STUDENT ID: 21170677

NAME: BIGYAN KARKI

Major Prediction Report- SAP

1. INTRODUCTION

AFL (Australian Football League) is a popular professional Australian rules football league nationwide. Data and statistics have become an integral part of the language of the AFL. Each year, statisticians crunch massive amounts of match data to help teams gain a competitive edge (Analytics in the AFL - The Most Data Rich Sport on Earth, 2016).The unpredictability of the matches and their outcomes, combined with the passion of the supporters, makes it an intriguing subject for data analysis and prediction.

The major prediction report required implementation of minimum two predictive models for determining whether a team would win or lose, as well as forecasting the margins of the final-8 teams during each week of 2023 AFL season. In addition, simulating the results of each week in the final-8 matches of 2023 AFL season and ultimately implementing them in a R-shiny Application was also a key part of the assessment.

2. METHODS

I have implemented ELO ratings with logistic regression model as my first model for predicting the win-loss and forecasting the margins. For my second model, random forest has been implemented and the ELO ratings with optimal parameters has been used for simulations for this assessment. Following includes further details of the entire methods and workflows undertaken to achieve all the tasks:

2.1. Data Sources

The dataset was extracted from the fitzRoy package in R, which provides a comprehensive collection of AFL tables for all the rounds of the season 2023 (fitzRoy (version 1.3.0), 2023). This was the common dataset used for training and testing all the models along with the simulations itself. Amongst all the features in the dataset, features such as: Round, Home. Team, Home. Points, Away. Team, Away. Points and Margin were chosen for all models. The round_number and score columns were computed for convenience.

	Round	Home.Team	Home.Points	Away.Team	Away.Points	Margin	round_number	score
1	R1	Richmond	58	Carlton	58	0	1	0.5
2	R1	Geelong	103	Collingwood	125	-22	1	0.0
3	R1	North Melbourne	87	West Coast	82	5	1	1.0
4	R1	Port Adelaide	126	Brisbane Lions	72	54	1	1.0
5	R1	Melbourne	115	Footscray	65	50	1	1.0
6	R1	Gold Coast	61	Sydney	110	-49	1	0.0
7	R1	GWS	106	Adelaide	90	16	1	1.0
8	R1	Hawthorn	65	Essendon	124	-59	1	0.0
9	R1	St Kilda	67	Fremantle	52	15	1	1.0
10	R2	Carlton	90	Geelong	82	8	2	1.0
11	R2	Brisbane Lions	93	Melbourne	82	11	2	1.0
12	R2	Collingwood	135	Port Adelaide	64	71	2	1.0
13	R2	Adelaide	76	Richmond	108	-32	2	0.0
14	R2	Footscray	41	St Kilda	92	-51	2	0.0
15	R2	Fremantle	72	North Melbourne	73	-1	2	0.0

Figure 1: Sample Dataset for all models

Also, minimal data cleaning and wrangling tasks were performed by checking for empty values and mutating a new column with integer round numbers for convenient data retrieval using the following code:

```
#Checking for any empty or missing values & Cleaning data dor models
afl_data_final[afl_data_final == ""] <- NA
any(is.na(afl_data_final))

afl_data_final <- afl_data_final %>%
  mutate(round_number = as.numeric(str_extract(Round, "\\d+")))
```

Figure 2: Data Cleaning and Wrangling R code

2.2. MODELS & IMPLEMENTATIONS

2.2.1. ELO ratings

In this assessment, Elo ratings model has been implemented to determine the ratings of each teams in terms of the opponent teams in the AFL,2023. These ratings were used to further train the Logistic Regression method to determine win-loss and forecast margin for final 8 matches of the season. The Elo rating system is a method for calculating the relative skill levels of players, team or both. A player's or teams Elo rating is represented by a number which increases or decreases based upon the outcome of games between rated players. After every game, the winning player takes points from the losing one. The difference between the ratings of the winner and loser determines the total number of points gained or lost after a game. This technique has been really useful in many sports and has been adapted to determine ratings in the NBA and NFL as well (Elo Rating System, n.d.). In this assessment elo package was used to calculate the Elo ratings. The confusion matrix was used to determine the accuracy which was 62.5%. The data split and hyper parameter tuning for Elo were performed as follows: The dataset was split in terms of rounds i.e. less than 13 rounds for training and remaining as testing dataset. Then, hyperparameter tuning was performed to tune the parameters "**Initial Elos**" and "**K-Factor**". For hyperparameter tuning, a parameter grid was established to systematically search over potential hyperparameter combinations, with initial_elos of range 1000 to 3000, in increments of 50 and k-values from range 10 to 50, in steps of 5.

