Journal of Advanced Research in Applied Sciences and Engineering Technology 51, Issue 2 (2025) 86-96

SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage: https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Use of Natural Language Processing for the Detection of Hate Speech on Social Media

Mehedi Hasan Shohan[1], Kazi Rifat Ahmed[1], Nur Farhan Kahar[2,3], Nusrat Jahan[1,2], Md. Maruf Hassan[1,2,*], R. Badlishah Ahmad[2,3], Naimah Yaakob[2,3], Ong Bi Lynn[2,3], Nadira Islam[1]

[1] Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh
[2] Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia
[3] Centre of Excellence for Advanced Computing (AdvComp), Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia

ARTICLE INFO

Article history:
Received 23 November 2023
Received in revised form 12 May 2024
Accepted 21 August 2024
Available online 18 September 2024

Keywords:
Hateful speech; natural language processing; twitter; offensive language; deep neural network

ABSTRACT

Our society's communication patterns have fundamentally changed as a consequence of the emergence of social media platforms. One effect of these changes is a rise in unpleasant behaviours like making rude and derogatory comments online. Speaking harshly or disrespectfully to someone in person may be difficult. However, online abuse and posting of improper material are considered to be acceptable. Hate speech has the potential to hurt a person or a group of people. Inappropriate material must be identified, in order to be filtered or banned from the web. CNN is a type of deep machine-learning model that has been suggested for such identification, because it performs better than conventional techniques in resolving text categorization problems. Our goal investigates how hate speech may be detected using NLP. In addition, a recent technique has been used in this field to a dataset. This classifier is assigned in each tweet to one of the three Twitter dataset categories of hatred, foul language, or neither. This model's performance has been assessed with accuracy. The Naïve Bayes, the Decision Tree, KNN, Linear Regression, and the Random Forest are five algorithms that have been used. Of these, Linear Regression provided the greatest accuracy of 94%. It should be noted that when looking at each class separately, many hateful tweets have been mislabelled. It is advisable to look at the outcomes and faults in much detail, in order to comprehend the misclassification. Our analysis shows a better outcome in detecting hateful speech in social media.

1. Introduction

Social networks have seen a rise in nefarious behaviour along with a surge in usage of social media over the last 10 years. Hate speech is among those actions that users of YouTube, Facebook, Twitter, and other social media platforms need to protect themselves against, since it has the potential to cause maximum damage. Using a combination of ML and NLP, a method for predicting hate speech in social media platforms, and websites has offered and analysed in this study. Instances of hate

[*] *Corresponding author.*
E-mail address: ancssf@gmail.com

https://doi.org/10.37934/araset.51.2.8696

speech are stemmed, tokens are separated, characters are cleaned, and inflections are removed before the hate speech identification technique is finished. The acquired data is then analysed utilizing an advanced natural language processing optimization ensemble deep learning technique [1]. According to Al-Makhadmeh *et al.,* [2] the primary duties of NLP are semantics and syntax. There are several activities that come under the heading of syntax, including stemming, lemmatization, morphological segmentation, tagging speeches, parsing, and breaking sentences. Additionally, social media organizations like Twitter, Facebook, and YouTube have been chastised for failing to do more to combat hate speech proliferating their services and have been ordered to take action [2]. In fact, according to Shanita Biere *et al.*, [3] the government of Germany puts 50-million-euro penalties for social networks on the table annually if they don't take action against offensive statements within a week. Semantics includes tasks like sentiment analysis, entity recognition, language creation, language comprehension, character identification, and related tasks. These logical and statistical computing methods are used in a range of syntactic and semantic investigations. According to Davidson *et al.,* [4-6] hate speech is defined as declaration or conduct which targets an individual or a group of individuals due to that individual or group's gender, nationality, race, skin colour, ethnicity, sexual orientation, or place of residence. In the last 10 years, interactions on online social networks have increased significantly, which has resulted in a considerable surge in research on social media safety and security. Unfortunately, there has been a parallel surge in criminality that profit off this vast web of online social interactions. Because of the mobility and anonymity, often offered by social media platforms, hate speech has increased in volume and spread. Guang Xiang *et al.,* [7] social platforms have the outburst communication to each other, due to the quick growth of technology. Most users use Twitter to follow different people, and participate in different social activity on the platform, and provide comments on what they are thinking. Because of the anonymity provided by social media, it is easy for individuals to employ hate speech or other derogatory words during these digital conversations.

