2023_Spring_Analyzing_Suicidal_Ideation_in_Online_Communities_using_Natural_Language_Processing_Techniques

by SADID ISLAM

Submission date: 09-May-2023 01:01PM (UTC+0530)
Submission ID: 2083187255
File name: ine_Communities_using_Natural_Language_Processing_Techniques.pdf (421.62K)
Word count: 3692
Character count: 20800

Analyzing Suicidal Ideation in Online Communities using Natural Language Processing Techniques

Sadid Islam, Mubashira Rahman, Riead Hasan Khan
Department of Computer Science and Engineering
Brac University, Dhaka, Bangladesh
Email: sadid.islam@g.bracu.ac.bd, mubashira.rahman@g.bracu.ac.bd
rieadhasankhan@bracu.ac.bd

***Abstract*—This study uses several machine learning models to analyze posts from the r/SuicideWatch subreddit in order to identify patterns associated with suicide ideation. A qualitative analysis of the language and themes expressed in these posts was conducted. The models used were the Support Vector Machine Classifier, Multinomial Naive Bayes Classifier, Random Forest, and Multilayer Perceptron Classifier. The performance of these models was compared, and the SVM classifier had the most accuracy with a score of 93.22%. Additionally, this paper highlights the potential of social media platforms as a valuable source of data for studying suicide ideation and the effectiveness of machine learning techniques for analyzing large volumes of data in this area of research.**

***Index Terms*—Suicide ideation, Natural Language Processing, F-Net,**

I. Introduction

Suicide has become a global public health crisis, affecting millions worldwide. Suicide rates are on an upward trajectory in numerous regions of the world despite growing awareness and attempts at tackling this global problem, causing it to be a significant and continuous concern for medical professionals, legislators, and mental health activists. Together with the 703 000 suicides per year, many more people make suicide attempts. Every suicide is a tragedy that profoundly impacts the survivors, including the families, communities, and whole nations. Suicide occurred at any age and was the fourth most prominent cause of death worldwide for those aged 15 to 29 in 2019 (Suicide, 2021). [7] Finding and comprehending the components that lead to suicidal thoughts is an essential field of research since it is a severe issue that needs attention and help.

Social media platforms like Reddit have become an important source of data for researchers studying this phenomenon. Five meta-analyses yielded associations of general use of social network sites (SNS use) with higher levels of adolescent ill-being that ranged from very small to moderate (Vahedi Zannella, 2021). [11] By examining the language and themes covered in posts, researchers might learn more about the causes of suicidal thoughts and spot those who might be at risk for engaging in suicidal conduct. Utilizing machine learning approaches has greatly improved our capacity to examine and comprehend information from social media that relate to suicidal thoughts. Even when people are not declaring their intention to kill themselves, these strategies can help researchers find linguistic patterns that are suggestive of suicidal thoughts or acts. The research of suicidal thoughts on social media sites is becoming more common, and machine learning techniques like sentiment analysis and natural language processing (NLP) are being employed more often. With the use of these tools, enormous volumes of data may be quickly and accurately analyzed, yielding insightful information on the emotional states and linguistic patterns connected to suicidal thoughts.

In this study, we aim to use several machine learning models, such as the Multinomial Naive Bayes Classifier, Random Forest Classifier, Support Vector Classifier and finally Multilayer Perceptron Classifier to evaluate posts from the r/SuicideWatch subreddit, a community where people may share their suicide ideation difficulties and get help and assistance. Many deep learning models and transformer architectures have been proven to be very accurate in this field of classification, but they are quite resource-intensive. So,our goal is to find the effectiveness of Multinomial Naive Bayes Classifier, Random Forest Classifier,Support Vector Classifier and finally Multilayer Perceptron Classifier for the task of suicidal ideation detection as an alternative to resource-intensive processes. .

