International Journal of
Mechanical Engineering Research and Technology

ISSN 2454-535X

Email ID: info.ijmert@gmail.com or editor@ijmert.net

Int. J. Mech. Eng. Res. & Tech 2024

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X
http://www.ijmert.com
https://doi.org/10.62651/ijmert.2024.v16.i3.pp206-214
Vol.16 Issue 3, 2024

HYBRID FEATURE BASED PREDICTION OF SUICIDE RELATED ACTIVITY ON TWITTER

[1] Mr.K. Obulesh,[2] M. Akshara,[3] P. Shruthi,[4] P. Guru Priya
[1]Assistant Professor,[234]Students
Department Of CSE
Malla Reddy Engineering College for Women

ABSTRACT

Globally, the number of suicide deaths rises year, which is a troubling public medical concern. Through this experiment, casual, inactive individuals were organically disengaged from online life on Twitter and from sharing self-destructive ideas. Initial emotional evaluation of the idle locations was followed by a thorough comparison with random factors suggested by space experts. With the increasing normalcy of long-distance interpersonal communication platforms, clients have come to rely on these spaces for very sensitive conversations, including thoughts of suicide. The tweets are important for research because of the high frequency of information appearing in them and the considerable capacity and time constraints that computations using them must meet. Emoticons and synonyms features can now be distinguished, and the ngram model—a combination of Unigram, Bigram, and Trigram with half breed word reference—is used to calculate scores. Using machine learning methods, this model uses the informal points to predict how sincere the postings will be. In this study, we also contrast several methodologies such as SVM, NB, and RF.

I. INTRODUCTION

Suicide, the act of demonstrating one's own death, ranks as the tenth leading cause of death in the United States and is estimated to cost $44.6 billion annually. This minimizes the gravity of the problem with regard to every attempted suicide [1]. Self-destructive ideation includes a vast range of thoughts, from stray ideas to extensive planning or insufficient attempts. Receiving it is necessary for overall health due to the severity and impact of this emotional wellness problem [4]. Many individuals often bring up common symptoms while discussing their ideations, such as feeling helpless, alone, very tired, lacking confidence, having the impression that their mind is racing, or placing an inordinate emphasis on uninteresting goals [6].

Knowing the common subjects of self-destructive ideation will help us comprehend the motivations underlying these thoughts, which will ultimately lead to intervention and therapy [12]. Several risk variables have been identified by clinical study aimed at comprehending suicide. Mental illnesses including depression, schizophrenia, alcoholism, and drug addiction all anticipate a significant factor. Furthermore, the emotional strain resulting from emotional abuse, interpersonal relationships, and accounting are all

ISSN 2454 – 535X
http://www.ijmert.com
https://doi.org/10.62651/ijmert.2024.v16.i3.pp206-214
Vol.16 Issue 3, 2024

important factors [14]. Regardless, these portrayals of self-destructive thoughts often capture a clinical viewpoint [19]. As online informal groups have grown in complexity and recognition, individuals who are contemplating suicide have become more comfortable sharing their thoughts of self-destruction via tweets, web forums, and other social media platforms [18]. The result is an incredibly harmonious representation of ideas and motivations around suicide. As of right now, this article shows how to remove informal, inactive themes from this data. This model is an AI method for removing data from a variety of data sources [22]. Using this research, one may find common themes in online messages, such as brutality, medication usage, or discouragement. Various terms that are linked to "sadness" may also be used to describe it, such as "torment", "emotions", "dread", "stress", and "languishing" [19]. The model used in this article analyzes offline and online tweets linked to suicidal thoughts that are gathered from the Twitter API [8].

II. LITERATURE SURVEY

Predicting behavior connected to suicide on social media sites such as Twitter has grown in importance as a field of study for mental health monitoring. To increase the precision and dependability of these forecasts, hybrid feature-based prediction models include many feature types (such as textual, behavioral, and network-based characteristics). The literature review that follows summarizes important methods, conclusions, and difficulties in this field:

1. Introduction Significance: Social media is often used as a venue for people to communicate their sorrow, and suicide is a serious global public health problem. Early interventions may be possible if suicide-related behavior on Twitter is identified.

