iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** II **Month of publication:** February 2025

DOI: https://doi.org/10.22214/ijraset.2025.66791

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue II Feb 2025- Available at www.ijraset.com

Hate Speech Detection Using LSTM and Machine Learning Models

Ms. Purvi Rawal [1], Ms. Anisha Asirvatham [2],
[1]*Student, Department of MSc.IT, Nagindas Khandwala College, Mumbai, Maharashtra, India*
[2]*Assistant Professor, Department of Computer, and Information Science, Nagindas Khandwala College, Mumbai, Maharashtra, India*

Abstract*: The hate speech detection system in the paper employs machine learning and deep learning to improve performance accordingly on aspects such as classification accuracy and generalization. Some of the tried machine learning models include Logistic Regression, SGD Classifier, Decision Tree Classifier, Random Forest Classifier, and XGB Classifier, whose metrics are accuracy, precision, recall, F1-score, and the highest performance measured was 88% for the Random Forest Classifier because of ensemble learning itself. A deep learning model using the architecture of the Bidirectional LSTM has been implemented with embedding layers to handle semantic representation, and a Bidirectional LSTM layer captures the contextual relations of the text data. The proposed model was optimized through RMS Prop, using categorical cross-entropy loss, is robustly trained in terms of accuracy and loss curves, has smooth convergence without any trace of overfitting, hence ensuring good generalization on unseen data. The accuracy was further supported by using an evaluation consisting of a confusion matrix, test predictions, where two most important metrics were well classifying between "None" and "Offensive and Hate Speech." Accordingly, the system strikes an excellent balance between model complexity and interpretability, then applies to content moderation, policy enforcement across various social and online platforms.

Keywords: LSTM, Machine learning, hate speech, deep learning

I. INTRODUCTION

Hate speech detection with deep learning involves the application of sophisticated neural network architecture to classify text as hate speech or non-hate speech. Some common deep learning models, including LSTM networks, are widely used because they can grasp complicated patterns in textual data. These models process text data through numerical representations, such as word embeddings, including but not limited to Word2Vec, GloVe, or fast Text, which preserve semantic meaning. Among them, especially LSTMs, demonstrate exceptional performance in modelling sequential and contextual properties of languages, hence becoming highly effective in hate speech detection, where context plays an important role. The deep learning pipeline of our system starts with the pre-processing of data-text cleaning, tokenization, and padding. It allows the creation of a solid framework for the detection of hate speech that considers both the intricacies of language and scalability for enhanced performance and applicability to the real world. Deep learning methods for hate speech detection marry traditional machine learning approaches, such as logistic regression, random forest, and XG Boost, with modern complex architectures represented by LSTMs. LSTM is a kind of deep learning for sequence data, intuitively very good at observing contexts and temporal dependencies, and thus works well in all analysis tasks related to hate speech. LSTM and the structured prediction capability of machine learning classifiers for high accuracy and robustness in detection of hate speech for research purposes and in practical applications.

II. PROPOSED SYSTEM

The proposed hate speech detection system integrates the power of machine learning and deep learning to get the most robust and accurate results on diverse data. Several machine learning classifiers, including but not limited to Logistic Regression, SGD Classifier, Decision Tree Classifier, Random Forest Classifier, and XGB Classifier, were preliminary evaluated in terms of their accuracy, precision, recall, and F1-score. Logistic Regression was the model that stood for average performance, the SGD Classifier for efficiency when dealing with large data was a bit lower on accuracy with 86%, the Decision Tree Classifier offered good interpretability and feature importance insights while overfitting and ultimately yielding the same accuracy. It exhibited the best performance among these models, with an 88% accuracy rate: a Random Forest Classifier was ensemble learning that showed good handling different data patterns with well-balanced precision and recall and was good for minimizing both false positives and negatives. Secondly, this paper used the XGB Classifier, which used a gradient boosting method, as part of model selection. The model was then compiled with categorical cross-entropy loss and showed strong learning.

