International Journal of Research in Engineering, Science and Management
Volume 5, Issue 7, July 2022
https://www.ijresm.com | ISSN (Online): 2581-5792

Hate and Sarcasm Classification Using Machine Learning and Deep Learning Techniques: A Survey

Preethi[1*], Radhakrishna Dodmane[2]
[1]Student, Department of Computer Science and Engineering, N.M.A.M. Institute of Technology, Nitte, Karkala, India
[2]Associate Professor, Department of Computer Science and Engineering, N.M.A.M. Institute of Technology, Nitte, Karkala, India

***Abstract*: This survey paper confers a study on hateful and sarcasm classification techniques. The importance of hateful and sarcasm classification in social media is due to, the amount of online hate speech, and sarcasm speech growing day by day, and it affects human emotions. For this reason, hate and sarcasm speech has become a real problem in modern society and this needs to be controlled by using classification techniques to detect hate and sarcasm. The proposed survey describes the comparative study on datasets available, data pre-processing techniques used, and methods used for classification, which is used to automatically recognize hateful and sarcasm detection using both machine learning and deep learning techniques. The findings mainly addressed the Sentiment Based Features, Semantic-Based Features, N-grams, and Pattern-Based Features, which are most frequently and preferably used by the researchers in this field. Also, relative analysis on different machine learning and deep learning methods used for hate and sarcasm detection is analyzed in this work which helps the researchers in this field to choose better models for the specific data sets used.**

***Keywords*: Twitter, Hate speech, Sarcasm speech, Classification, Machine learning, Deep learning.**

1. Introduction

Online social networks and websites have become increasingly popular in recent years, and individuals have begun to use them extensively. The people attract more in these areas where they can get to post their day-to-day activities while using stories, posts, and tweets by using social media like Twitter, Instagram, and Facebook. It is found that Twitter had 187 million monetizable daily active users and Instagram had roughly one billion monthly active users worldwide. These sites have given the right to create a profile for the users and get connected with people and friends. It has become easy to keep in touch and share information by publishing their activities, and interests by using tweets, photos, and videos. These social media applications have also given the right to comment and people started misusing this feature of social media by using trolling, hateful and sarcastic messages everywhere. It has become most trending on the internet and people are expressing opinions by using tweets and posts. It is easy to comment and make a tweet by spreading negative thoughts by using hateful and sarcastic messages. Nowadays, the amount of social media users is increasing at a very high rate. Services such as those offered by Twitter, Facebook, and Instagram offer the flexibility to make profiles to possess a listing of peers to act with and to post and browse what others have denoted. Their contents area unit is rapidly growing, constituting an attention-grabbing example of the supposed massive information. Cyberbullying, sexual predation, self-harm practices incitement area unit a number of the effective results of the dissemination of malicious data on Social Network Sites(SNSs). The target of the trolls area unit typically selected victims. However, in some circumstances, the hate may be directed towards wide teams of people, discriminated for a few options, like race or gender. Such campaigns might involve a sizable amount of haters in that area unit self-excited by hateful and critical discussions. Such hate and wittiness would possibly find you with physical violence or violent actions. Thus sexists, non-secular fanatics, political extremists massively use SNSs to foster hate against specific individuals/organizations, by inflicting a sounding board result, which can critically injure the targets of the hate campaign, by mistreatment each psychological and physical violence.

Through all the datasets collected and performed a study from the previous researchers, we discovered that most microblogging social media, specifically, Twitter is a frequently used dataset. The reason Twitter is frequently used because of its distinguishable properties compared to another kind of dataset. According to Fabio del vigna [3], the Facebook dataset is comparatively less frequently used and it is used only in the detection of hate but not in sarcasm. Hence Facebook dataset is considered as not supporting dataset for both hate and sarcasm detection. The dataset is also publicly available on Crowdflower Hajime Watanabe [1]. In this paper, we use the terms hate and sarcasm speech. We decided in the favour of using this term since it is considered as a term for different kinds of insulting messages. Hate and sarcasm speech is also the most frequently used expressions for this fact and is even legal term in several countries.

