DEPRESSION DETECTION USING SOCIAL MEDIA STREAMS

Implementation Report

GitHub Repository: https://github.com/gargi91/Depression-Detection-Using-Social-Media

Oct 13, 2021

2019059 Gargi Chaurasia
2019221 Ayush Dubey

Abstract

Social media analysis has shown promising results for public health assessment and monitoring. In this research, we explore the task of automatically analysing social media textual data using Natural Language Processing (NLP) and Machine Learning (ML) techniques to detect signs of depression, a mental health disorder that needs attention. In our proposed models, we use feature engineering with supervised machine learning algorithms (such as NB, SVM, RF, and KNN), and we compare their results with those of deep learning algorithms. Adopted the CNN model for depression classification that obtained the highest F1-score on the test dataset 0.95 recall. This model is later used to detect the depression of a given text.

Further, we compare our results from the research paper Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences. This paper was having few limitations which we have tried to eradicate in our implementation such as better data preprocessing and use of recently published language modelling methods like BERT.

Problem Statement

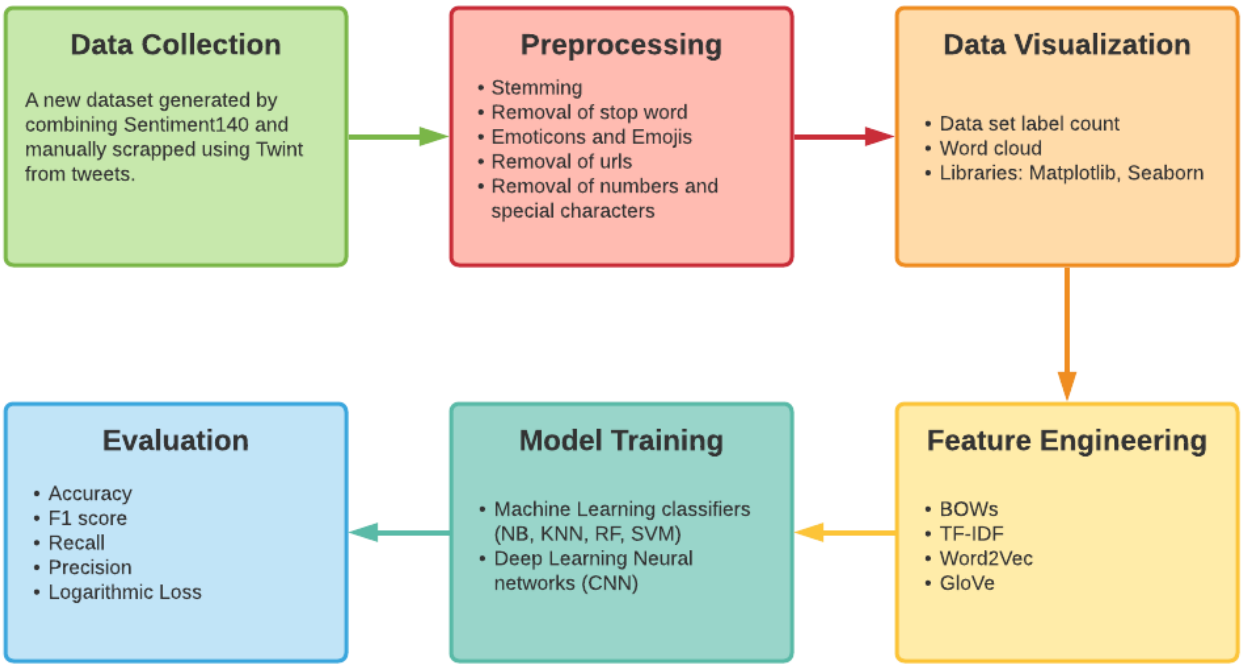
Our focus in this paper is to build a generalizable classification model using the techniques of Natural Language Processing (NLP) and Machine Learning Algorithms

(ML) that can identify the depressed users through their social media posts e.g tweets. We have used different methods to select the best features that lead the model scale up. Conventional Machine Learning algorithms and neural networks will be applied to reach this objective.

We will experiment with different features, including word embeddings, to distinguish between depressed and nondepressed social media users and probe textual features to identify people at high depression risk.

