# *NFL Predictions*

Final Report

Team Xipher

Akhil K, 01FB15ECS026
5th Semester, Department of CSE
PES University
blitzkrieg.akhil@gmail.com

Arnav Deshpande, 01FB15ECS052
5th Semester, Department of CSE
PES University
arnav.deshpande6@gmail.com

Athul Pai, 01FB15ECS057
5th Semester, Department of CSE
PES University
athulpai05@gmail.com

***Abstract*— This literature survey proposes to cover previous literature about a model to predict certain parameters (like the win percentage, importance of a player to the result and the ticket price of a game) in the NFL (National Football League). For this purpose, Google scholar articles and papers, IEEE Xplore articles on data visualizations and some other relevant papers from the World Wide Web have been referred to. The current work constitutes of two models, a regression model and a classification model. The regression model is used to predict match ticket prices. Using standard regression techniques, a weighted average of the ticket price in dollars for the two teams in a game is calculated. The classification model is used to predict the result of a match as a win or a loss. This uses a neural network to predict a match result based on raw attributes. A weighted average of this with the player rating (based on player momentum, team momentum, player health, etc.) gives the result prediction.**

***Keywords—NFL ; predictions ; neural networks***

## I. INTRODUCTION

Football is a very popular American sport attracting huge investment in the sport. The worst teams on paper often stun the best of teams, with the right momentum. This makes it one of the most exciting and unpredictable sports ever. Investors are always at a dilemma about the team to bet their money on since NFL games are hugely based on momentum and not solely on facts on paper. The proposed model considers momentum to get a more accurate result than a vanilla classifier. With the help of a custom switch (like the ones used in LSTMs (Long Term Short Memory) to update memory), a combination of player ratings (an abstraction of momentum) and match predictions, better accuracy is achieved. With this improved accuracy, investors and fans are well informed about whom to bet their money on.

## II. Summary of Literature Survey

The work by Andrew D. Blaikie et. al [1] used multilayer artificial neural networks for predictions related to NFL results. A committee of machines approach was used for the prediction model. Top models were selected based on the mean square error. The paper states that a network with 8 neurons in the first hidden layer and 4 neurons in the second hidden layer did not perform well for the prediction.

### *Summary of the results reported*

The work by Andrew D. Blaikie et. al [1] used computer based simulations on www.thepredictiontracker.com to compare their efficiency with those models. The results show that the data reduction techniques usually work better than taking complete data and a large ANN. Also, none of the models performed well constantly over all the seasons. However, their models perform in the top half of prediction models compared to other models accounted in their work indicating that they have created a generally well performing model for NFL predictions compared to other models at that time.

### *Limitations of the previous works*

A few limitations were noticed in the work of Andrew D. Blaikie et. al [1]. Due to many conferences in NCAA compared to NFL, their model found it difficult to adjust the predictions with conferences. When a team from a strong conference plays a match with a team from a weak one, their model predicted an even probability of winning. But this is not the case, a team from a strong conference would have hustled more to stay alive in their conference and thus have an edge over the team from the weak conference. Also, a strong team with an incredible win stretch would have racked up attributes like rushing yards, touchdowns, points, etc. Then the team would have amazing statistics on average over the whole season and wouldn't depict the state of the team at a given point of time in the later stages of the season. A relatively weak team in a tough conference would have had tough wins and hence lower statistics than the former team. Intuitively, this is an even match but statistically, it shows the former team holds the edge over the latter. The results from their model predict that a purely statistical model doesn't depict the actual state of a team (momentum).

| Mean Absolute Error | | | | | |
|---|---|---|---|---|---|
| Model | 2007 | 2008 | 2009 | 2010 | Mean |
| On the Field | 11.00 | 11.31 | 11.68 | 10.99 | 11.24 |
| Efficiency | 11.00 | 11.70 | 11.69 | 11.25 | 11.41 |
| PCA | 11.21 | 11.12 | 11.48 | 11.20 | 11.25 |
| Every Statistics | 11.26 | 11.24 | 11.68 | 11.18 | 11.34 |
| LRCO | 11.20 | 11.10 | 11.51 | 11.16 | 11.24 |

| Legend |
|---|
| Upper Quartile |
| Top Half |
| Bottom Half |
| Lower Quartile |

Results from [2]

### *Lacuna in previous approaches*

The model in Andrew D. Blaikie et. al [1] uses 8 neurons in the first hidden layer and 4 in the second hidden layer did not perform well. A lacuna in the paper is that they did not explain why it is so. They did not explain the effect or consequence of doing so as well. They could have studied the effect on overall accuracy of the network by varying the size of the network. This might not be necessarily done, as part of the current work. It is also evident that some specific reasoning behind opting for the chosen network of 4 neurons in the first hidden layer and 6 neurons in the second hidden layer is lacking.

