Advances and Applications in Mathematical Sciences
Volume 21, Issue 8, June 2022, Pages 4645-4652
© 2022 Mili Publications, India

A STUDY OF HATE SPEECH DETECTION IN SOCIAL MEDIA FORSENTIMENT ANALYSIS

N. SOLOMON PRAVEEN KUMAR[1] and M. S. MYTHILI[2]

[1]Research Scholar, [2]Associate Professor
Department of Computer Science
Bishop Heber College (Autonomous)
Affiliated to Bharathidasan University
Tiruchirappalli – 620017, Tamil Nadu, India
E-mail: solomon@bhc.edu.in
mythili.ca@bhc.edu.in

Abstract

Due to the advancement in technology and the explosion of the information age, people communicate with each other indirectly via using online social networks (OSNs), similar as Facebook, Snapchat, Instagram, and Twitter. Users of OSNs have the ability to post whatever they wish without any sort of control. This leads to the spread of hateful and offensive content, thereby increasing crimes, murders, and terrorism among users. Hate Speech Detection (HSD) using sentiment analysis is the objective of this article. Emotions in humans can either be positive or negative. Despite this, there are still more categories such as joy, sorrow, disgust, surprise, depression, frustration, anger, fear, confidence, trust, anticipation, shame, kindness, love, friendship, faith, and wonder. It lays the groundwork for methodologies for computational analysis in these areas, such as Natural Language Processing and Machine Learning algorithms. A variety of sentiment analysis methodologies and algorithms are outlined and compared, as well as its limitations and possible future directions. The purpose of this paper is to analyze and compare algorithms, approaches, and features used in machine learning to compute sentiment classification where Support Vector Machine has better accuracy, precision and recall compared to Naive Bayes, Logistic Regression, Random Forest, Ada Boost and Neural Network.

1. Introduction

The exponential growth of data on the web sphere accelerated the need of

2020 Mathematics Subject Classification: 03B70.

Keywords: Sentiment Analysis, Natural Language Processing, Machine Learning, Online Social Networks, Support Vector Machine.

Received March 24, 2022; Accepted April 16, 2022

extracting meaningful information from it. This information can be used for better decision making. The automatic generation of sentiments from the text is called as sentiment analysis (SA). It is a collaborative process of natural language processing and data mining. [1] Sentiment Analysis is a task of Natural Language Processing (NLP) that aims to extract sentiments and opinions from texts [2, 3]. Sentiment Analysis makes use of three terms in order to fetch the sentiment. That is object and feature, opinion holder, opinion and orientation. Sentiment Analysis deals with several technical challenges such as object identification, opinion orientation classification, and feature extraction. Usually sentiment analysis can be performed using supervised and unsupervised learning such as naïve Bayes, Neural Networks, and Support Vector Machine. Among these three techniques SVM is considered to be more suitable for sentiment Analysis. Sentiment classification can be performed in three stages such as Document level, Sentence level, and Feature level.

In document and sentence level the sentiment analysis make use of only a single object and extracts only a single opinion from the single opinion holder. But this type of assumptions is not suitable for many situations. Extracting sentiment for entire document/blog will not be efficient as extracting sentiment by considering aspects of each subject in the particular sentence. Every minute of the day, a tremendous amount of data is generated by social media networks. Social media like YouTube, Facebook, Twitter, LinkedIn, WhatsApp, Reddit, or any product website are available online around the globe. When people share their thoughts through social media, they express their emotions directly or indirectly. The process of analyzing this expression is called Sentiment Analysis [4]. Abusive and offensive language is the prime concern of technical companies now a days due to exponential growth in number of Internet users around the world and since these people are from different walks of life and different culture. There is a fine line between hate speech and offensive language, and to detect and differentiate among them is a big challenge. [6]

In related work, researchers generally classify the text into three classes such as Hateful, Offensive, and Clean

Definition of hate speech

According to Paula Fortuna and Sergia Nunes "Hate speech is language that attacks or diminishes, that incites violence or hate against groups, based on specific characteristics such as physical appearance, religion, descent, national or ethnic origin, sexual orientation, gender identity or other, and it can occur with different linguistic styles, even in subtle forms or when humour is used "[7].

2. Related Work

Natural Language Processing is playing huge role in detecting hate speech. It is very efficient to detect hate speech with NLP.

1. Shoven Ahammed et al. [2019] have used three major steps to detect hate speech in Bangla Language. Since the Bangla dataset is not available. The author has collected data from Facebook. The steps are formatting the dataset. The author selected hate speech from Facebook and incorporated it into machine learning. A maximum F1 score of 0.71 was obtained via the SVM algorithm and 0.73 was obtained via the Naive Bayes algorithm [8].

2. The study by Sean MacAvaney et al. [2019] identifies and explores the challenges faced by online automatic approaches to detect hate speech in text. The author sums up leading hate speech from a variety of sources. The authors propose a multi-view SVM as a simpler and more understandable alternative to neural methods [9].

