HATE CLASSIFY A SERVICE FRAMEWORK FOR HATE SPEECH RECOGNITION ON SOCIAL MEDIA NETWORK

Uppitarla Shireesha [1*], Dr. S. Shiva Prasad[2]

[1] PG Student, Department of CSE, Malla Reddy Engineering College Autonomous institution, Hyderabad, Telangana.

[2] Professor (HOD), Department of CSE, Malla Reddy Engineering College Autonomous institution, Hyderabad, Telangana.

Abstract—

Certainly, it is difficult for current machine learning methods to distinguish information that promotes hatred from that which just offends. Among the most common explanations fordue to these methods' limited accuracy in detecting hateful content. Hate categorization methods see this issue as one with several classes. As we'll see here, the turn hate labelling in online communities into a multi-label conundrum. And this to this aim, we offer a CNN-based service architecture, dubbed "Hate Classify," enabling tagging and classifying content. hate speech, provocative, or safe material on social media. Results show that CNN-based multiclass classification accuracy techniques, sequential CNN (SCNN) in particular, is very competitive and sometimes even compared to certain state-of-the-art classifiers. Even more so, in the multilabel sorting Among other solutions, theSCNN demonstrates adequate high performance. procedures based on convolutional neural networks. In light of the findings, it is clear that multilabel categorization Hate speech detection was boosted by as much as 20% without the use of multiclass classification.

Index Terms— CNN, SCNN, multilabel, multiclass, hate classify

DOI Number: 10.14704/NQ.2022.20.15.NQ88531 **NeuroQuantology2022;20(15): 5280-5285**

I. INTRODUCTION

As a result of its widespread use, social media has evolved into an excellent channel for expressing one's innermost thoughts and sentiments. Sadly, in the name of free speech, the growing use of social media has indeed led to the propagation of hateful information. By some estimates, the volume of online hate speech surged by 900% between 2014 and 2016. Statistics show that over three-quarters (73%) of Internet users have witnessed online abuse, and that nearly one-third (40%) have suffered online harassment themselves. To "spread, incite, promote, or justify racial hatred, xenophobia, antisemitism or any other forms of hatred predicated on intolerance, such as intolerance conveyed by aggressive nationalism as well as ethnocentrism, discrimination and hostility against minorities, migrants, as well as people of immigrant origin," according to the Council of Europe's Protocol to the Convention on Cybercrime, is the definition of "hate speech."

However, inside this United States, hateful speech is protected through the First Amendment's guarantee of free speech. To answer the question, "what is hate speech?" on their individual platforms, Google, Facebook, and Twitter each have their own rules to follow. When it comes to dealing against hate speech and other forms of harmful content, social media platforms aren't on the same page.

Twitter is the only major social media platform that does not prohibit the expression of hate speech. Twitter makes a distinction between hate speech and explicit, direct threats. Twitter only takes into consideration "one-sided" reports of

hostile conduct from accounts in which the only objective is toattack other users. Despite Twitter's assertion that "no one is above the rules," the business has come under fire for its allegedly nebulous policy guidelines. A voluntary code of behaviour to eliminate hate speech is defined by the European Union has indeed been agreed upon by Facebook, Twitter, Google's YouTube, and Microsoft as of May 31, 2016. More recently, the CEO of Facebook was asked about the company's policy regarding the flagging and identification of hate speech or offensive material, which brought the topic of hate speech on social media to the forefront. Based on his comments, it's clear that Facebook's present method for identifying hostile material is insufficient for distinguishing between mild, moderate, and extreme expressions of passion. Reason being, many people have various ideas on what constitutes hate speech. Offensive and hate speech has been identified as a concern in a number of earlier investigations, including Del Vigna et al.1. Davidson et al.2 distinguished hate speech from just offensive speech, nevertheless. The study's authors stated that foul language is often used in everyday life. This led to the formulation of the issue of hate speech categorization as a multiclass classification task with categories such as hate, offensive, and neutral language. As Davidson et al.2 have classified several types of speeches, we find that these descriptions are appropriate. In contrast to previous works, we see the hate speech issue as a multilabel problem rather than a multiclass one. The line among offensive & hate speech seems blurry at best, and experts on both sides of the debate have struggled to draw it. Thus, identifying just one group as the cause of a disagreement would never be fruitful. Our findings show that posing the issue as a multilabel one improves the reliability of hate speech detection. Hate Classify, according proposed service architecture, utilizes a hybrid of crowd-sourcing and machine learning methods to identify instances of offensive and hateful language on various social networking websites. The following are the article's most significant contributions.