## III. Problem Statement

The current work involves the prediction of NFL match results by using a custom model incorporating a deep neural network and ticket prices for the match using regression modelling. This work considers the most important factor in NFL, the momentum. The prediction model has the capability to get better accuracy than previous models reported. All predictions are made for Week 17 by learning from Week 1 – 16's data.

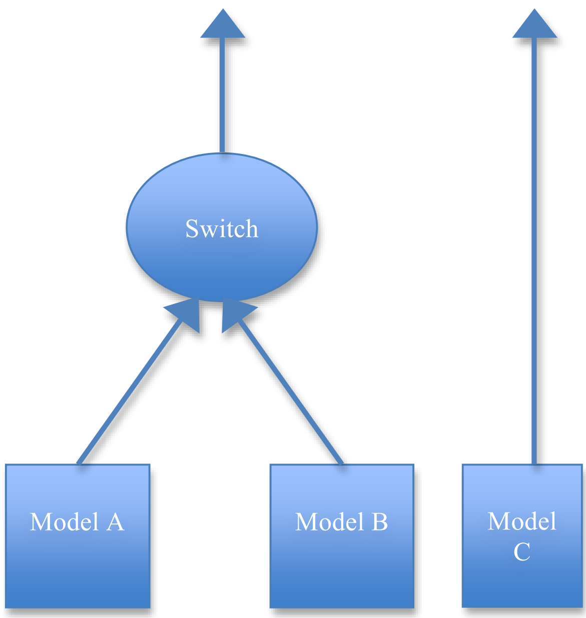
The data source is an online dataset from IBM Watson Analytics page. It is for the 2014 NFL Regular season. The datasets [3] are originally provided by Sports Data LLC.

The data sources for ticket prices [4] are from FiveThirtyEight's 2014 datasets on Github.

Due to lack of freely available data, our model is still in its budding form. Also, if the data has outliers (in the test data w/o cleaning), our model fails to achieve the claimed accuracy.

## IV. PROPOSED SYSTEM

The proposed system constitutes of four parts.




The four parts are:

1. Model A is our Neural Network Prediction model that predicts the result of a game between two teams.
2. Model B is the Team Rating model that gives a Week 17 rating for every team based on their Week 1 – 16 statistics.
3. The Switch essentially regulates the impact of Model A and Model B on the final prediction. This is the core element of our model.
4. Model C is a regression model that predicts ticket prices based on factors such as locality of the team, fan relationships, etc.

## V. DETAILED EXPLANATION OF THE PROPOSED SYSTEM

### MODEL A

Model A is used to predict the win and loss status of each team in the league.

This model uses the following dataset to predict the win and loss status.

The dataset contains a total of 89 attributes that contribute to the win and loss status but some of them outweigh the other and contribution by some of the features is negligible. Hence removing such features is a done as a preprocessing step before building the model.

To remove the features, we find the co-relation between win and loss feature and other features. The feature that has co-relation below a defined threshold is removed from the dataset.

After preprocessing the dataset, we start building the model. The preprocessed dataset is now passed in to a neural net that finds the respective weights for each of the feature in the dataset using back propagation technique.

The above weights constitute our model and it is used for predicting the win and loss status for a team.

### MODEL B

First, we are extracting data only for the regular NFL season and refining the dataset to remove rows/columns containing NAs.

It doesn't make sense to make a model on players who haven't played on Week 17. It will give a weird bias effect. So, we remove all those players who are "irrelevant".

Now, to find the team rating, we sum all the attributes column-wise for a team on a particular week. This converts individual player ratings to team rating. Once summed up, we normalize this data to differentiate the top teams from the rest, etc.

The ratings are padded with 1000 to have positive ratings for every team. It is visually and mathematically easier to infer results from these padded ratings.

To predict Week 17 ratings, I have used a custom weightage rating where later weeks' performance/ rating has a higher impact than earlier rating. Also known as the *momentum* of the game.

Once Week 17 predictions are produced, the switch uses them intelligently to produce more accurate match predictions.

### SWITCH

The switch is used for further increasing the accuracy of model A in predicting the win and loss accuracy.

Model B is like a black box that helps in further refining the output of the Model A.

Model B predicts the rating of each of the team playing in the season.

Using the weights of Model A, we calculate how probable a team is going to win or lose the match and then using the ratings predicted by model B we see whether the two teams playing are equally good or one team has an edge over the other.

If one team has an edge over the other we try to assign more weightage for that teams win and then again precompute the weights for Model A, by doing this we are able to achieve more accuracy and more refinement in the prediction of the win and loss status.

