

Schizophrenia Prediction using Multi-Class Classification Incorporating Neural Networks

Dr. M.S. Anbarasi¹ , Mr. Mohan.C² , Mr. Navaneetha Krishnan.P² and Mr. Yadhava Prasanna.G²

Assistant Professor¹, B.Tech Final Year Student²

Department of Information Technology,

Puducherry Technological University,

Puducherry, India

Abstract— Schizophrenia is a severe and chronic Mental Disorder. This Disorder is marked with disturbances in thoughts , perceptions and behaviours. Due to these disturbances the patient can trigger themselves and their friends/relatives moving close to them to commit suicide or attempt to do so. These Schizophrenia patient can not be discriminated from the normal person in the general population. Schizophrenia is also difficult to diagnose as there is no physical test till date to analyze it's symptoms since it is very similar to several other Mental Disorders like severe depression. Mostly of these people isolate themselves as they avoid socializing. So they spend their time in the social media to share their thoughts. Since they spend most of their time social media, most of their information can be collected from social media like Twitter. Twitter is a rich source for information from minute-to-minute trends to general discussions around topics. In the previous phase of the project, data have been Scattered from Twitter, cleaned and pre processed with Lexical Analysis and applied the Sentiment Analysis technique to identify Schizophrenia patient from normal person data using classification models . This project aims to analyse and predict the levels of depression using Multi-Class Classification incorporating Neural networks.

Keywords — *Sentiment Analysis, Schizophrenia Analysis , Multi-Class Classification , Neural networks.*

I. INTRODUCTION

Schizophrenia is a serious Mental Disorder in which people interpret reality abnormally. Schizophrenia may result in some combination of hallucinations, delusions, and extremely Disordered thinking and behavior that impairs daily functioning, and can be disabling. People with Schizophrenia require lifelong treatment. Early treatment may help get symptoms under control before serious complications develop and may help improve the long-term outlook.

Sentiment Analysis, which is also called opinion mining, uses social media analytics tools to determine attitudes toward a product or idea. Real-time Twitter trend Analysis is a great example of an analytics tool because the hashtag subscription Model enables you to listen to specific keywords (hashtags) and develop Sentiment Analysis of the feed. Twitter is the best site to carry Sentiment Analysis on written texts due to limitations in the length of posts. Sentiment Analysis is the automated process of identifying and classifying subjective information in text data. This might be an opinion, a judgment, or a feeling about a particular topic or product feature. The most common type of Sentiment Analysis is 'polarity detection' and involves classifying statements as Positive or Negative , which reflects their state of mind. Machine learning approaches are increasingly used for predicting Schizophrenia from social media posts. In Multi-Class Classification incorporating Neural networks, the data would be divided into sub-classes predicated on the polarities. This study aims to determine whether Multi-Class Classification incorporating Neural networks could be effectively used to predict signs of Schizophrenia in social media users by analyzing their social media texts.

II. LITERATURE SURVEY

Prasoon Gupta [1] applied collected data to eight supervised machine learning techniques with different grams of text after annotation and preprocessing. They have observed the best performance with the LinearSVC classifier and unigram. The combination gives us an accuracy of 84.4%, which is best in all the combinations which we have executed on our data set. They have consolidated the performance by calculating precision, recall, F1-Score, and tenfold cross-validation for all the combinations, and they got the best results with LinearSVC and unigram. So, they executed the sentiment analysis of tweets by the public during lockdown using this combination and found that almost half of the population (48.69%) is talking positive about the lockdown, 29.81% are neutral, and 21.5% of the people are feeling negative due to some reason.

Dr. M.S. Anbarasi¹ , Mr. Mohan.C² , Mr. Navaneetha Krishnan.P² and Mr. Yadhava Prasanna.G²

Vanlalawmpuia.R [2] highlighted and revealed the user's mental health status and condition by analyzing each user's data. Analyzing data involves training each user's data to get an output and also have a test set to get efficiency and accuracy. They have brought out many depressive indicative words which play an important role to bring this study success. We also proposed some few methods that have been carried out in this study. Analyzation for efficiency has been performed using this proposed method which containing packages of emotional words and can also increase the accuracy, efficiency and the scale of analyzation lapse.

Irvin Dongo [3] implemented a framework as an extension of Google Chrome consisting of a front-end and a back-end, which performs the credibility analysis in real-time and can be configured to use either web scraping or Twitter API to gather the needed data to feed the credibility model. The credibility model, previously proposed, computes post's credibility based on Text Credibility, User Credibility, and Social Credibility. Results show that a robust normalization process on the text obtained by the extraction methods, produces identical credibility results. Moreover, the number of followers obtained by the extraction methods have a minor difference for famous accounts since the number of followers is constantly growing. Additionally, web scraping is faster than Twitter API since for this latter, the use of web scraping is required to obtain the user_id and tweet_id before to perform the API request.

