Twitter Sentiment Analysis Surrounding Sports Events

Paul Kim, Abdullah Basulaib, Edwin G. Dos Santos, Amir Younesi

Motivation

The goal of our project is to analyze textual content of tweets that use hashtags related to sports events to map out average sentiment over the course of the event.

Data and Processing Dimensions

- Sports API to get event information
    - Match start/end, half start/end
    - Match key event (substitution, foul, goal, etc)
    - Home and Away team hashtags
    - Neutral hashtag (event hashtag)
- Twitter API (real-time streaming)
    - Gather (english) tweets in time intervals relative to sporting event
    - Variable length intervals
    - Three batches (one for each hashtag)
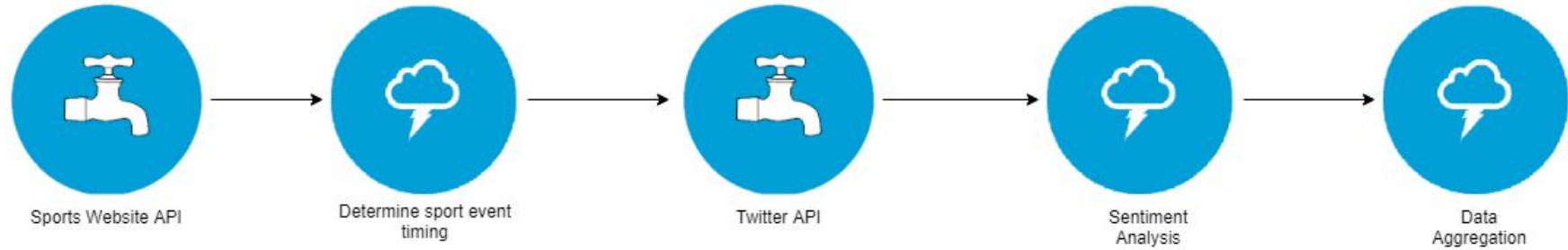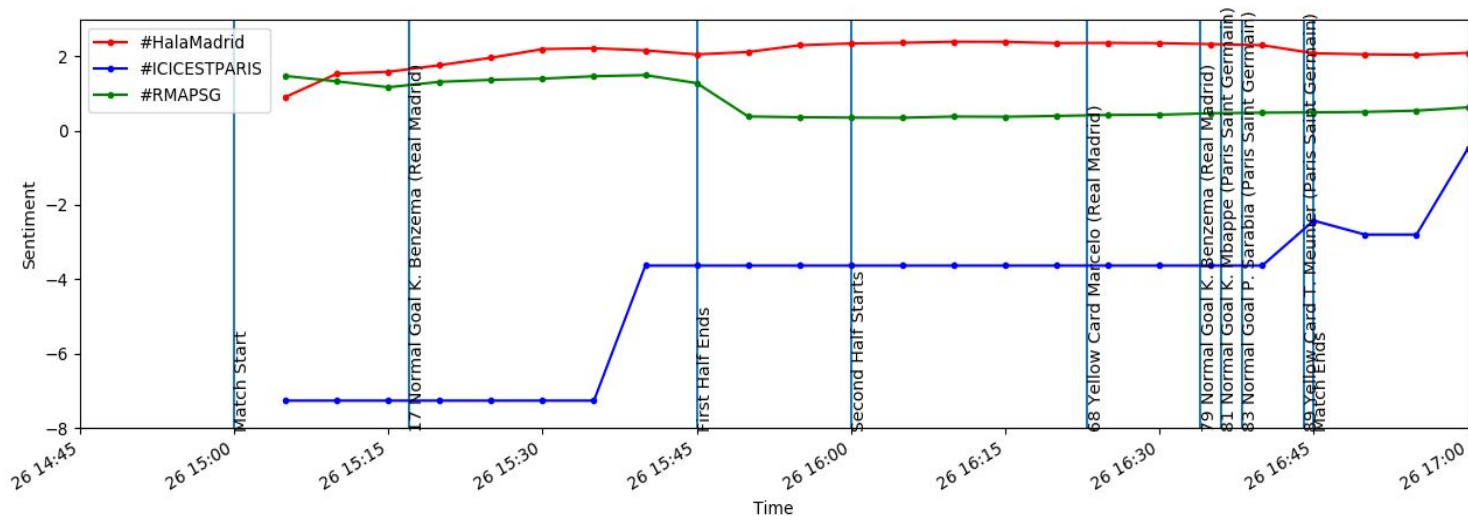
YORK U
UNIVERSITÉ
UNIVERSITY