In contrast to social media platforms wherein hate speech regulations are governed by the particular organizations, we describe a framework for detecting hate and offensive speech as just a service to social media firms that uses a crowd-sourced technique for hate speech identification. Hate speech detection was framed as a multilabel classification issue, and an adequate level of classification accuracy is reached. Social media hate speech detection is enhanced by 20% thanks to the multi-label categorization utilized throughout the HateClassify framework.

II. RELATED WORK

The primary goal of hate speech detection studies is to identify the most efficient features to incorporate in to other text classification algorithms. Even though n-grams and Bag-of-Words seem to be simple to implement as well as produce trustworthy results, those who are typically the primary characteristics chosen by researchers (BoW). Warner et al. [3] proposed using a small set of highly frequent terms to categories hostility toward different groups of people. Syntactic rules, like a user's writing style, were incorporated into n-grams by Chen et [4]. Combining the number of comments on each image to n-grams was a technique employed by Hossein mardi et[5]. Waseem and Hovy[6] combined the n-grams with other factors, including the tweet's length, the tweeter's location, and their gender, to detect hate speech. Researchers have also taken an interest in identifying the grammatical use of hate material. Sentiment characteristics, together with n-grams as well as the BoW, were utilized by Van Hee et[7] to investigate and identify hate speech. In order to investigate online bullying traces, Xu et [8] used n-grams and Part-of-Speech tagging (POS tagging). Every tweet's TF-IDF weighted unigram, bigram, trigram, emotion score, hashtag count, retweet count, URL count, character count, word count, and syllable count were utilized as features by Davidson et[2]. Many academics have turned to the notion of word generalization to deal with the sparsity that arises from the relatively brief duration of texts like tweets or online comments throughout hate detection. Brown Clustering was used by Warner and Hirschberg[3] to generalize words. Latent Dirichlet allocation (LDA) predicts the likelihood of words in distinct clusters, while Brown Clustering distributes words to precisely one cluster. Xiang et al. [9] generalized words using the help of the LDA. For word generalizations, various recent distributed word representations have indeed been devised; these are known as word embedding's. With the massive text as input, word embedding creates a vector space of words. We cluster words with comparable meanings together by moving them closer together in the word vectors. As part of their method for identifying hate speech, Zhong et [10] combined the BoW with both the hate effectiveness rating and word2vec (a word embedding approach). D juric et al. [11] compared the BoW method to another word embedding method, paragraph2vec, for detecting hate speech.

When compared to previous methods used for hate speech identification, state vector machine [12,3-5,7-9] or logistic regression (LR) [2,6,9] have shown superior performance in categorization. This regression model developed by Vowpal Wabbit was favored by Nobata et al. [13]. Models based on recurrent neural networks (RNNs) have been applied to the problem of identifying hate speech by researchers like Mehdad and Tetreault[14]. In this piece, we developed a system for detecting hate speech on social media platforms that makes use of crowdsourcing and neural networks. This same proposed service architecture incorporates word vector embedding as input characteristics and employs CNN models based classification. Also, prior research has viewed the hate speech issue as a multiclass categorization challenge. We have defined the issue as just a multi-label classification problem as well as presented a solution.

III. METHODOLOGY

The framework consists of two parts, both of which are necessary for the framework to perform its intended functions: (i) an offline training module, and (ii) an online hate or offensive speech recognition module. Step 1 and Step 2 in Figure 1 depict the offline training process, which is a recurring operation that takes the tweets and identifies the tweets tagged by various persons. To learn the features within the tweets, an deep neural network is trained offline. The fresh tweets' labels are predicted by an online process that makes use of existing model developed offline. The automated labelling might be agreed upon or disagreed upon by the social media members. In Figure 1, Steps 4, 5, and 6 of the online process are shown. The automated labelling job is optimised by re-training the algorithm using both the previously classified tweets from the online method and fresh tweets labelled by Twitter users.