The logical techniques make use of guidelines for word extraction from different sentences and word-mapping, with particular criteria assigned for differentiating across languages. From very large linguistic databases, patterns are derived using statistical techniques. A language can be expressed on different levels, including sound, word, syllable, conversation, and phrase, despite the fact that NLP is a fantastic tool for hate speech detection. This review delves into hate speech detection, exploring its types (racism, misogyny, or religious hate-speech), and all the methods to combat it. The paper identifies challenges and suggests solutions, also emphasizing the significance of monitoring hate speech found on public-internet [8]. Despite substantial debate in legal circles, the First Amendment of social media security has not received much attention, especially in light of speech limitations on college campuses. In several countries, including the UK, Canada, and France, hate speech is prohibited by law. Communication that makes racial or ethnic minority the target and has the potential to incite violence or cause social instability is often called and known as hateful speech. If found guilty of using hateful speech, a purveyor typically faces severe punishment, including probable jail time. Many websites have put in place their own anti-hate speech guidelines due to the fact that these laws extend to social media and the internet. The number of people that frequently use social media is close to 3.484 billion, around 7.7 billion people. Social media has such a large user base that any material may spread there very quickly. Content on different racial or religious backgrounds, genders, locations, or mindsets may be found on Twitter platform. People express their viewpoints out of flexibility and "freedom of speech", without taking into account the views of others. Sometimes posts are so offensive or disrespectful that the victims take them to be a real assault on themselves. This happens when someone's emotions or sense of self are hurt. Repression of minorities, terrorist acts, and suicide attempts might happen as a consequence [9].

This approach study points to a research gap in the efficiency of current NLP methods for identifying hate speech, especially when it comes to correctly categorizing offensive tweets. Although the research uses a prior method on a Twitter dataset, which achieved high overall accuracy but mislabelled many hate-related tweets, it considers CNNs as a possible option owing to their better performance in text classification. The work makes a contribution by concentrating on examining these misclassifications in order to comprehend the shortcomings of existing techniques and provide guidance for the creation of more precise hate speech detection systems. Improved NLP models that successfully handle the expanding problem of online hate speech may be made possible by this greater comprehension.

2. Related Works

An overview of earlier research in which the two subjects have been integrated follows a description of the different Natural Language Processing methodologies. Even for humans, figuring out if a document includes hate speech is a difficult procedure. It is essential to characterize hate speech before deploying the technology to detect it using machine learning. When a correct definition is given, approaching the problem becomes easier. An automated method for categorizing hate speech on Twitter is used. Each tweet is assigned to one of four predetermined categories by the classifier: non-hate speech, racism, sexism, or both (racism and sexism). Four convolutional neural network models, respectively, were trained. Word vectors encompassing a variety of sources, such as those generated with word2vec, character 4-grams, randomly created word with vectors, and the fusion of word with vectors with attributes n-grams. The features of all the sets of the networks were streamlined through tweet classification, using max-pooling in combination with a SoftMax algorithm. The word2vec embedding model are found better than the others in the 10-fold cross-validation test with a percentage of 78.3% as F1-score and greater accuracy than recall [10].

