

RACISM DETECTION BY ANALYZING OPINIONS THROUGH SENTIMENT ANALYSIS OF TWEETS FROM DIFFERENT IMMIGRATION CRISIS.

Friday, April 1, 2022

OVERVIEW

1. Project Background and Description

Twitter has become a dominating socio-political platform allowing live news updates outpacing news outlets by allowing witnesses and participants to quickly share the events on the platform. Twitter also allows people around the world to voice their opinions on current crisis. Amongst these voices, misuse of the platform has seen emerging racism and discrimination in various forms. This ranges from memes to openly racist remarks that incite social instability and violence. As seen through the recent Ukraine immigration crisis, twitter users have used the platform to voice their opinions, but it has seen various racist tweets from individuals to media outlets. Online Social Networks (OSNs) have noticeably played a significant role in such crisis and have proven to be a reliable and fruitful source of data for analysis and studying political discourse. Social media being the leading source of racism opinions dissemination should be monitored and racism remarks should be detected and blocked timely. Sentiment analysis provides a powerful tool to mine such data to analyze emotions and opinions. Moreover, the different type of discussions on racism illustrates geographical variability in racial attitudes and sentiment. Therefore, analyzing the details of how people, events, and circumstances are represented provides insights on user communication and problems surrounding racism can be investigated.

Automatic hate speech detection using machine learning algorithms is still new and requires extensive research efforts from both industry and Academia. While social media platforms have enforced strict policies about integrity and hateful content, the problem remains a difficult one, as it involves several layers of complexity: computational

complexity, since the volume of the content is huge, along with the subtleties and cultural aspects of each language, the problem of low resource languages and the inherent ambiguity in natural language. The objective of this research is to perform sentiment analysis on tweets from both the Afghanistan and Ukraine crisis then implement various learning models to detect racism in tweets. My research will target tweets that contain hateful and racist speech aimed at refugees and migrants. Therefore, I will try to improve on existing models throughout this paper.

2. Dataset

A Dataset for the Russo-Ukrainian Crisis has been extracted by researchers since the 1st week of the crisis and is being updated daily and publicly available on GitHub (Haq et al., 2022). However, due to twitter guidelines the repository only contains twitter IDs, but the tweets can be extracted using open-source tools such as Twarc, Tweepy, or Hydrator. The Afghanistan dataset will be scraped similarly. Twitter's API provides access to twitter objects that contain "a long list of 'root-level' attributes, including fundamental attributes such as *id*, *created_at*, and *text*" (Twitter).

```
Index(['id', 'conversation_id', 'created_at', 'date', 'time', 'timezone',  
      'user_id', 'username', 'name', 'place', 'tweet', 'language', 'mentions',  
      'urls', 'photos', 'replies_count', 'retweets_count', 'likes_count',  
      'hashtags', 'cashtags', 'link', 'retweet', 'quote_url', 'video',  
      'thumbnail', 'near', 'geo', 'source', 'user_rt_id', 'user_rt',  
      'retweet_id', 'reply_to', 'retweet_date', 'translate', 'trans_src',  
      'trans_dest'],  
      dtype='object')
```

3. Method

According to previous research analysis could be done through exploring and combining both text and image modalities as opposed to the traditional text only (Goutsos). Specifically, users often use messages encoded in images to avoid NLP-based hate speech detection systems.

An ensemble model will be proposed that makes use of recurrent neural networks. For performance comparison, several well-known machine learning models are implemented using the optimized parameters such as decision tree (DT), random forest (RF), logistic regression (LR), k nearest neighbor (KNN), and support vector machines (SVM). Term frequency-inverse document frequency (TF-IDF) and bag of words (BoW) are studied as feature extraction techniques (Lee, E).

