

# Depression Management Exerting Sentiment Analysis

Ark Pandey

Department of Computer Science and Engineering  
RNS Institute of Technology Bangalore, India

**Abstract**— The paper embarks on predicting the sentiments of a human being and identifies the ones who are facing the state of Depression, following with the cure of this misery using some machine learning algorithms from a scientific perspective. The study suggests that out of every 3 individuals 2 suffers from depression and the major cause of this depression are the THREE VILLAINS of the current generation, and they are “overthinking”, “lack of confidence” and the most important is the “unknown fear”. These 3 factors make a healthy one depressed and their mind feels inferior to everyone. How can someone or some technology stop this from happening? It’s easy, there’s a study which states that over 50% don’t even know why they are a victim of this disease. And this paper is for those individuals who don’t know why they have those “unknown fear” which results in nothing but anxiety. There’s an emerging technology that can predict their cause of depression and even suggest the way to cure this deadly disease and help the victim, live a fruitful life. Nowadays people talk to their phones or their personal computers more than any other individual or even by themselves. So, this research is a modern one and will cover only the current techniques, a regular soul understands the most. The model will keep the track of the contents they post on their social media account, and the pages or handles he/she follows on the respective account. Which will cause a huge dataset about their personal life. The details got from those sources are more genuine and accurate than the data retrieved by communicating with the person himself/herself. The model discussed further will consider all the data obtained by the subscriber and train itself for the work it is assigned for and give the result, whether or not the subscriber is depressed (classification). And also states the reason for depression. Which will help him/her to conquer the depression and can live a happy life ever after.

**Keywords:** Sentiment Analysis, Depression, Classification, Twitter, Cure

## I. INTRODUCTION

Depression, the ordinary definition with an extraordinary significance. The concept of this research evolved after witnessing the wrath of this deadly disease in the youth, which is left undiagnosed, even the individual suffering from depression doesn’t know about his/her mental health. Depression could be in any form it could be because of a loss in business, failure in any work, death of some special ones and much more.

If the problem is mostly interconnected with the youth, then the source of the dataset used for the sentiment analysis should be from the site where modern youth spends most of their time on, Yes! It is social media sites and what could be better than twitter for this purpose. Twitter users express their mindsets, their ideas, their sentiments in the form of tweets so, is there a way to read a lot of tweets at once and analyze the sentiments behind those clusters of

words effectively and accurately with a computer algorithm? Not only that, but the algorithm should also give the remedy for the problem. It should be a one-stop destination for both the diagnostics and the remedial purposes. If the algorithm does both the tasks effectively and efficiently, then there is a need for a solution like this. Else there is no need for its existence.

The paper comprises 7 different sections.

- 1) It all commences with the history of other studies in the corresponding subject.
- 2) The post initial section demonstrates how depression could be endured using sentiment analysis.
- 3) This section reveals the fundamental method of sentiment analysis.
- 4) The section expresses the proposed solution, it commences from the twitter data extractions to the feature extraction.
- 5) The following section reports the cure for depression using a machine learning algorithm.
- 6) This part involves the applications and challenges of sentimental analysis.
- 7) The last section is all about the conclusion and future enhancements.

## II. LITERARY ASPECT

The humongous data extracted from the social media sites, which was once used by some random folks online, to express their emotions, sentiments, feelings, ideas, opinions and much more, are now used by the data scientists to analyse what exactly is running in the current generations’ head. They proposed many machine learning algorithms to analyse their sentiments and help them find what they need. There are many research papers out there which helps the customers to “shop well”, to “visit the company’s products”, to “help the community by giving them perfect feeds in their social media accounts”, and much more.

The researchers who previously stepped foot in this field faced some serious issues, as it is a tough job to handle this size of data with some pre-existing algorithms. So they studied in this aspect and try to give their contribution in the field of sentiment analysis.

They proposed it in the paper[1], the data set used here was a repository of Kaggle and UCI archive across 5 domains. The work was done by comparing most of the ML with some hybrid algorithms. The results show that the hybrid algorithms were higher in accuracy than that of the regular ones.

After a lot of study and surveys, I figured out that there is an interesting paper[2], which has a great idea and if executed correctly, could change the course of the entire machine learning future. Some machine learning algorithms fail, not because they are not good enough, but they don’t have the capabilities to handle redundancies and to fix this issue they used the technology of the Synthetic Minority Oversampling Technique or (SMOTE). Which will reduce

the redundancies in the data set provided for the machine learning model.

