IDENTIFYING SIGNS OF DEPRESSION IN TODAY'S WORLD THROUGH THE APPLICATION OF MACHINE LEARNING ALGORITHMS

BY

SAKIB ROKONI
ID: 191-15-12961

AND

MD TARIQUL ISLAM
ID: 191-15-12963

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Abu Kaisar Mohammad Masum
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Dr. Sheak Rashed Haider Noori
Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
2nd February 2023

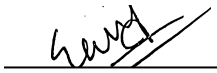
APPROVAL

This Project titled **"Identifying Signs of Depression in Today's World Through the Application of Machine Learning Algorithms"**, submitted by Sakib Rokoni ID No: 191-15-12961 and Md Tariqul Islam ID No: 191-15-12963 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 02/02/2023.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Sheak Rashed Haider Noori — **Internal Examiner**
Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Md. Sazzadur Ahamed — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Sazzadur Rahman — **External Examiner**
Associate Professor
Institute of Information Technology
Jahangirnagar University

DECLARATION

We hereby declare that, this thesis has been done by us **Sakib Rokoni Id No: 191-15-12961** and **Md Tariqul Islam Id No: 191-15-12963** under the supervision of **Abu Kaisar Mohammad Masum, Lecturer,** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

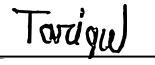
Abu Kaisar Mohammad Masum
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Dr. Sheak Rashed Haider Noori
Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Sakib Rokoni
ID: 191-15-12961
Department of Computer Science and Engineering
Daffodil International University



Md. Tariqual Islam
ID: 191-15-12963
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to Allah the Almighty for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound to our supervisors **Abu Kaisar Mohammad Masum, Lecture, Department of Computer Science and Engineering, Daffodil International University, Dhaka.** Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning"* to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan, Professor and Head, Department of Computer Science and Engineering, Daffodil International University, Dhaka.** For his kind help to finish our project and also to other faculty members and the staffs of Computer Science and Engineering Department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Depression is a serious mental disorder that may lead individuals to endure continual fluctuations in mood, in addition to sensations of melancholy. It is possible for people to have these symptoms as a result of it. It is also possible for people to isolate themselves from other people as a result of it is now well acknowledged that it is a sickness that poses a threat of mortality to people in every area of the world. At present moment, everyone, from young children to elderly adults, is feeling depression; nevertheless, the great majority of people are not even aware of how terrible their mental state actually is. This is despite the fact that everyone is now experiencing depression. Everyone has a responsibility to always be vigilant in keeping a careful watch on the state of their own mental health. As a direct result of this, we are going to make use of a technique for detection that is founded on the concept of machine learning. During the work of our inquiry, we made use of a wide variety of algorithms, some of which include the following: the Naive Bayes (NB), K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), K-Means and Random Forest (RF) approaches. For us, the results supplied by the Support Vector Machine classifier were the most accurate, and the accuracy of the SVM classifier was 97.67%. The results that we got from using this classifier were the best we could get.

.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

The average age of humans is increasing all throughout the world. The proportion of the world's population that is aged 60 and above is rising. This is the case for emerging countries as well as industrialized ones. In the year 2000, there were 607 million persons over the age of 60 living around the world; however, by the year 2015, that number had increased to 901 million. [1] The United Nations projects that by the year 2050, the global population will have grown by 2.1 billion, reaching a total of 3.1 billion people. According to the World Health Organization (WHO), a person is considered to be healthy if they are both physically fit and mentally sound.

Depression, which is also known as disappointment, is an issue that is widely seen in today's society. It is a broad and unpleasant medical illness that effects people's lives and can be found all over the globe. There are about 280 million individuals throughout the globe who are affected by depression [2]. Although depression is categorized as a mental condition, it may also have a negative impact on a person's physical health and well-being. People of all ages are adversely affected by depression to a significant degree. It is estimated that 60.80% of people throughout the world are experiencing moderate to high levels of depression [3]. It is estimated that 4.40 percent of the world's population is suffering with challenges associated to depression, which is a substantial number of individuals considering that this percentage is equivalent to the global population. According to the World Health Organization (WHO), depression is going to be a major factor for various different illnesses [4]. According to one of the studies that was included in the UNICEF study, mental illness is responsible for around $390 billion worth of annual harm to the economy, which may result in young people being disabled or even passing away [5] [6]. Depression is the most prevalent ailment in the world, affecting roughly 3.8 percent of the population [2], with 5.0 percent of those affected being adults and 5.7 percent of those affected being over 60 years of age.

In today's world, there are a lot of individuals who aren't thinking about their health. People are unconcerned about the problems with their mental health, and they are unsure how the sickness will manifest itself in their bodies. Some individuals are not interested in going to the doctor, which is one of the reasons why people of younger ages are also falling into significant ailments at that time.

There are some individuals who will seek therapy for the first stages of their mental health problems, but beyond that point, they are not serious about continuing their treatment in any capacity. They continue to take the same old drugs, and as a result, on a deeper level, they slip into a major sickness condition while being completely unaware of it. Because of the physical limitations they have, elderly

are unable to visit a medical facility, thus they never go in for their regular checkups. On the other hand, there are those individuals who are economically disadvantaged and, for financial reasons, are unable to see a physician when they need to.

As a solution to this issue, we decided to build a model that can identify the signs of depression. With that, we're through with those five depression level in the context of those inquiries. This has allowed us to gather our data and begin working with various algorithms, such as Logistic Regression, Naive Bayes, Support Vector Machine, Decision Tree, Random Forest Tree, K-Nearest Neighbors, Linear Regression, and K-Means. in order to choose the most effective classifier for practical use, we computed a wide variety of performance assessment criteria and compared the results.

1.2 Objectives

The fundamental objective of this research is to use machine learning techniques in order to develop a strategy for the early prediction of depression. The ability to make precise diagnoses and provide effective therapy is one of the most challenging aspects of contemporary medical practice. The vast majority of people in this day and age are oblivious to their own health issues. They do not have any means of knowing for certain what ailment they are dealing with since there is no method to diagnose it.

Before a doctor can make a diagnosis of a disease for a patient, the patient must undergo a number of different medical exams. People's health suffers as a direct result of the detrimental impacts of excessive pondering, and as a result, their lives are a mess. As a consequence of this, we have been working on developing a technique for the rapid diagnosis of the mental agony experienced by people. As a consequence of this, the system will first determine the state of the affected individual's health before proceeding to make a depression forecast. In addition to this The strategy might be used by regular people in order to guarantee that they are adhering to their routines. Sometimes, the sudden onset of people's mental health issues might be reason for alarm. When faced with such a scenario, people have the opportunity to simply and discretely evaluate how they feel about their own mental health. Therefore, it is feasible to immediately evaluate the degree of the patient's depression, which enables the physician to begin treatment as soon as it is possible to do so. When effective medications are made available, it also helps to minimize the costs associated with treatment. As a result, the prediction made by this approach may be useful for informing both members of the medical community and members of the general public about the mental health of individuals.

1.3 Motivation

The provision of high-quality treatment at affordable prices is one of the most significant challenges that mental healthcare organizations must face today. Because of the prohibitively expensive nature of therapy, the majority of individuals in Bangladesh do not give their mental health state any thought at this point in time. Occasionally, all they do is go to the doctor, but since checkups are so expensive, they don't take them too seriously. In such instance, they do not have any information about their genuine mental health. People in Bangladesh may also be incredibly sluggish about their day-to-day checks, which is one of the reasons why they might develop severe mental health problems without ever realizing it. Laziness and the high expense of therapy are the two main reasons why people disregard their mental disorders. After seeing a situation of this kind, we were motivated to design a system that would provide an expedited method for determining whether or not a person is depressed.

There are just a few studies that have been done in this study sector, which is an essential supplementary piece that we discovered.

And with those five stages of despair, we have completed our task. As a result, our study brings completely novel techniques to this research subject; for this reason, we believe that machine learning is a particularly fruitful area in which to make a contribution.

One other reason why depression is bad is that it may be hazardous, it can have an effect on well-being, and it can prevent a person from getting the supplements that are essential for the correct growth and progress of a human being. Following the observation of such occurrences, we pose the following significant question: "How can we transform machine approaches into a system that can forecast depression of by people's and may assist mental healthcare doctors in making better decisions?" This is the major impetus for doing this research.

1.4 Rational of the Study

The area of medicine is one that makes extensive use of machine learning technology. Everyone is struggling through their task with a sense of sadness. However, we have completed our research using five individuals whose symptoms suggested a depression degree of questionable. As a consequence of this, the work that we do will provide a novel perspective on the field of medicine. We put in our best effort and were successful in enhancing the current procedure for the creation of machine learning. It's possible that machine learning may be an interesting field all by itself. It is possible that, if combined with another exciting field of study, the resulting work of energy will be genuinely remarkable.

1.5 Research Questions

The results of our research will help machine learning techniques tackle the challenging issues they face. To get the best possible outcome, it is essential to develop a classifier that is both professional and practical. Building a bank of research questions that will be addressed after the study is complete is the first step in every successful investigation.

