www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 16(4), Sep-Dec 2025, 505-514. ISSN: 2007-1558
https://doi.org/10.61467/2007.1558.2025.v16i4.582

Four Supervised Models for Identifying Suicide Indicators in Text Data

Ricardo Ulises Caballero Hidalgo, Mireya Tovar Vidal, Meliza Contreras González
[1] Faculty of Computer Science, Benemerita Universidad Autonoma de Puebla, Puebla, Mexico.
rickycab888@gmail.com, mireya.tovar@correo.buap.mx, meliza.contreras@correo.buap.mx

Abstract. The increasing prevalence of mental health disorders like depression and anxiety often leads to suicide, with individuals frequently expressing such thoughts on social media. Utilizing machine learning techniques to analyze social media texts would work towards preventing these outcomes, even though predicting suicide risk remains a challenge. In this study machine learning classifiers were developed aiming to detect suicide indicators using the Kaggle Suicide and Depression Detection dataset (Komati, N. Suicide Watch). Four models—Multinomial Naive Bayes, Gradient Boost Machine (GBM), Random Forest and Support Vector Machines—were tested, yielding promising results: Among the four models presented here SVM with a 0.95 Precision and 0.94 F1 score showed the best results.

Keywords: Suicide, Supervised Learning, Machine Learning, Naïve Bayes, Gradient Boosting Machine, Random Forest, Support Vector Machines.

Article Info
Received November 29, 2024
Accepted December 4, 2024

1 Introduction

Suicide is defined as the deliberate act of ending one's own life. Although the reasons behind this act can be varied and complex, its impact is undeniably devastating. Statistics indicate that suicide being one of the most important causes of death throughout the world, underscoring the urgent need to address this issue comprehensively.

The impact of suicide extends beyond the individual committing suicide, it affects a whole environment deeply affecting families, friends, and communities. The emotional and psychological damage is long-lasting, leaving loved ones in a state of unimaginable pain and distress.

Society as a whole also bears the consequences, facing the loss of talent, human potential, and valuable contributions. The underlying causes of suicide are multifaceted, varying considerably from person to person. Factors such as mental disorders, traumatic experiences, health issues, hard interpersonal relationships, economic difficulties, and lack of social support among others can play significant roles in the decision to end a person's life.

Addressing these factors comprehensively is vital to preventing suicide and promoting mental health. This study examines the behavior of four widely used supervised machine algorithms, applying preprocessing and data cleaning techniques on the Kaggle Suicide and Depression dataset, along with cross validation and hyperparameter search. The study also discusses class balancing techniques and their results. This paper is structured in such a way that following the introduction comes a review of related studies followed by the methodology addressed here. Next the results, discussion, conclusions, and future research proposals are stated.

Over the past decade, mental health diseases have increased worldwide, and stress is playing a vital role in suicidal ideation. Despite the rise in the suicide rate, this phenomenon has been relatively under-researched. Today, technological advances make it possibly get to a large number of people with this problem. According to a study conducted by 'We Are Social' up to 2017, out of a total of 7.53 billion people in the world, 4.54 billion have access to the Internet. As of January 2020, the number of people using social media is 3.8 billion, having grown by more than 9% since 2019 (Valeriano et al., 2020).

The easiness of social media information spread allows people to quickly gather information and convert that into a useful tool worldwide in the prevention of global health, different known platforms like Facebook, Twitter, Instagram, and YouTube are excellent communication mediums (Kelil et al., 2022).

Numerous studies exist on these classifiers for sentiment analysis, in (Spasić et al., 2012), they utilized *Naive Bayes* with *Weka* to detect signs of suicide in 600 notes, achieving an F_1 score of 53%, surpassing a majority of models that had obtained an F_1 score of 50%. They also argue in (Spasić et al., 2012) that this type of tool does not necessarily require a large amount of training data to achieve good results, which is a known problem in many cases when training classification models (Spasić et al., 2012).

