Project Proposal: Football World Cup matches Sentiment Analysis

Team Information:

Name: Football Fan
Member Susovan Samanta, susovan2@illinois.edu

Topic of this project is to create a sentiment analysis of Football World Cup 2022 matches using Twitter trends. I will use sentiment analysis on available tweets for a particular match to find out which players from either team had performed good or bad.

Based on this analysis of the sentiment of these players, we can understand which player is performing good or bad for a long period of time. This is interesting because with the result of this analysis we can understand if a player is in good form.

Main task for this project is to collect tweets from a completed match and analyze the tweets and categorize words to understand the sentiment.

Approach for this project would be to take a sample match which has already been completed in recent years and then using text retrieval and text mining to collect all the tweets related to that specific match. We also need to check if those tweets mentioned any player's name form either of the team. Once we gather all the related tweets, we can do sentiment analysis for each tweet to understand if that tweet is of positive sentiment or negative sentiment for a particular player mentioned in that tweet.

After completing sentiment analysis for all the tweets, we can then summarize the sentiment for a specific player. It will be simply the aggregate of all the sentiments collected from all the tweets with the player's name mentioned.

Expected outcome of this project is to generate a detail report for a given match. In this report it should show the top performer of that match and the players with poor performance in that match.

I will be using the following tools to develop my projects:
Python (Programming Language), Tweepy (Python library to access Twitter API), Text Blob (Python Library for processing textual data), NumPy(Python library used for working with arrays), BM25Okapi (Ranking Function)

I will evaluate my work by determining how accurate my project is by comparing my test results with sports expert reviews, sports journals analysis and sports channel's analysis.

I will be working alone in this project and below is my estimated workload and distribution

Initial Research	4 hours
Learn tweepy and other tools	4 hours
Text Retrieval, Cleanup data	4 hours
Implementation	16 hours

Testing, fixes, improvements	10 hours
Final Project Documentations and Demo	8 hours
Total	46 hours