EARLY DETECTION OF DEPRESSION IN TEXT SEQUENCES USING LINGUISTIC METADATA

by

MADHANGI R 2015103062
SHRUTHI R 2015103572
APOORVA N P 2015103003

A project report submitted to the

FACULTY OF INFORMATION AND COMMUNICATION ENGINEERING

in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ANNA UNIVERSITY, CHENNAI – 25

MAY 2019

BONAFIDE CERTIFICATE

Certified that this project report titled **EARLY DETECTION OF DEPRESSION IN TEXT SEQUENCES USING LINGUISTIC METADATA** is the *bonafide* work of **MADHANGI R (2015103062)**, **SHRUTHI R (2015103572)** and **APOORVA N P (2015103003)** who carried out the project work under my supervision, for the fulfillment of the requirements for the award of the degree of Bachelor of Engineering in Computer Science and Engineering. Certified further that to the best of my knowledge, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on these or any other candidates.

Place: Chennai
Date:

Dr. V. Mary Anita Rajam
Professor
Department of Computer Science and Engineering
CEG, Anna University, Chennai – 25

COUNTERSIGNED

Head of the Department,
Department of Computer Science and Engineering,
CEG, Anna University Chennai,
Chennai – 600025

ACKNOWLEDGEMENT

We express our deep gratitude to our guide, **Dr. V. Mary Anita Rajam**, Professor of the Department of Computer Science and Engineering for guiding us through every phase of the project. We appreciate her thoroughness, tolerance and ability to share her knowledge with us. We thank her for being easily approachable and quite thoughtful. Apart from adding her own input, she has encouraged us to think on our own and give form to our thoughts. We owe her for harnessing our potential and bringing out the best in us. Without her immense support through every step of the way, we could never have it to this extent.

We are extremely grateful to **Dr. S. Valli**, Head of the Department of Computer Science and Engineering, Anna University, Chennai 25, for extending the facilities of the Department towards our project and for her unstinting support.

We extend our sincere thanks to the panel of reviewers **Dr. P. Uma Maheswari**, Associate Professor, **Dr. Geetha Palanisamy**, Professor and **Dr. K. Selvamani**, Associate Professor of the Department of Computer Science and Engineering for their valuable suggestions and critical reviews throughout the course of our project.

We thank our parents, family, and friends for bearing with us throughout the course of our project and for the opportunity they provided us in undergoing this course in such a prestigious institution.

Madhangi R **Shruthi R** **Apoorva N P**

ABSTRACT

Depression is ranked as the largest contributor to global disability and is also a major reason for suicide. Still, many individuals suffering from forms of depression are not treated for various reasons. Previous studies have shown that depression also has an effect on language usage and that many depressed individuals use social media platforms or the internet in general to get information or discuss their problems. This document proposes the early detection of depression using text sequences and Machine Learning.

This approach has text processing and Machine Learning methods imbued in it so as to predict whether a user is depressed or not. The input consists of a text sequence and the final outcome predicts the probability of the user being depressed. The text sequences are pre-processed so as to derive maximum information from it and they are fed as input to Machine learning algorithms, mainly the Naive-Bayes, The Convolutional Neural Network and the Support Vector machine algorithms, which in turn predict the possibility of the user being depressed.The results are measured using F1 scores, which is the harmonic average of precision and recall, and Early Risk Detection Error scores.

ABSTRACT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CBOW Continuous Bag Of Words

CNN Convolutional Neural Network

ERDE Early Risk Detection Error

LIWC The Linguistic Inquiry and Word Count

LR Logistic Regression

LSTM Long Short-Term Memory

NB Naive Bayes

NLP Natural Language Processing

POS Parts Of Speech

RSDD Reddit Self-reported Depression Diagnosis

SVM Support Vector Machine

TF-IDF Term Frequency - Inverse Document Frequency

CHAPTER 1

INTRODUCTION

1.1 PROJECT OVERVIEW

Natural Language Processing is a field of computer science which deals with interactions between the computer and human languages. It deals with making the computer to understand and interpret human language. Computational linguistics is the term used in Natural Language Processing in which the techniques of computer science are applied to the analysis and synthesis of language and speech. It is an interdisciplinary field dealing with the statistical or rule-based modelling of a natural language from a computational perspective. Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. It focuses on the development of computer programs that can access data and use it learn for themselves.

The input to the system is the user data collected in the form of Reddit posts which is then pickled and cleaned. The cleaned data is then pre - processed so as to remove the unwanted data and only the relevant data is retained. The modified data is then subjected to the Word Embeddings module, including Skip Gram, Continuous Bag of Words and GloVe and is then fed as input to the Neural Network models, namely Naive Bayes,Convolutional Neural Network,Support Vector Machine classifiers. Logistic Regression is used for improving the accuracy of the predictions.It is the appropriate regression analysis to

conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Finally, the F1 score and ERDE score are calculated for different classifiers which are trained with different language features like count vectorizer, TF - IDF and word embeddings. Finally the results are aggregated and compared for the different models' accuracy and ERDE error rate.

1.2 PROBLEM OBJECTIVE

Detection of depression is a crucial task, as people tend to hide their mental health from others because of the taboo associated with mental illness and therapy in the society. Thus, it becomes a hard task in making people talk about their illness and depression in general. In the case of social media, people find it comfortable sharing their feelings with the outside word. This makes it possible to detect depression in individuals from their posts and comments in social media. Given user data as text input in the form of Reddit posts, comments and self posts, the system learns and outputs the probability that the user whose data is given is depressed. This is done with the help of various Machine Learning algorithms, including Convolutional Neural Network, Support Vector Machine, Naive Bayes, Logistic Regression and Long Short-Term Memory neural network.

1.3 SCOPE

There are over 753 million people in the world and studies have shown that more than a tenth of the people suffer from depression. Depression, when left untreated can lead people to spiral deeper and deeper into the darkness, until ultimately, they kill themselves. Although rates of diagnosing mental illness have improved over the past few decades, many cases remain undetected. Symptoms associated with mental ill-

ness are observable on Twitter, Facebook, and web forums, and automated methods are increasingly able to detect depression and other mental illnesses. This approach followed here helps us find out the depressed people out in the world, with the help of their posts on the Reddit platform, which, as of today is one of the biggest social media platforms. The early detection and diagnosis of depression is useful in the sense that the people out in the world can seek medical help so as to get it treated.

1.4 CONTRIBUTION

Though conventional methods exist so as to detect depression, this approach is the first of its kind used for the 'Early' detection of depression. Studies have shown that depression leads to suicide mainly because the depressed people don't seek professional help. If the depression is correctly diagnosed early, the affected people can be urged to seek professional help.

1.5 SWOT ANALYSIS

1.5.1 Strengths

The strength of the project lies in that depression can be detected early using this method. A lot of people suffer from depression all alone and let out their pain in the social media hoping it would help. Most of the times, depression can't be self diagnosed and hence, timely help cannot be sought. This approach helps in that it detects depression using social media as a medium and the depressed people can be urged to seek professional help. It is socially relevant in today's stressful world.

1.5.2 Weakness

The approach used considers only language features for prediction and that is a weakness in this project. Using features such as voice and face can lead to better prediction rates. The second weakness is the small size of the data set available. Most of the available data is privatized and hence leads to difficulty in accessing it.

1.5.3 Opportunity

This approach provides us with an opportunity to delve deep into the relatively unexplored social media in relation to depression. It also helps dissipate the social stigma that social media is predominantly a waste of time.This approach when combined with image features can help provide a better accuracy which may aid us in helping people seek help.

1.5.4 Threats

Though the model accommodates for scalability, the tightest upper bound of the level of scalability is still unknown. Training machine learning models with a huge data set is extremely time consuming and still might not give accurate results.

1.6 ORGANISATION OF THESIS

The rest of the thesis is organised as follows:

- Chapter 2 presents the existing approaches for the detection of depression. It also discusses about the advantages and disadvantages of every approach.
- Chapter 3 discusses the requirements analysis of the system. It explains the functional and non-functional requirements, constraints and assumptions made in the implementation of the sys-

tem and the various UML diagrams. It describes the overall system architecture and the design of various modules along with their complexity.

• Chapter 4 explains the implementation details of each module, describing the algorithms used. It also elaborates on the results of the implemented system and gives an idea of its efficiency. It also contains information about the dataset used for testing and other the observations made during testing.

• Chapter 5 concludes the thesis and gives an overview of its criticisms. It also states the various extensions that can be made to the system to make it function more effectively.

CHAPTER 2

LITERATURE SURVEY

This section describes the previous research of language aspects of depression and anxiety. We describe the word count based approaches and tools, and describe the text classification methods applied in mental health diagnosis prediction.

2.1 LANGUAGE

The previous research of depression has shown that depression affects the language people use and, in particular, how it differs from healthy individuals. It has been observed that depressed people often use first person singular pronouns, word ”I”, verbs in past tense and absolutist words. This research lead to the development of software and tools for language feature based analysis and prediction such as Linguistic Inquiry and Word Count (LIWC) and Differential Language Analysis Toolkit (DLATK) [1].

LIWC program reads the input text and counts the words that reflect various emotions, thinking styles and social concerns. It has a predefined number of word categories developed by cross-domain researchers and specialists but also allows adding new custom categories. LIWC categorization has been also widely used to analyze social media posts and detect suicidal and self-harm ideation, and other social risk factors . Although, LIWC based features have proved their efficiency the LIWC software is a commercial language-dependent product which limits its usage.

2.2 DEPRESSION STIGMA

One of the reasons we are concerned with previous authors not removing depression-related text from their data is because we are concerned about stigma leading many depressed users to be silent about their depression.[9] Latalova suggests that stigma-related effects are an important factor preventing depression-related help-seeking among men and that a complex relationship exists between masculinity and depression. Through a narrative review of the research on stigma, they find that masculinity is both a cause of depression and a cause of reduced-help seeking, exemplified by gender norms like boys dont cry.

Similarly, after having conducted a survey of a random sample of college students from 13 American Universities, Eisenberg suggest that social-norms are a leading cause of perceived public stigma and, in turn, personal stigma. They found that higher self-stigma is associated with lower reported comfort seeking help and that self-stigma was highest among male students, Asian students, young students, poor students and religious students.In a random sample people from the general Australian public, Barney found this same pattern: higher reported self-stigma scores result in increased hesitation about seeking help for depression.[5] Major sources of this hesitation included personal embarrassment at having depression and the perception that others would respond negatively. This last finding is in contrast to Schomerus, who found that among a sample of the German public anticipation of discrimination by others did not prevent help seeking behavior.[7]

Our view is that given the consistent findings that self-stigma reduces help-seeking, depression detection efforts using social media and natural language processing have a unique opportunity to reach these individuals. If models can be trained to identify not just the depressed and open about it, but the depressed and hesitant, help could be directed

to individuals who would otherwise neglect to seek it. In this study, our aim is to approximate the scenario where the users are hesitant to post about depression. This is because, failure and delays in treatment seeking were generally greater in developing countries, older cohorts, men, and cases with earlier ages of onset. These results show that failure and delays in initial help seeking are pervasive problems worldwide. Interventions to ensure prompt initial treatment contacts are needed to reduce the global burdens and hazards of untreated mental disorders.

2.3 MACHINE LEARNING

This work belongs to Natural Language Processing (NLP) field and text classification as a particular task. Text classification task is one of the most researched tasks in NLP. It is aimed on predicting the dependent target variable (class label) using the features extracted from text messages which are treated as independent variables. In general, the previous research in text classification has been related to various domains where machine learning and deep learning methods showed amazing results mainly because of computational power available these days. One of example applications is movie review sentiment classification where authors applied Naive Bayes and Support Vector Machines (SVM) to categorize the movie reviews as positive and negative.[2]

The research then moved towards the extraction of population-based health information from social network insights an ever-growing source of data. Particularly, M. De Choudhury used Twitter messages to estimate the depression on a population level in US also applying SVM as a machine learning method. In general, Twitter social network has become a popular source of data for similar analysis. Additionally, another scientific research experimented with N-gram language models to predict post-traumatic stress disorder, depression, bipolar disorder, and

seasonal affective disorder.[3] Similarly to the previous paper, they analyzed Twitter messages and showed that language models outperformed LIWC and claimed that they model the language better than count-based approaches.[8]

On the other hand, the alternative feature engineering approaches are aimed on building classifiers on top of the topic-based features that compress input texts to a fixed number of non-overlapping topics. The other research group integrated the hidden topic features obtained from Latent Dirichlet Allocation (LDA) topic model to classify short and sparse texts. Moreover, Blei in his paper argued that SVM classifier with LDA-based document features performed better than simple bag-of-words features. They also pointed out that LDA can be considered as a dimensionality reduction technique that provides meaningful results which correlate with the underlying text data structure and is often well interpretable.

The other researchers have widely used LIWC categories in natural language analysis and, for instance, augmented LIWC features with LDA in to predict neuroticism and depression in students and showed promising results. Resnik has claimed that topic features has improved the precision and preserved recall to decrease. Despite the fact that we have both long and short text pieces we also applied topic modeling to see how various classifiers will work on this compressed data representation and analyze whether the topics discussed by clinical and control subjects are similar or not.

CHAPTER 3

PROPOSED SYSTEM ARCHITECTURE

3.1 FUNCTIONAL REQUIREMENTS

This system detects depression from the RSDD dataset as input. The requirements of the system are as follows:

- The output of the system should be accurate and precise.
- The output should have limited fallacy.
- The system should be optimized for time and space complexities.
- The system must be able to provide a classification, given any input from the dataset.

3.2 NON-FUNCTIONAL REQUIREMENTS

3.2.1 Hardware

No special hardware interface is required for the successful implementation of the system.

3.2.2 Software

- Operating system : Linux
- Programming language : Python
- Packages : numpy, pandas, os, spacy, datetime, nltk, glove, gensim.models, sklearn.decomposition, matplotlib, keras, tensorflow, xgboost, textblob, string.

3.2.3 Performance

The system must be optimized, reliable, consistent and available all the time.

3.3 DATASET DESCRIPTION

The Reddit Self-reported Depression Diagnosis (RSDD) dataset consists of Reddit posts for approximately 9,000 users who have claimed to have been diagnosed with depression (”diagnosed users”) and approximately 107,000 matched control users. All posts made to mental health-related subreddits or containing keywords related to depression were removed from the diagnosed users’ data; control users’ data do not contain such posts due to the selection process.

The group of diagnosed users is made of users who

1. have a post containing a high-precision diagnosis pattern (e.g., ”I was diagnosed with”) and a mention of depression,
2. do not match any exclusion conditions.

