JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets

Prof. S. L. Farpat 1) Prerna Suresh Nikam 2) Sayali Pradip Patil 3) Nandkishor N. Wankhade 4) Sh Altab Sh Rafik.
Computer Science Engineering Department, Padm. Dr. VBKCOE, Malkapur, Maharashtra, India.

ABSTRACT *- The primary objective of this project is to develop an effective system for detecting racism in social media content, specifically focusing on tweets. The system aims to analyze differential opinions expressed in tweets through sentiment analysis, identifying instances of racist language or sentiments. By doing so, the project seeks to contribute to the prevention and mitigation of online racism, which has become a significant social issue in today's digital age.*

Furthermore, the project aims to leverage deep learning techniques to enhance the accuracy and efficiency of racism detection in tweets. Specifically, it intends to employ a stacked ensemble deep learning model composed of a gated recurrent unit (GRU), a convolutional neural network (CNN), and a gated convolutional recurrent neural network (GCR-NN). By integrating these advanced deep learning architectures, the project aims to improve the system's ability to extract relevant features from raw text and accurately identify racist content in tweets.

Another objective is to conduct comprehensive experiments and performance evaluations to assess the effectiveness of the proposed GCR-NN model in comparison to other machine learning and deep learning approaches. Through rigorous testing and analysis, the project seeks to demonstrate the superiority of the GCR-NN model in accurately detecting tweets containing racist comments, achieving high levels of accuracy, precision, recall, and overall performance.

Overall, the project's objectives revolve around developing an advanced system for racism detection in tweets, leveraging deep learning techniques to enhance accuracy, and conducting thorough evaluations to validate the effectiveness of the proposed approach. By achieving these objectives, the project aims to contribute to the promotion of a safer and more inclusive online environment, free from the harmful effects of racism and discrimination.

