Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 11, November-2023

# Destabilization of Behavior Through Machine Learning

Dr. Mohammed Abdul Waheed, Vasant Kumar
Computer Science and Engineering, VISVESVARAYA TECHNOLOGICAL UNIVERSITY
Regional Center, Kalaburagi 585 105, Karnataka. INDIA

***Abstract*— Everyone's everyday life are enriched by the presence of online social networking platforms. Facebook ranks high upon list of most visited social networking sites. Facebook users are able to share their thoughts, feelings, and ideas via the use of text and emoticons. There are various perspectives, postings, & comments which could insult sensitivities of others. It's important to keep an eye upon such content in online communities and organizations, but doing so may be challenging since all people have their own opinions & ideas to share. Recent online discussions on depression constitute major cause for concern. Proposed technique tries to detect sad folks by recognizing mood underlying information published on Facebook. Sentiment analysis is used to identify the user's positive (e.g., delighted, surprised, and pleased) & negative (depressed, anxious, tense, sad, and angry) emotions. Next, we see breakdown of negative emotions into their component parts, and we're given an overview of depression severity. This research is useful for monitoring how people feel. For those who are feeling down, you may also provide this solution.**

***Keywords*—** Facebook, depression, anxiety, tension, sadness, and anger

## I. INTRODUCTION

Data mining & machine learning are foundations upon which this project stands. Machine Learning is an area of computer science which enables computers to learn without being specifically programmed. Providing algorithms which may be taught to carry out a job is core focus of machine learning. Data mining is practice of applying techniques from machine learning, statistics, & database systems to discover patterns in large data sets. Main goal of data mining, which spans fields of statistics & computer science, is to take a collection of raw data & extract useful information from it utilizing intelligent approaches. Understanding and categorizing the feelings and attitudes represented in written material is the focus of sentiment analysis. Social media platforms are generating an immense volume of data that is abundant in emotional content. Knowing the collective mindset is greatly aided by doing sentiment analysis on this user-generated content. Sentiment analysis on Twitter is more challenging than on other platforms because of the prevalence of slang and typos. Introduced a novel feature vector to separate tweets into good, negative, and opinion-extracting categories. [1]

Data mining is practice of discovering patterns and insights in massive datasets. There is little use in collecting massive amounts of data if we cannot utilize it to learn anything new or to anticipate future events according to known patterns[10].This has to be accomplished more quickly, easily, and precisely. Therefore, data mining is now crucial in the ICT sector. Among numerous uses for data mining[11] is in field of sentiment analysis. Text documents and social media may all be subjected to Sentiment Analysis. Sentiment Analysis enables users to get insights into opinions conveyed in tweets and categorizes them into negative & positive classifications. Businesses might use this to ascertain reception of their products among consumers; with this knowledge, they may customize their future offerings in response to feedback.

## II. LITERATURE SURVEY

Chiara Zucco, et.al,[1] 2017, suggested use of sentiment analysis and emotional computing approaches to depression identification and monitoring are described and explored. An initial plan for multi-modal system for depression monitoring is also offered, one that makes use of sentiment analysis & influential computation methods. In particular, article details key concerns & difficulties associated with design of such system.

N. Azmina, Nasiroh Omar[2] 2017, uses opinion mining & sentiment analysis to isolate Facebook's English & Malay terms. Emotional clusters are formed from retrieved textual information. This effort starts with translating unstructured material into relevant lexicons following extracting Facebook's contents. All of relevant lexicons are kept in database once human identifications have been performed. Sentiment analysis divides feelings into three categories: positive (happy), negative (angry), & neutral (no emotion). Findings are given with proportion of emotion categories in order that it may be deduced whether or not particular Facebook post has garnered favorable or negative feedback from people. An issue with published exam results is used to analyze students' reactions to the case study. This research is important because it will allow administrators and businesses to keep tabs on any topic of conversation that might lead to improvements in quality of their services. Julia Neidhardt, Hannes Werthner,[3] 2018, Travelers and businesses alike may benefit from wealth of information available from tourism-related websites. More and more businesses in the tourism industry realize user information is valuable resource. In addition, landscape of empirical

