



DETECTING DANGEROUS WEBPAGES BASED ON THE ANALYSIS OF SUICIDAL CONTENT USING MACHINE LEARNING ALGORITHM

AISHWARYA S¹, Mr. J. JAYAPANDIAN²

Krishnasamy College Of Engineering and Technology, Cuddalore¹

Asso.Prof, Krishnasamy College Of Engineering and Technology, Cuddalore²

Abstract: Suicide is a significant global issue, with approximately 800,000 people taking their own lives every year. Detecting individuals at risk of suicide remains a challenging task, as highlighted in numerous suicide studies. However, with the widespread use of social media platforms, we have observed that individuals often express their suicidal thoughts or experiences in public on these networks. Therefore, it is crucial to identify people who may be prone to suicide at an early stage.

In this paper presents a novel approach for detecting suicidal content in social media platforms using natural language processing (NLP) and machine learning techniques. The proposed method combines keyword-based detection, sentiment analysis, and NLP-based approaches to identify posts that may indicate suicidal ideation. By analyzing the language and sentiment used in social media posts, the method aims to identify content that suggests a person may be at risk of suicide. The method is developed by training it on a carefully curated dataset of labeled posts, which includes examples of both suicidal and non-suicidal content. Through rigorous evaluation using metrics such as precision, recall, and F1-score, the effectiveness of the proposed method is assessed. The results demonstrate that the method achieves a high level of accuracy in identifying suicidal content.

The implications of this research are significant. Social media platforms can incorporate the proposed method as an automated tool to flag potentially concerning content for further review. Trained mental health professionals can then examine the flagged posts and provide appropriate support and intervention to individuals in need. By leveraging this technology, timely interventions can be initiated to prevent suicides and offer assistance to those who may be at risk.

Overall, the proposed method offers a promising solution to the challenge of detecting suicidal content in social media platforms. By leveraging NLP and machine learning techniques, it provides a proactive approach to identify individuals who may be in distress, enabling timely intervention and potentially saving lives.

I. INTRODUCTION

Suicide is a complex and pressing issue that affects individuals worldwide, and social media platforms have become an important source of information for identifying individuals at risk. Detecting suicidal content in social media data requires advanced techniques such as natural language processing (NLP) and machine learning (ML) to analyze the vast amounts of textual data generated by users. Researchers and experts in the field have made significant progress in developing algorithms that can automatically detect and flag posts containing suicidal ideation or related content. These algorithms utilize various NLP and ML techniques, such as sentiment analysis, topic modeling, and pattern recognition, to identify patterns and signals indicative of suicidal thoughts. One key aspect of these algorithms is the ability to distinguish between general discussions about mental health and actual expressions of suicidal intent. This requires careful analysis of the language used, the context in which it is expressed, and other contextual cues such as images or emojis. Machine learning models are trained on labeled datasets that include examples of both suicidal and non-suicidal content to learn and generalize patterns effectively. The detection of suicidal content in social media data has the potential to be a valuable tool in suicide prevention efforts. Social media platforms can leverage these algorithms to identify at-risk individuals and intervene by providing resources, support hotlines, or connecting them with mental health professionals. Early detection and timely intervention can make a significant difference in saving lives and providing individuals with the help they need. However, it is important to approach the detection of suicidal content with caution and sensitivity. Privacy concerns and ethical considerations must be carefully addressed to ensure the responsible use of user data. Additionally, human



oversight and involvement are crucial in the process to validate algorithmic predictions and provide appropriate support and intervention. In conclusion, the development of NLP and ML algorithms for detecting suicidal content in social media data is a significant step forward in suicide prevention efforts. By leveraging these technologies, we can identify individuals at risk, offer timely support, and ultimately save lives. Continued research, collaboration between experts, and responsible implementation are key to further advancing this important field.

