ISSN: 2278-4632
Vol-13, Issue-09, September 2023

DETECTING RACIST TWEETS USING MACHINE LEARNING AND DEEP LEARNING

Dr. M Ramchander[1], Tadi Naga Praveen Reddy[2]
[1]Assistant Professor, Department of MCA, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India
[2]MCA Student, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India

ABSTRACT

Social media has witnessed the birth of numerous new and old types of racism due to its significance in the geopolitical environment. On social media, racism has taken many different forms, both overt and covert. Racist ideas have been made overt by being communicated under false names, and have been concealed by the use of memes in order to incite hatred, violence, and societal upheaval. Despite typically being associated with ethnicity, racism is increasingly pervasive on the basis of race, national origin, language, culture, and—most significantly—religion. Social, political, and cultural stability are all seriously at risk when racial tensions are incited on social media. As a result, racist utterances should be identified and outlawed as soon as feasible. Social media is the main channel via which racist ideas are spread. This project aims to find racist tweets using sentiment analysis. Long Short-Term Memory (LSTM) and Graph Convolutional Neural Network (GCN) are combined to create the LSTM + GCN with BERT model because of the improved performance of deep learning. Initially, started comparing different Machine Learning and Deep Learning Models. After final examination of accuracy, We found LSTM has improved 99% accuracy and better performance.

KEYWORDS: Racist, Racism, online abuse, Twitter, deep learning, machine learning, sentiment analysis.

I.INTRODCUTION

Our opinions and behaviours are frequently dictated by social media, which has assumed a dominant place in sociopolitical potential. Due to the widespread use of social media platforms and the freedom of expression, a number of vices, including racism, have increased recently. For example, prejudice and the stress it produces seem to thrive in Twitter's brand-new environment. With 1.3 billion accounts, 336 million active users worldwide, 90% of whom have public profiles, and 500 million tweets sent out each day, Twitter has a 1.3 billion user base. Currently, 22% of US citizens utilise the social media network. Tweets can be replied to and participated in by posting them on their profiles (retweeting), tagging other users, hitting the "like" button, or leaving a comment for the tweet's author. Tweets are publicly available until they are made private. The raw data for sentimental analysis is based on the expression of sentiments, emotions, attitudes, and perspectives on Twitter. Social media platforms have grown in popularity, which has encouraged widespread use of them for different sorts of racism throughout history and in the present. Through memes and the posting of racist Tweets under fictional accounts, racism is represented on social platforms in both overt and covert ways. Racism is increasingly common on the basis of race, national origin, language, culture, and—most significantly—religion, even though it is frequently connected with ethnicity. Inciting racial tensions on social media has been viewed as a serious threat to global peace as well as social, political, and cultural stability. Since social media is the main source of racist ideas, it should be closely watched, and any racist statements should be discovered and immediately removed.

Racist comments and tweets on social media have been associated with a number of physical and mental disorders, which have had a severe impact on people's health [1–5]. Three categories of racism on social media can be identified: institutionalised, personally mediated, and internalised [6]. Racism can be personally experienced through racial discrimination or uneven treatment, as well as through awareness of prejudice towards family members and acquaintances. Racism in society therefore has a negative impact on people and causes a variety of psycho-social stresses, which typically increase the risk of chronic diseases [7]-[9]. Racist groups and individuals also use sophisticated tactics and higher-level skills to disseminate racism online [10]. Special attention has been given to the field of sentiment analysis in order to analyse text from social media platforms for a range of purposes including hate speech identification, sentiment-based market prediction, and racism detection, among others.

II.LITERATURE SURVEY

Hate crimes are on the rise as a result of social media's broad use and users' ability to remain anonymous online. There are many overlapping and converging forms and purposes behind the troublesome situation of abusive content and sophisticated stuffing on social media [11]. Online users have negative emotions when they read about harassment and abuse, which leads them to communicate those emotions in an impolite manner. Due to their negative impacts on society, hate speech and cyberbullying are two examples of abusive language that have piqued scholars' interest recently. It is imperative that these components be decontaminated. Numerous research have been done for this goal to automatically identify the grating hate speech and messages on social media, among other topics. It still needs more study from both industry and academia to automatically detect hate speech using machine

Model	Dataset	Accuracy
Varients of BERT and Resnet	https://github.com/kperi/ MultimodalHate-SpeechDetection	0.97
resnet18 + nlpaueb/greek-bert, Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF) with TF-IDF, profile-related and emotion-related features.	3696 tweets, Self-made	0.913
Random Forest (RF) with TF-IDF and profile related features, Naïve Bayes, Logistic Regression, XGBoost and TF-IDF features	de Gibert, O. a. (2018). Hate Speech Dataset from a White Supremacy Forum. Association for Computational Linguistics, Github	XGBoost with TF-IDF, Recall:0.83, Precision:0.82
BERT,CNN,GRU and the ensemble of CNN and GRU (CNN+GRU)	selfmade	F1 score: 0.79 CNN
Distributed Bag of words (DBoW), Distributed Memory Mean (DMM), and Word2Vec CNN	1st dataset: university of Maryland, 2nd dataset: self-made 25000 tweets	1st dataset=96.67%, 2nd dataset=97.5%, Neural Network with 3 hidden layers with Doc2Vec
Naïve Bayes,Multilayer Preceptron,AdaBoost classifier,Support Vector Machine	Self-made tweeter dataset 4002 tweets	83.4%, MLP with SMOTE 71.2%, AB, MNB, BNB
Multinomial Naïve Bayes,Linear SVM, Random Forest and RNN	Self-made, Youtube	0.9464 for the first experiment and 0.857 for the second experiment
NB, RF,LR,DT, SVM and deep learning models	Self made : tweeter	SVM 74.6%
XGBoost,SVM,LR,NB,and FFNN	YouTube dataset (ICWSM 18 SALMINEN), Reddit dataset (ALMEREKHI 19), Wikipedia dataset (KAGGLE 18), Twitter dataset (DAVIDSON 17 ICWSM)	F1 score =0.92, XGBoost

