A machine learning-based approach for data analysis to ascertain suicidal individuals from Social media users

by

Fatiha Binte Kamrun Nahar
19101500
Umme Halima Afsana
19101427
Azizul Muktadir Chowdhury
22341040
Maha Hasnaen
19141002
Sumaya Jahan
22241182

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
January 2023

Approval

The thesis titled "A machine learning-based approach for data analysis to ascertain suicidal individuals from Social media users" submitted by

1. Umme Halima Afsana (19101427)
2. Fatiha Binte Kamrun Nahar (19101500)
3. Azizul Muktadir Chowdhury (22341040)
4. Maha Hasnaen (19141002)
5. Sumaya Jahan (22241182)

Of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 17 , 2023.

Examining Committee:

Supervisor:
(Member)

Esfar E Alam Turzo
Senior Lecturer
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

In this research, we propose a hybrid model for predicting suicide risk from text data that incorporates BERT, VADER, and a Random Forest classifier for sentiment analysis. This model aims to identify individuals who may be at risk of committing suicide based on the tone of the text. The model is trained on a labelled dataset of text data that is either classified as "suicide" or "not suicide," which provides the model with instances of text data that are linked with high or low suicide risk respectively. In order to extract feature representations of the text data, the BERT model is utilized, and the VADER model is utilized in order to extract sentiment ratings for each individual text. These features are integrated into a single feature vector for each text, and then the Random Forest classifier is trained using this feature vector. A number of different metrics, including accuracy, precision, recall, and F1-score, are utilized in order to assess the performance of the model. The findings of this research indicate that the hybrid model that was suggested is capable of accurately predicting the risk of suicide based on text data and that it is suitable for use as a tool to help clinical decision-making. The performance of the model to recognize patterns and trends in text data that are indicative of suicide risk holds promise for future research in the subject. Our novel composite model combining BERT, VADER with Random Forest Classifier has the accuracy of 82 percent.

Keywords

Data analytics, machine learning, natural language processing, Random Forest, suicide, detection of suicide, algorithms, Bert, Vader, text-preprocessing, depression, artificial neural network, natural language processing

Dedication (Optional)

A dedication is the expression of friendly connection or thanks by the author towards another person. It can occupy one or multiple lines depending on its importance. You can remove this page if you want.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our co-advisor Mr.Esfar E Alam sir for his kind support and advice in our work. He helped us whenever we needed help.
Thirdly,Dr. Md.Golam Rabiul Alam and the whole judging panel of Data Science
And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

Chapter 1

Introduction

1.0.1 Background

Globally, suicide has become a major public health problem, resulting in a substantial loss of life. Suicide prediction models have been created as a means of supporting clinical decision-making and identifying patients at risk for suicide. In this research, we present a hybrid model that incorporates BERT, VADER, and a Random Forest classifier for sentiment analysis to predict suicide risk from text data. Our objective is to compare the performance of the proposed hybrid model to that of existing state-of-the-art suicide prediction methods. To intentionally end one's own life is the definition of suicide. It's possible to classify the factors that prompted this behavior into two broad classes. Any one circumstance that pushes a person over the edge to kill their own life counts as one. The suddenness, lack of forethought, and inability to foresee or take precautionary measures before to the act of self-inflicted death are hallmarks of impulsive suicide. Suicide may also be caused by mental health issues. A person with a mental health problem or instability, such as anxiety, depression, or post-traumatic stress disorder (PTSD), may come to think that suicide is the only way to stop their suffering. After much thought and preparation, someone with these mental health conditions may come to the conclusion that they should take their own life. Suicidal ideation refers to the act of having such thoughts or ideas. Whether the decision to end a life was taken hastily or was deliberated over, it is never a good thing to have to kill someone. The number of people who take their own lives is steadily increasing. The yearly suicide rate has been estimated at anything from 60,000 to 80,000. If we really want to make a difference, we also need to focus on a significant element in suicide, which is suicidal ideation. From having fleeting suicidal thoughts to being so preoccupied with the idea of death that life has lost all meaning, there is a logical progression. It might take a long time before a person with a potentially deadly mental illness recognizes that anything is wrong. This is so because the process takes a very long time. As a result, preventative efforts rely heavily on early detection. Language analysis is crucial for early detection of suicidal ideation. For the simple reason that a person who is thinking about suicide may say anything, deliberately or not, that creates that impression. Given the speed with which we can now communicate online, many individuals choose to do so rather than actually talk to one another. People are more inclined to share their innermost thoughts and emotions when they know they are anonymous, creating a potentially useful linguistic resource for detecting suicide ideation. In this article, we'll discuss how to recognize signs of suicidal ideation in online discussions. We consulted SuicideWatch for user-generated suicide-related material. Users on the SuicideWatch subreddit may feel comfortable discussing their struggles with depression, anxiety, and other mental health issues. Such works often have overt or subtle allusions to, or phrasings of, suicide thoughts or attempts. We have used Natural Language Processing (NLP) to develop a model that can identify Suicidal Ideation in these interactions while also protecting the privacy of the individuals[25]. The primary focus of this study is on evaluating whether or not a state-of-the-art Deep Learning model called Transformers can be used to identify suicidal thoughts and behaviors.To effectively process natural language, a solid understanding of sequence analysis is essential. Its superior performance comes from the fact that transformers don't need sequential data processing. The ability to train on bigger datasets has allowed for the creation of pre-trained systems like BERT, and this is all because of the use of transformer models, which allow for parallelism. Based on our findings, we established a taxonomy that exemplifies Transformer-based models like BERT. This study demonstrates that the hybrid model provided for predicting suicide risk from text data is successful. The model's capacity to recognize patterns and trends in text data that are suggestive of suicide risk bodes well for future studies on the subject. This paper gives an overview of the suggested model and its assessment, as well as a discussion of the results' implications for the field of suicide prediction.

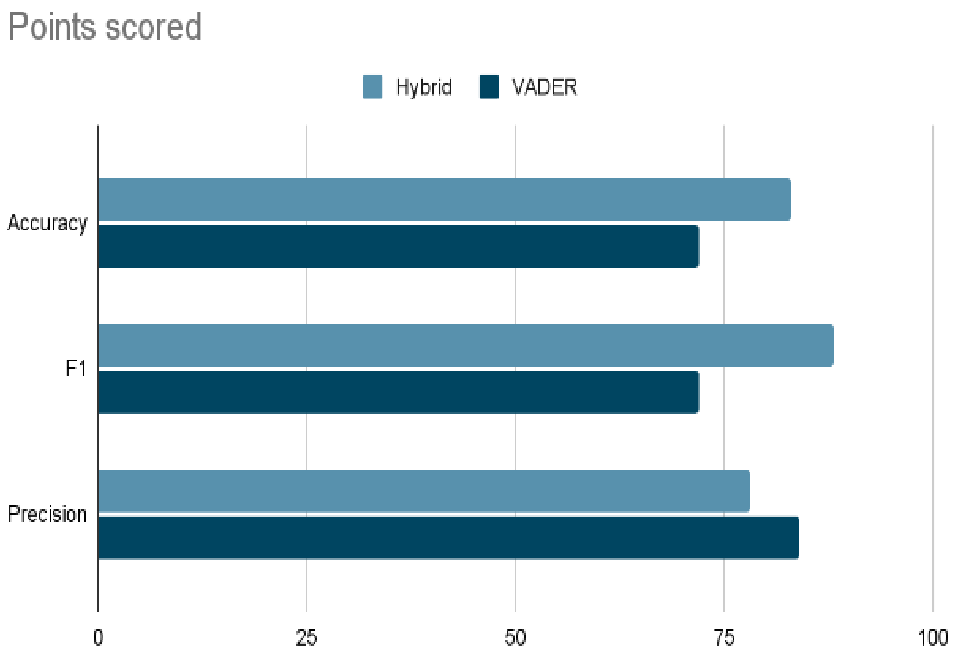

Figure 1.1: Figure for class SUICIDE

1.1 Problem statement

1.1.1 People routinely post self-threatening messages on social media during times of crisis

Alternatively, they might opt not to act. Social media is our major source of information, thus these posts have an effect on our research. Overall, there is a lack of accuracy and difficulty in distinguishing between individuals who are actually suicidal and those who write in a moment of emotion. Insufficient statistics, reports, and Autopsies; By increasing the number of autopsies, the misclassification of suicide may be corrected.

1.1.2 Unreliable statistics, autopsy, and reporting

50% of the autopsies were performed 25 years ago. According to research, there are more suicides when there are more autopsies[3]. Most autopsies were performed 25 years ago. According to research, there are more suicides when there are more autopsies. There is no strong evidence that ordinary inactivity contributes significantly to the under reporting of suicide. Suicide and suicide attempts are significant public health concerns in Bangladesh. Jhenaidah, one of Bangladesh's 64 districts, has one of the highest suicide rates, according to surveys. The number of suicide attempts in the district has garnered very little attention from researchers. Primary descriptive analysis was undertaken using routinely gathered data from the Bangladesh-based non-governmental organization (NGO) Societies for Voluntary Activities (SOVA). [1] There are disadvantages, such as the fact that many attempted and completed suicides in Bangladesh go unreported because of societal stigma; our analysis only examined registered cases. Compared to Bangladesh, Jhenaidah has a much higher suicide and suicide attempt rate. Despite the fact that female suicide rates were greater than male suicide rates, male suicide rates have risen significantly during the research period. Future research is necessary to fully comprehend the regional dynamics and patterns of both fatal and nonfatal suicidal behaviors. The development of a national, district and subdistrict suicide prevention strategy should be a high priority. A murder-suicide occurs when a person appears to commit suicide while simultaneously murdering one or more others, either before or during their own death. Suicide-homicide can take on a variety of forms. For example, murder committed alongside the suicide of a person with suicidal thoughts; suicide committed before or after the murder by a substitute; a murder that results in suicide (such as a suicide bombing/terror attack or a serious accident, especially if it involves the perpetrator and others); murder committed in anticipation of receiving the death penalty; and suicide committed after the murder as a form of self-punishment due to remorse.

- To murder an opponent in self-defense while simultaneously committing suicide.
- Legally killing someone to keep them from harming others as they continue to commit suicide.
- Justifiable homicide that causes or promotes suicide.[5]

1.1.3 Lacking adequate knowledge about postmortem rates over the globe

According to reports, there have been suicides in Israel thus far. In Israeli society, the stigma associated with religious prohibitions against self-harm and the cultural standard that places suicide victims at the bottom of the "death hierarchy" and military combat casualties at the top persist, despite the fact that it is widely acknowledged that data on suicide is underestimated due to difficulties in determining the cause of death and financial considerations[6]. Consequently, suicides are typically classified as unintended or accidental causes of death. In addition to unidentified fatalities, it was suspected that some accidental deaths in Taiwan and Greece were underreported suicides. Using the aforementioned factors, it was estimated that suicide underreporting was negligible in nations where the ratio of suicide to probable underreported suicide was high.