Exclusion conditions apply at both the user level and at the post level. At the user level, any users who did not have at least 100 posts in non-mental health subreddits before their self-reported diagnosis post were excluded. Two exclusion conditions applied at the post level.

1. The mention of depression was required to be no more than 80 characters away from the part of the post matching a diagnosis pattern; posts that did not satisfy this constraint were ignored.
2. A mention of depression is defined as the occurrence of one of the following strings: ”depression”, ”depresion”, ”depressive disorder”, ”major depressive”, or ”mild depressive”. Posts that matched a negative diagnosis pattern were ignored (e.g., ”mother was diagnosed with”).

3.4 CONSTRAINTS AND ASSUMPTIONS

3.4.1 Constraints

- It is only a matter of time after which, a much accurate predictor or classifier will be modelled and this might become obsolete.
- Compromise of results might happen if someone disrupts the system while training or during prediction.

3.4.2 Assumptions

- It is assumed that the dataset analyzed, contain both control user groups and groups diagnosed with depression.
- Only people who post mental health related are assumed to have depression in this case, other sickness which can induce depression is not considered.

3.5 SYSTEM MODELS

3.5.1 Use Case Diagram

Diagram 3.1 depicts the use case diagram for this approach towards the detection of depression.

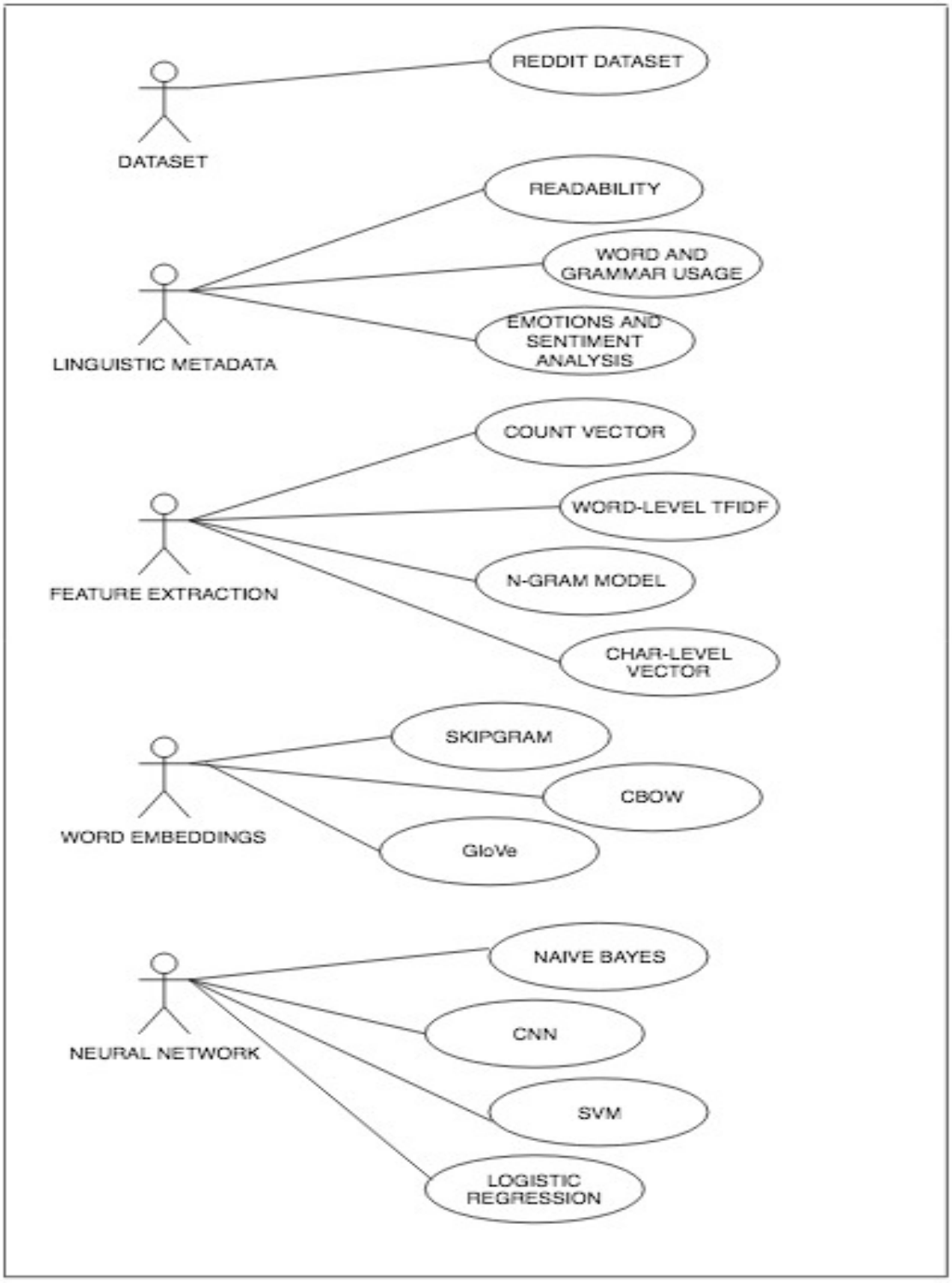

Figure 3.1 Use Case Diagram for Overall System

3.5.2 Sequence Diagram

Diagram 3.2 depicts the sequence diagram for this approach towards the detection of depression.

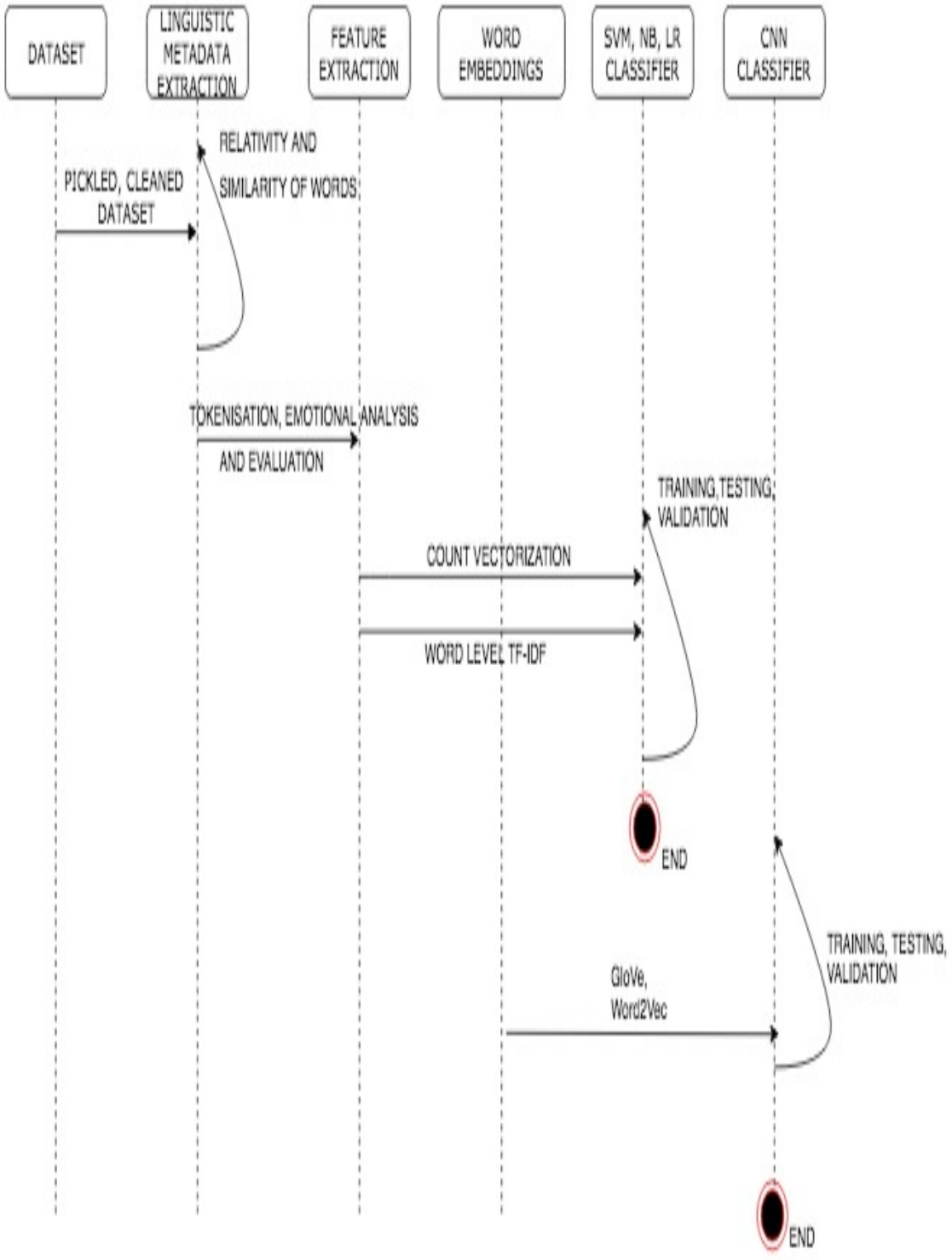

Figure 3.2 Sequence diagram

3.6 SYSTEM ARCHITECTURE

The block diagram for the entire system is shown below. The entire project right from pre-processing to classification is done using Python 2.7, with various add-on packages. A system for early detection of depression has been analyzed.

The system aims at early prediction of depression, by analyzing text sequences to find if a person is depressed or not. The dataset that is a part of this project is the Reddit Self-reported Depression Diagnosis (RSDD) dataset. This dataset has around 1,07,000 matched control groups with around 9,000 users claimed to have depression.

Next step involves adding user-level meta data with the help of the Linguistic Inquiry and Word Count (LIWC). This includes word and grammar usage, readability, emotions and sentiment analysis. Following this, is the addition of neural word embeddings to model words for text classification. This is done by building a Skip-Gram model, Continuous Bag of Words (CBOW) and the GloVe model.

The actual classification is then done with the help of Neural Network classifiers like Convoluted Neural Network (CNN), Support Vector Machine (SVM), Naive Bayes (NB) and Logistic Regression to predict it. They were chosen to compare the accuracy of detection and prediction against each other.

3.7 MODULE DESIGN

3.7.1 Dataset

The RSDD dataset consists of Reddit posts of approximately 9000 users who have claimed to have been suffering from or being diagnosed with depression and 107,000 control users. The diagnosed users were split into three parts as:

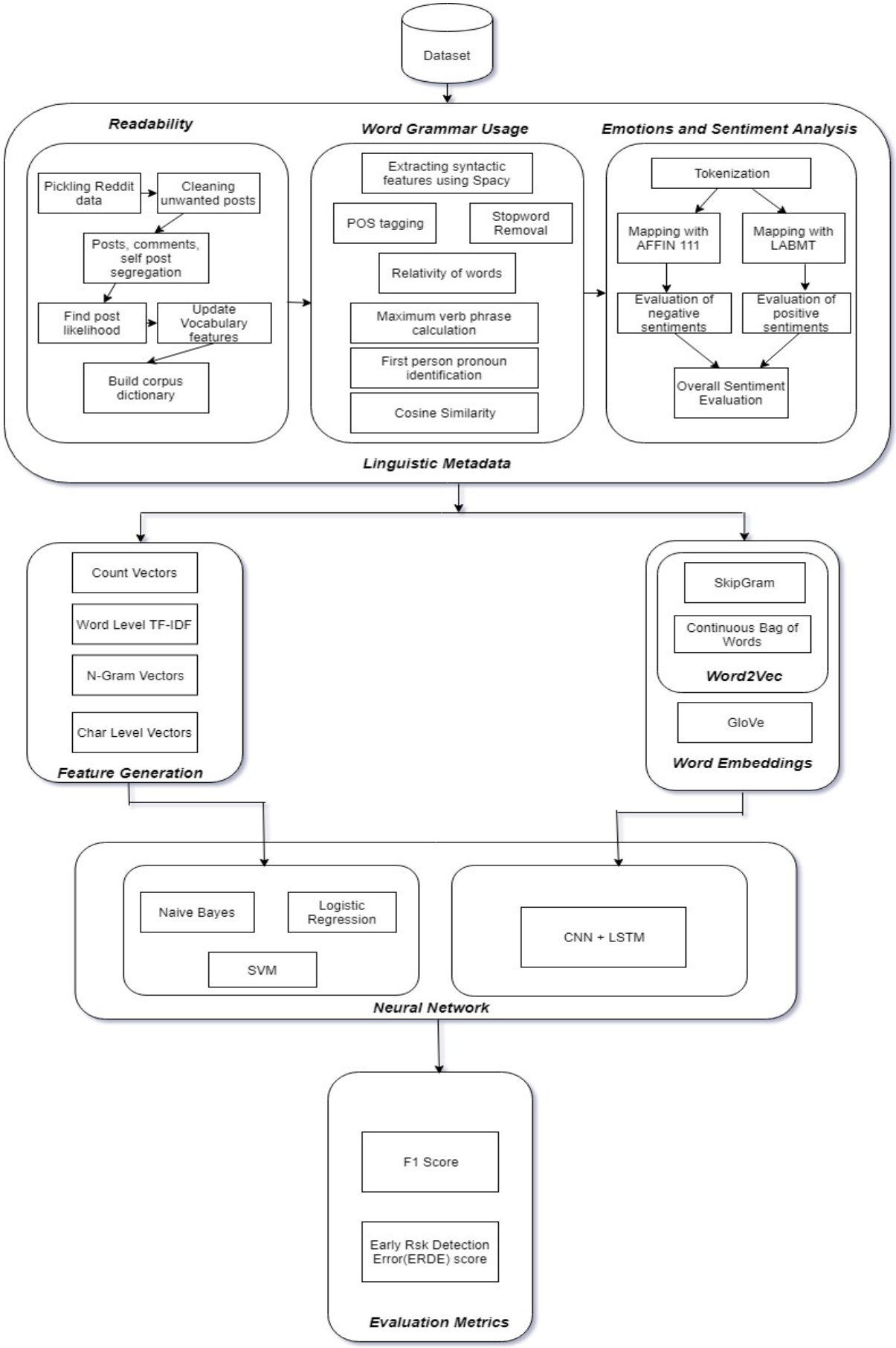

Figure 3.3 Architecture Diagram

- Have a post regarding high-precision diagnosis pattern.(e.g. I was diagnosed with)
- Do not match any exclusion conditions
- People with MH posts (Mental health posts).

3.7.2 Linguistic Metadata

Adding user-level metadata with the classification of user text sequences is dealt in linguistic metadata section. The Linguistic Inquiry and Word Count (LIWC) software helps in evaluating the written texts in several categories based on word counts. The metadata feature categories include the following as described below.

