

Identification of HATE speech tweets in Pashto language using Machine Learning techniques



Aftab Alam Janisar
aftab.janisar@gmail.com
Bahria university
Islamabad, Pakistan

Hammad Afzal
hammad.afzal@mcs.edu.pk
National University of Sciences and
Technology, Pakistan

Ganesh Kumar
gmrathi67@gmail.com
Universiti Teknologi
PETRONAS
Malaysia

ABSTRACT

From the last few years, researchers are very much attracted to sentiment analysis, especially towards hate speech detection systems. As in different languages procreation of hate speech has compelling and symbolic consideration on social media. Hate speech has a great impact on society, using hate words harms others dignity. Hate speech detection systems are important to stop the transformation of hate words into crimes. In this research, a framework is developed for hate speech detection system in the Pashto language. A dataset is created for which data is collected from Twitter. Because there is no related data available. Most of the research work has been done in this domain for other languages, and it's very mature in the context of detecting hate speech. But when it arrives at the morphological languages not much work has been done especially in the Pashto language.

This research aimed and collected data from Twitter, Tweets related to ethnicity and religion. The data collected from twitter has been annotated manually and categorized the data as hate or not by comparing it with the offensive content. For hate speech detection systems to view the impact of different features/attribute this study performed experiments on the existing classifiers i.e., SVM, Naïve Bayes, Decision tree and KNN. SVM produced the highest result at dataset of 500 i.e., 74% among all the classifiers. KNN and Decision Tree produced same result at dataset of 1500 i.e., 65.0%. Dataset of 2800 Decision Tree produced the highest result i.e., 72% and SVM produced 71.9%.

Key words : Sentiment Analysis, HATE speech, Natural Language Processing, Pashto Tweets

1. INTRODUCTION

Twitter, Facebook, and other social media platforms have an important role in today's life. These platforms are used to connect people across the world. These platforms are not just developed for communicating with the closed ones but to share free thoughts and ideas with people freely. [1] Freedom of speech has enabled these social media platforms and thus, provided a way for the individuals spreading hate speech orientation. Under the law of freedom of expression, one cannot propagate abusive speech because hate speech may transform into a criminal offense later. The importance of hate speech detection systems is to stop such unhealthy conversations which later leads to crime [2]. Studies have shown that how one can use any social media platform and

spread such hate content against not only the minorities but also against other major communities [3].

Pakistan has also the similar law which is [4] Article 153-A Pakistan's penal code enforces the imprisonment of up to 5 years and charges the one who has prompted the hate on one's protected characteristics that is religion, language, caste etc. A code of conduct has been released by [5] European commission to tackle down such activities regarding hate speech.

As we are living in a modern era in which advancement in information technology and detonation of social media posing's created a surrounding that can easily be exploited. There are multiple sources of news and a large quantity of data are yielded daily on different social media networking platforms. It is hard to believe because with no efforts it can be manipulated for different purposes like it could be exploited for profit purpose of any organization or any other political advantage, so sentiment analysis is an important step for the detection of deceptive news [6]. Sentiment analysis is basically the study of examining texts, reviews. Sentiment analysis is the art of classifying text to a class, Different social media platforms are the sources. Sentiment analysis is important because of Scalability, Real-time analysis, Consistent criteria. The sentiment analysis process is involved in many fields such as Artificial Intelligence, machine learning, Computer Science etc. Advancement has been made in this field, but artificial intelligence still working hard on identifying hate speech. Basically, the purpose of the hate speech is to make references about the content, tone and nature of speech. The core functionality of hate speech detection systems can be further divided into sub-tasks such as manual labeling of the datasets, model selection for cross-validation, deployment, download new messages/texts and predict [8].

The research on extracting opinions has been mature for some languages, the languages that include Latin alphabets are English, French, and Spanish. However, morphological complex languages such as Arabic, Urdu and Pashto state of the art are still not pleasing. Part of speech and named entity recognitions not up to the mark when applied to such languages.

