Analysis of suicide ideation documents posted on Twitter using an NLP classifier

MSc Research Project
Data Analytics

Rachana Swamy
Student ID: X21111413

School of Computing
National College of Ireland

Supervisor: Dr Christian Horn

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Rachana Devnur Swamy		
Student ID:	X2111413		
Programme:	Data Analytics (MSCDAD_JAN22B_O)	**Year:**	2022
Module:	Research Project		
Supervisor:	Dr Christian Horn		
Submission Due Date:	15th December 2022		
Project Title:	Analysis of suicide ideation documents posted on twitter using an NLP classifier		
Word Count:	6008 **Page Count** 21		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Rachana Devnur Swamy

Date: 15-12-2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Analysis of suicide ideation documents posted on Twitter using an NLP classifier

Rachana Swamy

X21111413

Abstract

One of the top causes of mortality, suicide currently accounts for 828,000 fatalities worldwide every year, up from 712,000 in 1990. As a result, suicide is currently the ninth leading cause of death worldwide. The number of research suggesting that social media and the Internet may have an impact on actions related to suicide is also growing. To assess whether or not a text contains suicidal ideas, researchers created a very simple suicidal ideation classification using Natural Language Processing (NLP), a subfield of Machine Learning (ML). Social networks, which allow users to contact friends and family, are crucial tools for learning people's viewpoints on many subjects. In recent years, the prominence of the study issue in the domains of NLP and psychology has greatly increased: the identification of suicidal thoughts through online social network analysis. The complex early signs of suicidal thoughts may be recognised with the appropriate use of social media data, possibly saving countless lives. Everyone was able to contact and share their thoughts and emotions with millions of people worldwide thanks to the quick growth of social media websites and technology. Social media websites like Google+, Instagram, Facebook, Twitter, and LinkedIn have become essential communication conduits. Users of these websites may produce, distribute, and receive information among a large group of individuals. Although these platforms have benefits, there are certain user safety concerns related to how they are established and how disclosing suicidal ideas are handled. Machine learning models have been employed for the classification of the tweets posted as positive or negative tweets. The models are evaluated using the evaluation parameters. The Support Vector Machine (SVM) classifier outperforms other conventional ML methods in terms of model performance, with an accuracy score of 92% and a 0.92 F1 score. The performance of the Stochastic Gradient Descent (SGD), and Logistic Regression (LR) classifiers was somewhat worse than that of the SV classifier, which achieved an accuracy score of between 91% and 91% and an F1 score of 0.91.

1 Introduction

1.1 Background

One of the most serious mental health problems that people experience worldwide is suicidal thoughts. Suicide can result from a variety of factors (Faryal *et al.*, 2021). Despondency, anxiety, and social isolation are their most prevalent and important risk factors. Early risk factor identification can reduce or even stop suicide. Twitter and other online social networking platforms are being used by more people as safe spaces where they can express themselves without worrying about being rejected by others (Aldhyani *et al.*, 2022). This study provides a technique and experiment for more effectively identifying suicidal thoughts and reducing the likelihood that someone would experience this terrible mental disease. The World Health Organization (WHO) estimates that over 800,000 individuals commit suicide annually, with one doing so on average every 40 seconds. 135,000 of the fatalities attributed to suicide occurred in India. Suicide is the sixth most common cause of death for adults and the first cause of death for teenagers, according to the WHO. Additionally, there have been 20 times more suicide attempts, which has a detrimental emotional and financial impact on families. Several suicide risk factors have been identified by the American Foundation for Suicide Prevention (AFSP). Desperation, substance misuse, depression, schizophrenia, as well as social factors like isolation, mourning, unemployment, harassment, or abuse, as well as aspects related to traumatic life events like illness, emotional issues, and a history of previous suicide attempts, can all be factors. Suicide is sometimes said to be a "permanent solution to a transitory problem" (Aldhyani *et al.*, 2022). Despite the increase in suicide instances, by recognising risk factors early in the suicidal process, the probability of suicide can be somewhat reduced. Nowadays, Twitter is seen as an emerging area for social science research. There is a searchable record of people's beliefs, deeds, and opinions. Twitter is one of the most popular social media sites for real-time communication. Tweets may include up to 280 characters. Twitter is available everywhere and has no age limits, except in China, North Korea, and Iran (Faryal *et al.*, 2021). It is identified that these posts accurately represent people's genuine sentiments since they freely express their emotions on social media without worrying about being rejected by others. Anyone can view material that has been published by others on Twitter as many accounts are open to the public. 500 million tweets are being sent out every day on this platform, which is used for social networking by approximately 23% of all internet-using people. Twitter developed a method to detect particular tweets and notify the user of the crisis scenario after realising that individuals do in fact post about suicidality. The interpretation and judgement of networked users many of whom might not be able to comprehend the whole risk—are the only sources of this warning message. It has been observed that individuals who are thinking of taking their own lives use a range of language strategies to convey their feelings of despair and hopelessness (Aldhyani *et al.*, 2022). If these tweets are carefully and ahead of time examined, lives could be saved. The purpose of this study was to investigate if it was feasible to discriminate between tweets that indicated a desire to commit suicide and those that did not by evaluating information from social media platforms. Human programmers split the tweets into two groups. To match the accuracy of human annotation, the tweets were then digitally classified using machine learning and aggregation approaches. Models were assessed using recall and accuracy metrics. Users may express their opinions and engage in conversation with one another using web-based solutions on platforms such as Twitter and Instagram (Faryal *et al.*, 2021). Although these websites are helpful, they have a detrimental effect on those who are considering suicide. Several academics have looked at the connection between social connections and people with suicidal ideation. Nowadays, a lot of individuals make it a point to read text messages on social media.

1.2 Research Objectives

- To analyze the suicidal ideation among the people from Twitter.
- To understand the root causes and particulars about kind of people to think about committing suicide.