II. LITERATURE SURVEY

1. "Automated Detection of Suicidal Ideation in Web-Based Text Through Machine Learning: A Systematic Review" conducted by Coppersmith et al. in 2018 focused on examining previous research that utilized machine learning methods to automatically detect suicidal ideation in web-based text. The study aimed to shed light on the different machine learning techniques employed for this purpose and emphasized the significance of identifying suicidal content online to prevent suicide and enhance mental health.

The authors conducted a thorough review of existing literature in this domain and found that several machine learning algorithms had been applied, including support vector machines, logistic regression, and deep learning. These techniques were used to analyze web-based text, such as social media posts, forum discussions, and online chat conversations, in order to identify signs of suicidal ideation.

By employing machine learning models, researchers were able to develop automated systems that could effectively detect patterns and linguistic cues indicative of suicidal thoughts. The study emphasized the importance of accurate and timely identification of individuals at risk, as it enables intervention and support to be provided promptly.

Overall, this systematic review highlighted the advancements made in utilizing machine learning for the automated detection of suicidal ideation in web-based text. The findings emphasized the potential of these techniques to contribute to suicide prevention efforts and improve mental health outcomes by enabling early identification and intervention for individuals in distress.

2. "Automatic Detection of Suicide-Related Posts in Twitter Data: An Application of Machine Learning," conducted by Aladağ et al. in 2020, aimed to investigate the effectiveness of machine learning techniques in identifying suicide-related posts on Twitter. The study focused on utilizing classifiers, including support vector machines, decision trees, and random forests, to categorize tweets as either suicidal or non-suicidal.

By leveraging these machine learning algorithms, the researchers were able to develop a system capable of automatically detecting suicide-related content in tweets. The classifiers were trained on a labeled dataset consisting of tweets that exhibited signs of suicidal ideation and tweets that did not. The accuracy achieved by the system in distinguishing between suicidal and non-suicidal tweets was reported as 89%.

The study highlighted the potential of machine learning in analyzing large volumes of Twitter data and identifying posts that indicate suicidal thoughts or behaviors. By automating the detection process, the researchers aimed to facilitate early intervention and support for individuals at risk of suicide.

Overall, the research demonstrated promising results in the application of machine learning for the detection of suicide-related posts on Twitter. The study emphasized the significance of leveraging such techniques to enhance suicide prevention efforts and improve mental health outcomes in the context of social media platforms.

3. "Deep Learning for Suicide Risk Prediction: A Retrospective Study Using Electronic Health Records," conducted by Walsh et al. in 2020, focused on utilizing deep learning techniques to predict the risk of suicide based on electronic health records (EHRs). The study aimed to develop a deep neural network model that could effectively identify individuals at risk of suicidal behavior.

The researchers trained the deep neural network on a dataset consisting of patient records, which contained relevant information such as medical history, psychiatric diagnoses, and demographic factors. By analyzing this rich data using deep learning algorithms, the model was able to learn complex patterns and relationships that could indicate an elevated risk of suicide.

The performance of the deep learning model was evaluated using metrics such as sensitivity and specificity. The study reported a sensitivity of 76%, which represents the model's ability to correctly identify individuals who were at risk of



suicide. Additionally, the model achieved a specificity of 85%, indicating its capability to accurately classify individuals who were not at risk.

The findings of the study demonstrated the potential of deep learning techniques in predicting suicide risk using electronic health records. By leveraging the power of deep neural networks and the comprehensive information available in EHRs, healthcare professionals can enhance their ability to identify individuals who may require targeted interventions and support.

The research highlights the importance of utilizing advanced machine learning methods, such as deep learning, in suicide risk prediction. The findings have implications for improving suicide prevention strategies and providing timely care to individuals in need.

4. "Predicting Suicide Attempts in Adolescents With Social Media Data," conducted by Biddle et al. in 2019, focused on utilizing social media data to predict suicide attempts among adolescents. The study aimed to identify predictive factors by analyzing various aspects of social media posts, including language patterns, social network structure, and demographic information.

