ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 2, July 2024

ENSEMBLED MACHINE LEARNING METHODS AND FEATURE EXTRACTION APPROACHES FOR SUICIDE-RELATED SOCIAL MEDIA

Merinda Lestandy[1], Abdurrahim Abdurrahim[2], Amrul Faruq[3]
M. Irfan[4], Novendra Setyawan[5],

[1,3,4,5] Departement of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Malang
[2] Informatics Master Program, Faculty of Industrial Technology, Universitas Islam Indonesia

email: merindalestandy@umm.ac.id[1], 22917002@students.uii.ac.id[2], faruq@umm.ac.id[4]
irfan@umm.ac.id[4], novendra@umm.ac.id[5]

Abstract

Suicide is a pressing public health concern that affects both young people and adults. The widespread use of mobile devices and social networking has facilitated the gathering of data, allowing academics to assess patterns, concepts, emotions, and opinions expressed on these platforms. This study is to detect suicidal inclinations using Reddit online dataset. It allows for the identification of people who express thoughts of suicide by analyzing their postings. The method addresses and evaluates different machine learning classification models, namely linear SVC, random forest, and ensemble learning, along with feature extraction approaches such as TF-IDF, Bag of Words, and VADER. This study utilised a voting classifier in our ensemble model, where the projected class output is selected by the class with the highest probability. This approach, typically known as a "voting classifier," employs voting to forecast results. The results collected suggest that employing ensemble learning with the TF-IDF 2-grams approach yields the highest F1-score, specifically 0.9315. The efficacy of TF-IDF 2-grams can be determined to their capacity to capture a greater amount of contextual information and maintain the order of words.

Keywords : Suicide, Media Social, Feature Extraction, Ensemble Machine Learning

Received: 09-11-2023 | **Revised:** 04-03-2024 | **Accepted:** 06-03-2024
DOI: https://doi.org/10.23887/janapati.v13i2.70016

INTRODUCTION

Committed suicide is a significant issue that needs urgent attention from the public health sector. According to the data, approximately 0.7 million individuals die each year because of suicide. The risk is heightened in the age range spanning from young to middle-aged adults [1], to the point where it is regarded as the second most widespread cause of death among individuals aged 10 to 34 [2]. The risk is heightened in the age range spanning from young to middle-aged adults [1], and it is even categorized as the second most common cause of death among those aged 10 to 34 [2]. This highlights the crucial need for mental health intervention, prevention, and counseling programs that explicitly target this susceptible population.

Suicidal ideation can profoundly impact individuals globally, generating emotions such as surprise, anger, sadness, depression, or anxiety. Suicidal ideation refers to intricate contemplations about deliberately terminating one's own life. The etiology of suicide is frequently intricate and influenced by a multitude of events that can be classified into three primary categories: health issues, environmental variables, and factors associated with personal history, such as experiences of childhood maltreatment or prior suicide attempts [3].

Common risk factors associated with suicide encompass mental illness, physical ailments, drug abuse, sexual assault, bullying, relationship difficulties, and stressful living circumstances. Given the complexity of the phenomenon, it is not possible to accurately predict suicide based on any individual risk factor [4]. While depression and suicide are often associated, it is crucial to acknowledge that a depression diagnosis alone does not adequately predict suicide. The COVID-19 pandemic exacerbates suicidal ideation [5], particularly because to the social isolation resulting from stringent measures that heighten the risk of suicide [6].

Suicide-related social media analysis is a relatively new and crucial area of research due to its potential for early detection and intervention in mental health crises. However, there is still an issue in having the ability to promptly identify individuals who may be susceptible to suicide and to respond promptly in order to avoid suicidal behavior [7]. Researchers have recently evaluated community mental health issues using two main techniques [8]. One approach is grounded in a sociological standpoint that depends on clinical interactions between healthcare practitioners and patients, utilizing conventional surveys to evaluate suicidal ideation. However, a disadvantage of this method is that individuals may encounter emotions of shame or hesitation when it comes to seeking aid from psychologists or counselors. As a result, individuals may suppress their emotions and fail to communicate their suicidal intentions prior to acting upon them [7], [8]. Nevertheless, the field of suicide prevention has multiple challenges, including societal stigma, restricted availability of specialist assistance, and insufficient training for healthcare practitioners who handle individuals at risk of suicide [9]. The convergence of these factors presents a fragmented barrier, leading to significant delays in mental health assessments [9].

Recently, social media has emerged as a potent instrument for investigating matters related to mental health and overall wellness, particularly among the youth. The website offers a platform for anonymous participation in several online communities that foster discussions on themes that are often seen as socially taboo. According to statistics, almost 20% of individuals who try to end their own lives leave a suicide note behind [10]. These notes are crucial evidence, and people who disclose them should be thoroughly questioned in an unbiased manner about their beliefs. The field of computational linguistics has progressively directed its attention on mental health forums on social media. This establishes the groundwork for enhancing technical methodologies that can effectively identify and prevent suicides [11].

Earlier studies have revealed that social media platforms such as Facebook, Twitter, and Reddit have evolved into channels where individuals can openly share their suicidal ideation or seek assistance [12]–[14]. There is a growing trend indicating that young people who have thoughts of suicide are more likely to post suicide notes on well-known social media platforms such as Facebook, Twitter, and Reddit [15], [16]. While the exact correlation between online expressions and suicide risk as assessed by medical experts remains uncertain, some research have revealed a connection between sharing suicidal ideas online and psychological assessments of suicide risk [17], [18]. Currently, research is mostly concerned with utilizing machine learning and natural language processing techniques to identify suicide and other mental health problems in social networks and online forums.