We've developed a list of inquiries that we want to answer with this project. They constitute,

1. Can may machine learning techniques be used in parallel for testing and evaluation?
2. To achieve the best possible outcome, how do we decide which classifier to use?
3. Will evolutionary approaches outperform conventional approaches?

1.6 Expected Outcome

There has been a meteoric rise in the use of machine learning, a subset of AI, in the modern digital world. These days, machine learning is present in just about every aspect of healthcare. As a result of our efforts, individuals of all ages will be able to feel safe and secure when receiving medical care. The following is a description of our results.

- Diverse techniques to machine learning are applied to improve speed and precision.
- Optimized machine learning outcomes will significantly improve performance.
- Repeatedly testing machine learning algorithms to choose the best classifier output.
- The accuracy of the classifier will improve.
- The human can know their body condition.
- Get proper advice for mental health.
- People of any age can do their mental health check-ups.

1.7 Layout of the Report

We explained how crucial Machine Learning is via the lens of data mining. The reason for this is also detailed here. Data mining and the use of machine learning enabled us to compile our dataset. The way we categorized our inquiry. To kick off this chapter, let's talk about the internal structure and how we arrived at that structure. In Chapter 2, we discussed the history of this system. Here are several works from the past in this field that each focus on a certain medical issue.

Even though some of the literature we've read on the topic is rather old, it has served us well in gaining insight into the tasks at hand. Some context will be provided so that you can follow along with our work.

In this paper's methodology section, we outline the complete process. Everything you need, from data collection to algorithm design, is taken care of by us. Everything is broken out into excruciating detail right here.

In the last section, we discussed the testing procedure and its outcomes. We have also discussed the general methodology and results of the exam. In addition, the outcomes were explained in depth.

Final chapter of our report is included; it describes our activities, media consumption, and plans for the future.

CHAPTER 2
BACKGROUND STUDY

2.1 Introduction

Recently, friction has increased once again in this sector. Two areas of artificial intelligence work that are seeing growth are machine learning and data mining. In this part, we will examine the relevant work that has been done in the past to help set the stage. Some context is provided here to set the stage for the investigation. When everything was said and done, many difficulties were covered.

2.2 Related Works

This sort of situation has a number of evolving problems. Therefore, in this part, we analyzed a variety of papers to see whether there is a gap in past research.

Machine learning techniques such as Random Forest Tree (RFT), Support Vector Machine (SVM), and Convolution Neural Network may be used to provide accurate forecasts of emotional states such as anxiety and depression (CNN). In order to encode the text, many different approaches, such as topic modeling, Bag-of-Words (BOW), and Term Frequency-Inverse Document Frequency, were used (TF-IDF). Python was used in the development of the study models, and in terms of accuracy (78%), as well as recall (0.72), the CNN performed better than other classifiers [7]. For the purpose of classification, machine learning strategies such as Logistic Regression, Cat Boost, Naive Bayes, RFT, and Support Vector Machines (SVM) were only few of the approaches that were used [8].

A total of 470 seafarers were interviewed to collect data on their jobs, families, and health using a set of 16 variables that included age, education, monthly income, employment status, body mass index, length of service, family makeup, marital status, history of hypertension, diabetes, or heart disease, occupational classification, organizational rank, ship type assigned, and dummy variables for academic achievement and gender.

Cat Boost was shown to have the highest accuracy and precision (82.6% and 84.1%, respectively) among all classifiers studied [7].

The diagnosis of depression by the use of medical data, such as fMRI signatures [9], depression questionnaire results [10, 11], or clinical criteria for depression as established in the DSM-5 and ICD-10 [12], is a significant field of study. The Support Vector Machine (SVM), Multilayer Perceptron (MLP), Logistic Regression (LR), Decision Tree (DT), Naive Bayes (NB), Maximum Entropy (ME), K-Nearest Neighbors (KNN), Adaptive Boosting (AB), Random Forest (RF), Gradient Boosting (GB), Bagging Predictors (BP), and various other single and ensemble models were used to develop the majority of the models [4, 9, 10, 12, 13, 14, 15, 16]

The use of social networking as a diagnostic tool for a wide variety of ailments is rapidly gaining popularity in the medical community. Saha et al. [17] selected their respondents and the psycholinguistic qualities they had based on entries made in Live Journal. After that, they were inputted into a collaborative modeling framework with the intention of classifying the mental problems that were discussed in internet forums, with a particular emphasis on depression. The current single-task learning (STL) and multi-task learning (MLT) baselines were defeated by the recommended joint modeling paradigm, and the research indicated that talks in online communities went beyond depression symptoms. There have also been applications of deep learning strategies, such as Long Short-Term Memory (LSTM) [18] and Convolutional Neural Networks (CNN) [19, 20]. In addition, a number of investigators have developed their own unique detectors [14, 21–22].

Du et al. [23] used psychiatric pressures to gather real-time data from Twitter and analyzed suicidal tweets. If you want to find tweets that include suicidal ideas, the Convolution Neural Network (CNN) has a greater accuracy (78%) than other approaches (Support Vector Machine [SVM], extra trees [ET], and others).

Reece et al. [24] investigated factors that may foreshadow the development of depression and post-traumatic stress disorder. Hidden Markov Model (HMM) was used to detect developing PTSD probability.

Nearly one-third (31.4%) of the whole group showed symptoms of depression, while nearly one-quarter (24.0%) showed symptoms of PTSD. Braithwaite et al. [25] utilized decision tree classification to examine the tweets of 135 Amazon Mechanical Turk (MTurk) employees to find indicators of suicidal intent. Accuracy in predicting future suicide rates was found to be 92%.

Both [26] and [27] discussed ensemble machine learning algorithms. [26] employed imbalance classification. It has been found that the performance of various machine learning algorithms varies depending on the context; however, there has not been a single algorithm that has been determined to be the most appropriate in all instances. Various researchers have used various machine learning algorithms to predict psychological diseases. In light of this, the present analysis made use of each of the machine learning algorithms in order to recognize depression symptoms.

TABLE 1: RELATED WORK OF DEPRESSION

Reference	Object(s)	Size of sample	Algorithms	Accuracy
[28]	Depression	1000	RF	96.00%
[29]	Depression	1000	DISVM	86.15%
[30]	Depression	35,887	CNN	66.45%
[31]	Depression	150	SVM	74.00%
[32]	Depression	10,000	NB	83.00%
[33]	Depression	10,000	MNB	86.00%
[34]	Depression	887	SVM	63.00%
[35]	A chatbot	7666	CNN	75.00%
[36]	Speech signals of persons	7356	Multi-layer perceptron	81.57%
[37]	Depression	1145	CNN	87.96%

2.3 Background Information

2.3.1 Machine Learning

Machine learning is a subfield of modeling algorithms that is always undergoing development. Its goal is to imitate human intelligence by acquiring information from its surroundings. It is well-known for its use in a variety of domains, including email filtering, computer vision, predictive analytics, medical applications, and many more. In addition, machine learning yields positive outcomes in all of these areas, which is one reason why the applications of machine learning are expanding on a daily basis. In today's world, radiation is often suggested by oncologists to cancer patients as a part of their overall course of treatment. The capacity of machine learning algorithms to generalize their learning from the immediate environment and to do invisible work will make it possible to improve the radiation practice in terms of both its safety and its efficacy, ultimately leading to improved outcomes. Therefore, machine learning plays a much larger role as an assistant in the area of medicine. The many forms of machine learning are demonstrating promising outcomes in their performance. They are extremely significant contributors to the overall improvement of the accuracy of the classifiers' job. The forms of machine learning may be broken down into three distinct phases. That is what:

The fundamental model used in machine learning is known as supervised learning. When it comes to supervised machine learning, machine learning algorithms need to be trained just like labeled data. If this isn't done, the technique won't be able to function correctly without the precise labeled data. When used appropriately, supervised learning has the potential to be an exceptionally powerful tool. Supervised machine learning algorithms will constantly aim to better themselves by discovering new forms in their environment and creating new relationships based on the data they have been given to train themselves on.

Learning on unsupervised machines has the distinct advantage of being able to work with data that has not been tagged. In this category, the training of algorithms does not need them to wait for labels to be pre-assigned.

This indicates that there is no need for human labor in this situation for training the dataset to be machine-readable. The application makes it possible for enormous data sets to function independently.

Learning without supervision enables easy data adaptation, which in turn makes it feasible to alter the structure that was previously concealed. Compared to methods for supervised learning, this one provides additional opportunities for deployment development. One of the approaches to machine learning known as reinforcement learning involves setting up a pattern of choices to be carried out at regular intervals.

The way in which people learn from experience and data in their everyday lives serves as an inspiration for the field of reinforcement learning. Therefore, each and every time, this algorithm learns via interacting with the surroundings that are around it. After then, it either bestows a favorable or an unfavorable award on the individual depending on the deed.

