NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

EXPLORATORY DATA ANALYSIS AND PREDICTION FOR UNEMPLOYMENT AMONG INDIVIDUALS WITH MENTAL HEALTH CHALLENGES

MASTER THESIS

Adeoluwa ATANDA

Nicosia

January, 2023

ADEOLUWA ATANDA
EXPLORATORY DATA ANALYSIS AND PREDICTION FOR UNEMPLOYMENT AMONG INDIVIDUALS WITH MENTAL HEALTH CHALLENGES
2023

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

EXPLORATORY DATA ANALYSIS AND PREDICTION FOR UNEMPLOYMENT AMONG INDIVIDUALS WITH MENTAL HEALTH CHALLENGES

MASTER THESIS

Adeoluwa ATANDA

Supervisor

Prof. Dr Nadire CAVUS

Nicosia

January, 2023

Approval

We certify that we have read the thesis submitted by Atanda Adeoluwa Fiyinfoluwa titled **"Exploratory Data Analysis and Prediction for Unemployment Among Individuals with Mental Health Challenges"** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of MSc in Computer Information Systems.

Examining Committee	**Name-Surname**	**Signature**
Head of the Committee:	Assist. Prof. Dr. Sahar Ebadinezhad	....
Committee Member:	Assoc. Prof. Dr. Nuriye Sancar	....
Supervisor:	Prof. Dr. Nadire Cavus	....

Approved by the Head of the Department

24../.04/2023

Prof. Dr. Nadire Çavuş
Head of Department

Approved by the Institute of Graduate Studies

NEAR EAST UNIVERSITY INSTITUTE OF GRADUATE STUDIES

..../..../2023

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

Declaration

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

Adeoluwa Atanda
20/01/2023

Acknowledgments

I would like to express my sincerest gratitude to my supervisor, Prof. Dr. Nadire CAVUS, for her unwavering support, guidance, and patience throughout my thesis journey. Her invaluable supervision and mentorship played a crucial role in helping me to complete this research to the best of my abilities. I am deeply grateful for her wisdom, guidance and for steering me in the right direction throughout my graduate program.

I would also like to extend my heartfelt appreciation to my family, for their love, encouragement, and unwavering support throughout my academic journey. Their constant prayers and belief in me have been my source of strength and motivation.

I am also grateful for my friends, who have been a constant source of support and inspiration throughout this journey. Their genuine encouragement and support have been invaluable.

Adeoluwa Atanda

Abstract

Exploratory Data Analysis and Prediction for Unemployment Among Individuals with Mental Health Challenges

Atanda, Adeoluwa Fiyinfoluwa

MSc, Department of Computer Information Systems

Prof. Dr. Nadire Cavus (Supervisor)

January, 2023, 115 pages

This thesis presents a machine learning approach to the exploratory data analysis and prediction of unemployment among individuals with mental health challenges. The goal of the research is to understand the factors that contribute to unemployment among this population and to develop models that can accurately predict unemployment among individuals with mental health challenges. The purpose of the study is to address the gap in the existing literature by examining the relationship between mental health issues and unemployment. Specifically, the study aims to explore this relationship through the use of exploratory data analysis and various machine learning models, including logistic regression, random forest, support vector machine (SVM), naive bayes, decision tree classifier, and k-nearest neighbours. The study will seek to identify patterns and trends in the data related to unemployment among people with mental health issues. By utilizing multiple machine learning models, the study can evaluate which models perform best in predicting unemployment among this population. The study employed a comprehensive data analysis and pre-processing technique to prepare the dataset for modelling. The models were trained and evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. The results were compared, and the best model was selected based on the highest overall performance. The study found that SVM model performed the best with an accuracy of 91%, followed by Random Forest with an accuracy of 89%. The results of this study provide valuable insights for policymakers and practitioners in the field of mental health and unemployment by demonstrating the effectiveness of machine learning models in predicting unemployment among individuals with mental health challenges. This study also makes a significant contribution to the field by providing a more comprehensive understanding of the relationship between mental health and unemployment through the use of multiple advanced machine learning models. The benefits of this study include providing a more comprehensive understanding of the relationship between mental health and unemployment, identifying an effective machine learning model for predicting unemployment among individuals with mental health challenges, providing valuable insights for policymakers and practitioners in the field of mental health and unemployment, and making a significant contribution to the field of mental health and unemployment research. This study can be used as a reference for future research on this topic, and also, it can be used by the policymakers to make a decision and also help the practitioners on how to approach this issue.

Keywords: Unemployment, mental health, exploratory data analysis, machine learning, prediction

Özet

Zihinsel Sağlık Sorunu Yaşayan Bireyler Arasındaki İşsizlik İçin Keşifsel Veri Analizi Ve Tahmin

Atanda, Adeoluwa Fiyinfoluwa
Yüksek Lisans, Bilgisayar Enformatik Sistemleri Bölümü
Prof. Dr. Nadire Çavuş (Danışman)

Ocak, 2023, 115 sayfa

Bu tez, zihinsel sağlık sorunları olan bireyler arasındaki işsizliğin keşfedici veri analizine ve tahminine yönelik bir makine öğrenimi yaklaşımı sunmaktadır. Araştırmanın amacı, bu popülasyonda işsizliğe katkıda bulunan faktörleri anlamak ve ruh sağlığı sorunları olan bireyler arasındaki işsizliği doğru bir şekilde tahmin edebilecek modeller geliştirmektir. gelişmiş bir makine öğrenimi yaklaşımı kullanıyor. Çalışma, lojistik regresyon, rastgele orman, destek vektör makinesi (SVM), naive bayes, karar ağacı sınıflandırıcı ve k-en yakın komşular dahil olmak üzere çok çeşitli makine öğrenimi modellerini dahil ederek literatürdeki boşluğu özellikle ele alıyor. Çalışmanın amacı, ruh sağlığı sorunları olan bireyler arasında işsizliği tahmin etmede hangi modelin en etkili olduğunu belirlemektir. Çalışma, veri setini modellemeye hazırlamak için kapsamlı bir veri analizi ve ön işleme tekniği kullandı. Modeller, doğruluk, kesinlik, geri çağırma ve F1 puanı gibi standart performans ölçütleri kullanılarak eğitildi ve değerlendirildi. Sonuçlar karşılaştırıldı ve en yüksek genel performansa göre en iyi model seçildi. Çalışma, SVM modelinin %91 doğrulukla en iyi performansı gösterdiğini, ardından %89 doğrulukla Random Forest modelinin geldiğini buldu. Bu çalışmanın sonuçları, akıl sağlığı sorunları olan bireyler arasında işsizliği tahmin etmede makine öğrenimi modellerinin etkinliğini göstererek, ruh sağlığı ve işsizlik alanındaki politika yapıcılar ve uygulayıcılar için değerli bilgiler sağladı. Bu çalışma ayrıca, çoklu gelişmiş makine öğrenimi modellerinin kullanımı yoluyla ruh sağlığı ve işsizlik arasındaki ilişkinin daha kapsamlı bir şekilde anlaşılmasını sağlayarak alana önemli bir katkı sağlamaktadır. Bu çalışmanın faydaları arasında ruh sağlığı ve işsizlik arasındaki ilişkinin daha kapsamlı bir şekilde anlaşılmasını sağlamak, akıl sağlığı sorunları olan bireyler arasındaki işsizliği tahmin

etmek için etkili bir makine öğrenimi modeli belirlemek, ruh sağlığı alanında politika yapıcılar ve uygulayıcılar için değerli bilgiler sağlamak yer alarak ruh sağlığı ve işsizlik araştırmaları alanına önemli bir katkı sağlamaktır. Bu çalışma, gelecekte bu konuda yapılacak araştırmalar için bir referans olarak kullanılabilir ve ayrıca politika yapıcılar tarafından karar vermede kullanılabilir ve uygulayıcılara bu konuya nasıl yaklaşılacağı konusunda yardımcı olabilir.

Anahtar Kelimeler: İşsizlik, ruh sağlığı, keşifsel veri analizi, makine öğrenimi, tahmin

Table of Contents

Approval ... i
Declaration ... ii
Acknowledgements ... iii
Abstract ... iv
Özet ... v
Table of Contents ... vi
List of Figures ... viii
List of Tables ... ix
List of Abbreviations ... x

CHAPTER I

INTRODUCTION
Background ... 1
Statement of the Problem ... 3
The Contribution to the Field of Study ... 4
Purpose of the Study ... 5
Significance of the Study ... 5
Limitations ... 6
Definition of Terms ... 7
Thesis Overview ... 8

CHAPTER II

LITERATURE REVIEW
Theoretical Framework ... 10
Unemployment ... 10
Mental Health Challenges ... 14
Relationship Between Unemployment and Mental health ... 16
Data Mining ... 19
Machine Learning ... 21
Label Encoding ... 36
Data Standardization ... 37
One Hot Encoder ... 37
Related Research ... 38

CHAPTER III

METHODOLOGY

Research Design 42
Dataset 43
Using an Existing Dataset 44
Data Access Permission 44
Data Analysis 48
Data Pre-processing 48
Research Procedure 49
Hardware & Software Specification 51

CHAPTER IV

RESULTS AND FINDINGS

Library Imports 53
Data Preparation 54
Exploratory Data Analysis Results.. 54
Machine Learning 65
Pre-processing for Machine Learning 66
Results for the Proposed Model 68
Confusion Matrix 70
Evaluation Metrics 72
Final Predictions Using the Best Model 73

CHAPTER V

DISCUSSION

Comparison with Previous Studies 74

CHAPTER VI

CONCLUSION and RECOMMENDATIONS

Conclusion 77
Recommendations 78
Recommendations for Employers 78
Recommendations for Governments 79
Recommendations for Researchers 80

REFERENCES 82

APPENDICES 99

Appendix A: Ethics Committee Approval 100

Appendix B: Turnitin Report ……….. 101

List of Figures

Figure 2.1: Negative effects of unemployment ………………………………….……. 13

Figure 2.2: Data mining processes ………………………………….………….... 20

Figure 2.3: Machine learning as a subset of artificial intelligence ……………..…. 22

Figure 2.4: Machine learning techniques ……………..…………………...…… 36

Figure 3.1: Research design ……….……………….……………..……………..… 43

Figure 3.2: Data access statement …………………………………..……………... 45

Figure 3.3: A simple block diagram for the proposed model ………..…………..… 51

Figure 4.1: Library imports ……………………………..…..…………………... 54

Figure 4.2: Count-plot of education levels …………………………..……..…... 55

Figure 4.3: Count-plot of mental health state ……….………………………..…… 56

Figure 4.4: Pie chart for education level distribution according to mental health state ……………………………………………………………….…...… 57

Figure 4.5: Boxplot to compare the distributions of annual income for people with and without mental illness …………………………………...…………… 58

Figure 4.6: Violin plot to compare the distributions of annual income for people with and without mental illness ………………………………………………… 59

Figure 4.7: Bar chart for unemployment rate for each mental health Issue ………………………………………………………………………….. 60

Figure 4.8: Bar chart for unemployment rate for individuals in each of these age ranges who reported having a mental illness ……………………......……. 61

Figure 4.9: Unemployment rate by gender for people with mental health issues …………………………………………………………………………... 62

Figure 4.10: Unemployment rate by education level …………….…………..…. 63

Figure 4.11: Correlations and plot heatmap for the dataset ………..………..…… 64

Figure 4.12: Correlations and plot heatmap for people with mental illness …..…. 65

Figure 4.13: Confusion matrix for the best proposed model (SVM) ……………… 71

Figure 4.14: Final prediction on the entire dataset using our best model ….……. 73

List of Tables

Table 2.1: Supervised versus unsupervised learning ……..………..…………….. 33

Table 3.1: Summary of categorical and numerical attributes present in the dataset .. 46

Table 3.2: Summary of binary attributes present in the dataset ……………..……… 47

Table 4.1: Results of the models ……..……………………..……………………. 73

Table 5.1: Comparison from previous studies ………..…………………………..… 76

List of Abbreviations

EDA – Exploratory Data Analysis

CNN – Convolutional Neural Network

RF – Random Forest

SVM – Support Vector Machine

ML – Machine Learning

KNN – K-Nearest Neighbour

SMI – Severe Mental Illness

MH – Mental Health

QoL – Quality of Life

TP – True Positive

TN – True Negative

CHAPTER ONE
INTRODUCTION

The introduction chapter in this study provides a background on the topic and highlights the statement of the problem. The chapter also identifies the contribution, purpose, and significance of the study, as well as the limitations and definition of terms. A thesis overview is also presented to give a brief outline of the following chapters.

1.1 Background

Unemployment is a significant social and economic problem that disproportionately affects individuals with mental health challenges. According to the World Health Organization (WHO, 2018), people with mental health issues are at a higher risk of being unemployed, and tend to face longer spells of joblessness compared to those without mental health issues. This is concerning, as unemployment has been linked to negative consequences for mental and physical health, as well as financial wellbeing (WHO, 2018). It has been challenging to determine how the pandemic has affected employment since conventional employment measurements are insufficient to account for pandemic activities, such as the temporary shutdown of some industries and the shift to fewer hours (if any at all) owing to the furlough system. An increase in unemployment is anticipated since pandemic assistance programs are scheduled to terminate soon in most countries (Eisenberg-Guyot et al., 2020). Both in industrialized and emerging nations, there are today more unemployed individuals than ever before (Kuhn et al., 2018).

There are several factors that contribute to the high rates of unemployment among individuals with mental health challenges. Many individuals with mental health conditions continue to experience stigma and discrimination in the workplace, which can take various forms, such as being unfairly passed over for employment, receiving unfair treatment or being subjected to negative stereotypes, ultimately leading to difficulties in securing and maintaining employment (Mental Health America, 2021). Additionally, individuals with mental health conditions may encounter various barriers related to their specific conditions, such as cognitive difficulties or challenges with social interactions, which can contribute to their difficulties in securing and maintaining employment. Given the crucial role that employment plays in both mental and physical health and financial wellbeing, it is imperative to develop effective strategies that can address these barriers and improve employment outcomes for this population (WHO, 2018). One potential approach is the use of machine learning techniques for exploratory data analysis and prediction, which can help to identify patterns and trends in data that may be relevant to employment outcomes (Huang et al., 2019). One potential approach is the use of machine learning techniques for exploratory data analysis and prediction, which can help to understand trends and pattern in the data which may be relevant to employment outcomes.

A kind of artificial intelligence known as “machine learning” uses algorithms to learn from data and make predictions or choices (Cavus et al., 2023; Jordan & Mitchell., 2015; Lawan et al., 2023). It has been applied successfully in a variety of contexts, including healthcare (Adweb et al., 2021; Ghassemi et al., 2014; Lawan & Cavus, 2022), finance (Lopez de Prado, 2016), banking (Cavus et al., 2022), and marketing (Kohavi et al., 2009). In the context of mental health and employment, machine learning can be a useful tool because it can examine big datasets to find patterns or trends that may be pertinent to employment outcomes, like success in job applications or resumes. It is now possible to identify specific patterns that are indicative of employment success by training machine learning algorithms on these datasets which can then guide the development of efficient strategies to improve the employment outcomes for people with mental health conditions (Huang et al., 2019).

There have been a few studies that have used machine learning to examine the relationship between mental health and employment, but there is a need for more research in this area (Huang et al., 2019; Kim et al., 2016). The current study aims to fill this gap by using a machine learning approach to explore the relationship between mental health and unemployment among individuals with mental health challenges, and to develop predictive models that can help to identify those at risk of unemployment. The study will use a dataset that includes information on mental and physical complaints, demographic information, and employment status. The dataset was collected on an internet website and 334 respondents participated. The study will apply six different machine learning models (logistic regression, random forest, SVM, naive bayes, decision tree classifier, and KNN) to predict unemployment and compare the results of these models. The results of the study will be used to develop predictive models that can be used to identify individuals at risk of unemployment and to develop strategies for improving employment outcomes.

Overall, the current study aims to contribute to the growing body of knowledge on the use of machine learning for exploratory data analysis and prediction in the context of mental health and employment, with the ultimate goal of improving employment outcomes for individuals with mental health challenges. By identifying patterns or trends that are predictive of employment success, the study has the potential to inform the development of targeted interventions and strategies that can help individuals with mental health challenges achieve better employment outcomes.

1.2 Statement of the Problem

Despite the numerous efforts to address mental health issues in society, individuals with mental health challenges continue to experience disproportionately high rates of unemployment (Kopelovich et al., 2020). This can have serious consequences for the individual, including financial strain, social isolation, and a decrease in overall quality of life (Griep et al., 2015). There is a lack of research on the specific factors that contribute to unemployment among individuals with mental health challenges and how these factors can be predicted and potentially mitigated (Bartelink et al., 2019).

Traditional methods of analysis and prediction, such as linear regression and decision trees, may not be sufficient for accurately identifying and understanding the complex relationships at play in this issue (Giugiario et al., 2011). Machine learning algorithms, on the other hand, have the ability to process and examine huge amounts of data in order to recognize patterns and relationships that may not be immediately apparent (Ge et al., 2017). However, the use of machine learning approaches raises concerns regarding their accuracy and fairness, as well as ethical and privacy issues (Jain et al., 2021).

There is a need for more robust and accurate methods for predicting unemployment among individuals with mental health challenges, as this can inform targeted interventions and support for this vulnerable population (Harvey et al., 2017). Machine learning approaches have the potential to deliver valuable insights into the complex relationship between unemployment and mental health, but there has been limited research on the application of these methods in this context (Jain et al., 2021). Therefore, the problem statement of this thesis is to explore the use of exploratory data analysis and machine learning approaches to predict unemployment among individuals with mental health challenges and address the limitations and concerns associated with their application.

1.3 The Contribution to the Field of Study

This study aims to make a significant contribution to the field of mental health and employment by using machine learning techniques to explore and predict unemployment among individuals with mental health challenges. To the best of our knowledge, there has been limited research on this topic using machine learning methods, and this study represents a unique and innovative approach to understanding and addressing this important issue. The use of machine learning allows for the analysis of large datasets and the identification of patterns and relationships that may not be immediately apparent using traditional methods of analysis. The results of this study have the potential to inform policy and practice in the area of mental health and employment, and to improve the lives of individuals with mental health challenges. The findings of this study may also have broader implications for other fields of study, such as labour economics and public policy, as they relate to employment and mental health. In addition, this study addresses important ethical and practical considerations surrounding the use of machine learning in this context, including issues of privacy and fairness.

These contributions make this study a valuable addition to the field of mental health and employment, and have the potential to inform future research and practice in this area. Overall, the goal of this study is to provide a better understanding of the factors that contribute to unemployment among individuals with mental health challenges and to identify potential strategies for addressing this issue. By doing so, this study aims to make a meaningful contribution to the field of mental health and employment, and to improve the lives of individuals with mental health challenges.

1.4. The Purpose of the Study

The purpose of this study is to investigate the relationship between unemployment and mental illness among individuals and to explore the potential of machine learning models to accurately predict unemployment based on mental illness and other relevant features

The research questions of the study are as follow:

- What are the relationships between unemployment and mental illness among individuals in the given data set?
- How do different machine learning models compare in terms of their ability to predict unemployment based on mental illness and other relevant features?
- How can the results of this analysis be used to better understand and address the relationship between unemployment and mental illness among individuals with mental health challenges?

1.5 The Significance of the Study

The significance of this study on exploratory data analysis and prediction for unemployment among individuals with mental health challenges using a machine learning approach lies in its potential to provide valuable insights and inform decision-making related to the employment of individuals with mental health challenges. Mental health conditions can often present significant challenges to employment, and understanding the factors that contribute to unemployment among this population can help identify potential interventions and strategies to improve employment outcomes. By using machine learning techniques to analyse data and make predictions about unemployment, the study may be able to identify patterns and trends that can inform policy and practice. Moreover, the use of machine learning in this context can provide a more objective and accurate assessment of the factors contributing to unemployment among individuals with mental health challenges, as compared to traditional statistical methods that may be subject to bias or error. Overall, this study has the potential to make a significant contribution to the field of mental health and employment by providing valuable insights and inform decision-making related to the employment of individuals with mental health challenges.

1.6 Limitations

One of the limitations of this study is the imbalanced dataset that we used, with 74.25% of the respondents being employed and 25.75% being unemployed. This imbalance may have affected the performance of the models and led to high predictive accuracy. It is important to note that the imbalance of the dataset does not invalidate the results of the study, but it should be taken into account when interpreting the results and drawing conclusions. Further research is needed to address this limitation and to improve the performance of the models. Another limitation of this study is the sample size of 334 respondents. While the sample size is large enough to provide meaningful results, it may not be representative of the larger population. Therefore, it is important to consider this limitation when generalizing the results to a larger population. Additionally, the data used in this study were self-reported, which may introduce bias and affect the accuracy of the results. In future studies, it would be beneficial to use other sources of data, such as administrative data or objective measures, to reduce bias and improve the accuracy of the results.

Furthermore, this study is based on a single dataset, and it would be beneficial to replicate the results using different datasets. This would help to ensure the robustness and generalizability of the findings. The dataset used in this study includes a limited number of features, and it would be beneficial to include more features in future studies to improve the performance of the models. Moreover, the study is based on one country, and it would be interesting to replicate the study in different countries to see if the results are consistent. Another limitation of this study is that individuals may be shy to reveal their mental states, as we found that some respondents who claimed not to have been diagnosed with mental health issues exhibited many symptoms of mental health issues. This may lead to underreporting of mental health issues and affect the accuracy of the results. This highlights the importance of considering self-reported data as a limitation and the need for further research to investigate the extent to which individuals are shy to reveal their mental states. The results of this study should be interpreted with this limitation in mind, and future studies could consider using other methods to collect data, such as diagnostic interviews, to reduce bias and improve the accuracy of the results.

In summary, this study is not without limitations, but it provides valuable insights into the relationship between unemployment and mental illness among individuals with mental health challenges and demonstrates the potential of machine learning models to accurately predict unemployment based on mental illness and other relevant features.

1.7 Definition of Terms

Mental health: Mental health is a critical aspect of overall well-being and refers to the state of being mentally and emotionally well. It encompasses a wide range of emotional, psychological, and social well-being. In this study, mental health is used to refer to the mental well-being of individuals with mental health challenges.

Unemployment: Unemployment is a measure of the number of individuals who are actively seeking employment but are currently without a job. In this study, unemployment is used to refer to the state of being without a job among individuals with mental health challenges.

Individuals with mental health challenges: This study focuses on individuals with mental health challenges, which refers to individuals who have been diagnosed with a mental health condition or who have experienced symptoms of a mental health condition. This population is particularly vulnerable to unemployment due to the nature of their condition.

Machine learning: Machine learning is a type of artificial intelligence that allows computers to learn and make predictions without explicit programming. In this study, machine learning techniques are used to predict unemployment among individuals with mental health challenges.

Exploratory data analysis: Exploratory data analysis (EDA) is the process of examining and summarizing data in order to gain a better understanding of it. In this study, EDA is used to understand the relationship between mental health and unemployment among individuals with mental health challenges.

Prediction: Prediction refers to the process of using data to make informed guesses about future events or outcomes. In this study, prediction models are developed using machine learning techniques to predict unemployment among individuals with mental health challenges.

1.8 Thesis Overview

The thesis is divided into six chapters, which are organized as follows:

Chapter 1: Introduction - Provides an overview of the research topic, including the background, research question, and objectives of the study. It sets the stage for the rest of the thesis, providing a clear and concise introduction to the topic and its significance.

Chapter 2: Literature Review - Offers a comprehensive examination of the existing literature on the topic of unemployment among individuals with mental health challenges. This chapter provides a detailed review of the theories, studies, and research that have been conducted in this area, and it helps to establish the context and background for the research being presented in the thesis.

Chapter 3: Research Methodology - Outlines the research methods and techniques that were employed in the study. This chapter provides a detailed description of the data collection and analysis methods, including the sampling strategy, data preparation, and statistical analysis. It also provides an overview of the machine learning models used in the study and the feature selection algorithms that were applied.

Chapter 4: Results - Presents the findings of the research, including the results of the statistical analysis and the performance of the machine learning models. This chapter provides a detailed examination of the results, including tables, figures, and visualizations that help to clarify the findings and make them easier to understand.

Chapter 5: Discussion - Involves a critical examination of the results and their implications. This chapter provides an in-depth analysis of the findings, including a discussion of the limitations of the study and the implications of the results for policymakers, practitioners, and future research. It also includes a comparison of the performance of different machine learning models used in the study and the most important features for predicting unemployment.

Chapter 6: Conclusion and Recommendations - Offers conclusions, observations, and recommendations for future research. This chapter provides a summary of the main findings of the study, and it highlights the contributions of the research to the field of unemployment and mental health challenges. It also provides recommendations for future research and for policy and practice, which can help to improve the understanding of this complex topic.

CHAPTER TWO
LITERATURE REVIEW

The literature review chapter in this study examines the theoretical framework of the study and explores the relationship between unemployment and mental health. The chapter also reviews important tools for data analysis such as data mining, machine learning, and data preparation techniques. Additionally, the chapter presents a comprehensive review of related research in the field, which helps to identify gaps in the current knowledge and guide the research design.

2.1 Theoretical Framework

2.1.1 Unemployment

Regular studies have shown that being unemployed, which is defined as being without a job but actively seeking one, negatively affects a number of health outcomes (O'Campo et al., 2015). There are many ways that being unemployed could be bad for your health, including the financial strain, job insecurity, and reduced future earning potential (Burr., 2021). Unemployed people may also suffer from stress and low self-esteem due to the stigma associated with being unemployed, which over time harms their health (Kalleberg., 2011). The high rate of unemployment among individuals with mental health issues is a significant problem, with rates of unemployment among this population being significantly higher than the general population. This represents a significant loss of productivity and talent, as well as a burden on the healthcare system, as individuals with mental health issues who are unable to find work are more likely to require mental healthcare services (Barr et al., 2012).