learning algorithms [12]. Here, a couple of recent works have been discussed that are similar [13, 14]. The detection and analysis of hate speech has greatly benefited from machine learning techniques [15].

The authors of [16] offer a multimodal hate speech detection algorithm designed specifically for Greek social media. The study focuses on Greek-language tweets that criticise immigrants and refugees, particularly those that employ racist and xenophobic language. On the gathered dataset, the ensemble model, transfer learning, and fine-tuning of the BERT and Resnet bidirectional encoder representations are used. The highest accuracy was

TABLE 1. Summary of the discussed research works

reported with nlpaueb/greek-bert for text modality and 0.97 with resnet18+ nlpaueb/greek-bert for text+image modality. Different variations of the BERT and Resnet are employed. [17] proposes a comparable state-of-the-art machine learning-based approach for the automated identification of hate speech in Arabic social media networks. As various emotional states are collected, numerous feature sets are used for analysis. The study uses four different machine learning approaches, including Naive Bayes (NB), DT, SVM, and RF using TF-IDF, profile-related, and emotion-related data. By combining TF-IDF and profile-related information, RF was able to get the highest accuracy, which was 0.913. In a similar vein, [18] uses attributes collected from content including true and fake news to categorise fake news and hate speech propaganda. The study makes use of TF-IDF characteristics with NB, LR, and XGBoost. With a recall score of 0.83, XGBoost shows that the model incorrectly identified 17% of the data as containing hatred. Furthermore, XGBoost attains a precision value of 0.82, meaning that the model mistakenly classified 18% of the data as hateful. The topic of hate speech in the Saudi Twitter community is investigated by authors using a range of deep learning approaches [19]. Several experiments using BERT, CNN, GRU, and the ensemble of CNN and GRU (CNN+GRU) are conducted on two datasets. The CNN model, according to the results, achieves an F1 score of 0.79 and an area under the receiver operating curve (AUROC) of 0.89.

The automatic identification of cyberbullying is examined in study [20]. The authors employ two distinct datasets to compare deep learning and machine learning techniques. Online racism is categorised using a variety of word embedding approaches, including distributed BoW (DBoW), distributed memory mean (DMM), and Word2Vec CNN. A neural network with three hidden layers using Doc2Vec features achieves an accuracy of 96.67% for one dataset and 97.5% for the second dataset. Similar to this, study [21] investigates the automatic recognition of racist or hateful tweets in Indonesia. The authors employ machine learning methods including SVM, Multilayer Perceptron (MLP), AdaBoost (AB) classifier, and Multinomial NB (MNB). As an upsampling technique, synthetic minority oversampling technique (SMOTE) is utilised, and experiments are run on both SMOTE and non-SMOTE features. According to the results, MNB has 71.2% accuracy for non-SMOTE features and 83.4% accuracy for MLP with SMOTE features. Work on detecting hate speech on social media is done by Ching She et al. in [22]. In order to conduct research, audio data from videos is taken out and translated to text using a speech-to-text converter. In the experiments, MNB, Linear SVM, RF, and RNN are employed. The classification of the video into normal and hateful movies is the subject of the first of two sets of studies, while the classification of the video into normal, racist, and sexiest classes is the subject of the second. Results indicate that RF performs better than other methods in terms of accuracy, achieving accuracy values

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

of 0.9464 for the first set of experiments and 0.857 for the second set.

[23] is another study of a similar nature that examines anti-Islamic hate speech on social media. The study develops an automated technique that can discern between content that is not anti-Islamic, content that is mildly anti-Islamic, and content that is very anti-Islamic. Different machine learning methods are applied, including deep learning models, NB, RF, LR, DT, and SVM. Results indicate that SVM achieves a testing accuracy of 72.17 percent. The effectiveness of SVM is further assessed using 10-fold cross-validation, which demonstrates a balanced accuracy of 80.7% and an accuracy of 74.6%. A novel technique is suggested in study [24] to identify hate speech on several social media platforms, including Reddit, YouTube, Twitter, and Wikipedia. These social media platforms are used to create a sizable dataset with 80% of the content classified as not being hateful and 20% as being hateful. BoW, TF-IDF, Word2Vec, BERT, and their combinations were used to test a number of machine learning algorithms, including XGBoost, SVM, LR, NB, and feed-forward neural networks. With a 0.92 F1 score and all features, XGBoost performs better than any other models. BERT traits have a significant impact on predictions, according to a feature importance analysis. This study uses the deep learning ensemble model to identify racist tweets on Twitter while taking into account the findings from deep learning models that have been previously published. High classification accuracy is what the study seeks to achieve using stacked recurrent neural networks. Sentiment analysis is used to identify racist tweets, with the ratio of tweets with negative sentiment serving as a marker.

