What features were used?

The features (information) that I used to try to predict matches can be classified into 6 broad categories. For each game (combination of home and away team), we can calculate difference in team ranking, form line, venue experience, key game statistics, player ranking and line up and fatigue.

*1.Team ranking:*

This was simple the different in ladder position, difference in number of ladder points, and difference in rolling 5 game percentages.

*2.Form-line*

This was simply the difference in the win-loss record over the last 5 rounds.

*3.Venue experience( Home ground advantage or HGA):*

This was the difference in wins at a particular venue between the home and away team over the last two years

*4.In-game Statistics differential*

These are the differences in in-game statistics between the home and away team, averaged over the last 5 rounds. These include: win loss form, score, percentage, kicks, handballs, contested possession, tackles, hit outs, rebound 50’s, inside 50s; free-kicks, clangers, marks inside 50’s, goal assists, bounces, time-on-ground.

Along with differences in the mean, the difference in the variance variance was also taken into account.

*5.Player information*

Along with team statistics, player performance was also taken into account. I used [u/JgreaterthanK](https://www.reddit.com/u/JgreaterthanK/)’s player rating formula and took the average player rating over the last 5 rounds. I then looked at the line up and based on the average player rating over the last 5 rounds for each player, calculated the expected total player rating for the upcoming match.

*6.Team information*

I also looked into team line-up before the match. For each match, I calculated the top performer for goals, clearances, goal assists, tackles, contested marks, and rebound 50s for the last 5 matches for both the home and away and determined whether or not that player was playing in the current match.

*7. Fatigue*

I modeled a fatigue factor as which third of the season we are in (1-8 = Beginning, 9-16 = Middle, 17-23 = end). I also had a feature in there indicating whether it was Round 1 or not so that the machine could learn any differences between round 1 predictions and the rest of the season.