There are several ways to address the relationship between mental health and unemployment. One approach is to provide mental health support and accommodations in the workplace, such as flexible scheduling and access to mental health resources (O'Campo et al., 2015). This can make it easier for individuals with mental health challenges to succeed in the workplace and reduce the stigma surrounding mental health (Schneider et al., 2019). Another approach is to provide job training and support for individuals with mental health issues. This can include vocational rehabilitation programs and job placement services, which can help people with mental health issues learn the skills they need to succeed in the workplace (O'Campo et al., 2015). Additionally, addressing the stigma surrounding mental health is crucial in reducing unemployment among people with mental health issues. This can be done through public education campaigns and efforts to promote understanding and acceptance of mental health issues (Waghorn & Hielscher., 2014). People affected by health issues are often less likely to be employed compared to their healthy counterparts. In the UK, 300,000 people abandon their jobs each year and turn to health benefits, resulting in a £13 billion cost for the government and £9 billion for employers (O'Campo et al., 2015). The adverse effects of poor mental health on employment are particularly concerning, with only 35% of people struggling with mental health having jobs (Modini et al., 2016) and a 40% gap in employment rates between those with and without mental health problems (Thomeer et al., 2015). The World Health Organization (WHO) (2008) indicates that mental health disorders make up 13% of the total global disease burden, while the OECD/EU (2018) reports that over one in six individuals in the EU experience mental health problems in a given year. Despite its significant impact, mental health is often neglected in terms of healthcare resources and hidden in the workplace due to stigma and discrimination (WHO, 2013). It is important to distinguish between unemployment and being out of the labour force when studying the relationship between employment and mental health. Unemployment refers to individuals who are without paid employment but are actively seeking work, while those who are not actively looking for work are classified as "inactive". Some forms of inactivity may have a similar impact on mental health as unemployment, particularly for individuals who sincerely want to work but have become disheartened and stopped looking for employment (Sonnentag, 2018).

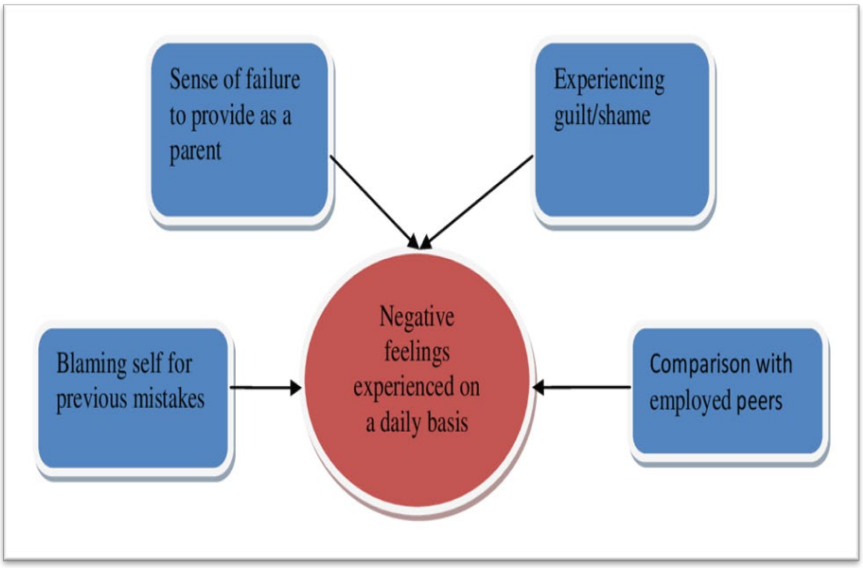
Unemployment is a significant social and economic problem that affects individuals and communities around the world. According to the International Labour Organization, the global unemployment rate was 5.7% in 2020, with nearly 210 million people unemployed worldwide. Unemployment is defined as the condition of people who are able, willing, and actively seeking work, but who are unable to find it (ILO, 2021). The consequences of unemployment can be severe for individuals, leading to financial difficulties, social detachment, and harm to both mental and physical health (Kopelovich et al., 2020). There are various reasons that result in unemployment, including economic factors, technological advancements, and demographic changes. The COVID-19 pandemic has made it mandatory for all businesses to start digital transformation (Cavus and Sancar, 2023). The COVID-19 pandemic has also had a significant impact on unemployment rates, with many people losing their jobs or experiencing reduced hours due to the economic downturn caused by the pandemic (ILO, 2021).

Unemployment can have significant consequences for individuals, communities, and societies. For individuals, unemployment can lead to financial hardship, as people who are unemployed often have limited access to income and may have difficulty paying for basic necessities such as food, housing, and healthcare (Thomeer et al., 2015). Unemployment can also have negative impacts on mental health, as people who are unemployed may experience feelings of despair, isolation, and a loss of self-worth. At the community level, high rates of unemployment can lead to social unrest and instability, as people may feel disenfranchised and disconnected from society (ILO, 2021). High levels of unemployment can also have negative impacts on the economy, as people who are unemployed are not contributing to economic growth and may be reliant on social welfare programs (Linardatou et al., 2014).

It appears that unemployment has an impact on people's psychological well-being. Particularly, it appears that those without jobs have a twofold increased risk of mental illnesses such mood disorders, anxiety, sadness, dysthymia, low self-esteem, and suicide risk (Suijkerbuijk et al., 2017). Notably, among men in particular, the relative risk of suicide plummeted by 37% among unemployed individuals with mental health related issues who were reintegrated into the workforce (Nazarov et al., 2015). While being unemployed can be the root of a mental condition, financial distress has the capacity to be a facilitator of the negative effects of unemployment on an individual's private health (Jansson-Fröjmark et al., 2008 & Linardatou et al., 2014).

Addressing unemployment involves several strategies, such as implementing policies and programs that generate employment opportunities, offering job training and education, and providing support for the unemployed (Davidson et al., 2012). Data science can play a role in understanding unemployment and developing effective strategies for addressing it. By analyzing large datasets on unemployment, data scientists can identify patterns and trends that may be relevant to understanding the causes of unemployment and developing effective interventions. For example, data scientists can use machine learning techniques to identify factors that are predictive of unemployment and to develop models that can help policymakers and practitioners design targeted interventions to support people who are unemployed. Similarly, to the progress made in the study of job quality (Giugiario et al., 2011), there has been a shift towards a more comprehensive understanding of various types of employment along the job quality spectrum (Vanroelen., 2019). The terms and conditions of an employment agreement are known as employment quality (Van Aerden et al., 2014; Vanroelen., 2019), and together with the nature and structure of work described by the employment terms, they make up job quality (Warhurst et al., 2017). Overall, unemployment is a significant social and economic problem that affects individuals and communities around the world (International Labour Organization., 2021). Data science can play a role in understanding and addressing unemployment by analyzing large datasets and developing predictive models that can inform the development of targeted interventions (Thieme et al., 2020). Numerous research studies have explored the connection between mental health and unemployment (Mokona et al., 2020). Findings from these studies suggest that both inactivity and unemployment can have negative consequences on mental health (Mayer & Hollederer, 2022). The majority of studies reviewed in this area have concluded that unemployment has a negative impact on mental health (van Oorschot, 2013). These negative effects may worsen over time and can impact physical and mental health, as well as life satisfaction (Jin & Mosweu, 2017).We try to depict the effects with the simple diagram below;

Figure 2.1.
Negative effect of unemployment (Willemse, 2020)



2.1.2 Mental Health Challenges

Mental health issues, often referred to as mental illnesses, encompass a broad range of disorders that are defined by disruptions in mood, thought, and behaviour. Some common examples of mental illnesses include depression, anxiety disorders, schizophrenia, eating disorders, and addictive behaviours (Miller, 2022). Thomeer et al. (2015) reported that only 35% of individuals with mental health problems are employed and there exists a 40% disparity in employment rates between those with and without mental health issues. Mental health problems are a crucial public health concern impacting individuals and communities globally. Mental health disorders are the top source of disability worldwide, affecting approximately 1 in 4 people at some point in their life. Mental health challenges come in various forms, including mood disorders such as depression and bipolar, anxiety disorders, and psychotic disorders like schizophrenia (WHO, 2020).

Mental health challenges can have serious consequences for individuals, including impaired functioning, decreased quality of life, and an increased risk of premature death (Kopelovich et al., 2020). They can also have significant impacts on communities and societies, as people with mental health challenges may be less able to participate in work, education, and other social activities. This is particularly concerning given the importance of mental health for overall wellbeing, as good mental health is essential for leading a fulfilling and productive life. Mental health issues can be caused by a variety of causes, including genetics, past experiences, environmental factors, and social factors (WHO, 2020). Given that some mental health illnesses have a hereditary component, genetics may have a role in the development of mental health problems. Events from life, such as trauma, abuse, or grief, can also contribute to mental health issues. A person's mental health may also be affected by social and environmental factors including poverty, prejudice, and a lack of social support (National Institute of Mental Health, 2021).

Individuals with mental health problems have a lower rate of employment compared to those without such issues, and even when employed, their wages tend to be lower (Levinson et al., 2010). Work has a positive effect on overall health and socio-economic outcomes for the general population (Barr et al., 2012). Work also has a positive correlation with economic, psychological, and clinical benefits for individuals with mental health issues (Cook et al., 2008). Several studies have also shown that work can result in reduced costs for mental health treatments in the short term (Schneider et al., 2019). Tracking the employment disparities based on mental health has become a crucial public health issue. Three significant national developments are expected to have impacted the labour force participation in the US: the significant number of individuals with mental illness enrolling for disability benefits (Burkhauser & Daly, 2011); the elevated unemployment rates resulting from the recent economic downturn (Bureau of Labour Statistics, 2013); and evidence-based psychosocial programs that support the employment aspirations of people with severe mental health conditions (such as schizophrenia) (Suijkerbuijk et al., 2017). With a noticeable increase in disability claimants who have mental health disorders, policymakers are paying closer attention to monitoring employment rates based on mental health status (Bureau of Labour Statistics, 2013).

According to the U.S. Department of Health and Human Services (2010), a disparity in mental health refers to a form of health inequality that is directly tied to social or economic disadvantage and affects groups of people who typically face greater social or economic barriers to health. The Centres for Disease Control and Prevention (CDC) defines this as a mental health disparity. Unemployment has been known to lead to poor mental health outcomes such as depression, anxiety, and suicidal thoughts (Kessler et al., 2010). The financial insecurity and difficulty in meeting basic needs that often result from job loss can further exacerbate the negative impact of unemployment on mental health (van Oorschot, 2013). These findings highlight the need for interventions that address both the economic and psychological consequences of unemployment.

Data science can play a role in understanding and addressing mental health challenges by analysing large datasets and developing predictive models that can inform the development of targeted interventions (Archenaa & Anita, 2015). By analysing data on factors such as genetics, life experiences, and social and environmental factors, data scientists can develop models that can help to identify individuals who are at risk of developing mental health challenges and to develop targeted interventions that can help to prevent or mitigate these challenges. Data scientists can use machine learning techniques to analyze large datasets and identify patterns that are predictive of mental health challenges (Shatte et al., 2019; Thieme et al., 2020). These patterns may include genetic markers, life experiences, or social and environmental factors that are associated with increased risk of mental health challenges (Thieme et al., 2020). By identifying these patterns, data scientists can develop predictive models that can help to identify individuals who are at risk of developing mental health challenges and to develop targeted interventions to support these individuals (Archenaa & Anita, 2015).

In addition to developing predictive models, data science can also be used to evaluate the effectiveness of interventions for mental health challenges (Thieme et al., 2020; Nakagawa et al., 2017). By analysing data on the outcomes of different interventions, data scientists can identify which interventions are most effective at improving mental health outcomes and can inform the development of evidence-based practices for addressing mental health challenges (Kotsiantis, 2011). For example, a study by Nakagawa et al. (2017) found that a combination of pharmacotherapy and cognitive behavioural therapy was effective in improving symptoms of depression in a sample of adults. This type of evidence can be used to inform the development of treatment guidelines and recommendations for individuals with mental health challenges (Shatte et al., 2019).

2.1.3 Relationship between Unemployment and Mental Health

Unemployment is a major public health concern that can have significant impacts on an individual's mental health (Xiong et al., 2020). Unemployment refers to the state of being without a job and can be caused by a range of factors, including economic recession, technological advancements, and changes in the job market (Mata et al., 2021). The loss of income, social support, and sense of purpose and identity that work can provide can have negative effects on an individual's mental health (Shatte et al., 2019). Studies have consistently found that unemployment is associated with increased rates of depression, anxiety, and other mental health problems. A review of studies conducted found that unemployment was associated with a two to three times greater risk of developing depression compared to those who were employed (Xiong et al., 2020). Similarly, a study by Mata et al., (2021) found that unemployment was associated with a higher risk of mental health problems in a sample of Norwegian adults.

According to research, people with serious mental health conditions such as schizophrenia, major depressive disorder, or bipolar disorder can experience a significant improvement in their employment rates when they receive evidence-based supported employment services, such as Individual Placement and Support (Modini et al., 2016). This is achieved by integrating vocational specialists into the mental health team and promptly finding job opportunities for patients (Becker et al., 2011). However, losing a job, whether voluntarily or involuntarily, has been shown to have a negative impact on mental health (Mayer & Hollederer., 2022). This can be confusing as losing a job often results in reduced income, but at the same time, it could potentially lead to less stress (de Vries et al., 2017). The loss of earnings may predominate, especially in cases of involuntary unemployment, which might result in poorer mental health consequences (Marshall et al., 2014). In either event, research has shown that people who see their jobs as giving them daily structure may suffer from a loss in mental health when they lose their jobs (Mokona et al., 2020). There is a large body of research exploring the link between mental health and employment. Many studies investigate the reasons for the generally observed negative impact of poor mental health on employment outcomes (Waghorn & Hielscher, 2014). In the workplace, mental health may affect the job-matching process, as someone who is depressed may not be able to enter the job market in the first place or perform well in an interview if they do (Tefft., 2011). Vornholt et al. (2017) found that negative symptoms, such as deficits in personality, and symptoms of depression are linked to lower employment rates among adults diagnosed with schizophrenia. Poor mental health can also lead to increased sick leave and reduced productivity both due to decreased performance and the need to manage symptoms at work. However, bipolar disorder may have a positive association with productivity as people with this illness are more likely to work in creative jobs and engage in creative activities (Tremblay et al., 2010). People with mental diseases have been found to benefit from employment, which lowers the expenses of having to treat mental illnesses (Davidson et al., 2012). Additionally, it has been connected to a rise in social support, self-esteem, motivation, and self-perception (Jin & Mosweu, 2017). Despite these advantages, estimates of the national unemployment rate for those getting public mental health care range from 50 to 80%. (National Alliance on Mental Illness., 2010). In response to this issue, there have been a number of employment programs developed for individuals with disabilities, including the quota system established by Italian law 68/1999 (Ministry of Labour and Social Policy, 1999). These programs have been shown to be effective in helping individuals with disabilities to secure employment in the non-competitive market (Nazarov et al., 2015).

The study conducted by Suijkerbuijk et al. (2017) aimed to determine if individuals with severe mental illnesses were more likely to lose their jobs during economic downturns, but no strong evidence was found to support this. Despite its importance, it is challenging to accurately determine the impact of mental health on employment. The onset of poor mental health is believed to decrease the probability of finding employment by 1.6%, which is about 10% of the employment gap. Recent advances in research methods have allowed for the use of fixed effects estimates, which are unbiased in the context of mental health and work, as the selection into mental health is mainly based on unchanging factors (Mark et al., 2022).

Further research is needed to distinguish the interplay between employment and mental health. While there is a connection between joblessness, under-employment, and poorer mental health, cross-sectional studies cannot establish the direction of this relationship (Hayes et al., 2008). For individuals who have been unemployed for a prolonged period, the negative impact on mental health may be exacerbated (van Rijn et al., 2016). Long-term unemployment refers to a prolonged period of unemployment, typically lasting six months or longer (Griep et al., 2015). The extended period of joblessness can lead to feelings of hopelessness and a loss of self-esteem, which can contribute to mental health problems (van Rijn et al., 2016). The relationship between unemployment and mental health is complex and can be influenced by a range of factors (Bartelink et al., 2019). For example, an individual's social support network can play a role in mitigating the negative effects of unemployment on mental health (Hughes et al., 2014). Those who have a strong support system of family and friends may be more able to handle with the challenges of unemployment and may have a lower risk of developing mental health problems (Pelzer et al., 2014).

In addition, an individual's coping strategies can also influence the relationship between unemployment and mental health (Schmitz, 2011). Those who engage in healthy coping strategies, such as seeking social support and maintaining a healthy lifestyle, may be more likely to cope with the stress of unemployment and may have a lower risk of developing mental health problems (Jain et al., 2021). Pre-existing mental health conditions may also influence the relationship between unemployment and mental health (Bartelink et al., 2019). Those who have pre-existing mental health conditions may be more vulnerable to the negative effects of unemployment and may be at an increased risk of experiencing mental health problems as a result of job loss (Hughes et al., 2014). It is crucial to grasp the potential negative impact of unemployment on mental health and to provide support to those who are unemployed (Pelzer et al., 2014). This may include interventions such as counselling, job training programs, and social support to help individuals cope with the challenges of unemployment and improve their mental (Schmitz, 2011).

2.1.4 Data Mining

Data mining is a key discipline in the field of big data analytics, and its applications are widespread across various industries. The fundamental purpose of data mining is to uncover valuable and hidden patterns and knowledge within large datasets. This iterative process involves examining a dataset, identifying features of interest, and discussing these with experts, before returning to the data with new insights (Ge et al., 2017). The resulting knowledge gained from a data mining session can take various forms, such as a set of rules, a tree, a graph or network, or one or more algorithms (SAS Institute, 1998).

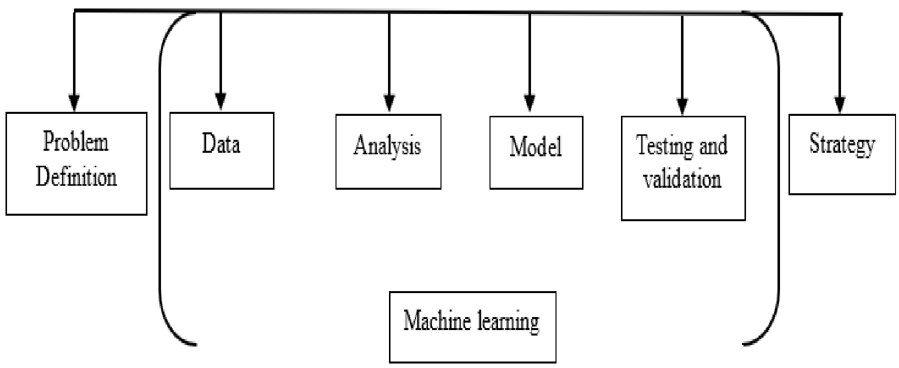
Data mining can be divided into two main methods: descriptive and predictive (Han et al., 2012). Descriptive data mining involves deriving meaningful patterns for decision making and outlining the properties of the data, while predictive data mining involves using multiple variables in a database to forecast the value of other unknown or future variables in order to predict future behaviour (Han et al., 2012). In general, data mining techniques are used for communication, prediction, analysis, and division of data (Zaki & Meira, 2014).

In the context of our study on predicting unemployment among individuals with mental health challenges, data mining techniques will be used to identify patterns and relationships in the data that may be relevant to predicting unemployment outcomes. This may involve both descriptive and predictive data mining methods, depending on the specific objectives and hypotheses of the study. By using data mining to uncover relevant patterns and knowledge, we aim to contribute to the development of interventions that can improve employment outcomes for individuals with mental health challenges, and to provide a better understanding of the factors that contribute to unemployment in this population.

The process of determining solutions to business problems by first defining the problem, collecting and analysing data to develop a model, testing and validating the model, and creating strategies based on the results is called data mining. The techniques, processes, and approaches used in collecting, analysing, testing, and validating data are collectively referred to as machine learning. The steps involved in data mining are shown in Figure 2.2.

Figure 2.2

Data mining processes (Nwaogu, 2019)



In our study on predicting unemployment among individuals with mental health challenges, we will utilize data mining and machine learning techniques to identify patterns and relationships in the data that may be relevant to predicting unemployment outcomes. This will involve collecting and analysing data on unemployment among this population, and using machine learning algorithms to map out the most appropriate model to predict unemployment. Once the model has been determined, it will be tested and validated using a validation set of data, and strategies will be drafted based on the results.

The use of data mining and machine learning in our study will allow us to identify factors that may be predictive of unemployment among individuals with mental health challenges and to develop targeted interventions to improve employment outcomes for this population. By defining the problem of unemployment in this population and applying a systematic approach to solving it, we hope to make a meaningful contribution to policy and practice in this area.

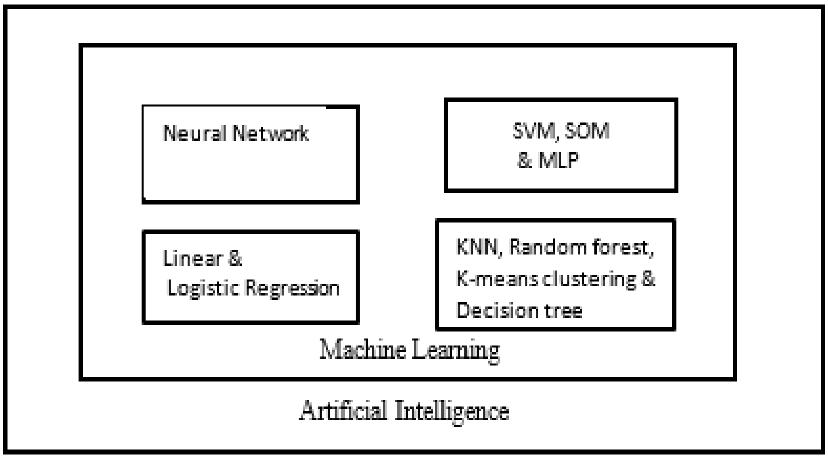
2.1.5 Machine Learning

In order for a system to continuously improve at a given activity, a subset of artificial intelligence called “machine learning” employs algorithms and statistical models (Peng et al., 2021). It entails using a lot of data and computing power to train models that can then act or predict depending on the input data (Pedregosa et al., 2011). As a result, data drives learning, with knowledge gained through the ability to make successful decisions based on the nature of the learning signal or input (Panesar., 2019). There are many distinct types of machine learning algorithms, including supervised learning, unsupervised learning, and reinforcement learning (Pedregosa et al., 2011). The system is trained on a labeled dataset in supervised learning, where the right result is delivered for each input. The program then employs this knowledge to forecast outcomes based on fresh, unforeseen inputs. Because it is not provided with labeled data when using unsupervised learning, the algorithm must independently identify patterns and relationships in the data (Priya et al., 2020). Using rewards and punishments, reinforcement learning teaches the algorithm to take actions that maximize a given reward (Pedregosa et al., 2011).

The capacity of machine learning to get better over time is one of its main benefits. The algorithm can learn from its errors and enhance its performance as it is exposed to more data and experiences. This allows it to make more accurate predictions or take more effective actions. There are several real-world uses for machine learning, such as speech and image detection, natural language processing, and predictive analytics. It is also utilized in a variety of sectors, including manufacturing, healthcare, and finance (Sharma et al., 2017).

Training models to generate predictions or take actions based on data is the process of machine learning, a subset of artificial intelligence (AI) (James et al., 2017). This is done using a variety of algorithms, which are essentially sets of instructions that tell the model how to process the data and make predictions (Priya et al., 2020). Machine learning is an iterative process that involves testing many different models in order to find the one that performs best on a given problem (Murphy, 2012). Obtaining desirable outcomes with this methodology frequently requires repeated experimentation and a substantial amount of data (Shalev-Shwartz et al., 2014). The aim of machine learning is to develop models that can learn from data and make precise forecasts or decisions without being explicitly programmed (James et al., 2017).

Figure 2.3.
Machine Learning as a subset of Artificial Intelligence (Nwaogu, 2019)



There are various techniques that can be used in machine learning, each with its own set of advantages and limitations. Here are some common machine learning techniques:

1. *Supervised learning:* This approach of machine learning uses labeled training data, which consists of a collection of input-output pairs, to train an algorithm to predict the outcome of hypothetical situations (Guo et al., 2016). Learning a function that can forecast the output value from an input value is the objective (Ribeiro et al., 2016). Various supervised learning techniques exist, including;

- Regression algorithms are a type of supervised learning algorithm that are used to predict a continuous output (such as a price or a probability). They work by finding the relationship between a set of input features and a continuous target variable, and then using this relationship to make predictions for new input data (James et al., 2017). There are several types of regression algorithms, including:

    I. *Linear regression:* This is a simple regression algorithm that assumes a linear relationship between the input features and the target variable. It works by finding the line of best fit through a set of data points and using this line to make predictions for new input data (Hastie et al., 2020).

II. *Polynomial regression*: This is a type of regression algorithm that can model relationships between the input features and the target variable that are more complex than a simple linear relationship. It works by fitting a polynomial function to the data and using this function to make predictions (Liang et al., 2021).

III. *Logistic regression*: This is a type of regression algorithm that is used to predict a binary output (such as a yes/no or 0/1). It works by using a logistic function to model the probability of an event occurring, and then making predictions based on this probability (Pedregosa et al., 2011).

IV. *Ridge regression*: This is a type of regression algorithm that is used to address the problem of overfitting in linear regression. It works by adding a penalty term to the objective function that penalizes the size of the model coefficients (Guo et al., 2016).

V. *Lasso regression:* This is a type of regression algorithm that is similar to ridge regression, but it uses a different penalty term that encourages the model coefficients to be exactly equal to zero (Hastie et al., 2020).

Regression algorithms are widely used in a variety of applications, including predictive modeling, risk assessment, and economics (Geron, 2017). They are relatively simple to implement (Ribeiro et al., 2016) and can provide valuable insights into the relationships between different variables (Hastie et al., 2020). However, they may not be suitable for modeling highly non-linear relationships (Kotsiantis, 2011) or for handling large amounts of categorical data (James et al., 2017).

- Classification algorithms are a type of supervised learning algorithm that are used to predict a discrete output (such as a class label or a binary output). They work by learning a decision boundary from labeled training data and then using this boundary to make predictions for new input data (Alpaydin., 2010). There are several types of classification algorithms, including:

I. *Logistic Regression:* This is a type of machine learning algorithm that is commonly used for binary classification tasks, where the goal is to predict a value that can only take on one of two possible values (Murphy, 2012). This makes it well-suited for tasks where the outcome is binary, such as predicting whether a customer will churn (leave the company) or not, or whether a patient has a certain disease or not (Geron, 2017). One of the key advantages of logistic regression is that it is a simple and interpretable model (Ribeiro et al., 2016). This means that the model can be easily understood and explained to non-technical stakeholders, and the coefficients of the model can be used to determine the importance of each input variable in predicting the outcome (Archenaa & Anita, 2015). Another advantage of logistic regression is that it is a fast and efficient algorithm. It is less computationally intensive than other machine learning algorithms, such as support vector machines or neural networks, and therefore can be trained on large datasets relatively quickly (James et al., 2017). Logistic regression can also be extended to handle multi-class classification tasks, where the goal is to predict a value that can take on one of multiple possible values (Hastie et al., 2020). In these cases, the algorithm will learn multiple binary classifiers and combine them to make predictions for each possible value. This allows the model to handle more complex classification tasks, where the outcome is not necessarily binary. One potential limitation of logistic regression is that it assumes that the input data is linearly separable. This means that the model assumes that a straight line can be drawn to separate the different classes in the data (Alpaydin., 2010). If this assumption is not met, then the model may not be able to accurately predict the outcome. In these cases, it may be necessary to use a different algorithm, such as a support vector machine or a neural network, that can handle non-linear data (Murphy, 2012). Overall, logistic regression is a useful and widely-applicable machine learning algorithm for binary and multi-class classification tasks. It is a simple and interpretable model that can be trained quickly on large datasets, and can provide insights into the relative importance of each input variable in predicting the outcome. (Geron, 2017).