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 11, November-2023

research is shifting as data is constantly created & always accessible. In this research, we examine how users behave and communicate when traveling. In particular, user anticipation and excitement for planned excursions is examined to better understand their interconnected nature. Analyzing user interactions using social network analysis. Emotions are measured by analyzing free-form language users provide in travel discussion forum using text mining & sentiment analysis methods. Results of experiment offer credence to theory as network influences users' emotions. S. Alami, O. Elbeqqali [4], 2017, Internet forums, instant messaging services, and other forms of digital communication have all benefited from exponential growth of technology for communication and information. However, criminals make utilize this technical advance in negative way. Users create a wide variety of dubious content (text, picture, video...) and share it with one another on social networking platforms. Since vast majority of social media data is saved in plain text, we will limit our analysis to that. This new mode of communication poses a formidable opportunity for academic inquiry, yet text mining offers a powerful tool for addressing this issue. Text analysis employs a similarity method to identify potentially malicious social media messages. Our method is tested in wild using authentic comments.

## III. PROPOSED SYSTEM

Hallmark symptom of mental stress is having an abundance of negative thoughts instead of enough positive ones, and this may be understood with implementation of machine learning algorithms. NLP, & CNN for text content processing are all incorporated in suggested system. Purpose of this research was to perform comprehensive evaluation of current literature to determine where SA stands in terms of health & happiness. We zeroed upon user-generated information, rather than content created by medical experts, to reflect viewpoint of people whose well-being & health are at stake.

## IV. METHODOLOGY

Logistic Regression In field of machine learning, logistic regression is common technique which falls within supervised learning paradigm. In this way, we may make accurate predictions about the future using just a small set of independent factors. Predicting outcomes of organized variables is done via logistic regression. The output ought to be a number that can be easily categorized or counted. It's possible to get values of 0 and 1, although probabilistic values ranging from 0 to 1 are returned instead of precise values like 0 & 1.

$$e^{(\beta 0+1x)}/1 + e^{(\beta 0+\beta 1x)}$$

K nearest neighbour classifier Supervised Learning is basis of K-Nearest Neighbor method, a basic machine learning algorithm. In K-NN method, there will be little difference between old cases & new cases/data. KNN is an example of a non-parametric method; it does not make any assumptions about the data it highlights or the distribution of that data. Multiple classes may make use of it without issue.
Annotation of a decision tree In the field of data mining, the decision tree is common supervised machine learning approach. A decision tree is diagram that individuals utilize to illustrate statistical likelihood or to indicate sequence of events, actions, or consequences..

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Random forest classifier Supervised Learning is basis of K-Nearest Neighbor method, a basic machine learning algorithm. In K-NN method, there can be little difference between old cases & new cases/data. KNN is example of non-parametric method; it does not make any assumptions about the data it highlights or the distribution of that data. Multiple classes may make use of it without issue.
Annotation of decision tree In the field of data mining, the decision tree is common supervised machine learning approach. A decision tree is diagram that individuals utilize to illustrate statistical likelihood or to indicate sequence of events, actions, or consequences.

$$K_k^{cc}(\mathbf{x}, \mathbf{z}) = \sum_{k_1,\ldots,k_d,\sum_{j=1}^{d} k_j=k} \frac{k!}{k_1! \cdots k_d!} \left(\frac{1}{d}\right)^k \prod_{j=1}^{d} \mathbf{1}_{\lceil 2^{k_j} x_j \rceil = \lceil 2^{k_j} z_j \rceil},$$
$$\text{for all } \mathbf{x}, \mathbf{z} \in [0,1]^d.$$

An assignment involving classification prediction may have four different outcomes.
Whenever observation is accurately identified as falling to specific class, we refer to prediction as "true positive."
A true negative is a prediction that an observation isn't related to a class when the observation in question does not truly belong to that class.
Labeling observation as belonging to given class while it really does not is instance of false positive.
An instance of incorrectly identifying observation as unrelated to class, despite it being connected, is known as false negative.
Recall, precision, & accuracy are three basic measures utilized to assess classification model.
Accuracy refers to the percentage of instances in which the test data is properly anticipated. For getting the accuracy rate, divide the number of correct possibilities by the total amount of guesses.

Accuracy=correct predictions/all predictions

Precision is proportion of instances that really belong to class to which they were assigned, as opposed to total number of instances for which that class was assigned.

Precision=true positives/true positives+false positives

Recall refers to the proportion of properly recognized cases belonging to certain class, represented as percentage of the total amount of instances in that class.

