Comparison of naïve bayes, logistic regression and support vector machine for predicting suicidal tendency from social media content

Prajwal Rai, Kumar Prasun, Gajendra Sharma, Yubraj Bhattarai

developer.prajwal007@gmail.com

Kantipur City College, Kathmandu, 44600 , Nepal

ABSTRACT

The objective of this study is to employ Machine Learning methodologies in order to forecast the chances of depression and suicide among individuals in Nepal by analyzing their social media engagement. The dataset consisting of 2200 entries was subjected to analysis using the Naïve Bayes, Logistic Regression, and Support Vector Machine techniques. The accuracy of the Support Vector Machine was found to be 95.45%. The timely identification of suicidal and depressive incidents had a crucial role in addressing depression and diminishing suicide rates. This study assesses the sensitivity and accuracy of various Machine Learning algorithms in the context of early and late detection.

Keywords: Depression; Social media; Detection; Suicide; Machine learning

1. INTRODUCTION

A) Background

Identifying persons who do not show signs of depression on social media is difficult since their mental condition is inherently unpredictable. Depression is a significant public health issue, frequently associated with disability and suicide. The symptoms encompass diminished energy levels, alterations in appetite, heightened anxiety, impaired concentration, and challenges in making decisions. Although there have been improvements in preventing, diagnosing, and treating depression, the occurrence of depression is still increasing, especially as a result of the COVID-19 pandemic, which has worsened anxiety and depressive disorders worldwide. The worldwide incidence of depressed symptoms witnessed a notable rise, escalating from around 193 million individuals to 246 million, indicating a growth rate of approximately 28%. [1].

The prevalence of anxiety disorders on a global scale has shown a notable rise of 25%, primarily attributed to the widespread utilization of social media platforms such as Facebook, Twitter, and Instagram in emerging nations. These platforms facilitate the expression of individuals' views, emotions, and experiences, hence promoting the establishment of relationships and facilitating communication. Nevertheless, this phenomenon has the potential to give rise to mental health disorders and addiction, hence engendering emotional susceptibility. Machine learning algorithms have the capability to analyze typical mental states such as happiness, sorrow, anger, and anxiety-depression by analyzing shared messages and photographs. Assessing depression levels and elevated indicators can help evaluate the risk of suicide in persons with cardiovascular illness.

B) STATEMENT OF PROBLEM

Approximately 60% of individuals who die by suicide are believed to have had a mood condition, such as depression or bipolar disorder [2]. The findings of the study indicate a heightened prevalence of suicide among adolescents and young adults, with poisoning emerging as the predominant method of self-inflicted harm. The study identified various psychosocial and economic factors that contribute to suicide and deliberate self-harm among women in Nepal. These factors include abuse, interpersonal conflicts, marital disputes, relationship issues, adjustment problems, unpaid loans, and financial losses. Additionally, mental health conditions such as mood disorder, adjustment disorder, and substance abuse disorder were also found to be significant contributors to these occurrences. [3]. Excessive utilisation of social media can result in mental health issues and reliance, especially among persons who are highly sensitive. Family dynamics in Nepal are influenced by socio-cultural and economic issues, which in turn impact the psychological and social well-being of women. Previous research on suicide among women has been limited, with a focus on several factors such as poverty, social isolation, gender inequity, education, and traditional systems. The utilisation of machine learning techniques has the potential to facilitate the identification of emotional states within social media posts, hence augmenting comprehension of depressive disorders and their associated consequences. This has the potential to enhance the identification, mitigation, and support provided to persons encountering mental health difficulties.

C) SCOPE OF STUDY

- Analyzes data from social media platforms to predict early signs of suicidal tendencies.
- Data categorized into positive, negative, or neutral.
- Uses Naïve Bayes, Logistic Regression, and SVM algorithms for classification.
- Maintains data confidentiality.

D)OBJECTIVES

- To analyze accuracy of Machine Learning algorithms (Naïve Bayes, Logistic Regression and Support Vector Machine algorithm) for prediction of suicidal tendency using social media content.
- To be able to compare accuracy of the Machine Learning algorithms (Naïve Bayes, Logistic Regression and Support Vector Machine algorithm).

2. LITERATURE REVIEW

Nepal's suicide rates have surged by 72% over the past decade, raising concerns about mental well-being, as reported by the Epidemiology and Disease Control Division [4]. Mental health issues are causing a significant rise in suicides globally, including in Nepal due to inadequate infrastructure and understanding, and Lesotho, with the highest suicide rate in the world [5]. Antigua & Barbuda, located in the Caribbean Sea, has the world's lowest suicide rate of 0.4 suicides per 100,000 inhabitants annually [6]. The study on machine learning predicts suicidal ideation, planning, and attempt among Korean adults found satisfactory classification performance with sensitivity and accuracy ranges of 0.808-0.853 and 0.843-0.863, respectively. Four machine learning classifiers were used, with extreme gradient boosting showing superior performance. The research found that over 80% of individuals at risk for suicidal thoughts and planning could be accurately predicted [7]. The study uses machine learning and Natural Language Processing (NLP) techniques to analyze depression and suicide thoughts in Reddit comments and posts. The algorithms, including Logistic Regression, Naïve Bayes, Support Vector Machine, and Random Forest, achieve high accuracy rates, enhancing comprehension of multidisciplinary fields related to suicide [8]. The Support Vector Machine and Logistic Regression models were used to detect suicide intent in Spanish language social networks, achieving

74% and 79% accuracy respectively. The model aims to provide timely alerts and early treatments for suicide thoughts [9].