Readability module

The data obtained from scraping blogs cannot be used to find if depressed write more or less compared to the control group. Hence, this criteria cannot be used as a prominent feature for classification. It has been found that controlled group often post very short messages like headlines, single sentence or a single word or emoticon. Considering these constraints the readability standard measures such as Gunning Fog Index (FOG), Flesch Reading Ease (FRE), Linear Write Formula (LWF) and New Dale-Chall Readability. For the readability module, the following steps are undertaken

- Check if the id of the data frame of dataset is not null and initialize it.
- Initialize the comments of each posts if they are not deleted.
- Prepare the data frame as an augmentation of posts and comments.
- If the post is null, then initialize features like verb phrase length, maximum verb phrase length, maximum height, noun chunks and subordinate conjunctions to null.

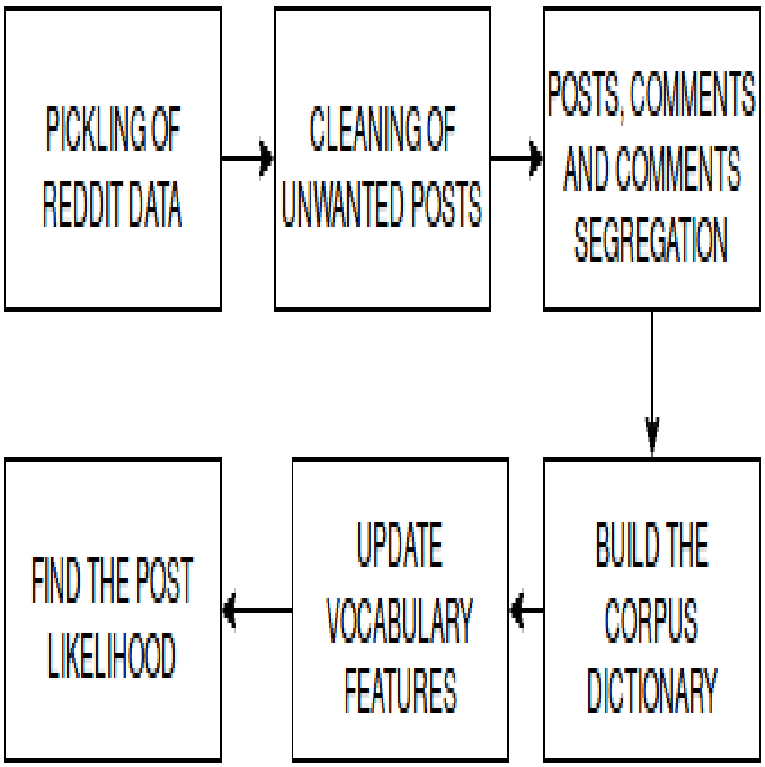

Figure 3.4 Readability

Word and Grammar Usage

The word and grammar usage metadata considers the effects of depression on the grammar usage which are very explicit when a depressed person has any form of communication. One good example is the increased usage of personal pronouns I and verbs in past tense. There are very few explicit phrases like I was diagnosed with depression are less aimed for early detection but are used at predicting the obvious cases. The posts containing phrases like my depression , my anxiety and my therapist and some common antidepressants names are used as strong indicator of positive cases. This metadata feature is summed up over all documents of a particular user since this is a very strong feature for prediction even if it found in few or one document.

- Get each record from Reddit post.
- Set all features to null if the post is null.

- For each record, get verb phrase length, maximum verb phrase length, maximum height, noun chunks and subordinate conjunctions.

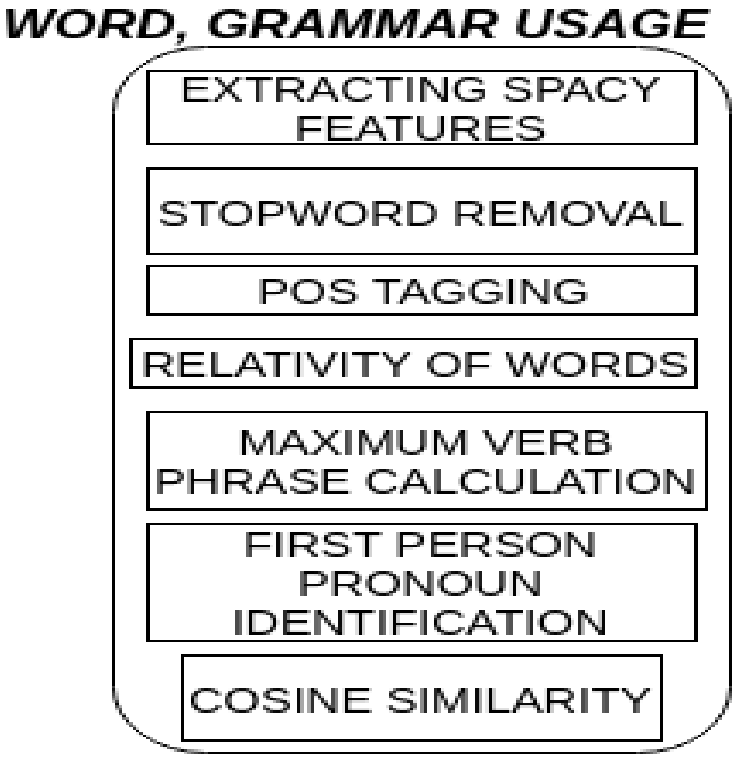

Figure 3.5 Word and Grammar Usage

Emotions and Sentiment Analysis

Sentiment analysis is an area that can be very helpful for finding the emotional statements in the field of mental health text classification. The important indicator that has been considered is how the authors express towards their personal hysteria. This is achieved with the help of LIWC tool that includes two features for positive and negative emotions and separate features for anxiety, anger or sadness. Apart from the LIWC tool, the NRC Emotion Lexicon, Opinion Lexicon and VADER Sentiment lexicon are also used for analysing the emotional state of the user.

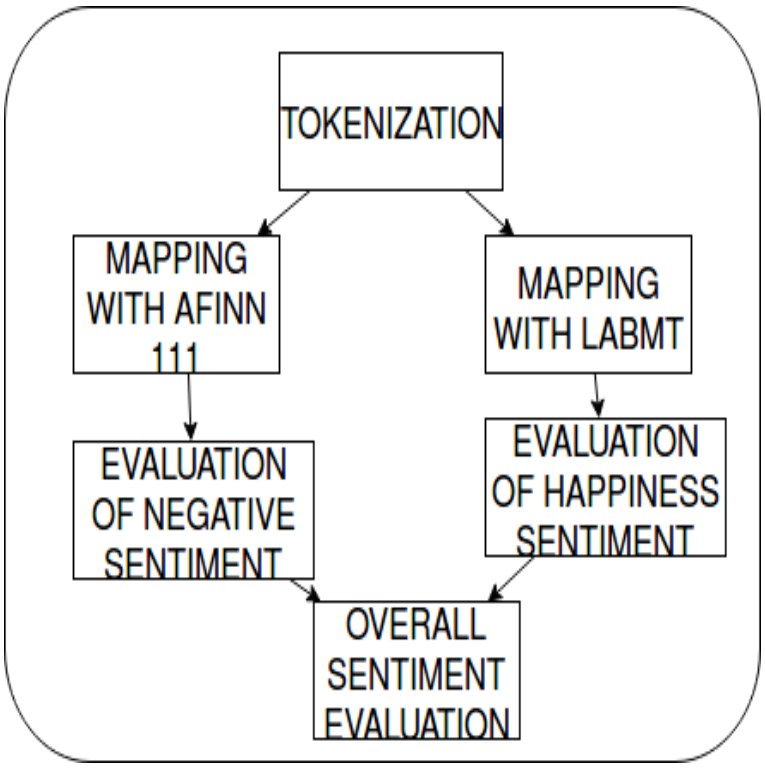

Figure 3.6 Emotions and Sentiment Analysis

- Tokenize the words from post.
- Find the positive and negative scores by mapping the tokenized words with sentiment dictionary.
- Update the score to the data frame.

3.7.3 Word Embeddings

Neural word embeddings is an efficient way to model words for text classification tasks. In contrast to the customary way, the concept does not handle each word with a single neuron, but uses several neurons to represent a word and each neuron has a part to play in the description of several words. This helps in better understanding of the language instead of just individual words or sentences.There are packages in python which builds GloVe and Word2Vec word embedding models. The following methods as described below are used for word embeddings.

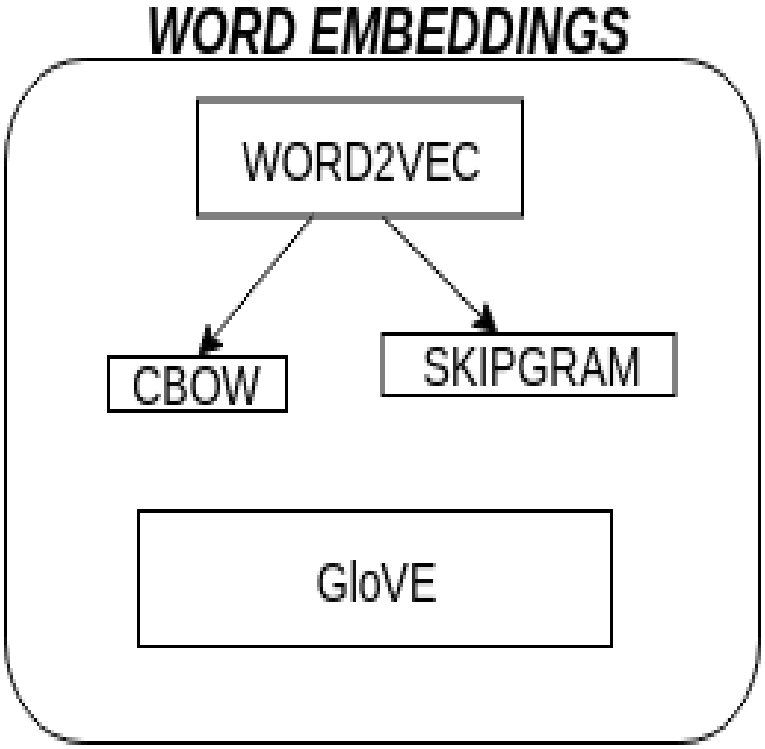

Figure 3.7 Word Embeddings

Skip Gram

The skip-gram model tries to predict the surrounding words given a target word. The skip-gram model aims to predict the context from the target word. It uses a context window based on which decides how many surrounding words for the context is to be predicted.

- Tokenize the words from post.
- Build a text corpus.
- Apply skip-gram to the sentences and list the model.
- Save the model.

Continuous Bag Of Words

The CBOW model is often considered as the reverse process as that of the skip-gram model process. The input to this model will be a continuous array of words which precedes or succeeds the expected

word of a sentence. The output of the model will be the current word. Hence, CBOW can also be considered as a task as predicting the word given its context.

GloVe Model

The Global Vectors for Word Representation model tries to capture the count of overall statistics of how often a word appears. GloVe works similar to Word2Vec. While Word2Vec is a predictive model that predicts the content given a word, GloVe learns by constructing a matrix that counts how many times a word appears in a context. The fastText model finds the closest neighbours of the depression dataset and the nearest words are grouped according to cosine similarity.

- Build a text corpus.
- Apply GloVe model by choosing the number of components and learning rate.
- Fit it to the corpus and add dictionary.
- Save the model.

3.7.4 Neural Network Model

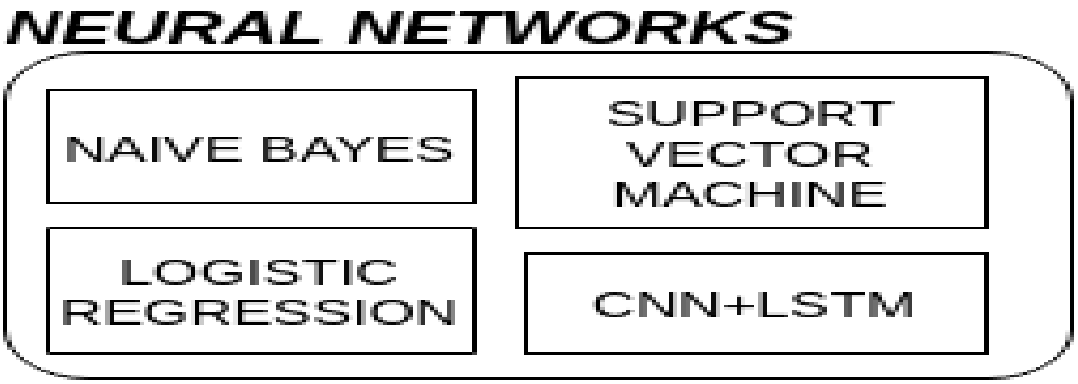

Figure 3.8 Neural Network

Convolutional Neural Network

The CNN model is a state-of-art model in image processing and classification tasks. These are also efficient in text and sentence classification tasks. It consists of neurons with learnable weights. A single neuron represents a region of a input sample (text in our case). CNN has set of layers namely the input layer, hidden layers and the output layer. The hidden layer includes the convolutional, pooling, normalization and fully-connected layers. A neural network is considered convolutional if it has at least one convolutional layer.

- Initialize the convolutional layer count.
- Build a model and train it.
- Apply max-pooling algorithm to the trained model.
- Repeat the same steps till the layer count becomes zero.

Support Vector Machines

The SVM is a supervised machine learning technique that is being utilized for both classification and regression tasks. The SVM works by dividing a linearly separable data into two classes with the maximum between-class distance(also termed as margins). The SVM finds an optimally distant line that lies as far from the nearest class points as possible. If the input data is not linearly separable, Kernels can be used to map the data into higher-dimensional space to make it linearly separable. The most popular non-linear kernels are Gaussian kernels, sigmoid kernels and Polynomial kernels.

- Compute kernel distances between data points.
- Assemble the constraint set as matrices to solve.
- Pass these matrices to the solver.
- Identify the support vectors as those that are within some specified

distance of the closest point and dispose of the rest of the training data.

- Compute the inner product of the test data and the support vectors.
- Perform the classification returning either the label (hard classification) or the value (soft classification).

Naive Baiyes Classifier

Naive Bayes classifier is a probabilistic classifier with strong (naive) independence assumptions between the features. Multinomial NB is the one of the most common classifier used for classification of text into labelled classes.

- Extract the text.
- Count the text.
- If the counted text is in class, find prior of class.
- Concatenate all text and apply conditional probability.

