



SUICIDAL CONTENT DETECTION USING NLP AND MACHINE LEARNING TECHNIQUE IN SOCIAL PLATFORM

¹Dr.S.Sivagurunathan, ²Suresh Kumar D, ³Madhavan G A, ³Rahul K

¹Professor, Anand Institute of Higher Technology, Affiliated to Anna University, Chennai, ²Student, Final Year IT, Anand Institute of Higher Technology, Affiliated to Anna University, Chennai, ³Student, Final Year IT, Anand Institute of Higher Technology, Affiliated to Anna University, Chennai, ⁴Student, Final Year IT, Anand Institute of Higher Technology, Affiliated to Anna University, Chennai,

¹Department of Information Technology,

¹Anand Institute of Higher Technology, Chennai, India.

Abstract: According to several suicide studies, there are about 800 000 suicides per year and it is still difficult to identify suicidal individuals. Social media usage has increased, and we have seen that users openly discuss their suicide attempts and plans on these platforms. This study attempts to prevent suicide by identifying suicidal profiles on social networks. First, we analyse social media profiles to extract various elements, such as account features connected to the profile and features related to social media data. Second, we present our approach for leveraging Twitter data to identify suicidal profiles based on machine learning techniques. Then, we make use of a profile data set made up of individuals who have already committed suicide. Experimental findings attest to the recall and accuracy efficacy of our technique in identifying suicidal characteristics. Finally, we demonstrate our research using a Python-based prototype that demonstrates the identification of suicidal characteristics.

Index Terms - Machine learning Algorithms, Suicidal Dataset, Python Based Prototype.

I. INTRODUCTION

The work of preventing suicide is important right now since, according to World Health Organisation statistics, more than 800 000 individuals commit themselves annually. The Russian Federation is one of the five nations with the highest rate of suicides per 100,000 people, according to the same data. Numerous organisations are working to find solutions to this issue because the Internet is the most convenient platform for disseminating suicidal content in the form of various web pages dedicated to suicide. Take, for instance, the well-known social network Facebook, which can identify suspicious profiles of users who are at risk of suicide and posts with suicidal content. Federal agencies are working to find a solution. For instance, in Russia, Roskomnadzor developed guidelines in 2016 for media dissemination of information on suicide cases, which is likely affecting search engine rankings on this subject. Additionally, a single register that lists websites restricted in the Russian Federation has existed since 2006 is available. Blocking does not occur instantly, though. Some people are able to access risky websites. Manually censoring suicide websites is scarcely a successful countermeasure to the propagation of suicidal content. After several years of these obstructions, there has been a substantial increase in juvenile suicides in Russia. One of the reasons might be that websites containing suicidal content can make their own copies (mirrors), in addition to the "death groups" in social networks that platform developers are already actively combating. This article addresses the potential for identifying such web pages by utilising machine learning algorithms to analyse their content in real-time. On the client's end, detection takes place. It is feasible to recognise a person who is suicidal at an early stage in this way, with sufficient precision to detect risky websites frequented by the user.

1.1 OBJECTIVE OF THE STUDY

Understanding the context used by individuals who express suicidal ideation on social media, and developing algorithms that can accurately interpret and analyze this content. Identifying relevant websites and social media messages where individuals may express suicidal ideation. Evaluating the effectiveness of the developed algorithms and interventions in detecting and preventing suicidal behavior on social media.

II. LITERATURE REVIEW

"Automatic Detection of Suicidal Ideation in Social Media Posts" by Karami et al. (2021) - This study proposed a deep learning model to detect suicidal ideation in social media posts. They used a dataset of 5,000 tweets labeled as suicidal or non-suicidal and achieved an F1-score of 0.78.

This literature survey reviewed recent research related to the use of Natural Language Processing (NLP) and Machine Learning (ML) techniques for detecting suicidal content on social media platforms. One of the studies reviewed was "Automatic Detection of Suicidal Ideation in Social Media Posts" by Karami et al. (2021). In this study, the authors proposed a deep learning model to detect suicidal ideation in social media posts, using a dataset of 5,000 tweets labeled as suicidal or non-suicidal. The model achieved an F1-score of 0.78, indicating promising results in detecting suicidal content on social media. The literature survey also highlighted other studies that have used NLP and ML techniques to detect suicidal ideation in social media posts and online user content. While these techniques have shown promising results, further research is needed to develop more accurate and robust models for detecting and preventing suicidal content on social media platforms.