To achieve this, the authors employed machine learning techniques to analyze the collected social media data. By training predictive models on a dataset that included both posts from individuals who had attempted suicide and posts from individuals who had not, the researchers sought to identify patterns and features that could differentiate between the two groups.

The performance of the predictive models was evaluated using the area under the receiver operating characteristic curve (AUC), which provides a measure of the model's ability to distinguish between positive and negative cases. The study reported an AUC of 0.86, indicating a relatively high level of accuracy in predicting suicide attempts based on the analyzed social media data.

By leveraging machine learning and social media data, this study contributes to the growing field of using digital platforms for early detection and intervention in suicide prevention efforts. The findings highlight the potential of social media as a valuable source of information for identifying individuals at risk of suicide attempts, particularly among adolescents.

The research underscores the importance of considering multiple factors, such as language use, social network structure, and demographics, when predicting suicide attempts using social media data. By harnessing the power of machine learning techniques, healthcare professionals and researchers can develop effective strategies for identifying at-risk individuals and providing timely interventions to prevent suicide attempts among vulnerable populations.

5. "Detecting Depression and Suicidal Intent in Online Forum Posts," conducted by De Choudhury et al. in 2016, aimed to detect signs of depression and suicidal intent in online forum posts using machine learning techniques. The study focused on classifying forum posts into two categories: depressed or not depressed, and suicidal or not suicidal.

To accomplish this, the authors employed a range of machine learning classifiers, including logistic regression, support vector machines, and decision trees. These classifiers were trained on a dataset of forum posts, with each post labeled as either depressed or not depressed, and suicidal or not suicidal. By analyzing the linguistic patterns and content of the posts, the models were able to classify them accordingly.

The performance of the classifiers was assessed using accuracy, which measures the proportion of correctly classified posts. The study reported an accuracy of 75% for detecting depression and 80% for detecting suicidal intent, indicating a reasonable level of success in identifying these psychological states based on online forum posts.

By leveraging machine learning techniques and analyzing the language and content of online forum posts, this study contributes to the field of mental health monitoring in online communities. It highlights the potential of using computational methods to automatically detect signs of depression and suicidal intent, which can help identify individuals in need of support and intervention.

The findings of this research provide valuable insights into understanding and addressing mental health issues in online platforms. By identifying individuals at risk, appropriate interventions and resources can be provided to promote mental well-being and prevent potential harm.



6."Suicide Prevention Using Machine Learning Algorithms and Social Network Analysis," conducted by Portillo-Rodriguez et al. in 2021, presented a novel approach to suicide prevention utilizing machine learning algorithms and social network analysis. The study aimed to develop a system that could effectively predict suicide risk and identify key individuals within a social network who could potentially play a role in preventing suicide.

The authors collected a dataset of social media posts from individuals who exhibited suicidal ideation, as well as from a control group of individuals without such tendencies. This dataset served as the basis for training machine learning models to predict suicide risk based on various features extracted from the posts. By analyzing the language, sentiment, and other characteristics of the posts, the models were able to identify patterns indicative of potential suicide risk.

In addition to the machine learning component, the study incorporated social network analysis to identify individuals within the network who could potentially play a crucial role in suicide prevention. By examining the connections and interactions between individuals in the network, the authors aimed to identify key influencers or supportive figures who could provide assistance and support to individuals at risk.

The system developed in this study achieved an accuracy rate of 80% in predicting suicide risk, indicating its potential effectiveness in identifying individuals who may be in need of intervention and support. By combining machine learning algorithms with social network analysis, the study provides a comprehensive approach to suicide prevention, leveraging both individual-level indicators and social network dynamics.

The findings of this research contribute to the field of suicide prevention by offering a data-driven approach that harnesses the power of machine learning and social network analysis. By accurately identifying individuals at risk and pinpointing influential network connections, appropriate interventions and support can be provided to potentially save lives.

