CHAPTER 1

INDTRODUCTION

1.1 Rationale

The World Health Organization (WHO) in 2023 reported that more than 700 000 people die due to suicide every year, and for every suicide that happen many more people attempt suicide. Suicide is the fourth leading 77% of global suicides occur in low- and middle-income countries. The process of suicide starts with suicidal thoughts or ideation. It then matures to a suicidal attempt and finally to a completed suicide. Despite the growing number of suicidal cases, it can be prevented by doing primary identification and detection[15].

Social media has become an everyday part of our lives. Nowadays, many people are active on several popular social media networks such as Reddit to share their daily lives and experiences, good or bad. Recently, individuals have broadcast their suicidality on social media sites such as Reddit, indicating that Social Media may have the potential for use as a suicide prevention tool.

Reddit is a website that serves as a hub for sharing news, rating content, and engaging in discussions. Users, often referred to as " Redditors," contribute types of content like links, text posts, images, and videos. The community members have the power to vote on these contributions, either by giving them a thumbs up or down. The posts are organized into categories known as "communities" or "subreddits," which are created by users themselves. Contributions that receive upvotes rise to the top of their subreddit and can even make it to the site's front page if they gather enough support. Social media has recently been used as a domain for research related to the detection of suicidal thoughts[16][13][1]. However, there are challenges associated with conventional suicide-risk screening methods, and finding and comprehending patterns of suicidal ideation represent a challenging task[5]. At the same time, individuals contemplating suicide are increasingly turning to social media and online forums, such as Reddit, to express their feelings and share their struggles with suicidal thoughts[21] which create a natural language problem such as informal language, spelling errors, and grammar mistakes.

1.2 Theoretical Framework

Binary classification falls under the umbrella of supervised learning in machine learning. The main objective of this method is to sort input data into two distinct categories accurately. These categories can take various forms, including text, images, or numerical data. Often referred to as 'positive' and 'negative', or '1' and '0', the goal is to determine a precise decision boundary that separates the two categories. This is achieved by training

the algorithm on a dataset already labeled with these categories. Once the algorithm is trained, it can be applied to predict the category of new, unknown data. An array of algorithms can be utilized for binary classification, such as logistic regression, decision trees, and Support Vector Machine[4][3].

The general process of a classification problem involves several steps. The first is collecting data that will be used to train and test the classification model. Second, the data is cleaned and transformed to make it suitable for analysis. This process may involve handling missing values and normalizing and encoding the categorical variables. Third, the features are enriched to make them relevant to the problem and the most relevant features are selected to be used to train the classifier. Fourth, an appropriate or the best-performing classifier is selected by evaluating a model . The evaluation is carried out by testing the trained model on a separate dataset to evaluate its performance. This evaluation is also important to ensure the model does not overfit the training data.

1.3 Conceptual Framework/Paradigm

Suicidal ideation detection (SID) involves the determination of whether an individual entertains suicidal thoughts or ideations, which is included in the classification problem[21]. It can be achieved through various data sources like tabular data, electronic health records, social media posts, or suicide notes. Machine learning and deep learning models have recently been employed in handling textual content from different sources, including social media platforms such as Twitter and Reddit, for early identification of signs of suicide risks and prevention[16][13][1][19]. Primarily, the exploitation of user-generated content, mainly text posted by users, has been hailed for showing potential when it comes to a proactive approach to warning people about the likelihood of committing suicide. Additionally, this field employs techniques including sentiment analysis and feature engineering. Furthermore, they also evaluated the performance of different machine learning models like SVM, Logistic Regression, Random Forest, and Extra Trees models with promising outcomes in recognizing suicidal thinking with textual cues. In general, research on this topic underscores that leveraging diverse data sources coupled with sophisticated analytics is crucial when addressing this urgent problem.

This study concentrates on enhancing features in the context of identifying suicidal ideation using Reddit data. The implemented system aims to find and comprehend patterns of suicidal ideation and alleviates the common vocabulary discrepancies observed in Reddit data. The feature enrichment strategies employs encompass Content-based Features, Linguistic Inquiry and Word Count (LIWC) Features, and feature expansion. Content-based features pertain to language characteristics, including words, sentence structure, and grammar[2]. On the other hand, LIWC Features involve features derived from text through LIWC analysis tools. The LIWC software analyzes individual language usage to

identify potential patterns indicative of depression[11]. Lastly, feature expansion involves identifying unweighted words in the post representation and substituting the value of that word representation with semantically related words found in the word vector[9].

