have been modified to be numeric attributes. We will start with a few machine learning libraries that have shown to be effective when working with this type of datasets. Additionally, Machine learning technique was employed to function effectively with numerical attributes, in order to achieve the highest prediction result results when using Support Vector Machine (SVM), Random Forest (RF) and Logistic Regression (LR) algorithm, attributes must generally be in numerical form and be normalized. Prior to being fed into the prediction model, the values of the output data attributes will be altered and normalized into ranges (either [0, 1] at the status column.

3.3.4 Data normalization

Normalization is one of the suggested pre-processing procedures before training the model using certain types of prediction or classification algorithms. Due to the fact that scaling to these ranges typically produces superior results, this approach advises setting the instances' values to either [0,1].

3.4 Description of the selected model

3.4.1 Support Vector Machines (SVM)

This is an algorithm used in supervised machine learning models to address majorly two groups of classification problem, After obtaining sets of training data for each category that have been labelled, a SVM model may categorize incoming text, based on the input and output on the dataset compared to more contemporary methods like neural networks, SVM have this advantages of better performance and also increased of speed.

In Figure 8 below, a labelled training data using a plane to provides a simple example to understand the foundations of Support Vector Machines and how they work. Considering a scenario where our data has two features, x1 and x2, and two tags, green and blue. We need a classifier that can identify different classes of object as indicated in Figure 8, a red or blue object from a pair of (x1, x2) coordinates. A SVM, which creates best-performing hyperplane (a two-dimensional line) to partition the tags, is fed these data points, this line will be marked with a border, and anything that crosses the other.

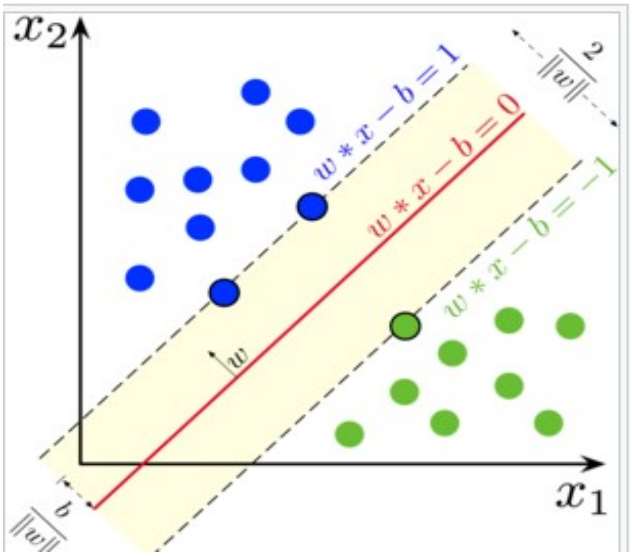


Figure 8: Support Vector Machines Illustration

3.4.2 Random Forests

Random Forest is an algorithm that belong to supervised machine learning technique which can be applied for prediction or classification purpose, it contain a group of Decision Trees that is normally trained using the bagging method, usually with maximum samples set.

Making a Bagging Classifier instead of passing, or using a Decision Tree Classifier, which is less practical and not usually optimized, therefore uses of Random Forest Classifier class is a better option. A Random Forest Regressor class is also available for regression task, when building trees, the Random Forest method adds additional unpredictability instead of focusing on finding the best feature overall, we search for the best feature among a random selection of features. More trees are produced as a result, which (again) lowers the variance in exchange for a larger bias and generally results to a stronger model.

3.4.3 Logistic Regression

Logistic regression is a supervised machine learning algorithm which is majorly used for predicting the probability of a target variable, due to its binary nature of its variable factors that the output will depend on, it can only give two viable classes with data recorded as either (1) which can represent YES or SUCCESS and (0) which can represent NO or FAIL. Logistic regression model can forecasts the value of M (P=1) as a function of Q mathematically. It is among the simplest Machine Learning approaches, If the estimated probability is larger than 50%, that can either forecast to indicate the instance is a member of the class (denoted as positive class, designated "1") or that it does not (i.e., that it belongs to the negative class, labelled "0") if the estimated probability is less than 50%. Hence, it is a binary classifier or The most fundamental kind of Binary logistic regression limits the kinds of the objective or dependent variable to one of two possibilities —1 or 0—of 1. We can simulate the relationship between a binary or binomial target variable and several predictor variables using this technique. The sigmoid curve is depicted in Figure 9 by the graph below. The y-axis values, which range from 0 to 1, are visible and intersect the axis at 0.5.

One can categorize the classes as either positive or negative, the result falls into the likelihood of the positive class if it is between 0 and 1. In this project design every outcome of the theory values is seen as a way to assess the as positive if it is equal to or greater than 0.5 and as negative if it is not. In order to evaluate the algorithm's performance using the weights on functions, we must additionally define a loss function.

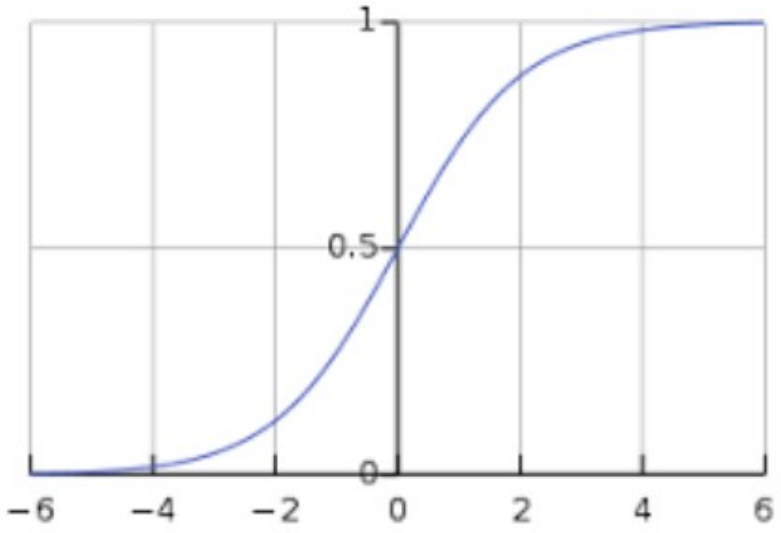


Figure 9: Logistic Regression Illustration

3.5 Details of the experiment

We give a brief overview of the models used in this research study in this section. Additionally, we go into great depth on the exploratory setup we employ for each prediction job as well as the evaluation decisions we make that aim to as nearly resemble the real-world environment as feasible.

We determine the complement of this flag, which indicates whether or not the student's academic achievement is good. In all of our prediction tasks, we refer to this as complementing parameters and utilize it for answer parameter. Thus, the task might be phrased as the outcome variable in order to address the issue of recognizing student performance in relation to their attendance.

3.5.1 Experimental setup

STEP 1: Dataset tuning

Our dataset collection consists of groups of students GPA and attendance records. In recent studies addressing the issue of forecasting student performance, the models were validated using cross validation. By using training data from earlier groups, the aim is to properly predict future outcomes.

STEP 2: Model

During the research investigation, experiment was carried out with Logistic Regression (LR), Random Forests (RF) and Support Vector Machines (SVM), to

predict students' academic performance. In this modelling phase, during this stage, three of the aforementioned modelling methods were selected and used, with their parameters calibrated to optimal levels, there are limitations on the kinds of data that can be used with some techniques. As a result, going back to the data preparation stage is typically necessary. The steps conducted in this phase are described above. The anticipated outcomes become somewhat non-deterministic as a result. We perform numerous runs with to account for this prediction.

STEP 3: Training and Testing data

To split the dataset for supervised learning tasks like classification, for example, the models' accuracy was evaluated by looking at the error rates. For the purpose of developing the model on the train set and evaluating its quality on a different test set. The whole dataset was divided into training (70%) and test data (30%). Building a model under this step the model tool was able to run on the prepared dataset to achieve the aim of this project.

STEP 5: Evaluation

The procedure of evaluation is used to guarantee the validity and accuracy of the results, if the outcomes contain flaws, the process allowed for a review of the first stage to determine why the results were incorrect. On a data science project, the data are often split into training and testing sets. The objective of this stage, which followed the modelling process, was to test the accuracy of the three models using testing data, there were many approaches used depending on the task and the environment. For instance, using the confusion matrix in the context of supervised learning to check the outcomes of the task of classifying objects. Other evaluation metrics that was used in the experiment are F1-score, Recall, Precision and Accuracy.

STEP 6: Deployment

Deployment is the process of making the outcomes useful and understandable, and by doing so, the project had to succeed in reaching its objectives. Without access to the results, a model is not very useful. The FLASK API was used to deploy the machine learning model as micro services.