©IJRASET: All Rights are Reserved

III. LITERATURE REVIEW

Al-Makhadmeh. Z, and Tolba. A, (2020) This study shows in aiming for feature value conformance during hate speech prediction from different social media websites, it employs a hybrid method using state-of-the-art natural language processing with any machine-learning technique. Following the collections of hate speech, for instance, steaming of samples, token splitting, deletion of characters, and smoothing out inflection are undergone before the hate speech. The paper discussed here puts hate speech from social media sites through some effective learning processes and obtained accuracy of 98.71%.

Aluru. S.S, et al (2020) In this research work, we perform the first large-scale study of multilingual hate speech. We employ deep learning models to develop classifiers for multilingual hate speech classification using 16 datasets from 9 languages. We conduct many experiments under various conditions, such as low and high resource, monolingual, and multilingual settings for several languages. We did a large-scale analysis. We observe that in low resource setting, simple models such as LASER embedding with logistic regression performs the best, BERT-based models perform better.

Alshalan. R, and Al-Khalifa. H, (2020) This paper showed that with the rising of the hate speech phenomena in Twittersphere, quite a significant volume of research effort has been given to develop an automated solution that identifies hate speech-right from simple ones based on the machine learning model to models based on deep neural network architecture. In conducting experiments, for hate speech that contains 9316 labelled tweets. The best performance of the CNN model could be underlined. The F1-score equals 0.79 with 0.89 AUROC for dataset.

Al-Hassan. A, and Al-Dossari. H, (2022) The Author had mentioned about the people are communicating through social networks everywhere. Whatever the reason, it is noticeable that verbal misbehaviours like hate speech is now propagated through the social networks. This research targets the identification and classification classes. The results indicate that all 4 deep learning models outperform the SVM model in detecting hateful tweets. Although the SVM achieves an overall recall of 74%, the deep learning models. While adding a layer of CNN to LTSM enhances the of detection with 72% precision, 75% recall, and 73% F1 score.

Alatawi. HS, et al (2021), The author had stated that supremacist hate speech is one of the latest harmful contents observed in social media. The important influence of these radical groups is no longer confined to social media. Hence, there is a need to automatically detect this kind of speech and do it timely. First is the BiLSTM model combined with domain-specific word embeddings pre-trained from a white supremacist corpus to grasp semantic capture of white supremacist slang and coded words. For both these models, and as for BERT, this had a 0.80 F1 score. Both models have been tested on a combined balanced dataset originating from Twitter and Stormfront dataset, compiled from the white supremacist forum.

Ali. R, et al (2022) The Research shows that it because the huge amount of data generated on these platforms reaches the far corners of the world in the blink of an eye. The developers of these platforms are facing a few challenges to keep the cyber space as inclusive and healthy as possible. As a matter of fact, there is a need that an automated technique is designed which detects and removes offensive and hateful comments before their harmful impacts. We then explore transfer learning by experimenting with four variants of BERT: we show that BERT can achieve 0.69 encouraging F1-scores respectively on our multiclass classification task.

Cao. R, et al (2020) The Author had mentioned about the cohesiveness of online social communities and even raises public safety concerns in our societies. Motivated by this rising issue, several traditional machine learning and deep learning methods have been developed to detect hate speech in online social platforms automatically. In this paper, we propose Deep Hate, a novel deep learning model that fuses multi-faceted text such as word embeddings, sentiment platforms. We also perform case studies to provide insights into the salient features that best aid in detecting hate speech in online social platforms.

Faris. H, et al (2020) The study done by the author had shown that Cyber hate speech is a critical serious problem worldwide. Very few studies have investigated the problem of hate speech detection over online networks. This paper interpreted hate speech detection on Twitter, where A dataset is collected and pre-processed using the NLTK library. This research brought to light such uncharted challenges for further research. For example, the need for more comprehensive lexical resources of abusive, offensive expressions as well as to go deeper into deep learning approaches but investigate the optimization within the deep learning.