### MODEL C

The dataset used for the ticket price prediction contained 16 attributes. It was reduced down to 12 attributes using PCA (Principal Component Analysis) which explain about 96 percent of the variablility in the dataset. All of the attributes were used for prediction.

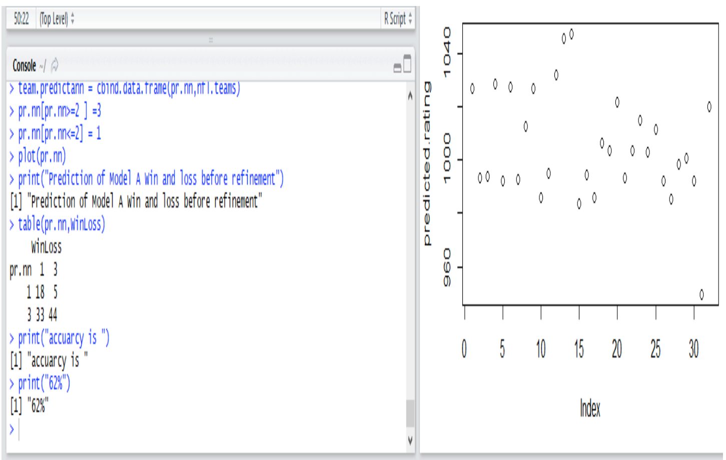
Two models were studied. The first one was ordinary multiple linear regression and the second one was a neural network. The Mean Squared Error (MSE) was computed in both cases in order to assess the model. The linear regression model in the first case turned out to be better as it yielded a lower MSE value.

Hence, due to the relative simplicity of the dataset, multiple linear regression was used to predict average ticket prices of NFL teams. This can be extended to predicting the average ticket price for every game. A weighted average of individual team prices would serve that purpose.

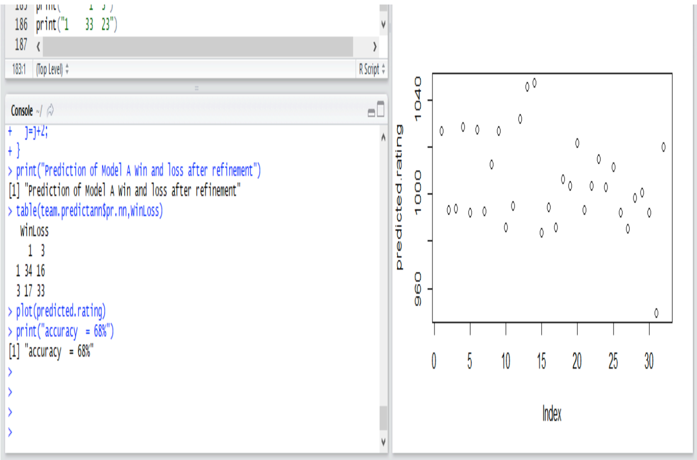
## VI. RESULTS

### RESULTS OF MATCH PREDICTION

Using just Model A we are able to achieve an accuracy of 60- 65%




After refining the model A by the using the results of Model B we are able to reach an accuracy level of 65- 70%.




Using our final model, we able to achieve an accuracy of 65 – 70%.

## VII. CONCLUSIONS

This project proposes a Rating driven approach that uses ratings (and momentum by extension) to predict NFL results with more accuracy. Experiments on this proposed model demonstrate the effectiveness of our approach.

Players, investors and fans can use this model to better understand teams and with it, their strengths and weaknesses. Weaker teams can use these models to correct their strategies, for improvement whereas elite teams can use this to look for further improvement. All in all, models such as ours emphasize on making teams better in their game.

## VIII. CONTRIBUTIONS

Athul Pai, 01FB15ECS057 – Model A and Switch.

Akhil K, 01FB15ECS026 – Model B.

Arnav Deshpande, 01FB15ECS052 – Model C.

# IX. REFERENCES

[1] Blaikie, Andrew D and Abud, Gabriel J and David, John A and Pasteur, R Drew (2011), Denison University, "NFL & NCAA Football Prediction using Artificial Neural Networks", Paperback publication, Proceedings of the 2011 Midstates Conference on Undergraduate Research in Computer Science and Mathematics.

[2] Kahn, J. (2003) "Neural Network Prediction of NFL Football Games," World Wide Web electronic publication, homepages.cae.wisc.edu/~ece539/project/f03/kahn.pdf.

[3] SportsData datasets at IBM Analytics Communities, https://www.ibm.com/communities/analytics/watson-analytics-blog/analyze-nfl-data-for-the-big-game/

[4] FiveThirtyEight datasets for Ticket Price Predictions, https://github.com/fivethirtyeight/data/blob/master/nfl-ticket-prices/national-average.csv