A Study on Sentiment Analysis of Mental Illness Using Machine Learning Techniques

Pradeep Kumar Tiwari¹, Muskan Sharma³, Payal Garg², Tarun Jain¹, Vivek Kumar Verma¹ and Afzal Hussain³

¹Manipal University Jaipur, Dehmi Kalan, Off Jaipur-Ajmer Expressway, Jaipur, Rajasthan -303007 India
²G L Bajaj Institute of Technology and Management, Greater Noida, India
³Maulana Azad National Institute of Technology, Bhopal, Madhya Pradesh, India

E-mail: pradeeptiwari.mca@gmail.com

Abstract. In the digital age, social media plays a crucial role in society. Social media provides a platform to youth for exchanging their views on public issues and express their personal issues. Hence online media can be used for studying the behavior of people. Applying sentiment analysis on the data obtained timely from social networking sites (here Twitter), depression, anorexia, and other similar mental illness can be predicted among youth. The importance of detecting depression is that it is the root cause of a plethora of diseases. Early prediction can also mitigate the number of suicides. This work is to detect depression and PTSD (Post Traumatic Stress Disorder) among the Twitter users. Analysing the tweets, how likely a person is to suffer from any of the aforementioned diseases can be discovered.

1. Introduction

Sentiment Analysis or opinion mining or emotion AI is the process of extracting subjective information, opinions, and attributes from the text. It is a field within NLP (Natural Language Processing). It is a widely growing research area in computer science [1][2][3]. There are numerous applications of sentiment analysis.

- Study the reviews of products, movies, airlines, and hotels.
- Prediction of election results.
- Feedback on medicine and drugs.
- Analyze flaws in a product.
- Compare similar products.
- Analyzing mental health issues.

The main purpose of this work is to detect depression & PTSD among the Twitter users. Analysing the tweets, how likely a person is to suffer from any of the aforementioned diseases can be discovered. This data can be of great use for the doctors for treatment of patients. Also, this data can be used by the forensic experts to perceive if someone committed a suicidal or if someone is suicidal. Furthermore, this data can act
as an alert for the families of the affected people. Greater negative posts and lowered social activities are the key elements of such a person [4][5][6][7][8][9].

The functions that must be a part of the system to be a good sentiment analyser specifically are as follows

- The system is able to stream tweets and filter according to the required keywords.
- The system is able to process these tweets by different algorithms and mention their accuracy level.
- Also, it displays the confusion matrix for all the algorithms.
- It also gives the user a chance to enter his own sample tweet [1][2][3][4][10][11][12].

2. Literature Review

The author Patwa et al. (2020) presents the results of SemEval-2020 Task 9 on Sentiment Analysis of Code-Mixed Tweets (SentiMix 2020). We also publish and describe our Hinglish (Hindi-English) and Spanglish (Spanish-English) corpora annotated with word-level language identification and sentiment labels at the sentence level. These corpora, respectively, consist of 20 K and 19 K examples. The labels for sentiment are positive, negative, and neutral. SentiMix attracted a total of 89 applications including 61 teams that participated in the Hinglish contest and 28 systems submitted to the Spanglish contest. The best performance achieved was Hinglish F1 score of 75.0 percent and Spanglish F1 of 80.6 percent. We observe that the most common and successful approaches among participants are the BERT-like models and ensemble methods [1].

De Choudhary et al., (2013) achieved the goal of measuring depression through the use of a CES-D (Center for Epidemiological Studies Depression) scale, SMDI (Social Media Depression Index), PCA (Principal Component Analysis), and SVM (Support Vector Machine) classifier. While another method also proposed by same authors in 2014, PHQ-9 (Patient Health Questionnaire) and LIWC (Linguistic Inquiry and Word Count) was used to predict postpartum depression using Facebook data [2][3].

Socher et al. (2014) used deep recursive models to predict emotions. The methods used were RNTN (Recursive Neural Tensor Network) and MV-RNN (Matrix-Vector), of which RNTN could provide 80.7% accuracy [4].

