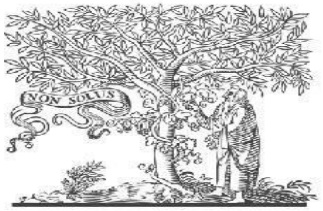
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 05th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/27

Title **ANALYSIS OF DEPRESSION TWEETS USING CONVOLUTION NEURAL NETWORKS**

Volume 12, ISSUE 04, Pages: 206-211

Paper Authors

Mr.N.Durga Rao, Shaik Ayaz Ahmed, Vallapureddy Naga Sandeep Reddy, Vankayalapati Ramesh, Regandla Dheeraj Living Stone

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Analysis of Depression Tweets Using Convolution Neural Networks

Mr.N.Durga Rao[1], Assistant professor, Department of ECE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

Shaik Ayaz Ahmed[2], **Vallapureddy Naga Sandeep Reddy[3]**, **Vankayalapati Ramesh[4]**, **Regandla Dheeraj Living Stone[5]**
[2,3,4,5] UG Students, Department of ECE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1] durgarao.467@gmail.com , ayazshaik9390@gmail.com
v.sandeepreddy143@gmail.com , dheerajlivingstone@gmail.com
vankayalapatiramesh08@gmail.com

ABSTRACT

This Paper aims to detect the depression in tweets using Machine Learning models which are taken from twitter, In particular by using a Convolutional Neural Network which is evaluated based on the various word embeddings. This well-trained Machine Learning model, if deployed on social media platforms like Facebook, Twitter and Instagram can help in detection of depression in tweets based upon the text used in posts by implementation of text-vectorization method.

Keywords: Detection, Depression, social media, Machine Learning, CNN, text-vectorization.

I.INTRODUCTION

According to the World Health Organization (WHO) survey, approximately 322 million people suffer from depression, with the number continuing to rise. So, as a part of the initiative, our aim is to detect the depression in its early stages which would stop people from committing suicide. Many people who are suffering from depression don't even try to consult a doctor as they feel it is too scary for society. Also, the people who are suffering from depression are mostly adults. The World Health Organization states that around 5% of adults worldwide are suffering from depression. Studies on mental health have been said to benefit from using social media since it offers an impartial collection of people's language usages and habits. Furthermore, details from Social media has the promise of enhancing conventional survey methods by offering more precise assessments of behaviour across time while significantly increasing population sample sizes. People do post about their depression and their remedies on social media, according to preliminary research. And multiple research have shown that it is feasible to utilise data mining and machine learning techniques to construct models to find

potential occurrences of depression on social media based on the symptoms and markers of depression. A CNN is a convolutional neural network made up of a network of artificial neurons that are linked and include biases and weights that can be learned. These neurons communicate with one another. The connections have adjusted numerical weights. Such that a properly trained network would react appropriately when provided with an image or pattern to identify throughout the training phase. Several layers of neurons that can recognise features make up the network. Many neurons in each layer react to various combinations of inputs from the layers below. As seen in Figure 1, The layers are constructed such that the input contains a number of primitive patterns that the first layer recognises, followed by patterns of patterns that the second layer recognises, and patterns of those patterns that the third layer recognizes. This paper proposes a solution based on convolutional neural networks that can detect the depression of a person based upon the text posts they make on social media platforms. The existing system uses CNN, text vectorization is included in this paper and reducing the number of layers used in the architecture. In this paper plan is to achieve an accuracy of 79%.

II.LITERATURE SURVEY

Depression Indications in Text Sequences Marcel Trotzek et al. IEEE [1] Prior research has indicated that language usage is impacted by depression, and that many sad people use social networking sites or the internet in general to acquire information or talk about their issues. This article discusses the early identification of depression utilising artificial intelligence models based on social media postings. In specifically, a classification based on user-level linguistic information is compared to a convolutional neural network based on various word embeddings. In a recent early detection test, it is demonstrated that an ensemble of both techniques may produce state-of-the-art outcomes. Also, the shortcomings of the widely used ERDE score as a criterion for early detection systems are highlighted when applied to shared activities.

A Machine Learning Approach to detect Depression and Anxiety using Supervised Learning Anamika Ahmed et al. [2] a model that determines the various degrees of these mental diseases using a common psychological evaluation and machine learning methods. The suggested approach employs supervised learning algorithms to forecast the degree of sadness and anxiety. It is utilised to determine how to score evaluate the patient's level of depression. Based on the scores of the responses, this machine learning-based algorithm categorised the severity levels of depression into four categories.

