DOI: http://dx.doi.org/10.21123/bsj.2020.17.4.1328

Detection of Suicidal Ideation on Twitter using Machine Learning & Ensemble Approaches

Syed Tanzeel Rabani * *Qamar Rayees Khan* *Akib Mohi Ud Din Khanday*

Department of Computer Sciences, Baba Ghulam Shah Badshah University -185234, Rajouri, J&K, India.
*Corresponding author: syedtanzeel@bgsbu.ac.in , [2] rayees.dcs@gmail.com, [3]akibkhanday@gmail.com
*ORCID ID: https://orcid.org/0000-0001-7596-5964 ,https://orcid.org/0000-0002-7628-173X, https://orcid.org/0000-0001-6804-4905

Received 11/5/2020, Accepted 2/9/2020, Published 1/12/2020

Abstract:

Suicidal ideation is one of the most severe mental health issues faced by people all over the world. There are various risk factors involved that can lead to suicide. The most common & critical risk factors among them are depression, anxiety, social isolation and hopelessness. Early detection of these risk factors can help in preventing or reducing the number of suicides. Online social networking platforms like Twitter, Reddit and Facebook are becoming a new way for the people to express themselves freely without worrying about social stigma. This paper presents a methodology and experimentation using social media as a tool to analyse the suicidal ideation in a better way, thus helping in preventing the chances of being the victim of this unfortunate mental disorder. The data is collected from Twitter, one of the popular Social Networking Sites (SNS). The Tweets are then pre-processed and annotated manually. Finally, various machine learning and ensemble methods are used to automatically distinguish Suicidal and Non-Suicidal tweets. This experimental study will help the researchers to know and understand how SNS are used by the people to express their distress related feelings and emotions. The study further confirmed that it is possible to analyse and differentiate these tweets using human coding and then replicate the accuracy by machine classification. However, the power of prediction for detecting genuine suicidality is not confirmed yet, and this study does not directly communicate and intervene the people having suicidal behaviour.

Key words: Ensemble Learning, Machine learning, Suicidal Ideation, Text classification, Twitter, Weka.

Introduction:

Suicide is one of the significant public health concerns consuming a lot of lives. According to the statistic of the World Health Organisation (WHO) (1), around one million people die due to suicide each year, and on average, suicide occurs every 40 seconds. Among the total suicide-related deaths, 135000 deaths occurred in India alone (2). WHO further mentioned that suicide is the primary cause of death among teenagers and the sixth leading cause among adults. Furthermore, there are 20 times more suicidal attempts disrupting the families emotionally and economically. American Foundation for Suicide Prevention (AFSP) has identified various risk factors associated with suicide. The factors include personal issues like hopelessness, substance abuse, anxiety, schizophrenia; social factors like isolation from society, loss of loved ones, unemployment, bullying or abuse, or factors related to negative events in life like illness, emotional disorders, and history of previous suicide attempts. It is a common saying that suicide is a permanent solution for dealing with temporary problems. Despite the growing numbers of suicidal cases, it can be prevented to some extent by understanding the risk factors related to suicidal behaviour in the early stages of the suicidal process. The process of suicide starts with suicidal thoughts or ideation. It then matures to suicidal attempt and finally to the completed suicide. Prevention can be done by reducing the risk factors or by reducing the obstacles to mental health resources. However, the social stigma related to this mental illness obstructs psychiatric health professionals to counsel and treat those people and thus emerge as one of the important aspects of this research.

With the widespread upsurge of "social web 2.0" and internet technology, there is a growing inclination of people to form online communities and interact with each other. Research has revealed that people feel comfortable to talk about suicidal ideation online rather than to talk about it in face to face settings (3)—the reason being a sense of self-control and the anonymous feature of social media. Various cases exist, where suicide victims also revealed their final feelings before their death on social media (4–6).

Due to the non-availability of laboratory tests to predict the suicidal ideation, It is believed that social media could offer an opportunity to analyse the behaviour of people by identifying the risk factors and warning signs well before, to prevent the deaths. The major causes of suicidal ideation are depression, anxiety, hopelessness, stress that slowly progresses to suicidal (7–9). If these risk factors are adequately analysed on social media, it can help in preventing to some extent the potential suicidal victims.

In today's age, Twitter is regarded as an emerging platform for social science research. People's attitudes, beliefs and activities are a searchable archive. Twitter is one of the most popular social media sites that help users to share their thoughts and feelings in a real-time. The Tweets has a maximum limit of 280 characters. Twitter is functioning in all countries except China, North Korea and Iran and requires no criteria for age. As people discuss their feelings openly on social media without worrying about social stigma, it can be assumed to represent people's genuine feelings. A large number of twitter accounts are public, making it possible for anyone to view each other's content. It is estimated that 23 % of all the adults who are online use this site for social networking, and as many as 500 million tweets are tweeted every day (10). After recognizing that people do post about suicidality, Twitter created a feature to report those posts and notify the user about the crisis situation. This notifying content relies only on the decision and interpretation of networked users, many of whom might not be able to understand the genuine risk (11).

It is being observed that persons suffering from suicidal ideation use different features in their language to portray about their distress feelings/sufferings which if analysed carefully and in advance can be used to save precious lives (12). This study aimed to find the feasibility in detecting and differentiating the suicidal tweets from non-suicidal tweets by analysing the data from social media platforms. Human coders were instructed to categorize the tweets into two classes. After that, machine learning and ensemble methods were applied to automatically classify the tweets to replicate the accuracy of human annotation. The models were examined using precision and recall metrics.

This research paper makes novel contributions to the already existing literature in the following ways.

- Our work extracts a number of relevant features to differentiate between suicidal and non-suicidal tweets by the help of a novel feature engineering technique.
- This paper builds a novel dataset by extracting the tweets related to suicide and non-suicide from Twitter. The dataset is annotated and thereafter used to train our model for binary differentiation (suicidal and non-suicidal) of the tweets.
- This work reveals the hidden knowledge from the perspective of data mining.
- Benchmarking is done on various machine learning and ensemble methods to distinguish their effectiveness.

The paper below is organized under various sections starting from the literature survey from the already existing literature related to the proposed work followed by the proposed methodology using data extraction, pre-processing & feature extraction and classification algorithms. In corresponding sections, the findings, experimental results and future scope of the study will be elaborated.

Related Work:

Text classification task has been studied enough on the shorter text like tweets (13,14). However, the classification of suicide-related text from non-suicidal text using various machine learning techniques is still in its infancy. The research related to the classification and differentiation of suicidal users from non-suicidal users typically revolved around suicide notes. So this problem has been studied mainly on the parameters of psychology and psychiatry (15). Other methods to address this problem are using questionnaires to assess the risk of potentially suicidal individuals through clinical methods (16). As the data scarcity is a significant issue involving this research, thus drawing the attention of researchers towards machine learning techniques to understand the language of suicidal individuals from the online user-generated content (9).

The motive of text-classification is to observe whether suicidal individuals can be distinguished based upon the posts they share. Such methods include the filtering of keywords and phrases related to suicide (12,17,18).

The problem has also been analysed through NLP and machine learning to some extent. Researchers investigated the correlation between suicidality and social media (8,12). Jashinsky et.al. (12) detected suicide-related tweets using a keyword-based approach. Some filtering terms like accidentally, shaving, cutting myself were used to remove the tweets with no risk of suicide. The geo-located tweets were collected and matched with arbitrary users in the United States. A strong correlation was found between the suicidal tweets and actual state suicide rates. The study raised concern for more research on social media to help vulnerable individuals.

Bart Desmet et.al.(19) investigated a method to automatically classify suicide-related Dutch forum posts. The researchers used Naïve Bayes (NB) and Support Vector Machine (SVM) for classification. The model was optimized through various Genetic algorithms which attained the F score of 92.69%.

Bridianne O'Dea et al. (20) used suicide-related keywords to collect the tweets with the help of Twitter API. Dataset generated from twitter was manually annotated into three levels of concern. The features supplied to the machine learning algorithms were token unigrams and bag-of-words. The study was first of its kind that could replicate human accuracy with the help of machine learning techniques.