A critical review of social media research on suicide prevention found that machine learning techniques, including decision trees, achieved an F1 score of 95%-97%, demonstrating high accuracy in identifying and predicting suicide outcomes [10]. The LSTM algorithm outperformed other machine learning approaches in predicting suicidal tendency in social media posts, with an accuracy of 92.4%, using a dataset of 232,074 posts from Kaggle and sentiment classification techniques [11]. A multi-modal feature-based technique using a Logistic regression classifier achieved 87% accuracy in detecting suicide ideation from online social media. The system identified six distinct categories of online behaviors, enabling early identification of suicidal thoughts and clinical indicators [12]. A deep learning technique using LSTM-Attention-CNN combined model has achieved an accuracy rate of 90.3% in detecting suicidal ideation from social media posts, using natural language processing to gather behavioral and textual characteristics [13]. The study uses deep learning and machine learning models to detect and analyze suicidal ideation on social media. The model achieved 95% accuracy using textual features, with a 91.5% accuracy using XGBoost. The system uses a publicly accessible dataset [14]. Supervised Learning for Suicidal Ideation Detection in Online User Content achieved 95% accuracy using GBDT, SVM, MLFFNN, and LSTM. Using statistical, linguistic, word embedding, and theme features, the method demonstrated feasibility and practical use [15]. The study demonstrates an ensemble method for detecting suicidal ideation in social media, achieving 80.61% accuracy and 79.20% F1 score, suggesting that feature combinations can improve classification performance [16]. The study explores the use of machine learning and natural language processing techniques for detecting suicidal ideation on Reddit. It suggests a model using BERT embedding and Softmax layers, producing a macro F1-score of 0.477, and comparing its performance with Bi-LSTM and GloVe. The transformer-based universal sentence encoder outperformed baseline models with 94.16% accuracy, while the RoBERTa-based model achieved the highest accuracy at 95.21%. [17]. The rule-based approach on Twitter predicts suicidal risk behavior based on age distribution. Male users have a 32% higher suicide risk than female users under 37. The model shows superior accuracy, with rates ranging from 79% to 90% for 7 distributions and above 90% for 25 out of 33 nodes. Experimental research indicates that the average accuracy of suicide detection in online social networks is 70%. [18]. From Extracting psychiatric stressors for suicide from social media using deep learning achieved 74% of accuracy using CNN algorithm, 70.3% of accuracy using SVM, 68.9% of accuracy using ET, 66.5% of accuracy using RF, 69.9% of accuracy using LR and 72% of accuracy using Bi-LSTM. The study explored deep learning methods and transfer learning strategies to identify suicide-related mental stresses from Twitter. It utilized an existing annotation dataset from clinical literature to enhance the learning process [19]. The study demonstrates a machine learning-based approach for detecting suicidality among opioid users on Reddit. The model achieved 95% accuracy in Fast Text and 94.4% in Recurrent neural network. The study found that opioid usage is linked to higher unintentional overdose rates and suicide likelihood. The model uses datasets labeled with unique criteria to forecast and classify out-of-sample targets [20]. The study presents a deep learning framework for estimating suicide risk and mental health using social media. The model achieved an AUC of 0.848 and a True Positive Rate of 0.846, with a false positive rate of 0.1. The model also improved the identification of anxiety and bipolar illness, reducing the error rate by up to 11.9% compared to a single-task model [21]. The study used a Logistics Regression classifier to detect suicide ideation in Twitter users using multiple feature analysis. The goal was to identify characteristics in tweets that could indicate suicidal thoughts, using machine learning and natural language processing techniques [22]. The COVID-19 pandemic has significantly impacted Nepal's economic and mental health, leading to public health concerns. The government implemented lockdown measures to promote social distancing, but these measures have negatively affected the emotional well-being of the Nepalese population. The closure of movie halls, gyms, health clubs, and museums, as well as restrictions on cultural, social, and religious gatherings, have disrupted traditional grief procedures. These measures have instilled fear, worry, and uncertainty among the Nepalese population, affecting their overall well-being in various aspects of life [23]. A sociological study in Kathmandu Valley reveals an increasing trend in suicide rates each year, with women having a higher tendency to engage in suicide attempts. This is due to disparities in mental health symptoms, coping strategies, and societal expectations, such as gender-based violence and limited opportunities for education and work. Causes of suicide include mental diseases, despair, neurological problems, poverty, limited career opportunities, societal prejudice, gender-based violence, alcohol and drug misuse, cancer, and HIV infection. Females are more vulnerable, with early marriage and domestic violence being major causes [24].