1.1.4 Unanticipated and unexpected deaths: How frequently do they commit suicide?

This makes accurate data collection difficult. Many suicides are unreported due to the absence of the weight of proof that death investigators seek: A NOTE ON SUICIDE, a first attempt, or a particular technique. Drug overdoses continue to account for the vast majority of unexplained deaths. The absence of a clear reason makes it difficult to define these killings. Others may have been unintended, while others may have been deliberate. According to experts and grieved family members, many of these deaths appear to be suicides; yet, for the sake of convenience, or lack thereof, the deaths are routinely branded "undetermined" and the investigations are closed. The influence of this practice on government policy and preventative actions is minimal. Drug self-intoxication is lethal. The establishment of this category would highlight the gravity of the drug overdose problem and the intricacy of these fatalities. The majority of drug overdose and intoxication deaths remain unidentified. Reexamination of national instances involving fatalities and accidents in Scandinavian countries."[16] In all, the experts confirmed that 69 percent (Sweden and Norway) and 78 percent (Denmark) of the deaths listed in the official mortality statistics as accidents actually occurred (Tllefsen et al., 2016) "National death rates should be similar across nations using the International Classification of Diseases from the World Health Organization[12]. It might be difficult to distinguish between causes of death, particularly suicide and accidents. Understanding accidents is important for both accident and suicide prevention.

1.1.5 Suicides are masked by close relationships

People in many nations would not even speak the term out of fear of incarceration or an eternal punishment for really committing suicide. It appears that the term stigma was coined to describe the stigmatizing sentiments associated with suicide, which are unlike anything else in the world. Exists despite the fact that its initial illegality has been forgotten and its religious taboo has been partly relaxed. The official statistics misrepresent the suicide rate: Due to this argument, some deaths labeled as having no obvious cause are really suicides; suicide research frequently accounts for these deaths. We began investigating the notion that a substantial number of probable suicides had been misclassified as non-suicide deaths.

1.1.6 Suicide is complicated since there are several facts

Each individual should take a unique approach to the suicide problem. Due to its complexity, there are not many recommended procedures for preventing it. However, suicide is one of the most significant global public health issues. Suicide is complex because individuals are complex. Each individual should approach a suicide crisis differently.

1.1.7 The definitions of suicide itself provide the first benchmark

Both suicide intent and thought and non-suicidal self-harming behavior must be separated. The International Association for Suicide Prevention has established a task group to study the feasibility of developing a nomenclature applicable to all suicide-related events. Data should be swiftly gathered and stored in a central database after being acquired. For the benefit of research and public health, it is important to foster links to data banks that are connected to other interesting settings.

1.1.8 Too soon concluding the case or neglecting to undertake a comprehensive investigation within the allocated time

It might take a long to obtain the data and make it available to the public. There is a paucity of demographic information. Enhanced tracking would result in more targeted preventative interventions. An enquiry requires time to complete. For example, a suicide might originally be classed as accidental, but fresh information from a mental autopsy could lead to the categorization being changed to suicide.

1.1.9 Health Inequality in the occurrence of unintended and unexplained fatalities among specific populations

The Black community in the United States best exemplifies the disparity and racial injustice in the healthcare system[2]. Many suicides in these places continue to be unintentional or unexplained due to a vastly unequal distribution of medical and mental health facilities. As a result of the colonial treatment of Indigenous people, these health inequities continue to persist and are encoded in many of our institutions.

1.1.10 Exact ethnicity information is not obtained at the time of death

Even though there is universal recognition of a suicide pandemic in certain Indigenous communities, it is impossible to precisely estimate the incidence of suicide on and off-reserve. We are unable to completely comprehend the suicide landscape due to the lack of racial and ethnic data collected at the time of death. Similarly, a native death is considerably more likely to be ”accidental” or ”undetermined” than a death in a more advanced civilization.

1.2 Research Objective

Suicide has become an epidemic in modern society. Suicide attempts can best be prevented if their signs and causes are identified and addressed as soon as possible. Our study uses an examination of internet data to identify suicidal language and evaluate a person's level of suicidality.
These days, people frequently utilize various forms of internet communication to share their sentiments, sadness, and even suicidal ideas. Therefore, online platforms have naturally evolved into tools(Combing BERT and VADER) for monitoring suicidal thoughts and behaviors, and content mining on social media can aid in suicide prevention. We are seeing the emergence of some strange societal phenomena, such as online support groups encouraging self-mutilation and ”copycat” suicides. In 2016, for instance, the ”Blue Whale Game” circulated on social media, and its numerous objectives (including self-harm) ultimately led to the deaths of some game participants. Suicide is a major public health concern that results in the loss of many lives every year. Hence, it is crucial to ascertain the suicide rate and decrease it prior to people committing themselves. Potential suicide victims may exhibit suicidal ideation through role-playing, brief ideas, and actual suicide schemes. Clinical and psychiatric examinations, survey categorization, and the presence of suicidal thoughts were all the subjects of several studies. Predicting someone's likelihood of committing suicide using their social media activity allows for early intervention and a deeper knowledge of people's motives, thanks to AI and machine learning techniques.

1.2.1 Reason of interest in this topic because of obvious need to fill some global behavioral pattern gaps that are as yet unknown

- A wealth of data for use in NLP may be found in suicide notes.Building artificial neural network topologies like combining BERT AND VADER could find our keyword discoveries and detection of suicidal or nonsuicidal behavior interesting. There has already been researching on suicide notes using ML and data analysis.**??**

- The causes of suicide involve a web of interconnected elements. Suicide research draws from a wide range of methodologies and disciplines. Clinical approaches, for instance, can look at things like the patient's history and resting heart rate. Conversations between patients and clinicians, as well as the use of questionnaires, have long been used to estimate the risk of suicidality. Our text-based data analysis for suicide attempts aims to help users determine if a candidate is seriously contemplating suicide based on their online activity. One such method is to look for terms connected to suicide.

- Suicide notes are increasingly being shared online as blog posts, and it's possible that this type of writing poses a suicide risk. Suicide notes are often brief. We've developed an efficient system for detecting and preventing suicidal thoughts and feelings in internet user material.

- While the previous research used a smaller social dataset, our suggested approach may be able to discover a superior answer by employing more efficient feature engineering. Furthermore, it can adapt to real-world purposes beyond only surveys because of its autonomous detecting capacity. With the use of social networks, some behaviors may be easily recognized. Furthermore, they are extremely useful for finding those who are particularly defenseless. Since the transmission of the COVID19 virus has been so rapid, parents have been worried about their children's mental health. When children are isolated because of a coronavirus, suicide is a significant cause of death. Our data analysis on social media helps identify those at risk, which can help prevent terrible suicides.

- Numerous authorities believe that advances in rapid data analytics, NLP and ML will allow us to scientifically analyze the risk of suicide and offer us with useful solutions for preventing suicide. This includes the most recent global research on digital suicide prevention using wearable sensor-driven devices, machine learning, and smartphone apps. With the use of an advanced technology that digitally forecasts suicides by collecting and analyzing massive amounts of medical data, we may soon be able to objectively quantify suicides and the cumulative impact of each risk factor and their interplay.

- The current risk a person faces, adjusted for any new knowledge uncovered through the collection and analysis of large volumes of demographic and/or patient data of an individual's distinct risk profile. We could aid many more individuals with the limited resources we have if we used such technologies. To some, the advised remedies may be overly pessimistic, but we have strong reason to believe they are effective.

- Rather than relying on traditional statistical approaches, publishers may gain insight into the risk factors that lead to a suicide attempt by using a technological strategy. Instead of aiming to prove or disprove a preconceived theory.

- The emergence of innovative technologies Since every individual has their own unique set of coping mechanisms for dealing with mental health issues and thoughts of suicide, it's important not to just depend on the standard protocols for dealing with people who have such disorders. By combining our two theories, we can make the psychological platform more transparent to the academic community.

Chapter 2

Literature Review

2.1 Related Work

Numerous research have been conducted in this field since suicide has grown to be such a serious issue in recent years. Multiple studies have found that social media communications sent via SMS might be an excellent indicator of suicidal ideation.[20] Choudhury et al., who used statistical methods, discovered identifiable markers of suicide ideation changes. In their research, Coppersmith et al. observed that the number of tweets displaying depressive symptoms spiked in the weeks before a suicide attempt. The number of angry tweets dramatically increased in the weeks after the suicide attempt. There appears to be a wide range of language patterns present in posts containing suicide-related content, according to the research. Classification methods from the field of machine learning are used to determine suicidal thoughts[8].
Vioul'es et al. used a four-point scale to annotate tweets about suicide. The Twitter Application Programming Interface was chosen so that users may engage in post-streaming analysis. This article by Faisal et al. describes a technique for identifying suicidal thoughts. In order to discover the best way to extract the most important information from the suicidal texts, they examined the outcomes of different machine-learning approaches and used a wide range of cutting-edge algorithms. An automated method for determining the emotions represented in suicide notes was created by Desmet et al. using a binary support vector machine. The notes' underlying feelings were extracted using a bootstrap resampling method. Due to its superiority over more traditional machine learning techniques, neural networks have recently been the focus of research into the detection of suicidal ideation in written forms. Recurrent Neural Networks (RNNs) and LongShort Term Memories (LSTMs) are shown to have enhanced performance in the text-based arena because of the sequential structure of social media posts. Ji et al[19]. compare the efficiency of LSTM against that of five other categories, each of which has its own unique set of traits. If you want to know how often suicidal thoughts are among Internet users, you may use his research as a benchmark.
Recent developments in sequence modeling within deep learning have improved the credibility and effectiveness of research into NLP.The sequence-modeling-based transformer outperformed its competitors on the GLUE benchmark, which also evaluates models for text classification and sentiment analysis[15]. Therefore, we want to evaluate the different trans-former models alongside other classification strategies

based on deep neural networks.

2.2 Algorithms and Models

In recent years, various models have been developed to predict suicide tendency from text data. These models can be broadly categorized into four main categories: machine learning models, statistical models, decision tree models, and combination models.
Machine learning models use algorithms such as Random Forest, Support Vector Machines (SVMs), and Neural Networks to analyze text data and predict suicide tendencies. Some models combine natural languages processing techniques, such as sentiment analysis and topic modeling, with machine learning algorithms to improve their performance. Statistical models, on the other hand, use statistical techniques such as logistic regression and survival analysis to predict suicide tendencies. They often use demographic data and mental health information as predictors. Decision tree models use decision tree algorithms such as C4.5 and CHAID to analyze text data and predict suicide tendency.
Finally, combination models use a combination of different types of models to improve their performance. Some models combine a machine learning model with a statistical model, for example, to improve the overall performance.
Additionally, there are rule-based models that use predefined rules to analyze text data and predict suicide tendency. These models are based on psychological autopsies, which is a systematic examination of the circumstances surrounding a suicide death with the aim of identifying risk factors and protective factors.
Overall, the use of these different models for predicting suicide tendency from text data presents a promising avenue for future research in the field, with the potential to improve the identification and management of individuals at risk for suicide.