II. Literature Review

The combination of social media data with machine learning and natural language processing offers a singular chance to spot and help people who are at risk of suicide and has the potential to have a big influence on efforts to prevent suicide. Researchers can create more potent suicide detection and response efforts by spotting these tendencies in social media data. In a study Agarwal and Dhingra (2021) present a novel approach based on deep learning to detect suicidal intent in literature featuring Hindi-English code mixing. Their study introduces the Hinglish Suicide Ideation Corpus (HSIC), a unique collection of social media posts that combine Hindi and English languages and have been annotated for suicidal ideation. The proposed technique utilizes a deep neural network architecture, incorporating contextual information, semantic and syntactic elements, and language embeddings, to determine the presence of suicide intent in each post. Results indicate that code-mixed posts achieve an overall F1 score

of 0.55, while English and Hindi posts achieve F1 scores of 0.60 and 0.50, respectively. [12] These findings demonstrate the potential effectiveness of the suggested method in identifying suicidal thoughts within Hindi-English code-mixed literature, representing a significant advancement in this domain. However, the study acknowledges limitations related to the generalizability of the model and the size and representativeness of the corpus. Nonetheless, the results provide valuable groundwork for further research on identifying suicidal thoughts in code-mixed text. Another paper suggests a unique method for identifying suicidal thoughts in Chinese microblogs using psychological lexicons. After manually annotating a dataset of Chinese microblogs from Sina Weibo, they used a set of psychological lexicons to extract semantic features from the text to find microblogs showing suicidal ideation. They then compared their method to other machine learning and natural language processing techniques to gauge its effectiveness. This strategy works well at spotting people in danger of acting suicidally and giving them the right treatments. The psychological lexicon-based strategy fared better than the other strategies, as evidenced by its F1 score of 0.72. With their method, the authors offer a fresh viewpoint on how psychological lexicons might be used to identify suicidal thoughts in Chinese microblogs. More study is required to assess the generalizability and scalability of this technique to other social media platforms and languages. Additionally, the study includes limitations that may restrict how broadly the findings may be applied, such as using a single dataset and manual annotation. However, the suggested method is a step in the right direction for identifying suicidal thoughts and can help prevent and treat suicidal conduct (Huang et al., 2014). [5]

According to Yeskuatov and his team, [9] due to the unstructured and cluttered nature of the material, identifying suicidal thoughts on social media is a difficult process. They have suggested machine learning and natural language processing (NLP) methods to deal with this problem. This work developed a classifier using mood, emotion, and language variables using a dataset of Reddit postings on suicide. The F1-score of 0.83 obtained by the classifier indicates great accuracy. Once more, they developed a machine learning classifier using the dataset of Reddit postings about mental health using a combination of linguistic, structural, and temporal variables. This classifier's F1-score of 0.86 demonstrates the strategy's success. In addition, they classified Reddit postings on suicide using a deep learning strategy based on convolutional neural networks (CNNs). The strategy outperformed conventional machine learning methods with an F1-score of 0.87. In all, it has been discovered that a mixture of lexical and semantic data, together with deep learning-based methodologies, is efficient in identifying suicidal intent on Reddit. Another research by Ji.S et al.(2020) offers an overview of machine learning techniques and programs for identifying suicidal thoughts on social media sites. [8] The authors looked at papers that suggested several strategies based on multimodal, linguistic, and semantic aspects. They highlighted several possible applications for suicide prevention and intervention while also summarizing the advantages and disadvantages of various machine-learning approaches. The authors outlined the advantages and disadvantages of the various machine learning methods for identifying suicidal thoughts. High accuracy and sensitivity were among the positive traits, whereas low specificity and generalizability were the negative traits. The authors additionally identified several uses for machine learning tools for spotting suicidal ideation, such as real-time surveillance and assistance for those who could act suicidally. A thorough examination of the state of the art is not provided, and the review is constrained by the inclusion of just a portion of the literature already available on the subject. The research still offers a helpful overview of the many machine-learning approaches and applications for identifying suicidal ideation, which can help create successful suicide prevention and intervention techniques. The article (Chatterjee et al., 2021) suggests a unique method for identifying suicidal thoughts on Twitter by using several feature analyses. In order to extract information from a collection of tweets involving suicidal thoughts, the authors employed linguistic, semantic, sentiment, network, and temporal aspects. The suggested method successfully identified suicidal thoughts on Twitter with excellent accuracy, sensitivity, and specificity. [10] After comparing it to those approaches, the authors stated that their approach outperformed existing machine learning techniques. The model can capture the many facets of the tweets due to the merging of several variables, including linguistic, semantic, sentiment, network, and temporal information. The study does, however, have specific areas for improvement. The study only used a dataset from Twitter; therefore, the suggested method might not work with other social media sites or languages. Furthermore, the study did not consider how user factors, including age and gender, may affect the identification of suicidal thoughts.