II. *Support Vector Machines (SVMs):* The supervised learning model known as Support Vector Machines (SVMs) may be applied to both classification and regression applications (Wang & Gu, 2021). Generally speaking, SVMs are most effective at classifying complex but small- to medium-sized datasets (Chang & Lin, 2011). One of the key advantages of using an SVM is that it can efficiently handle high-dimensional data, such as images, text data, and other types of data that have many features (Scholkopf & Smola, 2002). This is because SVMs use a technique called the kernel trick, which allows them to transform the data into a higher-dimensional space in a way that makes it easier to find a hyperplane that can separate the different classes in the data (Hofmann et al., 2008). Additionally, SVMs are very effective when the data has a clear separation between the different classes, which means that the classes can be easily distinguished from each other. This is because SVMs try to find the hyperplane that maximally separates the different classes, which can result in very good classification performance (Westreich et al., 2010).

III. *Decision trees*: Decision trees are a machine learning algorithm that can perform both classification and regression (Kotsiantis, 2011). They build a tree-like model of decisions based on the features of the data (Gourisaria et al., 2020). The algorithm divides the data at each internal node based on a decision rule, with each branch resulting in a different outcome (Cervantes et al., 2020). The leaves of the tree show the final classification or prediction. Decision trees are widely used due to their simplicity, ability to handle both continuous and categorical data (Shamrat et al., 2021), speed in training and prediction, and handling of large amounts of data. However, they can overfit, especially if the tree is too deep (Gourisaria et al., 2020). To prevent overfitting, it is common to prune the tree by removing unimportant nodes (Kotsiantis, 2011). Another disadvantage is that decision trees can be sensitive to small changes in the data, and they may not be as accurate as some other machine learning algorithms in certain situations (Cervantes et al., 2020).

IV. *Random forests*: These are classification algorithms that work by training multiple decision trees on different subsets of the training data and then aggregating the predictions made by each tree (Biau et al., 2016). This can improve the generalization performance of the model and reduce the risk of overfitting. Random forests are a type of ensemble learning method, which means that they are composed of multiple decision trees that work together to make predictions (Gregorutti et al., 2016). Ensemble methods are often more accurate than individual decision trees, because they can average out the biases of the individual trees and reduce overfitting (Żbikowski et al., 2021). One of the key advantages of using a random forest is that it can handle high-dimensional data, such as images, text data, and other types of data that have many features (Li et al., 2020). This is because random forests use a technique called bagging, which trains each decision tree on a different subset of the data (Biau et al., 2016). This can help to reduce overfitting and improve the performance of the model (Żbikowski et al., 2021). Another advantage of random forests is that they can handle missing values in the data, which is a common problem in many real-world datasets (Li et al., 2020). Random forests can handle missing values by using a technique called imputation, which estimates the missing values using the other available data (Shaik et al., 2019). In summary, random forests can be a good choice for regression and classification tasks when you have a large dataset with many features and some missing values (Gregorutti et al., 2016). They are particularly well-suited for handling high-dimensional data, and can provide good performance even with relatively little tuning of the model parameters (Ziegler et al., 2014).

V. *Neural networks:* Classification algorithms inspired by the human brain are called neural networks. A neural network is a machine learning algorithm modeled after the structure and function of the brain (Kriegeskorte et al., 2019). It consists of interconnected "neurons" that process and transmit information. Neural networks can learn and adapt from the data they receive and can be applied to various tasks, such as classification, regression, and clustering (Kriegeskorte et al., 2019). There are various types of neural networks, including feedforward, convolutional, and recurrent neural networks (Canziani et al., 2016). In a feedforward neural network, information flows in one direction from the input layer to the output layer. Convolutional neural networks, a type of feedforward network, are designed to process data with a grid-like structure, such as images (Bergstra et al., 2015). Recurrent neural networks allow information to circulate in loops, enabling them to handle sequential data like time series or natural language. Although neural networks are flexible and powerful, they can be computationally demanding to train and may require a large amount of data to work effectively (Canziani et al., 2016). They can also be difficult to interpret, as it can be hard to understand how the network is making its predictions (Kriegeskorte et al., 2019).

Classification algorithms are widely used in a variety of applications, including natural language processing, computer vision, and predictive modeling (Gregorutti et al., 2016). They are particularly effective for tasks that involve making decisions based on input data, and they have been successfully applied to tasks such as spam filtering, image classification, and credit fraud detection (Bergstra et al., 2015). However, they may not be suitable for tasks that require a lot of interpretation or for modeling complex relationships between input and output data (Westreich et al., 2010).

- Sequence labeling algorithms*:* These algorithms are a type of supervised learning algorithm that are used to predict a label for each element in a sequence of input data. They are commonly used for tasks such as part-of-speech tagging, named entity recognition, and language modeling (Sutskever et al., 2014). There are several types of sequence labeling algorithms, including:

  I. *Hidden Markov models (HMMs):* These are probabilistic graphical models that are used to model sequences of observations. They work by representing the probability distribution over a sequence of hidden states, and then using this distribution to predict the most likely sequence of labels for a given input sequence (Mor et al., 2021).

II. *Conditional random fields (CRFs):* These are probabilistic graphical models that are similar to HMMs, but they model the dependencies between the labels of adjacent elements in the sequence rather than the dependencies between the hidden states (Song et al., 2015).

III. *Recurrent neural networks (RNNs):* These are neural network architectures that are designed to process sequential data. They work by using feedback connections to allow information to be passed from one element in the sequence to the next (Graves, 2013).

IV. *Long short-term memory (LSTM) networks:* These are a variant of RNNs that are particularly effective at modeling long-term dependencies in sequential data. They work by using gating mechanisms to control the flow of information through the network (Graves, 2012).

Sequence labeling algorithms are widely used in a variety of applications, including natural language processing and speech recognition. They are particularly effective for tasks that involve modeling the dependencies between elements in a sequence, and they have been successfully applied to tasks such as part-of-speech tagging, named entity recognition, and language modeling (Sutskever et al., 2014). However, they may not be suitable for tasks that require a lot of interpretabilities or for modeling complex relationships between input and output data.

One of the key advantages of supervised learning is that it can learn complex relationships between input and output data. However, it requires a large amount of labeled data in order to be effective, and the quality of the predictions made by the model is directly related to the quality of the labeled data used for training (Miyato et al., 2018). In addition, supervised learning algorithms may not be able to adapt to changes in the underlying data distribution, and they may perform poorly on tasks that require a lot of generalization.

Steps involved in a supervised learning algorithm: Supervised learning algorithms are trained using labelled data, which consists of a set of input data and the corresponding correct output labels. The goal of the training process is to learn a function that maps the input data to the correct output labels (Mahesh, 2020). There are several steps involved in a supervised learning algorithm and they include;

I. *Collect and prepare the data:* The first step is to collect a sufficient amount of labeled data to train the algorithm (Mahesh., 2020). The data should be cleaned and prepared for input into the model.
II. *Choose an appropriate model and learning algorithm:* The next step is to choose a model and learning algorithm that is suitable for the task at hand. This will depend on the type of data and the nature of the problem being solved (Miyato et al., 2018).
III. *Split the data into training and validation sets:* The labeled data is frequently divided into a training set and a validation set. The validation set is used to assess the model's performance during training, while the training set is used to fit the model.
IV. *Train the model:* In order to reduce the difference between the anticipated output and the accurate output labels, the model's parameters are adjusted when input data from the training set is fed into it. Usually, an optimization algorithm like gradient descent is used for this procedure (Adi et al., 2019).
V. *Evaluate and fine-tune the model:* After the model has been trained, its effectiveness is assessed using data from the validation set. If the model's performance is unsatisfactory, its hyperparameters can be changed or more training data can be gathered to improve the model.
VI. *Test the model:* To assess how well the model performs with unobserved data, it is lastly evaluated on a different test set. This is the final step before deploying the model in a real-world setting.

Supervised learning is widely used in a variety of applications, including natural language processing, computer vision, and predictive modeling (Mahesh., 2020). It has been successfully applied to tasks such as speech recognition, image classification, and predicting stock prices. In general, supervised learning is a powerful tool for solving many types of problems and has the potential to improve the efficiency and accuracy of decision-making in a wide range of industries (Gourisaria et al., 2020).

2. Unsupervised learning: A model is trained with unlabeled data in this machine learning approach (Adi et al., 2019). The model is then asked to explore the data for patterns and correlations without being instructed on what to look for. Numerous unsupervised learning methods exist, including:

- Clustering algorithms: These algorithms are a type of unsupervised machine learning algorithm that divide a dataset into groups, or clusters, of similar items (Kuchaki et al., 2012). These algorithms are used to discover patterns and relationships within data that may not be immediately obvious. There are many different types of cluster algorithms, including:

  I. *K-means clustering*: K-means clustering is a well-known technique that splits a dataset into a predetermined number of groups (k) by minimizing the distance between points inside each cluster. According to the article "K-Means clustering for data mining" by Hossain et al. (2019), K-means is one of the most widely used clustering algorithms and it is simple and easy to understand.

  II. *Hierarchical clustering:* Hierarchical clustering creates a tree-like structure (called a dendrogram) that shows how the data can be divided into clusters (Kuchaki et al., 2012). Hierarchical clustering may be divided into two primary categories: agglomerative, which begins with individual points and groups them into clusters, and divisive, which begins with the complete dataset and divides it into smaller clusters. According to Guo et al. (2020), hierarchical clustering is a powerful technique for finding patterns in data, especially when the number of clusters is not known in advance.

  III. *Density-Based Spatial Clustering of Applications with Noise (DBSCAN):* DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is an algorithm that is designed to identify clusters of high density, and can also identify outliers (points that do not belong to any cluster) (Kuchaki et al., 2012). According to Campello et al. (2015), DBSCAN is a density-based algorithm that is particularly well-suited for identifying clusters of arbitrary shape.

IV. *Gaussian mixture model (GMM):* Gaussian mixture model (GMM) is an algorithm that assumes that the data is generated from a mixture of several different Gaussian distributions (Rauber et al., 2016). It can be used to identify complex, non-linearly separated clusters. According to Jordan and Mitchell (2015), GMM algorithm is a powerful tool for clustering and density estimation that is particularly well-suited for identifying complex, non-linearly separated clusters.

Cluster algorithms are commonly used in a variety of applications, including marketing, customer segmentation, image analysis, and fraud detection

- Dimensionality reduction algorithms: These methods are an example of a machine learning method that reduces the number of dimensions of a high-dimensional dataset while preserving as much information as feasible (Rauber et al., 2016). Dimensionality reduction aims to reduce the number of dimensions without sacrificing valuable information or patterns. Different dimensionality reduction methods exist, including:

  I. *Principal component analysis (PCA):* By maximizing the variance along the new axes, this linear dimensionality reduction strategy places the data in a lower-dimensional space (Kuchaki et al., 2012).

  II. *t-distributed stochastic neighbor embedding (t-SNE*): With the help of this non-linear dimensionality reduction method, high-dimensional data may be mapped to a lower-dimensional space while still maintaining the local structure of the data (Adi et al., 2019).

  III. *Autoencoders*: These dimensionality reduction methods employ neural networks to learn how to compress input data into a lower-dimensional representation, which is followed by the reconstruction of the original data from the representation (Cunningham & Yu., 2014).

  IV. *Linear discriminant analysis (LDA):* This supervised dimensionality reduction method maximizes class separability while projecting the data into a lower-dimensional environment (Goldstein & Uchida, 2016).

Dimensionality reduction algorithms are often used as a pre-processing step in machine learning pipelines to reduce the complexity of the data and improve the performance of predictive models (Cunningham & Yu, 2014). Since it might be challenging to show data in more than three dimensions, they are also helpful for high-dimensional data visualization (Goldstein & Uchida., 2016).

- Anomaly detection algorithms: These algorithms are a type of machine learning algorithm that are used to identify unusual or unexpected patterns in a dataset. These algorithms are used to identify data points that are significantly different from the rest of the data and may indicate an issue or problem that needs to be addressed (Goldstein & Uchida., 2016). There are several types of anomaly detection algorithms, including:
  I. *Statistical algorithms:* These algorithms use statistical techniques, such as mean and standard deviation, to identify anomalies by identifying data points that are significantly different from the rest of the data (Danks & London, 2017).
  II. *Density-based algorithms:* These algorithms identify anomalies by identifying data points that are in low-density regions of the data (Reddy & Bindu, 2017).
  III. *Distance-based algorithms:* These algorithms identify anomalies by identifying data points that are significantly different from their neighbors (Van de Velden et al., 2019).
  IV. *Supervised algorithms:* These algorithms are trained on a dataset with labeled anomalies and use this training to identify anomalies in new data (Sathe & Adamuthe, 2021).

Anomaly detection algorithms are commonly used in a variety of applications, including fraud detection, intrusion detection, and quality control. In conclusion, unsupervised learning has a number of potential applications, including data compression, data visualization, and density estimation. It is particularly useful for tasks where labeled data is scarce or difficult to obtain. However, unsupervised learning can be challenging because the model has no guidance on what to look for in the data, and the results may be difficult to interpret. In addition, unsupervised learning algorithms may be sensitive to the presence of noise or outliers in the data.

Table 2.1.

Supervised versus unsupervised learning (Sathe & Adamuthe, 2021)

Characteristics	Supervised Learning	Unsupervised learning
Methods	There are provided input and output variables.	Only input data is given
Goals	To identify the precise functions that will be used to predict output when a new dataset is provided	To model the hidden patterns in a dataset in other to learn about the data
Class	Machine learning problems, data mining problems and neural network	Machine learning, data mining problems and neural networks
Examples	Classification Regression Linear regression Support vector machine	Cluster Association k-means Association
Use	Export systems in image recognition, speech recognition forecasting, financial analysis and training neural networks and decision tree	Pre-processing of data, during exploratory analysis as well as pre-training supervised learning algorithms

3. Semi-supervised learning: Semi-supervised learning techniques are frequently employed when labeling a big dataset requires a lot of money or time, or when the proportion of instances that have been labeled to the total number of examples in the dataset is low (Van & Hoos, 2020). By leveraging the additional unlabeled data to understand the data's underlying structure, these techniques may also be utilized to enhance the performance of supervised learning algorithms (Enguehard et al., 2019). There are several types of semi-supervised learning algorithms, including:

• *Graph-based algorithms*: These algorithms use the relationships between data points, as represented by a graph, to propagate labels from the labelled examples to the unlabeled examples. An example of this type of algorithm is label Propagation (de Sousa, 2015).

• *Generative models:* These algorithms use a generative model to estimate the distribution of the data and use this estimate to label the unlabelled examples. An example of this type of algorithm is the Generative Adversarial Networks (GANs) (Goodfellow et al., 2014)

• *Self-training*: This is a simple semi-supervised learning approach that involves training a supervised learning algorithm on the labelled examples and then using this algorithm to label the unlabelled examples. The labelled examples and the predicted labels are then used to train a new model.

Semi-supervised learning algorithms are commonly used in a variety of applications, including natural language processing and image classification. In natural language processing, they have been used for tasks such as text classification (Krawczyk., 2016), sentiment analysis and named entity recognition (Venkateswarlu et al., 2018). In image classification, they have been used for tasks such as object recognition (Chapelle et al., 2006) and image segmentation (Enguehard et al., 2019).

4. Reinforcement learning: A kind of machine learning called reinforcement learning teaches an agent to interact with its surroundings in a way that maximizes a reward (François-Lavet et al., 2018). In reinforcement learning, the objective is to discover a policy that maximizes the predicted reward over time. The agent receives a reward signal for every action it performs (Stadie et al., 2015). According to François-Lavet et al. (2018), The following phases make up the loop that defines reinforcement learning:

- The agent monitors the condition of the environment.
- Based on its policy, the agent decides which action to take.
- The action is carried out by the agent, who is rewarded by the environment.
- In light of the reward it got, the agent changes its policy.

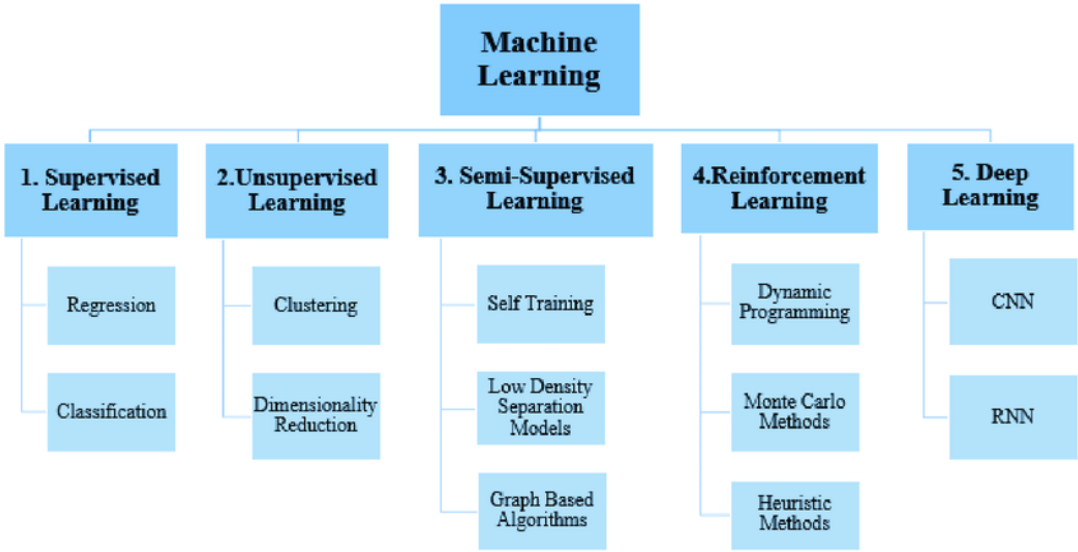
Reinforcement learning algorithms can be classified into two main categories: value-based algorithms and policy-based algorithms (Saravanan & Sujatha, 2018). Value-based algorithms learn a function that estimates the expected reward for each state or action, an example of this is Q-learning (Wang et al., 2018). while policy-based algorithms directly learn a policy that maps states to actions, an example of this is REINFORCE algorithm (Kelleher., 2015). Reinforcement learning has been applied to a variety of problems, including control problems, games, and natural language processing. It has also been used to train autonomous agents, such as robots and self-driving cars (Saravanan & Sujatha, 2018).

5. Deep learning: Artificial neural networks are used in deep learning, a sort of machine learning, to discover intricate patterns and connections in data (LeCun et al., 2015). These neural networks are made up of numerous layers of linked nodes, or "neurons," and are modeled after the composition and operation of the human brain (Stadie et al., 2015). Deep learning algorithms are trained by processing large datasets and adjusting the neural network's weights and biases through backpropagation in order to minimize a loss function (Saravanan & Sujatha, 2018). This is usually achieved using optimization algorithms such as stochastic gradient descent (Qiu et al., 2016). Unlike traditional machine learning algorithms, deep learning algorithms have the advantage of automatically learning and extracting useful features from data, allowing them to perform well on complex and large datasets such as image and text data (Stadie et al., 2015). This has led to state-of-the-art results in tasks such as image classification (Krizhevsky et al., 2017), natural language processing (Collobert et al., 2011), and speech recognition (Hinton et al., 2012).

One of the key challenges of deep learning is the need for large amounts of labelled data to train the model (Kelleher., 2015). This can be a significant barrier to entry for organizations that do not have access to large datasets or the resources to label them (Saravanan & Sujatha, 2018). Additionally, deep learning algorithms can be computationally intensive, requiring specialized hardware such as graphics processing units (GPUs) to train and run them efficiently (Qiu et al., 2016). Despite these challenges, deep learning has the potential to revolutionize many industries and has already had a significant impact on a wide range of applications (Goodfellow et al., 2016). As the amount of data available continues to grow and the hardware needed to train and run deep learning algorithms becomes more accessible, it is likely that deep learning will continue to play a major role in the development of new machine learning applications (Schmidhuber, 2015).

Figure 2.4

Machine Learning Techniques (Stadie et al., 2015)



2.1.6 Label Encoding

Label encoding is a technique for converting categorical variables, which have a limited number of potential values, into numerical values that machine learning algorithms may use (Saravanan & Sujatha, 2018). This is necessary because most machine learning algorithms expect numerical input data. To perform label encoding, each unique category in a categorical variable is assigned a unique integer. For example, if a categorical variable has three categories: “red”, “green”, and “blue” they could be encoded as 0, 1, and 2, respectively (Saravanan & Sujatha, 2018).

Label encoding is a simple and efficient method for encoding categorical variables, but it has some limitations (Witten et al., 2016). One limitation is that it implies an ordinal relationship between the categories, which may not be appropriate if the categories do not have an inherent order. For example, in the example above, the encoding suggests that "red" is less than "green" and "green" is less than "blue," which may not be the case. In this situation, it may be better to use one-hot encoding, which creates a new binary column for each category.

It is important to note that label encoding should only be used for categorical variables that do not have any meaningful numerical relationship (Witten et al., 2016). For example, it would not be appropriate to use label encoding for a categorical variable representing the size of a T-shirt (small, medium, large), as there is a clear numerical relationship between the categories. In this case, it would be better to use the actual numerical values (e.g., small=1, medium=2, large=3) as the input to the model.

2.1.7 Data Standardization

Data standardization is a process that is used to transform variables in a dataset so that they have a mean of zero and a standard deviation of one. Standardization is often used as a pre-processing step for machine learning algorithms, as it can help to ensure that the different variables in the dataset are on the same scale, which can improve the performance of some algorithms (Witten et al., 2016). To standardize a variable, you first need to calculate the mean and standard deviation of the variable. You can then subtract the mean from each value in the variable and divide the result by the standard deviation (Saravanan & Sujatha, 2018). This will transform the variable so that it has a mean of zero and a standard deviation of one. Standardization is typically applied to continuous variables, such as those that represent measurements or quantities. It is not usually applied to binary or categorical variables, as these types of variables do not have a meaningful mean or standard deviation.

Standardization can be useful in situations where the variables in the dataset have different scales and units, as it can help to remove the effects of these differences (Witten et al., 2016). However, it is important to note that standardization should only be applied to variables that are approximately normally distributed. If the variables are not normally distributed, standardization may not be appropriate.

2.1.8 One Hot Encoder

One-hot encoding is a method of representing categorical variables as numerical data. It creates a binary column for each unique category in a categorical variable, with the value 1 indicating the presence of the category in a particular observation and 0 indicating its absence. This representation is useful for feeding categorical data into machine learning models that expect numerical inputs (Witten et al., 2016).

2.2 Related Research

Mental health challenges have been shown to be a risk factor for unemployment, with studies finding that individuals with mental illness are more likely to be unemployed or have lower rates of employment compared to the general population (Harvey et al., 2017; Vornholt et al., 2017). In order to help us acquire information that will aid in tackling the issue, other researchers have previously used a number of methodologies to analyze and draw relevant conclusions about the association between unemployment and mental health. In order to gain a handle on our work, and to help us choose the algorithm we will use for our machine learning prediction, related research is being undertaken in order to describe some of their deductions.

One study that examined the relationship between mental health and employment outcomes used logistic regression to predict unemployment in a sample of individuals with severe mental illness (SMI) (Suijkerbuijk et al., 2017). The study found that factors such as lower levels of education, being male, and having a longer duration of SMI were significantly associated with unemployment. Another study used decision tree models to predict employment status among individuals with SMI and found that factors such as being married, having a higher level of education, and receiving social security benefits were significantly associated with employment (Kaufman et al., 2016).

A systematic review of employment interventions for individuals with mental illness found that vocational rehabilitation programs, which often involve job coaching and training, were effective in improving employment outcomes (Joyce et al., 2015). In addition to vocational rehabilitation programs, other employment interventions that have shown promise for individuals with mental health challenges include supported employment programs, which provide ongoing support to individuals as they transition into the workforce, and individual placement and support programs, which involve matching individuals with appropriate job opportunities and providing ongoing support (Marshall et al., 2014).

In 2018 research by Yedida et al. (2018), a prediction system employing the k-Nearest Neighbours algorithm is used to provide predictions on whether or not an employee would quit a firm. The number of years a person has worked for a firm, the average amount of hours worked each month, and other factors are taken into consideration while conducting the analysis. To obtain the prediction, well-known classifiers including Naive Bayes for each of the new samples, they applied the Bayes' rule to calculate P(Y|X), the necessary output, by first estimating the values of P(X|Y) and P(Y) in order to make the algorithm, which assumes that the values associated with each class are distributed according to a Gaussian distribution. Then they used Logistic Regression utilizing methods of regularization to avoid overfitting. In their work, an L2 regularized model was employed. They also used Multi-layer Perceptron (MLP), a learnt non-linear transformation is used to transform the input data into a space where it is linearly separable in an artificial neural network (ANN) model, which has many layers of nodes, each completely linked to the next and employs backpropagation for training and lastly, they used k- Nearest Neighbours (K-NN) where Various distance metrics, including the Euclidean distance, Manhattan distance (which was utilized in their study), Minkowski distance, etc., may be used to determine the distance from neighbours. A comparison analysis shows that k-Nearest Neighbors performs significantly better than the other classifiers discussed.

A different study also in 2022 by Porkodi et al. (2022), deployed a few other classifiers to the human resource data to reduce work - life conflict, incorporating Decision Tree, Logistic Regression, Support Vector Machines (SVM), KNN, Random Forest, and Naive Bayes approaches. After applying feature selection techniques to the dataset, analysis produced findings that will aid in preventing employee occupational stress. Logistic regression technique was also used to propose a mechanism to prevent job stress in Srivastava and Agarwal (2020) study. A group of high-risk employees was identified by aggregating statistical information and employees. For the purpose of lowering mental stress, these personnel are given special care.

