SUICIDE TEXT-CLASSIFICATION

[1]Shubha Jain, [2]Ajay Gaur, [3]Amit Sabharwal, [4]Mohit Singh, [5]Prabhat Pandey, [6]Vijataswa Awasthi, [7]Waseed Mohamad Khan
Department of Computer Applications, Axis Institute of Higher Education, Kanpur, Uttar Pradesh, India

ABSTRACT
People often share their thoughts and feelings with friends and family on social media, providing valuable insights into their mental state. Recognizing the importance of this, recent studies in NLP and psychology have focused on identifying suicidal thoughts through the analysis of online social network activity. By effectively utilizing Data from social media platforms, coupled with the intricate early indicators of suicidal ideation, can be identified, potentially averting numerous fatalities. This research delves into the application of machine learning and deep learning techniques for the detection of suicide intent across platforms such as Twitter, WhatsApp, and Facebook.. The principal objective is to improve the performance of the model in surpassing the outcomes of preceding studies, thereby more precisely detecting early warning indicators and averting suicide attempts. This is accomplished through the integration of deep learning and machine learning models with feature extraction techniques such as word embedding and count vectorizer.

KEYWORD: NLP (Natural Language Processing), LR (Logistic Regression), SGD (Stochastic Gradient Descent)

I. INTRODUCTION

Suicidality represents a significant public health concern, attributed to an estimated one million fatalities annually. It stands as the primary cause of mortality among young individuals and ranks as the sixth leading cause of death for adults aged between 20 and 59 years. Unsuccessful suicide attempts occur ten to twenty times more frequently than successful ones, often resulting in severe emotional and financial consequences. Furthermore, a study conducted in Belgium revealed that a percentage of 10% among men and 15% among women between the ages of 15 and 24 have reported having suicidal thoughts. Despite these concerning figures, suicide is frequently perceived as a tragedy that can be averted, irrespective of the stage of suicidal ideation the individual may be in.

The advent of "social" Web 2.0 has significantly transformed human communication, allowing people to connect and form online communities. These modifications have impacted the manner in which individuals engage in discussions surrounding suicidal behavior. The realm of online communication offers a semblance of control and confidentiality, which in turn, fosters a higher degree of self-disclosure and reduced inhibition. Social networking platforms have emerged as arenas wherein individuals are afforded the opportunity to articulate their thoughts and emotions pertaining to suicide. While peers can recognize and respond to such behaviors, their responses may be inadequate, delayed, or absent. Therefore, it is beneficial to involve trained website staff to monitor and intervene when necessary, provided.

II. RELATED WORK

Suicide significantly impacts individuals, families, communities, and entire nations. Among adolescents, it ranks as the second-leading cause of mortality, surpassed only by diabetes, liver disease, dementia, and various other diseases. Due to the stigma surrounding mental illnesses, over 40% of people seeking primary care are reluctant to address their depressive symptoms. Immediate medical attention is crucial for suicidal thoughts and behaviors, as there is no reliable method for monitoring, evaluating, or preventing suicide. EHRs contain unstructured data, such as clinical notes, which are challenging to exploit in data mining applications. Another study used natural language processing Natural Language Processing (NLP) methodologies are employed to derive structured outputs from Electronic Health Record (EHR) notes and to integrate these outputs effectively. It employs clinical codes to identify potential associations between medications (such as antidepressants) and psychosocial stressors (such as depression and eating disorders). Machine learning models face limitations, particularly when analyzing exceedingly re-tweeted Twitter conversations, as most tweets are infrequent. Automatic detection of suicidal content on social media has seen limited progress Additionally, Research has investigated the utilization of social media platforms for the prediction of suicide rates at the population-wide level. The findings suggest that social media data holds considerable potential for the conduct of surveillance, thereby facilitating the prediction of suicide rates to the development and verification of prediction models incorporating social media data.