The study by Patel et al.(2021), analyzes internet health discussion boards to find mental and physical problems connected to COVID-19. According to the report, the two most prevalent mental ailments are anxiety and depression, while the three most prevalent physical disorders are exhaustion, respiratory issues, and loss of taste and smell. The authors emphasize the potential of NLP methods for examining digital healthcare communities and spotting linguistic patterns and trends among forum users. The study may not reflect the opinions and experiences of the larger public because it relies solely on the information provided by individuals from online health forums. [2] Kabir et al. (2022) detected depression severity from Bengali social media texts using natural language processing techniques. [1] They used text-based data in Bengali from blogs and open postings to create a procedure for creating annotated corpora and extracting textual data from Bengali literature for predictive modeling. The author used machine learning and deep learning models to assess the degree of depression from texts. They utilized DSM-5 to properly diagnose and Selenium to scrape data from social media. They employed four separate classified labels to distinguish between the texts. They employed models

such as the random forest, support vector machine, logistic regression, k-nearest neighbor, and Naive Bayes for preprocessing and data modeling. In comparison to previous models, the authors claim that the recurrent neural network model successfully predicted the severity. Priya et al., (2020) discuss the purpose of the study, data were gathered by utilizing the DASS-21 questionnaire and text analysis from social media, which resulted in predictions of anxiety, depression, and stress created using a machine learning algorithm. They employed various machine learning algorithms and evaluated which one provided the highest level of accuracy. For instance, when they used the CNN and SVM algorithms to forecast stress and anxiety, the CNN algorithm performed with an accuracy of around 79% whereas the SVM algorithm performed with only 58%. [3]They did this by comparing each algorithm's output for each prediction and selecting the algorithm with the best accuracy for predicting the outcome. Lastly, another study used assessments of eight fundamental emotions as characteristics from Twitter tweets throughout time, including a temporal analysis of these features, to create a novel technique for identifying users who have or are at risk of developing depression. With the use of descriptive statistics from the emotion time series as inputs, they improved the training procedure's accuracy. They measured emotions using the EMOTIVE technology and Ekman's core emotion model. For differentiation, the data set was divided into sets of temporal and nontemporal features. Statistical and mathematical techniques were applied, including mean, standard deviation, entropy, mean momentum, and mean differencing. They contrasted the outcomes of temporal and nontemporal data sets generated by the suggested systems(Chen et al., 2018). [6]

III. Dataset

The dataset compiles postings from the Reddit platform's "SuicideWatch" and "depression" subreddits. The Pushshift API is used to gather the posts. All posts made to "SuicideWatch" between Dec. 16, 2008 (when it was created) and Jan. 2, 2021 were gathered, whereas postings made to "depression" between Jan. 1, 2009, and Jan. 2, 2021, were gathered. [4] This dataset has 232074 unique values divided and labeled into 2 categories, suicide or non-suicide. The texts are basically people sharing their feeling publicly on social media. The mention of suicide was evident in the texts. We used 20 thousand data from this dataset in this research

IV. Data Prepocessing

The dataset was divided into testing and training sets with a fixed random state of 42, preserving an 80:20 ratio for consistency. Different deep learning and machine learning models, such as Multi-layered Perceptron Classifier Random Forest, Support Vector Classifier, and Multinomial Naive Bayes were trained using the training data. To improve each model's performance, appropriate hyperparameters were added.

The quantity of the dataset made it unfeasible to use an automated method to find the best hyperparameters due to the length of time it would take. An alternative method involved manually adjusting the hyperparameters and assessing how it affected the model's accuracy was chosen. By identifying local accuracy maxima using this technique, the most effective hyperparameters might be chosen.

V. Methodology

A. *Model Training and Evaluation*

The dataset was divided into testing and training sets with a fixed random state of 42, preserving an 80:20 ratio for consistency. Different deep learning and machine learning models, such as Multi-layered Perceptron Classifier Random Forest, Support Vector Classifier and Multinomial Naive Bayes were trained using the training data. To improve each model's performance, appropriate hyperparameters were added.

The quantity of the dataset made it unfeasible to use an automated method to find the best hyperparameters due to the length of time it would take. An alternative method involved manually adjusting the hyperparameters and assessing how it affected the model's accuracy was chosen. By identifying local accuracy maxima using this technique, the most effective hyperparameters might be chosen.