Jahan. MS, and Oussalah. M, (2023) The paper had review that through social media platforms that offer anonymity, ease of access, the creation of online communities, and online debate, the issue of detection and tracking of hate speech is turning out to be an ever-increasing challenge for society, individuals, policymakers, and researchers alike. This paper systematically reviews the literature in this field, focusing on NLP and deep learning technologies for terminology, processing pipeline, core methods used, with focus deep learning architecture. The existing surveys, their, and future directions of research have been widely discussed in the sequel.

Kapil. P, et al (2020) The Author had said that the phenomenal growth of the internet, social media networks, and messaging platforms has provided great opportunities for the intelligent system to build on immense opportunities.

©IJRASET: All Rights are Reserved

In recent times, although a few benchmark datasets have come up for hate speech detection, these are limited in volume and do not follow any uniform annotation schema. In this paper, it is proposed that the deep multi-task learning framework takes leverage of useful information among the multiple related classification tasks with the improvement in an individual task's performance. Experiments on the 5 datasets testify that the proposed framework behaves very well both on Macro-F1 and Weighted-F1.

Khan. S, et al (2022) The study shows that the dominating approaches for the detection of hate speech are the machine learning-based approaches in the existing set of approaches with deep CNN and Hierarchical Attention-based deep learning model for tweet representation learning toward hate speech detection. This proposed model takes the tweets as input and passes through a BERT layer followed by an attention-aware deep convolutional layer. From this analysis, we observe that the deep convolutional layer exclusion bears the highest impact on the performance of the proposed model. We further investigated different embedding techniques, activation functions, batch sizes, and optimization algorithms for effectiveness.

Mossie. Z, and Wang. JH, (2020) The research paper had stated that volume of significant research attention has been paid to automated hate speech detection. Though hate speech may affect all population groups, some population groups are more vulnerable to its impact compared to others. Feature extraction is done through Word n-grams, and the Word embedding techniques used include Word2Vector. Among our experimental results, feature extraction using word embedding techniques such as Word2Vec performs among which GRU achieves best result. It thus forms the vulnerable group that has been subjected to hatred, hence it's in the removal of contents that psychological harm and physical conflicts. Mullah. N.S, and Zainon. W.M.N, (2021) This paper had reviewed the machine learning algorithms and techniques for hate speech detection in social media. The problem of hate speech is usually modelled as a text classification task. In this paper, we investigate the basic baseline components of hate speech classification using ML algorithms. Three-fold is the contribution of this work in providing the information necessary in critical steps done in the detection of hate speech using the ML algorithms. Lastly, some research gaps and open challenges have been identified. A review of the different variants of ML techniques is done: classical ML, ensemble approach, and deep learning methods. Roy. P.K, et al (2020) The recent studied had targeted how the problem could be solved by using machine learning and deep learning methods. In one regard, the methodology has involved the use of DCNN for hate speech detection. A proposed DCNN model utilizes tweet text combined with GloVe embedding vectors to capture the semantic meaning of tweets. This proposed model is able, through convolution operations, to effectively extract patterns indicative of hate speech. In future studies, improvement of model performance could be proceeded, and the adapted approach can move to more applications toward cyber safety.

Zhou. Y, et al (2020) This research paper shows that most of the current work of hate speech detection is implemented in three kinds of principles: text classification method-ELMo, BERT, and CNN-putting into raise the performance through the following two ways of fusion. First one is the fusion of ELMo, BERT, and CNN classification results, and the second is the fusion of three different parameter classifications which is based on the simple neural network. So, the results have shown that fusion processing is one promising way to enhance hate speech detection performance.

IV. METHODOLOGY



Fig 1. Methodology Model

A. Data Collection

Data collection has been taken from twitter according to the users and the other features as well so that it can be further divided for analysis what kind of hate comments have been extracted through positive, negative and neutral tweets and therefore formed a very integral aspect of this research in developing a high-quality and representative dataset for, say, the detection of hate speech on Twitter. Data is collected primarily from Twitter through utilisation of tools such as the Twitter API, which allowed the collection of textual data in large volumes and real-time. Additionally, for collecting data, some annotated and publicly available datasets mention the dataset names whenever relevant were added into theirs to increase diversity and scale by expanding the scope of this present dataset. Data annotation was performed for labels in supervised learning.