To detect this hateful and sarcastic speech, this paper provides a brief overview of the datasets, data preprocessing techniques, and methods that are most frequently and suitable for detecting hate and sarcasm. Since the many authors worked on this area by using different datasets and techniques, in this paper we highlighted mainly types of feature extractions that are suitable for other classification methods. Identified

*Corresponding author: preethig67@gmail.com

commonly used data preprocessing techniques by referring to several papers, and also considered the most effective feature extraction techniques those are namely Sentiment Based Features, Semantic-Based Features, N-gram, Pattern Features. Also specified about Facebook and Twitter dataset, which were commonly used in hate and sarcasm detection. We Worked on a qualified review of machine learning and deep learning methods used in the detection of hate and sarcasm with representation by using the graphical chart. Section II discusses the feature extraction for data pre-processing, section III discusses the classification methods, and section IV discusses the results and future work.

2. Feature Extraction for Data Pre-Processing

Based on the survey, Figure 1 shows the frequently used features in hateful and sarcasm detection methods. The N-gram is the most frequently used feature in both hateful and sarcasm detection. The PoS-tagger is also used effectively in hateful and sarcasm compared to other features. As the study shows that the Bag-of-Word is used preferably in the detection of hate speech than sarcastic detection. The following section discusses the various types of feature extraction techniques used for data pre-processing.

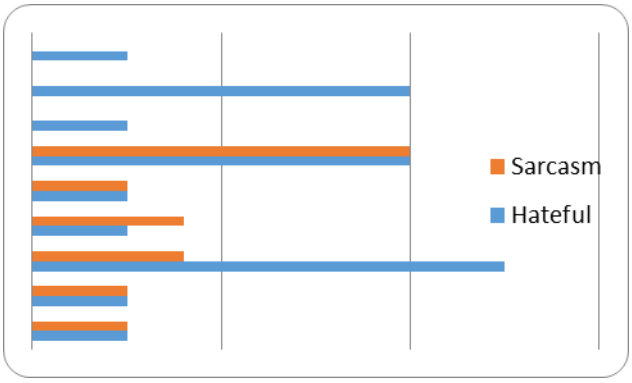

Fig. 1. Commonly used features in hateful and sarcasm detection

A. Sentiment-Based Features

The most basic features used to detect hatred and sarcasm speech are sentiment-based characteristics. This is because “negative” tweets are more likely to contain hate and sarcasm than “positive” tweets. This feature is used to extract the features to determine if the tweet is positive, negative, or neutral [1]. The SentiStrength tool is used to extract positive words and negative words [1], [2]. In the research paper [2], the authors thought of reasonable inconsistency between sentiments of words yet as alternative elements inside the tweet. Owing to adjectives, verbs and adverbs have higher emotional content than nouns [9] thus positive and negative words are connected to PoS-tag. Axel Rodríguez [17] considered the Sentiment feature as one of the powerful filters to remove the unrelated texts, also help to identify and categorize the opinions expressed in a piece of text determining towards a particular topic is positive, negative, or neutral.

B. Semantic-Based Features

The usage of punctuation capitalized words, and interjections by internet users are shown through semantic-based characteristics. Considering the number of exclamation marks, question marks, full stop marks, all-capitalized words, quotes, interjections, happy expressions, words within the tweet is useful to discover the hate speech [1]. The punctuation-related options and customary expressions area unit thought-about within the critical issue, correlating these expressions with punctuation is feasible to decide whether what is said is sarcastic or not [9]. In the detection of sarcasm, the noisy set of tweets are used with the hashtag, and then the patterns are extracted which are appeared more than ten times. Njagi Dennis Gitari [16] worked on generating the lexicon of sentiment expressions victimization linguistics and sound judgment options. Later these options are used to produce a classifier for the detection of hate speech. To build a lexicon of hate-connected words, authors used a rule-based technique and subjective options known from the sentences, and the linguistics options are learned directly from the corpus. Ilham Maulana Ahmad Niam [18] used enhanced extension appropriate improvements to identify emotions expressed through text, namely Latent Semantic Analysis (LSA), which is a statistical and mathematical method used to extract the meaningful condition words and to identify the relationships between words in the text. They used LSA on detecting the hate speech present in the image.