Conventional Methods: The main methods utilized by early models were sentiment analysis, keyword-based identification, or simple machine learning models that just looked at the content of tweets.

2. Feature-Based Hybrid Models

Definition: To improve prediction accuracy, hybrid feature-based models use many feature types, usually integrating textual data with behavioral and network-based characteristics.

Types of Features:

Textual Features: Topic modeling, sentiment polarity, word embeddings, and n-grams are included. These encapsulate the tone and vocabulary used in tweets.

Behavioral Features: Take into account the regularity of tweets, the time at which they are sent, and any changes in behavior over time.

Network Features: In order to comprehend the social context and impact, include analyzing user activities such as retweets, mentions, and follower-following connections.

3. Important Research and Methods

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
https://doi.org/10.62651/ijmert.2024.v16.i3.pp206-214
Vol.16 Issue 3, 2024

Multi-Modal Approaches: To enhance prediction, studies often use multi-modal data. Combining linguistic characteristics (like LIWC categories) with metadata (like device kind and location) and user behavior (like interaction patterns) is one example.

Deep Learning Integration: To capture temporal connections and contextual subtleties in tweets, recent research has made use of deep learning models like as LSTMs, CNNs, and attention processes.

Models that combine both: To build a more reliable prediction system, several strategies combine many classifiers (such as SVM, Random Forests, and deep learning models) using ensemble techniques.

4. Performance Metrics and Evaluation Metrics: F1-score, ROC-AUC, accuracy, precision, recall, and recall are common metrics used to assess prediction models. Certain studies additionally assess the models' predictive timeliness—that is, their capacity to foresee events before they happen.

Findings: In every case, hybrid models perform better than single-feature or single-model methods. For instance, combining textual analysis with behavioral and network variables often results in noticeably higher recall and accuracy rates for the detection of material associated to suicide.

5. Predictive Data Privacy and Ethics Challenges: The moral use of personal data is one of the main issues. Researchers must follow ethical standards and strike a compromise between the need for precise prediction and the safeguarding of user privacy.

Data Imbalance: Unbalanced datasets result from the comparatively small number of tweets pertaining to suicide as compared to the overall amount of tweets. To solve this, methods including undersampling, oversampling, and creating synthetic data (like SMOTE) are often used.

Contextual Understanding: Words and phrases may take on multiple meanings depending on the context, thus it's important to understand the context of tweets. For instance, sarcasm detection and discerning between remarks made in jest and those with suicidal purpose are persistent problems.

6. Uses and Prospects

Real-Time Monitoring: Using these models for real-time monitoring is gaining traction. They may be integrated with crisis intervention programs to provide at-risk people prompt assistance.

Cross-Platform Analysis: Adding data from several social media platforms (like Facebook and Reddit) to the models might provide researchers a more complete picture of a user's mental health.

tailored Models: Upcoming studies could concentrate on creating tailored models that take into consideration individual variations in social behavior, language use, and mental health background.

7. Impact of Conclusion: A possible method for spotting suicide-related behavior on Twitter is to use hybrid feature-based prediction models, which may help with preventative mental health treatment. Nonetheless, issues with context awareness, real-time deployment, and data ethics must be resolved.

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

https://doi.org/10.62651/ijmert.2024.v16.i3.pp206-214

Prospective Studies: The continuous improvement of these models is expected to improve their accuracy and practicality in real-world scenarios, particularly by including multi-modal data and sophisticated deep learning algorithms.

