UNIVERSITY OF TORONTO

DEPARTMENT OF ENGINEERING SCIENCE

FINAL THESIS REPORT
APRIL 2021

Time Dependency in NFL Against the Spread Predictions

Author
Michael LIZZI
1003079704

Supervisors
Jimmy BA
Harris CHAN
Michael ZHANG

Abstract

Against the spread betting is a form of sports prediction aimed at the lucrative betting market in the National Football League. While recent revelations in the machine learning field have allowed for research in the application of neural networks for sports prediction, this paper aims to design new models meant specifically for increasing a bettor's likelihood of profitability when placing against the spread bets. The paper attempts to improve bettor profitability by accounting for the passage of time during model creation. Through experimentation and analysis, this paper demonstrates the difficulties of developing against the spread models and suggests how this form of betting is often not in a bettor's favour.

Acknowledgement

Thank you to my supervisors for helping support this project. Specifically, thank you Harris Chan and Michael Zhang for all the teaching, guidance, and suggestions and all your machine learning knowledge and expertise. I could not have done it without you. Here's to hoping we cash in on these algorithms one day!

Table of Contents

Table of Figures

Table of Tables

1 Introduction

The American football betting market is currently the largest of any major North American sport [1], with an estimated $6 billion wagered on the 2019 National Football League (NFL) Super Bowl alone [2]. One of the most common forms of wagers placed in the NFL is known as betting *against the spread* [3]. These types of bets involve choosing a team to win by a particular point margin, colloquially known as *the spread* or *the line*. The line is applied to help even the playing field between teams and encourage evenly distributed bets for the favourite and the underdog [4]. In recent years, the development of classification models using machine learning has allowed for an alternative statistical approach to sports prediction. If a suitable model can be trained using historical NFL data, a bettor may be able to increase their expected profitability. Furthermore, knowledge extracted from these models are desired by coaches and owners to help formulate strategies needed to win matches [5].

Current work has already been done to assess the efficacy of using machine learning models for sports prediction. However, these models fail to intelligently address the time series nature of the problem. Over time, previous statistics used to train a model may potentially become less relevant as teams reorganize rosters, develop players, or make changes to coaching personnel. Furthermore, current models take into account an arbitrary timeline for statistics, such as using the previous 3 games for data. The model may then become susceptible to outliers if a team recently underperformed or if star players were out with injuries. Lastly, previous works often do not attempt to make against the spread predictions. Against the spread betting allows for a possible statistical advantage over other betting methods because spreads are set based on public action, but the average bettor has been proven to hold a recency bias when placing against the spread bets [6]. Thus, the goal of this thesis is to design models for against the spread predictions and investigate how particular statistics affect performance. By applying different parameters or preprocessing steps in our dataset, we can investigate how models perform over different periods of time. I hypothesize that by creating models across different time periods, against the spread prediction rates will improve.

2 Background

2.1 Literature Review

Currently, most literature regarding sports prediction revolve around the use of supervised learning. Supervised learning is a machine learning technique in which each instance in a dataset uses the same set of features and is labelled with correct and known outputs [7]. In sports prediction, features such as yards gained, points scored, or fouls committed are data points which are repeatable across all games of a particular sport. Furthermore, games are easy to label with a win or a loss after the game has been played. This makes supervised learning one obvious candidate for sports prediction.