3. RESEARCH METHODOLOGY

This chapter examines the research technique employed in this dissertation, highlighting the significance of choosing a suitable approach for achieving a successful project. The components of utmost importance in a dissertation are the structure, system integration framework, prototype design, parameters, and data gathering strategies. The suggested approach confers advantages upon various stakeholders, including mental health practitioners, community organisations, government agencies, educational institutions, and the general public, through the mitigation of stigma and the facilitation of open conversations surrounding mental health.

A) Methods of data collection
This paper employs a data collection methodology, systematic data collection from various sources, based on the study's characteristics, using appropriate tools.

B) Primary data
The dissertation utilized comment picker and export comments for data collection, with limited data from social media due to daily access restrictions.

C) Secondary data
The secondary data sources were obtained via Kaggle.

D) Data Management and Analysis
A total of 2200 comments were analyzed using Machine Learning Models (Naïve Bayes, Logistic Regression, Support Vector Machine) and categorized as positive, negative, or neutral.

E) Research process
The Research process is a organized procedure that encompasses gathering, evaluating and comprehend facts in order to address specific issue.

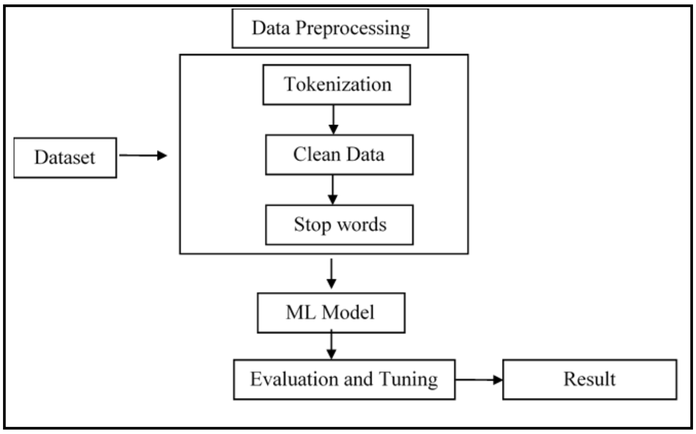

Figure 1: Research Process

Dataset

A dataset of 2200 comments, labeled with positive, negative, and neutral parameters, was collected and stored in an Excel file and saved in CSV format.

	A	B	C
1	username	text	sentiment
2	abiral_bishnu_official78	I`d have responded, if I were going	positive
3	anupriyagcpantha	Sooo SAD I will miss you here in Kathmandu!!!	positive
4	bestman23alive	my boss is bullying me...	negative
5	bikkygrg1	what interview! leave me alone	positive

Figure 2: Dataset

Data preprocessing

In machine learning involves cleaning and organizing unprocessed data for training and development, using tokenization, data cleansing, and stop words for analysis.

Tokenization

The initial data preparation stage in natural language processing, dividing text into smaller tokens like words or letters for easy computer processing and evaluation.

	A	B	C
1	username	text	sentiment
2	abiral_bishnu_official78	['responded', 'going']	positive
3	anupriyagcpantha	['Sooo', 'SAD', 'miss', 'Kathmandu']	positive
4	bestman23alive	['boss', 'bullying']	negative
5	bikkygrg1	['interview', 'leave', 'alone']	positive

Figure 3: Tokenization

Clean data

Data clean is the procedure of eliminating duplicate terms in order to enhance their quality.

Stop words

Stop words are a set of commonly used words in a language like ‘a’ , ‘the’ , ‘I’ , ‘am’ , ‘he’ , ‘she’ etc. are often filtered during the preprocessing .

Role of Sentiment analysis

Sentiment analysis is crucial for understanding emotions in written material, monitoring social media, and conducting market research. It aids businesses in understanding public sentiment, guiding decision-making, and maintaining brand reputation.

4. **RESULT**

A distribution plot is referred as displot, a type of data visualization that displays the distribution of a dataset. If sentiment score range from -1 (negative) to 1(positive), a displot could illustrate how many text samples fall into different sentiment polarity bins.

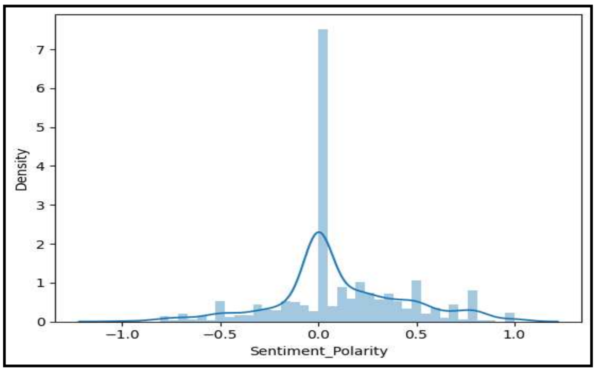

Figure 4: Distribution plot

A bar plot sentiment polarity visually represents the distribution of sentiment scores for different data points. Sentiment polarity scores typically range from negative to positive with zero representing neural sentiment.

