

A close-up, low-angle shot of a brown leather basketball with black lines, resting on a polished wooden basketball court floor. The lighting is dramatic, coming from the side, casting a long shadow of the ball across the floor. The background is dark and out of focus.

NBA Twitter Correlation

Andres Fandino

Josh Harkness

Mina Sobol

November 30, 2020

Introduction

- Introduction
- Preprocessing - Loading Data
- Preprocessing - Cleaning Data
- Data Transfer
- Analysis - Sentiment
- Analysis - Emotions
- Analysis - Embeddings
- Analysis - Combinations
- Experiments & Results
- Related Works & Way Forward
- Demonstration

Introduction: Off Court Blues

THE CUT

MAR. 23, 2017

Has Tinder Boosted the Performance of NBA Players?

By Jesse Singal

PRO BASKETBALL TALK | NBC SPORTS

Report: Heat suspended Dion Waiters due to Instagram boat photo while he was supposedly out sick

By Dan Feldman

Dec 15, 2019, 9:00 PM EST

Heavy.com > Sports > Soccer

Diego Maradona's Drug Problem: Cocaine Binges Between Matches Nearly Killed Him

Introduction: Platform

Kevin Durant ✓

@KDTrey5

IM ME, I DO ME, AND I CHILL. @35ventures @boardroom

📍 At the crib dog 🔗 apple.co/theetcs 📅 Joined April 2009

1,467 Following 18.6M Followers

SHAQ ✓

@SHAQ

VERY QUOTATIOUS, I PERFORM RANDOM ACTS OF SHAQNESS

📍 Orlando, FL 🔗 exercisethatvote.org/shaq 📅 Joined November 2008

690 Following 15.5M Followers

LeBron James ✓

@KingJames

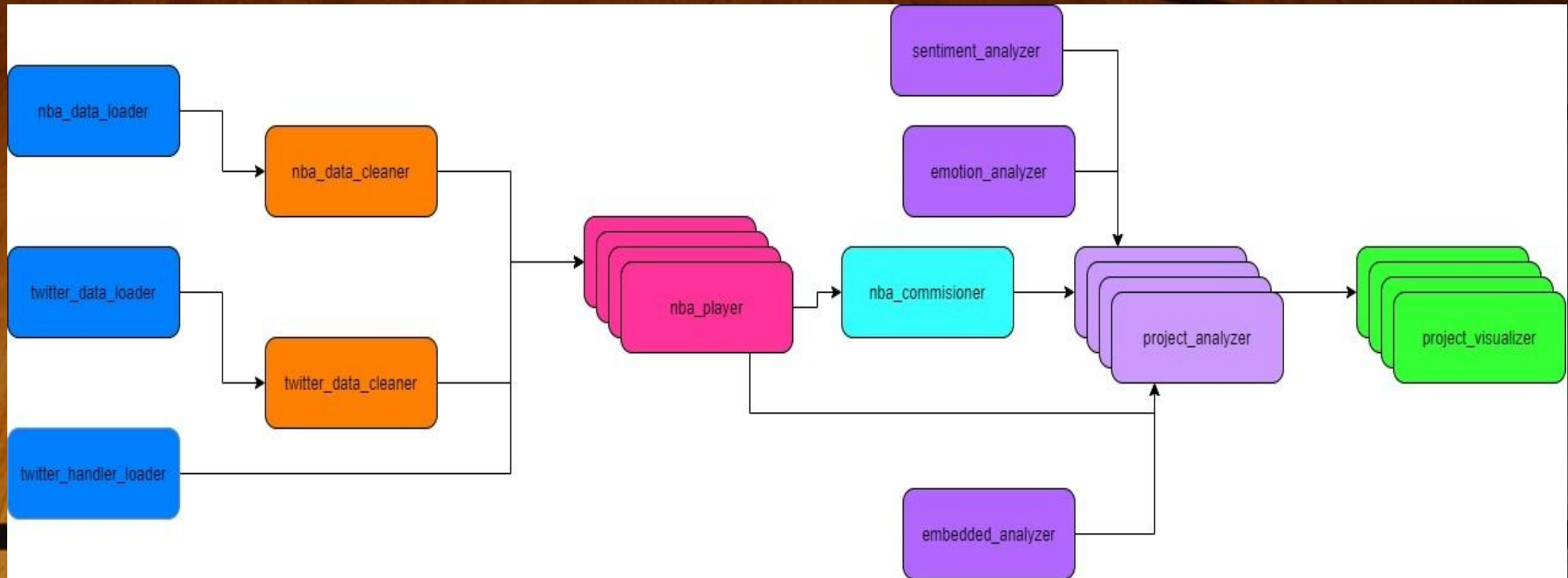
EST. AKRON - ST.V/M Class of '03 LeBronJamesFamilyFoundation.org #IPROMISE

📍 Amongst La Familia! 🔗 LeBronJames.com 📅 Joined March 2009

191 Following 48.4M Followers

Introduction: Analysis

- Collect text from twitter for top players (predictor variables)
- Collect regular season statistics from NBA website (target variable)
- Clean the data
- Extract features
- Use regression models to make predictions



Preprocessing - Loading Data

Twitter API

- Required Twitter Handles.
- All tweets within a season or within 7 days from game played.

