

Analyzing Racist Tweets on Soccer Players

Ridhuvarshan Natarajan
Computer Science
Binghamton University
Binghamton, New York
rnatar2@binghamton.edu

Chethana Gopinatah
Computer Science
Binghamton University
Binghamton, New York
cgopina1@binghamton.edu

Aravind Padmanabhan
Computer Science
Binghamton University
Binghamton, New York
apadman1@binghamton.edu

ABSTRACT

Hate speech is all over the web, especially on platforms such as Twitter, Facebook, Instagram, whose goal is to propagate free speech. As users on the internet tend to be very forthcoming with, they want to express, there is a lot of hurtful and insensitive comments floating around. This kind of derogatory speech is prevalent when emotions are high especially during soccer games when one's favorite team loses.

Our major goal is to single out such hate speech, especially, racist comments focused on soccer players on Twitter.

We are interested in collecting the tweets about soccer players and soccer games and analyze if there have been any racism comments targeted on a particular player after their poor performance in a match.

Implementation

We are interested in collecting the tweets about soccer players and soccer games and analyze if there has been any racism comments targeted on a particular player after their poor performance in a match.

To analyse the tweets for racist content (on soccer players), we are planning to use the WEB APIs provided by Twitter. This can be achieved by creating a python application that periodically runs the APIs fetches the data and stores in the database. For data storage we have decided to use the MongoDB. For integration of the python application to the MongoDB a wrapper interface like PyMongo can be used. For the periodic job scheduling, a scheduler like CronJobs can be used.

The following API (standard) is for searching tweets posted in the last 7 days. If requirement arises, an advanced premium API can be used to search earlier tweets.
<https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets>

Though the mentioned search API allows search access to tweets upto a week old, the script to run it must be made more frequent as hateful/racist content is removed quickly by Twitter moderation. The frequency of running a script to call this API will be decided after running a small test and will be

based on parameters including but not limited to volume of tweets, change in volume/severity during weekends or post-match or during the match itself, location of the match, location of the account tweeting racist stuff etc.

There is another API provided by Twitter which is used to filter real time tweets.

<https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter>. Twitter also provides a sampling real time tweets API that can be used to get real time tweets in general without filter (<https://developer.twitter.com/en/docs/tweets/sample-realtime/api-reference/get-statuses-sample>). Based on the sample analysis of data fetched, the decision can be made on what API(s) can be used to fetch data from Twitter.

To identify what words are racist in the first place, we plan to get a list of the basic racist words in English. These words can be tweets can either be filtered in Twitter itself or in the python application. This depends upon the API we choose. As the project progresses, it can be enhanced by using the APIs of structured repositories like HateBase.

It should also be noted that not all tweets that contain racist words can be classified racist. For example, the n word if used by a black person in a casual conversation might not be racist [citation needed]. Other challenges include quote detection, sarcasm detection etc. These can be approached by using sentiment analysis tools.

REFERENCES

<https://developer.twitter.com/en/docs/tweets/search/overview/standard>
<https://hatebase.org/>
<https://github.com/hatebase/Hatebase-API-Docs>
<https://knightlab.northwestern.edu/2014/03/15/a-beginners-guide-to-collecting-twitter-data-and-a-bit-of-web-scraping/>