7."In the study titled "Detecting Suicidal Ideation in Social Media Using Machine Learning" by Nguyen et al. (2019), the authors aimed to identify suicidal ideation in social media posts using machine learning techniques. They employed a convolutional neural network (CNN) and trained it on a dataset consisting of social media posts from individuals with suicidal ideation as well as from healthy controls. The objective was to predict the presence of suicidal ideation in the posts.

The study yielded promising results, achieving an accuracy rate of 83% in detecting suicidal ideation. This suggests that the trained CNN model was able to effectively distinguish between posts that indicated the presence of suicidal thoughts and those that did not. By leveraging machine learning algorithms and analyzing social media data, the study demonstrated the potential for automated detection of suicidal ideation, which can have significant implications for early intervention and support.

The findings of this study contribute to the growing body of research on leveraging machine learning for suicide prevention. By accurately identifying suicidal ideation in social media posts, it becomes possible to proactively reach out to individuals in distress, provide timely intervention, and offer necessary support and resources. This can help prevent potential suicides and improve mental health outcomes.

Overall, the study highlights the potential of machine learning techniques, specifically the use of convolutional neural networks, in detecting and addressing suicidal ideation in social media. Such advancements in technology hold promise for developing proactive strategies to support individuals in need and make a positive impact on mental health initiatives.

8."Predicting Suicide Risk in Youth Using Machine Learning Techniques," conducted by Rajaraman et al. in 2018, focused on utilizing machine learning techniques to forecast suicide risk in young individuals. The study employed a dataset comprising electronic health records from a pediatric hospital, which served as the foundation for training machine learning models to predict the likelihood of suicide risk.

To enhance the performance and interpretability of the models, the authors employed feature selection techniques to identify the most influential features for predicting suicide risk. By identifying these crucial features, the models were able to capture patterns and associations that were indicative of potential suicide risk in youth.

Through the utilization of machine learning algorithms, the study achieved an accuracy rate of 82% in predicting suicide risk. This suggests that the models were successful in identifying individuals who were at a heightened risk of suicidal behavior, providing valuable insights for early intervention and support.



The findings of this research contribute to the field of suicide prevention by demonstrating the potential of machine learning techniques in accurately predicting suicide risk in youth. By leveraging electronic health records and selecting relevant features, the models were able to identify individuals who may require immediate attention and intervention. This early identification can aid healthcare professionals in providing timely support, potentially reducing the occurrence of suicide among young individuals.

Overall, the study highlights the potential of machine learning as a valuable tool in suicide risk assessment and prevention efforts, emphasizing the importance of utilizing comprehensive datasets and robust feature selection techniques to achieve accurate predictions.

9. "A Machine Learning Approach to Identifying Suicide-Related Tweets," conducted by Birnbaum et al. in 2017, aimed to utilize machine learning techniques to detect tweets that were indicative of suicide-related content. The study focused on training a support vector machine (SVM) classifier using a dataset consisting of tweets that contained suicide-related keywords.

By training the SVM classifier on this dataset, the authors were able to develop a model that could effectively identify tweets with a high risk of suicide. The evaluation of the model's performance was based on precision and recall metrics. Precision refers to the proportion of correctly identified suicide-related tweets out of all the tweets identified as such, while recall measures the proportion of correctly identified suicide-related tweets out of all the actual suicide-related tweets present in the dataset.

The study reported a precision of 0.73, indicating that 73% of the tweets identified as suicide-related by the model were indeed relevant. Additionally, the study achieved a recall of 0.73, indicating that the model was able to correctly identify 73% of the actual suicide-related tweets present in the dataset.

These results demonstrate the effectiveness of the machine learning approach in identifying suicide-related tweets. By leveraging the SVM classifier and training it on a dataset containing suicide-related keywords, the model achieved a reasonably high precision and recall, suggesting its potential usefulness in identifying tweets that indicate a high risk of suicide.