NBA API

- Retrieved Seasonal Stats based on player.
- Only the last 5 seasons were collected.

Preprocessing - Cleaning Data

Twitter Data

- Removed URLs, mentions, hashtags, punctuation, newlines, stopping words, excess spacing, and emojis.
- Converted everything to lowercase.
- Lemmatized remaining words.
- Squashed Tweets from one player's single game into a single string .

NBA Statistical Data

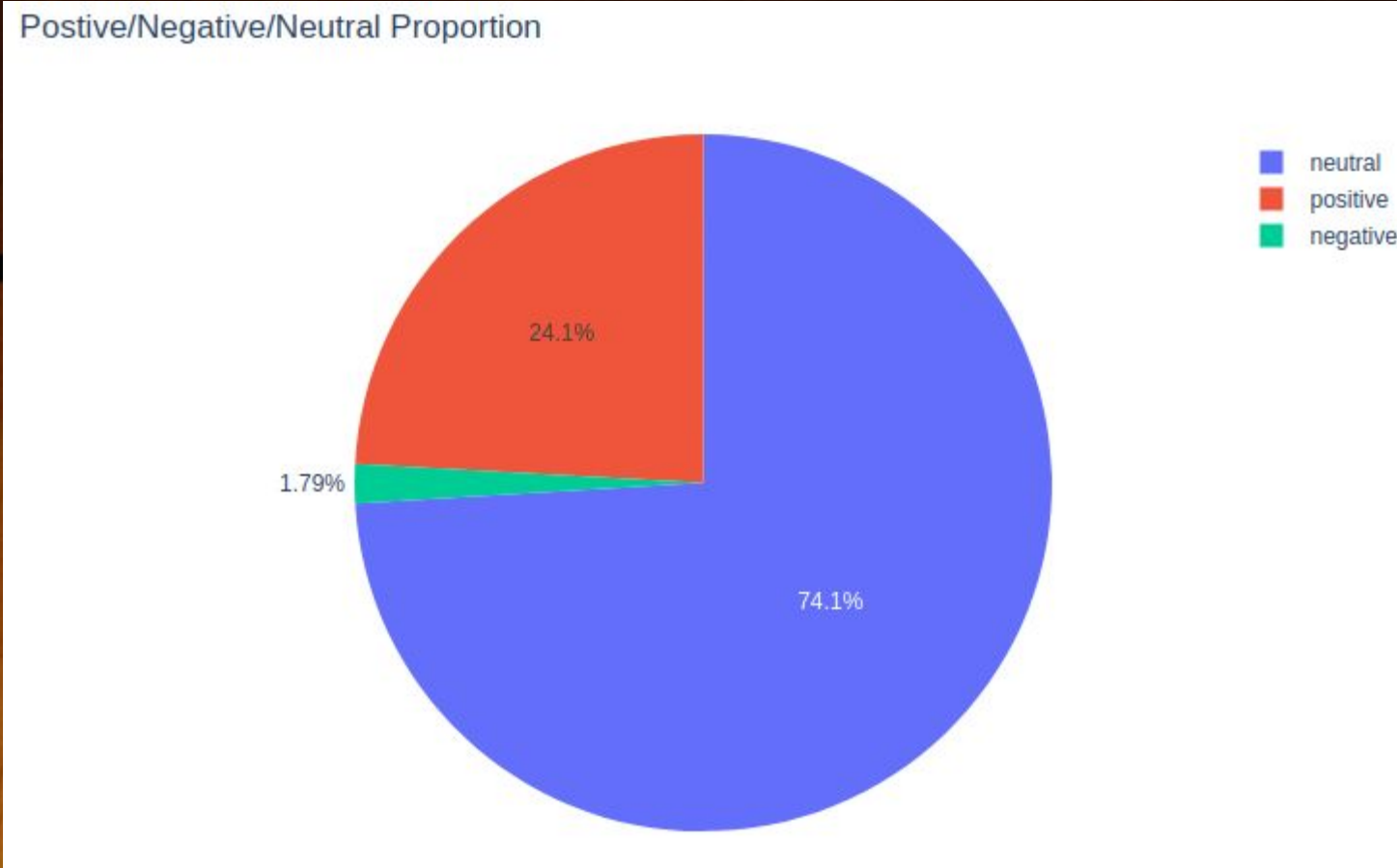
- Removed all stats other than points scored and attempts vs. made shots.
- Manually determined shot accuracy.
- Calculated +/- averages for each player to use as the Y variable.