The goal outlined in (Kelil et al., 2022) is to detect suicidal intentions expressed by Spanish-speaking individuals. Humans generated a dataset by annotations, and this was used later to obtain a model for future studies, which is being made available for future studies. Within the set of procedures, tests are conducted with different types of text vectorization. For constructing the classification model of all the collected tweets, the authors selected exactly 2,068 tweets to be annotated. Out of this total, 498 were labeled as tweets with suicidal tendencies and 1,570 as tweets without suicide risk. Given that the categories were imbalanced, 500 tweets were randomly selected from the 1,570 without suicidal tendencies, so that there were 500 non-suicidal tweets and 498 suicidal tweets to balance the database. Subsequently, 20% of the balanced data was allocated to testing, and the remaining 80% for training the algorithms, achieving an approximate F_1 score of 0.75 for support vector machines and logistic regression.

In (Rai et al., 2024), it is suggested that identifying individuals showing signs of depression on social media is challenging, as their mental state is inherently unpredictable. This study compares the accuracy of machine learning algorithms using 3 categories of sentiments (positive, negative, and neutral), where an accuracy of 89, 89, and 95% was achieved for each model, respectively.

In a study conducted in (Rabani et al., 2021), text was collected from the *Twitter (Now X)* and *Reddit APIs*, and 12,534 records were used from a total of 40,000 initially collected. The posts were classified into 3 categories: low risk, medium risk, and high risk. Among all algorithms, the best of all was the decision tree, with an F_1 score of around 0.96.

Another study (Ao et al., 2021) utilizes the dataset employed in this study to compare various models such as *LSTM, GRU*, *Bidirectional LSTM*, *Logistic classifier*, *SVM, XGBoost, and LGBM* for sentiment classification. By comparing the advantages and disadvantages of each model, it was found that unidirectional *LSTM* achieved an accuracy of 92.4%. No other metrics used in the mentioned study are known.

In (Desmet & Hoste, 2018), genetic algorithms are used to detect signs of suicide in notes in Dutch obtained from online networks. This study places particular emphasis on *feature selection* and *hyperparameter tuning*. The results indicated that text analysis for detecting potential suicide cases is a viable and promising technique for detecting these signs, achieving an F_1 score of 93% in relevant messages and 70% in severe messages. They also suggest that one of the main obstacles to properly training supervised machine learning models is obtaining data, and more in the case of suicide. *Support Vector Machine and Naïve Bayes* were commonly used algorithms for text classification and sentiment analysis, especially Support Vector Machine when properly optimized they can be compared to more complex algorithms and are more than enough to solve practical problems.

It is proposed in (Burgueño et al., 2023) a novel detection approach that uses deep learning, essentially, the study analyzes raw natural language information coming from a variety of sources, such as social media, and detects suicidal thoughts through text. The study mentions that the *BERT* model can achieve an optimal classification result. The dataset used comes from the Criminal Research Department of Kenya and other social media sources. In this study, various F_1 score values were obtained with an average of 0.64 for *Naive Bayes* and 0.90 for *BERT.*

The application of *Logistic Regression*, *Random Forest, Support Vector Machine, and Neural Networks*, has been explored for detecting individuals at risk of self-harm on social media like *Reddit*. Recent studies have shown that *Logistic Regression* and *Random Forest* perform better in this task, while *Neural Networks* tend to have more false positives (Henrique et al., 2023).

In the context of suicide, prediction is still a significant challenge. In (Roza et al., 2023), different algorithms and models are developed to identify mental disorders in the Brazilian population. Using clinical and sociodemographic data from a community sample in Brazil (n = 4,039), several classification models were implemented, highlighting Random Forests with an *AUC ROC* of 0.814. The presence of depressive symptoms was crucial for identifying suicide risk. These findings underline the utility of machine learning techniques in the early detection of suicide risk and the importance of preventive interventions (Murphy, 2006).

In (Sanderson et al., 2020), the performance of different predictive models was compared to identifying suicide risk. Using a dataset covering 2000 to 2016, with 3,548 suicide cases, 35,480 controls, and 4,040 predictors per person, *Recurrent Neural Network* models, *One-dimensional Convolutional Neural Networks*, and *Gradient-boosted Decision Trees* were evaluated. The results showed that these models outperformed *Logistic Regression* in terms of area under the curve, with the *XGB (Extreme Gradient Boosting)* model standing out for its superior discrimination and calibration.