In the realm of research, one of the hottest trend topics right now is machine learning. All of the recently developed methods of machine learning are always searching for new development opportunities. Even for those who are well-versed in the subject, keeping up with the pace and intricacy of this discipline when it comes to newly developed methods may be challenging. Therefore, let's talk about the many approaches that may be taken to make machine learning less mysterious and to provide a learning path for those who are just getting started with the concept. The following strategies are discussed, each of which offers an overview and is based on what you are capable of building based on your knowledge and abilities about machine learning: Tracking Patterns, Classification, Association, Outlier Detection, Clustering, Regression, and Prediction are some of the topics that will be covered in this course. We will be able to learn how data mining experts better understand their carriers by using strategies and technologies that come from the intersection of database management statistics and machine learning if we are familiar with the techniques that they use, which we will be able to do if we know the techniques. What is the process behind machine learning? To begin answering this issue, we must first have an understanding of the components that make up machine learning.

The following are the three components that make up machine learning: The computational algorithm serves as the primary foundation for diagnosis; variables and qualities participate in the decision-making process; the best available information simplifies the process of understanding the system and teaches it how to learn.

Recently, things have become less complicated as a result of developments in machine learning and improvements in predictive analytics technology. In the process of creating predictions about human diseases, machine learning demonstrated tangible improvements. Machine learning typically contributes certain implementations by using the data and training their algorithms for the purpose of creating predictions, so revealing a solution vision within the context of how data mining works. In the section on predictions, machine learning has already shown a great number of successful works with the data and the application of algorithms to the area of data mining. In addition, over the course of our study, we made extensive use of several machine learning methods. After the data have been labeled, we apply the classifiers to the data in order to determine which classifier will provide the most accurate results. We are able to determine which classifier functioned the best and provided the most accurate results by using measurements. The machine learning process makes it easy to recognize patterns and trends in the most recent work in a certain sector. There is no need for the participation of people in the task that they do. As a result of ongoing research and development efforts, machine learning is rapidly becoming more prominent in the eyes of the global community. It can handle multidimensional and multi-variety data. Learning by machine results in a vast array of different applications across several fields. The use of machine learning is gaining more and more attention, and early results in the field of predictive analytics are promising. The primary justification for this is because it utilizes the data flow in order to train its algorithms, which it then combines with data mining initiatives. As a consequence of this, it has the potential to perform better than other approaches.

2.3.2 Data Mining

Data mining is the process of identifying irregularities in a collection of information in order to forecast the outcomes and establish a connection. Data mining, also known as the knowledge finding of databases, is a technique that could be used to acquire information.

In addition to that, it includes the newly developed technique of finding helpful patterns and establishing connections with a significant quantity of data. In order to conduct an analysis of the vast majority of information, this area of study brings together tools from the fields of statistics, artificial intelligence, and machine learning with the management of databases. The majority of applications for data mining may be found in the commercial sector, scientific research, and government security. Data mining is an empowered activity that uses tactics and tools to forecast future movement and obtain additional knowledge about the company sector. You can utilize data to improve earnings, lower prices, lower risks, improve customer interactions, and a great many other things if you apply a broad variety of data mining approaches. It's a method that businesses employ to transform their raw data into useful information for the company. The method of data mining and the works produced by those involved are well-known for producing satisfactory outcomes in the relevant subject. Data mining will feel more complex leader type in the next generation purpose, in addition to other variables and their relationships designed for any model, and further refinement is possible through research that will create new methods for determining which characteristics are the most interesting. The mining of data is based on three interrelated scientific fields, which are as follows:

Statistics, which is the numerical gathering of data, is now being performed for the purpose of work. presumably dealing mathematically with the analysis and version of gathering numerical data with the use of data mining approaches. There have already been a great number of productive studies completed in this data mining for statistics segment.

There are a lot of useful works that are already taking up big placements in our world that make use of artificial intelligence in combination with data mining techniques.

Within the realm of artificial intelligence, human-like intellect is mostly used to accomplish work, while machines or software are used to show such work. The purpose of the data mining task, therefore, was to have the real, solid data ready for the working procedure.

Work opportunities in machine learning and data mining processing are abundant right now. The process of machine learning involves running their algorithms in order for them to be able to learn from data and use what they have learned to make predictions. In the meanwhile, improved results from machine learning are dependent on the process of data mining. This procedure labels all of the data and provides the real data to classifiers so that they may verify their accuracy.

Data mining is the process of investigating and analyzing large blocks of data in order to collect useful patterns and structures. Data mining has a wide range of applications, including spam email filtering, the identification of fraud, the prediction of illness, database marketing, and many more. The reach of its works continues to expand exponentially. In the area of artificial intelligence, it is possible for it to be integrated with other subfields of artificial intelligence if it gives excellent results in the prediction work of other subfields. Therefore, now that we are familiar with their works in chronological order, we can observe how the process works. The process of data mining may be broken down into five distinct stages, which are as follows: The system will first gather data, and then it will pack it away into the various repositories. The information that was gathered should then be controlled and kept safe. After that, get access to the data and choose how it should be organized. The utilization program then checks the user's input and arranges the data according to the user's preferences. Finally, the end-user provides the information in a format that is simpler to distribute.

The number of jobs that combine the processes of data mining and machine learning is steadily growing. The combined efforts of machine learning and data mining are yielding a significant amount of successful payoff in the globe. In this research work, we have also done our work with data mining and machine learning, where data mining plays a critical part to raise our accuracy of the job.

In this research work, we have also done our work with data mining and machine learning. From the very beginning, the first step in data mining is to gather all of the data, and then the data are stored. After that, they labelled the data in order to ensure that they had the real proper data to apply for the purposes of testing. The same as in our study work, we have gathered a total of 503 different pieces of data to utilize in this work. These data are used to achieve this task. After that, some of the data is used for the training of the classifier, while the remaining data is put to use for testing reasons. Therefore, the sorted data are sent to the classifier in order to ensure the highest possible level of accuracy in the task. The significance of data mining is expanding in many areas, including the development of forecasts, market choices, and others. Nowadays because of rapidly expanding data businesses need ever more data to support their operations, a growth in that industry's overall demand for data analysts is a direct result of this trend.

With the assistance of data mining, we are able to conduct an analysis of the reactions and insights provided by our company clients' consumers' behaviors. This results in an outstanding accomplishment and propels the data types industry forward by one step. Its primary use is in the prediction type of approach, which requires the data mining process to do a wide variety of laborious tasks. Data mining has also been a very successful endeavor in the medical industry, where its applications are widespread. Even after gaining such a substantial amount of knowledge about data mining, you may still find yourself wondering, "Why is data mining so important?" In order to make this more clear, we will place some emphasis on that goal. We are able to filter through all of your information, which is jumbled and creates a repetitive cacophony, by using data mining. Determine which of these are relevant, and then choose the right facts in order to evaluate which outcomes are most likely to occur. Reduce the amount of time needed to arrive at informed and precise conclusions. The effects of data mining are becoming more widespread on a daily basis, and as a result, it will be extremely useful in their future endeavors. If this is the case, then the outcomes of data mining will be beneficial if it will investigate new niches in global business-related domains, and if advertising will target prospective buyers of new choices.

We use these tools to locate the greatest bargain on lawnmowers as well as the most affordable airline and the phone number of a long-lost classmate.

In addition, we use these resources to discover the most affordable flight and the phone number of a long-lost classmate. Assume that intelligent beings have lost their grasp on the facts that pertains to medical research and atomic particle physics. The use of computers may lead to the discovery of innovative treatments for illnesses as well as new perspectives on the nature of the cosmos. The explanation makes it abundantly evident that data mining is achieving an ever-increasing level of success in its most recent endeavors by making use of innovative approaches.

2.4 Challenges

Coming up with questions to ask was the most challenging aspect of this project. For the purpose of the study, information was gathered with the use of the PHQ-9 (Depression Questionnaire). The Patient Health Questionnaire, sometimes known as the PHQ-9, is a set of nine items that measures several elements of clinical depression.

The PHQ-9 is the depression module, and it assigns a point value from 0 (not at all) to 3 (very much) to each of the nine DSM-IV criteria (nearly every day). It has been shown to be safe and effective for usage in primary care settings. Although it is not a screening test for depression, it is used in determining the degree of depression as well as the patient's reaction to therapy.

In order to compile the dataset, we polled 1150 individuals and collected their responses to a series of questions. There are 650 individuals who do not suffer from depression, whereas there are 500 people who do. In the course of our endeavor, we have encountered a significant number of challenges. Following the collection of everyone's information, the data set underwent certain modifications. In order to ensure that the findings of our experiment are accurate, we will need to exercise extreme caution.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

Numerous well-known research organizations have begun the process of developing improved methods of machine learning and data mining. The next step in their process was to incorporate these findings into their research works. Combining machine learning with data mining as a method for categorization is now seeing a surge in popularity. Predictive analytics is the evolutionary algorithm that is most often used in conjunction with machine learning. Other types of evolutionary algorithms are described.

