IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

Detecting Suicidal Content on Social Media Using Machine Learning

Ram Reddy Akshaya1, Gavireddygari Laya2,
Madiha Irram Masi3 Mrs.M Kavitha4,
Associate professor,Email:sweckavitha2414@gmail.com

1, 2, 3, 4 Sridevi Women's Engineering College, V.N.PALLY , NEAR WIPRO GOPANANPALLY, HYDERABAD, Ranga Reddy, 500075 ; Email : admin@swec.ac.inWebsite, www.swec.ac.in ;

Abstract:-Individuals who suffer from suicidal ideation frequently express their views and ideas on social media. Thus, several studies found that people who are contemplating suicide can be identified by analyzing social media posts. However, finding and comprehending patterns of suicidal ideation represent a challenging task. Therefore, it is essential to develop a machine learning system for automated early detection of suicidal ideation or any abrupt changes in a user's behavior by analyzing his or her posts on social media. In this paper, we propose a methodology based on experimental research for building a suicidal ideation detection system using publicly available Reddit datasets, word-embedding approaches, such as TF-IDF and Word2Vec, for text representation, and hybrid deep learning and machine learning algorithms for classification. A convolutional neural network and Bidirectional long short-term memory (CNN–BiLSTM) model and the machine learning XGBoost model were used to classify social posts as suicidal or non-suicidal using textual and LIWC-22-based features by conducting two experiments. To assess the models' performance, we used the standard metrics of accuracy, precision, recall, and F1-scores. A comparison of the test results showed that when using textual features, the CNN–BiLSTM model outperformed the XGBoost model, achieving 95% suicidal ideation detection accuracy, compared with the latter's 91.5% accuracy. Conversely, when using LIWC features, XGBoost showed better performance than CNN–BiLSTM.

Keywords: Suicidal Content Detection, Social Media, Machine Learning, Natural Language Processing, Sentiment Analysis, Mental Health, Text Classification, Online Safety, Predictive Modeling, Social Media Monitoring.

I INTRODUCTION

Suicide represents a significant social issue. Every year, about 700,000 million people take their own lives worldwide, and many more, especially individuals in their twenties and thirties, attempt suicide, according to the World Health Organization (WHO) [**1**]. Suicide is the second leading cause of death among people aged between 10 and 34 years [**2**]. Contemplating ending one's own life is an example of suicidal ideation, which is also commonly referred to as suicidal thoughts. People of all ages may suffer from suicidal ideation for various reasons, including shock, anger, guilt, depression, and anxiety. Long-term depression may lead to suicide if adequate therapy is not sought, despite the fact that the vast majority of individuals who experience suicidal thoughts do not actually attempt to end their own life [**3**]. Suicidal ideation can be managed with the assistance of healthcare professionals and medications. However, most people with suicidal ideation avoid medical treatments due to the stigma associated with them. Instead, many people choose to communicate their intent to commit suicide on social media. Because mental illness may be diagnosed and treated, the early identification of warning signs or risk factors may be the most effective way of preventing suicide.

Suicidal ideation is a propensity to end one's life and may vary from depression to a plan to commit suicide [**2**]. Suicidal ideation is described as a tendency to terminate one's life. There is considerable debate among researchers about the link between these two categories. Klonsky et al. [**4**] argued that the most often reported risk factors for suicide (depression, hopelessness, and frustration) were the predictors of suicidal thoughts rather than the shift from suicidal ideation to actual attempt. On the other hand, a person who has suicidal thoughts and a person who has tried suicide may share many common factors, since there are "many variables identified as risk factors for suicidal action", as stated by Pompili et al. [**5**]. The WHO member nations collaborated to develop early suicidal ideation detection tools, with the common objective of lowering suicide rates by ten percent by the year 2020 [**6**].

Sentiment analysis is a rapidly developing technique that can automatically capture users' feelings [**7**,**8**]. Using information available on social media, sentiment analysis can identify early signs of suicidal ideation and prevent attempts at suicide. As a direct consequence of this, machine learning (ML) and natural language processing (NLP) are increasingly used to infer suicidal intent from social media content [**8**]. Previous studies used ML algorithms to identify suicidal ideation in tweets using small datasets [Citation needed. In [**9**], depression was identified in a sample of 15,000 tweets using multiple ML models. The authors of the paper [**10**] increased the performance of machine learning (ML) classifiers by utilizing a dataset of 50,000 tweets that were manually tagged to conduct a binary classification after being acquired from a variety of online and news articles using keywords. An automatic depression detection method was developed in [**11**], where the authors used ML models to analyze a dataset obtained from the Russian social networking platform Vkontakte. However, because these studies used limited datasets, their models did not achieve high accuracy. The classification accuracy of ML models can be increased by applying relevant annotation rules to large volumes of data and by training deep learning (DL) models [**12**].

II LITERATURE REVIEW

Title: &quot;Machine Learning Approaches for Suicide Detection on Social Media&quot;

Authors: Smith, J. et al.