Data Transfer

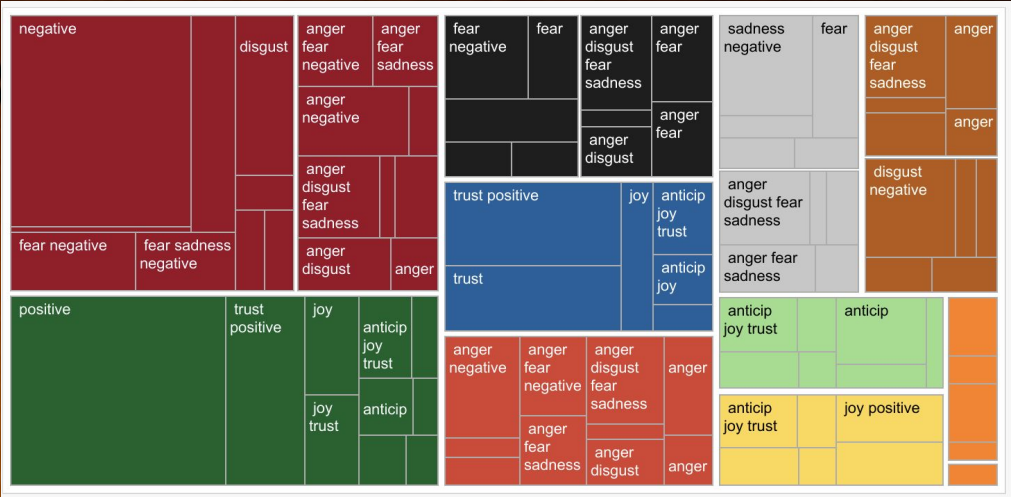
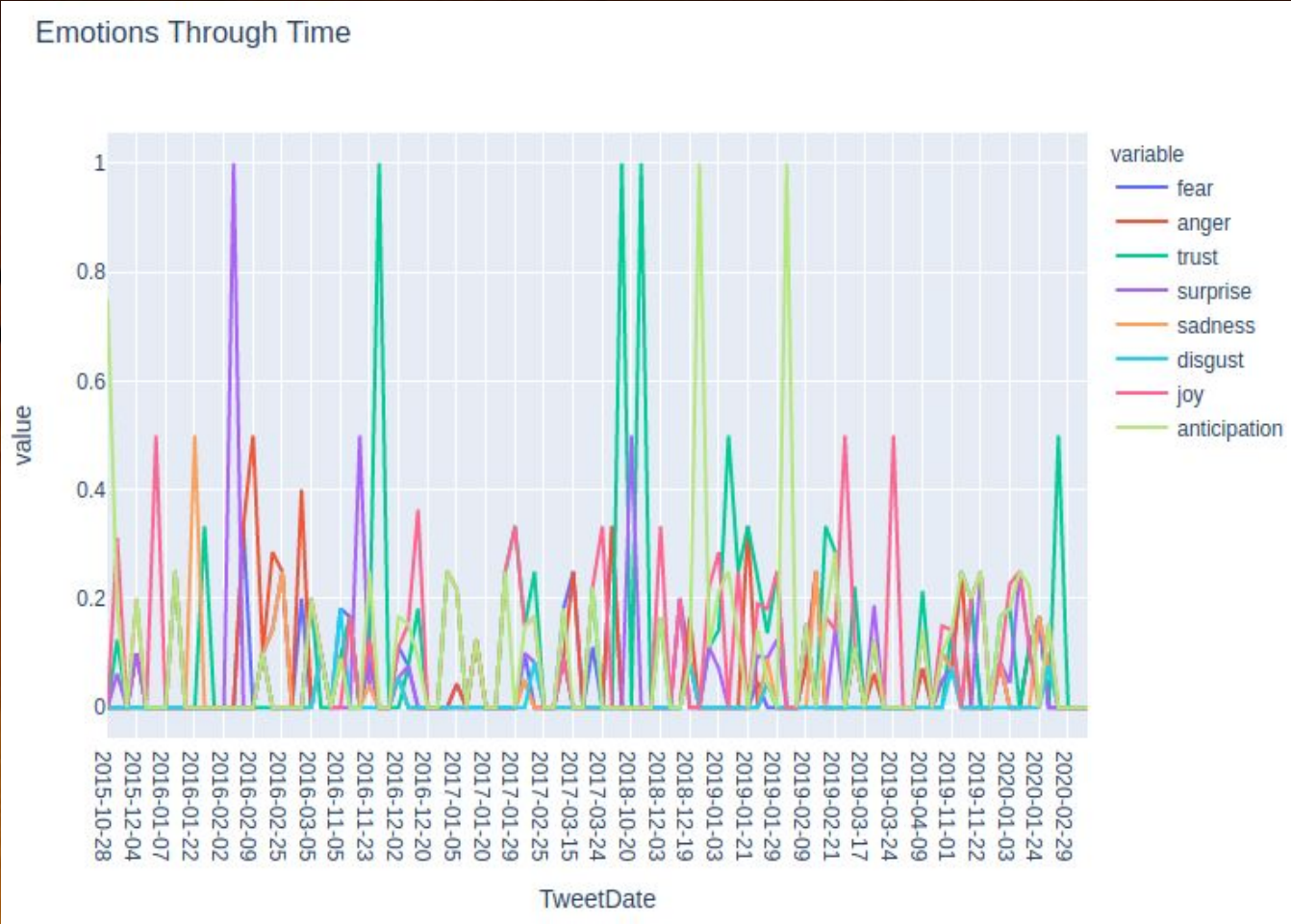
NBA Player Object

