Machine learning algorithms based early Recognition of Depression from Social Media Posts

[1] Durga prasad battina, [2] A. N. L Kumar

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

durgananna143@gmail.com

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

Abstract: *Different social media systems are cutting-edge amongst all age groups of humans. They publish their each day sports regarding the matters which have came about to them. People additionally explicit their emotions which can be of any kind, together with depressive, sarcastic, irony, and many extra. Identifying despair from those social media posts could be very difficult work. This work has accrued a dataset containing depressive and non-depressive tweets from Twitter and investigated unique traditional machine-learning classifiers. Among all classifiers, the Support Vector Machine (SVM) plays better than the remaining and acquired an F1-rating of 0.89.*

KEY WORDS- Machine-learning classifiers, SVM, Twitter, Depression

I. INTRODUCTION

The smooth get entry to of distinct interpersonal interaction destinations has empowered all people to make results easily, specific, and offer their mind, musings, conclusions, and emotions approximately something with billions of others around the world. With the improvement of improvements, it is totally viable to proportion your contemplation approximately something through on-line media levels, as an instance, Twitter, Wikipedia, Google, Face book, Instagram, and so forth.

Our paintings is based at the information collected from Twitter approximately Depression. We observed exclusive situations where someone can go into the depression kingdom smoothly but coming out of its miles very hard without professional advice or consulting a psychiatrist. Even though it is a vital mental issue, now not precisely 50% of the people who've this wistful difficulty accessed psychological properly-being administrations. Sorrow has gotten one of the main psychological wellbeing troubles. This will be a consequence of several elements, which include having nonattendance of mindfulness about the illness. One solution might be to create a gadget that might locate someone's depression even early. It will assist create consciousness amongst human beings to preserve proper mental fitness. There can be unique reasons behind a person getting into the depression nation, including now not getting the preferred activity, due to a family problem, abusive relationship, regular disappointment within the examinations, no longer getting a healthy operating environment, the death of a cherished one's, some other character troubles and, intake of excessive medicinal drugs also leads to depression.

Depression is a splendid hassle in our network and has continuously been a trending area for sentiment evaluation researchers. It is especially a intellectual sickness wherein humans emerge as sad without understanding the cause at the back of their disappointment. People start forming poor thoughts of their minds; they could not give attention to their paintings efficaciously, which creates a sad environment. Depression can also purpose intellectual problems additionally. It's an extreme crippling disorder that might negatively have an effect on people from all generations, main to unhappiness, feeling lonely, and lack of ability to sleep. It is considered the most important issue in global incapacity and a key cause for suicide. Depression frequently leads human beings to dedicate suicide due to the fact they cannot discover a solution. And if it isn't handled, it affects people's everyday lives surrounded by means of the character who is genuinely depressed, as in a circle of relatives, within the workplace, or maybe in our societies. As per the World Health Organization (WHO) examine in 2018, over 350 million individuals experienced despair, and just about 1 million individuals with wretchedness ended their lives every 12 months. As in step with WHO, four.4% of people are going thru a kingdom of depression that is more common in females than males. The use of social networks has increased with the upward thrust in populace and communications generation, and they are being used for lots one of a kind functions. Here we've got used those social networking sites as a supply to accumulate records on despair. In these studies, numerous methods of depression prediction are mentioned in depth. The strategies contain the collection of a dataset through social media published texts. From extracted records, the end result is obtained.

Here, by using figuring out and extracting emotions from the textual content posted via social media (Twitter), the use of system-mastering strategies, and herbal language processing (NLP) techniques, we gift a person's degree of melancholy.

Machine-learning strategies would possibly provide a few highlights that could help with reading the interesting examples included up in virtual channels and cycle them to find the kingdom of mentality (for example, 'pleasure', 'bitterness', 'outrage', 'uneasiness', discouragement) amongst interpersonal companies' clients. In this observe, we aim to take a look at the put up and decide if the user is depressed. We can in addition detect different intellectual issues and might be capable of shape a mechanism that might help us with distinguishing and cut off despondency dispersion in interpersonal organizations. This exam abuses Twitter's data over 10,000 tweets. Different traditional device-gaining knowledge of classifiers are applied to understand the melancholy stage, of which Support Vector Machine (SVM) suggests the highest consequences, with an exactness of 0.Ninety one and F1-rating of 0.89. This paper's ultimate portion is as follows: Section 2 illustrates the associated paintings. Section three explains the dataset. Section four describes the exclusive conventional device-getting to know classifiers to come across melancholy. Section 5 tells us about category method. Section 6 displays the end result of the work. Section 7 is about the discussion. Finally, Section eight outlines the conclusions and destiny scope of the look at.