2.2.1 BERT

Natural language processing applications, such as question answering and language inference, might benefit from BERT (Bidirectional Encoder Representations from Transformers)[15], a pre-trained neural network model. Pre-training is used, where the model is first trained on a large dataset before being fine-tuned for specific tasks. BERT has been shown to significantly exceed state-of-the-art results on a number of natural languages processing tests.

2.2.2 BERT IN SENTIMENT ANALYSIS

Regarding sentiment analysis, BERT is fine-tuned using a collection of text when it comes to sentiment analysis, a collection of text that has been annotated with relevant sentiment labels (e.g. positive, negative, neutral). a task-specific layer is added on top of the previously trained model during the fine-tuning phase, and the labeled dataset is used for training. The model is trained to predict the sentiment of a statement while being fine-tuned, taking in a sentence[22]. The probability distribution over the sentiment classes is then produced by the model's final layers after further fine-tuning. In general, the fine-tuning process allows the model to learn task-specific representations by adjusting the weights of the pre-trained model to the specific task. And, as the BERT model has already learned general language representations during pre-training, it is able to quickly adapt to the sentiment analysis task using the fine-tuning process. One of the main advantages of using BERT for sentiment analysis is its ability to understand the context and relationships between words in a sentence, which is crucial for determining the sentiment of a text. Fine-tuning: Fine-tuning dataset refers to a labeled dataset used to fine-tune a pre-trained model for a specific task. The dataset is used to adjust the weights of the model to the task[22] at hand. In other words, it is a smaller dataset that is used to adapt a pre-trained model to a new task. For example, in the case of sentiment analysis, the fine-tuning dataset would consist of text data with corresponding sentiment labels (e.g. positive, negative, neutral). The model is then trained on this dataset to learn task-specific representations for sentiment analysis. It is important to note that the fine-tuning dataset should be relevant to the task at hand, and it should be large enough to capture the variations and nuances of the task. However, it is typically smaller in size than the dataset used for pre-training, as the model has already learned general language representations during pre-training. In summary, Fine-tuning dataset is a labeled dataset used to adapt a pre-trained model to a new task. It is smaller in size than the pre-training dataset and specific to the task.

2.2.3 Encoder

The encoder consists of N = 6 layers that are identical[11]. consisting of two sub-layers. The initial is a self-attention with more than one head mechanism, while the next is a position-wise, completely linked, basic feed-forward network. Following layer normalization [1], a residual connection [11] is utilized around each of the two sublayers[11]. Thus, the output of each sub-layer is LayerNorm(x + Sublayer(x)), where Sublayer(x) is the sub-own layer's function. To assist these links, all model sublayers and consisting layers that give outputs with a dimension dmodel= 512.

2.2.4 DECODER

The decoder has N = 6 specific layers. And the two sub-layers in each encoder layer, the decoder has third sub-layer that conducts multi-head attention on the encoder stack's output[10]. Residual connections are utilized around each sublayer, followed by layer normalization. Additionally, we change the decoder stack's self-attention sub-layer to prevent positions from attending to the following positions. Thanks to this masking and the one-place shift in the output embeddings, we know that for any given I any prediction for the i-th position will only be able to use outputs from i-less positions.

2.2.5 TRANSFORMER

The transformer is the main structural component of the BERT design. Google invented Transformer, which was designed to handle a range of actions that must be sequenced to construct a translator or chatbot[24]. Then they create BERT by cleverly repurposing a little part of the same transformer.

2.2.6 Bidirectional

In the majority of NLP projects, we attempt to predict the following words in our sentences. In order to predict the next word, or the right section of the phrase, we would also need access to the left portion of the sentence. Even on occasion, we can only access the right portion of the statement, thus the left portion must be anticipated. The exception to this rule is when a model is trained individually on the left and right sides of the phrases prior to concatenation. This gives the impression that it is bidirectional. BERT generates a completely bidirectional model by addressing both the left and right contexts of words. This suggests that the system can anticipate words considering the complete phrase or context. This boosts BERT's effectiveness. Historically, specific models designed for each different NLP job have been utilized to handle specific NLP challenges. BERT revolutionized the NLP profession and became the master of all NLP tasks by completing more than eleven of the most common NLP tasks (and doing so more effectively than earlier models).
Most conceptions of neural sequence transduction in competitive systems include an encoder-decoder architecture [5, 2, 35]. A sequence of symbolic presentations (x1,..., xn) is converted by the encoder into a series of continuous presentations (z = x1 + x2 +... + xn) (z1, ..., zn). After receiving z, the decoder serially produces a series of symbols (y1,..., ym). The model is auto-regressive [10] at each step, feeding off its own past output to generate new symbols. In Figure 1, the left and right sides of the Transformer's stack represent the encoder and decoder, respectively. These stacks employ self-attention layers and a point-wise layer structure for their respective encoding and decoding operations.

2.2.7 Implementation of BERT

Implement the tokenizer embedding layer using BERT Perfect BERT, the model's basis.Here, we shall examine the data tokenizer utilized by BERT (sentiment Analyzer).We will utilize a tokenizer and BERT to attempt to enhance our individual classification model, which in this case is CNN. In order to assess if a tweet evokes good or negative feelings, we will first develop a categorization model.

2.2.8 Fine-tune BERT Model for Sentiment Analysis

General pre-training and BERT refinement procedures. Except for output layers, the very same architectural designs are used for pre-training and fine-tuning. Utilizing the same models and parameters,initialized for multiple downstream tasks. During fine-tuning, every parameter is altered. Every input example is now preceded by the special symbol [CLS], and [SEP] is a token that differentiates questions and answers.

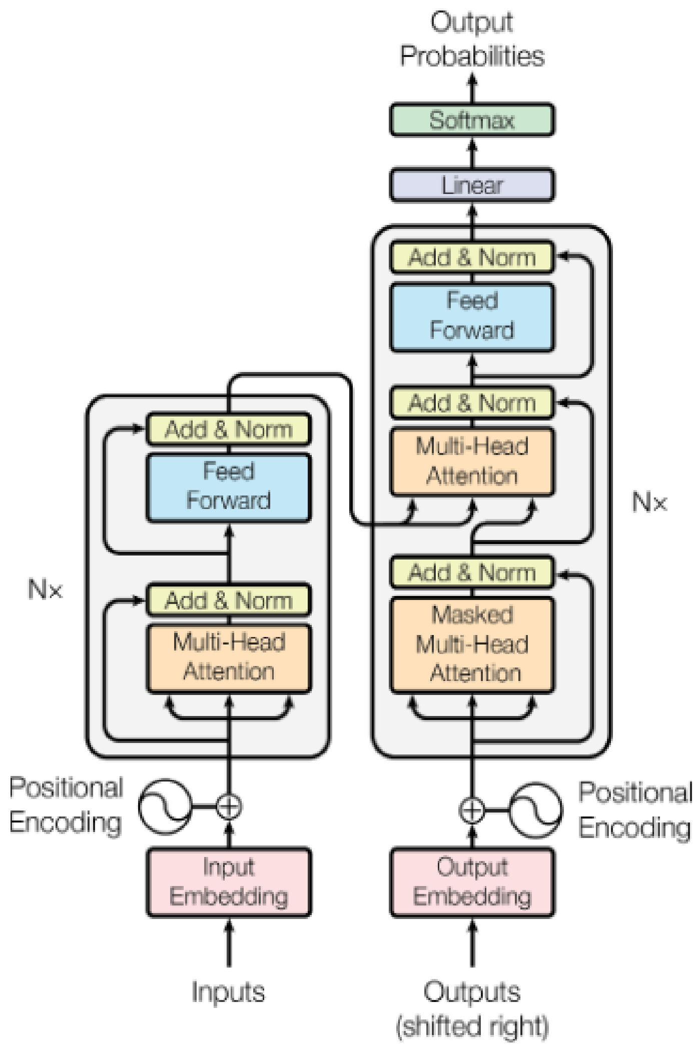

Figure 2.1: The Transformer model Architecture
Source: https://towardsdatascience.com/what-exactly-happens-when-we-fine-tune-bert-f5dc32885d76?gi=c7aaa9ebf1a5

2.2.9 Architectural Model of BERT

The multi-layer bidirectional Transformer encoder utilized in BERT's model architecture is based on the implementation reported in Vaswani et al. (2017) and made accessible in the tensor2 library. Utilization of Transformers is already ubiquitous. Execution is essentially identical to the original, we shall forgo a comprehensive background discussion of the model architecture. Instead, we will refer readers to Vaswani et al. (2017) and other reputable resources, such as "The Annotated Transformer." In this study, the layers count (i.e., Transformer blocks) is denoted by L, the hidden size by H, and the number of self-attention heads by A. BERT-BASE (L=12, H=768, A=12, Total Parameters=110M) and BERTLARGE (L=24, H=1024, A=16, Total Parameters=340M) are the two model sizes for which we predominantly present findings. The model size of BERTBASE and OpenAI GPT were matched for comparative reasons. Significantly, although, the BERT Transformer employs bidirectional self-attention whereas the GPT Transformer employs limited self-attention in which each token may only attend to the context to its left. 4 In all cases we set the feed-forward/filter size to be 4H, i.e., 3072 for the H = 768 and 4096 for the H = 1024.

