DSO 569 Report

Suicide Detection

Introduction

As people continue to experience unprecedented levels of stress due to various factors such as work pressure, financial instability, and personal challenges, there has been an increase in the rates of depression, anxiety, and other mental health problems, causing a surge in suicide rate. It's crucial to address these concerns and prioritize mental health as an essential component of overall well-being.

However, addressing mental health concerns has always been a challenging task, as individuals who are suffering from depression may not be willing to seek help due to the fear of judgment. Currently, there are limited effective methods available identifying individuals who have been contemplating suicide at the earliest possible stage, which hinder timely interventions aimed at preventing them from engaging in suicidal behavior.

The objective of this project is to leverage natural language processing techniques to develop a suicide prediction tool that advances our ability to predict future suicidal behavior. We aim to evaluate and contrast various machine learning models including XGBoost, RNN, Bidirectional GRU, and Bert, by analyzing a dataset consisting of over 230,000 authentic records to identify the optimal algorithm that achieves the highest predictive accuracy while maintaining reasonable computational costs.

Data Description

The dataset is a collection of posts from "SuicideWatch" and "depression" subreddits of the Reddit platform. The posts are collected using Pushshift API. All posts that were made to "SuicideWatch" from Dec 16, 2008 (creation) till Jan 2, 2021, were collected while "depression" posts were collected from Jan 1, 2009 to Jan 2, 2021. There are 232,074 observations.

The dataset displays a text column and a class column (suicide, non-suicide) as shown below:

	text	class
2	Ex Wife Threatening SuicideRecently I left my ...	suicide
3	Am I weird I don't get affected by compliments...	non-suicide
4	Finally 2020 is almost over... So I can never ...	non-suicide
8	i need helpjust help me im crying so hard	suicide
9	I'm so lostHello, my name is Adam (16) and I'v...	suicide

Figure 1: Raw Data

Data Preprocessing & Feature Engineering

In the data cleaning section, it consists of the following 4 parts:

Translate emoji to English Description

Emoji are graphical symbols that represent emotions, ideas, or concepts, and are commonly used in text messages, social media posts, and online comments. By converting emojis into their English descriptions or labels, we can incorporate their meaning into the NLP models more effectively, and improve the accuracy and interpretability of the models.

Remove stopwords

Stopwords are words that are commonly used in natural language, such as "the", "a", "an", "in", "of", "to", and so on. They do not carry significant meaning on their own, and their frequent occurrence in text can add noise and make it more difficult to identify the important words that are more indicative of the content. Therefore, we removed NLTK (Natural Language Toolkit)'s default list of english stopwords and some unmeaningful punctuations like "." ","

Regex cleaning

Regex cleaning involves using regular expressions to identify and replace patterns of text that are not relevant to the analysis or modeling tasks. For example, we replace the word "jobs", "career", "intern", "internship" into " _WORK_" since these words have the same meaning in a high level. This technique helps us reduce noise, and improve accuracy.

Lemmatization

Lemmatization is the process of reducing words to their base or dictionary form. For example, the lemma of the word "mice" is "mouse", and the lemma of the word "swimming" is "swim". The goal of lemmatization is to group together different forms of a word so they can be analyzed as a single term, which can help improved generalization, and reduced Dimensionality.

Topic modeling

We utilized Bigram and Trigram Topic models to examine common themes across different labels.

Bigram

non-suicide	suicide
end life	want die
want die	like shit
suicidal thoughts	want end
want live	want kill
want talk	need help
…	…

Non-suicide: negative thoughts but want live and talk
Suicide: nearly all negative words

Trigram

non-suicide	suicide
smile face sunglass	face pensive face
cool smiling face	downcast face sweat
loudly crying face	face tears joy
face rolling eyes	face steam nose
face tears joy	crying face loudly
…	…

Non-suicide: lots of positive emojis
Suicide: lots of negative emojis

Figure 2: Topic Models

Upon analyzing the reviews, it is evident that the non-suicide reviews contain a higher frequency of positive emojis and an abundance of negative thoughts, yet the individuals express a desire to

continue living and engaging in conversation. On the other hand, the suicide reviews predominantly comprise negative vocabulary and negative emojis.

Model Exploration

We began by constructing a baseline model using the Logistic regression algorithm. We then evaluated the performance of various models, including XGBoost, RNN, Bidirectional GRU, and Bert, to identify the optimal solution.

Prior to constructing our models, we utilized a pre-trained Bert Tokenizer to tokenize the preprocessed data. Subsequently, we partitioned the dataset into training and testing sets in a 7:3 ratio.

Model	Accuracy (%)	Processing Time
Logistic (Baseline)	71.74	11.2 s
XGBoost	85.45	56.8 s
RNN	94.93	7 min 38 s
Bidirectional GRU	96.49	2 h 40 min
Bert	97.40	28 min 23s

After the comparative analysis of several machine learning models, we can find that the baseline logistic model has shown the lowest accuracy but the fastest processing speed, taking only 11.2 seconds. The XGBoost algorithm significantly improves accuracy while taking a modest amount of time, only 56.8 seconds. The RNN model with a 0.5 dropout rate outperforms XGBoost with an accuracy of approximately 95% on the validation set, and a processing time of around 7 minutes, which is reasonable and acceptable. We also evaluated the performance of Bidirectional GRU model with 1 hidden layer and 0.4 dropout rate, which slightly improves the accuracy to 96.49%, but its processing time is relatively longer. Finally, the BERT model outperforms all other models with the highest accuracy of 97.4% and reaching the best accuracy and lowest loss in the first epoch, requiring only 28 minutes for processing. Although the improvement over the RNN model is about 3%, it is a significant accomplishment when handling a vast amount of data.