Many of the researchers tried to combine two or more machine learning algorithms to get even better results. But during my research period, I came across a great paper[3], this research tried to do things differently, instead of combining the ML algorithms, they combined the features itself. So the data provided to the machine learning algorithm will be even better and will cause less pain for the ML model to access the data or to manipulate the data.

There are some times when we have access to all the algorithms and can apply everyone, but it is still difficult to find the best algorithm for our work, and the burden of this issue was understood by a paper[4], which compared almost all sort of machine learning algorithms and given the results of the best algorithm of them all.

Another work was done with the focus on analysing short and informal text like tweets This work concentrated on casual sentences. The researchers developed an emotional feature generator using a fused method to figure out the sentiment of the text. The developed method was measured resistant to the majority common techniques in this area; Maximum Entropy, Naive Bayes and SVM. The results show that using sentiment features rather than conventional text analysis can yield higher accuracy.

Sentiment could also be analysed with shopping, a random person does. After receiving the product he/she could give three types of comments, positive, negative or neutral. And in case of online shopping, these are the exact 3 things a person can comment on. So how can a model predict the quality of the comment? There will be no man reading millions of comments each day, so the paper[5] discusses the special algorithm called SENTI which can do this job precisely, it can classify the comment given on an online platform, as positive, negative or neutral.

But do all the algorithms accomplishes the work they need to do? That was the question that led to this paper. The model presented in this paper should do all the processing perfectly to make the machine learning model efficient and robust.

### III. WHAT IS SENTIMENT ANALYSIS? AND COULD DEPRESSION BE MANAGED BY THE VIRTUE OF SENTIMENT ANALYSIS?

Piloting depression is an important step to be taken soon. This is the disease that, most of the time left unnoticed and is not taken seriously in today's world. Popular actors, sportsperson, politician and thousands of common individuals commit suicide with the impact of depression. They not only waste their energy but also wastes their valuable time, the time which could do something productive, which could change the course of the lifestyle they used to have. But depression spoils this all. It leads the protagonist to lose all of its energy and act lethargic which kills all of its potentials and ability to do any work, even follow its daily routine. After having great research on the aspect I came across much other research work and concluded that, Yes! We could manage Depression using the method of sentiment analysis.

A technique known as Natural Language Processing (NLP) can analyse sentiments of a twitter user each user tweets his/her own opinion on a random topic and that tweet could detect the sentimental status of the tweeter. How this could be done, is explained later in this paper. But as of now the nature of the sentiments could be any of the three, Positive, Negative or Neutral. If the tweet is Positive or Neutral, then it is well and good but if the nature is Negative, then there is a scope for a mental health issue. Most of the time the negative tweets are just representing a thought on a topic, But what about some tweets like the one shown in the following figure?

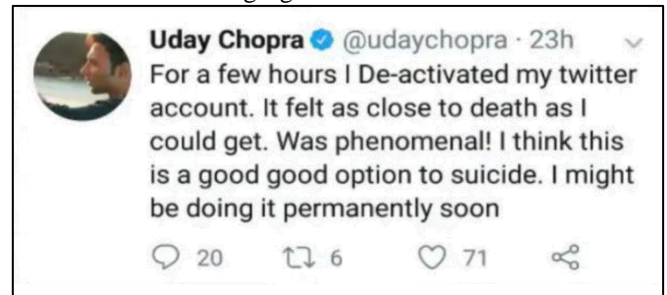


Fig. 1: Tweet by a celebrity

This is the tweet of a popular celebrity, named 'Chopra'. He posted this tweet on his twitter handle. By analysing this tweet one can conclude that he is in depression. But how can a machine learning model detect this tweet as a "depressed tweet". And this is the step where sentiment analysis comes in the picture, the process of sentiment analysis (discussed later) could declare the above tweet as a negative tweet, but how a depressed tweet? The negative tweets classified by the machine learning model are further classified as a Depressed or a Non Depressed tweet using some machine learning algorithms. All the algorithms proposed, further collaborates and give the selected tweets out of the lot which was posted by a depressed user.

### IV. THE FUNDAMENTAL METHOD

The algorithm for the sentiment analysis roams around the built-in library of the programming language, Python and it is called TextBlob. It offers an API to access its methods and perform basic NLP tasks The library takes the entire statement. Sentiment analysis is the process of determining the point of view or the emotion of the writer, i.e., whether it is positive or negative or neutral. The sentiment function of textblob returns two features, polarity, and subjectivity.