- The Elo-scoring function then was computed with the ranges of values until the possible optimal parameters were obtained.
- Finally, the Brier score measure was computed, which provided mean squared difference between predicted probabilities and actual outcomes for the optimal parameters obtained. And the differences of elo ratings obtained with a model run of optimal parameters were used to train the logistic regression model.

2.2.2. Logistic Regression

Logistic regression is a widely used statistical method to predict the probability of a binary outcome based on one or more predictor variables. Logistic regression is a process of modelling the probability of a discrete outcome given an input variable. The most common logistic regression includes a binary outcome; something that can take two values such as true/false, yes/no, and so on. Multinomial logistic regression can model scenarios where there are more than two possible discrete outcomes (Logistic Regression, 2007).

As aspects of this assessments are a classification problem, which is the reason for implementation of this model. In the context of our AFL dataset, we utilized logistic regression to predict match outcomes, which is denoted as score based on the predictor variable D. Rate, which is the difference of the Elo ratings obtained.

	Round	Home.Team	Home.Points	Away.Team	Away.Points	Margin	round_number	score	D.Rate
1	R1	Richmond	58	Carlton	58	0	1	0.5	0.0000000
2	R1	Geelong	103	Collingwood	125	-22	1	0.0	-15.0000000
3	R1	North Melbourne	87	West Coast	82	5	1	1.0	15.0000000
4	R1	Port Adelaide	126	Brisbane Lions	72	54	1	1.0	15.0000000
5	R1	Melbourne	115	Footscray	65	50	1	1.0	15.0000000
6	R1	Gold Coast	61	Sydney	110	-49	1	0.0	-15.0000000
7	R1	GWS	106	Adelaide	90	16	1	1.0	15.0000000
8	R1	Hawthorn	65	Essendon	124	-59	1	0.0	-15.0000000
9	R1	St Kilda	67	Fremantle	52	15	1	1.0	15.0000000
10	R2	Carlton	90	Geelong	82	8	2	1.0	22.1762493
11	R2	Brisbane Lions	93	Melbourne	82	11	2	1.0	0.6472000
12	R2	Collingwood	135	Port Adelaide	64	71	2	1.0	15.0000000
13	R2	Adelaide	76	Richmond	108	-32	2	0.0	-22.1762493

Figure 3: Training Data Set for Logistic Regression

Then, for my assessment I performed round-by-round predictions of the win-loss for final 8 matches of the season i.e. for rounds QF, EF, SF, PF and GF. Finally, confusion matrix was implemented to determine the performance of the model.

Further, the Receiver Operating Characteristic (ROC) curve was plotted to visualize the model's performance for win/loss predictions. Similar processes were performed for the margin prediction. In addition, for margin predictions part, its effectiveness was evaluated by the Root Mean Square Error (RMSE) and R-Squared (R^2) value, which provided further insight on the percentage of variability explained by our model.

2.2.3. Random Forest

Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fuelled its adoption, as it handles both classification and regression problems (What is Random Forest, n.d.).

Following steps describes the tasks performed for this model:

- The dataset as in figure-1, were used to train the model and the tidymodels package in R was used, wherein the Random Forest model was setup for regression mode with the ranger engine.
- The model was trained on the entire dataset and evaluated on its performance. The R-squared value was computed for both training and testing data to measure the proportion of the variance for the win-loss outcome that's explained by the predictors.
- To avoid overfitting, cross-validation with multiple folds of the training data was employed.
- Hyperparameter tuning was performed, where a grid search approach was used to tune the parameters mtry (number of variables randomly sampled as candidates at each split when building a tree.), trees (number of trees in the Random Forest) and min_n (minimum node size). The best model was then selected based on the root mean square error (RMSE).
- The random forest model was then trained with the optimal values of trees and mtry.
- The model was used to predict win-loss outcomes for the final-8 rounds of the season and confusion matrix was also computed to evaluate the model's win/loss classification performance.
- For the margin prediction, same processes were performed in the model and Actual vs. predicted margins were plotted, also performance metrics like RMSE and R-squared were calculated to assess the accuracy and goodness-of-fit of the model.

