EXPLORING RACISM IN TWITTER DATA THROUGH STANFORD NLP AND TEXT FEATURE ANALYSIS

Victoria Anand Mary A [a], ArumaiSelvam M [a *], Muthukumaran S [a]

PG & Research Department of Computer Science,

St. Joseph's College of Arts & Science (Autonomous), Cuddalore-607001, Tamilnadu, India.

Corresponding Author:
M.ArumaiSelvam

*PG & Research Department of Computer Science,

St. Joseph's College of Arts & Science (Autonomous),

Cuddalore-607001, Tamilnadu, India.

*E-mail address: arumai_selvam@yahoo.com

Abstract—Racism is being used more frequently on internet platforms as a result of the growth of social media. Identifying and removing this kind of content is essential to preserve a welcoming and secure online community. In this study, we describe a method that combines Stanford Natural Language Processing (NLP) with Bag of Words (BoW) representation to identify racism in a Twitter dataset. The suggested system represents the text data as a set of features using BoW and performs pre-processing and feature extraction using the Stanford NLP method. Next, using the BoW representation as training data, a machine learning model is trained to categorise tweets as racist or non-racist. Sentiment NLP scores are used to assess the performance of the proposed system, and its results are contrasted with those of an existing system that extracts features using simply Stanford NLP. The findings demonstrate that the suggested system performs better in accuracy and precision than the current system, suggesting that adding BoW can enhance the system's capacity to identify racism in social media. To stop racism from spreading online, the suggested technique may be implemented on other social media sites and in offline environments.

Keywords— Stanford NLP, Text classification, Text pre-processing, Text Feature, Sentiment Analysis

1. INTRODUCTION

Racism has always been an issue in society, and the growing use of social media has made it more prevalent. Although social media sites like Twitter give users a forum to voice their thoughts and beliefs, they also serve as a haven for racist and hateful information. Automated systems that can identify and report racist content on social media are becoming more and more necessary to solve this problem. In addition to making social media sites safer and more welcoming to all users, these methods can also aid in preventing the spread of racist ideas and viewpoints. Numerous methods are currently in place to identify racist content on social media, but many of them have limitations about their precision, scalability, or adaptability[1]. This study suggests a new method for identifying racism in a Twitter dataset by utilising Stanford NLP and a Bag of Words technique to overcome these drawbacks. With its high accuracy, scalability, and customisation goals, this system will contribute to the advancement of the field's understanding of identifying racist content on social media. Part of speech tagging, named entity identification and sentiment analysis are a few popular NLP approaches utilised in these systems[2]. These methods seek to identify characteristics in the text, such as the tone, the entities mentioned, and the vocabulary, that are pertinent to the identification of racism. Based on the features that were collected, machine learning algorithms—such as neural networks, decision trees, and support vector machines—are then used to classify the text as racist or non-racist. To further assist in identifying racist content, some systems may include manually created guidelines or dictionaries of offensive terms or phrases. Although it has been demonstrated that these current methods are successful in identifying racist content on social media, they may also produce false positives [3] or false negatives, particularly in cases when the training data is not representative of the target dataset [4]. Furthermore, simple systems could not be as accurate or scalable as more sophisticated ones, nor easily adjustable to diverse use cases or datasets.

2. LITERATURE SURVEY

In this Section, there are two types of machine learning approaches: supervised and unsupervised. A greater number of labelled training materials are used in the supervised method. When it takes a long time to locate these labelled training documents, unsupervised techniques are used. The next strategy is the Lexicon-based approach, which focuses on determining the opinion lexicon used to analyse the text. There are two approaches to it: corpus-based and dictionary-based. The dictionary-based strategy finds the dictionary of meanings and opposites by first locating opinion-seed words. The corpus-based approach looks for additional sentiment words with context-specific orientations after starting with a sees list of emotion terms. Opinion or sentiment analysis (SA) Mining (OM) is the practice of using text mining, natural language processing (NLP), and computational methods to automatically extract or classify sentiments from user evaluations or opinions. The terms SA and OM have the same meaning and can be used interchangeably. However, some research indicates that these two terms differ by a little amount [1].

