International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.144

Uncovering Racist Sentiments on Twitter: A Stacked GCR-LSTM with BERT Approach for Analysis of Differing Opinions.

Sudipta Shaw
Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur, India
ss5996@srmist.edu.in

Sanjay krishna
Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur, India
sk5387@srmist.edu.in

K. ShanthaKumari
Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur, India
shanthak@srmist.edu.in

Shyam Mehta
Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur, India
shyammehta016@gmail.com

Tanisha Bisht
Computing technologies
SRM Institute of Science and Technology
Kattankulathur, India
tb8439@srmist.edu.in

Parth Shah
Electronics and Instrumentation
SRM Institute of Science and Technology
Kattankulathur, India
pc9110@srmist.edu.in in

Abstract— For all its benefits social media has also enabled the darker aspects of human nature like racism to flourish - often hidden behind seemingly benign posts or anonymous accounts. Such practices have become especially common when dealing with issues related to skin colour, culture ,language origin or religion .This trend is not only harmful but threatens our cultural stability ,social fabric and peace on an international level . To counteract this situation a recent research suggests the use of deep learning tools such as gated recurrent units (GRU) convolutional neural networks (CNNs) with LSTM ,and ensemble modelling along with sentiment analysis aimed at locating racist tweets For those interested in machine learning, its worth noting the capabilities of Gated Convolutional Recurrent Neural Networks with LSTM and BERT (GCR-NN+lLSTM+BERT). Within this framework the GRU component is adept at extracting specific characteristics from unprocessed text while CNN can provide necessary context for RNN to more accurately predict outcomes alongside the usage of LSTM and BERT.

I. INTRODUCTION

Social media has become a breeding ground for racism, where individuals spread hate speech, bigotry, and prejudice against people based on their race, religion, language, culture, and national origin. Racism on social media can take various forms, including the use of memes, false identities, and overt expressions. Unfortunately, one of the negative consequences of this widespread use of social media is the rise of vices such as racism. Twitter, for example, has emerged as a platform where racism and its associated tensions are prevalent. With over a million words and contexts and memes followed by different opinions and content,Twitter's reach and influence are vast. Additionally, 90% of Twitter users have a public profile, making it easier for harmful sentiments to spread quickly. Today, 22% of American citizens utilise the social media network. Twitter users can respond to and participate in tweets by publishing them on their profiles (retweeting), tagging other users, and clicking the like button. Tweets are publicly available until they are set to be private. The foundation of sentimental analysis is the expression of sentiments, emotions, attitudes, and ideas on Twitter. Social media platforms have become increasingly popular, which has led to widespread use of them for several contemporary and historical kinds of racism. Racism is depicted on these platforms in both overt and covert ways, such as through memes and the publication of racist Tweets under fictitious names. Racism is not only limited to ethnicity, but also targets people based on their race, national origin, language, culture, and religion. Racial tensions incited on social media are considered a major threat to global peace and stability, as well as social, political, and cultural stability.Racist statements should be quickly identified and prohibited from social media, which is the main source of racist ideas.

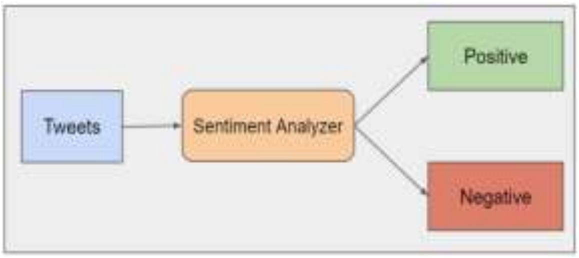

Fig.1: Example figure

Racist comments and social media tweets have been connected to a number of physical and mental disorders, which have a harmful impact on one's health [7–12]. Three categories of racism on social media can be identified: institutionalised, personally mediated, and internalised [13]. Racism can be experienced at a personal level through discriminatory or unfair treatment, as well as by being aware of bias towards loved ones. This has negative effects on individuals, leading to various forms of psychosocial stress that can increase the risk of chronic illnesses. Additionally, cyber-racism is being spread by racist groups and individuals who employ advanced strategies and skills. As a result, sentiment analysis has become a significant area of research to analyze social media text for tasks such as identifying hate speech, making sentiment-based market predictions, and detecting racism, among others.