1.3 Research Questions

1. How to enhance suicide ideation prediction through tweets?
2. What techniques can be used to assess individuals who are actually suicidal?

2 Related Work

2.1 Detection of Suicide Imagination Using Machine Learning

There is a wealth of studies on how to classify shorter messages like tweets. It is still in the early stages of development, nevertheless, to use different machine learning algorithms to distinguish text linked to suicide from material not related to suicide. Studies aiming at classifying and dividing suicidal users from non-suicidal users typically focused on suicide notes. The majority of this study has therefore depended on psychiatric and psychological standards. Another tactic is to use surveys to assess the likelihood of suicidal individuals using clinical approaches. Because data scarcity is a big issue for this study, researchers are employing machine learning algorithms to understand the language of suicidal persons from user-generated content found on the internet.

Haque *et al.* (2022) compiled tweets with keywords connected to suicide using the Twitter API. Three degrees of anxiety were manually applied to an online Twitter dataset. Token unigrams and bag-of-words were the features that the machine learning algorithms were given. The research used artificial intelligence techniques for the first time to replicate human precision. Haque *et al.* (2022) investigated how common major depression (MDD) is among Twitter users. A database of People on Twitter with MDD that use the Centre for Epidemiologic Studies Depression Scale was created utilizing crowdsourcing approaches (CES-D). Depending on how well or poorly they scored on the CES-D for depression, the users were divided into groups. In recent years, several studies have examined the possibilities of identifying suicidal thoughts in persons who use social media. Haque *et al.* (2022) used a statistical method centred on a score-matching model to create markers of the change from mental health discussion to suicidal ideation. The early stage is characterized by anxious thoughts, a sense of helplessness, and despair. The second stage is characterized by a decrease in social cohesion and self-esteem. The third stage is characterised by hostility and a desire to commit suicide. In the weeks preceding a suicide attempt, there was a clear increase in tweets indicating melancholy. According to Rabani *et al.* (2021), long short-term memory (LSTM) and convolutional neural networks (CNN) are Deep Learning (DL) techniques that significantly advance the area of NLP as word embedding gains popularity. Machine Learning (ML) techniques cannot be employed for all applications because of a variety of limitations, including dimension inflation, data sparsity, and excessive processing times. They may significantly improve by incorporating DL techniques, which may extract key features from input data, into traditional machine learning techniques. To obtain great accuracy, a model's layer count might be increased. As a result, the model will offer a more reliable and accurate categorisation.

2.2 Effects of Various Factors of The Suicidal Process

The factors that directly affect suicide behaviour include hormonal imbalance, psychological disorders such as the inability to handle problems, substance abuse, and other neurological issues like depression. These circumstances could lead to the formulation of a suicide plot, an attempted, and finally action if they hold. According to Rabani *et al.* (2021), the other two factors that might considerably increase the chance of suicide are the availability of key suicide procedures and the spread of the idea. The social contagion makes suicide more likely to occur, which promotes it as a valid treatment for mental problems. Protective factors and other components are also helpful in preventing suicide. According to Rabani *et al.* (2021), a few of these are religion, the need for children, and a sense of direction in life. In low-income countries, there are 0.1 psychiatrists available for every 100,000 people, according to statistics from the WHO's Global Health Observatory. In India, there are 0.75 psychiatrists for every lakh inhabitants. It is a well-known fact that most people find it difficult or uncomfortable to discuss their suicidal thoughts with a psychiatrist or other form of counsellor. Since stigma considerably impacts suicidality, clinical therapy for persons at risk of suicide becomes exceedingly challenging to administer on a large scale. According to Chatterjee *et al.* (2022), another issue with this mental disorder is the lack of diagnostic testing. At-risk individuals cannot be identified without carefully observing the behaviour of mental health examiners, who may be relatives, friends, siblings, or psychiatrists. Due to the sharp increase in suicides, researchers are looking at this mental health issue from a variety of perspectives. The traditional approach involves clinical patient engagement and the use of clinical methods. Utilizing machine learning to analyse the suicide messages posted on social media is another tactic. People are more likely to communicate their feelings on social media since it is anonymous, which has boosted involvement in human media studies for preventing suicide. There are presently more social networking sites available. The many social media activity are scheduled and location-specific, and this information may be used to ascertain the key specifics of the event.