Abusive speech or content is legal as well as ethical issues and different countries have established laws against these hate speeches and protect their community from harassment, violence, and offensive speech.

Social media (Facebook and Twitter etc.) have online hate content and abusive content has speeded and multiple government authorities have the political pressure on these companies to fight against these hate content.

The focus of the current frameworks in sentiment analysis techniques is basically on traditional methodologies. Hate speech techniques are used for the Extraction of opinion and it has also been done for English, Urdu, and Arabic for other languages [18]. The previous frameworks fulfill the requirements of hate speech for the Pashto language. In this research work, Pashto language corpus and lexicon are created which has included slur words and hate words. But before this research work, there is no corpus or hate speech dataset available for the Pashto language [11] [12]. Which is the largest spoken language in Afghanistan and the second-largest language in Pakistan.

In hate speech detection systems, previous work is done in most of the major categories like religion, ethnicity, and gender etc. This work aimed to detect hate speech of Pashto language in major categories as well as define to the level of hate is conveyed. A study is conducted in Turkey [9] derived guidance from it and use their framework. According to the machine learning aspects, no work yet has been done in this regard, and to best of our knowledge, no work has yet been done in hate speech detection systems in the Pashto language.

Structure of this paper is as follows: In section 2 an overview of the existing literature on Pushto language followed by research methodology and experiment process in section 3 that includes gathering, development and pre-processing of dataset and Tokenization process. Section 4 explains about the experimentation and result.

2. LITERATURE REVIEW

This section presents an overview of existing literature on Pushto language. Sentiment analysis and hate speech are well-studied areas; a lot of research work has been done in the area. From the last few years, researchers are very much attracted to sentiment analysis and especially towards hate speech detection [16]. But the previous research was primarily focused on other languages mainly on the English language. This research is mainly focusing on the traditional approach. As in this approach classifiers, features are passed manually. Traditional approaches are based on SVM, Naïve Bayes, decision tree and many others. There are some modern approaches used for hate speech detection systems such as deep learning approaches, GRU, Natural language processing, skipped CNN and neural networks [14]. Modern approaches are CNN, RNN, LSTM and many others [15] [7].

In context traditional approaches reference [17] has proposed a system for Urdu language text analysis using SentiUnits. Their methodology towards notion examination

depends on the recognizable proof and extraction of SentiUnits from the given content, utilizing shallow parsing. SentiUnits are the articulations, which contain the supposition data in a sentence. We use a supposition commented on vocabulary-based methodology.

In hate speech detection systems, a huge amount of work has been done in it, which was mainly centered on the English language as it is an international language. In the Pashto language, some work has been done in the context of an End-to-End OCR System for Pashto Cursive Script [13], Pashto language stemming algorithms and rule-based part of speech tagging for the Pashto language. But to the best of my knowledge, no work has yet been done in hate speech detection system of Pashto language.

Sentiment analysis and hate speech are a well-studied area [11] presented The Lexicon-based Approach for Hate Speech Detection. The presented research considers centers around the extraction of conclusions from advanced conceived Pashto content. The investigation included the production of different class classifiers by adjusting the philosophy of message level undertaking utilizing investigation of Tweets in expansion to this, word-assumption vocabularies with tokenization of sentences and interpretation of existing English dictionaries were created. The findings demonstrate that lexical highlights-based Pashto assessment examination extricates assumptions with high precision. This research is based on opinion examination on content in the Pashto language. The significant commitments of this, think about incorporate building a Pashto corpus, vocabulary list and a summed-up classification system dependent on lexical highlights to foresee conclusions from a given content. The present examination is in view of the uni-gram approach and can be upgraded further by presenting the n-gram approach. They additionally mean to improve the corpus and make it openly accessible.