3.2 Research Subject and Instrumentation

The primary subject of this research is an online forum, which was analyzed and ranked according to the amount of free-flowing, thought-provoking information it offers. As a direct consequence of this, we decided to concentrate our efforts for the whole of our study on both data mining and machine learning. We gathered information not just for the purpose of operating the company, but also for the purpose of developing models, gathering data, and training the model. This meant that we were collecting information for many purposes at the same time. The use of cutting-edge machinery and the development of unique procedures by our own team are the two additional aspects that make up this system. We used the Windows platform, the Python programming language, and a wide range of extra libraries like Pandas, NumPy, Seaborn, and a great number of others. We were successful in all of the training and testing procedures that necessitated the use of the operating system in which we deployed Anaconda and Google Collab, in addition to the utilization of extra tools such as weka and other application programs. The following sequence, which is open source and may be used in Python applications that apply machine learning, is often used in these sorts of projects.

3.3 Workflow

Primary 4 Steps:

- Data Collection
- Data Preprocessing
- Implement algorithm
- Performance Evaluation

3.4 Implementation Procedure

The aim of this effort is to develop a method for predicting depressive episodes. In order to make an accurate prognosis, a number of crucial criteria, particularly depressive symptoms, are taken into consideration. The several steps that we went through to carry out this project are shown in Figure 3.4.1. To begin, a questionnaire consisting of nine questions and centered on depressive symptoms has been developed. After then, with the assistance of this survey, we were able to collect responses from a very large number of individuals. In the next step, we fed this data into the classifier by using a few different preparation processes.

Following the completion of any necessary preprocessing, our data is then split into a training set and a testing set. In this particular example, the training process made use of seventy percent of the whole data set. Only for purposes of testing was the remaining thirty percent of the data set used. This is a totally arbitrary method of division. After that, the classifiers were trained with the help of the training data. After training the classifiers, we used the testing data to make a prediction about the patient's current state of depression. This section has been used to calculate a few of the metrics that go into the performance assessment.

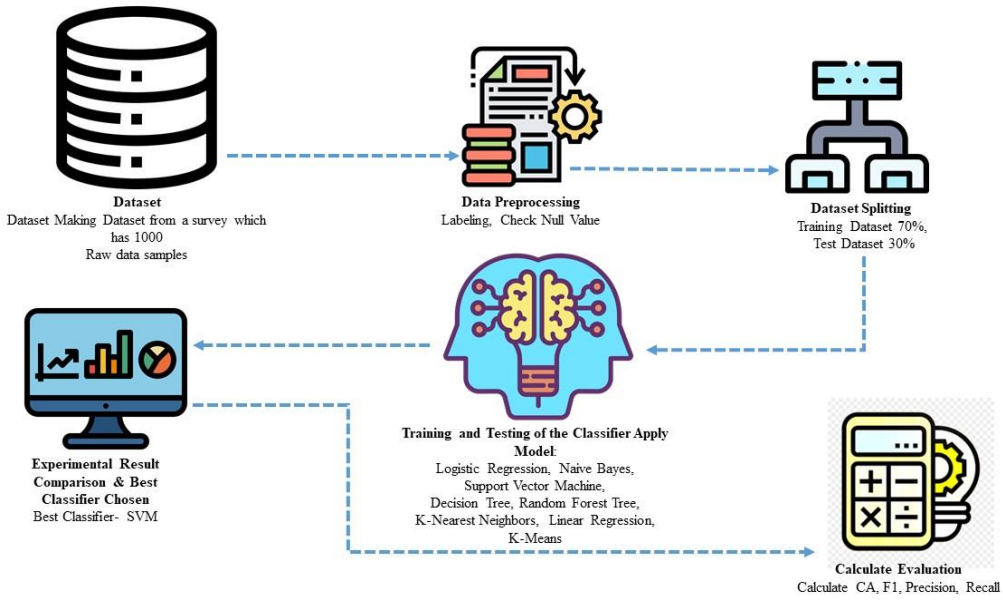


Figure 3.4.1: Working Procedure

Using these variables, we were able to determine the optimum classifier to use for making predictions in this setting. Using the equations that are provided below, a number of performance metrics expressed as a percentage have been computed based on the confusion matrix that was generated by the classifier.

$$\text{TotalData} = \frac{\text{TrueDepressed} + \text{TrueNondepressed}}{\text{Total Data}} \times 100$$

$$\text{Recall} = \frac{\text{TrueDepressed}}{(\text{TrueDepressed} + \text{FalseNonDepressed})} \times 100$$

$$\text{Precision} = \frac{\text{TrueDepressed}}{(\text{TrueDepressed} + \text{FalseNonDepressed})} \times 100$$

$$\text{Specificity} = \frac{\text{TrueNonDepressed}}{\text{FalseDepressed} + \text{TrueNondepressed})} \times 100$$

$$\text{F1} - \text{Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \times 100$$

3.5 Implementation Requirements

- Hardware:
    - Processor: Core i5
    - Ram: 16GB ▪ GPU: GTX 1660
- Software:
    - Google Colab, Anaconda, Weka, Orange
    - Language: Python 3.8
    - Libraries: Numpy, seaborn, pandas, sklearn, matplotlib, label encoder.

3.6 Training the Model

In general, after the completion of the preprocessing step, we divided our data for the task into a training set and a testing set. In this particular case, the data set was used for training purposes, which accounted for 70% of the total. For the purposes of testing, the remaining 30% of the whole data set was used.

We avoided using any data that was either unnecessary or insufficient by just using the data that was required for our work. In addition, we use a wide range of data mining methods to the information included in this collection. After then, the best classifier was found by training it with the data that had been obtained.

3.7 Data Description and Analysis

The dataset that is being utilized in this study was acquired via the use of a survey. The majority of the survey is made up of 9 questions. The symptoms of the depression that we are dealing with served as the inspiration for these questions that we have devised. And the answers to each question were derived and examined based on the primary symptoms of depression. These nine questions take into account aspects related to the individual as well as level of depression.

The Questionnaires on depression and the Possible values are detailed in Table 2. A total of one thousand different records had to be utilized in order to accomplish this assignment. Seventy percent of the data is employed during the training process of the classifier, while the remaining thirty percent is utilized during testing.

TABLE 2: QUESTIONNARIES ON DEPRESSION AND THE POSSIBLE VALUES

Questionnaires on depression	Possible value
1. Little interest or pleasure in doing things	Not at all [0] Several days [1] More than half the days [2] Nearly every day [3]
2. Feeling down, depressed, or hopeless	Not at all [0] Several days [1] More than half the days [2] Nearly every day [3]
3. Trouble falling or staying asleep, or sleeping too much	Not at all [0] Several days [1] More than half the days [2] Nearly every day [3]
4. Feeling tired or having little energy	Not at all [0] Several days [1] More than half the days [2] Nearly every day [3]
5. Poor appetite or overeating	Not at all [0] Several days [1] More than half the days [2] Nearly every day [3]
6. Feeling bad about yourself - or that you are a failure or have let yourself or your family down	Not at all [0] Several days [1] More than half the days [2] Nearly every day [3]

7. Trouble concentrating on things, such as reading the newspaper or watching television	Not at all [0] Several days [1] More than half the days [2] Nearly every day [3]
8. Moving or speaking so slowly that other people could have noticed Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual	Not at all [0] Several days [1] More than half the days [2] Nearly every day [3]
9. Thoughts that you would be better off dead, or hurting yourself in some way	Not at all [0] Several days [1] More than half the days [2] Nearly every day [3]

In the end, each participant's response was given a numeric value between 0 to 3, and the total score was calculated by adding together the values for each question. Results were classified as None, Mild, Moderate, Moderately Severe and Severe once the final severity level was obtained. Table 3 describe the severity levels of depression.

TABLE 3: SEVERITY LEVELS OF DEPRESSION

Depression	
None	0 - 4
Mild	5 - 9
Moderate	10 - 14
Moderately Severe	15 - 19
Severe	20 - 27

We have worked on eight different classes of individuals, including male and female persons (*Male = 0 and Female = 1*), where we have taken each class and stored it in a variable while also utilizing encoder labels to number these people. Label encoding refers to the process of converting the names of dataset columns into numerical values. so that the machine can understand what it is reading.

And computers that use machine learning may come to more informed conclusions about how to deal with such labels. In supervised learning, this phase of the pre-processing procedure is very critical for structured datasets. The class name and encoder label are shown in Table 4.

TABLE 4: CLASS NAME AND THEIR ENCODEER LEVEL

Class	Label Encoder
Student	0
Service Holder	1
Unemployed	2
Business	3
Housewife	4
Teacher	5
Doctor	6
Medical Students	7

3.8 Classifier Description

3.8.1 Logistic Regression

Both categorization and predictive analytics benefit from the use of logit models. The likelihood of an event, such as voting or not voting, is calculated using logistic regression. As the outcome is a probability, the dependent variable can take on values between

0 and 1. The odds (probability of success/probability of failure) are logit transformed in logistic regression.