2.3 Using Multi-Modal Feature-Based Techniques to Detect Suicide Ideation

Individuals have begun blogging, tweeting, and publishing in online forums about their suicidal thoughts more often since the invention of social media. Particularly teenagers use social media to announce their plans to kill themselves, seek advice on how to do it in online forums, and even participate in suicide pacts. According to Chatterjee *et al.* (2022), since online communication allows for anonymity, people are free to publicly express their worries and fears about the status of the world. These websites make it easier to spot suicidal tendencies and stop them before it happens. Twitter, which is becoming more and more popular, is one of the social media sites that are most frequently employed for sentiment analysis research. Men and women differ in young people's attitudes about suicide. The poll also shows which gender tries suicide the most often. Another study that analysed student data from around the globe found several traits influenced by the environment and connected to students' tendency for suicide. Worry, insecurity, and loneliness were highlighted by the authors as significant factors that can affect students' mental health. Additionally, they provided achievable coping methods to help prevent similar circumstances. In an article, the authors of Chatterjee *et al.* (2022) outline the many signs and symptoms of suicidal behaviour. The authors looked at the epidemic, suicide risk factors, and safety measures. Another study by Manisha *et al.* (2019) found that young people had particular health risks for this problem. The authors present information on suicide attempts and suicidal thoughts. Manisha *et al.* (2019)'s outstanding work on the topic focuses on using NLP to stop suicide ideation and attempts. In order to recognize suicide endeavours in a database of acute psychiatric cases, they evolved a rule-based system for classifying suicidal thinking and employed a hybrid machine-learning technique. Another research examines teenage suicide behaviour in mental health institutions using machine learning and "natural language processing (NLP)" on electronic health histories. A technique for evaluating the suicidality of "Spanish-speaking social media users" was developed by the authors of Manisha *et al.* (2019) by analysing behavioural, structural, and "multimodal data from several social platforms" and expanding algorithms to identify people with suicidal intent. Abdulsalam and Alhothali (2022) tested the hypothesis that the amount of anxiety in a tweet about suicide could be inferred only from the tweet's text utilizing both human assessors and a "machine learning classifier". By extracting six feature groups of characteristics thought up of depression-related character traits from clinical and digital social behaviours employing Twitter tweets, the writers of Abdulsalam and Alhothali, (2022) developed a multimodal anxiety vocabulary knowledge prototype that may be employed to acknowledge depressed Twitter users. Machine learning, which provides a device with the ability to learn on its own, act on data without previous training, and autonomously access data, is one of the numerous applications of artificial intelligence. According to Abdulsalam and Alhothali, (2022), machine learning applications in engineering and computer science have rapidly increased during the past several years. It has been used for a variety of tasks, including the development of pharmaceuticals, the detection of fraud, online search queries, and online recommender systems. The most efficient ML approach is classification, which includes dealing with the evaluation of new attributes.

3 Research Methodology

3.1 Data Collection and Description

The analysis procedure starts with data collecting since the machine learning algorithms are prone to train and test based on the collected dataset. One of the biggest challenges in the field of suicidal ideation detection is the lack of a public dataset. Due to societal stigma, traditionally, it has been difficult to obtain data that is connected to mental diseases or suicide ideation. But more and more individuals are using the Internet to express their annoyance, look for support, and talk about mental health problems. Since Twitter has been proven to be useful in evaluating mental illnesses like suicidal thoughts, the suicidal dataset is collected from the Kaggle open-source site and it is publicly accessible. The main goal of this project is to compile various tweeter posts that relate to suicide but do not convey suicidal intent in a straightforward manner. Following the ethical guidelines, the tweets are not labelled with any identifiable information. This protects the privacy and data protection of the individuals. In place of the personal information of the individuals, an ID is given for each tweet post.

The CSV file collected from the Kaggle has been imported to the Python environment using Pandas. The dataset includes only two data columns and the total number of rows in the dataset is 9118. The first column is the "Label" column and the second column is the "Tweet" post. Using the tweet post column, the search of the posted tweets has been recognized using tokenized phrases. First, a number of tweets on suicide are retrieved from the tokenized texts. Reading tweets exposed us to phrases suggesting suicide ideation, such as "wanna die," "slave," "meaning less life", "kill myself" and so forth. An effort is made to attempt to compile the terms typically used to denote suicidal thoughts or ideas. Numerous published papers and journals were collected and analyzed that are related to suicides. Going through the numerous paper and articles gave significant knowledge on suicidal expressions (Thangaraj, and Sivakami, 2018). Using the final keyword list generated after the tokenization of the tweet post, the collection of the tweets is performed manually. There is manual procedure is performed to collect and change the labels and remove the non-frequent texts from the list. The following are a few tokenized texts that are recognized and associated with suicide ideation as per the selected dataset: "wanna die," "slave," and "meaning less life", "kill myself" and so forth.

3.2 Data Pre-Processing

In order to increase the effectiveness of a classifier, raw data must be transformed into a more useful format. Since most tweets had high noise, the dataset was appropriately cleaned in this study before the job of detecting suicidal thoughts was carried out. The process of data pre-processing for textual analysis is shown in the below flow diagram