Another research by Chekroud et al. (2016), used a variety of data mining approaches to predict fatigue. These approaches including Random Forest, SVM, Gradient Boosted Classifier, and Logistic Regression. Extreme Gradient Boosting appears to outperform classifiers in the experimental analysis presented in this research for attrition prediction tasks.

Machine learning approaches were also employed by researchers to identify mental issue (Garcia-Ceja et al., 2018). Periods of rest and stress were classified into two categories. SVM and Bayesian networks both offered respectable accuracy, coming in at 82.7% and 84.6%, respectively. Another Systematic Review conducted by Zulfiker et al. (2021), conducted a systematic review of studies that applied machine learning approaches to predict unemployment and mental health outcomes.

They found that decision tree, neural network, and support vector machine algorithms were commonly used.

The Random Forest algorithm, 10-fold cross-validation, and oversampling were used by Gonçalves et al. (2020) in their research on the prediction of mental disease linked to unemployment. The best model had an accuracy of 94.24%, sensitivity of 98.78%, and specificity of 89.66%, and their findings demonstrated effective forecast with accuracy, sensitivity, and specificity values above 90%.

Using data from the World Health Organization, Abbas and Raza (2021) performed comparative research to forecast unemployment and mental health results using a variety of machine learning algorithms, including decision tree, neural network, and support vector machine.

While these interventions have been found to be effective in improving employment outcomes for individuals with mental health challenges (Waghorn & Hielscher, 2014; Peiró, 2012), there is still a need for research on predicting unemployment specifically among this population. In this study, we aim to use machine learning techniques to predict unemployment among individuals with mental health challenges. In conclusion, the use of machine learning techniques for predicting unemployment among individuals with mental health challenges is a promising area of research that has the potential to inform the development of effective interventions to improve employment outcomes for this population (Jain et al., 2021).

Further research is needed to fully understand the factors that contribute to unemployment among individuals with mental health challenges and to identify the most effective machine learning algorithms for predicting unemployment in this population (de Vries et al., 2017).

We will utilize exploratory data analysis to identify relevant predictors of unemployment and to understand the relationship between these predictors and unemployment outcomes. We will also examine the effectiveness of different machine learning algorithms in predicting unemployment among individuals with mental health challenges. By identifying predictors of unemployment and examining the effectiveness of different machine learning algorithms in predicting unemployment, this study aims to contribute to the development of interventions that can improve employment outcomes for individuals with mental health challenges.

Additionally, this study aims to provide a better understanding of the factors that contribute to unemployment among individuals with mental health challenges, which can inform policy and practice in this area.

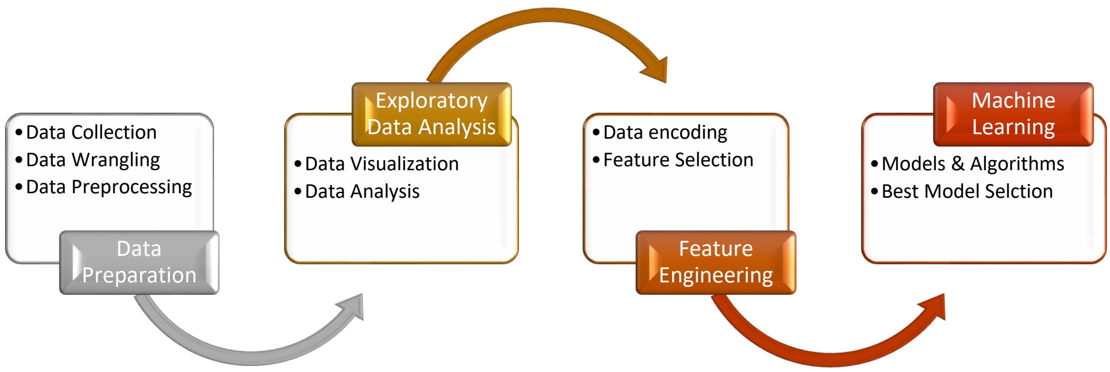
CHAPTER 3
METHODOLOGY

The methodology chapter of this study outlines the research design and procedures used for the study. The chapter describes the dataset used for the study and the process for accessing it, including data access permissions. The chapter also discusses the data analysis process, including data pre-processing techniques, which help to clean and prepare the data for analysis. Furthermore, the chapter outlines the research procedures, including the steps taken to analyze the data, generate models, and validate the results. In addition, the methodology chapter specifies the hardware and software specifications used for the study, which include the tools and platforms used for data analysis and modeling. Overall, the methodology chapter provides a clear and comprehensive overview of the research design and procedures, which enables the study to be replicable and valid.

3.1 Research Design

The research design for this study will be a mixed-methods approach, combining both quantitative and qualitative data analysis methods. The main focus of the study will be the exploratory data analysis and prediction of unemployment among individuals with mental health challenges. The study will utilize both primary and secondary data sources. Primary data will be collected from Kaggle dataset online where paid surveys administered to a sample of individuals with mental health challenges who are currently unemployed or have experienced unemployment in the past has been collected and made available for research purpose. The survey included questions about demographic characteristics, mental health status, employment history, and other relevant factors that may impact unemployment. Secondary data will be collected from publicly available sources, including government statistics on unemployment and mental health, as well as relevant literature on the topic. The quantitative data we utilized will be analyzed using descriptive statistics and machine learning techniques. Exploratory data analysis will be conducted to identify patterns and trends in the data, and to identify key predictors of unemployment among individuals with mental health challenges. Based on the results of the exploratory data analysis, a prediction model will be developed using machine learning algorithms to predict the likelihood of unemployment among individuals with mental health challenges. The findings of this study will contribute to a better understanding of the factors that contribute to unemployment among individuals with mental health challenges, and will provide insights into the potential for using machine learning techniques to predict and prevent unemployment in this population. The results of the study will have implications for policymakers, mental health professionals, and individuals with mental health challenges, and may inform the development of interventions and support systems to address unemployment among this population.

Figure 3.1.
Research design



3.2 Dataset

The dataset consists of 334 rows, with each row representing a collection of 31 features. Tables 1.1 and 1.2 list the dataset's properties as well as their descriptions. Table 1.2 displays all of the binary features in the dataset, whereas Table 1.1 displays all of the non-binary attributes. ‘Unemployed’ is regarded the target variable for classification among numerous qualities. Pre-processing procedures are used after the dataset has been gathered. For this, a multi-step pre-processing procedure was used acquiring a cleaned dataset All 'nan' values in the dataset are replaced with zero. The attribute 'Education' contains numerous values that all mean the same thing. For example, the attribute values completed undergraduate and some undergraduate both correspond to a Bachelor's degree. As a result, they are replaced. Following the application of pre-processing techniques, training and testing datasets are obtained by bifurcating the pre-processed dataset in the ratio of 67:33. Because the attribute Unemployed' is the dependent variable in this classification, it is not included in the test dataset. The classifier is taught using the training dataset, and the testing dataset is then tested.

3.2.1 Using an Existing Dataset

When using a ready-made dataset in an article, it is important to properly credit the source of the dataset. This can typically be done by providing a citation that includes the name of the dataset, the creator or publisher of the dataset, and the date that the dataset was accessed. Additionally, it is important to provide any relevant information about the dataset, such as its size, scope, and limitations. In some cases, the dataset may have specific terms of use or licensing agreements that must be followed. If this is the case, it is important to include this information in the article and ensure that the dataset is used in accordance with these terms.

Finally, it is important to consider the ethical implications of using a ready-made dataset. This includes ensuring that the data is accurate, relevant, and not being used in a way that could be harmful to individuals or groups. By considering these factors, it is possible to use a ready-made dataset in a responsible and ethical manner in an article.

3.2.2 Data Access Statement

This is a funded research study by Corley (2019) looking at the relationship between unemployment and mental illness. Numerous studies by NAMI have shown the high unemployment rate among those with mental illnesses, but this is the only study to date to focus on causality (why they are unemployed). The variance's statistical significance has already been established by earlier, bigger samples.

Figure 3.2.
Data access statement

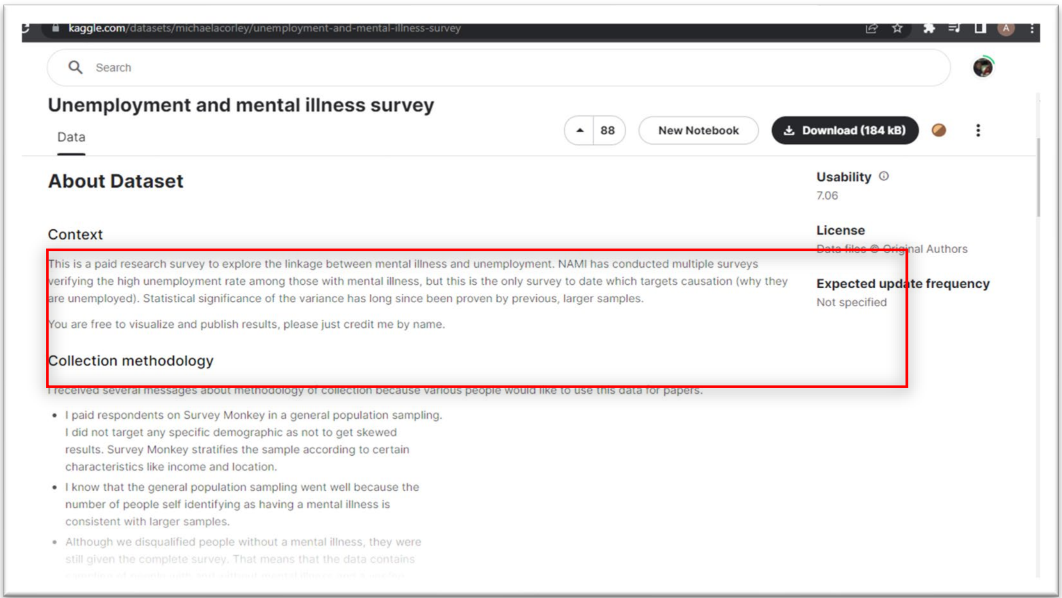


Table 3.1.
Summary of categorical and numerical attributes present in collected dataset

Attribute Name	Description	Values Present
How many days were you hospitalized for your mental Illness	Number of days of Hospitalization	0-100 and contains nan values
Education	Pursued Education	High School or GED, Some Phd, Completed Undergraduate, Some Undergraduate, Some\xa0Masters, Completed Masters, Completed Phd, Some Highschool
Total length of any gaps in my resume in months.	Length of gaps present in Resume	0-100
Annual income (including any social welfare programs)	Annual Income of individual in USD	0-100
How many times were you hospitalized for your mental illness	Number of times of hospitalization due to mental illness	0-100
Age	Age of respondent	'30-44', '18-29', '45-60', '> 60'
Household Income	House hold income of the person in USD	'$25,000-$49,999', '$50,000-$74,999', '$150,000-$174,999', '$0-$9,999', '$100,000-$124,999', '$125,000-$149,999', 'Prefer not to answer', '$10,000-$24,999', '$75,000-$99,999', '$200,000+', '$175,000-$199,999'
Region	Belongs to which region	'Mountain', 'East South Central', 'Pacific', 'New England', 'East North Central', 'South Atlantic', 'Middle Atlantic', 'West South Central', 'West North Central', nan
Device Type	Use of device by the person	'Android Phone / Tablet', 'MacOS Desktop / Laptop', 'Windows Desktop / Laptop', 'Ios Phone / Tablet', 'Other'

Table 3.2.

Summary of binary attributes present in collected dataset

Attribute Name	Description	Values Present
I am currently employed at least part-time	Whether an employee was engaged as part-timer or not	0-No 1-Yes
I identify as having a mental illness	Whether the individual has mental illness or not	0-No 1-Yes
I have my own computer separate from a smart phone	Whether the person occupies computer other than smart phone or not	0-No 1-Yes
I have been hospitalized before for my mental illness	Whether the person has hospitalization record	0-No 1-Yes
I am legally disabled	Whether the person has legal issues or not	0-No 1-Yes
I have my regular access to the internet	Whether the person has access to internet or not	0-No 1-Yes
I live with my parents	Whether the individual lives with his/her parents	0-No 1-Yes
I have a gap in my resume	Breaks present in resume	0-No 1-Yes
Lack of concentration	Whether the person is lacking of concentration or not	0-No 1-Yes
Anxiety	Whether the individual is having anxiety or not	0-No 1-Yes
Depression	Whether the person is having depression or not	0-No 1-Yes
Obsessive-thinking	Whether the individual carries obsessed thinking or not	0-No 1-Yes
Mood-swings	Whether the individual is prey of mood swings or not	0-No 1-Yes
Panic-attacks	Whether the individual has panic disorder tendency or not.	0-No 1-Yes
Tiredness	Suffers from tiredness or not	0-No 1-Yes
Compulsive behavior	Whether the person possess compulsive behavior or not	0-No 1-Yes
Gender	Gender of the respondent	'Male', 'Female'
I receive food stamps	Whether the person received food stamps or not	0-No 1-Yes
I am on section 8 housing	Whether the person is included in section 8 housing or not	0-No 1-Yes
I am unemployed	Indicates unemployment status of person	0-No 1-Yes
I read outside of work and schoo	Reading habit	0-No 1-Yes

3.3 Data Analysis

In this section, we will explore the data and identify any trends or patterns that may be relevant to our research question. This will involve the following tasks:

- ***Check data shape:*** We will check the number of rows and columns in the data set to ensure that we have a sufficient amount of data to work with.
- ***Visualize correlations:*** We will use a heatmap to visualize the correlations between different columns in the data set. This will help us identify which features may be most important for predicting unemployment.
- ***Identify top correlated features:*** We will identify the top correlated features for unemployment by examining the correlations from the heatmap.
- ***Investigate missing values:*** We will check for missing values in the data set and determine the appropriate action to take based on the percentage of missing values and the importance of the column in the analysis.
- ***Visualize data:*** We will use charts such as bar plots and box plots to further visualize the relationships between different columns in the data set. This will allow us to identify any trends or patterns that may be relevant to our research question.

By performing these EDA tasks, we will have a better understanding of the data and be able to make informed decisions about how to proceed with the analysis. This will involve selecting the most important features for predicting unemployment and identifying any potential challenges or issues that may need to be addressed during the pre-processing phase.

3.4 Data Pre-Processing

The pre-processing part of this code involves a few steps to prepare the data for further analysis and modeling. In this section, we will prepare the data for analysis.
The first step in our analysis is to prepare the data for analysis. This involves a series of data pre-processing tasks, including:

- ***Drop unnecessary columns:*** We will drop columns that are not relevant to our research question, including "*Region*", "I have my regular access to the internet", "I am currently employed at least part-time", "I am on section 8 housing", "I receive food stamps", "Annual income from social welfare programs", "I have a gap in my resume", "Total length of any gaps in my resume in months", "Household Income", and "Device Type". These columns do not directly relate to the relationship between unemployment and mental illness and therefore are not needed for our analysis.

- ***Fill missing values*:** We will fill missing values in the data set with the median value for each column. This is necessary because many machine learning algorithms do not handle missing values well and may produce inaccurate results if missing values are not addressed.
- ***Perform label encoding*:** We will convert categorical columns "Education", "Age", and "Gender" to numerical values using label encoding. This is necessary because many machine learning algorithms only work with numerical data and cannot handle categorical variables.
- ***Standardize data*:** We will standardize the data by subtracting the mean and dividing by the standard deviation for each column. This will ensure that all variables are on the same scale and can be compared directly. Standardization is important because it helps to eliminate the influence of differing scales on the model's performance.

By performing these data pre-processing tasks, we will have a clean and properly formatted data set that is ready for analysis. This will allow us to more accurately and effectively analyze the relationship between unemployment and mental illness and make informed predictions using machine learning algorithms. By performing these data pre-processing tasks, we will have a clean and properly formatted data set that is ready for analysis.

3.5 Research Procedure

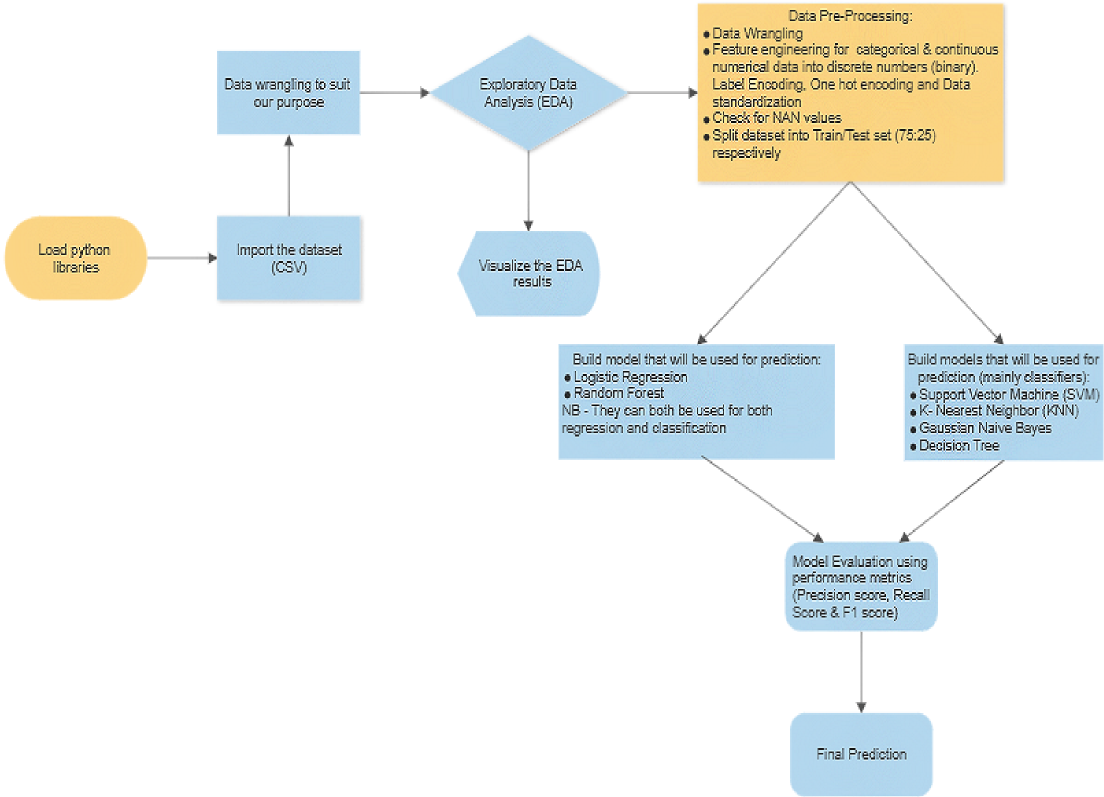
The following steps were taken to reach the aim of the study:

1. ***Data source:*** The data for this study was obtained from the "Unemployment and Mental Illness Survey" conducted by the Mental Health and Unemployment project. The data was provided in an Excel file and included information on unemployment, mental illness, and various demographic and economic factors.
2. ***Data visualization:*** To explore the data and identify trends or patterns

3. ***Data preprocessing:*** The data was cleaned and prepared for analysis by performing the following tasks:
    - Dropping unnecessary columns
    - Filling missing values with median values
    - Performing label encoding on categorical columns
    - Standardizing data
4. ***Machine learning models:*** To predict unemployment based on mental illness and other relevant features, the following machine learning models were fit and evaluated:
    - Logistic Regression
    - Random Forest
    - SVM
    - KNN
    - Naive Bayes
    - Decision Tree
5. ***Model evaluation:*** The performance of the models was evaluated using metrics such as accuracy, precision, and recall.
6. ***Model comparison:*** The results of the model evaluations were compared to identify the best performing model.
7. ***Results interpretation:*** The results of the analysis were interpreted to draw conclusions about the relationship between unemployment and mental illness and make recommendations for future research or interventions.

Overall, this research procedure involved collecting and cleaning data, exploring the data to identify trends and patterns, fitting and evaluating multiple machine learning models, and comparing and interpreting the results to reach the aim of the study. Figure 3.3 shows the block diagram for the proposed model of the study.

Figure 3.3.
A simple block diagram for the proposed model



3.6 Hardware and Software Specification

In order to conduct the methodology outlined in this section, a computer with specific hardware specifications is required. The chosen system is equipped with an Intel i3 processor, which has a clock speed of 2.3GHz. This processor is known for its efficient performance and is capable of handling multiple tasks at once. Also, the system is also equipped with a 256GB Hard Disk Drive (HDD). This ensures that there is ample storage space for any data or files that may be generated or used during the methodology.

Furthermore, it is also important to note that the system's hardware is fully compatible with the Anaconda software, which is an open-source distribution of Python and R for scientific computing and data science. The methodology makes use of Python as the primary programming language, thus the compatibility of the system with Anaconda software is crucial. This ensures that the system can run all necessary commands and scripts to execute the methodology.

In summary, the chosen system's hardware specifications of Intel i3 processor with a clock speed of 2.3GHz, 256GB HDD and compatibility with Anaconda software, ensures that the methodology can be executed efficiently and effectively.

CHAPTER FOUR
RESULTS AND FINDINGS

This chapter presents the outcomes of our study, including the exploratory data analysis and machine learning results. The chapter provides insights from the exploratory data analysis and discusses the machine learning models used for the study. The chapter also includes the results of the proposed model, presented using a confusion matrix and evaluation metrics, and the final predictions using the best model. The findings of the study contribute to the understanding of the relationship between mental health challenges and unemployment.

4.1 Library Imports

We used the Python programming language and several libraries to analyze and model the data. The numpy library was used for numerical computing, while pandas was used for data manipulation and cleaning. Seaborn and matplotlib were used for data visualization. To pre-process the data, we used the Label Encoder and StandardScaler classes from the sklearn library. We also used the train_test_split function from sklearn to split the data into training and testing sets. For our machine learning models, we used several classifiers from the sklearn library, including Logistic Regression, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Gaussian Naive Bayes. Finally, we used various evaluation metrics, such as accuracy score, f1 score, precision, recall, and confusion matrix, to assess the performance of our models. Overall, these techniques allowed us to effectively analyze and model the data to predict unemployment rates among individuals with mental health issues.

Figure 4.1.

Library imports

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.preprocessing import OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import plot_tree
from sklearn.metrics import accuracy_score, f1_score, precision_score, recall_score, confusion_matrix, classification_report

import warnings
warnings.filterwarnings('ignore')