Burmapet et al. (2015) dug deeper into the identification and classification of suicide-related tweets. An overall F-measure of 0.728 and especially 0.69 for suicide-related cases was obtained using TF-IDF (Term Frequency-Inverse Document Frequency), LIWC, PCA, SVM Classifier, rule-based, Naïve Bayes, J48 decision tree and random forest [5].

Braithwaite et al. (2016) used DSI-SS (Depressive Symptom Inventory-Suicide Subscale), ACSS (Acquired Suicide Scale) and (INQ) Interpersonal Needs Questionnaire, an updated version of LIWC, Scikit-learn library, Decision tree learning to predict the risks involved in suicides [6].

Saravia et al. (2016) analyzed and detected mental illness via social media that helped predict depression. The CES-D Scale, TF-IDF, Sentiment 140API, PLF, Random Forest Classifier were also used. Developed an online system that produced minimal results in future for efficient prediction of user behavior [7].

Kang et al. (2016) extracted depression tweets to identify depressive users using an SVM Classifier. A lexicon was also built using Visual Sense Ontology and Sent strength dictionaries, K-means clustering latent fusion and LIWC. Multimodal analysis provided more efficient results than existing methods [8].

Aldarwish et al. (2017) used BDI-II Questionnaire to predict the level of depression via social media posts. A depression model was also created using RapidMiner, Naïve Bayes as well as SVM classifiers [9].
Benton et al. (2017) predicted depression using Multitask Learning approach Feedforward multilayer perceptron Single Task Learning and Logistic Regression. The proposed model delivered better performance than Logistic regression models [10].

3. Proposed Work Implementation
Figure 1 depicts the architecture specification

3.1. Data Extraction

Login to Twitter with ID & password. Sign in in case of a new user.

Then, apply for Twitter’s developer’s access. On receiving the access, create an application. Clicking on the application icon, there is a section of keys.
With the help of these credentials & Twitter streamer, tweets can be extracted.
Save your credentials in file named ‘twitter_credentials’.
Define a function stream_tweets. The purpose of this function is to handle account authentication and the builds connection with Twitter Streaming API.
Also, call the filter function (defined later) in this function.
Next, make a class StdOutListener. This is a basic listener that just prints received tweets to stdout.
Authenticate and connect to Twitter Streaming API
Filter the tweets according to the following keywords, "stress", "depression", "upset", "dejected", "suicide", "suicidal", "trauma", "PTSD".
Store the filtered data in a .csv file.

3.2. Preprocessing
Import the libraries nltk, string, re, time & pandas.
Start the function getdata. Retrieve the .csv file to work on the dataset of tweets.
The next function processdata recovers the tweets from the .csv file. In this function the emoticons and punctuation marks are removed. Only the text & ID of the user is stored for future use.
Then read the dictionary and select the columns which have the word and polarity.
After the dictionary is prepared, the sentiment of words is calculated by comparing the dataset with the dictionary words.
After finding the sentiment calculate the polarity of the tweet
Save the ID with the corresponding polarity in another .csv file.

3.3. Testing and Training
Retrieve the tweets and the processed data obtained as a result of preprocessing in the last step.
Define a function for representing the confusion matrix for all the classifiers.
A total of five algorithms, Naïve Bayes, Decision Tree, Support Vector Machines, k nearest neighbours & Random Forest are used to predict.
Accuracy, time taken to complete & confusion matrix for each classifier is found.
The tweets with more negative thoughts are likely to suffer from depression, the ones having neutral polarity may or may not be prone to depression while the positive ones are quite unlikely to be depressed.

3.4. Sample Tweet
In addition to this, a sample tweet can be inputted by the user for testing.
The user’s tweet is analysed by the most accurate algorithm.