B. Data Pre-processing

Preprocessing in cleaning up the text from its noisy raw state into a structured, clean one required for eventual analysis and modelling.The below fig.2 shwon the sentimental between which are positive, negative as well as neutral. Such data is bound to arise when the origin of texts is social media due to informal language use, so solving some tasks, as inconsistent formatting of text elements, slang expressions, abbreviations, were at stake. The preprocessing involves tokenization to break down the text into single words, removing the stop words unless they are domain-specific.

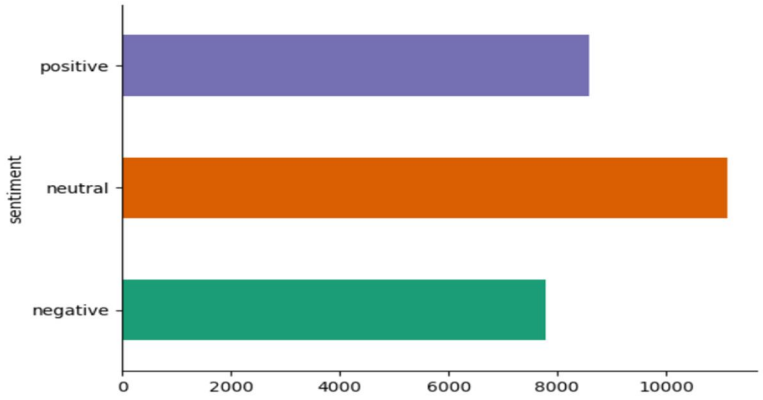


Fig. 2 Sentimental Tweets

C. Feature Extraction

Feature extraction was an important step to turn pre-processed text into meaningful numerical representations that could be analysed by machine learning models. Based on the process flow diagram, the embedding of words into dense vectors using GloVe would represent both semantic relationships and context based on co-occurrence in a large corpus. These word-level embeddings are further combined at sentence level by using techniques such as average pooling or concatenation into one feature vector for every tweet that captured the overall meaning. Besides, some domain-specific features were extracted such as frequency of offensive words, sentiment polarity, and tags that provided additional insight nature of text.

D. Modelling

This step involves the training and testing of the machine learning model that will classify the tweets correctly into categories such as hate speech and neutral using extracted features. First, some classical machine learning models, including logistic regression, random forest, and XG Boost, are trained based on the extracted feature vectors from the embeddings of GloVe along with domain-specific attributes:

1) Random Forest: The Random Forest algorithm, to which the above process flow diagram was applied, implements an ensemble learning technique by combining many decision trees. It is the improvement in the accuracy of classification with added robustness due to a collection of many trees, each one trained on the bootstrapped subset.
2) Logistic Regression: It used a sigmoid activation function to map the linear output into a range of 0 to 1, thus providing probabilistic predictions that were easily interpretable. Hyperparameters such as regularization strength-L1 or L2-were tuned to prevent overfitting and enhance generalization.
3) XG Boost: It was done cautiously to arrive at an appropriate balance between accuracy and computational efficiency. In particular, handling missing values and emphasis on informative features makes XG Boost very powerful in classifying tweets into hate speech and neutrality; it outperformed lot of traditional models in area.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue II Feb 2025- Available at www.ijraset.com

E. Training and Evaluation

Preprocessing, cleaning, and data transformation in accordance with what will be needed by the model after the training is complete, perform the evaluation test dataset to establish the performance metrics in the form of accuracy, precision, recall, mean squared error, etc. Another important aspect is the robustness of the model, which can be assured by cross-validation. Basically, there are a few critical steps for training and testing any LSTM model. First, pre-processing should be done on the data. LSTM layers followed by subsequent dense layers for the overall output prediction. The performance metrics evaluate the model on accuracy.