Flask is a Python-based Web framework used for frontend and backend application configuration as well as for quickly and easily creating web applications, it gives programmers complete control over data access. The core of Flask is the Jinja templating engine and Werkzeug's (WSGI) toolkit. When building a blog website or

any other type of commercial website, Flask offers a number of libraries, tools, modules, and functions including handling user requests, routing, sessions, form validation, and more. Flask is designed for easy implementation of REST APIs. (*Explain What Flask Is and Its Benefits? | I2tutorials*, n.d.). The only things that Flask gives are those that are absolutely necessary, developers who work with other capabilities, such as data management, must create separate modules or employ extensions and because Flask has so many extensions, its functionality has been increased.

Flask is a well-known Python web framework on top of which Flask API is created, it makes use of Flask's simplicity and adaptability to offer a compact and user-friendly framework for creating APIs. Flask API is a popular option for data-driven apps because it works nicely with other Python modules like NumPy and Pandas etc. Developers may quickly create API endpoints that support different HTTP methods including GET, POST, PUT, and DELETE using the Flask API. In order to secure their APIs and manage access to critical resources, developers can use the authentication and authorization procedures supported by the Flask API. A robust and adaptable tool for creating RESTful APIs in Python is Flask API. For programmers creating data-driven apps and micro services, it is a popular choice due to its simplicity, flexibility, and interoperability with other Python modules(*The Flask Web Framework: A Beginner's Guide*, n.d.).

It has various advantages, development that is simpler and more scalable, flexibility, functionality and modularity.

CHAPTER IV

4.0. RESULT AND DISCUSSION

In this section of our experiment, we concentrate on this context by forecasting student academic performance using use the information accessible at the end of session. Specifically, we make use of grade point GPAs, and attendance records (Attend hours), as it was plotted against each other as show in the Figure 10.

In Figure 11, the login page where the admin/lecturer can have access to the overall attendance record via the lecturer email and password that has been done with one time registration. After successful login the a list of overall attendance will be displayed as shown in Figure 12, with a button at the end of the page to download the attendance record in csv file.

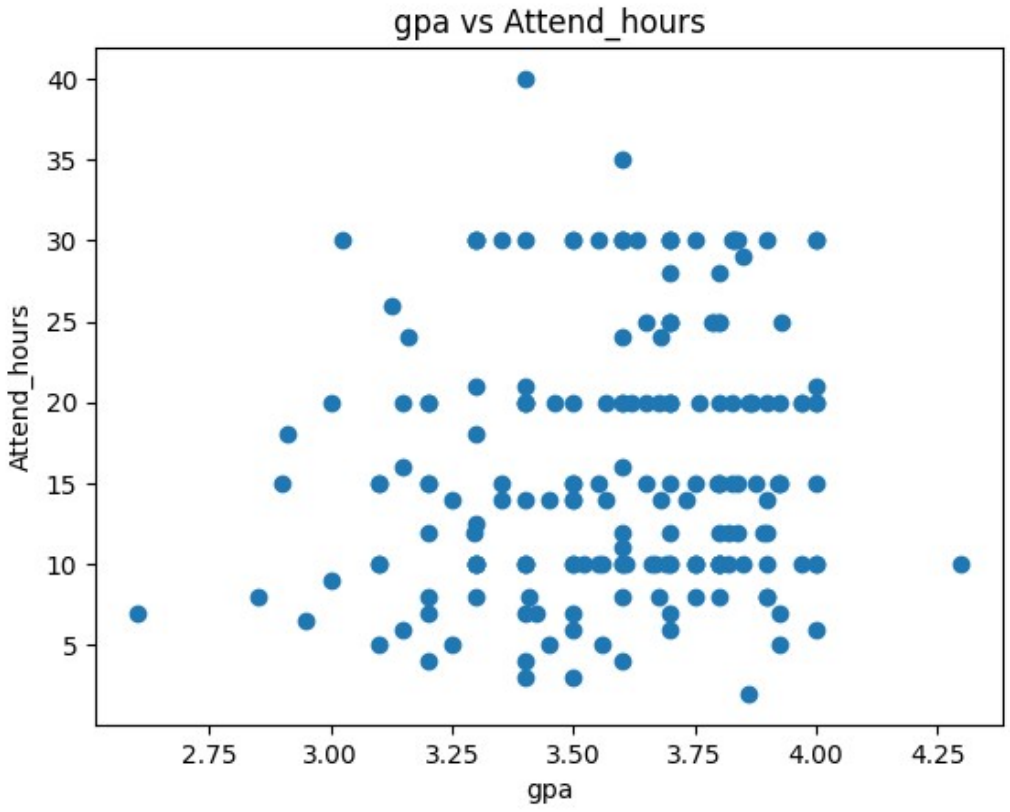


Figure 10: Plotting of student Attendance record (attend hours) against students GPA

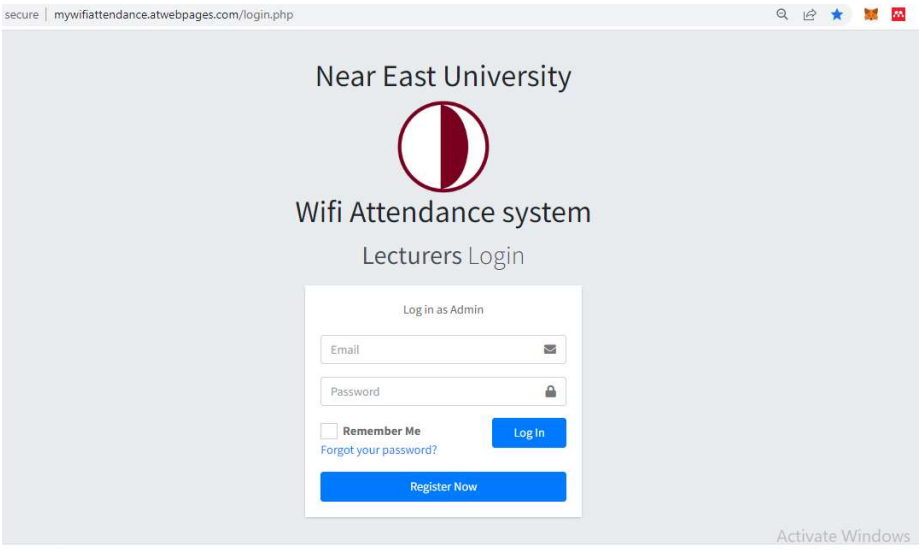


Figure 11: Login page for admin/lecturer to have access to the attendance record.

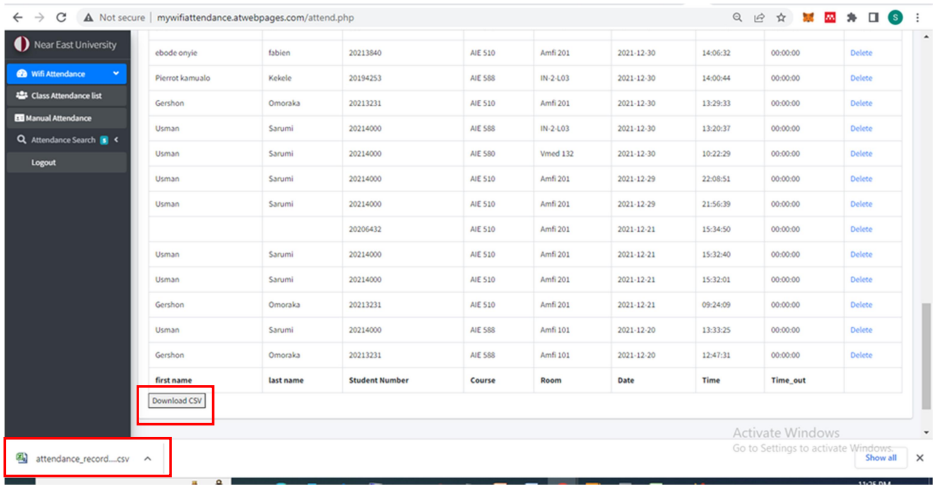


Figure 12: Overall attendance list on the webpage with download button for data generation in CSV format

After the CSV file has been generated successfully, data preparation was done in this stage some data was removed such as student number, student name etc. and Grade point assessment GPA was added to make the complete feature selection needed for our project before deployment into model.

We answer the following queries in this section:

• How well does the various models fare when tested using criteria like accuracy, Precision, Recall, F1-Score, and Confusion Matrix? To demonstrate that the probabilities and confidence score estimates generated by different algorithms are accurate so that schools can consistently implement interventions?

• How can we demonstrate the robustness of the findings while comparing the goodness of such estimates?

4.1 Algorithm performance measurement using evaluation metrics

The evaluation metrics (ACCURACY, PRECISION, F1-SCORE, and RECALL) and CONFUSION MATRIX were also employed to do a secondary check on the significance performance or believability of the produced dependable output rates in comparison with the evaluation metrics of all the algorithms used in this project. The results, which are displayed in the Tables and graphs below, highlight the importance of each model algorithm used and their performance results in the evaluation metrics for predicting students' academic performance.

