

# A Sentiment Analysis of Suicidal Notes using Machine Learning

Farziz Aktar Ahmed  
Student

Faculty of Engineering and  
Technology Assam Down Town  
University Guwahati, India

Junumoni Khakhlari  
Student

Faculty of Engineering and  
Technology Assam Down Town  
University Guwahati, India

Nitumani Sarmah  
Assistant Professor

Faculty of Engineering and  
Technology Assam Down Town  
University Guwahati, India

## ABSTRACT

Suicide becomes an unavoidable issues for the modern society. Every year 800000 peoples loses their lives worldwide. Detecting sentiments of these peoples is a new challenge for many researchers. It can be achieved by analyzing their statements made by the victims before they commit any suicidal actions, with the help of machine learning, natural language processing etc. In this paper, we have used a dataset which contains 232074 unique values collected posts from “Suicide Watch” and “depression” subreddits of the Reddit platform, to develop different machine learning model to analyze the sentiments of these data. We developed several types of machine learning model to compare the accuracy and find out the best and suitable algorithm for the project of detecting people's sentiment. The accuracy we able to achieved, SVM 57.24%, Naive Bayes (Gaussian) 54.69%, Random forest 67.67%, Decision tree 70.95. Along with these algorithms we have also developed different versions of Naive Bayes model algorithm where Naive Bayes (Bernoulli), Naive Bayes (Multinomial) and Naive Bayes (Gaussian) able to achieve an accuracy of 49.92%, 51.65%, 54.69% accordingly. Here we have found that Decision Tree is providing best accuracy compare to another model algorithm. In addition, among all the versions of Naive Bayes model algorithms Bayes (Gaussian) is providing the best accuracy.

## Keywords

Sentiment Analysis, Suicide attempts, Decision Tree algorithm, Machine Learning, Support Vector Machine, Random Forest, Naive Bayes, Deep Learning.

## 1. INTRODUCTION

Suicide is one of the main general health concerns around the world consuming 8,00,000 lives each year, and normal suicide happens at regular intervals [1], [2]. Among the complete worldwide passing because of suicide, very 1,35,000 are from India alone [3]. World Health Organization (WHO) and the American Foundation for Suicide Prevention (AFSP) recognizes different risk factors that destruction decline the limit for suicide. Some basic risk factors are sadness, stress, a sleeping disorder, loss of relatives, sadness, and past suicidal endeavors. Suicide is a typical strategy viewed as an extremely durable answer for a temporary problem. In the beginning phases of the suicidal life-cycle, what begins with suicide idea, if the risk factors are being understand, and the patient is referred for treatment, many lives could be save.

This research is done by collecting the various data from the various social media platforms, where people share their different thoughts among all the people of the world. The social media platform helps us to detect the sentiment of the people from their post. Moreover, the automatic detection of suicidal comments

plays a key role to understand the sentiment of people and it will help to prevent the suicidal behavior of the people.

## 2. LITERATURE REVIEW OF PREVIOUS WORK

The utilization of social media has filled suddenly in these most recent couple of years. Every one individual is associated through social media those are the people who are utilizing cell phone. In different social media platform individuals are imparting through messages, offering their viewpoint through remarks and furthermore, they are also sharing their everyday life occurrences. However when the assessment, remarks, messages are in tremendous numbers like in billions it becomes difficult to examine the opinions of those information physically and here the ideas emerge to dissect the opinions of individuals that what really, they need to communicate. Are their thoughts is reflecting positive opinion, negative feeling, or neutral feeling. Various researcher utilize various techniques to precisely break down the feeling to an ever-increasing extent. For example, Naive Bayes, Support Vector machine (SVM), Decision Tree, Random Forest and so on. These are a few most famous calculations for opinion examination among the Researchers, and utilizing these calculations they can accomplished a very decent precision in their explores.

Go et.al.[15] in their research paper “Twitter Sentiment Analysis” expressed that they have analyzed their feeling detection on a bunch of twitter information to break down the opinions of client. They gathered their own information utilizing twitter API, every one of those twitter messages that have feelings. They have utilized Naive Bayes classifiers which they have worked without any preparation and Third-party library were utilized for Maximum Entropy and Support Vector Machine (SVM). Utilizing these techniques, they are capable get a precision of 73.913% with Support Vector Machine and with Naive Bayes they get an exactness of 44.9% and for Maximum Entropy it is did not contribute a lot to get a higher exactness.

Pak et. Al. [16] in their studies “Twitter as a Corpus for Sentiment Analysis and Opinion Mining” they have showed a programmed assortment of corpus that they can use to prepare feeling classifier. In this paper they have utilized multinomial Naive Bayes classifier that utilizes N-gram and POS-labels as highlights to group the substance. Additionally, they have involved TreeTagger for POS-labeling to noticed the distinction in circulations among positive, negative, and neutral sets.

