International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

DEPRESSION DETECTION USING MACHINE LEARNING APPROACH

M.M Mane[1], Sarthak Raj[2], Satyam Arya[3,] Pragya Chaudhary[4]

[1]*Assistant Professor, Dept. Of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India*

[2,3,4]*Student, Dept. Of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India*

Abstract - *Due to the increasing popularity of communications occurring in social networks, an inconsiderable part of young people's interpersonal interactions is performed online. New risks and threats emerge through the introduction of novel communication media, such as cyber bullying, -stalking and online grooming. In this context, the main affected group are children aged between 12 and 17 years. This can be daunting to explore things you care about. The danger of web bullying and intimidation suggests that people are avoiding sharing themselves and giving up searching for alternative viewpoints. Platforms struggle to facilitate conversations effectively, leading many communities to restrict or shut down user comments altogether. Depression is one of the main issues impacting mental health. Bullying, violence, and depression are the cornerstones of rising suicide worldwide. Therefore, this project seeks to build a method for determining whether or not a given post is suicidal or harassing.*

***Key Words*: Social network, Emotions, Depression, Sentiment analysis**

1.INTRODUCTION

Internet is taken into consideration as a primary storehouse of records in today's international. No single work can be achieved without assistance. It has ended up one of the most important methods of verbal exchange. Out of all communication techniques available, Social Media is one of the maximum common bureaucracies. Social media has modified the arena. The number of internet users globally in 2019 is 4.388 billion, up 9.1% 12 months-on-year. The range of social media customers global in 2019 is 3.484 billion, up 9% year on year. The sale of societal evils has been blamed on social media. People have a want to match in to be visibly praised in this new virtual world. Cyberbullying, harassment, and depression are just a few of the negative effects of social media. There's growing evidence that there's a correlation between social media and depression. It is an almost indisputable fact that access to a cloak of anonymity and a big, massive microphone brings the worst out in some people. Today, that is particularly actual on social media. Online harassment and hate- messaging are a developing fashion and a growing security challenge for all people on social media these days. In truth, 40% of Internet customers say they have in my view experienced online harassment. Harassment on the Web has additionally taken its toll on celebrities and public figures, a number of that has therefore signed off social systems for rights. Hence, we have provided you with this assignment wherein we will be an enveloping gadget that enables in detecting depressive and harassment textual content.

1.1 Problem Definition

Due to the lack of right systems for detecting despair and harassment, humans in recent times are careworn on social media. People expression despair are also now not identified and as a result they get lonelier and worse. The top obstacle of enforcing a Depression and Harassment Detection on Social Media System is the social and psychological help of the individuals themselves. Today, that is specifically authentic on social media. Online harassment and hate-messaging are a developing trend and a developing safety situation for everybody on social media nowadays. In truth, 40% of Internet customers say they have in my opinion experienced on-line harassment.

1.2 Motivation

Depression is a curable disorder. Early detection and intervention would shorten the remedy route. Unfortunately, the charge of accessibility to treatment is relatively low. It became stated that much less than 50% of those who have this intellectual illness gained get entry to mental health carriers. The boundaries encompass a lack of information and cognizance in despair, having bad perceptions about mental health offerings, and a limited number of mental health professions. To help boom the rate of accessibility to intellectual health services, it is necessary that advanced technology and a proactive approach must be used.

2. LITERATURE SURVEY

2.1 Existing Systems

We have surveyed various papers and systems which have carried out or researched on text type, herbal language processing, depression or mood or harassment detection and evaluation and discussed this underneath. This will help one benefit a comprehensive idea of all the one-of-a-

kind sorts of gadget proposed or developed and might help one develop an efficient system primarily based on formerly proposed or evolved systems

a. Mood Path: A Depression

Mood direction gives an evaluation of mental health that allow you to make the enchantment for judgment to explore professional treatment. Unlike different checks that ask you to reflect on the beyond weeks, Mood path asks you questions about your emotional properly-being at the moment over a chain of 14 days. The app is designed to facilitate conversations with a professional, but more than 150 physical games and tools can also be found in the app to paintings on your intellectual health. (1, n.d.)

b. Youper:

Youper is an AI chatbot developed to help users recognize and music their mind and feelings, and manage them. The key functions of this app are: Youper Ai chat, journal logs, mood logs, and many others. The application asks consumer to provide an explanation for how they experience with the option of asking Youper for guide in describing their emotions. (2, n.d.)

