ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 11, Issue.1 Feb 2023

HYBRID FEATURE BASED PREDICTION OF SUICIDE RELATED ACTIVITY ON TWITTER

Mrs N.Bhargavi[1],A.Pooja[2],G.Srija[3],K. Bhoomika[4],R.Manisha[5]

[1]Assistant professor, Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

[2,3,4,5]UG Students,Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

Abstract :

Suicide remains a concerning global health issue, with fatalities on the rise annually worldwide. This study focuses on identifying and analyzing suicidal ideation expressed on the online platform Twitter. Initially, we filtered out non-contributing users and identified potentially relevant tweets. We then conducted a comprehensive comparison of these tweets with risk factors outlined by domain experts. With the widespread adoption of social media platforms, users have increasingly utilized them to discuss highly personal topics, including thoughts of suicide. The abundance of data on these platforms presents challenges in terms of processing efficiency and resource constraints. To address this, we implemented a feature extraction approach involving emoticons and synonyms, along with an n-gram model combining Unigram, Bigram, and Trigram with a hybrid dictionary for scoring. Leveraging these techniques, our model predicts the severity of posts containing suicidal ideation using machine learning algorithms. Furthermore, we conducted comparative analyses involving support vector machines (SVM), Naive Bayes (NB), and Random Forests (RF) to evaluate the effectiveness of our approach.

I. INTRODUCTION

Suicide-related activity expressed on social media platforms, particularly Twitter, has garnered significant attention due to its potential implications for mental health interventions and suicide prevention efforts. The pervasive nature of social media enables individuals to openly share their thoughts, emotions, and experiences, including those related to suicidal ideation. With millions of users engaging on Twitter daily, the platform serves as a rich source of real-time data that can offer valuable insights into

mental health trends and public sentiment surrounding suicide. Detecting and understanding suicide-related activity on Twitter presents numerous challenges, including the sheer volume of data, the diversity of language used, and the nuanced nature of suicidal expressions. To address these challenges, researchers have increasingly turned to computational methods, particularly machine learning and natural language processing techniques, to analyze and identify patterns indicative of suicidal ideation in tweets. However, existing approaches often face limitations in accurately capturing the complexity and context of suicide-related discussions.

In this study, we propose a novel approach for the prediction of suicide-related activity on Twitter, leveraging a hybrid feature-based methodology. Our approach integrates a variety of features extracted from tweets, including linguistic, syntactic, semantic, and contextual cues, to enhance the predictive performance of our model. By combining multiple types of features, we aim to capture the multidimensional nature of suicide-related discourse on Twitter and improve the accuracy of our predictions.

Furthermore, our research seeks to contribute to the broader field of computational suicide prevention by developing a more robust and effective methodology for identifying individuals at risk of suicide based on their online activity. By harnessing the power of machine learning and natural language processing, we strive to create a proactive approach to suicide prevention that leverages the vast amount of data available on social media platforms like Twitter. Through our hybrid feature-based prediction model, we aim to provide valuable insights that can inform targeted interventions, support systems, and public health initiatives aimed at reducing the incidence of suicide.

II.EXISTING SYSTEM

The existing systems for predicting suicide-related activity on Twitter often rely on simplistic approaches that may overlook important nuances in language and context. These systems typically employ basic keyword-based filters or rule-based algorithms to flag tweets containing specific terms associated with suicidal ideation. However, such approaches tend to suffer from limited accuracy and effectiveness, as they may fail to capture subtle variations in

language or account for the broader context in which tweets are posted. Additionally, existing systems may struggle to differentiate between genuine expressions of distress and unrelated content, leading to false positives and negatives in their predictions. Moreover, these systems may lack scalability and flexibility, making them ill-suited for handling the dynamic and evolving nature of social media data.

III.PROPOSED SYSTEM

In contrast, our proposed system for the hybrid feature-based prediction of suicide-related activity on Twitter offers several advantages over existing approaches. By integrating a diverse set of features extracted from tweets, including linguistic, syntactic, semantic, and contextual cues, our system enhances its predictive capabilities and accuracy. This holistic approach allows our model to capture the multidimensional nature of suicide-related discourse on Twitter more effectively, enabling it to identify patterns and signals indicative of suicidal ideation with greater precision. Additionally, our system is designed to be adaptable and scalable, allowing it to handle large volumes of data and adapt to changes in language and user behavior over time. Moreover, our approach leverages advanced machine learning techniques to continuously learn and improve its performance, ensuring that it remains up-to-date and effective in detecting suicide-related activity on Twitter. Overall, our proposed system represents a significant advancement in the field of computational suicide prevention, offering a more nuanced and accurate approach to identifying individuals at risk of suicide based on their online activity.

IV.MODULES

Data Collection Module:

- This module is responsible for collecting tweets from Twitter's API or other sources.
- It may involve filtering and preprocessing the data to remove noise and irrelevant information.

Feature Extraction Module:

- This module extracts various features from the collected tweets, including linguistic, syntactic, semantic, and contextual features.

- Features may include sentiment analysis, word embeddings, syntactic parse trees, topic modeling, and user metadata.

Machine Learning Module:

- This module trains and evaluates machine learning models on the extracted features to predict suicide-related activity.
- Various algorithms such as Support Vector Machines (SVM), Naive Bayes (NB), Random Forest (RF), and neural networks may be used.