Moin Nadeem [4] established the feasibility of consistently detecting, identifying, and pursuing the diagnosis of individuals Twitter posts, henceforth referred to as 'tweets', Using solely these tweets, we aim to design and implement an automated computational classifier which may be able to parallel the performance and precision of a concerned human individual. The feasibility of this automated predictions will be crossvalidated and critiqued through standard Precision, Recall, and F1 scores, as well as Receiver Operating Classification curves.

III. PROPOSED SYSTEM

In this proposal social media data are collected and pre processed using lexical analysis. In pre processed data sentiment analysis technique is applied to discriminate between normal person and schizophrenia patient. Then Multi-Class Classification incorporating Neural networks is applied to find the level of Mental Disorder to segregate schizophrenia from lower level disorder patient. Data is collected by scraping Twitter by extracting relevant tweets using Twitter API for finding Schizophrenia and other Mental Health Disorder. The extracted data is converted to CSV data format. The obtained data is cleaned and pre processed using lexical analysis. Sentimental Analysis technique is applied to segregate schizophrenia patient data from cleaned and pre processed data set. Multi-Class Classification incorporating Neural networks technique is applied to find the best accuracy score. Based on accuracy score from Multi-Class Classification incorporating Neural networks technique, we predict whether the user is a schizophrenic patient.

A. Pre-processing and applying lexical analysis techniques for schizophrenia twitter data

For collecting the twitter data, we have to do some data mining process. With the help of twitter API, we have collected a large number of the dataset. The API variable is now our entry point for most of the operations we can perform with twitter. The API provides features to access different types of data. In this way, we can easily collect tweets data and store them in the system. The data pre-processing steps perform the necessary data pre-processing and cleaning on the collected dataset. We have applied an extensive set of pre-processing steps to decrease the size of the feature set to make it suitable for learning algorithms. The cleaning method is based on dictionary methods.

B. Polarity checking with sentimental analysis

After pre-processing pass the tweets fetched for each user to the trained model to predict the polarity. The fetched tweets are tokenized and padded to the same shape as embedding. Once the tweets are padded it is given to the trained model to get the polarity. The trained model is given one tweet at the time for prediction. If the value returned by the model is 0 then the polarity of the tweet is negative, else if 1 is returned then the polarity of tweet is positive.

C. Determining the accuracy score using multi-class classification

Based on the dataset we are using Multi-Class Classification and compare the performance of each model. We use a pre-categorized training dataset to train the model so that it is capable of categorizing any of the new data provided to it. It determines the accuracy score based on the classified tweets of the probability given to tweets belong to that particular class. Performance measures are used to assess how well these models perform in a given context. These performance metrics include accuracy score, precision score, recall score and F1 score. Because it helps us understand the strengths and limitations of these models when making predictions in new situations. Precision is the fraction of retrieved documents that are relevant to the query. Recall is the probability that a relevant document is retrieved by the query. F1 score is the harmonic mean of precision and recall. By using confusion matrix and classification report, we verify the model's accuracy.

D. Enhancement with the neural networks model

LSTM with CNN is built using Keras to determine whether social platform users are schizophrenia patient based on their Twitter posts. The model takes in an input and then outputs a single number representing the probability that the tweet indicates depression. The model takes in each input sentence, replace it with its embeddings, then run the new embedding vector through a convolutional layer. CNNs are excellent at learning spatial structure from data, the convolutional layer takes advantage of that and learn some structure from the sequential data then pass into a standard LSTM layer. Last but not least, the output of the LSTM layer is fed into a standard Dense model for prediction. The accuracy of the model is evaluated.

V. RESULT ANALYSIS AND DISCUSSION



Fig.1 Multi-Class Classification accuracy scores

By applying classification models to data set we have got accuracy values. Naive bayes gives 93.7% accuracy, Decision tree gives 98.5% accuracy, Support vector machine gives 93.6% accuracy, KNN gives 81.4% accuracy and Random forest gives 46.9% accuracy.

```
In [48]: labels_pred = model.predict(data_test)
labels_pred = np.round(labels_pred.flatten())
accuracy = accuracy_score(labels_test, labels_pred)
print("Accuracy: %.2f%%" % (accuracy*100))

Accuracy: 98.91%
```

Fig.2 Neural networks model

Neural networks model takes in each input sentence, replace it with it's embeddings, then run the new embedding vector through a convolutional layer. The convolutional layer learn some structure from the sequential data then pass into a standard LSTM layer. The output of the LSTM layer is fed into a standard Dense model for prediction.This model gives 98.91% accuracy.