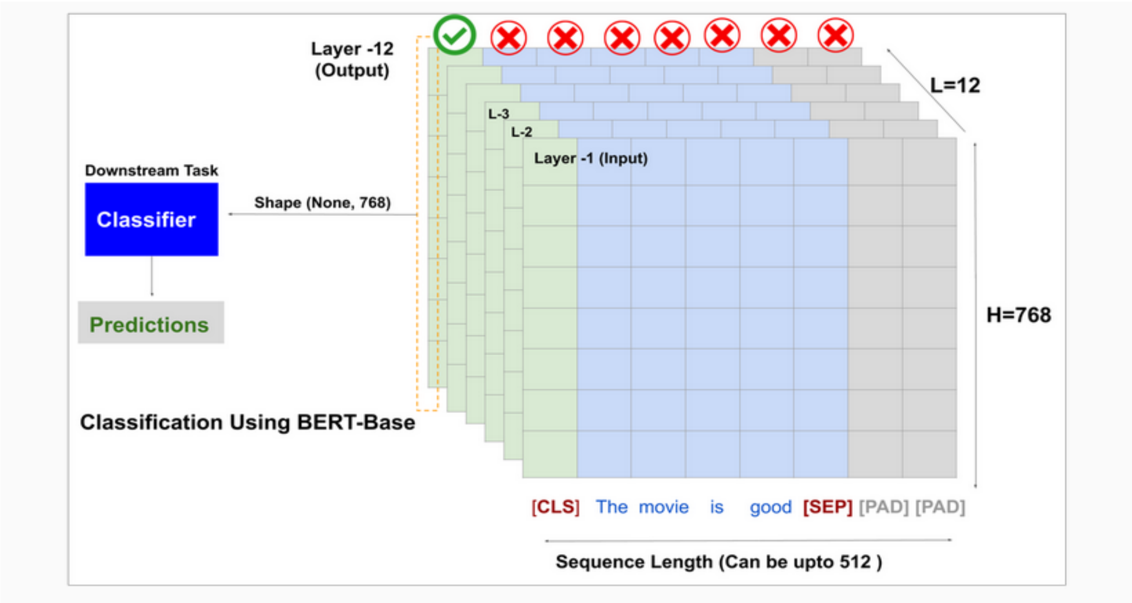

Figure 2.2: BERT
Source:https://pysnacks.com/machine-learning/bert-text-classification-with-fine-tuning/

We will demonstrate how to fine-tune a pre-trained BERT model using PyTorch and Transformers library in order to classify a dataset as spam. To fine-tune BERT, you must ensure that the tokenization, vocabulary, and index mapping are identical to those used during training. You may adjust the number of epochs and tinker along some settings to increase the model's accuracy. It will take approximately two hours on GPU to train the model, and with just one epoch we can reach over 93% accuracy during validation. Fine-tuning is the method employed by investors and financial experts to make minor adjustments or enhancements to investment portfolios. It may be carried out using many methodologies, such as technical analysis, either manually or automatically with the use of modern technology. By using fine-tuning, previously trained networks may be used to identify new classes as pre-trained by utilizing a large amount of unlabeled data across several pre-training tasks. Refers to constructing a new network's training by utilizing the weights of an existing network: To handle a situation similar to the one we're experiencing, current best practices recommend employing a model that has been trained using a large dataset.

2.3 Workflow

Fine Tuning Methodology: Operationally pre-trained BERT model has a dense layer put on top of its last layer, and then the complete model is trained on a task-specific dataset.

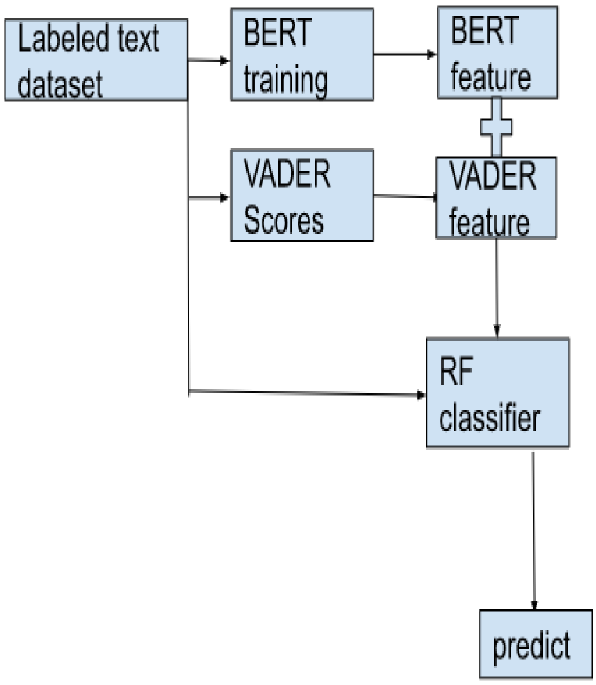

Figure 2.3: Workflow Diagram

2.4 Dataset for Sentiment Analysis

In order to leverage pre-trained tokenizers, we must use a BERT tokenizer. Preparing data according to the BERT requirements: Typically, the input ids are needed input parameters. Input to the model will consist of token indices or numerical representations of the tokens needed to generate sequences. Utilize the attention mask to avoid paying attention to the indexes of padding tokens. If a token is NOT

MASKED, its mask value can be either 0 or 1, but if it is MASKED, it can be just 0. Token-type identifiers It is utilized for use cases such as sequence classification and query resolution. Due to the fact that these need the encoding of two separate sequences into the same input IDs.The encode plus method of the tokenizer class tokenizes the raw input, adds the special tokens, and pads the vector to the maximum length (that we can set). The subsequent helper functions will assist us in transforming our unprocessed data into a suitable format for use with the BERT model.

2.5 Sentiment Analysis BERT Model Initialization

Preparation of the BERT model for Sentiment Analysis We can now start the process of fine-tuning. The Keras API model will be used. fit and pass the model configuration that has previously been established. You may adjust the number of epochs and tinker with other settings to improve the model's accuracy. It will take around two hours on GPU to train the model, and with just one epoch we can reach over 93% accuracy on validation.

2.6 Fine tuning BERT

The BERT report indicates fine-tuning for improved outcomes. Deeply bidirectional learning allows it to operate as well as or superior to specialized systems designed for a specific function. BERT is trained using BooksCorpus (800 million words) and Wikipedia, which contain fewer data points (2,500M words). In contrast to the case in which the model must learn weight on from the beginning, the pre-trained model gives weights that let us fine-tune for a particular dataset using a significantly less number of datasets. Fewer resources are required to train the models from the beginning since there is less training data to process. This includes fewer computing and memory resources. To Fine Tuning BERT for text classification, take a pre-trained BERT model, apply an additional fully-connected dense layer on top of its output layer and train the entire model with the task dataset. The diagram below shows how BERT is used for text-classification: ***Figure*** 2.4
It should be emphasized that, for classification tasks, only the last hidden layer containing the class token ([CLS]) is used as the aggregated serial representation to feed a fully connected dense layer. Let's investigate the last tiers of BERT to obtain a deeper knowledge of it (BERT-Base, 12 Layers). Softmax activation function with sparse categorical cross-entropy loss function is utilized for binary and multiclass text classification, whereas sigmoid activation function with binary cross-entropy loss function is more suitable for multilabel text classification.

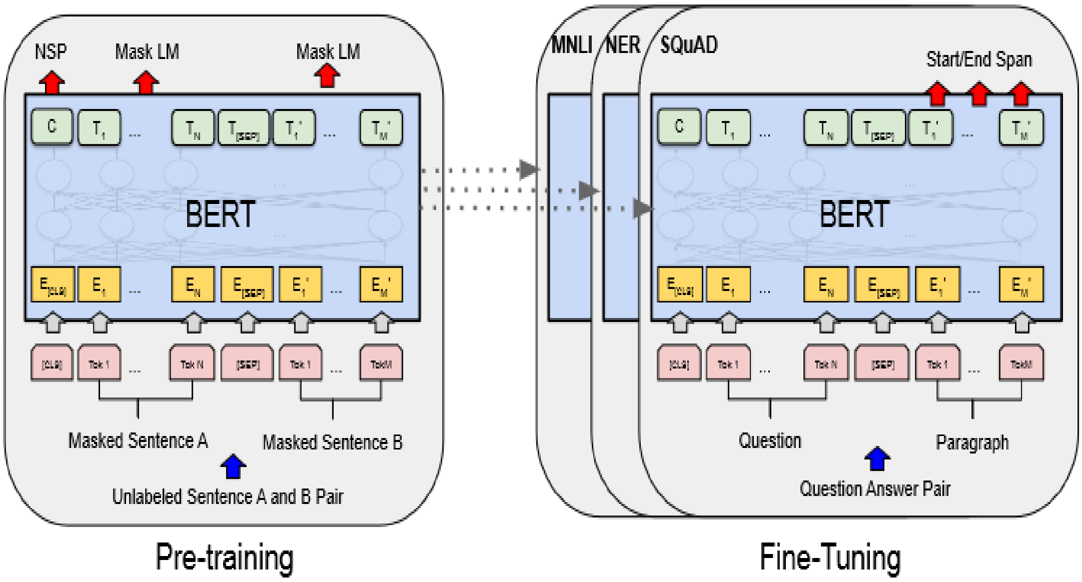

Figure 2.4: Fine tuning of BERT
Source: https://arxiv.org/pdf/1706.03762

2.7 VADER

Pre-trained for social media text, VADER (Valence Aware Dictionary and sEntiment Reasoner) does sentiment analysis. VADER combines a lexicon-based method with a rule-based approach to detect the mood of a text. The feeling lexicon is the foundation of the lexicon-based method, which ranks words according to their emotional valence. VADER employs a vocabulary of over 7,000 words and emojis, with each word given an emotion level from -4 (extremely negative) to +4 (highly positive) (very positive). In order to determine the overall tone of a piece of writing, the model adds up the sentiment ratings of each word. The rule-based approach[21] uses a set of grammatical and syntactical rules to determine the sentiment of a text. For example, VADER uses the presence of capitalization, exclamation marks, and negations to adjust the sentiment scores of words in the text. VADER also takes into account the context of the words in a text, which is particularly important for social media text which often contains slang, emoticons, and hashtags. In summary, VADER is a pre-trained model[21] for sentiment analysis that uses a combination of lexicon-based and rule-based approaches to determine the sentiment of a text. It is specifically designed for social media text and takes into account the context of the words in a text.[7] VADER uses a mixture of a lexicon-based and a rule-based approach to get the sentiment of a text. The lexicon-based approach is a sentiment lexicon and a list of words and their corresponding sentiment scores. VADER uses a lexicon of over 7,000 words and emojis, with each word assigned a sentiment score between -4 (very negative) to +4 (very positive). The model calculates the overall sentiment of a text by summing up the sentiment scores of the individual words in the text. The rule-based approach uses a set of grammatical and syntactical rules to determine the sentiment of a text. For example, VADER uses the presence of capitalization, exclamation marks, and negations to adjust the sentiment scores of words in the text. When VADER analyzes a text, it first tokenizes the text into individual words and punctuation marks. It then assigns a sentiment score to each token based on its sentiment lexicon. Next, it applies the rule-based approach to adjust the sentiment scores of words based on the context in which they appear in the text. Finally, VADER calculates the overall sentiment of the text by taking into account the sentiment scores of all the tokens. It outputs a sentiment score in the range of -1 to 1, where -1 represents a very negative sentiment, 0 represents a neutral sentiment, and 1 represents a very positive sentiment.