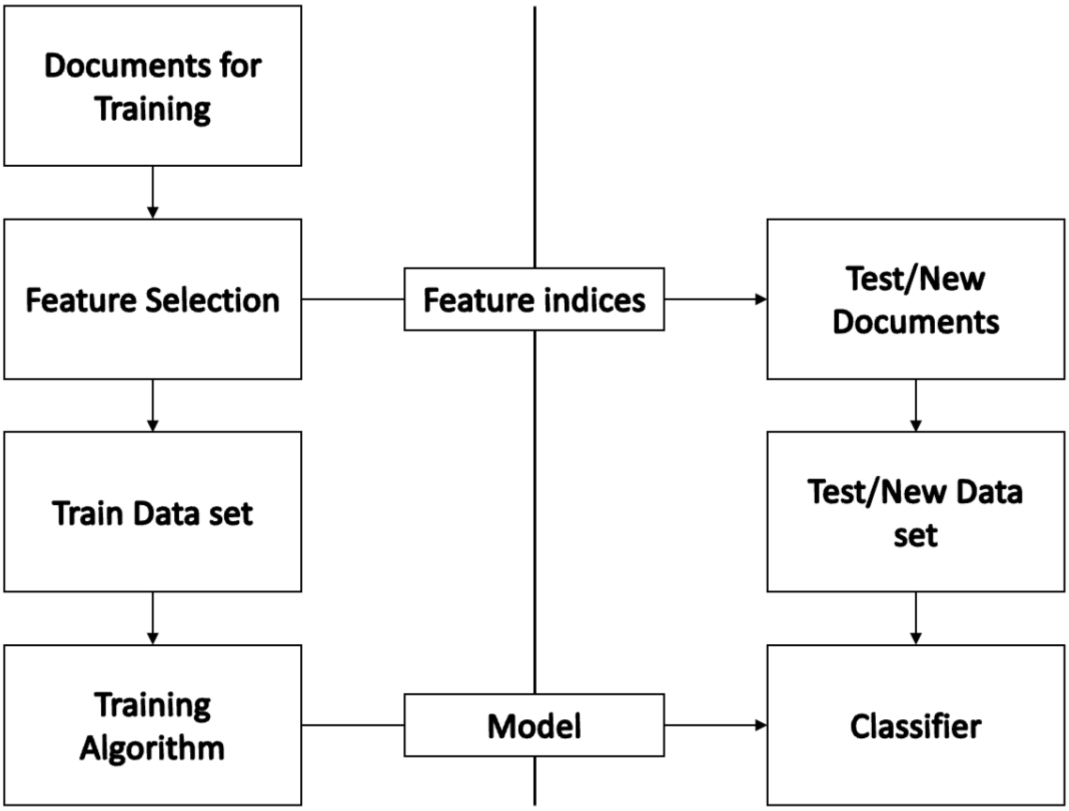

Figure 1: Flow Diagram of Textual Analysis

Word Transformation

Short conversational phrases and contractions make up the majority of the tweets in the chosen sample. To tokenize the text and replace it with its whole form in order to transform it into meaningful words, word segmentation has been used.

Eliminating Unnecessary Instances

Nonsensical characters are incomprehensible to ML models. They must be removed from tweets since their inclusion makes the text too loud. Using regular expressions, emojis, URLs, punctuation, whitespace, numbers, and user references were removed from the text.

Stemming and Lemmatization

A word-shortening technique called stemming aims to get to the term's fundamental meaning. Similar to stemming, lemmatization uses morphological and vocabulary analysis to return words to their dictionary forms. In order to execute stemming and lemmatization, the study makes use of NLTK's Porter Stemmer and Wordnet Lemmatizer, which increased the accuracy of text classification.

Stop-Words Elimination

A stop words list is a collection of inconsequential, often occurring words that have little to no grammatical significance for classifying text. In order to reduce the amount of unimportant information in our text and focus more on the important information, we have removed them using the NLTK's stop words corpus. We also removed less common terms from the tweets.

Feature Extraction

Feature extraction, which converts higher dimensional data into a collection of low dimensional feature sets, is one method used in ML for dimensionality reduction. Extraction of useful and important traits enhances ML model performance while lowering computational complexity (Onan, 2021). In order to convert text into a matrix (or vector) of possibilities, feature extraction has been carried out. Among all feature extraction techniques, word embedding and word count vectorizer are two of the most effective techniques widely used in supervised ML for text classification.

3.3 ML Model Training and Count Vectorization

It is necessary to transform source documents into a vector representation in order to classify text. By converting the source texts into vector representations with the same length as the tweets and an integer count of the number of times a word appeared in each tweet, CountVectorizer has been created to vectorize our tweets. Following count vectorization, a list of 36,121 unique words representing all of the tweets was created. These words were then loaded into the model vectorizer so that the machine-learning models could perform classification (Yadav, et al 2020). The original set of lexicons was divided into training and testing, which were 80% and 20%, respectively, to train the ML classification model. In this paper, various ML techniques for categorizing suicidal thoughts are described. The working of implemented classifier in this work has been discussed in the design specification section.

3.4 Evaluation of Classifiers

To assess the models, we chose to employ the common classification measures, including accuracy, precision, recall, f1 score, AUC, and confusion matrix. Such metrics are straightforward and may be generated using Equations demonstrated below for textual classification tasks:

$$Classification\ or\ Prediction\ Accuracy = \frac{TP + TN}{TP + FP + FN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$= \frac{2 * (Recall + Precision)}{Recall + Precision}$$

Where:

TP (True Positive): represents the number of correctly predicted positive outcomes,

FP (False positive): represents incorrectly predicted positive outcomes,

FN (False Negative): represents incorrectly predicted negative value,

TN (True Negative): represents correctly predicted negative outcomes.

4 Design Specification

4.1 Support Vector Classifier

A Support Vector Machine (SVM), an algorithm, determine the best way to distinguish between vectors that belong to a certain group (or category) and those that do not. It may be applied to any form of vector that encodes any kind of data. This means that in order to utilize the capabilities of SVM text classification, texts must be translated into vectors (Madireddy, 2018). A vector in SVM is a collection of grid points represented by a list of numbers. The classifier chooses where to create the best "line" (or the best hyperplane) while establishing the largest margin, to divide the space into two distinct subspaces, one for the vectors that belong to the specified category and one for the vectors that do not. Therefore, humans will be capable of applying the SVM approach to text classification issues and receive excellent results provided the model can discover vector representations which encode as much information from sample texts as feasible.

Some popular use of the SVM Classifier are listed below

- Predicting the class of an Email (Spam or Not Spam)
- Analysis of Review Dataset using the concept of Sentiment Analysis
- For various ML based Classification Tasks

Terminology in SVM

Hyper-Plane

A hyperplane in SVM is a decision boundary that separates the two classes. A data point may fall into a different class depending on which side of the hyperplane it is on. The size of the hyperplane depends on the number of input features in the dataset.

Support-Vectors

The hyper-location plane's orientation is influenced by the support vectors, or data points, that are closest to the hyper-plane. For the selected hyperplane, the margin, or the distance between the support vectors and the hyperplane, must be maximum. Even a minor disruption in the position of these support vectors can change the hyper-plane.

4.1.1 Working of SVM

The sample instances are transformed into a high-dimensional feature space and then the SV classifier is used to categorize those feature instances either using a linear or non-linear kernel function. The sample instances are transformed to allow for the hyperplane separator after a boundary between each class of the target variable is determined. As a consequence, by examining the characteristics of sample instances, it is feasible to determine the group to which the new features belong or will be assigned to the different class of response variables (Luo, 2021). In SVM a mathematical function is employed for classification that is referred as the SVM. Some popular kernels of the SVM classifier are:

- Polynomial Kernel
- Linear Classifier
- RBF (Radial Basis Function) Kernel
- Sigmoid Kernel

In case when the linearly separable data instances are available then the linear kernel is used. In other situations, non-linear kernel functions such as sigmoid or polynomial are used. To select the best model for each circumstance, you will need to experiment with them because they each use different settings and procedures.