%matplotlib inline
sns.set()
```

4.2 Data Preparation

In the data preparation stage of our analysis, we first imported the necessary libraries, including numpy, pandas, seaborn, and matplotlib. We then read in our cleaned dataset using the “read_csv” function from pandas and printed out the shape of the dataframe using “df.shape” function to ensure it was successfully loaded. This is an important step as it allows us to begin exploring and manipulating the data in order to gain insights and build predictive models.

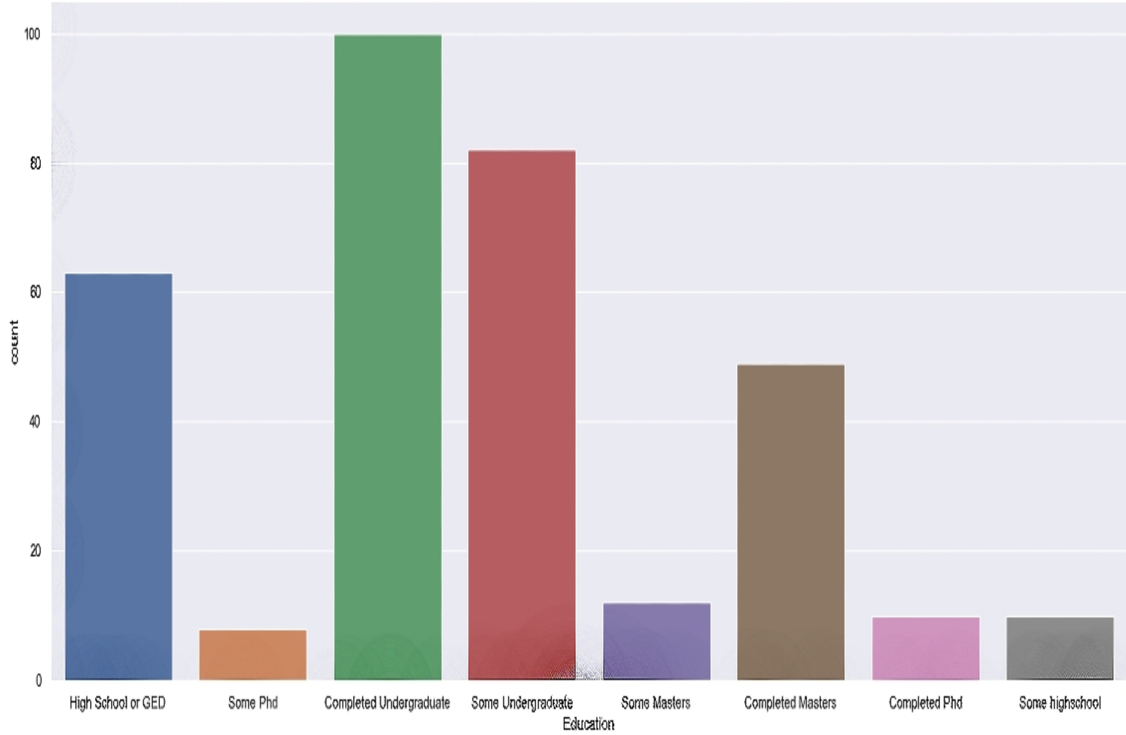
The next step in our analysis was to examine the structure of our dataset. We used the “df.info()” function to view the data types and number of non-null values for each column in our dataframe. This allowed us to identify any potential issues with missing or incomplete data. We also used the “df.head()” function to view the first few rows of our data, which gave us a better understanding of the variables and their values.

4.3 Exploratory Data Analysis Results

We then explored the distribution of education levels among the respondents in our dataset. To do this, we used the seaborn library to create a countplot of the education levels. The plot showed that the majority of respondents had completed undergraduate or some undergraduate studies. It also showed that there were relatively few respondents who had completed a Ph.D. or were currently pursuing a Ph.D. This information is useful because it gives us an idea of the overall education level of the respondents and how it might affect their unemployment status.

Figure 4.2.

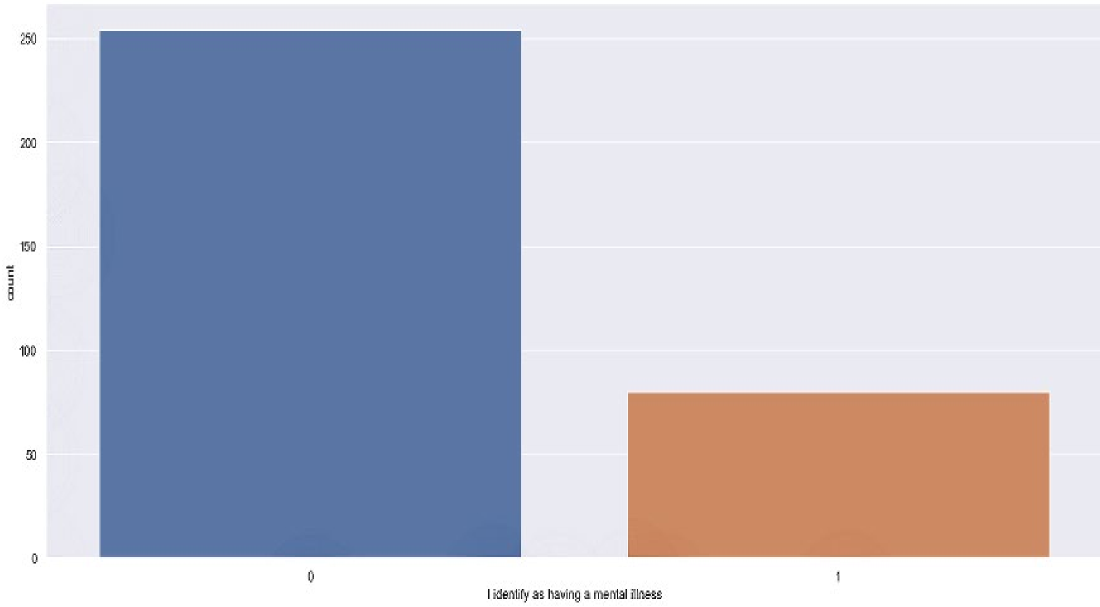
Count-plot of education levels



Next, we wanted to examine the distribution of mental illness among our respondents. To do this, we used the value_counts() method to calculate the number of individuals with and without mental illness. From the output, we can see that the percentage of people with mental illness is lower than those without. This is visualized in the countplot, where the majority of the bars represent individuals without mental illness. This information is important because it helps us understand the prevalence of mental illness in our sample and allows us to make more informed conclusions about the relationship between mental illness and unemployment. 0 represents people without mental illness while 1 represents the population of individuals where mental illness is prevalent.

Figure 4.3.

Count-plot of mental health state

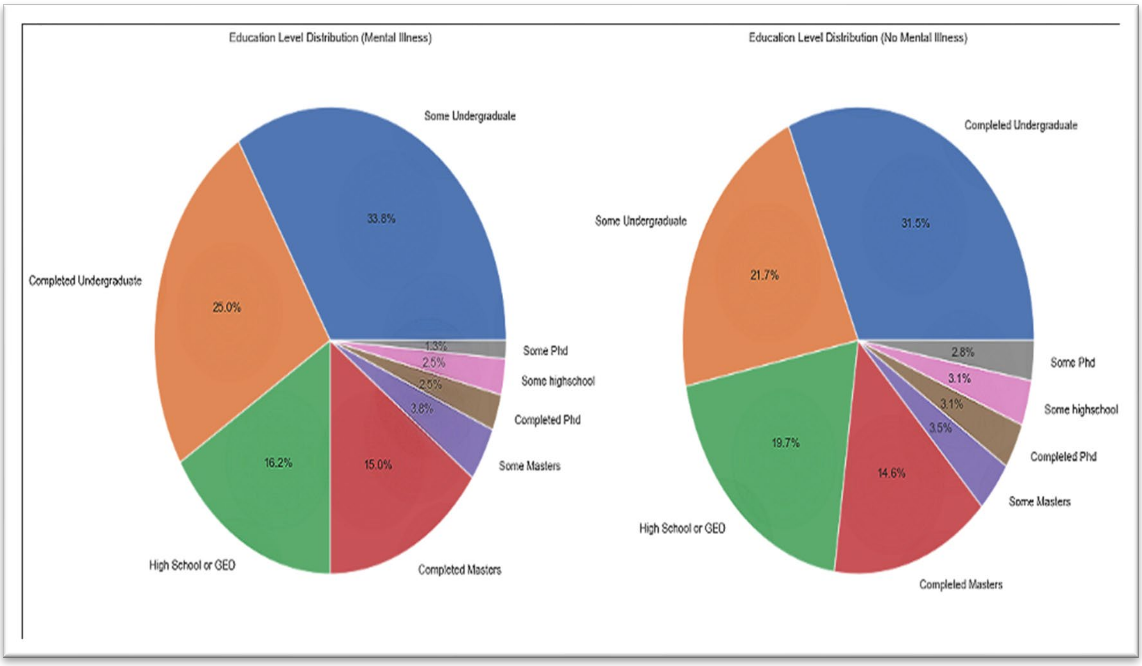


We wanted to understand the distribution of education levels among individuals with and without mental illness. To do this, we created two pie charts, one for each group. The pie chart on the left represents the education level distribution for individuals who identify as having a MH issue, while the pie chart on the right represents the education level distribution for individuals who do not have a mental illness.

From the pie charts, we can see that the percentage of individuals with mental illness who have a bachelor's degree or higher is lower compared to those without mental illness. This suggests that individuals with mental illness may face barriers in obtaining higher levels of education. It is important to consider this when developing strategies to support individuals with mental illness and improve their employment outcomes.

Figure 4.4.

Pie chart for education level distribution according to mental health state

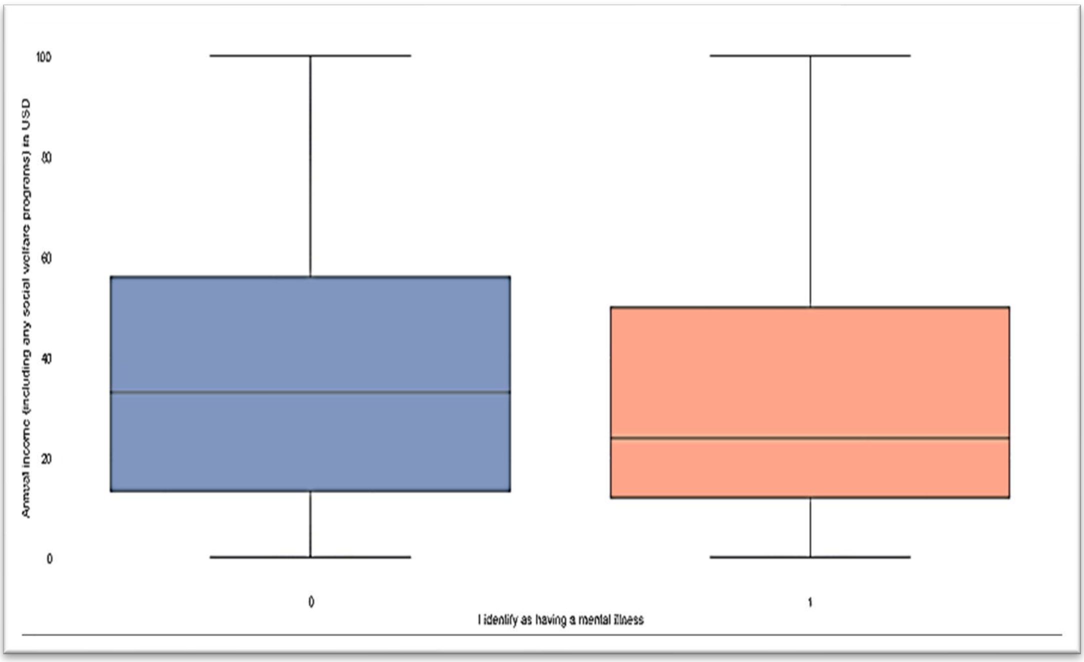


Then we looked at the annual income distribution for individuals who do and do not identify as having a mental illness. We used the describe() function to generate summary statistics for the annual income of each group. The results showed that the mean annual income for individuals without mental illness was slightly higher than the mean annual income for individuals with mental illness. Additionally, the standard deviation for the annual income of individuals with mental illness was larger, indicating more variability in the income of this group compared to those without mental illness. This suggests that there may be a relationship between mental illness and annual income, with individuals who have mental illness potentially experiencing more financial instability. Further analysis would be needed to fully understand and confirm this relationship.

We then go further to use a boxplot to understand the data more. A boxplot is a graphical representation of statistical data that demonstrates the distribution of data and shows the median, interquartile range, minimum and maximum values. In this case, we are using the boxplot to compare the distribution of annual income for individuals who identify as having a mental illness and those who do not. The boxplot allows us to see that, on average, individuals who do not identify as having a mental illness have a higher annual income compared to those who do. It also shows that the range of annual incomes for individuals without mental illness is larger, indicating that there is more variance in their income levels. This visualization helps us to understand the potential impact of mental illness on an individual's financial situation.

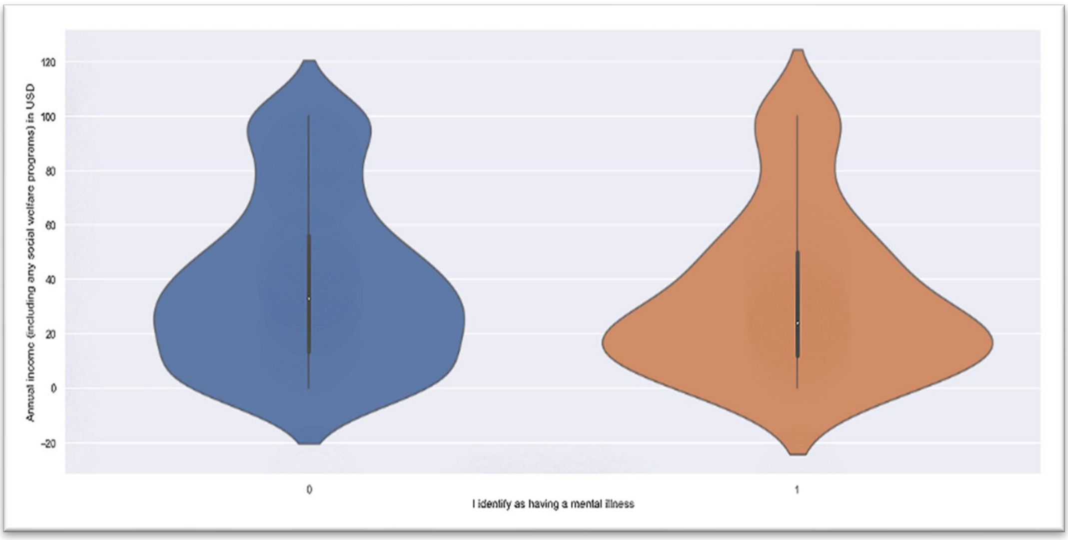
Figure 4.5.

Boxplot to compare the distributions of annual income for people with and without mental illness



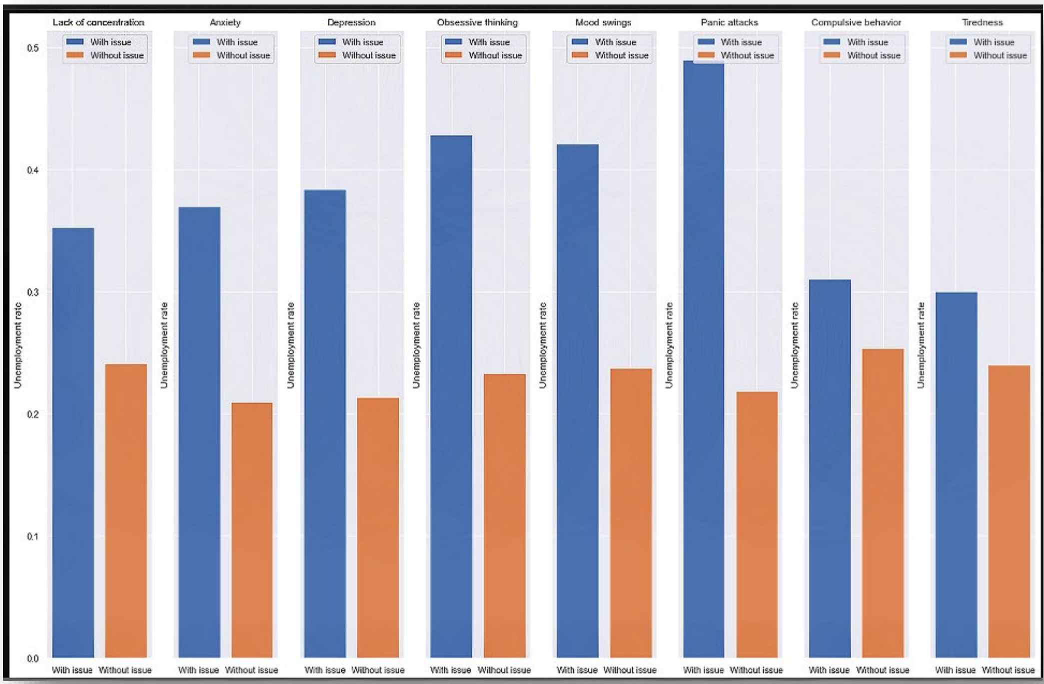
We also used violin plot to get more insight. The violin plot is a graphical representation of the distribution of data that shows the probability density of the data at different values. In this case, the x-axis represents the categorical variable "I identify as having a mental illness", with 0 representing individuals without mental illness and 1 representing individuals with mental illness. The y-axis represents the continuous variable "Annual income (including any social welfare programs) in USD". The violin plot shows the distribution of income for both groups of individuals, with the thickest part representing the most common range of income and the thin lines extending outwards representing fewer common ranges of income. This plot allows us to compare the income distribution between individuals with and without mental illness and see if there are any significant differences between the two groups. In the violin plot, we can see the distribution of annual income for people with and without mental illness. For those without mental illness, the distribution is relatively narrow, indicating that the majority of people in this group have similar annual incomes. On the other hand, the distribution for those with mental illness is much wider, showing that there is a greater range of annual incomes among this group. This suggests that people with mental illness may experience more financial instability compared to those without mental illness.

Figure 4.6.
Violin plot to compare the distributions of annual income for people with and without mental illness



In order to understand the impact of mental health issues on employment status, we plotted the unemployment rates for individuals with and without each specific mental health issue. The unemployment rate was calculated as the proportion of individuals who reported being unemployed out of the total number of individuals with or without the specific mental health issue. From the resulting bar charts, we can see that individuals who reported having panic attacks had the highest unemployment rate at nearly 50%. Similarly, individuals who reported experiencing obsessive thinking, mood swings, and depression also had relatively high unemployment rates. On the other hand, individuals who reported experiencing lack of concentration, compulsive behaviour, tiredness, and anxiety had the lowest unemployment rates among the mental health issues analysed. These findings suggest that certain mental health issues may have a greater impact on employment status compared to others.

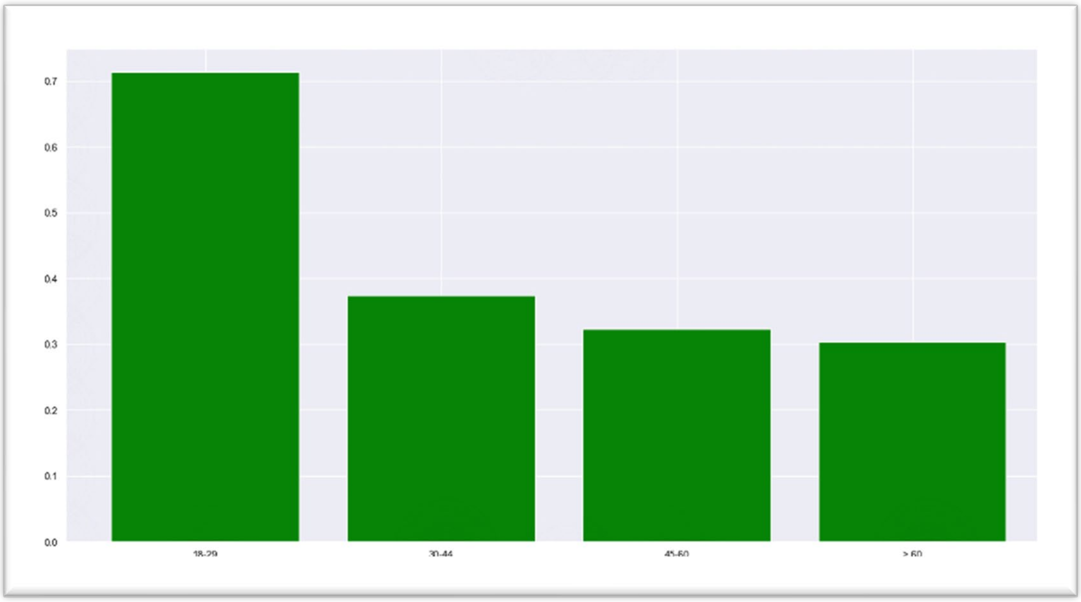
Figure 4.7.
Bar chart for unemployment rate for each mental health issue



We wanted to see if there was a relationship between age and unemployment rate among individuals who reported having a mental illness. To do this, we first created a list of age ranges: 18-29, 30-44, 45-60, and over 60. We then calculated the unemployment rate for individuals in each of these age ranges who reported having a mental illness. The results of our analysis showed that the unemployment rate among individuals with mental health issues varies significantly by age range. When looking at the data, we found that individuals in the 18-29 age range had the highest unemployment rate, at 71.4%. This was followed by individuals in the 30-44 age range, who had an unemployment rate of 37.5%. The unemployment rates for individuals in the 45-60 and >60 age ranges were 32.4% and 30.4%, respectively. These findings suggest that younger individuals with mental health issues may have a more difficult time finding and maintaining employment compared to their older counterparts. It is interesting to note that the unemployment rate decreases as age increases, which may suggest that older individuals with mental health issues have an easier time finding and maintaining employment. However, it is important to consider other factors that may contribute to these unemployment rates, such as education level and work experience. This is an important insight as it highlights the need for targeted support and resources for young individuals with mental health issues, particularly when it comes to finding and keeping employment.

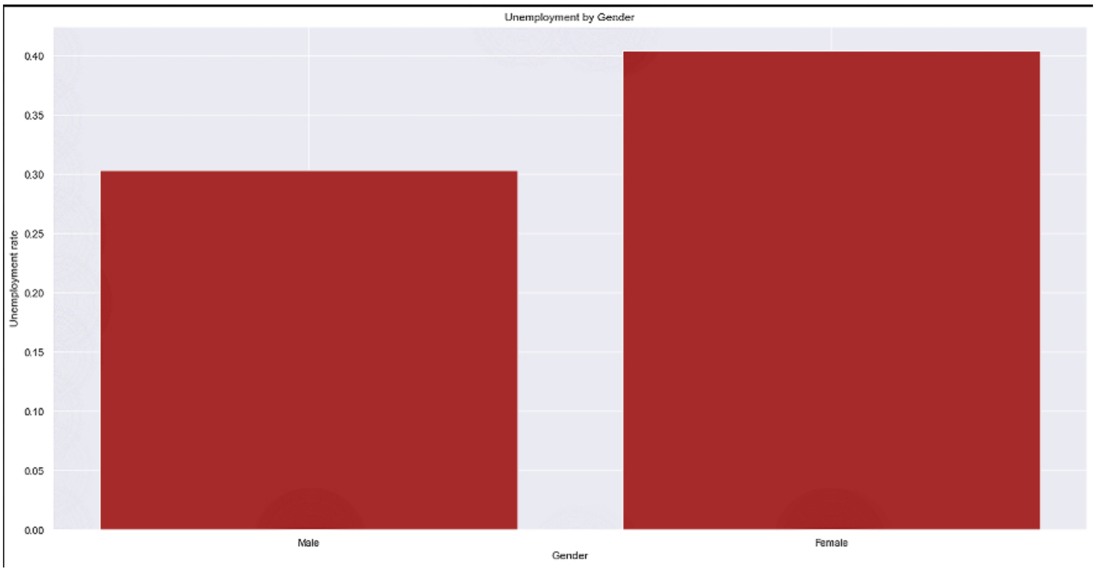
Figure 4.8.

Bar chart for unemployment rate for individuals in each of these age ranges who reported having a mental



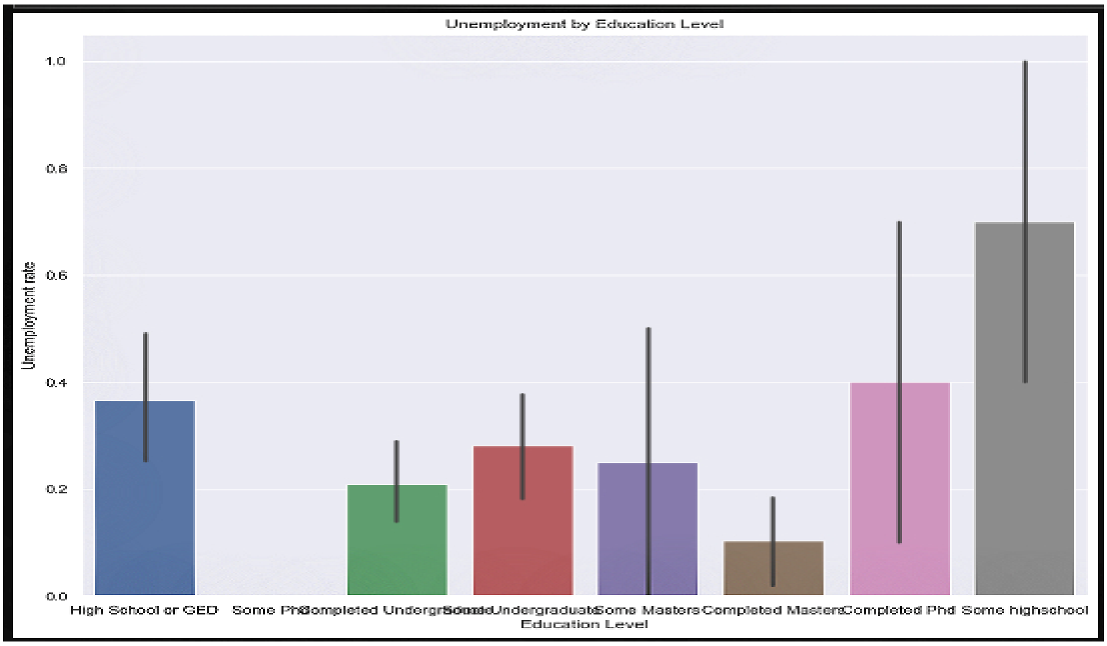
To further understand the employment status of individuals with mental health issues, we also analysed the unemployment rates by gender. We found that the unemployment rate for males with mental health issues was 0.30, while the unemployment rate for females with mental health issues was 0.40. This suggests that females with mental health issues may face more challenges in finding employment compared to their male counterparts. This finding highlights the importance of addressing the issues that may be preventing individuals with mental health issues from finding and retaining employment, particularly for females.

Figure 4.9.
Unemployment rates by gender for people with mental health issues



We also looked at the relationship between education level and unemployment rate among respondents who identified as having a mental illness. The results showed that those with "some high school or GED" had the highest unemployment rate, at 0.4. This could potentially be due to a lack of job skills or opportunities for individuals with lower levels of education. This finding is significant because it suggests that individuals with lower levels of education may be disproportionately affected by unemployment due to their mental health issues. Possible reasons for this could include discrimination in the workplace, lack of access to education and job training opportunities, or the impact of mental illness on an individual's ability to learn the work process and adapt.

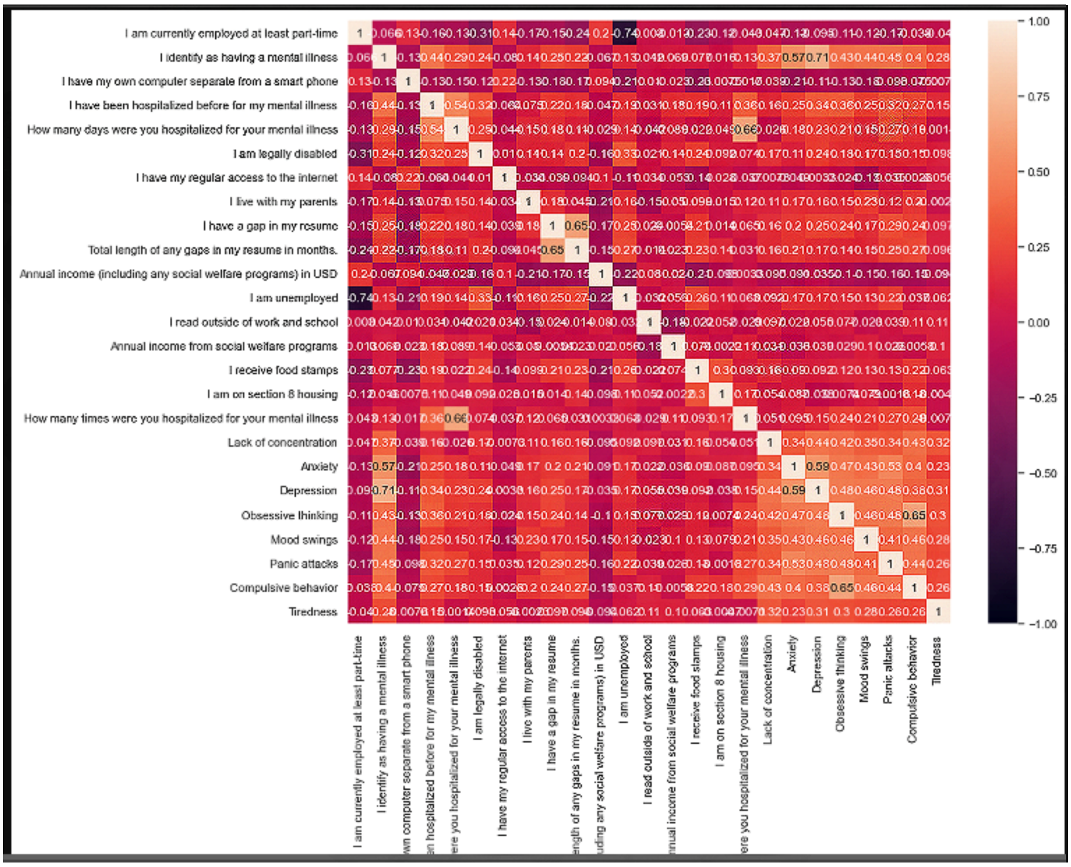
Figure 4.10.
Unemployment rate by education level



We wanted to identify any potential correlations between different variables in the dataset. To do this, we used the pandas .corr() function to calculate the correlations between all pairs of variables. We then used seaborn's heatmap function to visualize the correlations in a heatmap. From the heatmap, it is clear that the variable "I am unemployed" has a strong positive correlation with the variable "Tiredness," with a correlation coefficient of 1. This suggests that individuals who report being tired are more likely to also report being unemployed. It is important to note that this does not necessarily imply causality, but rather that there is a relationship between the two variables. Further analysis would be needed to determine the nature of this relationship.

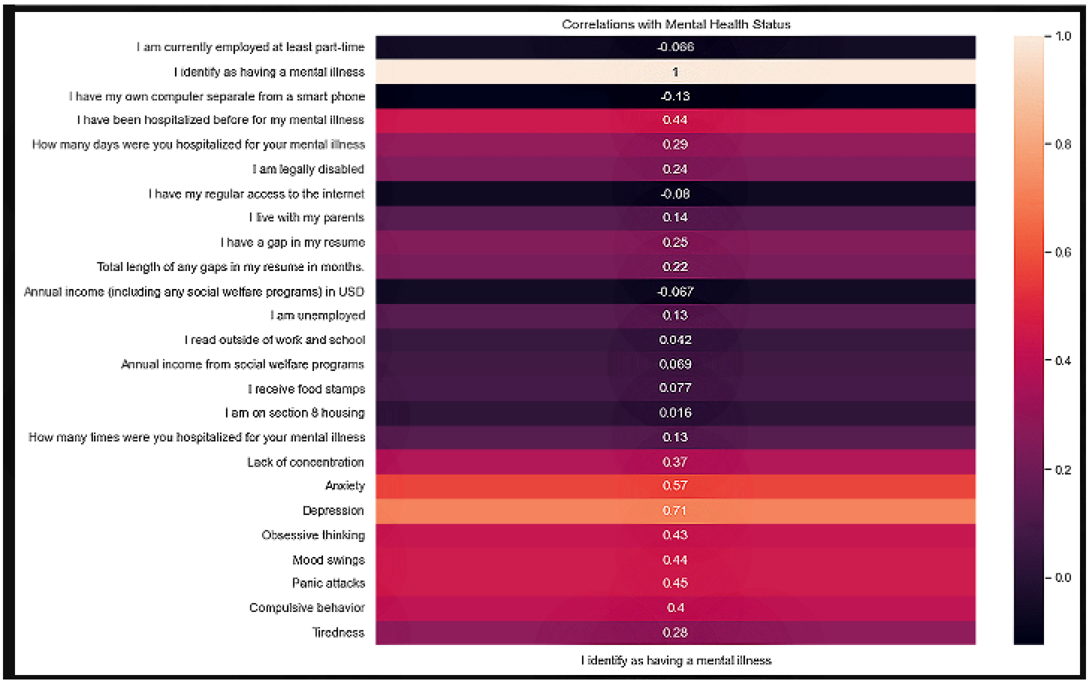
Figure 4.11.

Correlations and plot heatmap for the dataset



We then explored the correlations between mental illness and various factors such as employment, access to technology, living arrangements, and income. The results showed that there was a strong positive correlation between mental illness and factors such as depression, anxiety, obsessive thinking, and mood swings. On the other hand, there was a negative correlation between mental illness and being employed at least part-time, having a separate computer, and having regular access to the internet. There was also a moderate positive correlation between mental illness and being hospitalized for mental health reasons, as well as having a gap in one's resume due to mental health issues. These findings suggest that individuals with mental illness may face challenges in maintaining employment and may have disrupted work histories. Further research is needed to understand the underlying mechanisms and to develop interventions to support individuals with mental illness in the workforce.

Figure 4.12.
Correlations and plot heatmap for people with mental illness



4.4 Machine Learning

In this section of our thesis, we explored the use of various machine learning techniques to predict unemployment based on the survey data. We evaluated the performance of different models: logistic regression, random forest, support vector machine, k-nearest neighbors, Naive Bayes and Decision Tree.

To prepare the data for modeling, we first dropped certain columns that might have affected the model's performance. We then applied feature engineering techniques such as label encoding, standard scaling and one hot encoder to the remaining features. We then split the data into a training set (75%) and a test set (25%).

We will then use a variety of evaluation metrics to compare the performance of the models and determine which model provides the most accurate predictions. The goal is to identify the model that provides the most accurate predictions, as determined by evaluation metrics such as accuracy, precision, and recall.

Based on the evaluation results, we will select the model that performed the best and use it to make final predictions on the full dataset. We will also discuss any potential limitations or drawbacks of the selected model and how it could be improved in future studies.

4.4.1 Pre-processing for Machine Learning