Chiang developed a proactive method on social media platforms such as Facebook to detect signs of suicidal thoughts among users, allowing psychologists to help quickly [19]. In addition, a case study found that clinical health experts and psychologists raised concerns about the recent changes resulting from the advent of social media platforms such as Facebook. Users exhibiting suicide ideation were identified by analysis of their posted textual content and promptly offered help [20].

Although the process of training single-machine learning models may face challenges in determining accurate criteria for classification. The classifier may display substantial bias, which can result in overfitting, or significant variability, which can lead to underfitting. In order to avoid this, it is advisable to train many classifiers simultaneously and make use of their combined predictions. The integration of ensemble learning with bagging techniques [21] yields good outcomes.

The aim of this study is to develop a model of detection for identifying suicidal thoughts using data obtained from social media posts, namely from the Reddit site. Reddit is an expansive social media platform comprised of several subreddits, each devoted to a specific theme or subject. This study will primarily examine subreddits that are dedicated to offering mental health support, with a special focus on the subreddit "SuicideWatch". This subreddit aims to assist persons who articulate suicidal ideation and seek support from other members.

Effective risk assessment and prevention strategies for suicide-related content on social media are critical. This work explores how ensembled machine learning methods can enhance existing approaches to identify at-risk individuals and intervene in a timely manner. The study employs ensemble learning methodologies together with a diverse range of machine learning algorithms in its model. Data pre-processing encompasses the utilization of several strategies that cleanse the data before proceeding to the modeling phase. The study aims to identify pre-manifestation suicidal ideation and provide early treatments and support to persons requiring assistance.

METHOD

The first step of this study involved acquiring a dataset from Reddit, which was subsequently categorized into two distinct groups: label 1 indicating the prevalence of suicidal

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 2, July 2024

ideation and label 0 indicating its absence. Subsequently, the dataset underwent pre-processing techniques before to being fed into the machine learning model. The pre-processing stage involved removing special characters, converting text to lowercase, performing lemmatization, and eliminating stop words.

By conducting data processing to eliminate any inconsistencies or errors, the data is subsequently partitioned into distinct sets for the purposes of training and testing. The feature extraction approach utilizes a range of techniques, such as TF-IDF, TF-IDF with 2-gram, Bag of Words, VADER, and feature selection using the SVC classifier. This method employs L1-norm Support Vector Machine (SVM), which effectively decreases the number of irrelevant or redundant features to a quantity smaller than the number of samples. By employing the L1 norm, it is feasible to produce sparse solutions, leading to a substantial decrease in the number of features in a large feature set [22]. To decrease the number of dimensions in the dataset and avoid overfitting, many techniques The last phase involves the classification of data using many techniques, such as linear Support Vector Classifier (SVC), random forest, and ensemble learning. The evaluation of classification results is conducted by utilizing measures like accuracy, precision, recall, and F1 score. The objective of the model is to offer mental health practitioners enhanced assistance in formulating treatment protocols for persons who manifest suicidal thoughts. Furthermore, it serves as a crucial initial measure in the endeavor to avert undesirable outcomes. Figure 1 shows the stages of the proposed method.