De Choudhury et al. (8) explored social media to investigate Major Depressive Disorder (MDD) among twitter users. They applied crowdsourcing techniques to build a database of Twitter users suffering from the MDD using the Centre for Epidemiologic Studies Depression Scale (CES-D). The users were classified into groups based on scoring high and low for depression using CES-D. The study found that the science of depressive behaviour also replicates on social media. The users having a high-end score for depression were seen posting content at night hours. They also tend to use first-person pronouns, and network interaction is found to be very less, resulting in social isolation.

The recent study by De Choudury et.al. (21) investigated the individuals suffering from suicidal ideation on social media. The linguistic features were found to indicate the shift from stress to suicidal ideation. Various features characterizing this shift include hopelessness, anxiety and isolation.

Coppersmith et.al. (22) explored and examined the tweets posted by Twitter users prior to their suicidal attempt. Their language was analysed to investigate the pattern. It was found that there was an increase in the percentage of sadness related tweets before some weeks of the suicide attempt. It was followed by an increase in anger-related tweets after the attempt. The researchers interested an in-depth investigation of the use of SNS for suicide prevention can refer to the systematic review by (23).

After studying the literature, it was found that dataset used in the research was very low as data is not freely available due to its ethical considerations and the studies which were already done in this area of research have not achieved much accuracy and recall. In this research paper, a classifier is built that automatically distinguishes between suicidal and non-suicidal tweets using the machine and ensemble approaches with good accuracy and recall value.

Proposed Methodology:

An elaborated analytical and experimental methodology is used based on the extensive literature survey and the corresponding discussion with the mental health experts, psychologists on the problem. Based on the various factors that influence the performance of the model, real-time data was collected from twitter, and the data is submitted to the model, and the final results are generated. The overall methodology comprises of five main steps: (i) Corpus collection (ii) Data annotation (iii) Pre-processing and feature extraction (iv) Machine classification (v) Evaluation and validation.

The complete picture of the proposed methodology that is used in this work is depicted below in Figure 1.

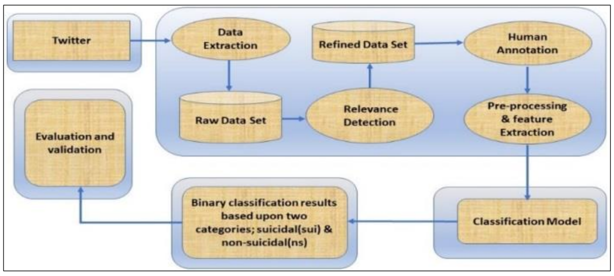
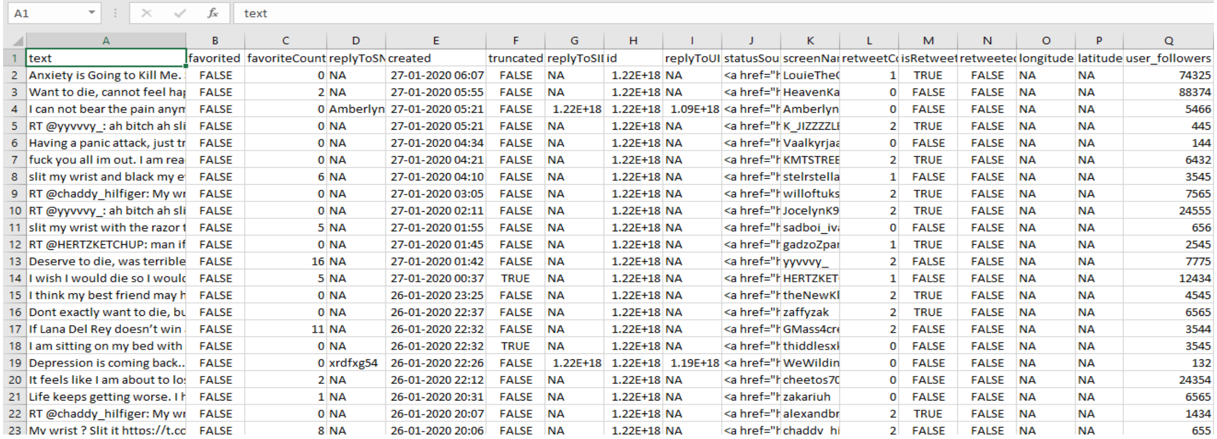

Twitter
Data Extraction
Refined Data Set
Human Annotation
Raw Data Set
Relevance Detection
Pre-processing & feature Extraction
Evaluation and validation
Binary classification results based upon two categories; suicidal(sui) & non-suicidal(ns)
Classification Model

Figure 1. Methodology to identify the suicidal tweets on Twitter

Corpus Collection:

To collect the relevant data, different keywords and phrases were collected that were used in previous papers and also from various sites and forums (20,24). Twitter Application Programming Interface (API) was used to access public data and extract the tweets. The extracted dataset consists of various fields implying text, favorited, favourite count, created, id, replyToUID, status source, screen name, retweet count, is retweet, retweeted, longitude, latitude, user followers etc. A sample tweet set is shown in Fig. 2 below. The total of 18756 tweets were extracted. Out of 18756 tweets, a total of 4266 tweets were selected for training and testing the machine learning model based on the relevancy of tweets.



A1 | text

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	text	favorited	favoriteCount	replyToSN	created	truncated	replyToSII	id	replyToUI	statusSou	screenNa	retweetC	isRetweet	retweete	longitude	latitude	user_followers
2	Anxiety is Going to Kill Me.	FALSE	0	NA	27-01-2020 06:07	FALSE	NA	1.22E+18	NA	<a href="h	LouieThe	1	TRUE	FALSE	NA	NA	74325
3	Want to die, cannot feel ha	FALSE	2	NA	27-01-2020 05:55	FALSE	NA	1.22E+18	NA	<a href="h	HeavenKa	0	FALSE	FALSE	NA	NA	88374
4	I can not bear the pain anym	FALSE	0	Amberlyn	27-01-2020 05:21	FALSE	1.22E+18	1.22E+18	1.09E+18	<a href="h	Amberlyn	0	FALSE	FALSE	NA	NA	5466
5	RT @yyvvvy_: ah bitch ah sli	FALSE	0	NA	27-01-2020 05:21	FALSE	NA	1.22E+18	NA	<a href="h	K_JIZZZZLI	2	TRUE	FALSE	NA	NA	445
6	Having a panic attack, just tr	FALSE	0	NA	27-01-2020 04:34	FALSE	NA	1.22E+18	NA	<a href="h	Vaalkyrjaa	0	FALSE	FALSE	NA	NA	144
7	fuck you all im out. I am rea	FALSE	0	NA	27-01-2020 04:21	FALSE	NA	1.22E+18	NA	<a href="h	KMTSTREE	2	TRUE	FALSE	NA	NA	6432
8	slit my wrist and black my e	FALSE	6	NA	27-01-2020 04:10	FALSE	NA	1.22E+18	NA	<a href="h	stelrstella	1	FALSE	FALSE	NA	NA	3545
9	RT @chaddy_hilfiger: My wr	FALSE	0	NA	27-01-2020 03:05	FALSE	NA	1.22E+18	NA	<a href="h	willoftuks	2	TRUE	FALSE	NA	NA	7565
10	RT @yyvvvy_: ah bitch ah sli	FALSE	0	NA	27-01-2020 02:11	FALSE	NA	1.22E+18	NA	<a href="h	JocelynK9	2	TRUE	FALSE	NA	NA	24555
11	slit my wrist with the razor t	FALSE	5	NA	27-01-2020 01:55	FALSE	NA	1.22E+18	NA	<a href="h	sadboi_iv	0	FALSE	FALSE	NA	NA	656
12	RT @HERTZKETCHUP: man if	FALSE	0	NA	27-01-2020 01:45	FALSE	NA	1.22E+18	NA	<a href="h	gadzoZpar	1	TRUE	FALSE	NA	NA	2545
13	Deserve to die, was terrible	FALSE	16	NA	27-01-2020 01:42	FALSE	NA	1.22E+18	NA	<a href="h	yyvvvy_	2	FALSE	FALSE	NA	NA	7775
14	I wish I would die so I woulc	FALSE	5	NA	27-01-2020 00:37	TRUE	NA	1.22E+18	NA	<a href="h	HERTZKET	1	FALSE	FALSE	NA	NA	12434
15	I think my best friend may h	FALSE	0	NA	26-01-2020 23:25	FALSE	NA	1.22E+18	NA	<a href="h	theNewKl	2	TRUE	FALSE	NA	NA	4545
16	Dont exactly want to die, bu	FALSE	0	NA	26-01-2020 22:37	FALSE	NA	1.22E+18	NA	<a href="h	zaffyzak	2	TRUE	FALSE	NA	NA	6565
17	If Lana Del Rey doesn't win	FALSE	11	NA	26-01-2020 22:32	FALSE	NA	1.22E+18	NA	<a href="h	GMass4cre	2	FALSE	FALSE	NA	NA	3544
18	I am sitting on my bed with	FALSE	0	NA	26-01-2020 22:32	TRUE	NA	1.22E+18	NA	<a href="h	thiddlesxl	0	FALSE	FALSE	NA	NA	3545
19	Depression is coming back..	FALSE	0	xrdfxg54	26-01-2020 22:26	FALSE	1.22E+18	1.22E+18	1.19E+18	<a href="h	WeWildin	0	FALSE	FALSE	NA	NA	132
20	It feels like I am about to los	FALSE	2	NA	26-01-2020 22:12	FALSE	NA	1.22E+18	NA	<a href="h	cheetos70	0	FALSE	FALSE	NA	NA	24354
21	Life keeps getting worse. I h	FALSE	1	NA	26-01-2020 20:31	FALSE	NA	1.22E+18	NA	<a href="h	zakariuh	0	FALSE	FALSE	NA	NA	6565
22	RT @chaddy_hilfiger: My wr	FALSE	0	NA	26-01-2020 20:07	FALSE	NA	1.22E+18	NA	<a href="h	alexandbr	2	TRUE	FALSE	NA	NA	1434
23	My wrist ? Slit it https://t.cc	FALSE	8	NA	26-01-2020 20:06	FALSE	NA	1.22E+18	NA	<a href="h	chaddy_hi	2	FALSE	FALSE	NA	NA	655