Polarity is a float that is in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. The amount of polarity determines the emotion and the sentiment of the writer. For example, a person has written a review for some random movie as "Very bad acting and screenplay". Suppose the polarity of this sentence would be -0.63. It is clear from the values that it is a negative emotion and against the brand image of the Movie. This is how the polarity can find the end sentiment behind the statement.

Subjectivity helps in determining personal states of the speaker including Emotions, Beliefs, and opinions. polarity alone is not efficient enough to handle complex text sentences. Sometimes the sentence needs more attribute analysis to check whether it is describing features or

opinions on some object. It has values from 0 to 1 and a value closer to 0 shows the sentence is objective(i.e. based on facts) and a value closer to 1 shows the sentence is subjective(i.e. not based on facts) Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float that is in the range of [0,1].

For example, there are two sentences:

The House has 6 rooms.

This House is really big.

In the first sentence, we have provided the house with an objective approach and described its features. On the other hand, the second sentence provides an opinion based on how the person found the home to be.

#### A. An Example

Let us consider the word 'SUICIDE', and proceed with the features of textblob. The following histogram is the Number vs polarity graph.

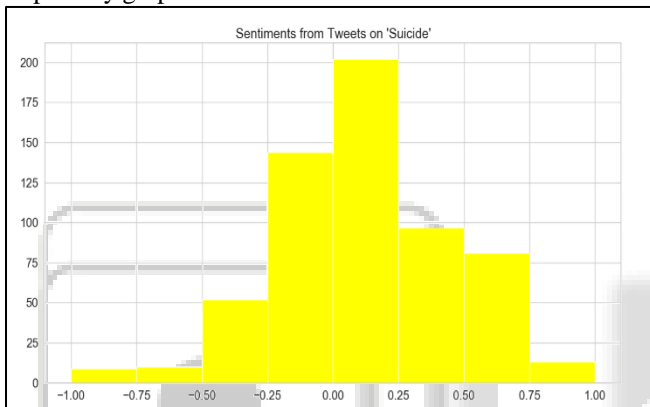


Fig. 2: The X-axis is for the polarity The Y-axis is for the number of tweets This graph depicts a histogram of polarity values for tweets on the word "suicide". And it is visible from the plot that the word 'suicide' is treated differently within a definite group of users.

### V. DETECTION OF DEPRESSION

#### A. Data Extraction:

The data adapted in this research is extracted legally via Twitter Application Program Interface. It consists of more than 5000 tweets that were entered online by general masses around the world. The users who tweet regularly are generally unaware of the impact they are creating. Their tweets, not just ventilate their personal view on the particular topic but also expresses their emotions, their sentiments, their struggles and most importantly their mental health. The data extracted from this platform will help this research to manage and fix the hornets' nest of Depression. This is the only set of data that will be used in all the algorithms proposed later.

#### B. Preprocessing the Data:

Why do we need to pre-process the data before training and testing the machine learning model? The answer is candid, to help the machine learning model work effortlessly and make them even more efficient. What can we do with the data before it could be sent to the machine learning model

for classification? We can do a lot to help our model, there are a series of steps involved in this pre-processing. They could be written as-

#### 1) Tokenisation:

This will break the tweet in atomic words and then process this tokenized data into further steps.

For Example1- God is great!

This tweet could be tokenized as- GOD, IS, GREAT and!.

Example2- Did you liked the movie?

And this could be tokenized as- DID, YOU, LIKED, THE, MOVIE and?.

#### 2) Cleaning the data:

We can clean the data by eliminating the symbols which will do nothing but confuse our machine learning model and the symbols are not used for expressing the emotions, they are just used to improve the way people speak or write.

For the above Example1: we can remove the "!" sign to clean the data. And in Example2: we can remove "?" To clean the data.

#### 3) Removing the Stop Words:

To remove the stop words, we need to figure out, what stop word is? They are nothing but the words used just to connect the words in the sentence. And these stop words have no impact on the sentiment analysis of the data at all, but it will eat up the machine learning model's time and would also reduce its efficiency.

Example1: we have GOD, IS and GREAT. In this sentence, the word "IS" is a stop word and it will have no impact in the sentiment analysis of a tweet. So, we have to remove it.

Example2: similarly in this example, the words "DID","YOU", "THE" are the stop words, and we will also remove them.

So how can we pre-process the data? We can perform some of the sampling algorithms to do so.

We have two types of sampling Under-Sampling and Over-Sampling, and they could be used to pre-process the data

The algorithm used for Over-Sampling is Synthetic Minority Over-sampling Technique(SMOTE) and for Under-Sampling we use Near miss algorithm.