The work of Purucker [8] was one of the first machine learning models implemented for NFL prediction. Purucker implemented four different models, including 3 unsupervised and one supervised. For each model, he compiled 5 statistics from each team up until week 14 of the 1994 season to be used as features in his input vectors. Each input vector was then constructed by running a cumulative total of each statistic over a team's past 3 games. The statistics compiled were: number of victories, yard differential, rushing yard differential, turnover margin, and possession time margin. The supervised model was the most effective and consisted of a continuous-input back-propagation network with one hidden layer. This model was trained by producing exemplar target and defeat-profile vectors based on the statistics of the top 4 and bottom 4 teams respectively. An additional element representing a win or loss was also included at training time. A variety of network parameters were tested including different model architectures and epoch size in order to find the optimal model. Ultimately, the back-propagation model correctly predicted 64.3% of week 15 games as compared to a <64% prediction rate by a football expert in the Pittsburgh Post-Gazette. Purucker also experimented by including two additional features in the input vectors, those being the Las Vegas line and a schedule difficulty factor. The Las Vegas line proved to increase the correct prediction rate while the schedule difficulty factor actually had negative results. Using the line for predicting games in week 15 and week 16, the model successfully predicted 71.4% and 78.6% of games respectively. Purucker's work outlines the feasibility of using machine learning for NFL prediction. However, his models take into account an arbitrary timeline for statistics in where he only used the previous 3 games for data. These models may then become susceptible to outliers if a team recently underperformed or if star players were out with injuries. Furthermore, Purucker addresses himself that improved performance may be achieved through better data encoding, additional testing of network architectures, or including additional features in input vectors to allow for better training.

In Kahn's paper [9], a similar technique to Purucker was used to create a backpropagation neural network. Kahn made use of the same features as in Purucker's study but also included a home or away feature. Weeks 1 through 13 were used as the training set and weeks 14 and 15 were used as the test set. In order to improve on results, Kahn experimented with different learning rates, momentum, and network architecture to optimize the accuracy of the network on the training set. At test time, Kahn created two prediction sets; one contained season-to-date averages and one containing a three week average for each statistic. In the end, the season-to-date set beat the three week average set for both weeks 14 and 15, achieving an accuracy in both weeks of 75%. In these 2 weeks, ESPN experts achieved a correct prediction rate of 57% and 87% respectively. This suggests that the model is accurate and has the potential to compete with expert opinions. However, the repeatability of the model is questionable, as it was only tested on 2 weeks. Further tests on different weeks would help to prove the robustness of the model.

David, Pasteur, Ahmad, and Janning [10] developed a method of using a committee of machines to make NFL predictions. In this approach, multiple models were trained using different sets of the data and the top models were chosen. The top models were then combined to make a final prediction. Data collected was similar to Purucker and included scoring, passing and rushing yards, turnovers, and win-loss record. A home-field advantage parameter was also included in the data. To describe each team, season-to-date averages were computed and used in the input vectors. A weighted average of previous season statistics were used in the first 5 weeks of the season as not enough data was present in the season yet for accurate averages to be calculated. Using the 2008, 2009, and 2010 seasons for testing, the models were able to correctly predict between 61% and 67% each year. Once again, an arbitrary time period (this time one season) was used to compute estimations for future games. Areas of shortfalls were also addressed including the need for additional statistics and the need for a more robust estimation tool for early season game statistics.

Further literature outside of prediction models was reviewed to help inform on future strategies for optimizing prediction rates. Firstly, while current models show progress towards predicting the results of a game, most bets places involve betting against the spread. In a paper by Vergin [6], it was discovered that the betting market overreacts to recent positive performance but is unaffected by recent negative performance. This results in Vegas creating spreads that favour these recent positive performance teams. These results are important for the future works of the thesis as it proves that averaging statistics over the most recent games introduce bias to the models. This complements the findings of Kahn who witnessed decreased performance using only recent data.

Secondly, Hoffer and Pincin attempted to predict how valuable different players were towards influencing the Vegas line. According to their models, every one of the 32 players who were calculated to have some influence on the Vegas line were quarterback. Similar results were found in 2013 where 25 of 28 influential players were quarterbacks. These results show have important quarterbacks are to their team. This is useful information as it helps inform how this thesis should incorporate parameters to check if a team's first quarterback is actually playing. Models should be able to account for if a quarterback is injured, traded, or retires as these events have drastic consequences on the potential outcome of future games. The results also illustrate how other positional players may not be necessary to include in prediction models as their influence towards the Vegas line is minimal.

