

Identification and Synthesis of Unstructured Data of Violence on Facebook Using a Mood and Emotion Analysis System

NIROOP, SRAVANTHI KALAL

AMC ENGINEERING COLLEGE

Abstract

Hate speech is a type of expression that targets an individual or a group based on their race, ethnicity, religion, sexual orientation, or other characteristics. Although it may be conveyed in a variety of ways, both online and offline, the growing popularity of social media has expanded both its usage and intensity significantly. As a result, the goal of this study is to find and analyse unstructured data from chosen social media posts that attempt to promote hatred in the comment sections. To address this problem, we offer FADOHS, a novel framework that combines data analysis and natural language processing methodologies to alert all social media providers to the prevalence of hatred on social media. To be more specific, we employ sentiment and emotion analysis algorithms to analyse recent postings and comments. Posts suspected of having dehumanising language will be screened before being submitted into the clustering algorithm for further analysis. According to the experimental results, the suggested FADOHS framework outperforms the state-of-the-art technique by around 10%. Social media allows users to freely express themselves online; yet, when individuals post harsh and opinionated comments that target specific persons or communities, this may lead to animosity against them. Because of the widespread criticism of obesity (fatness), a great deal of fat-shaming material has been made available online. The project "Framework for Detection and Integration of Unstructured Data of Hate Speech on Facebook Using Sentiment and Emotion Analysis" intends to create a software framework in Java for detecting and integrating hate speech on Facebook.

Key words: FADHOS, Speech, Hate Speech, Facebook, Social Network

I. INTRODUCTION

Hate speech is a major issue on social media, with potentially serious ramifications for individuals and society as a whole. The propagation of hostile and discriminatory messages has been assisted by the advent of social media, which may foster prejudice, intolerance, and even violence. To solve this problem, we suggest the FADOHS framework, which stands for "Framework for Automatic Detection of Hate Speech." Through sentiment and emotion analysis, FADOHS hopes to detect hate speech on Facebook sites and combine unstructured data. The system suggested is divided into four stages: data collection and preprocessing, sentiment and emotion analysis, hate speech grouping, and assessment. The technology takes data from Facebook sites in the first step and preprocesses it to eliminate noise and useless information. The algorithm then does sentiment and emotion analysis in the second step to comprehend the text's underlying emotions and sentiments.

The technique groups the text in the third stage based on the degree of hate speech expressed. Finally, in the fourth step, the system assesses the efficiency of various data analysis and natural language processing methodologies. The system offers two sorts of applications: those that use a dataset and those that allow users to provide comments. The application's administration section Overall, the FADOHS system offers a powerful tool for detecting and addressing hate speech on social media platforms. By identifying the most significant factors from the unstructured data, the system can effectively cluster information into different groups according to the degree of hatred being expressed. The system has the potential to make a significant contribution towards creating a more respectful and inclusive online community.

II. LITERATURE SURVEY

This article maps and explores current advancements in the study of racism and hate speech in the area of social media research, building on Jessie Daniels' 2013[1] assessment of scholarly work on race and racism online. Systematically examining 104 articles, We respond to three research queries: What geographic contexts, media outlets, and study techniques do scholars use when researching racism and hate speech on social media? How much does academic research examine how systematic racism is (re)produced on social media using critical racial perspectives? What are the main methodological and ethical issues facing the discipline? In order to understand racism on social media, the paper concludes that there is a lack of regional and platform diversity, a lack of researchers engaging in reflective dialogue with the subject of their study, and minimal engagement

with critical racial perspectives. There is a need for deeper examinations of the interactions between platform politics and user behavior that shape modern racism. [1]