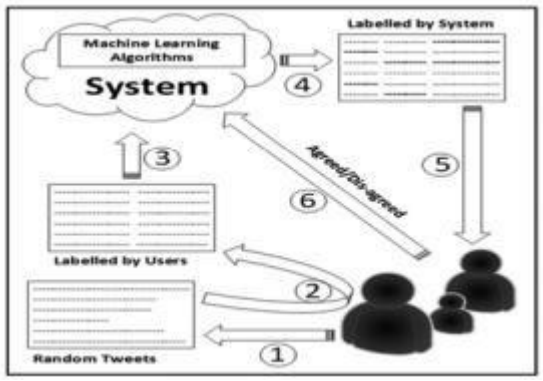

Fig1 : System architecture

3.1 Crowd-Sourced Policy

In contrast to current social media platforms, wherein hate speech standards are governed by individual social media companies, the proposed framework allows social media consumers to make the call on whether or not a certain tweet constitutes hate speech. Voting and training the computer to decide on hate speech is promoted among the populace. Tweets may also be made visible or hidden within a certain location based on majority vote determinations of whether or not they constitute hate speech. Using this strategy to alter social media platforms will not impose a group's prejudice on users. People of various geographical locations will be allowed can train these machines within their regions according with their liking and laws within a democratic fashion. Unfortunately, we were only able to carry out a single cycle of that methodology's recommended iteration. No further votes or label updates were used to training the models. Instead of retraining the models without the possible addition of our personal bias via voting, we

looked for learning most stable model which could consistently outperform with our bias across the multiple datasets. Models with the ability to learn from data and consistently produce good results will be better able to account for regional voting biases over time.

3.2 CNN Model

We present a reliable CNN model towards hate categorization based on sequential convolutional neural networks (SCNN). The layers of a SCNN are as follows: an embedding layer, a Dropout layer, a Flatten layer, as well as a dense layer.
We employed the method of dividing the dataset into such a training set, an evolution set, as well as a test set. Sixty percent of the dataset was put to use as training data, twenty percent as validation data, and twenty percent as test data. For hyperparameter tweaking, the test dataset is used, whereas the test set was put to good use for testing and comparing the model to others.

3.3 Tweets Classification for Hate Speech Detection

We investigated other machine learning models & compared them to the SCNN strategy. All of these methods were compared: (i) SVM over n-grams[12] (ii) LR using a consolidated set of features[2] (iii) long short-term memory (iv) CNNLSTM (v) CNN-nonstatic (vii) ATTCNN (viii) ATTCNN with max. In order to make a fair comparison, we employed n-grams through to n=4 in the initial two schemes with filter sizes ranging from 2 to 4 in neural networks. For this purpose, we use n-grams using SVM and LR with such a long number of features as our benchmark models. Here are brief descriptions of the above-mentioned models used in comparison.

n-grams with SVM: Davidson et al. [2] and Burnap and Williams [12] describe the method. It is based on the trigrams, bigrams, and unigrams. This SVM classifier is given the features. We compared the other models' multiclass classification results to this basic model.

LR with multiple features list: Davidson [2] describes a method for using LR with such a list of several characteristics. Each tweet's TF-IDF weighted unigram, bigram, and trigram as well as its sentiment score, hashtag count, retweet count, URL count, character count, word count, and syllable count are used as characteristics in just this method.

LSTM: The LSTM is indeed a popular RNN design for text categorization. After embedding a layer with such a single dense layer, we constructed a model of a single-layer LSTM to use as a benchmark.

CNNLSTM: By inserting an LSTM layer even before dense layer, CNNLSTM is a variant of the aforementioned model.

CNN-nonstatic: According to the model described by Davidson [15], CNN-nonstatic exists. Sentiment analysis throughout text was indeed the original application of the method. An embedding layer plus three convolutional 1-D maxpooling ID and Flatten layers were chained prior to the output dense layers in just this method. Changes were made so that three classes would be used instead of two throughout the method. To improve the model's classification performance, the weights of both the vectors inside the embedding layers have been tweaked for each task individually.

CNN2D: As a result of updating the model proposed by Davidson [15] to make use of convolutional 2-D neural layers rather than 1-D layers, the resulting model is known as CNN2D. To ensure that the output for embedding layers may be utilised inside the convolutional 2-D layers, every output from embedding layers are altered.

ATTCNN: As reported by Kim[16], ATTCNN incorporates an attention mechanism into the convolutional layer of the network. The model is comprised of a convolutional layer that pays close attention, a flattening layer, as well as a dense layer. In ATTCNN, Researchers improve upon the ATTCNN model via adding a second maxpooling layer following the attention convolutional one. The model is composed of a convolutional layer, the maxpooling layer, a flatten layer, as well as a dense layer.