In **(**Cheng et al., 2017**)**, an online survey was conducted with Chinese social media users (specifically, Weibo) analyze factors that influence suicide, including the likelihood of suicide, suicide communication on Weibo (WSC), depression, anxiety, and stress levels. They use *Support Vector Machine* to carry out the classification tasks; according to the study, *Support Vector Machine* models are well-known and highly effective in sentiment analysis in computer science.

In (Ruiz et al., 2019), the authors developed structured feature sets from longitudinal free-text posts, constructed six machine learning based models, and three of them were tested on a test dataset provided by the organizers of the CLPsych2019 event. The results demonstrated that these models identified users with tendencies to commit suicide because of posts on Reddit. The *Support Vector Machine* model showed the best average macro F_1 score for classifying four categories of suicide risk, which is attributed to its hyperparameter space and nonlinearity. Meanwhile, the *Naive Bayes* model yielded good average F_1 scores for binary group classification.

The *subreddit 'SuicideWatch'* is used by individuals with suicidal thoughts and provides important signals of suicidal behavior. Reddit presents itself as an ideal complement to traditional public health systems, due to its timeliness in exchanging ideas, versatility in expressing emotions, and compatibility for using medical terms. The effectiveness of the model is evidenced by the following results: an accuracy of 77.29% and an F_1 score of 0.77 with *Logistic Regression*, an accuracy of 74.35% and an F_1 score of 0.74 with *Naive Bayes*, an accuracy of 77.12% and an F_1 score of 0.77 with *Support Vector Machine*, and an accuracy of 77.298% and an F_1 score of 0.77 with *Random Forest* (Jain et al., 2022)

In (Saravanan et al., 2024), the study seeks to develop a reliable suicide risk assessment system through machine learning by comparing the *Random Forest* and *Support Vector Machine* algorithms to achieve higher accuracy levels. The objective is to develop a tool that effectively identifies individuals at risk of self-harm based on clinical, behavioral, and demographic data. Results show that *Random Forest* excels in suicide prevention analysis, with accuracy, sensitivity, and F_1 metrics of 0.85, highlighting the value of data-based methods and accessible interfaces to improve mental health in real practice.

In (Su et al., 2023), *Random Forests* are used aggregating *Decision Tree* ensembles using recursion and bootstrapping for feature selection. A group of Australian children were studied, and *Random Forest* outperformed the historical approach in predicting self-harm or suicide attempts. The findings in this study are in accordance with previous studies found in literature, assessing the effectiveness of predictive machine learning models.

After reviewing significant works related to detecting signs of suicide in texts, it is evident how machine learning models can perform this crucial task for society, health, and technological development in general.

In this study, four classifiers*, Naive Bayes, Gradient Boosting Machine, Random Forests and Support Vector Machine*, are used for detecting suicide from texts, the models obtained are compared with themselves and other studies.

3 Proposed Solution

The steps for implementing the proposed solution based on (Rai et al., 2024) are outlined next.

3.1 Data Collection

As mentioned above, a Dataset from the Kaggle platform is used, Table 1 shows the format of the dataset.

Table 1. Dataset format.

Unnamed	Text	Class
1	Ex wife threatening Suicide recently….	suicide
2	Am I weird I don't get affected by compliments…	non-suicide
3	Finally, 2020 is almost over…	suicide
4	I'm so lost hello, my name is ...	suicide

3.2 Data Preprocessing

The data needs to be preprocessed to prepare it in suitable formats for the subsequent model construction. Informal texts, which are common on social media, tend to require preprocessing and cleaning. At this stage of the process, the methodology used in **(**Dus & Nefedov, 2023) is applied.