Social networking sites facilitate contact and information sharing, but they are often used to start damaging campaigns against particular people and organizations. Some of the serious consequences of large-scale internet offensives include cyberbullying, encouragement of self-harm behaviors, and sexual predation. Additionally, attacks against groups of victims are possible, and they have the potential to escalate into physical violence. One of the strongest and most well-liked data mining techniques in the scientific community is the k-means algorithm. Nevertheless, despite its popularity, the technique has certain drawbacks, such as issues with random initialization of the centroids that 2 random initialization of the centroids which leads to a surprising convergence. Additionally, the number of clusters needed for this type of clustering algorithm must be predetermined, which accounts for the various cluster forms and outlier effects. The k-means algorithm’s inability to accommodate different data formats is a basic issue. Research on the k-means algorithm has been conducted to overcome these problems, and the results are presented in this work. Verifying the accuracy[4] of news and spotting social bots are currently social media platforms’ two top priorities, but there aren’t many gold-standard data sources accessible right now. This study represents a major step forward in the development of a substantial, feature-rich gold-standard dataset. During September 2016, nine news organizations collectively posted a compilation of news articles to Facebook. [2]

Each term in a short[5] text can potentially convey emotional meaning. Facebook comments and shared posts often convey human biases, which play a pivotal role in information spreading and content consumption. Such bias is at the basis of human-generated content, and capable of conveying contexts which shape the opinion of users through the social media flow of information. Starting from the observation that a separation in topic clusters, i.e. sub-contexts, spontaneously occur if evaluated by human common sense, this work introduces a process for automated extraction of sub-context in Facebook. Basing on emotional abstraction and valence, the automated extraction is exploited through a class of path-based semantic similarity measures and sentiment analysis. [3]

A Mobile-Based Application for Animal

III. EXISTING SYSTEM:

The authors of “Hate Me, Hate Me Not: Hate Speech on Facebook” have proposed several classification methods to distinguish among different types of hate speech. More specifically, they leverage morpho-syntactic features, sentiment polarity, and Italian has two classifiers that were created and implemented using word-embedded lexicons. Long short-term memory (LSTM) networks and support vector machines (SVMs) are both used in their framework. This study was premised on the concept outlined in Del Vigna et al.’s study and our understanding of hate speech.

In the existing system first they recognize a set of pages from American-based websites, known to discuss controversial topics such as immigration, race, and religion. We use these “seeds”, or Facebook IDs, to crawl the Facebook graphs and construct a small network using the “follow” relationship. Leveraging graph analysis techniques, we identify the most influential pages spreading hate speech and crawl their latest posts and comments. We then apply sentiment and emotion analysis algorithms to recognize posts with highly negative tones, specifically those suspected of instigating hatred. Finally, we convert each post into a document by concatenating all comments, and using the K-means algorithm to create clusters of posts based on the topics they discuss

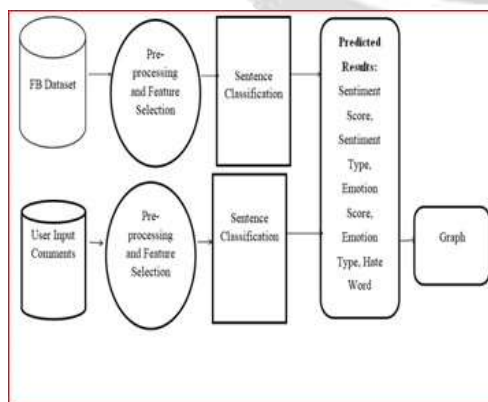


Fig 1: Architecture diagram

IV. PROPOSED SYSTEM:

In this proposed system we aim to create a framework (FADOHS) that can first detect hate speech on Facebook and then integrate unstructured data through clustering using sentiment and emotion analysis. It identifies the most significant factors from the unstructured data of posts and comments on Facebook pages that allegedly promote hate speech. By applying the four stages of FADOHS, we will perform experiments to assess the effectiveness of various data analysis and natural language processing methodologies. Ultimately, we can demonstrate the strong effectiveness of hate-speech clustering using hybrid methods of data analysis and natural language processing to identify and categorize information into different groups according to the degree of hatred being expressed. In our proposed system we develop 2 types of application where the first one use dataset and the second one using comments posted by the user. In the first type the facebook comments which are referred from kaggle website, are uploaded into the system. Then the preprocessing is done and each and every comment given in the dataset is processed one by one and the SentimentScore is calculated and the Sentiment type is predicted for the each dataset record. Additionally it computes the Emotion Score, Emotion Type (Happiness, Sadness, Anger, Fear, Disgust, Surprise), Hate Word and Label the dataset record completely. Finally a static Graph is plotted with the results which we received.