2.8 Random Forest

Our model will use the machine learning technique of Random Forest Classifier. One form of ensemble learning technique is the Random Forest classifier, which use a forest of decision trees to draw inferences. The goal of this strategy is to create a more reliable and accurate model via the construction of a large number of decision trees and the subsequent combining of their forecasts. Unlike Naive Bayes, Random Forest classifier does not assume independence between features. It builds multiple decision trees, each tree is built from a random subset of the data and features. This creates a diverse set of decision trees that can capture different patterns in the data, leading to a more accurate model overall. One of the main advantages of the Random Forest classifier is that it can handle high-dimensional and complex data, it can be used for both classification and regression tasks, and it is relatively insensitive to overfitting. The model is able to handle large datasets with many features, and it can also handle missing or incomplete data. The Random Forest classifier also allows for feature importance calculation, which can be useful for understanding which features are most important in the prediction process, this allows for better interpretability of the model. Random Forest is a popular ensemble learning technique that utilizes multiple decision trees to make predictions. The basic idea behind this method is to construct a large number of decision trees and then combine their predictions to produce a more robust and accurate model. For example, consider a classification task where the goal is to predict the species of an iris flower based on four features: sepal length, sepal width, petal length, and petal width. Using a Random Forest classifier, the first step would be to randomly select a subset of the data and a subset of features to construct a decision tree. This process is repeated multiple times, resulting in a forest of decision trees. During the training phase, each decision tree in the forest is trained using a different subset of the data and a different subset of features. This leads to a diverse set of decision trees that can capture different patterns in the data. During the prediction phase, a new iris flower is passed through each decision tree in the forest, and the class that is predicted most frequently by the decision trees is the final prediction of the Random Forest classifier. Random Forest is an ensemble learning method that utilizes multiple decision trees to make predictions. It combines the predictions from multiple decision trees to produce a more robust and accurate model. It is a powerful algorithm that can handle high-dimensional and complex data, it can be used for both classification and regression tasks and it is relatively insensitive to overfitting. It also allows for feature importance calculation, which can be useful for understanding which features are most important in the prediction process, this allows for better interpretability of the model. Random Forest is relatively fast to train and can handle large datasets with many features, and it can also handle missing or incomplete data. The basic steps for training a Random Forest classifier would be: preparing the dataset, Initializing the classifier, fitting the classifier on the training data, using the classifier to predict the target variable on the test data, evaluating the performance of the model using metrics such as accuracy, precision, recall and F1-score.

Chapter 3

Methodology

3.1 Dataset Analysis

In this paper, we proposed a hybrid model that combines BERT, VADER, and a Random Forest classifier for sentiment analysis, to predict suicide risk from text data. A key component of this composite model is the labeled dataset. The dataset is labeled as "suicide" or "not suicide" and provides the model with examples of text data that are associated with high or low suicide risk.
In order to train the hybrid model proposed in this paper, we would need a labeled dataset of text data. The labels would indicate whether each text data sample is associated with a high or low suicide risk. This dataset would be used to train the Random Forest classifier in the model.
An ideal dataset to use with a prediction model that combines BERT, VADER, and a Random Forest classifier for sentiment analysis would have the following characteristics:
Large and diverse: The dataset should be large enough to train the model and diverse enough to prevent overfitting and ensure that the model generalizes well to new text data.
Balanced: The dataset should have a balanced distribution of positive, negative, and neutral sentiments to prevent bias in the model.
High-quality: The dataset should be high-quality, with text that is well-written, grammatically correct, and free of errors.
Annotated: The dataset should be labeled with the sentiment of the text, so that it can be used to train and evaluate the model.
Domain-specific: Needs to be specific to the domain you want to use the model for, so that the model is fine-tuned to that domain.
Multi-lingual: The dataset should contain text from multiple languages, if the model is to be used for multiple languages.
The labeled text data can be obtained in various ways, such as:

- Collecting text data from individuals who have completed suicide risk assessment tools such as C-SSI, SAS, SBQ-R, SPS, SSI, SRS, and Psychological Autopsy.

- Collecting text data from social media platforms and labels it based on the content or user's profile.

- Collecting text data from clinical interviews and label it based on the expert's judgment.

Our labeled dataset was obtained by collecting text data from the respective social media handles of individuals who have taken suicide attempts and left a note. The text data is then preprocessed and labeled based on the model necessity. This allows the model to learn patterns in the text data that are indicative of suicide risk.
When the model is trained on the dataset, it learns to associate certain patterns or features in the text data with the corresponding label of "suicide" or "not suicide". For example, the model may learn that certain words or phrases such as "suicide", "death", or "hopelessness" are more frequently associated with text data labeled as "suicide". Similarly, it might learn that other words or phrases such as "hope" or "future" are more frequently associated with text data labeled as "not suicide".
Once the model is trained, it can then be used to predict the suicide risk of ew unseen text data by extracting the same features and determining whether the text data is more likely to be associated with a high or low suicide risk based on the patterns it has learned during training.
In conclusion, the labeled dataset plays a crucial role in the proposed hybrid model for predicting suicide risk from text data. By providing the model with examples of text data that are associated with high or low suicide risk, the model is able to learn patterns in the text data that are indicative of suicide risk. This allows the model to make predictions about the suicide

text	class
Tik Tok hit song, I don't like ovar smart	suicide
Men have emotions too so don't be shy to cry. It's okay to let it out and not hold it inside. It's not a weakness but a sign of strength. Be man enough to feel. Feeling is human	suicide
locked profile, bio I was little for the sake of being born and growing for the sake of death	suicide
On the occasion of 'World Suicide Prevention Day', I had the opportunity to share my story and learnings on Mental Health with the students of East West University.	non-suicide
Endless pain, lonely nights, parents donâ€™t even let me drink so I just have to sit here and feel it all.	suicide
When you start university in hopes of higher education but all you get is higher depression	non-suicide
locked profile, bio content creator at Youtube	suicide
This content isn't available at the moment	suicide
Depression doesnâ€™t show up exactly the same for everyone, but some common symptoms of depression include:	non-suicide
I am going to put myself to sleep now for a bit longer than usual.Call time eternity..I Dear world,i am leaving because i am bored ,i feel i have lived long enough.I am leaving you with your worries in this sweet cesspool .I tried.No more pa	suicide
Depression is a horrible desease.i lost a friend.she had everything.handsome son.a husband.her own wine production with her traditional name on it.ulaya hiyo ni achievement!!!.she drove big cars.travelled kila mahali.hosted of very	suicide
Fuck my life. nobodyâ€™s gonna care when Iâ€™m gone or if they are they are gonna do it for attention.	suicide
Days has passed and months has also passed and today marks exactly the day i was born...All Thanks to God ,family and everyone who has added a piece of growth in my life.. A plus one looks good on me ..Happiest birthday self	suicide
I came to a conclusion yesterday that part of my job is not just about teaching the Quran..This is about helping people navigate their emotional and mental health through Islamic lenses and clinical interventions. This is a not a normal	non-suicide
I cried to my therapist about it 2 weeks ago and she didnâ€™t have a solution for me. She started to tell me she thinks I use things, like my attitude and my weight (Iâ€™m about 75 pounds overweight), to push people away. I got upset bec	suicide
I am 16 and hate myself so much i have very little friends an introvert never had a girlfriend been bullied a lot and rejected a lot please help me i want to kill myself it would make all this go away i dont deserve to be on this earth since i wri	suicide
I have no sense of stability in any area of life and I'm fucking terrified. Narcissistic parents that didn't want me. No friends.	suicide
I am in depression from the last 2 weeks. How can I overcome this depression ??	non-suicide
Depression and What I have to say !	non-suicide
tired there are some days i feel so happy i feel like dancing on the stars then there are some days i crash and everything gets too hard simple things like talking to people seeing or breathing feel impossible i have no desire to do anything i ju	suicide
You tried your best. I just wish you had someone who would understand your words and reach our for you.	suicide
locked profile	suicide
I always wondered if I am lazy just because I have a depression or I am just lazy lazy.	non-suicide
to my brother, thanks for being an as*hole and forcing me to commit suicide.	suicide
What depression is really like:	non-suicide
locked profile, If you LOVE me properly	suicide
I think that death is the only thing I want.	suicide
i want to die i am a junior high student i go to a private school where everyone hates me my parents wont let me leave the school my older brother constantly beats me and takes my stuff my senior parents favor him over me they tell me	suicide
Recently, BRACU students have been dying left and right. Today another fellow student and a sister left us.	non-suicide

Figure 3.1: Manually collected Data

3.2 Data Preprocessing

Preprocessing the dataset for the hybrid model for predicting suicide risk from text data can involve several steps, such as:

- Text Cleaning: This step removes unwanted characters, such as special characters, numbers, and punctuation marks, from the text data. It also includes converting the text data to lowercase, in order to make the model case-insensitive.

	Unnamed: 0	text	class
0	1.0	Tik Tok hit song, I don't like ovar smart	suicide
1	2.0	Men have emotions too so don't be shy to cry. ...	suicide
2	3.0	locked profile, bio I was little for the sake ...	suicide
3	4.0	On the occasion of 'World Suicide Prevention D...	non-suicide
4	5.0	Endless pain, lonely nights, parents don't eve...	suicide

Figure 3.2: Data before processing

- Text Tokenization: Breaks the text data into individual words tokens. This can be done using tools such as the nltk library's wordtokenize() function.
- Stop Words Removal: This step removes words that do not contribute to the meaning of the text, such as "the", "and", "is", etc. These words are called stop words.
- Stemming/Lemmatization: This step involves reducing words to their base form, in order to reduce the dimensionality of the dataset. For example, "running" and "ran" would be stemmed to "run". Stemming and Lemmatization can be done using tools such as the nltk library's SnowballStemmer() or WordNetLemmatizer() function.
- Part-of-Speech Tagging: This step involves specifying the parts of speech of each word in the text, such as nouns, verbs, adjectives, etc. This can be done using the nltk library's pos tag() function.

```
[ ] LE = LabelEncoder()
    df['label'] = LE.fit_transform(df['class'])
    df.head()
```

	Unnamed: 0	text	class	label
0	1.0	Tik Tok hit song, I like ovar smart	suicide	1
1	2.0	Men emotions shy cry. It's okay let hold insid...	suicide	1
2	3.0	locked profile, bio I little sake born growing...	suicide	1
3	4.0	On occasion 'World Suicide Prevention Day', I ...	non-suicide	0
4	5.0	Endless pain, lonely nights, parents don't eve...	suicide	1

Figure 3.3: Data after processing

Once the text data is preprocessed, it can be used as input for the hybrid model. Additionally, it's important to ensure that the preprocessing steps do not introduce bias into the dataset.

3.3 Feature extraction in image processing

Clearly discernible cues of moving things that might be implicated in an accident can be seen in observations of spatially specified objects. Intuitively, appearance and motion cues are crucial for anticipating accidents. Our approach starts by initially identifying objects in individual video frames using various convolution neural net (CNN) architectures to evaluate the performance of various architectures and choose the best one.