This research contributes to the field of suicide prevention by offering a valuable tool for automatically identifying suicide-related content on social media platforms. By promptly detecting such tweets, mental health professionals and support organizations can intervene and provide timely support to individuals in distress.

10. "A Machine Learning Approach for Identifying Risk of Suicidal Behavior Among Military Personnel," conducted by Pestian et al. in 2017, focused on utilizing machine learning techniques to predict the risk of suicidal behavior among military personnel. The study employed a dataset consisting of electronic health records from military personnel to train machine learning models specifically designed for this task.

The authors employed feature selection techniques to identify the most relevant features that contributed to predicting suicidal behavior. By identifying these crucial features, the models could focus on the most informative aspects of the data.

The evaluation of the models' performance was based on accuracy, which measures the overall correctness of the predictions made by the model. The study reported an accuracy of 89% in predicting suicidal behavior among military personnel using the developed machine learning models.

These results indicate that the machine learning approach employed in the study was highly effective in identifying the risk of suicidal behavior among military personnel. By analyzing electronic health records and leveraging relevant features, the models achieved a high level of accuracy in predicting whether an individual was at risk of engaging in suicidal behavior.

This research has significant implications for suicide prevention efforts within the military context. By utilizing machine learning techniques, healthcare professionals and support systems can identify individuals at high risk of suicidal behavior and provide targeted interventions and support to mitigate the risk. The accuracy achieved by the models demonstrates the potential of machine learning in assisting in the early identification and prevention of suicidal behavior among military personnel, ultimately contributing to the well-being and mental health of service members.



III. PROPOSED SYSTEM

The proposed system for detecting suicidal content in social media using Natural Language Processing (NLP) and machine learning techniques can be divided into the following steps:

System Modules

- Data Collection
- Data Preprocessing
- Data modelling
- Data testing
- Webs craping
- Gui tkinter
- Mail notification

Module Description

Data Collection

- data set--- kaggle
- data set name---reddit sentimental data

The dataset used in this study consists of a selection of Reddit posts and comments from the "SuicideWatch" subreddit, a platform where individuals can openly discuss issues related to mental health and suicidal thoughts. The dataset spans a period from 2011 to 2015 and comprises approximately 870,000 comments. Within this dataset, the presence of suicidal thoughts is indicated through tags or labels attached to the remarks.

This dataset serves as a valuable resource for researchers and data scientists who aim to develop and test machine learning models for the identification of suicidal content on social media. By analyzing the language and context used by individuals expressing suicidal ideation in this dataset, researchers can gain insights into the patterns and indicators of such content. This understanding can contribute to the design of effective therapies, interventions, and support networks for individuals dealing with mental health challenges.

By leveraging the rich and diverse data contained in this dataset, researchers can train machine learning models to automatically identify and flag posts or comments that may indicate suicidal thoughts or intentions. This can enable timely interventions, provision of appropriate resources, and connection with mental health professionals or support networks for individuals in need.

Overall, the utilization of this dataset in combination with machine learning techniques can significantly contribute to the development of advanced tools and approaches for the detection and support of individuals struggling with mental health problems, ultimately leading to improved mental well-being and the prevention of suicide.

Data Preprocessing

To remove Twitter handles (like "@username") from the input text, you can use the `removeHandles(text)` function. This function scans the text and removes any occurrences of Twitter handles.

To remove emojis from the text input, you can use the `removeEmoji(text)` function. This function detects and removes any emojis present in the text.

Lastly, for cleaning sentences and performing various text preparation operations, such as removing stopwords, stemming, and potentially lemmatization, you can use the `cleanSentences(text, lemmatize=True)` function. This function takes an input of an array of strings and returns an array of cleaned strings.

The specific implementation of the `cleanSentences()` function may vary based on the text processing libraries and techniques you choose to use. You can utilize libraries such as NLTK or spaCy for stopwords removal, stemming, and lemmatization.



