ISSN [Online]: 2583-2654

Early Detection of Depression in Text Sequences

Sophia S[1] Infancia R[2], Savitha Shree M[3], Sandra Carolin S[4], Maria Rebecca D[5]
[1]Department of Computer Science Engineering,
Dr T Thimmaiah Institute of Technology,
[2]Department of Computer Science Engineering,
Dr T Thimmaiah Institute of Technology,
[3]Department of Computer Science Engineering,
Dr T Thimmaiah Institute of Technology,
4Department of Computer Science Engineering,
Dr T Thimmaiah Institute of Technology,

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 project addresses the early detection of depression using machine learning models based on messages on a social platform. This project is therefore focused on ways to classify indications of depression in written texts as early as possible based on machine learning methods. The project combines natural language processing and machine learning to detect depression level from text messages. Natural language processing is used to create a word embedding model which can convert text sentences to feature vectors. Machine leaning using Neural Networks is used to classify the features vector of sentences to depression level.

Keywords: Depression, early detection, linguistic metadata, convolutional neural network, word embeddings, machine learning, natural language, feature vectors.

I. Introduction

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

The work presented in this paper is structured as follows: Section 2 gives an overview of related work concerning depression, its influence on language, and natural language processing methods. Section 3 describes the dataset used in this work, analyzes the evaluation metric of the corresponding task, and proposes an alternative. Section 4 introduces the user based metadata features used for classification, while Section 5 describes the neural network models utilized for this task. Section 6 contains an experimental evaluation of these models and compares them to published results. Finally, Section 7 concludes this work and summarizes the results

II. Methodology

Design a word embedding model to convert sentences to feature vector. Train a Convolutional Neural Network model to detect the depression level from the feature vector. Measure the accuracy of depression detection against CLEF 2017 dataset.

Data Collections

- https://www.kaggle.com/ywang311/twitter-sentiment
- We use the sentiment 140 dataset.
- It contains 1,600,000 tweets extracted using the twitter api. The tweets have been annotated (0 = negative, 4 = positive) and they can be used to detect sentiment.

ISSN [Online]: 2583-2654

Natural Language Processing

Natural Language Processing The work described in this paper belongs to the area of Natural Language Processing (NLP) and text classification in particular. The origins of text classification tasks can be found in early research to automatically categorize documents based on statistical analysis of specific clue words in 1961 . Later, similar research goals lead to rule based text classification systems like CONSTRUE in 1990 and finally the field began to shift more and more to machine learning algorithms around the year 2000. In addition to text categorization, machine learning was also a driving force in other text-based tasks like sentiment analysis, which is focused on extracting opinions and sentiment from text document . It was first used in combination with machine learning to find positive or negative opinions in movie reviews and was then extended to other review domains , as well as completely different areas like social media.

Pre-processing

- target: the polarity of the tweet (0 = negative, 2 = neutral, 4 = positive)
- ids: The id of the tweet (2087)
- date: the date of the tweet (Sat May 16 23:58:44 UTC 2009)
- flag: The query (lyx). If there is no query, then this value is NO_QUERY.
- user: the user that tweeted (robotickilldozr)
- text: the text of the tweet (Lyx is cool)

Feature Extraction

- From the sentences , stop words of English (is, the, was etc) are removed.
- The words are converted to feature vector.
- Feature is in form of term vs frequency of occurrence in the sentence.

Training the Model

- Convolutional neural network model was trained with the feature vector as input and the label (Neutral/Depressed/ Not Depressed) as output.
- A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates.

Algorithm-Training

- Input : Labelled dataset.
- Output : Neural Classifier.
- Step1 : Vectorize training dataset (remove stop words).
- Step2 : Create a Term Frequency(TF) vector for each sentence in dataset.
- Step 3 : Create training set with TF as input and label (neutral, depressed, not depressed) as output.
- Step 4: Train Neural Network with TF as input and label as output.