The processing consists of the following stages:

1. Remove accented words with the same meaning: This also serves to reduce vector size by containing fewer words, which would otherwise increase processing time and model complexity.
2. Expand contractions: On social media, informal language often involves contractions, so this step expands contracted words.
3. Convert characters to lowercase: Words retain their meaning.
4. Remove symbols, special characters, internet addresses, unnecessary spaces: These characters are irrelevant to the model and increase its complexity.
5. Eliminate words with repeated letters, for example, words like "loveee" where the letter "e" repeats.
6. Remove stop words: These are commonly used words that hold little importance or meaning, like "an," "that," and "the." Although "no/not" are generally included in common stop word lists, they are retained here as they imply a negative tone, which can be helpful in classifying suicidal posts.
7. Spelling correction with *Symspell* (Symmetric Spelling): Used for quick spell-checking, *Symspell* can handle corrections across multi-word strings, such as separating words joined by missing spaces. It is an efficient, fast algorithm aimed at separating joined words.
8. Lemmatization: This involves converting verbs to their base forms, for example, changing "jumping" or "jumped" to "jump." This reduces semantic inconsistency, facilitates model interpretation, and other benefits.

3.3 Data Cleaning

Data cleaning involves the removal of irrelevant words, empty entries, as well as rows with a high word count. The methodology used here follows that employed in **(**Dus & Nefedov, 2023).

Data Cleaning consists of the following stages:

1. Removal of irrelevant words: As outlined in **(**Dus & Nefedov, 2023), data exploration revealed words such as "filler," which appeared 55,442 times, contributing no meaningful content to the texts and seeming more like noise due to data collection processes.
2. Removal of empty rows: Some empty rows appeared following preprocessing and were subsequently removed.
3. Elimination of atypical rows with high word counts: A word count was obtained for the preprocessed posts. To optimize model training in subsequent stages, we continued with the methodology of **(**Dus & Nefedov, 2023), where texts with more than 62 words were removed, increasing the speed and efficiency of model training.

3.4 Data Splitting

The dataset was divided into two parts: one for training the models and the other for testing, using 75% and 25% respectively. This practice is frequently used in machine learning tasks to conduct cross-validation of the models. Other studies, such as (Henrique et al., 2023; Rabani et al., 2021), have used this dataset to train classifiers, employing a similar 75%-25% split between training and testing data in both cases.

3.5 Model Training

Once the previous steps are completed, model training is carried out. At this stage, hyperparameter optimization is also conducted on those models that are sensitive to these parameters like Gradient Boosting Machine, Support Vector Machine and Random Forest. More details about the hyperparameters tuning are explained in later sections.

3.6 Cross Validation

The basic idea behind cross-validation is to expose the model to new information and assess its predictive capacity. As outlined in (Berrar, 2019), a central issue in supervised learning is the generalization ability of the resulting model, which is commonly referred to as overfitting.

Building a model that fits the given dataset perfectly is often straightforward, but the challenge lies in ensuring it generalizes well to new, unseen data. Ideally, the model's performance would be tested on fresh data from the same population as the training set. Though conducting new independent validation studies is rarely practical. Before committing resources to external validation, it's more efficient to get an estimate of the model's predictive performance upfront. This is commonly done using data resampling techniques like cross-validation (Berrar, 2019).

3.7 Discussion and Result Analysis

As the final step, conclusions are drawn and the results obtained from training and cross-validation are analyzed, the different metrics are assessed.

The results obtained from training the models are shown below and also details regarding the results obtained from the preprocessing and cleaning procedures. Each model's metrics are presented, along with the technique used to address data imbalance.

4.1 Results on The Dataset

Initially, the dataset contained 232,074 entries, which were reduced to a total of 174,968 entries in which 107,066 entries were labeled as non-suicidal and 67,902 cases labeled as suicidal. This reduction occurred due to data preprocessing, as explained in previous sections.

As observed, the dataset is imbalanced, even though not critically, it may affect the performance of the models. The approach to addressing this imbalance during the training phase is explained in later sections.

4.2 Experimental Results

Next the experimental results from training the models using the aforementioned dataset are presented. Metrics for each case are analyzed.