V. RESULTS

Results shown in the below fig. 3 represent form of various metrics for different machine learning classifiers are a very apt basis on which to build a description of these and those of their implications: Logistic Regression, SGD Classifier, Decision Tree Classifier, Random Forest Classifier, and XGB Classifier. Performance metrics such as accuracy, F1-score, precision, and recall act as very important pointers to the effectiveness of various machine learning models when classifying. Logistic Regression, in turn, has average performance and serves as a good baseline to compare with more complex classifiers that manage to show accuracy at the level of 87%, while the SGD Classifier reaches slightly lower metrics of accuracy 86 % and F1-score. Interpretable and apt to identify feature importance. The Random Forest Classifier is an ensemble learning method in which numerous decision trees are combined for reducing overfitting, hence generalizing better. Its accuracy is 88% against that of the Decision Tree. XGB Classifier by gradient-boosting algorithms demonstrates relatively good competitiveness, too, with an accuracy of only 86% taken together, these results suggest that model complexity and interpretability are bound by trade-offs in performance.

```
LogisticRegression
        Fold 0
        Fold 1
        Fold 2
        Fold 3
        Acc:  0.87146 | F1:  0.88062 | Precision :  0.89482 | Recall:  0.87146
SGDClassifier
        Fold 0
        Fold 1
        Fold 2
        Fold 3
        Acc:  0.86558 | F1:  0.87663 | Precision :  0.89446 | Recall:  0.86558
DecisionTreeClassifier
        Fold 0
        Fold 1
        Fold 2
        Fold 3
        Acc:  0.86760 | F1:  0.87697 | Precision :  0.88973 | Recall:  0.86760
RandomForestClassifier
        Fold 0
        Fold 1
        Fold 2
        Fold 3
        Acc:  0.88912 | F1:  0.88946 | Precision :  0.89004 | Recall:  0.88912
XGBClassifier
        Fold 0
        Fold 1
        Fold 2
        Fold 3
        Acc:  0.86038 | F1:  0.87546 | Precision :  0.90463 | Recall:  0.86038
```

Fig 3: Evaluation metrics for classification algorithms.

The architecture is implemented using the Sequential API of Keras as shown in the below fig.4. This API allows neural network components to be stacked together layer by layer. First, there is an embedding layer, which projects input text data into a dense vector representation. The embeddings carve out semantic relationships among words and are hence vital for further modelling. Next to this embedding layer, the bidirectional LSTM layer captures in both directions the information, being thus crucial for retrieving contextual information. This stands for the great fit of the bidirectional approach when it comes to hate speech detection tasks that understand how words relate within a given context. First, the first dense layer consists of 16 units, using the Reu activation function that permits the model to learn nonlinear combinations of the features extracted by the LSTM. Coming to optimizer, Respro is utilized with a learning rate of 0.0001 that adjusts this value through training in its search for effective convergence towards a better optimum. Finally, as metric accuracy takes place in evaluating the portion of properly classified instances.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Model: "sequential_2"

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	?	0 (unbuilt)
bidirectional_2 (Bidirectional)	?	0 (unbuilt)
dense_4 (Dense)	?	0 (unbuilt)
dense_5 (Dense)	?	0 (unbuilt)

Total params: 0 (0.00 B)
Trainable params: 0 (0.00 B)
Non-trainable params: 0 (0.00 B)

Fig 4: Layer-wise Summary of the Sequential Model

Firstly, the training and validation curves have been used to show the performance and generalization of the LSTM model during training. From the accuracy graph (left), in fig. 5 one can understand that both the training and validation accuracy increased quite smoothly with an increase in the number of epochs. This is further complemented by the loss graph on the right, showing a continuing decrease in both training and validation losses with time. The decrease in loss is faster within the training set, which, of course, is expected because the model keeps optimizing its parameters based on this data. The validation loss is following the same trend-the model does not overfit the training data, and hence generalizes well on the validation set. The closeness of the training and validation curves on both accuracy and loss underlines that the model is robust; it shows that most probably the hyperparameters, including learning rate, batch size, and dropout, had been tuned properly. Slight gaps between these two are natural and may be contributed to by noise and fluctuations in the data. However, further evaluation on an independent test set is recommended to confirm the model's generalization capabilities.