B. *Multinomial Naive Bayes*

We have set the parameter 'alpha' as 0.1 Multinomial Naive Bayes classifier and denoted 'fit prior' as true. After the algorithm was tuned to the training dataset, predictions on the test dataset were obtained.

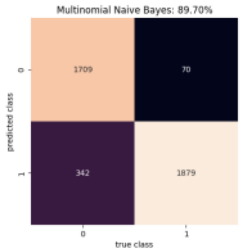

Figure 1: (a) Confusion Matrix of Multinomial Naive Bayes

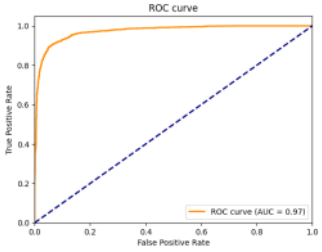

Figure 1: (b) ROC curve of the Multinomial Naive Bayes

As seen in Figure 1, (a) shows the confusion matrix, and (b) for the ROC curve of Multinomial Naive Bayes classifier ROC curve. In the test dataset, 89.7% of the cases had their labels predicted correctly. 90.47% of the anticipated positive events were accurately classified. The model's high level of accuracy

in differentiating between positive and negative classifications is demonstrated by the area under the curve (AUC), which is 0.97.

C. Support Vector Classifier

We have used the "linear" kernel parameter was used to build the SVM classifier.c=1 is set as the regularization value The "probability" parameter was also enabled by choosing "True." Reproducibility was ensured by setting the "random state" parameter to 42.

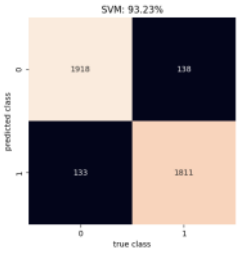

Figure 2: (a) Confusion Matrix of Support Vector Classifier

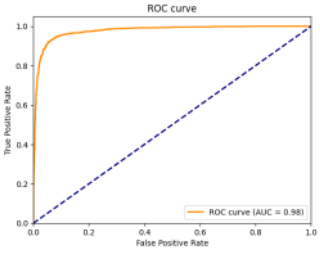

Figure 2: (b) ROC curve of Support Vector Classifier

The results are shown in Figure 2, where (a) is the confusion matrix and (b) is the support vector classifier's ROC curve. The Support Vector Machine classifier's area under the curve is calculated to be 0.98. The classifier's overall accuracy was 93.22%.

D. Random Forest Classifier

For the Random Forest Classifier, we set the number of trees to 500, the maximum depth of each tree to 100, the minimum number of samples needed at a leaf node to 2, and the random seed value to 42. The training data were then fitted to the Random Forest classifier.

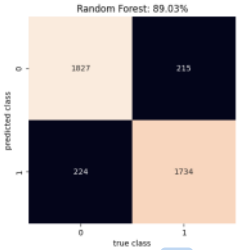

Figure 3:(a) Confusion Matrix of Random Forest Classifier

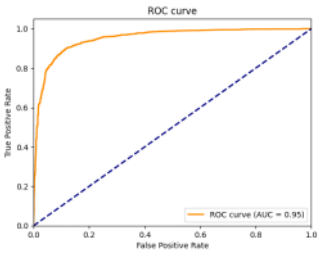

Figure 3:(b) ROC curve of Random Forest Classifier

As seen in Figure 3, (a) represents the ROC curve and (b) represents the Confusion matrix. The area under the curve score for the Random Forest classifier is 0.95. The classifier had an overall accuracy of 0.8902, which means it correctly predicted the label of 89.02% of the cases in the test dataset.

E. Multilayer Perceptron classifier

For the Multilayer Perceptron classifier, we used a hidden layer with 50 neurons and the 'ReLU' activation function. We also used the 'adam' solver. For L2 regularization, the classifier employed an alpha value of 0.0001 and an initial learning rate of 0.001. The random state was set to 42, and the maximum number of iterations was set at 500.