2.2 Model Goals

In order to assess our models, a measure of success is necessary. Statistically, due to Vegas book makers taking a standard 10% cut off the top of an against the spread bet, beating Vegas requires a correct prediction rate of 52.381% [11]. A model that achieves this rate would result in profit for a bettor. Therefore, our definition of a successful model is one that is able to beat Vegas by achieving a profitable correct prediction rate of 52.381%. In addition, the website `www.thepredictiontracker.com` contains information regarding the prediction rates of many other models. This list will be use as a comparison for how well our models perform relative to other pre-existing models.

3 Methods

3.1 Data Collection

In order to develop more robust models, a large dataset was required. The level of granularity was decided to be at a team level. Diving further into specific player statistics would involve an even larger data set that may prove to be unruly. Therefore, a new web scraper was developed from scratch in order to pull relevant team data from different NFL games.[1] The data was pulled from `pro-football-reference.com`, an online resource that contains detailed statistics from every NFL game throughout history. A list of the statistics compiled for each game is shown in Table 1. Data from every game since

[1] All scripts mentioned in this project, including the neural network and datasets used, can be found in the following GitHub repository: https://github.com/MLizzi/NFL-Prediction

1970 was collected and saved. These collection of games were chosen as they constitute what is colloquially known as the Super Bowl era as a result of the completion of the AFL-NFL merger in 1970.

Table 1: Statistics compiled using Python web scraper

Date	Net Pass Yards
Week	Net Rush Yards
Year	Net Total Yards
Season Number	Turnovers
Teams	Spread
Coaches	Over/Under
Roof	Time of Possession
Surface	Third Down Conversions
Starting Lineups	Fourth Down Conversions
First Downs	Record
Sacks	Penalties

Preprocessing on the data was required. Firstly, binary targets were assigned to each game to label them with the against the spread winner. A value of 1 denoted a profitable bet if placed for the favourite, while a value of 0 denoted a profitable bet if placed for the underdog. Games in which a tie (known as a push) occurred were excluded from the dataset as a bet for either the favourite or the underdog would have resulted in the bet being returned and therefore would not contribute to a bettor's profits or losses. Additionally, some games were excluded from the dataset when prior games could not be found as a result of a new team entering the league. Further preprocessing included the use of z-score normalization to ensure that all statistics were transformed into an appropriate range for the models.

Validation of the dataset was necessary to ensure that all required data was present and valid. Any missing statistics were found by contacting `pro-football-reference.com` directly as the data was often available, but had yet to be implemented on their website properly.

3.2 Dataset Creation

Firstly, the dataset used during training was built by taking data from 5 years of NFL seasons. This span of time was used to provide the model with an ample amount of data during training while also not being too large of a period such that older games started to become irrelevant. The data was split into a training and validation set at a 80/20 ratio, where games were shuffled prior to splitting.

Since game statistics of an unplayed game are unavailable to us when we want to deploy our model, a simple method using average statistics was used. Two datasets were built and used for all experiments. The first dataset, known as the independent location dataset, was built by taking an average of each statistic over a teams previous games. For example, a previous game count of 2 implies that the statistics used for the model (i.e. yards, turnovers, current wins, etc.) was the average from that team's past 2 games regardless of the location of the game being played. The second dataset, known as the location dependent dataset, used statistics from the home and away teams' respective past home or away games. For example, for a game with the Dallas Cowboys at home against the New England Patriots, the statistics used for Dallas were found by using averages over their past home games, and similarly New England's statistics were averaged over their last away games. For all experiments, these two types of datasets were built with a previous game count of 1, 2, 3, and 4 for a total of 8 different datasets. Also note that for games that are played at the beginning of a season, and therefore no previous games have been played yet, the time period spilled over into the last weeks of the previous season.