In this research [10] has proposed a system for recognition of Pashto letters using Zoning features. Their research work shows a shrewd acknowledgment framework proposed for Pashto letters. Nonetheless, manually written character acknowledgment was tried because the varieties fit as a fiddle furthermore, style. Notwithstanding that, these characters normally fluctuate among people. The recognizable proof turns out to be notwithstanding overwhelming due to the absence of standard datasets including engraved Pashto letters. This work structured a database of moderate size, which incorporates a sum of 4488 pictures, coming from 102 distinctive examples for every one of the 44 letters in Pashto. Besides, the acknowledgment structure concentrates zoning highlights pursued by K-Nearest Neighbor (KNN) and Neural Network.

3. RESEARCH METHODOLOGY

In this section the research study explained research methodology which involves gathering, development and pre-

processing of dataset and Tokenization process [20] [19]. The figure below represents our research methodology.

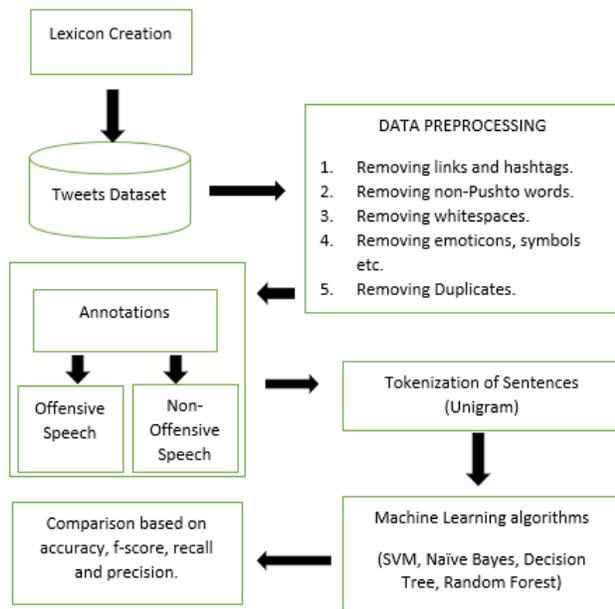


Figure 1: Dataset collection and Pre-processing

3.1 DATASET

This section describes the difference between hate speech and offensive speech and describes the process in detail about hate speech corpus and its annotation in the Pashto language.

3.2 HATE SPEECH

Although some social media communities are notoriously known as Facebook and twitter etc. where people have started to raise their opinions and which people have used as a platform to convey ideas and thoughts, keeping in mind about their issues to be heard and addressed to the global community, have also tried to put into these highlighted aspects.

Hate speech is the exchange of verbal attack towards the inherited and protected aspects of an individual, the aspects which a person doesn't have control to select or adopt. e.g. religion, race, ethnicity, disability, origin, nationality and sexual orientation. Attacks, as per the social community are defined as victimizing or underestimating speech towards these personal characteristics. Attempting verbally to make someone feel Disgusted, inferior, or excluded is considered as hate speech.

People call something as offensive that is triggering anyone's emotions, making them feel insulted regardless of the aspect that is marked to make comments on. Offensive speech may also include personal attacks towards people such as bullying. people always mix up hate speech with offensive language, which is true, but every offensive speech is not considered to be a hate speech, because hate speech is directed towards the inherited characters of people upon who they really are.

3.2.1 HATE SPEECH CLASSES

This research focuses on differentiating the types of hate speech categorically, i.e., religious attachment and ethnicity. This study also categories these three categories into further types identified by statistical hate speech conducted in Turkey. Symbolizing, enmity, attribution, and insult are four types that will be considered in our future work.

3.2.1.1 Symbolizing Symbolizing refers that discourse about anyone's inherited aspects as a basis of hate or humiliation.

3.2.1.2 Enmity/Attribution Negative discourse about any individual, group of individuals, a regional community or a region and expressing them, to be of a certain type. Aggravating some certain negative aspects referring to a certain event or a certain event occurred once and marking the whole community, area, people as same. for example, marking a whole nation as a terrorist based on a certain event happened once in a time.

