Detection of Depression Tendency Using Hard Voting Classifier With Feature Engineering

Radha Kori[1], Prof. Anshul Khurana[2]
[1]Dept of CTA
[2]Professor, Dept of CSE
[1, 2] Shri Ram Institute of Technology Jabalpur, Madhya Pradesh, India

Abstract- *Depression is ranked as the largest contributor for suicides and global disability. According to WHO more than 300 million people worldwide are suffering from depression. While depression may lead to social isolation and withdrawal. Researchers found that social media platforms are increasingly used by affected individuals. People suffering from depression share their experiences and support each other. Studies show that peer-to-peer social Medias increase the likelihood to seek professional help. Studies have shown that depression has also effect on language usage. Many individuals use social media platforms for expressing their views and sharing their problems. This gives us opportunity to work on early depression detection using Machine Learning methods. Our research focuses on implementing ML and Deep learning models to detect depression from social media like Twitter.*

Keywords- Depression and Language, Deep Learning, Machine Learning, Hard Classifier, Twitter, Feature Extraction.

I. INTRODUCTION

Precisely defining depression is not an easy task, not only because several sub-types have been described and changed in the past [1][2], 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 [3]. 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 [4]. 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 2018 [5].

Although the severity of depression is well-known, only about half of the individuals affected by any mental disorder in Europe get treated [6]. The proportion of individuals seeking treatment for mood disorders during the first year ranges between 29–52% in Europe, 35% in the USA, and only 6% in Nigeria or China [7]. In addition to possible personal reasons for avoiding treatment, this is often due to a limited availability of mental health care, for example in conflict regions [8]. Via a telephone survey in Germany [9], 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. Another study amongst people with severe mental illness in Washington D.C. showed that stigma and discrimination indeed exist, while they are not “commonly experienced problems” but rather “perceived as omnipresent potential problems”
[10].

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 [11], [12]. 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 [13]. A similar study in the USA [14] 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.

People frequently express their depression over social media which can be analyzed to identify the causes of their depression. Depressed people become more vulnerable over time and can commit any type of crimes, including suicide and killing others. As social media are becoming more and more popular, people, including the depressed individuals, are interacting with social media increasingly. If their social media interactions can be analyzed, proper steps can be taken in time. Research community has been trying to analyze social media data to discover sentiments of human beings. Depression analysis is the process of identifying whether people are going through depression from their textual activity on social media. Identification of depression from social media has been framed as a classification problem which is in the domain of Natural Language Processing (NLP). In this work we study NLP approaches that can successfully extract information from textual data to enhance identification of depression. These NLP approaches perform different feature extraction to build document representations.

1.2Early depression detection

Even though multiple studies have attempted to predict or analyze depression using machine learning techniques, before Losada and Crestani [4] , no one had attempted to build a public dataset in which a large chronological collection of writings, leading to this disorder, were made available to the research community. This is mainly due to the fact that text is often extracted from social media sites, such as Twitter or Facebook, that do not allow redistribution. On the other hand, in the machine learning community, it is well known the importance of having publicly available datasets to foster research on a particular topic, in this case, predicting depression based on language use. That was the reason why the main goal in Losada and Crestani was to provide, to the best of our knowledge, the first public collection to study the relationship between depression and language usage by means of machine learning techniques. This work was important for ADD, not only for creating this publicly-available dataset for EDD experimentation but also because they proposed a measure (ERDE) that simultaneously evaluates the accuracy of the classifiers and the delay in making a prediction. It is worth mentioning that having a single measure combining these two aspects enabled this dataset to be used as a benchmark task in which different studies can be compared in terms of how “early-and-accurate” their models are. Both tools, the dataset and the evaluation measure, were later used in the first pilot task of eRisk in which 8 different research groups submitted a total of 30 contributions.

Most of the existing psychological researches are based on questionnaires and academic interviews and their participants are often limited to small geographical regions. Technical details of depression analyses using Machine Learning and Deep Learning approaches for textual data are very rare as the researchers started to employ these techniques recently. Our purpose is to find words used in tweets that can help predict if a user is expressing negative sentiment or experiencing depression. By analysing these tweets, we can form a predictive model that can be applied to detect similar sentiment in other tweets. Using the results of our experiments, this will give us a deeper understanding of the relationship between the semantic used and the mental status of the user.