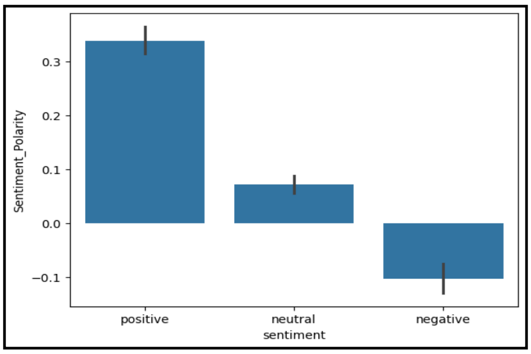

Figure: 5: Bar plot sentiment polarity

Naïve Bayes

The confusion matrix for the validation data serves a similar purpose to the testing data which helps assess generalization performance and it is often used during hyper parameter tuning to avoid overfitting.

- True Positive for class A : 151
- False positive for class B, but actual for class A : 7
- False positive for class C, but actual class is A : 8
- False negative for class A, but predicted class is B : 2
- True Positive for class B : 103
- False positive for class C, but actual class is B : 4

- False negative for class A, but predicted class is C : 14
- False negative for class B, but predicted class is C : 6
- True positive for class C : 101

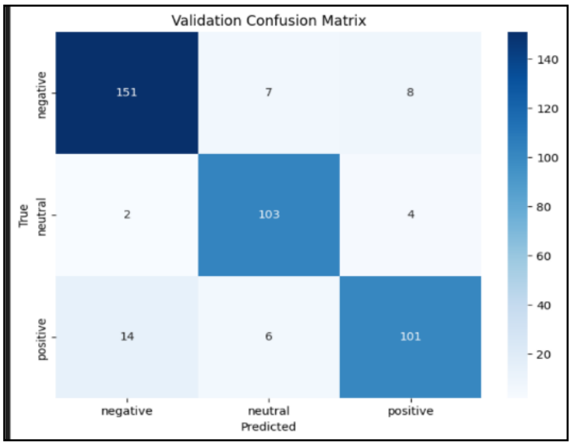

Figure 6: Validation confusion matrix for Naïve Bayes

The confusion matrix for the test data evaluated the model's performance on a set of data that is has not seen during training.

- True Positive for class A : 7
- False positive for class B, but actual for class A : 0
- False positive for class C, but actual class is A : 0
- False negative for class A, but predicted class is B : 1
- True Positive for class B : 20
- False positive for class C, but actual class is B : 0
- False negative for class A, but predicted class is C : 2
- False negative for class B, but predicted class is C : 0
- True positive for class C : 14

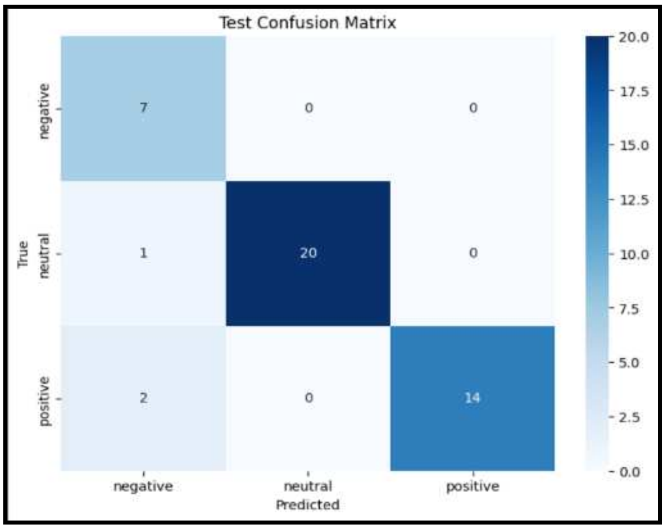

Figure 7: Test confusion matrix for Naïve Bayes

The model achieved a total accuracy of 89.65% during the validation phase, demonstrating strong performance across all classes and strong generalization capabilities when applied to new, untested data.

```
Validation Accuracy: 0.8964646464646465

Validation Classification Report:
              precision    recall  f1-score   support

    negative       0.90      0.91      0.91       166
     neutral       0.89      0.94      0.92       109
    positive       0.89      0.83      0.86       121

    accuracy                           0.90       396
   macro avg       0.90      0.90      0.90       396
weighted avg       0.90      0.90      0.90       396
```

Figure 8: Validation Accuracy for Naïve Bayes

The test phase yielded a 93.18% accuracy rate, demonstrating the model's ability to predict neutral and positive classes with high accuracy, recall, and F1-score metrics, but a diminished accuracy in negative classes.