3.2.1.3 Insult Targeting an individual, a group of individuals or a community by swearing, abusing or cursing. E.g. Abusive words as cursed, Stoner, Greedy, Kefirs (Disbelievers of Islam), Prostitutes, etc.

3.3 DATASET CREATION

The research goal was to collect data from multiple sources like different social media platforms but marked twitter as primary source for data collection because twitter allows the information from the past seven days which has metadata, and that metadata is very useful for further data processing. One reason for selecting Twitter was that people mostly communicate via twitter and it is freely available and accessible.

For corpus creation, the primary source is twitter and extracted the data from it, but it was very difficult to find the desired data from twitter as there was not much work done in Pashto language, Tweets were available but the desired hate speeches were difficult to find out. It was very difficult to find the desired tweets having slur words. Created the list of slur words and exploited a list of words to get and fetch data from twitter, but results were fewer or there were no results. But spending a lot of time collected data from twitter using Pashto swear and slang words. Swear words list contained around 65 words and added some common slur words into it. In that list, added some common, religious groups as well as ethnic words.

3.3.1 LEXICON CREATION

Primary source is twitter and extracted the data from it, but it was very difficult to find the desired data from twitter as there was not much work done in Pashto language, Tweets were available but the desired hate speeches were difficult to find out.

3.3.2 PRE-PROCESSING OF DATASET

For the Pashto language, there are different styles of writing in any language. in Pashto language, words are usually used with

the suffixes and prefixes. Sometimes words are used with the white spaces to complete the meaning of that word. In this study created carpus in which such words with white spaces, removed extra whitespaces.

- Tweets were cleaned by removing ‘URLs’, ‘RT’, ‘Emoji’s’, ‘punctuations’ and ‘@usernames’.
- Most of the hate tweet’s keywords were mentioned with the hashtags, a tweet with a hashtag is difficult to determine whether it is hated or normal tweet.
- In pre-processing hashtags were not removed.
- Duplicate tweets are removed from the data set.
- Removing extra junk text e.g., Smiles.
- For the normalization process, it is not required to split the words with underscore or hashtag, but by replacing and removing the punctuation with white spaces it normalizes hashtag automatically.
- Line breaks are replaced with white spaces.
- Part of Speech process and splits the words process followed.
- To analyze the sentiment of the text in this study it is removed the stop words.

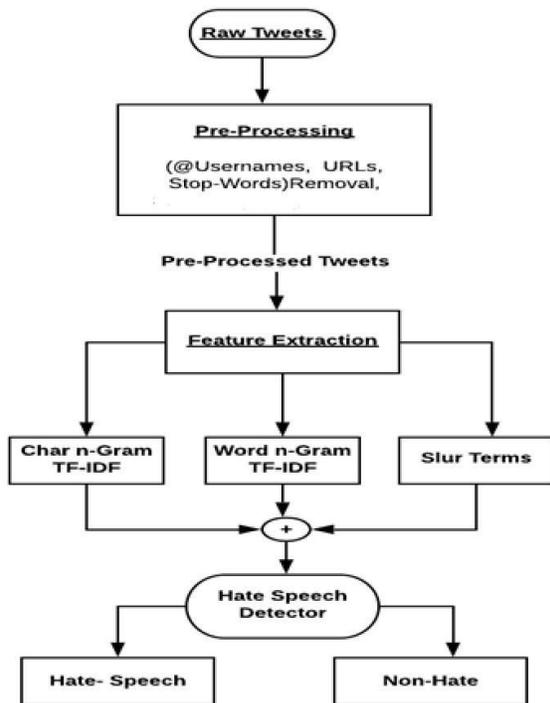


Figure 2: Hate speech Detection Framework

3.4 FEATURE EXTRACTION

Experiments are performed using these following given below features set.