A Hold-Out Method test was utilized to gauge each algorithm's performance in order to look at how well it performed when it was used in the project model. The process divides data split into two parts: the hold-out testing set (30%) and the training set (70%), where the testing set is used to validate and test the model, and the training set is used to train the model.

A confusion matrix evaluates the effectiveness of a classification model by contrasting the actual target values with those predicted by the machine learning model. Accuracy, Precision, Recall, and F1-Score were employed as performance measurement. The following are definition of these metrics:

4.1.1 Confusion matrix

Confusion Matrix creates a matrix, as the name implies, and highlights the method's general suitability for model performance. The Confusion Matrix serves as the foundation for all other measurements, although it's not consider as a metric, its a crucial element that may be utilized to evaluate how effectively the machine learning classification model is working. confusion matrix is a two-dimensional table with actual values and expected values. Mostly used for binary classification. In this

project our expected result should be GOOD or POOR, 1 or 0, etc. Furthermore, we created a classifier that forecasts a class based on an input and output features.

Table 1: Description of confusion matrix

	PREDICTED YES	PREDICTED NO
ACTUAL YES	TP = 40	FN = 3
ACTUAL NO	FP = 2	TN = 13

The accuracy of the matrix can be determined by an average of each value along the "main diagonal," There are 4 crucial words when describing confusion matrix:

True Positives TP: Situations where we correctly predicted YES and the actual result was YES. True Negatives (TN): Situations in which we predicted NO but the result was NO. False Positives (FP) are instances in which we predicted a YES result but the actual result was a NO result. False Negatives (FN) are situations in which we projected a NO result but the final result was a YES. In Table 1 above, True positive (TP) = 13, True Negative (TN) = 40, False Positive = 2, and False Negative = 3.

Table 2: Confusion matrix result in each model algorithm

Model	**RANDOM FOREST**	**SVM**	**LOGISTIC REGRESSION**
CONFUSION MATRIX	(40 , 2) (3 , 13)	(39 , 3) (5 , 11)	(39 , 3) (5 , 11)

In Table 2 the confusion matrix of the entire three machine learning model algorithm used in this project was indicated.

4.1.2 Accuracy

This is the proportion of accurate predictions amidst all the algorithm's predictions. It measures the proportion of accurate predictions to all input samples as shown in equation 1.

Accuracy = (TP+TN) / (TP+FP+TN+FN)

$$\text{Accuracy} = \frac{\text{No of correct predictions}}{\text{Total no of prediction made}} \quad (1)$$

It's one of the most used categorization measures, accuracy spans from 0% to 100% or 0 to 1, is very simple to learn and use, it is also available for any classification model with a score approach on machine learning libraries like Scikit-learn. By averaging the values along the matrix, it is possible to assess the matrix's correctness, the "main diagonal the accuracy of the model algorithm used for this research is calculated for random forest, support vector machine and logistic regression below manually in Equation 2-4 respectively.

i.e.

$$\text{Accuracy of Random Forest} = \frac{40 + 13}{40+2+3+13} = 0.913793103 \quad (2)$$

$$\text{Accuracy of Support Vector Machine} = \frac{39 + 11}{40+2+3+13} = 0.862069 \quad (3)$$

$$\text{Accuracy of Logistic Regression} = \frac{39 + 11}{40+2+3+13} = 0.862069 \quad (4)$$

For example above is the accuracy of Random forest algorithm which has the highest accuracy compare to Support vector machine and Logistic Regression with lowest accuracy as it was indicated in Table 3 and graphical representation in Figure 13.

Table 3: Accuracy result in each model algorithm

Model	**RANDOM FOREST**	**SVM**	**LOGISTIC REGRESSION**
ACCURACY	0.913793	0.862069	0.862069

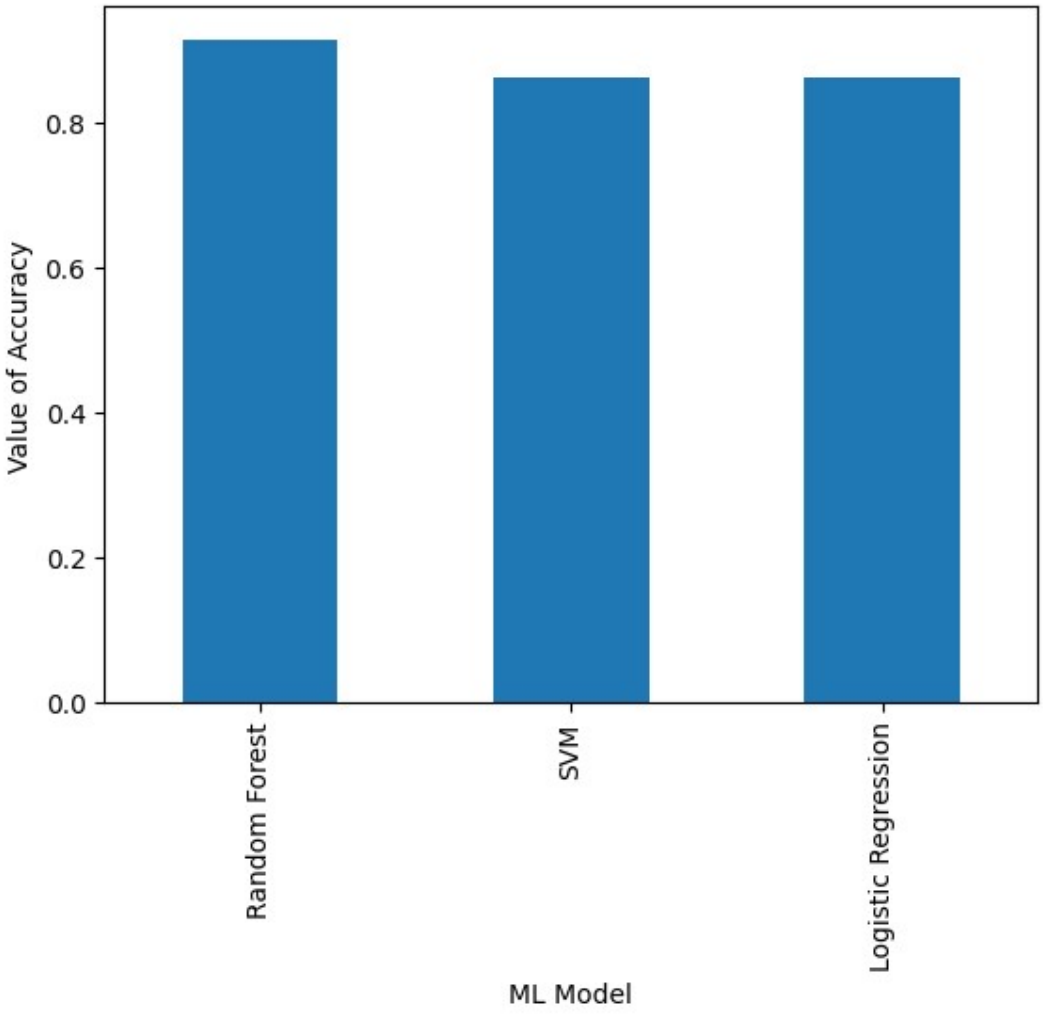


Figure 13: Bar chart representation of the accuracy for each model

4.1.3 Precision

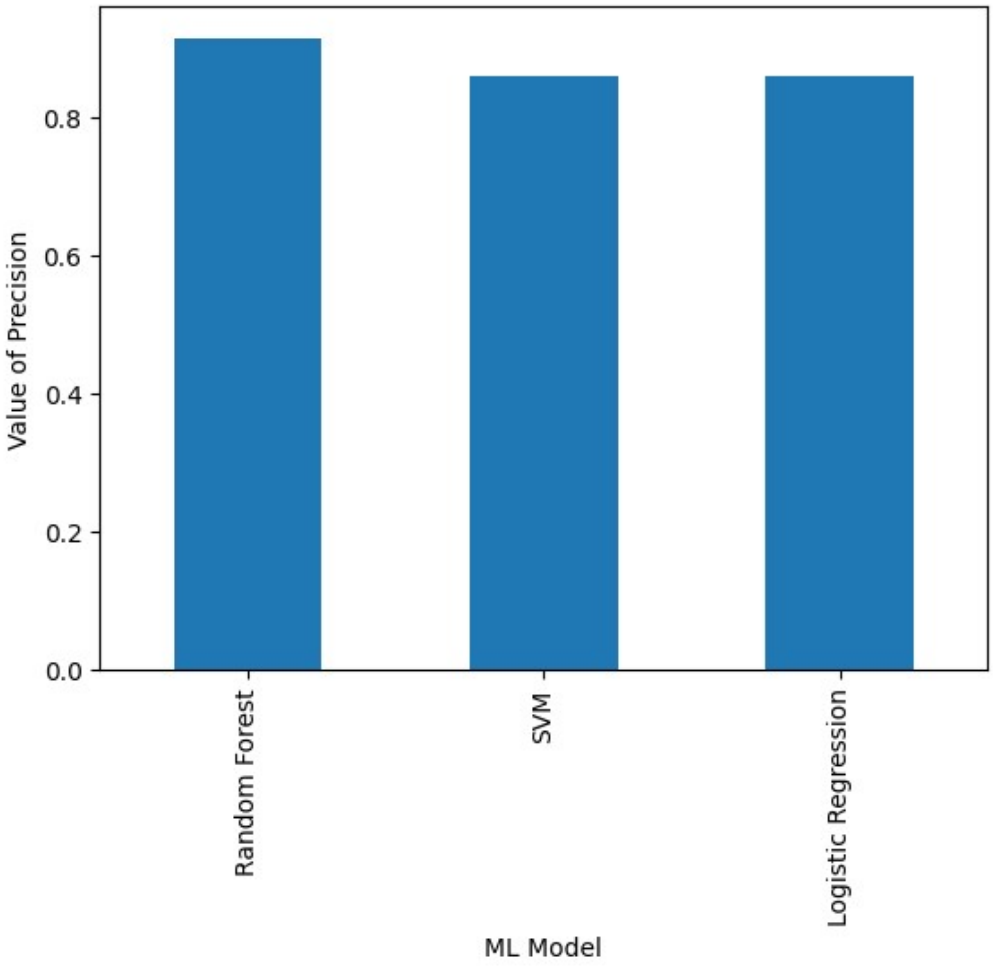
In any algorithm's, precision is the proportion of accurately classified cases among all actually classified instances. Out of all the accurate forecasts the program produced, the proportion of accurate positive predictions, as shown in Equation 5.
Precision = (TP) / (TP + FP):