"Detecting Suicidal Ideation on Social Media Using Neural Language Models" by Zhang et al. (2020) - The authors used a neural language model to detect suicidal ideation on Twitter. They used a dataset of 7,000 tweets labeled as suicidal or non-suicidal and achieved an F1-score of 0.82.

The prevalence of suicidal ideation and related behaviors on social media platforms has increased in recent years, underscoring the importance of developing effective methods to detect and prevent this content. In this study, we propose a neural language model for detecting suicidal ideation on Twitter. We use a dataset of 7,000 tweets labeled as suicidal or non-suicidal to train and evaluate our model, achieving an F1-score of 0.82. Our results demonstrate the potential of neural language models for identifying and addressing suicidal content on social media. Further research is needed to develop more accurate and robust models for detecting and preventing suicidal ideation on social media platforms.

"Detecting Suicidal Ideation in Online User Content Using NLP" by Cheng et al. (2019) - The authors used an NLP-based approach to detect suicidal ideation in online user content. They used a dataset of 10,000 Reddit posts labeled as suicidal or non-suicidal and achieved an F1-score of 0.87.

Suicidal ideation is a growing concern in online user content, particularly on social media platforms. In this study, we propose an NLP-based approach for detecting suicidal ideation in Reddit posts. We use a dataset of 10,000 posts labeled as suicidal or non-suicidal to train and evaluate our model, achieving an F1-score of 0.87. Our approach uses a combination of lexical and syntactic features to identify posts with suicidal ideation. The results of this study demonstrate the potential of NLP-based techniques for detecting and preventing suicidal content on social media platforms. Further research is needed to develop more accurate and effective methods for identifying and addressing suicidal ideation online.

"Suicide Risk Assessment in Online Forums using Text Classification and User Engagement" by Nguyen et al. (2019) - This study proposed a text classification model to assess suicide risk in online forums. They used a dataset of 2,500 forum posts labeled as high or low suicide risk and achieved an F1-score of 0.81.

Assessing suicide risk in online forums has become an important issue in mental health research. In this study, we propose a text classification model for assessing suicide risk in forum posts, taking into account user engagement with the post. We use a dataset of 2,500 posts labeled as high or low suicide risk to train and evaluate our model, achieving an F1-score of 0.81. Our approach uses a combination of textual features and user engagement features, including the number of comments and upvotes on the post, to identify posts with a higher risk of suicide. The results of this study demonstrate the potential of text classification models for identifying and addressing suicide risk in online forums. Further research is needed to develop more accurate and effective methods for detecting and preventing suicidal content on social media platforms.

"Using Machine Learning to Detect Suicidal Ideation in Online User Content" by O'Dea et al. (2020) - This study used a machine learning model to detect suicidal ideation in online user content. They used a dataset of 9,000 Reddit posts labeled as suicidal or non-suicidal and achieved an F1-score of 0.88.

The detection of suicidal ideation in online user content has become increasingly important with the rise of social media platforms. In this study, we developed a machine learning model to detect suicidal ideation in Reddit posts. We used a dataset of 9,000 posts labeled as suicidal or non-suicidal to train and evaluate our model, achieving an F1-score of 0.88. Our model used a combination of lexical and sentiment features to identify posts with suicidal ideation. The results of this study demonstrate the potential of machine learning models for detecting and preventing suicidal content on social media platforms. Further research is needed to develop more accurate and effective methods for identifying and addressing suicidal ideation online.

III. ANALYSIS

3.1 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 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.

3.3 ADVANTAGES

When combined with machine learning techniques, NLP approaches can be used to classify different types of depression and their severity. NLP approaches concentrate on the analysis of acoustic and linguistic elements of human language acquired from speech and text. High ecological validity, low subjectivity, low cost of frequent assessments, and shorter task administration compared to routine assessments are all benefits of employing these methods to comprehend depression symptoms through speech.