Saif et. al. [17] in the paper "Semantic Sentiment Analyze of Twitter" the analyst has used Stan passage Twitter Sentiment Corpus (STS), Health Care Reform (HCR), Obama-McCain Debate (OMD) Datasets to characterize the feeling of individuals. In their paper, they have presented a methodology for adding semantics as an extra component into their preparation set for

detecting the feeling of peoples. They have applied this methodology for the expectation of feeling for various Twitter datasets. Likewise, they have approach for Naïve Bayes calculation to perform their examination on opinion. In this examination paper by performing Stanford Twitter Sentiment Corpus, they got precision of 80.7%, with Health Care change they can get the exactness 71.1% and furthermore by drawing nearer to Obama-McCain technique the researcher can get the accuracy of 75.4%.

Sohn et.al. [18] in this diary paper "A Hybrid Approach to Sentiment Sentence Classification in Suicide Notes" the scientist has utilized datasets as 600 real suicide notes. The researcher has utilized two calculation named as NAÏVE BAYES and RIPPER.

Pestian et. al. [19] in their paper "sentiment Analysis of suicide notes: A shared Task" expressed that they have utilized 1319 individuals suicide notes (1950-2011, CHRISTINE) as a dataset in their paper. They have done their exploration involving a common errand in Biomedical domain which incorporates two elements one is the Anonymized clinical texts and explained suicide notes and the other one is it requires classification enormous arrangement of marks. In this paper they have depict about the difficulties to group the feelings found in notes abandoned by the people who have died by suicide in 2011. Altogether 106 researchers who have involved 24 groups answered the require the cooperation. This paper's outcomes were introduced at the Fifth i2b2/VA/Cincinnati Shared-Task and Workshop: Challenges in Natural Language Processing for Clinical Data in Washington, DC, on October 21-22, 2011, as an American Medical Informatics Association Workshop.

McCart et. al. [20] in their examinations "Utilizing Ensemble Models to Classify the Sentiment Expressed in Suicide Notes" in their diary they have utilized the dataset comprised of 900 self-destruction notes gathered north of a 70-year duration (1940-2010). Their best accommodation utilized an outfit of the two guidelines and STM models to accomplish a miniature found the middle value of F1 score of 0.5023, over the mean from the 26 groups that contended (0.4875). Likewise, they have involved calculation as Decision trees, KNN, SVM to sift through the opinion of individuals from different proclamation which are posted in various web-based entertainment stages.

Neri et. al. [21] in this diary paper "Feeling Analysis on Social Media" the researcher has utilized the dataset as 1000 posts-by center creeping of Facebook. In this paper the analyzer has utilized Recall and Precision algorithm to filter the feeling of the remarks which are posted in social media platforms by the people to recognize which are positive, negative, or neutral. They have performed their Analysis on different social media platform posts about reports, they looked at the opinion for "Rai - the Italian broadcasting service and many more dynamic private companies La7, etc. Their review maps concentrate on the outcomes with perceptions which is made by the Osservatorio di Pavia, an Italian foundation of exploration which is well versed in the media analysis at the hypothetical and exact level. The analyst got 87% exactness by utilizing recall algorithm and they got exactness of 93% by utilizing Precision algorithm.

Graves et. al. [22] in this journal paper "Use of Sentiment Analysis for Capturing Patient Experience from Free-Text Comments Posted Online" have mentioned that they have utilized NHS Choices Datasets to detect the sentiment of people. They have used machine learning techniques to extract the sentiment from 6412 online comments about hospitals on the English National Health Service website in 2010 using Weka data-mining software. In this paper [16] they have used Naïve Bayes Multinomials algorithm they got the accuracy of 88.6% and using

Decision trees the accuracy they got is 80.8%. Using Bagging they got the accuracy 82.5% and lastly by using AVIM they got 84.6%.

Munezero et. al. [23] in their paper "Exploiting Sentiment Analysis to Track Emotions in Students' Learning Diaries" referenced that they present a useful framework for detecting and picturing understudies feelings communicated in their journals. This framework permits the teachers to remove the feelings which were mentioned in students' diaries. They utilized a dataset of students Diaries from the Newman. Additionally, they have utilized "Stop word expulsion strategy", "Potter's stemming algorithm".

Altrabsheh et. Al. [24] in their paper "SA-E: Sentiment Analysis for Education" they have done investigate on another point opinion investigation on schooling. For this they have gathered understudies' inputs from social media platform like twitter, to break down the feelings of understudies to comprehend whether the understudies are good, bad, or having some other feelings. In this study they have introduce that Naïve Bayes and Support Vector Machine (SVM) can be joined to dissect the understudies' criticism in Real-time, and it holds an extraordinary potential. Alongside this they have presented another framework design named as System Analysis for Education (SA-E).

Sarlan et.al. [25] the paper named "Twitter Sentiment Analysis" expressed that they have involved twitter information as dataset, likewise they utilized Natural Language Processing (NLP), Case-Based Reasoning (CBR), Artificial Neural Network (ANN), Support Vector Machine (SVM) algorithm to extract the information of opinions whether its positive, negative, or neutral. By utilizing Support Vector Machine (SVM) they got the precision as 81.3%.