II. LITERATUREREVIEW

A. *Using social media to understand and guide the treatment of racist ideology*

DOI: rp-9788770040723.144

The paper "Using social media to understand and guide the treatment of racist ideology" proposes using sentiment analysis to detect racist tweets and guide the treatment of racist ideology. The authors highlight the growing prevalence of racism on social media and its impact on societal, political, and cultural stability. The study combines gated recurrent units (GRU) and convolutional neural networks (CNN) to create a stacked ensemble deep learning model called GCR-NN.Several previous studies have also addressed the issue of racism on social media. For instance, a study by Zhou et al. (2018) proposed using machine learning techniques to detect racist tweets and assess their impact on users. Another study by Ngo et al. (2019) used sentiment analysis to detect hate speech and racist content on social media. The authors highlighted the importance of detecting and addressing such content to prevent its negative impact on individuals and society.Moreover, other studies have examined the impact of social media on racism and prejudice. For example, a study by Correa et al..

B. Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery

The paper "Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery" explores the potential benefits and challenges of using social media for health research. The authors highlight the increasing popularity of social media platforms as a means of communication and data sharing, and how they can be utilised for health research purposes. The paper discusses various methodological and ethical considerations that researchers must take into account when using social media for recruitment and intervention delivery. For instance, the authors highlight the importance of ensuring participant confidentiality, privacy, and informed consent. The study also notes the challenges of sample representativeness, potential selection bias, and the reliability of self-reported data collected through social media.Several previous studies have also examined the use of social media for health research. For example, a study by Laranjo et al. (2015) explored the potential of social media for health promotion and intervention delivery, highlighting the benefits of social media as a cost-effective and accessible means of communication with participants. Another study by Peng et al. (2017) examined the feasibility of using social media for patient recruitment in clinical trials, and found that social media platforms can be effective in reaching a diverse and geographically dispersed sample.Overall, the paper highlights the potential benefits of using social media for health research, but also underscores the need for careful consideration of methodological and ethical issues when utilizing these platforms.

C. Online networks of racial hate: A systematic review of 10 years of research on cyber-racism

The paper "Online networks of racial hate: A systematic review of 10 years of research on cyber-racism" provides a comprehensive analysis of research on cyber-racism, focusing on online networks of racial hate. The authors examine the prevalence, forms, and impact of cyber-racism, and highlight the need for effective strategies to counter its negative effects.The paper reviews a range of studies on cyber-racism published over a 10-year period, highlighting the various forms of racist content, including hate speech, discrimination, and harassment. The authors also discuss the social and psychological impact of cyber-racism on individuals, such as anxiety, depression, and decreased self-esteem. The study also notes the potential for cyber-racism to have wider social and political consequences, including the normalization of racist attitudes and the promotion of extremist ideologies.Several previous studies have also addressed the issue of cyber-racism. For example, a study by Tynes et al. (2012) examined the experiences of racial and ethnic minority youth on social media, highlighting the prevalence of racist content and the negative impact on their mental health. Another study by Ziegele et al. (2018) explored the role of social media in the dissemination and normalization of extremist ideologies, including racist and anti-immigrant attitudes.Overall, the paper highlights the need for effective strategies to counter cyber-racism, including education, community-building, and policy interventions. The authors emphasize the importance of taking a multi-disciplinary and multi-level approach to addressing the problem of cyber-racism, and call for further research to better understand the complex dynamics of online networks of racial hate.

C. Reducing racial inequities in health: Using what we already know to take action

The paper "Reducing racial inequities in health: Using what we already know to take action" provides a comprehensive review of the literature on racial inequities in health and highlights evidence-based strategies for reducing these inequities.The authors review a range of studies on the social determinants of health and the impact of racism and discrimination on health outcomes. They note that racial and ethnic minorities experience higher rates of chronic disease, disability, and premature death compared to their White counterparts, and that these disparities are driven by a range of factors, including poverty, inadequate healthcare access, and exposure to environmental toxins.