Twitter data can be quite large and to analyse the entire dataset can introduce a measure of complexity that requires more sophisticated hardware and processing methods which some other experiments have relied on. Instead of analysing a large dataset or focussing on a smaller dataset, our method here is to first identify users which had expressed highly negative sentiments based on a set criteria and then applying further processing to obtain the words that of use to us. This trimming of the data to be processed allows us to not only focus on the more negative users but also greatly simplifies its processing.

We looked at both sentiment and stress analyses whereas some research have focused on a single aspect of a person’s tweets. Both sentiment and stress are related which gives us a complementary perspective of a user’s mental state and in combination allows us to see this as a longer term pattern of behaviour. As people are not always negative or stressful, our analysis also looked at the positive sentiment and relaxation scores as this gives us a good insight of the overall behaviour of the user over a period of time.

II. RELATED WORK

Depression is one of the leading causes of suicide worldwide. However, a large percentage of cases of depression go undiagnosed and, thus, untreated. Previous studies have found that messages posted by individuals with major depressive disorder on social media platforms can be analysed to predict if they are suffering, or likely to suffer, from depression. This study aims to determine whether machine learning could be effectively used to detect signs of depression in social media users by analysing their social media posts—especially when those messages do not explicitly contain specific keywords such as ‘depression’ or ‘diagnosis’. To this end, we investigate several text preprocessing and textual-based featuring methods along with machine learning classifiers, including single and ensemble models, to propose a generalised approach for depression detection using social media texts. We first use two public, labelled Twitter datasets to train and test the machine learning models, and then another three non-Twitter depression-class only datasets (sourced from Facebook, Reddit, and an electronic diary) to test the performance of our trained models in other social media sources. Experimental results indicate that the proposed approach is able to effectively detect depression via social media texts even when the training datasets do not contain specific keywords (such as ‘depression’ and ‘diagnose’), as well as when unrelated datasets are used for testing. Figure 2.1 below shows the steps of Machine Learning for classifying depression and no depression tweets.

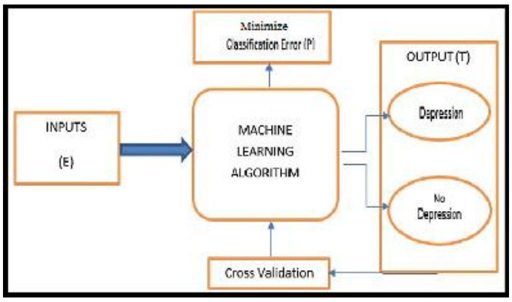

Figure 2.1: Machine Learning and Depression.

The use of sentiment analysis and machine learning on social media posts for detecting mental health problems is present in current research. This subsection shows some of the previous works on that matter.

Author’s in [15] proposed a model for classifying Facebook comments as depressive indicative or not, using features that were divided into three different categories: emotional variables, temporal categories, and standard linguistic dimensions. These features were then fed to different KNN algorithms, both individually and combined between each other. The best F1 measure was obtained with the Coarse KNN algorithm, using the emotional variables with a value of 0.71.

Author’s in [16] extracted 37 sentiment categories using the tool LIWC and examined how the variables were correlated with a user Center for Epidemiological Studies–Depression (CESD) score. CES-D is a survey that asks the frequency of depression-related symptoms that the patient has suffered over the past week. The final score ranges from 0 to 60, with higher scores indicating more severe depression symptoms [16]. A total of 18 sentiment predictors were found out to be reasonably correlated with the topic. The predictor, example words for each factor, and the coefficients of a multiple regression model for the CES-D score are detailed.

Numerous studies on automatic detection of the symptoms of depression have been carried out using artificial intelligence methods such as ML. A major stream of research involves depression detection using medical data, such as fMRI signature [17], results of depression questionnaires (such as DASS21 and DASS42) [18], or clinical criteria for depression as defined in DSM-5 and ICD-10 [19]. Data from clinical interviews, using systems such as Distress Analysis Interview Corpus-Wizard of Oz (DAICWOZ [20]) [21], has also been collected. Data from DAIC-WOZ includes videos, speeches, and text transcriptions of the participants, who could be either distressed or non-distressed. In recent years, researchers have also focused on depression detection using text messages from social media platforms, such as Twitter, Facebook, Reddit, and WeChat [22] in the hope that social media texts can help detect depression even when the individual is unaware of their depression or is in denial.