Birjali et. al. [26] in this paper "Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks" they utilized 892 TWEETS (Using Twitter4J API) as a dataset to characterize the sensations of individuals. Additionally, they have utilized calculation IB1, J48, CART, SMO, NAÏVE BAYES.

Aladağ et.al. [27] in their diary paper "Identifying Suicidal Ideation on Forums: Proof-of-Concept Study" they said that they have utilized the dataset as Reddit dataset (2008 - 2016). In their paper they have said that they have utilized Logistic relapse, arbitrary woods, SVM, Baseline ZeroR calculations to play out the arrangement of opinion of individuals whether the feeling of the remark is happy, sad, or angry. They have involved technique as a sum of 508,398 Reddit posts were posted somewhere in the range of 2008 and 2016 on Suicide Watch, it has longer than 100 characters in their posts. In their paper Depression, Anxiety, and Shower Thoughts subreddits were downloaded from the openly accessible in Reddit dataset. 10,785 posts were haphazardly chosen and 785 were physically commented on as suicide or non-suicide in their paper. The researcher has the precision of 80% by utilizing Logistic relapse, involving arbitrary woods they got the exactness as 92%, they got the precision of half by utilizing Support Vector Machine (SVM), finally by utilizing gauge ZeroR calculation they got the exactness of 66%.

RamyaSri et. al. [28] in their work "Feeling Analysis of Patients' Opinions in Healthcare using Lexicon-based Method" they have done explore on feeling examination. To perform the analysis, they have utilized patients' perspective information from 92 webpages where it contains assessments of patients from Southern California Orthopedic Institute which is situated in California, USA. To perform feeling examination on this information they have utilized Lexicon Based technique. Opinion examination apparatuses like "VADER" and "TextBlob" are utilized for arrangements. Utilizing these strategies, they can accomplish

exactness as 71.9% utilizing VADER dictionary-based approach and 73.0% in the TextBlob vocabulary-based approach in their work.

Mbarek et. al. [29] in the paper "Suicidal Profiles Detection in Twitter" in this paper the researcher has expressed that they have utilized 115 suicide profiles, 172 not suicide profiles (Using TWITTER HEREAFTER Site) as a dataset. They have utilized Random Forest, BayesNet, Adaboost, J48, SMO these calculations to sort out the proclamation which are positive, negative, or neutral. By utilizing Random Forest calculation, the analyst has 77% precision, utilizing SMO the researcher has 74% exactness.

Qaiser et. Al. [30] in their journal paper named "Feeling Analysis of Impact of Technology on Employment from Text on Twitter" they attempted to examine the opinion of individuals about the effect of innovation on joblessness and specialized head ways. In these examinations, they have viewed that as 65% individuals have a negative feeling about the Effect of innovation on joblessness and specialized progression. In this paper they have prepared Naïve Bayes AI classifier to characterize the information as per the clients' opinions. Utilizing Naïve Bayes and Support Vector Machine (SVM), they ready to accomplished a precision of 87.18% and 82.05% respectively.

Glenn et.al. [31] in the paper named "Can Text Messages Identify Suicide Risk in Real Time? A Within-Subjects Pilot Examination of Temporally Sensitive Markers of Suicide Risk" the researcher has expressed that they have utilized DIGITAL TEXT DATA (social media) as their datasets. They have utilized Machine Learning algorithm whether the statements are positive, negative, or neutral.

Ji et.al. [32] in this paper "Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications" the researcher has expressed that they use datasets as TEXT DATA (Reddit, Twitter, Reach Out), EHR, Mental Disorders (surveys III-A suicide notes III-C suicide sites III-C electronic wellbeing records III-B online social texts III-D). Likewise, they have involved algorithm as Machine Learning and DEEP LEARNING to remove the opinion of the information which are accessible in friendly stage.

Sharma et.al. [33] in their paper named "Analyzing the depression and suicidal tendencies of people affected by COVID-19's lockdown using sentiment analysis on social media websites" the researcher has used the Twitter information as a dataset. Additionally, they have involved the algorithm as Machine Learning (Unsupervised) to filter the feeling of the statement of individuals which are expressed in different social media platform to arrange whether its happy, sad, or angry.

George et. al. [34] in their examination paper "Application of Aspect-based Sentiment Analysis on Psychiatric Clinical notes to Study Suicide in Youth" they said that they have involved 1559 suicide notes as a dataset in their paper to break down. They have involved Weka as an instrument of data mining that can remove valuable information from Twitter which information are gathered by Twitter4J for better investigation. Likewise, they said that they have utilized Logistic regression, Random Forest algorithm to classify the feeling of individuals by the suicidal notes.

### 3. METHODODLOGY

To analyze the sentiment of a statement whether the sentiment is positive or negative, in terms of our paper is the statement have a suicidal tendency behind it or the statement is suicidal or non – suicidal. It combines machine learning and natural language processing to achieve this. In this paper we are going to perform sentiment analysis using some traditional methods as well as

Ensemble methods and try to find out the best algorithm that provides maximum accuracy.