3.8 COMPLEXITY ANALYSIS

3.8.1 Time Complexity

Word and Grammar Usage

The time complexity of word and grammar usage :

$$[(nlp\ \ time) + P * C + 2T] \tag{3.1}$$

nlp time - spacy's nlp language model time for processing

P - No. of Part-Of-Speech Tagger

C - Time for counter function

T - No. of tokens

N - No. of posts

Sentiment Analysis

The time complexity of sentiment analysis :

$$N * [2 * [T * L]] \tag{3.2}$$

N - No. of posts

T - No. of tokens

L - Look-up time

Naive Bayes Algorithm

The time complexity of naive bayes algorithm :

$$O(N * d) \tag{3.3}$$

N - No. of training examples

d - No. of features

Support Vector Machines Algorithm

The time complexity of support vector machine algorithm :

$$O(N^2 * d) \tag{3.4}$$

N - No. of training examples

d - No. of features

Convolutional Neural Network

The time complexity of convolutional neural network algorithm :

$$M * N * F * F * K * K \tag{3.5}$$

M is number of input channels

N is number of output channels

F is feature map size

K is filter size

3.8.2 Complexity of the Project

Numerous complexities were faced throughout the course of the project, out of which some of the issues are outlined below.

- The training phase takes considerable amount of time which only gets increased as we add more training data into it.
- Datasets for depression are generally hard to obtain.
- Since this project require depression posts data protected by laws, it makes it a difficult task to obtain data.
- Time taken in counting the first person pronouns is high.
- Feature extractions along with GloVe and Word2Vec embedding matrix takes a considerable amount of time as it goes through each and every word in the post.

float lmodern[linesnumbered]algorithm2e >

CHAPTER 4

IMPLEMENTATION AND RESULTS

The described system consists of various modules such as the Metadata pre-processing module, the Word embeddings module, the Neural Network module and the Evaluation module. The overview of the modules is described as follows :

4.1 PROTOTYPE ACROSS THE MODULES

The input and output to various modules in the system are described as follows:

- **Dataset Curation** This module takes the Reddit Self reported Depression Diagnosis dataset as input. It consists of Reddit posts of approximately 9000 users who have claimed to have been suffering from or being diagnosed with depression and 107,000 control users. The diagnosed users were split into three parts as :
    1. Have a post regarding high-precision diagnosis pattern (e.g. I was diagnosed with)
    2. Do not match any exclusion conditions
    3. People with MH posts(Mental health posts)
- **Dataset Pre-Processing** Adding user-level metadata with the classification of user text sequences is dealt in linguistic metadata section. The Linguistic Inquiry and Word Count (LIWC) software helps in evaluating the written texts in several categories based on word counts. The metadata feature categories include :
    1. Readability

2. Word Grammar Usage
3. Emotion and Sentiment Analysis

- **Word Embeddings** Neural word embeddings is an efficient way to model words for text classification tasks. In contrast to the customary way, the concept does not handle each word with a single neuron, but uses several neurons to represent a word and each neuron has a part to play in the description of several words. This helps in better understanding of the language instead of just individual words or sentences. There are packages in python called Word2Vec and GloVe which helps in creating word embedding models. The following methods as mentioned below are used for word embeddings :
  1. Skip Gram
  2. Continuous Bag of Words
  3. Global Vectors for Word Representation
- **Neural Network** A combination of machine learning algorithms are used for training and testing the system, including :
  1. Naive Bayes Classifier
  2. Logistic Regression
  3. Convolutional Neural Network
  4. Support Vector Machine
- **Evaluation Metrics** This module takes the trained model as input and produces the accuracy based on F1 score and ERDE score as output.

4.2 LINGUISTIC METADATA ALGORITHM

This section involves the algorithms for pre-processing of the dataset acquired.

4.2.1 Readability Algorithm

It takes as input the RSDD dataset and returns a cleaned Python dataframe.

Algorithm 1: Readability Algorithm

```
input : RSDD dataset
output : Python dataframe after noisy data removal
    prepareDataFrame (dataFrame)
    if dataframe.parentID is not null
        posts = dataframe. dataframe.text is not deleted
            comments = dataframe
    preparedDataframe = posts or comments
    return preparedDataframe
```

Algorithm 2: Readability Algorithm

```
input : RSDD dataset
output : Python dataframe after noisy data removal
    postText = getTextFromRecord (redditPost)
    if postTest is null
        syntacticFeatures.maxHeight = null
        syntacticFeatures.nounChunks = null
        syntacticFeatures.maxVerbPhraseLength = null
        syntacticFeatures.subordinateConjunctions = null
    return syntacticFeatures
```

4.2.2 Word Grammar Usage Algorithm

It takes as input a Python dataframe and returns a panda series containing the value of depressed language usage in each post.

Algorithm 3: Word Grammar Usage algorithm

input *: Python dataframe*
output *: Panda series containing value of depressed language usage*

```
getSyntacticFeatures (input redditPost)
postText = getTextFromRecord (redditPost)
if (postText == null)
        syntacticFeatures.maxHeight = null
        syntacticFeatures.nounChunks = null
        syntacticFeatures.maxVerbPhrase = null
        syntacticFeatures.subordinateCon = null
else
        documentData = s pacy.NLP (postText)
        nounChunks = documentData.nounChunks
foreach (sentence in sentences)
    subordinateConj += token -≻ token.type ==
subordinateConj    subordinateConj/ = sentences.size()
foreach (sentence in sentences)
        sentenceHeight = getMaximumTokenHeight ()
        maxHeight = max (sentenceHeight, maxHeight)
foreach (sentence in sentences)
        verbPhraseLength = sentence.getVerbPhraseLength ()
        maxVerbPhraseLength = max (verbPhraseLength,
        maxVerbPhraseLength)
syntacticFeatures.maxHeight = maxHeight-1
syntacticFeatures.nounChunks = nounChunks
syntacticFeatures.maxVerbPhraseLength =
maxVerbPhraseLength
syntacticFeatures.subordinateConj = s ubordinateConj
return syntacticFeatures
```

4.2.3 Emotion and Sentiment Analysis Algorithm

It takes as input a panda series containing the value of depressed language usage in each post and returns the sentimental value for each post.

Algorithm 4: Emotion and Sentiment Analysis algorithm

```
input : Panda series containing value of depressed language
 usage
 output : Sentimental value
      getSentimentFromText (input text)
                words = text.tokenize ()
                wordSentiments = words.map (word points to
          sentimentDictionary (word))
                sentenceSentiment = sqrt (wordSentiments.sum () /
          words.size ())
               return sentenceSentiment
       getSentimentScore (input text)
        if text is null
             return null
          else
              return getSentimentFromText (text)
      addEmotionalFeatureToDataFrame (input dataFrame)
          dataFrame.affin = getSentimentScore (dataFrame)
           return dataFrame as updatedDataFrame
```

4.3 WORD EMBEDDINGS ALGORITHM

This section involves the algorithms for fine-tuning the pre processed dataset.

4.3.1 Word2Vec Algorithm

It is a combination of the Skip Gram and the CBOW model.

Algorithm 5: Word2Vec Algorithm

input *: Pre-processed text*
output *: Probabilistic vector representation*

```
Word2Vec (text)
sentence = buildCorpus (text)
model = Word2Vec (sentences)
word = list (model.vocab)
model.save ()
model = Word2Vec.load ()
```

4.3.2 GloVe Algorithm

It takes as input the pre-processed text and returns the GloVe word embedded vectors.

Algorithm 6: GloVe Algorithm

input *: Pre-processed text*
output *: GloVe word embedded vectors*

```
corpus = BuildCorpus (text)
corpusFit (lines, windowSize)
glove = Glove (no of components, learning rate)
glove.fit (corpus)
glove.addDictionary (corpus)
glove.save ()
```

4.4 NEURAL NETWORK MODELS ALGORITHM

This section involves the training and testing of the system using various classifiers.

4.4.1 Naive Bayes Algorithm

Naive Bayes classifier is a probabilistic classifier with strong (naive) independence assumptions between the features.

Algorithm 7: Naive Bayes Algorithm

input *: Classes and text to be classified*
output *: Naive Bayes model with conditional probability*
TrainMultinomialNB (class,text)
V = ExtractVocabulary (text)
N = CountText(text)
for each C in class
do N1 = CountTextInClass (text,C)
prior[C] = N1/N
text1 = ConcatenateAllText (C)
foreach t in V
do condProb = ConditionalProb (t,C)
return condProb

4.4.2 Convolutional Neural Network Algorithm

The CNN model is a state-of-art model in image processing and classification tasks. These are also efficient in text and sentence classification tasks. It consists of neurons with learnable weights. A single neuron represents a region of a input sample (text in our case). CNN has set of layers namely the input layer, hidden layers and the output layer.

Algorithm 8: Convolutional Neural Network Algorithm

input *: Training files*
output *: CNN model*
buildCNN (input training files)
initialise layerCount
while (layercount $\succ$0)
intermediateModel=buildModel ()
trainedModel = trainModel (intermediateModel)
model = applyMaxPooling (trainedModel)
decrement layerCount
return model

4.4.3 Support Vector Machines Algorithm

The SVM is a supervised machine learning technique that is being utilized for both classification and regression tasks. The SVM works by dividing a linearly separable data into two classes with the maximum between-class distance(also termed as margins). The SVM finds an optimally distant line that lies as far from the nearest class points as possible.

Algorithm 9: Support Vector Machine Algorithm

---

input *: Training data*
output *: SVM linear model*
INITIALISATION
For the specified kernel, and kernel parameters, compute the kernel of distances between the datapoint.

TRAINING
Assemble the constraint set as matrices
pass these matrices to the solver
identify the support vectors as those that are within some specified distance of the closest point and dispose of the rest of the training data.

CLASSIFICATION
For the given test data z, use the support vectors to classify the data for the relevant kernel.
compute the inner product of the test data and the support vectors
perform the classification returning either the label (hard classification) or the value (soft classification).

4.5 TEST CASES

The table below discusses the test cases, for every possible end case and their outcomes, for modules as shown below.

MODULE NUMBER	INPUT	MODULE NAME	OUTPUT
1.	Selftext: "I wish to die soon" Post: NaN Comment: "He is not alone" ParentID=NaN(deleted) SelfText: NaN Post: "I wish to die" Comment: "Same here"	Readability	Body: "I wish to die. u' Comment: He is not alone. Such posts are ignored. Body: "I wish to die soon. Same here"
2.	Body: "I wish to die soon. u' Comment: You're not alone"	Word and Grammar Usage	[maxHeight, maxVerbPhraseLength , noun_chunks,subordinate Conjunctions,dtype]=[12, 41,2.2,9.8,object]
3.	"You're not alone"	Emotion and sentiment analysis	Emotion and sentiment value= 0.79056941
4.	["You're", "not, "alone"]	Word2Vec, GloVe, CV,TF-IDF	Vector representation of the corpus
5.	[0.77,0,0,0.68,.....]	Neural Networks	F1-score ERDE score

Figure 4.1 Test cases for the early detection of depression

4.6 DATASET INFORMATION

The Reddit Self reported Depression Diagnosis dataset consists of Reddit posts of approximately 9000 users who have claimed to have been suffering from or being diagnosed with depression and 107,000

control users. All posts made to mental health-related subreddits or containing keywords related to depression were removed from the diagnosed users' data; control users' data do not contain such posts due to the selection process. The diagnosed users were split into three parts as

1. Have a post regarding high-precision diagnosis pattern (e.g. I was diagnosed with)
2. Do not match any exclusion conditions
3. People with MH posts(Mental health posts)

Exclusion conditions apply at both the user level and at the post level. At the user level, any users who did not have at least 100 posts in non-mental health subreddits before their self-reported diagnosis post were excluded. Two exclusion conditions applied at the post level.

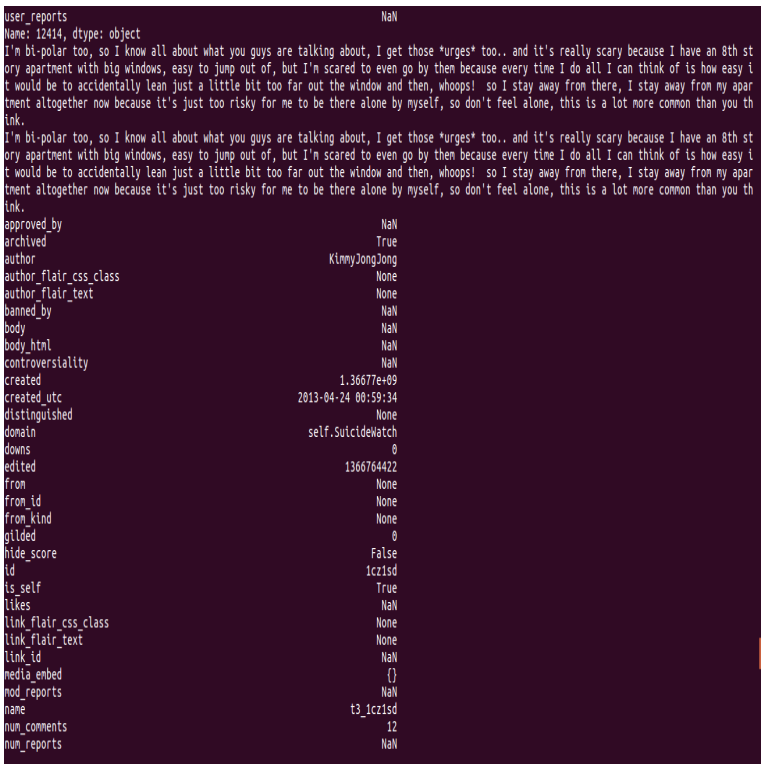

Figure 4.2 Data Extracted as a Panda Dataframe

4.7 OUTPUT OBTAINED AT VARIOUS STAGES

4.7.1 Dataset Pre-Processing

Adding user-level metadata with the classification of user text sequences is dealt in linguistic metadata section. The Linguistic Inquiry and Word Count (LIWC) software helps in evaluating the written texts in several categories based on word counts. The metadata feature categories include :

Readability

The readability process includes the basic preprocessing and cleaning of noisy data present in the dataset. For each row in the dataset, the body containing the depressed or suicidal message is extracted and returned as a unicode on which the further NLP processing takes place.

```
**************DOC CONTENT***************
Yeah I guess I did.
*************NOUNS******************
I
I
**************DOC CONTENT***************
If you've just started playing an instrument you can't be amazing. How many hours do you think every famous musician has poured into their work
? Learning takes so much time; are there any nearby musical schools or programs you could check out?
*************NOUNS******************
you
an instrument
you
you
every famous musician
their work
Learning
so much time
any nearby musical schools
programs
you
**************DOC CONTENT***************
You have nothing to be ashamed of, to be honest, so I don't see the problem.

It's the girls that should feel any shame, so just go to the cops and don't let anyone blackmail you.
*************NOUNS******************
You
nothing
I
the problem
It
the girls
any shame
the cops
anyone
you
**************DOC CONTENT***************
Have you been seeing a therapist? That would be the best place to start. What are your goals? Where would you like to be?
*************NOUNS******************
you
a therapist
the best place
What
your goals
```

Figure 4.3 Retrieval of Nouns from the processed unicode

Word Grammar Usage

The further language processing are done with the help of the Spacy which is an open-source python library for advanced Natural language processing. The English language model of Spacy is loaded and stored as a variable which can be used as a function for later use for language processing.