According to Yoon Kim *et al.,* [11], their proposed model combines an LSTM network model with CNNs, and a very high network attributes (RNN-LM) for the model. The proposed method is equivalent even though it has 60% less parameters than the most recent state-of-the-art English Penn Treebank corpus. In languages with very rich morphology, the method presented in the study significantly surpasses the performance of the LSTM baselines at both the word and morpheme levels, despite utilizing a minimal number of parameters. By learning multi-stage features and using Lp pooling on the SVHN dataset, the traditional ConvNet architecture was enhanced and attained a new state-of-the-art architecture which has an accuracy of 94.85%. Additionally, they look at the benefits of different pooling techniques and multi-stage ConvNet features. A basic neural language model that just takes character-level inputs is provided by Martins *et al.,* [12], among others. Predictions at the word level are still made. The recent studies describe the problem related to hateful speech on the social platforms through feature engineering and machine learning. It highlights the lack of comparative research on feature generation and ML algorithms for standard datasets. One study using SVM found that features of bigram achieved the best outcomes at 79% [13]. The home number digit detection was carried out by Sermanet *et al.,* [14]. An automated hate speech detection on platforms like Twitter has been used, emphasizing the importance of promoting a diverse range of opinions while preventing hate speech. Their approach highlights different NLP approaches and ML Algorithms used to categorize hate speeches, addressing the growing challenge of hate speech proliferation in social media platforms [15].

Rahman *et al.,* [16] have utilized text mining and sentiment analysis techniques to identify ongoing social crises from Twitter data, with promising results. This approach, with an 89% to 98% identification rate for the top 5 crises, offers a valuable tool for monitoring and addressing societal challenges. The author has recognized the growing use of social platforms for people to express their emotions and opinions. Several studies have explored ways to detect users' mental health status by analysing their social media posts, using different algorithms and these of approaches, the Support Vector Classifier algorithm stands out with impressive accuracy of 79.90%, precision of 75.73%, recall of 77.53%, and F1-factor of 76.61%, paving the way for future intelligent systems focused on mental health detection [17]. Recent research investigates the connection between emotional states and comfort food choices among students, using ML models. Data from a survey of 526 students were analysed, indicating a very significant relationship between food, and emotional choices. Notably, the Naïve Bayes model displayed a promising accuracy of 96.66% [18]. Sanoussi *et al.,* [19] has explored the vital challenge of identifying hate speech on social platforms, particularly within the circumstances of the Canadian population, where cyberbullying poses significant social challenges. Using a dataset which has 14,000 comments from Facebook in a mix of local languages such as Canadian and French, the study employees NLP techniques and various word embedded methods. It then applies ML algorithms, achieving a high accuracy of 95.4% for detecting insulting comments and 93.9% for identifying hate speech.

An attentional multi-channel convolutional-BiLSTM network is proposed for automatic hate speech detection. The model leverages word embeddings, multi-channel convolution for semantic extraction, attention-aware BiLSTM for context capture, and achieves superior performance on benchmark datasets compared to state-of-the-art methods [20]. BiCHAT, a novel deep learning model that uses BiLSTMs, CNNs, and hierarchical attention to learn tweet representations for hate speech detection. BiCHAT achieves state-of-the-art performance on three benchmark datasets, outperforming previous methods by 8% [21]. The proposed model, HCovBi-Caps, uses convolutional layers, BiGRUs, and capsule networks to capture contextual information and improve hate speech detection accuracy [22].

3. Methodology

In the proposed method, five algorithms have been included as part of methodology. These algorithms are Linear Regression (LR), K-Nearest Neighbor (KNN), Random Forest (RF), Naïve Bayes (NB), and Decision Tree (DT). The purpose of the proposed method was to look for an algorithm that could identify hate speech with the highest degree of accuracy.