Logit(pi) is the dependent or response variable in this logistic regression equation. The beta parameter is computed using greatest likelihood (MLE). This approach iteratively examines beta values to optimize log odds.

Logistic regression maximizes the log likelihood function to get the optimal parameter estimate. Once the ideal coefficient (or coefficients, if there are several independent variables) is discovered, conditional probabilities may be computed, recorded, and averaged to provide a forecast probability. Less than.5 predicts 0 and larger than 0 predicts 1. After computing a model, it's best to assess its goodness of fit, or ability to forecast the dependent variable. Hosmer–Lemeshow is a prominent model-fitting test.

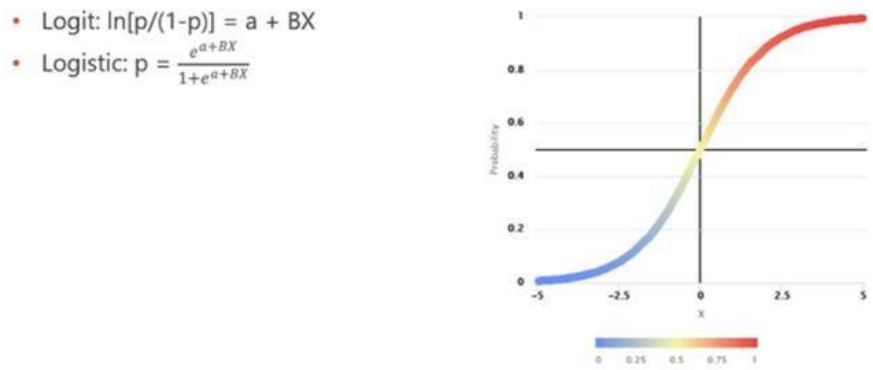


Figure 3.8.1 Logistic Regression Diagram

3.8.2 Naive Bayes

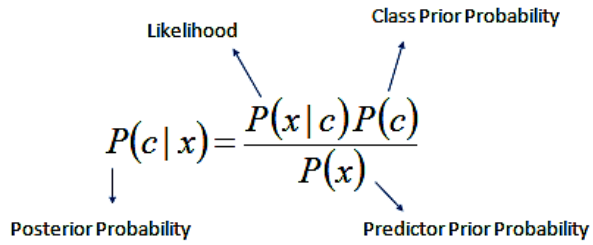
The Ignorant Naive and Bayes are the two terms that make up the Bayes algorithm. These two words may be defined as follows:

The term "naive" refers to the assumption that the presence of one aspect is unrelated to the occurrence of any other features. This is the reason why it is termed "naive."

For instance, if the fruit is identified based on its color, shape, and flavor, then a fruit that is red in color, round in form, and has a sweet taste is known to be an apple.

Therefore, the existence of any one quality is sufficient evidence to conclude that the item under consideration is an apple, and the presence of no attribute is necessary.

Bayes: This method is given the name Bayes due to the fact that it is based on the Bayes' Theorem premise.



$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Figure 3.8.2 Naive Bayes Diagram

3.8.3 Support Vector Machine

The "Support Vector Machine" (SVM) is a method for classification and regression that is used in supervised machine learning. On the other hand, it is often used in the context of solving classification issues [38]. The value of each feature is determined by the score that the SVM classifier achieves at a particular location, and every data point is represented as a point in an n-dimensional space (where n is the number of characteristics that we have). In order to choose the greatest number of nodes that will contribute to the formation of the hyperplane, a Support Vector Machine is used. A Support Vector Machine is the name of the method, and the support vectors are sometimes referred to as the maximum instances. Take a look at the image below to see how a decision hyperplane may be used to differentiate between two distinct groups of people. Next, in order to finish the identification process, we locate the hyperplane that most effectively differentiates the class labels (look at the below Fig. 3.8.3).

To put it another way, support vectors may also be thought of as the locations of each accuracy assessment. The Classification algorithm is a frontier that most successfully divides the two categories (hyper-plane and line).

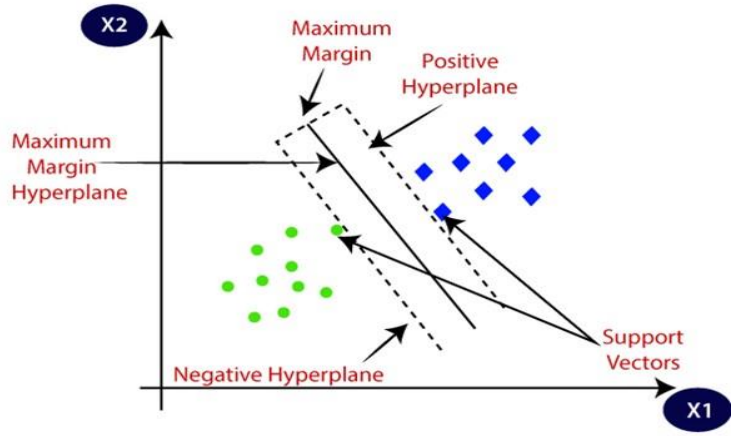


Figure 3.8.3 Support Vector Machine Diagram

3.8.4 Decision Tree

In machine learning, the process of classifying data is broken down into two stages: the learning stage, and the prediction stage. During the phase known as "learning," the model is formed by making use of the available training data. During the stage of prediction, the model is utilized to make projections about the response by basing those projections on the data that is supplied. The Decision Tree approach is one of the most used categorization systems because it is straightforward, easy to understand, and straightforward to interpret. The Decision Tree approach is an example of supervised learning, which is a category that exists within the area of machine learning. In contrast to other supervised learning approaches, the decision tree methodology may be used to problems involving both regression and classification. A Decision Tree is a kind of training model that is used to predict the class or value of the target variable by making use of fundamental decision rules that have been generated from previously collected data (training data).

When using Decision Trees to find the most probable category that a record falls into, we start at the very top of the tree. A comparison is made between the attribute values of the

root element and those of the record. When we compare two values, we pick the path that takes us to the next node in the tree depending on the difference between them.

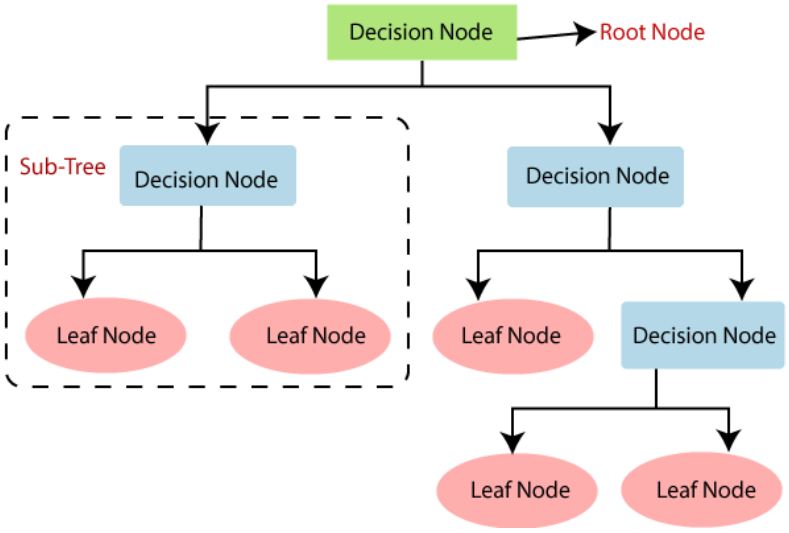


Figure 3.8.4 Decision Tree Diagram

3.8.5 Random Forest

For classification purposes, the Random Forest classifier is a supervised learning system that divides random trees into forest types. Classification and regression issues are both within its scope of application. It's a common strategy for addressing problems of categorization. It selects data points at random from a given collection of records. Decision trees are created from data samples, and these trees are then utilized to make predictions [39]. Then, they use a voting system to decide which option is best. The trees that are constructed with the random forest are more randomly generated. It looks for the best attribute from a given distribution, rather than the most important one, when splitting a node. This leads to a lot of variability, which ultimately improves the model. Random Forest performs better than a single decision tree since it is an ensemble learning strategy. When findings are averaged, the overfitting issue is mitigated.

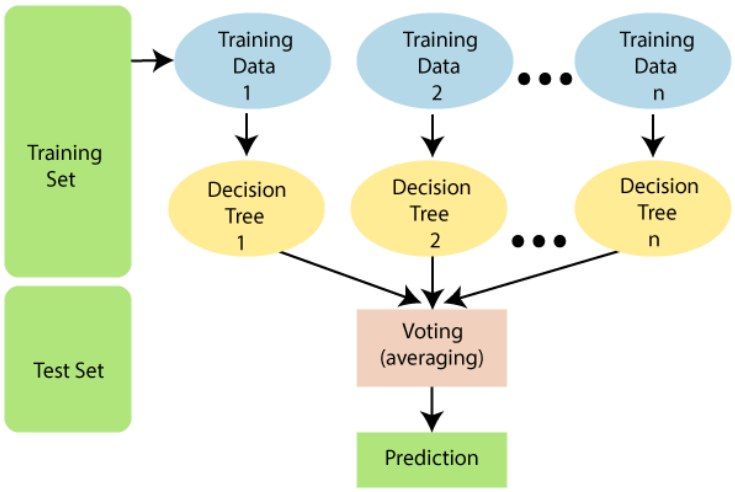


Figure 3.8.5 The Random Forest Diagram

3.8.6 K- Nearest Neighbor

One of the earliest and most basic approaches to machine learning is called the K-Nearest Neighbor method. It is based on an instructional methodology known as Supervised Learning. The KNN technique is used to solve difficulties relating to classification as well as regression. The key to success with the KNN approach is accurate feature matching. K Nearest Neighbor (KNN) is a technique to machine learning that is straightforward, easy to comprehend, and flexible.

