ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2 April 2024

MENTAL HEALTH TRACKING USING MACHINE LEARNING

[1]G. Geetha Devi, [2]Siri Boda, [3]Bommareddy Satya Priya reddy

Assistant professor in Department of information Technology Bhoj Reddy engineering College for women

UG Scholars in Department of information Technology Bhoj Reddy engineering College for women

Abstract

Depression has turned into a difficult issue in this ongoing age and the quantity of individuals impacted by melancholy is expanding step by step. In any case, some of them figure out how to recognize that they are confronting discouragement while some of them don't have any acquaintance with it. Then again, the tremendous advancement of web-based entertainment is turning into their "journal" to share their perspective. A few sorts of examination had been led to identify gloom through the client post via web-based entertainment utilizing AI calculations. Through the information accessible via virtual entertainment, the specialist can ready to know regardless of whether the clients are confronting misery. AI calculation empowers to arrange the information into right gatherings and distinguish the burdensome and non-burdensome information. The proposed research work expects to identify the downturn of the client by their information, which is shared via virtual entertainment. The Twitter information is then taken care of into two unique kinds of classifiers, which are Gullible Bayes and a half and half model, NBTree. The outcomes will be looked at in light of the greatest exactness worth to decide the best calculation to distinguish sadness. The outcomes show the two calculations perform similarly by demonstrating same exactness level.

I INTRODUCTION

Depression has become a serious problem and affecting mental health. Depression is also considered an important illness that affected more than 264 million people worldwide. There are many possibilities such as sudden changes of surroundings, changes in the brain's neurotransmitter level due to some feeling attacks, or even genetic features. Depression can be treated either by having a therapy session or medication. Even though many patients are undergoing the treatments there are still a high number of people who are undiagnosed cases of depression due to lack of knowledge of depression. Consequently, patients will suffer from bad thoughts such as self-isolation, wild behavior, suicidal thoughts, and dependent on medicines such as antidepressant pills. Depression can slow down the daily activity performances due to a lack of focus and interest. Eventually, it can harm the body and brain of the individual if depression is not detected.

Due to the advancement of technology, the number of social media users covers approximately 3.8 billion people. Evolvement and trendsetting of social media among people embrace the users to share their moods, thoughts, and opinion instantly. Social media has become a platform to express their feelings and emotion in social media which makes them satisfy themselves to share thoughts which keeps unsettling in their minds.

Social media application provides a space for their users to share or write about their opinion and random note which related to their feelings and emotions in words. The text shared on social media contains valuable insight can be used for various intelligent applications in the real world such as healthcare, entertainment; politics view communication, and tourism. The content or points created by the user is the data that is valuable for the researchers to analyze the state of mind. Along with the growth of social media applications, people start to share their mental health battles with the world via social media instead of keeping it private. The data which is available in the user account can be analyzed to figure out their level of mental health and an opportunity to help them to recover

Twitter is more known as a micro-blogging application that allows posting short stories not exceed 140 characters with simple interfaces. Almost every user's tweets are completely public and pullable by using their own Twitter API. Twitter's API allows doing complex queries like pulling every tweet about a certain topic .Twitter enables us to understand and supports more than 35 languages around the world. Twitter is a platform for analyzing and understanding trends and events happening around the world either on a scale of global or local. Twitter can extract value on a large scale with the help of Twitter API. As a result, researchers able to detect someone depressed or not based on the tweets which are available in public. Sentiment analysis is the process to identify the tone of the text is positive, negative, or neutral. A sentiment analysis system will assign a score for every text word based on the designed polarity. This can allow identifying the state of the user either they are in a positive mood or negative mood. The sentiment analysis technique is applied to each tweet to identify the sentiment score and labeled them as positive, negative, or neutral. Sentiment analysis in Twitter data quantifies the status of a tweet or comment by the user by calculating the sentiment scores. After the sentiment score on tweets via sentiment analysis technique, the machine learning algorithm enables to classify the tweets either it is depressive or not depressive based on the sentiment score labels on tweets. The researchers implement a machine-learning algorithm to detect depression based on the tweets data which was labeled with sentiment scores. The efficiency of detection is

evaluated based on the accuracy of the machine learning algorithm.