c. My Mobile Watch Dog

My Mobile Watchdog enables parents to music cellphone interaction from their youngsters and help them make good digital choices. You will set up rules for your child with over 20 parental controls, get warnings when uncommon behavior occurs and see wherein your child is always. While to be had on Android and IOS smartphones, sure apps don't work on iPhones and there have been reviews that a number of the protections put up via their dad and mom had been bypassed by means of teenagers. The web page of My Mobile Watchdog enjoys a variety of public coverage but a number of details is missing. The 8 capabilities described on the dashboard have very short reasons and offer little information on what capabilities are ignored for iPhone and whether jailbreaking is wanted. (3, n.d.)

d. Cyber Patrol Online Management

Cyber Patrol, which continues youngsters safe, comes for faculties in stand-alone models for the house and network fashions. The community model either runs on neighborhood location networks or thru proxy servers. Cyber Surveillance for schools is based totally on blocking off Network get entry to, and it is going by Microsoft's proxy service, Microsoft Internet Security and Optimization Service 2000, or Novell Border Boss.

e. Optimized deep Learning-based Twitter Cyberbullying Detection.

Cyberbullying is a criminal offense wherein a wrongdoer bullies a sufferer who is being abused and disliked online. Some solutions to cyberbullying prevention have been evolved but they have been often targeted on textual and interface interfaces. The bulk of the literature work geared toward enhancing the identity by way of adding new apps. Though the extraction and selection approaches of the app get tougher because the number of capabilities will increase. However, some other downside to those enhancements is that it is simple to supply such capabilities-as an example, person age-. In this paper authors propose automatic identity of Twitter cyberbullying based on the very complicated strategies of ML's deep-getting to know location.

2.1 Review of Existing Models and Approaches:

a. MoodPath:

The System isn't always based totally on textual information that facilitates to discover temper and mental health however as an alternative primarily based on question and solution machine. Also, the machine additionally takes 14 day to form an evaluation and get consequences.

b. Youper:

Youper is chatbot based totally gadget therefore the system simplest gets statistics approximately the questions being asked and it's far extra of mood tracker in place of depression or harassment detection.

c. My Mobile WatchDog:

Although blocking off a particular internet site is straightforward and receiving a caution must your toddler attempt to access anything banned, the filtering capabilities of My Smartphone Watchdog are not very sophisticated. You can't clear out domain classes; you need to block pages one by one, instead. The region functions had been deceptive. Although blocking off a certain internet site is simple and getting a warning must your infant attempt to get entry to whatever banned, the filtering abilities of My Smartphone Watchdog are not very sophisticated. You can't filter area classes; you have to block pages one by one, as an alternative. The area features are deceiving.

d. Cyber Patrol Online Management

- No faraway management or notification features.
- Clever children can beat time scheduler and IM clear out.
- IM filter may additionally block innocent phrases

3. OBJECTIVE

The proposed system aims at providing automation for detecting depression and harassment on social media platforms.

- The system will assist the authorities with certain features added in tackling the problem of depression and harassment that is going on in social media through alerts.

4. APPROACH

Our version basically makes a specialty of detecting depressive and harassment associated textual submit's which might be shared on social media. A hassle with modelling textual content is that it's miles messy, and techniques like system getting to know algorithms prefer well defined constant- length inputs and outputs. Machine mastering algorithms can't work with uncooked text at once; the textual content need to be transformed into numbers. Specifically, vectors of numbers. This version

accommodates of strategies like Bag of Words. The bag-of-phrases version is a simplifying representation utilized in herbal language processing and records retrieval (IR). A textual content (consisting of a sentence or a file) is represented in this paradigm as a bag (multiset) of its words, ignoring grammar and even phrase order while retaining multiplicity.

Hence, we first study the information that is available to us inside the csv documents. We load records from the ones more than one csv files. Then we carry out records merging to create a unmarried dataset on which we then carry out cleaning. We then visualize the dataset distribution and perform tokenization to symbolize text data for the model in shape of numbers. Then we perform series padding after which we split the records into three units. We then visualize the distribution of those units and take a look at if it's similar to the authentic distribution. Then we teach the ANN model and visualize the teach and validation result to stumble on underfit and overfit if they occur. Then we run the take a look at set and take a look at the outcomes.

4.1 Analysis/Framework/Algorithm:

ANNs are computational fashions inspired with the aid of an animal's vital frightened systems. It is able to system learning in addition to sample reputation. These presented as structures of interconnected "neurons" that could compute values from inputs. Stop through any EFT center, which will make the charge.

