International Journal of Future Generation Communication and Networking
Vol. 13, No. 04, (2020), pp. 1312-1323

A PRACTICAL APPROACH TO CATEGORIZE HATE SPEECH IN ONLINE SOCIAL NETWORKS

Chandra Sekhar Sanaboina[1*], Nagesh Cheedarla[2], Raghu Ram Kadambari[3]

[1]*Assistant Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Kakinada, India*

[2]*PG Student, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Kakinada, India*

[3]*Assistant Professor (C), Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Kakinada, India*

[1]*chandrasekhar.s@jntucek.ac.in,*[2]*nageshcheedarla@gmail.com,*[3]*kadambari.raghuram143@gmail.com*

Abstract

The main objective of this paper is to analyze the performance in terms of accuracy for various Machine Learning (ML) models such as Support Vector Machines (SVM), Logistic regression (LR), and ensemble classifier to compartmentalize the hate speech on Online Social Networks. It deals with a process centered on the automatically obtaining patterns and unigrams through training dataset. Those unigrams and patterns can be used afterward as a preparation (training) for algorithms in machine learning, among others. Sentiment Analysis is used to detect the polarity of the tweets if it is clean, hatred, or offensive by using sentiment analyzer. The system is implemented on a range of 24783 tweets. The results prove that the deployed model attained an accuracy rate of 88%, 72%, and 66% using SVM, LR, and Ensemble classifier respectively. The binary and ternary classification was implemented on tweets to categorize the hate, offensive, and neutral speeches.

Keywords*: Twitter, Speech Detect, Machine Learning, Sentiment Analysis, Support Vector Machines (SVM), Logistic Regression (LR), Ensemble, Word Cloud, Confusion Matrix.*

1. INTRODUCTION

With the increase of Social Networking sites communication between unknown people became more direct. This resulted in provoke of "Cyber" disputes among them. By this, more hate speech is generated between them until it became a serious problem. Here the hate speech refers to the use of a vocabulary of hatred, abuse, or offense that targets a band or who shares a similar aspect. The use of hate speech is prohibited by social networking sites. But the size of such social networks makes it extremely difficult to fully control everything. Hence the need to identify and eliminate hate speech of any kind emerges**.** Online Social Networks (OSN) and blogs for microblogging target internet users more than other websites. Services like Facebook, Instagram, and Twitter have become extremely popular among young people of diverse cultures, backgrounds, and values. These websites generate a lot of content on an hourly basis which cannot be controlled by systems. The data is generated called "Big Data".

With the rise of these interaction sites, cyber conflicts are taking place between individuals because different people come from various backgrounds thus each mindset is completely different from one another. Social Networks are made to connect people with the alike mindset. Social Networks help us in the easy sharing of photos, media, and whatever a user wants to share. One can post his/her feelings freely online without hesitation. People have become more and more attached to this online networking and cannot withdraw their attention from it.

ISSN: 2233-7857 IJFGCN

Facebook, Twitter, and other internet services become Internet destinations that are frequently visited. Such websites enable every individual to easily and quickly exchange photos, thoughts, links, and updates with each other. Another scenario is that social networking sites have outstripped written communications and increased instant messaging. Users can have 'phone calls' or 'private conversations' through their computers these days. It was made possible through the creation of a new platform such as Skype. In comparison to Twitter and Facebook, where the method of interaction is written, Skype involves a mode of communication that is direct. Introduction to webcams in the social networks made possible to generate less amount of verbal communication but increased in the demand for direct contact with the other person.

These networks are extremely large when compared to any other communication system. This system is also called the "Global Communication System" because it connects the world from one end to another. People can make use of this network for their advantage or they may use it for miscellaneous purposes. There are many scenarios where communication systems are dangerous to one individual. The distribution of too much information on these websites can be dangerous in two forms. Firstly, it might expose anything about you which is not intended to share and secondly, the shared information can sometimes put you at risk. This results in cyberbullying and information transfer from one place to another in a small amount of time. Criminals may use information about the birthday, venue, routine, hobbies, and interests of an individual to publicly humiliate a trustworthy partner or even persuade an offender to have all the rights to manipulate financial or personal details.