Fig 5: Training and Validation Accuracy of LSTM

The image of fig.6 shows the confusion matrix, probably one of the most usable forms of a visualization that one might do in machine learning. The confusion matrix normally would provide some performance results from the classifier. The matrix presented in this figure has a model that predicts on three classes: 0, 1, and 2. For instance, the cell on the far top left shows that none of the instances from the true class "0" were predicted by the model as "0.". Yellow-marked cells on the diagonals in this matrix represent the correct predictions for each class. In those cells, the numbers show how the model performed with respect to classifying the instances into their actual classes. By analysing the distribution of values within the confusion matrix, one can therefore gain insights into model strengths and weaknesses, understand possible biases, and areas for improvement.

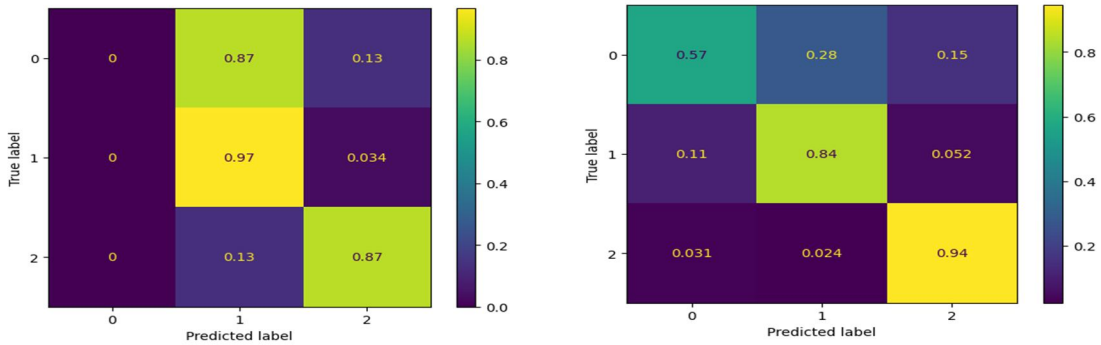


Fig 6: Classification Performance

The below fig.7 presents an example output of the hate speech and offensive language detector, which classifies the input text according to two classes: "None," representing the absence of hate speech or offensive language, and "Offensive and Hate Speech," upon the detection of either category. The output shows some test data with the model's predictions against ground truth labels. This tell about the performance of the system and how it can be improved. From this evaluation, key performance metrics such as precision, recall, and F1-score could be derived that quantify the effectiveness of the system in accurately detecting and classifying hate speech and offensive language.

```
Hate speech and offensive language detention algorithm.
Model prediction:
@CNN The weather forecast predicts sunshine and clear skies for tomorrow.   : predicted as None
I enjoy reading books and spending time in nature #selflove.                : predicted as None
Spending my weekends relaxing at home with a good movie with @sarah.        : predicted as None
I can't stand those immigrants ruining our country fuck them!!!            : predicted as offensive and hate speech
@melanie You damn broke bitch you should die LOL                           : predicted as offensive and hate speech
```