$$\text{Precision} = \frac{\text{No of actual/true positive predictions}}{\text{No of actual/true positive predictions} + \text{No of false positive predictions}} \quad (5)$$

We used precision in this project because it assesses the classifier's skill and shows how accurately it labels correct predictions. Precision always performs well when False Negatives must be avoided or are avoidable but False Positives cannot be ignored, as it was indicated in Table 4 and graphical representation in Figure 14.

Table 4: Precision result in each model algorithm

Model	**RANDOM FOREST**	**SVM**	**LOGISTIC REGRESSION**
PRECISION	0.912697	0.858598	0.858598



.

Figure 14: Bar chart representation of the precision for each model

4.1.4 Recall

This is the number of cases that were correctly classified is always divided by the total number of occurrences; nearly always, the recall value is the same as the correctly classified instance (CCI). Out of all the dataset's real positive cases, it is utilized to calculate the percentage of true positive forecasts, as shown in Equation 6.

Recall = (TP) / (TP+FN)

$$\text{Recall} = \frac{\text{No of actual/true positive predictions}}{\text{No of actual/true positive predictions} + \text{No of false negative predictions}} \quad (6)$$

Recall shows the ratio of precise positive forecasts out of all possible positive guesses divided by the sum of the dataset's True Positives and False Negatives; it is determined by dividing the total of True Positives by summation of all True positive and False Negatives. Recall provides this indication of missing positive forecasts, in contrary to the accuracy measurement we discussed earlier, it assesses how well the classifier performs in predicting all positive cases and shows how many valid positive labels the classifier has given, as it was indicated in Table 5 and graphical representation in Figure 14.

Table 5: Recall result in each model algorithm

Model	**RANDOM FOREST**	**SVM**	**LOGISTIC REGRESSION**
RECALL	0.913793	0.862069	0.862069

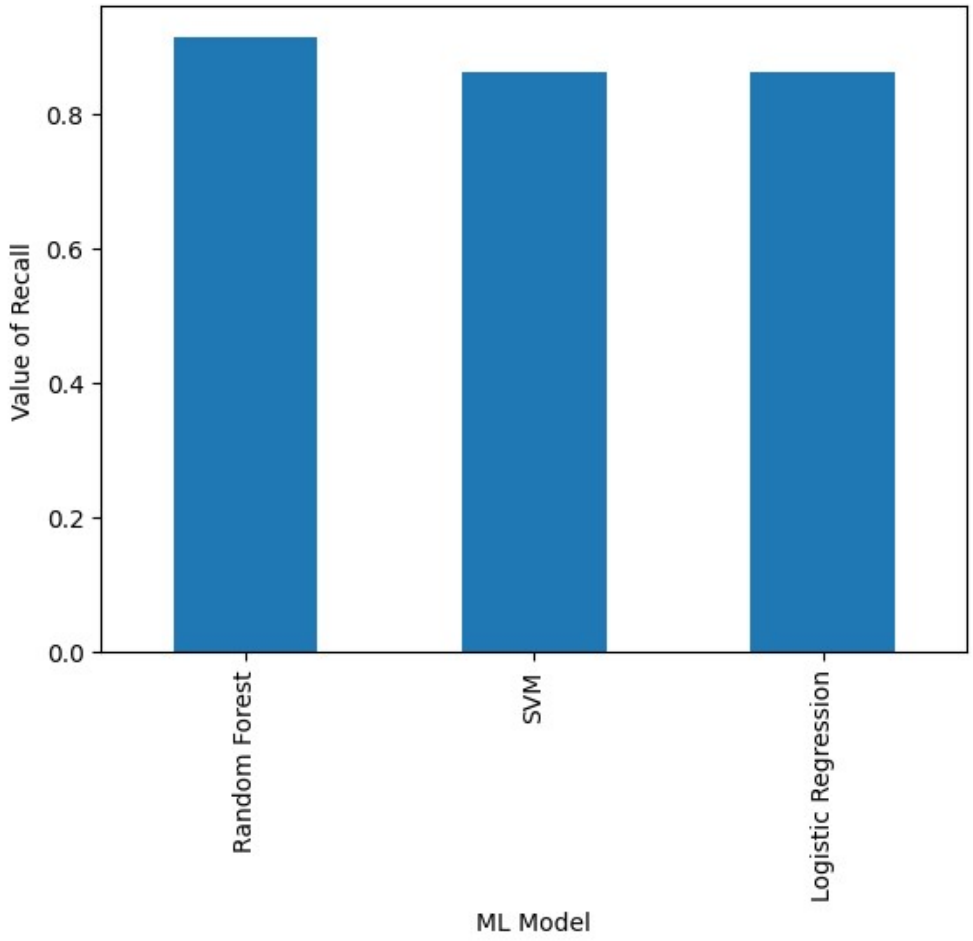


Figure 15: Bar chart representation of the precision for each model

4.1.5 F1-Score

The F1 score, which offers a single statistic to assess the effectiveness of the algorithm, is the weighted mean of recall and precision. It is calculated using the values of recall and precision (double precision multiplied by recall divided by the value of recall and precision added together) as shown in Equation 7. It uses the Precision and Recall an average of the two harmonics measurements, with the F1score falls within 0 and 1.

F1 score = (2TP) / (2TP+FP+FN)

$$\text{F1 score} = 2 * \frac{(\text{precision} * \text{recall})}{(\text{Precision} + \text{recall}).} \quad (7)$$

In an attempt to balance memory and precision, the F1 Score determines their harmonic mean. It serves as a measure of a test's precision, with 1 being the highest possible score, which indicated complete recall and precision as shown in Table 6 and graphically in Figure 16.