The paper also highlights evidence-based strategies for addressing racial health inequities, including policies that promote economic and educational opportunities, improve access to quality healthcare, and address structural racism and discrimination. The authors note the importance of community engagement and participation in efforts to promote health equity, and highlight the need for culturally sensitive interventions that address the unique needs and experiences of diverse populations.Several previous studies have also addressed the issue of racial health inequities. For example, a study by Williams et al. (2019) examined the impact of racism on health outcomes, highlighting the need for policies and interventions that address the root causes of these disparities. Another study by Braveman et al. (2017) identified a range of evidence-based strategies for promoting health equity, including community-based interventions and policies that address social determinants of health.Overall, the paperemphasises the need for a comprehensive, multi-disciplinary approach to addressing racial health inequities, and highlights the importance of using evidence-based strategies to guide policy and practice.

DOI: rp-9788770040723.144

The authors call for continued research and action to address these disparities and promote health equity for all populations.

III. METHODOLOGY

A. Overview

Racism on social media has become an increasingly prominent issue in recent years. Social media platforms provide a platform for individuals to express their opinions and beliefs, which can often include racist comments and hate speech. The anonymity of social media has made it easier for individuals to express racist views without fear of repercussions. Additionally, the speed and reach of social media mean that racist comments and hate speech can spread quickly, reaching a large audience.Machine learning and deep learning techniques have been used to address the problem of racism in social media. Sentiment analysis, a subfield of natural language processing, has been used to automatically identify and classify tweets with racist content. These techniques can analyze the language used in tweets to determine whether they contain negative sentiments, and identify the tweets that express racist opinions.However, there are also concerns about the use of machine learning and deep learning in detecting racism on social media. There is a risk of bias in the algorithms used to detect racism, which can lead to false positives or false negatives. Additionally, there are ethical concerns about the use of automated systems to monitor and censor social media content, which can infringe on freedom of speech. Overall, the issue of racism in social media is complex and requires careful consideration of both technical and ethical issues.

B. Disadvantage

1. Existing technology cannot be automatically identified and stopped to prevent future proliferation.
2. Poor performance.

Due to the improvement in performance of Deeplearning networks, To expand the stack for improved performance, gated recurrent units (GRUs), convolutional neural networks (CNNs), recurrent neural networks (RNNs) and LSTM have been combined in the gated convolutional recurrent neural network (GCR-NN) model. In this model, GRUs play a crucial role in extracting relevant and significant features from raw text, while CNNs extract essential aspects of RNN to ensure accurate predictions and on which we add LSTM to generate better output on the accurate predictions.

C. Advantage

1. The proposed GCR-NN with BERT proposes advanced and better performance with enhanced accuracy.
2. The proposed GCR-NN model can detect racist remarks in 97percent of total of tweets.

We developed the following modules in order to carry out the project mentioned earlier.

- Data input: With this module, we will enter data into the system.
- Processing:This module is intended for reading data to be processed.
- This module will be used to partition the data into testing and training models.
- Model creation techniques include GCN with BERT, LSTM, GRU, RNN, CNN, Ensemble Method LSTM + GCN with BERT, Logistic Regression, Random Forest, KNN, Decision Tree, Support Vector Machine, and Voting Classifier.
- Signing up as a user and logging in are required steps for using this module.
- User input: Prediction input will be produced by using this module.
- Forecast: The final forecasted figure will be shown. Logistic Regression, Random Forest, KNN, Decision Tree, Support Vector Machine, Voting Classifier.

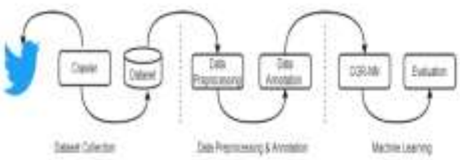

Fig.2: System architecture

IV. IMPLEMENTATION

Algorithms

GCN: For graph-structured data, a Graph Convolutional Network, or GCN, is a semi-supervised learning technique. It is founded on a powerful variant of convolutional neural networks that directly affect graphs.

BERT: BERT is an open source machine learning toolkit for interpreting natural language (NLP). Using the use of contextual information from the surrounding text,the objective of BERT is to assist computers in comprehending the context and meaning of ambiguous words in text.

