ISSN: 2347-971X (Print)
ISSN: 2347-9728(Online)

International Journal of Innovations in Scientific and Engineering Research (IJISER)

SPEECH AND HATE SPEECH RECOGNITION SYSTEM USING ML

[1*]N. Krishnamoorthy, [2]Preeti Gaikwad

[1*]Associate Professor Grade 2, Department of Software Systems and Engineering, School of Computer Science and Engineering ,Vellore Institute of Technology, Vellore, Tamil Nadu, India, Email: krishnamoorthy.n@vit.ac.in
[2] School of Information Technology and Engineering ,Vellore Institute of Technology,Vellore, Tamil Nadu, India. Email: preetigaikwad2316@gmail.com

Abstract- his research paper delves into the intricate realm of leveraging AI and machine learning methodologies to augment the capabilities of speech recognition systems. The investigation encompasses a wide array of techniques, with a particular emphasis on harnessing natural language processing to not only elevate the precision of recognition but also to fortify its resilience across various scenarios. A significant facet of this exploration involves a comprehensive analysis of Python libraries, notably the likes of pyttsx3. These libraries serve as instrumental tools for constructing sophisticated speech recognition systems that boast an assortment of functionalities. Among these features, the paper scrutinizes language support, affording the system the aptitude to comprehend and decipher an array of languages, thus making it accessible to diverse user bases. Furthermore, the paper delves into mechanisms for dynamically adjusting the volume of speech, enhancing the adaptability of the system to varying acoustic environments. In a broader societal context, the paper does not shy away from addressing the burgeoning concern of hate speech proliferation in online spaces. It underscores the pressing necessity for effective speech recognition systems to identify and counteract hate speech, thus cultivating a secure and welcoming digital milieu for users. As the research journey forges ahead, the paper also contemplates the promising trajectory of future research avenues. One particularly intriguing trajectory is the exploration of deep learning algorithms further to refine the precision and efficacy of speech recognition systems. This strategic direction reflects the continuous evolution of AI and machine learning techniques, propelling the field towards ever-greater achievements in understanding and interpreting human speech. The Decision Tree, KNN, and Logistic Regression methods considered for the experiment, out of which Logistic Regression has the highest accuracy.

Keywords: PYTTSX3, Speech Recognition, Machine Learning, KNN, Logistic Regression, Decision Tree

1 INTRODUCTION

The field of speech recognition has witnessed remarkable advancements in recent years, primarily driven by the rapid progress in Artificial Intelligence (AI) and Machine Learning (ML) techniques.[1] These advancements have revolutionized the accuracy and robustness of speech recognition systems, enabling their wide-ranging applications across various domains [2]. This paper aims to delve into the utilization of AI and ML techniques to enhance speech recognition performance, with a specific focus on the urgent need for effective hate speech recognition systems in the digital age[3-4]. The paper commences by exploring the major approaches employed in speech recognition, including the integration of natural language processing techniques, to enhance the overall accuracy and reliability of these systems. Additionally, it delves into the utilization of Python libraries, particularly the versatile pyttsx3 library, in the construction of robust speech recognition systems [5]. This library's various features and functionalities, such as support for multiple languages, voice selection, and volume adjustment, are analysed in detail to showcase their practical applications in virtual assistants and dictation software. However, as speech recognition technology continues to evolve, the proliferation of loathing on social media platforms and other online spaces has presented new challenges. The alarming rise of discriminatory and harmful language necessitates the development of effective hate speech recognition system that can swiftly identify and flag such content [6-10]. This paper underscores the imperative need to address this issue and highlights the potential of AI and ML techniques in building robust hate speech recognition systems. Furthermore, the paper emphasizes the significance of Python libraries and machine learning algorithms, particularly deep learning, in improving the accuracy and efficiency of hate speech recognition. By harnessing the capabilities of these technologies, researchers can contribute to creating a safe and inclusive online environment by proactively detecting and preventing hate speech [10-14].

