

**DEPRESSION DETECTION OF UNIVERSITY STUDENTS' USING
MACHINE LEARNING APPROACH**

BY

**MST. FAHMIDA MUNTASIM
ID: 191-15-12658**

**MD. RASHEDUL HAQUE ADOR
ID: 191-15-12578**

AND

**TAFSIRUL ISLAM NIBIR
ID: 191-15-12186**

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

Supervised By

Dewan Mamun Raza
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Md. Firoj Hasan
Lecturer
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

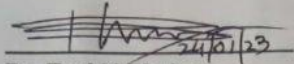
DHAKA, BANGLADESH

JANUARY 24, 2023

APPROVAL

This Project/internship titled “**Depression detection of university students’ using machine learning approach**”, submitted by Mst. Fahmida Muntsim ID: 191-15-12658, Md. Rashedul Haque Ador ID: 191-15-12578 and Tafsirul Islam Nibir ID: 191-15-12186 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 January 24, 2023.

BOARD OF EXAMINERS

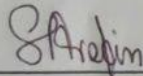


Dr. Touhid Bhuiyan

Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman

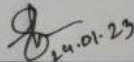


Dr. Mohammad Shamsul Arefin

Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

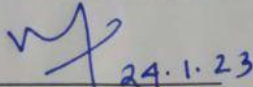


Md. Sabab Zulfiker

Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Ahmed Wasif Reza

Associate Professor

Department of Computer Science and Engineering
East West University

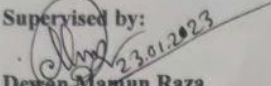
External Examiner

DECLARATION

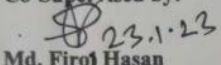
We hereby declare that, this project has been done by us under the supervision of **Dewan Mamun Raza, Senior Lecturer, Department of CSE Daffodil International University.**

We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

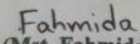
Supervised by:

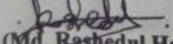

Dewan Mamun Raza
Senior Lecturer
Department of CSE
Daffodil International University


Co-Supervised by:


Md. Firoj Hasan
Lecturer
Department of CSE
Daffodil International University

Submitted by:


(Mst. Fahmida Muntasim)
ID: -191-15-12658
Department of CSE
Daffodil International University


(Md. Rashiedul Haque Ador)
ID: -191-15-12578
Department of CSE
Daffodil International University


(Tafsirul Islam Nibir)
ID: -191-15-12186
Department of CSE
Daffodil International University

©Daffodil International University

ii

ACKNOWLEDGEMENT

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

We really grateful and wish our profound our indebtedness to **Dewan Mamun Raza, Senior Lecturer**, Department of CSE 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 draft 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 CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE 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

In modern society, depression affects the majority of people. Many even commit suicide as a result of inadequate care. If depression is present in the patient in the early stages, it is simple to identify and cure. Since we don't know the depression level, we can't make the best choice at the appropriate time. A whole nation is built on its students. By teaching and improving the country, students represent it to the outside world. Depression is caused by a variety of factors, including challenges Bangladeshi adolescents have in their education. Determining the prevalence of depressive symptoms, their contributing variables, and strategies for reducing depression of university students are the objectives of our study. We analyzed the dataset with different samples from university students. We provide some question by a Google form, students are chosen the answer then set a range to find their depression level. About 1049 people's data were obtained from a Google form. In essence, the test was the data. In essence, the student has provided the data. We were able to determine the depression level through the analysis of that data. On this data, several algorithms have been applied. And have achieved the highest accuracy. With the help of this project, we can detect depression levels and administer the appropriate care or therapy. We're using some kind of algorithm to detect their depression level. They are five algorithms are chosen for this research. There are Decision Tree classification, Random Forest classifier, SVM, KNeighbors Classifier, and GaussianNB. Overall, the SVM algorithm prediction has the best performance. It gives 93% which is the best algorithm to prefer for this research.

TABLE OF CONTENTS

CONTENTS	PAGE No
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1-3
1.1 Introduction	1
1.2 Motivation	2
1.3 Rational of Study	2
1.4 Expected Outcome	3
1.5 Report Layout	3
CHAPTER 2: BACKGROUND	4-7
2.1 Introduction	4
2.2 Related Works	5-6
2.3 Scope of The Problem	7
2.4 Challenges	7
CHAPTER 3: RESEARCH METHODOLOGY	8-23
3.1 Introduction	8
3.2 Proposed Methodology	9

3.3 Data Collection and Processing	9-23
CHAPTER 4: EXPERIMENTAL ANALYSIS AND DISCUSSION	24-31
4.1 Introduction	24
4.2 Experimental Result	24-31
CHAPTER 5 : IMPACT ON SOCIATY AND ENVIRONMENT	32-33
5.1 Impact on Society	32
5.2 Impact on Environment	32
5.3 Ethical Aspects	32
CHAPTER 6: CONCLUSION AND FUTURE WORK	34
6.1 Conclusion	34
6.2 Future Work	34
REFERENCES	35-36

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.1: Proposed Methodology and Workflow	9
Figure 3.2: Age Distributions	11
Figure 3.3: Gender Distributions	12
Figure 3.4: Students Studying in a Particular Year	13
Figure 3.5: Before Data Preprocessing	13
Figure 3.6: After Data Preprocessing	15
Figure 3.7: How the Decision Tree Classifier Operates in Steps	17
Figure 3.8: Steps of How Random Forest Classifier Works	18
Figure 3.9: SVM Figure	19
Figure 3.10: Managing Nonlinear and Indistinguishable Planes	19
Figure 3.11: How SVM Works	20
Figure 3.12: How KNN Model Works	21
Figure 3.13: The Process of KNN Algorithm	22
Figure 3.14: How the Naive Bayes Operates	23
Figure 4.1: Decision Tree Classification Results	25
Figure 4.2: Random Forest Results	26
Figure 4.3: Support Vector Machine Results	27
Figure 4.4: KNeighborsClassifier Results	28
Figure 4.5: GaussianNB Results	29

LIST OF TABLES

TABLE NAME	PAGE NO
Tables 3.1 : Depression Severity	14
Tables 4.1 : Different Accuracy Results	30-31

CHAPTER 1

INTRODUCTION

1.1 Introduction

The term "depression" refers to a psychological condition that is currently quite widespread. This indicates a person's unfavorable psychological condition; depression typically results from a variety of common factors, including frustration, worry, and social as well as concerns with the family, health, and education. Depression often has a negative impact on a person's day-to-day functioning; a person's regular manner of living is impaired by his or her psychological or mental state. A person's growth in life and at work is typically hampered by this. Their mental condition prevents them from acting or concentrating appropriately. Students frequently experience psychological problems including depression. Due to poor communication or disagreements with family, friends, or other individuals, many people experience despair. Today, depression among students is a very serious issue that affects students all over the world. University students make up a special group of people who are going through a transitional period as they try to fit into their new academic lives and the social environments in which they now find themselves. Some students experience melancholy more frequently at this time. Many of them have had prior anxiety difficulties and some have other psychological problems such as communication troubles, which is a more often discovered condition that causes the majority of the students to experience despair. There isn't much research on depression in students in our country. When someone accepts verbal or physical humiliation, sadness is common. The loss of a close relative frequently raises the likelihood of depression. Sometimes Depression can happen when a person's brain is weak in certain biological chemicals. Unfortunately, many people do not understand or do not want to confess they have depression.

1.2 Motivation

An entire country is built on its students. Students represent the nation to the outside world by educating and developing it. However, if these youngsters leave university early, it poses a risk to all nations. The difficulties that Bangladeshi youth confront in their schooling, employment, relationships, drug use, and family matters all have a large or small role in depression. Therefore, the goal of our study is to identify the incidence of depressive symptoms, their contributing factors, and ways to lessen depression among university students. Depending on his life, a person might be sad in many different ways. A prevalent issue for people is depression disorder. The fundamental issue, though, is that few individuals in our nation are aware of mental health issues. The inability of people to comprehend their depression. They don't see a physician or a psychologist. We study the literature since there hasn't been a lot more research done in this field. This area really needs a lot more investigation. Accordingly, we are adopting the well-liked and widely-used machine learning technology to carry out this depression detection task. So it must determine if a person is sad or not based on their everyday activities or lifestyle. In order to determine a person's mental health status, we thus consider developing a method or system in which machine learning may be quite effective.

1.3 Rationale of the Study

Numerous studies on the prevalence of student depression have been conducted. Students from Bangladesh, on the other hand, have a significantly limited research field in depression. Research has often been conducted on university students. There hasn't been much research on high school and college students. But right now, for various reasons, the rate of sadness among all of these sensitive youngsters is increasing, which is dangerous for both the country and the nation. In our study, we collected university students to conduct research and employ machine learning to the prediction of depression.

1.4 Expected Output

Our work aims to make depression clearly distinguishable from other forms of conduct. Not being able to recognize if we are depressed is one of our main challenges. Our

approach makes it simple for someone to think about their mental condition. As a consequence, a sizable portion of the populace will no longer struggle with their mental condition.