1. Overview

Two approaches are used in this study: traditional machine learning and deep learning. The dataset was classified into depressed and not depressed tweets with label 1 and 0 respectively. Some prominent traditional machine learning algorithms leveraged for classification are: support vector machine (SVM), random forest, k-nearest neighbour and naive bayes classifier. Feature extraction is done using statistical features, term frequency inverse document frequency (TF-IDF), and topic modeling that is specifically used in deep learning models. For the deep learning approach, we have experimented with one classifier, namely convolutional neural network (CNN). Since the neural networks cannot deal with tweets directly, we used two well-known word embedding techniques called GloVe by Stanford and Word2vec by Google for the vector representation. The results showed that Random Forest with F1 Score .99 is the best out of the seven classifiers that we have experimented with, with an accuracy of 99.77% with features of BOW and TF-IDF.



An Overview of Machine Learning Process

1.1 Depression

According to the World Health Organization (WHO), more than 300 million people worldwide are suffering from depression, which equals about 4.4 percent of the global population. While forms of depression are more common among females (5.1 percent) than males (3.6 percent) and prevalence differs between regions of the world, it occurs in any age group and is not limited to any specific life situation. Depression is therefore often described to be accompanied by paradoxes, caused by a contrast between the self-image of a depressed person and the actual facts. Latest results from the 2016 National Survey on Drug Use and Health in the United States report that, during the year 2016, 12.8 percent of adolescents between 12 and 17 years old and 6.7 percent of adults had suffered a major depressive episode (MDE).

Precisely defining depression is not an easy task, not only because several subtypes have been described and changed in the past, but also because the term “being depressed” has become frequently used in everyday language. In general, depression can be described to lead to an altered mood and may also be accompanied, for

example, by a negative self-image, wishes to escape or hide, vegetative changes, and a lowered overall activity level [2, p. 8]. The symptoms experienced by depressed individuals can severely impact their ability to cope with any situation in daily life and therefore differ drastically from normal mood variations that anyone experiences.

At the worst, depression can lead to suicide. WHO estimates that, in the year 2015, 788,000 people have died by suicide and that it was the second most common cause of death for people between 15 and 29 years old worldwide. In Europe, self-harm was even reported as the most common cause of death in the age group between 15 and 29 and the second most common between 30 and 49, again in results obtained by WHO in 2015.

1.2 Social Media & Mental Health

Social media use has risen sharply over the last ten years. In 2020, over 3.6 billion people were using social media worldwide, a number projected to increase to almost 4.41 billion in 2025. Social media offers advantages over traditional data sources, including ease of access, reduced cost, and up-to-date data availability. We can learn about public health topics by passively analyzing existing social media data (Paul et al., 2016). Thus, analyzing social media posts can play an important alternative in monitoring trends instantly and identifying mental disorders throughout the population. The importance of this field emerges with the progressive rise in the number of depressed users due to the COVID-19 pandemic and an anticipated increase in the number of suicides (Sher, 2020), this field becomes even more promising in providing decision-makers, such as Public Health Agency of Canada (PHAC, with rapid tools to mitigate risks.

Although the severity of depression is well-known, only about half of the individuals affected by any mental disorder in Europe get treated. The proportion of individuals seeking treatment for mood disorders during the first year ranges between 29–52 percent in Europe, 35 percent in the USA, and only 6 percent in Nigeria or China. In addition to possible personal reasons for avoiding treatment, this is often due to a limited availability of mental health care.