Table 6: F1-Scorel result in each model algorithm

Model	**RANDOM FOREST**	**SVM**	**LOGISTIC REGRESSION**
F1-SCORE	0.912910	0.859075	0.862069

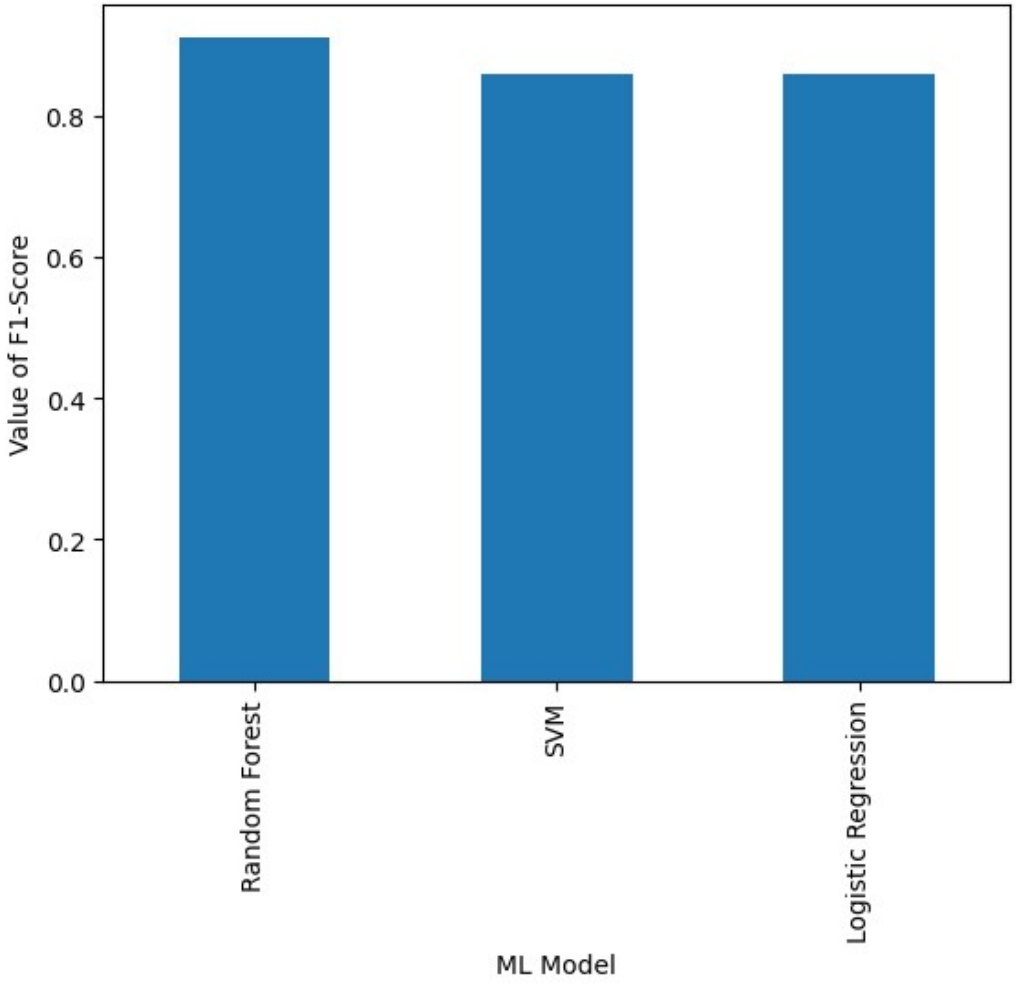


Figure 16: Bar chart representation of the F1-Score for each mode.

4.1.6 Error Bars

Error bars are representations that display the range of variability, in data. They are commonly used on graphs to indicate the error or uncertainty associated with a reported measurement. Error bars provide an indication of how precise a measurement's how close the true value (free from errors) might be to the reported value as illustrated in Figure 17. Typically error bars represent one deviation of uncertainty one standard error or a specific confidence interval like a 95% interval. It's important to state which measure was chosen in the graph or supporting text as these quantities are not interchangeable. When certain conditions are met error bars can be useful for comparing two quantities and determining if differences between them are statistically significant. They can also give insights into how a given function describes the data serving as an indicator of goodness of fit. In the sciences it is generally expected that scientific papers include error bars, on all graphs; however practices may vary across disciplines and each journal may have its own specific guidelines regarding their usage. Additionally research has shown that error bars can be utilized as a means to control algorithms for computation.

Alternatively error bars can be expressed using a minus sign (±) indicating both the lower limits of the associated errors.

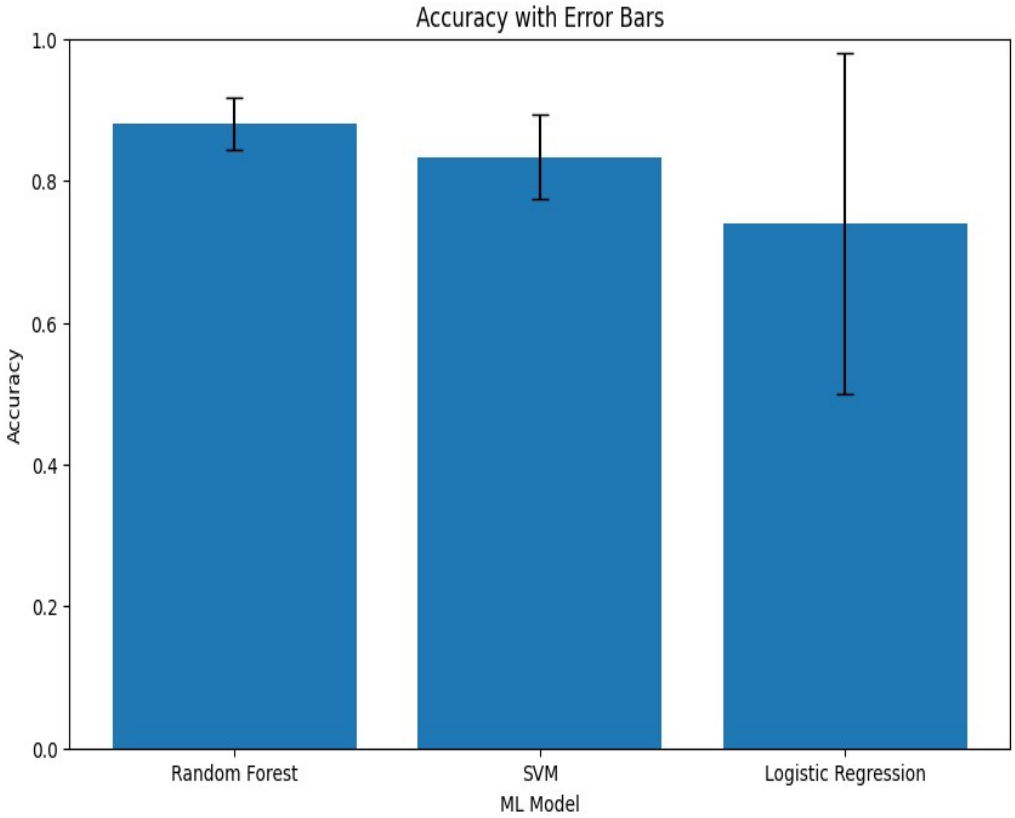


Figure 17: Error bar for each model

4.2 Discussion

Predicting student performance can be used as a base for preventative measures when it comes to potential student failure to reach learning objectives, and it can also be used to modify teaching strategies to accommodate student variety. The accessibility of student information that can be analysed to produce forecasts, such as behavioural information, the kind and frequency of activity engaged in (both in online and offline learning contexts), and the major reason attendance records were employed in this research study, provide additional support for this.

4.2.1 Comparing the model algorithm used in this project

Figure 18; show the comparison between the both models indicating percentage level of the evaluation metrics

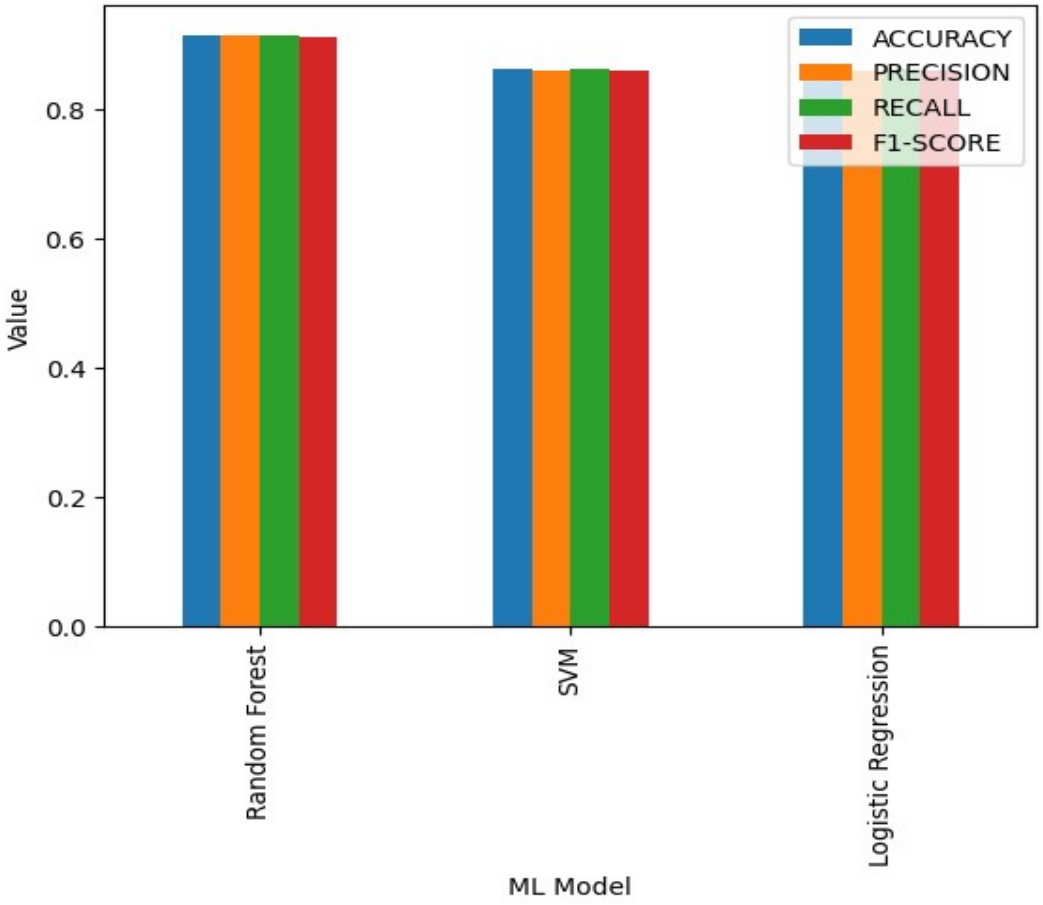


Figure 18: Bar chart representation comparing evaluation metric of all the models in graph