#### C. Classification of the Data:

The Data which was extracted from twitter and then could be used in the machine learning algorithm to classify the tweets into positive, negative or neutral content. So, if we dive a little deeper into the classification problem, then the first question a layman will ask is that, "How can we classify a comment into positive, negative or neutral?". The answer could hands-down be explained with an example. Let us consider 2 simple sentences/tweets.

S1-"I like the way it rains."

S2-"I hate the rainy season".

After all the pre-processing our S1 has the following words:

LIKE

WAY

RAINS

And these three words could be classified as positive, negative or neutral using some of the Natural Language Processing techniques, present in the python libraries. Now let us score the words of S1. The positive words will get a score of 1. The neutral will get a score of 0 and finally, the

negatives will receive a -1. After the results, we have the scores...

LIKE. 1. (A positive word)

WAY 0 (A neutral word)

RAINS 0. (A negative word)

By adding all of the scores, we have a score of 1, hence the score of this sentence is more than 0, hence we could classify as a positive tweet.

Similarly, for the S2 we have the scores like...

HATE -1

RAINY SEASON. 0

By adding all of the scores, we have a score of -1, hence the score of this sentence is less than 0, hence we could classify as a negative tweet.

And this is the way we are going to use for our classification problem, definitely by using some other criteria for the words. But with a similar essence.

What is the algorithm we are using to solve our issue?. Our research could be done by the virtue of a very popular and powerful machine learning algorithm, known as Support Vector Machine. Let us take a glance at this classification algorithm.

#### 1) Support Vector Machine (Supervised Learning):

Consider, we have 2 kinds of data in our data sets, the first one is the positive comment and on the other hand, we have a negative comment. A Support Vector Machine (SVM) is a discriminating classifier formally defined by a separating hyperplane. The output generally outputs the hyperplane which classifies two different content in its respective category.

In our case, we have a huge data set of lots of comments and all of those could be classified as a positive or negative comment. The SVM could be profitable for us, we can take the help of this powerful algorithm for the classification of our dataset in different categories (positive, negative and neutral comment).

There is a graph below which clearly defines what the algorithm does. It contains two axes, X and Y. Where the X-axis states the number of positive words in the tweet and the Y-axis states the number of negative words in the tweet. We can observe from the graph, the data-set closed to the Y-axis (black dots) are the negative tweets. And the dataset closed to the X-axis (blue squares) is the positive tweets.

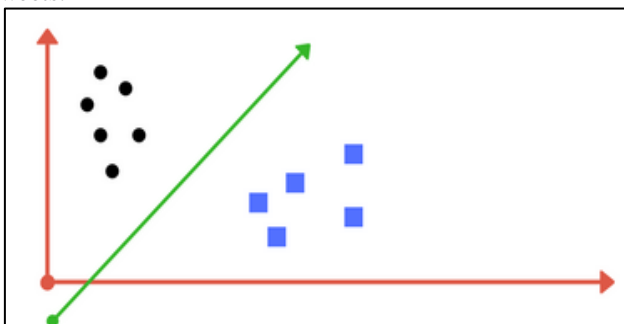


Fig. 3: Diagrammatic Illustration of SVM

Axis- X-No. of positive tweets

Axis- Y-No. of negative tweets

Using the line obtained by the SVM we can classify the tweets as negatives and positives. The dots and squares represent positive and negative sentiments.

But our research is mainly considered for analysing depression. There are some types of negative statements/tweets which could be even worse. And it could even convey the mental health of the user. I know there is a small number of such tweets available, but we will identify it by the same algorithm of SVM. Which will further classify the users who posted negative tweets as Depressed and Not Depressed. And how I am going to do this? The answer is predictable, just like I used words like "HATE", "PATHETIC" etc. for classifying the negative tweets.

Let us consider one tweet

Example3. "All my frendz hate me".

This is the tweet made by a high school student. The young student's sentiments are very clear that he is facing some sort of depression, it could be very little in scale, but ignoring this could be disastrous.

Humans could understand this. But when it comes to handling data in thousands, the human brain will give up, and so we need an algorithm to fix this issue. But we have done it before using SVM. And could be used again with some of the mutations in the data-sets. This time I'll use some more deep words to classify them as depressed and the words/statements could be "HATES ME", "SUICIDE", "DEATH OF DESIRE" and many more. In Example3 we have words like "HATES ME". Which will help the SVM find the depressed batch from the mass. And they could be helped further by the experts.