4.2.1 Hyperparameters and Cross validation

Since models like *GBM*, *Random Forest and SVM* are known to be sensitive to hyperparameters, these models were tuned for F_1 score optimization by using search hyperparameter techniques. The aim was placed on the mentioned metric, considering the overall prediction quality of the models. In most cases, since there are many hyperparameters that can be used to train models only those of higher importance are tested and used for experimentation. This is the part of machine learning that is more about trial and error than theoretical based knowledge.
In the case of *GBM* a set of initial hyperparameters to search from were used and then after performing *GridSearchCV,* a scikit learn functionality that allows to combine all the hyperparameters and train the models with them, the best model is obtained and so the corresponding hyperparameters. In Table 2 are shown the initial grid of hyperparameters and the those obtained from the best model. The same approach utilized with *GBM* was used with *Random Forest* and the initial and best parameters are shown in Table 3.

Table 2. GBM hyperparameters.

Parameters	Initial	Best
n_estimators	40, 50, 60, 100	60
max_depth	40, 60	60
Learning_rate	0.05, 0.1, 0.5	0.1

Parameters	Initial	Best
bootstrap	True	True
max_depth	None, 10, 20, 30	60
n_estimators	50, 100, 200	0.1
criterion	gini, entropy, log_loss	log_loss

To find the hyperparameters of SVM because of the computational cost of SVM and the size of the dataset, a different approach was utilized. In this case *HalvingRandomizedSearchCV* was used. This is an optimization technique that efficiently searches for the best hyperparameter combinations. It gradually refines the search by discarding fewer promising combinations. It is also known for its efficiency in obtaining good results with less computational effort (Kaminski, n.d).

In Table 4 the result of using this approach is shown by obtaining the best hyperparameters.

Table 4. SVM hyperparameters.

Parameters	Initial	Best
C	0.5, 1, 2	0.5
kernel	linear, rbf, poly	linear
gamma	scale, auto	auto
degree	2,3	3
class_weight	None, balanced	None

As mentioned above in every case cross validation was conducted as it is fundamental to assess for model overfitting. The technique utilized in this study for every model was *k*-fold cross validation. Using $k=5$ for all the models except for SVM in which $k=2$ was utilized mainly to reduce the computational cost of the algorithm.

4.2.2 Imbalance Data Results

The metrics for the trained models (see Table 5), *Naive Bayes, GBM, Random Forest and Support Vector Machine* on the imbalanced training data are shown below. The models were trained with 70% of the data and were evaluated against the Test Data which accounted for 30% of the imbalanced dataset.

Table 5. Model Evaluation Metric Results (Imbalanced Data).

Metrics	Naïve Bayes	GBM	Random Forest	SVM
Accuracy	0.91	0.87	0.91	0.93
Precision	0.85	0.86	0.89	0.93
Recall	0.94	0.80	0.87	0.88
F_1 *Score*	0.89	0.83	0.88	0.90

Undersampling involves randomly reducing the predominant class in the dataset to make it similar in size to the class with fewer entries, which in this case is the suicide class with 67,902 entries. Once this is done, the models are then trained. Similarly to what was mentioned above the metrics presented are those evaluated against the Test Data.

Table 6. Model Evaluation Metric Results (Undersampling).

Metrics	Naïve Bayes	GBM	Random Forest	SVM
Accuracy	0.91	0.87	0.90	0.92
Precision	0.88	0.88	0.88	0.93
Recall	0.94	0.85	0.92	0.90
F_1 Score	0.91	0.86	0.90	0.92

As observed in Table 6, the evaluation results with Naïve Bayes remained almost the same, with a 2% increase in accuracy, and a 3% increase in Precision which is the most important metric in this particular problem since it is equivalent to the model's capacity of detecting true suicide cases. Thus, undersampling had a positive effect in this regard.

In the case of GBM a general improvement of the model is noted, including a slight increase in accuracy like that achieved with Naïve Bayes. Additionally, the results for the other evaluation metrics improved, with *Recall* having the greatest increase with 5%, making the model more balanced at the time to predict cases of the two classes.