Hatred speech relates to a vocabulary that threatens a class of people or people who share similar feature or property. The property here refers to gender inequality and racial or ethnic group (i.e., discrimination) or culture or belief, etc. The Internet is a common source where a lot of hate speech is generated in a short time. Most of the websites are prohibited to use hatred speech but the scale of the network makes this nearly difficult to monitor anything. Hence the importance of automatically detecting such speech and filtering hateful language using machine learning algorithms is the need of the day.

Hate speech is the one that circulates faster through the communication medium and has no integrity at all(i.e., it may be true or false). The world has witnessed voluminous scenarios where false news took the lives of many innocent people. To eliminate such instances to some extent one can implement the ML approaches to filter out such content related to hateful and offensive. Hate speech recognition also serves as a prominent research area in today's world.

2. RELATED WORK

This segment discusses similar works related to hate speech detection. This section describes related works regarding hate speech detection with different techniques. These include SVM, SENTA Tool, Naive Bayes (NB), Logistic Regression (LR) and Random Forest, etc.

Mondher Bouazizi and Tomoaki Otsuki in their paper [1] implemented a model-based approach to Twitter sarcasm detection. The model uses several sets of features that cover the various types of sarcasm to categorize tweets as non-sarcastic and sarcastic. The implemented technique achieves 83.1% accuracy.

Mondher Bouazizi et al. in their paper [2] discussed a new approach that classifying tweets into unique categories namely "Positive," "Neutral" and "Negative" for ternary classification and "positive," "negative" for binary classification. The scope for classification is limited to seven different classes by using the SENTA tool created to help users pick from a broad range of features the way that suits their application the most. The proposed approach shows the accuracy rate of 60.2% for the classification of multiple-classes. Also, this model proved to be accurate of 70.1% for binary and ternary classification.

Rui Ren et al. in their paper [3] discussed Investor perception that performs a significant part in the share sector. The textual content created by the user on the Web is a valuable source for the reflection of investor sentiment and forecasts share prices as a contribution to stock exchange data. The

ISSN: 2233-7857 IJFGCN

proposed approach combines Sentiment Analysis with Machine Learning Techniques using SVMs. Its accuracy rate shows 89.93%, with just an 18.6% increase after the addition of sentiment variable. The proposed model helps the investor to make a wiser decision.

Samah Aloufi and Abdulmotaleb El Saddik in their paper [4] focused on assessing opinions expressed via Twitter by soccer fans. Such tweets mean a change in the fan's emotions as they watch football and respond to game stuff, such as goal scoring, penalties, and so on. This proposed solution includes the creation of a soccer-specific emotion dataset later manually labeled with this data to create a lexicon. Finally developed an emotion classifier that can identify the feelings conveyed in soccer conversations.

Mondher Bouazizi and Tomoaki Ohtsuki in their paper [5] concluded that Tri-class sentiment analysis, examines the recognition of the user's exact emotion, instead of the aggregate polarity of his/her message. This method introduced dynamically assigns different results to each feeling in a tweet and chooses the tweets with the highest grades. The result will be added to the SENTA Tool where the dataset was manually labeled. The results show F1 score equals 45.9%.

Salud Maria Jimenez-Zafra et al. in their paper [6] focused on negation in Spanish Sentiment Analysis. This approach is tested with a large corpus that is written in Spanish. The polarity classification is done based on the Lexicon based system. The results reveal that the final system's accuracy is significantly improved by 12.61% and the overall accuracy is 62.61%.

Zhao Jianquiang and Gui Xiaolin in their paper" [7] discussed the significance of pre-processing text in Twitter content. They were focused on extracting new sentiment features. The accuracy and the F1-measurement of the Twitter sentiment classifiers is enhanced while using the pre-processing techniques but hardly changes while deleting URLs and adding a few words. The classifiers Random Forest and Naïve Bayes are much more attentive than the classifiers SVM and LR if numerous pre-processing strategies were implemented.