Figure 2. Twitter data extracted through API

Exploration & Knowledge Discovery of Dataset:

For understanding the suicidal content at the first level, the frequently used word by suicidal individuals length of the tweets and their distribution are understood through Figure 3-5.

Word Cloud.

Word cloud is the visual representation of the data. It gives a visual understanding of most of the frequent words used in the dataset. As depicted in Fig. 3, the words like “life”, “suicide”, “die”, “kill”, “depression” are used frequently by the posts containing suicidal ideation. The words like “feel”, “want”, “think” also exist in these posts indicating the intension of the suicidal users. For example, some users having suicidal ideation write like “I want to end my life, I have no one here.”

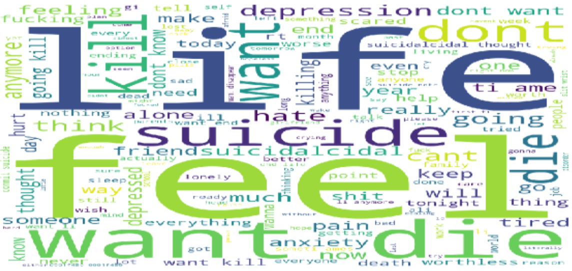


Figure 3. Word cloud of the suicidal tweets

Tweet Length and Distribution

The tweet-length describes the length of the tweet in characters used by the user. Fig. 4 provides a visual understanding of the no. of suicidal and non-suicidal tweets and their length. Fig. 5 below provides the tweet length distribution of the dataset used in this research. The tweet-length as analysed in Fig. 4 depicts that suicidal users want to share their intentions freely, so the length is larger than non-suicidal tweets,

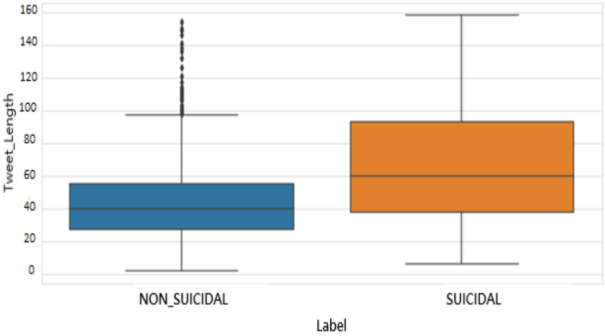


Figure 4. Visual representation of the length of used in our research

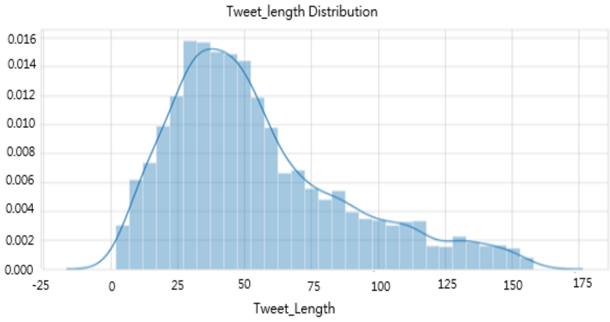
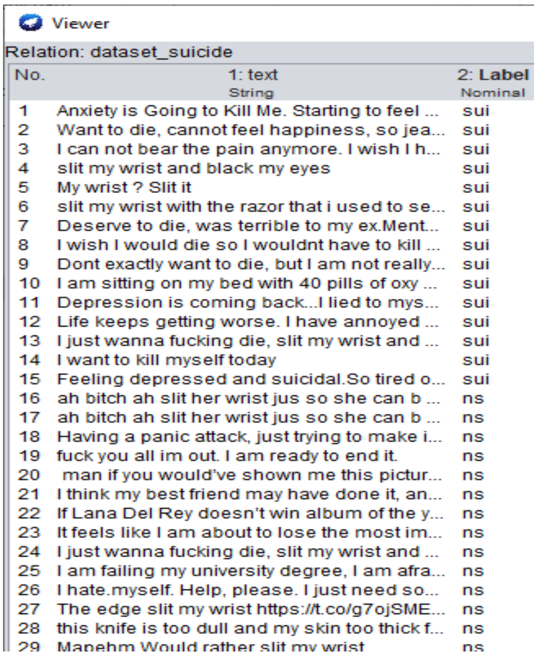


Figure 5. Tweet Length distribution of the dataset suicidal and non-suicidal tweets

Human Annotation:

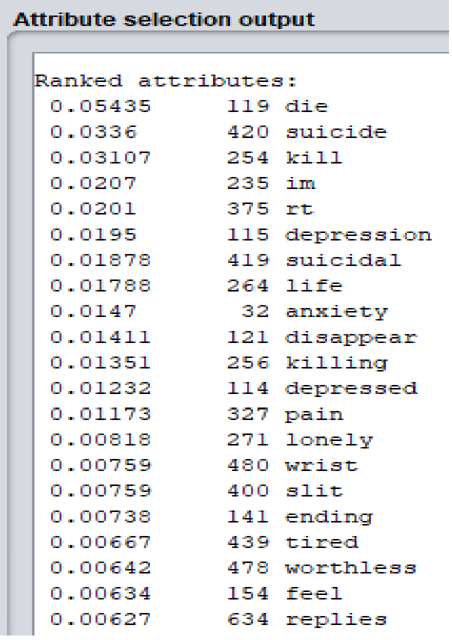
Human Annotation is an important step for supervised machine learning model. As this problem is related to mental health, the mental health practitioner and a psychologist is consulted for their valuable inputs and expert suggestions in labelling the tweets. Thus, the coding team consists of one mental health practitioner, a psychologist and two computer professionals. Mental health practitioner & Psychologist helped us in understanding the language of suicidal ideation by his expertise in dealing with the said patients. The coders were asked to categorize the tweets into two classes as suicidal(sui) & non suicidal(ns) by judging the context of tweets as well. The coders conceptualized thus the annotation scheme. The suicidal text contains those tweets that discuss about killing oneself, having anxiety, hopelessness, depression or any other tweets related to suicidal ideation. The non-suicidal class contains those tweets that discuss suicide in a formal way, refer to the second person's suicide, flippant references or other text that is not relevant to the suicide. Some manually annotated tweets are shown below in. Fig. 6.