The majority of research studies on depression detection using social media messages usually follow either a textual-based featuring approach or a person descriptive-based featuring approach. Textual based featuring focuses on the linguistic features of the social media text, such as words, POS, n-gram, and other linguistic characteristics [23]. In contrast, the descriptive-based featuring approach focuses on descriptions of the subject, such as age, gender, employment status, income, consumption of drugs or alcohol, smoking, and other details of the subject or patient [24]. These features are then input into the detection models. Most models for depression detection have been developed using ML classifiers, such as the Support Vector Machine (SVM), Multilayer Perceptron (MLP), Logistic Regression (LR), Decision Tree (DT), Naïve Bayes (NB), Maximum Entropy (ME), K-Nearest Neighbours (KNN), Adaptive Boosting (AB), Random Forest (RF), Gradient Boosting (GB), Bagging Predictors (BP), and other single and ensemble models. Deep learning methods, such as the Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) [25], have also been used. Additionally, several studies have constructed custom detectors.

In research on depression detection using social media texts, custom datasets are frequently created but not made publicly available. In contrast, in this study, we utilise five publicly available datasets. These include two binary-class Twitter datasets that are used for both training and testing, and three single-class datasets from Facebook, Reddit and an electronic diary for further testing. While our goal is a generalised approach for depression detection in social media texts, the datasets include only text messages and exclude any emoticons, emojis, pictures, videos and web links that are commonly part of social media messages. Additionally, we address the issue of overfitting that generally arises when collecting depression data from social media messages. A model might perform poorly on datasets it was not trained on due to overfitting. We also focus our efforts on overcoming the problem of imbalanced data samples, which can negatively impact the performance of classifier models.

III. PROPOSED WORK

We initially Social media analysis has shown promising results for public health assessment and monitoring. In this research, we explore the task of automatically analyzing 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, DT, and KNN). Proposed system contains data collection, pre-processing, feature extraction, training, evaluation and testing. Proposed system is represented in figure below:

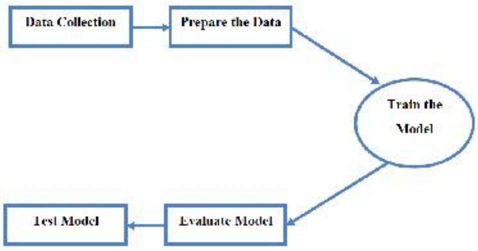

Figure 3.1: Proposed model.

3.2 Data Collection

Dataset

- Combining 10,314 tweets from Sentiment140
- Originally has over 1.6 million tweets
- For this implementation: 8,000 positive/neutral tweets & 2,314 depressive tweets.
- Labels: 0 (not depressive) and 1(depressive)

Dataset distribution is shown below:

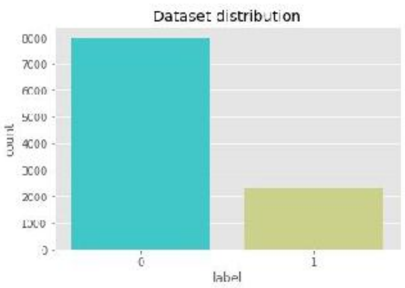

Figure 3.2: Dataset distributions

3.3 Data Pre-processing

The raw dataset procured contained tweets with urls, hashtags, user handles and stopword 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.

3.4 Feature Extraction

It includes:

- 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.

3.5 Proposed Learning Model

Proposed machine learning model is an ensemble model that combines Naïve Bayes, Decision Tree and K-

Nearest Neighbor method to provide final prediction by majority voting.

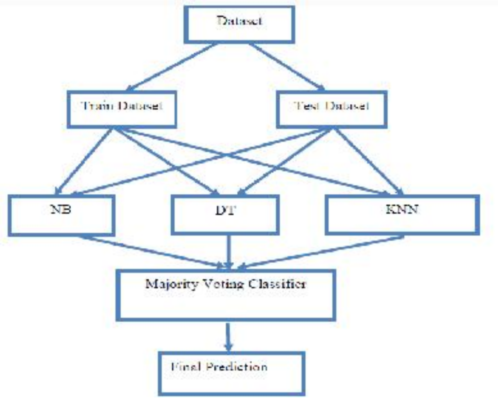

Figure 3.5: Proposed model.