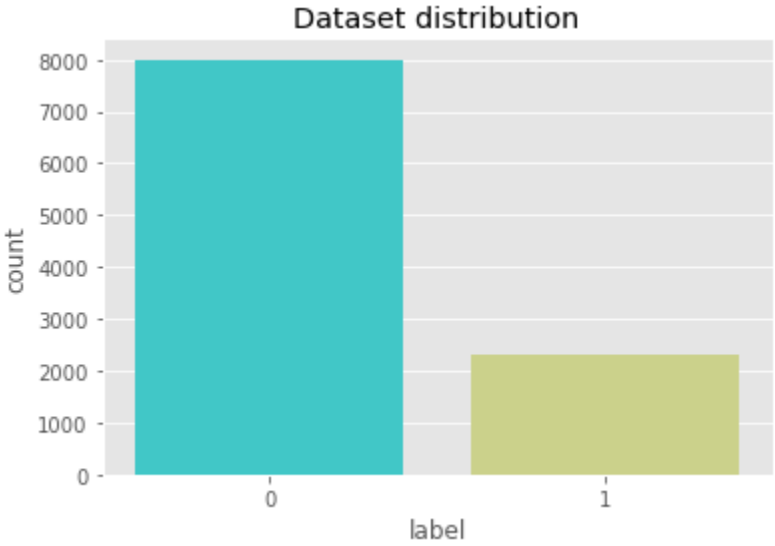
Researchers found out that shame and self-stigmatization seem to be much stronger reasons to not seek psychiatric help than actual perceived stigma and negative reactions of others. They further speculate that the fear of discrimination might be

relatively unimportant in their study because people hope to keep their psychiatric treatment secret.

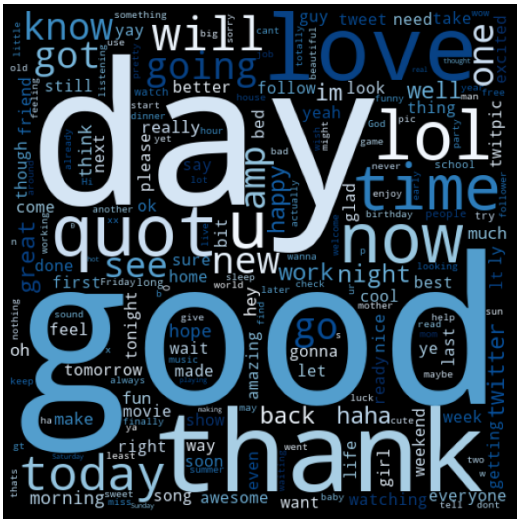
While depression and other mental illnesses may lead to social withdrawal and isolation, it was found that social media platforms are indeed increasingly used by affected individuals to connect with others, share experiences, and support each other. Based on these findings, peer-to-peer communities on social media can be able to challenge stigma, increase the likelihood to seek professional help, and directly offer help online to people with mental illness. A similar study in the USA came to the conclusion that internet users with stigmatized illnesses like depression or urinary incontinence are more likely to use online resources for health-related information and for communication about their illness than people with another chronic illness. All this emphasizes the importance of research toward ways to assist depressed individuals on social media platforms and on the internet in general.

2. Dataset

The dataset utilized in all experiments for the referred paper was first described in 2016 for research on depression and language use and then finally published as part of the CLEF 2017 conference eRisk pilot task on early detection of depression. It contains chronological sequences of posts and comments from reddit.com, collected for a total of 135 depressed users and a random control group of 752 users. Due to privacy and copyright issues the authors of the paper could not provide the dataset.



Therefore, a new dataset is used in the implementation, which has been generated by combining part of the Sentiment140 (8,000 positive tweets), and another one for depressive tweets (2,314 tweets), with a total of 10,314 tweets. This dataset has been provided in [this] github repository. Each data in the dataset consists of the *tweet* and *label*.



Wordcloud for label=0

Wordcloud for label=1

2.1 Why Tweets?

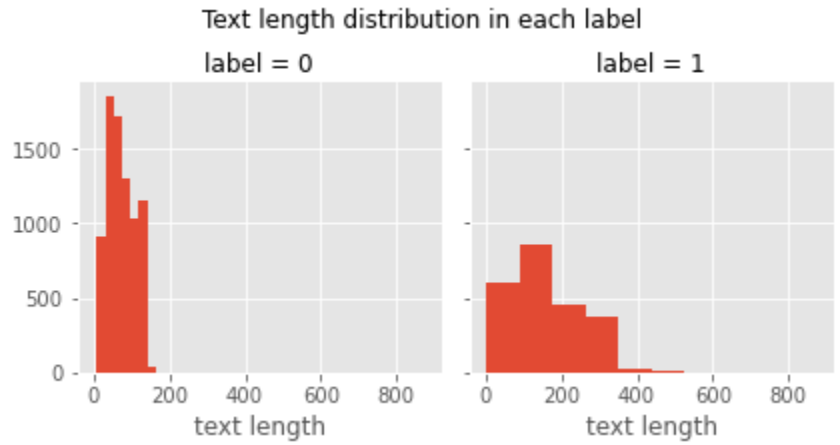
Twitter is the most attractive social network service among researchers. Twitter currently ranks as one of the world's leading social networks based on active users. Every update the user posts to his followers on Twitter is called a tweet. Tweets are mostly accessible to the public and can be obtained and analyzed, unless flagged by the user as "private". That is the reason behind choosing twitter over other social media platforms.
Tweets can be collected using Twitter API by searching the tweets for specific keywords, hashtags, or any defined query and can be limited to particular locations, hashtags and time periods.