Relation: dataset_suicide

No.	1: text String	2: Label Nominal
1	Anxiety is Going to Kill Me. Starting to feel ...	sui
2	Want to die, cannot feel happiness, so jea...	sui
3	I can not bear the pain anymore. I wish I h...	sui
4	slit my wrist and black my eyes	sui
5	My wrist ? Slit it	sui
6	slit my wrist with the razor that i used to se...	sui
7	Deserve to die, was terrible to my ex.Ment...	sui
8	I wish I would die so I wouldnt have to kill ...	sui
9	Dont exactly want to die, but I am not really...	sui
10	I am sitting on my bed with 40 pills of oxy ...	sui
11	Depression is coming back...I lied to mys...	sui
12	Life keeps getting worse. I have annoyed ...	sui
13	I just wanna fucking die, slit my wrist and ...	sui
14	I want to kill myself today	sui
15	Feeling depressed and suicidal.So tired o...	sui
16	ah bitch ah slit her wrist jus so she can b ...	ns
17	ah bitch ah slit her wrist jus so she can b ...	ns
18	Having a panic attack, just trying to make i...	ns
19	fuck you all im out. I am ready to end it.	ns
20	man if you would've shown me this pictur...	ns
21	I think my best friend may have done it, an...	ns
22	If Lana Del Rey doesn't win album of the y...	ns
23	It feels like I am about to lose the most im...	ns
24	I just wanna fucking die, slit my wrist and ...	ns
25	I am failing my university degree, I am afra...	ns
26	I hate.myself. Help, please. I just need so...	ns
27	The edge slit my wrist https://t.co/g7ojSME...	ns
28	this knife is too dull and my skin too thick f...	ns
29	Manehm Would rather slit my wrist	ns

Figure 6. Manual annotation of data as SUICIDAL



```
Ranked attributes:
 0.05435    119 die
 0.0336     420 suicide
 0.03107    254 kill
 0.0207     235 im
 0.0201     375 rt
 0.0195     115 depression
 0.01878    419 suicidal
 0.01788    264 life
 0.0147      32 anxiety
 0.01411    121 disappear
 0.01351    256 killing
 0.01232    114 depressed
 0.01173    327 pain
 0.00818    271 lonely
 0.00759    480 wrist
 0.00759    400 slit
 0.00738    141 ending
 0.00667    439 tired
 0.00642    478 worthless
 0.00634    154 feel
 0.00627    634 replies
```

Figure 7. Most Significant attributes(sui) & NON_SUICIDAL (ns)

Preprocessing and Feature Extraction:

The data collected from social media contains a lot of noise (25–28). The data needs to be filtered to remove the noise and prepare it for machine classification before the machine processes the data. The established methods (29,30) were used, and all those tweets were removed that had less than 75% of inter-annotator agreement. Thus, out of 4797, a total of 531 tweets were removed, leaving 4266 tweets for classification. The text is pre-processed by using various promising standard techniques of text mining (27,31). After removing the URLs, non-ASCII characters, the text was tokenized and stemmed using Tokenizer and Stemmer available in Weka Tool. Features were extracted using the Term Frequency Inverse Document Frequency (TFIDF) and BOW. In WEKA an unsupervised filter "StringToWordVector" is used as an implementation of TFIDF that converts our text corpus into string attributes. After applying the required filter on our data, a lot of attributes(features) are extracted. TFIDF removes the words having low importance in a corpus. For example, words like "a", "of", "the" occur frequently but have low semantic importance. The equation of TFIDF in context of our collected corpus would be

$$tfidf(t, tw, D) = tf(t, tw) \times idf(tw, D) \quad \ldots \quad (1)$$

$$idf(t, D) = \log \frac{|D|}{1|\{tw\epsilon D: t\epsilon tw\}|} \quad \ldots (2)$$

Where t is the word as a feature; tw denotes each tweet, and D denotes the Document space (set of all tweets). The tweets are also tokenised using WordTokenizer with the delimiters as .,;:'"()?!< and all the words for processing are converted into lower case tokens.

After applying TFIDF, the words which are least significant are removed, and the remaining are taking into account whose presence in the dataset seems to be critical. Most significant attributes are spotted using the information gain algorithm (InfoGainAttributeEval) with Ranker search method. The threshold for discarding the irrelevant features was set to 0.0025. Some of these significant features (attributes) are shown above in Fig. 7

Binary Classification Model:

The detection of suicidal ideation in tweets is viewed here as a supervised binary classification problem. After pre-processing of the tweets, only two columns, label and the title of tweets are left for training the machine learning model. The motivation to formulate the problem is provided by the researcher (32) as under.

Let the title be represented as t and label be represented as l.

Given a corpus $\{t_i, l_i\}_i^n$ consisting of a set of tweets $\{t_i\}_i^n$ and labels $\{l_i\}_i^n$, a machine learning model is trained so that the model learns suing the data and the labels provided. This type of model is called a supervised model. The model learns from the data provided with the help of supervisory function as follows

$$l_i = Fun(t_i) \quad \ldots \quad (3)$$

Where $l_i = 1$ which means that the expression t_i is the text of suicide (sui) and otherwise $l_i = 0$ means the text is non-suicidal text (ns). The objective of the classification model is to predict the class of the test data with minimum error. A loss function $Loss(l, fun(t))$ predicts the error where l is the accurate label and fun(t) is the predicted label.

WEKA is selected as a machine learning tool for binary classification of our data related to suicide. This tool has been developed by the University of Waikato in New Zealand. The extracted dataset from twitter consists of various fields many of which were dropped due to their irrelevancy in this research. Only text field was retained for analysis.

Various Supervised machine-learning approaches like Naïve Bayes, Decision trees, Multinomial Naïve Bayes, Logistic Regression and Support vector machine (SMO) were used to train the model. Besides those, various ensemble learning techniques like Bagging, Random Forest, AdaBoost, Voting and stacking were also applied

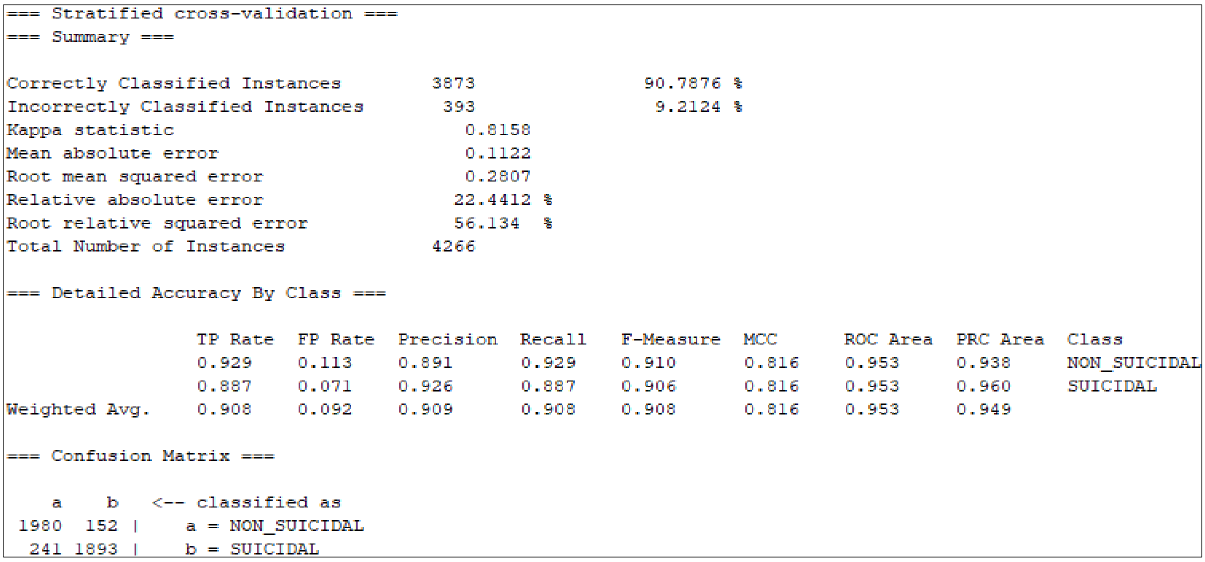
Results and Discussion:

Using Twitter API, a total number of 18756 tweets were extracted based upon the keywords and phrases relevant to the suicidal ideation. Out of the total extracted tweets, the irrelevant tweets and tweets having less than 75% of human agreement rate were discarded. Thus only 4266 tweets were annotated manually according to the proposed annotation scheme and supplied to the training and testing the machine learning model. Features in the dataset were extracted using TFIDF. A total number of 480 features is extracted. After that information gain algorithm (InfoAttributeEval) with ranker search method is applied for further discarding the less significant attributes. The threshold for discarding the less significant attributes was set to 0.0025. The most significant attributes (features) were supplied to the machine learning model, which was trained using 10-Cross validation technique. WEKA tool was used to implement various machine and ensemble learning algorithms. The five machine learning methods which were implemented in weka are Naïve Bayes (NB), Multinomial Naïve Bayes (MNB), Decision tree (REPTree and J48), Logistic Regression (LR) and Support Vector Machine (SMO). The Ensemble approaches implemented include Bagging (ZaroR, REpTree, SMO, LR, NB, MNB), Voting Ensemble (ZeroR, RepTree, SMO, LR, MNB), AdaBoost, Random Forest and Stacking (MNB, SMO, RepTree, ZeroR). Meta chlassifier used in stacking was SMO. Various algorithms produce different accuracy, precision and recall according to their working. Table 1. below compares the performance of various implemented machine and ensemble learning algorithms.