II RELATED WORKS

Alsagri et al. Added a novel technique based totally on the Linear SVM version. It makes use of Linguistic Inquiry and Word Count (LIWC), sentiment analysis, social activity, and synonyms and done an accuracy of eighty 2.5% at the Twitter dataset. Lin et al. Also done their study the usage of Reddit consumer's records. They hired blended features (LIWC+LDA+ BIGRAM) and N-grams to classify despair information the use of Multilayer Perception (MLP) version and scored an accuracy of 91%. Hassan et al. [4] performed their study on two datasets, the Twitter dataset, and 20newsgroups. In his research, by noticing and keeping apart emotions from the text, making use of emotion theories, machine getting to know techniques, and natural language processing techniques, he introduced how to discover a man or woman's depression stage. SVM shows the quality outcomes with an accuracy of ninety one% in assessment to Naive Bayes and Maximum

Entropy classifiers. Guntuku et al. delivered a new approach based at the Neural Network version wherein it utilizes N- grams capabilities and reached an accuracy of 70% at the Twitter dataset. Arun et al. showcase a novel method for identifying depression the use of scientific statistics from the on-going Mysore investigations of Natal results for Aging and Health (MYNAH). The proposed model turned into created, using XG Boost and an accuracy of ninety 7.8% via the usage of feature selection at the available facts of reviews with progressed reality. Khan et al. referenced a framework that predicts and calculates the Bengali textual content's sentiment that changed into acquired from Face book. They utilized system- mastering classifier algorithms to find the fine exactness and pick out the 2 styles of groupings as glad and sad. In the wake of pre-processing the data, they tokenized the statistics by utilizing Count Victimizer. They carried out six distinct algorithms to foresee the most noteworthy exactness from that point onward. Among them, the Multinomial Naive Bayes gave us the maximum fantastic accuracy of 86.Sixty seven%. Peng et al. proposed a multi- kernel SVM-primarily based version. It makes use of three characteristic categories: user micro-blog text, consumer profile, and user behaviours and gained an accuracy of 83.46%. Nadeem et al. Carried out their examine using 2.5M tweets. He employed capabilities including bow and sentiment analysis the use of the Naive Bayes model and scored an accuracy of 81%. Jain et al. Laboured on two unique depressive datasets. One is primarily based on the questionnaire and another on Twitter.

III DATA SET

Data Collection

To accumulate depressive tweets, we extract tweets with hash tags #melancholy and #sad fees the usage of Twitter and manually pick English tweets. We also used different unique phrases like 'distress', 'unhappiness', and 'sorrow' to acquire depressive tweets from this domain. Out of these gathered tweets, depressive and non-depressive tweets are further manually separated. To collect extra non-depressive tweets, we extracted tweets with keywords together with 'distress', 'disappointment', and 'sorrow' which do now not incorporate hash tags# despair and #unhappy prices. Further English tweets had been manually selected from them. Having simplest depressive or only non- depressive tweets from a selected do-most important may additionally lead to an independent class gadget; therefore, we made positive that there have been depressive and non-depressive tweets from every area.

Data Processing and Annotation

Tweets are annotated with the aid of a group of people fluent in English. Each tweet is manually annotated for the presence of despair.

Depression Annotation Each tweet is manually annotated for the presence of depression the usage of the tags' YES' and 'NO'. Tweets with the hash tags #melancholy are much more likely to comprise melancholy. Tweets that do not include these hash tags are then manually confirmed to control melancholy. Here is an instance of a tweet (with translation in English) that incorporates despair and one that does not:

Tweet: #depressive... I'm very disenchanted. Depression: YES

Tweet: #everyday rates... I'm very satisfied. Depression: NO

Hash tags #melancholy is randomly deleted from a few tweets which comprise despair in order that the dataset includes depressive tweets with the hash tags #despair and some without the hash tag.