Driven by the growing availability of data, for example through social media, and the technological advances that allow researchers to work with this data, ethical considerations are becoming more and more important in the field of Natural Language Processing (NLP). Based on these developments, NLP has changed from being mostly focussed on improving linguistic analysis towards actually having an impact on individuals based on their writings.

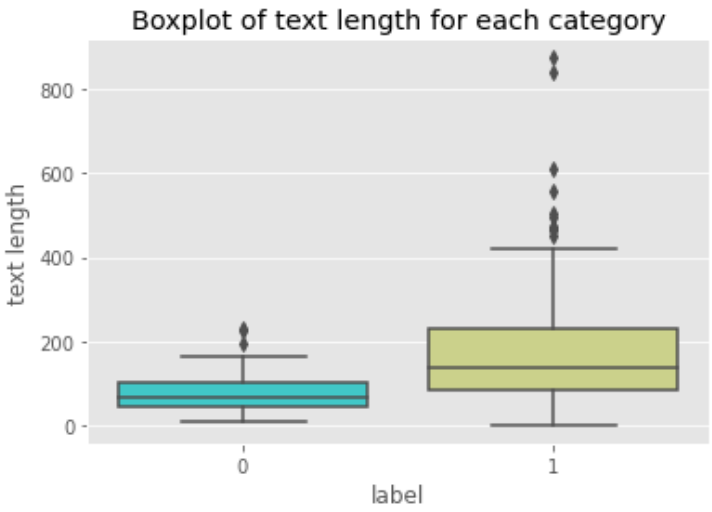
Unlike other social media platforms, tweets are posted publicly. Also, the user can make their profile after keeping it public, which also makes their posts private. So, procuring data from these platforms could raise privacy issues if user(s) decide to make their profile private after their posts are scrapped.

2.2 Preprocessing

The raw dataset procured contained tweets with urls, hashtags, user handles and stopwords which are immaterial in classification of depression.
Data cleaning is done, removing the mentioned text from each message in the dataset. Punctuations, numeric values, and special characters are also removed, since they contribute nothing to the classification task.



Also, previous studies have already shown that depression also has an effect on the language used by affected individuals. For example, a more frequent use of first person singular pronouns in spoken language was first observed in 1981. Therefore, first person singular pronouns are barred from stopwords corpus and kept in the dataset.



3. Depression Classification Methods

3.1 Feature Extraction

• Statistical Features: Statistical features include the number of words per tweet, word density, number of unique words used, number of characters per tweet. For exploratory reasons, the number of words per tweet is used for this phase of implementation.

• Tf-idf: The term frequency (tf) of a document is the normalized occurrence of a term within a document, but most used terms such as "the" and "is" will have high term frequency with no added value. On the other hand, the inverse document frequency (idf) measures the value that the term provides to the document, for that tf-idf represents the significance of a term in a document based on the frequency of its appearance in the current document and in the other documents in the corpus.

• Topic Modeling: Topic modeling is a probabilistic model for finding hidden semantic structures. It is an unsupervised method that considers the set of user posts as an aggregate of latent topics in which a topic is a distribution of co-occurring terms. Different terms which convey an associated aspect are grouped together.

• Word Embeddings: Finally, the linguistic features extracted are exploited using word embeddings. Word embedding represents a document’s vocabulary using language modeling and feature learning techniques in natural language processing (NLP) as

a dense vector that captures the terms' semantics. Using unsupervised learning approaches, vocabulary terms are initialized with fixed-length continuous-valued vectors then trained using a large corpus of text to calculate distributed representation of words using vector arithmetic based on the company it keeps. The resulting word embedding vectors can be either context-independent such as Word2vec and GloVe.