3.1 Data Collection

A significant quantity of Twitter data is required for the automated system to develop its tracking and hate speech recognition capabilities. The following Figure 1 represents the block diagram for the collection of data in this proposed research:

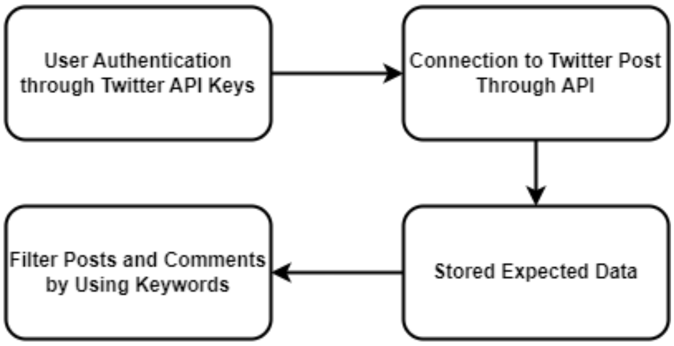

Fig. 1. Data Collection Block Diagram

For the proposed model, a total 24,783 instances of data have been gathered to implement the model. These mostly include hate speech and insulting language that is motivated by identity politics, sexual orientation, nationality, race, or religion. The dataset underwent a division in two distinct datasets: training and testing. For training data, 80% of the dataset was utilized, while 20% was reserved for testing. Leveraging this data, the machine learning system successfully classified the dataset into categories, distinguishing between hateful and non-hateful speech.

3.2 Import Libraries

Libraries in Python are a group of modules that may be used repeatedly in different projects without having to be written from start. However, every Python file ending in .py can be regarded as a module. Code that can be imported and used in other applications, such as functions, classes, and statements, is often found in such modules.

3.3 Import Datasets

To run python programmes for data analysis, datasets are required. A Python application may import external data in different form of file syntax all thanks to a number of modules that are included with Python. In this example, several types of data can be seen that have been imported into a Python programme. The dataset can be read by every row in the file using the CSV module, and a comma as a delimiter. Before the delimiter is given, the file is initially opened in read-only mode. A loop was used to read each element from the CSV file. Table 1 shows a few examples of imported datasets.

Table 1
Examples of tweets from the imported dataset

ID	Record	Hateful Speech	Obscene Language	None	Group	Tweet
0	4	1	1	6	4	!!! ST @mayslovely: Woman should not work...
1	4	1	6	1	2	!!!!! ST @mloow17: fat that hot...kill whs na...
2	4	1	6	1	2	!!!!!!! ST @BroadU Shit!!!! ST @67@chubby...
3	4	1	4	2	2	!!!!!!!!! ST @F_G_Anderson: @jhon_thoug she lo...
4	8	1	8	1	2	!!!!!!!!!!!!! ST @ShenikaRoberts: The fat shit you...

Table 2 shows significant information about many datasets, providing a quick summary of their properties and research importance.

Table 2
Data description of the imported dataset

ID	Column	Null/ non-null (0/1)	Count	Data Type
0	Unidentified: 0	0	24,783	int64
1	record	1	24,783	object
2	hateful_speech	0	24,783	int64
3	obscene_language	0	24,783	int64
4	none	0	24,783	int64
5	group	0	24,783	int64
6	tweet	0	24,783	object

Table 3 shows few examples of dataset descriptions.

Table 3
Few examples of Dataset Characteristic

	mean	mean	count	max	75%	50%	25%	min
Unidentified: 0	12681.192027	12681.192027	24783.0	25296	18995.5	12703	6372.5	0.0
record	3.243473	3.243473	24783.0	9.0	3.0	3.0	3.0	3.0
hateful_speech	0.280515	0.280515	24783.0	7.0	0.0	0.0	0.0	0.0
obscene_language	1.399459	1.399459	24783.0	9.0	3.0	3.0	2.0	0.0
none	1.113299	1.113299	24783.0	9.0	0.0	0.0	0.0	0.0
group	0.462089	0.462089	24783.0	2.0	1.0	1.0	1.0	0.0

3.4 Data Cleaning

The most important phase in the data analysis process is gathering, organizing, and preparing the data since inaccurate data might have long-term negative consequences if not handled properly. Data preparation, often known as data wrangling, is therefore required for accurate analysis. The goal of data preparation is to generate “clean text” that computers can read without making errors. Clean text is just human language that has been structured in a manner that computer models can understand. Simple Python code that eliminates stop words and Unicode words and breaks down complicated words to their base form may be used to tidy up the text.