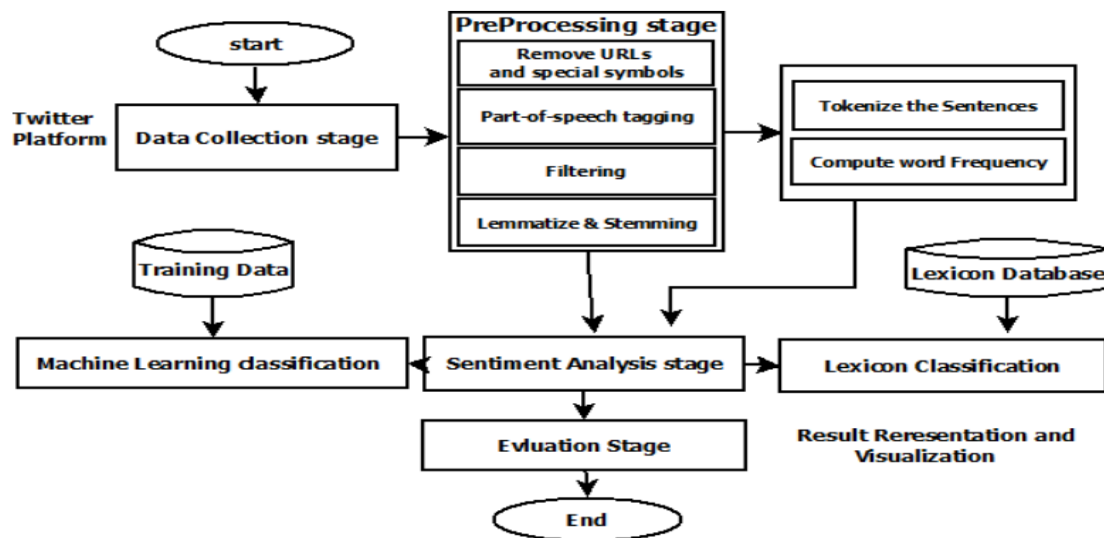
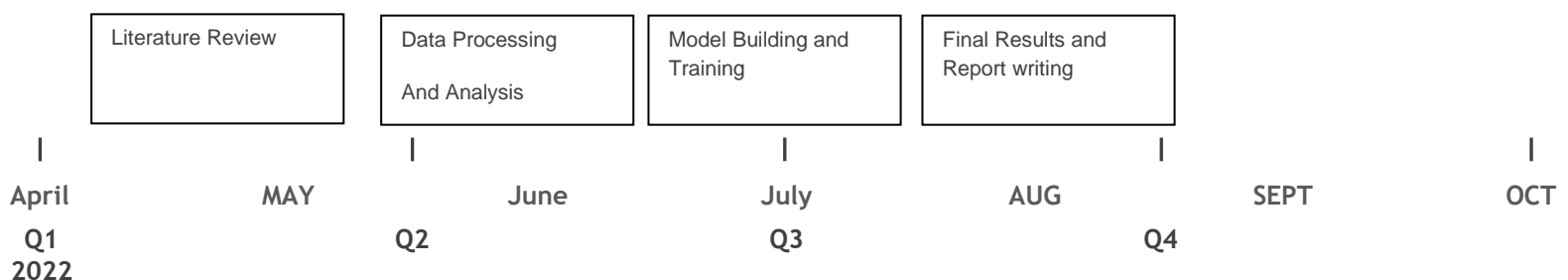


Figure 1. Approach follow diagram of using natural language preprocessing, text mining, and Sentiment Analysis techniques.

4. GANT Chart



Citations:

Twitter. (n.d.). *Tweet object / docs / twitter developer platform*. Twitter. Retrieved April 5, 2022, from <https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet>

Goutsos, D. (2021). Multimodal hate speech detection in greek social media. *Multimodal Technologies and Interaction*, 5(7), 34. doi:<http://dx.doi.org/10.3390/mti5070034>

Lee, E., Rustam, F., Washington, P. B., El Barakaz, F., Aljedaani, W., & Ashraf, I. (2022). Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets using Stacked Ensemble GCR-NN Model. *IEEE Access*.