Table 1. Comparison between various ensemble and machine learning algorithms

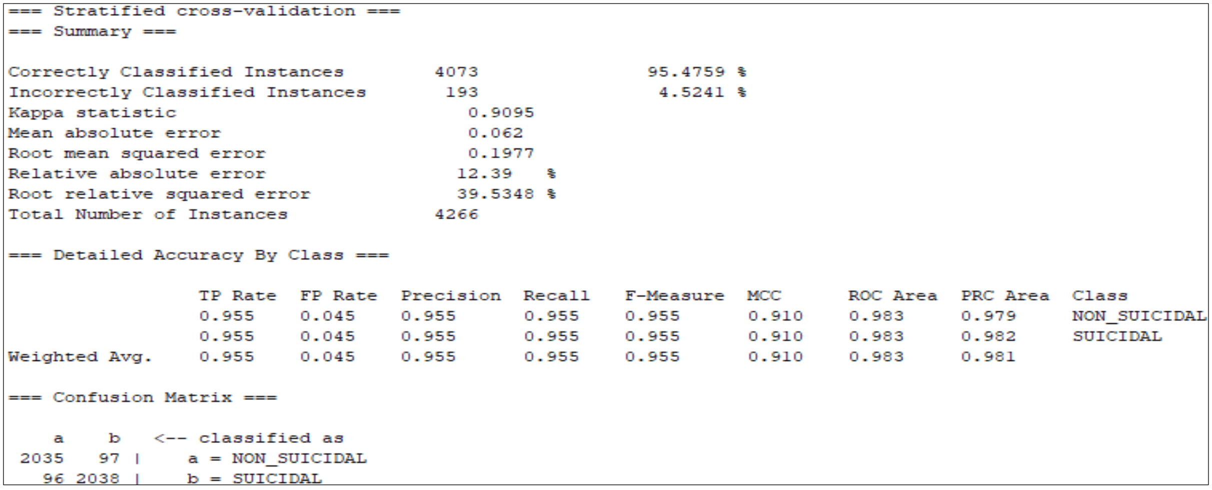
Machine Learning Algorithms			
Algorithms	**Accuracy**	**Precision**	**Recall**
Multinomial Naïve Bayes (MNB)	85.8%	80.2%	95.1%
Naïve Bayes (NB)	81.7%	84.2%	78.1%
Decision Tree (REPTree)	88.8%	89.5%	87.9%
Decision Tree(J48)	**90.7%**	**92.6%**	**88.7%**
Logistic Regression (LR)	**95.4%**	**95.5%**	**95.5%**
SMO	**93.5%**	**94.6%**	**92.3%**
Ensemble Methods of Machine learning			
Methods	Accuracy	Precision	Recall
Bagging	93.4%	94.9%	91.8%
Voting Ensemble	93.8%	93.1%	94.8%
AdaBoost	**95.9%**	**96%**	**95.9%**
Random Forest	**98.5%**	**98.7%**	**98.2%**
Stacking	93.5 **%**	94.6%	92.4%

Figures 8-10 provide the detailed result (screenshot) generated through those machine learning algorithms having an accuracy more than 90 %. Figures 11-12 provides the detailed result (Screenshot) of through ensemble learning approaches (AdaBoost & Random Forest) having an accuracy of more than 95%.



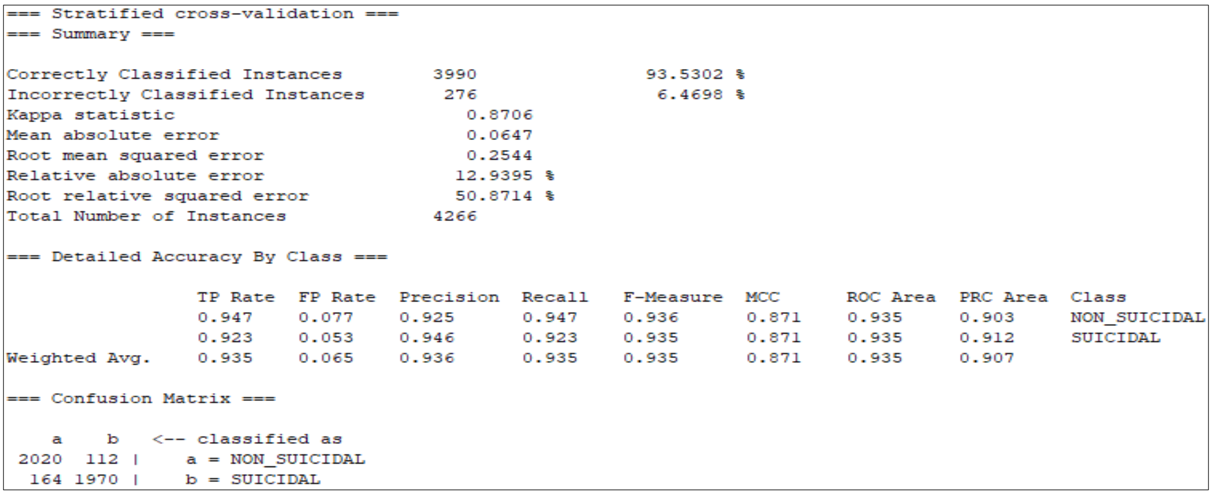
```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3873               90.7876 %
Incorrectly Classified Instances       393                9.2124 %
Kappa statistic                          0.8158
Mean absolute error                      0.1122
Root mean squared error                  0.2807
Relative absolute error                 22.4412 %
Root relative squared error             56.134  %
Total Number of Instances             4266

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.929    0.113    0.891      0.929    0.910      0.816    0.953     0.938     NON_SUICIDAL
                 0.887    0.071    0.926      0.887    0.906      0.816    0.953     0.960     SUICIDAL
Weighted Avg.    0.908    0.092    0.909      0.908    0.908      0.816    0.953     0.949

=== Confusion Matrix ===

    a    b   <-- classified as
 1980  152 |    a = NON_SUICIDAL
  241 1893 |    b = SUICIDAL
```

Figure 8. Result generated using Decision tree (J48) Algorithm



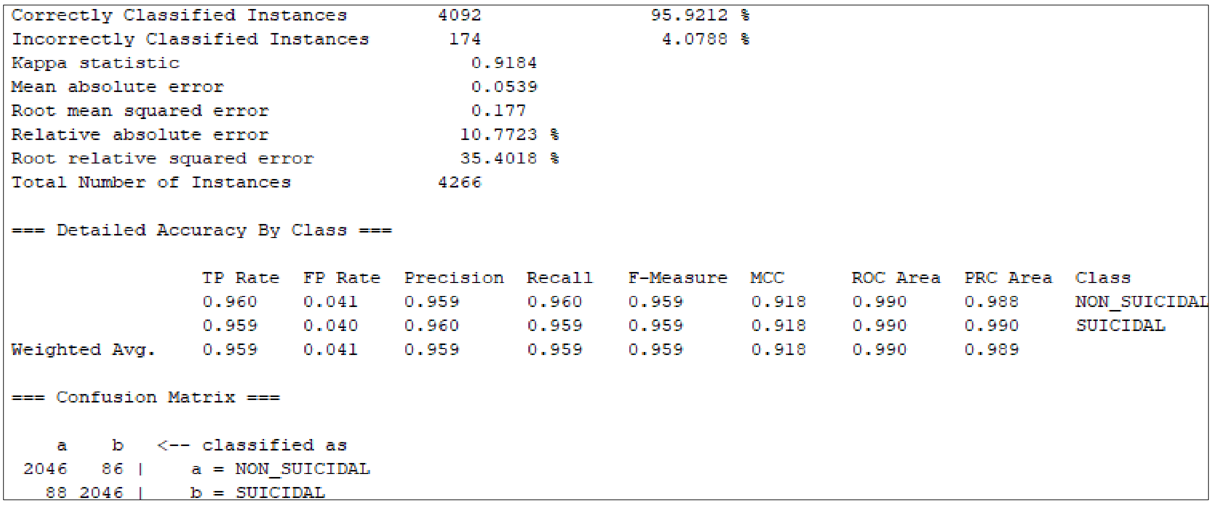
```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4073               95.4759 %
Incorrectly Classified Instances       193                4.5241 %
Kappa statistic                          0.9095
Mean absolute error                      0.062
Root mean squared error                  0.1977
Relative absolute error                 12.39   %
Root relative squared error             39.5348 %
Total Number of Instances             4266

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.955    0.045    0.955      0.955    0.955      0.910    0.983     0.979     NON_SUICIDAL
                 0.955    0.045    0.955      0.955    0.955      0.910    0.983     0.982     SUICIDAL
Weighted Avg.    0.955    0.045    0.955      0.955    0.955      0.910    0.983     0.981

=== Confusion Matrix ===

    a    b   <-- classified as
 2035   97 |    a = NON_SUICIDAL
   96 2038 |    b = SUICIDAL
```