Machine Learning is the domain of studies of different computer algorithms that can improve through experience. Simon has defined the Machine Learning as "the process of a change and enhancement in the behaviors through exploring latest information in time." It used for to solve the complex problems by using the previous data.[5] In machine Learning supervised learning methodologies are going perform in this project. The below mentioned machine learning algorithms are used to perform sentiment analysis in this paper.

**SUPERVISED LEARNING:** Supervised Learning is a kind of methodology where the process of providing input data and giving the correct output to the machine learning model. It also helps us to solve various real-world problems [4].

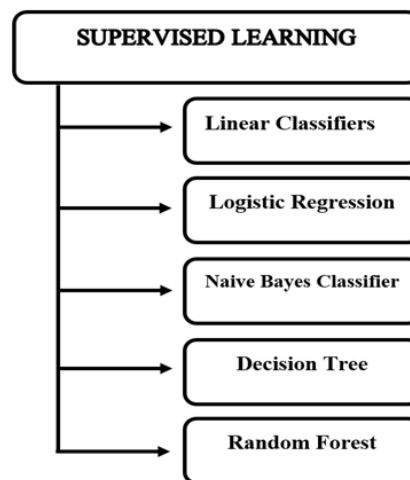


Figure 1: - Classification of Supervised Learning [4][35]

**DECISION TREE CLASSIFIERS:** - Decision tree is a supervised learning methods that creates a classification model by building a tree like structure. Decision tree has two nodes decision node and leaf node. [6] It has the capability to handle the large amount of information.[4].It has several types of Decision Tree algorithms such as:

1. Iterative Dichotomies 3 (ID3),
2. Successor of ID3 (C4.5),
3. Classification And Regression Tree (CART) [7],
4. CHI-squared Automatic Interaction Detector (CHAID) [8],
5. Multivariate Adaptive Regression Splines (MARS) [9],
6. Generalized, Unbiased, Interaction Detection and Estimation (GUIDE), Conditional Inference Trees (CTREE) [10],[11],
7. Classification Rule with Unbiased Interaction Selection and
8. Estimation (CRUISE), Quick, Unbiased and Efficient
9. Statistical Tree (QUEST). [12]

**SUPPORT VECTOR MACHINES:** - Support Vector Machines is associated to classical multiplayer perceptron neural networks. Support vector machines (SVMs), introduced by Vapnik and his coworkers in the early 1990's (Cortes, Vapnik 1995; Vapnik 1996, 1998), these are proved to be so effective techniques

for data mining (Peng et al. 2008; Yang, Wu 2006).[13][14]. Support Vector Machine divide the datasets in to classes to figure out a maximum marginal hyperplane (MMH) [4].

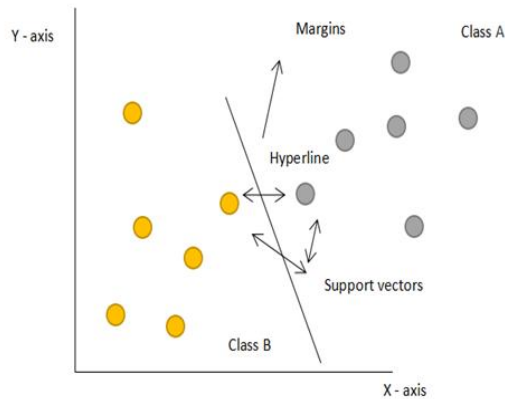


Figure 2: - Working graph of SVM [4][36]

- i. **NAIVE BAYES:** - Naïve Bayes is the simplest and most powerful machine learning algorithm which is used to solve classification problem. It is used for text classification. A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". Naive Bayes theorem explains the probability of an event occurring given the probability of another event that has already occurred [4]. The equation of Naive Bayes theorem is given below -

Equation (1):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where it defines A and B are the events and  $P(B) \neq 0$ .

**Gaussian Naive Bayes:** - Gaussian Naive Bayes is a different version of Naive Bayes which follows the Gaussian normal distribution and supports continuous data containing many non-prior knowledge. A way to deal with make a straightforward model is to expect that the information is described by a Gaussian method with no independent dimensions between dimensions.

**Bernoulli Naive Bayes:** - Bernoulli Naive Bayes is a different version of Naive Bayes. It is used for discrete data which works on Bernoulli distribution. It depends on the Bernoulli Distribution and acknowledges binary values, which is 0 or 1.

**Multinomial Naive Bayes:** - The Naive Bayes method is a strong tool for analyzing text input and solving problems with numerous classes. Multinomial Naive Bayes algorithm is a probabilistic learning strategy that is utilized in Natural Language Processing (NLP).

The algorithm depends on the Bayes theorem and predicts the tag of a text. It computes the probability of each tag for a given example and afterward gives the tag with the highest probability as result.

