Group Members:

Sam Barua, Tamer Abdelaal & Sam Portelli

Project Outline

We as an AFL Consultancy Company, have been commissioned by the AFL to prepare a high-level analysis of Football Game Statistics to understand team and player performance and to be able to predict outcomes in future matches. We have been requested to use machine learning to determine a prediction model and to provide visualisation tools to help understand game statistics which help predict the future outcomes of games using a number of factors, involving teams, players, rain, venue, and other.

Data Source

We will be looking at a data set that has been put together on the Kaggle website: <https://www.kaggle.com/stoney71/aflstats>. This data set has a wide set of statistics for seasons 2012 to 2021, whose original sources are from  [afltables.com](http://afltables.com/afl/afl_index.html) and [www.footywire.com](http://www.footywire.com/). This will provide a sound base to assist with our analysis. Please see article which provides suggestions for features to use in machine learning: <https://www.reddit.com/r/AFL/comments/8x17l4/machine_learning_in_afl_part_ii_its_all_about_the/>.   
Features suggested, based on rolling 5 game statistics (see article for details):

1. Team ranking
2. Form-line
3. In-game statistics differential
4. Player information
5. Team information
6. Fatigue

The metadata consists of three csv files:  
games.csv, players.csv and stats.csv.  
  
The statistics for each game can in the games.csv file can be combined with the statistics in the stats.csv file using the gameId and team as foreign keys. For analysis based on players, this can be derived on the basis of playerid for any particular game or set of games in a season.

**Games.csv has 12 fields of:**  
 gameId

year

round

date

venue

startTime

attendance

homeTeam

homeTeamScore

awayTeam

awayTeamScore

rainfall

The following 5 fields were added to assist with modelling:  
 winning team

winningmargin

winningmargin\_percent

hometeam\_points

awayteam\_points

**Players.csv has 7 fields of:** playerId

displayName

height

weight

dob

position

origin

**Stats.csv has 31 fields of:** gameId

team

year

round

playerId

displayName (player name)

gameNumber

Disposals

Kicks

Marks

Handballs

Goals

Behinds

Hit Outs

Tackles

Rebounds

Inside 50s

Clearances

Clangers

Frees

Frees Against

Brownlow Votes

Contested Possessions

Uncontested Possessions

Contested Marks

Marks Inside 50

One Percenters

Bounces

Goal Assists

% Played

Subs

Please see the ERD image created for original data:  
Table

Description automatically generated

Application Functionality

* To use past AFL game statistics to create a predictive model for AFL game outcomes in the future, using machine learning
* To demonstrate the model’s effectiveness by showing a past season’s predictions versus the actual results for a historical nominated round e.g. the actual versus prediction for Round 1 games of 2021, Round 20 of 2021.
* To prepare game predictions for Round 1 of the new season 2022, based on the fixture
* To prepare a game prediction for any two nominated teams for their next match in 2022.
* To show the top 10 players for 2021 based various statistics
* To create two dream teams made of the best players and to predict their game outcome.

Application Design

* To obtain AFL team information and statistics via CSV files from Kaggle.
* Use Postgres SQL to import, upload, convert data, creating summary data, which can be used for creating predictive models for machine learning.
* Use Python Pandas to develop the machine learning models to be used in the application, using data stored in the Postgres database (in Heroku)
* Use Flask to create a front end application, which will retrieve data from Postgres based on user options selected and present visualisations requested.

Final Design

Please see the following image for the design.

Github Repository

A link to the primary GitHub repository you’ll be housing your work in:  
**https://github.com/SamPortelli13/Final-Project.git**