II LITERATURE SURVEY

Depression is considered a mental health problem that becomes a common issue to discuss in everyday health issues. Ignoring the symptom of depression or untreated of depression lead to severe problems which can risks own live Depression occurs in the beginning stages start from complicated interaction between social, biological, and psychological factors. Big and complicated problems may develop the fact of depression.

There are 7 types of depression with two main headers which are clinical depression and bipolar disorder.

The depressive disorder will affect approximately 2 weeks and during the period the patient will feel

1. Feelings of guilt, worthlessness, helplessness
2. Less food consumption and weight loss or more food consumption and weight gain
3. Suicide trials and thought of death
4. Insomnia, early-morning awakening, or excessive sleeping
5. Consistent sad mood almost every day
6. Loss of interest and favorites in hobbies and activities

Manic depression is driven by moods that swing from severe highs (mania) or mild highs (hypomania) to severe lows (depression). The mood swings hold from days to weeks or longer and may be melodramatic. The individual can be either too energetic or too depressive. They will be excessively ecstatic, less sleep, talkativeness, racing thoughts, hyper energy, poor in judgment, and inappropriate social behavior in the situation if the individual is in a manic state.

Machine learning application on detect depression via social media:

Machine learning helps to identify interesting patterns and knowledge based on exploration in the dataset. Previous researchers have done depression analysis based on Facebook data which is available in the public domain. The researchers carried out the research based on the emotional and linguistic style of the word usage. Researchers also conducted the classification by using the SVM algorithm with different kernels and it shows that the algorithm outperforms with better accuracy.

Another researcher named Nadeem in 2016 has experimented on detects Major Depressive Disorder via Twitter data by using Naïve Bayes and SVM algorithm. The ending results show that Naïve Bayes outperforms SVM

Hybrid model machine learning is also implemented to detect depression via Twitter data. The naïve Bayes-SVM hybrid model

shows great accuracy for the sentiment analysis task.

Most of the researchers applied a single machine-learning algorithm to detect depression. Moreover, the researchers are also focusing on the accuracy value to determine which algorithm suits the problem. The researchers also applied a single one-size dataset for every algorithm. Overall, it shows that Naïve Bayes is outperformed in most of the research experiments with great accuracy. However, the limitation of each paper as listed above is the drawback and key to emphasize the algorithm for better performance.

III EXISTING SYSTEM

Previous work has found that automatic analysis of depression symptoms from texts can be applied in, for example, sentiment retrieval from suicide notes and detecting insulting or depressive words or sentences in conversations or blog posts [13–18]. However, there is still substantial untapped potential in research on extracting depressive symptoms from texts. Key challenges include portraying significant cues of depression from texts. Also, there is a substantial hurdle in detecting depression symptoms from short texts. To contribute towards solving these challenges, we aim to develop an automatic algorithm for detecting depression symptoms in texts, using a text-based sample of young people seeking advice about self-perceived depressive symptoms. We believe our automatic detection approach, describing the problems of the users in natural language, can be a substantial contribution to this research field. Hence, the current study focuses on how symptoms of depression are manifested through text in natural language using AI

IV PROPOSED SYSTEM

The framework begins with data collection by using Twitter scrapper tools and is stored in a .csv file. The raw data will be cleaned and start to do data pre -processing. The data will be tokenized, stemming, and lemmatization as a part of normalizing the data. Next, the data will be analyzed by using sentiment analysis to obtain a score of words. The data are fed into two different classifiers which are Naïve Bayes and NBTree. The data is split into the train and test set. The training data is used for model development to make sure the classifier learns. The test data will feed into the model once the model learned about the data for evaluation. The accuracy results of both classifiers will be compared to determine which algorithm outperforms

V IMPLEMENTATION

1. Gathering Data:

Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems.

In this step, we need to identify the different data sources, as data can be collected from various sources such as files, database, internet, or mobile devices. It is one of the most important steps of the life cycle. The quantity and quality of the collected data will determine the efficiency of the output. The more will be the data, the more accurate will be the prediction.

This step includes the below tasks:

- Identify various data sources
- Collect data
- Integrate the data obtained from different sources

By performing the above task, we get a coherent set of data, also called as a dataset. It will be used in further steps.

2. Data preparation

After collecting the data, we need to prepare it for further steps. Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training. In this step, first, we put all data together, and then randomize the ordering of data.