III. MATERIALS AND METHODS

A. PROPOSED METHODOLOGY

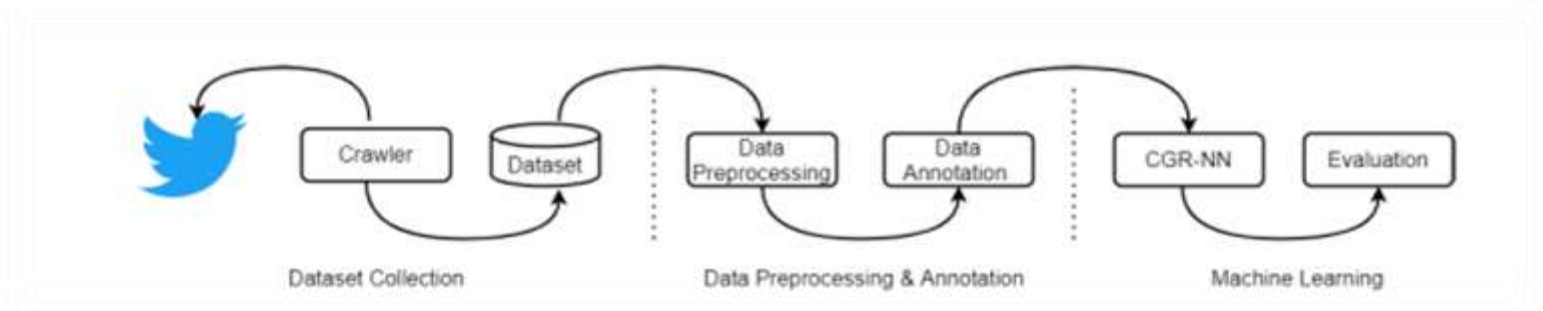

FIGURE 1. Architecture of the proposed methodology.

In tis paper, a method for detecting racism on social media platforms is proposed, using deep learning and machine learning techniques. The proposed approach's step-by-step flow is shown in Figure 1. Twitter is crawled in the first stage, then the data is cleaned up and processed, and then the data is annotated. After training and testing on the datasets, the proposed stacked ensemble model is compared to several different deep learning and machine learning models for performance.

B. DATASET DESCRIPTION

Twitter is where the dataset for racist tweets is gathered. Because Twitter is the most popular platform used by many people to communicate their sentiments, views, comments, and ideas, it has been the primary pick of the majority of researchers for text and sentiment analysis. This study specifically aims to investigate the racist trends found in Twitter tweets. Racist-related tweets have been gathered for data collecting. Several keywords are used for this, including "#racism," "#racial," and "#racist," among others. 31,962 tweets in total have been gathered that meet the requirements. In which, 2242 tweets are Racist , while 29,720 tweets are Non-Racist.

C. DATA PREPROCESSING

The data is cleaned in a number of processes at the preprocessing level. In order to properly train a model, the document must be properly prepped and cleaned. The reviews in this study were preprocessed using a combination of natural language processing (NLP) techniques utilising Python's NLTK.

- **Tokenization:** The process of dividing natural texts into tokens devoid of any white spaces is known as tokenization. It entails disassembling phrases into their component words. Despite appearing easy and uncomplicated, selecting the right tokens is a difficult task.
- **Stemming:** Different spellings of the same term are used throughout the text, which can complicate machine learning models. The altered variants of the word "go" include the words "gone," "going," and "go." Each word is stemmed into its root form, thus "gone" becomes "go," and "going" becomes "going." The Stemmer Porter algorithm is used to do stemming.
- **Lemmatization:** Although it follows a similar process to tokenization, the result is different. Tokenization only eliminates the final's' or 'es' from a word to transform it into its root form, which frequently yields incorrect terms or spelling. By taking into account the context in which a word is used, lemmatization preserves the term's root form. Additionally, it reduces the number of times similar words occur alone. The suggested strategy for word

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

preprocessing uses this method to reduce the number of unique occurrences of identical text tokens.

- **Stop Words Exclusion:** Stop words are terms that don't help the machine learning algorithms when they're being trained. Instead, they expand the feature space to add complexity. Stop words like a, am, and an, among others, are thus eliminated to improve the models' learning effectiveness in this study.
- **Case Normalization:** The text must be transformed to lowercase letters because specific words with different case requirements, such as "Racism" & "racism," must be handled similarly in all circumstances. Because it reduces the recurrence of features that differ only in case sensitivity, it is frequently referred to as data cleansing.
- **Noise Removal:** This stage eliminates any noise that can impair the classification's performance. In this stage, noise types such special characters, numeric data, id, and "#" signs, among others, are erased.

The preprocessed text from the sample tweets is provided in Table 1 after the procedures above.

TABLE 1. Sample text before and after the preprocessing

Before preprocessing	After preprocessing
@_LeBale racism is good	racism good
@manoutdoors4 @AJ_Lady_Liberty @FBIWFO @TheJusticeDept @FBI it is clear to hundreds of millions of people of all walks that this country has a severe problem with systemic racism. your denial is discussing. the world is changing , get on board or get left	clear hundr million people walk country sever problem system racism denial

D. DATA ANNOTATION

To annotate the dataset with positive and negative sentiments, this study uses the TextBlob library.