2.2.4. Simulations

For simulation, the same data set was run to the ELO model with optimal hyper parameters obtained from previous elo models, i.e. k_factor =15 and initial_elo_rating = 2000, to obtain the elo ratings of final 8 teams. For simulation purposes, it was necessary to establish all potential pairings between the top 8 teams. Utilizing the permutations function, a grid was created and the respective team's elo ratings were added and a data frame was formed.

```
> grid
# A tibble: 56 × 4
# Groups:   team, opp [56]
   team  opp   elo_team elo_opp
   <chr> <chr>    <dbl>   <dbl>
 1 ADL   BRI      2049.   2075.
 2 ADL   CAR      2049.   2040.
 3 ADL   COL      2049.   2095.
 4 ADL   GWS      2049.   2034.
 5 ADL   MEL      2049.   2040.
 6 ADL   ST       2049.   2004.
 7 ADL   SYD      2049.   2009.
 8 BRI   ADL      2075.   2049.
 9 BRI   CAR      2075.   2040.
10 BRI   COL      2075.   2095.
# ℹ 46 more rows
```

Figure 4: Simulation grid with ELO

For the next phase of simulation, week-by-week prediction of the winner and loser was performed, which involved a sampling process based on team Elo ratings, then aggregated to derive the prediction results. Furthermore, the result was compared with the actual 2023 AFL results to determine the performance of the simulations. Spanning from week-1 to week-4, a total of 4 simulations were conducted, with each week encompassing different rounds. To articulate the effectiveness of the simulations, accuracy metrics were determined via a confusion matrix. Additionally, RMSE values were utilized, and a ROC curve was plotted for visual representation.

3. RESULTS

After computation of the models in R, the following results were obtained for the Win-Loss and Margin Models:

3.1. ELO & Logistic Regression Results

From the ELO model, the difference in ratings between the teams in each match were obtained and further used to train the Logistic Regression Model,

team.A	team.B	p.A	wins.A	update.A	update.B	elo.A	elo.B	tip_elo
Richmond	Carlton	0.5000000	0.5	0.0000000	0.0000000	2000.000	2000.000	0
Geelong	Collingwood	0.5000000	0.0	-7.5000000	7.5000000	1992.500	2007.500	0
North Melbourne	West Coast	0.5000000	1.0	7.5000000	-7.5000000	2007.500	1992.500	0
Port Adelaide	Brisbane Lions	0.5000000	1.0	7.5000000	-7.5000000	2007.500	1992.500	0
Melbourne	Footscray	0.5000000	1.0	7.5000000	-7.5000000	2007.500	1992.500	0
Gold Coast	Sydney	0.5000000	0.0	-7.5000000	7.5000000	1992.500	2007.500	0
GWS	Adelaide	0.5000000	1.0	7.5000000	-7.5000000	2007.500	1992.500	0
Hawthorn	Essendon	0.5000000	0.0	-7.5000000	7.5000000	1992.500	2007.500	0
St Kilda	Fremantle	0.5000000	1.0	7.5000000	-7.5000000	2007.500	1992.500	0
Carlton	Geelong	0.5107917	1.0	7.3381246	-7.3381246	2007.338	1985.162	1
Brisbane Lions	Melbourne	0.4784267	1.0	7.8236000	-7.8236000	2000.324	1999.676	0
Collingwood	Port Adelaide	0.5000000	1.0	7.5000000	-7.5000000	2015.000	2000.000	0
Adelaide	Richmond	0.4892083	0.0	-7.3381246	7.3381246	1985.162	2007.338	0
Footscray	St Kilda	0.4784267	0.0	-7.1764000	7.1764000	1985.324	2014.676	0
Fremantle	North Melbourne	0.4784267	0.0	-7.1764000	7.1764000	1985.324	2014.676	0

Figure 5: Optimal ELO ratings

After running the logistic regression model, the predicted win-loss values obtained for the final rounds of AFL 2023 season was as follow, where in 1 represents win and 0 represents loss:

	ID	Round	Home_Team	Away_Team	predicted.classes
1	1	QF1	Collingwood	Melbourne	1
2	2	EF1	Carlton	Sydney	1
3	3	EF2	St Kilda	GWS	0
4	4	QF2	Brisbane Lions	Port Adelaide	1
5	5	SF1	Melbourne	Carlton	1
6	6	SF2	Port Adelaide	GWS	1
7	7	PF1	Collingwood	GWS	1
8	8	PF2	Brisbane Lions	Carlton	1
9	9	GF	Collingwood	Brisbane Lions	1

Figure 6: Predicted Win- Loss Logistic Regression

	Round	Home.Team	Away.Team	score
1	QF	Collingwood	Melbourne	1
2	EF	Carlton	Sydney	1
3	EF	St Kilda	GWS	0
4	QF	Brisbane Lions	Port Adelaide	1
5	SF	Melbourne	Carlton	0
6	SF	Port Adelaide	GWS	0
7	PF	Collingwood	GWS	1
8	PF	Brisbane Lions	Carlton	1
9	GF	Collingwood	Brisbane Lions	1