3.2 Traditional Models

Several traditional models are trained to perform classification. Since multiple feature extraction methods are used, each feature extraction is trained on each model.

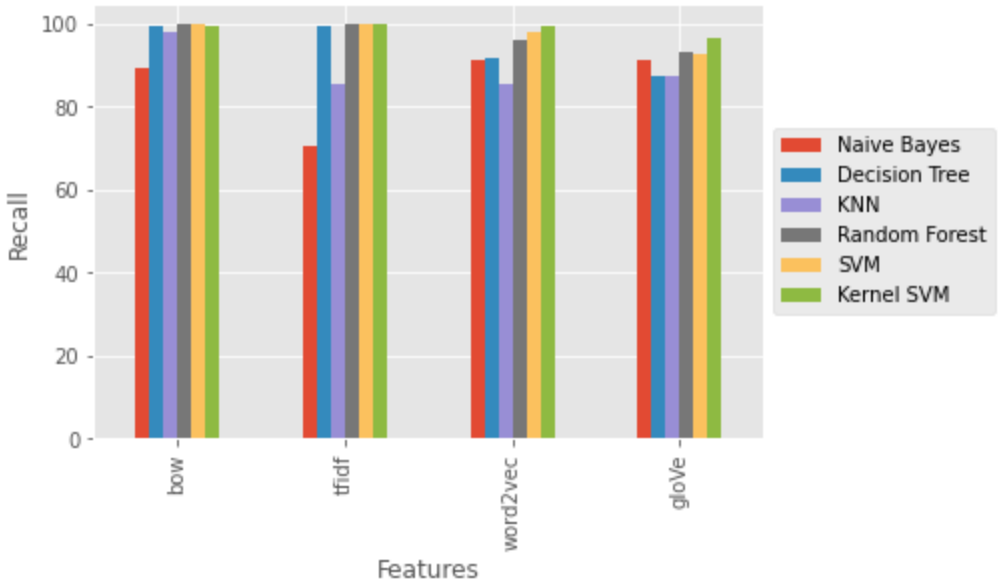
- Naive Bayes Classifier: A Naive Bayes classifier is a probabilistic machine learning model that's used for classification tasks. The crux of the classifier is based on the Bayes theorem.

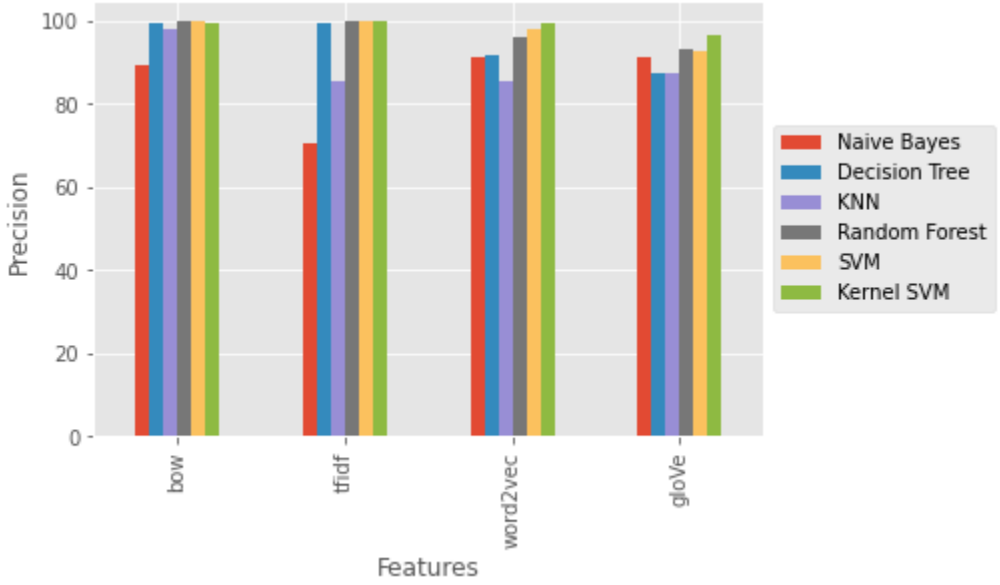
- Decision Tree: are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation.