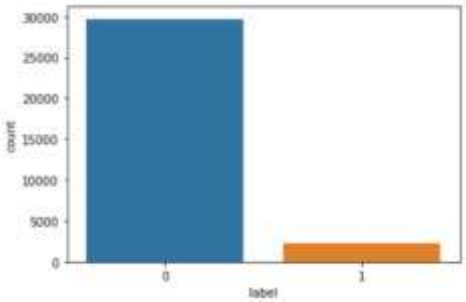

FIGURE 2. Ratio of sentiment in dataset.

In order to apply a sentiment label to a text, Textblob determines the polarity score for that text. The polarity score range for textblobs ranges from -1 to 1. According to Figure 2, the data is divided into positive and negative categories.

IV.MACHINE LEARNING MODELS

Machine learning algorithms have been employed for the purpose of detecting racism in tweets due to their superior performance than conventional methods. Some well-known models, including RF, LR, DT, SVM, and KNN, are briefly examined in this work to ensure completeness. Carefully changing a variety of hyperparameters improves the performance of these models.

1) RANDOM FOREST

A tree-based classifier called RF constructs its trees using a random vector that is drawn from the input vector. By first creating several decision trees using random features, RF constructs a forest. The conclusion from each decision tree is then combined to generate the final forecast, which is then voted on. Votes from decision trees with lower mistake rates are given more weight, and the opposite is also true. Reduces the likelihood of making an incorrect forecast by employing decision trees with low error rates [25]. The equations below can be used to define RF:

$$p = mode\{T1(y), T2(y), \ldots, Tm(y)\} \quad (1)$$

$$p = mode\{ Xm\ m=1\ Tm(y)\} \quad (2)$$

2) LOGISTIC REGRESSION

The statistical-based classifier LR is mostly used to analyse binary data when one or more factors are used to determine the outcomes. It is also used to assess the likelihood of a class relationship [33]. LR is particularly recommended for categorical data because to its better performance. It approximates the link between the dependent variable and one or more independent variables of the categorical data. LR approximates probability by means of a logistic function.[32]. A popular "S" sloping or sigmoid curve known as a logistic function or logistic curve is defined as

$$f(x) = \frac{L}{1+e^{-m(v-v_0)}} \quad (3)$$

3) SUPPORT VECTOR MACHINE

A well-known machine learning algorithm called SVM is frequently used to classify both linear and nonlinear data. It is the primary option for many academics when it comes to binary classification issues, and it is available in a variety of kernel functions [27]. In order to classify data points, the SVM classifier's main task is to estimate the hyperplane using a feature set [28]. The size of the hyperplane depends on the number of features. Because there are several possible hyperplane configurations in n-dimensional space, the problem is to build hyperplanes that maximise the margins between samples of classes. The following is the cost function used to determine the hyperplanes:

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

$$J(\theta) = \frac{1}{2}\sum_{j=1}^{n} \theta_j^2 \quad (4)$$

Such that,

$$\theta^T x^{(i)} \geq 1,\ y^{(i)} = 1, \quad (5)$$

$$\theta^T x^{(i)} \leq -1,\ y^{(i)} = 0, \quad (6)$$

4) K NEAREST NEIGHBOR

A simple and well-liked machine learning technique called KNN may be utilised to address classification and regression problems. KNN uses the concept of "neighbours" because it anticipates finding neighbouring data that is similar to its own. It determines the separation between the new data points and their neighbours using metrics for measuring distance, like Euclidean distance, Manhattan distance, Minkowski distance, etc. The KNN's K value determines how many neighbours are employed for prediction. Here [32] is a list of well-known metrics for measuring distance:

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}, \quad (7)$$

$$Manhattan\ Distance = \sum_{i=1}^{k} |x_i - y_i|, \quad (8)$$

$$Minkowski\ Distance = \left(\sum_{i=1}^{k} |x_i - y_i|^q\right)^{1/q}, \quad (9)$$

5) DECISION TREE

DT is a rule-based supervised machine learning method. The widely used and successful DT prediction model is capable of handling classification and regression problems. The most popular methods for attribute selection, which is the core problem in DT, are information gain and the Gini index [30]. Information gain is the rate of growth or reduction in the entropy of characteristics, where entropy indicates how homogenous a dataset is [31].

$$E(D) = -P(positive)log2.P(positive) - P(negative)log2.P(negative)$$

The entropy E of a dataset D that contains both positive and negative decision qualities is calculated using the equation above. The formula: is used to compute the gain of the attribute X.

$$Gain\ (attribute\ X) = Entropy(Decision\ Attribute\ Y) - Entropy(X, Y)$$

6) GCN with BERT

A potent method for processing graph-structured data with textual information uses GCN and BERT. We can capture both the structural dependencies and the semantic meaning of the text by presenting the data as a graph and using BERT to encode text attributes. Each node's text properties are encoded using BERT, the graph is convolutioned to spread information, and the node embeddings are improved iteratively. By utilising the improved node embeddings, this integrated model enables us to handle numerous downstream tasks, such as node categorization or link prediction. The integration of GCNs and BERT provides a comprehensive framework for effectively handling graph data with textual features, opening up possibilities for advanced analysis and understanding of complex data structures. The combination of GCNs with BERT offers a thorough framework for managing textual elements in graph data, opening up opportunities for sophisticated analysis and comprehension of intricate data structures.