Figure 2 represents a flow chart of the proposed method. It shows implementation of the dataset to different classifiers and their outcome.

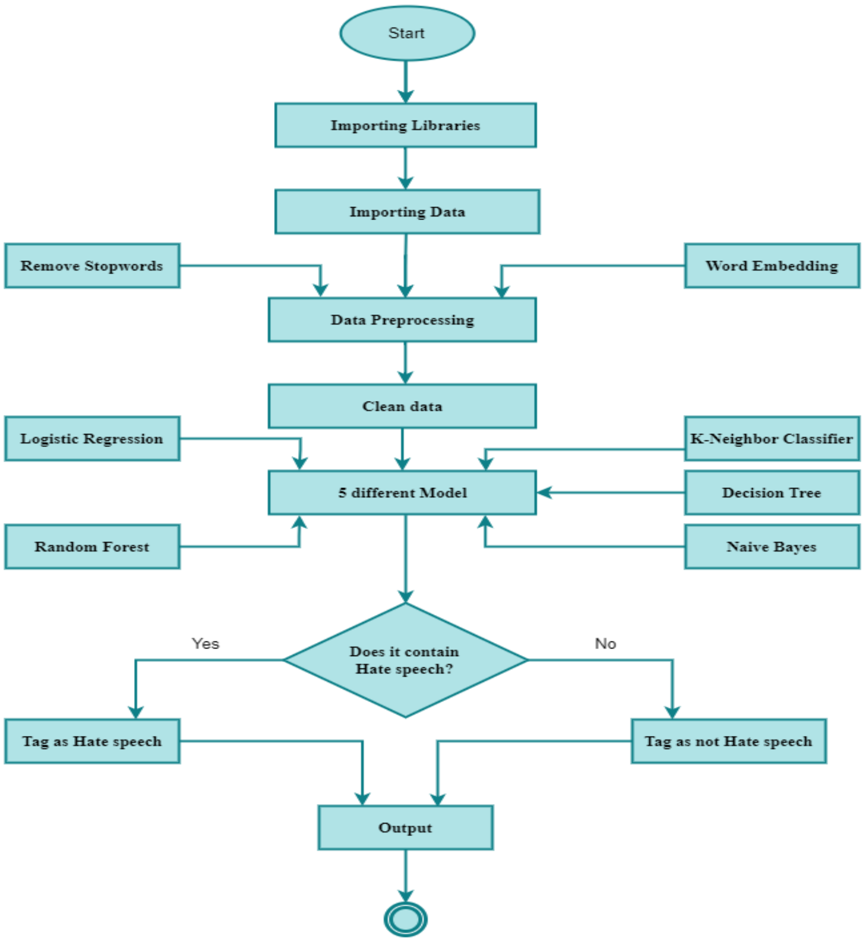

Fig. 2. The Proposed Model’s Flow Chart

3.4 Model Description
3.4.1 Naïve bayes

The well-known Bayes' theorem is the foundation of the Naïve Bayes classifier technique. This technique is more of a family of algorithms than a single approach, and they are all founded on the premise that each pair of characteristics being categorized is independent of the other.

3.4.2 Random forest

The mostly used ML technique known as random forest was developed by Leo Breiman and Adele Cutler, who combined the outcome of numerous decision trees to generate a single conclusion. Its popularity stems from its versatility and utility in resolving classification and regression problems.

3.4.3 Decision tree

For classification and regression applications, a non-parametric learning which is based on supervised learning is known as the decision tree. Its hierarchically arranged structure is made up of root, branches, internal, and leaf nodes.

3.4.4 K-nearest neighbor

KNN is the most fundamental and basic algorithm that is used in supervised learning in machine learning. K-NN algorithm classifies a new data point by comparing it to previously classified data. However, classification issues are where it is most usually utilized. It is a supervised learning method used to resolve both regression and classification problems.