3.4.1 Slur/swear words

In this study presented list of multiple slur words, and as a distinguished set of features, passed those slur words list which is used in the lexicon.

S = list of slur term

N = total number of words in the sentence F = feature vector (words) in the lexicon.

$$F1 = \{s1 | \leq x \geq n\}$$

3.4.2 N-Gram

N-gram is basically featuring an identification approach and it’s a sequence of words. In-text categorization words and character-based n-grams are mostly used.

3.4.3 Word n-grams

For capturing the sequential context, used word n-gram with the range ‘n’. ‘W’ words represented in a sentence.

‘t’ the total number of words represented in a sentence.

$$W = \{w1, w2, w3, \dots, wt\}$$

To weighted further n-gram by their TF-IDF score and can be represented as

$$F = W_i (tf \ idf)$$

3.5 TF-IDF

TF-IDF is an important technique which is specially used for information retrieval. This is also used to represent how specific a word or sentence in a document. The term TF-IDF stands for ‘Term frequency- Inverse document frequency’. TF-IDF plays with a raw dataset, but first, it converts the raw dataset in vector form. Each word is assigned to a vector.

If it is needed to extract some important information from a document this method of TF-IDF increases the proportion of words appearing in the document. TF-IDF is composed of two words ‘TF’ and ‘IDF’ and two statistical methods used by it.

3.5.1 Term frequency

This tells us about how many times a term appears in the document.

3.5.2 Inverse Document Frequency

IDF demonstrate the information about the given word has it. Basically, forgiven term measures the weight in the document. The equation inverse Document frequency is given below where |D| it shows the cardinality (elements number) in the document. df (d, t) to avoid zero division 1 is added in the equation, this is the number in which t appears in a document.

$$IDF(d, t) = \log\left(\frac{1+|D|}{1+df(d, t)}\right)$$

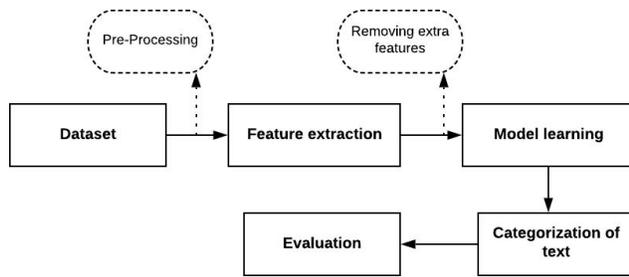


Figure 3: Useful Feature

4 RESULTS & DISCUSSION

For traditional classifiers, in this study compiled the results for each classifier with their individual features in Precision, Recall and F-Score. The results are mentioned below in bars Chart form with the Metric score at the dataset of 500.

Table 1: Metric score at Dataset of 500

Dataset of 500	Accuracy	Precision	recall	F1- Score
SVM	0.650	0.899	0.535	0.74
Naïve Bayse	0.616	0.647	0.621	0.696
Decision Tree	0.662	0.517	0.764	0.622
KNN	0.642	0.643	0.738	0.631

For all the classifiers results are yielded for each of the features. The below-mentioned results are for 500 tweets dataset, in which hate speech rate detection of Accuracy, Precision, Recall and F-Score are mentioned.



Figure 4: performance and Comparison on the dataset of 500

It can be seen from table 5.1 and figure 5.1 that SVM produces a high F-score from the mentioned dataset of 500. In table 5.1 it is visible that detecting hate speech using the slur terms produced results that are good enough. This study randomly selected 500 tweets from the entire dataset, which included common slur terms, religious hate terms, ethnic hate speeches etc and performed the experiments. The highest F-score results which are produced on SVM. The classification score for the dataset of 500 basically depends upon the number of training samples, whenever there is greater number of training samples, it produces greater F-score. These are the metric score on the dataset of 500. The above figure 5.1 shows the results on a dataset of 500 tweets, the F-score of all the classifiers is presented. SVM produced the highest F- score (74%). Naïve Bayse produced the F-score with (69.6%), Decision tree produced the F- score with (62.2%) and KNN has produced the results with F-score (63.1%).