We have to interact with a lot of individuals in society. We often see that some of our friends have wrapped themselves. They are no longer contributing to any social activities. The number of such persons in society is increasing, not decreasing. This essay will assist in the identification of people's mental states.

1.5 Report Layout

This chapter has introduced depression level detection, inspiration, the justification for the investigation, and the conclusion of the thesis. Then, the report's format is followed.

Chapter-1: In this chapter ,we describe about the depression in our university students.How can a students come out into the depression and what is the motive in this research.

Chapter-2: The background of our study topic will be covered.We discuss some others research and comparison with their work.Some authors are research different experiment in various ways.

Chapter-3: In this section , discuss data collection to final preprocessing. Here preprocess in our row dataset and analyze data then we using it machine learning technique and chosen desired model.

Chapter-4: The experimental results and discussion will be covered.All algorithm performance like precision, Recall, f1-score and support . Then we comparison all the algorithm .

Chapter-5: The Impact on Society, Environment, and Sustainability will be covered.

Chapter-6: We will discuss the resolution and our future steps in chapter.

CHAPTER 2

BACKGROUND

2.1 Introduction

Today's generation is highly acquainted with the term "depression." But depressed people experience a lot of challenging circumstances. In this section, we will discuss related research that has already been done by other researchers, a synopsis of their findings, the size of the issue, and the difficulties we encountered. In the section on comparable work, we include a few studies, related works, their applicable techniques, classifiers, and accuracy levels that are relevant to our study. We create a summary of all works in the research summary section, and then for easier comprehension, we present it as a table. When it comes to the problem's breadth, we talk about how we can assist or advance this effort. Finally, the Challenges section details the kinds of issues and obstacles we encountered throughout our study project and how we overcame them. Therefore, it is crucial to recognize and diagnose depression. Many studies have been conducted in the past to identify depression. On their own venues and in their own languages, several international researchers have been active. Sentiment analysis has received more attention than depression. Among these, there hasn't been much research on Bengali language depression detection. In this stage, we have attempted to highlight the study on the detection of sadness in social media comments or statuses from different scholars. Researchers' efforts to identify sadness using social media remarks and status updates are being discussed. In recent decades, data mining has become more popular as a method for examining scientific topics in the context of educational study. We'll discuss data mining in relation to data classifier research or theses in this chapter. In the first portion, previous related work will be described; in the second, the nature of the issues with our research-related work will be observed and in the third section, the challenges we encountered in our study will be covered.

2.2 Related Works

Researchers from studied first-year college students. They assessed concerns with mental health using the (GAIN-SS). Using generalized linear models, the associations between 12-month mental health disorders were calculated. Finally, they realized that Even less frequent were exclusive types of mental health difficulties, with internalizing problems accounting for 14.2% (SE=0.56), externalizing problems for 8.6% (SE=0.46), drug usage for 1.7% (SE=0.21), and antisocial problems for 0% [1]. Contains studies on the issues of mental health that Bangladeshi students face. They gathered 635 pieces of data in all, but only 590 were used for the final study. They use 11 distinct types of inquiries. They used the DASS-21 in Bangla for their project. The depression's Cronbach's alpha was 0.77. The Feature Set of their research has a size of 7. Their most accurate approach, logistic regression, has a 95% accuracy rate [2]. The authors of focused on student depression. They developed an online questionnaire system in the shape of a conventional psychometric tool. 121 characteristics might be extracted by them. The data of 466 students were evaluated using their methodology. They found that depression affects 25.32 percent of students [3]. Authors concentrated on first-year students. During their investigation, they obtained 400 pieces of information. Although they used a number of classifiers, logistic regression turned out to be the most effective. They divided the 14 factors they utilized into four groups in their surveys. According to their analysis, there are 52.3% male and 47.8% female applicants [4]. They examined four courses and discovered that 50.2% of the students suffered from moderate depression. Only 468 of the 523 students that were questioned by the authors of had their data analyzed. Although the dataset used in this research is not mentioned, it included a number of psychometric tests and socio-demographic questions. The final findings revealed that 27.1 percent of participants were anxious and 44% of participants were sad [5]. The authors of investigated alexithymia. The sample size was 492, and there were 14 features. The program (SPSS) version 21 was used to analyze the data. They evaluated stress, anxiety, and depression using a number of techniques. They observed that 24.6% of the subjects had alexithymia. The prevalence of depression, dread, and stress was 28.5, 38.4, and 22.6 percent, respectively [6]. From the Bengali depression dataset, Khan, M. R. H.

et al. conducted research on machine learning-based sentiment analysis. Data is gathered from social media postings, several poems, books, and quotes from different illustrious people. They use several algorithms, including decision trees, k-nearest neighbors, random forests, support vector machines, and xg boost. And using Multinomial Naive Bayes, they achieved an accuracy of 86.67%, which is the highest. By using strength and statistical indicators, Shukla, D. M. et al [7]. Suggested a technique for identifying depression in a person using speech or voice signals. They determine that by listening to their vocal signal, one may determine if a person is sad or not. This uses the RAVDESS audiovisual dataset from Ryerson, which includes emotive speech and music. They use the Multi-Layer Perceptron method. Cross-validation sets and training sets both had an average accuracy of 81.56%. Research on using emotional artificial intelligence to diagnose depression has been done by Deshpande, M. et al [8]. By looking through Twitter tweets, they may identify sadness. 10,000 Tweets were acquired using the Twitter API from Twitter tweets to create the training and test dataset. The methods of SVM and Naive Bayes classification were used to identify depression [9]. Support vector machines improve accuracy by 79%. The Multinomial Naive Bayes method, however, achieves the maximum level of accuracy. Multinomial Naive Bayes achieved the highest accuracy of 83%. In order to identify depression, Asad, N. A. et al. studied user postings on social media. The information is gathered from Facebook and Twitter. 150 people's Facebook data was personally acquired from Facebook with their consent using lovely soup to get information from Twitter. In this study, SVM, a Naive Bayes classification technique, is used. They use the well-known BDI-II questionnaire approach to determine the severity of their depression. It is assessed as being non-depressed, normal, mild, and borderline. A depression score of 1 to 55% is considered depressed. Naive Bayes has an accuracy of 74% and 100% precision, which is quite high [10].

2.3 Scope of the Problem

In essence, our study involves building a model by evaluating the provided data and using machine learning methods. The model we provide may identify depression. People in society will be significantly affected by this effort. In contemporary culture, there are a

lot of individuals who struggle with mental illnesses. However, they are unaware that they are sad. They isolate themselves and are unsure of what to do. They thus erred in their judgment at the time. He needs a method that will enable him to identify his issue and determine if he is really depressed or not.

2.4 Challenges

We use data from the survey for our task, which is collected by google from. Several difficulties had to be overcome throughout the data collection process via the online survey platform.

- Getting Data
- Getting Valid Data
- Justification of Data
- Preprocessing the row data set
- Chosen the desired Algorithm

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this part, we'll go through the complete research approach. Each analysis may be solved using a different strategy. We started by compiling the survey's data. We must remove certain null values and columns that are not pertinent to our study after building the dataset and processing the data. The machine learning algorithm is then selected. Since we've already said that we're using five different machine learning techniques, we must first construct a data set before we can build the model and fit the algorithm. The model is then trained using this collection of data. Feature selection is functioning on such a base. The information was then split into a training set and a testing set. It's also known as the train data set and the test data set. We then only get a significant percentage of the data needed to evaluate our model after fitting the data into different machine learning algorithm models and training it using a training data set. After then, the model's accuracy is assessed. We've been using our typical process flow chart to provide an overview, but to further understand how some of the algorithms work, we'll go through them using equations and diagrams. The whole research project's workflow is shown here, along with a brief summary of the entire study effort.

Subject of study and research components:

Today's culture has a serious issue with depression. By gathering information from a survey of the students, we attempted to identify depression. For our job, we used machine learning techniques. Our primary area of study was real emotion detection.