7) LSTM

The vanishing gradient problem is addressed by the LSTM, a form of RNN that can identify long-term dependencies in sequential data. To selectively store, forget, and output information, it makes use of memory cells and gating mechanisms. The information to be stored is decided by the input gate, the information to be deleted from the memory cell is decided by the forget gate, and the network information is controlled by the output gate. Language modelling, sentiment analysis, and machine translation are examples of sequential data analysis and generating jobs where LSTMs excel. Gradient descent and backpropagation across time are used to train them. For accurate predictions or generation, LSTMs are frequently utilised in a variety of fields where capturing long-term dependencies is essential.

8) LSTM + GCN with BERT

A complete model for processing graph-structured data containing textual and sequential information is provided by the combination of LSTM, GCN, and BERT. While GCN manages the structural interactions in the graph, LSTM allows the model to capture the sequential dependencies inside the textual data. Each node's textual properties are encoded using BERT to capture semantic meaning. The GCN spreads information throughout the graph, the LSTM processes BERT embeddings sequentially, and BERT offers rich contextualised representations. With the use of both sequential and structural information, this combined model enables a comprehensive understanding of graph data and makes it useful for a variety of tasks, including node categorization and link prediction.

9) GCN with BERT + LSTM

Graph-structured data including textual and sequential information can be processed and analysed using the GCN, BERT, and LSTM model. The model may encode textual properties of

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

each node and capture semantic meaning by using BERT. The GCN takes advantage of the graph structure to spread information and record node dependencies. In order to capture contextual information and sequential dependencies inside the text, the LSTM component sequentially processes the BERT embeddings. By including both textual and sequential information, this integrated model enables a thorough understanding of graph data, making it suitable for a variety of tasks such as node categorization, link prediction, and graph-level predictions.

V.RESULTS AND DISCUSSIONS

Experiments on sentiment analysis on racist tweets have been carried out on a Windows 10 machine with an Intel Core i7 of the 11th generation. On Jupyter, machine learning and deep learning models are built using the Tensor-flow, Kara's, and Sci-kit Learn frameworks. Performance of each model is evaluated using its accuracy, precision, recall, F1 score, number of correct predictions, and number of wrong predictions.

A. RESULTS USING DEEP LEARNING MODELS

For performance assessment and a fair comparison with the suggested ensemble deep learning model, a number of single deep learning models—including GRU, LSTM, CNN, and RNN—are also developed. Deep learning models' performance is maximised by modifying alternate topologies for various parameters, including the number of layers, loss function, optimizer, and neurons. The results show that deep learning models significantly outperform machine learning models. Due to the high data requirements of deep learning, the training and performance of these models are enhanced by assembling enormous datasets for racism detection. The accuracy of RNN is 0.95, whereas that of LSTM, GRU, and CNN is all 0.99. [FIGURE 3–8]