2 RELATED WORKS

Several recent studies have made notable contributions in the areas of speech recognition and hate speech detection. In the realm of speech recognition, several techniques have made a significant impact. Self-supervised pre-training [1] has proven to be effective in leveraging large amounts of unlabeled data to improve model performance. Structured temporal dependency [2] has helped capture long-range dependencies in speech sequences, enhancing the accuracy of recognition systems. Additionally, research on data augmentation [3] has shown its potential to augment training data and improve generalization. Architectural innovations have also played a crucial role. The Conformer Transformer architecture [4] and the Context Net architecture [5] have demonstrated their effectiveness in capturing complex acoustic patterns and contextual information. Techniques such as multi-head attention [6], the Jasper model [7], and the hybrid CTC-attention architecture [9] have further contributed to achieving state-of-the-art results in speech recognition. Evaluating techniques like Spec Augment [8] have provided valuable insights into improving the robustness of recognition systems. In hate speech detection, researchers have made strides in improving performance and mitigating biases. Adversarial training and auto-augmentation [10] have shown promise in enhancing the robustness of hate speech detection models. Transfer learning with CNN [11] and attention-based CNN [12] have leveraged pre-trained models and attention mechanisms to capture relevant features in text data. Multimodal fusion [13] has been explored to incorporate additional contextual information from multiple modalities.

Adaptive attention-based LSTM [14] has proven effective in dynamically attending to important features, while bias mitigation approaches [15] have aimed to address the challenges of biased datasets in hate speech detection. Various studies spanning machine learning and deep learning applications such as hate speech detection on Twitter [16], product recommendation systems [17], crime prediction [18], image classification using deep learning [19], disease prediction in agriculture [20], and medical image classification [21]. Additionally, the survey encompasses research on healthcare, network security, and emerging technologies, including heart disease detection [22], quantum key distribution [23], trust certificate distribution in mobile ad hoc networks [24], trust-aware routing protocols [25], secure ad hoc networking [26], signcryption schemes for IoT security [27], and collaborative filtering for recommendation systems [29]. Furthermore, it includes publications related to the application of AI, IoT, and cognitive technologies [28], E-collaboration during global crises [30], and management information systems [31]. Overall, this literature survey offers a comprehensive overview of recent research endeavors across various domains of computer science and technology.

3 PROPOSED WORK

The dataset used in this project, obtained from Kaggle, is a Twitter dataset comprising a training set (x_train: 19826 samples, 30588 features) and a corresponding set of labels (y_train: 19826 samples), as well as a test set (x_test: 4957 samples, 30588 features) with corresponding labels (y_test: 4957 samples) [16].

Tweet: Tweets contain text, hashtags, mentions of other users, and URLs. Each tweet typically contains several attributes, such as the actual text of the tweet, the timestamp of when it was posted, the username of the user who posted it, the number of retweets and likes it received, and other metadata associated with the tweet[17]. Labels: refer to the assigned categories or tags that are used to classify or annotate the tweets based on 3 categories and shown in Figure 1.

1) Hate speech
2) Offensive language
3) Neither

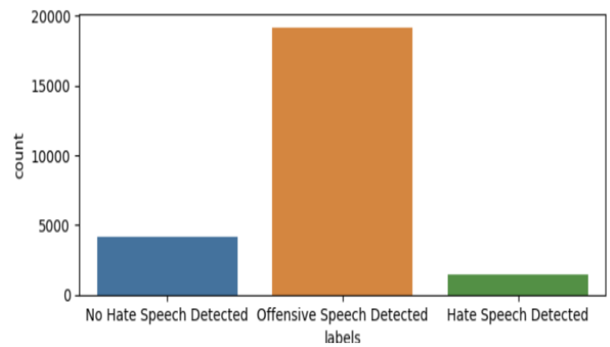

Figure 1 Distribution of categorical value from a dataset[18]

The development process involved leveraging key Python libraries such as NumPy, Seaborn, NLTK, Scikit-learn, and Matplotlib. Pandas facilitated data manipulation, NumPy handled numerical computing tasks, Seaborn enabled insightful data visualizations, NLTK provided natural language processing capabilities, Scikit-learn supported machine learning algorithms, and Matplotlib facilitated versatile data visualization. These libraries collectively enhanced data analysis, modeling, and interpretation [19-20].