- **Data exploration:**

It is used to understand the nature of data that we have to work with. We need to understand the characteristics, format, and quality of data. A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.

- **Data pre-processing:**

Now the next step is preprocessing of data for its analysis.

3. Data Wrangling

Data wrangling is the process of cleaning and converting raw data into a useable format. It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues.

- Missing Values
- Duplicate data
- Invalid data
- Noise

So, we use various filtering techniques to clean the data.

4. Data Analysis

Now the cleaned and prepared data is passed on to the analysis step. This step involves:

- Selection of analytical techniques
- Building models
- Review the result

ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2 April 2024

The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome. It starts with the determination of the type of the problems, where we select the machine learning techniques such as Classification, Regression, Cluster analysis, Association, etc. then build the model using prepared data, and evaluate the model.

5. Train Model

Now the next step is to train the model, in this step we train our model to improve its performance for better outcome of the problem.

We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and, features.

6. Test Model

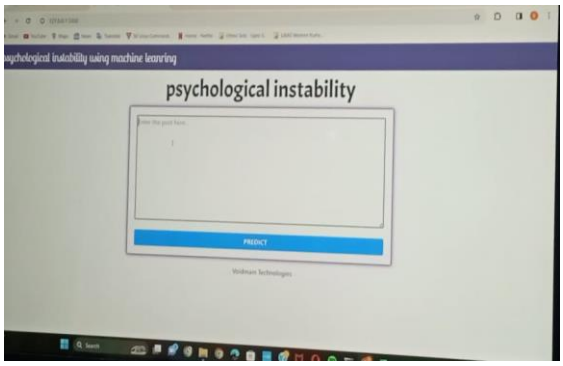
Once our machine learning model has been trained on a given dataset, then we test the model. In this step, we check for the accuracy of our model by providing a test dataset to it.

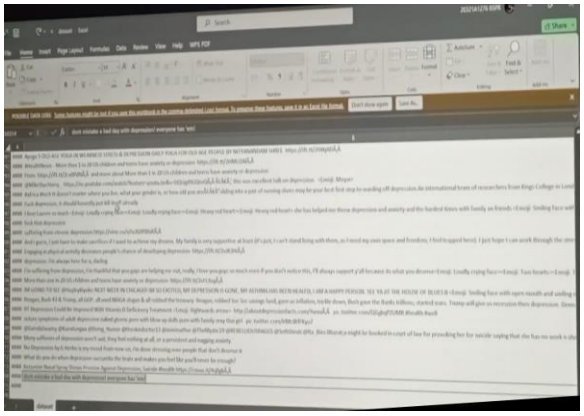
7. Deployment

The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system.

If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will check whether it is improving its performance using available data or not. The deployment phase is similar to making the final report for a project.

VI RESULTS

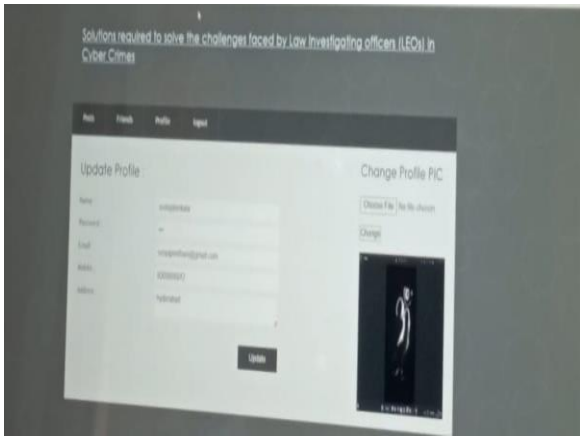




ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2 April 2024



VII CONCLUSION

In conclusion, depression has become a serious problem and affecting mental health. Besides that, the fast growth of social media enables an abundance of data in one's social media. Twitter is an application which is focused on 140 characters per tweets enables the user to share their opinion and thought shortly and directly. Each tweet enables the researchers to extract and analyze the information shared in the tweet. The sentiment analysis technique is applied to each tweet to identify the sentiment score and labelled them as positive, negative, or neutral. The labeled tweets are fed into a machine learning algorithm that enables the classification of the tweets into correct groups. Naive Bayes and NBTree which is the selected machine learning algorithm have been implemented on two different sizes of tweet datasets to find the accuracy of the algorithm on classifying the depressive and non-depressive tweets.