Figure 7: Original Win-Loss

The performance was analysed using a confusion matrix and the statistics of the matrix are as follow:

	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Precision	Recall	F1	Prevalence	Detection Rate	Detection Prevalence	Balanced Accuracy
1	1	0.3333333	0.75	1	0.75	1	0.8571429	0.6666667	0.6666667	0.8888889	0.6666667

Figure 8: WIN-LOSS Confusion Matrix Stat

For the Margin Prediction, the results obtained from the model and original margins are as follows along with the training model graph and error metrics:

	ID	Round	Home_Team	Away_Team	predicted.classes_margin
1	1	QF1	Collingwood	Melbourne	24
2	2	EF1	Carlton	Sydney	22
3	3	EF2	St Kilda	GWS	-4
4	4	QF2	Brisbane Lions	Port Adelaide	17
5	5	SF1	Melbourne	Carlton	6
6	6	SF2	Port Adelaide	GWS	12
7	7	PF1	Collingwood	GWS	33
8	8	PF2	Brisbane Lions	Carlton	28
9	9	GF	Collingwood	Brisbane Lions	18

Figure 9:Predicted Margin

Round	Home.Team	Home.Points	Away.Team	Away.Points	Margin
QF	Collingwood	60	Melbourne	53	7
EF	Carlton	74	Sydney	68	6
EF	St Kilda	77	GWS	101	-24
QF	Brisbane Lions	123	Port Adelaide	75	48
SF	Melbourne	71	Carlton	73	-2
SF	Port Adelaide	70	GWS	93	-23
PF	Collingwood	58	GWS	57	1
PF	Brisbane Lions	79	Carlton	63	16
GF	Collingwood	90	Brisbane Lions	86	4