Figure 9. Result generated using Logistic Regression Algorithm



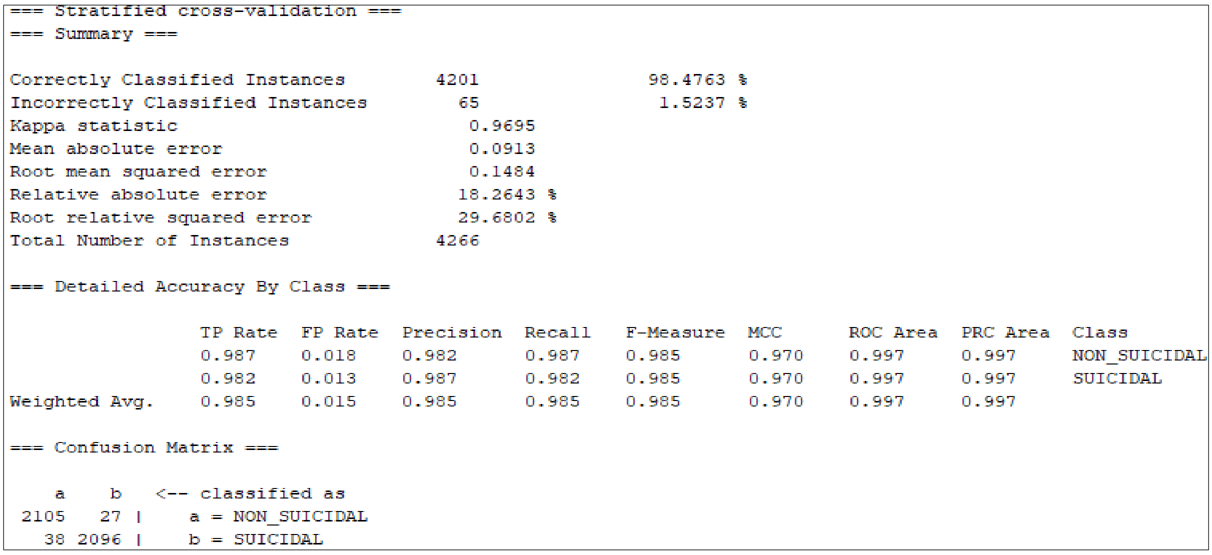
```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3990               93.5302 %
Incorrectly Classified Instances       276                6.4698 %
Kappa statistic                          0.8706
Mean absolute error                      0.0647
Root mean squared error                  0.2544
Relative absolute error                 12.9395 %
Root relative squared error             50.8714 %
Total Number of Instances             4266

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.947    0.077    0.925      0.947    0.936      0.871    0.935     0.903     NON_SUICIDAL
                 0.923    0.053    0.946      0.923    0.935      0.871    0.935     0.912     SUICIDAL
Weighted Avg.    0.935    0.065    0.936      0.935    0.935      0.871    0.935     0.907

=== Confusion Matrix ===

    a    b   <-- classified as
 2020  112 |    a = NON_SUICIDAL
  164 1970 |    b = SUICIDAL
```

Figure 10. Result generated using SMO Algorithm



```
Correctly Classified Instances        4092               95.9212 %
Incorrectly Classified Instances       174                4.0788 %
Kappa statistic                          0.9184
Mean absolute error                      0.0539
Root mean squared error                  0.177
Relative absolute error                 10.7723 %
Root relative squared error             35.4018 %
Total Number of Instances             4266

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.960    0.041    0.959      0.960    0.959      0.918    0.990     0.988     NON_SUICIDAL
                 0.959    0.040    0.960      0.959    0.959      0.918    0.990     0.990     SUICIDAL
Weighted Avg.    0.959    0.041    0.959      0.959    0.959      0.918    0.990     0.989

=== Confusion Matrix ===

    a    b   <-- classified as
 2046   86 |    a = NON_SUICIDAL
   88 2046 |    b = SUICIDAL
```

Figure 11. Result generated using AdaBoost



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4201               98.4763 %
Incorrectly Classified Instances        65                1.5237 %
Kappa statistic                          0.9695
Mean absolute error                      0.0913
Root mean squared error                  0.1484
Relative absolute error                 18.2643 %
Root relative squared error             29.6802 %
Total Number of Instances             4266

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.987    0.018    0.982      0.987    0.985      0.970    0.997     0.997     NON_SUICIDAL
                 0.982    0.013    0.987      0.982    0.985      0.970    0.997     0.997     SUICIDAL
Weighted Avg.    0.985    0.015    0.985      0.985    0.985      0.970    0.997     0.997