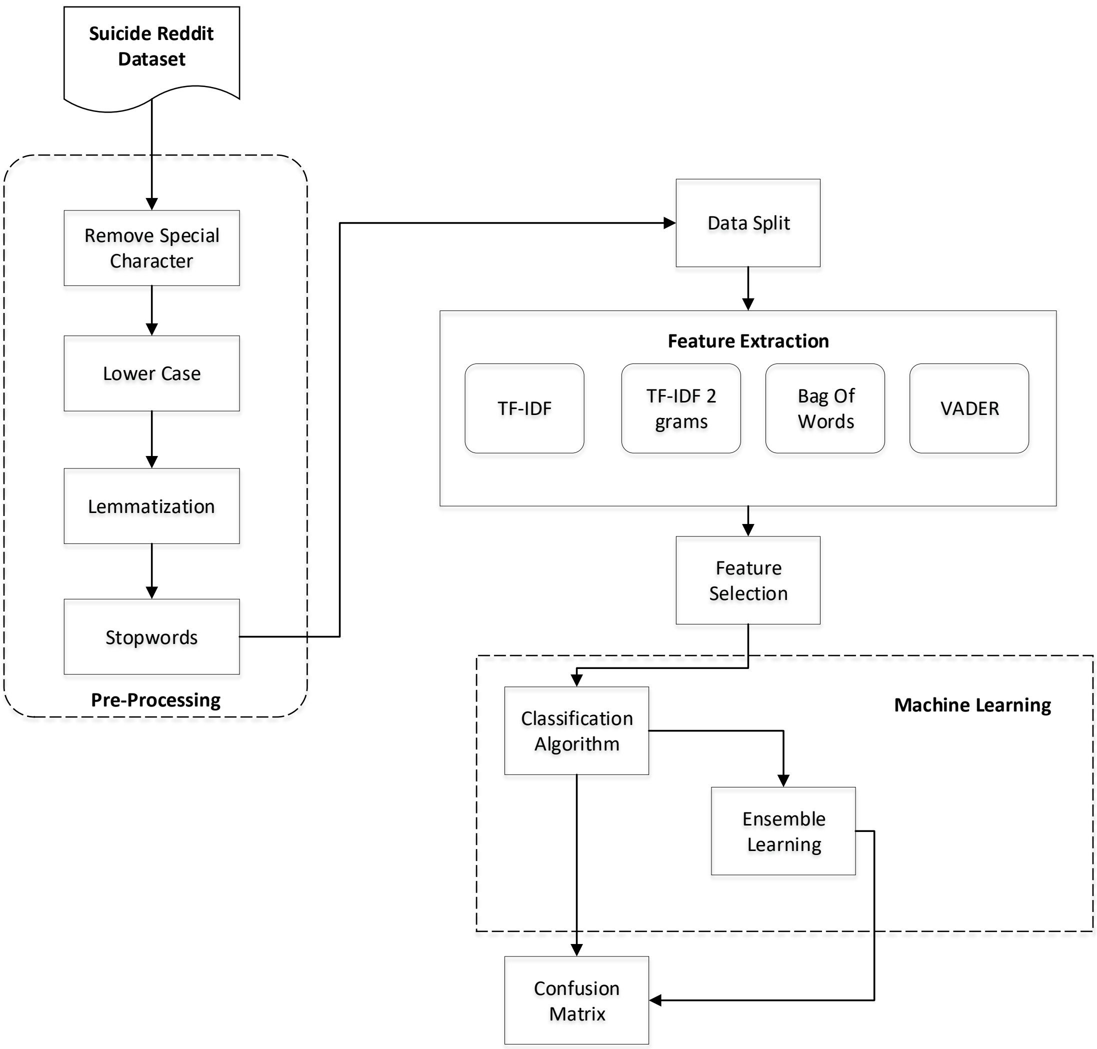

Figure 1. Proposed method using Ensemble Learning

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 2, July 2024

Dataset

The dataset originates from Kaggle and is available for access through the provided link: https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch. This dataset comprises two distinct labels, namely "suicide" and "non-suicide". The dataset was acquired by utilizing the Push shift API to gather posts from the "SuicideWatch" subreddit on the Reddit platform. Posts generated throughout the timeframe of December 16, 2008, to January 2, 2021, on the "SuicideWatch" subreddit are classified under the category of "suicide". Simultaneously, posts sourced from the "r/teenagers" subreddit were utilized to determine content that was irrelevant to suicide. Table 1 shows the example of suicide dataset obtained via Kaggle.

Table 1. Example Suicide Dataset

Suicide Post	Non-Suicide Post
i want to die i hate all of this. the covid, the everything is bad. I hate it al. I wish I was dead but i cant buy a gun i just dont think anything is worth it anymore. ive been trying to feel better by encouraging people live on this subreddit but i cant do that if i dont feel the same.	I'm wondering how some of you guys are going to celebrate Valentines Day. I I m going to celebrate it with my Lucina and Corrin plushies.
I have nothing to live for.My life is so bleak. I don't have any genuine friends, I feel like a stranger in my own family and I've scared off the only person I truly cared about. I have no future or goals to look forward to. My life is falling apart and I'm just letting it happen because I'm so, so exhausted. I just don't know if I want to be alive anymore.	Why does no one use the email function of reddit It's kinda sad, so underappreciated, edit- have y'all never used the email function, how uncultured
Is it worth it?Is all the trouble, work and anxiety really worth living for.	Guys I want friends That's it , I'm alone and don't talk to anyone dm me or anything, I'm just tired of only talking to my dad and sister, literally only my dad and sister , I like animated series but I'm flexible to anything the last series I watch was Santa Clarita's diet and the last animated series I watch was the hollow and shera (I was watching them at the same time), and I have a very extended music repertoire I can draw you anything you ask (or at least I'll try) you can text me at any hour of the day , I pretty much only do that , I have weird family anecdotes and stories I can tell you , but if you are not interested in anything I mentioned it doesn't matter I'm just here to be a friend

Pre-Processing

Even to start on word weighting and modeling, the initial phase entails pre-processing. This is the stage where data is processed by carrying out various operations, such as eliminating special characters, converting text to lowercase, lemmatizing, and eliminating stop words, all in accordance with the methods employed in this research.

a. Remove Special Character, in the first step of pre-processing, non-essential characters such as @, $, and * are removed, as they do not contribute any meaning or importance to the text. Including such characters in the data merely introduces extraneous information, thus necessitating their elimination to enhance accuracy [11]. Figure 2 shows the result of Remove Special Character step.

I'm wondering how some of you guys are going to celebrate Valentines Day. I I m going to celebrate it with my Lucina and Corrin plushies.	→	I'm wondering how some of you guys are going to celebrate Valentines Day. I I m going to celebrate it with my Lucina and Corrin plushies.

Figure 2. Remove Special Character Result

b. Lowercase, this feature serves the purpose of transforming all the letters in a word to lowercase [23]. Figure 3 shows the result of Lowercase step.

I'm wondering how some of you guys are going to celebrate Valentines Day. I I m going to celebrate it with my Lucina and Corrin plushies.	→	I'm wondering how some of you guys are going to celebrate Valentines Day. I I m going to celebrate it with my Lucina and Corrin plushies.

Figure 3. Lowercase Result

c. Lemmatization is a procedure like stemming, however it differs in approach. It employs both lexical and morphological analysis to reduce words to their base form in the dictionary. To improve the accuracy of our text, we employed NLTK's WordNet Lemmatizer in this study [24]. Figure 4 shows the result of Lemmatization step.

i'm wondering how some of you guys are going to celebrate valentines day. i i m going to celebrate it with my lucina and corrin plushies.	→	i'm wonder how some of you guy be go to celebrate valentine's day. i i be go to celebrate it with my lucina and corrin plushy.