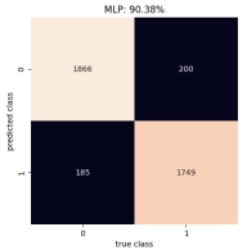

Figure 4: (a) Confusion Matrix of Multilayer Perceptron classifier

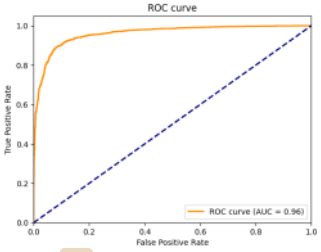

Figure 4: (b) ROC curve of Multilayer Perceptron classifier

As shown in Figure 4, (a) represents the confusion matrix, and (b) represents the ROC curve. The accuracy of the MLP classifier is 0.9037, or 90.37%. The area under the curve score is 0.96.

VI. Discussion

In this research, we have used precision, recall, F1 score for generating the output. After using the algorithms the results

were compared. The comparison between the results of algorithms gave much clarity to the work. In the confusion matrix "true class" was plotted in the X axis whereas "predicted class" was plotted on the Y axis. For the ROC curve graph, X axis was "False positive rate" and Y axis was "True positive rate". Firstly, Table 1 shows the classification report summarizing the performance of the Multinomial Naive Bayes classifier model.

	precision	recall	f1-score
non-suicide	0.960652	0.833252	0.892428
suicide	0.846015	0.964084	0.901199
accuracy	0.897	0.897	0.897
macro avg	0.903334	0.898668	0.896814
weighted avg	0.904795	0.897	0.896702

Table 1: Performance of Multinomial Naive Bayes

Then Table 2 the classification report summarizing the performance of the Support Vector Classifier model portrays the comparison.

	precision	recall	f1-score	support
non-suicide	0.932879	0.935154	0.934015	2051
suicide	0.931584	0.929194	0.930388	1949
accuracy	0.93225	0.93225	0.93225	0.93225
macro avg	0.932232	0.932174	0.932201	4000
weighted avg	0.932248	0.93225	0.932248	4000

Table 2: Performance of Support Vector Classifier

Table 3 contains the classification report summarizing the performance of the Random Forest Classifier model.

	precision	recall	f1-score	support
non-suicide	0.894711	0.890785	0.892744	2051
suicide	0.885598	0.889687	0.887638	1949
accuracy	0.89025	0.89025	0.89025	0.89025
macro avg	0.890154	0.890236	0.890191	4000
weighted avg	0.890271	0.89025	0.890256	4000

Table 3: Performance of Random Forest Classifier

Again, Table 4 contains the classification report summarizing the performance of the Multilayer Perceptron classifier model.

	precision	recall	f1-score	support
non-suicide	0.903195	0.9098	0.906485	2051
suicide	0.904343	0.897383	0.90085	1949
accuracy	0.90375	0.90375	0.90375	0.90375
macro avg	0.903769	0.903592	0.903668	4000
weighted avg	0.903754	0.90375	0.903739	4000

Table 4: Performance of Multilayer Perceptron classifier

Finally plotting the results of all the algorithms can be clearly seen the differences. The slopes were compared against a perfectly calibrated curve. In the X axis mean productive value and in Y axis fraction of positives were plotted.As seen from the calibration curve Support Vector Machine Classifier had the best performance with an accuracy of 93.22% .

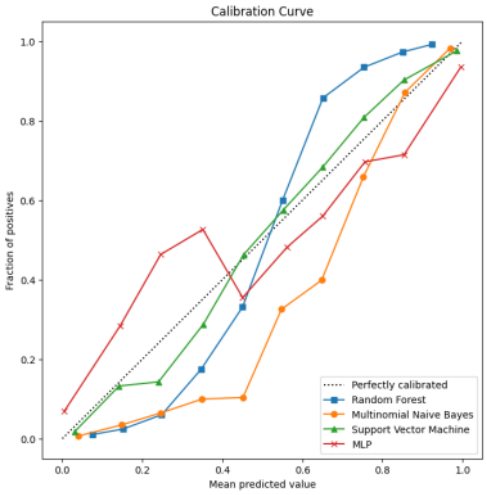

Figure 5: Comparison of all algorithms

VII. Future Work and Conclusion

A. Future Work

To improve upon our work, future work should focus on creating a larger dataset. Also, working with a multilingual dataset is more contextual, given that most of the posts are written in mixed languages. This would make it adaptable to different languages. Furthermore, improvements can be made by evaluating the performance of transformer models and other deep learning models in terms of accuracy, resource consumption, and training time and then comparing them with the models used in our research. So that we can identify the strengths and weaknesses of the models involved in creating a more efficient and accurate model for suicidal ideation detection.