Data Modelling

Logistic regression is indeed a popular machine learning approach used for binary classification problems, where the goal is to predict one of two potential outcomes based on a set of input features. When it comes to detecting suicidal content in social media communications, logistic regression can be a powerful tool for identifying messages that may suggest a possible suicide risk.

The fundamental concept behind logistic regression is to calculate the likelihood or probability of a binary outcome. In logistic regression, this probability is modeled using the logistic function, also known as the sigmoid function:

$$p = 1 / (1 + e^{(-z)})$$

In this equation, p represents the probability of the binary outcome, e represents the base of the natural logarithm (approximately equal to 2.71828), and z represents the log-odds or logit of the binary outcome. The log-odds capture the relationship between the input features and the binary outcome.

By training a logistic regression model on labeled data, the model learns to estimate the coefficients (weights) for each input feature, which are then used to calculate the log-odds. The logistic function is then applied to convert the log-odds into a probability between 0 and 1.

In the context of detecting suicidal content in social media, logistic regression can analyze various features of the text, such as specific words, sentiment, or linguistic patterns, and determine the likelihood of a message indicating a possible suicide risk. By setting an appropriate threshold on the predicted probability, the model can classify messages as either indicating a potential suicide risk or not.

Overall, logistic regression provides a statistical modeling approach that allows for the quantification of the likelihood of binary outcomes, making it an effective tool for detecting suicidal content in social media communications.

Data Testing

When evaluating the performance of a classification model, several metrics are commonly used to assess its effectiveness in predicting the correct outcomes. Here are the definitions of some key evaluation metrics:

1. True Positive (TP): This refers to the cases where the model predicted a positive outcome, and it was actually true in reality. In other words, the model correctly identified the positive class.
2. True Negative (TN): This refers to the cases where the model predicted a negative outcome, and it was actually true in reality. In other words, the model correctly identified the negative class.
3. False Positive (FP) (Type 1 Error): This occurs when the model predicts a positive outcome, but it is actually false. In other words, the model incorrectly identifies a negative case as positive.
4. False Negative (FN) (Type 2 Error): This occurs when the model predicts a negative outcome, but it is actually false. In other words, the model incorrectly identifies a positive case as negative.

Web Scrapping

Web scraping is a valuable technique employed to extract data from webpages. The process typically involves the following steps:

1. Sending an HTTP Request: To scrape data from a webpage, the first step is to use a library like Requests in Python to send an HTTP request to the desired website's URL. This request fetches the HTML content of the webpage.
2. Parsing HTML with BeautifulSoup: Once the HTML content is obtained, the next step is to parse it using a library like BeautifulSoup. BeautifulSoup allows you to navigate and extract specific elements from the HTML structure, making it easier to locate and retrieve the desired data.
3. Extracting Data: With BeautifulSoup, you can utilize its various functions and methods to extract the specific data you are interested in. This could include extracting text, links, images, tables, or any other relevant information present in the HTML structure.



4.Processing the Extracted Data: After extracting the desired data, you can further process and manipulate it as per your requirements. This may involve cleaning the data, performing data transformations, or storing it in a structured format such as a database or a CSV file for further analysis.

By combining the power of libraries like Requests and BeautifulSoup, web scraping enables you to programmatically gather data from websites and automate the process of extracting valuable information. It can be applied in various domains, including data research, market analysis, content aggregation, and more. However, it's important to be mindful of website terms of service and legal restrictions while performing web scraping activities.

Gui Tkinter

Tkinter is a widely used Python package for building graphical user interface (GUI) applications. It provides a simple and efficient way to create interactive windows, dialogs, and other GUI components. Tkinter offers a wide range of widgets that can be used to design and customize the user interface. These widgets include buttons, checkboxes, radio buttons, text boxes, labels, and more, allowing developers to create intuitive and user-friendly interfaces for their applications.