Figure 10: Original Margin AFL 2023

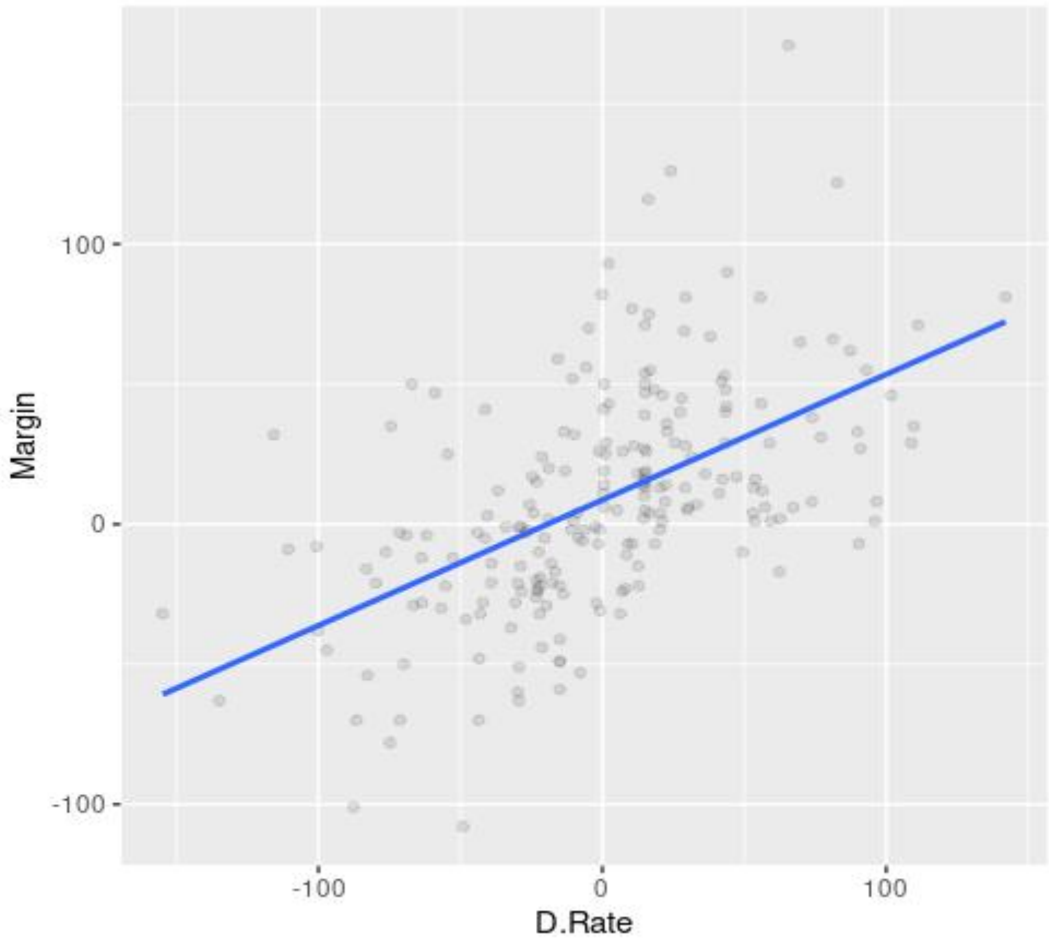

Figure 11: Logistic Regression Training Graph

ID	Measures	value
1	RMSE	22.5067891
2	R2	0.2161757

Figure 12: Error Metrics

3.2. Random Forests Results

The results from hyperparameter tuning obtained, resulted in value of number of trees as 353 and number of features as 6 and were used to train the model for further predictions.

The results obtained with the random forest model along with original data including performance measures are as follows:

	Round	Home.Team	Away.Team	score
1	QF	Collingwood	Melbourne	1
2	EF	Carlton	Sydney	1
3	EF	St Kilda	GWS	0
4	QF	Brisbane Lions	Port Adelaide	1
5	SF	Melbourne	Carlton	0
6	SF	Port Adelaide	GWS	0
7	PF	Collingwood	GWS	1
8	PF	Brisbane Lions	Carlton	1
9	GF	Collingwood	Brisbane Lions	1

Figure 13: Original Win-Loss Data

ID	Round	Home_Team	Away_Team	predicted.classes_margin
1	QF1	Collingwood	Melbourne	1
2	EF1	Carlton	Sydney	1
3	EF2	St Kilda	GWS	0
4	QF2	Brisbane Lions	Port Adelaide	1
5	SF1	Melbourne	Carlton	0
6	SF2	Port Adelaide	GWS	0
7	PF1	Collingwood	GWS	1
8	PF2	Brisbane Lions	Carlton	1
9	GF	Collingwood	Brisbane Lions	1

Figure 14:Predicted Win-Loss Random Forest

	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Precision	Recall	F1	Prevalence	Detection Rate	Detection Prevalence	Balanced Accuracy
1	1	1	1	1	1	1	1	0.6666667	0.6666667	0.6666667	1

Figure 15:Confusion Matrix Statistics for Win Loss

Round	Home.Team	Home.Points	Away.Team	Away.Points	Margin
QF	Collingwood	60	Melbourne	53	7
EF	Carlton	74	Sydney	68	6
EF	St Kilda	77	GWS	101	-24
QF	Brisbane Lions	123	Port Adelaide	75	48
SF	Melbourne	71	Carlton	73	-2
SF	Port Adelaide	70	GWS	93	-23
PF	Collingwood	58	GWS	57	1
PF	Brisbane Lions	79	Carlton	63	16
GF	Collingwood	90	Brisbane Lions	86	4

Figure 16:Original Margin

ID	Round	Home_Team	Away_Team	prediction_margin
1	QF1	Collingwood	Melbourne	11
2	EF1	Carlton	Sydney	7
3	EF2	St Kilda	GWS	-23
4	QF2	Brisbane Lions	Port Adelaide	52
5	SF1	Melbourne	Carlton	-11
6	SF2	Port Adelaide	GWS	-21
7	PF1	Collingwood	GWS	9
8	PF2	Brisbane Lions	Carlton	15
9	GF	Collingwood	Brisbane Lions	10