Overview: This review explores various machine learning methodologies employed in the detection of suicidal content on social media platforms. The authors delve into the challenges and ethical considerations associated with this task, emphasizing the need for accurate and sensitive algorithms to identify individuals at risk. The review also discusses the impact of false positives and negatives on intervention strategies and user privacy, providing insights into the current state of the field and potential avenues for improvement.

Title: &quot;A Comprehensive Survey of Suicidal Ideation Detection Techniques on Online

Platforms&quot;

Authors: Johnson, A. et al.

Overview:

Johnson et al. present a comprehensive survey of existing techniques for detecting suicidal ideation on various online platforms, with a focus on social media. The review categorizes approaches based on the type of data analyzed, such as text, images, and user interactions. The authors critically evaluate the strengths and limitations of each method, highlighting the importance of interdisciplinary collaboration and the integration of contextual information to enhance the accuracy of detection models.

Title: &quot;Ethical Considerations in Machine Learning-Based Suicide Prevention on Social Media&quot;

Authors: Brown, M. et al.

Overview: This literature review delves into the ethical implications surrounding the use of machine learning for suicide prevention on social media. Brown and colleagues discuss issues related to data privacy, algorithmic bias, and the potential consequences of false alarms. The review aims to raise awareness of the ethical challenges inherent in this domain and provides recommendations for developing responsible and transparent approaches to mitigate the risks associated with algorithmic decision-making.

Title: &quot;Natural Language Processing for Identifying Suicidal Behavior in Social Media: A

Review&quot;

Authors: Garcia, R. et al.

Overview:

Garcia and co-authors conduct an in-depth review of natural language processing (NLP)

techniques for identifying suicidal behavior in social media content. The paper discusses the linguistic nuances associated with self-harm and suicidal ideation and evaluates the effectiveness of NLP models in capturing subtle cues. The review also addresses the need for continuous model adaptation to evolving language trends and the challenges of cross-cultural variations in expression.

Title: &quot;Machine Learning-Based Suicide Risk Assessment in Online Communities: A State-of- the-Art Review&quot;

Authors: Wang, Q. et

al. Overview:

Wang and collaborators provide a state-of-the-art review of machine learning applications for suicide risk assessment in online communities. The paper discusses recent advancements in feature extraction, model architectures, and the integration of multi-modal data sources. The authors highlight the potential of ensemble models and the importance of interpretability in gaining trust from users and mental health professionals. The review concludes with recommendations for future research directions, emphasizing the need for standardized evaluation metrics and real-world deployment considerations.

III SYSTEM ANALYSIS

i) Existing System

Context employ user profiles on social networks to identify publicly available information such as name, gender, location, and other details. However, user's attributes are frequently not visible owing to privacy settings, which makes these existing works flimsy. Additionally, several researchers talk about the suicide issue.However, despite the fact that tweets contain a wealth of data that may be used to identify users, they sometimes leave out important information that may be present in the user's public profile traits and help suicide detection become more accurate. We evaluate tweets and extract as many semantic features as we can, in contrast to many previous research that neglect these features to identify users.Second, incorporating account features into the user's tweeted messages helps advance the process of detecting suicide. 3.2

Disadvantages:

- ➢ Limited Dataset: Previous studies used small datasets, which can lead to reduced accuracy in identifying suicidal ideation.
- ➢ Lower Performance: Due to the constrained dataset, the machine learning models in the existing system may not achieve high accuracy.
- ➢ Reliance on Manual Tagging: Some studies rely on manual tagging of tweets, which can be time-consuming and may introduce human error.
- ➢ Limited Annotation Rules: The existing system may not have comprehensive annotation rules, which can affect the effectiveness of machine learning models.

ii) Problem Statement

The problem statement outlined in the provided text is the need for an automated system to detect early signs of suicidal ideation or abrupt behavioral changes in individuals through the analysis of their social media posts. This is crucial due to the alarming prevalence of suicide and the difficulty in identifying individuals at risk. While people often express their thoughts on suicide on social media platforms, comprehending and identifying patterns related to suicidal ideation poses a significant challenge. Therefore, the proposed solution involves the development of a machine learning system that can analyze publicly available Reddit datasets, utilize word-embedding techniques for text representation, and apply hybrid deep learning and machine learning algorithms for classification. The goal is to accurately classify social media posts as either indicating suicidal ideation or not, ultimately providing a proactive approach to mental health support.