C. N-grams

An adjoining sequence of n items from a given sample set of speech is named as N-gram. For improved performance using n-gram and combining them with additional features further provides better results. Indeed, the unigrams and other n-grams are included in the feature sets by the majority of researchers, namely, Hajime Watanabe [1], Ika Alfina [2], Fabio del vigna [3], Tomas Ptacek [7], Mondher Bonazizi [9], Shubhodip Saha [10], Gamback [11], Badjatiya [12]. The bag of words (BOW) model is used to represent the text in three classes of features, namely, word n-gram, character n-gram, and negative sentiment and also have five features such a word unigram, word bigram in word n-gram and character trigram, character quadrigram in character n-gram and negative sentiment are used, the word n-gram outperformed character n-gram in hate speech detection [2]. Authors of research article [1] have used unigrams that have a part-of-speech (PoS) tag of noun, verb, the adjective is extracted from the training set and stored with a minimal number of occurrences. The result shows that using unigrams in binary classification presents lower frequency while performing hate speech detection. Using character n-gram with the minimum prevalence and character feature set containing 3-grams to 6-grams, and in n-grams using unigrams, bigrams, trigrams as binary options, conjointly using skip-grams to neglect absolute gaps and the result shows that using unigrams and bigrams as features achieves average performance in sarcasm speech detection [7]. Concatenating word2vec word and character n-grams in feature embedding improves the performance in hate speech detection [11].

When using TF-IDF and character n-gram method for feature extraction in the detection of hate speech, TF-IDF works better

than the character n-gram method [12]. Authors have used character n-gram, token n-gram, token unigram, and skip-grams on Amateur annotations and Expert annotations in the detection of hate speech on tweets [13]. Authors of the research article used items as part of the speech of the words employed in the sentences and also used trigrams of a part of speech as options for their model, they conjointly combined to perform words and a part of speech n-grams and used them as a single feature for classification, later these options completely captured the fashion-based mostly options [20].

D. Pattern Features

Pattern features are features that are extracted in the same way as unigrams. Dividing the words to tweet into two groups depending on the sentimental or non-sentimental word, later these are representing by using the PoS tag, and vectors are extracted from different tweets from different lengths and for binary classification and ternary classification pattern extraction are used [1]. The high-frequency words and content words in a pattern can be taken with the minimum prevalence, and enhancing the pattern options by exploiting the word shape also possible, Comparing to other features, the pattern and word-shaped pattern results in improved performance [7]. The syntactic and semantic features using in pattern-related features for extraction and pattern-related features are significantly high during cross-validation [9]. Authors of research article [19], the main feature type used based on surface patterns, and classified patterns as an ordered sequence of high-frequency words and slots for content words, while doing pattern extraction, provides hundreds of patterns in that some are too general or too specific, two criteria used to reduce the feature space, the first is to remove all patterns appearing in sentences originating from a single product (Amazon) and second is to remove all patterns appears in clearly sarcastic and in not sarcastic.

The author of the research article [1] has worked on collecting hateful and offensive expressions for detecting hate speech using Twitter datasets. A preliminary study on detecting hate speech in the Indonesian language using the Twitter dataset, collecting the tweets using Twitter Streaming API done by [2], the collected tweets were related to a political event and Jakarta Governor Election 2017. Considering the Facebook comments which are appeared on the set of public Italian pages are possible to detect hate speech [3]. Using different sets of features in data pre-processing techniques to detect the sarcastic on the Twitter dataset is possible [5],[7],[9],[10]. Many authors in the research article worked on Twitter datasets to detect the hate speech in tweets [11],[12],[14].

The survey based on the hate speech detection using Natural Language Processing (NLP) on Twitter datasets, mentioned that the generic features, such as a bag of words, embeddings, perform well in classifications, and also Character-level approaches work better than token-level approaches, they also identified some of the complex features are evaluated on the individual dataset; most of these are not publicly available such as considering bullying as ethnic minorities [4]. The study on different Machine Learning algorithms shows that machine learning algorithms are frequently used in sarcasm detection on Twitter datasets and also identified that Customized Machine Learning algorithms are not suitable for sarcasm detection [6]. The survey was conducted on comparing the commonly used features with datasets and also comparing the sarcasm classification with Machine Learning approaches such as supervised learning, semi-supervised learning, structured learning, hybrid approach, neural network, and Rule-Based approaches [8].