IV. RESULTS & DISCUSSIONS

To assess the effectiveness of the CNN-based strategy, a collection of tweets was compared to the current approaches mentioned in the "Framework for Hate Speech Detection" section: (i) Dataset 1 is a CrowdFlower dataset, (ii) Dataset 2 was previously used by Davidson, and (iii) Dataset 3 is Waseem and Hovy's Sexism against Racism dataset. The CrowdFlower dataset 1 has 14,509 tweets in total. There are a total of 24,783 tweets in Dataset 2. The 3rd dataset, Dataset 36 contains a total of 6492 tweets. Dataset 3 is the most uneven of the three, whereas Dataset 1 is the least lopsided. In Dataset 3, about 86% of the tweets fall into one of three categories. In Dataset 2, offensive tweets account for 77% of the total. However, in Dataset 1, the proportion of objectionable tweets is 50%, not 33%, and around 16% are labelled as hate speech. Experiments are carried out on the Amazon EC2 cloud using the Python packages Keras, Tensorflow, and Sklearn. The classification accuracy was assessed for both multiclass and multilabel classification. To quantify accuracy, precision, recall, and F-measure are utilised as evaluation measures.

4.1 Multiclass Classification Results

Tables 1–3 detail the outcomes of the multiclass categorization process. During our analysis, we noticed that, compared towards the baseline model n-grams only with SVM and indeed the LR with multiple features, this same neural-network-based models, with the exception of the RNN, significantly outperformed in precision scores when individual tells classes. This was especially true when it came to identifying this same hate class in Datasets 1 and 2, and indeed the Sexism class throughout Dataset 3. In contrast to the baseline, nevertheless, neural-network-based models' recall performance is worse in Datasets 2 and 3. Consequently, all of the neural-network-based models with the exception of RNN outperformed the baseline by a little margin. An further key finding of ours reveals that the F-measure scores for neural-network-based models were impacted in recognising the minority class to something like a greater extent the more imbalanced the dataset was. Dataset 2 shows that neural-network-based models that hate speech identification perform somewhat inferior to the baseline model, whereas Dataset 1 shows that LR with many features within F-measure scores works better. Nonetheless, they continue to outperform everyone else in terms of accuracy ratings across the board.

Recent studies have demonstrated that perhaps the model with only an attention mechanism inside the convolutional layers outperforms the model alone without attention mechanism when it comes to text categorization. 16 Throughout contrast, our research showed that utilising the attention mechanism inside the convolution layer, particularly with maxpooling, improves the accuracy score at the expense of the recall score. Consequently, attention convolutional models continue to have a poor average F-score.

Dataset 1									
	Precision			Recall			F-measure		
Classification Technique	Hate	Offensive	Neither	Hate	Offensive	Neither	Hate	Offensive	Neither
Multi-Features LR	0.39	0.95	0.7	0.53	0.92	0.83	0.419348	0.954759	0.758477
n-gram SVM	0.39	0.94	0.7	0.48	0.93	0.82	0.440345	0.984973	0.757263
RNN	0	0	0.51	0	0	1	0	0	0.676497
CNNLSTM	0.48	0.58	0.87	0.46	0.66	0.8	0.479787	0.657419	0.835533
SCNN	0.47	0.65	0.88	0.56	0.58	0.89	0.521068	0.623008	0.883972
CNN-non-static	0.43	0.63	0.94	0.72	0.78	0.7	0.528435	0.677021	0.801439
CNN2D	0.61	0.68	0.86	0.32	0.73	0.95	0.429785	0.764113	0.903762
ATTCNN	0.53	0.65	0.85	0.34	0.66	0.93	0.43	0.67	0.88
ATTCNN-with max	0.65	0.63	0.81	0.08	0.72	0.96	0.15	0.68	0.89