The first model we used to analyse the dataset is a Decision Tree, which is a machine learning algorithm used for classification tasks, including speech recognition. It works by constructing a tree-like model of decisions and their possible consequences [21]. The dataset is properly labeled and structured. Split the dataset into training and testing sets to evaluate the model's performance. Import the necessary libraries for Decision Tree (e.g., sci-kit-learn in Python). Instantiate a Decision Tree model with desired hyperparameters. Train the model on the training dataset using the fit() function [22]. During the training phase, the Decision Tree model learns to make decisions based on the provided features and corresponding labels. It splits the feature space based on various attributes and their thresholds, optimizing for the best classification performance After training, the model is evaluated on the testing dataset to assess its performance using metrics such as accuracy, precision, recall, and F1-score is given in Table 1. The model's hyperparameters can be optimized to enhance its performance through techniques like cross-validation or grid search [23].

Table 1 Classification Report for Decision Tree

	Precision	**Recall**	**F1-Score**	**Support**
Hate Speech Detected	0.34	0.30	0.32	465
No Hate Speech Detected	0.78	0.81	0.79	1379
Offensive Speech Detected	0.92	0.92	0.92	6335
Accuracy			0.87	8179
Macro avg	0.68	0.68	0.68	8179
Weighted avg	0.86	0.87	0.86	8179

Even though Decision Tree gave good results we worked with KNN for better results and to improve accuracy. KNN which is a machine learning algorithm applied for classification tasks. KNN works by finding the K nearest neighbors to a given data point based on a distance metric and assigning a class label based on the majority vote of those neighbors [24-25]. We imported the necessary libraries for KNN (e.g., scikit-learn in Python). Instantiate a KNN model with desired hyperparameters, such as the number of neighbors (K) and the distance metric [26]. Train the model on the training dataset using the fit() function. Evaluate the trained model on the testing dataset using appropriate metrics. Refer Table 2.

Table 2 Classification Report for KNN

	Precision	**Recall**	**F1-Score**	**Support**
Hate Speech Detected	0.38	0.34	0.36	290
No Hate Speech Detected	0.71	0.67	0.69	835
Offensive Speech Detected	0.90	0.92	0.91	3832
Accuracy	-	-	0.84	4957
Macro avg	0.66	0.64	0.65	4957
Weighted avg	0.84	0.84	0.84	4957

The results from the KNN model were not superior to Decision Tree so we decided to use another model named Logistic Regression to further increase the accuracy of the model [27]. Logistic Regression [LR] is a popular algorithm used for binary classification tasks, such as distinguishing between hate speech, offensive speech, and no hate speech. It models the probabilitty of a given input assosciated to a specific class. Perform text pre-processing tasks such as tokenization, lowercasing, and removal of stop words or special characters. Import the necessary libraries for Logistic Regression (e.g., sklearn.linear_model, sklearn.metrics, etc in Python).

A Logistic Regression model with desired hyperparameters, such as regularization strength [28]. Train the model on the training dataset using the fit() function. Evaluate the trained model on the testing dataset using appropriate metrics. Refer: Table 3.

Table 3 Classification Report for Logistic Regression

	Precision	**Recall**	**F1-Score**	**Support**
Hate Speech Detected	0.48	0.24	0.32	290
No Hate Speech Detected	0.84	0.85	0.84	835
Offensive Speech Detected	0.92	0.95	0.94	3832
Accuracy			0.90	4957
Macro avg	0.75	0.68	0.70	4957
Weighted avg	0.88	0.90	0.89	4957

Based on the evaluation of different classifiers, the logistic regression model achieved an accuracy of 90% with fair performance in hate speech detection. Decision tree classifier attained an accuracy of 87% with moderate performance in hate speech detection.

Figure 2 Learning Curve for Logistic Regression

The K-nearest neighbours (KNN) classifier achieved an accuracy of 84% with relatively lower performance in hate speech detection. To improve hate speech detection, a combined solution could involve using an ensemble method, such as a voting classifier, that combines the predictions of multiple models [29]. This approach leverages the strengths of each model to enhance precision, recall, and F1-score for hate speech identification.