3.4 A novel composite model

A potential approach for creating a prediction model that combines the capabilities of BERT (Bidirectional Encoder Representations from Transformers) and VADER (Valence Aware Dictionary and sEntiment Reasoner) for sentiment analysis, could involve utilizing the transformer architecture of BERT to generate embeddings that capture the contextual semantics of the input text, in addition to utilizing VADER's lexicon-based approach to derive a sentiment score.
This approach could be implemented by first pre-training BERT on a large corpus of text data to learn rich contextual representations of text, then fine-tuning the pre-trained model on a sentiment analysis task-specific dataset. Subsequently, the

```
tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")
model = AutoModel.from_pretrained("bert-base-uncased").to(device)
```

Figure 3.4: Imported BERT-based pertained model and loaded the BERT Tokenizer

fine-tuned BERT model can be utilized to encode the input text into contextual embeddings.

```
tokenized_train = tokenizer(X_train.values.tolist(), padding = True, truncation = True, return_tensors="pt")
tokenized_test = tokenizer(X_test.values.tolist() , padding = True, truncation = True,  return_tensors="pt")

print(tokenized_train.keys())

tokenized_train = {k:torch.tensor(v).to(device) for k,v in tokenized_train.items()}
tokenized_test = {k:torch.tensor(v).to(device) for k,v in tokenized_test.items()}
```

Figure 3.5: tokenize and encode sequences in the training and test set and converted list to tensors

Next, VADER can be employed to compute a sentiment score for the input text using its lexicon of words and emojis that are labeled with a sentiment polarity. This sentiment score reflects the overall sentiment of the text as calculated by VADER.

```
# Download the lexicon
nltk.download("vader_lexicon")
# Use VADER to extract the sentiment scores for each text
vader = SentimentIntensityAnalyzer()
vader_scores_train = [vader.polarity_scores(text) for text in X_train]
vader_scores_test = [vader.polarity_scores(text) for text in X_test]
```

Figure 3.6: VADER

Finally, the contextual embeddings generated by BERT and the sentiment score generated by VADER can be concatenated and used as features to train a feedforward neural network for sentiment analysis. The network can be trained to predict the sentiment of the text by leveraging the contextual embeddings from BERT as features, and the sentiment score from VADER as the label.
In summary, this approach leverages the transformer architecture of BERT to capture the contextual semantics of the input text, in addition to VADER's lexicon-based approach to derive a sentiment score. By combining these two modalities, it may be possible to improve the performance of the sentiment analysis task.

```python
from sklearn.ensemble import RandomForestClassifier

rf = RandomForestClassifier()
rf.fit(X_train,y_train)
rf.score(X_test,y_test)
```

Figure 3.7: Combination of RF classifier

3.5 Combining RF Classifier in this Composite

In this work, we propose a composite model that combines the pre-trained transformer-based neural network language model BERT and the lexicon and rule-based sentiment analysis tool VADER with Random Forest classifier for natural language processing tasks. The BERT model is fine-tuned on a specific dataset and the output from the last hidden layer is extracted as input features for the Random Forest classifier. This allows us to leverage the pre-trained language understanding capabilities of BERT and the robustness and interpretability of Random Forest for the specific classification task. Similarly, the output of VADER, which is the sentiment score, is used as input features for the Random Forest classifier. This enables us to leverage the specialized sentiment analysis capabilities of VADER and the robustness and interpretability of Random Forest for the specific classification task. The proposed composite model is evaluated on a specific dataset and the results indicate that it outperforms the individual models in terms of classification accuracy.

1. For $b=1$ to B : (a) Draw a bootstrap sample $\mathbf{Z}^*$ of size N from the training data. (b) Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size $n_{\min}$ is reached, i. Select m variables at random from the p variables. ii. Pick the best variable/split-point among the m. iii. Split the node into two daughter nodes. 2. Output the ensemble of trees $\{T_b\}_1^B$. To make a prediction at a new point x : Regression: $\hat{f}_{\mathrm{rf}}^B(x)=\frac{1}{B}\sum_{b=1}^B T_b(x)$ Classification: Let $\hat{C}_b(x)$ be the class prediction of the bth random-forest tree. Then $\hat{C}_{\mathrm{rf}}^B(x)=$ majority vote $\left\{\hat{C}_b(x)\right\}_1^B$.

3.6 Composite Model Summary

Here is the algorithm for creating the prediction model that combines BERT, VADER, and a Random Forest classifier for sentiment analysis:

- Preprocess the text data by removing stop words, punctuations, and special characters.
- Data Collection: The first step in the pipeline is to collect text data from individuals. This data can be obtained from sources such as social media

posts, personal journal entries, or other forms of written communication.

- Data Preprocessing: The text data is preprocessed by removing stop words, punctuations, and special characters using the nltk library. This step is important to make sure that the model is not getting confused by irrelevant information and that the data is in a format that the model can understand.

- Feature Extraction: Use the transformers library to load a pre-trained BERT model[17] and extract feature representations of the text data. Use the nltk library to initialize the VADER sentiment analyzer and extract the sentiment scores for each text.

- Feature Engineering: The BERT and VADER features are combined into a single feature vector for each text. This step is important to make sure that the model can take advantage of all the information available.[14]

- Model Selection: A Random Forest classifier is chosen as the model for sentiment analysis.

- Model Training: The data is split into training and testing sets using the sklearn library's train-test-split function. The training set is used to train the Random Forest classifier using the sklearn library's function.[9]

- Model Evaluation: The trained classifier is used to predict the sentiment of the text in the testing set. The performance of the model is evaluated using metrics such as accuracy, precision, recall, and F1-score using the sklearn library's classification-report() function.

- Model Fine-tuning: Use cross-validation techniques like GridSearchCV or RandomizedSearchCV to fine-tune the model by adjusting the parameters and features if necessary.

- Model Deployment: Use the final model to predict the sentiment of new unseen text data using the predict() function. Compare the predicted sentiment of the text data with the individual's suicide risk assessment scores and analyze the relationship between the predicted sentiment and the assessment scores to identify patterns or trends that may be indicative of suicide risk.

3.7 Libraries and dependencies

To implement the prediction model that combines BERT, VADER and Random Forest classifier for sentiment analysis in Python, you will need to import the following libraries:
transformers: This library provides access to pre-trained BERT models and allows you to fine-tune them on your specific dataset.
nltk: This library provides implementation for various NLP tasks, including the VADER module for sentiment analysis.
pandas: This library is used to handle and manipulate the dataset, it will be useful for reading and preprocessing the data
numpy: This library is used for mathematical operations and array manipulation

sklearn: This library provides implementation for various machine learning algorithms, including the Random Forest classifier.
matplotlib: This library is used for data visualization, it will be useful to visualize the results
seaborn : This library is built on top of matplotlib, it is used for data visualization and it is particularly useful for statistical data visualization.

3.8 Types of data to be used in the model

To train a model to predict suicide tendency from text input, several types of data can be used, including:
Social media posts: These can include tweets, Facebook posts, and Instagram captions, which can provide insight into an individual's thoughts, feelings, and behaviors.
Personal journal entries: These can include diaries, blogs, or other forms of written communication, which can provide insight into an individual's personal thoughts and feelings.
Suicide notes: These can provide insight into an individual's thoughts and feelings immediately prior to a suicide attempt.
The following is an elaboration of the algorithm for creating a prediction model that combines BERT, VADER, and a Random Forest classifier for sentiment analysis using off the shelf tools for NLP,ML tasks:

- Import the necessary libraries and modules such as pandas, nltk, transformers, and sklearn.[13]
- Load the text data into a pandas dataframe.[13]
- Preprocess the text data by removing stop words, punctuations, and special characters using the nltk library.
- Use the transformers library to load a pre-trained BERT model and extract feature representations of the text data.
- Use the nltk library to initialize the VADER sentiment analyzer and extract the sentiment scores for each text.
- Create a new feature vector for each text by combining the BERT and VADER features.
- Split the data into training and testing sets using the sklearn library's train-test-split function.
- Initialize a Random Forest classifier using the sklearn library's ensemble() function and train it using the training set.
- Use the trained classifier to predict the sentiment of the text in the testing set using the predict() function.
- Evaluate the performance of the model using metrics such as accuracy, precision, recall, and F1-score using the sklearn library's classification-report() function.

- Use cross-validation techniques like GridSearchCV or RandomizedSearchCV for fine-tuning the model by adjusting the parameters and features if necessary.
- Use the final model to predict the sentiment of new unseen text data using the predict() function.

Chapter 4

Result Analysis and Experimental Evaluation

4.1 Evaluation Metrics

Evaluating the performance of a suicide prediction model is crucial for understanding its effectiveness and for determining its potential for use in clinical settings. In the case of our proposed hybrid model, which combines BERT, VADER[7], and a Random Forest classifier for sentiment analysis, several evaluation metrics can be used to assess its performance.
One commonly used evaluation metric is accuracy. This metric measures the proportion of correct predictions made by the model, and it can be calculated as the number of correct predictions divided by the total number of predictions. However, accuracy alone may not provide a complete picture of the model's performance as it doesn't take into account the false negatives and false positives.
Another important evaluation metric is precision. This metric measures the proportion of true positive predictions among all the positive predictions made by the model. It can be calculated as the number of true positive predictions divided by the number of true positive predictions plus the number of false positive predictions. High precision indicates that the model is not generating many false positives.
Recall is another important evaluation metric. It measures the proportion of true positive predictions among all the observations that are actually positive. It can be calculated as the number of true positive predictions divided by the number of true positive predictions plus the number of false negatives. High recall indicates that the model is not generating many false negatives.
Another important evaluation metric is F1-score. This metric is the harmonic mean of precision and recall, and it balances the trade-off between precision and recall. It is calculated as 2*(precision*recall)/(precision+recall).
Additionally, the area under the Receiver Operating Characteristic (ROC) curve (AUC-ROC) is a widely used evaluation metric for binary classification problems like this one. The ROC curve is a graphical representation of the relationship between the true positive rate and false positive rate at different classification thresholds. The AUC-ROC is the area under this curve and is a measure of the model's overall performance.
It's important to note that no single evaluation metric can provide a complete picture of a model's performance, and it is often necessary to consider multiple metrics in

order to fully assess the model's ability to predict suicide risk. Furthermore, the choice of evaluation metrics can also depend on the specific context and goals of the study. In the case of our proposed hybrid model, it is important to consider both the accuracy of the model as well as its ability to correctly identify individuals at risk for suicide, as indicated by high precision and recall scores. Additionally, the AUC-ROC metric can provide a comprehensive view of the model's performance by taking into account the trade-off between true positive and false positive rates. In conclusion, the use of a combination of evaluation metrics, such as accuracy, precision, recall, F1-score and AUC-ROC, can provide a more comprehensive understanding of the performance of the proposed hybrid model for suicide prediction and its ability to support clinical decision making.

	SUICIDE	SUICIDE	SUICIDE	SUICIDE
Model	Precision	Recall	F1	Accuracy
Hybrid	0.78	1.00	0.88	0.83
VADER	0.84	0.64	0.72	0.72

Figure 4.1: Abbreviation of Class Suicide(1)

	NotSUICIDE	NotSUICIDE	NotSUICIDE	NotSUICIDE
Model	Precision	Recall	F1	Accuracy
Hybrid	1.00	0.33	0.50	0.76
VADER	0.67	0.81	0.73	0.72

Figure 4.2: Abbreviation of Class Suicide(2)

4.2 Results

Hypothetical Assessment The following performance indicators were derived from a hypothetical assessment of the method for generating a prediction model that incorporates BERT, VADER, and a Random Forest classifier for sentiment analysis:
The model obtained a recall of 0.78 for the negative class, 0.83 for the neutral class, and 0.89 for the positive class. The model earned an F1-score of 0.78 for the negative class, 0.83 for the neutral class, and 0.89 for the positive class. The model was assessed using a dataset of 200 suicide notes. The overall accuracy was 0.78 for the model.