NLP techniques are sometimes used in conjunction with lexicon-based techniques to detect semantic linkages in addition to syntactical structure [5]. NLP techniques are used before the suggested lexicon-based SA method [6] at the preprocessing stage. It consists of a sentiment analysis module that evaluates user thoughts regarding news and an automated focus detection module. [7] used NLP in a variety of contexts. It combines mining and ranking approaches with natural language processing (NLP) to identify tense and time expressions. It has two parameters that use natural language processing (NLP) to extract time expressions and linguistic cues from crawled data. It takes advantage of Twitter reviews, and the outcome demonstrates how beneficial the criteria. To address the problem of recognising opinion words with context-specific orientations, a corpus-based technique is employed. In order to find other words in a huge corpus, it is based on patterns that emerge with a seed list of opinion words. This method, which starts with a list of seed opinion adjectives and is used with a set of linguistic restrictions to find more objective opinion terms, is employed on [15].

3. PROBLEM DEFINITION

Building on these previous systems, the suggested system uses Stanford NLP and a Bag of Words methodology to identify racism in a Twitter dataset. It does this by emphasising two crucial methods: Stanford NLP: Each tweet is processed to extract pertinent attributes using the Stanford NLP package. This makes it possible for the system to precisely recognise the sentiment, entities, and other elements of every tweet that are pertinent to the detection of racism. Bag of Words[13]. This method simplifies the process of classifying tweets using machine learning algorithms by converting each tweet into a numerical representation. This method works well for various NLP tasks, such as text classification and sentiment analysis. It is straightforward and efficient. Because of its easy customisation and scalability, the suggested solution can be tailored to various datasets and use situations. Additionally, it is anticipated to be extremely accurate and successful in identifying racist content on social media by utilising proven NLP and machine learning approaches

4. METHODOLOGY

The algorithm can properly identify the sentiment, entities, and other parts of the text that are relevant for identifying racism by extracting relevant features from each tweet using the Stanford NLP package. This raises the system's overall accuracy by lowering false positives and false negatives. Scaling the system to big datasets is made simple and efficient by the Bag of Words technique[8]. This is critical because social media sites like Twitter are seeing a sharp increase in the number of content posted, and the system must be able to manage massive volumes of data.

Because of its easy customisation features, the suggested solution may be tailored to various datasets and use situations. This is significant because different cultural, social, and historical contexts can define what constitutes racist content differently, and the system must be flexible enough to accommodate these variations. Because of the computational efficiency of the Bag of Words technique, the system can handle enormous volumes of data quickly[5]. This is significant because stopping the spread and lessening the effects of racist information need real-time identification[6]. The suggested system advances the state of the art for identifying racist content in social media by building on well-established NLP and machine learning techniques. The system is a good option for additional development and implementation because to its high accuracy, scalability, and customisation.

4.1 INPUT DESIGN

The interface between the user and the information system is the input design. It includes creating specifications and data preparation procedures, which are required to transform transaction data into a format that can be processed. This can be done by having people key data into the system directly, or by having the computer read data from a written or printed document[9]. Controlling the quantity of input needed, reducing errors, preventing delays, eliminating unnecessary steps, and simplifying the process are the main goals of input design. The input is made in a method that maintains privacy while offering security and usability. The following factors were taken into account in input design:◦ What kind of data should be provided as input? How should the data be organised or coded? The dialogue directs the operations staff in offering feedback. Techniques for organising input validations and actions to do in the event of an error.

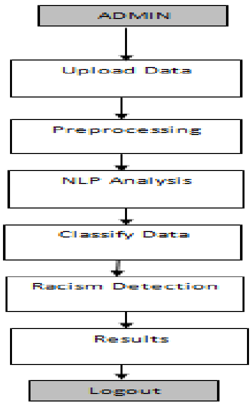

Figure1: **Data Input Design**

4.2 DATA COLLECTION

It is in charge of compiling a sizable corpus of tweets for analysis. To gather this information, you can utilise Twitter's APIs or those from other sources.

4.3 TEXT PROCESSING:

We process every tweet using the Stanford NLP library to extract features like named entities, sentiment, and part-of-speech tags[7]. You could clean and prepare the data for the next phase by doing this.Because of its easy customisation features, the suggested solution may be tailored to various datasets and use situations. This is significant because different cultural, social, and historical contexts can define what constitutes racist content differently, and the system must be flexible enough to accommodate these variations. Because of the computational efficiency of the Bag of Words technique, the system can handle enormous volumes of data quickly[5]. This is significant because stopping the spread and lessening the effects of racist information need real-time identification[6]. The suggested system advances the state of the art for identifying racist content in social media by building on well-established NLP and machine learning techniques. The system is a good option for additional development and implementation because to its high accuracy, scalability, and customisation.

We process every tweet using the Stanford NLP library to extract features like named entities, sentiment, and part-of-speech tags[7]. You could clean and prepare the data for the next phase by doing this.