## VI. RESULTS OF SVM

On performing the proposed methodology for, "I really want to commit suicide" The above sentence is fed in the algorithm for the analysis of tweets and classification of the tweets into further collections very minutely. The further collections are not only limited to positive, negative or neutral but also in the form of positive, weakly positive, strongly positive, negative, weakly negative, strongly negative and neutral.

The results are:

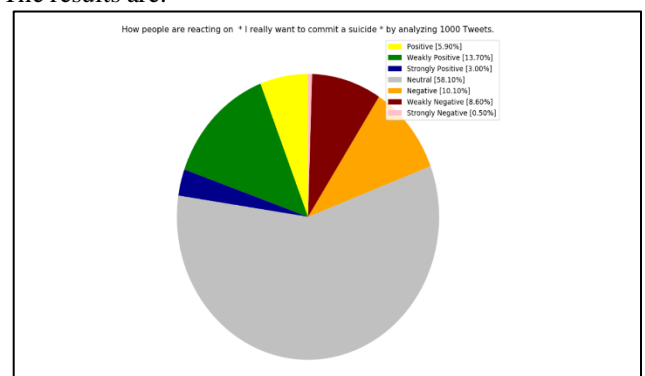


Fig. 4: Pie chart of the result

This result represents the sentiments of the tweeters, who expressed their views on the considered statement.

## VII. PROPOSED CURE FOR THE DEPRESSION

### A. Categorization of Users:

In the first part of the proposed solution, we found a depressed lot of the masses. But what's the point of finding the depressed lot if we can't help them. So, How are we

going to help them? This question has an amazing and unique answer.

And once again this would be done by one of the machine learning algorithms known as the “K-Nearest” algorithm.

We have already classified users as “Depressed User” and we could use the results obtained by the SVM to further classify depressed users according to their “Types of Depression”, which will help the experts to treat them if they know what the situation the patient is facing is. For this classification, we are not going to use the same SVM but this time we will take the help of the KNN algorithm to classify this. The very first question for using this algorithm will be

Why use KNN to classify, if we have SVM?

We use KNN in this classification instead of the traditional SVM because there are many types of depression, for example, property issues, money matters, love affairs, death of someone close and more. So this sort of classification is known as “Multi-Class Classification”. This is very difficult to implement using SVM because it will divide all the features into many binary features, and then proceeds, which will slow down the model. Here comes the KNN for rescue, it is the algorithm that is efficient for the multi-class classification.

### B. What is KNN?

KNN is also known as the lazy algorithm as it does not train the data multiple times. Each point (data-sets) will be placed only once at the time of prediction.

What is 'K' in K nearest neighbours?

K is a number used to identify similar neighbours for the new data point. Referring to our Example of friend circle in the School. We select 3 that we want to be very close friends based on common activities. In this case, K is 3.

KNN considers K nearest neighbours to decide where the new data point belongs to.

How does KNN work?

- 1) Step 1: Select the value for K. K should be preferably an odd number.
- 2) Step2: Find the distance of the new point from every training data.
- 3) Step 3: Find the K nearest neighbours to the new data point.
- 4) Step 4: For classifying the new point, count the number of data points among the k. A new data point will belong to the group that has the most neighbours.

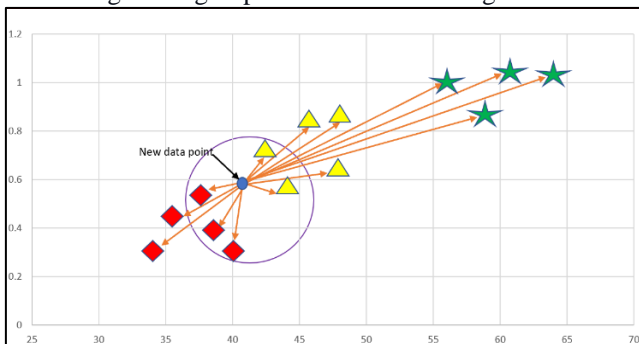


Fig. 5: Diagrammatic Illustration of KNN

This algorithm is the best tool for our research, It is the best way to categorize the depressed users within their types.

Till now, we have all the depressed Twitter users along with their type of depression. But this is just the classification of the problem, but we have to fix this solution.

How are we going to fix our problem of curing the depression problem?

The only way to fix the issue of depression is to distract the patient and how can we do this, by keeping them “BUSY” in something. But to “BUSY” someone we have to identify what are his/ her’s hobbies or what he/she likes to do. How could someone identify the hobbies of the individual who is not willing to live life?