Handwriting recognition, finance, political science, healthcare, image recognition, and video recognitions are just few of the fields that may make use of KNN's capabilities. Financial institutions make use of credit ratings in order to make educated guesses on the credit rating of their customers [40].

The KNN approach presupposes that the current particular instance and the former cases are similar, and it allocates the new case to the category that is the closest to the classifications that were previously used. KNN stands for the number of neighbors that are the most immediate. The amount of people that live nearby is the single most important consideration. When there are two courses, the number K is often an odd number. The procedure is referred to as the closest neighbor algorithm when K=1, the default value.

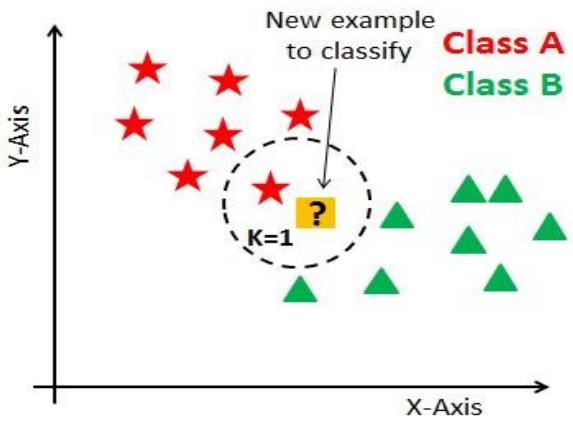


Figure 3.8.6 K- Nearest Neighbor Diagram

3.8.7 Linear Regression

Linear regression is a basic statistical approach used to predict the connection between continuous variables. Linear regression displays the linear connection between independent (X-axis) and dependent (Y-axis) variables. Simple linear regression has one input variable (x). Multiple linear regression involves more than one input variable. The linear regression model describes variable relationships using a slanted straight line. The graph shows linear relationships between dependent and independent variables. Increasing x raises y (dependent variable). Best-fit straight line is the red line. We draw a line that best fits the provided data points.

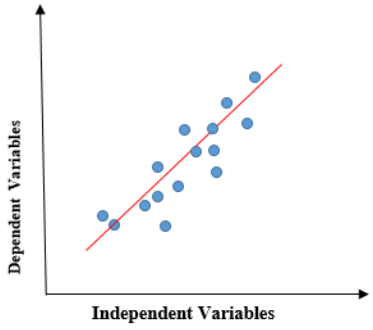


Figure 3.8.7 Linear Regression Diagram

3.8.8 K-Means

K-Means Clustering is an unsupervised learning approach used to solve clustering problems in machine learning and data science. Learn about the K-means clustering algorithm, how it works, and how it's implemented in Python within the scope of this article [41].

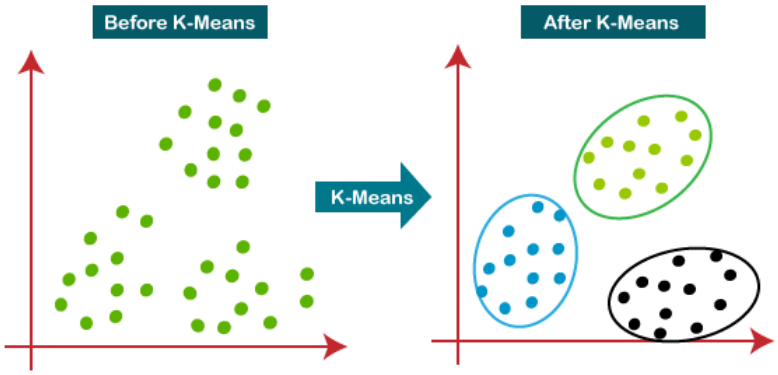


Figure 3.8.8 K-Means Diagram

When applied to an unlabeled dataset, the K-Means Clustering method creates distinct clusters. For this procedure, K specifies the number of clusters that must be formed in advance; for example, if K=2, only two clusters will be created, if K=3, only three, and so on. It provides a simple approach to discovering the categories of groups in the unlabeled dataset on its own, without the requirement for training, and enables us to cluster the data into distinct groups. Each cluster has a centroid that the algorithm uses to make decisions. This technique primarily works to reduce the total distance between individual data points and their respective clusters.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

In this part, we plan to describe the results that were obtained from the overall test. Using a number of different approaches to estimating, we will investigate how to assess execution and then examine the outcomes that follow from it.

4.2 Performance Evaluation

Due to the fact that this is largely an issue with several classes, the classifier produced a confusion matrix. In order to provide an accurate assessment of this work, we derive Accuracy, F1-score, Precision, and Recall from the Confusion Matrix. The findings of a wide variety of performance assessment criteria are shown in Table 5. Table 5 demonstrates, after doing an in-depth analysis of the findings, that the performance of the classifier is superior than that of the other seven classifiers. For all classes 81.16, 97.68, 91.01, 93.04, 88.7, 92.17, and 12.75 this classifier has the best accuracy of any and all classifiers. Using the Naive Bayes, Support Vector Machine, Decision Tree, Random Forest, K-Nearest Neighbor, and Logistic Regression algorithms, along with the K-Means method, we were able to determine that Support Vector Machine was the most effective classifier algorithm. The following is in the outcome:

TABLE 5: COMPARISON OF SEVEN CLASSIFIER'S PERFORMANCE

Classifier	Accuracy	Precision	Recall	F1 Score
Naive Bayes	81.16%	0.89	0.93	0.93
Support Vector Machine	97.68%	1.00	1.00	1.00
Decision Tree	91.01%	0.99	0.97	0.98
Random Forest	93.04%	1.00	0.98	0.99
K- Nearest Neighbor	88.7%	0.99	0.94	0.97

Logistic Regression	92.17%	0.99	0.98	0.99
K-Means	12.75%	0.20	0.53	0.29

We have used seven different classifiers on our dataset, and the results of these classifications are displayed in Table 5 as Accuracy, F1-Score, Precision, and Recall for each individual classifier.

4.3 Result Discussion

Because this is mainly a multiclass issue, the classifier produced a confusion matrix for the problem. The generated matrices for each of the classifiers are shown in Table 4. In order to assess the quality of this work, we compute Accuracy, F1-Score, Precision, and Recall using the confusion matrix that was shown before. Table 5 demonstrates, after looking at the findings as a whole, that the classifier has superior performance than the other seven classifiers. This classifier has the greatest accuracy of all the classifiers for all classes, with scores of 81.16, 97.68, 91.01, 93.04, 88.7, 92.17, and 12.75 respectively. In order to test the efficacy of the classification algorithm, we looked at a wide variety of quality evaluation criteria. We made the discovery that our classifier is superior to all other data mining techniques. Following the application of the Naive Bayes, Support Vector Machine, Decision Tree, Random Forest, K-Nearest Neighbor, and Logistic Regression algorithms, we came to the conclusion that Support Vector Machine was the best classifier approach.

CHAPTER 5
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

A typical mental illness is a depression. An estimated 5% of individuals worldwide are thought to experience depression. A substantial component of the overall global illness burden is depression, a primary cause of disability globally. Close friends and family may not always notice that a person is depressed, unlike with other illnesses, and even when they do, many are unsure of how to assist. Some would deny there is a problem in response to their loved one's increasingly depressed demeanor. Others could stigmatize the illness and view it as a sign of weakness, while others might feel responsible and support the patient too vehemently. Regardless of how someone decides to respond, the additional uncertainty, strain, and/or irritation may seriously damage even the healthiest of relationships. Even though the effects of the pandemic are not always evident, it is having a devastating impact on society as a whole. The disease may have the most obvious implications for those who suffer from depression and those who are close to them. This illness has probably had more of an influence on societies throughout the world than may be immediately apparent, from early school dismissal to detrimental effects on child development to increased healthcare expenses-for example: When compared to other young people, adolescents with depression have a 60% greater probability of dropping out of secondary school. There is a substantial correlation between teenage depression and early termination of schooling. Teenage pregnancy has increased due to depression in teenagers, and studies have shown that these moms are three times more likely to get government assistance paid for by tax dollars. An alarming and depressing statistic for many reasons, as well as a horrible loss of skill and production, is that suicide is the second-leading cause of death among teenagers between the ages of 15 and 19. Depression is a key risk factor for suicide.