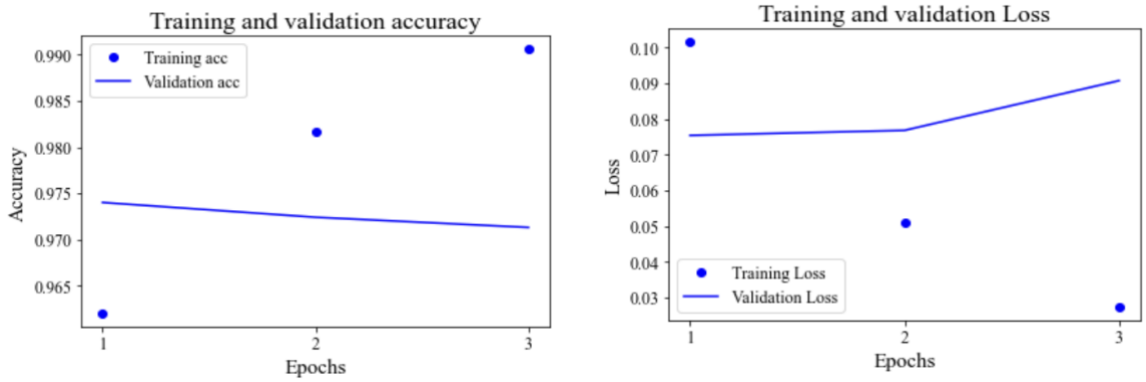

Figure 3: BERT Performance (3 Epochs)

Business Application / Implementation

This novel tool is specifically designed to mine words in online communication using text classification. It can be applied in different scenarios, serving as a suicide prediction detector that identifies words or phrases reflecting a strong tendency towards depression, anxiety, or hidden signs of mental illness that could potentially lead to suicidal behavior. Corporates or schools could insert this tool as a plug-in on the necessary websites of company forums or school discussion boards. By doing so, it can help identify individuals who are unaware of their problems, thereby raising their self-awareness and hopefully prompting them to seek medical help.

An intranet, a platform that is designed to facilitate communication and collaboration among employees within an organization, could be an ideal destination to use this tool. Employees share information, insights, and feelings on this open platform to seek motivation, gain recognition, and achieve emotional resonance. Hence, any suppressed feelings or depressive symptoms of the employees may be inadvertently expressed in the comments they post. In a similar way, a student forum is also an online community where students engage with each other and share their thoughts and opinions about their daily life. Having this tool as a plug-in on the virtual platform, word patterns from online text indicating signs of depression or preoccupation with death could be analyzed and flagged. The backend system will be notified if certain users have been red flagged multiple times.

Another effective strategy for utilizing this tool is through the implementation of employee surveys or investigations by corporate HR departments to monitor the mental health of their employees and identify potential suicide risks. These surveys incorporate a range of targeted questions that aim to assess an employee's mental health, stress levels, mood, overall satisfaction with their work, and other key indicators. Employees are encouraged to be open and honest in their responses, given that these surveys are usually anonymous and confidential. By analyzing the data collected from these surveys, our tool can accurately identify warning signs and risk factors associated with suicide. Once an employee is flagged as a high-risk individual with a strong tendency towards suicide, the HR department will be notified immediately, allowing them to take appropriate actions to support and protect the employee. These surveys will be conducted on a quarterly basis to ensure that the mental health of employees is continuously monitored and any potential risks are identified and addressed in a timely manner.

Addressing mental health concerns has always been a challenging task, as individuals who may be at risk of suicide or suffering from depression may not be willing to seek help due to the fear of judgement. However, our innovative tool is able to effectively target those who are struggling with depression by analyzing their online posts on intranet or student forums, as well as responses to surveys. This enables schools and corporations to take appropriate actions to support and protect those at risk. Creating a safe and supportive environment where students or employees feel comfortable seeking help is crucial for promoting mental health. Mental health training can be provided to each individual, which can help them recognize warning signs and offer necessary support to those in need. In addition, schools or corporations can provide access to professional counseling services, offering individuals the opportunity to speak with trained mental health professionals who can help them manage their feelings and save them from depression.

Conclusion

To sum up, Suicide Sentiment Analysis employs text classification to look for possible suicidal inclinations. This tool can be used in diverse contexts, such as intranet systems, academic forums, or workplace surveys, allowing organizations to identify early mental health warning signs and take suitable measures to help those in need. By creating a secure and supportive atmosphere, individuals are motivated to seek assistance and professional counseling. We compared several candidate models, including Logistic Regression, XGBoost, RNN, Bidirectional GRU, and BERT. The BERT model, with 97.4% accuracy, outperformed the others despite a longer processing time. Its implementation can encourage help-seeking, decrease suicide rates, and bolster suicide-prevention efforts in organizations, hence fostering better mental health for both employees and students.

Future Improvement

There are several future improvements that can be made in terms of the models and application.

To improve model efficiency, we would like to try fasttext, DistilBert, smf FastBert. To improve accuracy, we would like to try more advanced models which are derived from Bert such as RoBerta, DeBerta-V3. More advanced models could be employed to further enhance the tool's accuracy and efficiency, ultimately contributing to the development of a more effective suicide prediction and prevention system. We would like to find an acceptable trade-off between efficiency and accuracy. Also, automating workflows and creating pipelines allows to deal with streaming data and make real-time predictions.

In terms of the application, improvement could involve refining and expanding the scope of its text classification capabilities to include a broader range of mental health indicators, which allows for more comprehensive support for individuals' overall mental health in need. For instance, the algorithm can predict suicide tendency, as well as other mental disorders like Bipolar Effective Disorder, Obsessive-Compulsive Disorder (OCD) and so on. Additionally, if we can integrate the tool in social media platforms and other communication channels to gauge the suicide susceptible individuals, we can further extend its reach and utility, as well as improving prediction of suicide.