### C. Classifying the Interests:

We can, and for the last time, we will take the help of machine learning to classify the patients based on their hobbies. The hobbies of the patient could be anything. It could be sports, music, dance, reading, etc. But if he/she is depressed, it is very difficult for them to remember their interest, these are the activities which once caused them happiness. Now for this problem, we will ask them some questions about their life indirectly and figure out what their hobbies are.

And how are we going to do this?

Let us understand this approach with an example. A depressed Twitter user’s hobby is to play some sport, but due to some obvious reasons, he is not able to convey this to the doctor. Now the doctor will ask some of the indirect questions to him like.

Do you Brian Lara?

What is the best thing you like in movies

let’s dance tonight?

and many more questions could be framed by the doctor.

After getting the data in the forms of answers, the doctor will feed the data in the KNN algorithm and the algorithm will classify the patients according to their hobbies.

But how could KNN do this?

The algorithm will break the answers in the fundamentals and then process it. let us continue with the previous example. The answer given by the individual for the above questions were-

Yes! I like him.

I don’t watch movies.

Sorry.

this data when fed in the KNN algorithm. The result will be “he likes sports”. All the questions were answered weirdly but the first question was answered perfectly and with some enthusiasm.

With the help of KNN, we figured out the hobbies of the patient which could be used by the doctors to help them fight depression. They could engage the patient within their hobbies to divert their minds from the cause of distress to the pleasure gained by a craft. With all the mind diverted to the activity, the patient could emerge victorious from the depression.

## VIII. CHALLENGES IN SENTIMENT ANALYSIS

Apart from many advantages, the sentiment analysis of a human being can never be done splendidly by a machine or

a machine learning model, It will contain some loopholes. The flaws may be ignored but can never be dodged, they will keep disturbing the result at one's convenience. In our case, the tweets with an abundance of negative words were considered as a tweet posted by a depressed individual. But it could be false as some of the data expressed in this online platform could be "Sarcastic". Let's consider an example of a random tweet.



Fig. 6: A Sarcastic tweet

The person behind this tweet never meant about losing both of his/her kidneys but sarcastically jokes about iPhone's high price, The deal of apple's latest flagship back then was pretty expensive. The gag behind this tweet could be readily understood by the human surfing this platform, but when this tweet gets an interaction by the machine learning model, such as ours. It will get confused and will not be able to differentiate this tweet as a sarcastic tweet, it will consider this as a regular tweet made by a random user and declare the tweeter as the one with no kidney.

This was just one example of the level of sarcasm portrayed by the user, and hence the set of algorithms or a computer program fails in here. There could also be sarcasm in the tweets of a political organization, which could never be ignored and the sentiments behind it should be analysed by a human and not by an algorithm. The computer program analyses the sentiments based on the words the tweeter uses in the tweets. And the words of sarcasm in no way meant the same. Which could be effortlessly understood by the above example. This is the major challenge in the field of sentiment analysis, which needs to be concurred in future researches.

## IX. CONCLUSION AND FUTURE ASPECTS

This paper focuses on the management of depression using the sentiment analysis phenomenon. In this our proposed solution, not only classified the users as depressed or not but also attempted to help them cure their depression with some machine learning algorithms. The proposed solution helps the twitter user know that he/she is depressed or not and also attempts to help them overcome their problem by diverting their attention from the cause of depression to their hobbies which will relax their brains and also help them focus on the goals of their life. The remedy of the depression needs to be spread across the globe as this could help a lot of others to

fight with depression and mental illness. As future work, I am planning to explore the method with a diversity of datasets and also with more mass of data.

## REFERENCES

- [1] M. H. A. El-Jawad, R. Hodhod, and Y. M. K. Omar, "Sentiment Analysis of Social Media Networks Using Machine Learning," 2018 14th International Computer Engineering Conference (ICENCO), 2018.
- [2] A. C. Flores, R. I. Icoy, C. F. Pena, and K. D. Gorro, "An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set," 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST), 2018.
- [3] N. Iqbal, A. M. Chowdhury, and T. Ahsan, "Enhancing the Performance of Sentiment Analysis by Using Different Feature Combinations," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018.
- [4] K. Jain and S. Kaushal, "A Comparative Study of Machine Learning and Deep Learning Techniques for Sentiment Analysis," 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2018.
- [5] T. Karthikayini and N. Srinath, "Comparative Polarity Analysis on Amazon Product Reviews Using Existing Machine Learning Algorithms," 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), 2017.