Keywords *- Racism, social media, online abuse, Twitter, deep learning..*

INTRODUCTION *- Social media has become a dominant element in thesocio-political perspective and controls our minds and actions indifferent ways. Due to the widespread use of social media platformsand freedom of expression, some evils have appeared in recent years, one of themost important of which is racism. Social media sites such as Twitter represent anew environment where racism and related stress seem to thrive [1]. Twitter is currently used by 22% of US adults 4,444 [2], while Twitter has 1.3 billion accounts and 4,444,336 million active users worldwide, 90% of which have 4,444 public profiles, resulting in 500 million tweets per day [3] .If tweets are not made private, they are publicly available and Twitter users can react and engage with such tweets by sharing them on their profile (re tweeting), tagging someone's username, clicking the favorite button or replying to the author. from the tweet. tweet [4]. On Twitter, the expression of feelings, feelings, attitudes and opinions create the raw data for sentimental analysis [5] The growing popularity of social media platforms has led to their widespread use in many old and new racist forms. practices [6]. Racism is expressed on such platforms in various hidden forms, such as memes, and overtly such as sending tweets containing racist remarks under false identities. Although racism is often associated with ethnicity, it now develops on the basis of skin color, origin, language, culture, and especially religion. Social media opinions and remarks inciting racial differences were considered a serious threat to social, political and cultural stability and peace in various countries. Social mediais the main source of spreading racist view sshould be monitored and racist remarks detected and prevented in time. Racist comments and tweets in social media have been identified as a source of several diseases with negative mental and physical consequences [7]–[12]. Considering its use in social media, racism can be classified in to three groups: institutionalized, personally mediated, and internalized [13]. Personally mediated racism can be experienced through racial or ethnic discrimination or conscious discrimination by family and friends. Consequently, racist behavior in societies negatively affects individuals and creates a wide range of psychosocial distress, often leading to chronic disease risk [14]–[16]. In addition, racist groups and individuals perpetuate cyber-racism using increased skill and sophistication through various channels and strategies [5].Special attention has been paid to sentiment analysis to analyze text on social media platforms for various tasks, including hate speech detection, sentiment-based market forecasting and racism detection, etc. The use of social media is a potential source of information generation, which contains important information about people's attitudes, reactions, feelings and opinions about certain events, objects, personalities and entities. Sentiment analysis provides powerful tools to mine such data for sentiment analysis. The vast majority of Twitter streams are characterized less by coherent rational* *discourse and more by a flood of emotions and impressions, and can be used to divide stories into good and bad poles [17], [18]. Research shows that things can become less obvious than a shared sense of outrage and a compelling sense of consensus and that Twitter channels can be quite insular and node-specific [19]. Given its widespread use, social media has become an attractive source for understanding attitudes and analyzing communication on sensitive topics such as racism. In the United States, 4,444 race- and ethnicity-related discussions on Twitter were considered indicators of the current state of 4,444 relationships based on 4,444 races. In addition, 4,444 differences in the types of conversations about racism indicate geographic variation in racial attitudes and feelings [20]. Thus, by analyzing the details of how people, events and circumstances are presented, the dynamics of user interaction becomes clear, and many issues related to racism can be exposed on this platform. Due to the extreme and atypical racist attitude, the face of individuals is associated with personal characteristics and attitudes; one can be easily relativized, contextualized and thus depoliticized. This leads to a blurring of real and specific structural inequalities in society experienced by certain ethnic groups [21].Machine and deep learning approaches have proven their strength and superiority over traditional methods in several fields, such as image processing [22], [23], text classification [24], [25], and sentiment analysis is no exception. Several recent studies show that machine learning techniques perform better in sentiment analysis tasks [26], [27]. Therefore, this study uses machine learning and deep learning models to perform sentiment analysis on racism-related tweets, and makes the following contribution. As shown, machine and deep learning approaches have proven their strength and superiority over traditional methods in several domains. Processing [22], [23], text classification [24], [25] and sentiment analysis is no exception. Several recent studies show that machine learning techniques perform better in sentiment analysis tasks [26], [27]. Therefore, this study uses machine learning and deep learning models to perform sentiment analysis on racism-related tweets and makes the following contributions:• A general model using recurrent neural networks is proposed. For this purpose, gate recurrent unit (GRU), convolutional neural network and recurrent neural network are stacked to create a GCR-NN model for sentiment analysis. • A large number of tweets containing racist comments/text are. Indexed by Twitter for use by the research community. The dataset is labeled as positive, negative, and neutral emotionsTextBlob based on a polarity score.• Several well-known machine learning models using optimized parameters such as decision tree (DT), random forest (RF), logistic regression (LR), k nearest neighbor (KNN), and support vector machines (SVM) have been applied. To compare performance. Term Frequency-Inverse Document Frequency (TF-IDF)and Bag of Words (Bow) were investigated as feature extraction techniques.• For a fair comparison with the proposed approach, GRU, Long Short Term Memory (LSTM), CNN and Snare implemented as independent models. Similarly, the performance of several state-of-the-art models is compared with the proposed GCR-NN in terms of precision, accuracy, recall and F1 scores. The rest of the paper is organized as follows. Part II describes several important investigations related to current research. The proposed approach, dataset and description of machine learning algorithms are presented in Section III.Part IV contain the analysis and discussion of the results. Finally, Section V concludes...*

PROBLEM FORMULATION -

The growing prevalence of racism on social media platforms, particularly in the form of tweets, has become a pressing concern in contemporary society. Racism manifests itself in various ways, ranging from covert expressions hidden behind memes to overt comments inciting hatred, violence, and social unrest. This online racism poses a significant threat to social, political, and cultural stability, as it targets individuals based on attributes such as skin color, origin, language, culture, and religion. Existing on social media under the guise of fake identities and inflammatory comments, racism represents a serious challenge that demands attention and intervention.

The problem formulation in this research paper revolves around the need to develop effective methods for detecting and mitigating racism within the vast landscape of social media, with a specific focus on Twitter. The challenge lies in identifying tweets containing racist content, a task complicated by the diverse and evolving nature of online expressions. Traditional approaches often fall short in addressing this issue, necessitating the exploration of advanced techniques, particularly sentiment analysis, to uncover the differential opinions expressed in tweets.