Algorithm-Classification

- Algorithm : Classification
- Input : social media messages from user, Neural model
- Output : Neutral/Depressed/ Not Depressed
- StatusCount= [0,0,0]
- Stmsg=[“Neutral”, “Depressed”, “Not Depressed”]
- For each message in input
- TF<- vectorize the message
- status = Classify using neural model (TF)
- StatusCount[status]+=1;
- End
- Return Stmsg[Index(Max(StatusCount))]

III. RESULTS AND DISCUSSIONS

(Please see pages from 109 to 112)

IV. CONCLUSION

This work presented a word embedding model to detect depression by processing text sentences. Text sentences are converted to feature vectors. The features are classified to sentiments and from the distribution of sentiments, the depression status is detected. The proposed method was able to achieve 98% accuracy is classification of depression.

The work can be extended in following directions
1. Integrating some features other than text into classification of depression

ISSN [Online]: 2583-2654

2. Long term trend analysis in messaging behavior to remove bias.

REFERENCES

[1] Liu S, Tong J, Menga, Yang J, Zhao X, He F, et al. Study on an effective cross-stimulus emotion recognition model using EEGs based on feature selection and support vector machine. Int J Mach Learn Cybern 2016;9(5):721–6.

[2] Nakisa B, Rastgoo MN, Tjondronegoro D, Chandran V.Evolutionary computation algorithms for feature selectionof EEG-based emotion recognition using mobile sensors.Expert Syst Appl 2018;93:143–55.

[3] Mehmood RM, Lee HJ. A novel feature extraction method based on late positive potential for emotion recognition inhuman brain signal patterns. Comput Electr Eng 2016;53:444–57..

[4] Thammasan N, Moriyama K, Fukui K, Numao M. Familiarity effects in EEG-based emotion recognition.Brain Inform 2017;4:39–50.

[5] Li Z, Tian X, Shu L, Xu X, Hu B. Emotion recognition from EEG using RASM and LSTM. Commun Comput Inform Sci2018;819:310–8.

[6] Zhang Y, Zhang S, Ji X. EEG-based classification of emotions using empirical mode decomposition and autoregressive model. Multimed Tools Appl 2018;77 (20):26697–710.

[7] Momennezhad A. EEG-based emotion recognition utilizing wavelet coefficients. Multimed Tools Appl 2018; 77(20):27089–106.

[8] Zoubi OA, Awad M, Kasabov NK. Anytime multipurpose emotion recognition from EEG data using a liquid state machine based framework. Artif Intel Med 2018;86:1–8.

[9] Thejaswini S, Ravi Kumar KM, Rupali S, Abijith V. EEG based emotion recognition using wavelets and neural networks classifier. Cognit Sci Artif Intel 2017;101–12.

[10] AbdelAal MA, Alsawy AA, Hefny HA. EEG-based emotion recognition using a wrapper-based feature selection method. Adv Intel Syst Comput 2018;639:247–57

[11] Goshvarpour A, Abbasi A, Goshvarpour A. An accurate emotion recognition system using ECG and GSR signals and matching pursuit method. Biomed J 2017;40:355–68.

[12] Chai X, Wang Q, Zhao Y, Liu X, Bai O, Li Y. Unsupervised domain adaptation techniques based on auto- encoder for non-stationary EEG-based emotion recognition. ComputBiol Med 2016;79:205–14.