4.2 Logistic Regression

A statistical technique called logistic regression employs a logistic function to forecast binary classes. Here, the dependent and independent variables in the logistic model have values of either 0 or 1, and the logistic function or logistic curve is a sigmoid curve. As a result, logistics regression is a popular and practical regression technique for forecasting binary classes, or two classes.

The link between one dependent variable and independent factors is estimated using the logistic model. It employs the logarithm of odds, which is a linear association of predictor attributes in the sample dataset, as the dependent variable (value denoted "1") (Padurariu, and Breaban, 2019). The logistic function is the function that computed the log-odds transformation into likelihood. Logit, referred to as the short form for the logistic unit, is used as its measuring unit.

Although it may be used for binary text classification, it is not considered as a classifier. It only predicts the likelihood of a result given an input. However, it may be implemented for creating a classifier, for instance by selecting a threshold value and categorizing sample data for binary classification with likelihood over the threshold as one class label and another class below the threshold value. The maximum likelihood estimation technique is used to do the estimation.

4.3 Stochastic Gradient Descent Classifier

A linear classifier (logistic regression and SVC) with the moniker Stochastic Gradient Descent (SGD) Classifier has been improved using the SGD. SGD is not a separate classifier as like the SVM and other supervised classifiers. For optimizing or minimizing the loss function of each SVM and logistics regression classifier, SGD is used. The accuracy of categorization or prediction is increased or algorithmic losses are minimized using optimization techniques.

A cost function is minimized through gradient descent. One of the most widely used algorithms for optimization and by far the most popular method for optimizing neural networks is gradient descent. However, SGD may also leverage these sorts of techniques, including linear support vector machines and logistic regression, to enhance the linear classifier.

There are three popular gradient descents:

Batch gradient descent: In order to discover the minimum inside its basin of attraction, batch gradient descent computes the gradient utilizing the whole dataset.

SGD: By utilizing a single sample, stochastic gradient descent (SGD) calculates the gradient.

Mini-batch Gradient Descent: The best of both sort of gradient descents are combined in this technique, which updates each mini-batch of n training samples.

5 Implementation

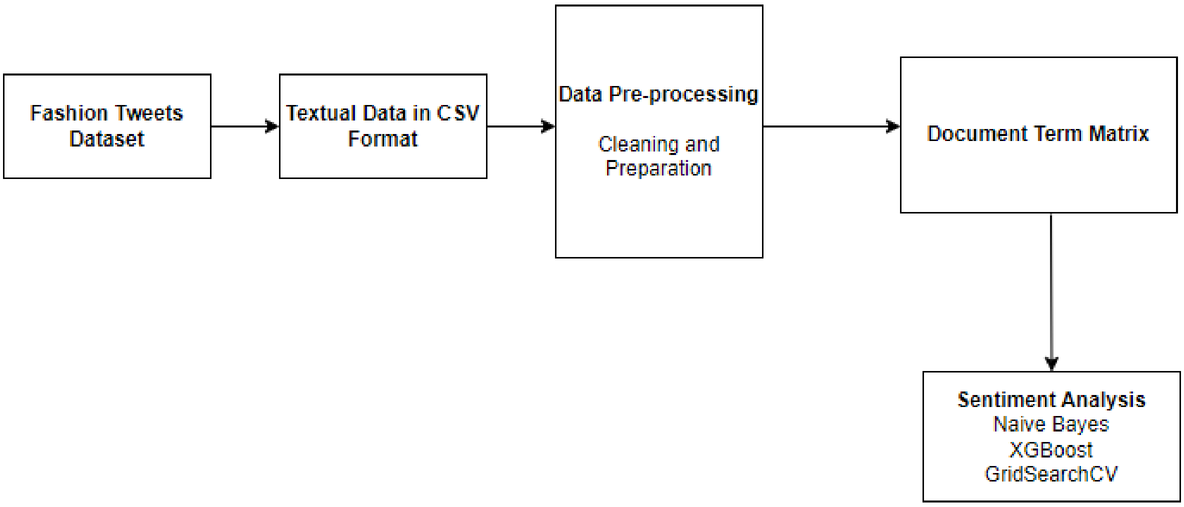

Figure 2: Flow Diagram of Implementation

The implementation of the proposed solution is done in the following manners

- Word Transformation
- Eliminating Un-necessary Instances
- Stemming and Lemmatization
- Stop-Words Elimination
- Feature Extraction

6 Evaluation

6.1 Data Analysis Results

To determine the frequency of suicidal thoughts and determine lexical differences, the entire pre-processed textual sample was examined. In the dataset of suicide tweets, all unigram frequencies have been calculated. Each sentiment type's word cloud (positive, negative, and neutral) is calculated and displayed below. In order to better understand the nature of each form of sentiment and how it relates to suicidal thoughts, the WordCloud visualization tool for Python additionally captures the world cloud for each type of sentiment.

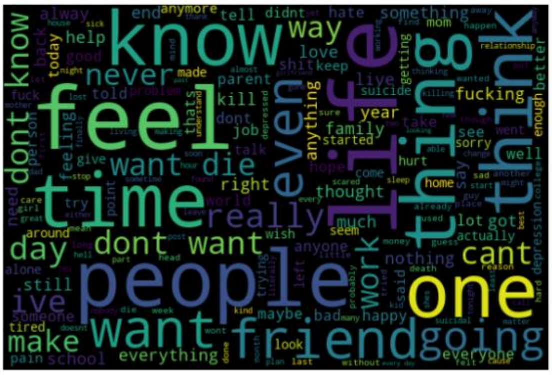

Figure 3:WordCloud of all Tokenized Tweets

The WordCloud of the suicidal class reveals that words like "feel" "end" "can't" "time" "life" and "never are often used in tweets with suicide intent. Then, the recognized words such that words like "death," "desire to die," and "kill" also signify the user's suicide intents.