Test Accuracy: 0.9318181818181818

Test Classification Report:

	precision	recall	f1-score	support
negative	0.70	1.00	0.82	7
neutral	1.00	0.95	0.98	21
positive	1.00	0.88	0.93	16
accuracy			0.93	44
macro avg	0.90	0.94	0.91	44
weighted avg	0.95	0.93	0.94	44

Figure 9: Test Accuracy for Naïve Bayes

Logistics Regression

The confusion matrix for the validation data serves a similar purpose to the testing data which helps assess generalization performance and it is often used during hyper parameter tuning to avoid overfitting.

- True Positive for class A : 126
- False positive for class B, but actual for class A : 8
- False positive for class C, but actual class is A : 14
- False negative for class A, but predicted class is B : 1
- True Positive for class B : 92
- False positive for class C, but actual class is B : 5
- False negative for class A, but predicted class is C : 2
- False negative for class B, but predicted class is C : 8
- True positive for class C : 96

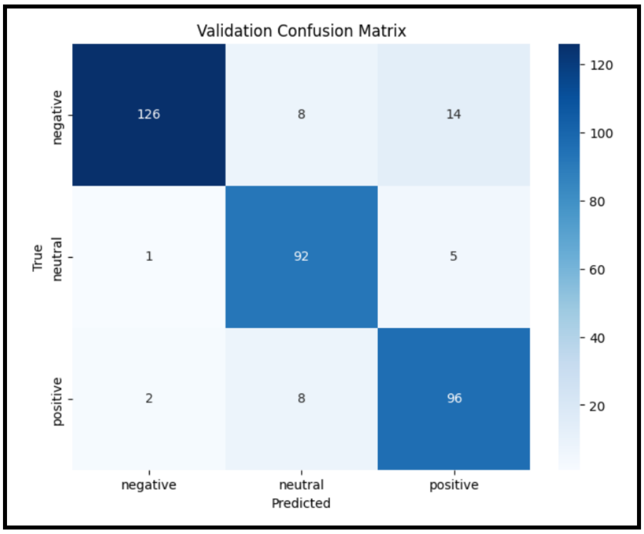

Figure 10: Validation Confusion matrix for Logistics regression

The confusion matrix for the test data evaluated the model's performance on a set of data that is has not seen during training.

- True Positive for class A : 23
- False positive for class B, but actual for class A : 0
- False positive for class C, but actual class is A : 2
- False negative for class A, but predicted class is B : 0
- True Positive for class B : 30
- False positive for class C, but actual class is B : 2
- False negative for class A, but predicted class is C : 0
- False negative for class B, but predicted class is C : 1
- True positive for class C : 30

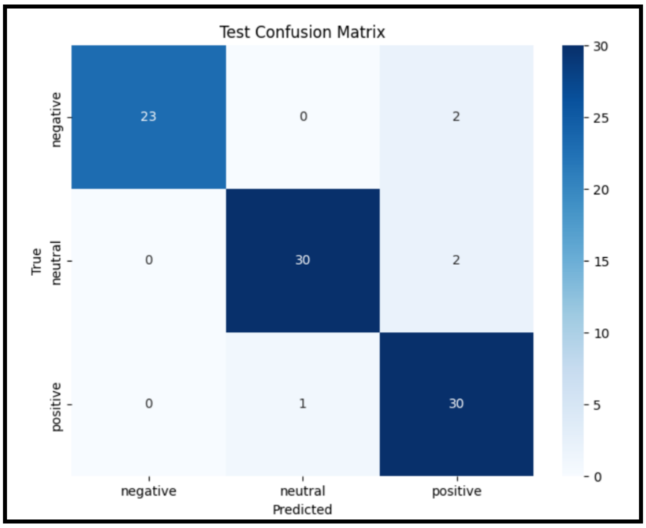

Figure 11: Test Confusion matrix for Logistics regression

The model achieved an accuracy of 89.20% during the validation phase, demonstrating strong performance in each class. The overall accuracy of 89.20% demonstrates the model's ability to generalize effectively to new and untested data.

```
Validation Accuracy: 0.8920454545454546

Validation Classification Report:
              precision    recall  f1-score   support

    negative       0.98      0.85      0.91       148
     neutral       0.85      0.94      0.89        98
    positive       0.83      0.91      0.87       106

    accuracy                           0.89       352
   macro avg       0.89      0.90      0.89       352
weighted avg       0.90      0.89      0.89       352
```

Figure 12: Validation accuracy for Logistics regression

The model achieved a 94.32% accuracy during testing, demonstrating strong performance across all classes. The classification report reveals high recall and accuracy, particularly for negative and neutral classifications. The model's overall accuracy of 94.32% demonstrates its ability to generalize effectively to new data.