As shown in Figure 2, based on the evaluation of different classifiers, the logistic regression model achieved an accuracy of 90% with fair performance in speech and hate speech detection [30]. Decision tree classifier attained an accuracy of 87% with moderate performance. The K-nearest neighbours (KNN) classifier achieved an accuracy of 84% with relatively lower performance [31-32].

4 RESULTS AND DISCUSSION

The provided table presents a comprehensive evaluation of three different machine learning models in the context of hate speech detection, no hate speech detection, and offensive speech detection. For each category, the table displays Precision (the accuracy of positive predictions), Recall (the ability to capture actual positive instances), F1-score (a balance between precision and recall), and Support (the number of instances in the evaluation dataset). In the Table 4, "Hate Speech Detected" category, the Logistic Regression model exhibits the highest precision at 0.48 but has a relatively low recall of 0.24, resulting in an F1-score of 0.32. In the Table 5,"No Hate Speech Detected" category, the Logistic Regression model again performs well with a precision of 0.84 and high recall of 0.85, resulting in an F1-score of 0.84. In the Table 6, "Offensive Speech Detected" category, all three models achieve strong performance, with F1-scores ranging from 0.91 to 0.94. Additionally, the Table 7 provides overall accuracy percentages for each model, indicating their accuracy across all categories, with KNN achieving 84%, Decision Tree achieving 87%, and Logistic Regression achieving 90%. These metrics collectively offer a comprehensive overview of the models' performance in various aspects of speech detection.

Table 4 Hate Speech Detected

	Precision	Recall	f1-score	Support
Decision Tree	0.34	0.30	0.32	465
KNN	0.38	0.34	0.36	290
Logistic Regression	0.48	0.24	0.32	290

Table 5 NO Hate Speech Detected

	Precision	Recall	f1-score	Support
Decision Tree	0.78	0.81	0.79	1379
KNN	0.71	0.67	0.69	835
Logistic Regression	0.84	0.85	0.84	835

Table 6 Offensive Speech Detected

	Precision	Recall	f1-score	Support
Decision Tree	0.92	0.92	0.92	6335
KNN	0.90	0.92	0.91	3832
Logistic Regression	0.92	0.95	0.94	3832

Table 7 Overall Accuracy

Method	Accuracy
KNN	84%
Decision Tree	87%
Logistic Regression	90%

5 CONCLUSION

The proposed study focused on employing machine learning techniques to enhance speech recognition, specifically within the context of hate speech detection. The Decision Tree model was employed initially, utilizing its tree-like structure to make informed classification decisions. With proper dataset labelling and partitioning, this model demonstrated promising results. However, to further refine accuracy, we transitioned to the K-Nearest Neighbours

(KNN) model. While KNN proved effective, it fell short of surpassing the Decision Tree's performance.

In pursuit of heightened accuracy, we explored the Logistic Regression model. This algorithm, designed for binary classification tasks, effectively modelled probabilities and demonstrated potential for improved hate speech detection. Evaluating each model's performance via accuracy, precision, recall, and F1-score metrics, the Logistic Regression model emerged as the most accurate, achieving a notable 90% accuracy in hate speech detection. The Decision Tree model followed with an accuracy of 87%, and the KNN model achieved 84% accuracy. To further elevate hate speech detection capabilities, we proposed a combined approach, considering ensemble methods like the voting classifier. Such an approach, leveraging the strengths of various models, holds promise in achieving enhanced precision, recall, and F1-score metrics. These endeavours in model selection and evaluation were visualized through classification reports demonstrating the models' performance on testing data. The proposed study underscores the iterative process of refining speech recognition systems for hate speech detection. The exploration of distinct models – Decision Tree, KNN, and Logistic Regression – highlighted the complexities of accuracy enhancement. While Logistic Regression led the accuracy race, the pursuit of comprehensive hate speech detection warrants continued investigation into ensemble methodologies and refined model architectures. Through these efforts, we endeavor to create a safer online environment, effectively curbing the proliferation of hate speech and promoting constructive digital interactions.