**Random Forest:** - Random Forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. Random Forest, like its name proposes, contains several numbers of individual decision trees that work in

coordinates.[4] Each-and-every individual tree in the irregular forest lets out a class forecast and the class with the most votes, which we select as our model for prediction.

#### 4. IMPLEMENTATION:

This project includes dataset where all the data is taken from the posts of "Suicide Watch" and "depression" subreddits of the Reddit platform etc., and the dataset contain the comment of the post, their own post. After importing the dataset first work to get the size of the dataset, for that we use ". shape" function and we get shape as (232074, 3). Since the dataset contain various emoji's and stop-words and individual character hence it is particularly important to clean the data in the first place. For that we use "neattext.functions" which is use to clean the text in column wise. Now we get the clean data and now it was ready for splitting. But as in this paper we are implementing different model algorithm so in the first place we must convert string dataset to an integer dataset. For that we must perform level encoding to convert the dataset. Now the dataset is ready for splitting. We split the data according to the calculation so that we get maximum accuracy. Now the main data set is split into 7:3 ratio, where 70% is training dataset and where 30% testing dataset. After that we remove the target column from both the part and store differently in the different dataframe and the remaining data on a separate dataframe.. As we are working on a label dataset so we implemented supervised learning model. In this dataset the target column having a binary value, for that we mistrusted classification model only. The first model we use is Decision Tree Algorithm where we use decision tree classifier as the data set was binary target column. Decision tree is structured tree classifier in which internal node represent the feature in the dataset, the branches represent the decision rule and leaf node represents the decision. In simple words root node represent the entire dataset, and the leaf node represents the output. As decision tree simple in decision making, so it is one of the fastest algorithms to work on. Using this algorithm, we got high accuracy in the model.

- a. Second model we use **Support Vector Classifier** which import from sklearn library. This model works on a linear kernel function to classification. It works exceptionally good with large data size. After implementing this model, we got high accuracy.
- b. Third algorithm we use in Naïve Bayes. In this algorithm we use Gaussian NB, as data is categorical and we must perform classification model. Naïve Bayes works on Probability distribution. Gaussian Distribution is also known as normal distribution. In this model we got near about 50% accuracy.
- c. Fourth algorithm we use is Random Forest. It consists of Several Decision tree algorithm. The algorithm operates by constructing a multiple of decision trees at training time and outputting the mean or mode of prediction of the individual trees. Hence it has the highest accuracy.

That is how we implement our coding part. Where we use multiple models to get the highest accuracy. As suicide is an important topic to work on so we want to get the highest accuracy.

#### 5. RESULT ANALYSIS

Suicide is one the burning topic in this fast-moving world, as society does not have time to discuss about it. So, we want to track suicidal thought on the various social media platform. Because of that we want to develop a model which can detect suicidal thought by looking at their comment, post, and text etc., and we develop the model. As it is very sensible topic to work on so accuracy is most important, that is why implement various model until we get satisfactory accuracy. The highest accuracy found in Decision

Tree model. The highest accuracy was achieved was 70.95%. It means if we implement this model in real life then we will be able save more than 70.95% human resource from suicide. Along with this some other algorithm also been tasted in this research work,

those as Support Vector Machine (SVM) 57.24%, Naïve Bayes (Gaussian)54.69%, Random Forest 67.67% etc. An accuracy comparison graph has been included [Figure 1]