Last but not least, the suggested system, FADOHS, provides a thorough framework for identifying hate speech on Facebook and incorporating unstructured data using sentiment and emotion analysis. The algorithm can efficiently categorize information into several groups based on the level of expressed hatred by finding the most important factors from the unstructured input. The most serious instances of hate speech can therefore be given priority, and resources can be directed there. In order to ensure that the system can be used

in a number of circumstances and that it can be tailored to different user needs, the system offers two different types of applications: one that uses a dataset and one that allows users to add comments. Monitoring of sentiment analysis and other outcomes is possible in the application's admin section, which contributes towards creating a more respectful and inclusive online community. In future studies, we plan to further utilize our framework not only on comments but also their replies, in an attempt to accurately identify individuals who are suspected of promoting hate speech. Long-term benefits can be extremely valuable because this may be able to detect cyberbullies and cyberterrorists. We would also like to perform a more in-depth examination of the emotion filtering and clustering findings to identify the most dependable setup for optimizing outcomes.

V. PROPOSED SYSTEM:

We create our suggested system to accept the input dataset that is referred from Kaggle and contains user comments from Facebook in the first module. There are no labels in this dataset. Utilizing natural language processing techniques, it analyzes the comments to find linguistic patterns that are suggestive of hate speech. Once recognized, the module flags the comments as potentially containing hate speech.

Preprocessing Model: This module is responsible for cleaning and preparing the data for analysis. It removes irrelevant information and transforms the data into a format that can be easily analyzed by other modules.

Sentiment Analysis: This module analyzes the sentiment of Facebook posts and comments using natural language processing techniques. It computes a sentiment score for each post and comment, as well as predicts the type of sentiment. (positive, negative, or neutral).

REFERENCES

- [1] Zuckerberg Refugee Crisis: Hate Speech Has, Place Facebook, StreetGuardian, Honolulu, HI, USA, 2010
- [2] Fortune. (2018). Facebook Removed 2.5 Million Pieces Hate Speech 1st Quarter. Accessed: Jul. 16, 2018. [Online]. Available: <https://fortune.com/2018/05/15/facebook-hate-speech-removals>
- [3] ILGA. (2018). Hate Crime Hate Speech. Accessed: May 6, 2018. [On-line]. Available: <https://www.ilga-europe.org/what-we-do/our-advocacy-work/hate-crime-hate-speech>
- [4] Facebook. (2020). Community Standards Home. Accessed: May 11, 2018. [Online]. Available: <https://www.facebook.com/communitystandards/>.
- [5] CNBC. (2020). Facebook's Artificial Intelligence Still Has Trouble Finding Hate Speech But it Finds a Lot of Nudity. Accessed: May 11, 2018. [Online]. Available: <https://www.cnbc.com/2018/05/15/facebook-artificial-intelligence-still-finds-it-hard-to-identify-hate-speech.html>
- [6] [6] S. Chinnasamy and N. A. Manaf, "Social media as political hatred mode in Ts 2018 general election," in SHS Web Conf., vol. 53, 2018, p. 2005.
- [7] [7] A. Matamoros-Fernández and J. Farkas, "Racism, hate speech, and social media: A systematic review and critique,"

Telev. New Media, vol.22, no. 2, pp. 205224, Feb. 2021.

- [8] [8] F. Del Vigna, A. Cimino, F. Dell-TOrletta, M. Petrocchi, and M. Tesconi, "Hate me, hate me not: Hate speech detection on Facebook," in Proc. 1st Italian Conf. Cybersecur. (ITASEC), Venice, Italy, 2017, pp.8695..