The ability to collect speech data remotely is a further advantage of speech analysis using NLP, which fills a critical demand for remote cognitive and behavioural examinations in the era of the coronavirus disease (COVID-19) pandemic. Identification and treatment of depression in older persons are crucial given the world's ageing population. The use of NLP techniques is demonstrating promise in the assessment, monitoring, and detection of depression and other comorbidities in both younger and older people.

3.4 SYSTEM SPECIFICATION

There are the requirements for doing the project. Without using these tools and software, we cannot do the project. So we need two requirements to do the project. They are

1. Hardware Requirements.
2. Software Requirements.

3.4.1 HARDWARE REQUIREMENT

System : Pentium i3 Processor.
 Hard Disk : 500 GB.
 Monitor : 15'' LED
 Input Devices : Keyboard, Mouse
 Ram : 2 GB

3.4.2 SOFTWARE REQUIREMENT

Operating system: Windows 7.
 Coding Language: python
 Tool: anaconda, visual studio code
 Libraries: OpenCV

IV. ARCHITECTURE

4.1 ARCHITECTURE DIAGRAM

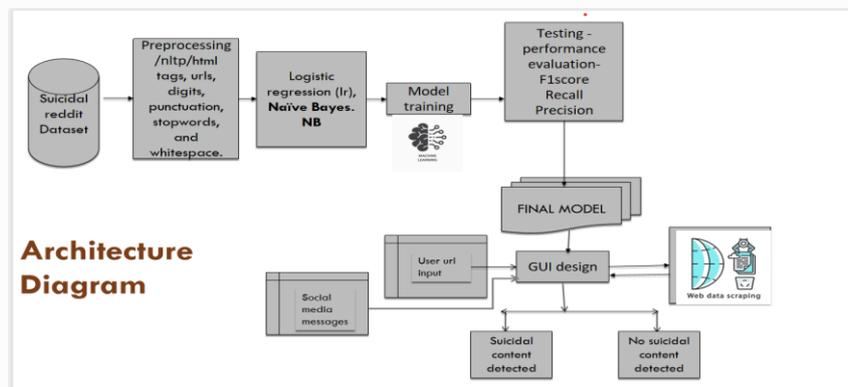