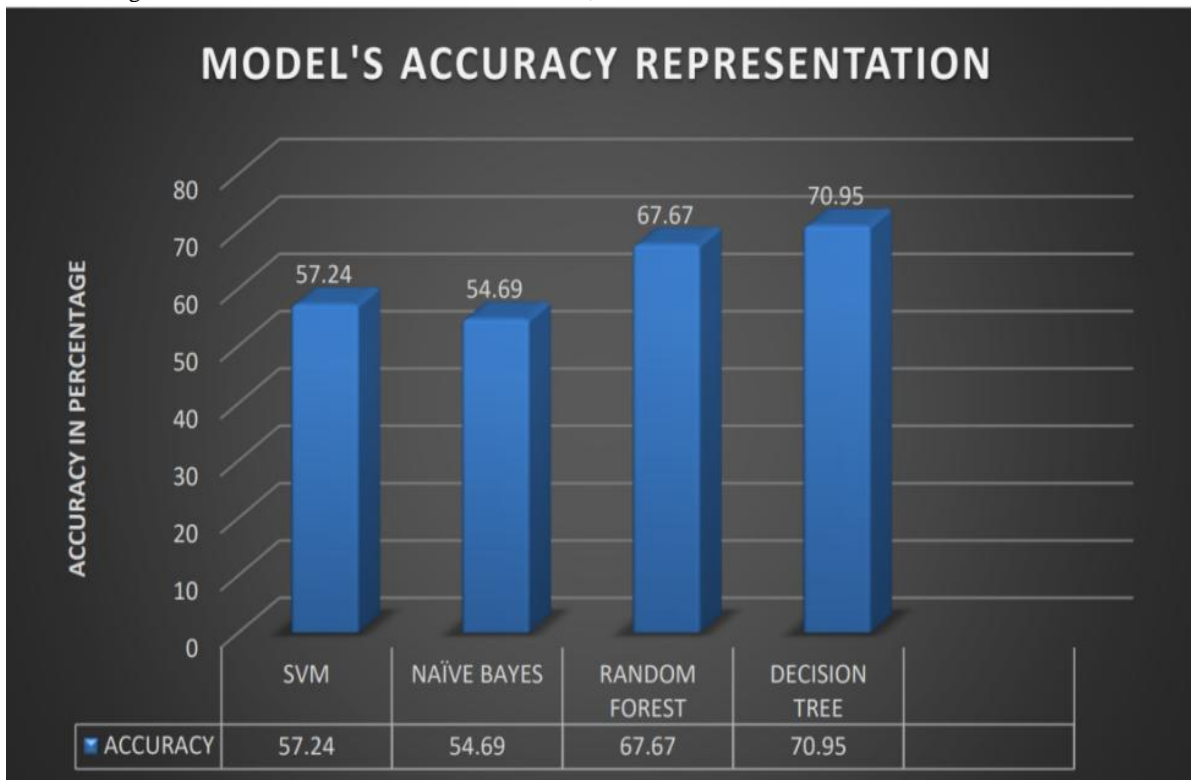


Figure 1: - Accuracy Representation of various model

**NAIVE BAYES:** Along with these above-mentioned algorithms, this research works explores some other versions of Naïve Bayes algorithm also. There are three distinct types of Naïve Bayes, Gaussian Naive Bayes, Bernoulli Naive Bayes, and

Multinomial Naive Bayes. This research work able to identify the better version of Naïve Bayes is Gaussian Naïve Bayes which records highest accuracy with 54.69% among all the versions of Naïve Bayes [Figure 2].