Table 7a: Comparing the evaluation metric of the model in table

Model	ACCURACY	PRECISION	RECALL	F1-SCORE	CONFUSION MATRIX
RANDOM FOREST	0.913793	0.912697	0.913793	0.912910	(40 , 2) (3 , 13)
SVM	0.862069	0.858598	0.862069	0.859075	(39 , 3) (5 , 11)
LOGISTIC REGRESSION	0.862069	0.858598	0.862069	0.862069	(39 , 3) (5 , 11)

In Table 7a and 7b, it was indicated that when considering the random forest model algorithm it has a HIGHEST in all the evaluation metrics when compared with Support vector machine and Logistic Regression in hierarchy, because Random forest contain a group of Decision Trees that is normally trained using the bagging method, usually with maximum samples set by adding an additional unpredictability instead of focusing on finding the best feature overall, it rather use the best feature among a random selection of features. More trees are produced as a result, which (again) lowers the variance in exchange for a larger bias and generally results to a stronger model.

Therefore accuracy in this case is

Table 7b: Comparing the accuracy in evaluation metric of the model in table

Model		**ACCURACY**
RANDOM FOREST	0.913793	**91%**
SVM	0.862069	**86%**
LOGISTIC REGRESSION	0.862069	**86%**

Therefore Random forest has the HIGHEST accuracy was used in this project for model deployment and implementation.

4.2.2 Model Comparison with other previous research studies as baseline

We contrasted our findings with those of a few earlier research that looked into the likelihood of pupils succeeding in their academic endeavours based on several factors for feature selection such as Assignment, Lab work, Age, gender, student Attendance , parental influence (mother and father education, job, wealth), Disability, location, transportation to school, information prior to attending college, such as the gap year between high school graduation and college enrolment, the university admission test, and the university entry categories), GPA and extra activities etc.

We discovered that the dataset utilized in earlier research (Yadav et al., 2012b) was imbalance in comparison with our dataset record. The size of the data set, the

variables that were assessed, and the type of methodology employed all have an impact on how accurately predictions of students' performance may be made. Table 8 introduces the techniques employed in the evaluated studies, including several ML algorithm used such as logistic regression, support vector machine, Random forest, OneR, J48, Random forest, multilayer Perceptron, Naïve Bayes, Decision tree, JRip, back propagation, mu0ltiboost classifier and deep neural network (Yadav et al., 2012b), (Watt et al., 2020), (Hassan et al., 2021), (Walia et al., 2020), (Kotsiantis et al., 2010), (Aziz et al., 2014) and ensemble method was also used in (Nabil et al., 2021), (Saleem et al., 2021), (Malini & Kalpana, 2021) and (Kumar et al., 2021) for training the model, obtaining model accuracy, and remarkable findings.

In addition, predictive models used in this research study are built using only traditional ML techniques as other research work done but we choose to use unique features (Student attendance record and student grade point assessment GPA) as selected feature to train our model. After applying ML techniques using our real dataset, we are able to achieve an accuracy of 91% with Random forest ML model algorithm that perform better due to it features contain a group of Decision Trees that is normally trained using the bagging method, usually with maximum samples set.

As shown in the comparison table in Table 8 that serves as a baseline, the experimental findings show that our suggested approach has produced the best accuracy in comparison to all other assessed techniques in this study and earlier research works. This will justify the method deploy in this research work give higher accuracy for model performance using student attendance record and GPA as selected features compared to other previous research work.

Table 8: Comparison with other previous research studies

Reference	**All Model used in the research**	**Dataset Preparation**	**Features selection**	**Result of best accuracy model in the research**	**Ensemble Method**	**Year**
(Yadav et al., 2012b)	Decision tree and random forest	NO	YES	**Decision tree - 66.9%**	NO	2012
(Watt et	OneR, J48,	YES	YES	**Naïve baiye and**	YES	2020

al., 2020)	Random forest , multilayer Perceptron, Naïve Bayes and support vector machine, Ensemble boosting,			**MLP - 86.19%**		
(Hassan et al., 2021)	Logistic Regression and Decision Tree	YES	YES	**Decision Tree - 78%,**	NO	2021
(Nabil et al., 2021)	deep neural network (DNN), decision tree, random forest, logistic regression, support vector classifier, and K-nearest neighbour	YES	YES	**Deep neural network (DNN) - 89%**	NO	2021
(Saleem et al., 2021)	Decision tree, Random forest, K-nearest neighbour,	YES	YES	**Ensemble stacking – 80%**	YES	2021

	Gradient boosting, Naïve Bayes, Ensemble (Voting, boosting, stacking and bagging)					
(Kumar et al., 2021)	J48, multilayer Perceptron, Naïve Bayes, Ensemble (Bagging, AdaBoostM1 ,and Random SubSpace)	YES	YES	**MLP - 80.33%**	YES	2021
(Walia et al., 2020)	NaiveBayes, Decision Tree, Random Tree, JRip	YES	YES	**JRip - 81.77%**	NO	2020
(Kotsiantis et al., 2010)	Decision Tree, Back Propagtion, Naıve Bayes, Neural Network. Logistic regression, Support	YES	YES	**Naïve Bayes - 72.48%**	NO	2010

	vector machine					
(Aziz et al., 2014)	Naïve Bayes, Rule Based (OneR) ,Decision Tree (J48)	YES	YES	**Rule Based (OneR), Decision Tree (J48) – 68.8%**	NO	2014
(Malini & Kalpana, 2021)	MLP, Bagging classifiers, and MultiBoost classifiers.	YES	YES	**Bagging classifiers - 88%**	YES	2021
This proposed research study	Random forest, Support vector machine and Logistic regression	YES	YES	**Random forest – 91%**	NO	2023

4.2.3 Conditions related to academics performance forecasting

Due to our findings on a significant role attendance record always play in academic institution and its impact in student academic performance, this is make us choose attendance records and student grade point assessment as a selected feature used for prediction in this project.

This project clearly buttress the vital role attendance play in making good academic performance, because it make it possible to monitor factors that contribute to excellent academic performance and student GPA already entails some other factor features, using those features again may cause usage of redundant data .

Due to our findings on how significant attendance record can be in structure of academic institution and impact it have in student academic performance , this make us to choose it in combination to student grade point assessment as a features factors to be used for prediction for this project model.

This project clearly buttress the vital role attendance play in making good academic performance, because it make it possible to monitor factors that contribute to excellent academic performance and student GPA already entails some other factor features, using those features again may cause usage of redundant data .

The following are some ways that attendance and GPA might be related to one another and improve academic performance.

a. Engagement in the classroom: Regular attendance frequently suggests participation in the learning process. Students can participate in conversations, ask questions, and get prompt feedback from teachers when they attend classes. Students who are actively participating in class are more likely to retain information and perform better on tests and assignments, which can increase their GPA.
b. Consistent attendance guarantees that students receive a steady stream of material throughout the course, which promotes ongoing learning. Students who fall behind because of missed classes may find it harder to catch up and comprehend later material, which could result in worse grades.
c. Graded Participation/Attendance: Some courses may include attendance or participation as one of the evaluation factors. In such circumstances, consistent attendance has a direct impact on the GPA. A student's total grade might be improved if they consistently receive participation or attendance points.
d. Building Relationships: By attending class, students can get to know their professors and fellow students. Strong connections can lead to increased academic success through enhanced communication, individualized support, and a clearer grasp of expectations.
e. Accountability and time management: A sense of accountability is developed via regular attendance. Students who regularly attend class are more likely to establish efficient time management techniques and adhere to a study

schedule, which might enhance their capacity to finish assignments and get ready for tests.

f. Reduced Absences During Critical Periods: Missing courses when critical topics are addressed or exams are being prepared for can have a major negative influence on a student's grade point average. Students who regularly attend class are more likely to obtain important information when they need it.

g. Motivation and Discipline: Students who emphasize attendance and understand that it is crucial to accomplishing academic goals are more likely to show discipline and motivation in their studies. A greater overall academic achievement may result from this approach.

It's crucial to remember that attendance is not the only determining factor, even though it can have a beneficial impact on GPA and academic achievement. A number of other factors, including study habits, time management, personal obstacles, and extracurricular activities, can also have a big impact. The key to getting the highest academic achievement is striking a balance between attendance, active involvement, and effective study habits.

4.3 Model Implementation

The model was developed as a web-based application utilizing the programming languages FLASK, HTML, CSS, JavaScript, and Python. The employed machine learning algorithm is Random forest because it has the highest accuracy. The specific steps taken by the algorithm to forecast students are displayed in the following screenshots, with Good or Poor performance.