3.4.5 Linear regression

A case model with just one independent variable is simple linear regression. The variable's dependency is determined using basic linear regression.

$$y = \beta 0 + \beta 1x + \varepsilon \tag{1}$$

where:

y = The response variable (Dependent)
x = The predictor variable (Independent)

In simple regression, the impact of independent variables is differentiated from the interaction of dependent variables.

4. Results

In the proposed method, the primary objective was to develop a system for classifying abusive and hateful language in Twitter data, employing NLP techniques with approaches such as a bag of words. The dataset initially used for this task was split into two different sets: the training dataset and the testing dataset, with an 80: 20 allocation ratios. This division allowed for the evaluation and validation of the model's performance. To accomplish this, five different ML algorithms were employed, and their analysis is provided below. Table 4 shows the Naïve Bayes model's accuracy and its results.

Table 4
The analysis report for Naïve Bayes

	0.0	1.0	Accuracy	Weighted_avg	Macro_avg
Support	4678	279	4957	4957	4957
F1-Score	0.58	0.12	0.44	0.56	0.35
Precision	0.96	0.07		0.91	0.51
Recall	0.42	0.68		0.44	0.55

Table 5 shows the Random Forest model's accuracy and its results.

Table 5
The analysis report for Random Forest

	0.0	1.0	Accuracy	Weighted_avg	Macro_avg
Support	4678	279	4957	4957	4957
F1-Score	0.97	0.30	0.94	0.93	0.63
Precision	0.97	0.42		0.93	0.69
Recall	0.98	0.24		0.94	0.61

Table 6 shows the Decision Tree model's accuracy and its results.

Table 6
The analysis report for Decision Tree

	0.0	1.0	Accuracy	Weighted_avg	Macro_avg
Support	4678	279	4957	4957	4957
F1-Score	0.96	0.31	0.92	0.92	0.63
Precision	0.98	0.29		0.93	0.62
Recall	0.95	0.33		0.92	0.64

Table 7 shows the K-Neighbor model's accuracy and its results.

Table 7
The analysis report for K-Nearest Neighbour

	0.0	1.0	Accuracy	Weighted_avg	Macro_avg
Support	4678	279	4957	4957	4957
F1-Score	0.58	0.12	0.44	0.56	0.35
Recall	0.42	0.68		0.44	0.55
Precision	0.98	0.07		0.91	0.51

Table 8 shows the accuracy of the Linear Regression model and its results.

Table 8
The analysis report for Linear Regression model

	0.0	1.0	Accuracy	Weighted_avg	Macro_avg
Support	4678	279	4957	4957	4957
F1-Score	0.99	0.25	0.94	0.94	0.61
Recall	0.98	0.18		0.95	0.58
Precision	0.94	0.44		0.93	0.70

Classification report of all the different algorithms have been given above. Next the process began by training each of these algorithms by using the training datasets. Subsequently, these algorithms which have the highest accuracy during training, were selected to further train and evaluate the test dataset on the model. The results of these evaluations were then recorded. Table 9 shows the analysis of these five algorithms that were implemented for better accuracy.

Table 9
Final Accuracy Analysis Results

Model	Accuracy
The Linear Regression (LR)	94%
The K-Nearest Neighbor (KNN)	93%
The Random Forest (RF)	93%
The Naïve Bayes (NB)	45%
The Decision Tree (DT)	91%

Initially, the Naïve Bayes algorithm was applied to the test data, yielding an accuracy rate of 45%. However, this was surpassed by the Decision Tree algorithm, which accomplished a remarkable accuracy of the 93%. Notably, both the Random Forest and KNN algorithms also exhibited the same accuracy rate of 93%. Of particular interest was the Linear Regression algorithm, which emerged as the most successful among all tested algorithms, achieving an accuracy rate of 94%. These findings are summarized, which provides a detailed, and organized analysis of the performance of these five implemented techniques, highlighting their respective accuracy rates.