Figure 4. Lemmatization Result

d. Stop words, in the field of text categorization, a compilation of words that contain little relevance yet regularly exist and do not considerably add to the interpretation of textual content is known as a list of stopwords. For the aim of our study, the NLTK stopword corpus is employed to exclude such words from the text, resulting in the elimination of redundant information and permitting closer scrutiny of essential data in the analysis [25]. Figure 5 shows the result of Stop Words step.

i'm wonder how some of you guy be go to celebrate valentine's day. i i be go to celebrate it with my lucina and corrin plushy.	→	wonder guys go celebrate valentine's day. celebrate lucina corrin plushy.

Figure 5. Stop Words Result

TF-IDF

The TF-IDF approach is a widely employed methodology in natural language processing that assesses the importance of a keyword in a set of documents. This approach quantifies the occurrence rate of a certain keyword in a single document compared to its occurrence rate over the full collection of documents [26]. The TF-IDF method distinguishes itself from the Bag-of-Words methodology by recognizing that specific keywords may possess higher importance than others, based on their frequency and rarity over the full collection of documents. The TF-IDF value of a specific keyword in a document is obtained by multiplying the frequency of the keyword's occurrence (TF) by its inverse document frequency (IDF) value. This methodology offers a more sophisticated technique for examining textual data. The inverse document frequency (IDF) is calculated by dividing the total number of documents in a collection by the number of documents that include a certain keyword.

Subsequently, the logarithm of the quotient is computed, employing a technique that ensures that frequently occurring keywords in multiple papers are assigned a lesser significance compared to those occurring in only a few.

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 2, July 2024

Therefore, keywords that are found in only a few papers suggest that they contain more significant information and are assigned a greater importance.

N-Gram

When implementing supervised machine learning algorithms, we utilize the feature extraction approach known as N-Grams. An N-gram is a consecutive sequence of tokens that can be found in written texts such as tweets, speeches, or blogs. This terminology refers to a unigram, bigram, and trigram, which correspond to values of n equal to 1, 2, and 3, respectively. For example, by choosing a bigram of N=2 from the sentence "Wonder guys go celebrate Valentine's Day, celebrate Lucina Corrin plushy," we can create word pairs like "wonder guys," "guys go", "go celebrate", "celebrate valentine's", "valentine's day", "day celebrate", "celebrate lucina", "lucina corrin" and "corrin plushy" In our proposed technique framework, we utilized a value of N = 2, specifically known as a Bigram [27].

Bag Of Words

The objective of this method is to eliminate characteristics from textual expressions to make them suitable for use in modelling, particularly in machine learning models. During this procedure, any details on the order or arrangement of words in the document are disregarded, leading to a representation of words as a "bag". The model focuses solely on the presence of common terms in the paper, disregarding their exact placement. The objective is to transform each page into a vector that may be utilized in machine learning models or extracted from them. An uncomplicated method for scoring involves assigning Boolean values to indicate the existence or nonexistence of words, with 0 representing nonexistence and 1 representing existence. Within the framework of this study, this feature extraction technique is integrated with a multitude of different techniques and algorithms [28].

VADER

VADER, also known as the Valence Aware Dictionary and Sentiment Reasoner, is a sentiment analysis tool that employs precise rules and lexicons. The main purpose of this system is to detect and classify emotions and sentiments conveyed in social media posts. The VADER technique integrates a sentiment lexicon with lexical elements, such as words, that are commonly classified according to their semantic orientation as either positive or negative.

The VADER analysis generates four sentiment measures by analyzing the vocabulary employed in the text. Three metrics measure the extent of positive, neutral, and negative attitudes, reflecting the proportion of the text that belongs to each of these categories. The fourth metric is the composite score, which is obtained by standardizing all lexical values and producing a value that falls within the range of -1 to 1 [29].

Support Vector Classifier

The Support Vector Classifier (SVC) is a specialized algorithm that aims to identify the ideal linear classification model that best represents the distribution pattern of the training samples. This is achieved by selecting two data points from distinct categories that have the minimum distance between them, known as a "support vector".

The Support Vector Machine (SVM) calculates the decision function by maximizing the distance (margin) between the decision boundaries in a feature space with many dimensions. By employing this classification technique, the occurrence of misclassification in the training data is reduced, leading to an enhancement in the ability to generalize. SVM and other approaches exhibit substantial disparities in their classification capabilities, particularly when confronted with constrained input data. Support Vector Machines (SVM) is a robust methodology employed in the tasks of data classification and regression analysis. A notable benefit of SVMs is their ability to obtain a subset of support vectors during the learning phase, typically representing only a fraction of the original data set. This set of support vectors, which represents a specific classification problem, is composed of a small data set [30].

Consequently, the objective of prediction may only yield a limited quantity of highly effective training examples derived from potentially vast or intricate data. This classifier is like a Support Vector Classifier (SVC) that utilizes the 'linear' kernel option. Nevertheless, it is implemented utilizing the linear library rather than the SVM library. This model offers enhanced flexibility in selecting penalties and consistently excels when handling datasets with a substantial number of samples [31].

Random Forest

The decision tree technique is a key component in ensemble learning methods for constructing prediction models [32]. The objective of this ensemble technique is to improve the dependability of predictions by combining the results of several baseline estimators that employ the decision tree algorithm. An example of this approach is the Random Forest process, which generates many decision trees known as "forests". When fresh data is being classified, each individual tree generates a prediction regarding the category. The final prediction is then chosen by the majority vote of all the trees. Typically, augmenting the quantity of trees in a random forest result in an elevated level of prediction precision.