The data used from each game were the total yards, turnovers, current wins, rushing touchdowns, and passing touchdowns for each home and away team. The spread was also used but was calculated via the previously discussed averaging methods since the spread is an available statistic prior to a game being played. These 11 stats were concatenated in the described order into one vector and used as an input to the model as shown in Table 2. This particular input vector used the location dependent dataset with a previous game count of 4 and is shown prior to normalization.

Table 2: Example input vector prior to normalization

Home Stats					Away Stats					Spread
346.75	1.5	1	1.25	1.25	293.25	2.5	0	0.75	0.5	-5.5

3.3 Model Creation

A fully connected neural network was built using 2 hidden layers of size 160 and 40 respectively. A 10% dropout was used for both hidden layers as anything larger was too dramatic and caused model instability. ReLU activation functions were used at the end of each layer except for the output, where a sigmoid activation was used instead. Since only 2 classes existed for this problem, a binary cross entropy loss function was minimized over 1000 epochs using full batch gradient descent, with momentum 0.9, learning rate 0.0001, and an L2 regularization term with weight decay coefficient $\lambda = 0.01$.

These hyperparameters were selected based on experimentation and analysis of many different loss and accuracy curves. Initial experimentation involved determining the time period to use when training the models. Using a larger time period implied that models would be generated based on data from many years ago, which may not be a good indicator of future performance. Conversely, training on too small of a dataset would result in overfitting as not enough information was being provided to the network. These experiments involved training the models on a certain time period while attempting to tune the hyperparameters until convergence was found. If a good convergence could not be found, then the time period was increased to allow more data to be used and then the process was repeated.

As an example, a model was trained using the 2004 and 2005 season statistics, a 2 year time period. The following year of 2006 was used as the validation set. When looking at the loss curve of this model as seen in Figure 1, the training loss does decay over the training epochs as desired, but the validation loss immediately begins to diverge. This implies that the model was unable to generalize well and was overfitting the training data. To accommodate this, the time period was increased until viable results were found.

Eventually, better convergence was found using a 5 year time period. An example of a loss curve for this experiment is shown in Figure 2 and shows how both the training and validation loss curves converge over time. It is important to note that in this plot, both loss curves barely decay and sit near a value of $ln(2)$. This implies that the model is making near random predictions. When looking at the accuracy curves, both the training and validation sets improve in accuracy over time, but tend to sit near 50%, which is in agreement with their respective loss curves.