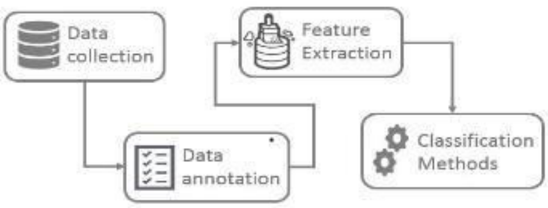
III. PROPOSED METHODOLOGY

To determine whether a user has suicidal tendencies, articles or blogs with suicidal content are categorized. Machine learning has been employed to address this issue. The classification process generally encompasses the extraction of features and the representation of text prior to the application of machine and deep learning models. Figure 6 depicts a prevalent methodology employed in numerous studies discussed within this article.

The initial phase entails the acquisition of data and the subsequent creation of a dataset from one or more social media platforms. In the annotation phase, these datasets are designated through the application of diverse labeling techniques. Subsequently, the third phase, which encompasses feature extraction, is executed prior to the implementation of machine learning and deep learning models..

All the modeling techniques are defined as shown in Figure1

Figure 1:



In their analysis, researchers employed the Ordinary Least Squares (OLS) regression model to examine the linguistic context surrounding the keyword "suicide" and its various forms across different time periods. The findings revealed a distinct pattern: the frequency of the use of keywords related to suicide was notably highest between the hours of 1 and 5 in the morning. Furthermore, the study identified a positive correlation between the number of suicide deaths and age groups, specifically for individuals between the ages of 15 and 44, whereas there was a negative correlation observed for individuals above the age of 45.

Feature Extraction

A multitude of methodologies have been employed to extract features from social media posts, with the objective of discerning whether these posts manifest suicidal ideations. The textual attributes of these posts have been scrutinized through the utilization of TF-IDF matrices, which serve to underscore the importance of specific terms in delineating between posts expressing suicidal thoughts and those that do not. Furthermore, the analysis of N-gram features has been utilized to examine the content of blogs, identify terms within the corpus of the blog, and ascertain the probability of the occurrence of n words within a specific document.

Classification Methods

Researchers have used various machine classification algorithms to examine and analyze social media content for suicide detection. Three main approaches have been the focus of research:

1. Time-Series Analysis: Structuring the problem as a time-series issue to identify changes in user behavior over time.

2. Text Classification (Supervised): Viewing the task through the lens of a text classification problem, aimed at identifying linguistic connotations associated with suicide.

3. Clustering (Unsupervised): Organizing user-generated content into various categories according to their characteristics, addressing the issue via unsupervised clustering techniques.

Temporal Behavior Analysis

The Multi-Layer Perception classifier was employed using the top 1,500 features out of 5,000. The classifier successfully categorized 90.2% of non-risky tweets, with only 9.0% incorrectly labeled. However, it successfully identified only 65.1% of dangerous tweets. In a research endeavor focused on the Japanese language, the term "kietai," which translates to "I wish to disappear," was juxtaposed with instances of suicide to delve deeper into the temporal patterns associated with suicidal expressions.

(i) Attribute Extraction

To assess anyway social web posts indicate melancholic notion, various approach have been employed to bring out relevant aspect. One common approach is analyzing textual attributes through TF-IDF matrices, which measure the importance of label in distinguishing between melancholic and non- melancholic posts [13,1,2,3]. In this context, n-gram features are utilized to examine blog gratified and single out key terms within the blog mass. These features also help estimate the threat of specific word sequences occurring in a given document [9,3,1]. N-grams are particularly useful in tweet sentiment analysis due to Twitter's character limit, which often leads users to use shorter n-grams [15]. Some studies integrate psycholinguistic attribute from LIWC with textual attribute [6]. LIWC not only measures word frequency but also categorizes words according to their syntactical roles (such as nouns, pronouns, and verbs).

(ii) Classification of Methods

To assess anyway social web posts indicate melancholic notion, various approach have been employed to bring out relevant aspect. One common approach is analyzing textual attributes through TF-IDF matrices, which measure the importance of label in distinguishing between melancholic and non- melancholic posts [13,1,2,3]. In this context, n-gram features are utilized to examine blog gratified and single out key terms within the blog mass. These features also help estimate the threat of specific word sequences occurring in a given document [9,3,1]. N-grams are particularly useful in tweet sentiment analysis due to Twitter's character limit, which often leads users to use shorter n-grams [15]. Some studies integrate psycholinguistic attribute from LIWC with textual attribute [6]. LIWC not only measures word frequency but also categorizes words according to their syntactical roles (such as nouns, pronouns, and verbs).