A neural network is an oriented graph. It consists of nodes which within the biological analogy represent neurons, related by means of arcs. It corresponds to dendrites and synapses. Each arc related to a weight at the same time as at every node. Apply the values received as input via the node and define Activation feature along the incoming arcs, adjusted with the aid of the weights of the arcs. A neural network is a machine studying set of rules based totally on the version of a human neuron. The human mind consists of millions of neurons. It sends and method signals in the form of electrical and chemical signals. These neurons are linked with a special structure known as synapses. Synapses allow neurons to bypass alerts. From huge numbers of simulated neurons neural networks forms. An Artificial Neural Network is an records processing method. It works just like the way human mind methods records. ANN consists of a massive range of connected processing gadgets that work together to method records. They also generate meaningful effects from it. We can observe Neural community now not most effective for classification. It can also follow for regression of non-stop goal attributes. Neural networks discover great application in records mining used in sectors. For example, economics, forensics, and many others. And for pattern reputation. It may be extensively utilized for statistics category in a massive amount of facts after careful training. The following three layers can be found in a neural network:

- **Input layer** – The activity of the input units represents the raw information that can feed into the network.
- **Hidden layer** – To determine the activity of each hidden unit. The weights on the links between the input and hidden units, as well as the actions of the input units. There may be one or more hidden layers.
- **Output layer** – The behavior of the output units depends on the activity of the hidden units and the weights between the hidden and output units.

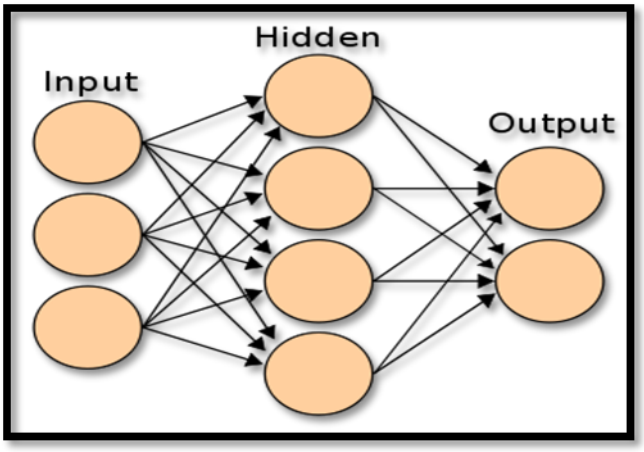

Fig -1: 3 Layers of Neural Network

5. TRAINING CLASSIFICATION MODEL

5.1 Preparing Dataset

Twint is a modern Python-written Twitter scraping software that helps you to retrieve Tweets from Twitter debts by utilizing Twitter's API. (5, n.d.) We used twint library to retrieve the depression and harassment dataset based on keywords and hashtags that reflect melancholy and harassment. Other than twint we've used sentiment one hundred forty dataset and toxic comment dataset to gather effective, bad and poisonous remarks to feature to our dataset. The labels we've a mentioned are zero for non-harassment and non-depression within the respective fashions and 1 for melancholy or harassment inside the respective fashions. We have completed sampling of those Kaggle datasets and our fetched datasets to gain a distribution.

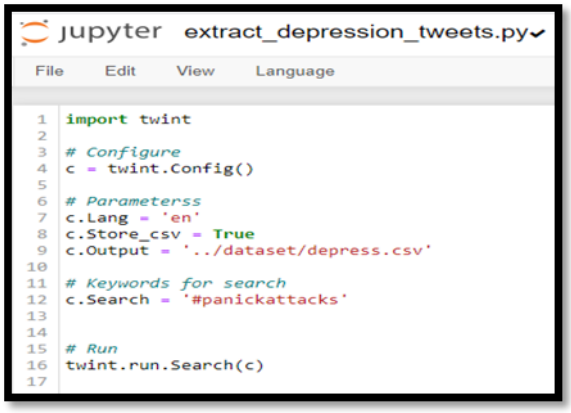

Fig -2: Data Extraction

5.2 Data Cleaning

Data cleaning is an essential tool for Data and NLP research. Text cleaning involves eliminating stop words (words that don't make much of a difference to the model) and these are usually pronouns, conjunctions etc. In addition to that code cleaning in NLP involves deleting unique characters such as @, # and trailing. Stemming is translating those terms that are identical to a specific root word so model will apply to all those cases where similar words are used, function better, and therefore minimize difficulty.