Additionally, the paper addresses the limitation of conventional sentiment analysis methods and emphasizes the adoption of deep learning models to enhance the accuracy and efficiency of racism detection. The formulation involves the assembly of a stacked ensemble deep learning model, combining a gated recurrent unit (GRU), a convolutional neural network (CNN), and a gated convolutional recurrent neural network (GCR-NN). This integration aims to harness the strengths of each component in extracting relevant and salient features from raw text, facilitating precise identification of racist comments within tweets.

In summary, the problem formulation revolves around the imperative to combat online racism, specifically on Twitter, by advancing sentiment analysis methodologies, leveraging deep learning models, and proposing a novel ensemble approach to enhance the accuracy of racism detection in tweets.

PROPOSE SYSTEM METHODOLOGY -

The proposed system methodology for racism detection through sentiment analysis of tweets involves a sophisticated combination of advanced techniques, focusing on the utilization of deep learning models. The overarching goal is to develop an effective and accurate system capable of identifying tweets containing racist content. The following steps outline the proposed methodology:

Data Collection:

Gather a diverse and representative dataset of tweets containing a mix of racist and non-racist content. This dataset will serve as the foundation for training and evaluating the proposed model.

Preprocessing:

Clean and preprocess the collected tweets to remove noise, irrelevant information, and ensure standardized formatting.

Tokenization, stemming, and other text processing techniques will be applied to prepare the text data for analysis.

Sentiment Analysis Models:

Implement and compare various sentiment analysis models to understand the sentiments expressed in tweets. Traditional sentiment analysis models, such as rule-based approaches and machine learning-based models, will be employed for baseline comparisons.

Deep Learning Architecture:

Develop a stacked ensemble deep learning model, combining a gated recurrent unit (GRU), a convolutional neural network (CNN), and a gated convolutional recurrent neural network (GCR-NN). This architecture aims to capture intricate patterns and dependencies within the tweet text, enhancing the discrimination between racist and non-racist content.

Training and Validation:

Train the deep learning model on the prepared dataset, using a subset for validation to fine-tune hyper parameters and prevent overfitting. The model will learn to recognize subtle linguistic nuances indicative of racist sentiments.

Performance Evaluation:

Evaluate the proposed model's performance using metrics such as accuracy, precision, recall, and F1-score. Compare the results with traditional sentiment analysis models to demonstrate the superiority of the deep learning approach in detecting racism in tweets.

Optimization:

Fine-tune the model based on the evaluation results, optimizing hyper parameters, adjusting the architecture, and incorporating feedback from the validation set to enhance overall performance.

Implementation:

Implement the optimized model into a practical application or tool capable of real-time or batch processing of tweets. This implementation will allow for the scalable and efficient detection of racist content within the dynamic context of social media.

User Interface (Optional):

If applicable, design a user-friendly interface that allows users to interact with the system, providing feedback and potentially improving the model through a feedback loop.

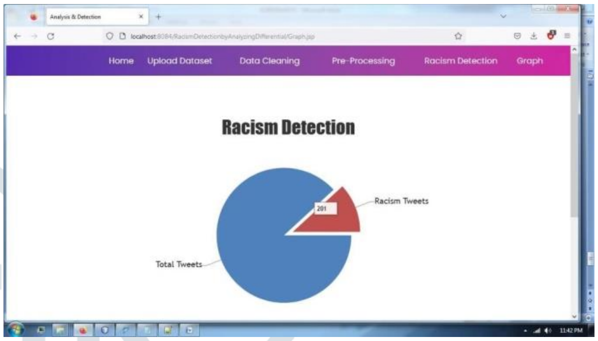
Documentation and Reporting:

Document the methodology, findings, and outcomes comprehensively. Generate a report detailing the system's capabilities, limitations, and potential areas for future improvement.
This proposed system methodology combines the strengths of traditional sentiment analysis, machine learning, and deep learning to tackle the nuanced challenge of racism detection in tweets, aiming for a high level of accuracy and efficiency.