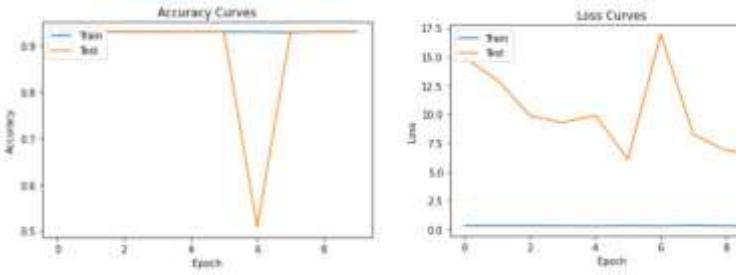

FIGURE 3. Accuracy & Loss curves of CNN Model

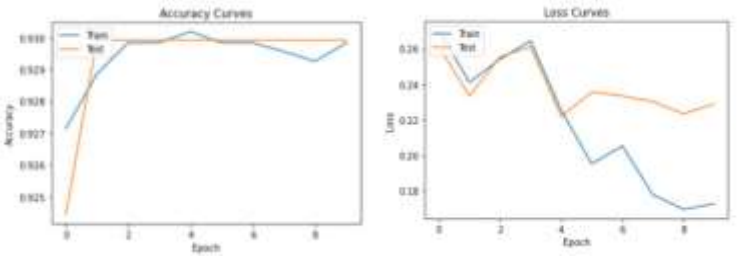

FIGURE 4. Accuracy & Loss curves of RNN Model

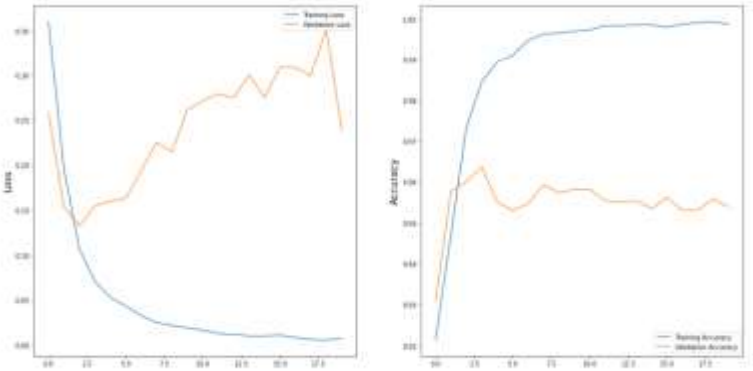

FIGURE 5. Accuracy & Loss curves of LSTM Model

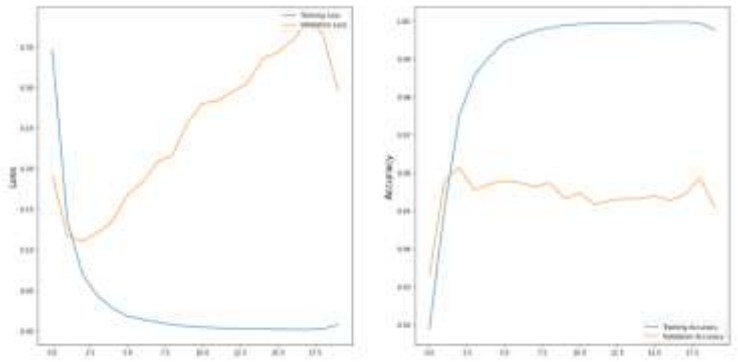

FIGURE 6. Accuracy & Loss curves of GRU Model

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

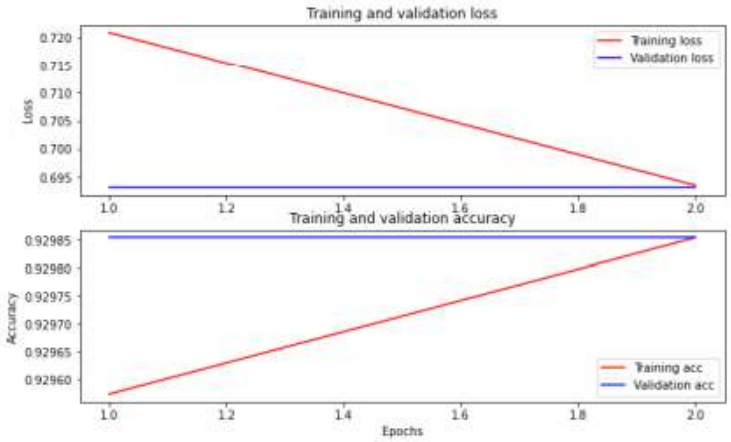

FIGURE 7. Accuracy & Loss curves of GCN-NN with BERT Model

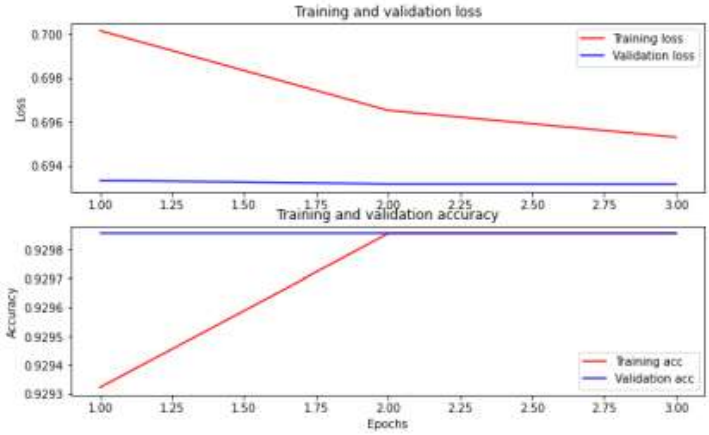

FIGURE 8. Accuracy & Loss curves of LSTM + GCN with BERT Model

B. COMPARISION WITH MACHINE LEARNING MODELS

Each model was assessed using a binary classification task that separated classes into Positive and Negative. The models' overall accuracy, which ranged from 0.92 to 0.95, was very high. The accuracy, precision, recall, and F1 score for the Positive class for both Logistic Regression (LR) and Random Forest (RF) were 0.95, 0.96, 0.99, and 0.98, respectively. Precision, recall, and F1 scores for the Negative class for LR and RF, however, were lower at 0.84, 0.51, and 0.63, respectively. K-Nearest Neighbours (KNN) maintained a high F1 score of 0.97 for the Positive class despite having a slightly lower accuracy of 0.93. KNN struggled with the Negative class, though, and its precision, recall, and F1 scores were lower. With an accuracy of 0.94, Decision Tree (DT) outperformed Support Vector Machine (SVM), which had worse precision and recall for both classes. The Voting Classifier outperformed LR and RF with an accuracy of 0.95 and increased performance for the Negative class. These findings emphasise the advantages and disadvantages of each model and the significance of taking into account particular metrics depending on the task and class distribution. Overview of machine learning model performance is shown in [Table 2].

C. DISCUSSIONS

This study's objective is to identify racist tweets using sentiment analysis. The dataset is classified into positive and negative classifications for this reason. Positive classes suggest that there is no racist content in these tweets, whereas negative classes suggest that these tweets are racist since they express unfavourable attitudes about racism. As a result, a distribution of accuracy and right and incorrect predictions is given here with regard to the negative class.

A total of 31962 tweets—29720 positive tweets and 2242 negative tweets—are included in the gathered dataset. Machine learning models by themselves are unable to provide the best accuracy, however LSTM+ GCN and BERT do so, with LSTM's accuracy increasing to 0.99.