4. Result Analysis
The results obtained are as shown in figure 2 to 6 and in table 1:
4.1. Naive Bayes
Accuracy : 87.1326950098962 %
Completion Speed : 3.23342

![Confusion Matrix for a Naïve Bayes Classifier](image)

**Figure 2.** Confusion Matrix for a Naïve Bayes Classifier

4.2. Decision Tree
Accuracy : 92.80124721162923 %
Completion Speed : 33.46576

![Confusion Matrix for a Decision Tree Classifier](image)

**Figure 3.** Confusion Matrix for a Decision Tree Classifier
4.3. **Support Vector Machine**

Accuracy: 64.49122937202857%

Completion Speed: 807.33403

**Figure 4.** Confusion Matrix for a SVM Classifier

4.4. **K-n Neighbors Classifier**

Accuracy: 76.62388146884058%

Completion Speed: 138.56859

**Figure 5.** Confusion Matrix for a k-nearest Neighbours’ Classifier
4.5. Random Forest

Accuracy: 51.350182789334795%

Completion Speed: 9.13764

Observing the accuracy & confusion matrix it can clearly be inferred that Decision Tree gives the most accurate results.

Hence, the sample tweet is checked by Decision Tree algorithm.
5. Conclusion

With the obtained results it can be inferred that out of the five algorithms used, Decision Tree gives the most accurate results. However, this project is not yet perfect and can be extended for additional features. Stop words can be used for improved accuracy of the results obtained. Using lexicons would categories the tweet according to the frequency hence removing the words with lower frequency. N-grams and POS (Part of speech) tags can be used for better results. Here emoticons have been removed, however they play an important role in determining the thoughts of the user hence they should be analyzed too. Also, this project cannot detect sarcasm. Hence more research can be done to teach the analyzer to understand the real intention of the user’s tweet.

6. References

[1] Patwa, P., Aguilar, G., Kar, S., Pandey, S., PYKL, S., Gambäck, B., & Das, A. (2020). Semeval-2020 task 9: Overview of sentiment analysis of code-mixed tweets. arXiv preprint arXiv:2008.04277.

[2] De Choudhury, M., Counts, S., & Horvitz, E. (2013, May). Social media as a measurement tool of depression in populations. In Proceedings of the 5th Annual ACM Web Science Conference (pp. 47-56). ACM.

[3] De Choudhury, M., Counts, S., Horvitz, E. J., & Hoff, A. (2014, February). Characterizing and predicting postpartum depression from shared facebook data. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (pp. 626-638). ACM.

[4] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 1631-1642).

[5] Burnap, P., Colombo, W., & Scourfield, J. (2015, August). Machine classification and analysis of suicide-related communication on twitter. In Proceedings of the 26th ACM conference on hypertext & social media (pp. 75-84). ACM.

[6] Braithwaite, S. R., Giraud-Carrier, C., West, J., Barnes, M. D., & Hanson, C. L. (2016). Validating machine learning algorithms for Twitter data against established measures of suicidality. JMIR mental health, 3(2), e21.

[7] Saravia, E., Chang, C. H., De Lorenzo, R. J., & Chen, Y. S. (2016, August). MIDAS: Mental illness detection and analysis via social media. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 1418-1421). IEEE.

[8] Kang, K., Yoon, C., & Kim, E. Y. (2016, January). Identifying depressive users in Twitter using multimodal analysis. In 2016 International Conference on Big Data and Smart Computing (BigComp) (pp. 231-238). IEEE.

[9] Aldarwish, M. M., & Ahmad, H. F. (2017, March). Predicting depression levels using social media posts. In 2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS) (pp. 277-280). IEEE.

[10] Benton, A., Mitchell, M., & Hovy, D. (2017). Multi-task learning for mental health using social media text. arXiv preprint arXiv:1712.03538.

[11] Suhara, Y., Xu, Y., & Pentland, A. S. (2017, April). Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In Proceedings of the 26th International Conference on World Wide Web (pp. 715-724). International World Wide Web Conferences Steering Committee.
[12] Khan, A., Husain, M. S., & Khan, A. (2018). Analysis of Mental State of Users using Social Media to predict Depression! A Survey. *International Journal of Advanced Research in Computer Science*, 9(Special Issue 2), 100.