WORKING ON LANGUAGES -

In summary, the chosen technology stack of Python (ES13), Anaconda with Jupyter, CSV files, and a web page server creates an efficient and effective environment for developing language processing solutions, emphasizing ease of use, data management, and user interaction.

ARCHITECURES –



RELATED WORKING -

The surge in hate crimes facilitated by the widespread use of social media and the shield of online anonymity has become a pressing concern [37]. The manifestation of abusive content on social media, characterized by harassment and maltreatment, triggers negative emotions among users, prompting discourteous expressions [31]. Cyber bullying and hate speeches, exemplifying abusive language, have garnered substantial attention from researchers due to their detrimental impact on society. The imperative to cleanse these contents has led to numerous studies exploring the automatic detection of hate speeches using machine learning algorithms [39].

A specific case study focuses on Greek social media, particularly Twitter messages containing racist speech and xenophobia directed at migrants and refugees [28]. The model employs an ensemble approach, transfer learning, and fine-tuning of bidirectional encoder representations from transformers (BERT) and Resnet on the collected dataset. The results reveal high accuracy, with the text modality achieving 0.944 accuracy using nlpaueb/greek-bert and 0.97 with resnet18+ nlpaueb/greek-bert using text+image modality. Another study addresses hate speech detection in Arabic social media networks, employing various machine learning algorithms such as Naïve Bayes, Decision Trees, Support Vector Machines, and Random Forests with different feature sets [29]. The highest accuracy of 0.913 is achieved using Random Forest with TF-IDF and profile-related features.

In the realm of fake news and hate speech propaganda, a study classifies content using extracted features from both fake and real news [30]. The models, including Naïve Bayes, Extreme Random Trees, Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), and an ensemble of CNN and GRU, achieve an F1 score of 0.79 and an AUROC of 0.89 using the CNN model. Additionally, the automatic detection of cyber bullying is investigated using different word embedding techniques and neural networks [32]. The study achieves an impressive accuracy of 96.67% for one dataset and 97.5% for *the second dataset using a neural network with three hidden layers and Doc2Vec features. Another study delves into the automatic detection of hate speech or racism in Indonesian tweets, utilizing machine learning models such as Multinomial Naïve Bayes, Multilayer Perceptron, Ada Boost classifier, and SVM with SMOTE as an up sampling technique [33]. The results highlight the efficacy of MLP with SMOTE features, achieving an accuracy of 83.4%, while AB and MNB achieve 71.2% accuracy for non-SMOTE features. These studies collectively contribute to the ongoing efforts in automatic hate speech detection, addressing the diverse linguistic and cultural nuances present in social media content.*

MATERIALS AND METHODS

A. PROPOSED METHODOLOGY

The proposed approach for racism detection on social media platforms employs machine learning and deep learning techniques. The process begins with the crawling of Twitter data, followed by data cleaning, preprocessing, and data annotation. The collected tweets, related to racist comments, are gathered using keywords like '#racism', '#racial', and '#racist' within the period of 29 July 2021 to 6 August 2021. A total of 169,999 tweets matching the criteria are collected using the 'Twint library,' and essential attributes such as 'username,' 'date,' 'location,' and 'content' are extracted. The dataset is then preprocessed to enhance the quality of the data for effective model training.

B. DATASET DESCRIPTION

The racism tweets dataset is sourced from Twitter due to its widespread usage and significant user engagement. The study focuses on analyzing racism trends based on Twitter posts, utilizing keywords to collect relevant data. A total of 169,999 tweets are collected within the specified time frame, and attributes such as 'username,' 'date,' 'location,' and 'content' are extracted for further analysis.

