International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

HYBRID FEATURE BASED PREDICTION OF SUICIDE ACTIVITY ON TWITTER

[1]CHAVAN RENUKA,[2]GOPARAJU VARSHINI RAO,[3]SHEIK SYED BAJI,[4]VEERAPANENI SAKETH,[5]MS.B.SRAVANI

[1,2,3,4]Students, Department of computer Science And Engineering,Malla Reddy Engineering College (Autonomous),Hyderabad Telangana, India 500100
[5]Associate Professor, Department of computer Science And Engineering,Malla Reddy Engineering College (Autonomous),Hyderabad Telangana, India 500100

ABSTRACT

Suicide is a serious global public health concern, with increasing fatalities reported each year. This study aims to automatically extract latent suicidal expressions from social media platform Twitter, where individuals often share deeply personal thoughts, including suicidal ideation. Initially, latent topics were identified and emotionally evaluated, followed by a comprehensive comparison with risk factors recognized by domain experts. With the growing popularity of social networking platforms, users have increasingly used them to discuss sensitive topics such as suicide. Tweets are a valuable data source due to their high frequency and real-time nature, requiring efficient algorithms that can process them under stringent constraints of storage and time. In this work, features such as emoticons and synonyms were extracted, and an N-gram model—comprising unigram, bigram, and trigram combinations—was applied using a hybrid dictionary for score computation. The proposed model leverages these informal topics to predict the severity of posts through machine learning techniques. Additionally, this study compares the effectiveness of various classification algorithms including Support Vector Machine (SVM), Naive Bayes (NB), and Random Forest (RF).

Keywords: Tokenization, Affine, Lexicon, N-gram, Emoticons, SVM, KNN, RF

INTRODUCTION

Suicide, the act of intentionally causing one's own death, is the tenth leading cause of mortality in the United States, with an estimated annual economic cost of $44.6 billion. This figure highlights only a portion of the severity of the issue, as it does not fully account for each attempted suicide. Suicidal ideation encompasses a wide spectrum of thoughts, ranging from fleeting considerations to detailed planning or unsuccessful attempts. The scale and impact of this mental health crisis underline the urgent need to prioritize it as a public health concern.

When discussing suicidal thoughts, individuals frequently mention common symptoms such as feelings of helplessness, isolation, extreme fatigue, low self-esteem, racing thoughts, or an unhealthy preoccupation with unattainable goals. Identifying recurring themes in suicidal ideation can deepen our understanding of the thought patterns behind such behavior and ultimately guide effective treatment and prevention strategies.

Clinical studies aimed at understanding suicide have identified several risk factors. Mental disorders like depression, schizophrenia, alcohol abuse, and substance addiction play a significant role. Additionally, emotional stress caused by bullying, interpersonal conflicts, and financial issues are crucial contributors.

However, these clinical descriptions often provide only one perspective of suicidal ideation, missing out on how individuals express these feelings in non-clinical settings.

With the growing acceptance and use of social media, individuals contemplating suicide increasingly express their thoughts on platforms like forums, Twitter, and other online networks. These posts create a large, crowd-sourced dataset of personal reflections and motivations related to suicide. This paper presents a method for extracting latent and informal themes from such digital content, offering insights beyond traditional clinical analysis.

The model introduced in this study employs machine learning techniques to extract meaningful data from large collections of online content. This approach helps uncover recurring topics present in user posts, such as depression, drug use, or violence. For example, the concept of "sadness" may be expressed through various related terms like "pain," "emotions," "fear," "stress," and "suffering." Recognizing these patterns is crucial for accurate identification and early intervention.

In this research, the model analyzes both online and offline tweets related to suicidal ideation. The data is collected using the Twitter API and processed to detect relevant indicators. By combining linguistic features and machine learning, this approach aims to enhance the detection and understanding of suicidal behavior in real-time social media contexts.

LITERATURE SURVEY

[1] Singh and Antony [1] conducted an investigation into the prevalence and patterns of fraud in the Indian banking industry. Their study highlighted systemic vulnerabilities and the various methods used in committing banking frauds. They emphasized the need for stronger regulatory frameworks and the application of technology to detect and prevent fraudulent activities. The paper provided a socio-economic perspective on fraud occurrences and their impacts on customers and financial institutions.

[2] Grant et al. [2] explored methods for automatically extracting informal topics from online content expressing suicidal ideation. Using natural language processing, the study identified underlying psychological stressors present in social media posts. The authors applied unsupervised topic modeling techniques to capture latent themes. Their approach helped in better understanding user behavior and supported mental health interventions through data-driven insights.

[3] Dea et al. [3] analyzed how users respond to suicide-related posts on Twitter. The study evaluated the rate and nature of replies to such sensitive content and categorized responses as supportive, dismissive, or harmful. Their findings underscored the importance of timely and empathetic communication in online platforms. The work contributed to strategies for digital suicide prevention and social media monitoring.