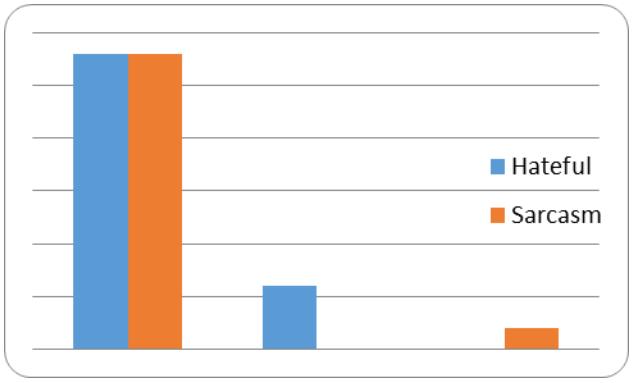

Fig. 2. Commonly used dataset in hateful and sarcasm detection

Figure 2 clearly shows that the Twitter dataset is a commonly used dataset for both Hateful and Sarcasm speech whereas the Facebook dataset is used comparatively less but better than Amazon. The authors of the research article [19] used the Amazon dataset in their work, they stated that Amazon reviews are much longer than tweets and they are more structured and grammatical, comes in a specific product. The unique concept of working on recognizing the image to detect whether it contains hate speech or not, the process of detecting hate speech on image done by translating the image into text and then doing the image selection process by cutting the only image area containing the text, this will reduce the time taking for the process of reading Optical Character Recognition [18]. Authors of the research article [21] collected over 872,428 Twitter profiles out of these 412,716 tweets that contained negative texts, grouped based on tweets and identified majority of tweets (60.4%) are on insulting women. The survey work briefly describes the Short Text, Long text, Transcripts & Dialogue, and Miscellaneous text forms [22].

3. Classification Methods

The algorithms are classified as follows: The approaches used in the research article organized into two specific areas, those are hate and sarcasm classification methods. In both, the area machine learning and deep learning methods are used. Based on the survey, found that machine learning methods are commonly used and more suitable for the detection of hate and sarcasm speech. Many authors are preferred to use machine learning because of the performances in hate and sarcasm speech detection and also some of the authors are suggested combining the machine and deep learning approaches for the better result.

A. Hate Detection

In the study, we found that many authors have tried different approaches in machine learning and deep learning in the detection of hate speech than sarcasm speech. The result which they got is considered acceptable. The Support Vector Machine (SVM), Random Forest Decision Tree (RFDT), J48 graft classifier, Naïve Bayes (NB), Logistic Regression (LR), Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), Gradient Boosted Decision Tree (GBDT) are commonly used in the detection of hate speech on Twitter and Facebook data. Bellow Figure 3 shows commonly used classification methods used by the authors in the detection of hate speech.

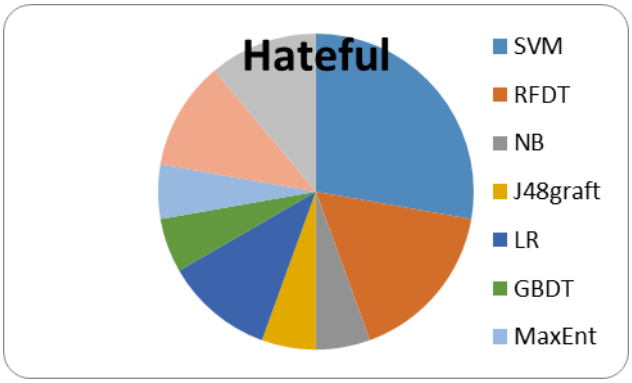

Fig. 3. Commonly used classification methods in hateful detection

The authors of the research article [1] have used binary and ternary classification, in the binary classification they combined "hateful" and "offensive" under one class and referred to as "offensive" and the other class is "clean". And in the ternary classification, they split into three classes namely, "hateful", "offensive" and "clean". For the classification, they have used "J48graft". The author worked on the Indonesian language done by [2] on Using the Twitter datasets of Indonesian language to detect the hate speech, by using a supervised learning approach such as NB, SVM, Bayesian Logistic Regression, and RFDT, in their work found that using different feature RFDT performs superior than the other methods. The Facebook data used to detect hate speech, and the classification is done by using two different learning algorithms, the two classifications of three categories contain Strong, Weak, and No hate, and in two categories contains Strong, Weak hate, Conducting a three-class experiment the result of using SVM and LSTM was not able to discriminate between three classes this is because of using a small number of Strong hate documents and a low level of annotator agreement, wherein two-class experiment the LSTM outperforms SVM [3]. The hate speech identification system architecture used CNN and LR, in each tweet in the dataset annotated by one Expert annotator and three Amateur annotators, when CNN applied on random vectors, word2vec, character n-grams, and word2vec with character n-grams, the precision and F1-score were better compared to other, and in LR with character n-grams recall was better than CNN [11]. The combination of LSTM classifier + Random Embedding + GBDT, were used in tweet embeddings and initialized to random vectors and back-propagation is used to train the LSTM [12], it is found that one of the best methods used in their work.