The word-clouds are also captured for each sorts of sentiment and they are shown below.

Word-Cloud for Positive Sentiment

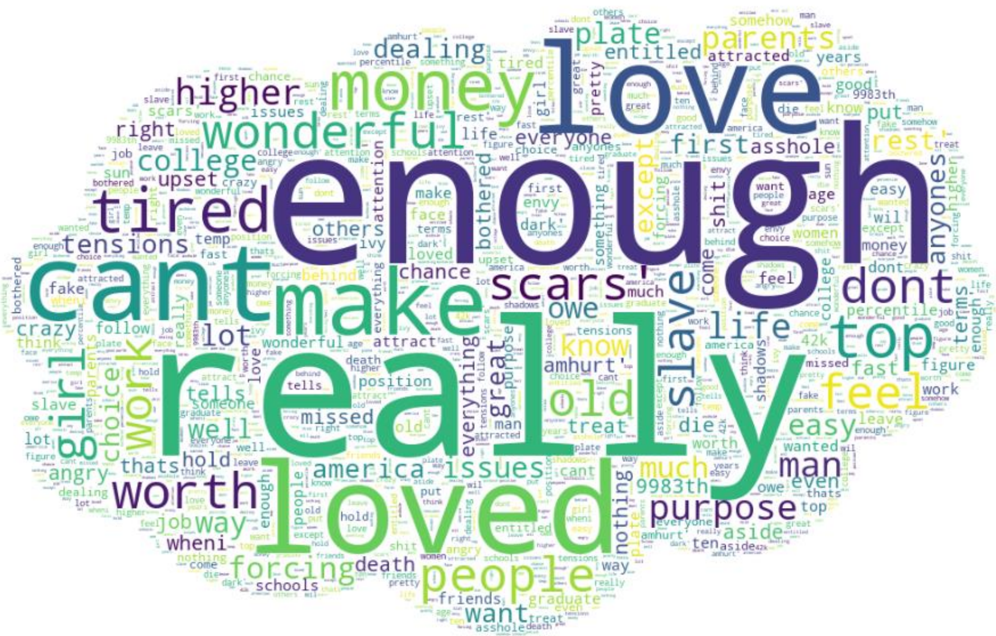

Figure 4: WordCloud for positive sentiment

The WordCloud of the positive sentiment class reveals that words like "enough" "really" "loved" "make" "money" and “wonderful” are often used in tweets with positive sentiment.

Word Cloud for Negative Sentiment

Figure 5: WordCloud for Negative sentiment

The WordCloud of the negative sentiment class reveals that words like "life" "end" "want" "horrible" "stop" and “meaning” are often used in tweets with positive sentiment.

Word-Cloud for Neutral Sentiment

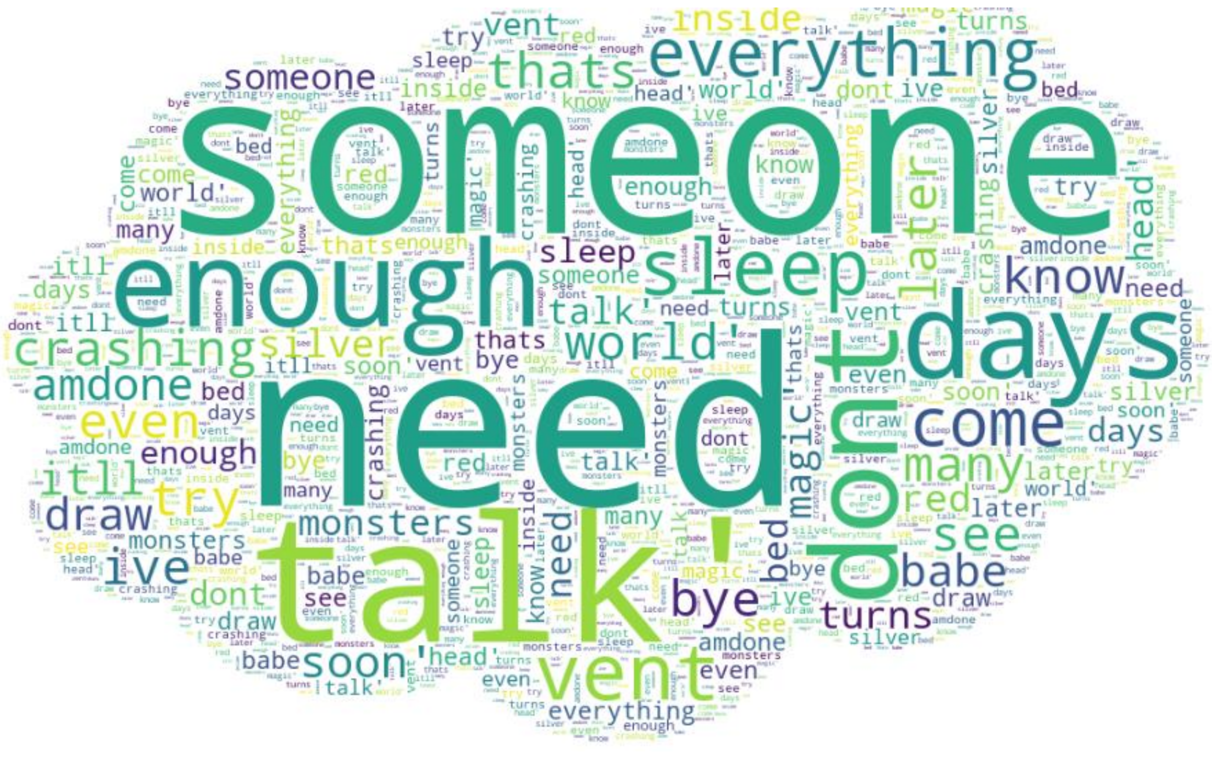

Figure 6: WordCloud for Neutral sentiment

The WordCloud of the neutral sentiment class reveals that words like "someone" "need" "talk" "enough" "dont" and “days” are often used in tweets with positive sentiment.