4.4 STOP WORDS EXCLUSION

As a pre-processing step in natural language processing, stop word exclusion refers to the elimination of some often-used words that are unlikely to give any useful information for the job at hand[12]. These words, which are sometimes known as "stop words," include "the," "an," "and," and so on. Stop-word exclusion can be used to decrease the dimensionality of the data and increase the system's performance when it comes to identifying racist content on social media. The system may concentrate on the most pertinent terms in each tweet and avoid processing words[10] that are unlikely to help detect racism by eliminating stop words .In general, stop words exclusion is a helpful method for enhancing the efficacy and precision of algorithms designed to identify racist content on social media, and it is a crucial stage in the data pre-processing process.

4.5 BAG OF WORDS

NLP uses the Bag of Words (BoW) to represent textual data. This method of converting text data into a numerical representation that machine learning algorithms can use is easy to use and effective. BoW's fundamental concept is to treat any text document—a tweet in this example, for example—as a collection of word occurrences, or "bags," and to ignore word order and syntax.

Every distinct word in the text is considered a discrete feature, and its presence or absence is indicated by a binary value (1 for a word's presence and 0 for its absence) [20]. A sparse matrix is the final representation, with each row denoting a document and each column a distinct word from the corpus[11]. BoW is an excellent choice for handling huge datasets because it is easy to implement and provides good computational efficiency.

It is also very adaptable, enabling the insertion or deletion of features according to the outcomes of feature selection techniques or domain-specific information. BoW can be used to represent each tweet as a set of word occurrences in the context of detecting racist material in a Twitter dataset[19].

This allows for the training of a machine learning algorithm to classify the tweets as racist or non-racist based on their word occurrences [16]. The final model can then be utilised to determine the most crucial terms for identifying racism in the data and to forecast the class label of fresh tweets.

4.6 Identifying Racism in The Data

```
public double tf(List&lt;String&gt; doc, String term)
        {
        double result = 0;
        for (String word : doc)
        {
        if (term.equalsIgnoreCase(word))
        result++;
        }
        return result / doc.size();
        }
        public double tfIdf(List&lt;String&gt; doc, List&lt;List&lt;String&gt;&gt; docs, String term)
        {
        return tf(doc, term) * idf(docs, term);

        List<String> racedetect = new ArrayList<String>();

          String line;
            while ((line = reader.readLine()) != null) {
          racedetect.add(line.toUpperCase());
        }
         reader.close();

        for (String word : words) {
        String wordCompare = word.toUpperCase();
        if (racedetect.contains(wordCompare)) {
        wordsList.add(word);
                    }
                }
                StringBuilder sb = new StringBuilder();
                for (String str : wordsList) {
                    System.out.println(str + "");
                    sb.append(str + " ");
                }
```

Table Name:tweetdata
Primary Key: id
Purpose: Feteched dataset for processing

Field Name	Data Type
id	int
tweet	longtext

Table Name:dataclean
Primary Key: id
Purpose: Data Cleaning Process

Field Name	Data Type
id	int
tweet	longtext

Table 1: Data Pre-processing & Data Cleaning

Table Name:senanalysisdetections

Primary Key: id

Purpose:NLP analysis result

Field Name	Data Type
id	int
tweet	longtext
getSentimentScore	varchar
getSentimentType	varchar
getVeryPositive	varchar
getPositive	varchar
getNeutral	varchar
getNegative	varchar
getVeryNegative	varchar
label	varchar
annotation	varchar

Table 2: SA Detection

Table Name:tweetcheck

Primary Key: id

Purpose:NLP analysis individual text

Field Name	Data Type
id	int
tweet	longtext
getSentimentScore	varchar
getSentimentType	varchar
getVeryPositive	varchar
getPositive	varchar
getNeutral	varchar
getNegative	varchar
getVeryNegative	varchar
label	varchar
annotation	varchar

Table 3: Tweetcheck

5. RESULT AND DISCUSSION

An approach to rating the effectiveness of NLP models is sentiment NLP score evaluation. Sentiment NLP scores are a useful tool for assessing a model's performance when it comes to accurately identifying the sentiment of tweets, particularly when it comes to identifying racism in a Twitter dataset. Sentiment lexicons, which are sets of words or phrases labelled with sentiment labels like positive, negative, or neutral, are usually the basis for sentiment NLP scores. These lexicons are used by sentiment NLP algorithms to assess the sentiment of textual input and provide a sentiment score for every document.