Figure 17: Predicted Margins Random Forest

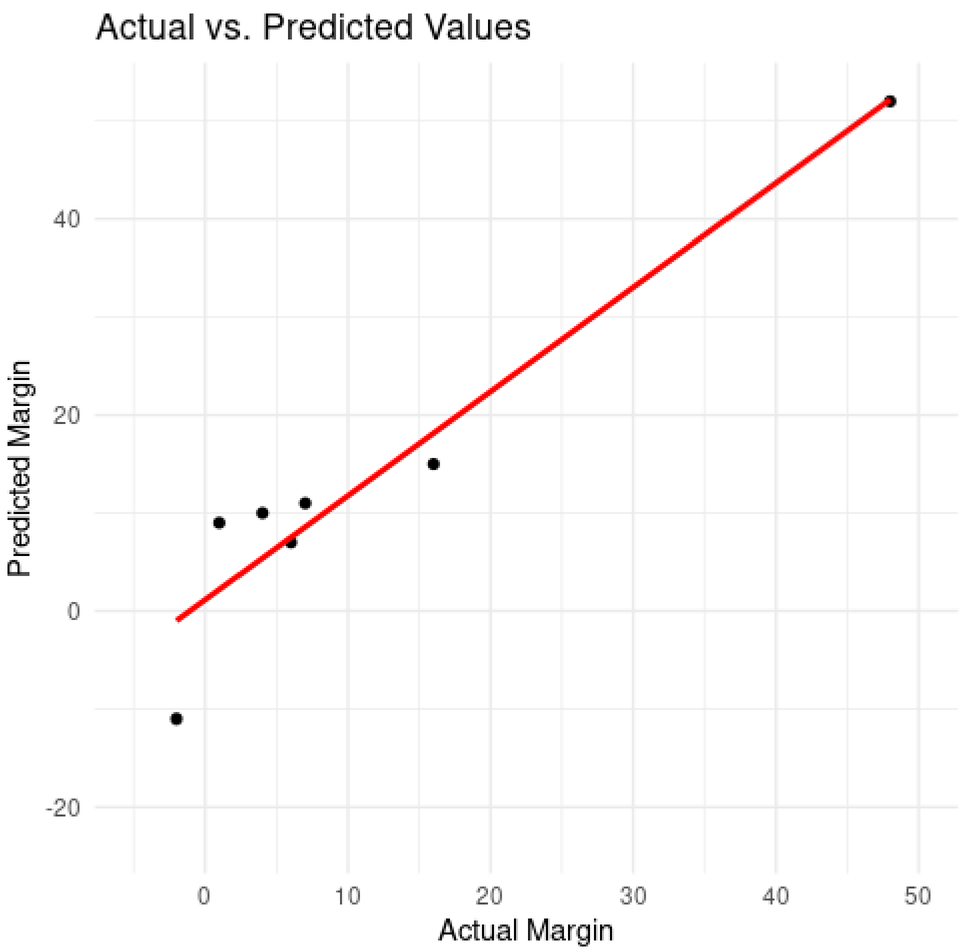

Figure 18: Actual vs Predicted Margin Random Forest

	ID	Measures	value
1	1	RMSE	4.9441323
2	2	R2	0.9544786

Figure 19: Error Metrics Random Forest

3.3. Simulation Results

Below are the original and final results obtained by performing simulations based on ELO and sampling:

finalID	team	opp	elo_team	elo_opp	winner	loser
QF1	COL	MEL	2095.389	2040.437	MEL	COL
QF2	BRI	ADL	2074.959	2049.061	ADL	BRI
EF1	CAR	SYD	2039.931	2008.904	CAR	SYD
EF2	ST	GWS	2004.406	2034.299	ST	GWS

Figure 20: Week-1 Predicted Simulation

finalID	team	opp	elo_team	elo_opp	winner	loser
SF1	MEL	CAR	2040.437	2039.931	CAR	MEL
SF2	ADL	GWS	2049.061	2034.299	GWS	ADL

Figure 21: Week-2 Predicted Simulation

finalID	team	opp	elo_team	elo_opp	winner	loser
PF1	COL	GWS	2095.389	2034.299	GWS	COL
PF2	BRI	CAR	2074.959	2039.931	BRI	CAR

Figure 22: Week-3 Predicted Simulation

finalID	team	opp	elo_team	elo_opp	winner	loser
GF	COL	BRI	2095.389	2074.959	BRI	COL

Figure 23: Grand Finals Predictions

The combined original and final simulations week by week are as follows:

finalID	team	opp	elo_team	elo_opp	winner	loser
QF1	COL	MEL	2095.389	2040.437	COL	MEL
QF2	BRI	ADL	2074.959	2049.061	BRI	ADL
EF1	CAR	SYD	2039.931	2008.904	CAR	SYD
EF2	ST	GWS	2004.406	2034.299	GWS	ST
SF1	MEL	CAR	2040.437	2039.931	CAR	MEL
SF2	ADL	GWS	2049.061	2034.299	GWS	ADL
PF1	COL	GWS	2095.389	2034.299	COL	GWS
PF2	BRI	CAR	2074.959	2039.931	BRI	CAR
GF	COL	BRI	2095.389	2074.959	COL	BRI