Test Accuracy: 0.9431818181818182

Test Classification Report:

	precision	recall	f1-score	support
negative	1.00	0.92	0.96	25
neutral	0.97	0.94	0.95	32
positive	0.88	0.97	0.92	31
accuracy			0.94	88
macro avg	0.95	0.94	0.94	88
weighted avg	0.95	0.94	0.94	88

Figure 13: Test Accuracy for Logistics regression

Support Vector Machines

The confusion matrix for the validation data serves a similar purpose to the testing data which helps assess generalization performance and it is often used during hyper parameter tuning to avoid overfitting.

- True Positive for class A : 147
- False positive for class B, but actual for class A : 5
- False positive for class C, but actual class is A : 14
- False negative for class A, but predicted class is B : 1
- True Positive for class B : 102
- False positive for class C, but actual class is B : 6
- False negative for class A, but predicted class is C : 3
- False negative for class B, but predicted class is C : 10
- True positive for class C : 108

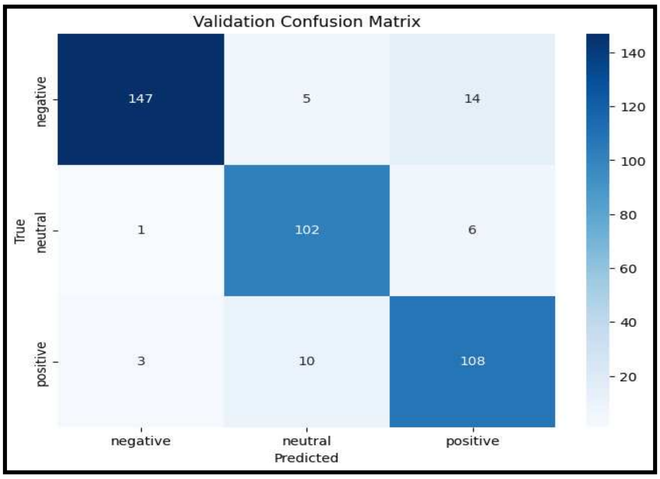

Figure 14: Validation Confusion matrix for Support Vector Machines

The confusion matrix for the test data evaluated the model's performance on a set of data that is has not seen during training.

- True Positive for class A : 7
- False positive for class B, but actual for class A : 0
- False positive for class C, but actual class is A : 0
- False negative for class A, but predicted class is B : 0
- True Positive for class B : 19
- False positive for class C, but actual class is B : 2
- False negative for class A, but predicted class is C : 0
- False negative for class B, but predicted class is C : 0
- True positive for class C : 16

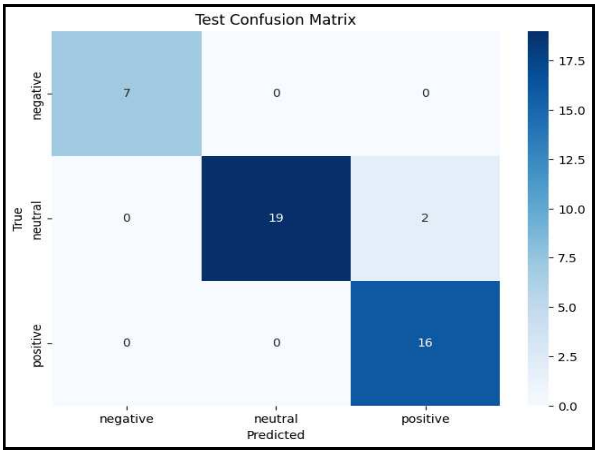

Figure 15: Test Confusion matrix for Support Vector Machines

The model's validation phase yielded a 90.15% accuracy, demonstrating strong performance in every class. The report evaluates accuracy, recall, and F1-score, demonstrating the model's ability to generalize to new data and demonstrate comprehensive performance.

```
Validation Accuracy: 0.9015151515151515

Validation Classification Report:
              precision    recall  f1-score   support

    negative       0.97      0.89      0.93       166
     neutral       0.87      0.94      0.90       109
    positive       0.84      0.89      0.87       121

    accuracy                           0.90       396
   macro avg       0.90      0.90      0.90       396
weighted avg       0.91      0.90      0.90       396
```

Figure 16 : Validation Accuracy of Support Vector Machine

The model achieved a remarkable accuracy of 95.45% during testing, with exceptional precision and recall for the negative class, good precision and recall for the neutral class, and great performance for the positive class, demonstrating its ability to effectively generalize to new data.

```
Test Accuracy: 0.9545454545454546

Test Classification Report:
              precision    recall  f1-score   support

    negative       1.00      1.00      1.00         7
     neutral       1.00      0.90      0.95        21
    positive       0.89      1.00      0.94        16

    accuracy                           0.95        44
   macro avg       0.96      0.97      0.96        44
weighted avg       0.96      0.95      0.95        44
```

Figure 17: Test Accuracy for Support Vector Machine

Summary of result

Table 1: Summary of accuracy

Classifier	Validation Accuracy	Test Accuracy
Naïve Bayes	89.65%	93.18%
Logistics Regression	89.20%	94.32%
Support Vector Machines	90.15%	95.45%

5. DISCUSSIONS AND ANALYSIS

The purpose of this dissertation study was to utilize social media data in order to forecast the inclination towards suicide. A total of 2200 datasets were gathered from various social media sites and stored in a CSV file. The file includes columns for usernames, text, and sentiment labels. The data was then divided into several sets for testing and validation. Three distinct algorithms are employed for comparison.