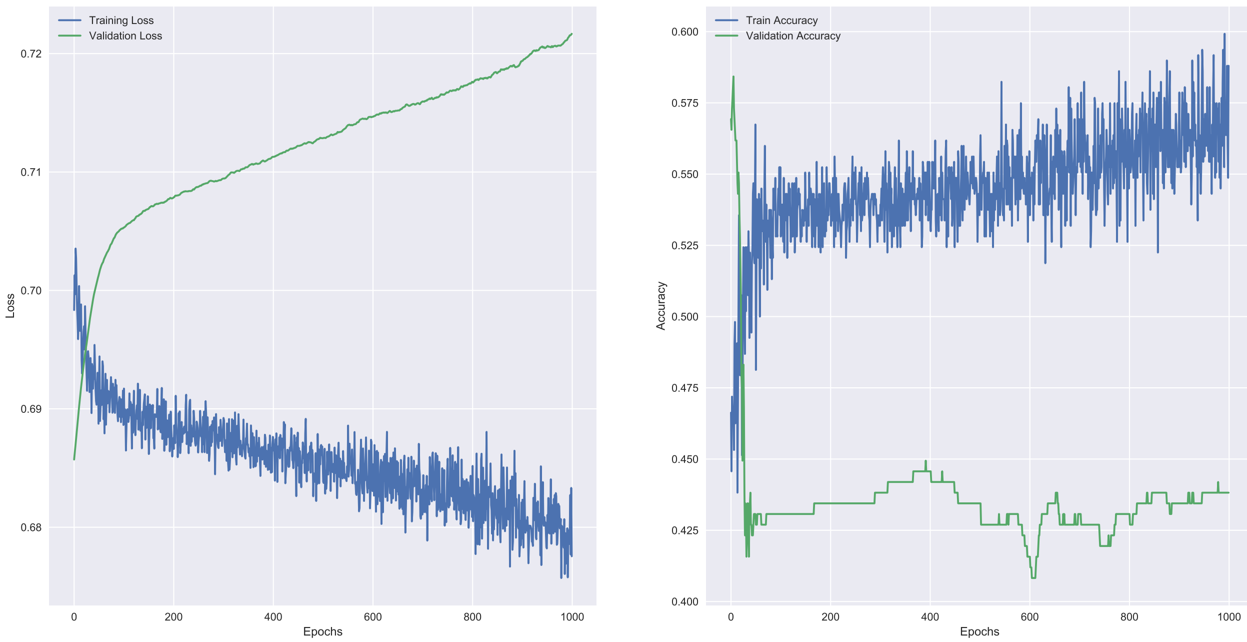

Figure 1: Loss (left) and accuracy (right) curves of a model using a 2 year period for the training data

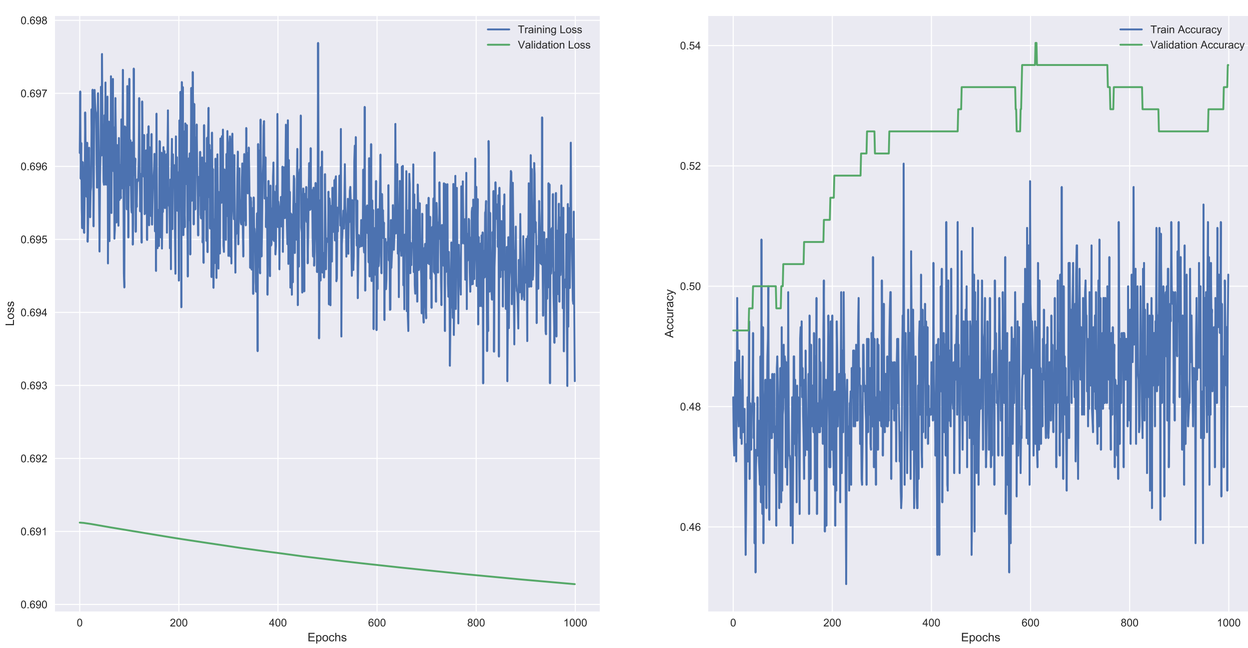

Figure 2: Loss (left) and accuracy (right) curves of a model using a 5 year period for the training data