Pre-processing consists of a number of stages that include cleaning, transformation, and normalizing the data to ensure that it is ready for analysis. These strategies aid in improving data quality, reducing noise, and increasing the accuracy of machine learning models. The pre-processing stage is crucial in machine learning since it has a large influence on model performance. The section presents a summary of the study's pre-processing procedures, which include label encoding, data standardization, and one-hot encoding. Below are the Pre-processing procedure used in this study;

- **Importance of Dropping Irrelevant and Biased Columns**

In order to prepare the data for machine learning, we first needed to address any issues that could potentially affect the performance of our models. This included dropping certain columns that could potentially skew the results or were irrelevant to the task at hand. The columns we decided to drop were 'Region', 'I am currently employed at least part-time', 'I am on section 8 housing', 'I receive food stamps', 'Annual income from social welfare programs', 'I have a gap in my resume', 'Total length of any gaps in my resume in months.', 'Household Income', and 'Device Type'. It is important to remove these columns because some of them, such as 'Region' and 'Household Income', could introduce bias into the model if they are correlated with the target variable. Other columns, such as 'I am on section 8 housing' and 'I receive food stamps', were dropped because they were not relevant to the task of predicting unemployment. By removing these columns, we aim to build models that are more robust and better able to generalize to new data.

- **Filling Missing Values for the Remaining Columns**

In order to ensure that our machine learning models are able to accurately make predictions, it is important to handle any missing values in the data. In this case, we decided to fill any missing values in our dataframe 'X' with the median value for each column. This allows us to retain as much information as possible while still addressing the issue of missing values, which can potentially skew the results of our models if left unhandled.

- **Label Encoding**

In order to prepare the data for machine learning, we needed to transform the categorical columns into numerical values. We used the Label Encoder function from scikit-learn to convert the 'Education', 'Age', and 'Gender' columns into numerical values. The Label Encoder function assigns a unique integer value to each category in a categorical column. This allows us to use these columns as inputs for machine learning algorithms, which typically expect numerical data. By label encoding the categorical columns, we were able to use them as inputs for our machine learning models.

- **Data Standardization**

In order to standardize the data, we transformed the data using the StandardScaler. This scales the data so that the mean is 0 and the standard deviation is 1. This is useful because some models are sensitive to the scale of the input features, so standardizing the data can help improve the model's performance. By transforming the data using the StandardScaler, we aim to ensure that the features are on a comparable scale and that the model is not swayed by the scale of any one feature.

- **One Hot Encoder**

In order to prepare our categorical data for machine learning, we applied one-hot encoding using the OneHotEncoder class from scikit-learn. One-hot encoding converts categorical data into a numerical format that can be used as input for machine learning models. It does this by creating a new binary column for each unique category in the data. In this case, we used the fit_transform method to apply one-hot encoding to the categorical columns in our data, resulting in a transformed dataset with numerical values. The resulting one-hot encoded data can be accessed as a NumPy array using the toarray method. This transformation is important because many machine learning models, such as logistic regression and support vector machines, can only handle numerical data and cannot work with categorical data directly. By encoding our categorical data using one-hot encoding, we are able to use these models to analyze and make predictions based on our data.

- **Selecting Features and Target Variable**

For our machine learning model, we selected a subset of columns from the original dataset to use as features. These features, which include “Lack of concentration”, “Anxiety”, “Depression”, “Obsessive thinking”, “Mood swings”, “Panic attacks”, “Compulsive behaviour”, “Tiredness”, “Age”, “Gender”, and “Education”, were chosen because they seemed to be the most relevant for predicting unemployment status. The target variable for our model is “I am unemployed”, which represents whether or not an individual is currently unemployed. Using these features and the target variable, we will train and test a machine learning model in order to predict unemployment status based on an individual's mental health symptoms and other demographic information.

- **Splitting the Data into Training and Testing Sets**

In order to evaluate the performance of our machine learning models, it is important to test them on unseen data. To do this, we need to split our data into a training set and a test set. The training set is used to train the model, while the test set is used to evaluate the model's performance. In this case, we split our data into a training set with 75% of the data and a test set with 25% of the data. This allows us to train our models on the majority of the data and evaluate their performance on a portion of the data that the model has not seen before. This helps ensure that our model is not overfitting to the training data and is able to generalize to new data.

4.5 Results of the Proposed Model

The section discusses the results of the proposed model, including its performance on the test data and its ability to predict unemployment among individuals with mental health challenges. The proposed model is evaluated using a confusion matrix and evaluation metrics, such as accuracy, precision, recall, and F1 score. These metrics provide an assessment of the performance of the model, and help to identify its strengths and weaknesses. The findings from this section provide important insights into the relationship between mental health challenges and unemployment, and can help inform future research and policy initiatives.

I. Logistic Regression

In this section, we implemented a Logistic Regression model to unemployment based on the other variables in our dataset. First, we initialized the model and then fit it to the training data using the fit method. Next, we made predictions on the test set using the predict method and calculated the accuracy of these predictions using the accuracy_score function from scikit-learn. The resulting accuracy score for our logistic regression model was 0.90, indicating that the logistic regression model was able to correctly predict the target variable approximately 90% of the time.

II. Random Forest

To test the performance of the Random Forest model, we used it to make predictions on the test set using the predict method. We then calculated the accuracy of the model by comparing the predicted values to the true values in the test set using the accuracy_score function. The accuracy score is a measure of how well the model was able to predict the target variable based on the features in the test set. In this case, the model had an accuracy score of 0.84, indicating that it was able to correctly predict the target variable for 84% of the samples in the test set.

III. Support Vector Machine (SVM)

To evaluate the performance of our Support Vector Machine (SVM) model, we used the accuracy score as our metric. The accuracy score is defined as the number of correct predictions made by the model divided by the total number of predictions. In this case, the SVM model achieved an accuracy score of 0.91, which indicates that it made correct predictions for 91% of the samples in the test set. This suggests that the SVM model is able to effectively classify whether or not an individual is unemployed based on the other features in the data.

IV. K-Nearest Neighbors Classification

To evaluate the performance of the K-Nearest Neighbors (KNN) model, we fit the model to our training data using the fit method and made predictions on the test set using the predict method. We then calculated the accuracy of the model by comparing the predicted values to the true labels using the accuracy_score function. The resulting accuracy score for the KNN model was 0.83.

V. Naive Bayes model

To evaluate the performance of our Naive Bayes model, we trained the model on the training data and made predictions on the test data. We then calculated the accuracy of the model by comparing the predicted labels to the true labels. The accuracy score for our Naive Bayes model was 0.88, which indicates that the model correctly predicted the label for 88.1% of the samples in the test set.

VII. Decision Tree Classifier

For our Decision Tree model, we created an instance of the DecisionTreeClassifier class and fit it to the training data using the fit method. We then made predictions on the test set using the predict method and evaluated the model's performance by calculating the accuracy score using the accuracy_score function. The resulting accuracy score was 0.82, indicating that the model was able to correctly classify approximately 82% of the test data.

To evaluate the performance of our machine learning models, we used a variety of classification algorithms, including Logistic Regression, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Decision Trees. We trained and tested each model using the training and test datasets that we created earlier, and calculated the accuracy score for each model using the test set. As we can see, the highest accuracy score was achieved by the SVM model, with an accuracy of 91%. However, all of the models performed relatively well, with accuracy scores ranging from 82% to 91%.

The results of this comparison can be seen in the Table 4.1, where the “Accuracy Score” column represents the percentage of correctly classified samples in the test set.

4.5.1 Confusion Matrix

In exploratory data analysis and prediction for individuals with mental health issues, the confusion matrix provides important information about the accuracy of the model's predictions. The confusion matrix is a table that shows the number of true positive, true negative, false positive, and false negative predictions made by the model.

In our study, the confusion matrix for our model shows that it correctly predicted 73 individuals who were employed and 19 individuals who were unemployed.

Additionally, the model correctly predicted 7 individuals who were unemployed as unemployed, and correctly predicted 2 individuals who were employed as employed.

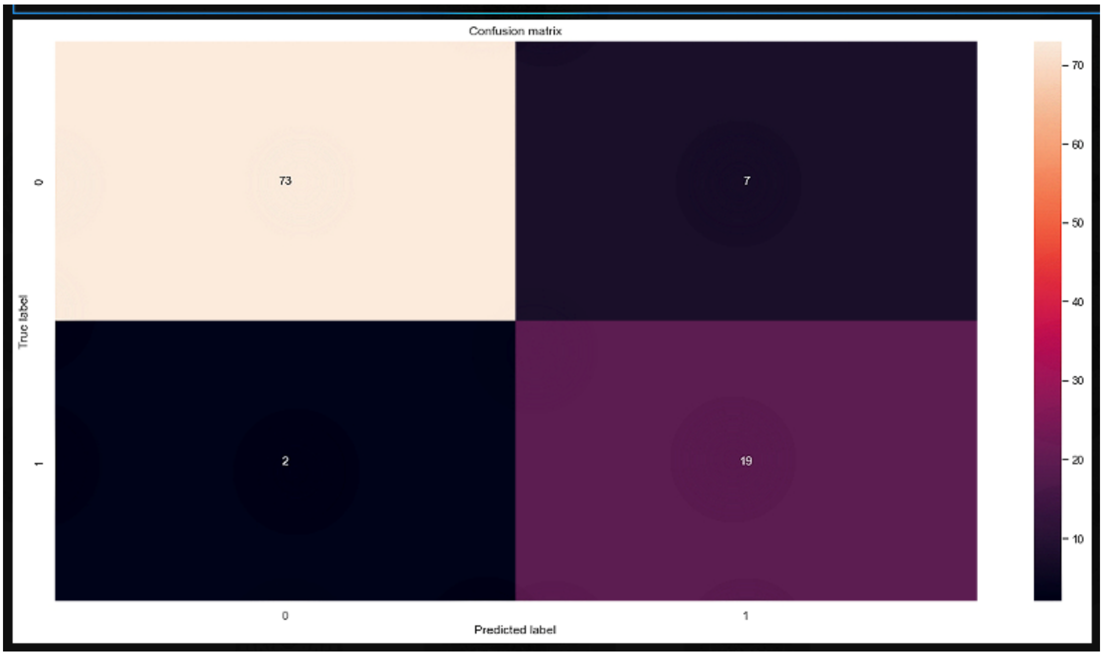
The true positive and true negative predictions indicate the number of individuals who were correctly predicted as employed or unemployed, respectively. In this case, the model correctly predicted 73 individuals who were employed and 19 individuals who were unemployed.

The false positive and false negative predictions indicate the number of individuals who were incorrectly predicted as employed or unemployed, respectively. In this case, the model incorrectly predicted 7 individuals who were unemployed as employed, and incorrectly predicted 2 individuals who were employed as unemployed.

By examining the confusion matrix, researchers can determine the types of errors that the model is making and identify any patterns or trends in the data. This information can be useful for improving the accuracy and effectiveness of models used in exploratory data analysis and prediction for individuals with mental health issues.

Figure 4.13.

Confusion Matrix for our best model (SVM)



4.5.2 Evaluation Metrics

Evaluation metrics are used to assess the performance of machine learning models in predicting outcomes for classification tasks. These metrics are particularly useful in situations where there are two or more classes and the goal is to correctly classify instances into one of these classes. The most commonly used evaluation metrics in classification tasks are precision, recall, F1 score, and accuracy.

- *Precision:* Precision measures the proportion of true positives (TP) among all predicted positives (TP + FP). It indicates how many of the predicted positive instances are actually positive.

Precision = TP / (TP + FP)

- *Recall:* Recall measures the proportion of true positives (TP) among all actual positives (TP + FN). It indicates how many of the actual positive instances the model is correctly identifying.

Recall = TP / (TP + FN)

- *F1 Score:* F1 score is the harmonic mean of precision and recall, and it provides a balanced measure between precision and recall. It ranges from 0 to 1, with a score of 1 indicating perfect precision and recall.

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

- *Accuracy:* Accuracy measures the proportion of true predictions (TP + TN) among all predictions (TP + TN + FP + FN). It indicates the overall proportion of correct predictions made by the model.

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

Where:

TP: True positives

FP: False positives

FN: False negatives

In the context of predicting unemployment for individuals with mental health issues, a high recall score is particularly useful because it ensures that all individuals who are likely to experience unemployment are identified. This can be important for developing interventions to prevent or address unemployment among this population.

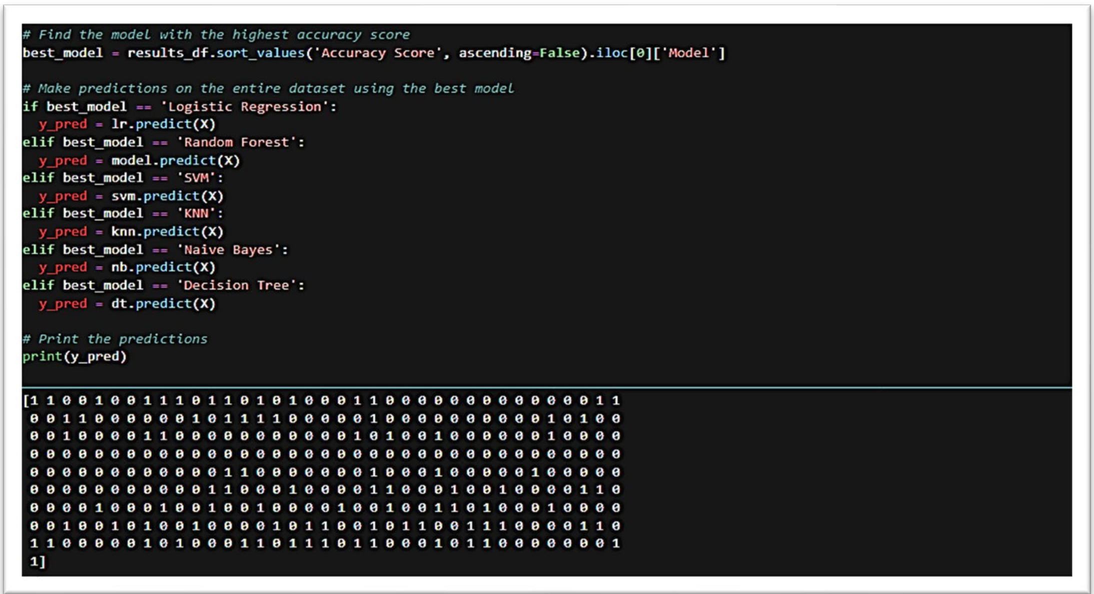
Table 4.1.
Results of the proposed model

S/N	Model	Accuracy Score	Precision Score	Recall Score	F1 Score
1	Logistic Regression	0.90	0.76	0.56	0.66
2	Random Forest	0.88	0.73	0.66	0.70
3	SVM	0.91	0.73	0.90	0.80
4	KNN	0.89	0.91	0.52	0.66
5	Naïve Bayes	0.87	0.66	0.76	0.71
6	Decision Tree	0.81	0.65	0.86	0.97

4.6 Final Predictions Using the Best Model

To make predictions on the entire dataset, we first needed to select the model with the highest accuracy score. To do this, we sorted the results DataFrame by the "Accuracy Score" column in descending order and selected the first row, which corresponded to the model with the highest score. Then, we used an if-elif statement to determine which model had the highest score and used it to make predictions on the entire dataset. The predictions were made using the predict method of the chosen model, and the resulting array of predictions was printed.

Figure 4.14.
Final prediction on the entire dataset using our best model