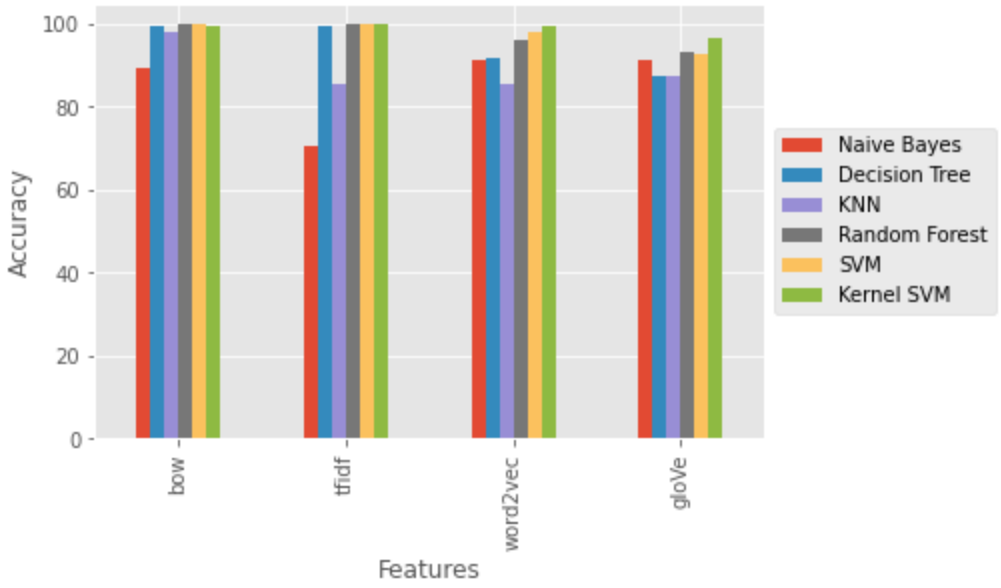
- K Nearest Neighbour: In k-NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.
  Using, elbow method various k-NN models with different k values were trained, to get the value of k with least error rate. Where for bag of words feature extraction, the optimum value of k was found out to be 38.
- Random Forest: Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees.

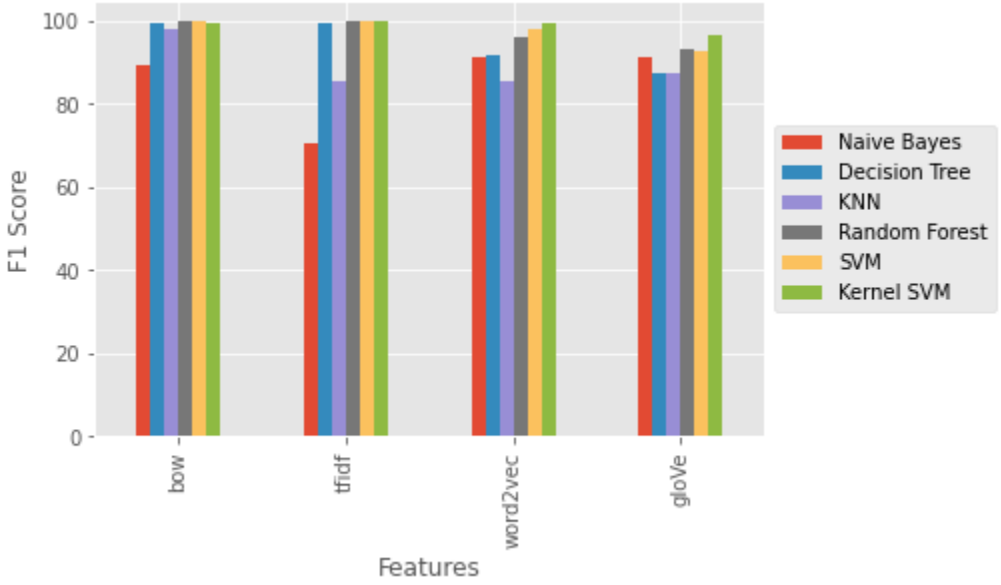
- SVM: In machine learning, support-vector machines are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis.