Dataset Analysis

The dataset includes 12,029 English tweets, out of which 5,529 tweets are labelled as depressive and six,500 non-depressive. The dataset includes sorts of tweets:

1. Tweets which might be depressive however do now not contain hash tags #despair.

2. Tweets containing hash tags but not taken into consideration as depressive.

This scarcity in the corpus also enables broaden a higher machine for depression detection. The average period of a tweet is 22.2 tokens in line with tweet. The specified descriptions of this dataset may be seen in table 2.

Table 2: Data Statistics

Class	User Collected Dataset
Depressive	5,529
Non-depressive	6,500
Total	12,029

IV DEPRESSION DETECTION SYSTEM

For the detection of depression in English tweets, we have used a baseline classification system in which word-primarily based functions became used to identify the extent or sort of despair. Further, these functions were discovered through gadget-studying techniques to come across despair.

Pre-processing

It is a standard practice through online media to make use of camel instances while writing hash tags. Along those lines, we extract the hash tags from each tweet and extract separate tokens from it through eliminating the '#' and making use of a hash tag deterioration technique, accepting it is written in camel case. For instance, we will get 'I', 'Am', and 'Depressive' from '#I Am Depressive'. URL's mentions prevent words, and punctuations are taken out from tweets for in addition processing.

Features

Word N-Grams: Word n-gram indicates having or now not having a continuous series of n-word or tokens in a tweet. Word n-grams had been demonstrated to be treasured features for melancholy detection in preceding experiments. We take into account all n-grams for estimations of 'n' from 1 to 5. We keep in mind just the ones n-grams for capabilities that take place as a minimum ten times inside the corpus to prune the feature space.

V CLASSIFICATION APPROACH

We have used seven unique traditional gadgets- getting to know classifiers such as Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbours (KNN), Logistic Regression (LR), Gradient Boosting (GB), and Naive Bayes (NB). We use the sickest-examine implementation of these strategies for melancholy detection.

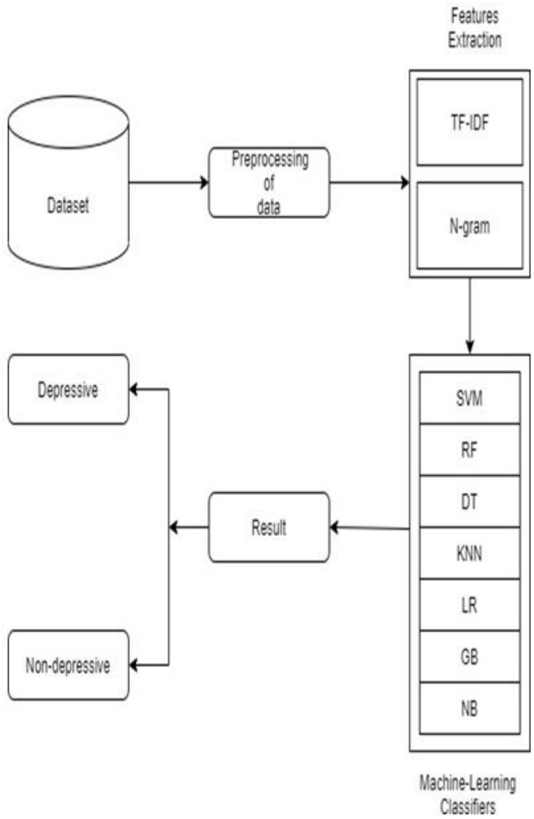

Figure 1: Structure of Depression Detection

VI RESULT

On doing substantial experiments at the gathered dataset, its miles determined that the distinctive traditional classifiers gain suitable performance. The Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbours (KNN), Logistic Regression (LR), Gradient Boosting (GB) and Naive Bayes (NB) achieves an F1-score of zero.89, zero.72, zero.79, 0.62, zero.88, 0.81 and 0.84.

The mathematical equations for the precision, do not forget, and F1-rating can be visible from following equations.