LSTMs are a type of neural network architecture that can be used in sentiment analysis. They are designed to help overcome the problem of vanishing gradients that can occur in traditional RNNs, allowing them to more effectively capture long-term dependencies in sequential data. In sentiment analysis, LSTMs can be used to model the sentiment of a text sequence by analyzing its context and identifying patterns that are indicative of positive or negative sentiment. By processing text in this way, LSTMs can be used to accurately classify the sentiment of a piece of text, making them a powerful tool for sentiment analysis tasks.

*GRU:*Kyunghyun Cho et al. developed the gated recurrent units (GRUs) recurrent neural network gating approach in 2014. The GRU has fewer parameters because it doesn't have an output gate, but it performs similarly to an LSTM with a forget gate.

RNN: RNNs are designed to handle sequential data, such as sentences or paragraphs, which is important for sentiment analysis since the sentiment of a sentence can

DOI: rp-9788770040723.144

often depend on the sentiment of the preceding sentence.In sentiment analysis, RNNs work by processing input text one word at a time and producing a hidden state that summarises the sentiment information learned so far. The hidden state is updated for each new word in the sentence, taking into account both the current word and the previous hidden state. This allows the RNN to learn the context of the words and how they contribute to the overall sentiment of the sentence.The final hidden state produced by the RNN can then be used to predict the sentiment of the sentence as a whole. This is typically done by passing the hidden state through a classifier, such as a fully connected neural network or logistic regression model, which outputs a probability distribution over the possible sentiment labels (e.g. positive or negative).RNNs are particularly useful for sentiment analysis tasks where the length of the input text can vary, since they are able to handle variable-length input sequences.

CNN: Convolutional Neural Networks (CNNs) are used in sentiment analysis to extract important features from textual data. They are commonly applied to the task of sentence classification where each sentence is treated as a sequence of words. CNNs employ filters that scan the sentence, and through convolution operations, they extract features such as n-grams, sentiment words, and word embeddings. These features are then fed into a fully connected layer for classification. CNNs have shown promising results in sentiment analysis due to their ability to capture local and global information from the input data. Additionally, they require less computational resources compared to recurrent neural networks.

Ensemble Method: Rather than using a single model, ensemble methods combine multiple models in an effort to increase model accuracy. The precision of the results is significantly increased by the integrated models. The popularity of ensemble techniques in machine learning has increased as a result.

Logistic Regression: Logistic regression is a statistical method for binary classification, which can also be applied to sentiment analysis. It uses a linear function to model the relationship between the input variables (in this case, the text data) and the binary output variable (positive or negative sentiment).

*Random Forest:*Random forest is a machine learning algorithm that can be used in sentiment analysis. It works by creating a forest of decision trees, where each tree makes a prediction about the sentiment of a particular input. The algorithm then aggregates the predictions from all the trees to make a final prediction. Random forest can handle large datasets with many features and has good accuracy. It can be trained on annotated data and then applied to new data to classify sentiment as positive, negative, or neutral.

K-Nearest Neighbours (KNN): KNN is a non-parametric machine learning algorithm that can be used for sentiment analysis. It is based on the idea that similar data points should have similar labels. In the context of sentiment analysis, KNN can be used to classify a given text by finding the k-nearest neighbours (i.e., texts with similar features) and using their labels to make a prediction.

Support Vector Machines (SVM): SVM is a popular machine learning algorithm used for classification tasks, including sentiment analysis of tweets. SVM works by identifying the best hyperplane (i.e., boundary) that separates the different classes in the data, such as positive and negative sentiment. It then uses this hyperplane to classify new data points based on which side of the boundary they fall on.

*Voting Classifier:*The voting classifier is a machine learning technique used in sentiment analysis of tweets on Twitter. It is an ensemble method that combines multiple algorithms to make a final prediction based on the outputs of the individual models. In this approach, multiple models are trained on the same dataset using different algorithms like decision trees, random forests, and support vector machines (SVM), each with its own strengths and weaknesses. Then, the voting classifier combines the predictions of all these models to make the final decision. This method has been shown to improve the accuracy of sentiment analysis on Twitter data and can be useful for analyzing large datasets with complex features.