Zhao Jianquiang and Gui Xiaolin in their paper [8] implemented word embedding acquired through unsupervised learning on broad Twitter companies using implicit semantic contextual associations and statistical co-occurrence features between each of vocabulary in tweets. They were coupled to n-gram characteristics with word sentiment polarity rate features to form a compilation of sentimental tweets. This feature set is built into a deep CNN for training and forecasting labels for emotion classification. Their analysis is made on five Twitter datasets with n-gram model, the results revealed that the method performed better on the classification of Twitter emotions with an increase in F1-score by 10.21% and improvement of accuracy by 8.23%.

Hajime Watanabe et al. in their paper [9] discussed recognizing hateful speech and messages from the Twitter Dataset. This project deals with an approach focused on the dynamically collected patterns matching and unigram features from the train data set. Among others, these patterns and unigrams would later be used as preparation for machine learning algorithms. Sentiment analysis is used to detect the accuracy rate of the tweets.

Rajesh Basak and Shamil Sural et al. in their paper [10] concluded that public bullying in online social platforms has serious implications for the lives of victims. To filter such content, they've proposed an approach that can detect public embarrassment on Twitter automated. The bullying tweets are broken down into six styles. Based on the classification and categorization of bullying tweets, a BlockShame web application is developed and implemented for an on-the-fly muting/filtering of prudes assaulting a victim on Twitter. Filtering all such content is important to eliminate user's fraud access to the system and protect the privacy of the users.

3. PROPOSED METHOD

This section describes the implementation of hate speech detection systems using SVM, LR, and Ensemble approach. For this purpose, a Twitter dataset is used to analyze the data content. The data set is described briefly in the next section. We make use of different algorithms for this process with

ISSN: 2233-7857 IJFGCN

default parameters. The present system consists of different modules for information about hate speech detection. The architecture of the proposed system is described in Fig 1.

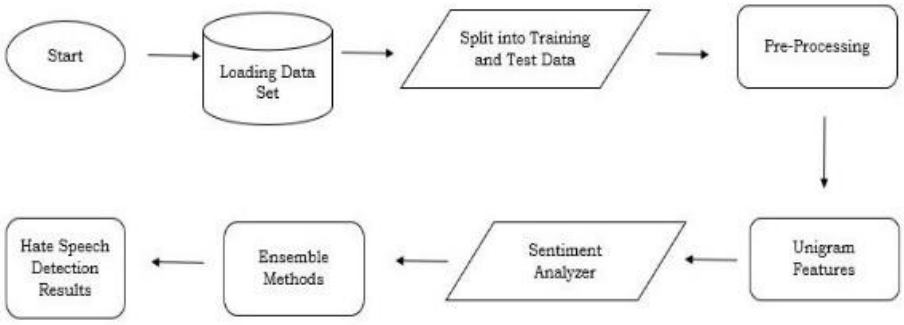

Figure 1: Proposed System

The dataset is divided into "training" and "testing" where 30% is given to testing and the remaining 70% priority is given to training the model. The classification accuracy and the confusion matrix are adopted as evaluation matrices.

3.1. ABOUT DATASET

The prior knowledge on the dataset such as its attributes, dimensions, etc is an important factor by which one can perform correct operations. For the system implementation, an offline dataset which is a publicly available online site called "Kaggle" is considered. The public dataset is a Twitter dataset which contains tweets of different users. The dataset is a composition of offensive, hateful and neutral languages. The dataset contains the following attributes like hate speech, offensive language, neutral, tweets, and also an attribute called "class" which determines the polarity of the tweets. The count attribute summaries the result of the tweet based on its prior dependencies.

These details of the dataset are aggregated in the following Table I.

Table I: About Dataset

Description	Value
Dataset	Twitter
No. of Tweets	24783
No. of Attributes	7
Types of Classes	0, 1, 2

3.2. STEPS INVOLVED IN PROPOSED HATE SPEECH DETECTION

There are various steps involved in the detection of hate speech. They are

Step 1: Import libraries – all the required libraries such as NumPy, SciPy, Scikit-learn, etc will be imported to the working environment.

Step 2: Load the data - the Twitter dataset will be loaded.

Step 3: Summarize the data - descriptive statistic features of the dataset will be displayed.

ISSN: 2233-7857 IJFGCN

Step 4: Split the data into test and train - the dataset will be split into two parts. One part as Test and another part as Train, where test data set size is 30% and training data set size as 70%.