A) Naïve Bayes

Naïve Bayes is a widely utilised supervised learning technique in the field of text categorization, renowned for its straightforwardness and efficiency. It has the capability to process datasets with a large number of dimensions using a small amount of training data. The Naïve Bayes models can be classified into three types: Gaussian, Bernoulli, and Multinomial. The Multinomial model is utilized to assess the probability of encountering various terms inside a document, with a specific focus on discrete data such as text. Nevertheless, the performance of the model may be influenced by the assumption of feature independence.

B) Logistics Regression

In the mathematical modelling of logistic regression, a logistic function is utilized to convert the linear combination of predictor variables with coefficients. The linear combination of the predictors in the equation represents the log-odds of the event y being 1. The administration and synchronization of the movement of commodities, data, and assets.

In the context of binary classification issues, logistic regression is a statistical technique that involves the transformation of predictor variables into linear combinations with coefficients. Its main purpose is to oversee and synchronize the movement of products, information, and resources. The algorithm that is interpretable yields outcomes that are readily comprehensible, enabling the estimation of probabilities and the handling of binary and multinomial findings.

C) Support Vector Machines

The Support Vector Machine (SVM) is a robust supervised learning technique commonly employed for computational tasks involving regression and classification. The method effectively classifies data points within N-dimensional environments, particularly in cases where the dimensions surpass the number of samples. The process of text vectorization involves the conversion of raw data into numerical representations. The Support Vector Machine (SVM) model underwent training using TF-IDF vectors, and subsequently, a classification report was prepared for the validation data. Our dissertation proposes the utilization of Support Vector Machines (SVM) in the domains of education, mental health therapies, and policy creation. This study identifies early suicidal tendencies by analyzing social media data, focusing on depressed individuals. It also examines social, cultural, and economic interactions, identifying mental health status and aiding in government policy development for mental health.

D) Limitation

- Less amount of dataset were used.
- Only textual data were used for prediction.
- The different individuals may interpret the expression differently so identifying the suicidal tendency based on textual data is inherently subjective.

6. CONCLUSION

- Establish ethical protocols for data collection and application.
- Include diverse demographics and cultural backgrounds for impartial predictive model.
- Initiate public awareness campaign for user education.
- Develop algorithms to prevent misinterpretations and enhance forecast accuracy.
- Implement user feedback mechanism for predictions.

ACKNOWLEDGEMNTS

I would like to express my sincere gratitude to my wonderful supervisor, Er. Kumar Prasun, for his constant support and motivation, which motivated me to work on this paper. However, without the kind support and advice, this achievement would not have been possible. I also want to sincerely thank Dr. Gajendra Sharma for support and guidance. Additionally, I would like to thank my parents and sister for moral support. Also my friend Yubraj Bhattarai

REFERENCES

[1] F. B. Mphc, E. C. Mphc, M. G. Mphc, E. L. Mphc, and C. Kebles, "Green Spaces and Potential Impacts on Mental Health: A Pilot Study Regarding College Students Ability to Cope with Stress and Anxiety," 2022, doi: 10.1016/S0140-6736(21)02143-7.

[2]"What Is the Connection Between Depression and Suicide? | JED." Accessed: Dec. 05, 2023. [Online]. Available: https://jedfoundation.org/resource/what-is-the-connection-between-depression-and-suicide/

[3]S. P. Kasaju, A. Krumeich, and M. Van der Putten, "Suicide and deliberate self-harm among women in Nepal: a scoping review," BMC Womens Health, vol. 21, no. 1, pp. 1–10, Dec. 2021, doi: 10.1186/S12905-021-01547-3/TABLES/2.

[4]"Nepal's mental health crisis." Accessed: Dec. 05, 2023. [Online]. Available: https://kathmandupost.com/columns/2023/08/23/nepal-s-mental-health-crisis

[5]J. Renaud et al., "Suicidal ideation and behavior in youth in low- and middle-income countries: A brief review of risk factors and implications for prevention," Front Psychiatry, vol. 13, p. 1044354, Dec. 2022, doi: 10.3389/FPSYT.2022.1044354/BIBTEX.

[6]"Suicide mortality rate (per 100,000 population) - Antigua and Barbuda | Data." Accessed: Dec. 05, 2023. [Online]. Available: https://data.worldbank.org/indicator/SH.STA.SUIC.P5?view=chart&locations=AG

[7]J. Lee and T. Y. Pak, "Machine learning prediction of suicidal ideation, planning, and attempt among Korean adults: A population-based study," SSM Popul Health, vol. 19, Sep. 2022, doi: 10.1016/j.ssmph.2022.101231.