- Kernel SVM: "Kernel" is used due to a set of mathematical functions used in Support Vector Machine that provides the window to manipulate the data. So, Kernel Function generally transforms the training set of data so that a non-linear decision surface is able to transform to a linear equation in a higher number of dimension spaces. Basically, It returns the inner product between two points in a standard feature dimension.


Recall
100
80
60
40
20
0
bow
tfidf
word2vec
gloVe
Features
Naive Bayes
Decision Tree
KNN
Random Forest
SVM
Kernel SVM


Precision
100
80
60
40
20
0
bow
tfidf
word2vec
gloVe
Features
Naive Bayes
Decision Tree
KNN
Random Forest
SVM
Kernel SVM





3.3 Deep Learning Model

Deep Learning algorithm has obtained superior results on multiple language-related tasks in comparison to conventional machine learning algorithms. And hence the target is to increase the recall and F1-score in the test dataset.
CNN model is applied on preprocessed dataset using word embedding (GloVe).

- CNN: In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery. For instance,

CNN is used for applications such as image classification , facial recognition , object detection etc.
Most recently, however, Convolutional Neural Networks have also found prevalence in tackling problems associated with NLP tasks like Sentence Classification, Text Classification, Sentiment Analysis, Text Summarization, Machine Translation and Answer Relations.

3.4 Results

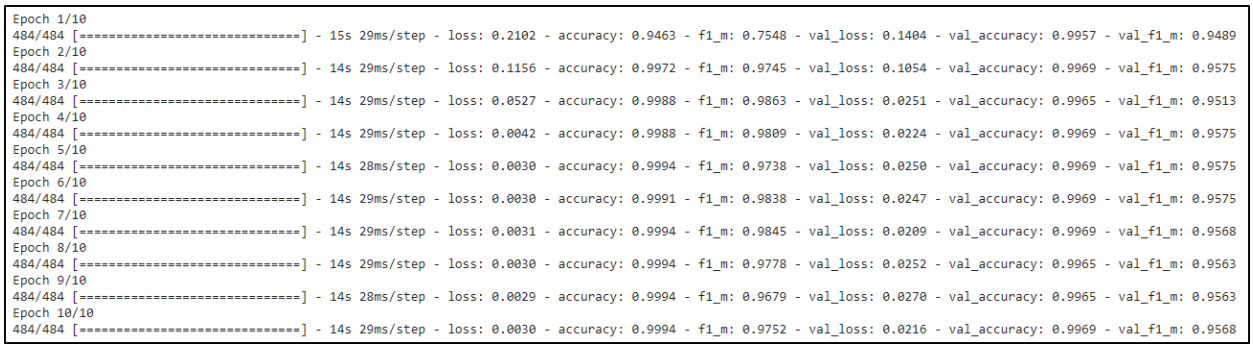
3.4.1 Reference Paper

In the reference research paper used, the maximum F score achieved was 0.73 against the Meta LR Wait model.

Model	$p >$	$ERDE_5$	$ERDE_{50}$	F_1	P	R
UNSLA [54]		13.66	9.68	0.59	0.48	0.79
FHDO-BCSGA [54]		12.82	9.69	0.64	0.61	0.64
FHDO-BCSGB [54]		12.70	10.39	0.55	0.69	0.46
TVT-NB [62]		13.13	8.17	0.54	0.42	0.73
TVT-RF [62]		12.30	8.95	0.56	0.54	0.58
GloVe W+N	0.5	12.95	7.57	0.63	0.56	0.73
GloVe Crawl	0.7	12.98	8.59	0.63	0.58	0.69
fastText Wiki	0.6	13.06	8.17	0.57	0.47	0.71
fastText W+N	0.55	13.11	7.95	0.60	0.49	0.77
fastText Crawl	0.6	13.01	8.60	0.64	0.60	0.67
fastText reddit	0.7	13.52	8.04	0.62	0.51	0.79
fastText reddit	0.8	12.71	9.23	0.56	0.63	0.50
Meta LR	0.35	12.65	8.57	0.66	0.59	0.73
Meta LR	0.55	12.35	9.86	0.65	0.72	0.60
Meta LR Wait	0.35	13.32	11.33	**0.73**	**0.77**	0.69
G W+N + Meta LR	0.45	12.34	8.93	0.71	0.72	0.69
fT Wiki + Meta LR	0.35	13.52	**7.29**	0.55	0.41	**0.85**
fT Wiki + Meta LR	0.5	**12.13**	8.77	0.71	0.71	0.71
fT reddit + Meta LR	0.55	12.46	8.77	0.67	0.69	0.65