The bi-gram and tri-gram are generated for assessing the frequency of tokenized texts. The generated bi-gram is shown below.

```
[('dont know', 1165),
 ('feel like', 1155),
 ('dont want', 1146),
 ('want die', 1064),
 ('want kill', 223),
 ('dont think', 197),
 ('suicidal thoughts', 178),
 ('high school', 178),
 ('best friend', 167),
 ('dont feel', 161),
 ('feels like', 161),
 ('want live', 158),
 ('really want', 153),
 ('years old', 148),
 ('feel likei', 145),
 ('dont care', 143),
 ('hate want', 143),
 ('year old', 141),
 ('years ago', 139),
 ('want end', 138)]
```

Figure 7: BI-Gram

The bi-gram for the tokenized texts is shown in the above figures. The word "don't know" is often used in tweets. The suicidal thoughts texts such as "want kill" "want die" "want end" and "suicidal thoughts" are also repeated significant numbers of times.

```
[('dont want die', 236),
 ('hate want die', 131),
 ('want die feel', 130),
 ('feel like consuming', 123),
 ('like consuming entire', 123),
 ('die feel blood', 121),
 ('feel blood boiling', 121),
 ('blood boiling skin', 121),
 ('boiling skin burning', 121),
 ('skin burning eyes', 121),
 ('burning eyes watering', 121),
 ('eyes watering heart', 121),
 ('watering heart drops', 121),
 ('heart drops feel', 121),
 ('drops feel like', 121),
 ('consuming entire body', 121),
 ('entire body evil', 121),
 ('body evil cruel', 121),
 ('evil cruel hate', 120),
 ('dont want live', 86)]
```

Figure 8: Tri-Gram

The tri-gram for the tokenized texts is shown in the above figures. The word "don't want die" is often used in tweets. The suicidal thoughts texts such as "want die feel" "die feel blood" "don't want live" and "consuming entire body" are also repeated significant numbers of times that depict the suicidal intent.

6.2 ML Classification Model Performance Analysis

The experimental approach for identifying suicidal thoughts using ML models has been constructed and assessed after the n-grams frequency analysis. The techniques used in this paper include feature extraction using CountVectorizer on ML models like SVC, Logistics Regression, and SGD. Evaluation matrices have been used to compare the effectiveness of each ML classification model after training each type of ML model. The pictures shown below and the table represents the performance metrics of each classifier.

	precision	recall	f1-score	support
0	0.91	0.96	0.93	1060
1	0.94	0.87	0.90	764
accuracy			0.92	1824
macro avg	0.92	0.91	0.92	1824
weighted avg	0.92	0.92	0.92	1824

	precision	recall	f1-score	support
0	0.89	0.95	0.92	1060
1	0.93	0.84	0.88	764
accuracy			0.91	1824
macro avg	0.91	0.90	0.90	1824
weighted avg	0.91	0.91	0.91	1824

	precision	recall	f1-score	support
0	0.89	0.96	0.93	1060
1	0.94	0.84	0.89	764
accuracy			0.91	1824
macro avg	0.92	0.90	0.91	1824
weighted avg	0.91	0.91	0.91	1824

Figure 9: Evaluation Metrics of SVM, Logistic Regression and SGD (Top to bottom)

Evaluation Metrics	SVM	Logistic Regression	SGD
Accuracy	0.92	0.91	0.91
Precision	0.94	0.91	0.91
Recall	0.87	0.91	0.91
F_1 Score	0.92	0.91	0.91

Although accuracy has a wide range of applications, it is not always the optimal performance measure to utilize, especially when the target variable classes in the dataset are unbalanced. As a result, other performance metrics are used, such as the F1 score, which determines an algorithm's effectiveness by taking accuracy and recall into consideration. The SVM classifier outperforms other conventional ML methods in terms of model performance, with an accuracy score of 92% and a 0.92 F1 score. The performance of the SGD, and LR classifiers was somewhat worse than that of the SV classifier, which achieved an accuracy score of between 91% and 91% and an F1 score of 0.91.

7 Discussion

It is well acknowledged that a text classifier has to have a growing quantity of contextual information in order to become better at learning. The LSTM processing chain may process inputs in both forward and backward time sequences thanks to the BiLSTM processing chain, which duplicates the LSTM processing chain. By including a second hidden layer, BiLSTM improves upon the unidirectional LSTM by enabling hidden-to-hidden linkages to propagate in the opposing temporal sequence. As a result, the model may make use of data from both the present and the future. For sentiment classification issues, it is desirable for a model to be aware of both the past and the future contexts. The method enables BiLSTM to take the future context into account (Mohammed, and Kora, 2022). Additionally, without keeping duplicate context information, its layer learns bidirectional long-term dependence between time steps in time series or sequence data. When we want the network to learn from the full-time series at each sampling interval while simultaneously providing access to contextual data, these connections are essential. As a result, it proved to be a great performance for our study. However, the BiLSTM model has the drawback of needing more training data and time than the other classifiers.