5.2 Impact on the Environment

In today's world, people face a wide range of challenges. Those who suffer setbacks, such as losing a career, a loved one, or their financial stability, sometimes spiral into depression. Those affected by this problem could become frustrated or feel like they don't matter. When individuals are furious, they create problems for other people, which has a negative impact on the environment. Those who have a low self-esteem have taken the tragic decision to end their lives. This book made matters worse by causing pain to the loved ones of unhappy people who believed, like the author, that suicide was their only alternative. To create a more upbeat atmosphere, people should reach out to those who seem down and offer to assist them.

5.3 Ethical Aspects

Depression, which is the main cause of disability, has a catastrophic impact on an individual's quality of life and imposes a substantial financial and social cost on society as a whole. Depression is a significant factor in the pandemic of suicide that claims the life of one American every 13 minutes. The clinical treatment of depression combines the urgency of a crisis with the intensely felt and pervasive feelings of sadness and hopelessness that are common among patients. This combination gives rise to a number of ethical concerns regarding the need for patient safety, the appropriate management of disease, and the restoration of individual self-agency. Autonomy, Beneficence, Justice, and Non-Maleficence are the Four Primary Ethical Principles Autonomy, Beneficence, Justice, and Non-Male Every patient has an unalienable right, as a matter of course, to exercise their autonomy in making choices that are considerate of their own values and convictions. Children need to be taught how to think ethically so that they do not grow up and do things that are immorally harmful to society.

5.4 Sustainability Plan

Because preventing depression is the most effective method for dealing with the condition, it is essential that all programs be developed with a preventive approach in mind.

The need of prevention and early intervention cannot be overstated, particularly for younger and more vulnerable groups who are difficult to reach via traditional medical treatment. However, in order to increase the number of people who get the appropriate therapy, there has to be more clarity on what early intervention truly includes when it is put into reality. Treatment for depression should focus heavily on reducing the risk of suicidal thoughts and behaviors. Because there are so many different factors that might lead to someone taking their own life, it is imperative that efforts be made to prevent suicide using an interdisciplinary approach. This must be reflected in the joint responsibility and obligation of the departments. People have a tendency to detest the phrase "mental illness," and since there has been a history of discrimination around mental health, they may be hesitant to seek treatment. By more effectively integrating mental health therapies with other types of services, it may be able to reach a greater number of patients. One other method for reducing the stigma connected with mental health services is to refer to them using a different language, such as "wellbeing," instead of the traditional term. With the use of digital health technologies, people's treatment choices may be enhanced, and these tools also have the potential to inspire more patient autonomy. On the other hand, the use of these tools could be met with resistance from a subset of patients and medical professionals who are unclear how to integrate them into their treatment plans. Better training of medical professionals is essential in order to implement blended care, which is a combination of traditional and digital health care services. Utilizing language in treatment that is derived from patients' own experiences in receiving care is a method that is very successful. Words like "rehabilitation" are very necessary in order to convey an air of hope and make the treatment's goals abundantly obvious. It is essential to encourage medical professionals who deal with depressed people to communicate in a language that can be comprehended by everyone and that is relevant to the persons' own life experiences. It will be easier to ensure that services are successful for their users if service users and caregivers are involved in the establishment of services and rules. For the purpose of fostering empowerment in others, services should also make use of the knowledge and abilities of individuals who have prevailed over challenges related to mental health.

One example of this would be peer support. The individual who is depressed is not the only member of the family who may be affected by depression; thus, therapy should concentrate on the whole family rather than just the one who is sad. It is extremely important to provide careers with services that may aid them in order to protect their health and welfare. The dangerously high percentage of depression among young people is continuing to rise.

Young people need our help to make better use of their own strengths and capabilities so that they may become more resilient and cope better with pressure. If young people who are depressed utilize digital tools to encourage themselves to obtain help, they won't have to feel as if they are alone in their struggle. Depression is one of the primary factors that contributes to poor performance at work, which is where most adults spend the bulk of their life. It is necessary to increase the number of workplace education programs in order to cultivate a healthy culture in the workplace, assist management in recognizing the signs and symptoms of depression, and provide support for individuals who are working through their depression or returning to work after a break. Real-world data is limited in one of the most important areas of study on depression. Increasing our spending on real-world data collection on the epidemiology of depression, service consumption, and the effectiveness of treatments and services is necessary if we are to improve our understanding of the present care gaps and bring about changes that are long-lasting.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FURTHER RESEARCH

6.1 Summary of the Study

This role mostly involves making a diagnosis based on an individual's symptoms and determining what condition they have. Several different techniques to machine learning are used in order to successfully attain this goal. The primary objective of this research is to develop an effective method for identifying depressive symptoms in humans via the use of machine learning and data mining techniques. In order to do this challenge, you will need to train the classifier using a total of 70% of the data and then test it using the remaining 30%. In order to assess the performance of the working classifier, we explored a wide range of performance evaluation criteria. We were surprised to see that the classifier outperforms every other data mining technique. And the Support Vector Machine turned out to be the best classifier we could find. The result from the seven classifiers may also be rather satisfactory. The K-Means classifier gave us a number that was the lowest possible, and it was for Accuracy (12.75%), F1 (0.29%), Precision (0.20), and Recall (0.53) This means that our score was below 50%. Therefore, we believe that the results of our study provide a superior output for the condition checking of human depression identification.

6.2 Conclusion

The majority of the work in this profession is diagnosing a patient based on their symptoms and establishing what illness the patient is suffering from. In order to accomplish this objective in a timely and efficient manner, many distinct approaches to machine learning are used. Through the use of machine learning and data mining strategies, the fundamental purpose of this investigation is to develop an efficient approach for recognizing the presence of depression symptoms in human subjects. In this investigation, seven distinct classifiers are used. The fact that the classifier performed better than any other data mining approach came as a complete surprise to us. And as it turned out, the Support Vector Machine was the most effective classifier that we could locate.

6.3 Implication for Future Study

In this project, we acquire and work with 1150 unique datasets, representing the raw data. And use various classifiers on those datasets. Seven different classifiers are used in our study. After we finish this project, we hope it will help people learn more about depression and lead to more research on the subject. If you want to find signs of depression, you should always look for more output levels. We couldn't get information from a wide range of people because of this. The sample size is also low, and the dataset is small.

In future we will give more efforts to collect data focus on adding more layers of categorization to datasets, it may be possible to get a bigger and more reliable set of data that we may use in future additional classifiers and data mining algorithms to a larger dataset, or get a deeper understanding of people's mental health so that we may better assist them or refer them to a psychologist. Finally, we can promise that in the future, we will keep at it until this is the greatest job possible and we will make a website that makes it easy to spot signs of depression will be built. The ability of a machine learning model to spot signs of depression could help people from all walks of life.

REFERENCES

[1] United Nations, "World Population Ageing," 2015. [Online]. Available at <<https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf>> last accessed on 22-10-22 at 10:00 PM

[2] S. Smys, *Computer Networks and Inventive Communication Technologies: Proceedings of Fifth ICCNCT 2022*. Springer Nature, 2022. Accessed: Jan. 05, 2023. [Online]. Available: https://books.google.com.bd/books?id=v_WVEAAAQBAJ&pg=PA819&lpg=PA819&dq=World+Economic+Forum

[3] A. R. Arusha and R. K. Biswas, "Prevalence of stress, anxiety and depression due to examination in Bangladeshi youths: A pilot study," *Children and Youth Services Review*, vol. 116, p. 105254, Sep. 2020, doi: 10.1016/j.childyouth.2020.105254.

[4] gergana007, "Depression: A neglected public health domain in Bangladesh," *Globally Minded*, May 02, 2018. <<https://www.globallyminded.org/home/depression-a-neglected-public-health-domain-in-bangladesh/>> last accessed on Jan. 05, 2023.

[5] "Depression News from Medical News Today," *Medical News Today*, 2019. <<https://www.medicalnewstoday.com/categories/depression>> last accessed on 22-10-22 at 10:00 PM

[6] H. S. ALSAGRI and M. YKHLEF, "Machine Learning-Based Approach for Depression Detection in Twitter Using Content and Activity Features," *IEICE Transactions on Information and Systems*, vol. E103.D, no. 8, pp. 1825–1832, Aug. 2020, doi: 10.1587/transinf.2020edp7023.

[7] Y. Tyshchenko,<<https://core.ac.uk/download/pdf/237085027.pdf>> last accessed on 20-10-22 at 10:00 PM

[8] A. Sau and I. Bhakta, "Screening of anxiety and depression among the seafarers using machine learning technology," *Informatics in Medicine Unlocked*, vol. 16, p. 100149, 2019, doi: 10.1016/j.imu.2018.12.004.

[9] J. R. Sato, J. Moll, S. Green, J. F. W. Deakin, C. E. Thomaz, and R. Zahn, "Machine learning algorithm accurately detects fMRI signature of vulnerability to major depression," *Psychiatry Research: Neuroimaging*, vol. 233, no. 2, pp. 289–291, Aug. 2015, doi: 10.1016/j.pscychresns.2015.07.001.