In conclusion, this method demonstrated that the Linear Regression algorithm was the most effective in classifying abusive and hateful language in Twitter data, showcasing the potential of NLP techniques and machine learning in addressing online content moderation challenges.

5. Conclusions

The goal of this study was to employ these NLP techniques to identify hate speech, acknowledging the need for comprehending the most diverse nature of hate speech definitions from multiple platforms. While identifying hate speech remains difficult owing to human biases, the outcome discovered that deep learning models, especially CNNs, show promise when larger and higher-quality datasets are available. Future research recommendations include the need for contributions that make hate speech recognition simpler, more versatile, and user-friendly. Furthermore, examining misclassifications may give information on the difficulties of forecasting hate speech, and investigating particular phrases to distinguish offensive language from hate speech can offer to be a fascinating topic.

Further research should look at the many ways hate is conveyed on Twitter, such as direct targeting, group dialogues, and random outbursts. Examining the differences between these tactics and investigating unique user characteristics and motives in expressing hatred might give significant insights for dealing with hate speech in different social media platforms.

Acknowledgement
The University of Malaysia Perlis's Centre of Excellence for Advanced Computing (ADVCOMP), which provided a portion of the funding for this study.

References

[1] Jahan, Md Saroar, and Mourad Oussalah. "A systematic review of Hate Speech automatic detection using Natural Language Processing." *Neurocomputing* (2023): 126232. https://doi.org/10.1016/j.neucom.2023.126232

[2] Al-Makhadmeh, Zafer, and Amr Tolba. "Automatic hate speech detection using killer natural language processing optimizing ensemble deep learning approach." *Computing* 102 (2020): 501-522. https://doi.org/10.1007/s00607-019-00745-0

[3] Biere, Shanita, Sandjai Bhulai, and Master Business Analytics. "Hate speech detection using natural language processing techniques." *Master Business AnalyticsDepartment of Mathematics Faculty of Science* (2018).

[4] Davidson, Thomas, Dana Warmsley, Michael Macy, and Ingmar Weber. "Automated hate speech detection and the problem of offensive language." In *Proceedings of the international AAAI conference on web and social media*, vol. 11, no. 1, pp. 512-515. 2017. https://doi.org/10.1609/icwsm.v11i1.14955

[5] Del Vigna12, Fabio, Andrea Cimino23, Felice Dell'Orletta, Marinella Petrocchi, and Maurizio Tesconi. "Hate me, hate me not: Hate speech detection on facebook." In *Proceedings of the first Italian conference on cybersecurity (ITASEC17)*, pp. 86-95. 2017.

[6] Malmasi, Shervin, and Marcos Zampieri. "Detecting hate speech in social media." *arXiv preprint arXiv:1712.06427* (2017). https://doi.org/10.26615/978-954-452-049-6_062

[7] Xiang, Guang, Bin Fan, Ling Wang, Jason Hong, and Carolyn Rose. "Detecting offensive tweets via topical feature discovery over a large scale twitter corpus." In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp. 1980-1984. 2012. https://doi.org/10.1145/2396761.2398556

[8] Parihar, Anil Singh, Surendrabikram Thapa, and Sushruti Mishra. "Hate speech detection using natural language processing: Applications and challenges." In *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1302-1308. IEEE, 2021. https://doi.org/10.1109/ICOEI51242.2021.9452882

[9] Fortuna, Paula, and Sérgio Nunes. "A survey on automatic detection of hate speech in text." *ACM Computing Surveys (CSUR)* 51, no. 4 (2018): 1-30. https://doi.org/10.1145/3232676

[10] Gambäck, Björn, and Utpal Kumar Sikdar. "Using convolutional neural networks to classify hate-speech." In *Proceedings of the first workshop on abusive language online*, pp. 85-90. 2017. https://doi.org/10.18653/v1/W17-3013

[11] Kim, Yoon, Yacine Jernite, David Sontag, and Alexander Rush. "Character-aware neural language models." In *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, no. 1. 2016. https://doi.org/10.1609/aaai.v30i1.10362