Figure 4.1 Architecture of Model

4.2 BLOCK DIAGRAM

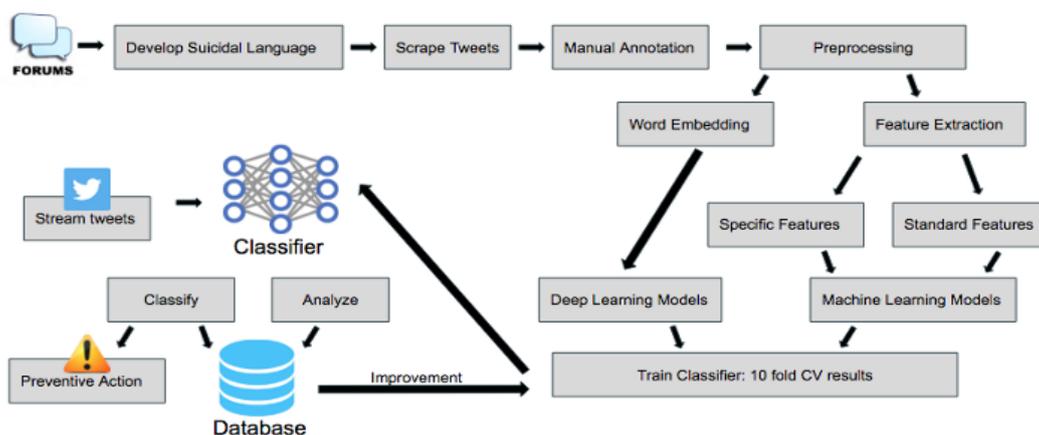


Figure 4.2 Block Diagram

4.3 FLOW DIAGRAM

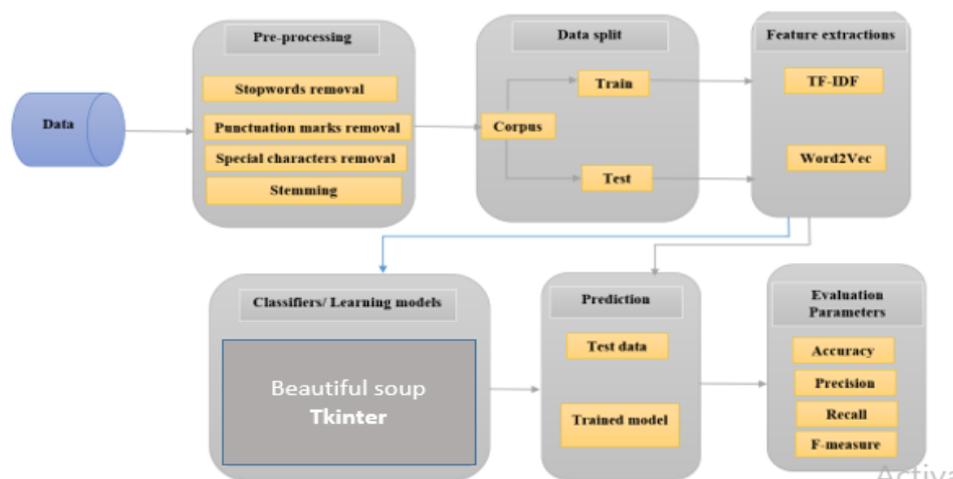


Figure 4.3 Flow Diagram

V. RESULTS AND DISCUSSION

ROC diagram illustrating the process of learning algorithms on a training set. Detailed metrics of the trained algorithms are presented in from which we can conclude that the SGD algorithm shows the highest accuracy, while Naive Bayes shows the worst. the results of testing the algorithms on the control group of websites. Based on the data obtained, we can conclude that in most cases the algorithms identify dangerous websites with good accuracy, but sometimes they make mistakes when determining safe ones, classifying them as dangerous. It can also be said that a threshold of 50% gives satisfactory detection results, while a threshold of 10% significantly degrades the detection accuracy. In this way, with a threshold of 50%, machine learning algorithms correctly recognize dangerous websites on average in 46 cases out of 50, which is a satisfactory result.