Step 5: Pre – Processing - deals with the pre-processing of the data. Where unwanted content will be filtered in this section.

Step 6: Unigram features - Unigram features are obtained. These unigrams are obtained based on the pre-processed data.

Step 7: Sentiment Analysis - Unigrams that are collected from Step 6 will be passed to the sentiment analyzer where the polarity of the tweets will be calculated based on sentiment analyzer. The outcomes will be passed to Ensemble. Various models of ML will be tested here.

Step 8: Results – results obtained from the model in the form of the confusion matrix and classification metrics such as accuracy, F1 score, etc.,

3.3. DATA PRE-PROCESSING

Pre-processing relates to the modifications performed to the information before the algorithm is loaded. Several algorithms of machine learning make assumptions about the data. It is often a very good idea to plan the data in such a way that the problem structure is better presented to the machine learning algorithms. Pre-processing the data is a method of transforming raw information into some kind of cleaner dataset. In other terms, whenever the information is gathered from various sources, it is processed in raw format, which is not feasible for evaluation. Raw data (real-world data) is always imperfect and cannot be passed through a model. That would trigger some errors. This data is obtained in the form of a huge dataset obtained from Twitter. The dataset is an integrated set with all kinds of language. This language is namely “hate speech”, “offensive language”, and “neutral language”. We need to pre-process the dataset to work with the Twitter dataset.

3.4. FEATURES FOCUSED ON SENTIMENT

While the objective of identifying hatred speech varies significantly, it always makes perfect sense to analyze feelings and define polarity using the feeling-based function as the minimal functionality which allows hate speech identification. It is because hate speech is seen more in a “negative” post, rather than in a “positive” post. To analyze the tweets and their behavior we will use the Vader Sentiment package from which Sentiment Intensity Analyzer is used. By using this feature, we can make use of the sentiment analyzer and can predict the polarity of the tweets.

When importing the sentiment analyzer to detect the polarity of the tweets which are pre-processed initially, these tweets will be assigned with two functions namely: FKg and Fre. Here, FKg stands for “Flesch-Kincaid grade level” test whereas Fre stands for “Flesch readability ease”. These two functions have their priority when working with the detection of tweets based on some constraints.

3.5. HATE SPEECH DETECTION USING SVM

SVM is a strong and versatile supervised algorithm of machine learning used for classification and regression. However, they are used commonly for classification problems. SVMs first developed in the 1960s but subsequently improved in 1990. Compared to other machine learning algorithms, SVMs have a unique way of implementation. They are extremely popular because of their ability to manage multiple constant and categorical variables. The SVM design is essentially a reflection of various categories inside a hyperplane in multidimensional space. SVM will generate its hyperplane incrementally so that the mistake could be reduced. SVM aims to split the dataset into groups to identify a Maximum Marginal Hyperplane (MMH). When a hyperplane is formed SVM categorizes the data. This algorithm’s principal task is to categorize the data. For a large dataset, this algorithm proves to be a reliable choice to attain efficiency in predicting the results. This algorithm outperforms the Random Forest (RF), as RF takes time to load and process the data. The current implementation chooses a definite parameter along this process to optimize the results. These parameters can be varied from one to another. Using SVMs can be efficient in case of high dimensions and is relatively efficient in memory.

ISSN: 2233-7857 IJFGCN

3.6. HATE SPEECH DETECTION USING LR

In case, if the subject variable is dichotomous (binary), LR is the correct regression method to perform binomial and multinomial classification. LR, as with all regression analysis is a predictive technique. It is often used to describe the data and also to explain the relation among a binary dependent variable and one or even more independent variables of an interval, ratio point, nominal or ordinal. In the current scenario, LR in our workspace is fed with a different type of parameters. These parameters are namely "solver", "multi-class" etc. The regression algorithm will be given a default number of iterations to operate. When the operation is initialized the algorithm performance can be visualized through certain operations.

3.7. HATE SPEECH DETECTION USING ENSEMBLE

E Ensembles will improve us through the combination of several models in the machine learning result. Ensemble structures generally comprise of many individually trained supervised versions of learning and these evaluations are combined to deliver better predictive performance than that of a single model. This ensemble mechanism is split into two sections:

- Ensemble Sequential approach
- Ensemble Parallel approach

The sequential ensemble indicates that the basic learners in these types of ensemble approaches are generated sequentially. The reason for these approaches is to manipulate the vulnerability of simple learners.