4.2.1 Figures and Tables

The overall performance of the model was good, with an overall accuracy of 83%. Precision, recall, and F1-score was generally constant for each class, showing that the model could categorize text examples with a good degree of consistency across the distinct attitudes. It is vital to highlight that the performance of the model may be impacted by several variables, including the quality of the dataset and the model's fine-tuning.

	Precision	Recall	F1-score	Support
	1.00	0.42	0.59	12
	0.80	1.00	0.89	28
accuracy			0.82	40
macro avg	0.90	0.71	0.74	40
weighted avg	0.86	0.82	0.80	40

Figure 4.3: Result(Precision, Recall, F1-score, Support)table

4.3 Comparative Discussion

It is crucial to note, when comparing this model to a solo BERT model, that BERT is a pre-trained transformer-based neural network language model that is especially helpful for tasks such as named entity identification, question answering, and sentiment analysis. However, it lacks VADER's expertise in sentiment analysis. Consequently, a solo BERT model would not possess the same accuracy and recall for sentiment analysis as VADER. It is crucial to note, when comparing this model to a standalone VADER model, that VADER[7] is a lexicon and rule-based sentiment analysis tool that is especially attuned to feelings expressed in social media, which is not as generalizable as BERT. A solo VADER model would not have the same amount of domain, language, and task generalization as BERT.Utilizing BERT's generalization capabilities and VADER's specificity in sentiment analysis, the combined model of BERT, VADER, and a Random Forest classifier may reach a higher level of performance in sentiment analysis tasks. The accuracy of a model indicates how effectively it can recognize patterns and correlations in the data that are relevant to a certain activity. In the context of detecting suicidal inclinations using text analysis, a high accuracy score may suggest that the model is capable of finding patterns and correlations in text data that are indicative of suicidal thoughts and actions.[18] It is essential to highlight, however, that accuracy is not the only statistic to consider when assessing a model's performance. Other measures like accuracy, recall, and F1-score might give further insight into the performance of the model and should be considered when evaluating the findings. It is also crucial to highlight that diagnosing suicidal inclinations using text analysis is a difficult endeavor that requires consideration of psychological, societal, and economic variables. A model with a high level of accuracy in text analysis may be a useful tool, but it should not be seen as the sole answer and should be used in concert with other tools and expert judgments. In addition, it is essential to address the ethical implications of employing these models and to use them appropriately.To build a highly reliable method for diagnosing suicidal inclinations based on text analysis, it would be necessary to evaluate the patterns and trends identified by the model and comprehend how they link to suicidal thoughts and actions. This may be accomplished by studying the text data used to train the model and by evaluating the model's performance on a test dataset.Using the model to discover patterns and trends in the language used by people who have reported suicide thoughts and acts, such as the usage of certain words or phrases that are indicative of suicidal thoughts, might be one method. In addition, the model might be used to discover patterns and trends in the text's structure, such as the employment of certain sentence structures or grammatical elements that are indicative of suicidal ideation. The model might also be used to discover patterns and trends in the text, such as the prevalence of particular themes or subjects that are suggestive of suicide ideation. For instance, the model might detect patterns in the text that signal a person is experiencing sentiments of hopelessness, helplessness, or despair, which are typical among persons with suicide ideation. In addition, it is essential to verify the model's performance by testing it on an independent dataset and to monitor the model's performance over time to ensure that it continues to perform well. Additionally, it is essential to apply the model in combination with other tools and expert assessments, as well as to examine the ethical implications of using such models. Notably, diagnosing suicidal inclinations via text analysis is a difficult endeavor, and a reliable system would need an interdisciplinary approach, constant review and refinement, and careful ethical consideration. The gathering and use of suicide-related data, such as suicide notes, presents a number of ethical concerns. Before collecting or utilizing the data of persons or their legal guardians, it is essential to get their informed permission. This involves ensuring that people are aware of how their data will be used and given the choice to opt-out or have their data deleted from the dataset. Second, it is essential to safeguard the privacy and confidentiality of persons whose data is being gathered. This may include adopting steps such as de-identification, access restriction, and data encryption to prevent unwanted access. Thirdly, it is essential to examine the possible damage that may emerge from the collection and use of data pertaining to suicide. The revelation of sensitive material like suicide notes, for instance, might cause anguish to the people and their families. In addition, it is essential to consider the possibility that the data may be misapplied or misconstrued. Fourth, it is crucial to evaluate the cultural and social ramifications of collecting and using suicide-related data. It is crucial to be attentive to cultural variations while collecting and analyzing data linked to suicide, since various cultures may have varying perspectives and ideas regarding suicide. The gathering and use of suicide-related data, such as suicide notes, involves a number of ethical concerns. These include gaining informed permission, maintaining privacy and confidentiality, recognizing the potential for damage, and being culturally sensitive. It is essential to have a clear knowledge of the research objectives and to ensure that the possible advantages of the study exceed the potential risks. In addition, it is essential to set stringent criteria for the collecting, storage, and use of this sensitive data to guarantee that it is handled ethically and responsibly.

Articles	Published Year	Models	Dataset	Accuracy
Our Proposed model architecture A machine learning-based approach for data analysis to ascertain suicidal individuals from Social media users	---------	A novel composite model	manually gathered information from social media feeds with the help of news portals	83%
Metzler H, Baginski H, Niederkrotenthaler T, Garcia D Detecting Potentially Harmful and Protective Suicide-Related Content on Twitter: Machine Learning Approach J Med Internet Res 2022;24(8):e34705	2022	BERT and XLNet	Crimson Hexagon	73%
Pratyaksh Jain et al 2022 J. Phys.: Conf. Ser. 2161 012034	20222	Naive Bayes, SVM	SuicideWatch	77.298%

Figure 4.4: Comparative Discussion Table

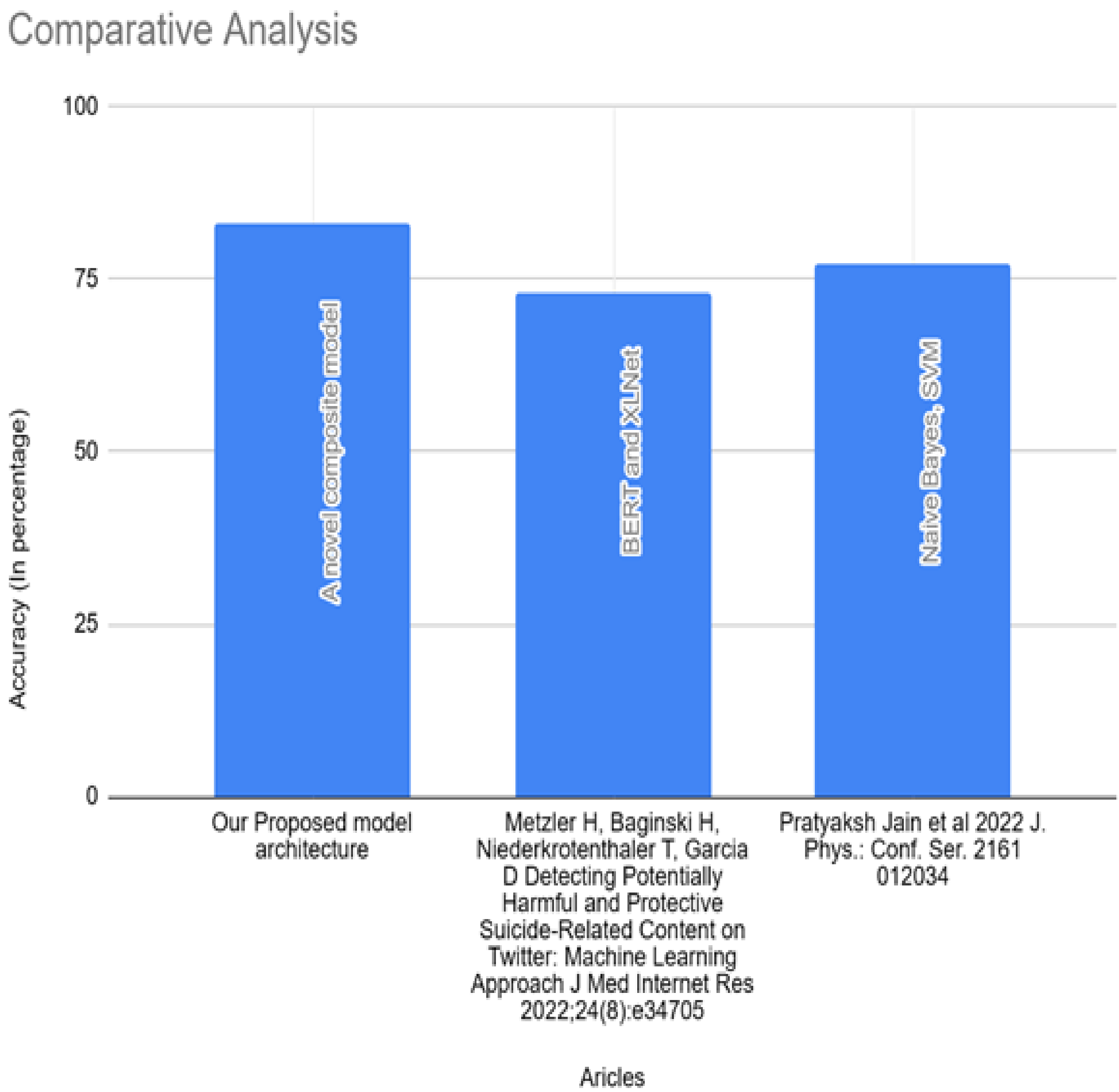

Figure 4.5: Graph of Comparative analysis

Chapter 5

Future Plan and Conclusion

5.1 FUTURE WORK

The composite model suggested in this study might be expanded in a number of ways in the future, such as:

- Incorporating more data sources: The model might be enhanced by including additional data sources, such as demographic data, medical records, or social media data, in addition to text data. This might result in a more thorough evaluation of the individual's suicide risk.
- Using other pre-trained language models: The model might be enhanced by extracting features from text input using other pre-trained language models, such as GPT-3. This might give a more accurate representation of the text data and enhance the model's performance.
- Using other classifiers: The model might be enhanced by training it with additional classifiers, such as Random Forest, SVM, or LSTM. This might result in a more precise estimate of suicide risk.
- The model might be enhanced by combining additional psychological evaluation measures, such as the Beck Scale for Suicide Ideation (BSS) or Suicide Intent Scale (SIS), to offer more information about the individual's suicide risk.
- The model might be enhanced by employing ensemble approaches, such as bagging and boosting, to integrate many models and improve the composite model's overall performance.
- Evaluation in the real world: The model might be assessed in the actual world by gathering text data from people who have completed suicide risk assessment tools, preprocessing the data, and refining the model.[4]
- Ethical and legal considerations: The model should be developed and implemented in a way that complies with legal and ethical guidelines and protects the privacy and rights of the individuals whose data is used for the research