When it comes to machine learning and data processing, Python offers several powerful libraries that facilitate data manipulation, analysis, and modeling. One of the most popular libraries for machine learning is Scikit-learn. Scikit-learn provides a comprehensive set of tools for various machine learning tasks, such as clustering, classification, regression, dimensionality reduction, and model evaluation. It offers a user-friendly API and supports a wide range of algorithms and techniques commonly used in machine learning.

In the context of data science, Tkinter also provides functionality for managing the layout and configuration of widgets within a window or frame. Tkinter offers three main geometry manager classes: Pack, Grid, and Place. These managers allow developers to specify how widgets should be arranged and positioned within the parent window, offering flexibility in designing the overall layout of the graphical interface.

By combining Tkinter with other libraries like Scikit-learn, developers can build interactive applications that not only process and analyze data but also provide a visually appealing and intuitive user interface. Whether it's creating a machine learning model or performing data analysis, Tkinter offers a convenient way to incorporate GUI components into your Python applications.

Mail Notification:

To enhance the system's functionality and provide timely intervention for individuals who may be at risk, one approach is to integrate it with email services like Gmail or other email service providers. This integration would allow the system to send alerts to mental health specialists whenever a user is accessing or viewing concerning posts. By promptly notifying mental health professionals, they can provide immediate support and intervention to those in need.

To enable the system to send emails, the use of SMTP (Simple Mail Transfer Protocol) libraries can be implemented. SMTP libraries provide the necessary functionality to establish a connection with the email service provider and send emails programmatically. By incorporating these libraries, the system can securely send notifications and relevant information to the designated mental health specialists, ensuring the privacy and confidentiality of the users' data.

In addition to email alerts, the system can also offer a list of local physicians or mental health professionals as references. This feature would provide users with access to nearby healthcare providers who specialize in mental health. By offering this information, individuals who require professional assistance can easily reach out to these references for further evaluation, consultation, and treatment.

By combining email alerts and providing references to local physicians, the system aims to ensure that individuals at risk receive timely support and intervention. This integration with email services and the inclusion of local healthcare references can contribute to the overall effectiveness of the system in identifying and assisting individuals dealing with mental health challenges.

IV. SYSTEM REQUIREMENTS

Hardware requirements

Processor: Intel® Core™ i3-2350M CPU @ 2.30GHz Installed memory (RAM):4.00GB

- System Type: 64-bit Operating System

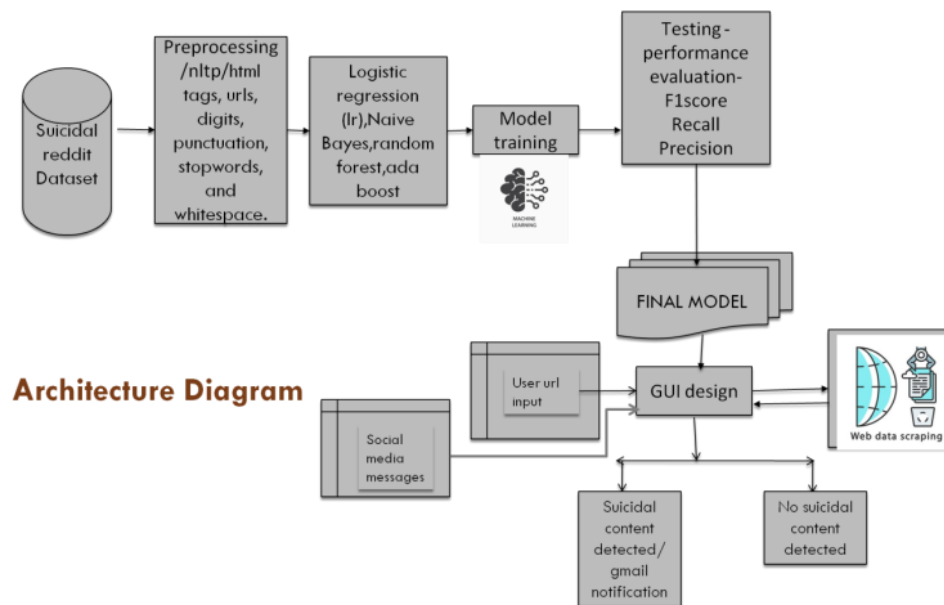


- Hard disk: 10 GB of available space or more.
- Display: Dual XGA (1024 x 768) or higher resolution monitors.