The parallel name implies that the base learners in this type of ensemble approach are generated in parallel. Within this ensemble method, we have a wide variety of methods called "Bagging", Boosting", and "Voting" ensemble methods. This group incorporates a meta-estimator that fits randomized decision trees and uses averaging to enhance predictive performance or accuracy and manage over-fitting on various dataset sub-samples.

While implementing this algorithm in our system we make use of different parameters and functions. The results of this system are quite reasonable. Using this approach, the hate speech can be depicted with a better performance followed with good accuracy.

4. RESULTS AND DISCUSSION

This section summarizes the results using the data visualizations and descriptive statistics for better understanding. Histograms, word cloud, scatter plots, density plots, and Correlation matrix plots are used for visualization purposes. Data will be grouped into bins by using the histograms and also this is the fastest way to get a glance at how each attribute is distributed in the system.

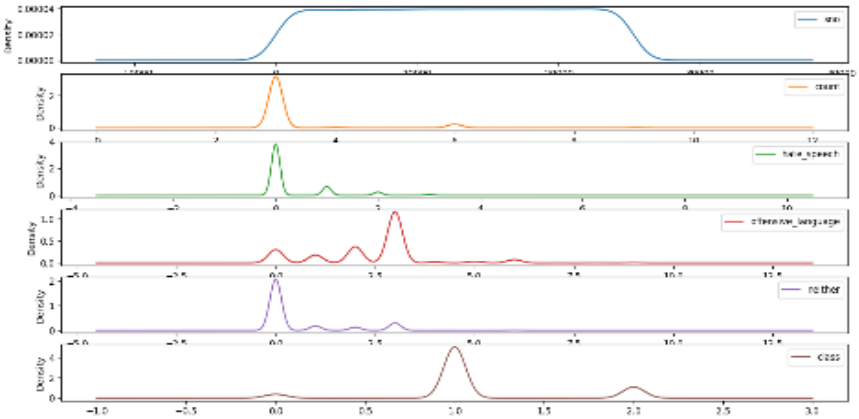

Figure 2: Density plot of the Twitter dataset

ISSN: 2233-7857 IJFGCN

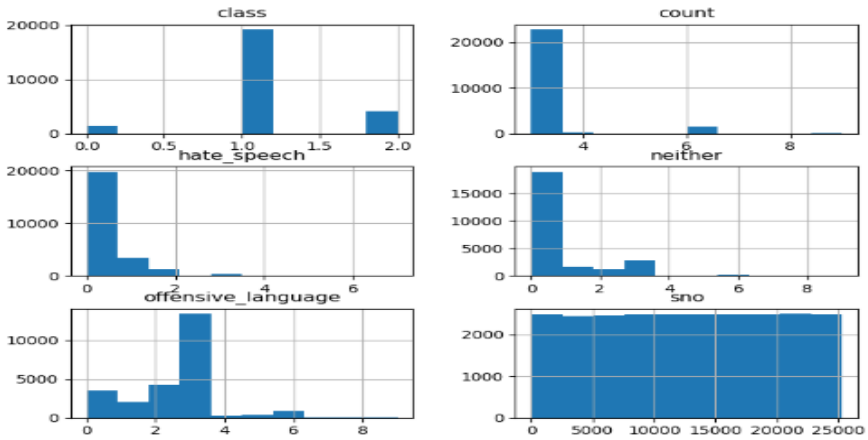

Figure 3: Histograms of the Twitter dataset

Fig. 2 and Fig. 3 describes the density plots and histograms respectively for the dataset. The Twitter dataset is visualized using these varieties of plots and can be obtained from the pre-defined packages. Both representations are on the same dataset. In general, density plots perform better than histograms to visualize many representations at once.

The pattern properties are collected the very same method we collect the unigrams. To collect the unigrams, we need to introduce the patterns first. The patters are obtained in the training data through the data cleaning process.

As the tweets are based on three sentimental categories. It tweet may be a hateful tweet, offensive tweet, or neither of it which is described as a neutral tweet. The following figures annotate the three categorical features of three different classes.