**Recall=true positives / true positives+false negatives.**

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 11, November-2023

## V. SYSTEM ARCHITECTURE

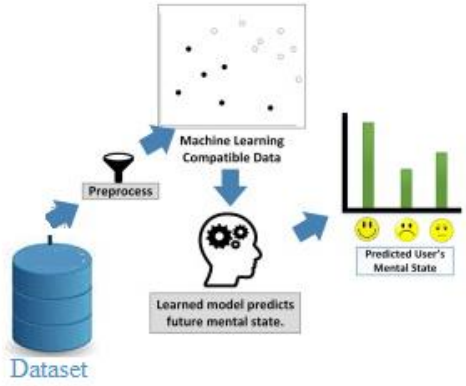
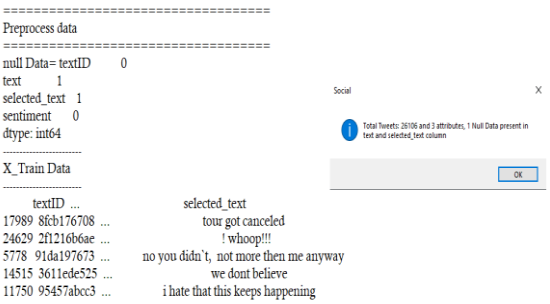



Figure 1: System Architecture

## VI. IMPLEMENTATION



```
==================================
Preprocess data
==================================
null Data= textID        0
text         1
selected_text   1
sentiment       0
dtype: int64
----------------------
X_Train Data
----------------------
        textID  ...                selected_text
17989  8fcb176708  ...                tour got canceled
24629  2f1216b6ae  ...                ! whoop!!!
5778   91da197673  ...  no you didn't,  not more then me anyway
14515  3611ede525  ...                we dont believe
11750  95457abcc3  ...  i hate that this keeps happening
```

Figure 2: Preprocessing
It preporcess the dataset in terms of train and test data

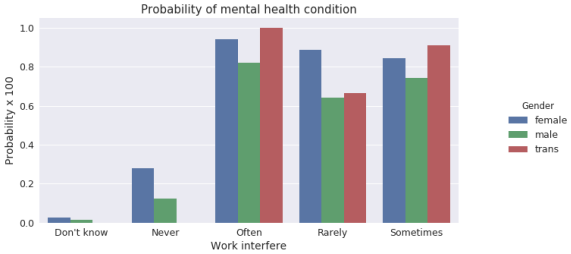



Figure 3: Mental health condition based on text

## VII. CONCLUSION

Program helps us locate feelings relating to a given subject, dispute, problem, issue utilizing keywords attached to topic of interest. The gathered and processed data represents the collective mood of social media users. Uses for this information range from analyzing consumer needs to determining how certain problems have affected customers. The gleaned insights may then be implemented for future user experiences.

## REFERENCES

[1] Neethu M S, Rajasree R," Sentiment Analysis in Twitter using Machine Learning Techniques"(2013)
[2] Shahid Shayya, Noor Ismavati Jaffar, "Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges"(2016)
[3] Young sub-Han," Sentiment Analysis on Social Media Using Morphological Sentences Pattern Model",(2017)
[4] Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining"(2017)
[5] Neethu MS And Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques"(2016)
[6] Chiara Zucco, Barbara Calabrese, Mario Cannataro, "Sentiment Analysis and Affective computing for depression monitoring." (2017)
[7] N. Azmina, Nasiroh Omar, "Sentiment Analysis: Determining people's emotions in Facebook" (2017)
[8] Julia Neidhardt, Hannes Werthner," Predicting happiness: user interactions and sentiment analysis in an online travel form"(2018)
[9] S. Alami, O. Elbeqqali "Detecting Suspicious Profiles Using Text Analysis Within Social Media" Journal Of Theoretical And Applied Information Technology,(2017)
[10] D.Rajeswara Rao, Vidyullatha Pellakuri, "Knowledge Based Information Mining on data using statistical approaches", International Journal of Pharmacy and technology, Volume 8, issue 4, December 2016, pg no:21961-21966.
[11] B.Sekhar Babu, P.Lakshmi Prasanna, P.Vidyullatha, "Personalized web search on e-commerce using ontology based association mining", International journal of Engineering and technology, Volume 7,issue 1.1, 2018,Pg no:286-289