Table 2: Metric score at Dataset of 1500

Dataset of 1500	Accuracy	Precision	recall	F1- Score
SVM	0.625	0.610	0.776	0.610
Naïve Bayse	0.678	0.610	0.628	0.623
Decision Tree	0.644	0.665	0.623	0.650
KNN	0.654	0.643	0.665	0.650

For all the classifiers results are yielded for each of the features. The below-mentioned results are for 1500 tweets dataset, in which hate speech rate detection of Accuracy, Precision, Recall and F-Score are mentioned.

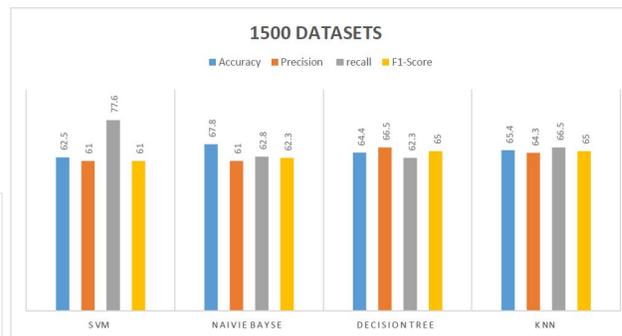


Figure 5: performance and Comparison on the dataset of 1500

It is clearly visible from table 5.2 and figure 5.2 that KNN and Decision tree produces high F- score as (65.0%) from the mentioned dataset. In table 5.2 it is visible that detecting hate speech using the slur terms produced results that are good enough. 1500 tweets are selected randomly from the entire dataset included slur terms, ethnic hate words, religious hate

speeches. Performed the experiments on a dataset of 1500. Two of the classifiers KNN and Decision tree produced the highest similar results with an F-score of (65%) on the dataset of 1500.

Classification score for the dataset of 1500 basically depends upon the number of training samples, whenever there is greater number of training samples, it produces greater F-score. The above figure 5.2 shows the results on a dataset of 1500 tweets, the F-score of all the classifiers is presented. SVM produced the highest F-score (61%). Naïve Bayse produced the F-score with (62.3%), Decision tree produced the F-score with (65.0%) and KNN has produced the results with F-score (65.0%).

Table 3: Metric score at Dataset of 2800

Dataset of 2800	Accuracy	Precision	recall	F1- Score
SVM	0.619	0.666	0.729	0.719
Naïve Bayse	0.654	0.661	0.653	0.641
Decision Tree	0.633	0.665	0.653	0.720
KNN	0.665	0.665	0.638	0.656

For all the classifiers results are yielded for each of the features. The below-mentioned results are for 2800 tweets dataset, in which hate speech rate detection of Accuracy, Precision, Recall and F-Score are mentioned.



Figure 6: performance and Comparison on the dataset of 2800

It is clearly visible from table 5.3 and figure 5.3 that SVM and Decision tree produces a high F-score as above 70% from the mentioned entire dataset of 2800. In table 5.3 pretty good results are yielded using slur terms for hate speech detection. 2800 tweets are selected randomly from the entire dataset included slur terms, ethnic hate words, religious hate speeches. Performed the experiments on a dataset of 2800 tweets. Two of the classifiers SVM and Decision tree produced the highest results with F-score with (71.9%) and

(72%) on the entire dataset of 2800. The classification score for the entire dataset of 2800 basically depends upon the number of training samples, whenever there is greater number of training samples, it produces greater F-score. The above figure 5.3 shows results on a dataset of 2800 tweets,

F-score of all the classifiers is presented. SVM produced the highest F-score (71.9%). Naïve Bayse produced the F-score with (64.1%), Decision tree produced the F-score with (72.0%) and KNN has produced the results with F-score (65.6%).