The NBTree algorithm gives an accuracy of 97.31% to classify the depressive and non-depressive tweets on the 3000 tweets dataset and 92.34% on the 1000 tweets dataset. On the other hand, Naïve Bayes show 97.31% on the 3000 tweets dataset and 92.34 % on 1000 tweets datasets.

Naïve Bayes and NBTree are equally efficient by giving the same accuracy value in the experiment. However, the project is limited to the text only. The work can be enhanced in future work by targeting a selective user and their tweets at certain time and determine either the status of depression

REFERENCES

[1] "Depression," World Health Organization. https://www.who.int/health-topics/depression#tab=tab_1 (accessed Sep. 09, 2020).

[2] L. Goldman, "Depression: What it is, symptoms, causes, treatment, types, and more," 2019. https://www.medicalnewstoday.com/articles/8933 (accessed Oct. 10, 2020).

[3] S. Kemp, "Digital 2020: 3.8 billion people use social media - We Are Social UK - Global Socially-Led Creative Agency," Jan. 30, 2020. https://wearesocial.com/blog/2020/01/digital-2020-3-8- billion-people-use-social-media (accessed Oct. 10, 2020).

[4] Y.-T. Wang, H.-H. Huang, and H.-H. Chen, "A Neural Network Approach to Early Risk Detection of Depression and Anorexia on Social Media Text."

[5] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," Heal. Inf. Sci. Syst., vol. 6, no. 1, pp. 1–12, 2018, doi: 10.1007/s13755-018-0046-0.

[6] A. Sistilli, "Twitter Data Mining: Analyzing Big Data Using Python | Toptal." https://www.toptal.com/python/twitter-datamining-using-python (accessed Feb. 04, 2021).

[7] Fontanella Clint, "How to Get, Use, & Benefit From Twitter's API." https://blog.hubspot.com/website/how-to-use-twitter-api (accessed Mar. 05, 2021).

[8]"Depression."https://www.who.int/news room/factsheets/detail/depression (accessed Oct. 10, 2020).

[9] N. Schimelpfening, "7 Most Common Types of Depression," Aug. 26, 2020. https://www.verywellmind.com/common-typesof-depression-1067313 (accessed Sep. 09, 2020).

[10] M. R. Islam, A. R. M. Kamal, N. Sultana, R. Islam, M. A. Moni, and A. Ulhaq, "Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique," Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2018, pp. 4–7, 2018, doi: 10.1109/IC4ME2.2018.8465641.

[11] H. S. ALSAGRI and M. YKHLEF, "Machine Learning-Based Approach for Depression Detection in Twitter Using Content and Activity Features," IEICE Trans. Inf. Syst., vol. E103.D, no. 8, pp. 1825–1832, 2020, doi: 10.1587/transinf.2020edp7023.

[12] M. Nadeem, "Identifying Depression on Twitter," pp. 1–9, 2016.

[13] M. Gaikar, J. Chavan, K. Indore, and R. Shedge, "Depression Detection and Prevention System by Analysing Tweets," SSRN Electron. J., pp. 1–6, 2019, doi: 10.2139/ssrn.3358809.

[14] M. M. Aldarwish and H. F. Ahmad, "Predicting Depression Levels Using Social Media Posts," Proc. - 2017 IEEE 13th Int. Symp. Auton. Decentralized Syst. ISADS 2017, pp. 277–280, 2017, doi: 10.1109/ISADS.2017.41.

[15] P. Kaviani and S. Dhotre, "International Journal of Advance Engineering and Research Short Survey on Naive Bayes Algorithm," Int. J. Adv. Eng. Res. Dev., vol. 4, no. 11, pp. 607–611, 2017.

[16] T. M. Christian and M. Ayub, "Exploration of classification using NBTree for predicting students' performance," Proc. 2014 Int. Conf.

Data Softw. Eng. ICODSE 2014, no. November 2017, 2014, doi: 10.1109/ICODSE.2014.7062654.

[17] S. Narkhede, “Understanding Confusion Matrix | by Sarang Narkhede | Towards Data Science,” May 09, 2018. https://towardsdatascience.com/understanding-confusion-matrixa9ad42dcfd62 (accessed Feb. 04, 2021).