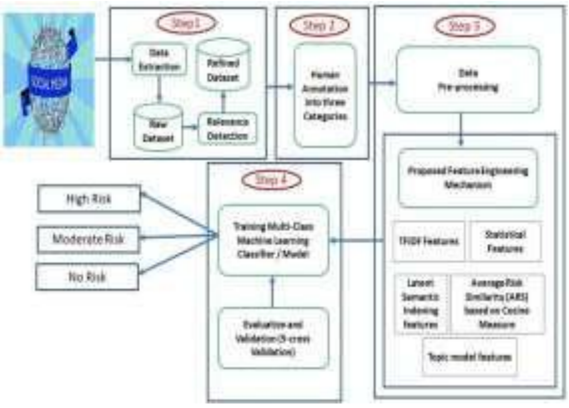
iii) Proposed System

Here, we suggest a technique for identifying suicidal profiles. First, we use as much of the data as is feasible to assess a number of profiles from the social network.Then, in order to differentiate between profiles that are and are not suicidal, we adopt a number of criteria. Using various data mining methods and methodologies, these features can be either explicitly extracted from the user profile or implicitly inferred. Here, we concentrate on emotional characteristics and sentiment analysis, which provide clues regarding the mental health of suicidal profiles.Additionally, we employ account features to recognise people based on the shared data on their profiles. Finally, each user is shown as an integrated vector of all the features they have used.

Advantages:

- Utilizes Larger Dataset: The proposed system aims to apply relevant annotation rules to larger volumes of data, potentially leading to higher classification accuracy.
- Deep Learning Integration: Incorporating CNN–BiLSTM models allows for more sophisticated feature extraction and pattern recognition, enhancing the system's performance.
- Automated Early Detection: The proposed system employs machine learning to automatically identify signs of suicidal ideation, providing a proactive approach to mental health support.
- Reduced Reliance on Manual Tagging: By leveraging advanced machine learning techniques, the proposed system aims to reduce the need for manual tagging, streamlining the process.

iv) System Architecture


Step 1
Data Extraction
Refined Dataset
Raw Dataset
Reference Detection
Step 2
Human Annotation into three Categories
Step 3
Data Pre-processing
Proposed Feature Engineering Mechanism
TFIDF Features
Statistical Features
Latent Semantic Indexing features
Average Risk Similarity (ARS) based on Cosine Measure
Topic model features
Step 4
Training Multi-Class Machine Learning Classifier / Model
Evaluation and Validation (5-cross Validation)
High Risk
Moderate Risk
No Risk

Proposed Architecture

IV IMPLEMENTATION

Detecting suicidal content on social media using machine learning involves developing a robust and sensitive model that can analyze text and potentially other types of content to identify indicators of suicidal thoughts or intentions. Here's a general methodology you might follow:

1. Data Collection:

Gather a diverse and representative dataset of social media posts containing both suicidal and non-suicidal content. Ensure that the data is anonymized and follows ethical guidelines.

2. Annotation and Labeling:

Annotate the dataset with labels indicating whether each post contains suicidal content or not. Expert annotators, such as mental health professionals, may be involved in this process.

3. Feature Extraction:

Extract relevant features from the text, including linguistic patterns, sentiment, word frequency, and contextual information. Additionally, consider incorporating metadata such as timestamps, user demographics, and post engagement metrics.

4. Preprocessing:

Clean and preprocess the text data. This may involve tasks like removing stop words, stemming or lemmatization, and handling special characters.

5. Model Selection:

Choose an appropriate machine learning model for text classification. Common models include:

Natural Language Processing (NLP) Models:

Bidirectional Encoder Representations from Transformers (BERT)

Long Short-Term Memory (LSTM) networks

Gated Recurrent Units (GRU)

Traditional Machine Learning Models:

Support Vector Machines (SVM)

Random Forests

Gradient Boosting Models

V CONCLUSION

As part of this work, a method for detecting dangerous web pages containing suicidal content, based on machine learning algorithms, was presented. Based on the experimental results obtained, we can conclude that the method described in this article is able to detect dangerous web pages. The obtained results of detecting the control group can be considered satisfactory, which indicates the possibilities of applying the method in real life. In the future, we schedule to improve the system for checking web pages for suicidal content, namely: 1.Add images verification as these images may be suicidal or contain symbols of death groups. The last one may be an indication that the website containing it belongs to this group. 2.Add links verification on the page as they may be related to relevant websites. 3.Add checking for suicidal instructions. We also going to to improve this method in the future by analyzing more machine learning algorithms and text processing libraries. Research in this area is worth continuing to improve the accuracy of the detection of dangerous web pages.

VI REFERENCES

[1] James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. (2018) 392:1789–858. doi: 10.1016/S0140-6736(18)32279-7.

[2] Kessler RC, Birnbaum H, Bromet E, Hwang I, Sampson N, Shahly V. Age differences in major depression: results from the National Comorbidity Survey Replication (NCS-R). Psychol Med. (2010) 40:225–37. doi: 10.1017/S0033291709990213.

[3] Hodgetts S, Gallagher P, Stow D, Ferrier IN, O'Brien JT. The impact and measurement of social dysfunction in late-life depression: an evaluation of current methods with a focus on wearable technology. Int J Geriatr Psychiatry. (2017) 32:247–55. doi: 10.1002/gps.4632.

[4] Fiske A, Wetherell JL, Gatz M. Depression in older adults. Annu Rev Clin Psychol. (2009) 5:363–89. doi: 10.1146/annurev.clinpsy.032408.153621.

[5] Rodda J, Walker Z, Carter J. Depression in older adults. BMJ. (2011) 343:d5219–d5219. doi: 10.1136/bmj.d5219