The promise and complexity of hybrid feature-based prediction models for suicide-related behavior on Twitter are highlighted in this research review. These models will be further developed in the future with the goal of providing more trustworthy resources for mental health therapies.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

- By 2018, Reilly N. Grant, AnaM. León, David Kucher, Daniela S. Raicu, Jonathan F. Gemmell, and Samah J. Fodeh had This article examines a user's social media postings that discuss suicide.Using the clustering method, they examine the post. After cleaning the data from Reddit, it is sent on to word2vec, which creates a word vector. We use a k-mean clustering approach to uncover workouts connected to suicide by aggregating comparative expressions of Reddit content. The goal of this work is to discuss the sociological and technical aspects of content mining research into Reddit material [1].
- Published in 2018 by Bridianne 'Deaa, Melinda R., Philip J. B., Mark E. Larsena, Alison. Calearc, and H. Christensena, this study proposes a stage for online networking. Some individuals have used Twitter to spread harmful thoughts and intentions about themselves. Observe the Twitter behaviour of users in relation to a chemical associated to suicide as compared to non-suicide related substances.
- The dataset was used to examine and analyse replies, retweets, and likes for both suicide-related and non-suicide-related messages. Reaction times to postings mentioning suicide were also significantly lower than those for posts unrelated to suicide. To find the answer's speed, one uses the mean and standard deviation [2]. Currently, Fatima Chiroma, Mihaela Cocea, and Han Liu (2018) assessed the performance of four popular machine classifiers—RF, DT, NB, and SVM—in categorising Tweets pertaining to suicide. Trial results showed that the correlation between suicide and flippant courses had the best execution, with an F-proportion of 0.78.
- It is necessary to examine and examine the display of various machine learning methods with test results in order to enhance the presentation of machine learning processes for categorisation of correspondence connected to suicide. According to another study, we may anticipate a more precise and robust portrayal of group learning in the near future [3].

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
https://doi.org/10.62651/ijmert.2024.v16.i3.pp206-214
Vol.16 Issue 3, 2024

- An article published in 2018 by Jingcheng Du, Jianhong Luo, Yaoyun Zhang, Cui Tao, Qiang Wei, and Hua Xu One of the primary causes of death in the United States, according to this article, was suicide. Suicide may be explained in part by mental health issues. Recognising that persons in danger may experience mental exhaustion might help prevent them from engaging in risky behaviours or taking their own lives. The critical endeavour to eliminate mental exhaustion from Twitter data by means of substantial learning-based strategies. The prevalence of major learning-based systems is shown by connection with typical AI estimates.
- CNN has been leading the pack in identifying tweets connected to suicide with an accuracy of 78.00% and an F1 extent of 83.000%, surpassing SVM, Decision Trees (DT), and others. In comparison to CRF, RNN-based mental fatigue affirmation achieves a better F1 extent of 53.25% accurate equal and 67.94% ambiguous facilitation. In addition, with the same level of care, the move-picking method for starting clinical notes on the Twitter corpus outperforms the planning method using the Twitter corpus alone, with an F1 extent of 54.90%. The results demonstrate the benefits of substantial learning-based approaches for the automated confirmation of online weariness [4].

Disadvantages

- The current implementation of the system does not provide for offline data analysis.
- Because no pre-processing is done on the data sets, this system performs poorly.

PROPOSED SYSTEM

A. Input Data

1) Online: - Through Twitter's Application Programming Interfaces (APIs), a large number of people from all around the globe may automatically access Twitter data. 2) Offline: https://github.com/npanwar/SuicideTweets_Analysis is a website from which we have collected offline tweets.

B. Pre - Processing

One part of data mining systems is data pre-processing, which involves turning raw data into usable information. Genuine information is oftentimes missing, contradictory, and repulsive in its categorical tendencies or practices, and it is also likely to include a number of errors. To address these issues, it is recommended to prepare the necessary information in advance [6, 26]. Data found on Twitter is not organized in a traditional database. It has to be taken care of before it can be used. As a result, the obtained tweets are sifted to remove irrelevant data and remove any inconsistencies that can hinder the determination of the concealed emotion [28, 23, and 24]. Information may be more easily processed in subsequent phases because of this.