Architecture

Fig. 1

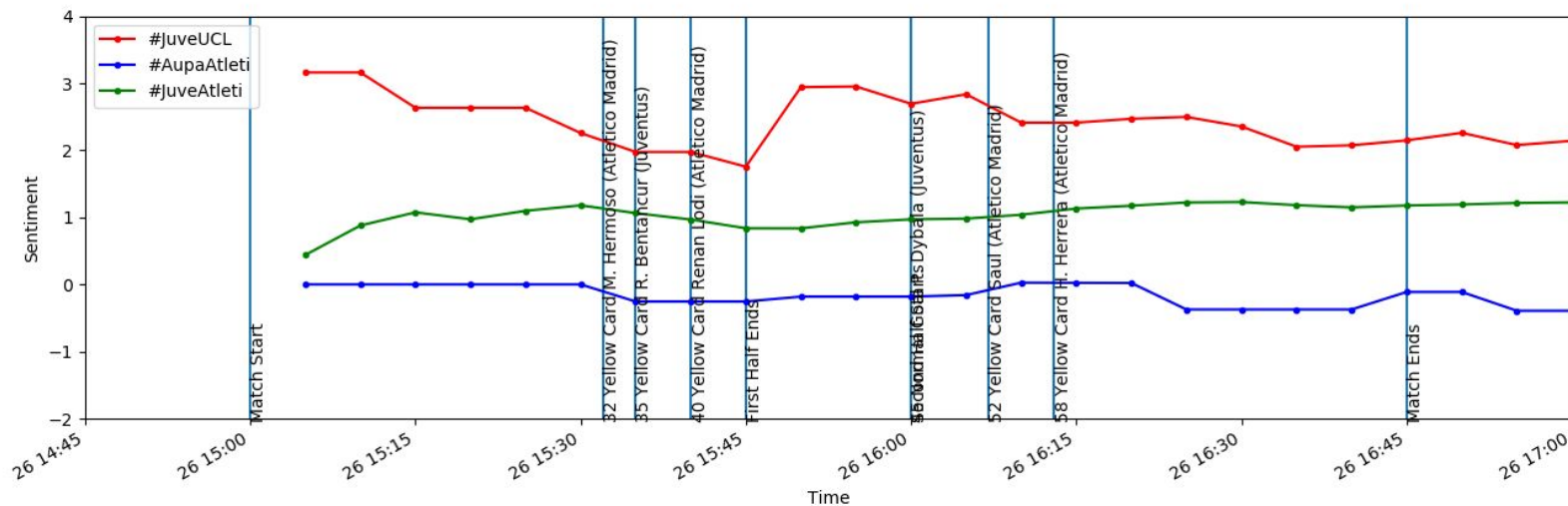Overall architecture and data flow

Analysis

- Questions:
  - After a key event, is there a spike in average public sentiment? Is it positive, negative or neutral?
  - Is there a trend of a particular event?
    - e.g. red card, average sentiment swap between teams?

- Objective
  - Understand how events change public sentiment over the course of a sporting event
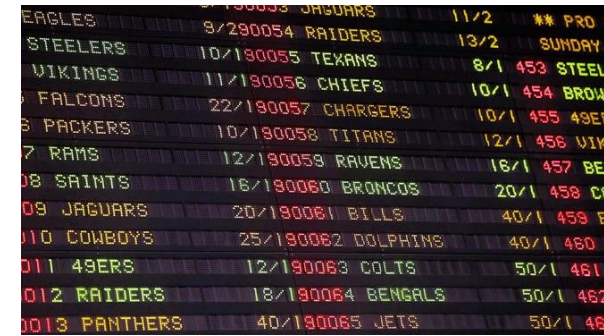
YORK U
UNIVERSITÉ
UNIVERSITY

Results: Real Madrid vs PSG (UCL)

Results: Juventus vs. Atletico Madrid (UCL)

Variations

- Real-time betting odds analysis
  - How do bets sway depending on tweets

- Tracking users of a political campaigns hashtag over the course of an election
  - Allows campaigns to better understand how significant events change sentiment
  - Compare sentiment to other campaigns

Limitations

- People use popular hashtags to gain more traction on tweets that might be unrelated to the event
- Data need to be cleaned further
- Not every user will tweet using the hashtag

YORK U
UNIVERSITÉ
UNIVERSITY
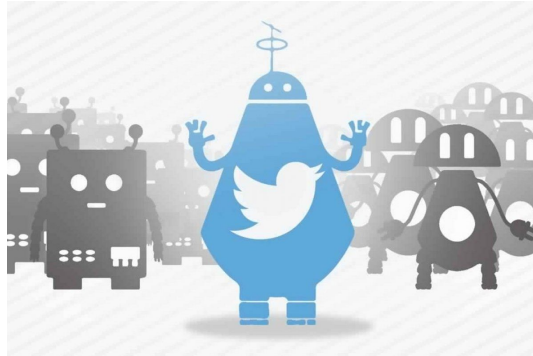
Limitations pt.2

- We can't analyze media (pictures, .gifs)
- Match events don't include controversial moments
- Text analysis does not account for sarcasm or inside jokes

Conclusion

- Overall, the goal is to understand how events influence public sentiment
- Can be applied to other sports as well (assuming appropriate API)
- Can be scaled horizontally
  - Cluster computing
    - Each cluster processes a league
    - Each node runs a game

Thank you