IV. RESULTS WORK

We began Steps of proposed model:

1. Import the required libraries.
2. Distribute the dataset into depressive and non-depressive.

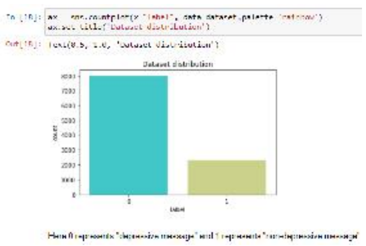

Figure 4.2: Depressive and znon-Depressive.

3. Clean the dataset.
 a. remove the urls.
 b. remove the hashtag.
 c. remove emojis.
 d. convert to lowercase.
 e. apply stemming to the tokens.
3. Finally get a cleaned dataset.
4. Create a WordCloud of depressive data.
5. WordCloud of non-depressive data.
6. Create the model using all basic classifiers like Naive Bayes = ‘NB’, Decision Tree = ’DTree’, Random Forest = ‘RF’, K Nearest Neighbours = ‘KNN’, SVM = ’SVM’ Kernal SVM = 'KSVM'.

In the table below, the results are analyzed for the existing method and proposed method. The results are analyzed by calculating accuracy of the model.

Random Forest 91.77
CNN with hybrid Model 97.04
Proposed Method 97.89

It is observed that proposed classifier achieved best accuracy due to the past tweet behavior features for which string patterns are created.

V. CONCLUSION

In this thesis, we proposed an algorithm to identify which Twitter accounts belong to bots and which Twitter accounts belong to humans. To do this, we also need to prepare a dataset that better represents the bot population on Twitter - incorporating bots from each of the three tiers. The two main contributions from this thesis are:

- The first contribution is the bot/human classification algorithm was developed.
- The second contribution from this thesis is the analysis and rationale for the features we use.
- Extending this point, we also contribute some possible features like string generation from tweets and re-tweet as well as reply of tweets, that we suspect will increase accuracy and better handle the bot population as a whole.
- All things considered, our algorithm’s performance meets the expectations we had at the set out of this thesis.

REFERENCES

[1] Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization, 2017.

[2] A. T. Beck and B. A. Alford, Depression: Causes and Treatment. Second Edition. University of Pennsylvania Press, 2009.

[3] Key Substance Use and Mental Health Indicators in the United States: Results from the 2016National Survey on Drug Use and Health. Rockville, MD: Center for Behavioral Health Statistics and Quality: Substance Abuse and Mental Health Services Administration, 2017.

[4] E. S. Paykel, “Basic concepts of depression,” Dialogues in Clinical Neuroscience, vol. 10, no. 3, pp. 279–289, 2008.

[5] “Mental health”, World Health Organization. [Online]. Available:http://www.who.int/mental_health/management /depression/en/ [Accessed: 26-Dec-2018].

[6] J. Alonso, M. Codony, V. Kovess, M. C. Angermeyer, S. J. Katz, J. M. Haro, G. De Girolamo, R. De Graaf, K. Demyttenaere, G. Vilagut et al., “Population level of

unmet need for mental healthcare in Europe," The British Journal of Psychiatry, vol. 190, no. 4, pp. 299-306, 2007.

[7] P. S. Wang, M. Angermeyer, G. Borges, R. Bruffaerts, W. T. Chiu, G. De Girolamo, J.Fayyad, O. Gureje, J. M. Haro, Y. Huang et al., "Delay and failure in treatment seeking after first onset of mental disorders in the world health organization's world mental health surveyinitiative," World Psychiatry, vol. 6, no. 3, p. 177, 2007.

[8] A. Rahman, S. U. Hamdani, N. R. Awan, R. A. Bryant, K. S. Dawson, M. F. Khan, M. M.-U.-H. Azeemi, P. Akhtar, H. Nazir, A. Chiumento et al., "Effect of a multi-component behavioral intervention in adults impaired by psychological distress in a conflict affected area of Pakistan: A randomized clinical trial," JAMA, vol. 316, no. 24, pp. 2609–2617, 2016.