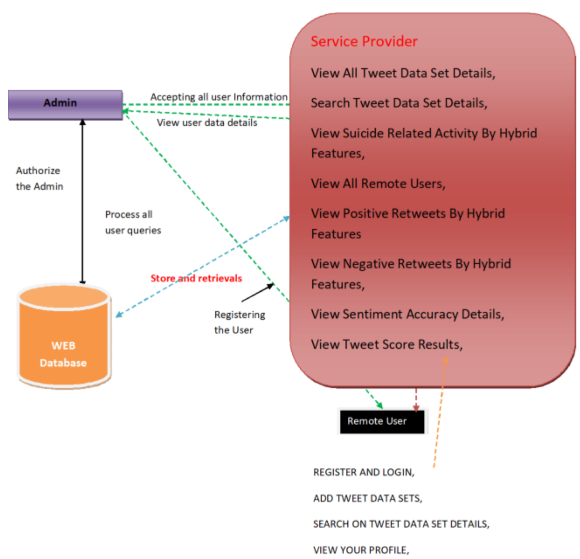
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024
https://doi.org/10.62651/ijmert.2024.v16.i3.pp206-214

The pre-planning system includes the following innovations:
1. Eliminating tweets that do not include English.
secondly, making all of the collected tweets lowercase [32].
The third step is to remove URLs from the tweets. This involves removing any text that represents a link or hyperlink.
4. Using @username instead of any usernames included in tweets, removing the username altogether and not including it into estimates [23].
5. Removing hashes from hash labels and replacing them with regular words is necessary since hash marks may provide some supporting information. As an example, Happy [5] may stand in for #Happy.
6. Getting rid of extra spaces, unnecessary characters, and the like [8].
In order to remove unnecessary terms from the tweets, the Twitter dataset undergoes pre-processing. To increase the accuracy of the examination's results, remove punctuation, stop words, and hash labels [23].

Advantages

- ❖ The algorithm in the suggested study processes both online and offline tweets pertaining to suicidal thoughts that are retrieved from the Twitter API.
- ❖ Because SVM, NB, and RF are present, the system is more effective.

IV. SYSTEM ARCHITECTURE



V. SYSTEM IMPLEMENTATION

Modules

Service Provider

A valid username and password are required for the Service Provider to access this module. The user will be able to do tasks like after a successful login, Explore Every Tweet Data Set, Search Every Tweet Data Set,See All Remote Users,

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
https://doi.org/10.62651/ijmert.2024.v16.i3.pp206-214
Vol.16 Issue 3, 2024

View Activities Relating to Suicide Using Hybrid Features,Check out the details of sentiment accuracy, see tweet scores, view sentiment results by hybrid features, view positive retweets, and view negative retweets.

View and Authorize Users

The admin can get a complete rundown of all registered users in this section. Here, the administrator may see the user's information (name, email, and address) and grant them access.

Remote User

All all, there are n users in this module. Registration is required prior to performing any operations. Details will be entered into the database after a user registers. He will need to log in using the permitted username and password when registration is completed. Upon successful login, users will be able to do activities such as adding Twitter data sets, searching for details about those sets, and seeing their profiles.