[8]M. Villalba et al., "Depression and Suicide Analysis Using Machine Learning and NLP You may also like Laser Treated Carbon Composites for the Determination of Suicide Biomarkers: Development of a Forensic Diagnosis Device Weather and Suicide of Construction Workers in Australia A Alashwal and A A Moustafa-Chernobyl cleanup workers from Estonia: follow-up for cancer incidence and mortality Depression and Suicide Analysis Using Machine Learning and NLP," J Phys Conf Ser, vol. 2161, p. 12034, 2022, doi: 10.1088/1742-6596/2161/1/012034.

[9]K. Valeriano, A. Condori-Larico, and J. Sulla-Torres, "Detection of suicidal intent in Spanish language social networks using machine learning," International Journal of Advanced Computer Science and Applications, vol. 11, no. 4, pp. 688–698, 2020, doi: 10.14569/IJACSA.2020.0110489.

[10]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 Journal of Science, vol. 62, no. 11, pp. 4092–4100, Nov. 2021, doi: 10.24996/ijs.2021.62.11.29.

[11]Z. Ao, J. Lai, W. Lu, and H. Mo, "Comparisons of LSTM and other Machine Learning Approaches on Predicting Suicidal Tendency Regarding to Social Media Posts," Proceedings - 2021 International Conference on Computers and Automation, CompAuto 2021, pp. 19–23, 2021, doi: 10.1109/COMPAUTO54408.2021.00011.

[12]M. Chatterjee, P. Kumar, P. Samanta, and D. Sarkar, "Suicide ideation detection from online social media: A multi-modal feature based technique," International Journal of Information Management Data Insights, vol. 2, no. 2, Nov. 2022, doi: 10.1016/j.jjimei.2022.100103.

[13]S. Renjith, A. Abraham, S. B. Jyothi, L. Chandran, and J. Thomson, "An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 10, pp. 9564–9575, Nov. 2022, doi: 10.1016/j.jksuci.2021.11.010.

[14]T. H. H. Aldhyani, S. N. Alsubari, A. S. Alshebami, H. Alkahtani, and Z. A. T. Ahmed, "Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models," Int J Environ Res Public Health, vol. 19, no. 19, Oct. 2022, doi: 10.3390/ijerph191912635.

[15]S. Ji, C. P. Yu, S. F. Fung, S. Pan, and G. Long, "Supervised learning for suicidal ideation detection in online user content," Complexity, vol. 2018, 2018, doi: 10.1155/2018/6157249.

[16]A. Li, J. Liu, M. Shi, and H. Jiang, "Detecting Suicidal Ideation in Social Media: An Ensemble Method Based on Feature Fusion," 2022, doi: 10.3390/ijerph19138197.

[17]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," International Journal of Environmental Research and Public Health 2022, Vol. 19, Page 10347, vol. 19, no. 16, p. 10347, Aug. 2022, doi: 10.3390/IJERPH191610347.

[18]E. R. Kumar and N. Venkatram, "Predicting and analyzing suicidal risk behavior using rule-based approach in Twitter data," Soft comput, pp. 1–9, Apr. 2023, doi: 10.1007/S00500-023-08095-Y/METRICS.

[19]J. Du et al., "Extracting psychiatric stressors for suicide from social media using deep learning", doi: 10.1186/s12911-018-0632-8.

[20]H. Yao, S. Rashidian, X. Dong, H. Duanmu, R. N. Rosenthal, and F. Wang, "Detection of Suicidality Among Opioid Users on Reddit: Machine Learning–Based Approach," J Med Internet Res 2020;22(11):e15293 https://www.jmir.org/2020/11/e15293, vol. 22, no. 11, p. e15293, Nov. 2020, doi: 10.2196/15293.

[21]A. Benton, M. Mitchell, and D. Hovy, "Multi-Task Learning for Mental Health using Social Media Text", Accessed: Dec. 07, 2023. [Online]. Available: https://www.nami.org/Learn-

[22]"Suicide Ideation Detection using Multiple Feature Analysis from Twitter Data | IEEE Conference Publication | IEEE Xplore." Accessed: Dec. 07, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9753295

[23]K. Poudel and P. Subedi, "Impact of COVID-19 pandemic on socioeconomic and mental health aspects in Nepal," International Journal of Social Psychiatry, vol. 66, no. 8, pp. 748–755, Dec. 2020, doi: 10.1177/0020764020942247/ASSET/IMAGES/LARGE/10.1177_0020764020942247-FIG2.JPEG.

IJRP.ORG
International Journal of Research Publications
ISSN: 2708-3578 (Online)

[24]R. Rajbahak, "SUICIDAL TENDENCY AMONG WOMEN (A Sociological Study of Suicide in Kathmandu Valley)".