The random forest technique involves creating numerous "simple" decision trees during the training phase and determining the majority (mode) among them during the classification phase. This selection technique addresses the issue of decision trees' tendency to overfit on training data, which is undesirable [33]. During the training phase, the random forest algorithm employs a widely used technique called bagging to each individual tree in the ensemble. Bagging is an iterative process that involves randomly choosing samples with replacements from the training set and constructing a decision tree based on those samples. Every tree undergoes growth without the need for a pruning procedure. The ensemble's tree count is a tunable parameter that can be automatically learned by the previously discussed out-of-bag mistake [34][35]. The Random Forest algorithm has experienced significant popularity in recent years due to its outstanding performance in various domains, such as bioinformatics and biological computing. The use of this technology can be applied to tasks such as identifying hate speech [36] and analyzing the characteristics of authors [37].

Ensemble Learning

Ensemble learning is a methodology that leverages the concept of combining predictions from several methods to achieve greater overall accuracy compared to the predictions made by individual members of the group. As a result, numerous ensemble models can be combined using techniques such as averaging, voting, or probabilistic approaches to reach conclusions. Ensemble learning methods exhibit adaptability and usefulness in many learning models and issues [38].

Ensemble learning is a technique employed to train numerous learning machines and aggregate their outputs, considering them as a committee of decision-makers. The concept is that the committee's decisions, through the appropriate amalgamation of individual predictions, should exhibit superior overall accuracy, on average, compared to each individual committee member. Empirical and theoretical research has consistently demonstrated that ensemble models frequently outperform individual models in terms of accuracy. Ensemble components have the capacity to forecast values in diverse formats, including real numbers, class labels, posterior probabilities, ranks, groupings, or other things. Consequently, their decisions can be amalgamated utilizing various techniques, such as averaging, voting, and probabilistic procedures. Ensemble learning approaches are versatile and can be employed across various learning models and problems [38].

This study utilised a voting classifier in our ensemble model, where the projected class output is selected by the class with the highest probability. This approach, typically known as a "voting classifier," employs voting to forecast results. Soft voting is a technique that amalgamates forecasts from many models to improve the overall performance of the ensemble. This strategy is employed when the underlying classifier in the ensemble is probabilistic, such as linear SVC or random forest. This suggests that a base classifier having a greater likelihood for a specific class would be assigned a greater voting weight for that class compared to a classifier with a lower likelihood [39].

Confusion Matrix

This study will employ regular metrics to assess the applied models, encompassing accuracy, precision, recall, and F1 score. For binary classification, the metrics will be calculated by the following steps (1) to (4). The term "accuracy" refers to the extent to which the expected value aligns with the actual value, as measured by equation (1). True Positives (TP) are examples where positive projections are correct, while False Positives (FP) are examples where optimistic predictions are incorrect. False Negatives (FN) refer to incorrect negative predictions, while True Negatives (TN) represent correct negative predictions.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+FP} \qquad (1)$$

Precision is a measure of quantity that evaluates the level of correctness in data by considering both the quality of the information and its capacity to make accurate predictions. Equation (2) is employed to express precision.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

The recall metric measures the model's ability to identify the right samples of one specific class. The computation can be executed using equation (3)

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

The F1-Score is a statistic that combines two essential elements of model evaluation: precision and recall. Precision is a measure that assesses the level of accuracy demonstrated by a model in making predictions about data. Recall is a statistic used to assess how well a model can accurately identify examples that belong to a specific class. The F1-Score provides a complete evaluation of the model's predicted performance by combining numerous metrics. The F1-Score is calculated using equation (4).

$$F1 - Score = \frac{2*(precision*recall)}{precision+recall} \quad (4)$$

RESULT AND DISCUSSION

The data collected from the suicide dataset was preprocessed to eliminate redundancy and noise using preset approaches [40], after extracting information from Reddit posts related to suicide and non-suicide. The research involved several processes, including the comparison of various feature extraction methods such as TF-IDF, TF-IDF 2-grams, Bag of Words, and VADER. The objective of feature selection is to identify a subset of characteristics that have several advantages, including reducing computational burden, preventing overfitting, and enhancing the accuracy of classification models [41].

The machine learning models used include linear SVC, random forest, and ensemble learning, which combines many classifiers such as random forest, logistic regression, stochastic gradient descent, support vector machine, and the bagging approach. The linear SVC classifier was built with a penalty parameter C of 1000, l1 regularization, and a training approach that allowed for a maximum of 500 iterations. The Bag of Words technique yielded the highest level of effectiveness, with accuracy results of 0.8690 and F1-scores of 0.8696. Table 2 shows the results of confusion matrix of Linear SVC.

Table 2. Confusion matrix results Linear SVC

Method	Accuracy	Precision	Recall	F1-Score
TF-IDF	0.8623	0.8688	0.8552	0.8619
TF-IDF 2 grams	0.8604	0.8729	0.8453	0.8589
Bag Of Words	0.8690	0.8698	0.8695	0.8696
VADER	0.6901	0.7007	0.6954	0.6915

The confusion matrix results generated by the random forest classifier are shown in Table 3, in the random forest technique, the max_depth parameter is set to 10 to restrict the maximum depth of each tree in the ensemble. This is done to avoid overfitting during the training phase. The Random Forest classifier's accuracy findings indicate that the TF-IDF 2 grams feature extraction method surpassed the other methods, earning a value of 0.8345. Additionally, the F1-score reached 0.8462.