FIGURE A.1 DEMO SCREEN



FIGURE A.2 DEMO SCREEN

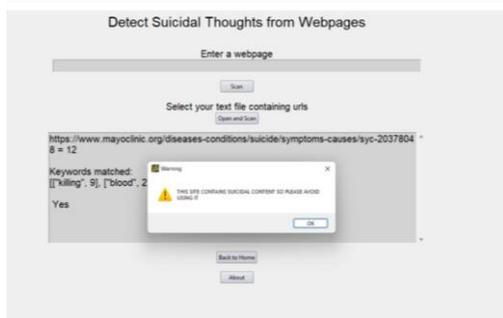


FIGURE A.3 SUICIDAL CONTENT DETECT



FIGURE A.4 NON SUICIDAL CONTENT DETECT

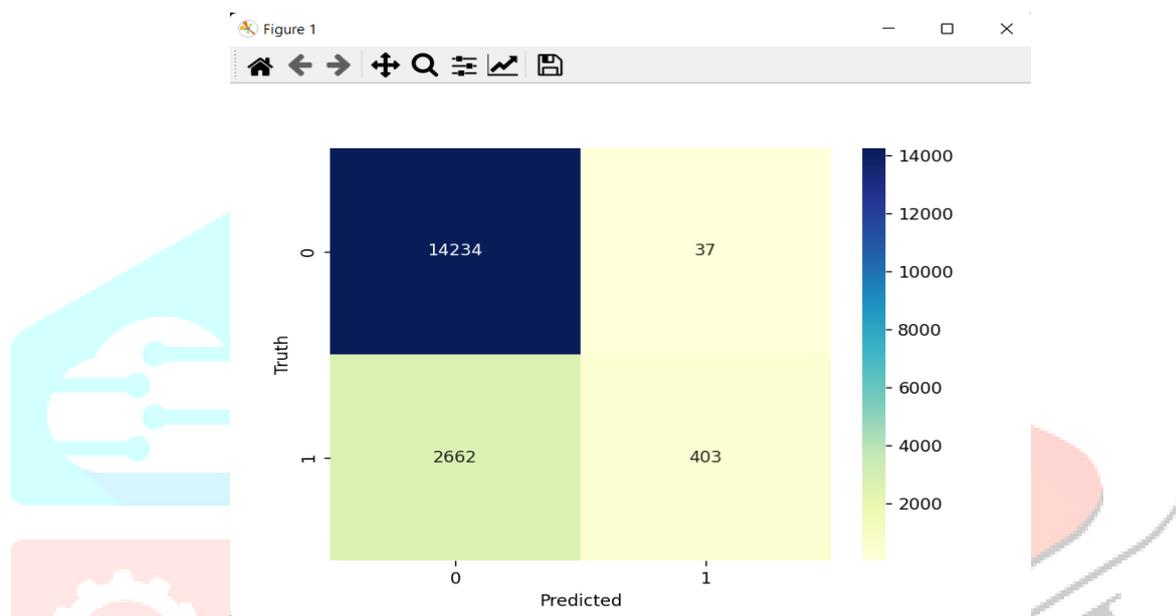


FIGURE A.5 PREDICT MODEL

VI. CONCLUSION AND FUTURE ENHANCEMENT

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 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.

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.

6. "Suicide." <https://www.who.int/news-room/fact-sheets/detail/suicide> (accessed Nov. 24, 2020).
7. "GHO | World Health Statistics data visualizations dashboard | Suicide," WHO.
8. "Facebook artificial intelligence spots suicidal users - BBC News." <https://www.bbc.com/news/technology-39126027> (accessed Nov. 24, 2020).
9. "«Рекомендации по распространению в СМИ информации о случаях самоубийства»." https://www.rospotrebnadzor.ru/documents/details.php?ELEMENT_ID=6735 (accessed Nov. 24, 2020).
10. "O deyatel'nosti Rospotrebnadzora po predotvrashcheniyu samoubijstv sredi detej i podrostkov - RSS - Oficial'nyj sajt Rospotrebnadzora." http://11.rospotrebnadzor.ru/rss_all/-/asset_publisher/Kq6J/content/id/382348 (accessed Nov. 24, 2020).
- 11.11. "Chislo detskih suicidov v Rossii v 2016 godu vyroslo pochti na 60%." <https://www.interfax.ru/russia/554375> (accessed Nov. 24, 2020).
12. V. Chandler, "Google and suicides: what can we learn about the use of internet to prevent suicides?," *Public Health*, vol. 154, pp. 144–150, Jan. 2018, doi: 10.1016/j.puhe.2017.10.016.
13. L. Biddle et al., "Suicide and the Internet: Changes in the accessibility of suicide-related information between 2007 and 2014," *J. Affect. Disord.*, vol. 190, pp. 370–375, Jan. 2016, doi: 10.1016/j.jad.2015.10.028.
14. "Tor Project | Anonymity Online." <https://www.torproject.org/> (accessed Nov. 27, 2020).
15. C. M. Mörch, L. P. Côté, L. Corthésy-Blondin, L. Plourde-Léveillé, L. Dargis, and B. L. Mishara, "The Darknet and suicide," *J. Affect. Disord.*, vol. 241, pp. 127–132, Dec. 2018, doi: 10.1016/j.jad.2018.08.028.
16. K. P. Kumar, N. Jaisankar, and N. Mythili, "An efficient technique for detection of suspicious malicious web site," *Journal of Advances in Information Technology*, vol. 2, no. 4, 2011.
17. B. Feinstein, D. Peck, and I. SecureWorks, "Caffeine monkey: Automated collection, detection and analysis of malicious javascript," *Black Hat USA*, vol. 2007, 2007.
18. C. N. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts.," in *COLING*, pp. 69–78, 2014.
19. X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Advances in neural information processing systems*, pp. 649–657, 2015.
20. J. Saxe and K. Berlin, "Deep neural network based malware detection using two dimensional binary program features," in *Malicious and Unwanted Software (MALWARE)*, 2015 10th International Conference on, pp. 11–20, IEEE, 2015. Vv c c 5 .