Machine learning algorithms are the most commonly used algorithms in the detection of hate detection, hence these types of algorithms are more suitable for the detection of hate speech. From bellow, Figure 4 clearly shows that SVM is the most frequently used algorithm comparing to other machine learning algorithms, The machine learning algorithms such as SVM, RFDT, and J48 graft classifier is used to perform hate detection after the parameters are optimized and re-running the classification again on the validation set, J48 graft classifier outperformed other classifiers, this is because the optimal value of confidence threshold for pruning(C) C = 0.04 and SVM deals with numeric features but whereas "J48graft" deals with binary features. [1]. Authors of the research article [2] have used Naïve Bayes, SVM, Logistic Regression, and RFDT classifier, in their work first they tried these algorithms with five different features separately, later they tried on all five features altogether. ANN is used to implement a mobile application for the detection of hate speech in Albanian language [30]. The two machine learning methods namely RFDT and NB used in the detection of hate, NB results better than RFDT [25]. Combined machine learning algorithms are used in [29], various popular classifiers like NB, ME, KNN, RF, and SVM are combined and applied on both unbalanced and balanced twitter datasets to detect hate speech.

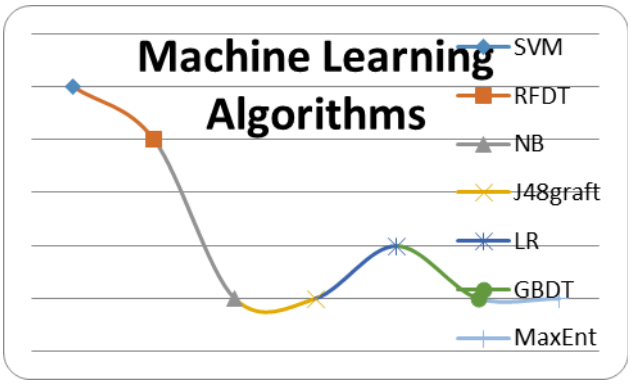

Fig. 4. Commonly used machine learning algorithms in hate detection

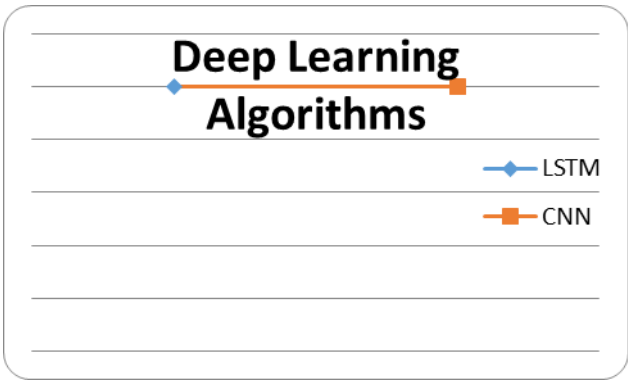

Fig. 5. Deep learning algorithms in hate detection

Fig. 5 shows that the authors of the research article have used deep learning algorithms in detecting hate. The Convolutional Neural Network, LSTM are the deep learning algorithms used

in detecting hate speech. The deep learning for Twitter hate speech text classification, created two CNN models based on different input vector sets, later that was used to fed neural networks, this classifier worked by assigning the hate tweets to four different predefined categories [11]. The tweets are categorized into racism, sexism, and neutral and performed a deep neural model on it [27]. Authors of research articles have used tweets to detect hate speech by using the ensemble deep learning approach [28].

B. Sarcasm Detection

Based on the research done by many authors, on sarcastic detection, using Twitter data for the detection of sarcasm speech and have used commonly machine learning algorithms than deep learning algorithms. Supervised learning is most preferred by the authors. In this survey, we found that SVM, RFDT, NB, K-NN, RNN, Maximum Entropy algorithms used for the detection of sarcasm speech on Twitter. Bellow Fig. 6 shows, how frequently these methods are used in the detection of sarcasm speech.