B. Conclusion

In conclusion, social media sites have become part and parcel of our lives, and with people venting their frustrations online, they have become a valuable source of data. With the increasing rate of suicide and social media platforms in mind, we worked to create an accurate model for suicidal ideation detection. Our research involved four different models- Support Vector Machine Classifier, Multinomial Naive Bayes Classifier,Random Forest and Multilayer Perceptron Classifier.Out of which Support Vector machine had an accuracy of 93.22

REFERENCES

[1] M. K. Kabir, M. Islam, A. N. B. Kabir, A. Haque, and M. K. Rhaman, "Detection of depression severity using Bengali social media posts on mental health: Study using natural language processing techniques," JMIR Formative Research, vol. 6, no. 9, e36118, 2022.

[2] Patel, R., Smeraldi, F., Abdollahyan, M., Irving, J., Bessant, C. (2021). Analysis of mental and physical disorders associated with COVID-19 in online health forums: a natural language processing study. BMJ open, 11(11), e056601.

[3] A. Priya, S. Garg, and N. P. Tigga, "Predicting anxiety, depression, and stress in modern life using machine learning algorithms," Procedia Computer Science, vol. 167, pp. 1258–1267, 2020.

[4] Suicide and Depression Detection. (2021, May 19). Kaggle. https://rb.gy/cmdli

[5] Huang, X., Zhang, L., Chiu, D., Liu, T., Li, X., Zhu, T. (2014, December). Detecting suicidal ideation in Chinese microblogs with psychological lexicons. In 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (pp. 844-849). IEEE.

[6] X. Chen, M. D. Sykora, T. W. Jackson, and S. Elayan, "What about mood swings: Identifying depression on twitter with temporal measures of emotions," in Companion Proceedings of the The Web Conference 2018, 2018, pp. 1653–1660.

[7] World Health Organization: WHO. (2021). Suicide. www.who.int. https://www.who.int/news-room/fact-sheets/detail/suicide

[8] Ji, S., Pan, S., Li, X., Cambria, E., Long, G., Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. IEEE Transactions on Computational Social Systems, 8(1), 214-226.

[9] Yeskuatov, E., Chua, S. L., Foo, L. K. (2022). Leveraging Reddit for Suicidal Ideation Detection: A Review of Machine Learning and Natural Language Processing Techniques. International journal of environmental research and public health, 19(16), 10347.

[10] Chatterjee, M., Samanta, P., Kumar, P., Sarkar, D. (2022, February). Suicide ideation detection using multiple feature analysis from Twitter data. In 2022 IEEE Delhi Section Conference (DELCON) (pp. 1-6). IEEE.

[11] Vahedi, Z., Zannella, L. (2021). The association between self-reported depressive symptoms and the use of social networking sites (SNS): A meta-analysis. Current Psychology, 40, 2174-2189.

[12] Agarwal, K., Dhingra, B. (2021, December). Deep Learning Based Approach For Detecting Suicidal Ideation in Hindi-English Code-Mixed Text: Baseline and Corpus. In Proceedings of the 18th International Conference on Natural Language Processing (ICON) (pp. 100-105).

2023_Spring_Analyzing_Suicidal_Ideation_in_Online_Commu...

ORIGINALITY REPORT

18%	9%	13%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to West Coast University Student Paper	1%
2	Patti M. Valkenburg, Adrian Meier, Ine Beyens. "Social Media Use and its Impact on Adolescent Mental Health: An Umbrella Review of the Evidence", Current Opinion in Psychology, 2021 Publication	1%
3	hdl.handle.net Internet Source	1%
4	www.mdpi.com Internet Source	1%
5	"Proceedings of International Conference on Intelligent Vision and Computing (ICIVC 2022)", Springer Science and Business Media LLC, 2023 Publication	1%
6	Dibyo Fabian Dofadar, Riyo Hayat Khan, Md. Golam Rabiul Alam. "COVID-19 Confirmed Cases and Deaths Prediction in Bangladesh	1%

Using Hidden Markov Model", 2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART), 2021
Publication