- Each player possessed their own unique dataframe.
- Their average points per game and shot accuracy were persisted through pickling.
- All analyzers used some variety of either individual player or all players' information mixed together.

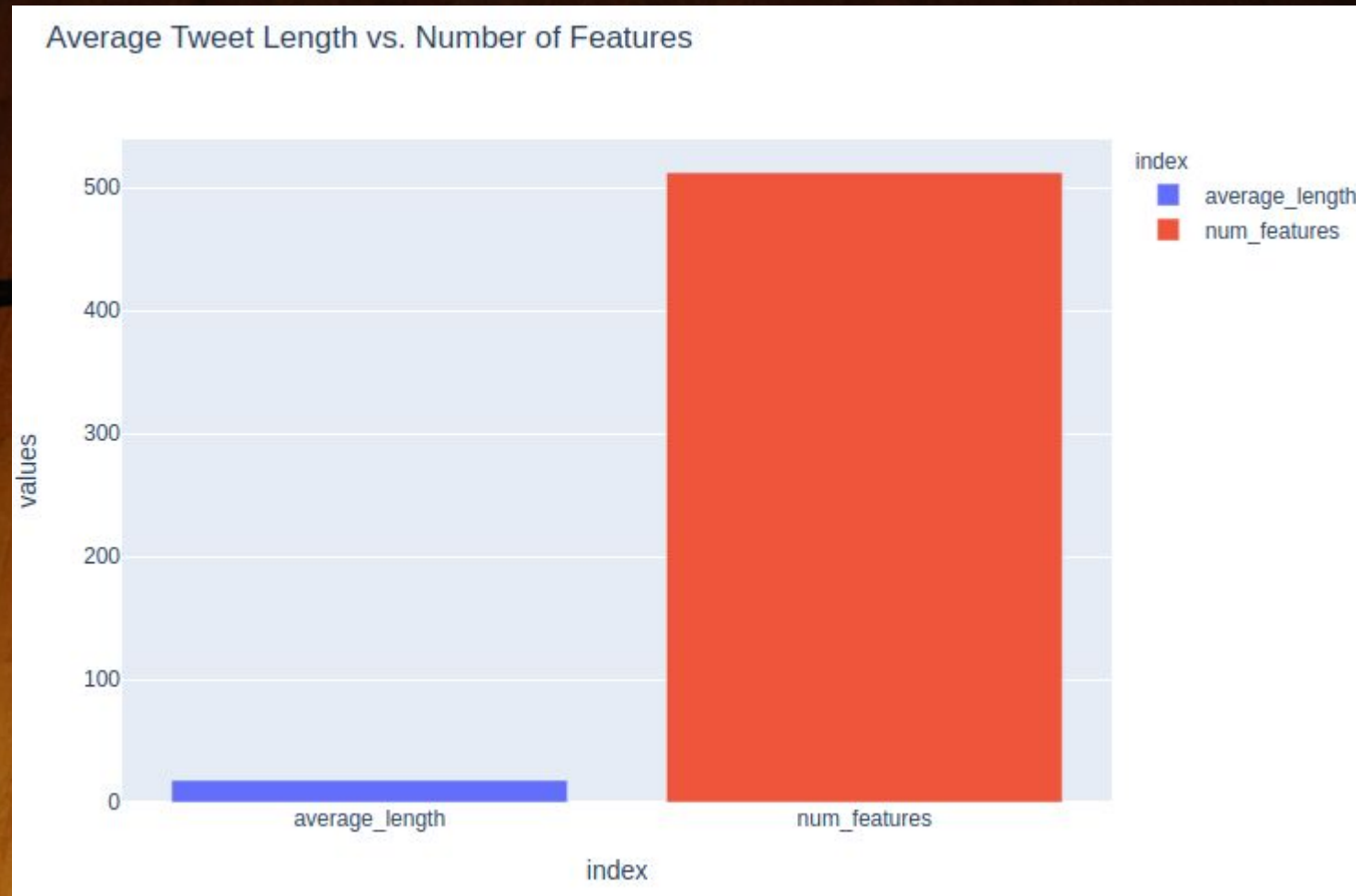
Analysis - Sentiment



Analysis - Emotions



Analysis - Embeddings



Experiments

Target Variable:

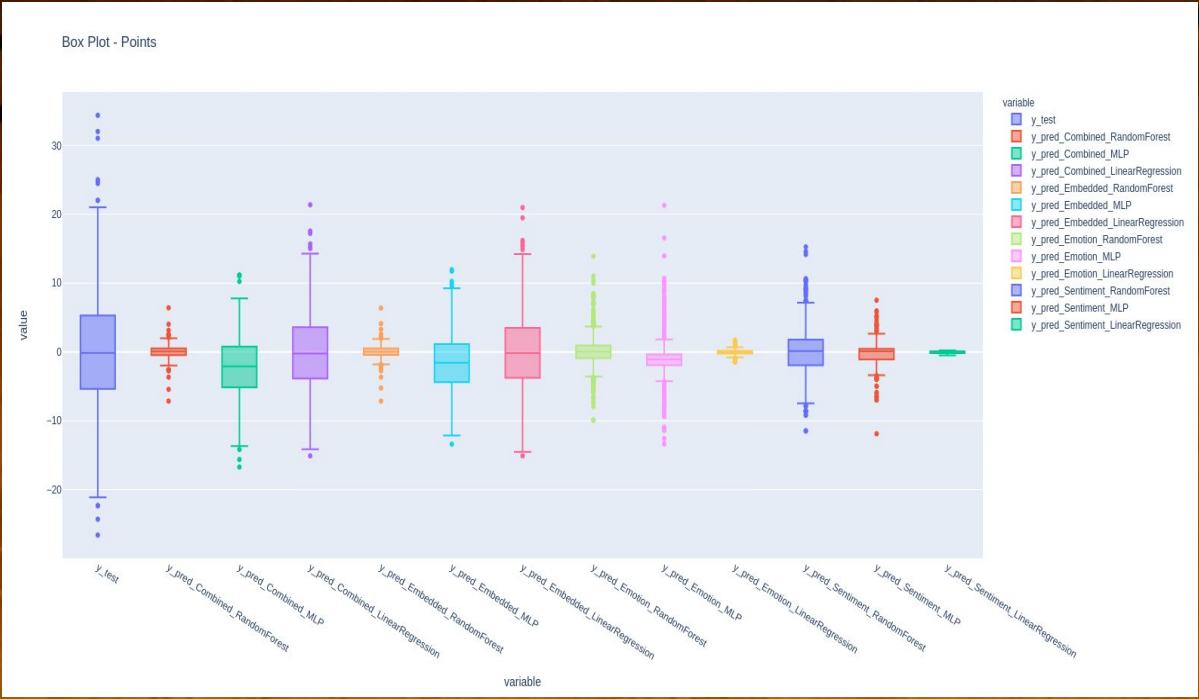
- PlusMinusPoints Or PlusMinusAccuracy

Predictor Variables:

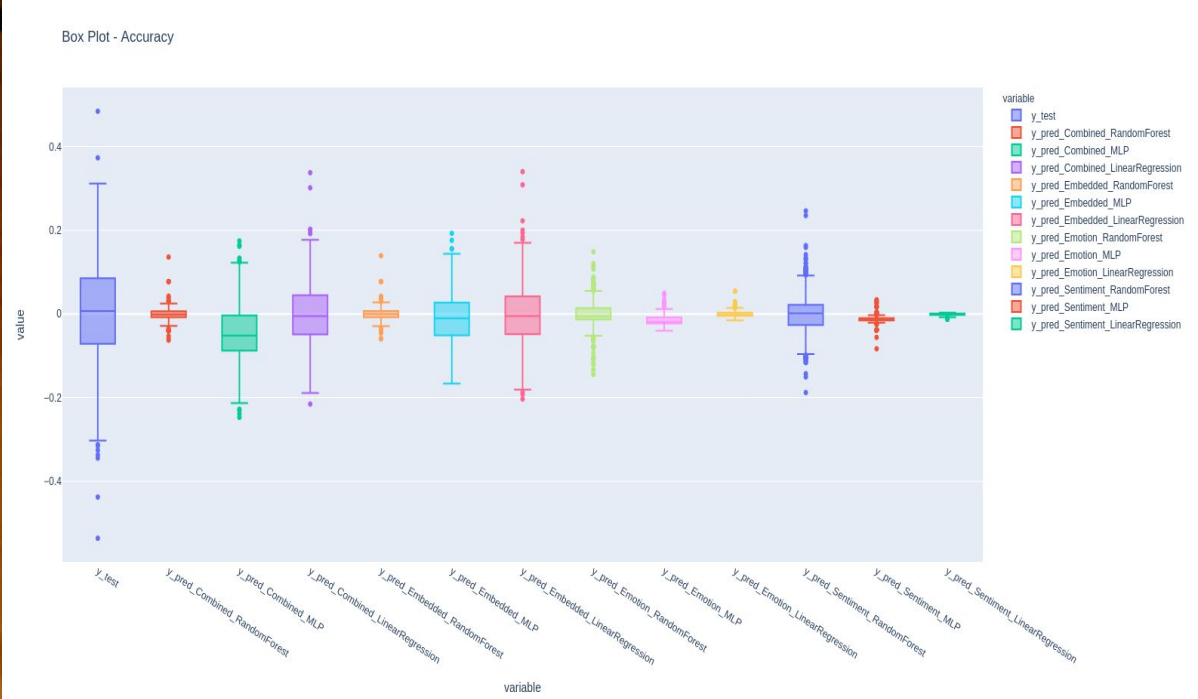
- Sentiments only (positive, negative, neutral)
- Emotions only (fear, anger, trust, surprise, sadness, disgust, joy, anticipation)
- Embeddings only
- Combined