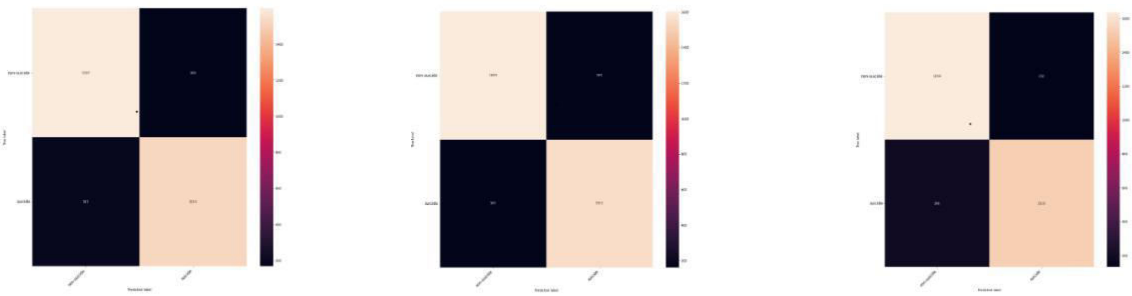
IV. RESULT

The model's Precision, Recall, and F1 scores are evaluated by comparing them to those of added mock-up, using their respective dubiety Matrices. Table 1 displays the outcomes.

Table 1: Results of various models on Test Set

Name	Precision	Recall	F1 Score
LR	0.90	0.89	0.89
SGD Classifier	0.92	0.91	0.91
Soft Voting	0.90	0.90	0.90

Natural Language Processing (NLP) is employed alongside other policy such as text sorting and sorting adapting. Data from various sources is collected, and multiple techniques like Logistic Regression (LR), Soft Voting, and Stochastic Gradient Descent (SGD) Classifier are used to verify the accuracy of the news. After evaluating all observations, the best results were achieved using the SGD Classifier.



Logistic Regression Soft Voting SGD Classifier

Figure 2: Confusion matrix of LR, Soft Voting, SGD Classifier

V. CONCLUSION

Various methods have been proffered to detect fake news, including record mining and social network analysis techniques. In this paper, we present several approaches to verify whether the collected news is authentic or fake. The approach we propose is named.

REFERENCES

[1] Baroni, M., Bernardini, S., 2004. BootCaT : Bootstrapping Corpora and Terms from the Web. In: Proceedings of the 4th International Conference On Language Resources and Evaluation (LREC '04). Lisbon, Portugal, pp.1313–1316.37

[2] Carletta, J., 1996. Assessing Agreement on Classification Tasks: The Kappa Statistic. Computational Linguistics 22 (2), 249–254.

[3] Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector ma-chines. ACM Transactions on Intelligent Systems and Technology 2 (3),1–27.

[4] Cortez, P., Peralta, J., 2014. Global and decomposition evolutionary sup-port vector machine approaches for time series forecasting. Neural Computing and Applications 25(5), 1053–1062.

[5] De Smedt, T., Daelemans, W., 2012. Pattern for Python. Journal of Ma-chine Learning Research 13, 2063–2067.

[6] Desmet, B., Hoste, V., Verstraeten, D., Verhasselt, J., 2013. Gallop Documentation. Tech. rep.

[7] Forman, G., 2003. An Extensive Empirical Study of Feature Selection Metrics for Text Classification. Journal of Machine Learning Research 3, 1289–1305.

[8] Goldberg, D. E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley.38

[9] [12] Goldberg, D. E., Deb, K., 1991. A Comparative Analysis of Selection Schemes used in Genetic Algorithms. In: Foundations of Genetic Algorithms. Vol. 51. Morgan Kaufmann Publishers, San Mateo, CA, USA, pp.69–93.