1.4 Statement of the Problem

Challenges associated with suicidal ideation detection in Reddit is to find and comprehending patterns of suicidal ideation, owing to their sparse layout from TF-IDF as feature extraction and the frequent utilization of abbreviations, slang, and a lack of grammatical correctness in Reddit or social media posts pose a unique challenge. The machine learning process regularly grapples with vocabulary mismatches when attempting to grasp the nuances of user expressions, particularly when employed to detect suicidal ideation, leading to suboptimal performance.

To address this issue, the introduction of feature enrichment becomes imperative, encompassing content-based and LIWC features coupled with feature extraction and expansion methodologies. This comprehensive approach is anticipated to significantly enhance the accuracy of suicidal ideation detection systems, leveraging SuicideWatch and Pushshift data from Reddit. The fundamental classification methods employed for this purpose include Support Vector Machines (SVM), Logistic Regression, Random Forest, and Extra Trees, collectively contributing to a more robust and nuanced analysis of Reddit content for improved suicidal ideation detection.

1.5 Objective and Hypotheses

The primary goal of this research is to examine the impact of feature enrichment on the performance of systems designed for detecting suicidal ideation. This exploration focuses on finding the pattern of suicidal ideation and the use of writing indicators or social media data to improve accuracy. To address the challenge of vocabulary disparities and comprehending the complex pattern, a multifaceted approach is adopted, incorporating content-based, LIWC features and feature expansion. By leveraging these diverse features, the study seek to mitigate language-related challenges commonly encountered in the context of suicidal ideation detection. Furthermore, the research aims to enhance the accuracy through the application of feature extraction and expansion techniques, acknowledging their potential to enrich the model, increase the related term, and reduce vocabulary mismatches contribute significantly to the refinement of detection systems.

Content-based and source-based features are used in previous research[18] to overcome the message credibility and discover the correlation between word and personalty. Categorized by Gupta[18] show some key observations about the content features which correlate with credibility, for instance, message with a large number of unique characters and contain URLs tend to be more trustworthy. LIWC is used by Agnes to count word categories in

the message and to find the correlation between each word and user personality[17]. The result of this study, LIWC can be used to provide a good prediction for user personality with an average r score of 79.4% for all personalities.

The word enriching or feature expansion method in previous research is proven to be able to overcome the problem of writing vocabulary by expanding words based on their meaning[9]. From this concept, the word enriching and feature expansion method is expected to be able to overcome the problem of limited keyword search terms and vocabulary mismatches on the topic of classifying suicidal ideation.

1.6 Scope and Delimitation

This study is centered on addressing the prevalent challenge of vocabulary mismatch, suicidal ideation patterns, and its consequential impact on the outcomes of suicidal ideation detection. The imperative nature of these issues arises from the complexity of finding and comprehending the pattern of suicidal ideation on social media using machine learning with common natural language processing techniques. Furthermore, the brevity and grammatical informality commonly observed in natural language contribute to the complexity of the problem. These factors collectively lead to the sparsity of the tf-idf data's dimensionality. In response to these challenges, Feature Enrichment emerges as a crucial strategy, aiming to enhance and augment these linguistic features. The ultimate objective is to mitigate the risk of misinterpretation by the classifier, ensuring a more accurate determination of whether an individual is writing a suicidal ideation or not. By enriching the features, the study endeavors to overcome the inherent limitations posed by the linguistic intricacies of natural language data in the context of mental health classification.

1.7 Significance of the Study

As discussed above, by integrating feature enrichment using content based, LIWC and feature expansion this study is expected to improve the Machine Learning model for the task of suicidal ideation detection. The contributions of this study include:

1. Using SuicideWatch and Pushshift data from Reddit, which is a sub-platform of the Reddit social media, new aggregation platforms consist of 643 users and 9085 of post from all users.

2. For the purpose of identifying suicidal messages, the study conduct experiments to test several methods of feature enrichment method using content-based feature, LIWC, and feature expansion.

3. Analyzing the suicide and non-suicide post in the dataset with the implemented features