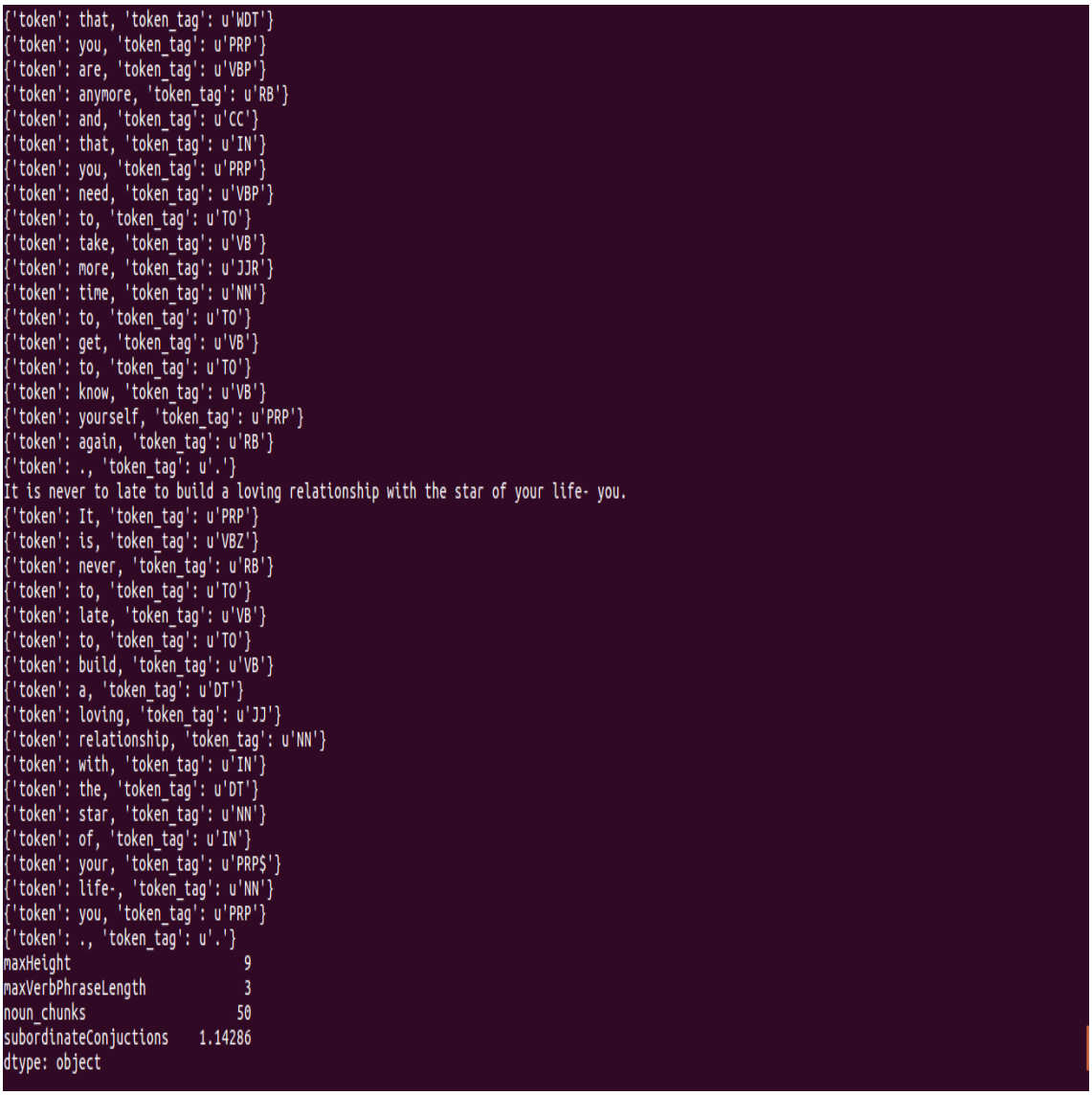

Figure 4.4 Relativity and verb,personal pronoun ratio identification

Emotion and Sentiment Analysis

The sentiment and emotional analysis of a particular user post is identified using the AFINN-111 dataset. The AFINN-111 data has a set of words which has a score of how much the word contributes towards negativity. The LABMT data has a set of words which has a score of how much the word contributes towards positivity.

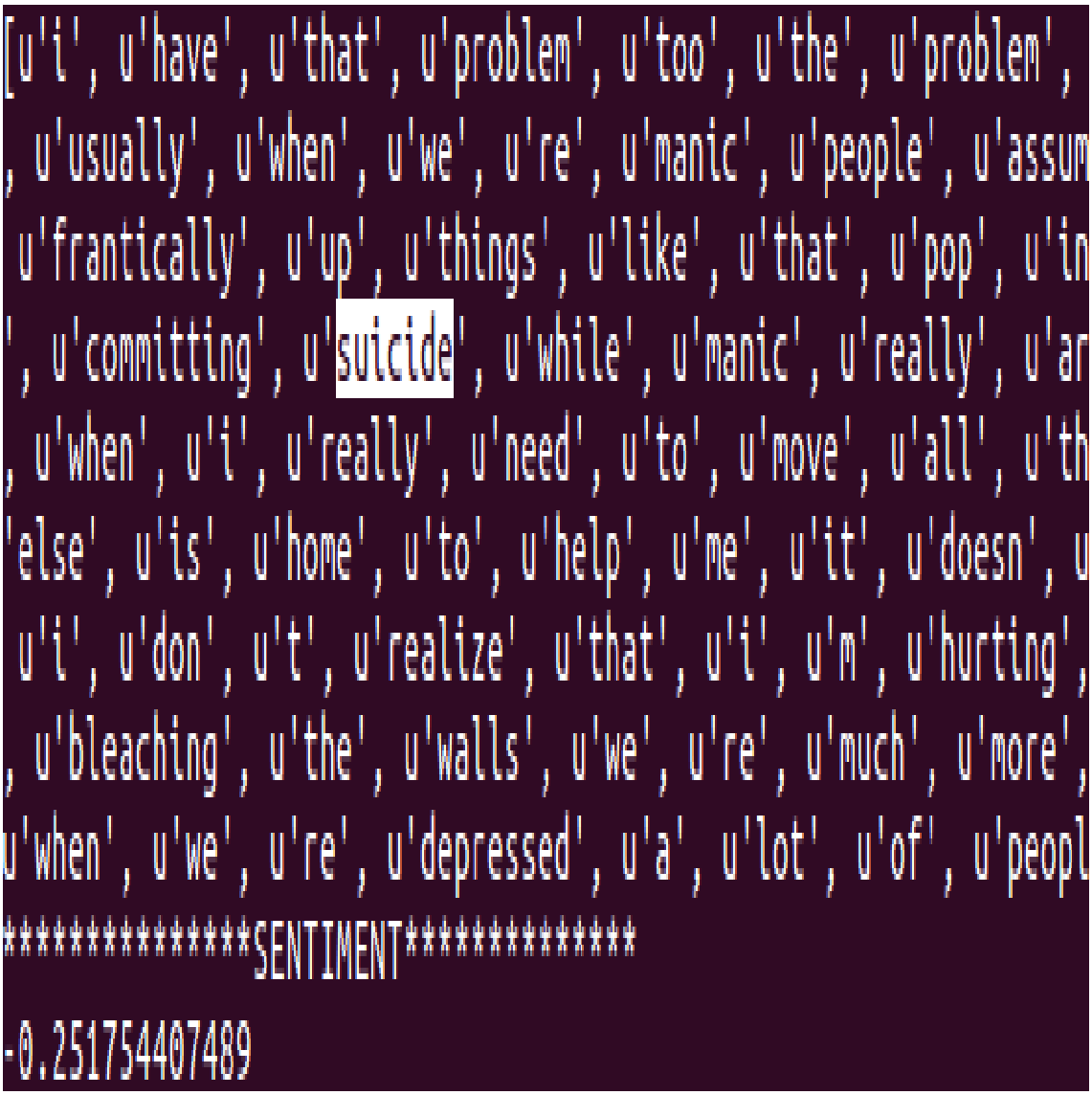

Figure 4.5 AFFIN - negative score

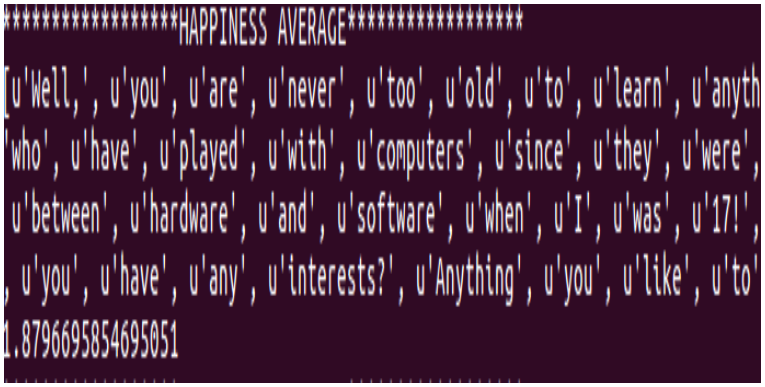

Figure 4.6 LABMT - positive score

4.7.2 Word Embeddings

Neural word embeddings is an efficient way to model words for text classification tasks. In contrast to the customary way, the concept does not handle each word with a single neuron, but uses several neurons to represent a word and each neuron has a part to play in the description of several words.The following methods as mentioned below are used for word embeddings :

Word2Vec

The Skip-gram model has been constructed for the data sample using the Word2Vec and the word embedded vector matrix has been constructed. We have also plotted how the words would be related for this small corpus using PCA.The way CBOW works is that it tends to predict the probability of a word given a context. A context may be a single word or a group of words.Together , they form the Word2Vec algorithm.

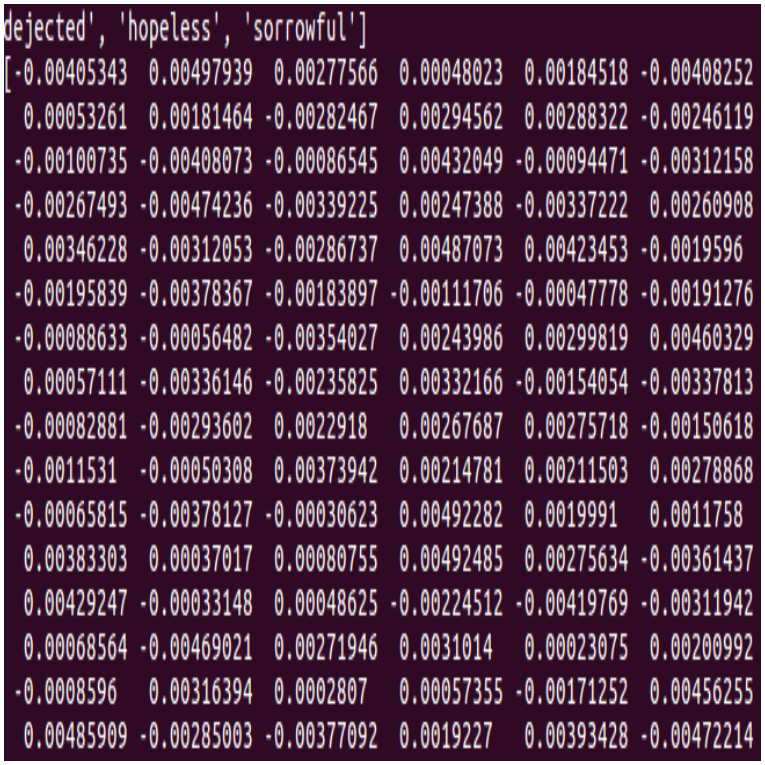

Figure 4.7 Word2Vec representation

GloVe

GloVe stands for global vectors for word representation. It is an unsupervised learning algorithm developed by Stanford for generating word embeddings by aggregating global word-word co-occurrence matrix from a corpus. The resulting embeddings show interesting linear substructures of the word in vector space.

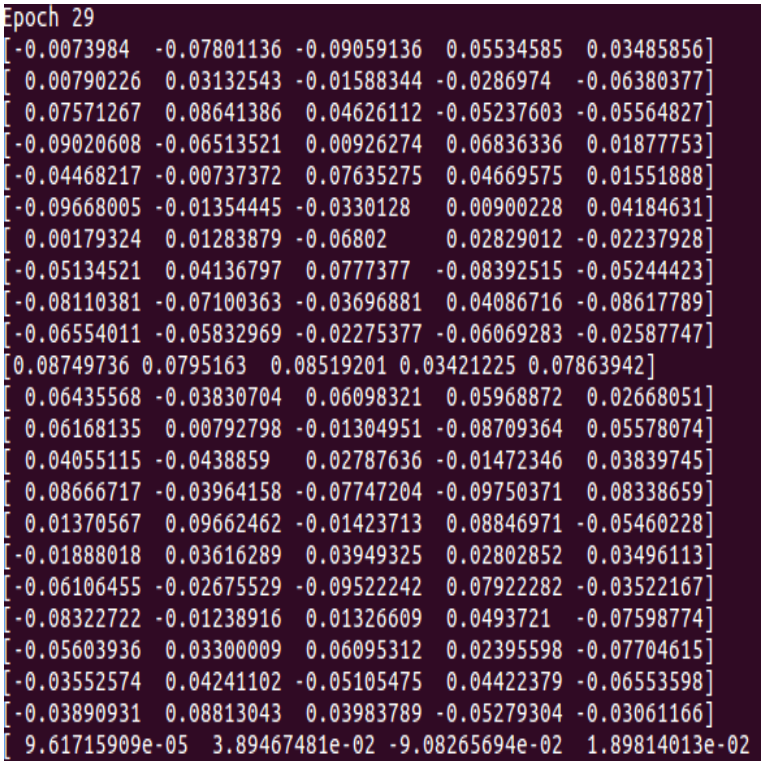

Figure 4.8 GloVe representation

4.7.3 Neural Network

A combination of Machine Learning algorithms are used for training and testing the system. The system is first subjected to logistic regression as to improve the predictions.