The below-mentioned tables are the complete results of the four traditional classifiers SVM, Naive Bayse, Decision tree and KNN. While in combined data, for hate speech detection the highest F-score is yielded by SVM and Decision tree compared against the other classifiers. In some cases, SVM with datasets of 500 and 2800 better performed than decision trees as well as better than naïve and KNN. In some cases, Decision trees with datasets of 1500 and 2800 yielded better results than other classifiers. Both the quantity as well as quality is important in machine. learning. Misinterpretation of hate speech occurs when the subject of offensive language is not clearly objected in the speech.

SVM

Table 4: SVM Results

SVM	Accuracy	Precision	recall	F1- Score
500	0.650	0.899	0.535	0.74
1500	0.625	0.610	0.776	0.610
2800	0.619	0.666	0.729	0.719

Table 4.4 shows the complete results for the SVM classifier on all the three datasets. According to the above table 5.4 SVM performed better than other classifiers on the dataset of 500 and on the entire dataset of 2800. Dataset of 1500 hundred, it produced an average outcome for hate speech detection in the Pashto language.

Naive Bayes

Table 5: Naïve Bayes Results

Naïve Base	Accuracy	Precision	Recall	F1-Score
500	0.616	0.647	0.621	0.696
1500	0.678	0.610	0.628	0.623
2800	0.654	0.661	0.653	0.641

Table 4.5 shows the complete results for the Naïve Bayse classifier on all the three datasets. According to the above table 5.5, Naïve Bayse produced better Results than the Decision tree and KNN on the dataset of 500. On other datasets, it yielded an average result for hate speech detection in Pashto.

Decision Tree

Table 6: Naïve Bayse Results

Decision Tree	Accuracy	Precision	recall	F1-Score
500	0.662	0.517	0.764	0.622
1500	0.644	0.665	0.623	0.650
2800	0.633	0.665	0.653	0.720

Table 4.6 shows the complete results for the Decision tree classifier on all the three datasets. According to the above table, 5.6 Decision trees performed better than other classifiers on the dataset of 1500 and on the entire dataset of 2800. Dataset of 500 hundred, it produced an average outcome for hate speech detection in the Pashto language.

K-NN

Table 7: Naïve Bayse Results

K - NN	Accuracy	Precision	recall	F1-Score
500	0.642	0.643	0.738	0.631
1500	0.654	0.643	0.665	0.650
2800	0.665	0.665	0.638	0.656

Table 4.7 shows the complete results for the KNN classifier on all the three datasets. According to the above table, 5.7 KNN produced average results on all three datasets. It yielded a result of more than 60% on each dataset for hate speech detection in Pashto.

5 CONCLUSION AND FUTURE WORK

This research work proposed a framework that detects hate speech in the Pashto language. From all around the globe research on hate speech detection has been mature for some languages but it’s still a problem for some morphological languages to detect hate speech, in this context modern machine learning techniques can help in countering this problem. The number of studies has been done in other languages for hate speech detection but as it is stated before some languages are still not entertained, and they are still neglected. This research work is basically focused on the Pashto language, which is the second biggest territorial language spoken in Pakistan and has the status of official

language in Afghanistan. This study developed the corpus of Pashto hate speech detection and manually annotate the corpus as the tweet is hate speech or no hate. This research aimed at the protected characteristics of one include ethnicity, and religion. This research study focused on useful features for classification of Pashto hate speech and using only existing traditional models for our results. Done these experiments on the supervised machine learning techniques are used. Experiments performed on SVM, Naïve Bayse, Decision tree and KNN Classifiers. Metric score at Dataset of 500 SVM scored the highest among all classifiers. Metric score at Dataset of 1500 KNN and Decision tree produces high F-score. Metric score at Dataset of 2800 SVM produces and Decision tree produces high F-score. For future work improvement should be made to a developed dataset and work more on it, furthermore, challenges must be addressed in future work as well as deep learning techniques should be applied. Sub-categories should be entrained in future work. Deep learning models and other traditional models should be trained in a way on more data to improve the results as well as improve the models for hate speech detection in the Pashto language.