Model	Accuracy	Class	Precision	Recall	F1 score	Support
LR	0.95	Positive	0.96	0.99	0.98	9806
		Negative	0.84	0.51	0.63	742
		Macro Avg.	0.90	0.75	0.81	10548
		Weighted Avg.	0.96	0.96	0.95	10548
RF	0.95	Positive	0.96	0.99	0.98	9806
		Negative	0.84	0.51	0.63	742
		Macro Avg.	0.90	0.75	0.81	10548
		Weighted Avg.	0.96	0.96	0.95	10548
KNN	0.93	Positive	0.94	0.99	0.97	9806
		Negative	0.75	0.21	0.33	742
		Macro Avg.	0.84	0.60	0.65	10548
		Weighted Avg.	0.93	0.94	0.92	10548
DT	0.94	Positive	0.97	0.98	0.97	9806
		Negative	0.67	0.53	0.59	742
		Macro Avg.	0.82	0.76	0.78	10548
		Weighted Avg.	0.94	0.95	0.95	10548
SVM	0.92	Positive	0.93	1.00	0.96	9806
		Negative	0.00	0.00	0.00	742
		Macro Avg.	0.46	0.50	0.48	10548
		Weighted Avg.	0.86	0.93	0.90	10548
Voting Classifier	0.95	Positive	0.96	1.00	0.98	9806
		Negative	0.96	0.40	0.57	742
		Macro Avg.	0.96	0.76	0.77	10548
		Weighted Avg.	0.96	0.96	0.95	10548

TABLE 2.COMPARING ACCURACY OF MACHINE LEARNING MODELS

VI.CONCLUSION

Racist remarks are more common on social media sites like Twitter and should be automatically identified and blocked in order to stop them from spreading. In this study, racism is detected using sentiment analysis to identify tweets that include racist content by identifying unfavourable feelings. The LSTM + GCN model is employed to produce sentiment analysis with greater performance.

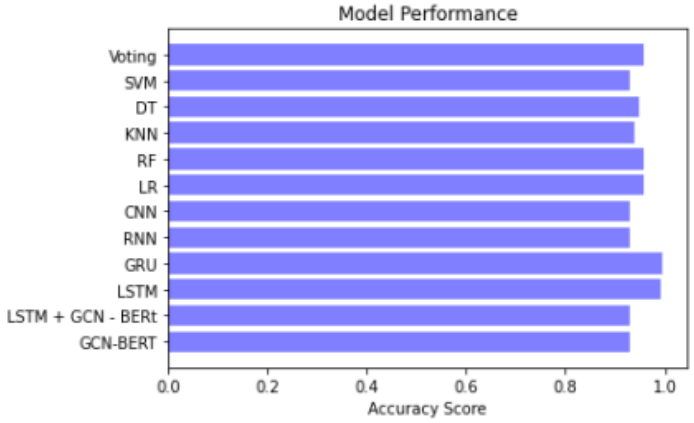

FIGURE 9. Performance chart of different models

We employ a sizable dataset of 31962 non-null tweets, of which 2242 are critical and 29720 are affirmative. The accuracy comparison of several deep learning and machine learning models is shown in [Fig]. The LSTM model clearly provides a higher accuracy score than other models.

VII.REFERENCES

[1] D. Williams and L. Cooper, ‘‘Reducing racial inequities in health: Using what we already know to take action,’’ Int. J. Environ. Res. Public Health, vol. 16, no. 4, p. 606, Feb. 2019.

[2] Y. Paradies, J. Ben, N. Denson, A. Elias, N. Priest, A. Pieterse, A. Gupta, M. Kelaher, and G. Gee, ‘‘Racism as a determinant of health: A systematic review and meta-analysis,’’ PLoS ONE, vol. 10, no. 9, Sep. 2015, Art. no. e0138511.

[3] J. C. Phelan and B. G. Link, ‘‘Is racism a fundamental cause of inequalities in health?’’ Annu. Rev. Sociol., vol. 41, no. 1, pp. 311–330, Aug. 2015.

[4] D. R. Williams, ‘‘Race and health: Basic questions, emerging directions,’’ Ann. Epidemiol., vol. 7, no. 5, pp. 322–333, Jul. 1997.

[5] D. R. Williams, J. A. Lawrence, B. A. Davis, and C. Vu,

''Understanding how discrimination can affect health,'' Health Services Res., vol. 54, no. S2, pp. 1374–1388, Dec. 2019.

[6] C. P. Jones, ''Levels of racism: A theoretic framework and a gardener's tale,'' Amer. J. Public Health, vol. 90, no. 8, p. 1212, 2000.

[7] S. Forrester, D. Jacobs, R. Zmora, P. Schreiner, V. Roger, and C. I. Kiefe, ''Racial differences in weathering and its associations with psychosocial stress: The CARDIA study,'' SSM-Population Health, vol. 7, Apr. 2019, Art. no. 100319.

[8] B. J. Goosby, J. E. Cheadle, and C. Mitchell, ''Stress-related biosocial mechanisms of discrimination and African American health inequities,'' Annu. Rev. Sociol., vol. 44, no. 1, pp. 319–340, Jul. 2018.

[9] A. T. Geronimus, M. Hicken, D. Keene, and J. Bound, '''Weathering' and age patterns of allostatic load scores among blacks and whites in the United States,'' Amer. J. Public Health, vol. 96, no. 5, pp. 826–833, 2006.

[10] A.-M. Bliuc, N. Faulkner, A. Jakubowicz, and C. McGarty, ''Online networks of racial hate: A systematic review of 10 years of research on cyberracism,'' Comput. Hum. Behav., vol. 87, pp. 75–86, Oct. 2018.

[11] R. Alshalan and H. Al-Khalifa, ''A deep learning approach for automatic hate speech detection in the Saudi Twittersphere,'' Appl. Sci., vol. 10, no. 23, p. 8614, Dec. 2020.