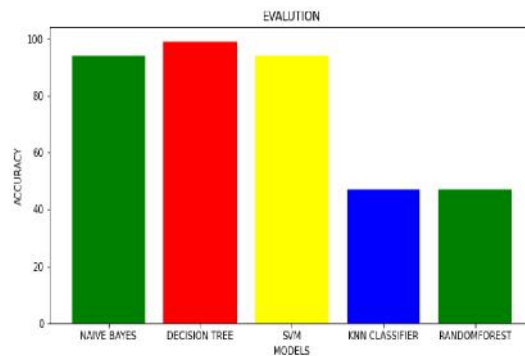


Fig. 3 Accuracy scores graph

The graph is plotted for accuracy scores of multi-class classification.

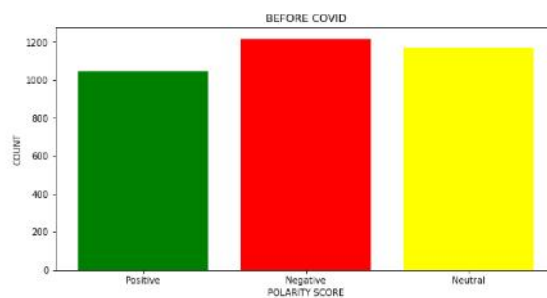


Fig.4 Before Covid Graph

Schizophrenia tweets before covid lockdown is represented in the graph.

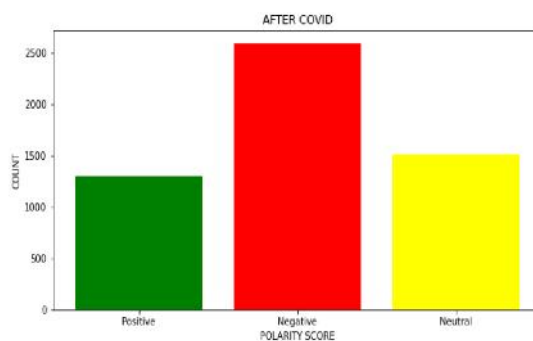


Fig.5 After Covid Graph

Schizophrenia tweets after covid lockdown is represented in the graph .

VI. CONCLUSION

Mental Health is extremely unpredictable in human community. Due to this impact one's thoughts behaviour and emotions can affect others surrounded them. Schizophrenia will be a major cause of major illness in the world due to their Mental Health to balanced their social life. Since there is no analytical tool to predict Schizophrenia, in this project schizophrenia prediction is done by using Multi-Class Classification incorporating Neural networks technique. For future enhancement various categories of schizophrenia can be predicted further.

ACKNOWLEDGMENT

We are deeply indebted to Dr. M.S.Anbarasi, Assistant Professor, Department of Information Technology, Puducherry Technological University, Puducherry, for her valuable guidance throughout the project work.

REFERENCES

1. R.Vanlalawmpuia,Mr.Lalhmingliana "Prediction depression in social network sites using data mining"Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020) IEEE Xplore Part Number:CFP20K74-ART; ISBN: 978-1-7281-4876-2.
2. Nirmal Varghese Babu and E. Grace Mary Kanaga "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review"SN Computer Science (2022) 3:74 <https://doi.org/10.1007/s42979-021-00958-1>.
3. Prasoona gupta,Sanjay kumar,R.R.Suman and S.Vinaykumar "Sentiment Analysis of lockdown in India during COVID-19:A case study on Twitter"IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, VOL. 8, NO. 4, AUGUST 2021.
4. T V Rampisela and Z Rustam "Classification of Schizophrenia data using support vector machine" IOP Conf. Series: Journal of Physics: Conf. Series 1108 (2018) 012044 doi :10.1088/1742-6596/1108/1/012044.
5. Enrique garcia ceja and Michael riegler "Motor activity based Classification of depression in unipolar and bipolar patients"2018 IEEE 31st International Symposium on Computer-Based Medical Systems.
6. Laszlo nemes and Attila kiss "Social media Sentiment Analysis based on COVID-19"JOURNAL OF INFORMATION AND TELECOMMUNICATION 2021, VOL. 5, NO.1.
7. Jung gu choi,Inhwan ko and Sanghoon han "Depression level Classification using machine learning classifiers based on actigraphy data"Digital Object Identifier 10.1109/ACCESS.2021.3105393.



Dr. M.S. Anbarasi

She received her M.E and Ph.D. degree from College of Engineering,Guindy Campus,Anna University. She is currently working as Assistant Professor in the Department of Information Technology. She has published several research papers in various refereed journals and international conferences. Her area of interest includes Data Mining, Big Data Analytics, Cloud computing and Software engineering.



Mohan .C

He is pursuing his B. Tech degree in the Department of Information Technology, Puducherry Technological University.



Navaneetha Krishnan.P

He is pursuing his B. Tech degree in the Department of Information Technology, Puducherry Technological University.



Yadhava Prasanna.G

He is pursuing his B. Tech degree in the Department of Information Technology, Puducherry Technological University.