Detection of depression-Related posts in Reddit social Media Forum Michael

M.Tadewsse et al. [3] Examining Reddit user postings to find any indicators that can point to relevant online users' views about depression is the main goal of this essay. They used machine learning techniques and Natural Language Processing (NLP) methods to train the data for this purpose and assess the effectiveness of our suggested methodology. They establish a glossary of concepts in this approach that are increasingly typical in stories of depression. The outcomes demonstrate that their suggested approach can greatly raise performance accuracy. This model recognises sadness by using postings from the Reddit social media Forum as input. using text categorization and NLP approaches.

Depression Detection Based on Deep Distribution Learning Wheidima Carneiro de Melo et al. [4] This work offers a deep learning architecture that uses distribution learning to predict depression levels with accuracy. It makes use of a novel expectation loss function to estimate the distribution of the underlying data over depression levels, where anticipated values are used as a basis. Of the distribution are adjusted to get closer to the levels of the truth. Even in the presence of label ambiguity, the suggested technique may generate precise forecasts of depression levels. suggested techniques for precise assessment of a subject's depression degree based on verbal and nonverbal data Using the faces recorded in films, it may gauge the degree of depression. In this model, the deep distribution is utilised. Learning.

III.PROBLEM IDENTIFICATION

It is known that the main cause for suicide in the recent years is because of depression. Nonetheless, for a variety of reasons, many people with depression do not receive treatment. Past research has indicated that language usage is impacted by depression, and that many depressed people use social media sites or the internet in general to acquire information or communicate. Talk about their issues. Using a convolutional neural network is the current system (CNN). where text from social media networks is used to train the model.

IV.PROPOSED METHODOLOGY

This paper aims to detect the depression using data collected from twitter which is a public web source. This paper suggests the CNN model as an effective and scalable technique to examine the impact of depression identification. The early identification of depression using machine learning models based on social media platform posts is the focus of this suggested approach. In specifically, a classification based on user-level linguistic information is compared to a convolutional neural network based on various word embeddings. In a recent early detection test, it is demonstrated that an ensemble of both techniques may produce state-of-the-art outcomes. An additional component of the proposed

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

system is a convolutional neural network (CNN), which will be trained using postings from social media sites.1.6 million tweets specifically from the Twitter API.

This paper is following a sequential model to implement the CNN architecture, which contains various layers such as an input layer, a vectorization layer, an embedding matrix, a dropout layer, a pooling layer, a dropout layer, and a dense layer as shown in Figure 1. Layers are created by using Keras from TensorFlow.

Text vectorization takes place for the input text given, and the respective token values are assigned to them by referring to the stored values obtained during the training process. If no such word exists in the stored values, the model tries to allocate the latest value based upon the word combinations.

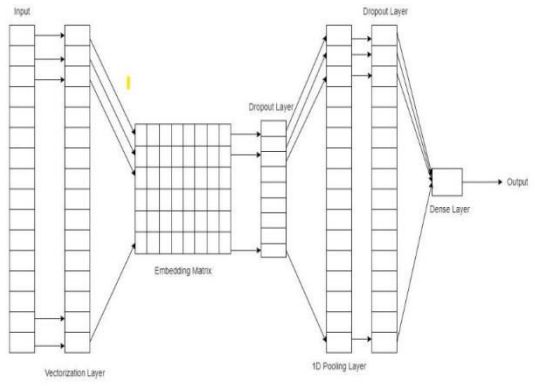


Figure 1: structure of CNN

V.IMPLEMENTATION

This paper has the following stages as shown in Figure 2, and each stage is explained in detail below.

A.DATA GATHERING

The dataset used in this paper is extracted from the website called Kaggle, which consists of 1.6 million tweets.

B.DATA PRE-PROCESSING

Data cleaning and splitting are two of the steps involved in data preparation. The data is the most important component of machine learning models, thus each of the operations that follow must be done properly.

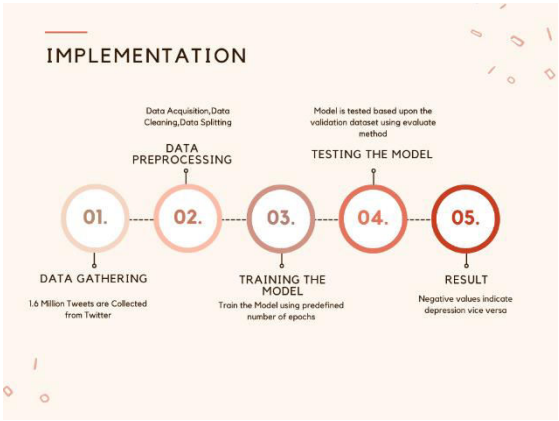


Figure 2: Process of Implementation

1.Data cleaning

Renaming the dataset and removing columns, among other steps, will help prepare the data for analysis using the single-variate, dual-variate, and multi-variate processes. To make data more valuable for analytics and decision-making, data cleaning's main objective is to find and remove mistakes and abnormalities.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

2.Data Splitting

Dataset splitting is essential since it lessens overfitting. The dataset is divided between training and testing groups in proportions of 80:20. The testing data is only used for testing, whereas the training data is used only for training.