7	"Intelligent Data Communication Technologies and Internet of Things", Springer Science and Business Media LLC, 2021 Publication	1 %
8	Submitted to Shri Mata Vaishno Devi University(SMVDU), Katra Student Paper	1 %
9	www.ijraset.com Internet Source	1 %
10	Muhammad Asif, Atiab Ishtiaq, Haseeb Ahmad, Hanan Aljuaid, Jalal Shah. "Sentiment analysis of extremism in social media from textual information", Telematics and Informatics, 2020 Publication	1 %
11	P.S. Wasan, M. Uttamchandani, S. Moochhala, V.B. Yap, P.H. Yap. "Application of statistics and machine learning for risk stratification of heritable cardiac arrhythmias", Expert Systems with Applications, 2013 Publication	1 %
12	www.scilit.net Internet Source	1 %

13	Submitted to Griffith College Dublin Student Paper	<1%
14	Md. Nurul Alam, Muhammad Masroor Ali. "Loan Default Risk Prediction Using Knowledge Graph", 2022 14th International Conference on Knowledge and Smart Technology (KST), 2022 Publication	<1%
15	link.springer.com Internet Source	<1%
16	Submitted to Multimedia University Student Paper	<1%
17	Submitted to The University of the West of Scotland Student Paper	<1%
18	Irfan Ali Kandhro, Muhammad Ameen Chhajro, Kamlesh Kumar, Haque Nawaz Lashari, Usman Khan. "Student Feedback Sentiment Analysis Model Using Various Machine Learning Schemes A Review", Indian Journal of Science and Technology, 2019 Publication	<1%
19	Submitted to Northcentral Student Paper	<1%
20	Submitted to University of KwaZulu-Natal Student Paper	<1%

21	Submitted to Westcliff University Student Paper	<1%
22	www.ijert.org Internet Source	<1%
23	"Big Data Intelligence for Smart Applications", Springer Science and Business Media LLC, 2022 Publication	<1%
24	Piyush Katariya, Vedika Gupta, Rohan Arora, Adarsh Kumar, Shreya Dhingra, Qin Xin, Jude Hemanth. "A deep neural network-based approach for fake news detection in regional language", International Journal of Web Information Systems, 2022 Publication	<1%
25	Yoshio Nakano, Toru Takeshita, Noriaki Kamio, Susumu Shiota et al. "Supervised machine learning-based classification of oral malodor based on the microbiota in saliva samples", Artificial Intelligence in Medicine, 2014 Publication	<1%
26	Ajay Kumar, David Zhang. "Chapter 85 Biometric Recognition Using Feature Selection and Combination", Springer Science and Business Media LLC, 2005 Publication	<1%

27	Anu Priya, Shruti Garg, Neha Prerna Tigga. "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms", Procedia Computer Science, 2020 Publication	<1 %
28	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
29	www.jetir.org Internet Source	<1 %
30	www.medrxiv.org Internet Source	<1 %
31	Lina Ahmad, Amjed Al-Mousa. "Identification of Donald Trump's Tweets Using Machine Learning", 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD), 2021 Publication	<1 %
32	Mark Goadrich. "The relationship between Precision-Recall and ROC curves", Proceedings of the 23rd international conference on Machine learning - ICML 06 ICML 06, 2006 Publication	<1 %
33	Submitted to University of Rwanda Student Paper	<1 %

34	Zhengping Li, Kai Zhang, Bokui Chen, Yuhan Dong, Lin Zhang. "Driver identification in intelligent vehicle systems using machine learning algorithms", IET Intelligent Transport Systems, 2018 Publication	<1 %
35	academic.oup.com Internet Source	<1 %
36	opus.lib.uts.edu.au Internet Source	<1 %
37	preprints.jmir.org Internet Source	<1 %
38	www.microsave.net Internet Source	<1 %
39	"Data Science and Security", Springer Science and Business Media LLC, 2021 Publication	<1 %
40	Akshma Chadha, Baijnath Kaushik. "Performance Evaluation of Learning Models for Identification of Suicidal Thoughts", The Computer Journal, 2022 Publication	<1 %
41	Lecture Notes in Computer Science, 2011. Publication	<1 %
42	Lecture Notes in Computer Science, 2012. Publication	<1 %

43	Moumita Chatterjee, Piyush Kumar, Poulomi Samanta, Dhrubasish Sarkar. "Suicide ideation detection from online social media: A multi-modal feature based technique", International Journal of Information Management Data Insights, 2022 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		