3.4 Model Deployment

In order to account for the time series nature of the problem, a sliding window was used to generate multiple models over the years. This means that a model was trained on five year period, and then tested on the subsequent year. The process is outlined visually in Table 3. The first sliding window began by using 2001 to 2006, and was shifted one year at a time until 2014 to 2019. For every window used, 3 models were created. These models used the exact same hyperparameters, but a different seed was used and so each model was initialized differently. This helped remove any randomness or luck in the results. Further, this whole process was repeated using the independent and dependent datasets as mentioned. Therefore, since 14 windows were used, each with 3 different initializations, a total of 42 models were made for both the independent and dependent datasets. The results of the model deployment could then be compared to assess which dataset resulted in better prediction for against the spread betting.

Table 3: Example of the sliding window architecture used for training and testing models. Cells coloured in orange denote training years, while cells coloured in green denote the tested year. Actual experiments began in 2001 and continued until 2019.

Window #	Model Years									
1	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
2	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
3	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
4	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
5	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010

3.5 Model Assessment

In order to assess what the models learned, interpretability analysis was performed. This was done by perturbing one statistic at a time in the input vector and analyzing the resulting output. For every game, one statistic was perturbed but only for the team in which a successful bet would have been placed. As an example, in a game where the Baltimore Ravens covered the spread against the Indianapolis Colts, 100 yards was added to the total yards of Baltimore. This new data was then fed through the model. This process was run for every statistic in order to inform on how sensitive the model was to each particular statistic.

The perturbation values used for each statistic are found in Table 4. These values were designed based on personal knowledge of the sport. Each perturbation value is set such that it would be within reason if a team had an above average performance, but not too large such that the game would be considered an outlier. This also implies that each perturbation type is equally as difficult to achieve. For example, gaining 100 yards would be as difficult as reducing your turnovers by 1 or scoring 1 additional passing touchdown. These are all reasonable assumptions.

Table 4: Perturbed statistic values

Statistic	Value Added
Total Yards	100
Turnovers	-1
Rushing TDs	1
Passing TDs	1
Current Wins	1

4 Results

4.1 Profitability of Models

The results of every individual model for both datasets can be seen in Appendix A.1. Since profitability is our main metric for a successful model, the test accuracy of each model was compared to the goal prediction rate of 52.381% and labeled as a profitable or not profitable model. The summary of the independent and dependent models are shown in Table 5 and Table 6 respectively. The amount of times the models chose an underdog or favourite are also included in the Appendix to show how the model does not just choose the same class every time.

The results show that the location independent models tended to be more profitable than the dependent models. Across all previous game amounts used, the dependent models were profitable only 19.64% of the time while the independent models were profitable 23.81% of the time. However, neither dataset was able to make consistently profitable predictions, with a majority of models for both datasets being unprofitable. The average correct prediction rate also tended to be slightly higher in the independent models but the difference is marginal.

Table 5: Profitability of models using location independent data

Prev. Games	# of Profitable Models	# of Unprofitable Models	Average Correct Prediction Rate
1	10	32	49.71% ± 0.43%
2	12	30	50.23% ± 0.48%
3	7	35	49.93% ± 0.53%
4	11	31	50.14% ± 0.48%

Table 6: Profitability of models using location dependent data

Prev. Games	# of Profitable Models	# of Unprofitable Models	Average Prediction Rate
1	9	33	49.75% ± 0.44%
2	7	35	49.70% ± 0.43%
3	8	34	49.83% ± 0.43%
4	9	33	49.80% ± 0.51%

Further, there does not seem to be a strong correlation between the number of previous games used and which models are profitable. The models using 1, 2, 3, and 4 previous games were found to have 19, 19, 17, and 20 profitable models respectively. Using 3 previous games seems to have performed the worst out of all the models, but again the difference is marginal.