3.2 Proposed Methodology

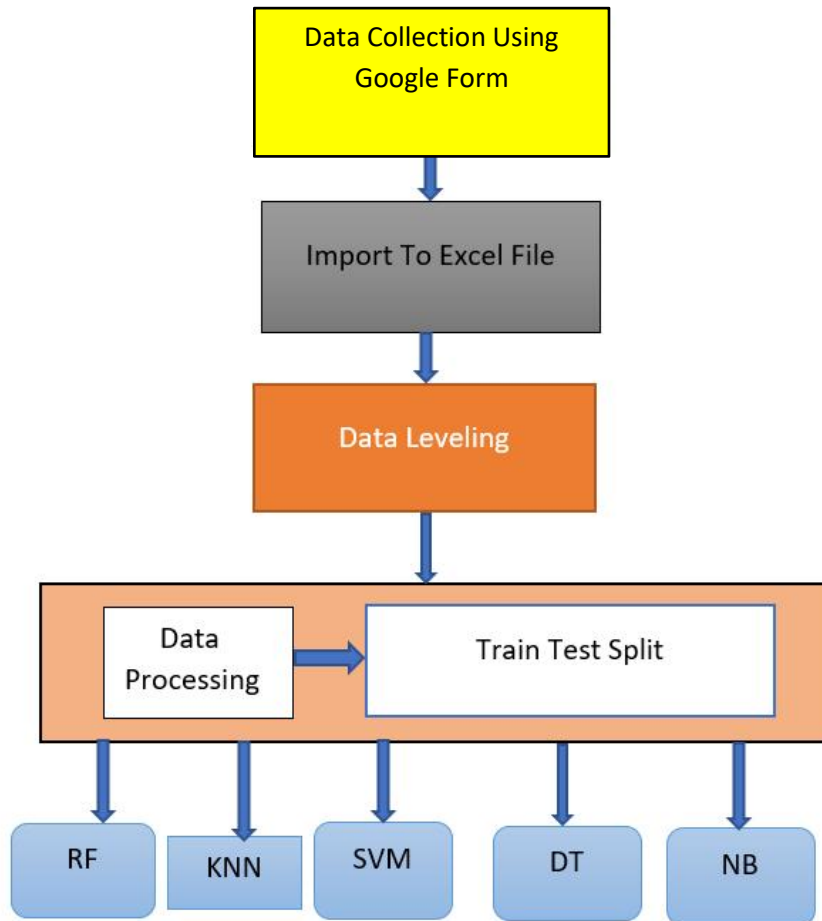


Figure 3.1: proposed methodology workflow

3.3 Data collection and Data Preprocessing

In order to meet our requirements, we made a few little adjustments to the data set after obtaining it through a survey. We provide some questioners to collect data in the university students. Its have 14 columns, those are maximum questions. We collected 1049 students opinion about their regular based activities. Then preprocess to raw dataset, in the preprocessing part we're using same technique to get perfect data to using machine learning algorithm. Our dataset have some column, some have combined text and integer value. We're using only integer value, thats why we remove text value from the every cell.

Then only integer value is using find out to depression range. Then this depression range are divided some categories and finally this range categories divided depression type from the dataset. The data set must then be used by each procedure. We must preprocess the data in order to properly fit the data set in our model, each of the five algorithms we utilize acts differently.

3.4.1 Data set

Using a Google form, we amassed 1049 student responses. Next, an excel file is created from the excel file. Prior to processing the data for our approach. The primary focus of the study is on this. We give some questionnaires about students' activities.

Age

Gender

3. Little interest or pleasure in doing things?
4. Feelings down, depression, or hopelessness?
5. Trouble falling or staying asleep or sleeping too much?
6. Feeling tired or having little energy?
7. Poor appetite or overeating?
8. Feeling bad about yourself or that you are a failure or have let yourself or your family down?
9. Trouble concentrating on things, such as reading the newspaper or watching television?
10. Becoming easily annoyed or irritable?
11. Thoughts that you would be better off dead or thoughts of hurting yourself in the same ways?
12. Worrying too much about different things?

3.4.2 Import Library

Import the module using an import statement in order to utilize its functionalities. The import keyword and the module name are what make up an import statement. This will be stated at the start of the code, behind any shebang lines or general comments, in a Python file. We use here pandas, numpy, matplotlib, seaborn and import some algorithm which is used in our project.

3.4.3 Exploratory Data analysis

In this section, we explore our data set in many ways saw the unique value, null value, data type and data set column.

3.4.4 Data visualization

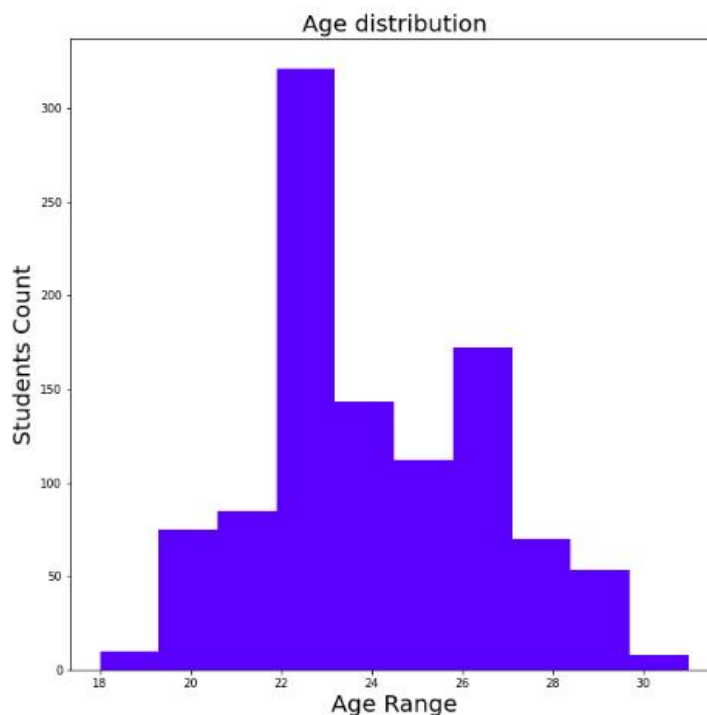


Figure 3.2: Age distributions

We provide age data in our dataset, and we found that students typically completed surveys between the ages of 20 and 35.

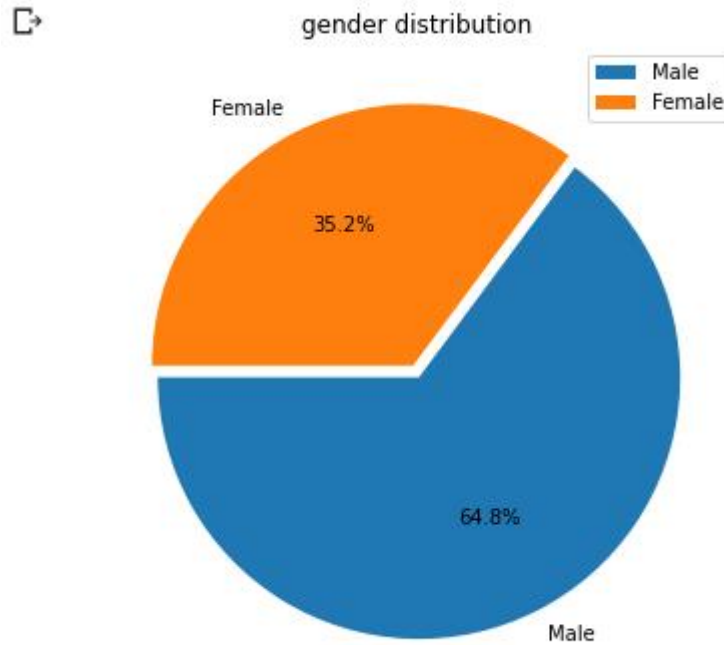


Figure 3.3: Gender Distribution

In our data set, which includes gender information, we discovered that students frequently completed questionnaires between the ages of 20 and 35. And in our dataset, the gender distribution is 35.2% female and 64.8% male. This pie graph from our dataset displays this value.

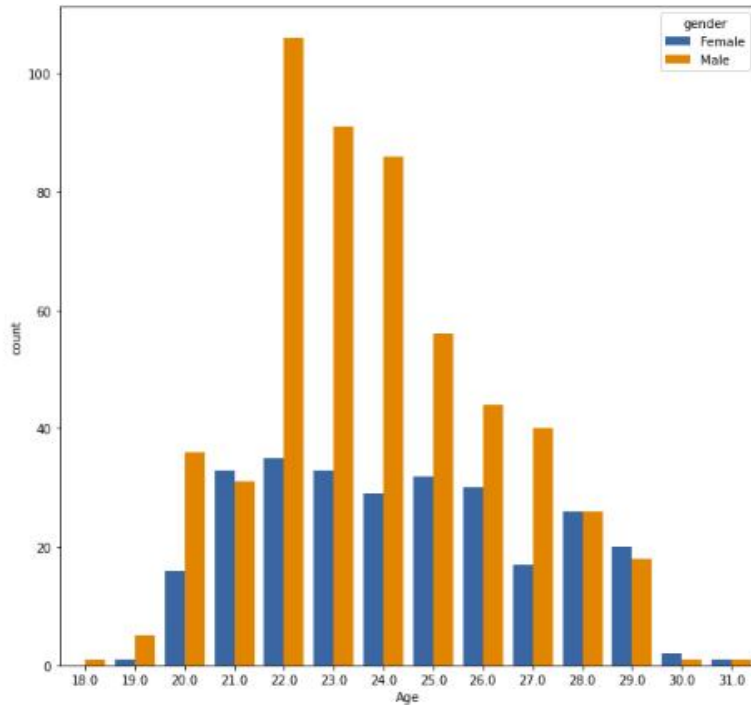


Figure 3.4: Students studying in a particular year

3.4.5 Data Pre-processing

Age	gender	3. Little interest or pleasure in doing things? (কাজ করার সামান্য আগ্রহ বা আলাপ ?)	4. Feelings down, depression, or hopelessness? (মন খারাপ, বিষ্ময়, বা আশাহীন ?)	5. Trouble falling or staying asleep or sleeping too much? (পড়ে বা ঘুমিয়ে থাকতে বা খুব বেশি ঘুমতে সমস্যা হয় ?)	6. Feeling tired or having little energy? (হালু বোধ করছেন বা অতিরিক্ত খাওয়া ?)	7. Poor appetite or overeating? (খুশমান্য বা অতিরিক্ত খাওয়া ?)	8. Feeling bad about yourself or that you are a failure or have let yourself or your family down? (নিজের সম্পর্কে খারাপ বোধ করছেন বা আপনি হার্ড হয়েছেন বা নিজেকে বা আপনার পরিবারকে হতাশ করছেন ?)	9. Trouble concentrating on things, such as reading the newspaper or watching television? (খবরপত্র বা টেলিভিশন দেখা ?)	10. Becoming easily annoyed or irritable? (সহজেই বিরক্ত বা উত্তেজিত হওয়া ?)	11. Thoughts that you would be better off dead or thoughts of hurting yourself in the same ways? (ছিন্ন হওয়া বা আত্মহত্যা করার চিন্তা ?)	12. Worrying too much about different things? (বিভিন্ন বিষয়ে খুব বেশি উদ্বেগ ?)
20.0	Female	0 not at all. (সেটেই না।)	2 More than half the days. (অর্ধেকেরও বেশি দিন।)	1 Several days. (কয়েক দিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	0 not at all. (সেটেই না।)	1 Several days. (কয়েক দিন।)	0 not at all. (সেটেই না।)	0 not at all. (সেটেই না।)	0 not at all. (সেটেই না।)	1 Several days. (কয়েক দিন।)
22.0	Male	0 not at all. (সেটেই না।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	2 More than half the days. (অর্ধেকেরও বেশি দিন।)	0 not at all. (সেটেই না।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	0 not at all. (সেটেই না।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)
24.0	Male	3 Nearly everyday. (প্রায় প্রতিদিন।)	1 Several days. (কয়েক দিন।)	0 not at all. (সেটেই না।)	0 not at all. (সেটেই না।)	0 not at all. (সেটেই না।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	0 not at all. (সেটেই না।)	0 not at all. (সেটেই না।)	0 not at all. (সেটেই না।)	0 not at all. (সেটেই না।)
21.0	Female	1 Several days. (কয়েক দিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	2 More than half the days. (অর্ধেকেরও বেশি দিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	0 not at all. (সেটেই না।)	3 Nearly everyday. (প্রায় প্রতিদিন।)
24.0	Male	2 More than half the days. (অর্ধেকেরও বেশি দিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	1 Several days. (কয়েক দিন।)	1 Several days. (কয়েক দিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	1 Several days. (কয়েক দিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	3 Nearly everyday. (প্রায় প্রতিদিন।)	2 More than half the days. (অর্ধেকেরও বেশি দিন।)

Figure 3.5: Before data preprocessing

Students complete the survey and then provide their feedback. There are four checkboxes and an integer number for each of the questions' parts. Student members tick one box. Every question has both English and Bengali words. We eliminate every word from this dataset during preprocessing and just keep integer values.

After subtracting a selected column, range the value according to the depression type.

For every question, there are some checked boxes.

0 = Not at all

1 = Several days

2 = More than half the days

3 = Nearly everyday

Interpretation of Total score:

Total Score	Depression Severity
0-4	Minimal Depression
5-9	Mild Depression
10-14	Moderate Depression
15-19	Moderately severe Depression
20-40	Severe Depression

Table 3.1: Depression Severity

After preprocessing the data with basic coding, we obtain the total score, which is then divided into ranges to produce the various forms of depression.

Age	gender	3. Little interest or pleasure in doing things? (কম আগ্রহ বা আনন্দ ?)	4. Feelings down, depression, or hopelessness? (দুঃখ, হতাশা, বা আশাহীনতা ?)	5. Trouble falling or staying asleep or sleeping too much? (ঘুম না আসা, ঘুমের অসুবিধা বা অতিরিক্ত ঘুম ?)	6. Feeling tired or having little energy? (লক্ষ্য কমে যাওয়া বা শক্তি হারা ?)	7. Poor appetite or overeating? (খাবারের অভাব বা অতিরিক্ত খাওয়া ?)	8. Feeling bad about yourself or that you are a failure or have let yourself or your family down? (নিজে সম্পর্কে খারাপ বোধ করা বা আপনার পরিবারকে হতাশ করা ?)	9. Trouble concentrating on things, such as reading the newspaper or watching television? (মনোযোগের অভাব, যেমন সংবাদপত্র পড়া বা টেলিভিশন দেখা ?)	10. Becoming easily annoyed or irritable? (সহজে বিরক্ত বা খিটখিটে হওয়া ?)	11. Thoughts that you would be better off dead or thoughts of hurting yourself in the same ways? (আপনি মৃত হতে উৎসাহিত হওয়া বা একই উপায়ে নিজেকে আঘাত করার চিন্তা ?)	12. Worrying too much about different things? (বিভিন্ন বিষয়ে অতিরিক্ত উদ্বেগ ?)	Total_score	Depression_Type
20.0	Female	0	2	1	3	0	1	0	0	0	1	8	Mild
22.0	Male	0	3	2	0	3	3	0	3	3	3	20	Sereve
24.0	Male	3	1	0	0	0	3	0	0	0	0	7	Mild
21.0	Female	1	3	3	3	3	3	2	3	0	3	24	Sereve
24.0	Male	2	3	1	1	3	1	3	3	3	2	22	Sereve

Figure 3.6: After data preprocessing

There are three types of depression type we get after preprocessing dataset. when working with the value range, we get this level. There are 71 students are severe depression type in according to the dataset , 72 students are minimal, 212 students are mild, 324 students are moderately Severe and last 370 students are moderate depression type.

Minimal Depression

Low-grade depression symptoms are present in mild depression. Even though many of the same signs of less severe depression, such as irritation, sorrow, and lack of motivation, are still present, they are frequently milder and less noticeable.

If a person exhibits two to four depressive symptoms over the course of two weeks, one of which must be either a low mood or a lack of interest or pleasure, that person is said to have a mild depressive disorder.

Mild Depression

People who suffer from mild, low-grade depression might not even be aware of their condition. In fact, their persistent melancholy and poor mood may have persisted for so long that they seem normal.

However, it is not typical to experience constant misery in life. Everyone occasionally

feels down because of upsetting or unpleasant life situations, but feeling down all the time doesn't have to be your life's narrative.

Moderate Depression

When compared to someone with mild depression, a person with moderate depression may experience symptoms that are more severe in terms of intensity and duration. Additionally, they could suffer greater symptoms compared to someone with mild depression. Major depressive disorder is frequently categorized when it is diagnosed by the severity (mild, moderate, or severe), as well as the presence of psychotic elements and a seasonal pattern. The symptoms of moderately severe depression can interfere with a person's ability to carry out daily tasks. In terms of the severity and frequency of symptoms, it may be different from mild depression.

Moderately severe Depression

Low mood and irritability on a daily basis, as well as a loss of interest or enjoyment in once-pleasurable activities, are typical characteristics of moderately severe depression. When someone has mild depression, these symptoms might range in severity and length.

Severe Depression

Symptoms of severe depression frequently include feelings of helplessness, heightened irritation, loss of enjoyment, difficulty concentrating or sleeping, as well as suicidal or death thoughts. Severe depression isn't a recognized mental health condition in the strictest sense. It actually refers to more severe depression. Depression is a significant condition that has an impact on a person's actions, emotions, thoughts, and physical well-being. Any age can experience depression. The symptoms of depression can be temporary or chronic, mild to severe, and range in intensity.

3.3 Implementation and Algorithms

3.3.1 Decision Tree

The most crucial consideration when creating a machine learning model is selecting the approach that is most suited to the data set and task at hand. The following are the two justifications for employing the decision tree:

Decision trees are frequently created to mimic the way that individuals think while making decisions, making them easy to understand. The decision tree's reasoning is straightforward because it exhibits a tree-like structure.

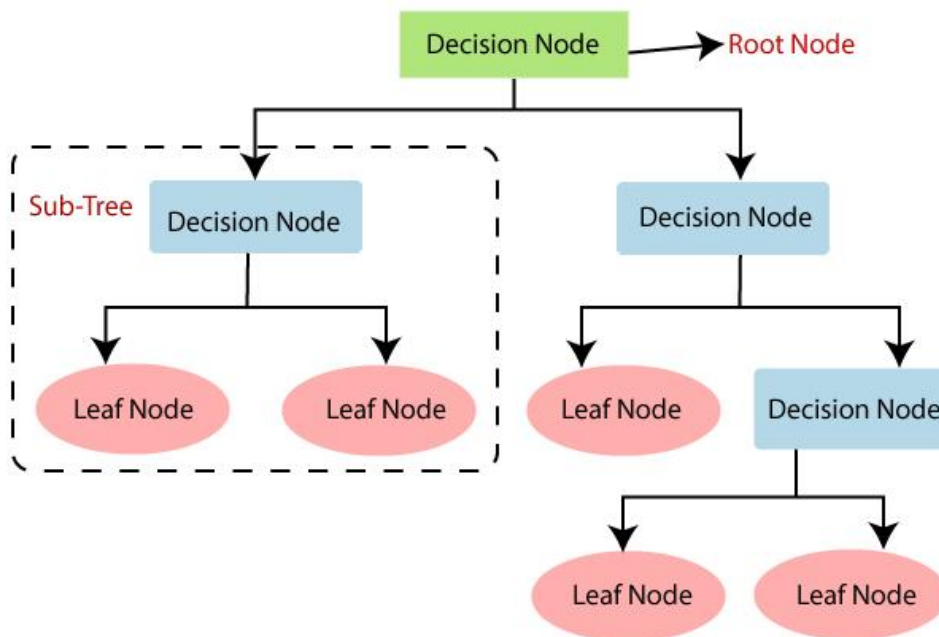


Figure 3.7: How the Decision Tree classifier operates in steps

The procedure in a decision tree starts at the root node and moves upward to predict the class of the supplied data set. By comparing the values of the record (real data set) attribute with those of the root attribute, this approach follows the branch and moves to the next node.

3.3.2 Random Forest

A technique for supervised learning is random forests. It has the capacity to be applied to both regression and classification. The algorithm is also the most flexible and user-friendly. A forest is made up of trees. The more trees a forest contains, the more robust it is regarded as being. From randomly selected data samples, random forests create decision trees, get forecasts from each tree, and then vote on the best choice. It also acts as a potent signal of the feature's worth.

Recommendation systems, picture classification, and feature selection are just a few applications for random forests. It can be used to spot fraud, classify dependable loan applicants, and predict illnesses. It serves as the foundation for the Boruta approach, which selects pertinent dataset features.

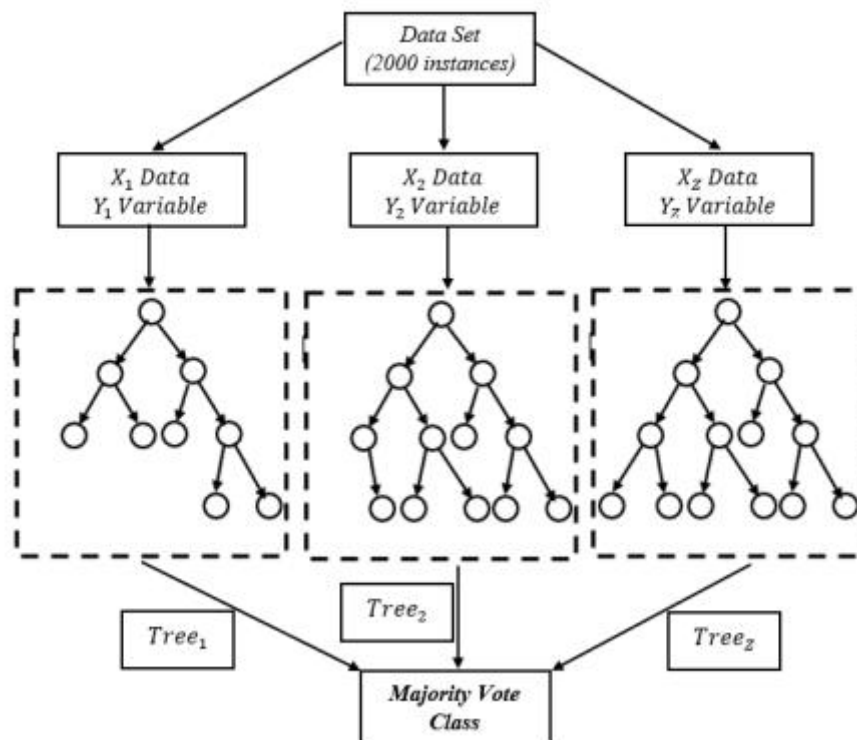


Figure 3.8: Steps of how random forest classifier works

Random forests is regarded as a particularly precise and resilient method due to the significant number of decision trees involved in the process. It is unaffected by the over fitting issue. This is done for the primary reason that any biases are eliminated by

averaging all of the projections. The method can be applied to problems involving classification and regression. Random forests can also be used to manage missing values. There are two ways to handle missing data: determining the proximity-weighted average or substituting continuous variables with their median values.

3.3.3 SVM (Support Vector Machine)

A popular Supervised Learning method known as the Support Vector Machine, or SVM, can be used to address classification and regression problems. But it's largely used in machine learning to solve classification problems.

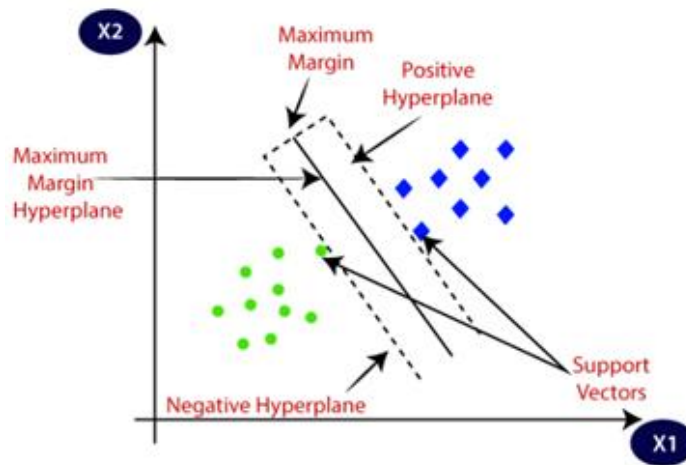


Figure 3.9: SVM figure

Linear SVM is a classifier that is used for linearly separable data, which denotes that a data set is linearly separable if it can be divided into two classes using a single straight line.

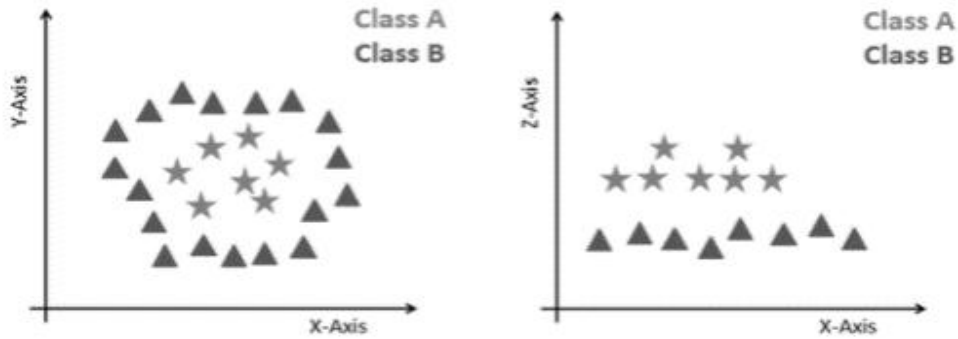


Figure 3.10: Managing nonlinear and indistinguishable planes

The goal of the SVM method is to determine the best decision boundary or line for classifying n-dimensional space into groups so that subsequent data points can be quickly assigned to the appropriate category. The border of the ideal choice is called a hyperplane. SVM is used to choose the extreme points and vectors that help build the hyperplane. The technology is referred to as a "Support Vector Machine," and support vectors are extreme examples. Take a look at the illustration below, which illustrates the use of a decision boundary or hyperplane to classify two separate categories:

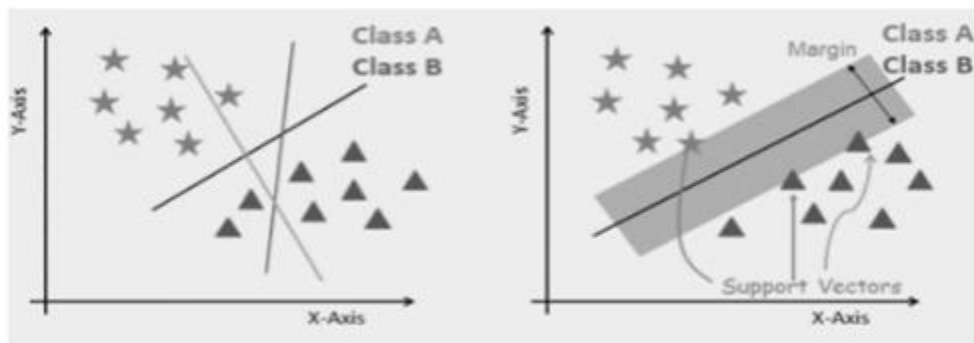


Figure 3.11: How SVM works

The kernel is used to implement the SVM algorithm. The data input space can be transformed by a kernel into the necessary format. A kernel trick technique is used by SVM. In this case, the kernel enlarges a constrained input region. In other words, it's feasible. It is said that the addition of additional dimensions transforms problems that at first glance appear unconnected into ones that can be divided and that this is especially

©Daffodil International University 20

helpful in problems involving non-linear separation. It is feasible to generate a categorization that is more detailed using the kernel technique. In a traditional dot product, any two actions can be used as a linear kernel. The product of the two vectors results from multiplying each pair of input values in total.

3.3.4 KNN

One of the most fundamental Machine Learning algorithms is the K-Nearest Neighbour method, which is based on the Supervised Learning methodology. It is a simple technique that can be applied to problems with regression and classification. The K-NN technique places the new case in the category that matches the current categories the most, presuming that the new case/data and old instances are equivalent. The K-NN method keeps all of the data that is currently available and categorizes incoming data points based on how similar they are to existing data. This means that new data can be quickly sorted into a clearly defined category using the K-NN technique.

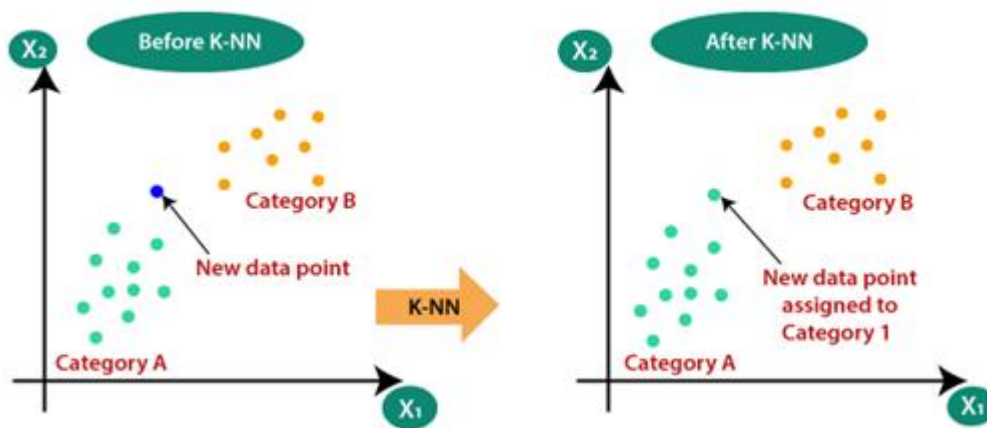


Figure 3.12: How KNN model works

Although the KNN method can be applied to both classification and regression, classification is where it is most frequently used. Because the KNN technique is non-parametric, it makes no assumptions about the data. Since it doesn't categorize the training set right away but instead keeps the data set, this method is also referred to as a

lazy learner. The KNN approach simply saves the data set during the training phase and classifies new data into a category that is very close to the previously trained data set.

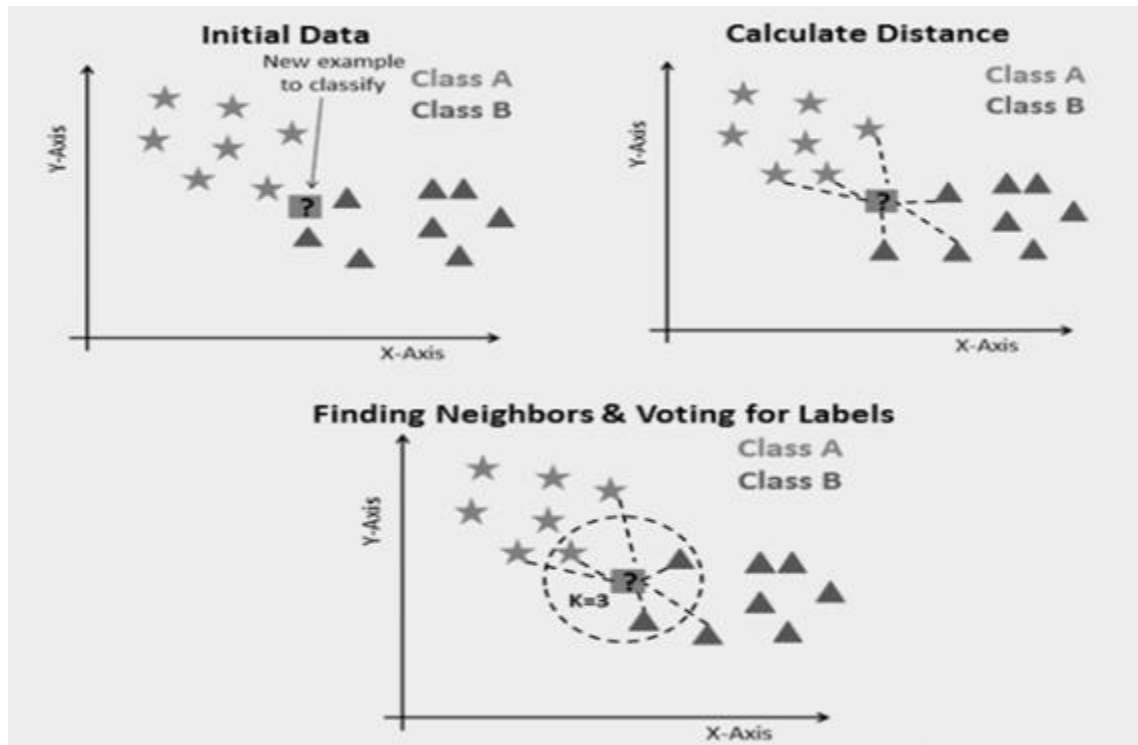


Figure 3.13: The process of KNN algorithm

K acting as a controlling factor in a projection model. There are no ideal neighbors in any data collection. Each dataset has a unique set of specifications.

3.3.5 Naive Bayes

The Naive Bayes machine learning model is recommended even when working with data that contains millions of records since it can handle large amounts of data. It performs NLP tasks like emotional analysis quite well. The classifying process is quick and easy.

To understand the naive Bayes classifier, we first need to understand the Bayes theorem. Therefore, let's start by discussing the Bayes Theorem.

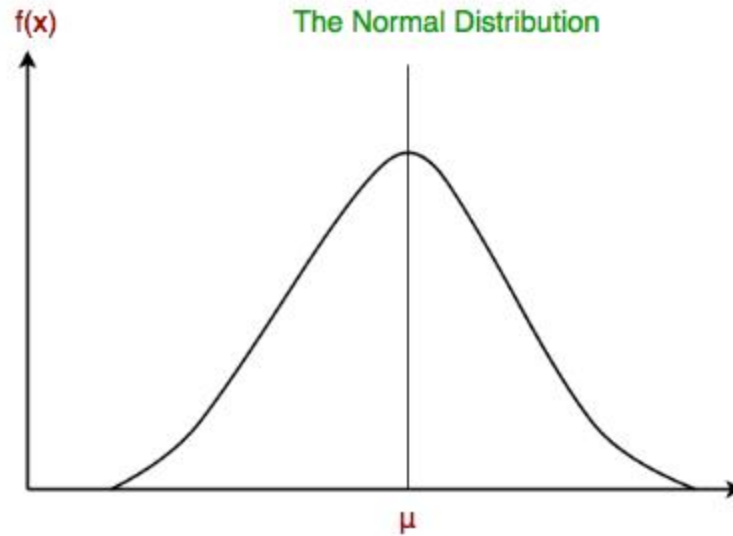


Figure 3.14: How the naive bayes operates

It is a theorem based on conditional probabilities. The conditional probability is the probability that something will occur given that something else has already occurred. Using our prior knowledge and conditional probability, we may calculate the likelihood of an event.

Naive Bayes algorithms are widely used in sentiment analysis, spam filtering, recommendation systems, and other fields. They are quick and easy to use, but their main flaw is the requirement for independent predictors. When the predictors are reliant, as they are in most real-world scenarios, the classifier performs worse.

CHAPTER 4

EXPERIMENTAL ANALYSIS AND DISCUSSION

4.1 Introduction

We'll discuss the outcomes of the experiment in this section. All except one of the five algorithms we'll be using have been described. Now we can assess how well those algorithms performed. We'll also evaluate the accuracy of each of the five approaches.

Some process takes to complete this research are follows:

Step-1: Collecting data set

Step-2: Import different libraries

Step-3: Data visualization

Step-4: Pre-processing

Step -5: Divide our data set

Step-6: Make modals for all five algorithms.

Step-7: Train with all five machine learning algorithms.

Step-8: Determine the accuracy of each algorithm.

4.2 Experimental Result:

We understand that no machine can create a flawless outcome. In a similar vein, we may adjust the parameters of our model during training to improve accuracy. The accuracy we find using several approaches is, nevertheless, rather high. The photographs that follow provide a quick rundown of our research activities. We are showing in these photos precision, recall, f1 score, support, accuracy, and heatmap.

Here, we use our data to identify the depression categories that are most prevalent.

	precision	recall	f1-score	support
Mild	0.68	0.84	0.75	62
Minimal	0.74	0.77	0.76	22
Moderate	0.83	0.66	0.74	103
Moderately Severe	0.74	0.79	0.76	100
Severe	0.62	0.57	0.59	28
accuracy			0.74	315
macro avg	0.72	0.73	0.72	315
weighted avg	0.74	0.74	0.74	315

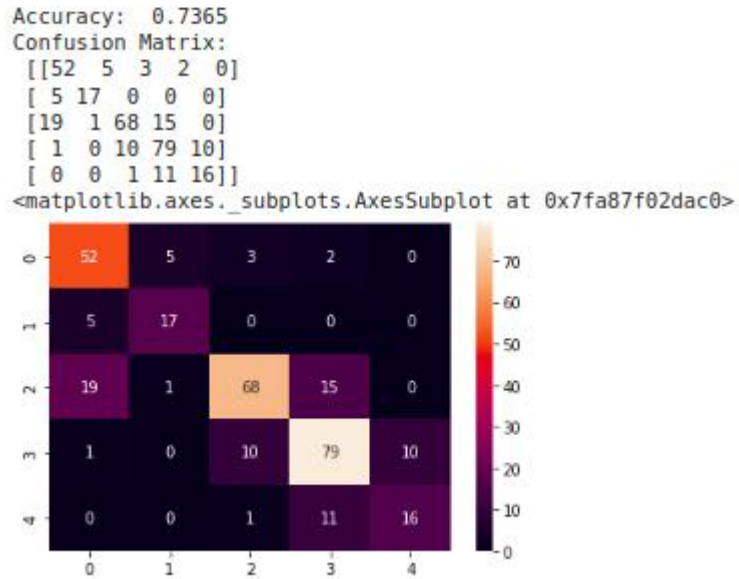


Figure 4.1. Decision tree Classification result

In the above 4.1 figure, we're showing all results from decision tree classification. It is showing 73% accuracy which is not good for best model performance. We also added a confusion matrix heatmap to show confusion matrix values.

	precision	recall	f1-score	support
Mild	0.66	0.73	0.69	62
Minimal	1.00	0.23	0.37	22
Moderate	0.69	0.79	0.74	103
Moderately Severe	0.68	0.81	0.74	100
Severe	1.00	0.21	0.35	28
accuracy			0.69	315
macro avg	0.81	0.55	0.58	315
weighted avg	0.73	0.69	0.67	315

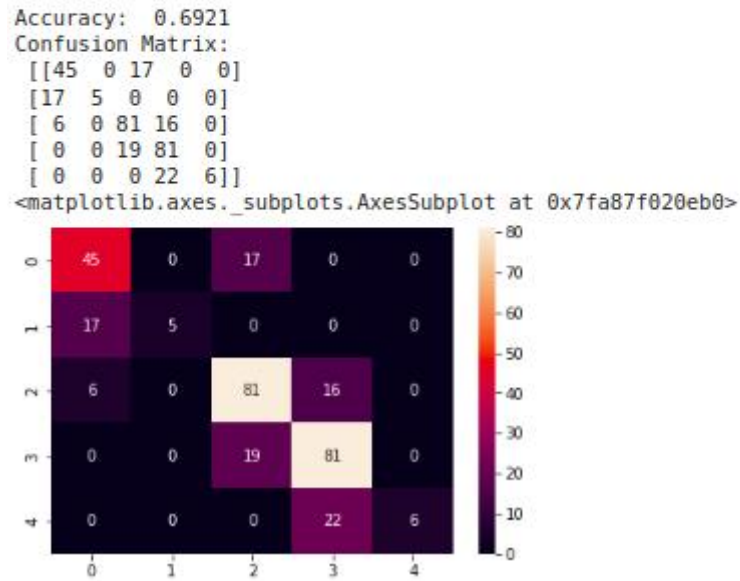


Figure 4.2 : Random forest Results

In the above 4.2 figure, we're showing all results from Random forest classification. It showing 69% accuracy .

	precision	recall	f1-score	support
Mild	0.88	0.97	0.92	62
Minimal	0.94	0.68	0.79	22
Moderate	0.96	0.98	0.97	103
Moderately Severe	0.92	0.97	0.94	100
Severe	1.00	0.71	0.83	28
accuracy			0.93	315
macro avg	0.94	0.86	0.89	315
weighted avg	0.93	0.93	0.93	315

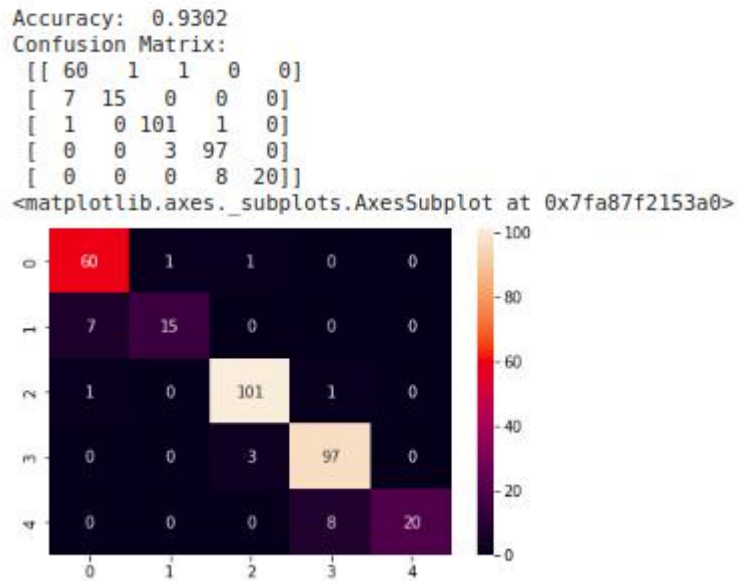


Figure 4.3 : Support Vector machine results

In the above 4.3 figure, we're showing all results from decision tree classification. It showing 93% accuracy which is good for best model performance.

	precision	recall	f1-score	support
Mild	0.78	0.87	0.82	62
Minimal	0.79	0.86	0.83	22
Moderate	0.77	0.83	0.79	103
Moderately Severe	0.76	0.77	0.77	100
Severe	1.00	0.36	0.53	28
accuracy			0.78	315
macro avg	0.82	0.74	0.75	315
weighted avg	0.79	0.78	0.77	315

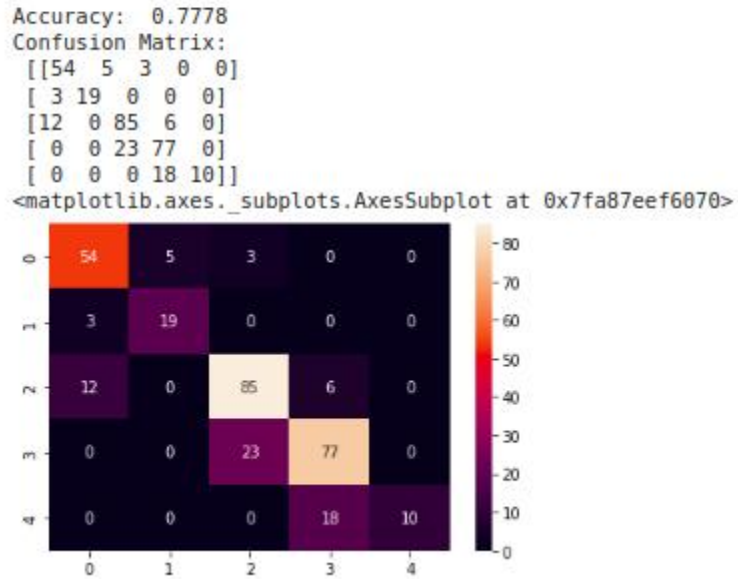


Figure 4.4: KNeighborsClassifier Results

In the above 4.4 figure, we're showing all results from KNeighbors classifier results. It showing 78% accuracy which is good for best model performance.

	precision	recall	f1-score	support
Mild	0.84	0.76	0.80	62
Minimal	0.85	0.77	0.81	22
Moderate	0.77	0.89	0.83	103
Moderately Severe	0.77	0.85	0.81	100
Severe	1.00	0.32	0.49	28
accuracy			0.79	315
macro avg	0.85	0.72	0.75	315
weighted avg	0.81	0.79	0.78	315

Accuracy: 0.7937

Confusion Matrix:

```
[[47  3 12  0  0]
 [ 5 17  0  0  0]
 [ 4  0 92  7  0]
 [ 0  0 15 85  0]
 [ 0  0  0 19  9]]
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fa87ee39eb0>

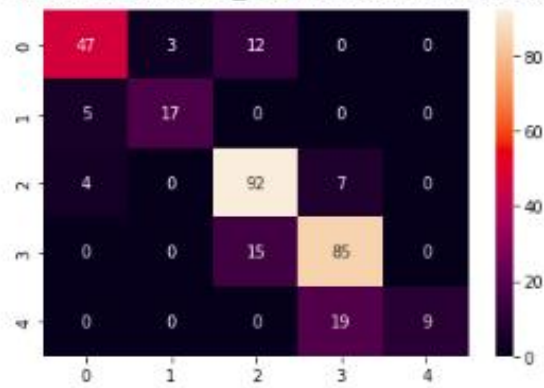


Figure 4.5: GaussianNB Results

In the above 4.5 figure, we're showing all results from GaussianNB Results. It is showing 79% accuracy, which is good for best model performance.

Accuracy Results

Algorithm Name	Type	Precision	Recall	F1-Score	Accuracy
SVM	Mild	0.88	0.97	0.92	0.93
	Minimal	0.94	0.68	0.79	
	Moderate	0.96	0.98	0.97	
	Moderately Severe	0.92	0.97	0.94	
	Severe	1.00	0.71	0.83	
Naive Bayes	Mild	0.84	0.76	0.80	0.79
	Minimal	0.85	0.77	0.81	
	Moderate	0.77	0.89	0.83	
	Moderately Severe	0.77	0.85	0.81	
	Severe	1.00	0.32	0.49	
KNN	Mild	0.78	0.87	0.82	0.78
	Minimal	0.79	0.86	0.83	
	Moderate	0.77	0.83	0.79	
	Moderately Severe	0.79	0.77	0.77	
	Severe	1.00	0.36	0.53	
Decision Tree	Mild	0.68	0.84	0.75	0.73
	Minimal	0.74	0.77	0.76	

	Moderate	0.83	0.66	0.74	
	Moderately Severe	0.74	0.79	0.76	
	Severe	0.62	0.57	0.59	
Random Forest	Mild	0.66	0.73	0.62	0.69
	Minimal	1.00	0.23	0.37	
	Moderate	0.69	0.79	0.74	
	Moderately Severe	0.68	0.81	0.74	
	Severe	1.00	0.21	0.35	

Table 4.1: Accuracy Results

We applied five different algorithms in our projects. In the above table 4.1 showing all algorithm accuracy. The SVM algorithm is given the best performance of all algorithms. It has given 93% accuracy that's is better than other algorithms.

CHAPTER 5

IMPACT ON SOCIETY AND ENVIRONMENT

5.1 Impact on Society

Students are the most important part of society. Public health concerns, such as frequent mental health problems, might cause students to engage in dangerous activities, including suicidal conduct in severe circumstances. Society is negatively impacted by mental health issues among students. A student who has an unanticipated psychological issue finds it difficult to concentrate on his educational work. Their daily lives are affected by this terrible illness. The students lose consciousness of all aspects of their existence. When a student has a psychological issue, he withdraws from his friends, family, and society in general. A student is therefore unable to effectively fulfill his civic obligation. If the majority of students have a socially accepted mental health condition, then the whole society will be affected and may experience an unanticipated event like suicide.

5.2 Ethical Aspects

This depression detection methodology does not infringe on human rights in any manner and is not immoral. There will be no privacy issue since the model does not gather any personal information, like name, identity, etc. This strategy helps to increase awareness rather than undermining someone's ability to appreciate or utilize something. Therefore, the model of depression detection may be easily handled by utilizing machine learning technology.

5.3 Impact on Environment

A person with depression has a loss of feeling of well-being compared to non-depressed persons. He harms not just himself, but the environment as well. A depressed individual is more likely than a normal person to develop a variety of addictions, which is dangerous for both him and the environment as a whole. A sad individual is also not very effective at saving the environment. Therefore, depression has negative effects on the body, the

environment, and even the mind. By identifying depression, our study will assist in returning a person to a healthy lifestyle. We also expect that the environment will benefit from this.

CHAPTER 6

CONCLUSION & FUTURE WORK

6.1 Conclusion

Our objective was to investigate the prevalence of depression, anxiety, and stress among Bangladeshi students, as well as the associated risk factors, and to provide them with a treatment program to aid them in overcoming this troubling illness. People now experience depression on a fairly regular basis. Today, it has developed into an intolerable and unanticipated mental health problem among Bangladeshi students. This research was done to determine the prevalence of depression among Bangladeshi students. We conducted surveys of Bangladeshi students in several locations in order to conduct thorough research of 1049 student records in the university.

6.2 Future Work

Modern technology, data science, and artificial intelligence have recently sped up, simplified, and improved every aspect of human existence. In the future, we plan to build our model as a web application or an Android application. We'll work to improve our models' accuracy in the future. We'll build a bigger database with a ton of data on individuals and add more classification layers to the datasets. Additionally, the website or mobile application created for the model may be made available to everyone, including physicians, by providing an immersive user-friendly GUI. The model may be made far more effective and practical by implementing new algorithms, adding new parameters, and adding more features. By gathering data from more diverse groups of individuals according to the area, age, and activities in the future, a robust dataset may be produced. Additionally, the Department of Mental Health may assist in expanding the concept.

Reference

1. R. Bruffaerts, P. Mortier, G. Kiekens, R. P. Auerbach, P. Cuijpers, K. Demyttenaere, J. G. Green, M. K. Nock, and R. C. Kessler, “Mental health problems in college freshmen: Prevalence and academic functioning,” *Journal of Affective Disorders*, vol. 225, pp. 97–103, 2018.
2. Mahmood, D. Y. (2017). Classification trees with logistic regression functions for network based Intrusion Detection System. *IOSR Journal of Computer Engineering*, 19(03), 48–52. <https://doi.org/10.9790/0661-1903044852>
3. “Social consequences of psychiatric disorders, I: Educational attainment” (1995) *American Journal of Psychiatry*, 152(7), pp. 1026–1032. Available at: <https://doi.org/10.1176/ajp.152.7.1026>.
4. Mamun, M. A., Hossain, M. S., & Griffiths, M. D. (2019). Mental health problems and associated predictors among Bangladeshi students. *International Journal of Mental Health and Addiction*, 20(2), 657–671. <https://doi.org/10.1007/s11469-019-00144-8>
5. Bhuiyan, M. A., Griffiths, M. D., & Mamun, M. A. (2020). Depression literacy among Bangladeshi pre-university students: Differences based on gender, educational attainment, depression, and anxiety. *Asian Journal of Psychiatry*, 50, 101944. <https://doi.org/10.1016/j.ajp.2020.101944>
6. Mei, G., Xu, W., Li, L., Zhao, Z., Li, H., Liu, W., & Jiao, Y. (2020). The role of campus data in representing depression among college students: Exploratory research. *JMIR Mental Health*, 7(1). <https://doi.org/10.2196/12503>
7. Masum, A. K., Abujar, S., Akter, S., Ria, N. J., & Hossain, S. A. (2021). Transformer based Bengali chatbot using general knowledge dataset. *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. <https://doi.org/10.1109/icmla52953.2021.00200>
8. Shukla, D. M., Sharma, K., & Gupta, S. (2020). Identifying depression in a person using speech signals by extracting energy and statistical features. *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. <https://doi.org/10.1109/sceecs48394.2020.60>
9. Deshpande, M., & Rao, V. (2017). Depression Detection Using Emotion Artificial Intelligence. *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. <https://doi.org/10.1109/iss1.2017.8389299>

10. Asad, N. A., Mahmud Pranto, M. A., Afreen, S., & Islam, M. M. (2019). Depression detection by analyzing social media posts of user. *2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON)*. <https://doi.org/10.1109/spicscon48833.2019.9065101>
11. Furlan, A.D. *et al.* (2009) “2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group,” *Spine*, 34(18), pp. 1929–1941. Available at: <https://doi.org/10.1097/brs.0b013e3181b1c99f>.
12. Ding, Y., Chen, X., Fu, Q., & Zhong, S. (2020). A depression recognition method for college students using deep integrated support vector algorithm. *IEEE Access*, 8, 75616–75629. <https://doi.org/10.1109/access.2020.2987523>
13. Dahiwade, D., Patle, G., & Meshram, E. (2019). Designing disease prediction model using machine learning approach. *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. <https://doi.org/10.1109/iccmc.2019.8819782>
14. Ignatiev, N., Smirnov, I., & Stankevich, M. (2022). Predicting depression with text, image, and profile data from Social Media. *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods*. <https://doi.org/10.5220/0010986100003122>
15. Hamaideh, S. H. (2017). Alexithymia among Jordanian University students: Its prevalence and correlates with depression, anxiety, stress, and Demographics. *Perspectives in Psychiatric Care*, 54(2), 274–280. <https://doi.org/10.1111/ppc.12234>

DEPRESSION DETECTION OF UNIVERSITY STUDENTS' USING MACHINE LEARNING APPROACH

by Fahmida Tafsirul_rashedul

Submission date: 22-Jan-2023 07:24PM (UTC+0600)

Submission ID: 1996905021

File name: Y_STUDENTS_USING_MACHINE_LEARNING_APPORACH_REPORT_plagiarism.pdf (744.58K)

Word count: 6720

Character count: 35898

DEPRESSION DETECTION OF UNIVERSITY STUDENTS' USING MACHINE LEARNING APPROACH

ORIGINALITY REPORT

22% SIMILARITY INDEX	14% INTERNET SOURCES	7% PUBLICATIONS	14% STUDENT PAPERS
--------------------------------	--------------------------------	---------------------------	------------------------------

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to Daffodil International University Student Paper	2%
3	Submitted to Waterford Institute of Technology Student Paper	2%
4	pada.psycharchives.org Internet Source	1%
5	Submitted to Gisma University of Applied Sciences GmbH Student Paper	1%
6	Submitted to Baze University Student Paper	1%
7	Submitted to University of Teesside Student Paper	1%
8	Submitted to University of Northampton Student Paper	1%