=== Confusion Matrix ===

    a    b   <-- classified as
 2105   27 |    a = NON_SUICIDAL
   38 2096 |    b = SUICIDAL
```

Figure 12. Result generated using Random Forest Algorithm

As little work has been done in classifying the data on social media related to suicidal ideation. The proposed work is compared with the most relevant and recent studies(20), (33),(34) regarding identification of suicidal Ideation. It is found that none of these studies achieved the accuracy closer to our work. Further, the studies done previously have used simple feature extraction techniques and data size used to train the model was also low leading to the bias and skewness in of the machine learning model. The imbalanced data also caused skewness in majority/minority class. Our machine learning achieved balance between precision and recall values due to the application of feature selection techniques for extracting relevant features. The most performing algorithm was Random Forest having an accuracy of 98.5%, precision of 98.7% and recall value of 98.2%. Some of the highlights of comparison is reflected in Table 2 below.

Table 2. Comparison of our Proposed work with some recent studies

Authors	**Dataset Size used**	**Machine learning used**	**Ensemble Approaches used**	**Features Used**			**Feature reduction technique used**	**Best Accuracy achieved**
				TFIDF	**BOW**	**Manual**		
Bridianne O'Dea et al (20)	2820	✓	×	✓	×	×	×	76 %
Fatima Chiroma (34)	1064	✓	×	✓	✓	×	×	80%
Akshama Chanda (33)	1897	✓	✓	×	×	✓	×	79.65 %
Proposed Work	**4266**	✓	✓	✓	✓	×	✓	**98.5%**

Evaluation and Validation of Proposed Model:

The performance of algorithms was mainly evaluated through confusion metrics and the corresponding ROC curve. ROC curve is the curve generated by comparing the true positive rate (TPR) with false-positive Rate (FPR). TPR represents the recall, and FPR indicates the probability of false alarm. The classification model is also evaluated through another metric called the area under the curve (AUC). The metric produces a value between 0 and 1. The closer to 1 indicates the better classification. AUC for various classification algorithms are as under

AUC for Naive Bayes = 0.893, AUC for Multinomial Naïve Bayes = 0.9593, AUC for REPTree = 0.9373, AUC for J48= 0.9527, AUC for Logistic Regression = 0.983, AUC for SMO = 0.9353, AUC for Bagging =0.9751, AUC for Voting = 0.9834, AUC for AdaBoost = 0.9896, AUC for Random Forest = 0.9972, AUC for stacking = 0.9355. Moreover, the model was validated through the 10 cross-validation technique. Through this technique, the data is partioned randomly into 10 equal subsamples. Out of 10 partitions, one subsample is used for testing while other 9 subsamples are used for training purposes. The cross-validation is then repeated 10 times, but each of the 10 subsamples is used only once for validating. The results drawn 10 times are then averaged to produce a single estimation. Confusion metrics work on four kinds of metrics as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).TP & TN imply the correct predicted values while as FP & FN imply otherwise. Based upon these confusion metrics, Precision & recall were calculated. Precision is the number of TP divided by the number of TP plus number Of FP. The recall is defined as number of TP divided by number of TP plus number of FN. From the experiments, it was found that logistic regression and SMO had a similar kind of accuracy as of Ensemble approaches. SMO has an accuracy of 93.5%, precision of 94.6% and a recall of 92.3%. On the other hand, Logistic Regression has an accuracy of 95.4%, precision of 95.5% and a recall of 95.5%. In Ensemble methods, Random Forest provides the highest accuracy of 98.5%, precision of 98.7% and recall of 98.2%. Other machine learning algorithms also performed fairly having the accuracy above 80%. The ROC curves of the best performing algorithms are depicted below in figures 13-17, and figure 18 provides the graphical view of comparative performance of the various machine and ensemble learning algorithms in classifying the tweets into two levels of concern.

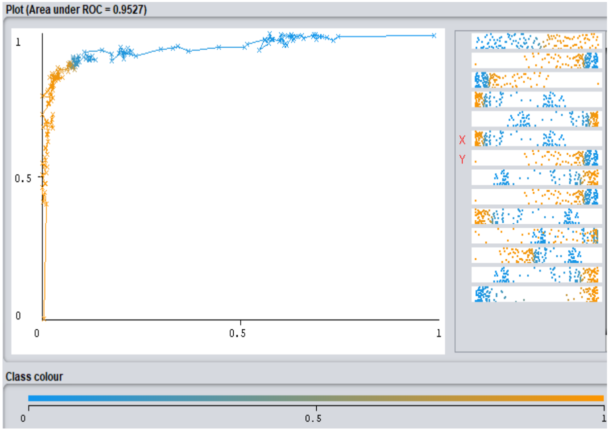


Figure 13. ROC Curve of J48 (AUC =0.9527)

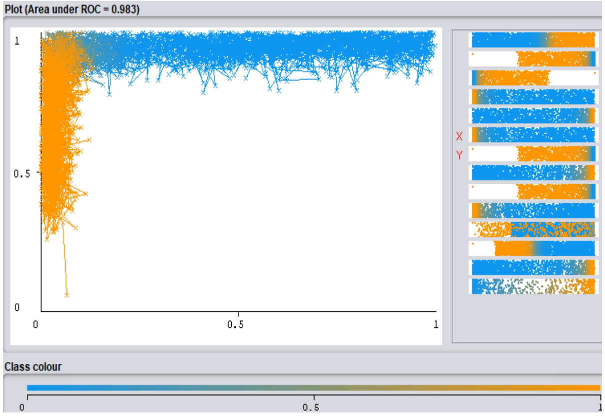


Figure 14. ROC Curve of LR (AUC=0.983)

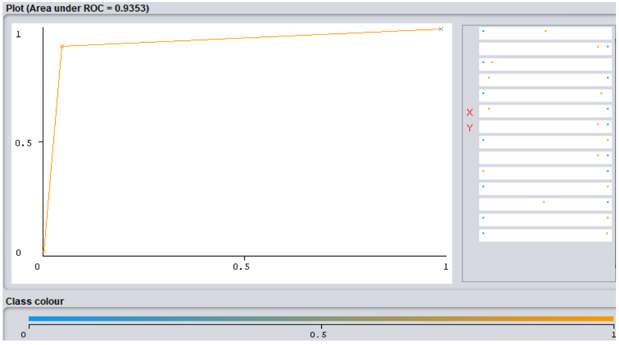


Figure 15. ROC Curve of SMO (AUC= 0.9353)

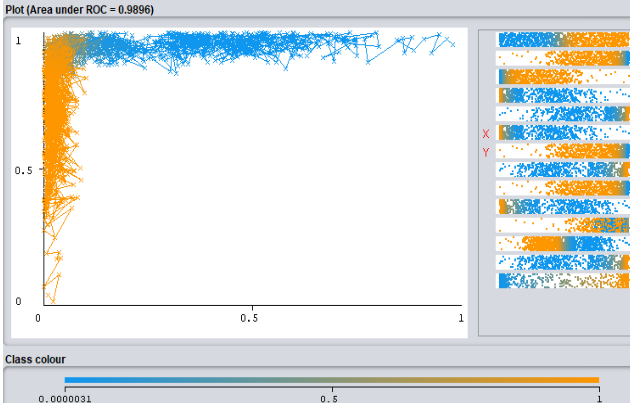


Figure 16. ROC Curve of AdaBoost (AUC = 0.9896)

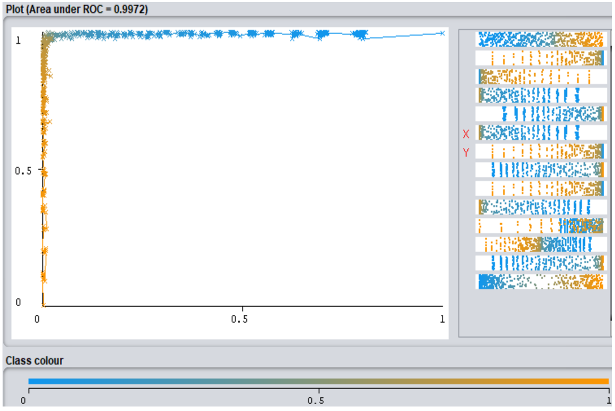


Figure 17. ROC Curve of RF (AUC = 0.9972)

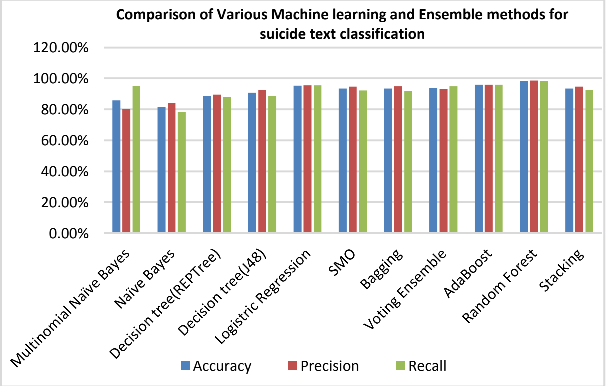


Figure 18. Comparative analysis of various Ensemble and Machine learning Algorithms

Conclusion and Future Work:

The amount of text on social media is growing on an exponential rate with the use of popular social networking sites (SNS), and the people posting their feelings & thoughts on these SNS. It is, therefore, a necessity to explore and develop new techniques for detecting posts with suicidal ideation with a hope to prevent potential suicide victims.

In this research paper, the possibility of detecting suicidality from the content generated through social media is explored and analysed. Most of the research in mental health was done by psychological experts with the help- of statistical tools. There are various limitations, like privacy and cost associated with obtaining the required data. Moreover, the social stigma associated with mental health hampers the findings.

In this paper, various machine learning algorithms and ensemble approaches like Decision trees, Naïve Bayes, Multinomial Naïve Bayes, Support vector machine (SMO), Regression, Bagging, Random Forest, AdaBoost, voting and Stacking are implemented for classifying the suicide-related tweets into two groups using the real tweets extracted through Twitter API. Random Forest was found to be the most effective and performs better with the accuracy of 98.5%, precision of 98.7% and recall value 98.2%. This research work will help our future researchers to work out on various other important aspects of the research and can contribute to this emerging problem with the new ways of solving the problem. Some of the future directions in this regard are:

In future, following points will be considered in our research.

- The relevant inputs captured through the questionnaires will be used in our machine learning algorithms for more accuracy.
- The connectivity between suicidal users will be explored.
- Blog posts will be investigated that will provide the rich knowledge in understanding the behaviour of suicidal users.
- Multi-class classification will be done to separate the different levels of distress.
- Deep learning algorithms like RNN, GAN and LSTM will be used to explore its feasibility in classifying the text data related to suicide.

Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Besides, the Figures and images, which are not ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in Baba Ghulam Shah Badshah University.

References:

1. Desmet B, Hoste V. Online suicide prevention through optimised text classification. Inf Sci (Ny). 2018;439–440:61–78.
2. Patel V, Ramasundarahettige C, Vijayakumar L, Thakur JS, Gajalakshmi V, Gururaj G, et al. Suicide mortality in India: A nationally representative survey. Vol. 379, The Lancet. Lancet Publishing Group; 2012. p. 2343–51.
3. Fu K, Cheng Q, Wong PWC, Yip PSF. Responses to a Self-Presented Suicide Attempt in Social Media. Crisis. 2013 Nov;34(6):406–12.
4. Gunn JF, Lester D. Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death. Suicidologi. 2015 Jun 1;17(3).
5. Kailasam VK, Samuels E. Can social media help mental health practitioners prevent suicides? Curr Psychiatr. 2015;14(2):37–39, 51.
6. Huang X, Xing L, Brubaker JR, Paul MJ. Exploring Timelines of Confirmed Suicide Incidents Through Social Media. Proc - 2017 IEEE Int Conf Healthc

Informatics, ICHI 2017. 2017;470–7.

7. Vioules MJ, Moulahi B, Aze J, Bringay S. Detection of suicide-related posts in Twitter data streams. IBM J Res Dev. 2018;62(1):7:1-7:12.
8. Choudhury M De, Gamon M, Counts S, Horvitz E. Predicting depression through social media. Proceedings of ICWSM. 2013.
9. Kiciman E, Kumar M, De Choudhury M, Coppersmith G, Dredze M. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In 2016.
10. How Twitter Users Compare to the General Public | Pew Research Center [Internet]. [cited 2020 Jan 1]. Available from: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/
11. Wolk-Wasserman D. Suicidal communication of persons attempting suicide and responses of significant others. Acta Psychiatr Scand. 1986;73(5):481–99.
12. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis. 2014;35(1):51–9.
13. Catal C, Nangir M. A sentiment classification model based on multiple classifiers. Appl Soft Comput J. 2017 Jan 1;50(50):135–41.
14. Kiritchenko S, Zhu X, Mohammad SM. Sentiment analysis of short informal texts. J Artif Intell Res. 2014 Aug 20;50:723–62.
15. Shiori T, NISHIMURA A, AKAZAWA K, ABE R, NUSHIDA H, UENO Y, et al. Incidence of note-leaving remains constant despite increasing suicide rates. Psychiatry Clin Neurosci. 2005 Apr;59(2):226–8.
16. Sikander D, Arvaneh M, Amico F, Healy G, Ward T, Kearney D, et al. Predicting risk of suicide using resting state heart rate. In: 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2016. Institute of Electrical and Electronics Engineers Inc.; 2017.
17. Huang Y-P, Goh T, Liew CL. Hunting Suicide Notes in Web 2.0 - Preliminary Findings. In Institute of Electrical and Electronics Engineers (IEEE); 2008. p. 517–21.
18. Varathan KD, Talib N. Suicide detection system based on Twitter. Proc 2014 Sci Inf Conf SAI 2014. 2014;785–8.
19. Desmet B. Automatic text classification for suicide prevention. 2014;1–205.
20. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on twitter. Internet Interv. 2015;2(2):183–8.
21. De Choudhury M, Kıcıman E. The Language of Social Support in Social Media and its Effect on Suicidal Ideation Risk. Proc . Int AAAI Conf Weblogs Soc Media Int AAAI Conf Weblogs Soc Media. 2017;2017.
22. Coppersmith G, Ngo K, Leary R, Wood A. Exploratory Analysis of Social Media Prior to a Suicide Attempt. Proc Third Work Comput Lingusitics Clin Psychol. 2016;106–17.
23. Joseph A, Ramamurthy B. Suicidal behavior prediction using data mining techniques. Int J Mech Eng Technol. 2018;9(4):293–301.
24. American Association of Suicidology – Suicide Prevention is Everyone's Business [Internet]. [cited 2020 Apr 8]. Available from: https://suicidology.org/
25. AL-Jumaili AS. A Hybrid Method of Linguistic and Statistical Features for Arabic Sentiment Analysis. Baghdad Sci J. 2020;17(1(Suppl.)):0385.
26. Bao Y, Quan C, Wang L, Ren F. The role of pre-processing in twitter sentiment analysis. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag; 2014. p. 615–24.
27. Anand N, Goyal D, Kumar T. Analyzing and Preprocessing the Twitter Data for Opinion Mining. In: Lecture Notes in Networks and Systems. Springer; 2018. p. 213–21.
28. Dar AR, Ravindran D. Fog computing resource optimization: A review on current scenarios and resource management. Baghdad Sci J. 2019;16(2):419–27.
29. Cash SJ, Thelwall M, Peck SN, Ferrell JZ, Bridge JA. Adolescent Suicide Statements on MySpace. Cyberpsychology, Behav Soc Netw. 2013;16(3):166–74.
30. Verma P, Khanday AMUD, Rabani ST, Mir MH, Jamwal S. Twitter sentiment analysis on Indian government project using R. Int J Recent Technol Eng. 2019 Sep 1;8(3):8338–41.
31. Burnap P, Colombo G, Scourfield J. Machine Classification and Analysis of Suicide-Related Communication on Twitter. 2015;
32. Ji S, Yu CP, Fung S-F, Pan S, Long G. Supervised Learning for Suicidal Ideation Detection in Online User Content. Hindawi Complex. 2018;2018:1–10.
33. Chadha A, Kaushik B. A Survey on Prediction of Suicidal Ideation Using Machine and Ensemble Learning. 2019;00(00).
34. Chiroma F, Liu HAN, Cocea M. Text Classification For Suicide Related Tweets. 2018 Int Conf Mach Learn Cybern. 2:587–92.

اكتشاف الميول الانتحارية على تويتر بأستخدام التعلم الالي وطرق المجموعة

سيد تنزيل رباني قمر ريس خان عكب محي الدين الدين خندي

قسم علوم الحاسوب، جامعة بابا غلام شاه بادشاه -185234 ، راجوري، جي أند كي، الهند

الخلاصة:

يعد التفكير في الانتحار من أخطر مشكلات الصحة العقلية التي يواجهها الناس في جميع أنحاء العالم. هناك عوامل خطر مختلفة يمكن أن تؤدي إلى الانتحار. من أكثر عوامل الخطر شيوعًا وأكثرها خطورة الاكتئاب والقلق والعزلة الاجتماعية واليأس. يمكن أن يساعد الاكتشاف المبكر لعوامل الخطر هذه في منع أو تقليل عدد حالات الانتحار. أصبحت منصات الشبكات الاجتماعية عبر الإنترنت مثل تويتر وريدت وفيس بوك طريقة جديدة للناس للتعبير عن أنفسهم بحرية دون القلق بشأن الوصمة الاجتماعية. تقدم هذه الورقة منهجية وتجربة باستخدام وسائل التواصل الاجتماعي كأداة لتحليل الأفكار الانتحارية بطريقة أفضل ، وبالتالي المساعدة في منع فرص الوقوع ضحية لهذا الاضطراب العقلي المؤسف. نجمع البيانات ذات الصلة عبر توترأحد مواقع الشبكات الاجتماعية الشهيرة (SNS) . ومن ثم تتم معالجة التغريدات يدويًا وإضافة تعليقات توضيحية لها يدويًا. وأخيرًا ، يتم استخدام أساليب التعلم الآلي المختلفة والمجموعات لتمييز التغريدات الانتحارية وغير الانتحارية تلقائيًا. ستساعد هذه الدراسة التجريبية الباحثين على معرفة وفهم كيفية استخدام الأشخاص للتعبير عن النفس في التعبير عن مشاعرهم وعواطفهم. وأكدت الدراسة أيضًا أنه من الممكن تحليل وتمييز هذه التغريدات باستخدام التشفير البشري ثم تكرار الدقة حسب تصنيف الماكينة. ومع ذلك ، فإن قوة التنبؤ للكشف عن الانتحار الحقيقي لم يتم تأكيدها بعد ، وهذه الدراسة لا تتواصل بشكل مباشر وتتدخل مع الأشخاص الذين لديهم سلوك انتحاري..

الكلمات المفتاحية : تعلم المجموعات، تعلم الآلة، أفكار انتحارية، تصنيف النص، تويتر، ويكا.