It is important to note that the majority of research only offers out-of-focus details on pre-processing techniques, which are crucial to text categorization. Our excellent accuracy score was largely due to the efficient pre-processing of the tweets utilizing a variety of NLP approaches. In order to reduce overfitting, the ML and DL models were both trained with the proper parameters. Due to the issue of class imbalance, models frequently appear to perform badly. Since there was no class imbalance in this study's usage of two evenly split classes, all of the evaluation matrices performed equally well. Although our experimental findings show that assessed matrices perform rather well, cross-validation on the DL classifiers was not carried out since it would have required a very time- and resource-intensive setup. Additionally, we used a single dataset of 9118 tweets for the experiment. The maximum character count for a Tweet is 298; this is sometimes insufficient to convey a person's suicidal intentions. A considerable difference between classifiers has been found when a big volume Twitter dataset has been utilized instead of the short tweets dataset.

8 Conclusion and Future Work

An essential and successful strategy for preventing suicide is the early identification of suicidal thoughts. Computer scientists have employed pattern design ML and DL-based representation learning, whereas psychologists have used statistical analysis for the majority of their work in this area. Medical professionals will be able to find and save a lot of lives if they can recognize early suicidal thoughts on micro-blogging websites like Twitter. The DL and ML techniques may provide fresh chances for enhancing early suicide prevention and suicidal thought detection. The WordCloud of the suicidal class reveals that words like "feel" "end" "can’t" "time" "life" and "never are often used in tweets with suicide intent. Then, the recognized words such that words like "death," "desire to die," and "kill" also signify the user's suicide intents. The SVM classifier outperforms other conventional ML methods in terms of model performance, with an accuracy score of 92% and a 0.92 F1-score. The performance of the SGD, and LR classifiers was somewhat worse than that of the SV classifier, which achieved an accuracy score of between 91% and 91% and an F1 score of 0.91. Only machine learning models have been implemented in this work for creating the classification model. In future, the ML, as well as DL models, will be implemented to achieve a model with the highest classification accuracy.

Acknowledgement

I would like to acknowledge and give my warmest thanks to my dear supervisor Dr Christian Horn, who made this work possible. I carried out his guidance, advice, and brilliant comments throughout all the stages of writing the paper. I also want to convey my deep gratitude for being so kind, and patient, and being available all the time when reached out until the final submission date. I would also like to extend my thanks to my fellow project group colleagues for being supportive and making this possible.

References

Abdulsalam, A. and Alhothali, A., 2022. Suicidal Ideation Detection on Social Media: A Review of Machine Learning Methods. *arXiv preprint arXiv:2201.10515*. https://arxiv.org/abs/2201.10515

Aldhyani, T.H., Alsubari, S.N., Alshebami, A.S., Alkahtani, H. and Ahmed, Z.A., 2022. Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. *International journal of environmental research and public health*, *19*(19), p.12635. https://www.mdpi.com/1660-4601/19/19/12635

Chatterjee, M., Kumar, P., Samanta, P. and Sarkar, D., 2022. Suicide ideation detection from online social media: A multi-modal feature based technique. *International Journal of Information Management Data Insights*, *2*(2), p.100103. https://www.sciencedirect.com/science/article/pii/S2667096822000465

Faryal, M., Iqbal, M. and Tahreem, H., 2021. Mental health diseases analysis on Twitter using machine learning. *iKSP Journal of Computer Science and Engineering*, *1*(2), pp.16-25. https://iksp.org/journals/index.php/ijcse/article/download/136/168

Haque, R., Islam, N., Islam, M. and Ahsan, M.M., 2022. A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning. *Technologies*, *10*(3), p.57. https://www.mdpi.com/2227-7080/10/3/57

Luo, X., 2021. Efficient english text classification using selected machine learning techniques. Alexandria Engineering Journal, 60(3), pp.3401-3409. https://www.sciencedirect.com/science/article/pii/S1110016821000806

Madireddy, V.R., 2018. Content Based Image Classification Using Support Vector Machine Algorithm. International Journal of Innovative Research in Computer and Communication Engineering, ISSN (Online), pp.2320-9801. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4289407

Manisha, M., Kodali, A. and Srilakshmi, V., 2019. Machine classification for suicide ideation detection on Twitter. *International Journal of Innovative Technology and Exploring Engineering*, *8*(12), pp.4154-4160. http://www.cse.griet.ac.in/pdfs/journals19-20/55.pdf

Mohammed, A. and Kora, R., 2022. An effective ensemble deep learning framework for text classification. Journal of King Saud University-Computer and Information Sciences, 34(10), pp.8825-8837. https://www.sciencedirect.com/science/article/pii/S1319157821003013

Onan, A., 2021. Ensemble of classifiers and term weighting schemes for sentiment analysis in Turkish. Scientific Research Communications, 1(1). https://scientificrc.org/index.php/src/article/view/30/6

Padurariu, C. and Breaban, M.E., 2019. Dealing with data imbalance in text classification. Procedia Computer Science, 159, pp.736-745. https://www.sciencedirect.com/science/article/pii/S1877050919314152

Rabani, S.T., Khan, Q.R. and Khanday, A.M.U.D., 2021. Quantifying Suicidal Ideation on Social Media using Machine Learning: A Critical Review. *Iraqi Journal of Science*, pp.4092-4100. https://ijs.uobaghdad.edu.iq/index.php/eijs/article/view/2877

Thangaraj, M. and Sivakami, M., 2018. Text classification techniques: A literature review. Interdisciplinary Journal of Information, Knowledge, and Management, 13, p.117. https://www.proquest.com/openview/b207fca7feb7fe6a69dca55c3e84e525/1?pq-origsite=gscholar&cbl=5314838

Yadav, B.P., Ghate, S., Harshavardhan, A., Jhansi, G., Kumar, K.S. and Sudarshan, E., 2020, December. Text categorization Performance examination Using Machine Learning Algorithms. In IOP Conference Series: Materials Science and Engineering (Vol. 981, No. 2, p. 022044). IOP Publishing. https://iopscience.iop.org/article/10.1088/1757-899X/981/2/022044/meta