ZKG INTERNATIONAL ISSN: 2366-1313

Machine Learning Algorithms based Depression Detection on Twitter Data

[1] Nelluri Lakshmi Pradeep, [2] Dr. B. Gohin

[1] MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

pradeepnelluri88@gmail.com

[2,] Associate Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

Abstract: *Depression has grow to be a severe problem on this present day technology and the variety of human beings stricken by despair is increasing every day. However, some of them control to acknowledge that they're facing despair at the same time as some of them do no longer realize it. On the opposite hand, the giant progress of social media is turning into their "diary" to percentage their nation of thoughts. Several types of studies were performed to hit upon melancholy through the consumer put up on social media using machine getting to know algorithms. Through the statistics to be had on social media, the researcher can capable of realize whether or not the customers are dealing with melancholy or no longer. Machine mastering algorithm allows categorising the information into correct companies and becoming aware of the depressive and non-depressive statistics. The proposed research work aims to stumble on the melancholy of the consumer by their information that is shared on social media. The Twitter information is then fed into unique types of classifiers, which are Naïve Bayes and a hybrid version, NB Tree. The results may be in comparison based totally on the best accuracy value to decide the first-rate set of rules to discover depression. The outcomes indicate both algorithms perform similarly by means of proving identical accuracy degree.*

KEY WORDS- *Depression, Social Media, Twitter, Classification, Hybrid, NB Tree, Naïve Bayes*

I. INTRODUCTION

Depression is now an issue of mental fitness that is the main concern. Depression is also believed to be a serious illness which affects more than 264 millions humans across the world 1. The causes are numerous, such as unanticipated changes to the surroundings, changes in neurotransmitter concentrations in the brain as a result of certain physical assaults or changes in genetics

[22]. Depression is treatable with treatments or drugs. While many sufferers do use medication however, there are plenty of patients suffering from depression who are not diagnosed due to ignorance about depression. In the end, patients will experience horrible feelings, an insular lifestyle, excessive behavior as well as suicidal thoughts and reliance on medication, including antidepressants. Depression may slow down efficiency in daily actions due to inattention and lack of interest.

The growth and development of traits that social media has for humans has made it easier for consumers to express their emotions as well as thoughts and opinions at the same time. Social media has evolved into a place for people's emotions and thoughts to express them, thereby forcing users to share their thoughts in a way which can be confusing.

Social media apps provide users the chance where they can write about their thoughts, as well as an unintentional note of the emotions and feelings they feel in the form of phrases. Ratings or content created by users can be used by researchers to examine the mindset of users. In addition to the increase in apps for social media, individuals begin to share their struggles in the realm of intellectual fitness online social networks rather than keeping the information private. Data in an account of a customer's account is analyzed in order to evaluate their cognitive fitness, and also the possibility to aid them in their recovery. Twitter is great micro-running software for blogs which allows you to set your short, concise testimonies not more than one 140 characters using easy interfaces. Most tweets from users are public, and can be posted using their own Twitter API. In the end, scientists can determine whether an individual is unsatisfied or not on the basis of the tweets that are publically available. The algorithm for sentiment evaluation can assign a score each textual word in accordance with the type of intensity of the text. It can help in determining the reputation of the user, determining whether they are in a good or terrible situation. Sentiment evaluation applies to every tweet in order to calculate the level of sentiment, and then label the tweet as good, poor or neutral. The analysis of sentiment in Twitter statistics measures the

ISSN: 2366-1313

nationality of a tweet or remark through the process formulating sentiment scores. Following the score of sentiment is calculated for tweets using the evaluation of sentiment is calculated, the system that gains knowledge of the algorithm can help classify tweets as depressing or not, in keeping to the labeling of sentiment score in the top of the tweets.

The aim of the paper is to uncover depression mostly through tweets and the use of system study algorithmic techniques, including Naive Bayes as well as the hybrid model known as NB Tree.