[12] A. Al-Hassan and H. Al-Dossari, ''Detection of hate speech in social networks: A survey on multilingual corpus,'' in Proc. 6th Int. Conf. Comput. Sci. Inf. Technol., vol. 10, 2019, pp. 1–19.

[13] A. Schmidt and M. Wiegand, ''A survey on hate speech detection using natural language processing,'' in Proc. 5th Int. Workshop Natural Lang. Process. Social Media, 2017, pp. 1–10.

[14] A. Alrehili, ''Automatic hate speech detection on social media: A brief survey,'' in Proc. IEEE/ACS 16th Int. Conf. Comput. Syst. Appl. (AICCSA), Nov. 2019, pp. 1–6.

[15] M. A. Al-garadi, M. R. Hussain, N. Khan, G. Murtaza, H. F. Nweke, I. Ali, G. Mujtaba, H. Chiroma, H. A. Khattak, and A. Gani, ''Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges,'' IEEE Access, vol. 7, pp. 70701–70718, 2019.

[16] K. Perifanos and D. Goutsos, ''Multimodal hate speech detection in Greek social media,'' Multimodal Technol. Interact., vol. 5, no. 7, p. 34, 2021.

[17] I. Aljarah, M. Habib, N. Hijazi, H. Faris, R. Qaddoura, B. Hammo, M. Abushariah, and M. Alfawareh, ''Intelligent detection of hate speech in arabic social network: A machine learning approach,'' J. Inf. Sci., vol. 47, no. 3, May 2020, Art. no. 0165551520917651.

[18] S. Goswami, M. Hudnurkar, and S. Ambekar, ''Fake news and hate speech detection with machine learning and NLP,'' PalArch's J. Archaeol. Egypt/Egyptol., vol. 17, no. 6, pp. 4309–4322, 2020.

[19] R. Alshalan and H. Al-Khalifa, ''A deep learning approach for automatic hate speech detection in the Saudi Twittersphere,'' Appl. Sci., vol. 10, no. 23, p. 8614, Dec. 2020.

[20] L. Ketsbaia, B. Issac, and X. Chen, ''Detection of hate tweets using machine learning and deep learning,'' in Proc. IEEE 19th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom), Dec. 2020, pp. 751–758.

[21] T. Putri, S. Sriadhi, R. Sari, R. Rahmadani, and H. Hutahaean, ''A comparison of classification algorithms for hate speech detection,'' IOP Conf. Ser., Mater. Sci. Eng., vol. 830, Apr. 2020, Art. no. 032006.

[22] U. Bhandary, ''Detection of hate speech in videos using machine learning,'' M.S. thesis, Dept. Comput. Sci., San Jose State Univ., San Jose, CA, USA, 2019.

[23] B. Vidgen and T. Yasseri, ''Detecting weak and strong Islamophobic hate speech on social media,'' J. Inf. Technol. Politics, vol. 17, no. 1, pp. 66–78, Jan. 2020.

[24] J. Salminen, M. Hopf, S. A. Chowdhury, S.-G. Jung, H. Almerekhi, and B. J. Jansen, ''Developing an online hate classifier for multiple social media platforms,'' Hum.-Centric Comput. Inf. Sci., vol. 10, no. 1, pp. 1–34, Dec. 2020.

[25] K. R. Kaiser, D. M. Kaiser, R. M. Kaiser, and A. M. Rackham, ''Using social media to understand and guide the treatment of racist ideology,'' Global J. Guid. Counseling Schools, Current Perspect., vol. 8, no. 1, pp. 38–49, Apr. 2018.

[26] F. Rustam, A. Mehmood, M. Ahmad, S. Ullah, D. M. Khan, and G. S. Choi, ''Classification of Shopify app user reviews using novel multi text features,'' IEEE Access, vol. 8, pp. 30234–30244, 2020.

[27] V. Rupapara, F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf, and G. S. Choi, ''Impact of SMOTE on imbalanced text features for toxic comments classification using RVVC model,'' IEEE Access, vol. 9, pp. 78621–78634, 2021.

[28] M. Mujahid, E. Lee, F. Rustam, P. B. Washington, S. Ullah, A. A. Reshi, and I. Ashraf, ''Sentiment analysis and topic modeling on tweets about online education during COVID-19,'' Appl. Sci., vol. 11, no. 18, p. 8438, Sep. 2021.

[29] L. E. Peterson, ''K-nearest neighbor,'' Scholarpedia, vol. 4, no. 2, p. 1883, 2009.

[30] P. H. Swain and H. Hauska, ''The decision tree classifier:

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

Design and potential,'' IEEE Trans. Geosci. Electron., vol. GE-15, no. 3, pp. 142–147, Jul. 1977.

[31] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, ''A performance comparison of supervised machine learning models for COVID-19 tweets sentiment analysis,'' PLoS ONE, vol. 16, no. 2, Feb. 2021, Art. no. e0245909.

[32] R. Ponnala and C. R. K. Reddy, "Software Defect Prediction using Machine Learning Algorithms: Current State of the Art," Solid State Technol., vol. 64, no. 2, 2021.

[33] Lakshmi Sreenivasa Reddy D and Ramchander M,"A Model for Improving Classifier Accuracy using Outlier Analysis Methods",ISSN:1687-4846, Delaware, USA ,December 2015