NFL Predictions

Literature Survey

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. PREVIOUS WORK

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.

J. Kahn et. al [2] used only a sample of data from NFL box scores assuming it represents the population. The work by J. Kahn et. al [2] uses a simple Artificial Neural Network (Back propagation multi-layer perceptron network) to account for the non-linearity in prediction of NFL games. The test data was applied to a variety of ANN configurations and the one with the best parameters was chosen. Each configuration varied by a single parameter with its previous one. They used supervised learning to generate accuracies for each of these configurations. It was discovered that a setup of 3 hidden nodes and a single hidden layer gives the highest accuracy for the data taken. Note that each model doesn't accurately fit data from every NFL source.

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

The work by J. Kahn et. al [2] tested their model on Week 14 and 15 of the 2003 NFL Season. For these weeks, the model achieved an accuracy of 75%. The model had an accuracy of 93.75% when tested on the test data.

Mean Absolute Error						Legend
Model	2007	2008	2009	2010	Mean	
On the Field	11.00	11.31	11.68	10.99	11.24	Upper Quartile
Efficiency	11.00	11.70	11.69	11.25	11.41	Top Half
PCA	11.21	11.12	11.48	11.20	11.25	Bottom Half
Every Statistics	11.26	11.24	11.68	11.18	11.34	Lower Quartile
LRCO	11.20	11.10	11.51	11.16	11.24	

Results of the model proposed in [1] for NFL

Table 2: Determination of learning coefficient and momentum parameters.

mu = 0.8				*mu=0*	
alpha	*0.01*	*0.05*	*0.1*	*0.01*	*0.05*
1	87.5	75	75	93.75	75
2	87.5	81.25	75	93.75	93.75
3	75	81.25	81.25	93.75	81.25
4	87.5	81.25	81.25	87.5	75
5	81.25	87.5	75	87.5	75
6	81.25	87.5	81.25	93.75	81.25
7	93.75	87.5	81.25	93.75	87.5
8	87.5	81.25	81.25	93.75	81.25
Average	85.15625	82.8125	78.90625	92.1875	81.25

Table 3: Determination of the number of hidden neurons.

	Number of Hidden Neurons					
	2	*3*	*4*	*5*	*6*	*8*
1	75	93.75	87.5	87.5	81.25	87.5
2	81.25	93.75	87.5	81.25	87.5	81.25
3	87.5	93.75	87.5	87.5	81.25	81.25
4	81.25	87.5	87.5	87.5	87.5	87.5
5	87.5	93.75	87.5	87.5	81.25	81.25
6	87.5	93.75	87.5	81.25	81.25	87.5
7	87.5	93.75	81.25	87.5	87.5	81.25
8	75	93.75	81.25	87.5	81.25	81.25
Average	82.8125	92.96875	85.9375	85.9375	83.59375	83.59375

Table 4: Determination of neural network structure.

	Network Structure	
	12-3-2	*12-3-3-2*
1	93.75	93.75
2	93.75	87.5
3	93.75	93.75
4	87.5	93.75
5	93.75	93.75
6	93.75	93.75
7	93.75	87.5
8	93.75	87.5
Average	92.96875	91.40625

Results from [2]

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).

Even though the work of J. Kahn et. al [2] had a decent accuracy of 75%, the model might face issues when fed with lots of data. This is due to their simplistic neural model as it doesn't have the capability of "learning" lots of complex attributes.

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. INFERENCES

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.

IV. PROBLEM STATEMENT

Prediction of NFL matches results (using a custom model incorporating a Deep Neural Network) and ticket prices for the game (using Regression modeling).

V. 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.