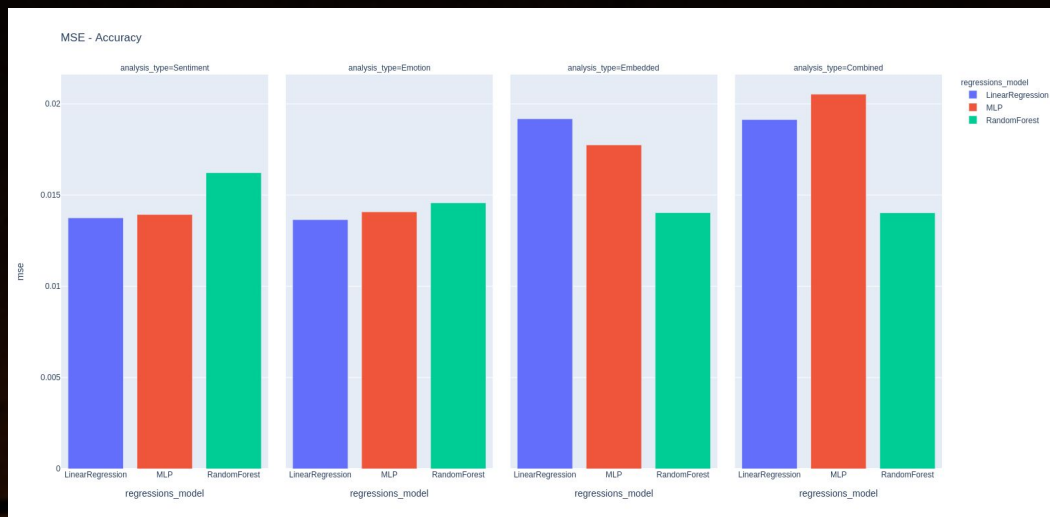
Results: Box Plots

Points

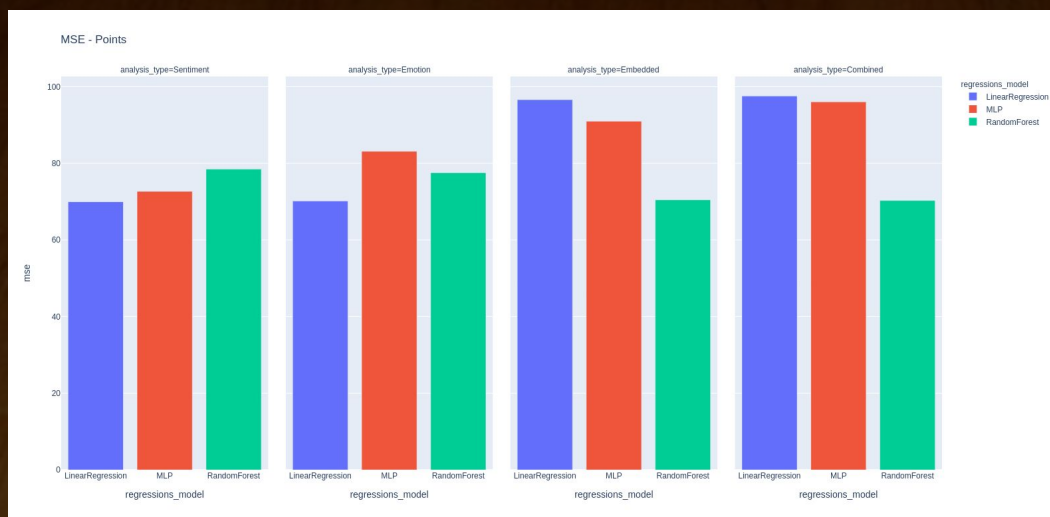


Accuracy





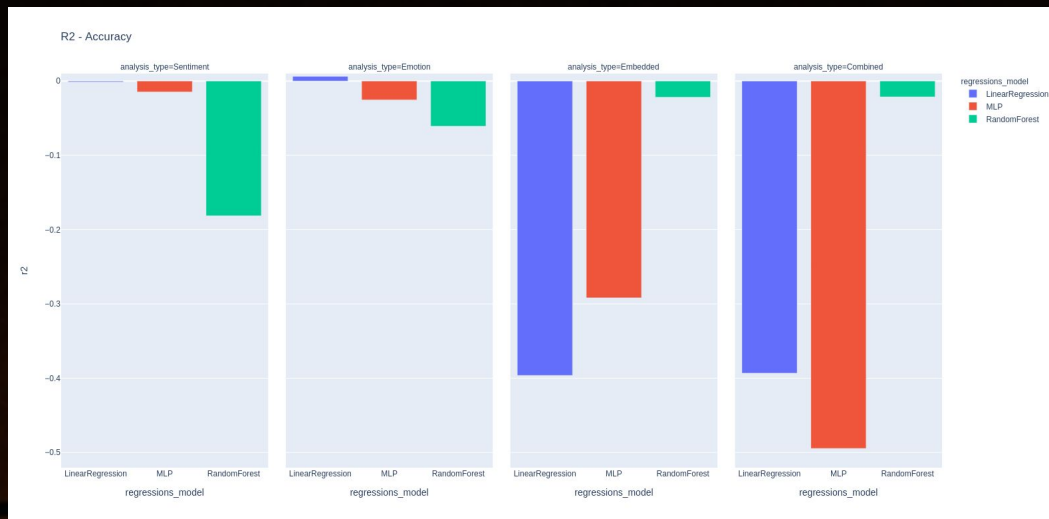
Accuracy - MSE



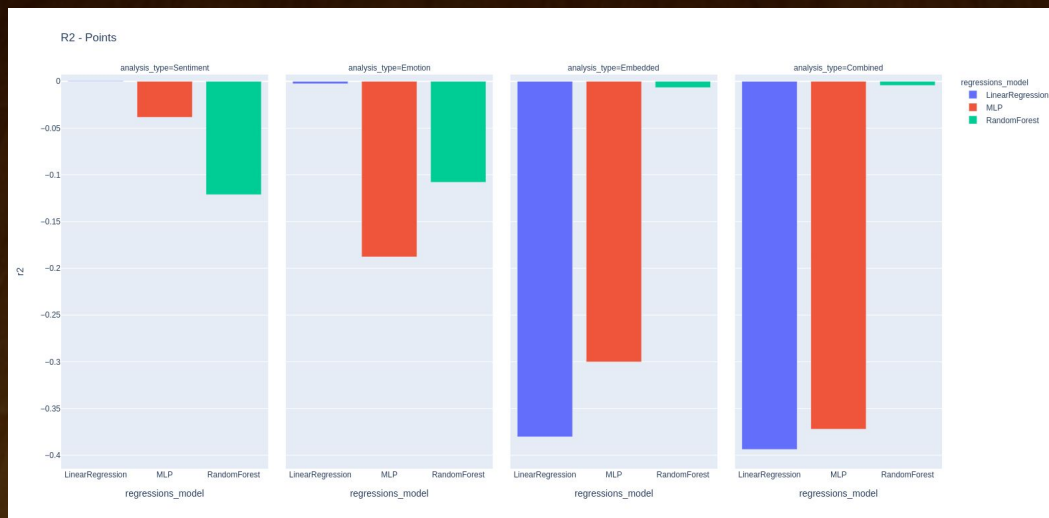
Points - MSE

Results

MSE



Accuracy - R-squared



Points - R-squared

Results

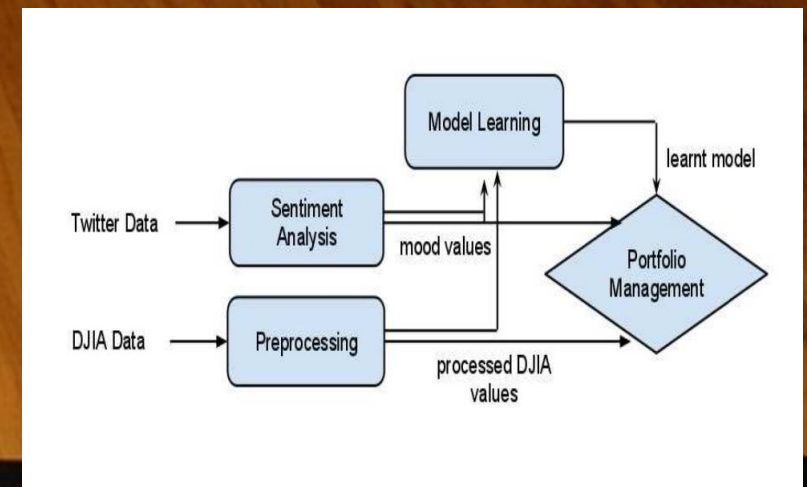
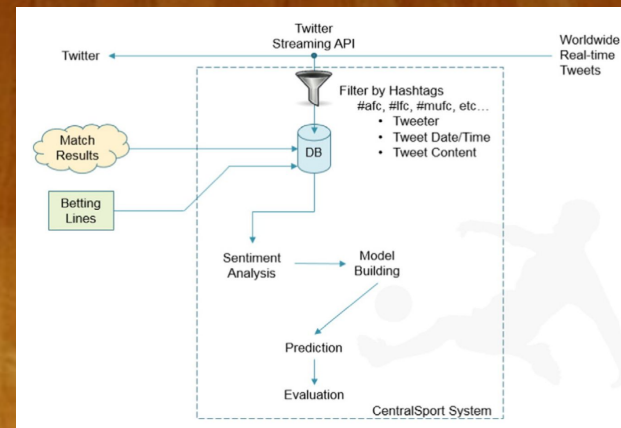
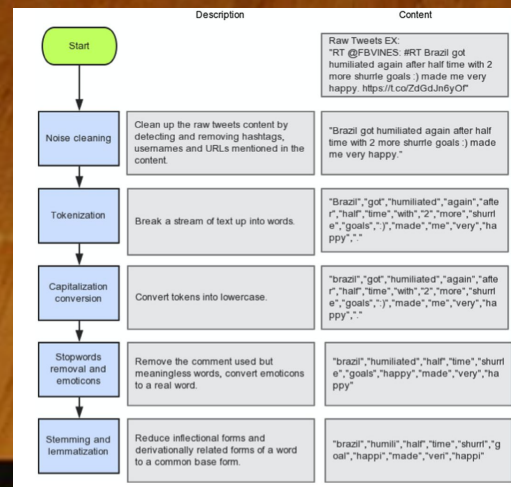
R-squared

Results: Conclusions

- R-squared results show that these models are not good at predicting the players performances
- Some factors contributing to these results are:
 - Small dataset (2k compared to 10M in comparable studies)
 - Curated content
 - Behavioral factors, players reacting differently to same feelings, and the ability to attain high performance levels even while experiencing bad emotions

Related Work

- World Cup 2014 in Twitter Word: A big data analysis of sentiments in U.S. sports fans tweets.
- Predicting wins and spread in the Premier League using a sentiment analysis of twitter. Decision Support Systems
- Stock prediction using twitter sentiment analysis



Improvements

Future Work

- Implement sarcasm detector using CNN classifier
- Implement third party detector
- Implement automation to extract Tweeter handles

Technical Challenges

- Dataset size
- Microsoft's Azure sentiment analyzer client

Demonstration

A photograph of a wooden basketball court floor. A thick, curved black line, likely a boundary line, arcs across the middle of the frame. The text "Thank you" is written in a large, bold, white sans-serif font, centered horizontally and partially overlaid by the black line.

Thank you