II LITERATURE REVIEW

1) A Neural Network Approach to Early Risk Detection of Depression and Anorexia on Social Media Text

Authors: Yu-Tseng Wang1, Hen-Hsen Huang1, and Hsin-Hsi Chen12

In latest years, the proliferation of textual content messages on social media systems which incorporates Twitter and Reddit has paved the way for the improvement of various applications aimed in the direction of detecting more than a few issues, together with influenza outbreaks, suicide hazard, and intellectual health problems. This look at focuses on our approach for early detection of melancholy and anorexia on social media, as part of the CLEF risk 2018 competition. To perceive functionality times of those intellectual ailments, our fashions leverage a aggregate of TF-IDF (Term Frequency-Inverse Document Frequency) data and convolution neural networks (CNNs) to research the content material of articles published with the resource of users. The professional evaluation of our fashions yielded promising effects, with a depression detection ERDE5 score of 10.Eighty one%, ERDE50 of nine.22%, and an F-rating of zero.37, and an anorexia detection ERDE5 rating of thirteen. Sixty five%, ERDE50 of 11.14%, and an F-score of zero. Sixty seven, showcasing the potential of utilizing social media records for early threat detection in intellectual health problems.

2.) Machine Learning Based Approach for Depression Detection on Twitter Using Content and pastime capabilities

AUTHOR: Hatoon S.AlSagri, Mourad Ykhlef

Social media, which include Face book, Twitter and Instagram, have modified our international all the time. People are now more linked to and tracking a type of proper identity. While social media sincerely has many top notch features, its

downsides also are simple. Recent research has discovered a link among heavy use of social media web sites and lengthy-time period hopelessness. The present take a look at ambitions to use device studies techniques to pick out a doubtlessly depressed Twitter user based totally on each in their community behaviours and tweets. For this motive, we functionally examined and tested classifiers to differentiate whether a client is depressed or no longer makes use of capabilities primarily based on his or her activities in the community and tweets. The consequences confirmed that the more capabilities used, the higher the accuracy and F-diploma ranking in detecting depressed users. This technique is a statistically based totally predictive technique for early detection of depression or different highbrow problems. The essential contribution of this act of examination is part of the exploration of talents and their impact at the detection of the extent of hopelessness.

3.) Depression Detection and Prevention System via Analysing Tweets

AUTHORS:Mrunal Gaikar, Jayesh Chavan, Kunal Indore, Rajashree Shedge

Social media systems like Twitter that could be a micro blogging device enable its clients to specific their feelings, feelings and critiques thru brief text messages. Detecting the feelings in a text can assist one perceive anxiety and melancholy of a character. Depression is a intellectual health trouble which can display up to anybody, at any age. There is a loss of systematic and green techniques to become aware of the mental kingdom of a man or woman. With extra than fifty eight millions tweets generated day by day, Twitter can be used with the intention to locate the sign of despair in a faster way. Recent studies have verified that Twitter may be used to prevent one from taking an excessive step. Our Proposed depression detection and prevention machine can locate any depression associated words or phrases from Tweets and additionally classify the sort of despair, if detected. This tool is proposed so you can diagnose depression and save you it. Proposed machine is using Support vector machine and Naïve Bayes classifier. This hybrid technique works well no longer only with shorter snippets but also with longer snippets.

III System Analysis

Existing System

The paper discusses past research the use of system reading strategies to come upon depression from social media facts,

especially Twitter posts. Previous research have implemented algorithms like Support Vector Machines (SVM), Decision Trees, K-Nearest Neighbours (KNN), Naive Bayes, and hybrid models on Twitter datasets to classify tweets as depressive or now not depressive. Research has shown Naive Bayes often performs well for sentiment evaluation on tweets, with accuracy over 90% in some times. Other research determined SVM with top notch kernels finished excessive accuracy. Some obstacles noted encompass reliance on older datasets, a focal point first-class on purchaser confessions as opposed to fashionable tweets, and potential over becoming. Overall, machine studying has examined promise for detecting indicators of despair from Twitter posts, with Naive Bayes growing as a constantly robust method based totally on past research.

Disadvantages of gift tool:

1. Over fitting - Complex device gaining knowledge of models like deep neural networks can over match the noise and idiosyncrasies of a small dataset, reducing generalizability.
2. Limited function engineering - Simple algorithms like Naive Bayes may not take whole gain of the wealthy linguistic and semantic facts to be had in textual content records.
3. Context dependence - Social media language may be very context precise, so fashions professional on Twitter won't transfer well to exceptional text property like clinical data.
4. Privacy troubles - Analyzing social media data for despair cues will increase ethical issues round consent and privacy protections.
5. Difficulty with sarcasm/humor - Sentiment assessment tools can struggle to come across nuances like sarcasm that invert the obvious emotion of a placed up.
6. Focus on type, now not prognosis - These systems classify textual content, however cannot definitively diagnose depression which requires scientific evaluation.
7. Lack of theoretically-grounded methods - Some device reading packages on this vicinity lack grounding in intellectual and linguistic idea.
8. Detection systems want normal retraining on new records to live applicable.