Figure 4: Top Words of Hateful Class

ISSN: 2233-7857 IJFGCN

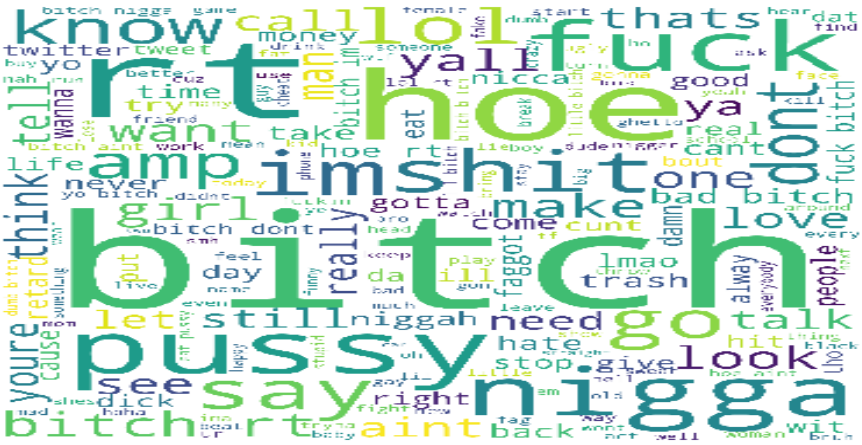

Figure 5: Top Words of Offensive Class

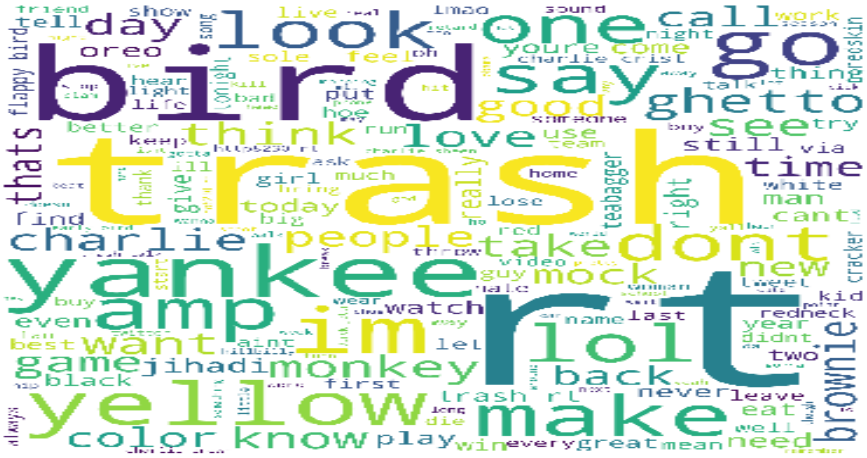

Figure 6: Top Words of Neutral Class

These are the top three classes and their top words which are derived from the dataset. This is obtained with the help of the word cloud feature which is included in our system implementation.

The filtered words are obtained from the tweets. These pattern matching features are the results of our data after the pre-processing. The data is visualized in such a way that all hateful words are categorized and visualized in Fig 4. In the same manner, offensive words and neutral words are visualized in Fig 5 and Fig 6 respectively.

4.1. CORRELATION MATRIX PLOT

The correlation matrix is plotted to demonstrate which variable has a high or low correlation concerning another variable. Fig. 7 represents the correlation matrix.