TABLE 1. Multiclass classification results on dataset 1

Dataset 2									
	Precision			Recall			F-measure		
Classification Technique	Hate	Offensive	Neither	Hate	Offensive	Neither	Hate	Offensive	Neither
Multi-Features LR	0.21	0.95	0.87	0.53	0.92	0.83	0.310811	0.935759	0.749529
n-gram SVM	0.3	0.94	0.88	0.48	0.93	0.82	0.359231	0.935973	0.948941
RNN	0	0.78	0	0	1	0	0	0.875404	0
CNNLSTM	0.37	0.91	0.74	0.28	0.92	0.75	0.328769	0.915973	0.754966
SCNN	0.2	0.9	0.71	0.45	0.86	0.67	0.286923	0.875545	0.67942
CNN-non-static	0.52	0.94	0.92	0.09	0.89	0.15	0.163443	0.915317	0.267944
CNN2D	0.58	0.91	0.8	0.16	0.96	0.8	0.260811	0.935332	0.81
ATTCNN	0.47	0.9	0.78	0.14	0.95	0.76	0.23	0.94	0.74
ATTCNN-with max	0.58	0.89	0.81	0.06	0.97	0.7	0.12	0.84	0.78

TABLE 2. Multiclass classification results on dataset 2

Dataset 3									
	Precision			Recall			F-measure		
Classification Technique	Sexism	Racism	Neither	Sexism	Racism	Neither	Sexism	Racism	Neither
Multi-Features LR	0.77	0.26	0.94	0.58	0.6	0.98	0.69	0.13	0.98
n-gram SVM	0.76	0.34	0.95	0.65	0.15	0.98	0.71	0.19	0.98
RNN	0.31	0	0.87	0.08	0	0.99	0.16	0	0.9
CNNLSTM	0	0	0.82	0	0	1	0	0	0.93
SCNN	0.68	0.34	0.88	0.18	0.16	0.99	0.28	0.23	0.93
CNN-non-static	0.82	0	0.95	0.45	0	0.92	0.56	0	0.93
CNN2D	0.79	0.34	0.99	0.49	0.17	0.99	0.62	0.24	0.96
ATTCNN	0.9	0	0.86	0.278	0	0.99	0.43	0	0.95
ATTCNN-with max	1	0	0.92	0.23	0	1	0.13	0	0.93

TABLE 3. Multiclass classification results on dataset 3

4.2 Multilabel Classification Results

Distinguishing between hate speech and offensive speech often becomes difficult for humans as well due to the same usage of the words and very slight distinction between the semantics. Table 4 presents comparison of results of different classifiers. The results demonstrate that the F-measure is highly affected by being tolerant to the false positives. The similar problem occurs in machine learning as well. Due to the overlapping nature of vocabulary used in all the three classes. The case is extreme relaxed on false positives and extreme strict on missing the labels. We obtained the precision score of 1 under all the models except RNN and CNNLSTM. This is due to the reason that we are too relaxed in false positive. Moreover, the recall scores of the convolutional neural network based models have shown significant improvement that has affected the F-measure scores as well. The results show an average increase of 0.095 in F-measure score for the convolutional neural network based models against all the classes. However, the hate class in Dataset 1 has shown the maximum increase of 0.2 in F-measure score as compared to the results in multiclass classification. It is clear from the results that the high number of similar words in the different classes and strict nature of convolutional neural network based models resulted in low recall in multiclass classification but still they predict the correct classes with probabilities higher than the 0.5. Overall, the SCNN has performed consistently well than the other models in the three datasets

Parameters Settings Dataset 1									
	Precision			**Recall**			**F-measure**		
	Hate	**Offensive**	**Neither**	**Hate**	**Offensive**	**Neither**	**Hate**	**Offensive**	**Neither**
RNN	0	1	1	0	1	1	0	1	1
CNNLSTM	1	1	1	0.45	0.65	0.79	0.62	0.79	0.88
SCNN	1	1	1	0.7	0.87	0.92	0.83	0.93	0.96
CNN-non-static	1	1	1	0.72	0.78	0.74	0.84	0.87	0.85
CNN2D	1	1	1	0.32	0.71	0.94	0.48	0.87	0.85
Dataset 2									
	Precision			**Recall**			**F-measure**		
	Hate	**Offensive**	**Neither**	**Hate**	**Offensive**	**Neither**	**Hate**	**Offensive**	**Neither**
RNN	0	1	0	0	1	0	0	1	0
CNNLSTM	1	1	1	0.22	0.88	0.37	0.33	0.93	0.55
SCNN	1	1	1	0.59	0.94	0.73	0.74	0.97	0.85
CNN-non-static	1	1	1	0.09	0.89	0.45	0.17	0.94	0.62
CNN2D	1	1	1	0.32	0.71	0.94	0.25	0.98	0.87
Dataset3									
	Precision			**Recall**			**F-measure**		
	Sexism	**Racism**	**Neither**	**Sexism**	**Racism**	**Neither**	**Sexism**	**Racism**	**Neither**
RNN	1	1	1	0.34	0.14	0.97	0.51	0.24	0.98
CNNLSTM	0	0	1	0	0	1	0	0	1
SCNN	1	1	1	0.44	0.09	0.98	0.61	0.17	0.99
CNN-non-static	1	0	1	0.45	0	0.88	0.62	0	0.94
CNN2D	1	1	1	0.5	0.09	0.97	0.67	0.17	0.99