C. DATA PREPROCESSING

Data preprocessing involves several steps to clean and prepare the dataset for training. Natural language processing (NLP) methods using the natural language toolkit (NLTK) of Python are applied for preprocessing. Key steps include tokenization, stemming, lemmatization, stop words exclusion, case normalization, and noise removal. Tokenization involves breaking sentences into constituent words, stemming reduces words to their root forms, lemmatization retains the root form based on context, stop words are excluded to enhance learning efficiency, case normalization converts text to lowercase, and noise removal eliminates unwanted characters and symbols.

D. DATA ANNOTATION

To annotate the dataset with positive, negative, and neutral sentiments, the TextBlob library is utilized. TextBlob calculates the polarity score for each text, which is then used to assign a sentiment label. The polarity score ranges from -1 to 1, and sentiments are categorized as positive, negative, or neutral based on the score. This annotation process helps in training the model to identify sentiments associated with the tweets, specifically focusing on racism detection.

RESULTS AND DISCUSSION

Experiments on sentiment analysis for racism tweets were conducted on an Intel Core i7 11th generation machine operating on Windows 10. The implementation of machine learning and deep learning models took place on Jupyter in Python, utilizing frameworks such as TensorFlow, Keras, and Scikit-learn. The performance evaluation metrics include accuracy, precision, recall, F1 score, the number of correct predictions, and the number of wrong predictions.

A. VISUAL REPRESENTATION OF SENTIMENT DISTRIBUTION

The distribution of the dataset is visualized based on the top four countries with the highest number of racist tweets. Figure 4a illustrates that the US has the highest number of tweets, followed by the United Kingdom (UK), Nigeria, and the Republic of South Africa (RSA). The sentiment distribution for each country reveals that the majority of tweets are neutral, accounting for 54%, 55%, and 43% in the US, UK, and RSA, respectively. RSA shows the highest ratio of negative tweets at 40%, while Nigeria exhibits the highest ratio of positive tweets, constituting 80% of total tweets from Nigeria. Figure 5 displays the word frequency in the dataset through a word cloud.

B. MACHINE LEARNING MODELS RESULTS USING BoW AND TF-IDF

Results for machine learning models using Bag-of-Words (BoW) and TF-IDF features are presented in Table 8. The performance of linear models is notably better, with SVM achieving the highest accuracy of 0.97 and LR attaining a score of 0.96. RF, employing an ensemble architecture with 300 Decision Trees under majority voting criteria, also exhibits good accuracy at 0.91. KNN, as a lazy learner, performs poorly, which aligns with expectations for larger datasets. The results emphasize the effectiveness of SVM and LR models, particularly in handling the large TF-IDF feature set of 125,461.

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] A. Perrin and M. Anderson. (2018). Share of U.S. Adults Using Social Media, Including Facebook, is Mostly Unchanged Since 2018. [Online]. Available: https://www.pewresearch.org/fact-tank/2019/04/10/share-ofu-s-adults-using-social-media-including-facebook-is-mostly-unchangedsince-2018/

[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] 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.

[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.

[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.

[16] 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.

[17] Z. Papacharissi, ''Affective publics and structures of storytelling: Sentiment, events and mediality,'' Inf., Commun. Soc., vol. 19, no. 3, pp. 307–324, Mar. 2016.

[18] G. Bouvier, ''How journalists source trending social media feeds: A critical discourse perspective on Twitter,'' Journalism Stud., vol. 20, no. 2, pp. 212–231, Jan. 2019.

[19] M. KhosraviNik, ''Social media critical discourse studies (SM-CDS),'' in The Routledge Handbook of Critical Discourse Studies. London, U.K.: Routledge, 2017, pp. 582–596.

[20] T. T. Nguyen, S. Criss, A. M. Allen, M. M. Glymour, L. Phan, R. Trevino, S. Dasari, and Q. C. Nguyen, ''Pride, love, and Twitter rants: Combining machine learning and qualitative techniques to understand what our tweets reveal about race in the US,'' Int. J. Environ. Res. Public Health, vol. 16, no. 10, p. 1766, May 2019.