Logistic Regression

The Logistic Regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

```
LR, Count Vectors:
Accuracy: 0.5950317735413057
f1_score 0.5917297612114152
recall_score 0.6033254156769596
precision_score 0.5805714285714285
ERDE score
*****ERDE accuracy score*******
ERDEtp 60.5550843872
ERDEtn 0
ERDEfp 107
ERDEfn 334
0.120734663594
```

Figure 4.9 Logistic Regression - Image 1

```
LR, WordLevel TF-IDF:
Accuracy: 0.6296938186019642
f1_score 0.6218289085545723
recall_score 0.6426829268292683
precision_score 0.6022857142857143
ERDE score
*****ERDE accuracy score*******
ERDEtp 62.8199399057
ERDEtn 0
ERDEfp 105
ERDEfn 293
0.136322095607
```

Figure 4.10 Logistic Regression - Image 2

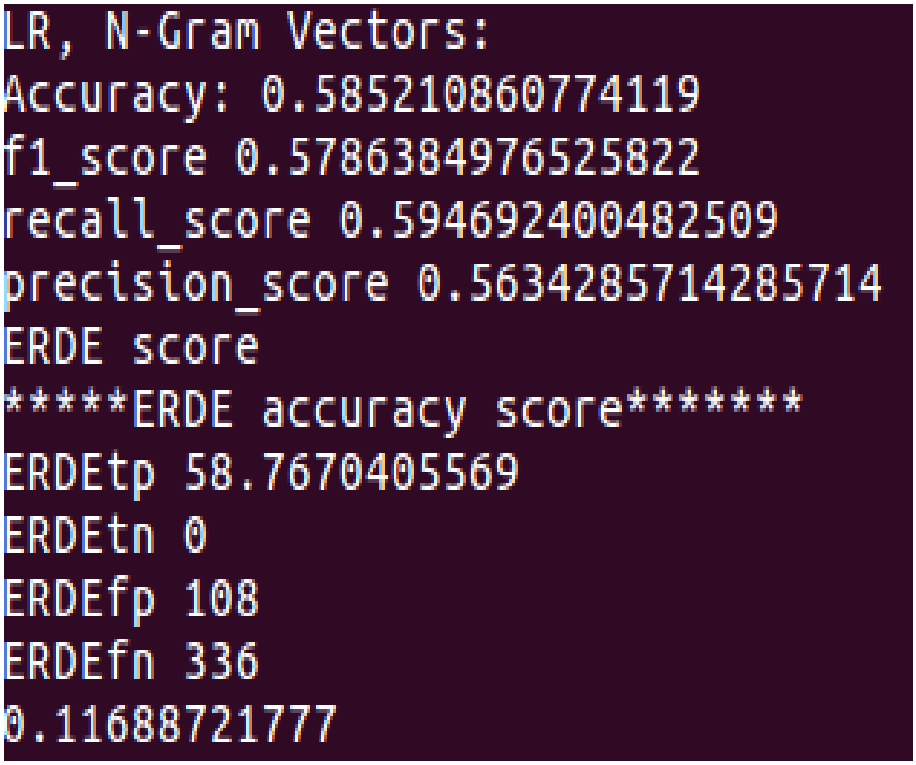

Figure 4.11 Logistic Regression - Image 3

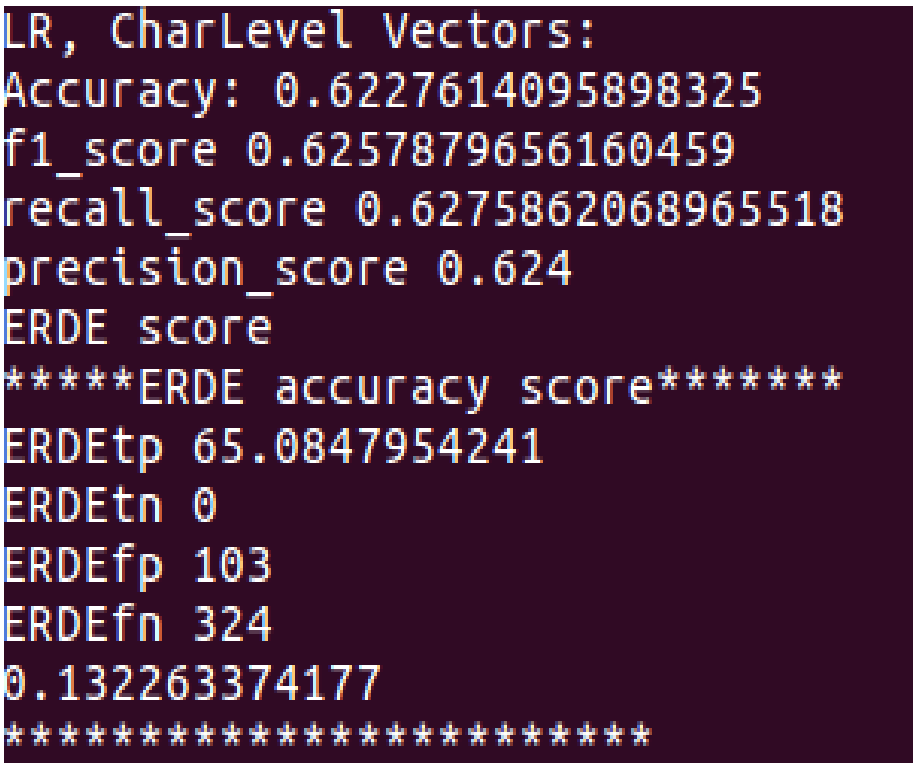

Figure 4.12 Logistic Regression - Image 4

Naive Bayes classifier

Naive Bayes is a family of algorithms based on applying Bayes theorem with a strong (naive) assumption, that every feature is independent of the others, in order to predict the category of a given sample.It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

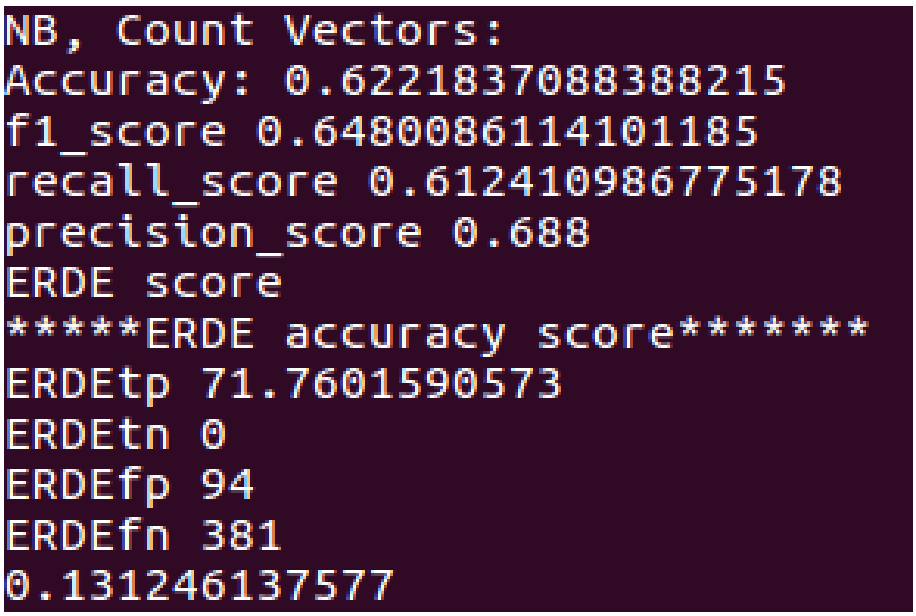

Figure 4.13 Naive Bayes classification - Image 1

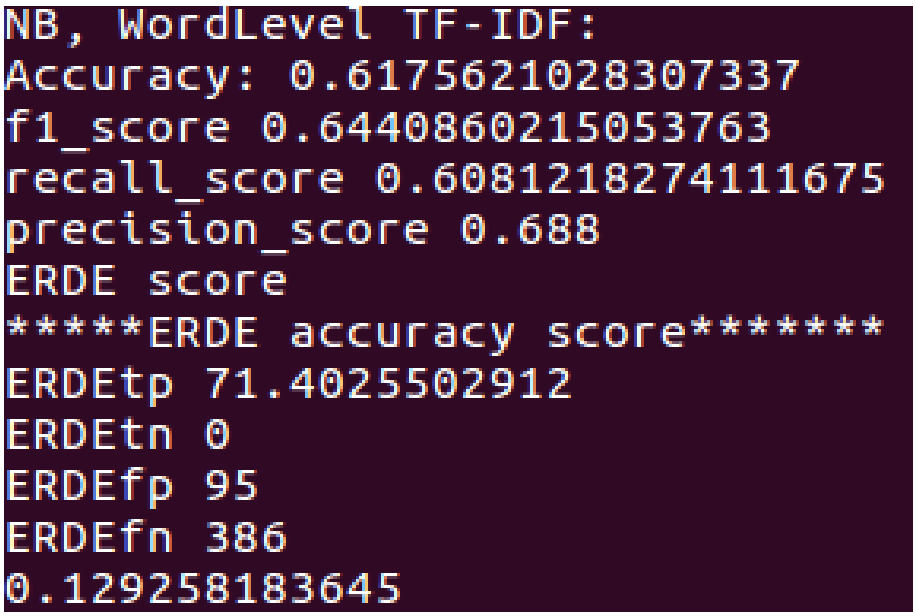

Figure 4.14 Naive Bayes classification - Image 2

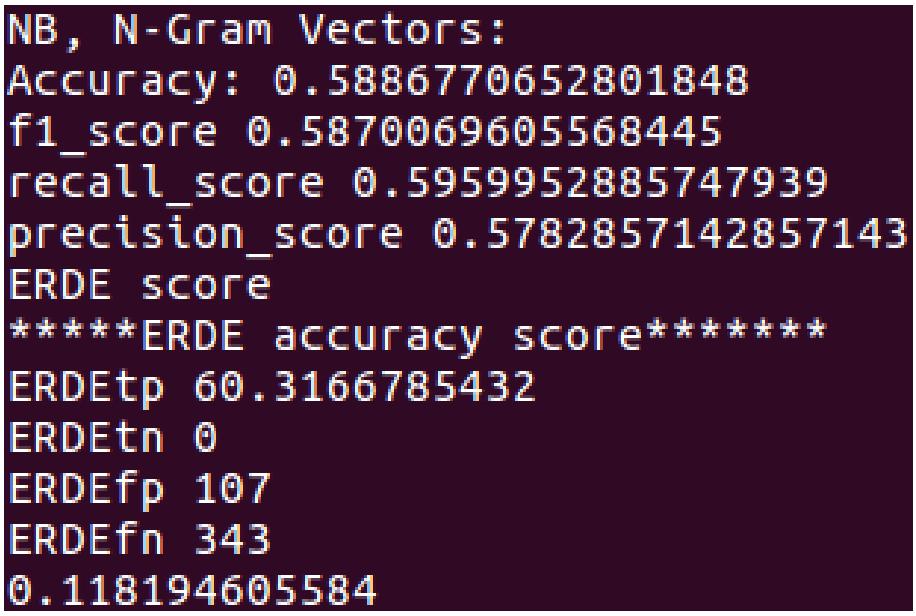

Figure 4.15 Naive Bayes classification - Image 3

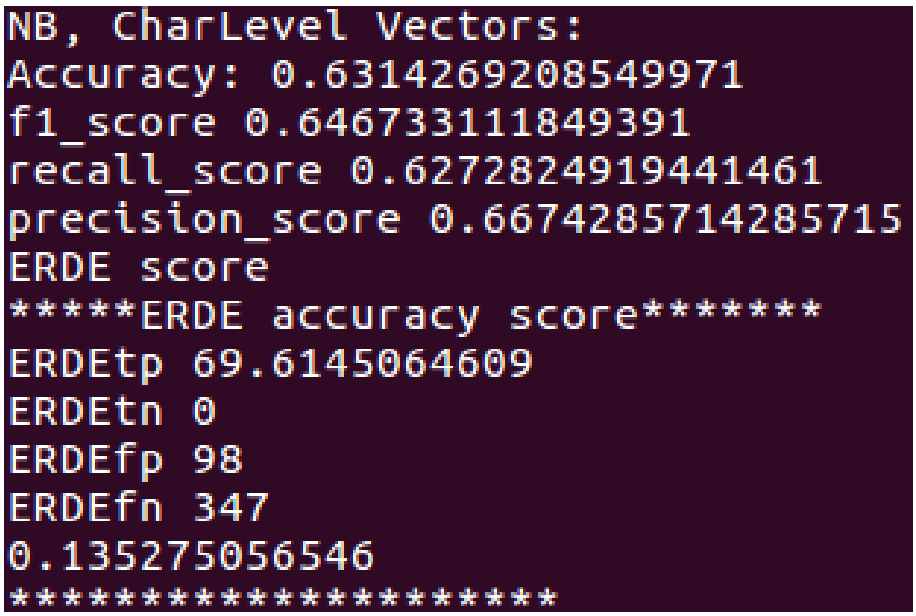

Figure 4.16 Naive Bayes classification - Image 4

Convolutional Neural Network

CNN has set of layers namely the input layer, hidden layers and the output layer including the convolutional, pooling, normalization and fully-connected layers. A neural network is considered convolutional if it has at least one convolutional layer.