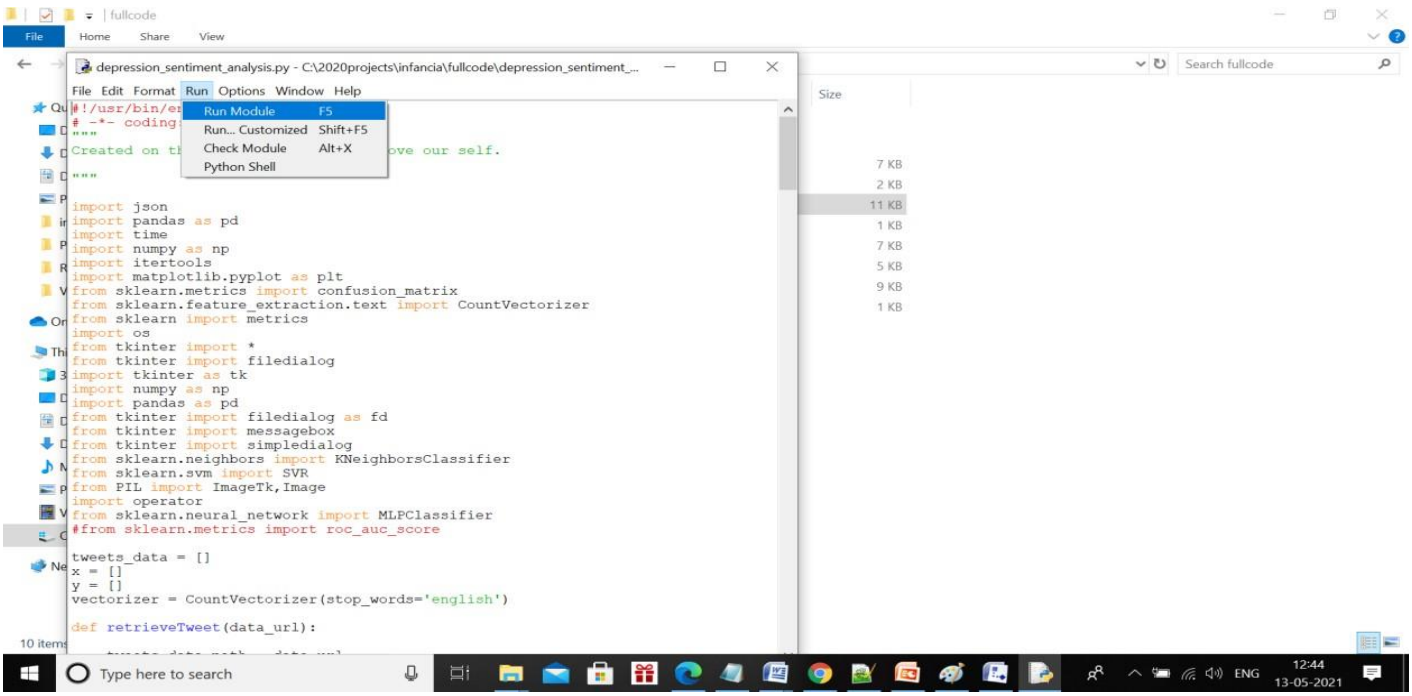
[13] Kaur B, Singh D, Roy PP. EEG based emotion classification mechanism in BCI.Proc. International Conference on Computational Intelligence and Data Science. Proc Comput Sci 2018;132:752–8.