3.4.2 Proposed Model(s)

- In the proposed model (CNN), the maximum F score achieved is 0.95.



```
Epoch 1/10
484/484 [==============================] - 15s 29ms/step - loss: 0.2102 - accuracy: 0.9463 - f1_m: 0.7548 - val_loss: 0.1404 - val_accuracy: 0.9957 - val_f1_m: 0.9489
Epoch 2/10
484/484 [==============================] - 14s 29ms/step - loss: 0.1156 - accuracy: 0.9972 - f1_m: 0.9745 - val_loss: 0.1054 - val_accuracy: 0.9969 - val_f1_m: 0.9575
Epoch 3/10
484/484 [==============================] - 14s 29ms/step - loss: 0.0527 - accuracy: 0.9988 - f1_m: 0.9863 - val_loss: 0.0251 - val_accuracy: 0.9965 - val_f1_m: 0.9513
Epoch 4/10
484/484 [==============================] - 14s 29ms/step - loss: 0.0042 - accuracy: 0.9988 - f1_m: 0.9809 - val_loss: 0.0224 - val_accuracy: 0.9969 - val_f1_m: 0.9575
Epoch 5/10
484/484 [==============================] - 14s 28ms/step - loss: 0.0030 - accuracy: 0.9994 - f1_m: 0.9738 - val_loss: 0.0250 - val_accuracy: 0.9969 - val_f1_m: 0.9575
Epoch 6/10
484/484 [==============================] - 14s 29ms/step - loss: 0.0030 - accuracy: 0.9991 - f1_m: 0.9838 - val_loss: 0.0247 - val_accuracy: 0.9969 - val_f1_m: 0.9575
Epoch 7/10
484/484 [==============================] - 14s 29ms/step - loss: 0.0031 - accuracy: 0.9994 - f1_m: 0.9845 - val_loss: 0.0209 - val_accuracy: 0.9969 - val_f1_m: 0.9568
Epoch 8/10
484/484 [==============================] - 14s 29ms/step - loss: 0.0030 - accuracy: 0.9994 - f1_m: 0.9778 - val_loss: 0.0252 - val_accuracy: 0.9965 - val_f1_m: 0.9563
Epoch 9/10
484/484 [==============================] - 14s 28ms/step - loss: 0.0029 - accuracy: 0.9994 - f1_m: 0.9679 - val_loss: 0.0270 - val_accuracy: 0.9965 - val_f1_m: 0.9563
Epoch 10/10
484/484 [==============================] - 14s 29ms/step - loss: 0.0030 - accuracy: 0.9994 - f1_m: 0.9752 - val_loss: 0.0216 - val_accuracy: 0.9969 - val_f1_m: 0.9568
```

- In the proposed model (ML), the maximum F score achieved is 0.99 by Random Forest and Support Vector Machine Classification models.

Traditional Models F-score

	Features			
Models	**BoWs**	**TF-IDF**	**Word2Vec**	**GloVe**
Naive Bayes	0.89	0.70	0.91	0.91
Decision Tree	0.99	0.99	0.91	0.87
k-NN	0.98	0.85	0.85	0.87
Random Forest	0.99	0.99	0.95	0.93
SVM	0.99	0.99	0.97	0.92
Kernel SVM	0.99	0.99	0.99	0.96

4. References

- https://ruor.uottawa.ca/bitstream/10393/42346/7/Skaik_Ruba_2021_thesis.pdf
- https://github.com/viritaromero/Detecting-Depression-in-Tweets
- https://medium.com/saarthi-ai/sentence-classification-using-convolutional-neural-networks-ddad72c7048c
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6111060/