VI. RESULT AND ANALYSIS

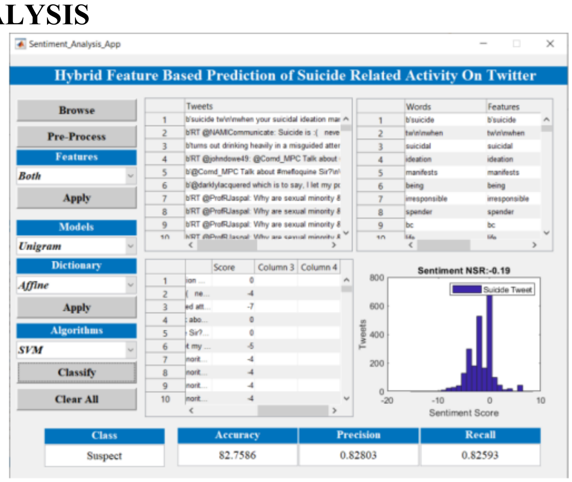
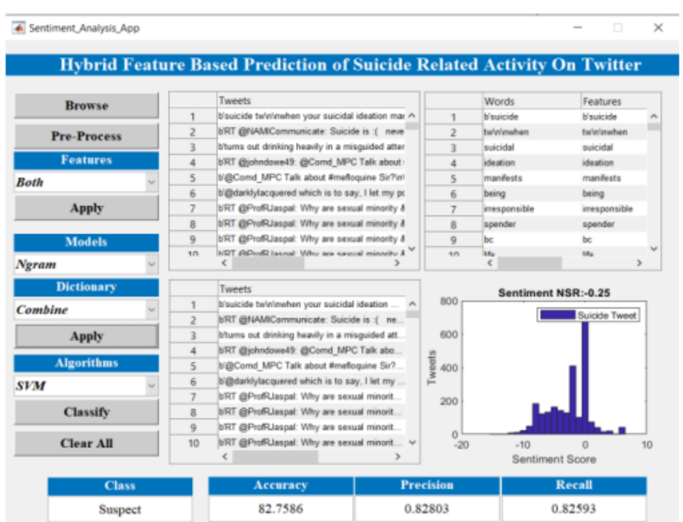


Fig: Unigram+ Affine Dictionary and SVM Classifier

Figure 1 shows the unigram model in action, with the use of an affine dictionary, to compute scores. Lastly, the Net Sentiment Score (NSR) plot is shown in the graphical user interface.



International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X
http://www.ijmert.com
https://doi.org/10.62651/ijmert.2024.v16.i3.pp206-214
Vol.16 Issue 3, 2024

Fig: N-grams+ Combine Dictionary and SVM Classifier

The N-gram model is shown in figure 8, and the scores are calculated using the affine and Lexicon (Combine) dictionaries. The Net Sentiment Score (NSR) plot is then shown in the graphical user interface. Finally, a suspect suicide event categorization using the SVM classifier.

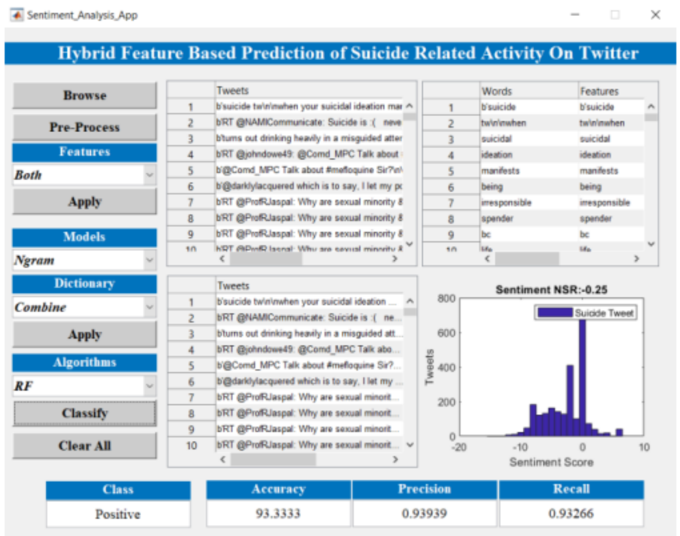


Fig. 9. N-grams+ Combine Dictionary and RF Classifier

Figure 9 shows the results of calculating the N-gram model's score using the affine and Lexicon (Combine) dictionaries. At last, the Net Sentiment Score (NSR) plot is shown in the graphical user interface. Using the RF classifier, the suicide event categorization is finally deemed affirmative.