Random Forest also improved its performance after applying undersampling, being *Recall* and thus F_1 score those with the greater change, although not that significant.

Support Vector Machine also increased its performance as the recall increased in being the most significant of all with a 2% increase, this is mainly due to the fact that the *Recall* also increased in 3% with respect to the imbalanced data. This translates into a more stable less biased model.

4.2.2 Oversampling Data Results

Oversampling is the opposite of Undersampling, it involves increasing the dataset size by randomly increasing existing data from the class with less data. Table 7 shows the corresponding metrics for the trained models.

Table 7. Model Evaluation Metric Results (Oversampling).

Metrics	Naïve Bayes	GBM	Random Forest	SVM
Accuracy	0.91	0.92	0.94	0.94
Precision	0.89	0.92	0.91	0.95
Recall	0.95	0.93	0.96	0.93
F_1 Score	0.92	0.92	0.94	0.94

Similarly to the case of applying Undersampling, Naïve Bayes did not show significant changes in evaluation results, though a slight improvement is evident.

In contrast, the GBM model demonstrates an overall improvement with significant increases across all the metrics.

Random Forest also increased its performance, especially the Recall metric, contrary to what happened with SVM in which Precision was the metric with the greatest increase.

5 Discussion

In this study, four classifiers, *Multinomial Naive Bayes, GBM, Random Forests and SVM* were trained to detect indicators of suicide in texts. The selection of the algorithms was based on the widespread application of the aforementioned algorithms and the growing use of ensemble models like GBM and Random Forest for classification tasks. During the literature review, no studies were found that utilized this dataset with the data preprocessing and cleaning methodology employed in this work in conjunction with the algorithms used here.

Every model except for Naïve Bayes was trained after searching for the best hyperparameters out of a group of starting hyperparameters considered as quite significant for each algorithm. Also, k-fold cross validation was conducted on each of the algorithms with $k=5$ in all but SVM in which $k=2$ was utilized. Every algorithm was trained on imbalanced data, followed using Undersampling and Oversampling techniques, respectively.

Several metrics were evaluated for each model, including Accuracy, Precision, Recall, and F_1 score. These metrics were assessed with respect to the portion of the dataset allocated for testing, providing a detailed understanding of the performance of the classifiers in terms of their ability to correctly predict positive suicide cases, which is the objective of the study.

The metrics showed a slight improvement when undersampling and oversampling techniques were used, compared to the results obtained when the data was imbalanced, with GBM, Random Forest and SVM showing a more significant improvement than Naïve Bayes. All the models showed good results in general and according to the literature review.

Starting with accuracy, every model exhibited excellent performances being Random Forest and SVM being the best in this regard. Precision, which is the metric of greatest importance in this study, since it represents the model's ability to detect true suicidal cases, we observe that SVM outperformed every model with 0.95, while the other models were really close to one another. This small disparity indicates that Naïve Bayes tends to be more conservative in its predictions, which may result in overlooking some potential suicide indicators compared to GBM.

Examining recall, we found that SVM and GBM were quite balanced in its predictions while Naïve Bayes and Random Forest tend to exhibit disparity in its results when compared with the other two models. This metric accounts for the model's ability to predict true non-suicidal cases. Which is important in terms of model balance but not as important as Precision since we do not want to miss true suicidal cases, represented by the precision metric.

Finally, we consider the F_1 score, which provides a balanced measure between precision and recall. Naïve Bayes and GBM both achieved a score of 0.92 whereas Random Forest and SVM both achieved 0.94, which are excellent outcomes. These results indicate significant parity in the predictive capacity of these models when trained as described in this work.

Table 8 presents the metrics of related works on text classification, demonstrating that the results obtained by applying the proposed solution are satisfactory and aligned with those found in other studies. In all analyzed cases, the results of this work exceed those of others in terms of F_1 score, which is a vital metric as it reflects the classification balance between classes.

Table 8. Comparison with other studies.