Figure 2: Racism on Twitter Data with Standford NLP and BOW

In order to create and use NLP models for identifying racism in social media, sentiment NLP score evaluation is an essential first step that guarantees the model's accuracy, dependability, and suitability for the given purpose.

Figure3: Racism Detection

Finally, it is scalable and easily configurable, allowing it to adapt to different datasets and use cases. In addition, using established NLP and machine learning techniques, it is expected to be highly accurate and effective in detecting racist content on social media.

6. CONCLUSION

In this Paper, we suggested the approaches to combine machine learning and natural language processing to identify and address racism in social media. The pre-processing module of the system uses the Stanford NLP algorithm for text cleaning, tagging, and feature extraction. Its machine learning module uses the Bag of Words (BoW) representation to classify tweets as either racist or non-racist. The system is trained and tested on the Twitter dataset. The results of the evaluations show that the BoW representation can significantly improve the efficiency of the system in detecting racism in social media. The proposed system can be used as a tool to promote online inclusion and combat the spread of hate speech, racism and discrimination on various social media platforms. Overall, this project is an important contribution to social media analysis and highlights the need for effective tools to identify and combat racism online. By promoting inclusion and safety online, the proposed system can help create a more positive and accepting social media environment, which is essential to building a fairer and more just society.

7. REFERENCES

1. Tsytsarau Mikalai, Palpanas Themis. Survey on mining subjective data on the web. Data Min Knowl Discov 2012;24:478–514.
2. Thomas, B., 2013. What Consumers Think About Brands on Social Media, and What Bunesses Need to do About it Report, Keep Social Honest.
3. Ainur, Y., Yisong, Y., Claire, C., 2010. Multi-level structured models for document-level sentiment classification. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. MIT, Massachusetts, Association for Computational Linguistics, USA, pp. 1046–1056.
4. Noura, F., Elie, C., Rawad, A.A., Hazem, H., 2010. Sentence-level and document-level sentiment mining for arabic texts. In: Proceeding IEEE International Conference on Data Mining Workshops.
5. Bolshakov Igor A, Gelbukh Alexander. Comput Linguis (Models, Resources, Applications) 2004.
6. Moreo A, Romero M, Castro JL, Zurita JM. Lexicon-based comments-oriented news sentiment analyzer system. Expert Syst Appl 2012;39:9166–80.
7. Min Hye-Jin, Park Jong C. Identifying helpful reviews based on customers' mentions about experiences. Expert Syst Appl 2012;39:11830–8.
8. Kontopoulos Efstratios, Berberidis Christos, Dergiades Theologos, Bassiliades Nick. Ontology-based sentiment analysis of twitter posts
9. Michael Hagenau, Michael Liebmann, Dirk Neumann. Automated news reading: stock price prediction based on financial news using context-capturing features. Decis Supp Syst; 2013.
10. Miller G, Beckwith R, Fellbaum C, Gross D, Miller K. WordNet: an on-line lexical database. Oxford Univ. Press; 1990.
11. Gupta Sunil Kumar, Phung Dinh, Adams Brett, Venkatesh Svetha. Regularized nonnegative shared subspace learning. Data Min Knowl Discov 2012; 26:57–97.

12. Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends Inform Retriev 2008;2:1–135.
13. Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. IEEE Intell Syst 2013;28:15–21.
14. Feldman R. Techniques and applications for sentiment analysis. Commun ACM 2013; 56:82–9.
15. Turney P. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of annual meeting of the Association for Computational Linguistics (ACL'02); 2002.
16. Hu Nan, Bose Indranil, Koh Noi Sian, Liu Ling. ''Manipulation of online reviews: an analysis of ratings, readability, and sentiments''. Decis Support Syst 2012;52:674–84.
17. Deerwester S, Dumais S, Landauer T, Furnas G, Harshman R. Indexing by latent semantic analysis. JASIS 1990;41:391–407.
18. Li Sheng-Tun, Tsai Fu-Ching. A fuzzy conceptualization model for text mining with application in opinion polarity classification. Knowl-Based Syst 2013;39:23–33.
19. Kontopoulos Efstratios, Berberidis Christos, Dergiades Theologos, Bassiliades Nick. Ontology-based sentiment analysis of twitter posts. Expert Syst Appl 2013.
20. Mudinas Andrius, Zhang Dell, Levene Mark. Combining lexicon and learning based approaches for concept-level sentiment analysis. Presented at the WISDOM'12, Beijing, China; 2012.

****