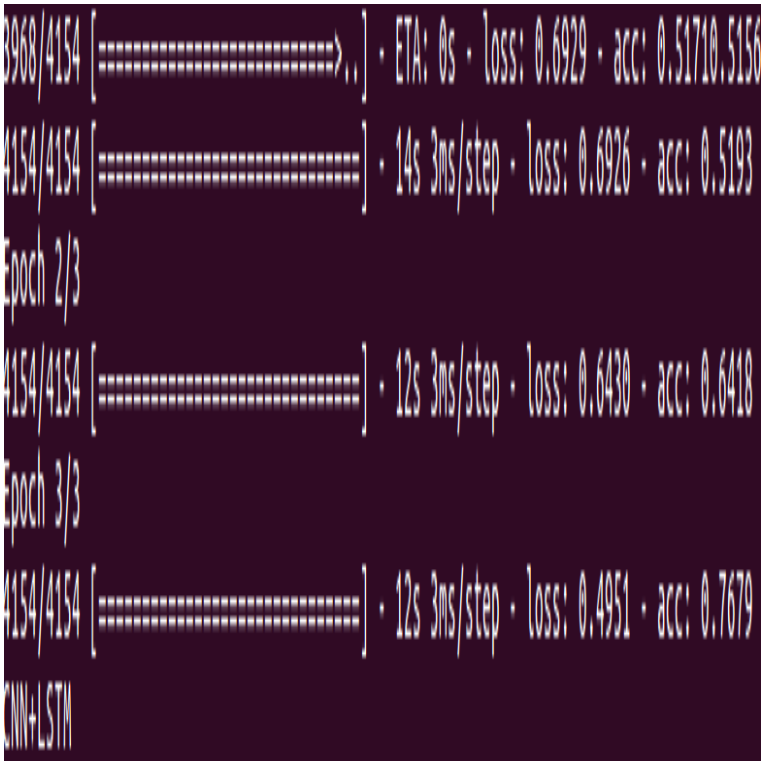

Figure 4.17 Convolutional Neural Network - Image 1

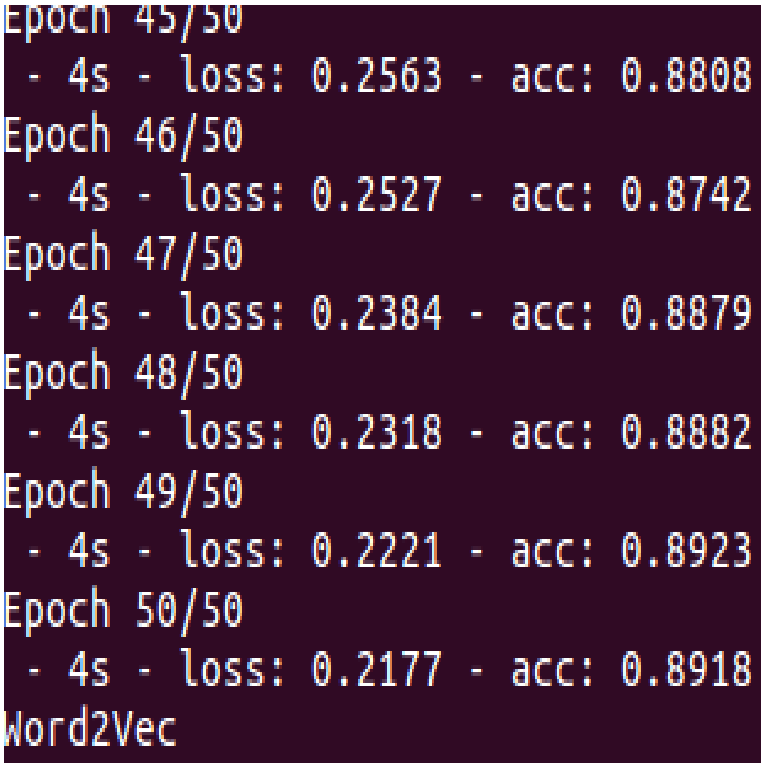

Figure 4.18 Convolutional Neural Network - Image 2

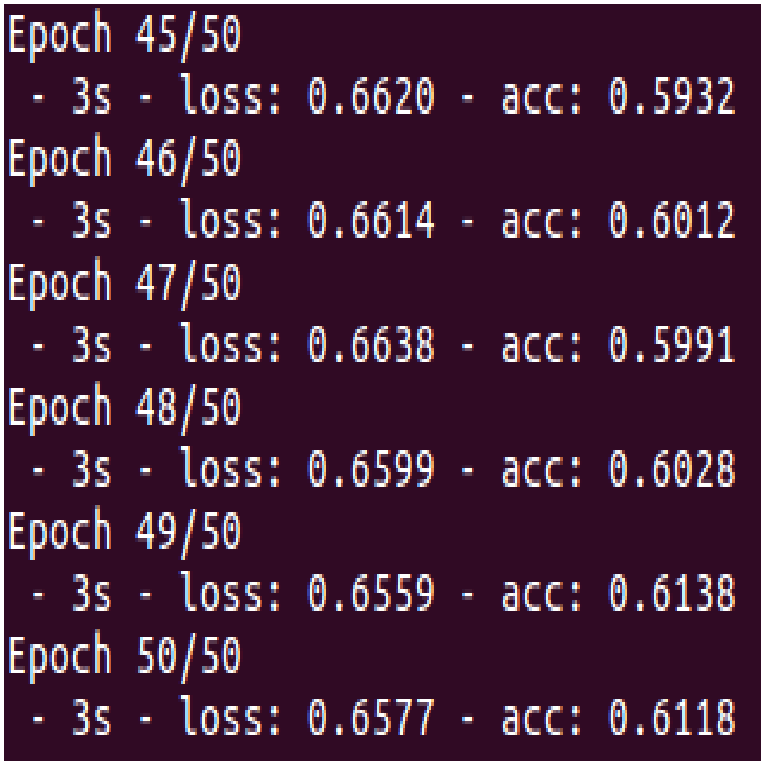

Figure 4.19 Convolutional Neural Network - Image 3

Support Vector Machine

The SVM works by dividing a linearly separable data into two classes with the maximum between-class distance (also termed as margins). The SVM finds an optimally distant line that lies as far from the nearest class points as possible.

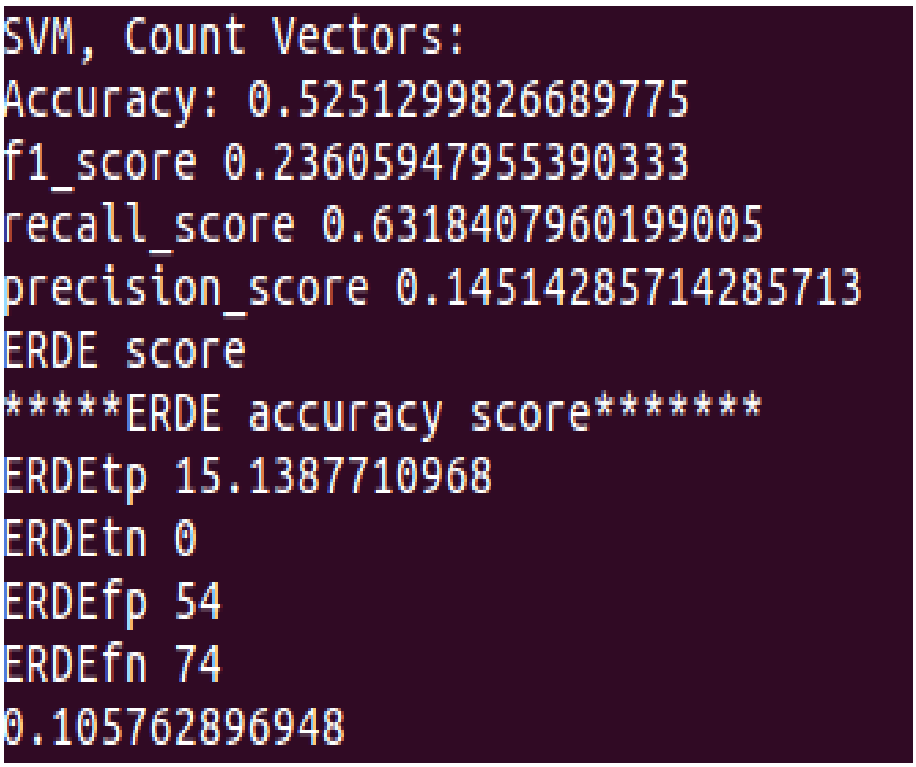

Figure 4.20 Support Vector Machine - Image 1

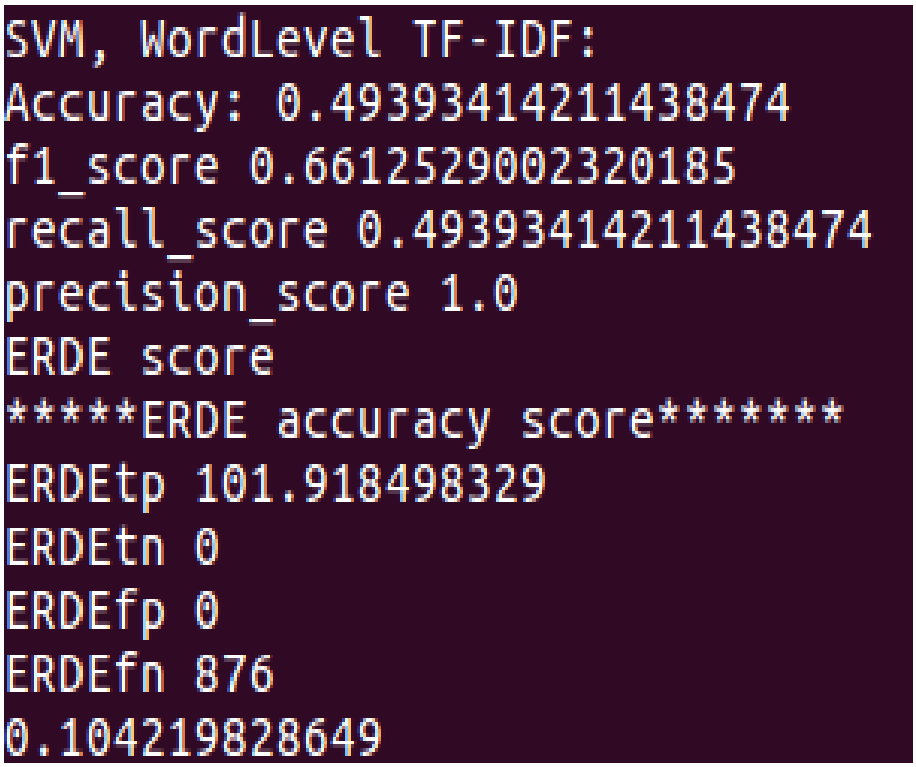

Figure 4.21 Support Vector Machine - Image 2

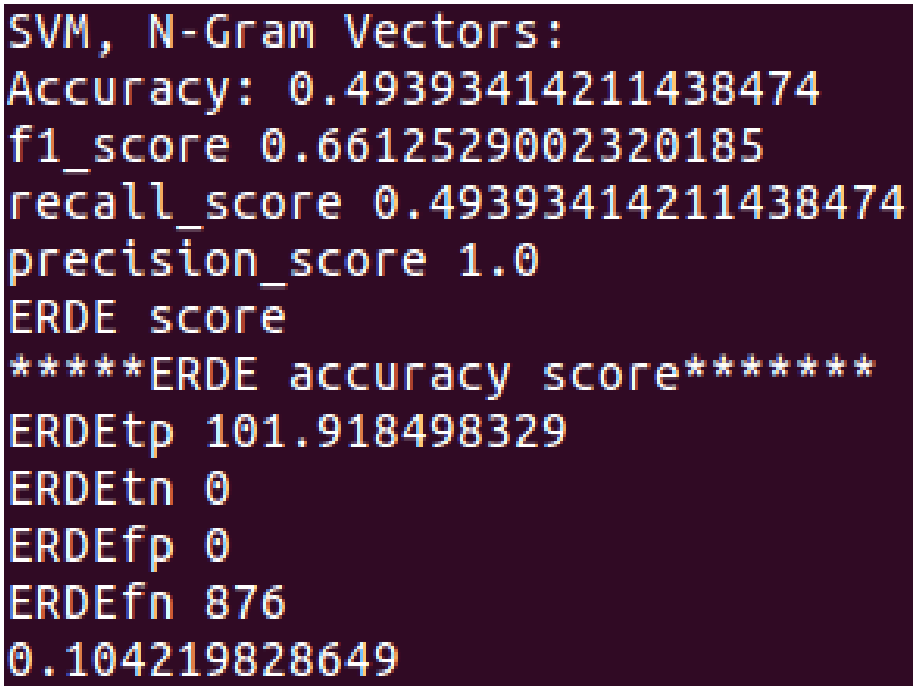

Figure 4.22 Support Vector Machine - Image 3

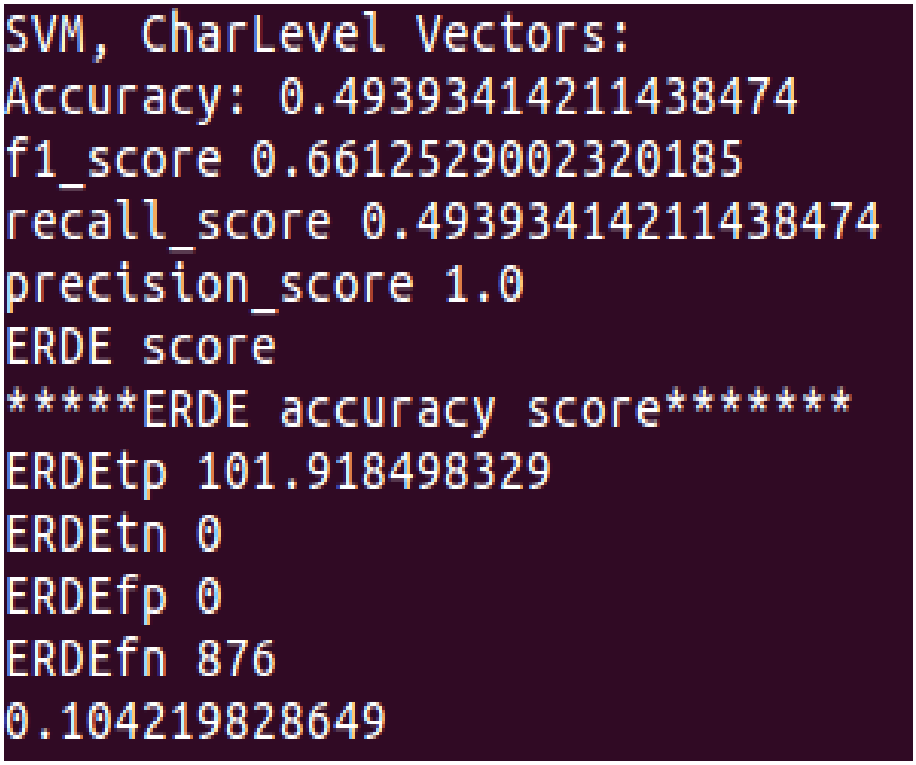

Figure 4.23 Support Vector Machine - Image 4

4.8 PERFORMANCE METRICS

The standard performance metrics of Precision and Recall will be used to evaluate the accuracy of the model. In the case of the detection of depression :

- True Positive - Actual number of depressed people who are correctly diagnosed.
- True Negative -Actual number of non depressed people who are correctly diagnosed
- False Postive - Number of non depressed people who are classified as depressed.
- False Negative - Number of depressed people who are classified as non depressed.

Precision is given as follows:

$$Precision = \frac{TruePositive}{TruePositive + FalsePostive} \tag{4.1}$$

Recall is calculated as follows:

$$Recall = \frac{TruePositives}{TruePositive + FalseNegative} \tag{4.2}$$

The F1 Score can be calculated as:

$$F1Score = \frac{2.Precision.Recall}{Precision + Recall} \tag{4.3}$$

When prediction is equal to false positive, ERDE score is calculated as :

$$ERDEScore = \frac{n.(TruePositive)}{n.(Users)} \tag{4.4}$$

When prediction is equal to false negative, ERDE score is calculated as :

$$ERDEScore = 1 \tag{4.5}$$

When prediction is equal to true positive, ERDE score is calculated as :

$$ERDEScore = 1 - \frac{1}{e^{(}delay - o)} \tag{4.6}$$

When prediction is equal to true negative, ERDE score is calculated as :

$$ERDEScore = 0 \tag{4.7}$$

Furthermore, a graph is drawn comparing each classifier with one another, using a set of features such as Count Vectors,Word TF-IDF, N-Gram TF-IDF and Char Vectors, demonstrated by Figure 4.24 and Figure 4.25.