Proposed System:

The paper proposes using machine studying algorithms - Naive Bayes to find out melancholy from Twitter records. The device collects tweets using the Twitter API, pre-processes the records through putting off noise and normalizing text, then performs sentiment evaluation to assign great, terrible or independent

ISSN: 2366-1313

sentiment scores to every tweet. The labelled tweet dataset is cut up into schooling and test gadgets to gather class models. Naive Bayes is a probabilistic set of regulations that calculates opportunities of every beauty based totally on characteristic occurrences. Naive Bayes is a hybrid version combining Naive Bayes and choice timber, the usage of Naive Bayes at the leaf nodes. Both algorithms are trained at the sentiment-categorized tweets and make predictions on the take a look at set. Their accuracy in classifying depressive and non-depressive tweets is evaluated and as compared. The motive is to determine which algorithm plays better for this text type venture on the Twitter datasets. The outcomes discovered Naive Bayes gather comparable immoderate accuracy, round ninety -ninety seven%, in detecting depressive tweets.

Advantages of Proposed system:

1. Simplicity of algorithms - Naive Bayes and NB Tree are significantly easy models tons much less liable to over fitting in comparison to deep neural networks.
2. Explain ability- The rules and possibilities generated with the resource of using Naive Bayes are more interpretable than black-field AI fashions.
3. Fast training - These algorithms may be skilled all of sudden on new statistics, allowing ordinary retraining to keep up with language shifts.
4. Established techniques - Both Naive Bayes and NB Tree are well-studied in NLP, presenting a robust technical foundation for the gadget.
5. Non-context based totally - Models educated on Twitter information might also want to generalize better to special text property than extra complicated models.
6. Public area statistics - Using publicly to be had, non-non-public Twitter posts reduces privateers worries related to despair detection from social media.
7. Comparative evaluation - Testing two techniques offers perception into the strengths and weaknesses of every set of suggestions for this software program.
8. Classification as a place to begin - The device goals to classes text, now not definitively diagnose melancholy, warding off overstating its abilities.

IV Data Analysis

Import Necessary Libraries

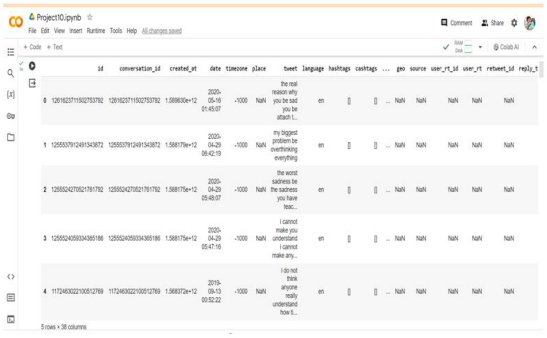
```
importnumpyas np
import pandas as pd
importmatplotlib.pyplotasplt
import seaborn assns
```

Download the data set from Kaggle

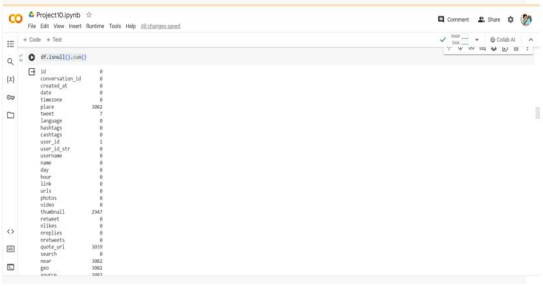
ISSN: 2366-1313

```
df =pd.read_csv('/content/depression.csv')

df.head()
```