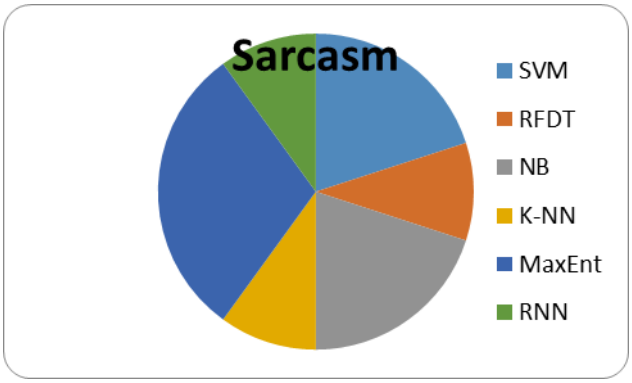

Fig. 6. Commonly used classification methods in sarcasm detection

The strong discrete baseline model built by using features from both target tweets and contextual tweets, later two local and contextual components are used in the neural network model to extract the dense real-valued features from the local and history tweets, using only local tweet features in discrete model achieves better performance than the neural model but using also context tweet features in the neural model the accuracy goes up in detection of sarcasm speech [5]. The sarcasm speech detection done on Czech language and English tweets, the results on the combination of the majority of options with Maximum Entropy outperforms SVM classifier on each balanced and unbalanced dataset distributions, additionally found best result on Czech dataset was achieved by the SVM classifier [7]. The sarcastic classification done on Twitter datasets used machine learning algorithms such as RFDT, SVM, K-NN, Maximum Entropy comparing these classifiers based on the result found that SVM only detecting one out of five sarcastic tweets and RFDT outperforms the other three classifiers [9]. The proposed approach to collect data from Twitter Archiver and aimed to classify sarcastic tweets as positive, negative, and neutral and performed by using NB and SVM classifiers, the most common tool RapidMiner and TextBlob, used for finding the polarity and subjectivity of the data, polarity and subjectivity confidence of the data, Weka used to find out the accuracy, precision, recall, and F-score, the SVM and NB showed lower accuracy but NB was better than SVM this is because of using a lesser number of tweets [10].

To detecting sarcasm, machine learning algorithms are commonly used algorithms; hence these types of algorithms are more suitable in the detection of speech. From bellow Figure 7 clearly shows that SVM is most frequently used. The 2250 tweets are used to perform sarcasm detection, for the classification machine learning algorithms SVM and Naïve Bayes classifier is used to determine the result [10]. The overall accuracy obtained reaches high using RFDT, outperforms other SVM, K-NN, and Maximum entropy classifiers [9]. The hyperbolic feature set is used on machine learning-based and context-based approaches [16]. The LR outperformed SVM with sequential minimal optimization [23]. The proposed SVM, Complementary NB, NB, and LR model on the ArSarcasm-v2 dataset is used, the performance of the proposed model has been compared with other models, shown that SVM outperformed another model [26].

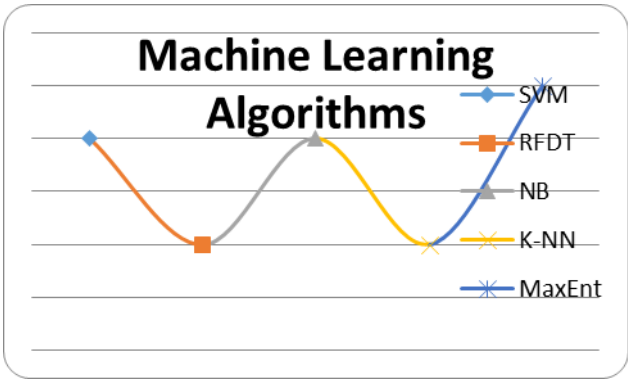

Fig. 7. Machine Learning algorithms in sarcasm detection

Table 1
Summary of hate and sarcasm classification along with different parameters

	Datasets			Features										Algorithms		Area	
	Twitter	*Facebook*	*Amazon*	*Sentiment*	*Semantic*	*N-gram*	*Pattern*	*Hashtag*	*PoS*	*Pragmatic*	*BoW*	*TF-IDF*	*Skip-gram*	*ML*	*DL*	*Hate*	*Sarcasm*
[1]	√			√	√		√							√		√	
[2]	√					√					√			√		√	
[3]		√				√			√					√	√	√	
[4]														√	√	√	
[5]	√							√							√		√
[6]														√			√
[7]	√					√	√		√					√			√
[8]																	√
[9]	√			√	√	√	√		√					√			√
[10]	√					√			√					√			√
[11]	√										√		√		√	√	
[12]	√					√					√	√		√	√	√	
[13]	√					√			√				√			√	
[14]																√	
[15]																√	
[16]					√				√					√			√
[17]		√			√	√						√		√		√	
[18]	√				√											√	
[19]	√		√	√			√	√									√
[20]	√					√			√					√			√