Input field for student attendance record and GPA will be filled, then the prediction tool that shows the anticipated achievement of a student is depicted in the screenshots, once prediction performance button is click as shown in Figure 19 - 20. For example In Figure 19, the student has attendance record of 70% and GPA score of 3.0, and the output result is excellent remark.

More so, in Figure 20, the student has attendance record of 10% and GPA score of 2.5, and the output result is warning remark.

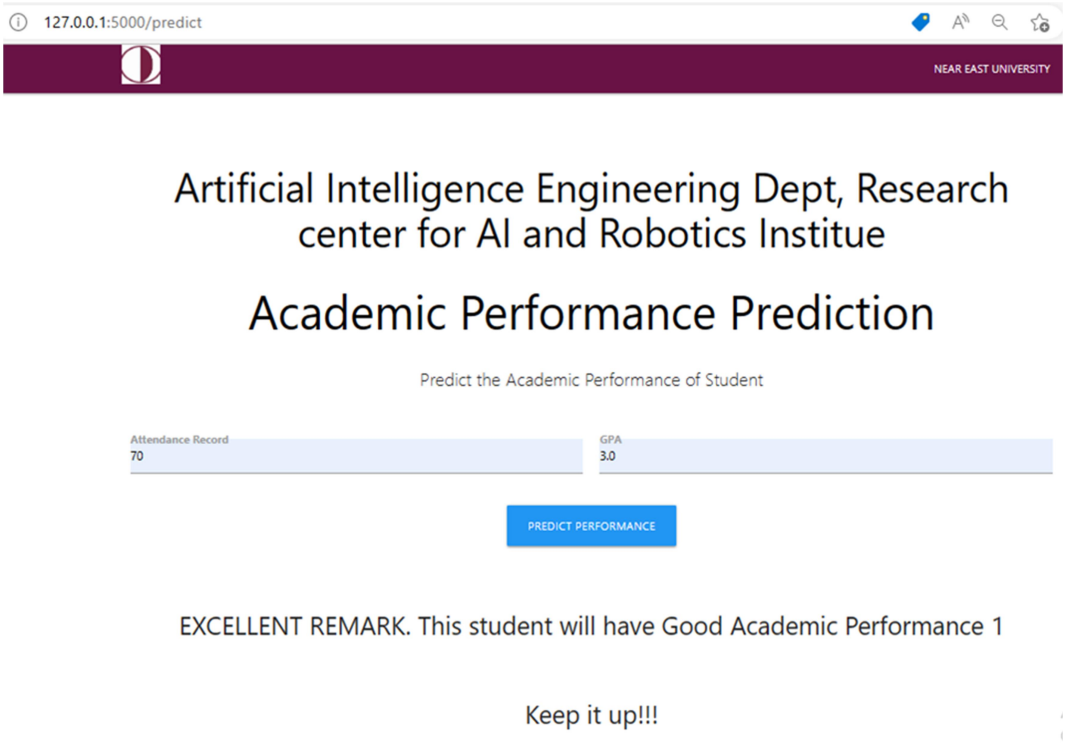


Figure 19: Testing output with excellent remark indicating good performance

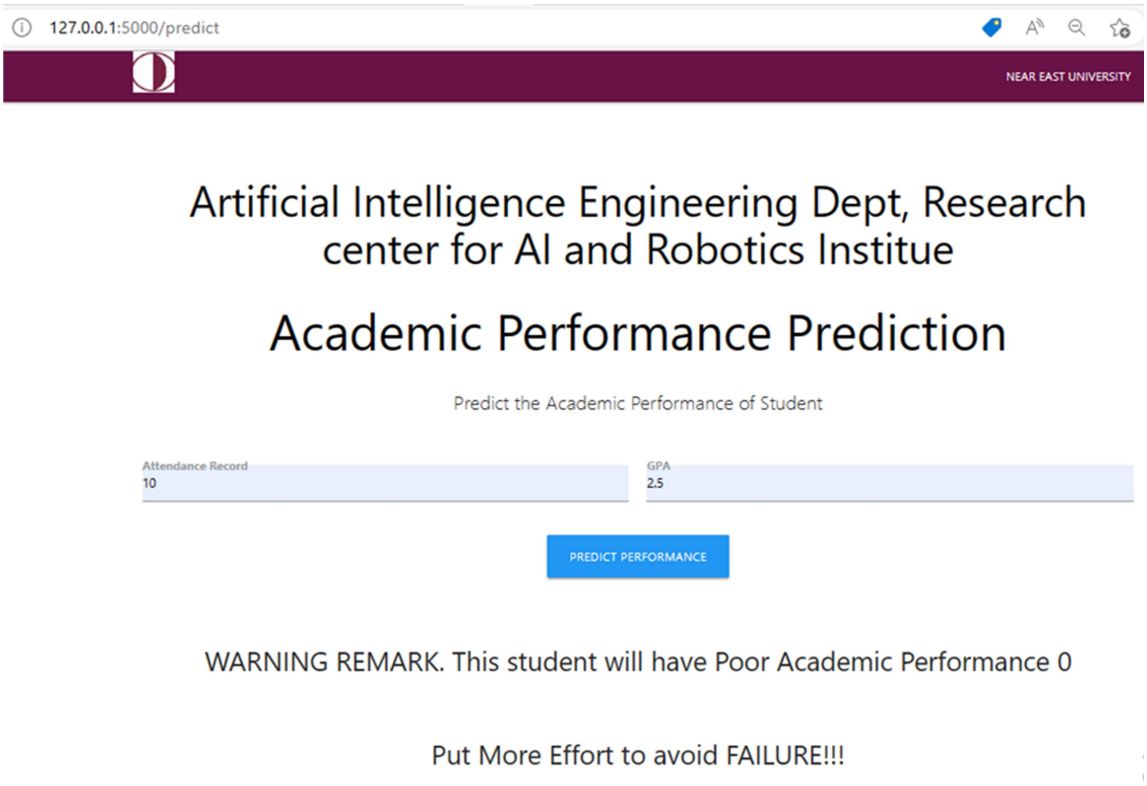


Figure 20: Testing output with warning remark indicating poor performance

CHAPTER V

5.0. CONCLUSION AND FUTURE WORK

Conclusion

The ability to forecast students' academic achievement is crucial in todays. In this study, we explored the feasibility of predicting student academic performance using a Wi-Fi IoT attendance system and students grad point assessment records. The results indicated that the attendance records obtained from the IoT system could be used as valuable features for predicting academic performance. We employed machine learning algorithms such as Random Forest, Logistic Regression, and SVM to build predictive models. Among these, the Random Forest algorithm demonstrated the highest accuracy of 91% in predicting student performance based on attendance records and GPA. The aim of this model algorithm is to forecast if the students will have either good or poor performance, in such that student who is at risk of failing in order for the educators and instructors to identify and carry out suitable and efficient correction techniques to aid at-risk students in achieving good performance and learning outcomes. By focusing on the future academic results in the following academic session, it also helps academic planners, both students and parents make better decisions on how to track student achievement on a semester-by-semester basis.

This study emphasizes how IoT-based attendance systems have the potential to offer perceptions into student behavior and its influence on academic achievement. These data-driven strategies help educational institutions spot early indications of student underperformance and take proactive steps to improve student success.

These results open the door to more investigation into how to improve predictive models and contribute to continuing initiatives to raise educational outcomes through data-driven insights. The relevant and practical issue in education by exploring the use of Wi-Fi IoT attendance systems for predicting student academic performance was addressed and this findings have direct implications for improving educational outcomes, through applicable and credible approach that allows approach allows for data-driven insights and demonstrates the potential of technology to enhance

decision-making in education with the help of algorithm works best for the specific context.

Future Work:

While the predictive capabilities of IoT-based attendance systems are well-understood by this study, there are still a number of opportunities for further research and development by incorporating extra features that can be obtained outside of the classroom, such as participation in extracurricular activities, social interactions, etc. Making composite features that capture many facets of student behaviour and performance and these additional variables can help to make forecasts more accurate. Since the quality of the data gathered has a significant impact on the accuracy of predictions. There is a chance of overfitting, which occurs when models perform well on training data but are unable to generalize to new data, depending on the size of the dataset and the complexity of the models.

Moreseo, Provision of front end interface that is linked with blockchain technology for prediction data reliability authentication by implementing smart contract to save and retrieve the data.

REFERENCE

A Review on Predictive Modeling Technique for Student Academic Performance Monitoring | MATEC Web of Conferences. (n.d.). Retrieved March 24, 2023, from https://www.matec-conferences.org/articles/matecconf/abs/2019/04/matecconf_eaaic2018_03004/matecconf_eaaic2018_03004.html

Abidemi, S. U., Mubarak, A. S., & Akanni, O. (n.d.). *Attendance System via Internet of Things , Blockchain and Artificial Intelligence Technology : Literature Review*.