References

[1]. Wang, C., Wu, Y., Chen, S., Liu, S., Li, J., Qian, Y., Yang, Z, “Self-supervised learning for speech recognition with intermediate layer supervision”, arXiv preprint arXiv:2112.08778, 2021.

[2]. Karthikeyan, T., Sekaran, K., Ranjith, D., Balajee, J. M, “Personalized content extraction and text classification using effective web scraping techniques”, International Journal of Web Portals (IJWP), Vol. 11, No. 2, pp. 41-52, 2019.

[3]. Praveen Sundar, P. V., Ranjith, D., Karthikeyan, T., Vinoth Kumar, V., Jeyakumar, B, “Low power area efficient adaptive FIR filter for hearing aids using distributed arithmetic architecture”, International Journal of Speech Technology, Vol. 23, No. 2, pp. 287-296, 2020.

[4]. Umamaheswaran, S., Lakshmanan, R., Vinothkumar, V., Arvind, K. S., Nagarajan, S, “New and robust composite micro structure descriptor (CMSD) for CBIR”, International Journal of Speech Technology, Vol. 23, pp. 243-249, 2020.

[5]. Han, W., Zhang, Z., Zhang, Y., Yu, J., Chiu, C. C., Qin, J., Wu, Y, “Contextnet: Improving convolutional neural networks for automatic speech recognition with global context”, arXiv preprint arXiv:2005.03191, 2020.

[6]. Shalini, A., Jayasuruthi, L., VinothKumar, V, “Voice recognition robot control using android device. Journal of Computational and Theoretical Nanoscience”, Vol. 15, pp. 2197-2201, 2018.

[7]. Kumar, V. V., Raghunath, K. K., Muthukumaran, V., Joseph, R. B., Beschi, I. S., Uday, A. K, “Aspect based sentiment analysis and smart classification in uncertain feedback pool”, International Journal of System Assurance Engineering and Management, pp. 1-11, 2021.

[8]. Basheer, S., Anbarasi, M., Sakshi, D. G., Vinoth Kumar, V, “Efficient text summarization method for blind people using text mining techniques”, International Journal of Speech Technology, Vol. 23, pp. 713-725, 2020.

[9]. Muthukumaran, V., Joseph, R. B., Uday, A. K, “Intelligent medical data analytics using classifiers and clusters in machine learning”, In Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies, pp. 321-335, 2021.

[10]. Ahmed, S. T., Kumar, V., Kim, J, “AITel: eHealth Augmented Intelligence based Telemedicine Resource Recommendation Framework for IoT devices in Smart cities”, IEEE Internet of Things Journal, 2023.

[11]. Maithili, K., Vinothkumar, V., Latha, P, “Analyzing the security mechanisms to prevent unauthorized access in cloud and network security”, Journal of Computational and Theoretical Nanoscience, Vol. 15, pp. 2059-2063, 2018.

[12]. Mahesh, T. R., Sivakami, R., Manimozhi, I., Krishnamoorthy, N., Swapna, B, “Early Predictive Model for Detection of Plant Leaf Diseases Using MobileNetV2 Architecture”, International Journal of Intelligent Systems and Applications in Engineering, Vol. 11, No. 2, pp. 46-54, 2023.

[13]. Perifanos, K., Goutsos, D, “Multimodal hate speech detection in greek social media”, Multimodal Technologies and Interaction, Vol. 5, No. 7, pp. 34, 2021.

[14]. Chinthamu, N., Gooda, S. K., Shenbagavalli, P., Krishnamoorthy, N., Selvan, S. T, “Detecting the Anti-Social Activity on Twitter using EGBDT with BCM”, International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 11, No. 4s, pp. 109-115, 2023.

[15]. Kumar, D., Swathi, P., Jahangir, A., Sah, N. K., Vinothkumar, V, “Intelligent speech processing technique for suspicious voice call identification using adaptive machine learning approach”, In Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies, pp. 372-380, 2021.

[16]. Garg, T., Masud, S., Suresh, T., Chakraborty, T, "Handling bias in toxic speech detection: A survey", ACM Computing Surveys, Vol. 55, No. 13s, pp. 1-32, 2023.