Table 3. Confusion matrix results Random Forest

Method	Accuracy	Precision	Recall	F1-Score
TF-IDF	0.8341	0.7925	0.9075	0.8461
TF-IDF 2 grams	0.8345	0.7938	0.9061	0.8462
Bag Of Words	0.8255	0.7840	0.9010	0.8385
VADER	0.7808	0.7848	0.7770	0.7809

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 2, July 2024

The confusion matrix outcomes of the ensemble learning classifier are displayed in Table 4, the matrix comprises various machine learning algorithms, including logistic regression, random forest, SVC, and SGD. The logistic regression is optimized using the lbfgs (limited-memory broyden-fletcher-goldfarb-shanno) method, while the random forest methodology employs decision trees with n_estimators = 100. The support vector classifier uses a kernel with a gamma value set to "scale", while stochastic gradient descent utilizes settings such as alpha=0.0001, max_iter=50, loss='log', and regularization penalty="elasticnet". In addition, the process was run simultaneously with n_jobs=-1. Our findings indicate that the utilization of ensemble learning in conjunction with TF-IDF 2 grams feature extraction yields superior outcomes compared to alternative approaches, achieving an F1-score of 0.9315 and an accuracy of 0.9306. In contrast, the TF-IDF technique yields an accuracy of 0.9303 and an F1-score of 0.9311.

Table 4. Confusion Matrix Results Ensemble Learning

Method	Accuracy	Precision	Recall	F1-Score
TF-IDF	0.9303	0.9242	0.9380	0.9311
TF-IDF 2 grams	0.9306	0.9246	0.9385	0.9315
Bag Of Words	0.9217	0.9074	0.9403	0.9235
VADER	0.7746	0.7704	0.7856	0.7779

The machine learning model achieved an accuracy of 0.9303 and an F1-score of 0.9315 through the utilization of ensemble learning using the TF-IDF 2-gram approach. On the other hand, the TF-IDF technique produced similar results, with an accuracy of 0.9303 and a virtually equal F1-score. The effectiveness of ensemble learning with TF-IDF 2 grams originates from its ability to effectively handle common words by assigning them low or negligible weight. The beneficial effects of TF-IDF 2-grams can be ascribed to their capacity to collect greater contextual information and maintain word order, particularly in tasks related to natural language processing [42][43]. TF-IDF 2-grams have proven useful in identifying distinctive and uncommon phrases that are crucial for understanding the text. This makes it a favored option for different classification tasks [44]. In contrast, the Bag of Words (BOW) method usually accords equal significance to words, as BOW merely converts words into binary digits. This finding aligns with prior studies that have demonstrated the effectiveness of ensemble learning in enhancing accuracy [45], [46], as well as the good results achieved when utilizing the TF-IDF approach as a feature extraction technique [28], [29].

CONCLUSION

The issue of suicide is serious particularly among those between the ages of 10 and 34, resulting in nearly 0.7 million suicides annually, predominantly affecting the youth and those in their middle years. The etiology of suicide encompasses many elements, including physical well-being, surroundings, and individual background, such as a history of childhood maltreatment or prior suicide endeavors. Nevertheless, it is imperative to employ efficient techniques for early identification of persons harboring suicidal ideation, to promptly implement preventive measures. Thus, the objective of this project is to create a proficient machine learning model and feature extraction techniques for identifying potential suicides. The findings indicated that the linear Support Vector Classification (SVC) machine learning model Yielded the most favorable outcomes when a certain approach was employed. It was followed by the random forest model using a different approach, and the bagging method employing yet another approach. Nevertheless, the most optimal outcomes were achieved by employing the TF-IDF 2-gram technique in conjunction with the ensemble learning model. The accuracy of this model was 0.93, and it achieved a high F1-Score. The enhanced performance of ensemble learning is attributed to its capability to integrate many algorithms. Additionally, the TF-IDF technique effectively handles common terms by assigning them low or zero weight. On the contrary, the Bag of Words (BOW) approach often assigns equal importance to words, as it solely transforms words into binary digits.

This research contributes to advancing the state-of-the-art in suicide prevention research and facilitates the development of more effective tools and interventions for identifying and supporting individuals at risk. This work also provides an overview of ensembled machine learning methods suitable for analyzing social media data, using voting-ensemble learning. Furthermore, it also discusses how these methods can be adapted and optimized for suicide-related content classification, risk assessment, and prediction tasks.

ACKNOWLEDGMENT

The research team would like to express their gratitude and appreciation to the Faculty of Engineering, Universitas Muhammadiyah Malang, for their support in the implementation of this work through the Engineering Study and Engineering Center scheme, Pusat Kajian dan Rekayasa Teknik-UMM (PUSKAREKATEK 2023).

REFERENCES

[1] World Health Organization, *Suicide worldwide in 2019: Global Health Estimates*. 2019. [Online]. Available: https://apps.who.int/iris/rest/bitstreams/1350975/retrieve

[2] A. Z. Ivey-Stephenson *et al.*, "Suicidal Ideation and Behaviors Among High School Students - Youth Risk Behavior Survey, United States, 2019.," *MMWR Suppl.*, vol. 69, no. 1, pp. 47–55, Aug. 2020, doi: 10.15585/mmwr.su6901a6.

[3] M. A. Reger, I. H. Stanley, and T. E. Joiner, "Suicide Mortality and Coronavirus Disease 2019—A Perfect Storm?," *JAMA Psychiatry*, vol. 77, no. 11, pp. 1093–1094, Nov. 2020, doi: 10.1001/jamapsychiatry.2020.1060.