4.2 Interpretability Analysis

As discussed, interpretability analysis allows for the investigation of which statistics the network pays most attention to when making its prediction decisions. Every model created for both independent and dependent datasets were tested and the results are summarized in Table 7 and Table 8 respectively. In looking at the tables, the results for which statistics were most likely to cause the model to become profitable were the same for both datasets. Turnovers, rushing touchdowns, and passing touchdowns caused the largest positive shifts in model profitabilities. This suggests that an increase in these statistics in a team's previous games would result in a model being able to more accurately predict the against the spread winner since these are the statistics that the model focuses most on. Influencing the total yards of a team had detrimental impact on the model, causing it to make less profitable predictions. Lastly, influencing the current wins had varying degrees of success. The amount of profitable models either changed slightly in either direction, or often did not change at all.

In addition, these results again show very little correlation between the amount of previous games used and the profitability of a model. Also, the majority of predictions made by these models continue to be unprofitable.

Table 7: Influencing statistics in the independent dataset for interpretability analysis

Prev. Games	Pre-Influenced # of Profitable Models	Post-Influenced # of Profitable Models	Stat Influenced	Amount Influenced
1	10	6	Total Yards	100
2	12	5	Total Yards	100
3	7	7	Total Yards	100
4	11	7	Total Yards	100
1	10	10	Current Wins	1
2	12	12	Current Wins	1
3	7	7	Current Wins	1
4	11	11	Current Wins	1
1	10	11	Turnovers	-1
2	12	12	Turnovers	-1
3	7	15	Turnovers	-1
4	11	14	Turnovers	-1
1	10	12	Rush TDs	1
2	12	16	Rush TDs	1
3	7	15	Rush TDs	1
4	11	11	Rush TDs	1
1	10	17	Pass TDs	1
2	12	11	Pass TDs	1
3	7	13	Pass TDs	1
4	11	13	Pass TDs	1

Table 8: Influencing statistics in the dependent dataset for interpretability analysis

Prev. Games	Pre-Influenced # of Profitable Models	Post-Influenced # of Profitable Models	Stat Influenced	Amount Influenced
1	9	7	Total Yards	100
2	7	7	Total Yards	100
3	8	4	Total Yards	100
4	9	6	Total Yards	100
1	9	7	Current Wins	1
2	7	7	Current Wins	1
3	8	10	Current Wins	1
4	9	11	Current Wins	1
1	9	10	Turnovers	-1
2	7	14	Turnovers	-1
3	8	10	Turnovers	-1
4	9	16	Turnovers	-1
1	9	10	Rush TDs	1
2	7	11	Rush TDs	1
3	8	10	Rush TDs	1
4	9	13	Rush TDs	1
1	9	12	Pass TDs	1
2	7	11	Pass TDs	1
3	8	16	Pass TDs	1
4	9	12	Pass TDs	1

5 Discussion

5.1 Model Profitability

The test results show that models are unable to repeatedly beat the profitability margin over time. Deploying these models for future predictions would have no guarantee of making money. The best model accuracy was 58.02%, which is only slightly profitable. On average, the models only make correct predictions 50% of the time. This can be compared with other algorithms found on `www.thepredictiontracker.com`. For the 2019 season, the models in this paper would be found in the middle of the pack, around 27th out of 66 total models. While this is a respectable placement, a better placement closer to profitability would have been preferred.

5.2 Difficulty of Against the Spread Prediction

When training these models, the training curves do not converge to low loss values as desired. The loss converges to a value of $ln(2)$ implies that the model is making near random predictions. This result was repeatedly seen regardless of the tuning done and may be an indication that making against the spread predictions is much more challenging then initially assumed. However, it is important to verify that the failure of the model was not due to the method of implementation. For this reason, an additional experiment was run where the model would attempt to predict the outright winner instead of predicting the against the spread winner as done previously. By using the exact same model and hyperparameters, the results of this experiment would demonstrate if there was an issue with the model's ability to learn or if it was in fact an issue with predicting against the spread. The experiment used a previous game count of 4 and was trained on data from 2001 to 2005, with testing done on the 2006 season. The resulting loss and accuracy curves are shown in Figure 3.