3. Amita Jain et al. have proposed to analyze large-sized text. Senti-NSetPSO comprises of two classifiers: binary and ternary based on hybridization of Particle Swarm Optimization (PSO) with Neutrosophic Set Senti-NSetPSO has been tested on large documents having more than 25kb in size. The proposed work reached 81.99% for the ternary classifiers and 95.3% for the binary classifiers. The author has collected Dataset from the Blitzer data-set [10].

4. Zafer Al-Makha et al. have developed a method for predicting hate speech from Twitter using a hybrid of natural language processing and machine learning techniques. Data is analyzed with the use of Killer Natural Language Processing Optimization Ensemble Deep Neural Network Learning

Approach (KNLPEDNN), which achieves a maximum prediction accuracy of 98.71% [11].

5. Yanling Zhore et al. have applied several famous machine learning methods for text classification such as Embeddings from Language Models (ELMo) and Bidirectional Encoder Representation from Transformers (BERT) to the data sets of SemEval 2019. A deep learning-based fusion approach is employed by the author for hate speech detection. According to the results, classification accuracy and F1-Score have significantly improved [12].

6. Oluwafemi Oriola et al. developed an English corpus from South African tweets and evaluated different machine learning techniques to detect offensive and hate speech. A character n-gram, word *n*-gram, negative sentiment, syntactic feature and their hybrid were extracted and examined using hyper-parameter optimization, ensemble and multi-tier meta-learning models of Support Vector Machine, logistic regression, random forest, and gradient boosting. According to the results of the experiment, Support Vector Machine, Random Forest, and Gradient Boosting meta-learning models were most consistent and balanced in the detection of offensive and hate speech, with true positive rates of 0.887 and 0.858 and an overall accuracy of 0.671 [13].

7. R. Srinivasan et al. have focussed on sentiment analysis with imbalanced class label distribution. The author has also focuses on “Code mixing”. Code mixed data consists of text alternating between two or more languages. The proposed work compares the performances of various machine learning approaches namely, Random Forest Classifier, Logistic Regression, XGBoost Classifier, Support Vector Machine and Naive Bayes Classifier using F1-Score [14].

8. According to Omar Sharif et al., an automated system can spot offensive text when it is mixed with multilingual code. The author employed two machine learning algorithms (LR, SVM), three deep learning algorithms (LSTM, LSTM + Attention), and three sets of transformers (m-BERT, Indic-BERT, XLM-R) for the tasks at hand. The proposed models achieved a weighed F1-Score of 0.76 [15].

9. In his study, Md Rabiul Awal et al. used a supervised approach that relied heavily on the annotated hate speech datasets, which are imbalanced

and often lacking in training samples for hateful content. A novel multitasks learning-based model, AngryBERT, is proposed to fill research gaps by simultaneously learning hate speech detection with sentiment classification and target identification. The AngryBERTcan accurately detect hate speech, identify its target and determine the emotion expressed [17].

10. Ishan Sanjeev Upadhyay et al. have experimented with two approaches. A first approach used contextual embeddings to train classifiers based on Logistic Regression, Random Forest, SVM, and LSTM. The second approach involved building a majority-voting ensemble of 11 models from fine-tunings of pre-trained transformer models (BERT, AL-BERT, RoBERTa, IndicBERT) with an output layer. The second approach was more efficient for English, Tamil, and Malayalam with weighted F1-Scores of 0.93, 0.75, and 0.49 for English, Tamil, and Malayalam, respectively [18].

4. Comparison Chart

Comparative table for different classification algorithms

Table 1. Comparison of different algorithms.

Algorithms or Hate Speech Detection Methods	Classification Accuracy	Precision	Recall	F1-Score
Support Vector Machine (SVM)-RBF	0.8581	0.8586	0.8581	0.858
Naïve Bayes	0.72	0.7	0.74	0.72
Logistic Regression	0.8768	0.8	0.82	0.81
Random Forest (RF)	0.722	0.711	0.722	0.713
Ada Boost (Adaptive Booster)	0.708	0.697	0.6708	0.701
Artificial Neural Network	0.596	0.548	0.596	0.549

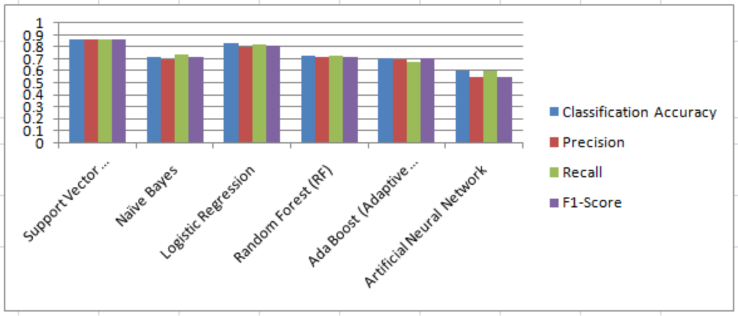

Figure 1. Comparison chart of different algorithms.