Model Evaluation Module:

- This module assesses the performance of the trained models using evaluation metrics such as accuracy, precision, recall, and F1-score.
- It may involve cross-validation techniques to ensure robustness and generalization of the models.

Integration and Deployment Module:

- Once the model is trained and validated, this module integrates it into a deployable system.
- It may involve developing APIs or interfaces for accessing the prediction system and deploying it on a server or cloud platform.

Monitoring and Maintenance Module:

- This module monitors the performance of the deployed system in real-time and handles any issues or errors that arise.
- It may involve periodic updates and retraining of the model to adapt to changing trends and patterns in social media data.

V.CONCLUSION

In conclusion, the "Hybrid Feature-Based Prediction of Suicide-Related Activity on Twitter" project addresses the critical need for early detection and intervention of suicide-related content on social media platforms. Through the integration of advanced machine learning techniques and feature extraction methods, the project aims to accurately identify and classify tweets containing suicidal ideation or related activity. By leveraging linguistic, syntactic, semantic, and contextual features extracted from Twitter data, the developed models exhibit promising performance in predicting and flagging potentially concerning content. The project underscores the importance of

harnessing technology to proactively identify and support individuals at risk of suicide, ultimately contributing to suicide prevention efforts in the digital age.

VI.FUTURE SCOPE

Looking ahead, there are several avenues for further enhancement and expansion of the project. Firstly, continuous refinement and optimization of the machine learning algorithms and feature extraction techniques can improve the accuracy and efficiency of suicide-related activity prediction. Additionally, incorporating real-time data streaming and sentiment analysis methods can enable more timely and dynamic detection of suicidal content. Furthermore, extending the scope of the project to include other social media platforms beyond Twitter, such as Facebook, Instagram, and Reddit, can provide a more comprehensive understanding of online suicide-related behavior. Collaborations with mental health professionals and organizations can facilitate the development of intervention strategies and support systems tailored to individuals identified through the predictive models. Overall, the future scope of the project involves ongoing research and innovation to enhance the capabilities of technology in combating suicide and promoting mental well-being in the digital realm.

VII.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", vol. 19, no. 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 Interv.*, vol. 13, pp. 105-107, July 2018.

4.U. Kingdom and U. Kingdom, "Text classification for suicide related tweets".

5.J. Du et al., "Extracting psychiatric stressors for suicide from social media using deep learning", vol. 18, no. 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", 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 Soc.Networks 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", *Comput. Commun.*, 2015.

10.G. Coppersmith, M. Dredze and C. Harman, "Quantifying Mental Health Signals in Twitter".

11.M. Gamon and S. Counts, "Predicting Depression via Social Media", vol. 2, pp. 128-137.

12.P. Arpitha, "Market Basket Analysis for Data Mining: Concepts and T echniques", *IJESC*, 2018.

13.Said A. Salloum and M. A.-E., "Mining Social Media Text: Extracting Knowledge from Facebook", *International Journal of Computing and Digital Systems*, 2017.

14.A Mostafa and K. S. Emran, "A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives", *ResearchGate.*, 2017.

15.D. G. Jyoti Ramteke, *Election Result Prediction Using Twitter sentiment Analysis*, 2016.

16.J. Allison and E. S. Lazard, "A text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat", *American Journal of Infection Control*, 2015.

17.J. K. Jasmina Smailovic, "Monitoring the Twitter sentiment during the Bulgarian elections" in , IEEE, 2015.

18.Gaurav Dudeja, "Sentiment Analysis of Mobile Reviews using Sentiwordnet", *Journal of Basic and Applied Engineering Research.*, 2015.

19.P. P. Alexander Pak, "Twitter As A Corpus For Sentiment Analysis And Opinion Mining", *C. Paper Preprocessing techniques for text mining J. Emerg. Technol. Web Intell.*, 2010.

20.*List of Text Emoticons*, July 2012.

21.P.F. Brown et al., "Class-based n-gram models of natural language", *Computational Linguistics*, vol. 18, no. 4, pp. 467-479, 1992.

22.J.C.B. Christopher, "A tutorial on support vector machines for pattern recognition", *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.

23.E.-H. Sam Han, G. Karypis and V. Kumar, Text Categorization Using

Weight Adjusted k-nearest Neighbor Classification, Springer, 2001.

24.I.B. Mohamad and D. Usman, "Standardization and its effects on K-means clustering algorithm", *Research Journal of Applied Sciences Engineering and Technology*, vol. 6, no. 17, pp. 3299-3303, 2013.

25.C. Silva and B. Ribeiro, "The importance of stop word removal on recall values in text categorization", *Neural Networks 2003 Proceedings of the International Joint Conference on*, vol. 3, no. 2003, pp. 1661-1666.

26.D.A. Hull et al., "Stemming algorithms: A case study for detailed evaluation", *JASIS*, vol. 47, no. 1, pp. 70-84, 1996.

27.Bruno Trstenjak, Sasa Mikac and Dzenana Donko, "KNN with TF-IDF based Framework for Text Categorization", *Procedia Engineering*, vol. 69, pp. 1356-1364, 2014.

28.M. James, Classification Algorithms, Wiley-Inter science, 1985.

29.H. Isozaki and H. Kazawa, "Efficient support vector classifiers for named entity recognition", *Proceedings of the 19th international Conference on Computational linguistics*, vol. 1, 2002.

30.L.S. Larkey and W.B. Croft, "Combining classifiers in text categorization", *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 289-297, 1996.