[4] J. C. Franklin *et al.*, "Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research.," *Psychol. Bull.*, vol. 143, no. 2, pp. 187–232, Feb. 2017, doi: 10.1037/bul0000084.

[5] G. Castillo-Sánchez, G. Marques, E. Dorronzoro, O. Rivera-Romero, M. Franco-Martín, and I. De la Torre-Díez, "Suicide Risk Assessment Using Machine Learning and Social Networks: a Scoping Review.," *J. Med. Syst.*, vol. 44, no. 12, p. 205, Nov. 2020, doi: 10.1007/s10916-020-01669-5.

[6] E. Yeskuatov, S.-L. Chua, and L. K. Foo, "Leveraging Reddit for Suicidal Ideation Detection: A Review of Machine Learning and Natural Language Processing Techniques.," *Int. J. Environ. Res. Public Health*, vol. 19, no. 16, Aug. 2022, doi: 10.3390/ijerph191610347.

[7] A. N. Weber, M. Michail, A. Thompson, and J. G. Fiedorowicz, "Psychiatric Emergencies: Assessing and Managing Suicidal Ideation," *Med. Clin. North Am.*, vol. 101, no. 3, pp. 553–571, 2017, doi: https://doi.org/10.1016/j.mcna.2016.12.006.

[8] S. T. Rabani, Q. R. Khan, and A. M. U. D. Khanday, "Quantifying Suicidal Ideation on Social Media using Machine Learning: A Critical Review," *Iraqi J. Sci.*, vol. 62, no. 11, pp. 4092–4100, 2021, doi: 10.24996/ijs.2021.62.11.29.

[9] M. Gaur *et al.*, "Characterization of time-variant and timeinvariant assessment of suicidality on Reddit using C-SSRS," *PLoS One*, vol. 16, no. 5 May 2021, pp. 1–21, 2021, doi: 10.1371/journal.pone.0250448.

[10] T. M. DeJong, J. C. Overholser, and C. A. Stockmeier, "Apples to oranges?: A direct comparison between suicide attempters and suicide completers," *J. Affect. Disord.*, vol. 124, no. 1, pp. 90–97, 2010, doi: https://doi.org/10.1016/j.jad.2009.10.020.

[11] M. Marks, "Artificial Intelligence-based Suicide Prediction," *Yale J. Heal. Policy, Law Ethics*, no. 1003774, p. 4, 2019, [Online]. Available: https://ssrn.com/abstract=3324874%0Ahttps://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8698029

[12] G. B. Colombo, P. Burnap, A. Hodorog, and J. Scourfield, "Analysing the connectivity and communication of suicidal users on twitter," *Comput. Commun.*, vol. 73, pp. 291–300, 2016, doi: https://doi.org/10.1016/j.comcom.2015.07.018.

[13] K. Daine, K. Hawton, V. Singaravelu, A. Stewart, S. Simkin, and P. Montgomery, "The Power of the Web: A Systematic Review of Studies of the Influence of the Internet on Self-Harm and Suicide in Young People," *PLoS One*, vol. 8, no. 10, p. e77555, Oct. 2013, [Online]. Available: https://doi.org/10.1371/journal.pone.0077555

[14] R. A. Fahey, J. Boo, and M. Ueda, "Covariance in diurnal patterns of suicide-related expressions on Twitter and recorded suicide deaths.," *Soc. Sci. Med.*, vol. 253, p. 112960, May 2020, doi: 10.1016/j.socscimed.2020.112960.

[15] S. Ji, S. Pan, X. Li, E. Cambria, G. Long, and Z. Huang, "Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 1, pp. 214–226, 2021, doi: 10.1109/TCSS.2020.3021467.

[16] B. Desmet and V. Hoste, "Emotion detection in suicide notes," *Expert Syst.*

Appl., vol. 40, no. 16, pp. 6351–6358, 2013, doi: https://doi.org/10.1016/j.eswa.2013.05.050.

[17] J. Jashinsky *et al.*, “Tracking Suicide Risk Factors Through Twitter in the US,” *Crisis*, vol. 35, no. 1, pp. 51–59, Jan. 2014, doi: 10.1027/0227-5910/a000234.

[18] G. Coppersmith, R. Leary, P. Crutchley, and A. Fine, “Natural Language Processing of Social Media as Screening for Suicide Risk,” *Biomed. Inform. Insights*, vol. 10, p. 1178222618792860, Jan. 2018, doi: 10.1177/1178222618792860.

[19] W.-C. Chiang, P.-H. Cheng, M.-J. Su, H.-S. Chen, S.-W. Wu, and J.-K. Lin, “Socio-health with personal mental health records: Suicidal-tendency observation system on Facebook for Taiwanese adolescents and young adults,” in *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services*, 2011, pp. 46–51. doi: 10.1109/HEALTH.2011.6026784.

[20] K. Lehavot, D. Ben-Zeev, and R. E. Neville, “Ethical Considerations and Social Media: A Case of Suicidal Postings on Facebook,” *J. Dual Diagn.*, vol. 8, no. 4, pp. 341–346, Nov. 2012, doi: 10.1080/15504263.2012.718928.

[21] L. Breiman, “Bagging predictors,” *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1007/BF00058655.

[22] M. Moon and K. Nakai, “Stable feature selection based on the ensemble L 1-norm support vector machine for biomarker discovery,” *BMC Genomics*, vol. 17, no. Suppl 13, 2016, doi: 10.1186/s12864-016-3320-z.