Figure 1 The results of the comparison of measurements of classification accuracy, precision, recall and F1 from cases of hate speech are shown in the Graph above where Support Vector Machine has better accuracy, precision and recall compared to Naive Bayes, Logistic Regression, Random Forest, Ada Boost and Neural Network.

5. Conclusion

Sentiment analysis (or) opinion mining play's significant role hate speech detection in social media. This study employed automated text classification techniques to detect hate speech messages. Moreover, this study discussed various machine learning and deep learning algorithms to classify hate speech detection. The objective of this research is to provide an overview of sentiment analysis from textual data, speech, and visual data, as well as to examine and compare methods, approaches, and features used in machine learning to compute sentiment classification where Support Vector Machine provides higher accuracy, precision, and recall than Naive Bayes, Logistic Regression, Random Forest, Ada Boost, and Neural Networks.

References

[1] S. Kaur and R. Mohana, A roadmap of sentiment analysis and its research directions, International Journal of Knowledge and Learning 10(3) (2015), 296-323.

[2] Liu, Bing, Sentiment analysis and opinion mining synthesis lectures on human language technologies 5(1) (2012), 1-167.

[3] I. Chaturvedi, E. Cambria, R. E. Welsch, F. Herrera, Distinguishing between facts and opinions for sentiment analysis, Survey and challenges, Inf. Fusion 44 (2018), 65-77. https://doi.org/10.1016/j.inffus.2017.12.006.

[4] V. Ahire and S. Borse, Emotion detection from social media using machine learning techniques: A Survey, In Applied Information Processing Systems Springer, Singapore (2022), 83-92.

[5] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed and M. T. Sadiq, Automatic detection of offensive language for urdu and roman urdu IEEE Access 8 (2020), 91213-91226.

[6] R. Pradhan, A. Chaturvedi, A. Tripathi and D. K. Sharma, A review on offensive language detection, In Advances in Data and Information Sciences Springer, Singapore (2020), 433-439.

[7] Paula Fortuna and Sèrgio Nunes, A survey on automatic detection of hate speech in text, ACM Computing Surveys (CSUR) 51(4) (2018), 1-30.

[8] Njagi Dennis Gitari and Zhang Zuping, Hanyurwimfura Damien, and Jun Long, A lexicon-based approach for hate speech detection. International Journal of Multimedia and Ubiquitous Engineering 10(4) (2015), 215-230.

[9] M. R. Awal, R. Cao, R. K. W. Lee and S. Mitrovic, AngryBERT: Joint Learning Target and Emotion for Hate Speech Detection, (2021). arXiv preprint arXiv:2103.11800

[10] Z. Al-Makhadmeh and A. Tolba, Automatic hate speech detection using killer natural language processing optimizing ensemble deep learning approach, Computing 102(2) (2020), 501-522.

[11] Y. Zhou, Y. Yang, H. Liu, X. Liu and N. Savage, Deep learning based fusion approach for hate speech detection, IEEE Access 8 (2020), 128923-128929.

[12] O. Oriola and E. Kotzé, Evaluating machine learning techniques for detecting offensive and hate speech in South African tweets, IEEE Access 8 (2020), 21496-21509.

[13] S. MacAvaney, H. R. Yao, E. Yang, K. Russell, N. Goharian and O. Frieder, Hate speech detection: Challenges and solutions, PloS one, 14(8) (2019), e0221152.

[14] I. S. Upadhyay, A. Wadhawan and R. Mamidi, Hopeful_Men@ LT-EDI-EACL2021: hope speech detection using Indic transliteration and transformers, (2021). arXiv preprint arXiv:2102.12082.

[15] S. Ahammed, M. Rahman, M. H. Niloy and S. M. H. Chowdhury, (Implementation of machine learning to detect hate speech in Bangla language, In 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART) IEEE 2019, 317-320.

[16] G. L. De La Peña Sarracén, Multilingual and multimodal hate speech analysis in twitter, In Proceedings of the 14th ACM International Conference on Web Search and Data Mining (2021), 1109-1110.

[17] O. Sharif, E. Hossain and M. M. Hoque, NLP-CUET@ DravidianLangTech-EACL2021: Offensive Language Detection from Multilingual Code-Mixed Text using Transformers, (2021). arXiv preprint arXiv:2103.00455

[18] R. Srinivasan and C. N. Subalalitha, Sentimental analysis from imbalanced code-mixed data using machine learning approaches, Distributed and Parallel Databases (2021), 1-16.

[19] Z. Fei, Z. Li, J. Zhang, Y. Feng and J. Zhou, Towards Expressive Communication with Internet Memes: A New Multimodal Conversation Dataset and Benchmark, (2021). arXiv preprint arXiv:2109.01839

[20] P. Jayasuriya, S. Ekanayake, R. Munasinghe, B. Kumarasinghe, I. Weerasinghe and S. Thelijjagoda, Sentiment classification of Sinhala content in social media, In 2020 International Research Conference on Smart Computing and Systems Engineering (SCSE) IEEE (2020), 136-141.