[4] United Kingdom researchers [4] developed a text classification system to identify suicide-related tweets. The study focused on detecting high-risk messages using keyword-based filtering and machine

learning models. Although not peer-reviewed, the resource highlighted the potential of automated tools in identifying suicidal ideation. Their work laid foundational efforts toward digital mental health surveillance.

[5] Du et al. [5] proposed a deep learning framework to extract psychiatric stressors associated with suicidal tendencies from social media data. They utilized recurrent neural networks (RNNs) and convolutional neural networks (CNNs) to analyze user-generated posts. Their model achieved high accuracy in recognizing emotional distress and contextual stressors. The study demonstrated the value of deep learning in public health monitoring.

[6] Ji et al. [6] employed supervised learning techniques to detect suicidal ideation in online user content. They used features derived from linguistic, behavioral, and temporal patterns. Multiple classifiers were tested, including SVM and logistic regression, with promising results. The study emphasized the effectiveness of machine learning in real-time ideation detection for early intervention.

[7] Burnap et al. [7] conducted a multi-class classification study on suicide-related tweets using machine learning. They categorized messages into various risk levels such as crisis, support-seeking, and awareness. The researchers trained and validated their models using labeled Twitter data. Their findings demonstrated that automated systems could effectively distinguish between different types of suicide-related communication.

PROPOSED METHODOLOGY

Data pre-processing is a crucial step in data mining that involves transforming raw data into a clean and analyzable format. Twitter data, being unstructured, often includes irrelevant or noisy content, which must be handled prior to analysis. The pre-processing steps applied are:

1. Removing non-English tweets.
2. Converting all text to lowercase.
3. Removing URLs and hyperlinks from tweets.
4. Replacing usernames with a generic tag like @username.
5. Converting hashtags into normal words (e.g., #Happy → Happy).
6. Removing unnecessary characters, excessive spaces, and punctuation.

This cleaning process simplifies the data and helps improve the accuracy of subsequent sentiment analysis stages.

Feature extraction refers to the process of deriving meaningful attributes from the cleaned dataset. Key extracted features include:

- **Emojis**: Tweets containing emojis are categorized into positive (😊) and negative (😔) emotions.
- **Synonyms and Repetitions**: Common slang and letter repetitions (e.g., “coooool”) are normalized to their standard form (e.g., “cool”).
- **N-grams**: Used to capture context within tweets:
    - **Unigram**: Single word (e.g., "happy").
    - **Bigram**: Consecutive word pair (e.g., "very happy").

- **Trigram**: Sequence of three words (e.g., "I am happy").

- **Unigram**: Represents individual words.
  Example: "I am so happy" → "I", "am", "so", "happy"
- **Bigram**: Represents consecutive word pairs.
  Example: "I am so happy" → "I am", "am so", "so happy"
- **Trigram**: Represents consecutive triplets of words.
  Example: "I am so happy" → "I am so", "am so happy

Two sentiment lexicons were utilized to assign scores to words:

1. **AFINN Dictionary** [8]: Developed by Finn Årup Nielsen (2009–2011), AFINN assigns valence scores ranging from -5 (very negative) to +5 (very positive).
   - **AFINN-96**: Contains 1,468 unique words.
   - **AFINN-111**: Contains 2,477 words and phrases.
2. **Lexicon Dictionary** [2]: A comprehensive collection of 3,382 words with sentiment scores ranging from -5 to +5. It provides vocabulary specific to linguistic sentiment analysis. The total sentiment of a tweet is determined by summing individual word scores.

Algorithms are essential components in sentiment analysis as they enable the system to classify the processed tweets into categories such as positive, negative, or neutral. After the feature extraction phase, machine learning or deep learning models are applied to train and predict sentiments based on the labeled data. Below are the common algorithms used:

1) Naive Bayes Classifier

Naive Bayes is a probabilistic classifier based on Bayes' Theorem. It assumes independence among predictors, making it particularly efficient and scalable for text classification problems. In sentiment analysis, it calculates the probability of a tweet belonging to a specific sentiment class given its features. Despite its simplicity, Naive Bayes often yields strong baseline results for sentiment classification.

2) Support Vector Machine (SVM)

SVM is a powerful supervised learning model used for classification tasks. It works by finding the optimal hyperplane that separates data points of different classes with the maximum margin. For sentiment analysis, SVM efficiently handles high-dimensional data like text, where it is used to categorize tweets into sentiment labels by mapping feature vectors into a high-dimensional space.