[23] A. K. Uysal and S. Gunal, “The impact of preprocessing on text classification,” *Inf. Process. Manag.*, vol. 50, no. 1, pp. 104–112, 2014, doi: 10.1016/j.ipm.2013.08.006.

[24] A. Jakhotiya *et al.*, “Text Pre-Processing Techniques in Natural Language Processing: A Review,” *Int. Res. J. Eng. Technol.*, pp. 878–880, 2022, [Online]. Available: www.irjet.net

[25] S. Sarica and J. Luo, “Stopwords in technical language processing,” *PLoS One*, vol. 16, no. 8 August, pp. 1–13, 2021, doi: 10.1371/journal.pone.0254937.

[26] J. E. Ramos, “Using TF-IDF to Determine Word Relevance in Document Queries,” 2003. [Online]. Available: https://api.semanticscholar.org/CorpusID:14638345

[27] M. Das, S. Kamalanathan, and P. Alphonse, “A Comparative Study on TF-IDF feature Weighting Method and its Analysis using Unstructured Dataset,” vol. 5571, pp. 0–2, 2021.

[28] A. Stephen, T. Lubem, and I. Adom, “Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets,” *Int. J. Inf. Technol.*, vol. 14, Sep. 2022, doi: 10.1007/s41870-022-01096-4.

[29] M. Chiny, M. Chihab, Y. Chihab, and O. Bencharef, “LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 7, pp. 265–275, 2021, doi: 10.14569/IJACSA.2021.0120730.

[30] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” *Neurocomputing*, vol. 408, pp. 189–215, 2020, doi: https://doi.org/10.1016/j.neucom.2019.10.118.

[31] L. R. Krosuri, R. Satish, E. Learning, S. Lin, S. Fan, and J. Yao, “Social Sentiment Analysis Using Classifiers and Ensemble Learning Social Sentiment Analysis Using Classifiers and Ensemble Learning,” 2019, doi: 10.1088/1742-6596/1237/2/022193.

[32] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[33] B. Zadrozny and C. Elkan, “Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers,” *Icml*, pp. 1–8, 2001, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.29.3039&rep=rep1&type=pdf

[34] P. Cunningham and S. Delany, “k-Nearest neighbour classifiers,” *Mult Classif Syst*, vol. 54, Apr. 2007, doi: 10.1145/3459665.

[35] P. D. Caie, N. Dimitriou, and O. Arandjelović, “Precision medicine in digital pathology via image analysis and machine learning,” *Artif. Intell. Deep Learn. Pathol.*, pp. 149–173, 2020, doi: 10.1016/B978-0-323-67538-3.00008-7.

[36] I. Alfina, R. Mulia, M. I. Fanany, and Y. Ekanata, “Hate speech detection in the Indonesian language: A dataset and preliminary study,” in *2017 International Conference on Advanced Computer*

Science and Information Systems (ICACSIS), 2017, pp. 233–238. doi: 10.1109/ICACSIS.2017.8355039.

[37] A. P. Garibay, A. T. Camacho-González, R. A. Fierro-Villaneda, I. Hernandez-Farias, D. Buscaldi, and I. V. M. Ruiz, "A Random Forest Approach for Authorship Profiling," 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:18882361

[38] G. Brown, "Ensemble Learning BT - Encyclopedia of Machine Learning," C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, pp. 312–320. doi: 10.1007/978-0-387-30164-8_252.

[39] A. Srinivas and J. P. Mosiganti, "A brain stroke detection model using soft voting based ensemble machine learning classifier," *Meas. Sensors*, vol. 29, p. 100871, 2023, doi: https://doi.org/10.1016/j.measen.2023.100871.

[40] N. Anand, D. Goyal, and T. Kumar, "Analyzing and Preprocessing the Twitter Data for Opinion Mining BT - Proceedings of International Conference on Recent Advancement on Computer and Communication," 2018, pp. 213–221.

[41] A. Yousefpour, R. Ibrahim, and H. N. A. Hamed, "Ordinal-based and frequency-based integration of feature selection methods for sentiment analysis," *Expert Syst. Appl.*, vol. 75, pp. 80–93, 2017, doi: 10.1016/j.eswa.2017.01.009.

[42] D. Cahyani and I. Patasik, "Performance comparison of TF-IDF and Word2Vec models for emotion text classification," *Bull. Electr. Eng. Informatics*, vol. 10, pp. 2780–2788, Oct. 2021, doi: 10.11591/eei.v10i5.3157.

[43] R. Roshan, I. A. Bhacho, and S. Zai, "Comparative Analysis of TF–IDF and Hashing Vectorizer for Fake News Detection in Sindhi: A Machine Learning and Deep Learning Approach †," *Eng. Proc.*, vol. 46, no. 1, 2023, doi: 10.3390/engproc2023046005.

[44] B. Jlifi, C. Sakrani, and C. Duvallet, "Towards a soft three-level voting model (Soft T-LVM) for fake news detection," *J. Intell. Inf. Syst.*, vol. 61, no. 1, pp. 249–269, 2023, doi: 10.1007/s10844-022-00769-7.

[45] B. Siswoyo, Z. A. Abas, A. Naim, C. Pee, R. Komalasari, and N. Suyatna, "Ensemble machine learning algorithm optimization of bankruptcy prediction of bank," vol. 11, no. 2, pp. 679–686, 2022, doi: 10.11591/ijai.v11.i2.pp679-686.

[46] M. Fayaz, A. Khan, J. U. Rahman, A. Alharbi, M. I. Uddin, and B. Alouffi, "Ensemble Machine Learning Model for Classification of Spam Product Reviews," vol. 2020, 2020.