Table 4.1 Accuracy and ERDE Score

Serial No.	Classifier	Feature	Accuracy	ERDE5
1	NB	CV	60.7	12.6
2	NB	Word-TF-IDF	61.3	12.9
3	NB	N-gram-TF-IDF	58.8	11.8
4	NB	Char-TF-IDF	63.1	13.5
5	LR	CV	59.5	12.07
6	LR	Word-TF-IDF	59.5	12.07
7	LR	N-gram-TF-IDF	62.9	13.6
8	LR	Char-TF-IDF	58.5	11.6
9	SVM	CV	52.5	10.5
10	SVM	Word-TF-IDF	49.3	10.4
11	SVM	N-gram-TF-IDF	49.6	10.7
12	SVM	Char-TF-IDF	51.5	11.2
13	LSTM+CNN	Without Word Embeddings	61.1	6.6
14	LSTM+CNN	GloVe	76.79	13.2
15	LSTM+CNN	Word2Vec	89.1	12.06

4.9 COMPARISON CHART

The comparison chart provides a visual representation for the comparison of various classifiers, using the Count Vector, Word TF-IDF, N-Gram TF-IDF and Character vectors.The accuracy of every classifier is plotted on the graph and it is seen that the Convolutional Neural Network with Word2Vec word embeddings gives the highest accuracy and lesser ERDE score compared to other classifiers, for the prediction of depression.

Count vectors count the number of times a word appears in a text and we might get high count vectors for one document and low count vectors for others.This can be a problem, so to solve this problem we use a technique called TF-IDF(Term frequency-Inverse term frequency). TF-IDF is a weighting factor which is used to get the important features

from the documents(corpus).It actually tells us how important a word is to a document in a corpus, the importance of a word increases proportionally to the number of times the word appears in the individual document, this is called Term Frequency(TF).The inverse document frequency(IDF) is a measure of how much information the word provides, that is, whether the term is common or rare across all documents.

In the fields of computational linguistics and probability, an N-gram is a contiguous sequence of N items from a given sample of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. The N-grams typically are collected from a text or speech corpus.Using Latin numerical prefixes, an n-gram of size 1 is referred to as a ”unigram”; size 2 is a ”bigram” (or, less commonly, a ”digram”); size 3 is a ”trigram”. English cardinal numbers are sometimes used, e.g., ”four-gram”, ”five-gram”, and so on. In computational biology, a polymer or oligomer of a known size is called a k-mer instead of an n-gram, with specific names using Greek numerical prefixes such as ”monomer”, ”dimer”, ”trimer”, ”tetramer”, ”pentamer”, etc., or English cardinal numbers, ”one-mer”, ”two-mer”, ”three-mer”.

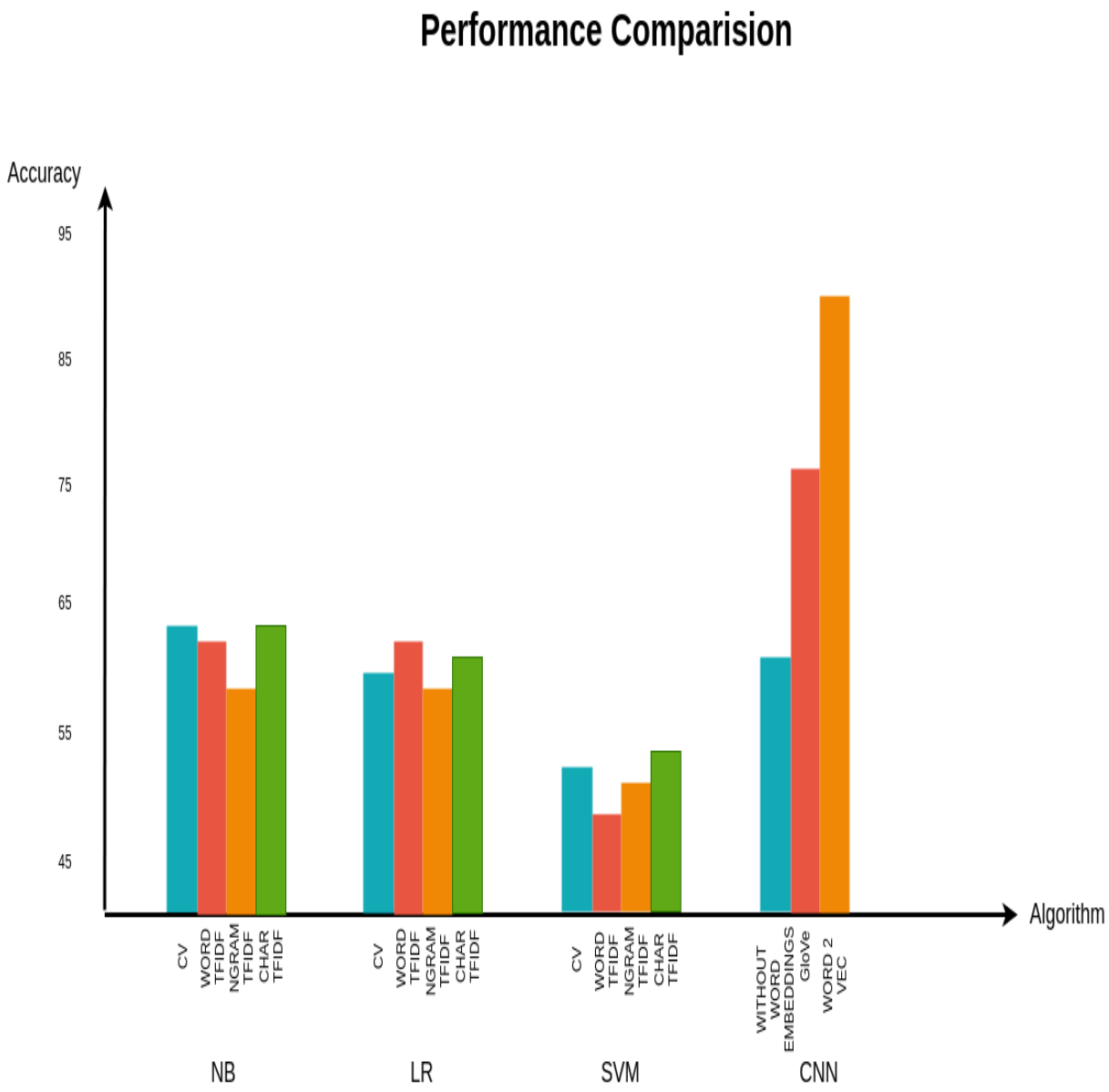

Figure 4.24 Comparsion Chart - Accuracy

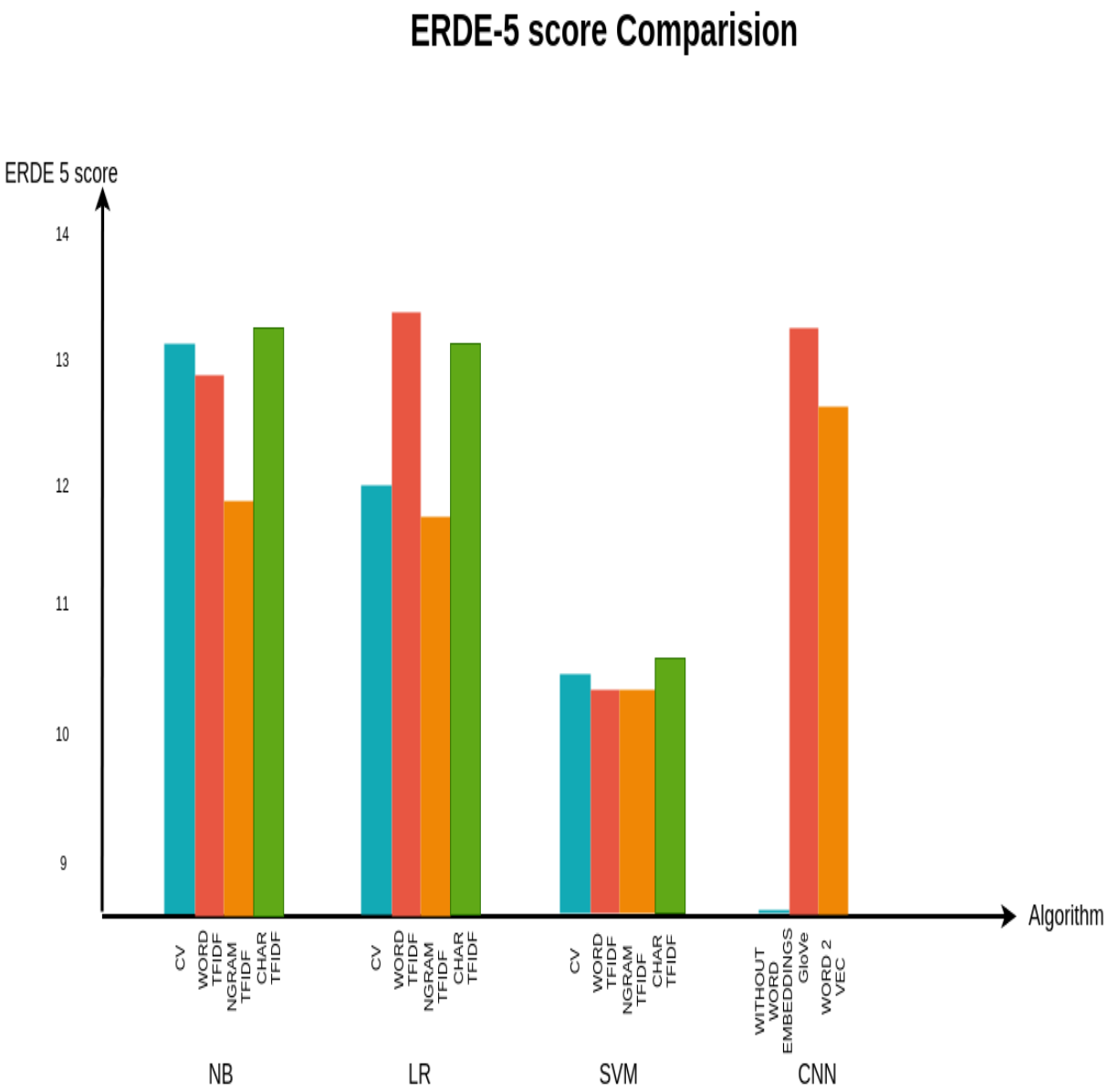

Figure 4.25 Comparsion Chart - ERDE Score

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 SUMMARY

This is a system for the early detection of depression and takes the RSDD data as input, which is then pre processed and used to train a host of classifiers using Machine Learning, which then output the accuracy of the predictions.The accuracy for different classifiers are compared to each other and since the dataset obtained is small, ERDE score is calculated so as to identify the best classifier. The pre-processed data is then subjected to the word embeddings module and then is used to train and test the Neural network.[10] The final evaluation metrics involve F1 scores and ERDE scores.

5.2 CRITICISMS

One of the major issues with the model is that the approach used considers only language features for prediction and that is a weakness in itself. Using features such as voice and face can lead to better prediction rates. The second weakness is the small size of the data set available. Most of the available data is privatized and hence leads to difficulty in accessing it.Though the model accommodates for scalability, the tightest upper bound of the level of scalability is still unknown. Training machine learning models with a huge data set is extremely time consuming and still might not give accurate results.

5.3 FUTURE WORK

One priority of future work in this area should be to agree on a new metric for early detection tasks like eRisk. Ethical issues in this area of research have been reviewed and should find more attention as well. Possibilities to publish the fastText model trained on reddit comments still have to be examined. Concerning the models presented in this work, additional experiments will be necessary to find better ways to integrate the metadata features directly into the neural network.

REFERENCES

[1] D. Xu F. Sadeque and S. Bethard, "Linear and recurrent models for early depression detection", In *Proceedings Conference and Labs of the Evaluation Forum CLEF 2017*, 2017.

[2] S. Hurst J. Mikal and M. Conway, "Ethical issues in using twitter for population-level depression monitoring: a qualitative study", In *BMC Medical Ethics*, 2016.

[3] Atreyee Mukherjee Zeeshan Ali Sayyed JT Wolohan, Misato Hiraga, "Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with nlp", *Proceedings of the First International Workshop on Language Cognition and Computational Models*, pp. 11–21, 2018.

[4] S. Counts M. De Choudhury and E. Horvitz, "Social media as a measurement tool of depression in populations", In *Proceedings of the 5th Annual ACM Web Science Conference.* ACM, 2016.

[5] T. Nguyen C. T. Mazumdar . H. Liu D. M.Hartley M. Torii, L. Yin and l N. P. Nelson, "An exploratory study of a text classification framework for internet-based surveillance of emerging epidemics", In *International Journal of Medical Informatics*, 2015.

[6] Sven Koitka Marcel Trotzek and Christoph M. Friedrich, "Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences.", *IEEE transactions on Knowledge and data engineering.*, 2018.

[7] G. Borges R. Bruffaerts W. T. Chiu G.De Girolamo J. Fayyad O. Gureje J. M. Haro Y. Huang P. S. Wang, M. Angermeyer, "Delay and failure in treatment seeking after first onset of mental disorders in the world health organizations world mental health survey initiative", In *World Psychiatry*. IEEE, 2017.

[8] Margaret L Kern Lyle H Ungar Sharath Chandra Guntuku, David B Yaden and Johannes C Eichstaed, "Detecting depression and mental illness on social media: an integrative review.", *Current Opinion in Behavioral Sciences*, pp. 43–49, 2017.

[9] R. Whitley and R. D. Campbell, "Stigma, agency and recovery amongst people with severe mental illness", In *Social Science Medicine*, 2016.

[10] Yang Ji Li Sun Xinyu Wang, Chunhong Zhang and Leijia Wu, "Depression detection model based on sentiment analysis in microblog social network", In *Language Processing and Knowledge in the Web*. Springer, 2015.

[11] Andrew Yates, Arman Cohan, and Nazli Goharian, "Depression and self-harm risk assessment in online forums.", *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 2958–2968, 2017.

[11] [7] [9] [10] [1] [2] [8] [3] [5] [4] [6]