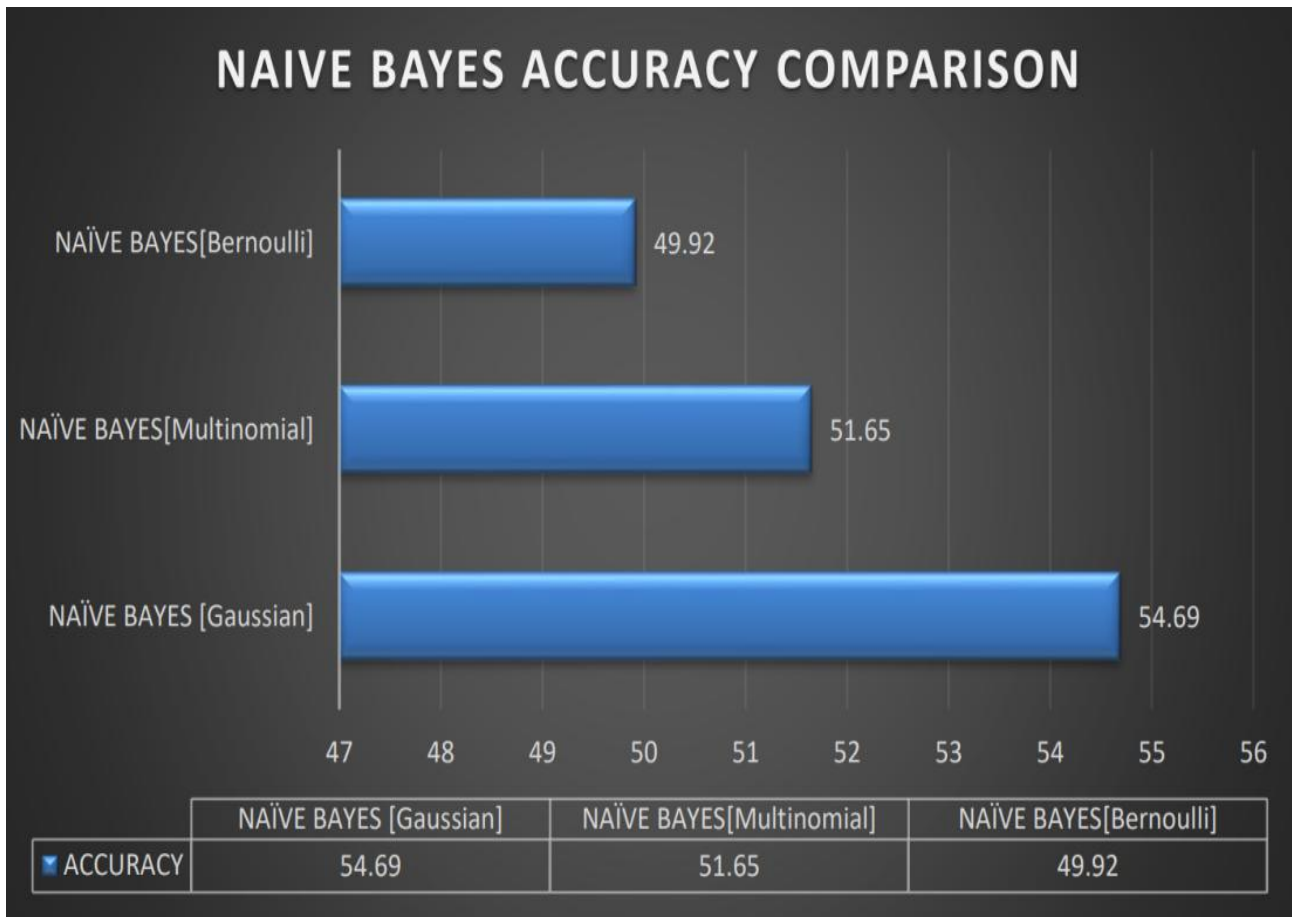


Figure 2: - Accuracy Representation of Naive Bayes

## 6. FUTURE WORK & CONCLUSION:

We implemented Decision Tree, Naïve Bayes (Multinomial) Naïve Bayes (BernoulliNB), Naïve Bayes (GaussianNB), Support Vector Machine, Random Forest, a Machine learning model, which are applied to detect suicidal tendency from different social media. Our implementation exhibited that our model performing Machine Learning models and Supervised learning approaches with a dataset containing 232074 unique values. After implementing these algorithms this research works has able to achieve a proficient level of accuracy.

There is a higher possibility to improve this accuracy score using some other machine learning model. The method proposed in this paper to improve the accuracy is Ensemble Model. Where multiple models combined to build the classifier which can lead this research works to a new achievement with a higher accuracy compare to the traditional models being used before. This research works prefers Ensemble Model as a future work to improve the accuracy.

## 7. REFERENCES

- [1] L. Sher and J.J. Peters, "The Impact of the COVID-19 Pandemic on Suicide Rates", QJM: An International Journal of Medicine, Vol. 113, No. 10, pp. 707-712, 2020.
- [2] B. Desmet and V. Hoste, "Online Suicide Prevention through Optimized Text Classification", Information Science, Vol. 439-440, pp. 61-78, 2018.
- [3] World Health Organization, "Suicide", Available: <https://www.who.int/news-room/fact-sheets/detail/suicide> Accessed at 2020.
- [4] "A Review On Analysis Of Suicidal Notes Sentiment Rich Data", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.9, Issue 5, page no.f21-f30, May-2022, Available :<http://www.jetir.org/papers/JETIR2205604.pdf>
- [5] Çelik, Özer. (2018). A Research on Machine Learning Methods and Its Applications. 10.31681/jetol.457046.
- [6] "A REVIEW PAPER ON STUDENT PERFORMANCE USING DECISION TREE ALGORITHMS", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.6, Issue 3, page no.336-341, March-2019, Available :<http://www.jetir.org/papers/JETIRAH06058.pdf>
- [7] C. E. Brodley and P. E. Utgoff, "Multivariate decision trees," Machine learning, vol. 19, no. 1, pp. 45–77, 1995.
- [8] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," Energy, vol. 32, no. 9, pp. 1761–1768, Sep. 2007, doi: 10.1016/j.energy.2006.11.010
- [9] S. Singh and P. Gupta, "Comparative study ID3, cart and C4. 5 decision tree algorithms: a survey," International Journal of Advanced Information Science and Technology (IJAIST), vol. 27, no. 27, pp. 97– 103, 2014.
- [10] L. Rokach and O. Maimon, "Top-Down Induction of Decision Trees Classifiers—A Survey," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE

- Transactions on, vol. 35, pp. 476– 487, Dec. 2005, doi: 10.1109/TSMCC.2004.843247.
- [11] T.-S. Lim, W.-Y. Loh, and Y.-S. Shih, “A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms,” *Machine learning*, vol. 40, no. 3, pp. 203– 228, 2000
- [12] Jijo, Bahzad & Mohsin Abdulazeez, Adnan. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*. 2. 20-28.
- [13] Akinsola, J E T. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology (IJCTT)*. 48. 128 - 138. 10.14445/22312803/IJCTT-V48P126.
- [14] Tian, Yingjie & Shi, Yong & Liu, Xiaohui. (2012). Recent advances on support vector machines research. *Technological and Economic Development of Economy*. 18. 10.3846/20294913.2012.661205.
- [15] Go, Alec. “Sentiment Classification using Distant Supervision.” (2009).
- [16] Pak, Alexander & Paroubek, Patrick. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of LREC*. 10.
- [17] Saif, Hassan & Alani, Harith. (2012). Semantic Sentiment Analysis of Twitter. *The Semantic Web--ISWC 2012*. 7649. 508-524. 10.1007/978-3-642-35176-1\_32.
- [18] Sohn, Sunghwan & Torii, Manabu & Li, Dingcheng & Wagholikar, Kavishwar & Wu, Stephen & Liu, Hongfang. (2012). A Hybrid Approach to Sentiment Sentence Classification in Suicide Notes. *Biomedical informatics insights*. 5. 43-50. 10.4137/BII.S8961.
- [19] Pestian, John & Matykiewicz, Pawel & Linn-Gust, Michelle & South, Brett & Uzuner, Ozlem & Wiebe, Jan & Cohen, Kevin & Hurdle, John & Brew, Chris. (2012). Sentiment Analysis of Suicide Notes: A Shared Task. *Biomedical informatics insights*. 5. 3-16. 10.4137/BII.S9042.
- [20] McCart, James & Finch, Dezon & Jarman, Jay & Hickling, Edward & Lind, Jason & Richardson, Matthew & Berndt, Donald & Luther, Stephen. (2012). Using Ensemble Models to Classify the Sentiment Expressed in Suicide Notes. *Biomedical informatics insights*. 5. 77-85. 10.4137/BII.S8931.
- [21] Neri, Federico & Aliprandi, Carlo & Capeci, Federico & Cuadros, Montse & By, Tomas. (2012). Sentiment Analysis on Social Media. 10.1109/ASONAM.2012.164.
- [22] Greaves, Felix & Ramirez-Cano, Daniel & Millett, Christopher & Darzi, Ara & Donaldson, Liam. (2013). Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online. *Journal of medical Internet research*. 15. e239. 10.2196/jmir.2721.
- [23] Suero Montero, Calkin & Munezero, Myriam & Mozgovoy, Maxim & Sutinen, Erkki. (2013). Exploiting sentiment analysis to track emotions in students' learning diaries. *ACM International Conference Proceeding Series*. 145-152. 10.1145/2526968.2526984.
- [24] Berardinelli, Nabeela & Gaber, Mohamed & Haig, Ella. (2013). SA-E: Sentiment Analysis for Education. *Frontiers in Artificial Intelligence and Applications*. 255. 10.3233/978-1-61499-264-6-353.
- [25] Sarlan, Aliza & Nadam, Chayanit & Basri, Shuib. (2014). Twitter sentiment analysis. 212-216. 10.1109/ICIMU.2014.7066632.
- [26] Birjali, Marouane & beni hssane, Abderrahim & Erritali, Mohammed. (2017). Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks. *Procedia Computer Science*. 113. 65-72. 10.1016/j.procs.2017.08.290.
- [27] Aladağ, Ahmet Emre & Muderrisoglu, Serra & Akbas, Naz & Zahmacioglu, Oguzhan & Bingol, Haluk. (2018). Detecting Suicidal Ideation on Forums and Blogs: Proof-of-Concept Study. *Journal of Medical Internet Research*. 20. e215. 10.2196/jmir.9840.
- [28] Ramyasri, V & Niharika, Ch & Maneesh, K & Ismail, Mohammed. (2019). Sentiment Analysis of Patients' Opinions in Healthcare using Lexicon-based Method. 2249-8958. 10.35940/ijeat.A2141.109119.
- [29] Mbarek, Atika & Jamoussi, Salma & Charfi, Anis & Ben Hamadou, Abdelmajid. (2019). Suicidal Profiles Detection in Twitter. 289-296. 10.5220/0008167602890296.
- [30] Qaiser, Shahzad & Yusoff, Nooraini & Ahmad, Farzana Kabir & Ali, Ramsha. (2020). Sentiment Analysis of Impact of Technology on Employment from Text on Twitter. *International Journal of Interactive Mobile Technologies (iJIM)*. 14. 88-103. 10.3991/ijim.v14i07.10600.
- [31] Glenn, Jeffrey & Nobles, Alicia & Barnes, Laura & Teachman, Bethany. (2020). Can Text Messages Identify Suicide Risk in Real Time? A Within-Subjects Pilot Examination of Temporally Sensitive Markers of Suicide Risk. *Clinical Psychological Science*. 8. 216770262090614. 10.1177/2167702620906146.
- [32] Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications Shaoxiong Ji, Shirui Pan, Member, IEEE, Xue Li, Erik Cambria, Senior Member, IEEE, Guodong Long, and Zi Huang
- [33] Sharma, Sparsh & Sharma, Surbhi. (2020). Analyzing the depression and suicidal tendencies of people affected by COVID-19's lockdown using sentiment analysis on social networking websites. *Journal of Statistics and Management Systems*. 24. 10.1080/09720510.2020.1833453.
- [34] George, Amy & Johnson, David & Carenini, Giuseppe & Eslami, Ali & Ng, Raymond & Portales-Casamar, Elodie. (2021). Applications of Aspect-based Sentiment Analysis on Psychiatric Clinical Notes to Study Suicide in Youth. *AMIA ... Annual Symposium proceedings. AMIA Symposium*. 2021. 229-237.
- [35] Source:-[https://www.researchgate.net/figure/Sentiment-classification-techniques\\_fig1\\_261875740](https://www.researchgate.net/figure/Sentiment-classification-techniques_fig1_261875740).
- [36] Source:-<https://litslink.com/blog/an-introduction-to-machine-learning-algorithms>