[10] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," *Procedia Computer Science*, vol. 167, pp. 1258–1267, 2020, doi: 10.1016/j.procs.2020.03.442.

[11] "Assessment of Anxiety, Depression and Stress using Machine Learning Models," *Procedia Computer Science*, vol. 171, pp. 1989–1998, Jan. 2020, doi: 10.1016/j.procs.2020.04.213.

[12] B. Ojeme and A. Mbogho, "Selecting Learning Algorithms for Simultaneous Identification of Depression and Comorbid Disorders," *Procedia Computer Science*, vol. 96, pp. 1294–1303, 2016, doi: 10.1016/j.procs.2016.08.174.

[13] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2017, doi: 10.1109/ictc.2017.8190959.

[14] S. G. Burdisso, M. Errecalde, and M. Montes-y-Gómez, "A text classification framework for simple and effective early depression detection over social media streams," *Expert Systems with Applications*, vol. 133, pp. 182–197, Nov. 2019, doi: 10.1016/j.eswa.2019.05.023.

[15] M. Daly and E. Robinson, "Psychological distress and adaptation to the COVID-19 crisis in the United States," *Journal of Psychiatric Research*, Oct. 2020, doi: 10.1016/j.jpsychires.2020.10.035.

[16] N. Jothi, W. Husain, and N. A. Rashid, "Predicting generalized anxiety disorder among women using Shapley value," *Journal of Infection and Public Health*, Apr. 2020, doi: 10.1016/j.jiph.2020.02.042.

[17] B. Saha, T. Nguyen, D. Phung, and S. Venkatesh, "A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1008–1015, Jul. 2016, doi: 10.1109/jbhi.2016.2543741.

[18] Y. Chen, B. Zhou, W. Zhang, W. Gong, and G. Sun, "Sentiment Analysis Based on Deep Learning and Its Application in Screening for Perinatal Depression," *IEEE Xplore*, Jun. 01, 2018. https://ieeexplore.ieee.org/abstract/document/8411894 (accessed Mar. 08, 2022).

[19] J. Kim, J. Lee, E. Park, and J. Han, "A deep learning model for detecting mental illness from user content on social media," *Scientific Reports*, vol. 10, no. 1, Jul. 2020, doi: 10.1038/s41598-020-68764-y.

[20] C. Gurrin *et al.*, "Introduction to the Third Annual Lifelog Search Challenge (LSC'20)," *Proceedings of the 2020 International Conference on Multimedia Retrieval*, Jun. 2020, doi: 10.1145/3372278.3388043.

[21] G. Shen *et al.*, "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution." [Online]. Available: http://hcsi.cs.tsinghua.edu.cn/Paper/Paper17/IJCAI17-SHENGUANGYAO.pdf

[22] H. Jung, H.-A. Park, and T.-M. Song, "Ontology-Based Approach to Social Data Sentiment Analysis: Detection of Adolescent Depression Signals," *Journal of Medical Internet Research*, vol. 19, no. 7, p. e259, Jul. 2017, doi: 10.2196/jmir.7452.

[23] J. Du *et al.*, "Extracting psychiatric stressors for suicide from social media using deep learning," *BMC Medical Informatics and Decision Making*, vol. 18, no. S2, Jul. 2018, doi: 10.1186/s12911-018-0632-8.

[24] A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with Twitter data," *Scientific Reports*, vol. 7, no. 1, Oct. 2017, doi: 10.1038/s41598-017-12961-9.

[25] S. R. Braithwaite, C. Giraud-Carrier, J. West, M. D. Barnes, and C. L. Hanson, "Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality," *JMIR Mental Health*, vol. 3, no. 2, p. e21, May 2016, doi: 10.2196/mental.4822.

[26] Xu-Ying Liu, Jianxin Wu, and Zhi-Hua Zhou, "Exploratory Undersampling for Class-Imbalance Learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, Apr. 2009, doi: 10.1109/tsmcb.2008.2007853.

[27] T. G. Dietterich, "Ensemble Methods in Machine Learning," *Multiple Classifier Systems*, vol. 1857, pp. 1–15, 2000, doi: 10.1007/3-540-45014-9_1.

[28] S. Biswas, Md. Mozahidul. Islam, U. Sarker, R. H. Hridoy, and Md. T. Habib, "Machine Learning-Based Depression Detection," *Computer Networks and Inventive Communication Technologies*, pp. 809–821, Oct. 2022, doi: 10.1007/978-981-19-3035-5_60.

[29] Y. Ding, X. Chen, Q. Fu, and S. Zhong, "A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm," *IEEE Access*, vol. 8, pp. 75616–75629, 2020, doi: 10.1109/access.2020.2987523.

[30] A. Mulay, A. Dhekne, R. Wani, S. Kadam, P. Deshpande, and P. Deshpande, "Automatic Depression Level Detection Through Visual Input," *IEEE Xplore*, Jul. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9210301/ (accessed Jan. 05, 2023).

[31] N. A. Asad, Md. A. Mahmud Pranto, S. Afreen, and Md. M. Islam, "Depression Detection by Analyzing Social Media Posts of User," *IEEE Xplore*, Nov. 01, 2019. https://ieeexplore.ieee.org/abstract/document/9065101 (accessed Feb. 14, 2022).

[32] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," *IEEE Xplore*, Dec. 01, 2017. https://ieeexplore.ieee.org/document/8389299 (accessed May 10, 2020).

[33] Md. R. H. Khan, U. S. Afroz, A. K. M. Masum, S. Abujar, and S. A. Hossain, "Sentiment Analysis from Bengali Depression Dataset using Machine Learning," *IEEE Xplore*, Jul. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9225511/references#references (accessed Aug. 18, 2021).

[34] M. Stankevich, V. Isakov, D. Devyatkin, and I. Smirnov, "Feature Engineering for Depression Detection in Social Media," *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*, 2018, doi: 10.5220/0006598604260431.

[35] F. Patel, R. Thakore, I. Nandwani, and S. K. Bharti, "Combating Depression in Students using an Intelligent ChatBot: A Cognitive Behavioral Therapy," *2019 IEEE 16th India Council International Conference (INDICON)*, Dec. 2019, doi: 10.1109/indicon47234.2019.9030346.

[36] D. M. Shukla, K. Sharma, and S. Gupta, "Identifying Depression in a Person Using Speech Signals by Extracting Energy and Statistical Features," *IEEE Xplore*, Feb. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9087223/ (accessed Jan. 05, 2023).

[37] A. Husseini Orabi, P. Buddhitha, M. Husseini Orabi, and D. Inkpen, "Deep Learning for Depression Detection of Twitter Users," *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 2018, doi: 10.18653/v1/w18-0609.

[38] "Support Vector Machine (SVM) Algorithm - Javatpoint," <<*www.javatpoint.com*. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>> last accessed on 28-12-22 at 11:00 PM

[39] “Classification Algorithms - Random Forest - Tutorialspoint,” *Tutorialspoint.com*, 2019. <<https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm>> last accessed on 28-12-22 at 11:10 PM

[40] JavaTpoint, “K-Nearest Neighbor (KNN) Algorithm for Machine Learning - Javatpoint,” <<*www.javatpoint.com*. https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>> last accessed on 28-12-22 at 11:15 PM

[41] “K-Means Clustering Algorithm - Javatpoint,” *www.javatpoint.com*. <<https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning>> last accessed on 28-12-22 at 11:20 PM

APPENDIX

Appendix A: Related Issues

In order to finish this task, we had to overcome various challenging obstacles. The collection of information from individuals was our primary challenge. Because of this, many respondents to our survey for the purpose of data collecting didn't submit their correct data or didn't give any data at all. As a result, a significant amount of time was lost on the data collecting. Another problem has arisen, and this one is connected to how we imagined we would go about predicting disease outbreaks in the future. After that, we came to the conclusion that, in order to carry out this task, we need concentrate on locating and recognizing cases of depression in today's world. And in addition to that, we are confronted with certain challenges, such as the need to number our accumulated data in order to apply the algorithm, as well as the use of label encoders in order to number the degrees of depression.

PLAGIARISM REPORT

IDENTIFYING SIGNS OF DEPRESSION IN TODAY'S WORLD THROUGH THE APPLICATION OF MACHINE LEARNING ALGORITHMS

ORIGINALITY REPORT

16%	8%	9%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	medical-data-models.org Internet Source	2%
2	Jabir Al Nahian, Abu Kaisar Mohammad Masum, Sheikh Abujar, Md. Jueal Mia. "Common human diseases prediction using machine learning based on survey data", Bulletin of Electrical Engineering and Informatics, 2022 Publication	1%
3	"Computer Networks and Inventive Communication Technologies", Springer Science and Business Media LLC, 2023 Publication	1%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	S. Sumathi, S. N. Sivanandam. "Introduction to Data Mining and its Applications", Springer Science and Business Media LLC, 2006 Publication	1%