REFERENCES

[1] Zhang, Ziqi, and Lei Luo. "Hate speech detection: A solved problem? the challenging case of long tail on twitter." *Semantic Web Preprint* (2018): 1-21.

[2] Eichhorn, Kate. "Re-in/citing linguistic injuries: speech acts, cyberhate, and the spatial and temporal character of networked environments." *Computers and Composition* 18.3 (2001): 293-304.

[3] Leets, Laura. "Responses to internet hate sites: Is speech too free in cyberspace?" *Communication Law & Policy* 6.2 (2001): 287-317.

[4] Pakistan, and Shaukat Mahmood. *The Pakistan Penal Code (XLV of 1860)*. Legal Research Centre, 1981.

[5] Jourová, Věra. "Code of Conduct on countering illegal hate speech online: First results on implementation." *European Commission*. [cit. 8. březem 2018] (2016).

[6] E. Elmurungi and A. Gherbi, "Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques," *DATA Anal.*, p. 9, 2017.

[7] Z. Zhang, D. Robinson, and J. Tepper, "Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network," in *The Semantic Web*, vol. 10843, A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, and M. Alam, Eds. Cham: Springer International Publishing, 2018, pp. 745–760.

[8] S. Aphiwongsophon and P. Chongstitvatana, "Detecting Fake News with Machine Learning Method," in *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Chiang Rai, Thailand, 2018, pp. 528–531.

[9] E.Kan, M. Nebioglu, S. zkan, F. Tekin, and G. Tosun, "Media watch on hate speech report januar-april 2018," 2018.

[10] S. Khan, H. Ali, Z. Ullah, N. Minallah, S. Maqsood, and A. Hafeez, "KNN and ANN- based Recognition of Handwritten Pashto

Letters using Zoning Features,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 10, 2018.

[11] U. Kamal, I. Siddiqi, H. Afzal, and A. U. Rahman, “Pashto Sentiment Analysis Using Lexical Features,” in *Proceedings of the Mediterranean Conference on Pattern Recognition and Artificial Intelligence - MedPRAI-2016*, Tebessa, Algeria, 2016, pp. 121–124.

[12] N. D. Gitari, Z. Zhang, H. Damien, and J. Long, “A Lexicon-based Approach for Hate Speech Detection,” *Int. J. Multimed. Ubiquitous Eng.*, vol. 10, no. 4, pp. 215–230, Apr. 2015.

[13] Ahmad, Riaz. "An End-to-End OCR System for Pashto Cursive Script." (2018).

[14] Z. Zhang and L. Luo, “Hate speech detection: A solved problem? The challenging case of long tail on Twitter,” *Semantic Web*, pp. 1–21, Nov. 2018.

[15] S. Biere, “Hate Speech Detection Using Natural Language Processing Techniques,” p. 31.

[16] Borum, Randy. "Radicalization into violent extremism I: A review of social science theories." *Journal of strategic security* 4.4 (2011): 7-36.

[17] Syed, Afraz Z., Muhammad Aslam, and Ana Maria Martinez-Enriquez. "Lexicon based sentiment analysis of Urdu text using SentiUnits." *Mexican International Conference on Artificial Intelligence*. Springer, Berlin, Heidelberg, 2010.

[18] Brill, Eric. "Discovering the lexical features of a language." 29th Annual Meeting of the Association for ational Linguistics. 1991.

[19] Aslamzai, Sebghatullah, and Saidah Saad. "Pashto language stemming algorithm." *Asia- Pacific Journal of Information Technology and Multimedia* 4.1 (2015).

[20] Rabbi, Ihsan, A. M. Khan, and Rahman Ali. "Rule-based part of speech tagging for Pashto language." *Conference on Language and Technology*, Lahore, Pakistan. 2009