Figure 24: AFL 2023 Final 8 Results

finalID	team	opp	elo_team	elo_opp	winner	loser
QF1	COL	MEL	2095.389	2040.437	MEL	COL
QF2	BRI	ADL	2074.959	2049.061	ADL	BRI
EF1	CAR	SYD	2039.931	2008.904	CAR	SYD
EF2	ST	GWS	2004.406	2034.299	ST	GWS
SF1	MEL	CAR	2040.437	2039.931	CAR	MEL
SF2	ADL	GWS	2049.061	2034.299	GWS	ADL
PF1	COL	GWS	2095.389	2034.299	GWS	COL
PF2	BRI	CAR	2074.959	2039.931	BRI	CAR
GF	COL	BRI	2095.389	2074.959	BRI	COL

Figure 25: Simulation Predicted by Model

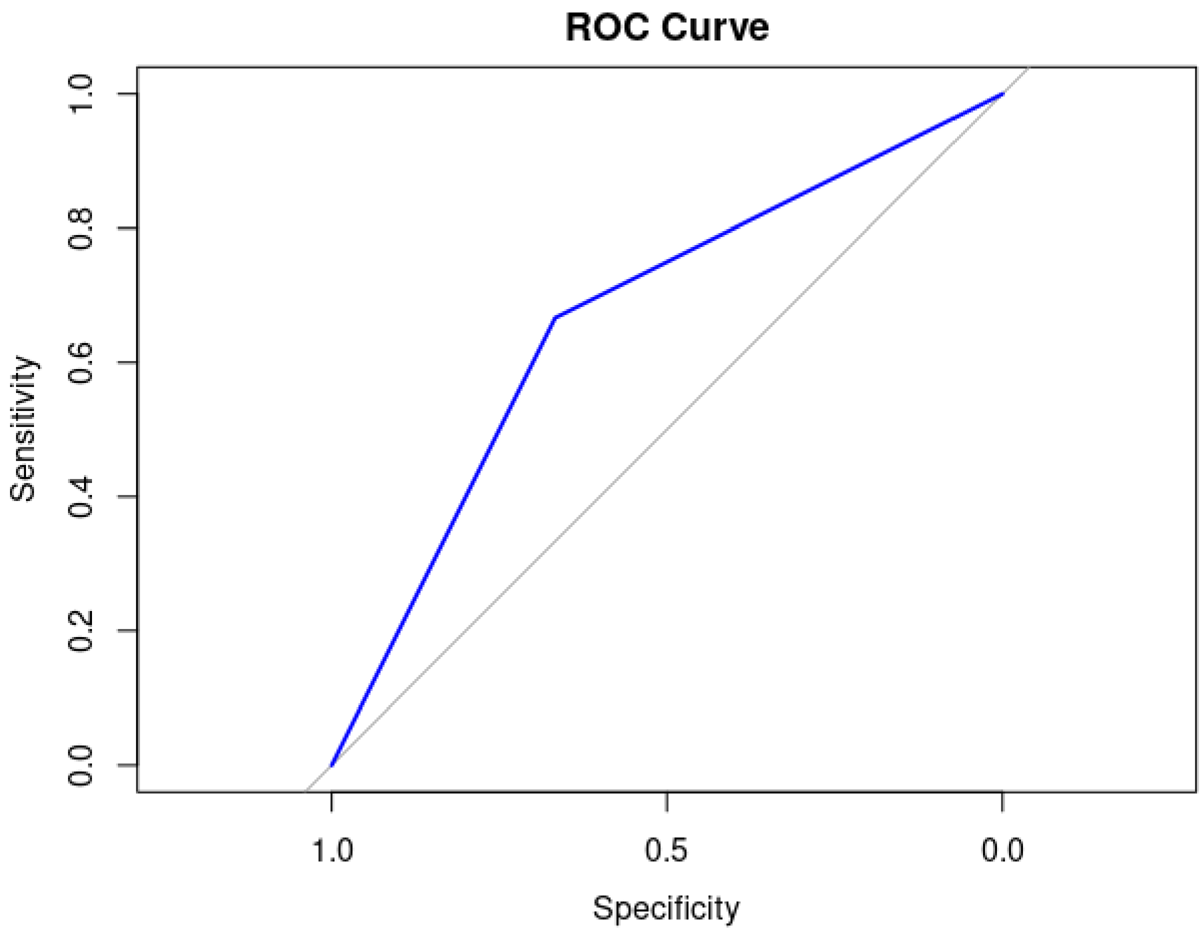

Figure 26: ROC Curve Final Simulation

Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Precision	Recall	F1	Prevalence	Detection Rate	Detection Prevalence	Balanced Accuracy
0.2	0.5	0.3333333	0.3333333	0.3333333	0.2	0.25	0.5555556	0.1111111	0.3333333	0.35

Figure 27: Confusion Matrix Statistics

4. DISCUSSION

From the models implemented, the win-loss and margin predictions were better predicted by the random forest than the Logistic Regression Model with ELO ratings. In my view, especially in terms of margins, the predictions are generally a regression problem rather than a classification problem. The logistic regression should usually work well in case of a classification problem contrary to the prediction of margins. Also, in my model the k-fold validation was not performed with the logistic regression and ELO models, but in case of random forests the implementation of k-fold validation for training and testing dataset and it's inherited feature engineering ability has led to better performance of both the win-loss outcomes and margins.