Reference	Algorithms	Accuracy	Precision	Recall	*F_1 Score*
Rabani, S. (2021)	Naïve Bayes	-	-	-	0.87
Henrique, R. T. (2023)	Naïve Bayes	0.89	0.89	0.89	0.89
Ao, Z. (2021)	LSTM	0.92	Unknown	Unknown	Unknown
This study	Naïve Bayes	0.91	0.89	0.95	0.92
This study	GBM	0.92	0.92	0.93	0.92
This study	Random Forest	0.94	0.91	0.96	0.94
This study	SVM	0.94	0.95	0.93	0.94

In summary, all models exhibited similar results and can be effectively used to predict texts indicating the possibility of a person committing suicide with a high probability of prediction. SVM is the best model as its precision is the highest and also shows the best balance among the metrics.

Table 9 presents a comparison with a study that employs the same data preprocessing and cleaning methodology using deep learning models. It can be observed that less complex algorithms like Naïve Bayes and GBM demonstrate high predictive capability in classification tasks when compared to more complex models such as CNN and LSTM.

Table 9. Comparison with other studies.

Reference	Algorithms	Accuracy	Precision	Recall	F_1 *Score*
Dus, Y., & Nefedov, G. (2024).	CNN	0.92	0.91	0.90	0.91
Dus, Y., & Nefedov, G. (2024).	LSTM	0.92	0.86	0.93	0.90
This study	Naïve Bayes	0.91	0.89	0.95	0.92
This study	GBM	0.92	0.92	0.93	0.92
This study	Random Forest	0.93	0.91	0.96	0.93
This study	SVM	0.94	0.95	0.93	0.94

6 Conclusions

In this study, four classifiers, Naïve Bayes, GBM, Random Forest and SVM were evaluated for suicide detection in texts. All models demonstrated promising performance in identifying potential cases of suicide, with results highlighting their respective strengths and weaknesses.

A comparison among the models was made and the metrics were analyzed. The SVM model was the best performing model showing the best Precision and the best balance among all the metrics. The data balancing techniques employed here were positive regarding prediction model improvement. Naïve Bayes did not show significant changes concerning the use of imbalanced data and the balancing techniques employed. Additionally, the importance of hyperparameter tuning in ensemble algorithms and SVM was evident.

Even though some studies claim that the imbalance data ratio is not considerable after a 2:1 class imbalance, in this case having a ratio of imbalance between the classes of 1.6:1 showed that it affected the performance of the algorithms as they improved their outcomes when balancing techniques were applied.

An important aspect of text classification work is the necessity of a labeled dataset, as obtaining appropriate texts is often not an easy task. Privacy and consent issues regarding the parties involved or organizations managing such information frequently arise.

Future work proposes to increase the search of hyperparameters and experiment with them. Also improving further, the data preprocessing and acquisition techniques, since natural language processing techniques pose a challenge for text classification problems, as data quality directly influences results and model predictive capability. The preprocessing and data cleaning methodology employed here removed a series of entries considered irrelevant, warranting a deeper analysis of the relevance of the deleted data. Overall, the results obtained were satisfactory, with metrics according to previous studies addressing text classification and suicide detection problems.

References

Ao, Z., Lai, J., Lu, W., & Mo, H. (2021). Comparisons of LSTM and other Machine Learning Approaches on Predicting Suicidal Tendency Regarding to Social Media Posts. In *2021 International Conference on Computers and Automation (CompAuto)* (pp. 19-23). IEEE Press. doi: 10.1109/CompAuto54408.2021.00011

Berrar, D. (2019). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology. 1*, 542-545. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

Burgueño, P. L. H., de León, R. I. D., & Arenas, E. F. A. (2023). Identifying Self-Harm Risk Subjects on Social Media. https://www.researchgate.net/publication/378938143

Cheng, Q., Li, T. M., Kwok, C. L., Zhu, T., & Yip, P. S. (2017). Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study. *Journal of medical internet research*, *19*(7), e243. doi: 10.2196/jmir.7276

Desmet, B., & Hoste, V. (2018). Online suicide prevention through optimised text classification. *Information Sciences*, *439*, 61-78. https://doi.org/10.1016/j.ins.2018.02.014