Precision (depressive)=

$$\frac{Number of accurately predicted depressive sentences}{Total number of predicted depressive sentences}$$

Recall (depressive)=

$$\frac{Number of accurately predicted depressive sentences}{Total number of actual depressive statements}$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Table 3: Result of Conventional Machine-Learning Classifiers

		Performance		
Models	Class	Precision	Recall	F1-score
SVM	Depressive	0.86	0.86	0.86
	Non-depressive	0.91	0.91	0.91
	Weighted	0.89	0.89	0.89
RF	Depressive	0.93	0.39	0.55
	Non-depressive	0.71	0.98	0.82
	Weighted	0.80	0.75	0.72
DT	Depressive	0.75	0.72	0.73
	Non-depressive	0.82	0.84	0.83
	Weighted	0.79	0.79	0.79
KNN	Depressive	0.91	0.23	0.37
	Non-depressive	0.66	0.98	0.79
	Weighted	0.76	0.69	0.62
LR	Depressive	0.86	0.84	0.85
	Non-depressive	0.90	0.91	0.90
	Weighted	0.88	0.88	0.88
GB	Depressive	0.87	0.63	0.73
	Non-depressive	0.79	0.94	0.86
	Weighted	0.82	0.82	0.81
NB	Depressive	0.75	0.91	0.82
	Non-depressive	0.93	0.80	0.86
	Weighted	0.86	0.84	0.84

VII CONCLUSION

Identifying depression from the textual contents is tough within the natural language processing region. The performance of the SVM outperforms numerous traditional system-studying classifiers. The cutting-edge research also can be prolonged to include the alternative modalities found in a social media submit, inclusive of snap shots, films, and audio clips. The inclusion of emoji's and other links present in a social media put up can also be demonstrated.

REFERENCES

1. M. M. Tadesse, H. Lin, B. Xu and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," in IEEE Access, vol. 7, pp. 44883-44893, 2019, doi: 10.1109/ACCESS.2019.2909180.

2. A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2017, pp. 138- 140, doi: 10.1109/ICTC.2017.8190959.

3. Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, Johannes C Eichstaedt,

Detecting depression and mental illness on social media: an integrative review, Current Opinion in Behavioural Sciences, Volume 18, 2017, Pages 43-49, ISSN 2352-1546, https://doi.org/10.1016/j.cobeha.2017.07.005.

4. Arun, P. V., M. Krishna, A. B.V., P. S.K. and S. V., "A Boosted Machine Learning Approach For Detection of Depression," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 41-47, doi: 10.1109/SSCI.2018.8628945.

5. M. R. H. Khan, U. S. Afroz, A. K. M. Masum, S. Abujar and S. A. Hossain, "Sentiment Analysis from Bengali Depression Dataset using Machine Learning," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-5, doi: 10.1109/ICCCNT49239.2020.9225511.arXiv:1607.07384v1 **[cs.SI]**

6. S. Jain, S. P. Narayan, R. K. Dewang, U. Bhartiya, N. Meena and V. Kumar, "A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter," 2019 IEEE Students Conference on Engineering and Systems (SCES), Allahabad, India, 2019, pp. 1-6, doi: 10.1109/SCES46477.2019.8977211.

7. N. A. Asad, M. A. Mahmud Pranto, S. Afreen and M. M. Islam, "Depression Detection by Analyzing Social Media Posts of User," 2019 IEEE International Conference on Signal Procesing, Information, Communication & Systems (SPICSCON), Dhaka, Bangladesh,2019, pp.13-doi: 10.1109/SPICSCON48833.2019.9065101.arXiv:1805.11869v1 **[cs.CL]**

8. E. Lunando and A. Purwarianti, "Indonesian social media sentiment analysis with sarcasm detection," 2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Bali, 2013, pp.195-198,doi:10.1109/ICACSIS.2013.6761575.

9. Mondher Bouazizi and TomoakiOhtsuki. 2015. Opinion Mining in Twitter How to Make Use of Sarcasm to Enhance Sentiment Analysis. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015

(ASONAM '15). Association for Computing Machinery, New York, NY, USA, 1594–1597. DOI:https://doi.org/10.1145/2808797.2809350

10. Prasadu Peddi (2018), "A STUDY FOR BIG DATA USING DISSEMINATED FUZZY DECISION TREES", ISSN: 2366-1313, Vol 3, issue 2, pp:46-57.