Also, footy is a sport of uncertainty and AFL has large sets of data and dependencies which ELO ratings would not have been able to distinguish and would result in poor performance along with logistic regression as compared to the Random forest model. This also states the fact that Elo and logistic performed minimal to explain the variability of the data as the R-Squared value was just 0.216 and less compared to random forests and the model lacked feature engineering as well. In terms of random forests, the R-square measure obtained was 0.9544 for margin predictions, which would state the since it is a non-linear model the dependencies were understood for better predictions without any correlation analysis of the features.

Amongst the final matches, the win-loss prediction by logistic regression for SF rounds were unusual as Melbourne lost against Carlton and Adelaide lost against GWS in 2023 AFL season. This would occur due to the limitation of ELO ratings, as the tournament progressed Carlton indeed had lower Elo ratings in previous rounds due to their loss against different teams. Due the lack of feature engineering, Elo's less sensitivity to the qualitative aspects of a team's performance improvements and lack of validation testing the logistic regression model could potentially have differed from the original result. This aspect is clear in the model R-Squared and RMSE values. In sporting terms, it can be related to as Sporting upsets as the performance of Carlton was improved into the second halves of the season as well as for GWS. The margins were way off for Logistic regression as it is a classification model and would perform with lesser accuracy for predictions, as a reference margin predicted for round EF2 was off by the value of 20 by this model.

For random forests, the performance of win-loss prediction was spot on as this is due to the fact that random forest has inbuilt feature engineering and accounts to the improvements and dependencies amongst the features for better prediction and explaining of the variability of the data which has been clearly shown in R-Squared value for margin predictions.

In terms of simulations, week-by week predictions were made as per the real matches occurred in the 2023 AFL season. The most surprising result occurred in the first week and round QF1 where in Collingwood lost against Melbourne as Collingwood are the 2023 champions. And also, in the grand finals Brisbane won against Collingwood as per the simulation model.

In my view, during the implementation of simulations, the updated elo ratings as per the progress of the tournament was not used, instead the hyperparameter tuned ELO ratings were used, which might have resulted to no being able to update the ELO ratings based on the season performance of team and thus led to poor outcomes. Also, only single loop of simulation was performed for each week, to get a more accurate prediction and understanding of the range of possible outcomes, more runs of the simulations would be required which can be a future task as well. Overall, the simulation performed well and aligned with the original results as the ROC curve for the simulation has its peak near 0.7.

For improvement, in terms of simulation as mentioned in above paragraph, more runs of simulations would be required and the most updated ELO ratings is required for sampling. Further, if random forests or other algorithms were used, would have been better for simulations rather than the sampling itself. In terms of logistic regression, the k-fold validation techniques should be implemented for better training and avoiding overfitting of the model and integration of different predictors(features) outside of data from the Fitzroy package should be implemented for better understanding of the dependencies and improvement of the overall performance.

Also, if the entire problem were indeed implemented with an AI machine learning models such as: Neural Networks, genetic algorithms in place of ELO ratings, Support Vector Machines (SVM) would potentially lead to better performance and predictions and could be a subject of future research in the context of AFL predictions.

References

Analytics in the AFL - The Most Data Rich Sport on Earth. (2016, September 1). Retrieved from VICE: https://www.vice.com/en/article/mg83jy/analytics-in-the-afl-the-most-data-rich-sport-on-earth

Elo Rating System. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Elo_rating_system

fitzRoy (version 1.3.0). (2023, March). Retrieved from RDocumentation: https://www.rdocumentation.org/packages/fitzRoy/versions/1.3.0

Logistic Regression. (2007). Retrieved from Logistic Regression - an overview | ScienceDirect Topics: https://www.sciencedirect.com/topics/computer-science/logistic-regression#:~:text=Logistic%20regression%20is%20a%20process,%2Fno%2C%20and%20so%20on

What is Random Forest. (n.d.). Retrieved from IBM: https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems