

**DEPRESSION DETECTION IN SOCIAL MEDIA COMMENTS DATA USING  
MACHINE LEARNING ALGORITHMS**

**BY**

**ZANNATUN NAYEM VASHA  
ID: 191-15-12939**

**BIDYUT SHARMA  
ID: 191-15-12720**

**MST. ISRAT JAHAN ESHA  
ID: 191-15-12921**

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

Supervised By

**Johora Akter Polin**  
Lecturer  
Department of CSE  
Daffodil International University

Co-Supervised By

**Ms. Nusrat Jahan**  
Sr. Lecturer  
Department of CSE  
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

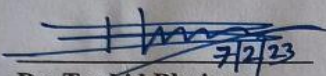
**DHAKA, BANGLADESH**

**JANUARY 2023**

## APPROVAL

This Project/internship titled “**Depression Detection In Social Media Comments Data Using Machine Learning Algorithms**”, submitted by Zannatun Nayem Vasha, Bidyut Sharma, and Mst. Israt Jahan Esha, ID No: 191-15-12939, 191-15-12720, and 191-15-12921 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment 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 29<sup>th</sup> January, 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**

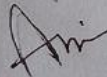


**Md. Abbas Ali Khan**

**Assistant Professor**

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

**Internal Examiner**



**Ms. Aliza Ahmed Khan**

**Senior Lecturer**

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

**Internal Examiner**



**Dr. Md. Sazzadur Rahman**

**Associate Professor**

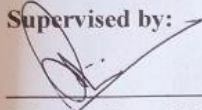
Institute of Information Technology  
Jahangirnagar University

**External Examiner**

## DECLARATION

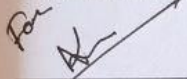
We hereby declare that, this project has been done by us under the supervision of **Johora Akter Polin, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

**Supervised by:**



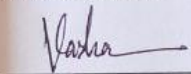
**Johora Akter Polin**  
Lecturer  
Department of CSE  
Daffodil International University

**Co-Supervised by:**

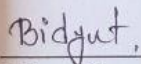


**Ms. Nusrat Jahan**  
Sr. Lecturer  
Department of CSE  
Daffodil International University

**Submitted by:**



**Zannatun Nayem Vasha**  
ID: -191-15-12939  
Department of CSE  
Daffodil International University



**Bidyut Sharma**  
ID: -191-15-12720  
Department of CSE  
Daffodil International University

Israt Jahan Esha

---

Mst. Israt Jahan Esha  
ID: 191-15-12921  
Department of CSE  
Daffodil International University

## 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 **Johora Akter Polin, 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 **Professor Dr. Touhid Bhuiyan, 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**

Nowadays, depression is a common and dangerous mental problem for our society, the country even the whole world. When a person is in a heartbreaking mood or going through an exquisite condition and it is not leaving him, trying to live alone, and giving him pain continuously is called depression. The last stage of depression is killing himself. According to WHO, currently, 4.4% of people worldwide suffer from depression. Many depressed people die almost every day. So, we will generate a model to find out who is suffering from depression and who is not. And finding depression is quite easy through our model. We collected huge data from Facebook, YouTube, and social media for the buildup models and learn to model and machine. Here we applied six classifiers to detect depression such as SVM, DT, LR, KNN etc. And when we are searching for which classifier gives the best accuracy then we see that the Support Vector Machine gives the best accuracy and which is 75%.

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	i
Declaration	ii
Acknowledgments	iv
Abstract	v
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-5</b>
1.1 Introduction	1
1.2 Motivation	3
1.3 Rationale of the Study	3
1.4 Research Questions	3
1.5 Expected Output	4
1.6 Report Layout	4
<b>CHAPTER 2: BACKGROUND</b>	<b>6-8</b>
2.1 Preliminaries/Terminologies	6
2.2 Related Works	6
2.3 Comparative Analysis and Summary	7
2.4 Scope of the Problem	8
2.5 Challenges	8
<b>CHAPTER 3: RESEARCH METHODOLOGY</b>	<b>9-20</b>

3.1 Research Subject and Instrumentation	9
3.2 Data Collection Procedure/Dataset Utilized	10
3.3 Statistical Analysis	11
3.4 Proposed Methodology/Applied Mechanism	13
3.5 Implementation Requirements	20
<b>CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION</b>	<b>21-28</b>
4.1 Experimental Setup	21
4.2 Experimental Results & Analysis	24
4.3 Discussion	26
<b>CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY</b>	<b>29-30</b>
5.1 Impact on Society	29
5.2 Impact on Environment	29
5.3 Ethical Aspects	30
5.4 Sustainability Plan	30
<b>CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE</b>	<b>31-33</b>
6.1 Summary of the Study	31
6.2 Conclusions	32
6.3 Implication for Further Study	33
<b>REFERENCES</b>	<b>34-35</b>



## LIST OF FIGURES

<b>FIGURES</b>	<b>PAGE NO</b>
Figure 3.1: Data collection	10
Figure 3.2: Comparison Models	11
Figure 3.3: Confusion Matrix for Logistic Regression	14
Figure 3.4: Confusion Matrix for Random Forest	15
Figure 3.5: Confusion Matrix for Decision Tree	15
Figure 3.6: Confusion Matrix for Support Vector Machine	16
Figure 3.7: Confusion Matrix for K Neighbors Classifier	16
Figure 3.8: Comparison of Accuracy, sensitivity, and specificity of different classifiers	18
Figure 3.9: Comparison Curve for Each Classifier Accuracy	18
Figure 3.10: Accuracy performance result from the dataset	19
Figure 3.11: Workflow diagram	20

## LIST OF TABLES

<b>TABLES</b>	<b>PAGE NO</b>
Table 4.2.1: F1 Measures of six classifiers	25
Table 4.2.2: Accuracy of Different Models in Percentage	26

# **CHAPTER 1**

## **Introduction**

### **1.1 Introduction**

Depression is a common and dangerous internal health condition that can affect people of all periods and walks of life. It's defined by patient passions of sadness, forlornness, and a lack of interest or pleasure in conditioning that was formerly enjoyed. Depression can also beget physical symptoms, similar to changes in appetite, sleep patterns, and energy levels, as well as difficulties with concentration, decision-making, and memory.

Depression is not simply a temporary feeling of sadness or "the blues," but rather a serious and often prolonged illness that requires treatment. It can interfere with a person's daily life, work, and relationships, and can lead to more serious health problems if left untreated. There are many different causes of depression, and it is often a combination of genetic, environmental, and psychological factors that contribute to the development of the disorder. Some people may be more at risk for depression due to a family history of the condition, a history of abuse or trauma, or certain medical conditions. If you think you may be experiencing depression, it is important to seek help from a mental health professional. Treatment for depression usually involves a combination of therapy, medication, and lifestyle changes, and can be very effective in reducing symptoms and improving quality of life. There are several different types of depression that a person may witness. Some common types of depression are given below:

**Major Depressive Disorder (MDD):** This is the most common type of depression and is characterized by persistent feelings of sadness, hopelessness, and a lack of motivation or enjoyment in activities.

**Persistent Depressive Disorder (PDD):** Also known as dysthymia, this type of depression is characterized by a long-term (two years or more) but less severe form of depression.

Bipolar Disorder: This type of depression is characterized by extreme mood swings, including periods of high energy and activity (called "manic episodes") and periods of low energy and activity (called "depressive episodes").

Atypical Depression: This type of depression is characterized by a range of symptoms, including an improved mood in response to positive events, excessive sleep, and weight gain.

It's important to note that depression can take many different forms and may have a variety of symptoms. It's also important to seek help from a mental health professional if you are experiencing symptoms of depression, as they can help you identify the type of depression you may be experiencing and develop a treatment plan.

There are many potential symptoms of depression, and they can vary from person to person. Here are some common symptoms that people with depression may experience:

- Persistent feelings of sadness, hopelessness, or emptiness
- Loss of interest in activities that were once enjoyable
- Difficulty concentrating or making decisions
- Changes in appetite and weight
- Changes in sleep patterns, such as difficulty falling asleep or staying asleep, or sleeping too much
- Decreased energy and fatigue
- Feelings of worthlessness or guilt
- Difficulty with relationships
- Thoughts of death or suicide

It's important to note that everyone experiences these emotions from time to time, and it's normal to feel down after a disappointment or a loss. However, when these feelings persist and interfere with daily life, it may be a sign of depression. If you are experiencing these symptoms and they are causing significant distress or impairment in your daily life, it is important to seek help from a mental health professional. They can help you determine the cause of your symptoms and recommend appropriate treatment options.

## **1.2 Motivation**

Motivation of our research work is to find out depressed people with the help of their posts on social media like Facebook, Instagram, etc. We collected raw data from social media and distributed them into two categories. By machine learning, we found & identified which post is depressive and which are non-depressive. And through that we can detect depressed people around us and can take necessary steps for the betterment of their minds so that we can grab them from depression & mental illness. There are several motivations for detecting depressive posts on social media. By detecting depressive posts, it may be possible to identify individuals who may be struggling with mental health issues and provide them with the necessary support and resources. Sometimes it may be possible to improve the quality of life for individuals who are struggling with the condition. By detecting them, it may be possible to create a more supportive environment that promotes mental well-being. It is important because depression is a common and potentially disabling condition that can have a significant impact on individuals and society as a whole. We worked on ten thousand of data & applied a machine learning process to continue the detection process. We trained 80% of the data and tested 20% to test the depressive & non-depressive posts.

## **1.3 Rationale of the Study**

Nowadays, most people like to spend their time on social media. People can't sleep, can't study, and can't work properly because of social media. Social media easily robs us of time. They suffer from depression when they use social media alone. Most people suffer from depression. We don't know who is suffering from depression or not. So, our research paper finds out who is depressed and who is not through machine learning.

## **1.4 Research Question**

1. Raw data collected
2. Depression detection easily
3. Section of Machine learning algorithms

#### 4. Model building using SVM

### **1.5 Expected Output**

Depression usually offers folks the time and want to replicate additional deeply regarding their lives and life path, providing them a chance to form some positive changes in their lives, Plant adds. Hopefully, they're going to realize some way to form necessary changes in their lives that may serve them higher over time. Depression could be a serious medical condition that's related to symptoms like melancholy, loss of enjoyment, loss of energy, problem concentrating, and dangerous thoughts. Depression is each an encephalopathy and a state of mind. Depression will render individuals disabled in their work life, family life, and social life. Left untreated, the emotional disorder is as pricey as cardiovascular disease or AIDS to the U.S. economy. Untreated depression is accountable for over two hundred million days lost from work every year. All of the problems we can recover slowly.

### **1.6 Report Layout**

Here we divided this report into six parts, which are listed below:

First of all, introduction part, this chapter discusses the study's introduction, motivation, expected outcomes, research questions, and project management and finance as well as the report's layout. After that, this chapter consists of comparative analysis and summary, preliminaries/terminologies, the scope of the problem, related work and the challenges of the research. Then we added Research Methodology, here firstly added research subject and instrument, after that we added data collection procedures or data utilization, then we added statistical analysis and the proposed methodology and finally, we added implementation requirements. Now experimental results and discussion portion, this portion is the most common and important part of our report. So, here firstly we are added to discuss the experimental setup then we have to add the experimental results analysis portion, and finally, we added the discussion portion. Our second last topic is impact on society, environment, and, sustainability. This topic is a normal part of our report and research. Firstly, here we discuss the impact on society, then we discuss the impact on the

environment, ethical aspects, and the sustainability plan. Our last and most important topic is the Summary, conclusion, recommendation, and implementation for future research. Here, we discussed within this portion a total summary of the study, conclusions, and implications for further study.

## **CHAPTER 2**

### **Background**

#### **2.1 Terminologies**

Anxiety disorders are among the foremost common and pervasive mental disorders within the United States. Depression could be a condition within which someone feels discouraged, sad, hopeless, uninterested, or far from life normally for quite a period, and once the sentiments interfere with daily activities. Depression could be a common ailment. Globally, it's calculable that fifty of adults suffer from depression. Depression could be a leading reason behind incapacity worldwide and could be a major contributor to the international burden of maladies.

#### **2.2 Related Works**

Several studies [6],[7] have emphasized the value of early diagnosis in enhancing treatment results for MDD. Early detection, as shown by Halfin's study [6], The emotional and financial consequences of this condition can be lessened with early detection, intervention, and therapy; in fact, Picardi et al. [7] found that participants who had received early screening had significantly less depressive symptoms and a higher quality of life. Early intervention for depression, as reported by Rost et al. [8], increases productivity and decreases absenteeism.

In the past decade, researchers have shifted their attention to studying social networks in order to better understand the prevalence of different diseases. Twitter, Inc. was proposed by Prieto et al. [19] as a means of automatically measuring the prevalence of a variety of health disorders. Chew and Eysenbach [21] utilized sentiment analysis on 2 million tweets to offer a supplementary intelligence technique, whereas Chunara et al [20] studied cholera-related tweets released during the first 100 days of the 2010 Haitian cholera outbreak. Alada et al. [22] analyzed messages for common linguistic patterns in an effort to avert suicide attempts. Development of cost-effective, acceptable, and population-focused therapies in depression is crucial, as shown by Rice et al [23]. Several online therapies (for both prevention and treatment) have been successfully piloted with young



people. Several research efforts [24–28] have examined the feasibility of using social media for the early diagnosis of mental health problems.

### **2.3 Comparative Analysis and Summary**

Depression has been the subject of several investigations. De Choudhury et al. [15] used social activity, mood, and linguistic signals displayed on Twitter, Inc. to establish a social media depression index, whereas Nadeem's bag of words study of Twitter, Inc. posts looked at the frequency of usage of my and me as a marker for depression. Additionally, utilizing topic modeling and rule-based approaches [31-33], participants in a task set at the Computational Linguistics and Clinical Psychology Workshop 2015 were able to successfully diagnose sadness and other mental health conditions in individuals using tweets from Twitter, Inc. Earlier diagnosis of depression has received less attention from researchers.

Although no approaches are offered by the authors, Ophir et al. [34] analyzed depression signals among teenage Facebook, Inc. users in the hopes of one day applying their coding methodology to early detection techniques. To create a tool for predicting and evaluating MDD in people, De Choudhury et al. [15] combined scores from the Center for Epidemiologic Studies Depression Scale [35] and BDI [36] with engagement patterns and language cues preceding a recent bout of depression among Twitter, Inc users. Several specific characteristics of posting activity were discovered in this study as being connected with the onset of depression. Less talking to other people, more time spent thinking and talking about one's own problems, and increased use of language describing despair. Unfortunately, the study relied on self-reported instances, as do the vast majority of studies that seek to predict depression, and methods that aim to detect people who are not yet aware that they have depression are still uncommon [28]. In addition, an early detection evaluation was not conducted by the authors of this study.

In the 2017 Conference and Labs for the Evaluation Forum session on early risk prediction on the internet (risk) [37], the authors offered a task on the early detection of depression using a time-aware technique and efficacy measures; this work is directly connected to our

study. The majority of the techniques relied on some combination of lexical, linguistic, semantic, and statistical aspects. We employed workshop techniques [13,37] and compared our results to those of the highest performing approaches [38,39]. In contrast to Villegas et al., who explicitly modeled partial information from the semantic representation of documents using learning algorithms like random forest (RF) or naive Bayes, Trozsek et al. [38] based their model on linguistic metainformation extracted from the subjects' writings and developed a classifier using recurrent neural networks. We use the same assessment approach as these other research, but our dual-model proposal and focus on different WFs sets it apart.

## **2.4 Scope of the problem**

You can facilitate outlining the scope of your downside by asking questions about what's operating and what is not operating. Scoping the downside facilitates the creation of a decent problem statement, that successively allows the creation of a testable hypothesis. This hypothesis forms the muse of the MVP, causing you down a path of repetitious learning and improvement toward the last word goal: finding a crucial drawback for your users.

## **2.5 Challenges**

In our study, the biggest challenge was the collection of the datasets and the learning machine. It was tough to learn dataset in machine.

## CHAPTER 3

### Research Methodology

#### 3.1 Research Subject and Instrumentation

Our research subject is depression detection in social media comments data using machine learning algorithms. Basically, we collect data from Facebook, Twitter, and YouTube. We are collecting data around 9230 where 5230 is depressed data and 4000 is non-depressed data. First of all, we kept all the data in ms excel. Then trained all the data through machine learning. Then we apply many algorithms such as naïve Bayes, logistic regression, KNN, decision tree, support vector machine, and random forest. And finally, these algorithms are done by google colab. Collaboratory, or “Colab” for brief, maybe a product from google analysis. Colab permits anybody to write down and execute discretional python code through the browser and is a particularly similar temperament to machine learning, information analysis, and education. With its straightforward and easy-to-use interface, Colab helps you start together with your information science journey with nearly no setup. If you are inquisitive about information science with Python, Colab could be a good spot to kickstart your information science comes without fear regarding configuring your setting. For the twenty-five GB choice to kick in, you initially ought to deplete all the session's memory. All twelve GB of it. Once you are doing that, the session can crash and Colab can show the subsequent at an all-time low on the screen: you've got exhausted all RAM! Google Colab is a wonderful tool for deep learning tasks. it's a hosted Jupyter notebook that needs no setup and has a wonderful free version, which supplies free access to Google computing resources like GPUs and TPUs. With its easy and easy-to-use interface, Colab helps you start together with your information science journey with virtually no setup. If you are inquisitive about information science with Python, Colab may be an excellent place to kickstart your information science comes without fear concerning configuring your setting. Colab professional limits RAM to thirty-two GB whereas Pro+ limits RAM to fifty-two GB. Colab professional and Pro+ limit sessions to twenty-four

hours. Colab professional doesn't offer background execution, whereas Pro will. Colab professional and Pro don't supply a full version of Jupyter Lab.

### 3.2 Data Collection Procedure/Dataset Utilized

5220	প্রকৃত সুখের সংজ্ঞা আজ আর্কি কেউ দিতে পারেনি, বিখ্যাত মনী	dep	1
5221	এক টুকরো সুখ খুঁজে পেতে মানুষ কত কিছুই না করে। সুখের জ	dep	1
5222	আমরা সবাই মূলত সুখের পৃথিবীতে বাস করি। কিন্তু, সে সুখ কে	dep	1
5223	হায়রে সুখ! তুমি কোথায় থাকো? তোমাকে খুঁজতে খুঁজতে আমি	dep	1
5224	সৃষ্টি কর্তা কাউকে সুখ দিয়ে পরীক্ষা করে। আবার কারো থেকে স	dep	1
5225	সারা দুনিয়ায় আজ পর্যন্ত কেউ সর্বোচ্চ সুখী হতে পারেনি।	dep	1
5226	দুনিয়ায় তো সর্বোচ্চ সোখ পাওয়া যায় না। এটা পেতে হলে পর	dep	1
5227	সুখ! আর কত সহস্র বছর কাটলে আমি তোমার দেখা পাব?- বল	dep	1
5228	সুখের জন্যই মরে কেউ, কেউ বা সুখের জন্যই বাঁচে। ওরে ভাই! শে	dep	1
5229	আমি সকল সুখ মিথ্যে মানি,	dep	1
5230	আমার অন্ধকার জগৎটা	dep	1
5231	খুব ভালো লাগলো	nondep	0
5232	আজও হাজারো ভালোবাসা বেঁচে আছে এই সমস্ত দুষ্টি মিষ্টি খন	nondep	0
5233	অসাধারণ	nondep	0
5234	অন্যরকম ভালোলাগায় মনটা ভরে গেল	nondep	0
5235	কারোর ভালোবাসার জীবনের কিছু অসাধারণ মুহূর্ত।	nondep	0
5236	দুষ্টি মিষ্টি রাগ, অভিমান এর কারণেই সম্পর্ক গুলি আরো শক্ত	nondep	0
5237	যেনো প্রেমিক প্রেমিকার মনের কথা কে ভাষায় রূপ দিয়ে অপ	nondep	0
5238	আমার প্রিয় ভালবাসার মানুষটির প্রিয় গান।	nondep	0
5239	এক কথায় অপূর্ব	nondep	0
5240	মন ছুঁয়ে যাওয়া	nondep	0
5241	হারিয়ে যাবো আমি তোমার সাথে।	nondep	0
5242	আহা উদ্ভাপ কত সুন্দর তুই খারমোমিটারে এ মাপলে	nondep	0
5243	পুরোটাই ভালোবাসা	nondep	0
5244	সুখের কোনো পরিসীমা নেই	nondep	0

Figure 3.1: Data collection

We collect all data from social media such as YouTube, Facebook, Instagram, and Twitter. First of all, we visit YouTube. YouTube has been verified to be a good instructional tool because it connects academicians, educators, and researchers from everywhere on the planet and supplies fascinating, knowledgeable, and fascinating content that has added a replacement dimension to education by creating it innovative still as inventive. We collect lots of data from YouTube. Here, we search for lots of depressing movies, lots of depressing songs, and lots of comments. Sometimes, we saw the full movie for some data which is based on depression. Most of the time, we listen to lots of songs, and we used to memorize song lyrics which is based on depression. Some depressed people leave comments in movies and song comment boxes. So, we collect data also comment boxes. Secondly, we visit Facebook. Students' interaction via Facebook will permit them to debate course contents at any time they have. Via Facebook, a student will post queries, share data and ask peers once facing any difficulties throughout their study time or throughout the

preparation of their assignments. Most of the data we collect is from Facebook because Facebook is very popular in our country. People easily share anything on Facebook. That's why most of the data we collect is from Facebook. Here, we collect data from different posts, pages, and videos.

And we also collect data from Instagram and Twitter. If you are looking for a social network, Instagram is a good community. folks visit and share photos and videos with folks in their online network. They additionally socialize by chatting on DMs in little teams. you ought not even to share a photograph or video to talk in an exceedingly DM and connect with folks in your network that approach.

Outside of the room, Twitter has provided several helpful extra benefits to students and academics alike, leaving skilled development and networking opportunities, co-curricular learning, support at conferences, larger data sharing, and overall accessibility.

### 3.3 Statistical Analysis

Models		Precision	Recall	F1 measure
Support Vector Machine	Depression	0.77	0.80	0.78
	Non-Depression	0.73	0.69	0.71
Logistic Regression	Depression	0.69	0.92	0.79
	Non-Depression	0.82	0.47	0.60
Decision Tree	Depression	0.71	0.69	0.70
	Non-Depression	0.62	0.63	0.62
Random Forest	Depression	0.74	0.79	0.77
	Non-Depression	0.71	0.65	0.68
Naïve Bays	Depression	0.69	0.92	0.79
	Non-Depression	0.82	0.47	0.60
K-Neighbors	Depression	0.66	0.77	0.71
	Non-Depression	0.63	0.49	0.55

Figure 3.2: Comparison Models

The closeness of 2 or additional measurements to every different is thought because of the preciseness of a substance. preciseness in scientific investigations is very important so as to confirm we tend to have gotten the proper results. Since we tend to generally use models or samples to represent one thing abundant larger, little errors are also exaggerated into massive errors throughout the experiment.

The recall is worked out because of the account dealings between diversity the quantum the composition of Positive samples duly classified as Positive to the whole number of Positive samples. The recall measures the model's capability to discover Positive samples. the upper the recall, a lot of positive samples were detected. Recall whereas high significance on breaking the number of false negatives, for illustration, positive cases that area units misclassified by the model as negatives. recall, in psychological science, is the act of retrieving info or events from the past whereas lacking a selected cue to assist in retrieving the data. an individual employs recall, as an example, once reminiscing a couple of vacations or reciting a literary work when hearing its title.

The recall is calculated because of the magnitude relation between variety the amount the number of Positive samples properly classified as Positive to the whole number of Positive samples. The recall measures the model's ability to discover Positive samples. the upper the recall, a lot of positive samples were detected. Recall places high importance on reducing the number of false negatives, for example, positive cases that area units misclassified by the model as negatives. recall, in psychological science, is the act of retrieving info or events from the past whereas lacking a selected cue to assist in retrieving the data. an individual employs recall, as an example, once reminiscing a couple of vacations or reciting a literary work when hearing its title.

Our trial has been supervised on fully dissimilar terms of point birth, victimization varied bracket algorithms (SVM, DT, LR, RF, NB, KNN). Our coaching and testing samples are depressed and non-depressed wisdom. For gaining optimum bracket results, the TF- IDF point synthesis is capitalized. Table one summarizes the fineness rate, Recall rate, and F1 Score for all algorithms. we have measured the fineness rate, Recall rate, and F1 Score for individual depressive & non-depressive knowledge.

### 3.4 Proposed Methodology/Applied Mechanism

A confusion matrix could be a table that's typically used to describe the performance of a classification model (or "classifier") on a group of data. The confusion matrix itself is comparatively easy to know, however, the connected nomenclature will be confusing. To see this rate, divide the entire range of negative outcomes you are expecting properly by the quantity of actual negative outcomes you get in your analysis. The name stems from the actual fact that it makes it straightforward to visualize whether or not the system is confusing 2 categories. A confusion matrix may be a figure of vaccination results on a bracket debit. The volume of accurate and wrong prognostications area units is epitomized with count values and countermined by every order. This is often the key to the confusion matrix. is confused once it makes predictions.

Confusion matrices are accustomed visualize necessary prognostic analytics like recall, specificity, accuracy, and exactitude. Confusion matrices are helpful as a result of they provide direct comparisons of values like True Positives, False Positives, True Negatives, and False Negatives. The confusion matrix is within the variety of a matrix wherever the column represents the particular worth's and also the row depicts the expected value of the model and contrariwise.

**Naïve Bayes:** Naïve mathematician is one every of the quick, straightforward milliliter algorithms to predict a category of collection of data. It may be used for Multi-class Classifications. It's the foremost widespread alternative for text classification issues.

**Logistic Regression:** Linear Regression could be a supervised regression model. supply Regression could be a supervised classification model. In regression toward the mean, we have a tendency to predict the worth by associate range number. In supply Regression, we have a tendency to predict the worth by one or zero. Here no activation operation is employed.

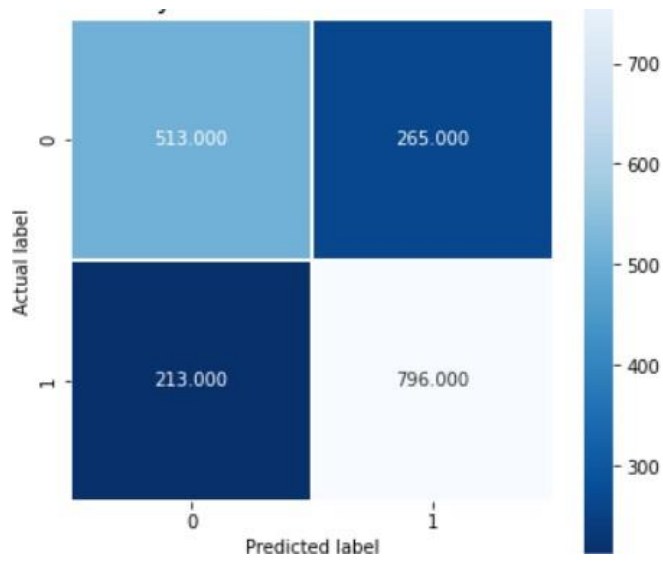


Figure 3.3: Confusion Matrix for LR classifier

Here, the TP rate is 513, the FP rate is 265, the FN rate is 213 and the TN rate is 796.

**Random Forest:** Random Forest is a supervised machine-learning process. Generally, it is used in classification and regression. Here, the TP rate is 523, the FP rate is 255, the FN rate is 197 and the TN rate is 812.

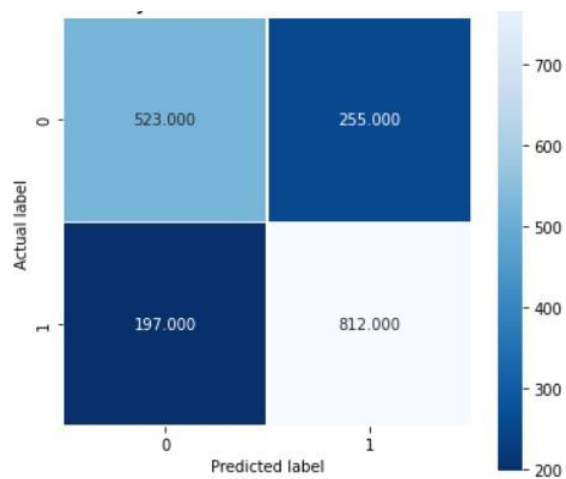


Figure 3.4: Confusion Matrix for RF classifier



**Decision tree:** A decision tree is a tree where it consists of so many nodes, links, and leafs. Here links define rules, nodes define and describes attributes, and leafs describes the outcome and categorical values.

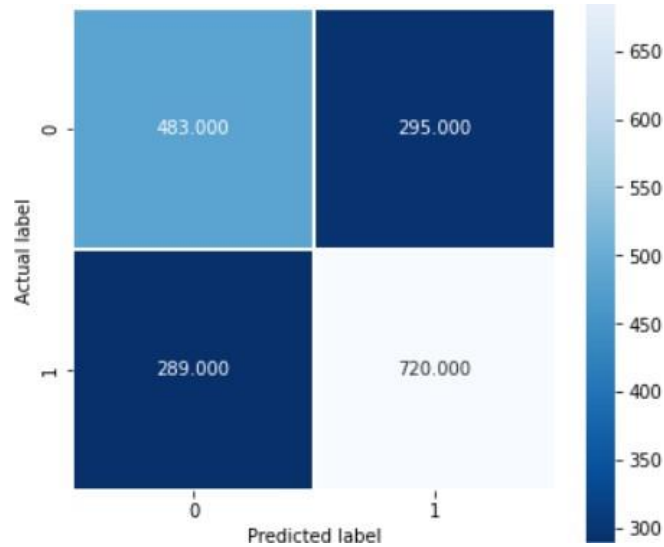


Figure 3.5: Confusion Matrix DT classifier

Here, the TP rate is 483, the FP rate is 295, the FN rate is 289 and the TN rate is 720.

**Support Vector Machine:** SVM square measure powerful however versatile supervised machine learning algorithms that square measure used each for classification and regression. However, typically, they're utilized in classification issues.

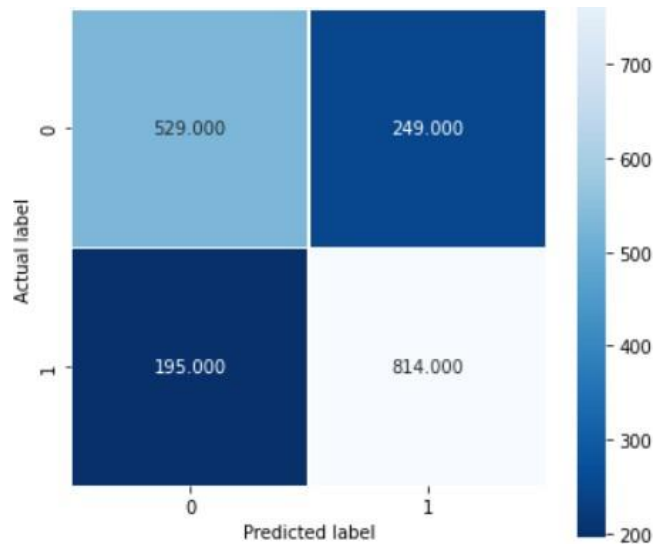


Figure 3.6: Confusion Matrix for SVM classifier

Here, the TP rate is 529, the FP rate is 249, the FN rate is 195 and the TN rate is 814.

**K Neighbors Classifier:** KNN classifier is an easy process to implement classifier and regression. It is also a supervised machine learning algorithm like as SVM, RF, LR classifiers.

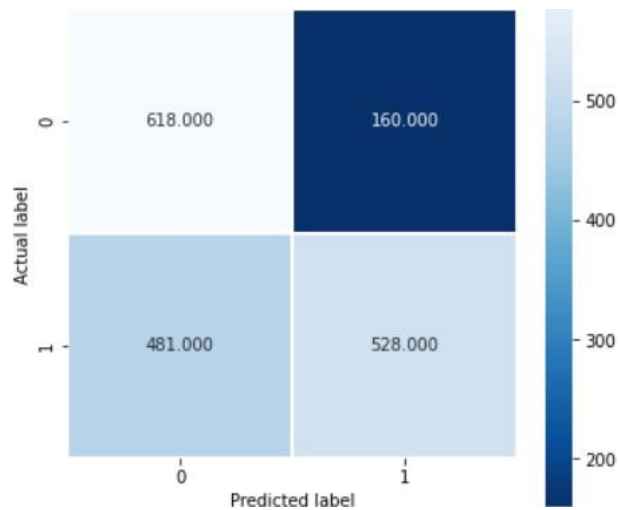


Figure 3.7: Confusion Matrix for KNN Classifier

Here, the TP rate is 618, the FP rate is 160, the FN rate is 481 and the TN rate is 528.

The 3 main metrics are used to judge a classification model: accuracy, precision, and recall. Accuracy is outlined because of the share of correct predictions for the task. It is often calculated simply by dividing the number of correct predictions by the number of total predictions. We have a tendency not to get 100 percent accuracy from your coaching dataset. If it does, it may mean that your model is overfitting.

Specificity itself may be delineated because of the algorithm/model's ability to predict a real negative of every class offered. In literature, it's additionally glorified merely because of the true negative rate. High specificity implies that the model is properly distinguishing most of the negative results, whereas an occasional specificity implies that the model is mislabeling tons of negative results as positive. Let's try to perceive this with the model used for predicting whether or not an individual is plagued by the illness. Specificity is that the proportion of verity negatives properly known by a diagnostic assay. It suggests however sensible the check is at distinguishing traditional (negative) conditions. Accuracy is the proportion of true results, either true positive or true negative, in a very population.

Here, below a chart is shown for accuracy, sensitivity, and specificity. Percentages are used along the y-axis and algorithms are used along the x-axis.

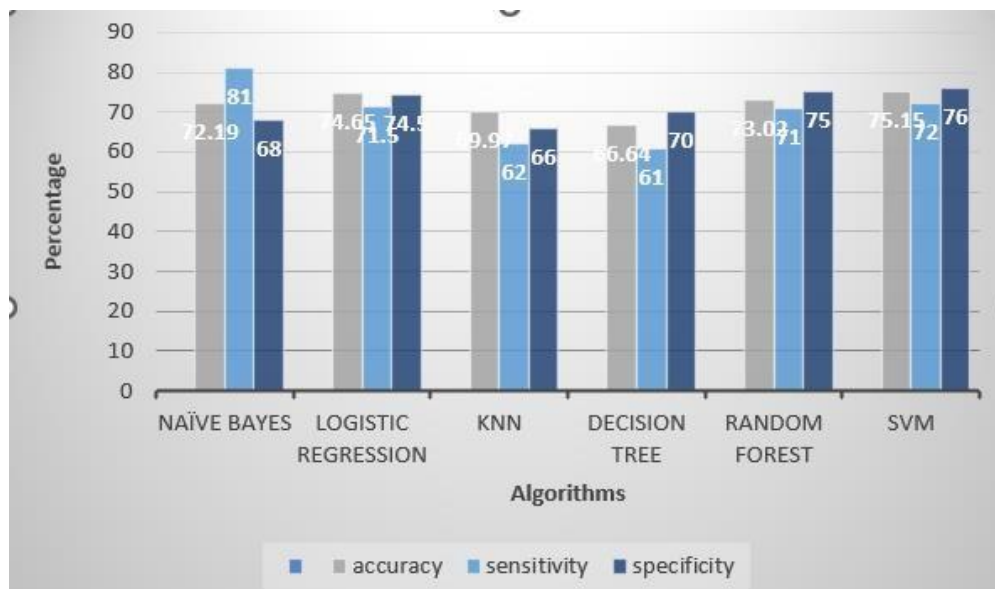


Figure 3.8: Showing the difference between Accuracy, sensitivity, and specificity  
 ©Daffodil International University

Here, we also added a curve and a bar chart for accuracy. We can easily say that from the curve and bar chart here the highest accuracy is 74.65% for Logistic regression and the lowest accuracy is 66.64% for the decision tree.

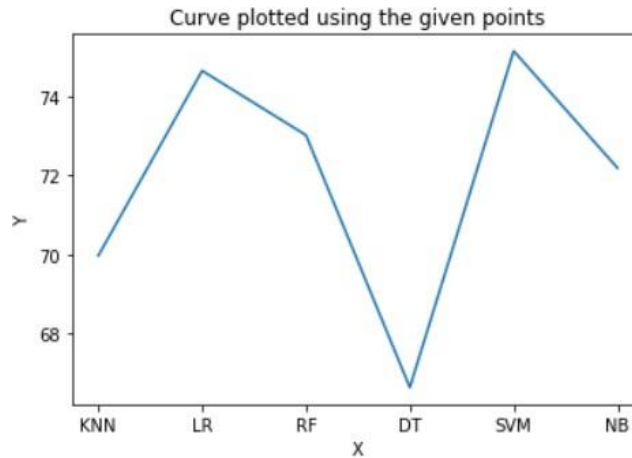


Figure 3.9: Comparison Curve for Each Classifier Accuracy

We also added curve and bar diagram for comparison of all classifiers. A simple chart is employed to represent information involving just one variable classified on a special, quantitative or temporal basis. during a straightforward chart, we tend to create bars of equal dimension however variable length, the magnitude of an amount is diagrammatical by the peak or length of the bars.

we can see that accuracy easily from a bar diagram or curve. Here, the accuracy of SVM is highest and the accuracy of the Decision Tree is the lowest.

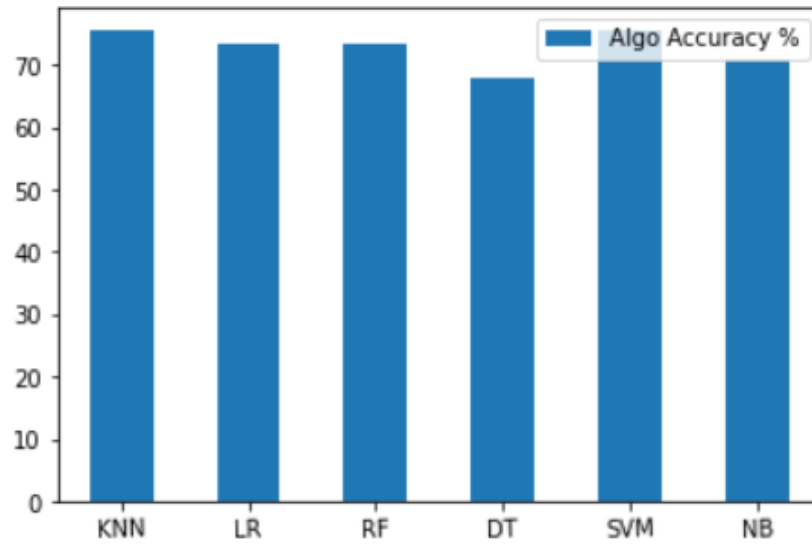


Figure 3.10: Accuracy performance result from the dataset

### 3.5 Implementation Requirements

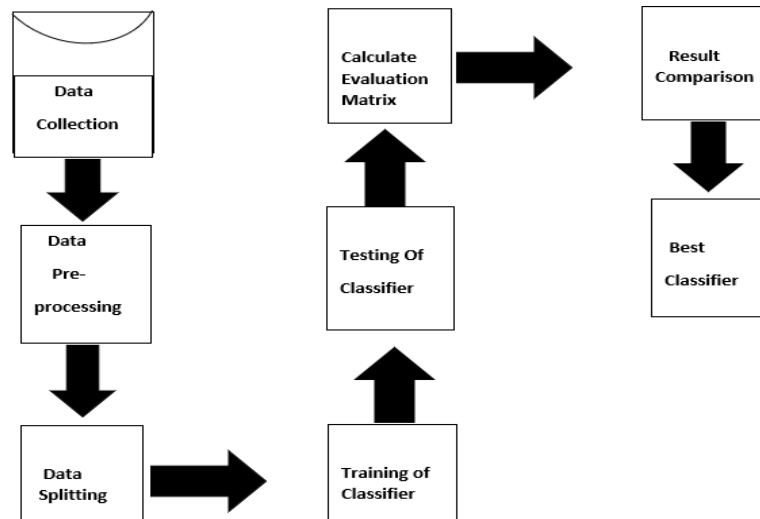


Figure 3.11: Workflow diagram

We've created a dataset of 10000 individuals' knowledge on depression and non-depression collected from Facebook and YouTube comments. Then we prepared for data pre-processing. Here we are using a TF-IDF vectorizer for data pre-processing. Data splitting works for translating Bangla data to binary data. After that, we are ready for training data, we are almost 80% data used for training and 20% data used for testing. Then we evaluate the confusion matrix for all classifiers. Then we compared the accuracy and found the best accuracy in all classifiers. And finally, we got the best classifier.

## CHAPTER 4

### Experimental Results and Discussion:

#### 4.1 Experimental Setup

We collected 10,000 independent and dependent variables on anxiety and quasi-anxiety from online text, responses, and status updates to generate a set of data. This represents one of the most well-known text analytics datasets.. There are some flaws in current work that aims to identify anxiety from social media comments.

1. Data gathering
2. Pre-processing
3. Split the data set
4. Training classifier
5. Compressor
6. Calculate
7. Testing

We also broaden the application of social media-based anxiety measures by describing the various characteristics of Facebook user reviews. We used machine learning techniques that could identify people who were depressed using those measurements.

Data gathering: We collected Facebook posts and comments totaling almost 10,000 in the Bengali language; we divided the data into training and testing, each receiving 80% of the data. Creating data for insights is the modeling's primary goal. Algorithms for machine learning are used to process data. The data is processed using a few models. To determine the effectiveness of large data sets, a distribution of data is used and from its output, we can calculate the values of recall, precision, and accuracy by using positive and negative dataset values. We had counted on one for depression and zero for no depression. These two datasets were combined for this study; if the combined dataset is unbalanced, the

missing values are tested, and the information is oversampled. After that, all the redundant data had been removed. In our dataset, about 300 duplicate records were discovered. The data was then divided into two sectors. Multiple data for machine learning is one area. Another segment is for test results. Training required 80% of the data and testing 20% of it. The quantity of instances with positive and negative data is first counted in the unbalanced dataset.

Data preprocessing: The dataset has gone through some preprocessing. To number the "Sentimental Column," we created a new column we've named "spam." We counted one when there was depression and zero when there was none. The data collection methods used by the social media platform present challenges for semantic analysis because they frequently contain errors or unnecessary text. Emoji processing is not required because the dataset we are using is emoji-free. Second, a classifier is used to perform the task of removing stop words. The list of stop words is available for download, and stemming is used to build systems by removing suffixes or prefixes that are used with the word. A snowball stemmer differs from a porter stemmer in the study because it can perform multiple language stemmers.

Modeling: The model's overall predictive performance will suffer as a result. Oversampling data techniques are used for imbalanced data modeling. The process of creating features that machine learning algorithms use to find patterns, and extract them to make them acceptable for algorithms for machine learning aid in prediction, is known as feature engineering. The term frequency (term frequency) and document term frequency were utilized by users and inverse document frequencies for feature extraction in the Data Preprocessing section. We labeled our set of data for machine learning using both vectors. Following that, data was labeled for machine learning in TF and IDF vectorizer. Next, we applied k-neighbors (K-nearest neighbors), multinomial naive Bayes, support SVM classifiers, logistic regression, logistic models, random forests, and decision tree classifiers to detect melancholy writings based on responses.



1. Decision tree: Splits are assessed using measuring entropy [14]. A decision tree is a decision-support tool that considers resource cost, utility, and event outcome probabilities. In statistical methods, machine learning, and predictive modeling, it is employed.
2. Random forest algorithm: The dataset is broken up into several distinct trees and organized into a class in the Random Forest Algorithm by similarities in their functionalities. The final class is then decided by a majority vote from among all the other classes[1]
3. logistic algorithm: By establishing a decision boundary, data can be categorized. A different type of linear regression model is logistics regression. It is comparable to ordinal logistic regression and ordinary logistic regression, but less understandably simple[1]
4. Naive Bayes: It is a very well-liked algorithm for text categorization and is very scalable. It is a very well-liked supervised classifier algorithm in the machine learning industry.[1]
5. Support vector machine: One of the best-known algorithms for supervised learning is called the support vector machine (SVM), and it is used to solve both classification and regression issues.

The information is represented within the context of a problem, typically a commercial problem, using a recognized mathematical equation. Creating data for insights is the main goal of modeling. Algorithms for machine learning process data. Some models are utilized to process the data.

Performance Measure: Dataset performance is measured using the confusion matrix, and the following results are obtained: validity, high accuracy, recall, F1 ratings, the confusion matrix, and curves. Precision is a measure of how accurate, sophisticated, and quick the model is at foretelling a specific section. The frequency with which the model was able to identify a particular category is measured by a recall. The precision and recall average make up the F1 score or measure.

Accuracy: Accuracy is one metric for rating classification models. Accuracy is the proportion of accurate estimates that our simulation tool makes. The total number of estimates and the number of correct predictions are equal.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1 Score} = \frac{2 * \text{P} * \text{R}}{\text{P} + \text{R}}$$

$$\text{TP Rate} = \frac{\text{TP}}{\text{AP}}$$

$$\text{FP Rate} = \frac{\text{FP}}{\text{AN}}$$

Confusion Matrix = T N F P F N T P

## 4.2 Experimental Results & Analysis

Using a variety of classification algorithms, our experiment was run on various feature extraction terms. Using the unbalanced datasets,

- 1 Performs depression detection.
2. The TF-IDF oversampling technique has been used to address the training dataset's class imbalance problem.
3. Both depressed and non-depressed data are included in our training and testing samples. The TF-IDF feature combination is utilized to achieve the best classification outcomes.

Table 4.2.1: F1 Measures of six classifiers

<b>Model</b>		<b>F1 measure</b>
Support Vector Machine	Depression	0.78
	Non Depression	0.71
Logistic Regression	Depression	0.79
	Non Depression	0.60
Decision Tree	Depression	0.70
	Non Depression	0.62
Random Forest	Depression	0.77
	Non Depression	0.68
Naïve Bays	Depression	0.79
	Non Depression	0.60
K-Neighbors	Depression	0.71
	Non Depression	0.55

Analysis: Sadness can be a dangerous medical disease, especially if it lasts for a long time and is mild to severe, in contrast to anxiety attacks brought on by routine tasks. Fortunately, despite being a quiet condition, depression leaves some signs due to excessive use of social media. These tips can be found in documents that have been shared on social media. Then, writing patterns on social media sites such as Facebook, Instagram, and Twitter were examined to see if they were associated with depressive episodes. Along with a binary classification, accuracy, precision, and recall measurements were employed to examine the

outcomes. Using various machine learning techniques, such as sentiment classification, sadness is detected through social media platforms like Twitter and other websites. Through platforms such as Facebook, Twitter, and other media sites, depression is evaluated using various machine learning approaches, including emotion analysis and processing, to determine whether a person is sad or nervous.

Table 4.2.2: Accuracy Different Models in Percentage

<b>Model</b>	<b>Accuracy</b>
SVM	75.15%
LR	74.65%
DT	66.64%
RF	73.02%
NB	72.19%
KN	69.97%

The highest taste rate and F1 results for the super-vector model for depressive and non-depressive data, respectively, are 0.78 and 0.77. There are several factors that affect the model's accuracy, and every characteristic affects classification in a different way. According to our findings, SVM methods on the same test data are the ones that provide the best accuracy.

### **4.3 Discussion**

Depression, sometimes known as depressive symptoms, is a serious medical disease that frequently has an impact on a person's feelings, thoughts, and actions. A chronic melancholy and a loss of interest are two features of depression, a mood illness. It is sometimes referred to as "depression," a mood disorder that impacts someone's thoughts, emotions, and behavior. Numerous emotional and physical problems may result from it.

The top three things that cause depression are as follows:

1. Disease
2. Alcohol and drugs
3. Personality

Despite mounting evidence that university students experience mental health problems, this topic has not received enough attention in low-and middle-income nations like Bangladesh. The attempt to evaluate depression and its health conditions made in this study may help close the knowledge gap since depression has been identified as the most significant factor influencing suicide.

Due to feelings of worthlessness, being worthy of death, or an inability to handle the pain of despair, thoughts of suicide are prevalent in serious depression—and suicide is a common form of self-harm. Major depression is most common in people aged 45 to 65. Center-aged adults have the highest rates of depression, though the very young and very elderly may have higher rates of severe depression than those in the middle of the bell curve.

In particular, if a recent stressful incident sets off the disease, those who have already experienced sadness are more likely to undergo another episode. Depression is a common side effect for people with chronic medical problems such as cardiac disease, diabetes, cancer, or chronic pain. It's critical to understand that depression is treatable. Frequently, simply recognizing it is the first step toward feeling better. That the issue is present Risk Elements Genetics If your family has a history of the illness, depression may be more likely to affect you. Sadness and grief are common responses to death or loss. Personal strife or disagreements with family or friends can cause depression. Physical, sexual, or emotional abuse in the past may also be a factor. Genealogy and genetics Depression is known to be significantly influenced by genetics.

2 or 3 times as likely as someone without a family history of the illness to experience depression if a parent, sibling, or other close relative has the illness. In depression psychopathology, patients' actions and thoughts can be influenced by their depressive mood, which in turn influences their moral judgments.

A widespread problem is depression. Even though millions of people experience depression, many do not seek treatment, placing a burden on public health. There are many

obstacles to treatment, such as distance from a therapist and the frequent requirement to try several medications before one is discovered to be effective. Eight strategies to enhance your mental wellness:

- Prioritize making in-person connections with people.
- Keep being active
- Contact someone
- Make use of your senses
- Start your relaxation routine
- Give thought and leisure top priority
- To keep your brain healthy, eating a brain-healthy diet
- Don't neglect your need for sleep

People can now express their emotions on a large scale through social media. Large data sets can be analyzed by Machine Learning algorithms, which can be used to perform additional computations and generate predictions or suggestions.

## CHAPTER 5

### Impact on Society, Environment and Sustainability

#### 5.1 Impact on Society

Depression could be a leading reason behind incapacity worldwide and could be a major contributor to the international burden of illness. A lot of ladies square measure plagued by depression than men. Depression will result in suicide. There is an effective treatment for gentle, moderate, and severe depression. Since depression is therefore current, all social staff, in spite of their follow setting and specialization, should be conversant in varied aspects of the syndrome. Depression has been a social drawback throughout history.

It may cause you to withdraw from your social relationships even from individuals to whom you're highest. folks that square measure depressed will realize it tough to feel on the brink of anyone, and might even believe that their friends don't like them. they'll additionally feel a burden to their friends and family, creating it more durable to raise - and settle for - facilitate

#### 5.2 Impact on Environment

Due to the Earth's rotation, this movement happens in an exceedingly volute manner. The movement of air around an unaggressive center is named depression. Such a depression ends up in severe downfall and thunderstorms. for instance, depression or misuse will cause state, which successively will cause the financial condition, poor nutrition, and every one of their associated environmental troubles. Similarly, mental state conditions like signboard may end up in environmental issues. Studies estimate that between 37-48% of vulnerability for depression stems from biology, going away from the setting to account for doubtless over 1/2 the remaining influence.

Spending time in nature has been found to assist with psychological state issues like anxiety and depression. for instance, analysis into ecotherapy (a form of formal treatment that involves doing activities outside in nature) has shown it will facilitate with delicate to

moderate depression. For instance, analysis studies reveal that rooms with bright lightweight, each natural and artificial, will improve depression and anxiety.

### **5.3 Ethical Aspects**

The clinical care of depression combines the associated degree of urgency of crisis with the deeply experienced and pervasive feelings of despair and unhappiness among patients, raising several moral problems associated with the necessity for patient safety, the acceptable treatment of sickness, and therefore the restoration of individual self-agency. Ethical problems with beneficence, nonmaleficence, confidentiality, altruism, justice and nondiscrimination, expertise, trust, and different on the face of it abstract ideas frequently emerge within the daily work of psychiatrists.

Similar to alternative branches of medication, medical specialty conjointly needs analysis to be conducted ethically, safeguarding the rights and privileges of the analysis participants. Ethics deals with the suitable manner of conduct of execs and provide the guiding principles of how choices ought to be taken.

### **5.4 Sustainability Plan**

Choosing activities that cause you to feel smart and doing stuff you relish will assist you to feel less depressed. pay time together with your pet, pay time engaged on comes, or visit a calming place area unit sample of this evidenced technique. 1. invariably take the medication as prescribed, daily. victimization standardized scales just like the GAD-7 for anxiety or the PHQ-9 for depression offer the North American nation the most effective likelihood of measuring whether or not treatment is functioning for a selected patient.



## CHAPTER 6

### Summary, Conclusion, Recommendation and Implication for Future

#### Research

##### 6.1 Summary of the study

Users connect with friends who share their self-interest in virtual communities and share thoughts, images, and videos that convey unique emotions, attitudes, and feelings. This offers a chance to check on user attitudes, thoughts, and feelings when they converse using all these software resources by looking at their feelings and emotions in social network data.

Our research aims to identify depressed people through social media comments, posts, or texts. We gathered nearly 10,000 pieces of information from people posting on social networking sites and posting their opinions on videos online, etc. Although the use of experimental datasets to diagnose depression has been widely accepted, there are still a number of unrecognized aspects. In our experiment, predictive analytics were employed to examine the information and calculate whether it was mentally unstable or otherwise. Data preparation Text processing, data extraction, and classification models are carried out gradually. We divided the nearly 10,000 posts and comments in the Bangla language that we collected into training and testing groups, with the remaining 20% of the data.

Under various conditions of feature extraction, our study has been running utilizing a range of classification algorithms (SVM, DT, LR, RF, NB, K-NN). The TF-IDF feature combination is used to achieve the best classification outcomes. We tested all the algorithms on the same datasets and our results show that if there is any tree that gives better accuracy, it is SVM. For both psychological distress and probably non-data, we individually determined the result, leading to significance, and recall value. Here, we found that the greatest F1 score and precision rate for the SVM classifier are 0.78 and 0.77, respectively. A model's precision is affected by several variables, and each feature has a unique impact on classification. On the same dataset, we also put an SVM model into practice. Finding effective treatments for Facebook users' mental health issues using machine learning techniques.

Social media has become a popular platform for people to express their feelings or any inner thoughts. People are increasingly affected by depression, which is an epidemic disease. A person's depression can be recognized or suspected in text from a variety of social media sites, such as online comments and responses, Youtube clips, as well as other weblogs, using algorithms for machine learning. There are preprocessing procedures carried out, such as features are extraction, data preprocessing, categorization, and classifier implementation. Classifiers correctly recognize depression when given 1000 data points from different posts and remarks in various categories. Machine Learning is a powerful tool that can be used to solve complex problems or produce useful information about the world around us.

## **6.2 Conclusion**

Depression frequently causes discomfort, interference with everyday functioning, an increased risk of suicide, an increase in medical costs, and a decrease in output. Whether they develop on their own or in combination with other general medical problems, depressions can be addressed. These comprise, among other preparatory procedures, data preparation, information labeling, feature extraction, and classifier implementation. The systematic extraction of previously undiscovered and genuinely significant knowledge from huge data sets is known as "data mining." Machine learning and data mining are becoming important in identifying depression. In our investigation, DMML will be applied to assess a subject's level of depression. We gather data from a variety of sources and use specialized knowledge gained via data mining in order to develop a more effective approach and identify depressed people on social media. computer learning, In general, machine learning is a method of systematizing the evaluation of data. It forecasts a person's mental health based on information from our survey. When we gather ten thousand pieces of information from various posts and comments made on a range of online dating websites, anxiety may be accurately diagnosed by classifiers. We used machine learning techniques in this study to bring a prospective viewpoint to the research mentioned above.

### **6.3 Implication for Future Research**

To expand the scope of our research and offer more precise loneliness diagnosis, deep learning techniques can be used in the future. Deep learning algorithms may help us diagnose depression in people more accurately and discover more about the complex interactions that lead to the disorder.

## Reference:

- [1] Tummala, R.K., Bhuvanewari, E., John, T.J., Karthi, S.P. and Arjun, K.P., 2021. Depression detection using data mining algorithms from social media context. *Materials Today: Proceedings*.
- [2] Zohuri, B. and Zadeh, S., 2020. The utility of artificial intelligence for mood analysis, depression detection, and suicide risk management. *Journal of Health Science*, 8, pp.67-73.
- [3] Narziev, N., Goh, H., Toshnazarov, K., Lee, S.A., Chung, K.M. and Noh, Y., 2020. STDD: Short-term depression detection with passive sensing. *Sensors*, 20(5), p.1396.
- [4] Liao, S.C., Wu, C.T., Huang, H.C., Cheng, W.T. and Liu, Y.H., 2017. Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns. *Sensors*, 17(6), p.1385.
- [5] Ilgen, M.A., Downing, K., Zivin, K., Hoggatt, K.J., Kim, H.M., Ganoczy, D., Austin, K.L., McCarthy, J.F., Patel, J.M. and Valenstein, M., 2009. Exploratory data mining analysis identifying subgroups of patients with depression who are at high risk for suicide. *The Journal of clinical psychiatry*, 70(11), p.6977.
- [6] CACHEDA, F., FERNANDEZ, D., NOVOA, F.J. and CARNEIRO, V., 2019. Early detection of depression: social network analysis and random forest techniques. *Journal of medical Internet research*, 21(6), p.e12554.
- [7] Saqib, K., Khan, A.F. and Butt, Z.A., 2021. Machine Learning Methods for Predicting Postpartum Depression: Scoping Review. *JMIR mental health*, 8(11), p.e29838.
- [8] F.J. and Carneiro, V., 2018. Analysis and Experiments on Early Detection of Depression. *CLEF (Working Notes)*, 2125.
- [9] Haque, U.M., Kabir, E. and Khanam, R., 2021. Detection of child depression using machine learning methods. *PLoS one*, 16(12), p.e0261131.
- [10] Huang, Z., Epps, J., Joachim, D. and Sethu, V., 2019. Natural language processing methods for acoustic and landmark event-based features in speech-based depression detection. *IEEE Journal of Selected Topics in Signal Processing*, 14(2), pp.435-448.
- [11] Nanomi Arachchige, I.A., Sandanapitchai, P. and Weerasinghe, R., 2021. Investigating Machine Learning & Natural Language Processing Techniques Applied for Predicting Depression Disorder from Online Support Forums: A Systematic Literature Review. *Information*, 12(11), p.444.
- [12] Vieira, E.R., Brown, E. and Raue, P., 2014. Depression in older adults: screening and referral. *Journal of geriatric physical therapy*, 37(1), pp.24-30.
- [13] Giuntini, F.T., Cazzolato, M.T., dos Reis, M.D.J.D., Campbell, A.T., Traina, A.J. and Ueyama, J., 2020. A review on recognizing depression in social networks: challenges and opportunities. *Journal of Ambient Intelligence and Humanized Computing*, 11(11), pp.4713-4729.
- [14] AlSagari, H.S. and Ykhlef, M., 2020. Machine learning-based approach for depression detection in twitter using content and activity features. *IEICE Transactions on Information and Systems*, 103(8), pp.1825-1832.
- [15] Tadesse, M.M., Lin, H., Xu, B. and Yang, L., 2019. Detection of depression-related posts in reddit social media forum. *IEEE Access*, 7, pp.44883-44893.
- [16] Feightner, J.W. and Worrall, G., 1990. Early detection of depression by primary care physicians. *CMAJ: Canadian Medical Association Journal*, 142(11), p.1215.
- [17] Clarke, F.M., Morton, H. and Clunie, G.J., 1978. Detection and separation of two serum factors responsible for depression of lymphocyte activity in pregnancy. *Clinical and Experimental Immunology*, 32(2), p.318.
- [18] Chatterjee, R., Gupta, R.K. and Gupta, B., 2021. Depression detection from social media posts using multinomial Naive theorem. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1022, No. 1, p. 012095). IOP Publishing.
- [19] Wu, C.T., Dillon, D.G., Hsu, H.C., Huang, S., Barrick, E. and Liu, Y.H., 2018. Depression detection using relative EEG power induced by emotionally positive images and a conformal kernel support vector machine. *Applied Sciences*, 8(8), p.1244.
- [20] Smith, M.V., Rosenheck, R.A., Cavaleri, M.A., Howell, H.B., Poschman, K. and Yonkers, K.A., 2004. Screening for and detection of depression, panic disorder, and PTSD in public-sector obstetric clinics. *Psychiatric services*, 55(4), pp.407-414.

- [21] Low, L.S.A., Maddage, N.C., Lech, M., Sheeber, L. and Allen, N., 2009, August. Content based clinical depression detection in adolescents. In *2009 17th European Signal Processing Conference* (pp. 2362-2366). IEEE.
- [22] Islam, M., Kabir, M.A., Ahmed, A., Kamal, A.R.M., Wang, H. and Ulhaq, A., 2018. Depression detection from social network data using machine learning techniques. *Health information science and systems*, 6(1), pp.1-12.
- [23] Kumar, A., Sharma, A. and Arora, A., 2019, March. Anxious depression prediction in real-time social data. In *International conference on advances in engineering science management & technology (ICAESMT)-2019, Uttarakhand University, Dehradun, India*.
- [24] Bailey, A. and Plumbley, M.D., 2021, August. Gender Bias in Depression Detection Using Audio Features. In *2021 29th European Signal Processing Conference (EUSIPCO)* (pp. 596-600). IEEE.
- [25] Karmen, C., Hsiung, R.C. and Wetter, T., 2015. Screening internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. *Computer methods and programs in biomedicine*, 120(1), pp.27-36.

***Conference/Journal Papers:***

Zannatun Nayem Vasha, Bidyut Sharma, Israt Jahan Esha, Jabir Al Nahian and Johora Akter Polin, "Depression detection in social media comment data using machine learning algorithms," *Bulletin of Electrical Engineering and Informatics* Vol. 12, No. 2, April 2023, pp. 987~996 ISSN: 2302-9285, DOI: 10.11591/eei.v12i2.4182.