Fig 7: Model Prediction

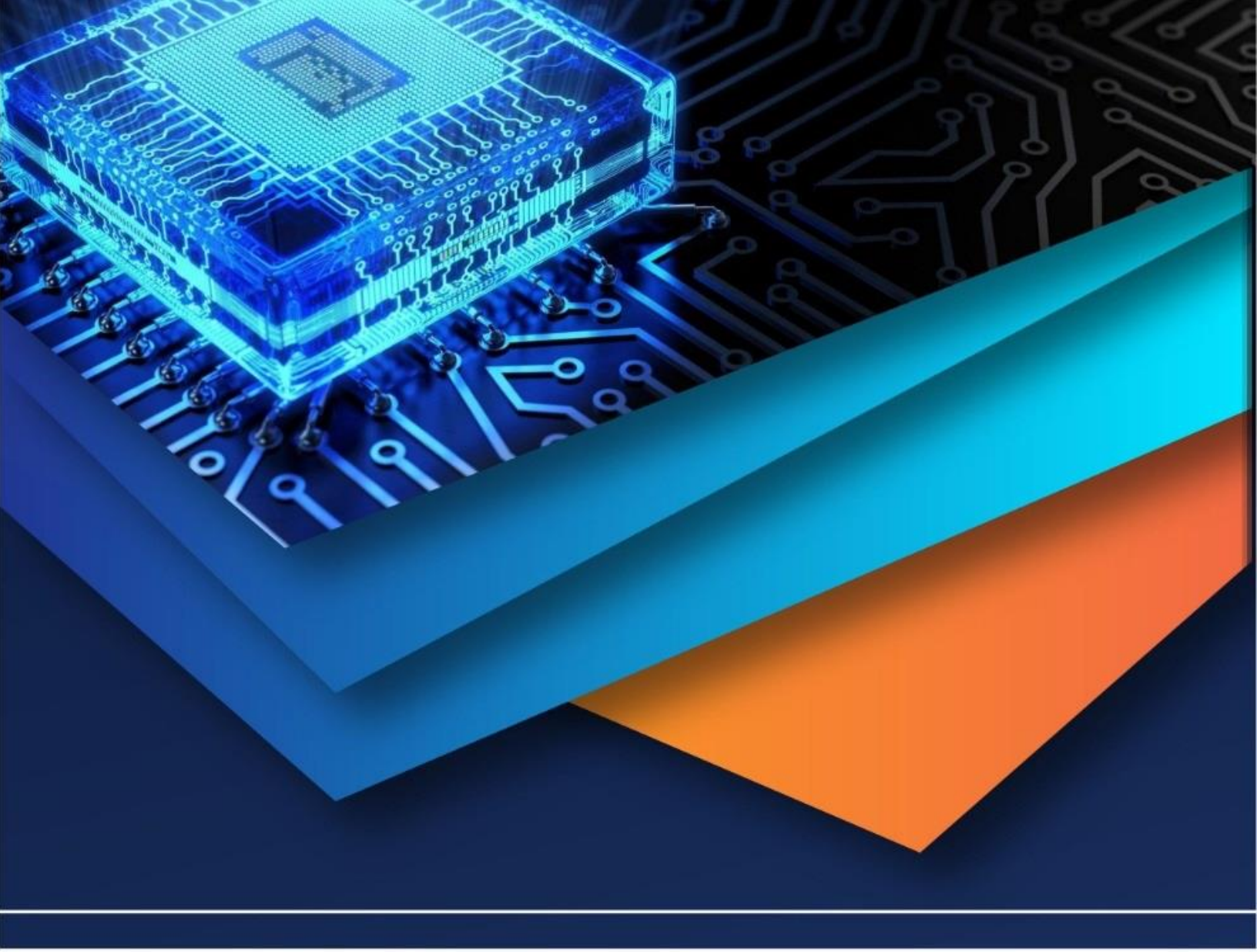
VI. CONCLUSION

In this paper, we developed a robust hate speech detection system that combines machine learning and deep learning techniques to achieve high accuracy and generalization. Among the different machine learning models explored, the Random Forest Classifier performed best with an accuracy of 88%, benefiting from ensemble learning, while the XGB Classifier was very competitive, especially on imbalanced datasets. To achieve better semantic understanding, a Bidirectional LSTM model with embedding layers was implemented that captured contextual relationships of text. Optimized by RMSprop and categorical cross-entropy loss, the model trained marvellously without overfitting and is believed to be reliable on unseen data. The detailed performance using accuracy, precision, recall, F1-score, and confusion matrix showed the ability of the system to classify "None" and "Offensive and Hate Speech" correctly. Considering the trade-offs between complexity and interpretability, this system scales to a reliable solution in detecting hate speech with good practical applications in content moderation and policy enforcement across different platforms