TABLE 4. Multilabel classification results for different parameters

V. CONCLUSION

Here, we introduced the HateClassify service architecture, which can identify online hate speech. To identify offensive textual material or speech, our HateClassify methodology uses a crowd-sourced technique that polls social media users. The experiments show that perhaps the classification accuracy produced by the CNN models, especially the SCNN, is considerably competitive and even better than numerous state-of-the-art methods, hence CNNs were used to assess the performance in terms of classification. Importantly, this study addresses the issue of hate speech categorization as the multilabel classification problem. Results from experiments using CNN methods for multiclass classification including multilabel classification are promising enough to suggest that these methods might work for identifying hate speech via social media.

REFERENCES

1. F. Del Vigna, A. Cimino, F. Dell'Orletta, M. Petrocchi, and M. Tesconi, "Hate me, hate me not: Hate speech detection on . Facebook," in Proc. 1st Italian Conf. Cybersecurity, 2017, pp. 86–95.
2. W. Warner and J. Hirschberg, "Detecting hate speech on the world wide web," in Proc. 2nd Workshop Lang. Social Media, 2012, pp. 19–26.
3.T. Davidson, D. Warmsley, M. Macy, and I.Weber,"Automated hate speech detection and the problem of offensive language," in Proc. 11th Int. AAAI Conf. Web Social Media, 2017, pp. 512–515
4. Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting offensive language in social media to protect adolescent online safety," in Proc. IEEE Int. Conf. Privacy, Secur., Risk Trust Int. Conf. Soc. Comput., 2012, pp. 71–80.
5. H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Analyzing labeled cyberbullying incidents on the instagram social network," Social Inform., T. Y. Liu, C. N. Scollon, and W. Zhu, Eds., 2015, pp. 49–66.
6. Z. Waseem and D. Hovy, "Hateful symbols or hateful people? predictive features for hate speech detection on Twitter," in Proc. NAACL Student Res. Workshop, 2016, pp. 88–93.
7. C. Van Hee et al., "Detection and fine-grained classification of cyberbullying events," in Proc. Int. Conf. Recent Adv. Natural Lang. Process., 2015, pp. 672–680.
8. J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore, "Learning from bullying traces in social media," in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol., 2012, pp. 656–666.
9. G. Xiang, B. Fan, L. Wang, J. Hong, and C. Rose, "Detecting offensive tweets via topical feature discovery over a large scale Twitter corpus," in Proc. 21st ACM Int. Conf. Inf. Knowl. Manage., 2012, pp. 1980–1984.
10. H. Zhong et al., "Content-driven detection of cyberbullying on the Instagram social network," in Proc. Int. Joint Conf. Artif. Intell., 2016, pp. 3952–3958.
11. N. Djuric, J. Zhou, R. Morris, M. Grbovic, V. Radosavljevic, and N. Bhamidipati, "Hate speech detection with comment embeddings," in Proc. 24th Int. Conf. World Wide Web, 2015, pp. 29–30.
12. P. Burnap and M. L. Williams, "Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making," Policy Internet, vol. 7, no. 2, pp. 223–242, 2015.
13. C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang, "Abusive language detection in online user content," in Proc. 25th Int. Conf. World Wide Web, 2016, pp. 145–153.
14. Y. Mehdad and J. Tetreault, "Do characters abuse more than words?" in Proc. 17th Annu. Meeting Special Interest Group Discourse Dialogue, 2016, pp. 299–303.
15. Y. Kim, "Convolutional neural networks for sentence classification," in Proc. Conf. Empirical Methods Natural Lang. Process., 2014, pp. 1746–1751.
16. Muhammad U. S. Khan, Assad Abbas, Attiqa Rehman and Raheel Nawaz, "HateClassify: A Service Framework for Hate Speech Identification on Social Media" in IEEE Internet Computing, vol. 25, no.1,pp. 40-49.

eISSN 1303-5150