[9] G. Schomerus, H. Matschinger, and M. C. Angermeyer, "The stigma of psychiatric treatment and help-seeking intentions for depression," European Archives of Psychiatry and Clinical Neuroscience, vol. 259, no. 5, pp. 298–306, 2009.

[10] R. Whitley and R. D. Campbell, "Stigma, agency and recovery amongst people with severe mental illness," Social Science & Medicine, vol. 107, pp. 1 – 8, 2014.

[11] K. Gowen, M. Deschaine, D. Gruttadara, and D. Markey, "Young adults with mental health conditions and social networking websites: Seeking tools to build community." Psychiatric Rehabilitation Journal, vol. 35, no. 3, pp. 245–250, 2012.

[12] J. A. Naslund, S. W. Grande, K. A. Aschbrenner, and G. Elwyn, "Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube," PLOS one, vol. 9, no. 10, p. e110171, 2014.

[13] J. A. Naslund, K. A. Aschbrenner, L. A. Marsch, and S. J. Bartels, "The future of mental health care: peer-to-peer support and social media," Epidemiology and Psychiatric Sciences, vol. 25, no. 2, pp. 113–122, 2016.

[14] M. Berger, T. H. Wagner, and L. C. Baker, "Internet use and stigmatized illness," Social Science & Medicine, vol. 61, no. 8, pp. 1821–1827, 2005.

[15] Islam, M.R.; Kamal, A.R.M.; Sultana, N.; Islam, R.; Moni, M.A.; Ulhaq, A. Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique. In Proceedings of the 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 8–9 February 2018; pp. 1–4.

[16] Park, M.; Cha, C.; Cha, M. Depressive moods of users portrayed in Twitter. In Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics, Beijing, China, 12 August 2012; pp. 1–8.

[17] Sato JR, Moll J, Green S, Deakin JF, Thomaz CE, Zahn R (2015) Machine learning algorithm accurately detects fMRI signature of vulnerability to major depression. Psychiatry Res 233 (2):289-291. doi:10.1016/j.pscychresns.2015.07.001.

[18] Kumar P, Garg S, Garg A (2020) Assessment of anxiety, depression and stress using machine learning models. Procedia Computer Science 171:1989-1998. Doi: https://doi.org/10.1016/j.procs.2020.04.213.

[19] Ojeme B, Mbogho A (2016) Selecting Learning Algorithms for Simultaneous Identification of Depression and Comorbid Disorders. Procedia Computer Science 96:1294-1303. doi:10.1016/j.procs.2016.08.174.

[20] Gratch J, Artstein R, Lucas G, Stratou G, Scherer S, Nazarian A, Wood R, Boberg J, DeVault D, Marsella S, Traum D, Rizzo S, Morency L-P The Distress Analysis Interview Corpus of human and computer interviews. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, and May 26-31, 2014. European Language Resources Association (ELRA), pp 3123-3128.

[21] Samareh A, Jin Y, Wang Z, Chang X, Huang S (2018) Detect depression from communication: How computer vision, signal processing, and sentiment analysis join forces.
IISE Transactions on Healthcare Systems Engineering 8 (3):196-208. doi:10.1080/24725579.2018.1496494.

[22] Chen Y, Zhou B, Zhang W, Gong W, Sun G Sentiment Analysis Based on Deep Learning and Its Application in Screening for Perinatal Depression. In: Proceedings of 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), 2018. pp 451-456. doi:10.1109/dsc.2018.00073.

[23] Chiong R, Budhi GS, Dhakal S (2021) Combining sentiment lexicons and content-based features for depression detection. IEEE Intelligent Systems 36 (6).

[24] Filho EMS, Veiga Rey HC, Frajtag RM, Arrowsmith Cook DM, Dalbonio de Carvalho LN, Pinho Ribeiro AL, Amaral J (2021) Can machine learning be useful as a screening tool for depression in primary care? J.Psychiatr Res 132:1-6.doi:10.1016/j.jpsychires.2020.09.025.

[25] Lin C, Hu P, Su H, Li S, Mei J, Zhou J, Leung H (2020) Sense-Mood: Depression Detection on Social Media. In: Proceedings of the 2020 International Conference on Multimedia Retrieval. Association for Computing Machinery, pp 407–411. doi:10.1145/3372278.3391932.