Abidemi, S. U., Oghenetega, O. G., Daniel, S. A., & Al-Turjman, F. (2022). Wi-Fi Attendance System in the IoT Era. *Lecture Notes on Data Engineering and Communications Technologies*, *130*, 19–29. https://doi.org/10.1007/978-3-030-99581-2_3/COVER

Abu Zohair, L. M. (2019). Prediction of Student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, *16*(1). https://doi.org/10.1186/s41239-019-0160-3

Aziz, A. A., Hafieza, N., & Ahmad, I. (2014). *First Semester Computer Science Students ' Academic Performances Analysis by Using Data Mining Classification Algorithms*. *September*, 15–16.

Carvalho Brito, A. E. S., European Technology Institute, Universidade do Porto, ECEC (15 2008.04.09-11 Porto), European Concurrent Engineering Conference (15 2008.04.09-11 Porto), FUBUTEC (5 2008.04.09-11 Porto), & Future Business Technology Conference (5 2008.04.09-11 Porto). (2008). *Using data mining to predict secondary school student performance*. https://repositorium.sdum.uminho.pt/handle/1822/8024

Dinh, H., To Loan Pham, T., Loan, L., Cu Nguyen, G., Thi Ha, D., Thi To Loan, P., Nguyen Giap, C., & Thi Lien Huong, N. (2020). An Empirical Study for Student Academic Performance Prediction Using Machine Learning Techniques Knowledge hub on Commercial Determinants of Health in Viet Nam View project My studies are on chemistry and bioactivities of natural products View project An Empirical Study for Student Academic Performance Prediction Using Machine Learning Techniques. *Article in International Journal of Computer Science and Information Security*.

https://www.researchgate.net/publication/340351415

Explain what Flask is and its benefits? | i2tutorials. (n.d.). Retrieved May 2, 2023, from https://www.i2tutorials.com/explain-what-flask-is-and-its-benefits/

For, S., Performance, S., Mutune, C., Ivy, C., David, P., Danny, S., & Bydon, S. (2021). *the University of Zambia the School of Education Department of Library and Information Science Performance Predictor : a Data Mining and Machine Learning*.

Gump, S. E. (2010). The Cost of Cutting Class: Attendance As A Predictor of Success. *Https://Doi.Org/10.3200/CTCH.53.1.21-26*, *53*(1), 21–26. https://doi.org/10.3200/CTCH.53.1.21-26

Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm. *Procedia Technology*, *25*, 326–332. https://doi.org/10.1016/J.PROTCY.2016.08.114

Hassan, R. J., Zeebaree, S. R. M., Ameen, S. Y., Kak, S. F., Sadeeq, M. A. M., Salih Ageed, Z., Al-Zebari, A., & Salih, A. A. (2021). Editor(s): (1) Dr. Dariusz Jacek Jakóbczak. *Asian Journal of Research in Computer Science*, *8*(3), 32–48. https://doi.org/10.9734/AJRCOS/2021/v8i330202

Hazaa, K. Al, Ismail, R., Al-tameemi, R. A. N., Romanowski, M. H., Ben, M., Rhouma, H., Elatawneh, A., Al, K., Ismail, R., Al-tameemi, R. A. N., Romanowski, M. H., Bensaid, A., & Ben, M. (2021). The effects of attendance and high school GPA on student performance in first-year undergraduate courses The effects of attendance and high school GPA on student performance in first-year undergraduate courses. *Cogent Education*, *8*(1). https://doi.org/10.1080/2331186X.2021.1956857

Imran, M., Latif, S., Mehmood, D., & Shah, M. S. (n.d.). *Student Academic Performance Prediction using Supervised Learning Techniques*. *d*, 92–104.

Ismail, N. A., Chai, C. W., Samma, H., Salam, M. S., Hasan, L., Wahab, N. H. A., Mohamed, F., Leng, W. Y., & Rohani, M. F. (2022). Web-based University Classroom Attendance System Based on Deep Learning Face Recognition. *KSII Transactions on Internet and Information Systems*, *16*(2), 503–523. https://doi.org/10.3837/tiis.2022.02.008

Kaunang, F. J., & Rotikan, R. (2018). Students' academic performance prediction using data mining. *Proceedings of the 3rd International Conference on*

Informatics and Computing, ICIC 2018. https://doi.org/10.1109/IAC.2018.8780547

Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2010). *PREDICTING STUDENTS ' PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES PERFORMANCE IN DISTANCE LEARNING USING MACHINE*. *9514*. https://doi.org/10.1080/08839510490442058

Kumar, M., Mehta, G., Nayar, N., & Sharma, A. (2021). EMT: Ensemble meta-based tree model for predicting student performance in academics. *IOP Conference Series: Materials Science and Engineering*, *1022*(1). https://doi.org/10.1088/1757-899X/1022/1/012062

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *2015-Augus*, 1909–1918. https://doi.org/10.1145/2783258.2788620

Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review. (2021). *TEM Journal*, *10*(4), 1919–1927.

Malini, J., & Kalpana, Y. (2021). *Analysis Of Factors Affecting Student Performance Evaluation Using Education Dataminig Technique Analysis Of Factors Affecting Student Performance Evaluation Using Education Dataminig Technique*. *12*(7), 2413–2424.

Mat, A., Zulkarnaen Khidzir, N., Deris, S., Nurul, N., Yaacob, H. M., Saberi Mohamad, M., & Safaai, S. S. (2019). A Review on Predictive Modeling Technique for Student Academic Performance Monitoring. *MATEC Web of Conferences*, *255*, 03004. https://doi.org/10.1051/MATECCONF/201925503004

Nabil, A., Seyam, M., & Abou-elfetouh, A. (2021). Prediction of Students ' Academic Performance Based on Courses ' Grades Using Deep Neural Networks. *IEEE Access*, *PP*, 1. https://doi.org/10.1109/ACCESS.2021.3119596

Ofelia San Pedro, M. Z., SJd Baker, R., Bowers, A. J., & Heffernan, N. T. (n.d.). *Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School*. Retrieved March 31, 2023, from http://www.studentclearinghouse.org

Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). Application of k Means

Clustering algorithm for prediction of Students Academic Performance. *IJCSIS) International Journal of Computer Science and Information Security*, *7*(1). https://arxiv.org/abs/1002.2425v1

Puah, S. (2021). *Predicting Students' Academic Performance: A Comparison between Traditional MLR and Machine Learning Methods with PISA 2015*. *January*. https://doi.org/10.31234/osf.io/2yshm

Qu, S., Li, K., Zhang, S., & Wang, Y. (2018). Predicting Achievement of Students in Smart Campus. *IEEE Access*, *6*, 60264–60273. https://doi.org/10.1109/ACCESS.2018.2875742

Rai, S., Shastry, K. A., Pratap, S., Kishore, S., Mishra, P., & Sanjay, H. A. (2021). Machine Learning Approach for Student Academic Performance Prediction. *Advances in Intelligent Systems and Computing*, *1176*, 611–618. https://doi.org/10.1007/978-981-15-5788-0_58/COVER

Saleem, F., Ullah, Z., Fakieh, B., & Kateb, F. (2021). Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning. *Mathematics*, *9*(17), 1–22. https://doi.org/10.3390/math9172078

Saluja, R., Rai, M., & Saluja, R. (2023). Designing new student performance prediction model using ensemble machine learning. *Journal of Autonomous Intelligence*, *6*(1), 1–12. https://doi.org/10.32629/jai.v6i1.583

Song, X., Li, J., Sun, S., Yin, H., Dawson, P., & Doss, R. R. M. (2021). SEPN: A Sequential Engagement Based Academic Performance Prediction Model. *IEEE Intelligent Systems*, *36*(1), 46–53. https://doi.org/10.1109/MIS.2020.3006961

Tao, T., Sun, C., Wu, Z., Yang, J., & Wang, J. (2022). Deep Neural Network-Based Prediction and Early Warning of Student Grades and Recommendations for Similar Learning Approaches. *Applied Sciences (Switzerland)*, *12*(15). https://doi.org/10.3390/app12157733

The Flask Web Framework: A Beginner's Guide. (n.d.). Retrieved May 2, 2023, from https://careerfoundry.com/en/blog/web-development/what-is-flask/

Walia, N., Kumar, M., Nayar, N., & Mehta, G. (2020). Student's Academic Performance Prediction in Academic using Data Mining Techniques. *SSRN Electronic Journal*, 1–5. https://doi.org/10.2139/ssrn.3565874

Yadav, S. K., Bharadwaj, B., & Pal, S. (2012a). *Data Mining Applications: A comparative Study for Predicting Student's performance*. 13–19. http://arxiv.org/abs/1202.4815

Yadav, S. K., Bharadwaj, B., & Pal, S. (2012b). Data Mining Applications: A comparative Study for Predicting Student's performance. *INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY & CREATIVE ENGINEERING*, *1*(12), 2045–2711. https://arxiv.org/abs/1202.4815v2