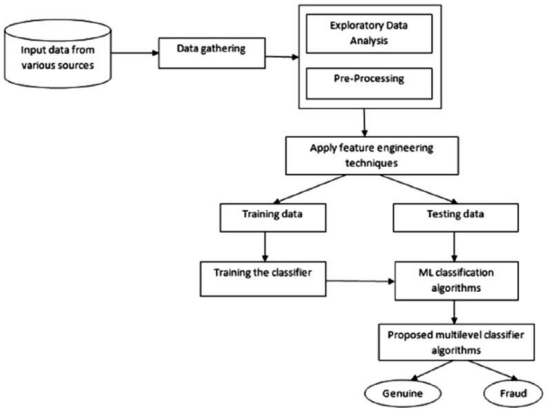


```
df.isnull().sum()
```



V Design

SYSTEM ARCHITECTURE



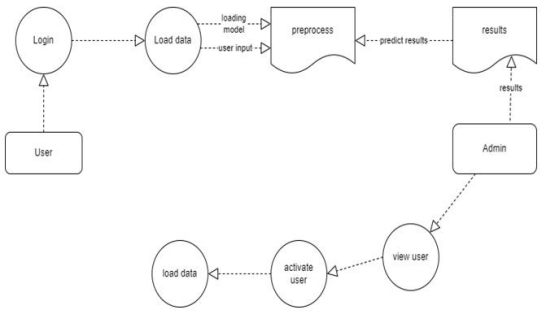
DATA FLOW DIAGRAM:

1. DFD is also referred to as bubble desk. This is a simple graphical formalism that may be used to symbolize a tool in terms of the input records to the device, the more than one processing used for these records and the output statistics generated by that tool.

2. Statistical waft diagram (DFD) is one of the maximum useful modelling tools. It is used to model device components. These objects encompass the device, the records utilized by the machine, the outside entity interacting with the machine, and the facts flowing via the system.

3. DFD indicates how facts flow through the system and how it's far transformed via the transformation chain. It is a graphical technique that shows the unfolding flow of statistics and changes seen as information movements from enter to output.

4. DFD is likewise referred to as bubble chart. A DFD may be used to represent a tool at any degree of abstraction. DFD can be divided into steps that constitute the advent of floating information and useful records.



ISSN: 2366-1313

VI MODULES:

- User
- Admin
- Data Pre-processing
- Machine Learning

-

MODULES DESCRIPTION:

User:

The User can sign within the primary. While registering he required a valid user e-mail and cellular for further communications. Once the user join up then admin can spark off the individual. Once admin activated the individual then consumer can login into our machine. User can add the dataset primarily based on our dataset column matched. For set of policies execution records need to be in float format. Here we took Employment Scam Aegean Dataset (EMSCAD) containing 18000 pattern dataset. User also can upload the today's statistics for gift dataset primarily based on our Django software. User can click on at the Classification within the net web page just so the statistics calculated Accuracy and macro avg, weighted avg primarily based totally on the algorithms. User can display the ml outcomes. Consumer also can show the prediction results.

Admin:

Admin can login along with his login records. Admin can prompt the registered clients. Once he prompts then best the client can login into our tool. Admin can view the general records inside the browser. Admin can click on the Results inside the net web page so calculated Accuracy and macro avg, weighted avg based totally on the algorithms is displayed. All algorithms execution whole then admin can see the general accuracy in internet web page. And moreover display the category outcomes.

Machine gaining knowledge of:

The Accuracy and macro avg weighted avg of the classifiers have emerged as calculated and displayed in my results. The classifier which luggage up the very high-quality accuracy can be determined because the splendid classifier.

VII CONCLUSION

Twitter is a utility that focuses on 140 characters consistent with tweet permitting the customer to percentage their mind and reviews in a concise and direct way. Each tweet allows researchers to extract and take a look at the statistics shared inside the tweet. Tagged tweets are assigned to a device studying set of rules that enables classify tweets into accurate corporations. The NB Tree set of guidelines presents 97.31% accuracy for classifying

depressive and non-depressive tweets in a dataset of three,000 tweets and ninety .34% in a dataset of one,000 tweets. Naïve Bayes and NB Tree are similarly powerful in imparting the equal accuracy cost as test. However, the paintings are restrained to the textual content. The artwork can be advanced in future artwork with the useful resource of focused on a selective individual of their tweets at a selected time and figuring out one or the opportunity depression state.

REFERENCES

1. "Depression," *World Health* 09. L. Goldman, "despair: What it is, symptoms, causes, treatment, types, and more," 2019. https://www.medicalnewstoday.com/articles/8933 (accessed Oct. 10, 2020).

2. 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).

3. 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."

4. 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.

5. A. Sistilli, "Twitter Data Mining: Analyzing Big Data Using Python | Total." https://www.toptal.com/python/twitter-data- mining-using-python (accessed Feb. 04, 2021).

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

7. N. Schimelpfening, "7 Most Common Types of Depression," Aug. 26, 2020. Of-depression-1067313 (accessed Sep. 09, 2020).

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

9. 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.

10 Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.