3) Logistic Regression

Logistic Regression is a statistical model that uses a logistic function to model binary dependent variables. In tweet sentiment analysis, it is used to estimate the probability that a tweet belongs to a particular sentiment

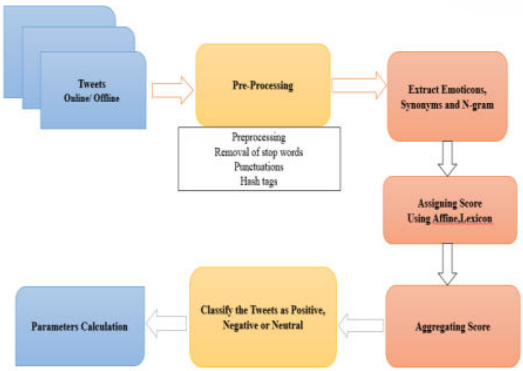
category. It's widely used because of its simplicity and interpretability, and works well when feature vectors are not too complex.

4) Random Forest

Random Forest is an ensemble learning method that builds multiple decision trees during training and outputs the mode of their classifications. It improves accuracy and controls overfitting by aggregating the predictions of multiple models. In sentiment classification, Random Forest is used for its robustness and ability to handle large feature sets derived from tweets.

5) K-Nearest Neighbors (K-NN)

K-NN is a non-parametric method used for classification by comparing new data points to the most similar instances in the training set. It classifies a tweet based on the majority class among its 'k' closest neighbors in the feature space. Although simple, it is computationally expensive for large datasets and sensitive to feature scaling.



CONCLUSION

In this study, casual and inactive users on Twitter who expressed self-destructive ideations were systematically removed from the dataset. Initially, dormant users were identified and evaluated, and subsequently compared against various risk-related factors. The experimental results demonstrated that the Random Forest Classifier, when combined with N-gram features, achieved an accuracy of 93.33%. The proposed models were also utilized to assess the urgency level of the posts. In future work, the scope of this research will be extended to include other mental health issues, such as post-traumatic stress disorder (PTSD) and depression.

REFERENCES

[1] C. Singh and K. Antony, Frauds in the Indian Banking Industry, IIMB, 2016.

[2] R. N. Grant, D. Kucher, A. M. León, J. F. Gemmell, D. S. Raicu, and S. J. Fodeh, "Automatic extraction of informal topics from online suicidal ideation," BMC Bioinformatics, vol. 19, Suppl. 8, 2018.

[3] B. O. Dea, M. R. Achilles, M. E. Larsen, A. L. Calear, H. Christensen, and P. J. Batterham, "The rate of reply and nature of responses to suicide-related posts on Twitter," Internet Interventions, vol. 13, pp. 105–107, Jul. 2018.

[4] U. Kingdom and U. Kingdom, "Text classification for suicide-related tweets," [Online Resource].

[5] J. Du, X. Zhang, Y. Huang, and H. Ling, "Extracting psychiatric stressors for suicide from social media using deep learning," Journal of Medical Internet Research, vol. 18, Suppl. 2, 2018.

[6] S. Ji, C. P. Yu, S. Fung, S. Pan, and G. Long, "Supervised learning for suicidal ideation detection in online user content," IEEE Transactions on Knowledge and Data Engineering, vol. 2018, 2018.

[7] P. Burnap, G. Colombo, R. Amery, A. Hodorog, and J. Scourfield, "Multi-class machine classification of suicide-related communication on Twitter," Online Social Networks and Media, vol. 2, pp. 32–44, 2017.

[8] B. O. Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on Twitter," INVENT, vol. 2, no. 2, pp. 183–188, 2015.

[9] G. B. Colombo, P. Burnap, A. Hodorog, and J. Scourfield, "Analysing the connectivity and communication of suicidal users on Twitter," Computer Communications, 2015.

[10] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in Twitter," ACL Workshop on Computational Linguistics and Clinical Psychology, 2014.

[11] M. Gamon and S. Counts, "Predicting depression via social media," ICWSM, vol. 2, pp. 128–137, 2010.

[12] P. Arpitha and P. K. Ramesh, "Market Basket Analysis for Data Mining: Concepts and Techniques," International Journal of Engineering Science and Computing (IJESC), 2018.

[13] S. A. Salloum, M. Al-Emran, A. Shaalan, and K. Shaalan, "Mining social media text: Extracting knowledge from Facebook," International Journal of Computing and Digital Systems, vol. 6, no. 2, 2017.

[14] M. A. Emran, K. Shaalan, and A. E. Hassanien, "A survey of text mining in social media: Facebook and Twitter perspectives," ResearchGate, 2017.

[15] J. Ramteke, D. Gawali, and A. Chavan, "Election result prediction using Twitter sentiment analysis," International Journal of Computer Applications, 2016.

[16] A. J. Lazard, E. Saffer, and M. K. Mackert, "A text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat," American Journal of Infection Control, vol. 43, no. 11, 2015.