Many authors have of research article experimented by using deep learning algorithms, unless not only adhering to the machine learning algorithms. Bellow Figure 8 shows that authors have used deep learning algorithms while detecting sarcasm on Twitter datasets. The Recurrent Neural Network, Maximum Entropy are the deep learning algorithms used in detecting sarcasm detection. The deep neural network under Recurrent Neural Network is used in constructing the deep neural network model, to perform the detection of sarcasm speech [5]. The authors of the research article proposed a deep neural model, observed that the novel sAtt-BLSTM convNet model outperformed bidirectional LSTM [24].

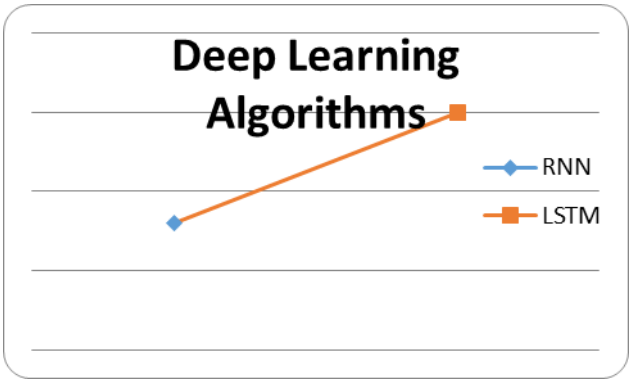

Fig. 8. Deep Learning algorithms in sarcasm detection

4. Conclusion and Future Work

In this paper, we presented a survey on the detection of hate and sarcasm speech. This task is usually done by using machine learning and deep learning algorithms. Using the Twitter datasets for the detection of hate and sarcasm improves performance. The accuracy of the imbalanced dataset is higher than those of the balanced one because the label bias in the imbalanced dataset decreases the performance of sarcasm detection. The most basic feature extraction such as sentiment, semantic, and pattern-related features on binary classification yields significantly higher accuracy. The frequently and most suitable models such as N-grams, BoWs, and PoS tags are highly preferred to use in feature extraction because using these will make it easy to detection hateful and sarcastic speech to get significant results. Both machine and deep learning work well in this area, supervised learning approaches provide better results. In future work, since the hateful and sarcasm detection is done only on Facebook and Twitter datasets, the researchers can also use Instagram, Youtube dataset in their work. Building a combined machine learning and deep learning methods to detect hate and sarcasm in the text would improvise to get a better outcome.

References

[1] Watanabe, H., Bouazizi, M., & Ohtsuki, T. (2018). Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection. IEEE Access, 6, 13825–13835.

[2] Alfina, I., Mulia, R., Fanany, M. I., & Ekanata, Y. (2017). Hate speech detection in the Indonesian language: A dataset and preliminary study. 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS).

[3] Fabio Del Vigna, Andrea Cimino, Felice Dell'Orletta, Marinella Petrocchi, Maurizio Tesconi "Hate me, hate me not: Hate speech detection on Facebook", Istituto di Informatica e Telematica, Pisa Italy, 2017.

[4] Schmidt, A., Wiegand, M.: A survey on hate speech detection using natural language processing. In: Proceedings of the Workshop on Natural Language Processing for social media, pp. 1–10. Association for Computational Linguistics (2017).

[5] Meishan Zhang1, Yue Zhang2 and Guohong Fu1, "Tweet Sarcasm Detection Using Deep Neural Network", School of Com Sci and Tech, Heilongjiang University, China.

[6] Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., & Wright, B. (2020). Sarcasm detection using machine learning algorithms in Twitter: A systematic review. International Journal of Market Research, 147078532092177.

[7] Tomas Ptacek, Ivan Habernal,Jun Hong, "Sarcasm Detection on Czech and English Twitter", Dept of Comp Sci and Engng, University of West Bohemia.

[8] Wicana, S. G., Ibisoglu, T. Y., & Yavanoglu, U. (2017). A Review on Sarcasm Detection from Machine-Learning Perspective. 2017 IEEE 11th International Conference on Semantic Computing (ICSC).