[17]. Banerjee, Saikat, Abhoy Chand Mondal, "A Region-Wise Weather Data-Based Crop Recommendation System Using Different Machine Learning Algorithms", International Journal of Intelligent Systems and Applications in Engineering, Vol. 11, No. 3, pp. 283-297, 2023.

[18]. Kumar, R. S., Saravanan, N. P., Devi, K. N., Jayanthi, P., Krishnamoorthy, N., Karthi, S, "Empirical Analysis on Crime Prediction using Machine Learning", In 2023 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-5, 2023.

[19]. Jayanthi, P., Krishnamoorthy, N., Sridharan, S., Tamilkumar, R., Yokesh, P, "An Enhanced Technique To Classify OCT Images Using Deep Learning", In 2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), pp. 1-5, 2023.

[20]. Krishnamoorthy, N., Prasad, L. N., Kumar, C. P., Subedi, B., Abraha, H. B., Sathishkumar, V. E, "Rice leaf diseases prediction using deep neural networks with transfer learning", Environmental Research, Vol. 198, 2021.

[21]. Krishnamoorthy, N., Asokan, R., Jones, I, "Classification of malignant and benign micro calcifications from mammogram using optimized cascading classifier", Current Signal Transduction Therapy, Vol. 11, No. 2, pp. 98-104, 2016.

[22]. Mahesh, T. R., Dhilip Kumar, V., Vinoth Kumar, V., Asghar, J., Geman, O., Arulkumaran, G., Arun, N, "AdaBoost ensemble methods using K-fold cross validation for survivability with the early detection of heart disease", Computational Intelligence and Neuroscience, 2022.

[23]. Kumar, V., Niveditha, V. R., Muthukumaran, V., Kumar, S. S., Kumta, S. D., Murugesan, R, "A quantum technology-based lifi security using quantum key distribution", In Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies, pp. 104-116, 2021.

[24]. Vinoth Kumar, V., Arvind, K. S., Umamaheswaran, S., Suganya, K. S, "Hierarchal Trust Certificate Distribution using Distributed CA in MANET. International Journal of Innovative Technology and Exploring Engineering, 8(10), 2521-2524.

[25]. Muthukumaran, V., Kumar, V. V., Joseph, R. B., Munirathanam, M., & Jeyakumar, B. (2021). Improving network security based on trust-aware routing protocols using long short-term memory-queuing segment-routing algorithms. International Journal of Information Technology Project Management (IJITPM), 12(4), 47-60.

[26]. Vinoth Kumar, V., & Ramamoorthy, S. (2018). Secure adhoc on-demand multipath distance vector routing in MANET. In Proceedings of the International Conference on Computing and Communication Systems: I3CS 2016, NEHU, Shillong, India (pp. 49-63). Springer Singapore.

[27]. Kumar, V. V., Muthukumaran, V., Ashwini, N., Beschi, I. S., Gunasekaran, K., & Niveditha, V. R. (2022). An efficient signcryption scheme using near-ring hybrid approach for an IoT-based system. International Journal of e-Collaboration (IJeC), 18(1), 1-31.

[28]. Zhao, J., & Kumar, V. V. (Eds.). (2021). Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies. IGI Global.

[29]. TR, M., Vinoth Kumar, V., & Lim, S. J. (2023). UsCoTc: Improved Collaborative Filtering (CFL) recommendation methodology using user confidence, time context with impact factors for performance enhancement. Plos one, 18(3), e0282904.

[30]. Zhao, J., & Kumar, V. V. (Eds.). (2022). Handbook of Research on Technologies and Systems for E-Collaboration During Global Crises. IGI Global.

[31]. Rajalakshmi, V., Muthukumaran, V., Koti, M. S., Vinothkumar, V., & Thillaiarasu, N. (2022). E-Collaboration for Management Information Systems Using Deep Learning Technique. In Handbook of Research on Technologies and Systems for E-Collaboration During Global Crises (pp. 398-411). IGI Global.

[32]. Kumar, V. V., Sikdar, B., Katina, P. F., & Ansari, I. S. (2022). Editorial Preface: Special issue on sustainable computing for cyber-physical systems. Sustainable Computing: Informatics and Systems, 35, 100733.