Table I: Analysis table

Classifier	Unigram+ Affine Dictionary				N-gram+ Combine Dictionary			
	Accuracy	Precision	Recall	NSR	Accuracy	Precision	Recall	NSR
SVM	82.76	82.80	82.59	-0.19	82.76	82.80	82.59	-0.25
RF	93.33	93.94	93.27	-0.19	93.33	93.94	93.27	-0.25

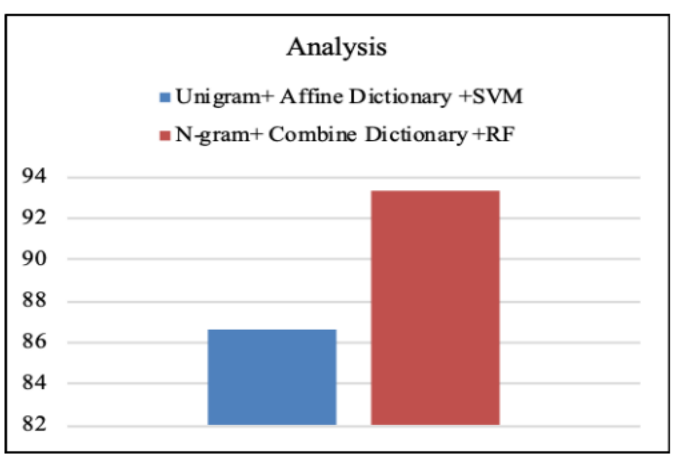


Fig: Analysis Graph

ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024
https://doi.org/10.62651/ijmert.2024.v16.i3.pp206-214

VII. CONCLUSIONS

Therefore, this research did not include sedentary adults who were tweeting about self-harm. After doing an abstract evaluation of the inactive people, we compared them to random components. There was a 93.33 percent success rate for the Random Forest Classifier while using the N-gram feature. Predicting the posts' urgency is another application of the proposed approaches. In the future, we want to broaden our studies to include a wider range of emotional wellness issues, such as major depressive disorder and severe stress.

REFERENCES

1. C. Singh and K. Ant ony, “ Frauds in t he Indian Banking Indust ry”,IIMB-2016
2. R. N. Grant , D. Kucher ,A. M. León,J. F. Gemmell, D. S. Raicu,and S. J. Fodeh, “ Aut omatic extraction of informal t opics from online suicidal ideat ion,” vol. 19, no. Suppl 8, 2018.
3. B. O. Dea, M. R. Achilles, M. E. Larsen,A. L. Calear, H. Christ ensen,and P . J. Bat t erham, “ T he rat e of reply and nat ure of responses to suicide-related posts on Twit ter,” Internet Interv ., vol. 13, no. July, pp. 105–107, 2018.
4. U. Kingdom and U. Kingdom, “ T ext classificat ion for suicide relat ed tweet s.”
5. J. Du et al., “ Ext ract ing psychiat ric st ressors for suicide from social media using deep learning,” vo l. 18, no. Suppl 2, 2018.
6. S. Ji, C. P . Yu, S. Fung, S. P an, and G. Long, “ Supervised Learning for Suicidal Ideat ion Detection in Online User Content ,” vol. 2018, 2018.
7. P. Burnap, G. Colombo, R. Amery, A. Hodorog, and J. Scourfield, “Mult i-class machine classification of suicide-related communication on Twitter,” Online Soc.Networks Media, vol. 2, pp. 32 –44, 2017.
8. B. O. Dea, S. Wan, P. J. Bat terham, A. L. Calear, C. Paris, and H. Christ ensen, “Detecting suicidalit y on T witter,” INVENT, vol. 2, no. 2, pp. 183–188, 2015.
9. G. B. Colombo, P . Burnap, A. Hodorog, and J. Scourfield, “ Analysing t he connect ivit y and communicat ion of suicidal users on twit t er,” Comput . Commun., 2015.
10. G. Coppersmit h, M. Dredze, and C. Harman, “ Quant ifying Ment al Health Signals in Twitter.”