[9] Bouazizi, M., & Otsuki Ohtsuki, T. (2016). A Pattern-Based Approach for Sarcasm Detection on Twitter. IEEE Access, 4, 5477–5488.

[10] Saha, S., Yadav, J., & Ranjan, P. (2017). Proposed Approach for Sarcasm Detection in Twitter. Indian Journal of Science and Technology, 10(25), 1–8.

[11] Gambäck, B., Sikdar, U.K.: Using convolutional neural networks to classify hate-speech. In: Proceedings of the Workshop on Abusive Language Online, pp. 85–90 (2017).

[12] Badjatiya, P., Gupta, S., Gupta, M., & Varma, V. (2017). Deep Learning for Hate Speech Detection in Tweets. Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion.

[13] ZeerakWaseem, "Are You a Racist or Am I Seeing Things? Annotator Influence on Hate Speech Detection on Twitter", University of Copenhagen Copenhagen, Denmark.

[14] Gitari, N. D., Zhang, Z., Damien, H., & Long, J. (2015). A Lexicon-based Approach for Hate Speech Detection. International Journal of Multimedia and Ubiquitous Engineering, 10(4), 215–230.

[15] Gangemi, A., Navigli, R., Vidal, M.-E., Hitzler, P., Troncy, R., Hollink, L., Alam, M. (Eds.). (2018). The Semantic Web. Lecture Notes in Computer Science.

[16] A, A., G, S., H R, S., Upadhyaya, M., Ray, A. P., & T C, M. (2020). Sarcasm detection in natural language processing. Materials Today: Proceedings.

[17] Rodriguez, A., Argueta, C., & Chen, Y.-L. (2019). Automatic Detection of Hate Speech on Facebook Using Sentiment and Emotion Analysis. 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC).

[18] Ahmad Niam, I. M., Irawan, B., Setianingsih, C., & Putra, B. P. (2018). Hate Speech Detection Using Latent Semantic Analysis (LSA) Method Based on Image. 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC).

[19] Davidov, D., Tsur, O., Rappoport, A., (2010). Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon. Proceedings of the Fourteenth Conference on Computational Natural Language Learning.

[20] Mukherjee, S., & Bala, P. K. (2017). Detecting sarcasm in customer tweets: an NLP based approach. Industrial Management & Data Systems, 117(6), 1109–1126.

[21] Lingiardi, V., Carone, N., Semeraro, G., Musto, C., D'Amico, M., & Brena, S. (2019). Mapping Twitter hate speech towards social and sexual minorities: a lexicon-based approach to semantic content analysis. Behaviour & Information Technology, 1–11.

[22] Joshi, A., Bhattacharyya, P., & Carman, M. J. (2017). Automatic Sarcasm Detection. ACM Computing Surveys, 50(5), 1–22.

[23] González-I, R., Muresan, S., & Wacholder, N. (2011). Identifying Sarcasm in Twitter: A Closer Look.

[24] Son, L. H., Kumar, A., Sangwan, S. R., Arora, A., Nayyar, A., & Abdel-Basset, M. (2019). Sarcasm Detection Using Soft Attention-Based Bidirectional Long Short-Term Memory Model with Convolution Network. IEEE Access, 1–1.

[25] Mossiz, Z., & Wang, J-Haur. (2018). Social Network Hate Speech Detection for Amharic Language.

[26] Allam, A. H., Abdallah, H. M., Amer, E., & Nayel, H. A. (2021). Machine Learning-Based Model for Sentiment and Sarcasm Detection.
[27] Pitsilis, G. K., Ramampiaro, H., & Langseth, H. (2018). Effective hate-speech detection in Twitter data using recurrent neural networks. Applied Intelligence.
[28] Al-Makhadmeh, Z., & Tolba, A. (2019). Automatic hate speech detection using killer natural language processing optimizing ensemble deep learning approach. Computing.
[29] Brawijaya, M. A. F, & Yuniarti, A. (2018). Ensemble Method for Indonesian Twitter Hate Speech Detection. Indonesian Journal of Electrical Engineering and Computer Science.
[30] Raufi, B., & Xhaferri, I. (2018). Application of Machine Learning Techniques for Hate Speech Detection in Mobile Applications. 2018 International Conference on Information Technologies (InfoTech).