CAB420 Project Proposal
Predicting the results of AFL matches

Shaun Sewell
N9509623

Project Objective

The objective of this project is to apply the methodology and model outlined by Aryan & Sharafat (2014) to AFL, replicating their application to the NBA datasets. Currently, there is no major publish work exploring the predictability of AFL match outcomes. This project aims to explore and test whether game outcomes can be predicted with any measure of accuracy.

Aryan & Sharafat (2014) model involved three steps: Statistic Prediction, Feature Formation and Result Prediction. Statistic prediction involves predicting the team's stats based on a statistical model, thus allowing the prediction of a future game's outcome. The four statistical models tested by Aryan & Sharafat (2014) are running averages, exponentially decaying averages, home and away averages, and cluster-based prediction. These statistical models were then applied in the Feature Formation step to form a single input vector that incorporates information about the teams involved in a match. Finally the Result Prediction is completed using linear regression, logistic regression and SVM using the feature vector as input.

Dataset

The data for this project is collected from the AFL statistics website footywire.com using a basic web page scraper. The data from each game of the past 7 AFL seasons (2012 - 2018), as well as the team records, were gathered. The table below shows the feature groups gathered for each match and season:

Disposals	Scoring Shots
Kicks	Goal Assists
Handballs	Inside 50s
Marks	Rebound 50s
Tackles	Contested Possessions
Hitouts	Uncontested Possessions
Clearances	Effective Disposals
Clangers	Disposal Efficiency
Frees For	Contested Marks
Frees Against	Marks Inside 50

Goals Kicked Behinds Kicked Rushed Behinds	One Percenters Bounces

Timeline

Task	Complete By
Submit proposal	Saturday 18th May
Introduction/Motivation, Related Works, Data and Methodology	Monday 20th May
Implement statistic prediction models	Friday 24th May
Implement feature formation	Friday 24th May
Implement ML algorithms (Linear Reg, Log Reg and SVM)	Friday 24th May
Gather prediction data	Saturday 25th May
Comparison, discussion, Conclusion and Future Works	Monday 27th May
Prepare presentation	Tuesday 28th May
Practice presentation	Wednesday 29th May
Presentation	Thursday 30th May
Submit final version of report	Sunday 2nd June

Aryan, O., & Sharafat, A.R. (2014). A Novel Approach to Predicting the Results of NBA Matches. http://cs229.stanford.edu/proj2014/Omid%20Aryan,%20Ali%20Reza%20Sharafat,%20A%20Novel%20Approach%20to%20Predicting%20the%20Results%20of%20NBA%20Matches.pdf