ISSN: 2233-7857 IJFGCN

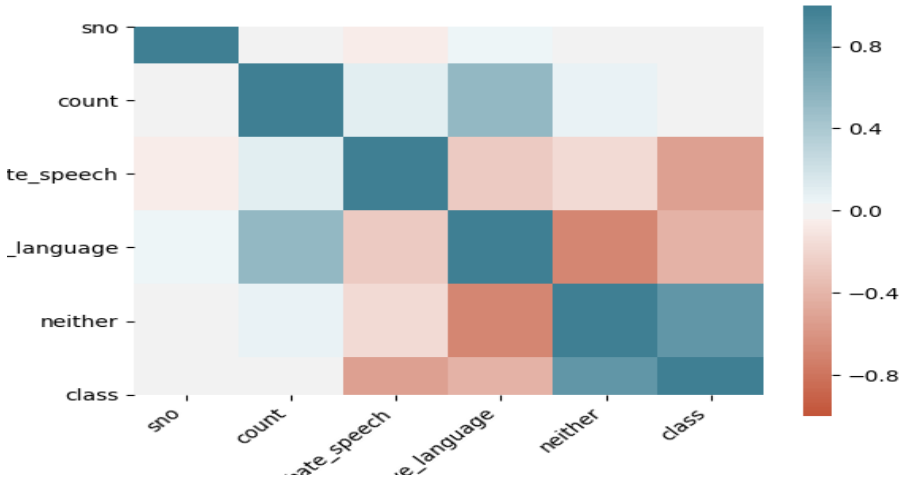

Figure 7: Correlation matrix plot

Each variable in the diagonal line from top left to right correlated perfectly positively with each other. Data must be prepared by removing the duplicates and verify that the dataset contains no null values. The dataset consists of different attributes with different scales. The data need to pre-processed to remove the repeated and unwanted content that is present in the tweets of the data. Cleaning each tweet of the dataset and representing it as clean data is important before feature extraction. Unigram features are extracted sensibly as per the user's interest in such a way that results are generated with higher accuracy. The extracted features are then assigned to the sentiment analyzer to find the polarity of each tweet. In the current dataset, classes like 0, 1, 2 are considered where 0 represents ‘hate language, 1 is ‘offensive language’, and 2 is ‘neutral language’. The class attribute of the dataset consists of these three values. The correct approach to use machine learning algorithms is to use different datasets for training and testing.

4.2. CONFUSION MATRIX REPRESENTATION

The confusion matrix explains the impact of a projection on a classification problem. The matrix of confusion is a basic representation of the accuracy of two or more groups of a model. Confusion matrix for the three classes which are namely hateful language, offensive language, and neutral language is depicted in Table II.

Table II: Confusion Matrix for Ternary Classification

Class	**Classified as**		
	Hateful	**Offensive**	**Clean**
Hateful	50	344	33
Offensive	42	5584	121
Neutral	7	456	798

In the ternary classification, a total of three classes and their validations concerning the accuracy of the used algorithm are represented. It is a challenging comparison to depict the accuracy of the confusion matrix with the ternary classification. For this purpose, we integrate both hateful and offensive categories into a single class as offensive and the confusion matrix for the same is shown in Table III.

ISSN: 2233-7857 IJFGCN

Table III: Confusion Matrix for Binary Classification

	Classified as	
Class	**Offensive**	**Clean**
Offensive	6483	63
Neutral	5185	5615

From Table III, offensive class 6483 cases are predicted as True Positive, 63 cases as False Negative and from the class of Neutral 5185 cases are predicted as False Positive and 5615 cases as False Negative.

4.3. ACCURACY OF THE PROPOSED MODEL

Eq. 1 represents Classification accuracy. It is indeed the combination of both the number of correct predictions and the total number of predictions made.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predctions\ made} \rightarrow ①$$

Accuracy for implemented algorithms is given in Table IV.

Table IV: Accuracy of the proposed approach

Algorithm Used	**Accuracy**
SVM	88%
LR	72%
Ensemble	66%

4.4. OUTCOMES OF IMPLEMENTED ML MODELS

This section highlights various output/classification metrics of the implemented ML models namely SVM, LR, and Ensemble that are obtained during the run time of the project. Each table has a unique performance outcome for each model. The tables depicted in this section are used to get an overview of each algorithm's performance based on their Precision, Recall, F1-Score, and Support.

The overview of the SVM, LR, and Ensemble methods along with their accuracy, micro average and weighted average, F1-score, Recall, Precision, and Support are given in Table V, Table VI, and Table VII respectively

Table V: Detailed overview of the SVM Algorithm.

	Precision	**Recall**	**F1-Score**	**Support**
0	0.14	0.50	0.21	115
1	0.96	0.91	0.93	6054
2	0.82	0.82	0.82	1266
Accuracy			0.89	7435
Micro Average	0.64	0.74	0.66	7435
Weighted Average	0.92	0.89	0.90	7435

ISSN: 2233-7857 IJFGCN

Table VI: Detailed overview of the LR Algorithm.