Proposed Architecture and Development

The proposed architecture is based on GCR-NN with LSTM And BERT Tokenisation where feature selection is done using BOW feature and TF score or the polarity score. Along with the proposed deep learning architecture we have used an ensemble method of using a voting classifier. In the deep learning proposed architecture the whole model is an stacked ensemble with various different layers which has GRU, followed by CNN ,on which RNN is stacked for output accuracy and then LSTM with Bert tokenisation is added to enhance the accuracy of the whole model. While in the voting classifier we have various different machine learning models which are then further processed to get an higher advantage on accuracy . Followed by which a comparison is done on which model provides better accuracy on the given sentiment analysis.

V. FUTURE ENHANCEMENTS

We can add this same model on localised dataset for example Bangla language racism detection which can be used to detect racism as in a certain organisation and area and can be further deduced to other languages and context.Also further this model and website could you be used for live analysation of social media on a day to day basis just as GPS works though the processing time could be of an greater issue.

VI. EXPERIMENTAL RESULTS

DOI: rp-9788770040723.144

Fig.3: Home

Fig.4: User registration

Fig.5: user login

Fig.6: Main page

Fig.7: Userpage input

Fig.8: Model result

VII. CONCLUSION

This study focuses on the problem of the growing frequency of racist language on social media websites and proposes the use of sentiment analysis as a means of detecting negative sentiment and identifying tweets containing racist content. The study utilizes deep learning techniques, as in the use of a stacked ensemble model of LSTM and GCR-NN with BERT tokenisation along with the usage of another ensemble such as voting classifier, to effectively carry out sentiment analysis. The proposed model is evaluated alongside other models using a large Twitter dataset annotated with TextBlobs. The findings indicate that 31.49% of the 169,999 tweets surveyed were racist, and the GCR+LSTM +BERT +NN model achieved an impressive average accuracy score of 0.98 for positive, negative, and neutral sentiment classification. Notably, the model demonstrated superior performance in detecting racist tweets, with a 97% accuracy rate and only a 3% error rate. The study also compared the performance of machine learning models in identifying racist tweets, with both models correctly identifying 96% and 95% of such tweets, respectively, and misclassifying 4% and 5% of tweets as racist. These results suggest that deep learning methods, such as the proposed GCR-NN+LSTM+BERT model,offer potential for identifying and obstructing racist language on social media sites.

REFERENCES

[1] 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.

[2] Rajesh, M., &Sitharthan, R. (2022). Image fusion and enhancement based on energy of the pixel using Deep Convolutional Neural Network. Multimedia Tools and Applications, 81(1), 873-885.

[3] M. Ahlgren. 40C Twitter Statistics & Facts. Accessed: Sep. 1, 2021. [Online]. Available: https://www.websitehostingrating.com/twitterstatistics/

[4] D. Arigo, S. Pagoto, L. Carter-Harris, S. E. Lillie, and C. Nebeker, ''Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery,'' Digit. Health, vol. 4, Jan. 2018, Art. no. 205520761877175.

[5] Pazhani. A, A. J., Gunasekaran, P., Shanmuganathan, V., Lim, S., Madasamy, K., Manoharan, R., &Verma, A. (2022).Peer–Peer Communication Using Novel Slice Handover Algorithm for 5G Wireless Networks.Journal of Sensor and Actuator Networks, 11(4), 82.

[6] M. A. Price, J. R. Weisz, S. McKetta, N. L. Hollinsaid, M. R. Lattanner, A. E. Reid, and M. L. Hatzenbuehler, ''Meta-analysis: Are psychotherapies less effective for black youth in communities with higher levels of anti-black racism?'' J. Amer. Acad. Child Adolescent Psychiatry, 2021, doi: 10.1016/j.jaac.2021.07.808. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0890856721012818

[7] 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.

[8] 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.

[9] 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.

DOI: rp-9788770040723.144
[10] D. R. Williams, ''Race and health: Basic questions, emerging directions,'' Ann. Epidemiol., vol. 7, no. 5, pp. 322–333, Jul. 1997.

[11] Z. D. Bailey, N. Krieger, M. Agénor, J. Graves, N. Linos, and M. T. Bassett, ''Structural racism and health inequities in the USA: Evidence and interventions,'' Lancet, vol. 389, no. 10077, pp. 1453–1463, Apr. 2017.

[12] 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.

[13] 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.

[14] 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.

[15] 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.