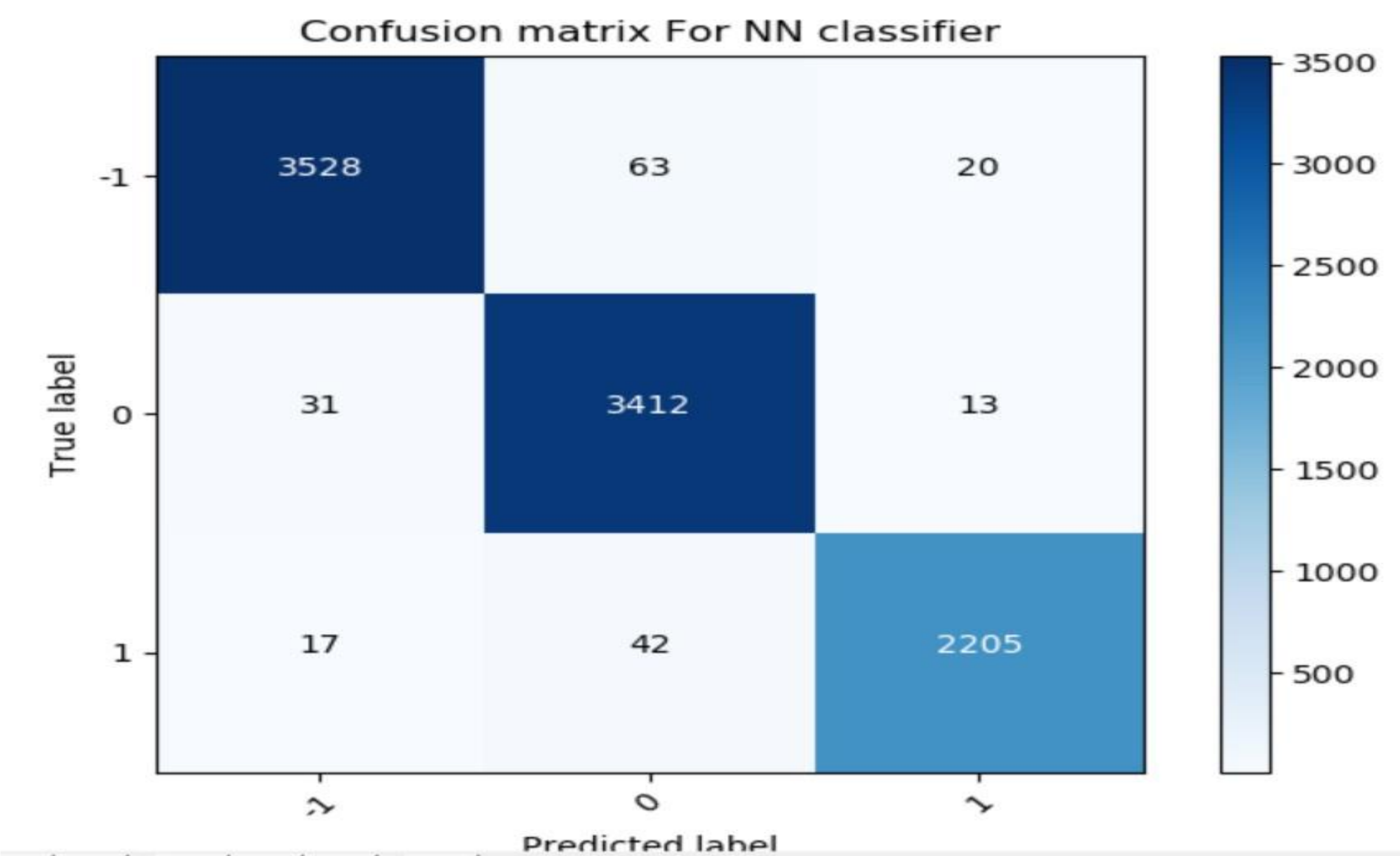
[14] Hemanth DJ, Anitha J, Son LH. Brain signal based human emotion analysis by circular back propagation and Deep Kohonen Neural Networks. Comput Electr Eng 2018;68:170–80.

[15] Feng H, Golshan HM, Mahoor MH. A wavelet-based approach to emotion classification using EDA signals. Expert Syst Appl 2018;112:77–86.