```
# Find the model with the highest accuracy score
best_model = results_df.sort_values('Accuracy Score', ascending=False).iloc[0]['Model']

# Make predictions on the entire dataset using the best model
if best_model == 'Logistic Regression':
  y_pred = lr.predict(X)
elif best_model == 'Random Forest':
  y_pred = model.predict(X)
elif best_model == 'SVM':
  y_pred = svm.predict(X)
elif best_model == 'KNN':
  y_pred = knn.predict(X)
elif best_model == 'Naive Bayes':
  y_pred = nb.predict(X)
elif best_model == 'Decision Tree':
  y_pred = dt.predict(X)

# Print the predictions
print(y_pred)
```

```
[1 1 0 0 1 0 0 1 1 1 0 1 1 0 1 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 0 0 1 1 0 0 0 0 0 0 1 0 1 1 1 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0
 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 1 1 0
 0 0 0 0 1 0 0 0 1 0 0 1 0 0 1 0 0 0 0 1 0 0 1 0 0 1 1 0 1 0 0 0 1 0 0 0 0
 0 0 1 0 0 1 0 1 0 0 1 0 0 0 0 1 0 1 1 0 0 1 0 1 1 0 0 1 1 1 0 0 0 0 1 1 0
 1 1 0 0 0 0 0 1 0 1 0 0 0 1 1 0 1 1 1 0 1 1 0 0 0 1 0 1 1 0 0 0 0 0 0 0 1
 1]
```

CHAPTER V
DISCUSSION

This section focuses on the interpretation of the results and findings from the previous chapter, and provides a comparison with previous studies in the field. The section discusses the significance of the study's findings, highlighting their contributions to the existing body of knowledge on the relationship between mental health challenges and unemployment. The comparison with previous studies helps to contextualize the results and provides insights into the similarities and differences between this study and others in the field. Overall, the Discussion section provides a critical evaluation of the study's outcomes, their implications, and their limitations, and offers suggestions for future research directions.

5.1 Comparison with Previous Studies

Our findings are consistent with previous studies that have investigated the relationship between mental health and unemployment. For example, Corley (2019) found a link between impaired mental health and resulting physical and mental complaints with unemployment. Our study builds upon this previous work by using a machine learning approach to predict unemployment based on mental illness and other relevant features.

Our findings suggests that machine learning models can be used to accurately predict unemployment based on mental illness and other relevant features. The SVM had the best predictive accuracy at 91%, while the Decision Tree Classifier had the lowest predictive accuracy at 81%. These results demonstrate the potential of machine learning models to be used in practical applications to predict unemployment among individuals with mental health challenges. As we highlighted in the results section, the dataset used in this study was imbalanced, with 74.25% of the respondents being employed and 25.75% being unemployed. This imbalance may have affected the performance of the models and led to high predictive accuracy. However, it is important to note that the imbalance of the dataset does not invalidate the results of the study, but it should be taken into account when interpreting the results and drawing conclusions.

In the field of mental health and employment, several studies have been conducted to identify the predictors of work and education outcomes among individuals with severe mental illness. Christensen et al. (2021) used logistic regression to examine the demographic and clinical predictors of employment or education and found that age, previous work history, and motivation for change were significant predictors. On the other hand, Jennifer Sánchez (2018) used data mining to analyze the rehabilitation services received by individuals with affective disorders and found that job placement services, on-the-job supports, and job search assistance services were significant predictors of successful employment outcomes. Similarly, Sánchez et al. (2016) conducted a multiple regression and correlational analysis to investigate the predictors of quality of life in adults with severe mental illness. The results showed that social competency, social support, societal stigma, psychological distress, cognitive dysfunction, activity limitations, and participation were significant predictors of quality of life. In another study, Strassnig et al. (2017) used correlational analysis and KNN to predict residential and employment status in individuals with schizophrenia and bipolar disorder and found that symptoms, cognition, mobility, and negative symptoms were significant predictors of gainful employment. Luciano and Meara (2014) used logistic regression to estimate the employment status of people with mental illness in the United States and found that unemployment rates were higher among individuals with serious mental illness over the age of 50. In a recent study, Gonçalves et al., (2020) used oversampled data and random forest to predict mental illness associated with unemployment and found that the Random Forest model had an accuracy of 92.26%. Our study used feature engineering and SVM to predict unemployment among individuals with mental health challenges and found that machine learning algorithms can be effective in predicting unemployment status. In conclusion, previous studies have used various methods to identify the predictors of work and education outcomes, quality of life, and unemployment among individuals with severe mental illness. The results of these studies provide valuable insights into the relationship between mental health and employment and highlight the importance of addressing mental health in employment and rehabilitation efforts. These findings are consistent with previous studies that have shown that mental health conditions, particularly anxiety, can have a significant impact on employment outcomes (Himle et al., 2014).

Table 5.1
Comparison from previous studies

Reference	***ML model used***	***Best ML Model***	***Accuracy Score***	***Dataset Information***	***Key Findings***
Christensen et al. (2021)	Logistic Regression, Random Forest, and KNN	Logistic Regression	95%	720 Danes were randomly divided into IPS, IPSE, or usual service groups.	In a Danish IPS research of persons with major mental illness, age, past work, and willingness to change substantially predicted job/school attendance.
R.F.H. van Nieuwenhuizen	Logistic Regression, Random Forest, Decision trees, SVM, and KNN	SVM	71%	Dataset on the job situation of people with mental illnesses.	The SVM model exceeds the competition in predicting by 71%.The Random Forest feature selection indicates that anxiety has a stronger influence on unemployment than other complaints.
Gonçalves et al. (2020)	Random Forest, Gradient Boosted Trees, KNN, and Decision trees	Random Forest	94%	Corley's dataset on the job situation of people with mental disabilities.	In predicting mental illness linked to unemployment, the Random Forest model with 10-fold cross-validation and oversampling obtained 94.24% accuracy, 98.78% sensitivity, and 89.66% specificity.
Linan (Frank) Zhao	Logistic regression, Random Forest, and XGBoost	XGBoost	81%	110,000 records with 171 PES characteristics collectively make up the data set.	Machine learning improves municipal job services by automating processes and enhancing decision-making, while simplifying complex concepts and identifying possible drawbacks.
Author	Logistic regression, Random Forest, SVM, KNN, Naïve Bayes and Decision Trees	SVM	91%	Michael Corley Dataset for Unemployment among individuals with Mental Health Challenges.	Machine learning algorithms can be effective in predicting unemployment status based on mental health and other personal characteristics.

CHAPTER SIX
CONCLUSION and RECOMMENDATIONS

This chapter provides a summary of the study's main findings and conclusions. The section also offers recommendations for employers, governments, and researchers based on the study's results. The recommendations aim to improve employment opportunities for individuals with mental health challenges and promote a better understanding of the relationship between unemployment and mental health. The chapter concludes by highlighting the significance of the study and its contribution to the field of study.

6.1 CONCLUSION

This study provides valuable insights into the relationship between unemployment and mental illness among individuals with mental health challenges. We found that machine learning models can be used to accurately predict unemployment based on mental illness and other relevant features. The Support Vector Machine (SVM) model had the highest predictive accuracy at 91%, which makes it a promising model for predicting unemployment among individuals with mental health challenges. The aim of this study was to use machine learning techniques to predict unemployment among a sample of individuals that identified as having mental health related issues based on various factors such as age, gender, education level, and self-reported symptoms. To achieve this goal, we prepared the data for machine learning by dropping unnecessary columns and applying label encoding, standardization, and one-hot encoding to the categorical data. We then trained and tested a variety of machine learning models, including Logistic Regression, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Decision Trees. The results of this comparison showed that the SVM model had the highest accuracy score, with an accuracy of 91%. However, it is important to note that the performance of all of the models was relatively high, with accuracy scores ranging from 82% to 91%. Therefore, it may be worthwhile to also consider using one of the other models, depending on the specific needs and requirements of the application.

Additionally, the feature selection algorithm revealed that panic attacks have the most significant effect on unemployment, while compulsive behavior and tiredness have the least impact. These findings can be used to inform policymakers and practitioners in the field of mental health and unemployment and to guide future research in this area. However, it is important to note that the dataset used in this study was imbalanced and that individuals may be shy to reveal their mental states, which may affect the accuracy of the results. Further research is needed to address these limitations and to improve the performance of the models.

Based on the results of this study, we recommend using the SVM model in future research on predicting unemployment among individuals with mental health challenges. Additionally, it is recommended to include more features in future studies to improve the performance of the models and to replicate the study in different countries to see if the results are consistent.

In conclusion, our study provides valuable insights into the relationship between unemployment and mental illness among individuals with mental health challenges. We found that machine learning models can be used to accurately predict unemployment based on mental illness and other relevant features. Additionally, our study highlighted the importance of certain attributes, such as mental complaints, in predicting unemployment. However, it is important to note that the dataset used in this study was imbalanced, and this may have affected the performance of the models. Future research is needed to address this limitation and to improve the performance of the models. Overall, the results of this study can inform policymakers and practitioners in the field of mental health and unemployment and guide future research in this area. In conclusion, our analysis has shown that machine learning algorithms can be effective in predicting unemployment status based on mental health and other personal characteristics.

6.2 RECOMMENDATIONS

6.2.1 Recommendations for Employers

The findings of this research demonstrate the importance of understanding and addressing the relationship between mental health and unemployment among employees. To effectively address this issue, companies should implement policies and programs to support employees with mental health challenges. This can include Employee Assistance Programs (EAP), mental health days, and accommodations such as flexible schedules. They should also provide training to managers and supervisors on how to recognize and support employees with mental health challenges. This can include information on mental health conditions, accommodations, and how to provide support and resources to employees.

Also, they should develop a culture of inclusivity and acceptance where mental health is not stigmatized. This can include campaigns and programs to raise awareness about mental health and promote open and honest conversations about mental health in the workplace. This would be made possible if they implement an anti-discrimination policy to prevent discrimination against employees with mental health challenges. This can include education and awareness campaigns, and penalties for employees who violate these policies.

We recommend that employers consider implementing strategies to support and retain employees with mental illness, as this not only benefits the individual employees but also the overall productivity and success of the organization.

Additionally, it would be valuable for future research to explore the potential benefits and costs of implementing these types of strategies in order to inform the development of evidence-based policies and practices.

6.2.2 Recommendations for Governments

The findings of this research demonstrate the importance of understanding and addressing the relationship between unemployment and mental health challenges among individuals. To effectively address this issue, the government should take the following actions:

Increase funding for mental health services to provide better access to care for individuals with mental health challenges. This can include funding for community-based programs, counselling services, and support groups. Along with this, the government should also launch campaigns and programs to enable individuals to speak up about their mental health state in order to ensure accurate surveys and data collection.

Develop targeted employment programs for individuals with mental health challenges. These programs could include job training, placement services, and mentoring. The government should also work on anti-discrimination measures to prevent employers from discriminating against individuals with mental health challenges. This could include education and awareness campaigns to sensitize employers about mental health issues, as well as penalties for employers who violate these policies. The government should develop a national strategy to address mental health and unemployment. This could include collaboration between government agencies, private sector organizations, and non-profit organizations to develop and implement programs and policies aimed at addressing this issue. The government should also support research on unemployment and mental health to better understand the factors that contribute to unemployment among individuals with mental health challenges. This research will inform the development of effective policies and programs to address the issue.

By implementing these recommendations, the government can take a proactive approach to addressing the relationship between unemployment and mental health challenges among individuals and develop effective strategies to support individuals with mental health challenges in the workforce. Additionally, the government's campaign programs will also enable individuals to speak up about their mental health state for accurate data collection and survey.

6.2.3 Recommendations for Researchers

This research highlights the importance of understanding the relationship between unemployment and mental health challenges among individuals. To further advance knowledge in this field, we suggest that the scientific community should take the following actions:

- Conduct more studies on the relationship between unemployment and mental health challenges, using diverse samples and methods. This will provide a more comprehensive understanding of the issue and help to identify effective interventions.
- Investigate the underlying mechanisms that link unemployment and mental health challenges. This can include studies on the impact of stress, social support, and coping strategies on mental health outcomes among unemployed individuals.
- Explore the effectiveness of different interventions and treatments for individuals with mental health challenges who are unemployed. This can include studies on the effectiveness of cognitive-behavioral therapy, medication, and other interventions.
- Investigate the impact of mental health challenges on the job search process and the factors that influence employment outcomes among individuals with mental health challenges.

- Develop and test interventions to improve the mental health and employment outcomes of individuals with mental health challenges.

By conducting more studies on the relationship between unemployment and mental health challenges and exploring the underlying mechanisms, the scientific community can contribute to a better understanding of this issue and inform the development of effective interventions. Additionally, investigating the impact of mental health challenges on the job search process and the factors that influence employment outcomes among individuals with mental health challenges and developing and testing interventions to improve the mental health and employment outcomes of individuals with mental health challenges is important for the field.

References

Adi, Y., Zeghidour, N., Collobert, R., Usunier, N., Liptchinsky, V., & Synnaeve, G. (2019). To Reverse the Gradient or Not: An Empirical Comparison of Adversarial and Multi-task Learning in Speech Recognition. *In Proceedings of the International Conference on Acoustics, Speech and Signal* Processing (pp. 3742-3746). IEEE. https://doi.org/10.1109/icassp.2019.8682468

Adweb, K. M. A., Cavus, N., & B. Sekeroglu. (2021). Cervical cancer diagnosis using very deep networks over different activation functions. *IEEE Access, 9*, 46612-46625. doi: 10.1109/ACCESS.2021.3067195

Archenaa, J., & Anita, E. M. (2015). A Survey of Big Data Analytics in Healthcare and Government. *Procedia Computer Science*, *50*, 408–413. https://doi.org/10.1016/j.procs.2015.04.021

Bartelink, V. H. M., Zay Ya, K., Guldbrandsson, K., & Bremberg, S. (2019). Unemployment among young people and mental health: A systematic review. *Scandinavian Journal of Public Health*, *48*(5), 544–558. https://doi.org/10.1177/1403494819852847

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., & Cox, D. D. (2015). Hyperopt: a Python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, *8*(1), 014008. https://doi.org/10.1088/1749-4699/8/1/014008

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bramer, M. (2013). *Principles of Data Mining, Second Edition (Undergraduate Topics in Computer Science)* (2nd ed. 2013). Springer.

Burr, H. (2021a). Monitoring trends in psychosocial and physical working conditions: Challenges and suggestions for the 21st century. *Scandinavian Journal of Work, Environment & Health*, *47*(5), 329–333. https://doi.org/10.5271/sjweh.3973

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Advances in Knowledge Discovery and Data Mining*, 160–172. https://doi.org/10.1007/978-3-642-37456-2_14

Canziani, A., Paszke, A., & Culurciello, E. (2016). An analysis of deep neural network models for practical applications. *ArXiv: Computer Vision and Pattern Recognition*. https://arxiv.org/pdf/1605.07678

Cavus, N., & Sancar, N. (2023). The importance of digital signature in sustainable businesses: A scale development study. *Sustainability, 15*(6), 5008, 2023. https://doi.org/10.3390/su15065008

Cavus, N., Mohammed, Y. B., Gital, A. Y., Bulama, M., Tukur, A. M., Mohammed, D. Tukur, A. M., Mohammed, D., Isah, M. L., & Hassan, A. (2022). Emotional artificial neural networks and gaussian process-regression-based hybrid machine-learning model for prediction of security and privacy effects on m-banking attractiveness. *Sustainability, 14*(10), 5826. https://doi.org/10.3390/su14105826

Cavus, N., Mohammed, Y. B.., Bulama, M., & Isah, M. L. (2023). Examining user verification schemes, safety and secrecy issues affecting m-banking: Systematic literature review. *SAGE Open, 13*(1), 2023. https://doi.org/10.1177/21582440231115237

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Chapelle, O., Scholkopf, B., & Zien, A. (Eds.). (2006). Semi-Supervised Learning. *MIT Press*. https://doi.org/10.7551/mitpress/9780262033589.001.0001

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, *3*(3), 243–250. https://doi.org/10.1016/s2215-0366(15)00471-x

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*(76), 2493–2537. http://pkuksa.org/~pkuksa/publications/papers/jmlr2011nlpfromscratch.pdf

Corley, M. (2019). Unemployment and mental illness survey: Exploring the causation of high unemployment among the mentally ill (Version 2). Retrieved November 10, 2022, from https://www.kaggle.com/michaelacorley/unemployment-and-mental-illness-survey.

Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, *17*(11), 1500–1509. https://doi.org/10.1038/nn.3776

Danks, D., & London, A. J. (2017). Algorithmic Bias in Autonomous Systems. *In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (pp. 4691-4697). https://doi.org/10.24963/ijcai.2017/654

Davidson, L., Bellamy, C., Guy, K., & Miller, R. (2012). Peer support among persons with severe mental illnesses: A review of evidence and experience. *World Psychiatry*, *11*(2), 123–128. https://doi.org/10.1016/j.wpsyc.2012.05.009

de Sousa, C. A. R. (2015). An overview on the Gaussian Fields and Harmonic Functions method for semi-supervised learning. In Proceedings of the *International Joint Conference on Neural Networks* (pp. 1-8). https://doi.org/10.1109/ijcnn.2015.7280491

de Vries, H., Fishta, A., Weikert, B., Rodriguez Sanchez, A., & Wegewitz, U. (2017). Determinants of sickness absence and return to work among employees with common mental disorders: A scoping review. *Journal of Occupational Rehabilitation*, *28*(3), 393–417. https://doi.org/10.1007/s10926-017-9730-1

Dehghan-nayeri, N., & Adib-Hajbaghery, M. (2011a). Effects of progressive relaxation on anxiety and quality of life in female students: A non-randomized controlled trial. *Complementary Therapies in Medicine*, *19*(4), 194–200. https://doi.org/10.1016/j.ctim.2011.06.002

Dehghan-nayeri, N., & Adib-Hajbaghery, M. (2011b). Effects of progressive relaxation on anxiety and quality of life in female students: A non-randomized controlled trial. *Complementary Therapies in Medicine*, *19*(4), 194–200. https://doi.org/10.1016/j.ctim.2011.06.002

Deleon, S. (2020). *Machine Learning Naive Bayes Base Belong To Us: Notebook Planner - 6x9 inch Daily Planner Journal, To Do List Notebook, Daily Organizer, 114 Pages*. Independently published.

Enguehard, J., O'Halloran, P., & Gholipour, A. (2019). Semi-supervised learning with deep embedded clustering for image classification and segmentation. *IEEE Access*, *7*, 11093–11104. https://doi.org/10.1109/access.2019.2891970

Farré, L., Fasani, F., & Mueller, H. (2018). Feeling useless: The effect of unemployment on mental health in the Great Recession. *IZA Journal of Labor Economics*, *7*(1). https://doi.org/10.1186/s40172-018-0068-5

Fioritti, A., Bassi, M., & de Girolamo, G. (2003). Italian psychiatry – 25 years of change. *International Psychiatry*, *1*(2), 14–17. https://doi.org/10.1192/s1749367600006470

Fioritti, A., Burns, T., Hilarion, P., van Weeghel, J., Cappa, C., Suñol, R., & Otto, E. (2014). Individual placement and support in Europe. *Psychiatric Rehabilitation Journal*, *37*(2), 123–128. https://doi.org/10.1037/prj0000065

François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning*, *11*(3–4), 219–354. https://doi.org/10.1561/2200000071

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K. J., & Tørresen, J. (2018). Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*, *51*, 1–26. https://doi.org/10.1016/j.pmcj.2018.09.003

Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data mining and analytics in the process industry: The role of machine learning. *IEEE Access*, *5*, 20590–20616. https://doi.org/10.1109/access.2017.2756872

Géron, A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. *O'Reilly Media*.

Giugiario, M., Crivelli, B., Mingrone, C., Montemagni, C., Scalese, M., Sigaudo, M., Rocca, G., & Rocca, P. (2011). Cognitive function and competitive employment in schizophrenia: Relative contribution of insight and psychopathology. *Social Psychiatry and Psychiatric Epidemiology*, *47*(4), 553–561. https://doi.org/10.1007/s00127-011-0367-7

GM, H., Gourisaria, M. K., Pandey, M., & Rautaray, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review*, *38*, 100285. https://doi.org/10.1016/j.cosrev.2020.100285

Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLOS ONE*, *11*(4), e0152173. https://doi.org/10.1371/journal.pone.0152173

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. *MIT Press EBooks*. https://dl.acm.org/citation.cfm?id=3086952

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2017). GAN（Generative Adversarial Nets). *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, *29*(5), 177–177. https://doi.org/10.3156/jsoft.29.5_177_2

Graves, A. (2012). Long short-term memory. *Studies in Computational Intelligence*, 37–45. https://doi.org/10.1007/978-3-642-24797-2_4

Graves, A. (2013). Generating Sequences with Recurrent Neural Networks. *ArXiv: Neural and Evolutionary Computing*. http://datascienceassn.org/sites/default/files/Generating%20Sequences%20With%20Recurrent%20Neural%20Networks.pdf

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2016). Correlation and variable importance in random forests. *Statistics and Computing*, *27*(3), 659–678. https://doi.org/10.1007/s11222-016-9646-1

Griep, Y., Kinnunen, U., Nätti, J., De Cuyper, N., Mauno, S., Mäkikangas, A., & De Witte, H. (2015). The effects of unemployment and perceived job insecurity: a comparison of their association with psychological and somatic complaints, self-rated health and life satisfaction. *International Archives of Occupational and Environmental Health*, *89*(1), 147–162. https://doi.org/10.1007/s00420-015-1059-5

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48. https://doi.org/10.1016/j.neucom.2015.09.116

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques. *Choice Reviews Online*, *49*(06), 49–3305. https://doi.org/10.5860/choice.49-3305

Harvey, S. B., Modini, M., Joyce, S., Milligan-Saville, J. S., Tan, L., Mykletun, A., Bryant, R. A., Christensen, H., & Mitchell, P. B. (2017). Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occupational and Environmental Medicine*, *74*(4), 301–310. https://doi.org/10.1136/oemed-2016-104015

Hastie, T., Tibshirani, R., & Friedman, J. H. (2020). The Elements of Statistical Learning: Data Mining, Inference, and Prediction 2nd Edition. *New York: Springer*.

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82–97. https://doi.org/10.1109/msp.2012.2205597

Holman, D. (2012). Job types and job quality in Europe. *Human Relations*, *66*(4), 475–502. https://doi.org/10.1177/0018726712456407

Hossain, M. Z., Akhtar, M., Ahmad, R., & Rahman, M. (2019). A dynamic K-means clustering for data mining. *Indonesian Journal of Electrical Engineering and Computer Science*, *13*(2), 521. https://doi.org/10.11591/ijeecs.v13.i2.pp521-526

Hughes, A., Kumari, M., McMunn, A., & Bartley, M. (2017). Unemployment and inflammatory markers in England, Wales and Scotland, 1998–2012: Meta-analysis of results from 12 studies. *Brain, Behavior, and Immunity*, *64*, 91–102. https://doi.org/10.1016/j.bbi.2017.03.012

International Labor Organization (ILO). (2021). *Global employment trends 2021: The role of technology in the future of work.* Retrieved November 27, 2022, from https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_795453.pdf

Ishwariya, M., & Issac, F. R. (2022). *A Big Data Standardization Technique for Data Management Application: A New Algorithm for Analyzing and Sequencing the Big Data*. LAP LAMBERT Academic Publishing.

Izadkhah, H. (2022). *Deep Learning in Bioinformatics: Techniques and Applications in Practice* (1st ed.). Academic Press.

Jain, T., Jain, A., Hada, P. S., Kumar, H., Verma, V. K., & Patni, A. (2021). Machine Learning Techniques for Prediction of Mental Health. *In Proceedings of the Third International Conference on Inventive Research in Computing Applications* (pp. 1606-1613). https://doi.org/10.1109/icirca51532.2021.9545061

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning. *Springer Texts in Statistics*. https://doi.org/10.1007/978-1-0716-1418-1

Jansson-Fröjmark, M., & Lindblom, K. (2008). A bidirectional relationship between anxiety and depression, and insomnia? A prospective study in the general population. *Journal of Psychosomatic Research*, *64*(4), 443–449. https://doi.org/10.1016/j.jpsychores.2007.10.016

Jin, H., & Mosweu, I. (2016). The societal cost of schizophrenia: A systematic review. *Pharmaco Economics*, *35*(1), 25–42. https://doi.org/10.1007/s40273-016-0444-6

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Joyce, S., Modini, M., Christensen, H., Mykletun, A., Bryant, R., Mitchell, P. B., & Harvey, S. B. (2015). Workplace interventions for common mental disorders: A systematic meta-review. *Psychological Medicine*, *46*(4), 683–697. https://doi.org/10.1017/s0033291715002408

Kalleberg, A. L. (2013). Good jobs, bad jobs : The rise of polarized and precarious employment systems in the United States, 1970s to 2000s. *Russell Sage Foundation EBooks*. http://ci.nii.ac.jp/ncid/BB07090115

Kelleher, J. D., Namee, M. B., & D'Arcy, A. (2020). *Fundamentals of Machine Learning for Predictive Data Analytics, second edition: Algorithms, Worked Examples, and Case Studies* (2nd ed.). The MIT Press.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, *62*(6), 617. https://doi.org/10.1001/archpsyc.62.6.617

Kopelovich, S. L., Monroe-DeVita, M., Buck, B. E., Brenner, C., Moser, L., Jarskog, L. F., Harker, S., & Chwastiak, L. A. (2020). Community mental health care delivery during the COVID-19 pandemic: Practical strategies for improving care for people with serious mental illness. *Community Mental Health Journal*, *57*(3), 405–415. https://doi.org/10.1007/s10597-020-00662-z

Kotsiantis, S. B. (2011). Decision trees: A recent overview. *Artificial Intelligence Review*, *39*(4), 261–283. https://doi.org/10.1007/s10462-011-9272-4

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232. https://doi.org/10.1007/s13748-016- 0094-0

Kriegeskorte, N., & Golan, T. (2019). Neural network models and deep learning. *Current Biology*, *29*(7), R231–R236. https://doi.org/10.1016/j.cub.2019.02.034