[12] Martins, Ricardo, Marco Gomes, Jose Joao Almeida, Paulo Novais, and Pedro Henriques. "Hate speech classification in social media using emotional analysis." In *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 61-66. IEEE, 2018. https://doi.org/10.1109/BRACIS.2018.00019

[13] William, P., Ritik Gade, Rup esh Chaudhari, A. B. Pawar, and M. A. Jawale. "Machine learning based automatic hate speech recognition system." In *2022 International conference on sustainable computing and data communication systems (ICSCDS)*, pp. 315-318. IEEE, 2022. https://doi.org/10.1109/ICSCDS53736.2022.9760959

[14] Sermanet, Pierre, Soumith Chintala, and Yann LeCun. "Convolutional neural networks applied to house numbers digit classification." In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, pp. 3288-3291. IEEE, 2012.

[15] Pariyani, Bhavesh, Krish Shah, Meet Shah, Tarjni Vyas, and Sheshang Degadwala. "Hate speech detection in twitter using natural language processing." In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, pp. 1146-1152. IEEE, 2021. https://doi.org/10.1109/ICICV50876.2021.9388496

[16] Rahman, Shoaib, Nusrat Jahan, Farzana Sadia, and Imran Mahmud. "Social crisis detection using Twitter based text mining-a machine learning approach." *Bulletin of Electrical Engineering and Informatics* 12, no. 2 (2023): 1069-1077. https://doi.org/10.11591/eei.v12i2.3957

[17] Lia, Rifat Jahan, Abu Bakkar Siddikk, Fahim Muntasir, Sheikh Shah Mohammad Motiur Rahman, and Nusrat Jahan. "Depression detection from social media using Twitter's tweet." In *Big Data Intelligence for Smart Applications*, pp. 209-226. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-030-87954-9_9

[18] Nasrin, Nadia, Biraj Saha Aronya, Nusrat Jahan, Imran Mahmud, Afsana Begum, and Zahereel Ishwar Abdul Khalib. "Impact of Emotional State on Food Preference by Students: A Machine Learning Approach." In *International Conference on Applied Intelligence and Informatics*, pp. 523-538. Cham: Springer Nature Switzerland, 2022. https://doi.org/10.1007/978-3-031-24801-6_37

[19] Sanoussi, Mahamat Saleh Adoum, Chen Xiaohua, George K. Agordzo, Mahamed Lamine Guindo, Abdullah MMA Al Omari, and Boukhari Mahamat Issa. "Detection of Hate Speech Texts Using Machine Learning Algorithm." In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0266-0273. IEEE, 2022.

[20] Fazil, Mohd, Shakir Khan, Bader M. Albahlal, Reemiah Muneer Alotaibi, Tamanna Siddiqui, and Mohd Asif Shah. "Attentional multi-channel convolution with bidirectional LSTM cell toward hate speech prediction." *IEEE Access* 11 (2023): 16801-16811. https://doi.org/10.1109/ACCESS.2023.3246388

[21] Khan, Shakir, Mohd Fazil, Vineet Kumar Sejwal, Mohammed Ali Alshara, Reemiah Muneer Alotaibi, Ashraf Kamal, and Abdul Rauf Baig. "BiCHAT: BiLSTM with deep CNN and hierarchical attention for hate speech detection." *Journal of King Saud University-Computer and Information Sciences* 34, no. 7 (2022): 4335-4344. https://doi.org/10.1016/j.jksuci.2022.05.006

[22] Khan, Shakir, Ashraf Kamal, Mohd Fazil, Mohammed Ali Alshara, Vineet Kumar Sejwal, Reemiah Muneer Alotaibi, Abdul Rauf Baig, and Salihah Alqahtani. "HCovBi-caps: hate speech detection using convolutional and Bi-directional gated recurrent unit with Capsule network." *IEEE Access* 10 (2022): 7881-7894. https://doi.org/10.1109/ACCESS.2022.3143799