ISSN [Online]: 2583-2654

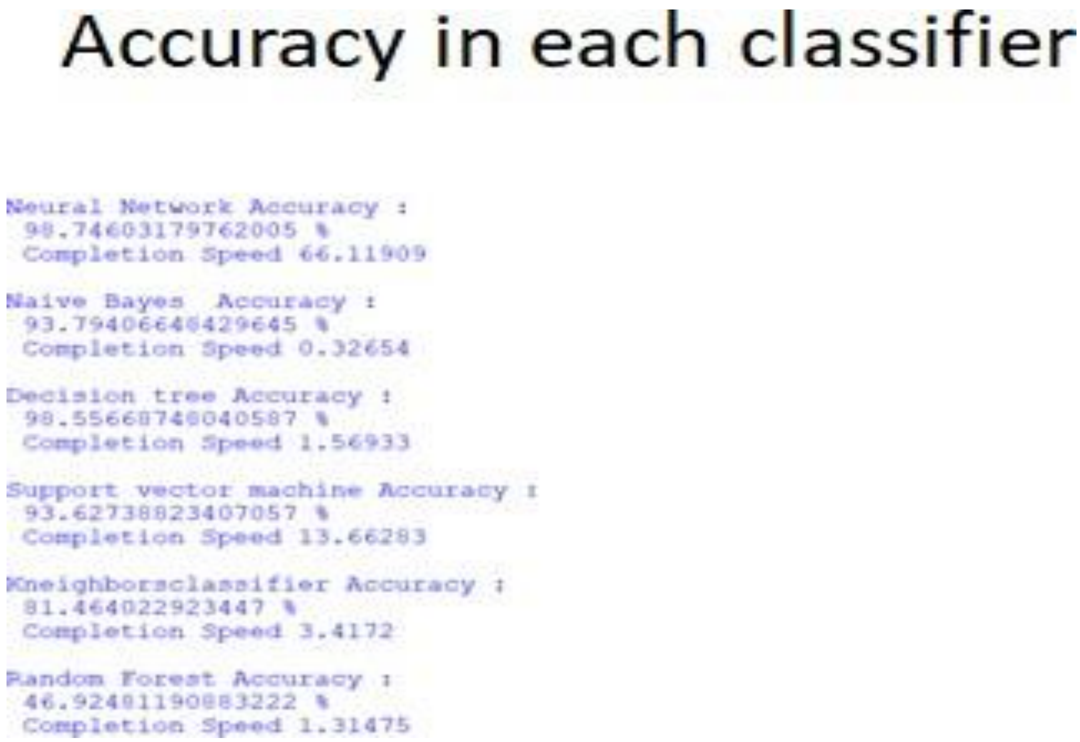
Snapshots



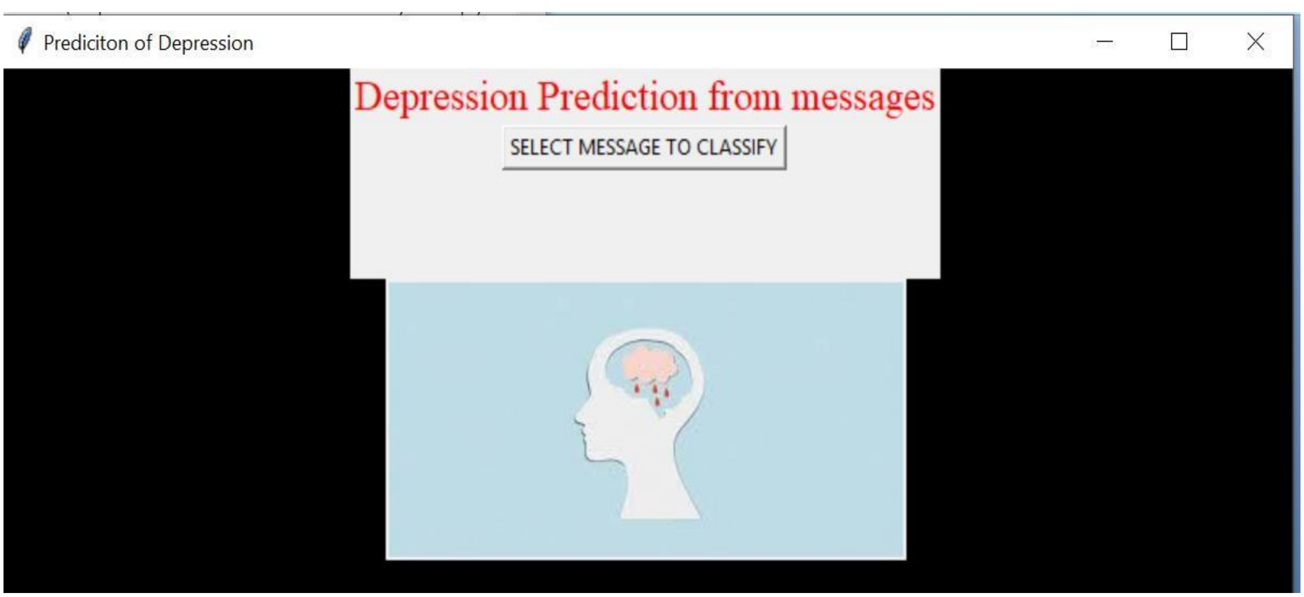


ISSN [Online]: 2583-2654

Accuracy is calculated for each classifier



Project GUI opens up



ISSN [Online]: 2583-2654

Input messages given to classification

```
Test - Notepad
File  Edit  Format  View  Help
I am feeling alone
Today i am happy
Sadness again stiking me
I am down
I don't know what to do with life
I am feeling very sad today
```

Each message is classified

```
The classificaiton result for  Today i am happy
 is

****************
Positive
****************
got result 1
The classificaiton result for  Sadness again stiking me
 is

****************
Negative
****************
got result 2
The classificaiton result for  I am down
 is

****************
Negative
****************
got result 2
The classificaiton result for  I don't know what to do with life
 is

****************
Neutral
****************
got result 0
The classificaiton result for  I am feeling very sad today is

****************
Negative
****************
```

ISSN [Online]: 2583-2654

Depression status is displayed