	Precision	Recall	F1-Score	Support
0	0.16	0.70	0.23	427
1	0.96	0.69	0.81	5747
2	0.77	0.85	0.81	1266
Accuracy			0.72	7435
Micro Average	0.63	0.75	0.63	7435
Weighted Average	0.89	0.72	0.78	7435

Table VII: Detailed Overview of Ensemble Extra Trees Classifier

	Precision	Recall	F1-Score	Support
0	0.51	0.10	0.17	97
1	0.97	0.64	0.77	6417
2	0.35	0.99	0.52	921
Accuracy			0.67	7435
Micro Average	0.56	0.71	0.59	7435
Weighted Average	0.92	0.86	0.88	7435

5. CONCLUSION

This paper presented a comparative study on the performance of SVM, LR, and Ensemble Classifier on the Twitter dataset. The performance analysis indicates that SVM increases the training speed and retains the accuracy by consuming lower memory. SVM is used to handle multiple types of data whereas LR is used to handle the large datasets. From the results, it can be observed that SVM performed better with good accuracy when compared with LR and Ensemble methods. These algorithms can handle both statistical and categorical attributes.

6. FUTURE SCOPE

The current hate speech detection system has been designed with default parameters using different algorithms. Future experimentations can be performed to tweak the default parameters and analyze the performance.

Further, the present system experimented on the offline dataset available at Kaggle. The system can be modified to work on live Twitter or OSN datasets and build a dictionary of unigrams with rich features to detect these hateful, clean, and offensive speeches.

REFERENCES

[1] M. Bouazizi and T. Otsuki, “A Pattern-Based Approach for Sarcasm Detection on Twitter,” IEEE Access, vol. 4, pp. 5477–5488, 2016, DOI: 10.1109/ACCESS.2016.2594194.

[2] M. Bouazizi and T. Ohtsuki, “A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter,” IEEE Access, vol. 5, pp. 20617–20639, 2017, DOI: 10.1109/ACCESS.2017.2740982.

ISSN: 2233-7857 IJFGCN

[3] R. Ren, D. D. Wu, and D. D. Wu, "Forecasting stock market movement direction using sentiment analysis and support vector machine," IEEE Syst. J., vol. 13, no. 1, pp. 760–770, 2019, DOI: 10.1109/JSYST.2018.2794462.

[4] S. Aloufi and A. El Saddik, "Sentiment Identification in Football-Specific Tweets," IEEE Access, vol. 6, no. c, pp. 78609–78621, 2013, DOI: 10.1109/ACCESS.2018.2885117.

[5] M. Bouazizi and T. Ohtsuki, "Multi-Class Sentiment Analysis in Twitter: What if Classification is Not the Answer," IEEE Access, vol. 6, no. c, pp. 64486–64502, 2018, DOI: 10.1109/ACCESS.2018.2876674.

[6] S. M. J. Zafra, M. Teresa Martin Valdivia, E. M. Camara, and L. Alfonso Urena Lopez, "Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter," IEEE Trans. Affect. Comput., vol. 10, no. 1, pp. 129–141, 2019, DOI: 10.1109/TAFFC.2017.2693968.

[7] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis," IEEE Access, vol. 5, no. c, pp. 2870–2879, 2017, DOI: 10.1109/ACCESS.2017.2672677.

[8] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," IEEE Access, vol. 6, no. c, pp. 23253–23260, 2018, DOI: 10.1109/ACCESS.2017.2776930.

[9] H. Watanabe, M. Bouazizi, and T. Ohtsuki, "Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection," IEEE Access, vol. 6, no. c, pp. 13825–13835, 2018, DOI: 10.1109/ACCESS.2018.2806394.

[10] R. Basak, S. Sural, N. Ganguly, and S. K. Ghosh, "Online Public Shaming on Twitter: Detection, Analysis, and Mitigation," IEEE Trans. Comput. Soc. Syst., vol. 6, no. 2, pp. 208–220, 2019, DOI: 10.1109/TCSS.2019.2895734.

ISSN: 2233-7857 IJFGCN