Dus, Y., & Nefedov, G. (2024). An Automated Tool to Detect Suicidal Susceptibility from Social Media Posts. *arXiv e-prints*. https://doi.org/10.48550/arXiv.2310.06056

Henrique, R. T., de Souza, S. G., Recamonde-Mendoza, Lotufo, P. A., Benseñor, I. M., Cavalcante, P. I., & Russowsky, B. A. (2023). Suicide risk classification with machine learning techniques in a large Brazilian community sample. *Psychiatry research*, *325*, 115258. https://doi.org/10.1016/j.psychres.2023.115258

Jain, P., Srinivas, K. R., & Vichare, A. (2022). Depression and suicide analysis using machine learning and NLP. In *Journal of Physics: Conference Series*, 2161(1), 012034. IOP Publishing Ltd. https://dx.doi.org/10.1088/1742-6596/2161/1/012034

Kaminski T. (n.d). Medium. Retrieved March 15, 2024, from https://medium.com/@tom.kaminski01/halvingrandomsearchcv-vs-gridsearchcv-f1cd42f54b3b

Kelil, T., Jaswal, S., & Matalon, S. A. (2022). Social Media and Global Health: Promise and Pitfalls. *RadioGraphics*, *42*(4), E109-E110. https://pubs.rsna.org/doi/10.1148/rg.220038

Murphy, K. P. (2006). Naive bayes classifiers. *University of British Columbia*, *18*(60), 1-8.

Rabani, S. T., Khan, Q. R., & Khanday, A. M. U. D. (2021). Quantifying Suicidal Ideation on Social Media using Machine Learning: A Critical Review. *Iraqi Journal of Science*, 62(11), 4092-4100. https://doi.org/10.24996/ijs.2021.62.11.29

Rai, P., Prasun, K., Sharma, G., & Bhattarai, Y. (2024). Comparison of naïve bayes, logistic regression and support vector machine for predicting suicidal tendency from social media content. *Comparison of naïve bayes, logistic regression and support vector machine for predicting suicidal tendency from social media content*, *145*(1), 18-18.

Ruiz, V., Shi, L., Quan, W., Ryan, N., Biernesser, C., Brent, D., & Tsui, R. (2019). CLPsych2019 Shared Task: Predicting Suicide Risk Level from Reddit Posts on Multiple Forums. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology* (pp. 162-166). Association for Computational Linguistics. doi = 10.18653/v1/W19-3020

Sanderson, M., Bulloch, A. G., Wang, J., Williamson, T., & Patten, S. B. (2020). Predicting death by suicide using administrative health care system data: Can recurrent neural network, one-dimensional convolutional neural network, and gradient boosted trees models improve prediction performance? *Journal of affective disorders*, *264*, 107-114. https://doi.org/10.1016/j.jad.2019.12.024

Saravanan, N., Moheshkumar, G., Shaid, V. M., Purushothman, S., & Sanjai, V. G. (2024). Accurate Prediction and Detection of Suicidal Risk using Random Forest Algorithm. In *2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN)* (pp. 287-292). IEEE Press. doi: 10.1109/ICPCSN62568.2024.00053

Spasić, I., Burnap, P., Greenwood, M., & Arribas-Ayllon, M. (2012). A Naïve Bayes Approach to Classifying Topics in Suicide Notes. *Biomedical informatics insights*, *5*, 87-97. https://doi.org/10.4137/BII.S894

Su, R., John, J. R., & Lin, P. I. (2023). Machine learning-based prediction for self-harm and suicide attempts in adolescents. *Psychiatry research*, *328*, 115446. https://doi.org/10.1016/j.psychres.2023.115446

Valeriano, K., Condori-Larico, A., & Sulla-Torres, J. (2020). Detection of Suicidal Intent in Spanish Language Social Networks using Machine Learning. *International Journal of Advanced Computer Science and Applications*, *11*(4). http://dx.doi.org/10.14569/IJACSA.2020.0110489