REFERENCES

[1] (Beddiar, 2021) Beddiar, D. R., Jahan, M. S., & Oussalah, M. (2021). Data expansion using back translation and paraphrasing for hate speech detection. Online Social Networks and Media, 100, 153. https://doi.org/10.1016/j.osnem.2021.100153
[2] (Corazza, 2020) Corazza, M., Menini, S., Cabrio, E., Tonelli, S., & Villata, S. (2020). Multilingual evaluation for online hate speech detection. ACM Transactions on Internet Technology (TOIT, 20). https://doi.org/10.1145/3377323
[3] (Das, 2021) Das, A., Al Asif, A., Paul, A., & Hossain, M. (2021). Bangla hate speech detection on social media using attention-based recurrent neural network. Journal of Intelligent Systems, 30(1), 578–591. https://doi.org/10.1515/jisys-2020-0060
[4] (Kovács, 2021) Kovács, G., Alonso, P., & Saini, R. (2021). Challenges of hate speech detection social media Computer Science https://doi.org/10.1007/s42979-021-00457-3
[5] (Luu, 2021) Luu, S. T., Nguyen, K. V., & Nguyen, N. L. T. (2021). A large-scale dataset for hate speech detection Vietnamese media texts. In Fujita, H., Selamat, A., Lin, J. C. W Intelligence:(Vol. 12798). https://doi.org/10.1007/978-3-030-79457-6_35
[6] (Mittal, 2023) Mittal, U. (2023). Detecting hate speech utilizing deep convolutional network and transformer models. International Conference Electrical, Electronics, Communication and Computers (ELEXCOM) (pp. 1–4). IEEE. https://doi.org/10.1109/ELEXCOM58812.2023.10370502
[7] (Modha, 2020) Modha, S., Majumder, P., Mandl, T., & Mandalia, C. (2020). Detecting and visualizing hate speech in social media: A cyber watchdog for surveillance. Expert Systems with Applications, 161, 113725. https://doi.org/10.1016/j.eswa.2020.113725
[8] (Mollas, 2022) Mollas, I., Chrysopoulou, Z., Karlos, S., & Tsoumakas, G. (2022). ETHOS: A multi-label hate speech detection dataset. Complex & Intelligent Systems, 8, 4663–4678. https://doi.org/10.1007/s40747-021-00608-2
[9] (Naseem, 2021) Naseem, U., Razzak, I., & Eklund, P. W. (2021). A survey of pre-processing techniques to improve short-text quality: A case on hate speech detection on Twitter. and Applications, 35239–35266. https://doi.org/10.1007/s11042-020-10082-6
[10] (Oriola, 2020) Oriola, O., & Kotzé, E. (2020). Evaluating machine learning techniques for detecting offensive and hate speech in South African tweets. IEEE Access, 8, 21496–21509. https://doi.org/10.1109/ACCESS.2020.2968173
[11] (Plaza-del-Arco, 2021) (2021). Comparing pre-trained language models for Spanish hate speech detection. Expert Systems with Applications, 166, 114120. https://doi.org/10.1016/j.eswa.2020.114120
[12] (Romim, 2021) Romim, N., Ahmed, M., Talukder, H., & Islam, M. S. (2021). Hate speech detection in the Bengali language: Proceedings of International Joint Conference on Advances in Springer. https://doi.org/10.1007/978-981-16-0586-4_37
[13] (Roy, 2022) Roy, P. K., Bhawal, S., & Subalalitha, C. N. (2022). Hate speech and offensive language detection in Dravidian languages using deep ensemble framework. Computer Speech & Language, 74, 101386. https://doi.org/10.1016/j.csl.2022.101386
[14] (Tontodimamma, 2021) Tontodimamma, A., Nissi, E., Sarra, A., & Santucci, L. (2021). Thirty years of research into hate speech: Topics of interest and their evolution. Scientometrics, 126, 157–179. https://doi.org/10.1007/s11192-020-03737-6
[15] (Velioglu, 2020) Velioglu, R., & Rose, J. (2020). Detecting hate speech in memes using multimodal deep learning approaches: Prize-winning solution to Hateful Memes Challenge. arXiv. https://doi.org/10.48550/arXiv.2012.12975

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)