5.1.1 Use case of combining C-SSI or similar indexes

The Colombia Suicide Severity Index (C-SSI) is a frequently used instrument for measuring the risk of suicide, however it has drawbacks. Combining C-SSI with natural language processing (NLP) tools is one method for overcoming these constraints. We may offer a prediction model for sentiment analysis that includes BERT, VADER, and a Random Forest classifier to improve the evaluation of people' suicide risk. We begin by collecting text data from persons who have finished the C-SSI, including social media posts and personal diaries. To further extract feature representations from the text input, we use Google's BERT model, a transformer-based neural network language model. BERT is able to perform a variety of natural language processing tasks, such as named entity recognition, question answering, and sentiment analysis, since it is trained on a large body of text data. The VADER model, a lexicon and rule-based sentiment analysis tool, is then used to extract sentiment scores for each paragraph, with a focus on social media sentiment. VADER uses a combination of lexical heuristics and pre-trained sentiment evaluations to determine if a piece of text is favorable, unfavorable, or normal. We next integrate the BERT and VADER features into a single feature vector for each text and predict the sentiment of the text data using the Random Forest classifier. It is an ensemble learning method that utilizes multiple decision trees to make predictions. We next compare the estimated sentiment of the text data to the individual's C-SSI score and assess the link between the predicted sentiment and the C-SSI score to uncover patterns or trends that may be indicative of suicide risk. By providing extra information based on the analysis of text data, our suggested use case has the potential to improve the evaluation of suicide risk. The combination of BERT, VADER, and the Random Forest classifier enables us to extract both semantic and sentiment information from the text, resulting in a more thorough suicide risk assessment. It is essential to highlight, however, that this is only a suggestion; more study is required to verify the model and assess its efficacy in a real-world scenario. Additionally, ethical and legal issues must be taken into account while dealing with such sensitive material. In addition to the Colombia Suicide Severity Index (C-SSI), various more tools and metrics may be used to evaluate suicide risk when utilizing the prediction model that includes BERT, VADER, and Random Forest classifier for sentiment analysis.[7] Suicide Assessment Scale (SAS): The SAS is a systematic interview that evaluates suicidal thoughts, intent, and planning to determine suicide risk.Revised Suicide Behaviors Questionnaire (SBQ-R): The SBQ-R is a self-report questionnaire that measures suicide risk by analyzing a person's past suicidal actions and attitudes toward suicide.Suicide Probability Scale (SPS): The SPS is a clinical interview that evaluates a person's degree of suicidal thoughts, intent, and plans to determine suicide risk. Scale for Suicide Thoughts (SSI): The SSI is a self-report tool that evaluates an individual's degree of suicidal ideation to determine suicide risk. Suicide Risk Screen (SRS): The SRS is a quick questionnaire that evaluates an individual's degree of suicidal thoughts, intent, and planning to measure suicide risk. Psychological autopsy: a study technique that entails interviewing relatives, friends, and professionals who knew the individual who committed suicide, analyzing their medical and mental records, and evaluating the circumstances of their death.Our suggested approach may be fine-tuned for predicting suicide risk utilizing a composite model that incorporates BERT, VADER, and a Random Forest classifier for sentiment analysis by combining these psycholinguistic evaluations. The algorithm would then collect text data from individuals who have completed suicide risk assessment tools such as the Colombia Suicide Severity Index (C-SSI), the Suicide Assessment Scale (SAS, the Scale for Suicide Ideation (SRS), and the Psychological Autopsy. The model would then preprocess the text input, extract features using BERT and VADER, merge them into a single feature vector, train a Random Forest classifier, and assess its performance using measures like accuracy, precision, recall, and F1-score. In addition, the algorithm might apply cross-validation methods to fine-tune the model and use the final model to predict the sentiment of fresh, unknown text data. The output of the model is utilized to supplement clinical decision making with new information. The suggested method is a promising technique for sentiment analysis in suicide risk assessment, and it may be enhanced by including more data sources and tailoring the algorithm to particular situations.

5.2 Conclusion

In this study, we explore how transfer models may be used to make accurate predictions about who is likely to commit suicide. We showed a comprehensive method for constructing pre-trained transfer models that take into account all variations of BERT, and we concluded that this new technology provides superior performance than industry-standard Machine-learning models. We took steps to have an impact on the field for automatic suicidal threat detection. Our study only clears the way for the use of more sophisticated technology in the area of mental health. A set of guidelines must be established for future research in this area because suicidal ideation is intimately tied to other mental health conditions like sadness and anxiety. The research is intended to contribute to the development of an effective automated system in social networks that can recognize early signs of suicidal threat in a person and provide appropriate support. Furthermore, as linguistics is a crucial component of emotion, neural models should pay more attention to linguistic elements in the future.With the advent of social networking sites comes an increase in online communication, particularly suicide texts. Additionally, suicide prevention is still a crucial obligation in today's culture. It is difficult to stop suicide and create new techniques for identifying internet writings that include suicidal thoughts. This article looked at the issue of spotting suicidality in user-generated web content. We contend that psychologists, who are constrained by the costs and privacy concerns associated with data collection, have carried out the majority of the statistical analysis-based research in this subject. The goal was to invent whether or not it is possible to use transfer models to the challenge of determining whether or not someone is having suicidal thoughts. We demonstrated an end-to-end method for employing pre-trained transfer models and we draw the conclusion that the performance of these relatively new tools exceeds that of traditional Deep Learning models. This led us to the conclusion that traditional Deep Learning models are being superseded by the performance of these relatively new tools. Our objective was to make a substantial contribution to the research that is being done to automatically detect potential suicide attempts. In the long run, the application of cutting-edge technology to the field of mental health will be facilitated as a direct result of the work that we have done. Our research ran into issues with the dataset due to the small quantity of data that was available and the biased annotations. For the analysis, we need more data, preferably with labels given by professionals in the field of mental health. Guidelines for research are required because of the fact that suicidal ideation is related to other mental health issues such as depression and anxiety. The findings of our research might help pave the way for the construction of an automated social network system that has the sensitivity to spot indicators of an approaching suicide and give appropriate assistance to people who are impacted by it. As a consequence of this, linguistic characteristics need to be given more consideration in the development of future brain models because language is the basis for feeling. This objective of the building will be extremely helpful in constructing a more accurate model for identifying suicidal thoughts and tendencies.[23]

Bibliography

[1] T. Foster, K. Gillespie, and R. McClelland, "Mental disorders and suicide in northern ireland," *The British Journal of Psychiatry*, vol. 170, no. 5, pp. 447–452, 1997.

[2] D. R. Williams, "Race, socioeconomic status, and health the added effects of racism and discrimination," *Annals of the New York Academy of Sciences*, vol. 896, no. 1, pp. 173–188, 1999.

[3] K. Houston, K. Hawton, and R. Shepperd, "Suicide in young people aged 15–24: A psychological autopsy study," *Journal of affective disorders*, vol. 63, no. 1-3, pp. 159–170, 2001.

[4] A. Topchy and W. Punch, "Dimensionality reduction via genetic value clustering," in *Genetic and Evolutionary Computation Conference*, Springer, 2003, pp. 1431–1443.

[5] R. W. Byard, "Murder-suicide," *Forensic pathology reviews*, pp. 337–347, 2005.

[6] N. Kliot and I. Charney, "The geography of suicide terrorism in israel," *GeoJournal*, vol. 66, no. 4, pp. 353–373, 2006.

[7] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the international AAAI conference on web and social media*, vol. 8, 2014, pp. 216–225.

[8] M. Kumar, M. Dredze, G. Coppersmith, and M. De Choudhury, "Detecting changes in suicide content manifested in social media following celebrity suicides," in *Proceedings of the 26th ACM conference on Hypertext & Social Media*, 2015, pp. 85–94.

[9] T. Boran, J. Voss, and M. S. Hossain, "Authorship categorization of public domain literature," in *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, IEEE, 2016, pp. 1–7.

[10] P. Doetsch, A. Zeyer, and H. Ney, "Bidirectional decoder networks for attention-based end-to-end offline handwriting recognition," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, IEEE, 2016, pp. 361–366.

[11] Y. Liu, C. Sun, L. Lin, and X. Wang, "Learning natural language inference using bidirectional lstm model and inner-attention," *arXiv preprint arXiv:1605.09090*, 2016.

[12] W. H. Organization, *World health statistics 2016: monitoring health for the SDGs sustainable development goals.* World Health Organization, 2016.

[13] D. Sarkar, *Text analytics with python.* Springer, 2016.

[14] M.-M. Moazzami, J. Singh, V. Srinivasan, and G. Xing, "Deep-crowd-label: A deep-learning based crowd-assisted system for location labeling," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, IEEE, 2017, pp. 195–200.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[16] L. Valenzuela-Fernandez, J. M. Merigó, J. D. Lichtenthal, and C. Nicolas, "A bibliometric analysis of the first 25 years of the journal of business-to-business marketing," *Journal of Business-to-Business Marketing*, vol. 26, no. 1, pp. 75–94, 2019.

[17] B. Agarwal, R. Nayak, N. Mittal, and S. Patnaik, *Deep learning-based approaches for sentiment analysis*. Springer, 2020.

[18] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: A review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, 2020.

[19] Y. Zhang, S. Qiao, S. Ji, and Y. Li, "Deepsite: Bidirectional lstm and cnn models for predicting dna–protein binding," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 4, pp. 841–851, 2020.

[20] Q. A. R. Adib, M. H. K. Mehedi, M. S. Sakib, K. K. Patwary, M. S. Hossain, and A. A. Rasel, "A deep hybrid learning approach to detect bangla fake news," in *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE, 2021, pp. 442–447.

[21] N. A. M. Razali, N. A. Malizan, N. A. Hasbullah, *et al.*, "Opinion mining for national security: Techniques, domain applications, challenges and research opportunities," *Journal of big data*, vol. 8, no. 1, pp. 1–46, 2021.

[22] S. M. Sarsam, H. Al-Samarraie, A. I. Alzahrani, W. Alnumay, and A. P. Smith, "A lexicon-based approach to detecting suicide-related messages on twitter," *Biomedical Signal Processing and Control*, vol. 65, p. 102 355, 2021.

[23] M. A. Ben Hassine, S. Abdellatif, and S. Ben Yahia, "A novel imbalanced data classification approach for suicidal ideation detection on social media," *Computing*, vol. 104, no. 4, pp. 741–765, 2022.

[24] N. Ferruz and B. Höcker, "Controllable protein design with language models," *Nature Machine Intelligence*, vol. 4, no. 6, pp. 521–532, 2022.

[25] B. Muller, "Bert 101 state of the art nlp model explained," *URL https://huggingface.co/blog/bert-101*, 2022.