Jupyter Notebook is used as our primary IDE; it runs on the Anaconda platform in a virtual environment. But while training the model, we moved to Google Colaboratory. And then the Model is trained repeatedly,

with different sample sizes and different epochs to find the optimized solution.

Testing is done with randomised, unique data to get accuracy and loss metrics. Testing is done using the inbuilt "evaluate" method in TensorFlow.

VI.RESULTS

Figure 3. illustrates results of this paper. It describes that, if the results obtained is negative, then the person who had written the tweet is said to be depressed and if the outcome is positive, it means the person is said to be not depressed.

This model achieved a 79% accuracy rate. Also, by giving the system additional time and a larger dataset to train on, we may attempt to increase accuracy.

```
In [21]: model.predict([["i am feeling alone"],["my thoughts always destroy my mood."]])
         1/1 [==============================] - 0s 128ms/step
Out[21]: array([[-1.9707662 ],
                [ 0.65351844]], dtype=float32)
```

Figure 3: sample Image of predicting output value

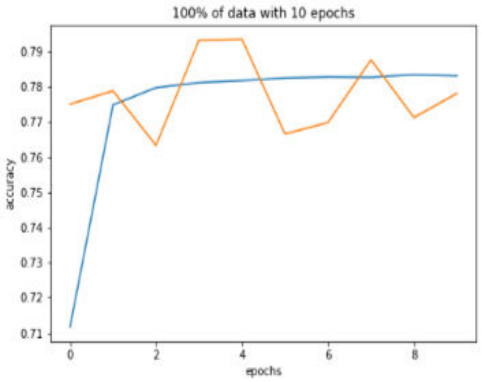


Figure 4: Training Accuracy statistics with 10 epochs & 100% data

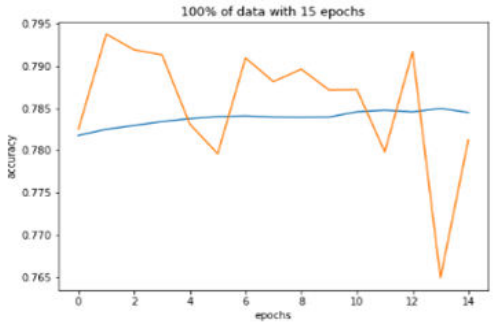


Figure 5: Training Accuracy Statistics with 15 epochs & 100% data

The training accuracy statistics are as shown in the below Figure 4 and Figure 5 when entire data is used in training with different epochs.

VII.CONCLUSION

Depression is one of the major reasons for suicides among youth in the recent years. But if we are successful in the detection of the depression, we can Help the people having depression from the other side. Our proposed system is to be developed using a convolutional neural network that

can detect the signs of depression by processing the text inside a post. This model can be plugged into the backend of a social media platform. Our model is having greater accuracy than the existing models and others.

VIII.REFERENCES

[1]. Marcel Trotzek, Sven Koitka, and Christoph M. Friedrich, Member, IEEE "Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences",2018

[2].AnamikaAhmed,RaihanSultana,MdTahmidur Rahman Ullas,MariyamBegom,Md. MuzahidulIslamRahi,Md. AshrafulAlam "A Machine Learning Approach to detect Depression and Anxiety using Supervised Learning",2020

[3]. Michael M.Tadesse , Hongfei Lin , Bo Xu , And Liang Yang. "Detection of Depression-Related Posts in Reddit Social Media Forum", 2019.

[4].Wheidima Carneiro De Melo,Eric Granger, AbdenourHadid. "Depression Detection Based on Deep Distribution Learning",2019

[5]. A. T. Beck and B. A. Alford, Depression: Causes and Treatment. 2nd ed. Philadelphia, PA, USA: Univ. Pennsylvania Press, 2009

[6]. Amanat, A.; Rizwan, M.; Javed, A.R.; Abdelhaq, M.; Alsaqour, R.; Pandya, S.; Uddin, M. Deep Learning for Depression Detection from Textual Data. Electronics 2022