Fig -3: Data Cleaning

5.3 Tokenization

Tokenization is a greater sophisticated form of data control which can suit and exchange exclusive textual content files. It might be the right alternative for big initiatives. The Tokenizer must be built to suit into either uncooked text or encoded text files in an integer. Next is sequencing and series padding wherein we remodel the tokens into sequences used for a sequential pattern. These sequences that modify in period, so we pad positive sequences to get all the statements or sequences to a general period. (4, n.d.)

Fig -3: Tokenization

6. SYSTEM DESIGN

There are two aspects to the project design.

6.1 Front-End Implementation

The front quit is a Web App which includes the subsequent modules:

• Home Page

The domestic web page acts as the touchdown page wherein the consumer comes while journeying our website online. It has a few statistics approximately what's depression and what is harassment.

• About Page

The about web page carries records about the group that developed the gadget.

• Check Page

The take a look at page is the principle center a part of the front-cease. It is the web page in which the person will input the text on which detection/ type is to be executed. It incorporates the form thru which the user will input statistics.

6.2 Back-End Implementation

Web Server

We used Flask Web Framework to broaden our Hyper Text Transfer Protocol (HTTP) server. This server is used to offer the following services:

●To provide a website for the consumer to get admission to.

●Accept text from the users to predict.

7. CONCLUSION

We have completed demanding situations indexed in despair and harassment detection on textual statistics on implementing, developing and trying out the system on diverse inputs. The machine performance is found to be tremendous in figuring out the type of text entered whether or not it belongs to elegance melancholy or class harassment or neither. We had been capable to reveal by way of enforcing the gadget on real time records and beneath real-time situations that generation can help clear up this hassle and provide an early indication of depression and harassment. Based from the locating in this literature evaluation, its miles found that the three maximum concerning troubles are regarding Ethical worries, Lack of information, Stigma and/or awareness of intellectual well-being. This study also gave a top-level view of the presently used techniques in text-primarily based depression detection. Even though it mentioned that using Classifiers, Support Vector Machine, and Probabilistic Classifier proves to be the most popular technique, it's miles interesting. Detection, however also suggest a new approach to cope with long sequences via summarizing the textual content before feeding it into the model. Current end result is higher than any text-based despair detection version, however still wishes optimization together with summarizer configuration or hyperparameter tuning to similarly improve the overall performance of this model.

ACKNOWLEDGEMENT

We wish to express our sincere gratitude towards our college and university for giving us this chance to work on a project. We would also like to sincerely thank our Project Guide Prof. M. M. Mane for her guidance and encouragement towards us while we were preparing this project and writing the research paper.

REFERENCES

1. (n.d.). MindDoc. Retrieved from https://mymoodpath.com/en/
2. (n.d.). Youper.ai. Retrieved from https://www.youper.ai/
3. (n.d.). my mobile watch dog. Retrieved from https://mymobilewatchdog.com/
3, j. T. (n.d.). Artificial Neural Network Tutorial. Retrieved from https://www.javatpoint.com/artificial-neural-network
4, t. d. (n.d.). Tokenization for Natural Language Processing. Retrieved from https://towardsdatascience.com/tokenization-for-natural-language-processing-a179a891bad4
5. (n.d.). Complete Tutorial On Twint: Twitter Scraping Without Twitter's API. Retrieved from https://analyticsindiamag.com/complete-tutorial-on-twint-twitter-scraping-without-twitters-api/

BIOGRAPHIES

Mrs. M. M. Mane is working as Assistant Professor at Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering Pune from last 8 years. She has been working in the field of statistical analysis, Data analytics for last five years and having good experience of machine learning and natural language processing projects.

Mr. Sarthak Raj is pursuing his Bachelor of Technology in Computer Engineering from Department of Computer Engineering of Bharati Vidyapeeth (Deemed to Be University) College of Engineering, Pune. At present he is in 8th Semester of his 4th Year. His areas of interest in the field are Machine Learning, Cloud and Backend.

Mr. Satyam Arya is pursuing his Bachelor of Technology in Computer Engineering from Department of Computer Engineering of Bharati Vidyapeeth (Deemed to Be University) College of Engineering, Pune. He has a keen interest in data analytics, Front End development and backend operations.

Ms. Pragya Choudhary is pursuing her Bachelor of Technology in Computer Engineering from Department of Computer Engineering of Bharati Vidyapeeth (Deemed to Be University) College of Engineering, Pune. At present she is in 8th Semester of her 4th Year. Her areas of interest in the field are Machine Learning and Python Programming.