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kuchaki Rafsanjani, M., Asghari Varzaneh, Z., & Emami Chukanlo, N. (2012). A survey of hierarchical clustering algorithms. *Journal of Mathematics and Computer Science*, *5*(3), 229–240. https://doi.org/10.22436/jmcs.05.03.11

Kühn, S. (2018). World employment social outlook: Trends 2018. *ILO*. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_615594.pdf

LARRY DAVIDSON, CHYRELL, KIMBERLY GUY, REBECCA MILLER

Lawan, A. A., & Cavus, N. (2022). A clinical validity-preserving machine learning approach for behavioral assessment of autism spectrum disorder. *OBM Neurobiology, 6*(3), 1-40. doi: 10.21926/obm.neurobiol.2203138

Lawan, A. A., Cavus, N., Abdulrazak, U. I., & Tahir, S. (2023). Neural Engineering Techniques for Autism Spectrum Disorder, vol 2: Diagnosis and Clinical Analysis. In Ayman S. El-Baz, Jasjit S. Suri (Eds.). *Fundamentals of machine-learning modeling for behavioral screening and diagnosis of autism spectrum disorder. Neural Engineering Techniques for Autism Spectrum Disorder*, pp. 253-268. https://doi.org/10.1016/B978-0-12-824421-0.00020-5

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, Q., Li, M., Guo, L., & Zhang, Z. (2020). Random forests algorithm based duplicate detection in on-site programming big data environment. *Journal of Information Hiding and Privacy Protection*, *2*(4), 199–205. https://doi.org/10.32604/jihpp.2020.016299

Liang, Y., Maeda, K., Ogawa, T., & Haseyama, M. (2021). Cross-domain semi-supervised deep metric learning for image sentiment analysis. *In the Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. IEEE. https://doi.org/10.1109/icassp39728.2021.9414150

Linardatou, C., Parios, A., Varvogli, L., Chrousos, G., & Darviri, C. (2014a). An 8-week stress management program in pathological gamblers: A pilot randomized controlled trial. *Journal of Psychiatric Research*, *56*, 137–143. https://doi.org/10.1016/j.jpsychires.2014.05.013

Liu, Y. (2019). *Python Machine Learning By Example: Implement machine learning algorithms and techniques to build intelligent systems, 2nd Edition* (2nd ed.). Packt Publishing.

Lott, S. F., & Banffy, R. (2022). *Functional Python Programming: Use a functional approach to write succinct, expressive, and efficient Python code, 3rd Edition* (3rd ed.). Packt Publishing.

Ma, Y., & Guo, G. (2014a). *Support Vector Machines Applications* (2014th ed.). Springer.

Marshall, T., Goldberg, R. W., Braude, L., Dougherty, R. H., Daniels, A. S., Ghose, S. S., George, P., & Delphin-Rittmon, M. E. (2014). Supported employment: Assessing thee. *Psychiatric Services*, *65*(1), 16–23. https://doi.org/10.1176/appi.ps.201300262

Mata, J., Wenz, A., Rettig, T., Reifenscheid, M., Möhring, K., Krieger, U., Friedel, S., Fikel, M., Cornesse, C., Blom, A. G., & Naumann, E. (2021). Health behaviors and mental health during the COVID-19 pandemic: A longitudinal population-based survey in Germany. *Social Science & Medicine*, *287*, 114333. https://doi.org/10.1016/j.socscimed.2021.114333

Mayer, D., & Hollederer, A. (2022). What helps the unemployed to stay healthy? A qualitative study of coping with the negative effects of job loss. *Journal of Public Health Research*, *11*(1), https://doi.org/10.4081/jphr.2021.2492

Mehedi Shamrat, F. J., Chakraborty, S., Billah, M. M., Das, P., Muna, J. N., & Ranjan, R. (2021). A comprehensive study on pre-pruning and post-pruning methods of decision tree classification algorithm. *In the Proceedings of the 5th International Conference on Trends in Electronics and Informatics (ICOEI)*. https://doi.org/10.1109/icoei51242.2021.9452898

Miller, B. F. (2022). Making progress on mental illness. *Science*, *375*(6582), 728–728. https://doi.org/10.1126/science.abn8143

Miyato, T., Maeda, S. I., Koyama, M., & Ishii, S. (2019). Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, *41*(8), 1979–1993. https://doi.org/10.1109/tpami.2018.2858821

Modini, M., Joyce, S., Mykletun, A., Christensen, H., Bryant, R. A., Mitchell, P. B., & Harvey, S. B. (2016). The mental health benefits of employment: Results of a systematic meta-review. *Australasian Psychiatry*, *24*(4), 331–336. https://doi.org/10.1177/1039856215618523

Modini, M., Tan, L., Brinchmann, B., Wang, M. J., Killackey, E., Glozier, N., Mykletun, A., & Harvey, S. B. (2016). Supported employment for people with severe mental illness: Systematic review and meta-analysis of the international evidence. *British Journal of Psychiatry*, *209*(1), 14–22. https://doi.org/10.1192/bjp.bp.115.165092

Mokona, H., Yohannes, K., & Ayano, G. (2020). Youth unemployment and mental health: prevalence and associated factors of depression among unemployed young adults in Gedeo zone, Southern Ethiopia. *International Journal of Mental Health Systems*, *14*(1). https://doi.org/10.1186/s13033-020-00395-2

Mor, B., Garhwal, S., & Kumar, A. (2020). A systematic review of Hidden Markov Models and their applications. *Archives of Computational Methods in Engineering*, *28*(3), 1429–1448. https://doi.org/10.1007/s11831-020-09422-4

Mucci, N., Giorgi, G., Roncaioli, M., Fiz Perez, J., & Arcangeli, G. (2016). The correlation between stress and economic crisis: A systematic review. *Neuropsychiatric Disease and Treatment*, 983. https://doi.org/10.2147/ndt.s98525

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective.* Cambridge. MIT Press.

Nakagawa, A., Mitsuda, D., Sado, M., Abe, T., Fujisawa, D., Kikuchi, T., Iwashita, S., Mimura, M., & Ono, Y. (2017). Effectiveness of supplementary cognitive-behavioral therapy for pharmacotherapy-resistant depression. *The Journal of Clinical Psychiatry*, *78*(8), 1126–1135. https://doi.org/10.4088/jcp.15m10511

Nazarov, Z., Kang, D., & von Schrader, S. (2015). Employment quota system and labour market outcomes of individuals with disabilities: Empirical evidence from South Korea. *Fiscal Studies*, *36*(1), 99–126. https://doi.org/10.1111/j.1475- 5890.2015.12047.x

Norris, D. R. (2016). *Job Loss, Identity, and Mental Health* (None). Rutgers University Press.

O'Campo, P., Molnar, A., Ng, E., Renahy, E., Mitchell, C., Shankardass, K., St. John, A., Bambra, C., & Muntaner, C. (2015). Social welfare matters: A realist review of when, how, and why unemployment insurance impacts poverty and health. *Social Science & Medicine*, *132*, 88–94. https://doi.org/10.1016/j.socscimed.2015.03.025

Paul, K. I., & Moser, K. (2009). Unemployment impairs mental health: Meta-analyses. *Journal of Vocational Behavior*, *74*(3), 264–282. https://doi.org/10.1016/j.jvb.2009.01.001

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Le Centre Pour La Communication Scientifique Directe - HAL - MemSIC*.

Peiró, J. M. (2012). Comentario al comentario Réplica al trabajo de Richard R. Jones: Single case experimental designs: Strategies for studying behavior change, de M. Hersen, & D. H. Barlow]. *Análisis Y Modificación De Conducta*, *4*(7). https://doi.org/10.33776/amc.v4i7.1491

Pelzer, B., Schaffrath, S., & Vernaleken, I. (2014). Coping with unemployment: The impact of unemployment on mental health, personality, and social interaction skills. *Work*, *48*(2), 289–295. https://doi.org/10.3233/wor-131626

Peng, G. C. Y., Alber, M., Buganza Tepole, A., Cannon, W. R., De, S., Dura-Bernal, S., Garikipati, K., Karniadakis, G., Lytton, W. W., Perdikaris, P., Petzold, L., & Kuhl, E. (2020). Multiscale modeling meets machine learning: What can we learn? *Archives of Computational Methods in Engineering*, *28*(3), 1017–1037. https://doi.org/10.1007/s11831-020-09405-5

Porkodi, S., Srihari, S., & Vijayakumar, N. (2022). Talent management by predicting employee attrition using enhanced weighted forest optimization algorithm with improved random forest classifier. *International Journal of Advanced Technology and Engineering Exploration*, *9*(90), 563-582. https://doi.org/10.19101/ijatee.2021.875340

Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*, *167*, 1258–1267. https://doi.org/10.1016/j.procs.2020.03.442

Punnoose, R., & Ajit, P. (2016). Prediction of employee turnover in organizations using machine learning algorithms. *International Journal of Advanced Research in Artificial Intelligence*, *5*(9). https://doi.org/10.14569/ijarai.2016.050904

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, *2016*(1). https://doi.org/10.1186/s13634-016-0355-x

Rauber, P. E., Falcão, A. X., & Telea, A. (2016). Visualizing time-dependent data using dynamic t-SNE. *In Proceedings of the VGTC Conference on Visualization*, 73–77. IEEE. https://doi.org/10.2312/eurovisshort.20161164

Reddy, K. S. S., & Bindu, C. S. (2017). A review on density-based clustering algorithms for big data analysis. *In Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. https://doi.org/10.1109/i-smac.2017.8058322

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. *ArXiv: Machine Learning*. https://arxiv.org/pdf/1606.05386

Saravanan, R., & Sujatha, P. (2018). A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification. *In Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS)*. https://doi.org/10.1109/iccons.2018.8663155

Sathe, T. M., & C. Adamuthe, A. (2021). Comparative study of supervised algorithms for prediction of students' performance. *International Journal of Modern Education and Computer Science*, *13*(1), 1–21. https://doi.org/10.5815/ijmecs.2021.01.01

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Schmitz, H. (2011). Why are the unemployed in worse health? The causal effect of unemployment on health. *Labour Economics*, *18*(1), 71–78. https://doi.org/10.1016/j.labeco.2010.08.005

Shaik, A. B., & Srinivasan, S. (2018). A brief survey on random forest ensembles in classification model. *In Proceedings of the International Conference on Innovative Computing and Communications*, 253–260. https://doi.org/10.1007/978-981-13-2354-6_27

Shalev-Shwartz, S., & Ben-David, S. (2015). *Understanding Machine Learning: From Theory To Algorithms*. New York: Springer.

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(09), 1426–1448. https://doi.org/10.1017/s0033291719000151

Song, D., Liu, W., Zhou, T., Tao, D., & Meyer, D. A. (2015). Efficient robust conditional random fields. *IEEE Transactions on Image Processing*, *24*(10), 3124–3136. https://doi.org/10.1109/tip.2015.2438553

Sonnentag, S. (2018). The recovery paradox: Portraying the complex interplay between job stressors, lack of recovery, and poor well-being. *Research in Organizational Behavior*, *38*, 169–185. https://doi.org/10.1016/j.riob.2018.11.002

Srivastava, S., & Agarwal, S. (2020). Workplace bullying and intention to leave: A moderated mediation model of emotional exhaustion and supervisory support. *Employee Relations: The International Journal*, *42*(6), 1547–1563. https://doi.org/10.1108/er-07-2019-0293

Stadie, B. C., Levine, S., & Abbeel, P. (2015). Incentivizing exploration in reinforcement learning with deep predictive models. *ArXiv: Artificial Intelligence*. https://arxiv.org/pdf/1507.00814.pdf

Suijkerbuijk, Y. B., Schaafsma, F. G., van Mechelen, J. C., Ojajärvi, A., Corbière, M., & Anema, J. R. (2017). Interventions for obtaining and maintaining employment in adults with severe mental illness, a network meta-analysis. *Cochrane Database of Systematic Reviews*, *2017*(9). https://doi.org/10.1002/14651858.cd011867.pub2

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Neural Information Processing Systems*, *27*, 3104–3112. http://cs224d.stanford.edu/papers/seq2seq.pdf

Thieme, A., Belgrave, D., & Doherty, G. (2020a). Machine learning in mental health. *ACM Transactions on Computer-Human Interaction*, *27*(5), 1–53. https://doi.org/10.1145/3398069

Thomeer, M. B., Mudrazija, S., & Angel, J. L. (2014). How do race and hispanic ethnicity affect nursing home admission? evidence from the health and retirement study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *70*(4), 628–638. https://doi.org/10.1093/geronb/gbu114

Tsiouli, E., Pavlopoulos, V., Alexopoulos, E. C., Chrousos, G., & Darviri, C. (2014). Short-term impact of a stress management and health promotion program on perceived stress, parental stress, health locus of control, and cortisol levels in parents of children and adolescents with diabetes type 1: A pilot randomized controlled trial. *EXPLORE*, *10*(2), 88–98. https://doi.org/10.1016/j.explore.2013.12.004

Van Aerden, K., Moors, G., Levecque, K., & Vanroelen, C. (2013a). Measuring employment arrangements in the European labour force: A typological approach. *Social Indicators Research*, *116*(3), 771–791. https://doi.org/10.1007/s11205-013-0312-0

van de Velden, M., Iodice D'Enza, A., & Markos, A. (2018). Distance-based clustering of mixed data. *WIREs Computational Statistics*, *11*(3). https://doi.org/10.1002/wics.1456 van Engelen, J. E., & Hoos, H. H. (2019). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373–440. https://doi.org/10.1007/s10994-019-05855-6

van Oorschot, W. (2013). Comparative welfare state analysis with survey-based benefit: recipiency data: The 'dependent variable problem' revisited. *European Journal of Social Security*, *15*(3), 224–248. https://doi.org/10.1177/138826271301500301

van Rijn, R. M., Carlier, B. E., Schuring, M., & Burdorf, A. (2016). Work as treatment? The effectiveness of re-employment programmes for unemployed persons with severe mental health problems on health and quality of life: a systematic review and meta-analysis. *Occupational and Environmental Medicine*, *73*(4), 275–279. https://doi.org/10.1136/oemed-2015-103121

Vanroelen, C. (2019a). Employment quality: An overlooked determinant of workers' health and well-being? *Annals of Work Exposures and Health*, *63*(6), 619–623. https://doi.org/10.1093/annweh/wxz049

Venkateswarlu Naik, M., Vasumathi, D., & Siva Kumar, A. P. (2018). An enhanced unsupervised learning approach for sentiment analysis using extraction of tri-co-occurrence words phrases. *In Proceedings of the Second International Conference on Computational Intelligence and Informatics* (pp. 17–26). https://doi.org/10.1007/978-981-10-8228-3_3

Vornholt, K., Villotti, P., Muschalla, B., Bauer, J., Colella, A., Zijlstra, F., Van Ruitenbeek, G., Uitdewilligen, S., & Corbière, M. (2017). Disability and employment – overview and highlights. *European Journal of Work and Organizational Psychology*, *27*(1), 40–55. https://doi.org/10.1080/1359432x.2017.1387536

Waghorn, G., & Hielscher, E. (2014). The availability of evidence-based practices in supported employment for Australians with severe and persistent mental illness. *Australian Occupational Therapy Journal*, *62*(2), 141–144. https://doi.org/10.1111/1440- 1630.12162

Wang, G., Ye, J. C., Mueller, K., & Fessler, J. A. (2018). Image reconstruction is a new frontier of machine learning. *IEEE Transactions on Medical Imaging*, *37*(6), 1289–1296. https://doi.org/10.1109/tmi.2018.2833635

Wang, X., Zhang, Y., Ren, X., Zhang, Y., Zitnik, M., Shang, J., Langlotz, C., & Han, J. (2018). Cross-type biomedical named entity recognition with deep multi-task learning. *Bioinformatics*, *35*(10), 1745–1752. https://doi.org/10.1093/bioinformatics/bty869

Westreich, D., Lessler, J., & Funk, M. J. (2010). Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology*, *63*(8), 826–833. https://doi.org/10.1016/j.jclinepi.2009.11.020

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools And Techniques (Fourth Edition). *Morgan Kaufmann Publishers Inc. EBooks*. https://dl.acm.org/citation.cfm?id=3086818

Xiong, J., Lipsitz, O., Nasri, F., Lui, L. M., Gill, H., Phan, L., Chen-Li, D., Iacobucci, M., Ho, R., Majeed, A., & McIntyre, R. S. (2020). Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of Affective Disorders*, *277*, 55–64. https://doi.org/10.1016/j.jad.2020.08.001

Zaki, M. J., & Meira, Jr, W. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms.* Cambridge University Press. https://doi.org/10.1017/cbo9780511810114

Żbikowski, K., & Antosiuk, P. (2021). A machine learning, bias-free approach for predicting business success using Crunchbase data. *Information Processing & Management*, *58*(4), 102555. https://doi.org/10.1016/j.ipm.2021.102555

Ziegler, A., & König, I. R. (2013). Mining data with random forests: current options for real- world applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *4*(1), 55–63. https://doi.org/10.1002/widm.1114

Zulfiker, M. S., Kabir, N., Biswas, A. A., Nazneen, T., & Uddin, M. S. (2021). An in-depth analysis of machine learning approaches to predict depression. *Current*

Research in Behavioral Sciences, *2*, 100044. https://doi.org/10.1016/j.crbeha.2021.100044

APPENDICES

Appendix A
Ethics Committee Approval

ETHICAL APPROVAL DOCUMENT

Date: 11/10/2022

To the Institute of Graduate Studies,

For the thesis project entitled as “EXPLORATORY DATA ANALYSIS AND PREDICTION FOR UNEMPLOYMENT AMONG INDIVIDUALS WITH MENTAL HEALTH CHALLENGES” the researchers declare that they did not collect any data from human/animal or any other subjects. Therefore, this project does not need to go through the ethics committee evaluation.

Title: Prof. Dr.
Name Surname: Nadire Çavuş

Signature:

Role in the Research Project: Supervisor

Appendix B

Turnitin Report

Adeoluwa ATANDA-Master Thesis

by Adeoluwa Atanda

Submission date: 05-Apr-2023 10:32PM (UTC+0300)
Submission ID: 2056863275
File name: control.docx (2.94M)
Word count: 24026
Character count: 134329

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

Exploratory Data Analysis and Prediction for Unemployment Among Individuals with Mental Health Challenges

M.Sc. THESIS

Adeoluwa ATANDA

Supervisor

Prof. Dr Nadire CAVUS

Nicosia

January, 2023

Adeoluwa ATANDA-Master Thesis

ORIGINALITY REPORT

17%	11%	8%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Liverpool John Moores University Student Paper	1%
2	www.science.gov Internet Source	<1%
3	www.mdpi.com Internet Source	<1%
4	E. Sally Rogers, Kim L. MacDonald-Wilson. "Chapter 4 Vocational Capacity among Individuals with Mental Health Disabilities", Springer Science and Business Media LLC, 2011 Publication	<1%
5	Sana Yakoubi, Isao Kobayashi, Kunihiko Uemura, Mitsutoshi Nakajima, Isoda Hiroko, Marcos A. Neves. "Recent advances in delivery systems optimization using machine learning approaches", Chemical Engineering and Processing - Process Intensification, 2023 Publication	<1%
	Submitted to Aston University	

6	Student Paper	<1 %
7	fiqey.medium.com Internet Source	<1 %
8	www.ncbi.nlm.nih.gov Internet Source	<1 %
9	Submitted to Harrisburg University of Science and Technology Student Paper	<1 %
10	informatika.stei.itb.ac.id Internet Source	<1 %
11	Submitted to University of East London Student Paper	<1 %
12	researchspace.ukzn.ac.za Internet Source	<1 %
13	Carina Gonçalves, Diana Ferreira, Cristiana Neto, António Abelha, José Machado. "Prediction of Mental Illness Associated with Unemployment Using Data Mining", Procedia Computer Science, 2020 Publication	<1 %
14	ebin.pub Internet Source	<1 %
15	www.surfactants.net Internet Source	<1 %

16	Submitted to University of Bolton Student Paper	<1 %
17	Submitted to Tilburg University Student Paper	<1 %
18	docs.neu.edu.tr Internet Source	<1 %
19	github.com Internet Source	<1 %
20	analytics4all.org Internet Source	<1 %
21	Henry G. Harder. "Disability Management Approach to Job Accommodation for Mental Health Disability", Work Accommodation and Retention in Mental Health, 2011 Publication	<1 %
22	academic-accelerator.com Internet Source	<1 %
23	helpcenterai.com Internet Source	<1 %
24	www.mtome.com Internet Source	<1 %
25	Kaitlyn M. Pederson, Elisa L. Toman, Holly A. Miller. "The relationship between mental health, risk, and community supervision outcomes", Journal of Criminal Justice, 2022	<1 %

Publication

26	www.tutorialspoint.com Internet Source	<1 %
27	Dunja Demirović Bajrami, Marija Cimbaljević, Yulia A. Syromiatnikova, Marko D. Petrović, Tamara Gajić. "Feeling ready to volunteer after COVID-19? The role of psychological capital and mental health in predicting intention to continue doing volunteer tourism activities", Journal of Hospitality and Tourism Management, 2023 Publication	<1 %
28	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
29	Submitted to Southampton Solent University Student Paper	<1 %
30	Jennifer Sánchez, David A. Rosenthal, Timothy N. Tansey, Michael P. Frain, Jill L. Bezyak. "Predicting quality of life in adults with severe mental illness: Extending the International Classification of Functioning, Disability, and Health.", Rehabilitation Psychology, 2016 Publication	<1 %
31	mdpi-res.com Internet Source	<1 %

IFoA

32 Publication <1%

33 Submitted to Rochester Institute of Technology
Student Paper <1%

34 Submitted to University of Dammam
Student Paper <1%

35 Submitted to University of Sunderland
Student Paper <1%

36 Submitted to University of Greenwich
Student Paper <1%

37 www.educba.com
Internet Source <1%

38 Submitted to University of Piraeus
Student Paper <1%

39 Submitted to Edith Cowan University
Student Paper <1%

40 Jonathan H. Chan, Chenqi Li. "Learning from imbalanced COVID-19 chest X-ray (CXR) medical imaging data", Methods, 2022
Publication <1%

41 dione.lib.unipi.gr
Internet Source <1%

42 morefeetontheground.ca
Internet Source <1%

43	www.nature.com Internet Source	<1 %
44	Submitted to University of Lancaster Student Paper	<1 %
45	www.ijraset.com Internet Source	<1 %
46	Submitted to University of Teesside Student Paper	<1 %
47	aquila.usm.edu Internet Source	<1 %
48	link.springer.com Internet Source	<1 %
49	ris.utwente.nl Internet Source	<1 %
50	www.freedomtoascend.com Internet Source	<1 %
51	www.igi-global.com Internet Source	<1 %
52	Submitted to Nottingham Trent University Student Paper	<1 %
53	ipfs.io Internet Source	<1 %
54	Submitted to Issaquah School District Student Paper	<1 %

55	Submitted to Melbourne Institute of Technology Student Paper	<1 %
56	eprints.soton.ac.uk Internet Source	<1 %
57	publications.polymtl.ca Internet Source	<1 %
58	www.ijmra.us Internet Source	<1 %
59	Submitted to Queen Mary and Westfield College Student Paper	<1 %
60	observatoriolgpd.com Internet Source	<1 %
61	www.mli.gmu.edu Internet Source	<1 %
62	medtrend.org Internet Source	<1 %
63	www.coursehero.com Internet Source	<1 %
64	www.gsconlinepress.com Internet Source	<1 %
65	Linan (Frank) Zhao. "Data-Driven Approach for Predicting and Explaining the Risk of Long-	<1 %

Term Unemployment", E3S Web of Conferences, 2020
Publication

66	Submitted to University of Cape Town Student Paper	<1%
67	kiss.kstudy.com Internet Source	<1%
68	Mabrouka Bensghaier, Wided Bakari, Mahmoud Neji. "Investigating the Use of Different Recurrent Neural Networks for Natural Language Inference in Arabic", Research Square Platform LLC, 2023 Publication	<1%
69	Piyush Kumar, Rishi Chauhan, Thompson Stephan, Achyut Shankar, Sanjeev Thakur. "A Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection", 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021 Publication	<1%
70	Submitted to University of Essex Student Paper	<1%
71	www.altmetric.com Internet Source	<1%

72	Emma Robson, Geoff Waghorn, Joanne Sherring, Adrienne Morris. "Preliminary Outcomes from an Individualised Supported Education Programme Delivered by a Community Mental Health Service", British Journal of Occupational Therapy, 2010 Publication	<1 %
73	Submitted to London School of Economics and Political Science Student Paper	<1 %
74	Radwan Qasrawi, Maha Hoteit, Reema Tayyem, Khlood Bookari et al. "Machine Learning Techniques for the Identification of Risk Factors Associated with Food Insecurity among Adults in Arab countries during the COVID-19 Pandemic", Research Square Platform LLC, 2023 Publication	<1 %
75	Submitted to University of Reading Student Paper	<1 %
76	www.pubfacts.com Internet Source	<1 %
77	Submitted to Sheffield Hallam University Student Paper	<1 %
78	academic.oup.com Internet Source	<1 %

79	assets.researchsquare.com Internet Source	<1 %
80	essenceproject.uk Internet Source	<1 %
81	ir-library.ku.ac.ke Internet Source	<1 %
82	www.biorxiv.org Internet Source	<1 %
83	Submitted to Nyenrode Business Universiteit Student Paper	<1 %
84	digitalcommons.calpoly.edu Internet Source	<1 %
85	emeritus.org Internet Source	<1 %
86	www.mic.com Internet Source	<1 %
87	"Abstracts presented at the Annual Scientific Meeting of the Australian Rheumatology Association and the Rheumatology Health Professionals Association, 2005.", Internal Medicine Journal, 2005 Publication	<1 %
88	Submitted to Oxford Brookes University Student Paper	<1 %

89	Submitted to University of Ulster Student Paper	<1 %
90	arrow.tudublin.ie Internet Source	<1 %
91	innervisionnc.org Internet Source	<1 %
92	pub.tik.ee.ethz.ch Internet Source	<1 %
93	research-api.cbs.dk Internet Source	<1 %
94	www.ukessays.com Internet Source	<1 %
95	www.sciencegate.app Internet Source	<1 %
96	Ethem Utku Aktas, Cemal Yilmaz. "Automated issue assignment: results and insights from an industrial case", Empirical Software Engineering, 2020 Publication	<1 %
97	Submitted to National College of Ireland Student Paper	<1 %
98	Submitted to Universiteit van Amsterdam Student Paper	<1 %

99	Submitted to University of Wales Institute, Cardiff Student Paper	<1 %
100	Submitted to University of Wales, Lampeter Student Paper	<1 %
101	Submitted to Cranfield University Student Paper	<1 %
102	Submitted to Middlesex University Student Paper	<1 %
103	Submitted to S.P. Jain Institute of Management and Research, Mumbai Student Paper	<1 %
104	Submitted to University of Lagos Student Paper	<1 %
105	Submitted to University of Sydney Student Paper	<1 %
106	hal.laas.fr Internet Source	<1 %
107	ir.lib.uth.gr Internet Source	<1 %
108	publications.idiap.ch Internet Source	<1 %
109	www.rnzcgp.org.nz Internet Source	<1 %

110	Flávio Lopes Francisco Bittencourt, Márcio Ferreira Martins, Marcos Tadeu D. Orlando, Elson Silva Galvão. "The proof-of-concept of a novel feces destroyer latrine", Journal of Environmental Chemical Engineering, 2021 Publication	<1 %
111	Submitted to Riga Graduate School of Law Student Paper	<1 %
112	Submitted to University of Strathclyde Student Paper	<1 %
113	Submitted to University of Texas Health Science Center Student Paper	<1 %
114	hdl.handle.net Internet Source	<1 %
115	super-chain.tech Internet Source	<1 %
116	www.beds.ac.uk Internet Source	<1 %
117	Submitted to CUNY, Hunter College Student Paper	<1 %
118	Submitted to Istanbul Aydin University Student Paper	<1 %
119	Megha Chakraborty, Darius Fenner, Wei Li, Johannes Faber, Kai Zhou, Georg Rümpker,	<1 %

Horst Stoecker, Nishtha Srivastava. " CREIME —A onvolutional ecurrent Model for arthquake dentification and agnitude stimation ", Journal of Geophysical Research: Solid Earth, 2022
Publication

120	apps.dtic.mil Internet Source	<1 %
121	computing.derby.ac.uk Internet Source	<1 %
122	www.dmh.ms.gov Internet Source	<1 %
123	Jennifer Sánchez. "Employment Predictors and Outcomes of U.S. State-Federal Vocational Rehabilitation Consumers with Affective Disorders: A CHAID Analysis", Journal of Affective Disorders, 2018 Publication	<1 %
124	Submitted to University College London Student Paper	<1 %
125	developers.google.com Internet Source	<1 %
126	www.emeraldinsight.com Internet Source	<1 %
127	www.icondata.org Internet Source	<1 %

128	"Deep Reinforcement Learning", Springer Science and Business Media LLC, 2020 Publication	<1 %
129	Debra Homa, David DeLambo. "Chapter 8 Vocational Assessment and Job Placement", Springer Science and Business Media LLC, 2015 Publication	<1 %
130	Submitted to Pentecost University College Student Paper	<1 %
131	Submitted to University of Suffolk Student Paper	<1 %
132	biblio.ugent.be Internet Source	<1 %
133	bmcbioinformatics.biomedcentral.com Internet Source	<1 %
134	cdr.lib.unc.edu Internet Source	<1 %
135	core.ac.uk Internet Source	<1 %
136	easychair.org Internet Source	<1 %
137	ijece.iaescore.com Internet Source	<1 %
138	krex.k-state.edu Internet Source	

		<1%
139	mentaldisorders.neurologyconference.com Internet Source	<1%
140	norma.ncirl.ie Internet Source	<1%
141	ses.wsu.edu Internet Source	<1%
142	www.fema.gov Internet Source	<1%
143	Advances in Intelligent Systems and Computing, 2016. Publication	<1%
144	Aveyard, Helen. "Doing a Literature Review in Health and Social Care: A Practical Guide", Doing a Literature Review in Health and Social Care: A Practical Guide, 2018 Publication	<1%
145	Deepa Chinnasamy, Saraswathi Devarajan. "Machine learning", Institution of Engineering and Technology (IET), 2020 Publication	<1%
146	Evan Auguste, Molly Bowdring, Steven W. Kasparek, Jeanne McPhee, Alexandra R. Tabachnick, Irene Tung, Chardée A. Galán. "Psychology's Contributions to Anti-Blackness	<1%

in the United States Within Psychological Research, Criminal Justice, and Mental Health", Perspectives on Psychological Science, 2023
Publication

147 Kirtida Naik, Bindu Garg. "Inferential Survey on oncology information systems using predictive analytics", Research Square Platform LLC, 2022 <1%
Publication

148 Serene Lin-Stephens. "Visual stimuli in narrative-based interventions for adult anxiety: a systematic review", Anxiety, Stress, & Coping, 2020 <1%
Publication

149 Shepherd, A.J.. "Ebook: Critical Analysis in Psychology: The Essential Guide to Interrogating Published Research", Ebook: Critical Analysis in Psychology: The Essential Guide to Interrogating Published Research, 2022 <1%
Publication

150 Youngoh Jung, James Schaller, James Bellini. "Predictors of Employment Outcomes for State-Federal Vocational Rehabilitation Consumers With HIV/AIDS", Rehabilitation Counseling Bulletin, 2010 <1%
Publication

151	cmdr.ubc.ca Internet Source	<1 %
152	doc.lagout.org Internet Source	<1 %
153	dspace.library.uvic.ca Internet Source	<1 %
154	jyx.jyu.fi Internet Source	<1 %
155	mts.intechopen.com Internet Source	<1 %
156	ndl.ethernet.edu.et Internet Source	<1 %
157	par.nsf.gov Internet Source	<1 %
158	scholar.uwindsor.ca Internet Source	<1 %
159	trace.tennessee.edu Internet Source	<1 %
160	trepo.tuni.fi Internet Source	<1 %
161	weightloss.global-summit.com Internet Source	<1 %
162	www.degruyter.com Internet Source	<1 %

163	www.hopkinsmedicine.org Internet Source	<1 %
164	www.socsci.ru.nl Internet Source	<1 %
165	Ellen S. Fabian. "Longitudinal outcomes in supported employment: A survival analysis.", Rehabilitation Psychology, 1992 Publication	<1 %
166	Melissa A. Barnett, Katherine W. Paschall, Olena Kopystynska, Shannon M. Warren, Melissa A. Curran. "Pathways Linking Parental Relationship Changes, Depressive Symptoms, and Parenting Behaviors to Young Children's Development", Family Relations, 2020 Publication	<1 %
167	Asier Uribeetxebarria, Ander Castellón, Ana Aizpurua. "Optimizing Wheat Yield Prediction Integrating Data from Sentinel-1 and Sentinel-2 with CatBoost Algorithm", Remote Sensing, 2023 Publication	<1 %
168	Azadeh, A.. "Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors", Energy Conversion and Management, 200808 Publication	<1 %

169	Ivan Navarrete, Iván La Fé-Perdomo, Jorge A. Ramos-Grez, Mauricio Lopez. "Predicting the evolution of static yield stress with time of blended cement paste through a machine learning approach", Construction and Building Materials, 2023 Publication	<1 %

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		