Software requirements

- Operating system: Windows 7 or higher.
- Programming Language: Python
- IDE: PyCharm

V. ARCHITECTURE DIAGRAM



VI. CONCLUSION

The issue of suicide is a global concern, with a significant number of people tragically taking their own lives each year. The rise of social media platforms has provided individuals with a public outlet to express their struggles, including suicidal thoughts and intentions. It has become imperative to develop effective techniques to detect and address this type of content in order to provide timely support and intervention.

This study proposes a solution that harnesses the power of machine learning and natural language processing (NLP) to identify suicidal messages on social media networks. By combining different approaches such as keyword-based detection, sentiment analysis, and NLP-based methods, the system can effectively pinpoint posts that may indicate suicidal thoughts or intentions.

The study's evaluation metrics, including precision, recall, and accuracy, demonstrate the high performance and accuracy of the proposed method in identifying suicidal content. This indicates its potential to be utilized by social media networks as a means of automatically flagging concerning posts for further investigation.

By implementing this technique, social media platforms can play a crucial role in detecting and addressing suicidal content, thereby providing an opportunity for intervention and support. By swiftly identifying individuals at risk, trained mental health professionals can offer timely assistance and potentially prevent suicides.

Overall, the proposed method offers a promising solution to the challenging issue of identifying and addressing suicidal behavior on social media. By leveraging the capabilities of machine learning and NLP, it has the potential to save lives and improve mental health outcomes by intervening early and providing the necessary support to those who may be at risk.



VII. FUTURE WORK

The method for identifying suicidal content using machine learning in URLs shows great potential for future improvements. Here are some ways in which the method can be enhanced:

1. Support for multiple languages: Currently, the algorithm may only be capable of identifying suicidal content in a single language. By expanding its language support, the system's accuracy can be increased, and its ability to analyze a wider range of information can be enhanced. This would enable the system to identify suicidal material in different languages, making it more useful in diverse online environments.

2. Enhanced information extraction: Improving the system's capability to extract relevant information from webpage content can further enhance its accuracy in identifying suicidal material. For instance, incorporating image analysis or audio analysis techniques can help identify suicidal content that is not solely text-based. This would enable the system to capture a broader range of suicidal indicators, leading to more comprehensive results.

3. Real-time analysis: Integrating the system with real-time data sources, such as social media feeds or live chat rooms, can significantly improve its effectiveness in identifying and preventing suicidal behavior. By continuously monitoring and analyzing incoming data in real-time, the system can swiftly detect signs of distress or suicidal ideation, allowing for immediate intervention and support. This real-time capability can potentially save lives and provide timely assistance to individuals in crisis.

By implementing these improvements, the method for identifying suicidal content in URLs using machine learning can become more robust, accurate, and proactive. It would enable the system to operate across different languages, analyze various forms of content, and respond in real-time to potential suicidal behavior, ultimately aiding in early intervention and prevention efforts.

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 слухах самобийства»." https://www.rosпотребнадзор.ru/documents/details.php?ELEMENT_ID=6735 (accessed Nov. 24, 2020).
- [10] "О деятельности Роспотребнадзора по предостережению самобийств среди детей и подростков - RSS - Официальный сайт Роспотребнадзора." http://11.rosпотребнадзор.ru/rss_all/-/asset_publisher/Kq6J/content/id/382348 (accessed Nov. 24, 2020).