In this experiment, the loss curve is much better than the ones studied previously. The loss continues to converge to lower values past the randomness value of $ln(2)$, which coincides with a much better validation accuracy. When applying this model against the test set, the test accuracy was found to be 54.68%. While this is less than the validation and training accuracies, this shows that the model is able to generalize the problem and come up with decent predictions. To further validate that the model can achieve good results, the hyperparameters were fine tuned to try and and obtain a higher validation accuracy. The new model was based on the previous model with the following changes:

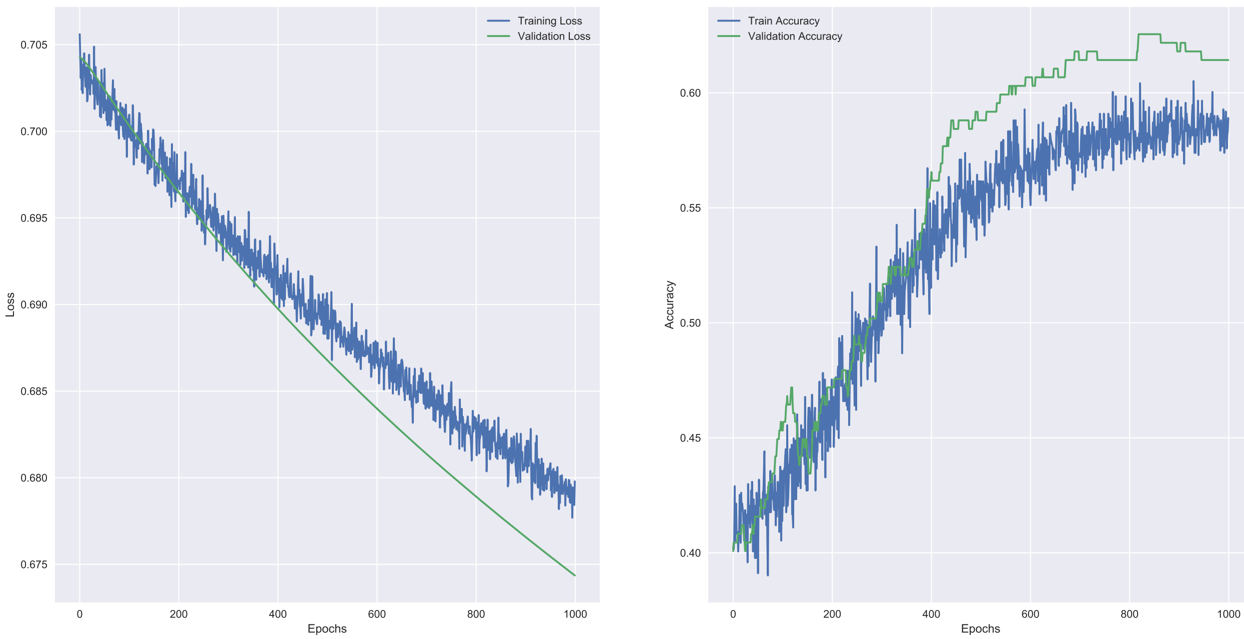

Figure 3: Loss (left) and accuracy (right) curves of a model attempting to predict the outright winner of a game. This model uses the exact same model hyperparameters as the models used in the against the spread predictions.

1. Learning rate changed from 0.0001 to 0.001
2. Weight decay coefficient changed from 0.01 to 0.001
3. Epochs trained over changed from 1000 to 2000

The results are shown in Figure 4. Here, the losses converge to an even lower value and the accuracies are much higher than the previous model without fine tuning. The test accuracy was found to be 72.66%. It is evident that with only a bit of fine tuning, predictions are quite good. These two experiments demonstrate that predicting the outright winner of a game is a much easier task than trying to predict against the spread. Therefore, our poor profitability results are likely not due to the model's ability to learn, but just exemplifies the difficulty in making against the spread predictions.

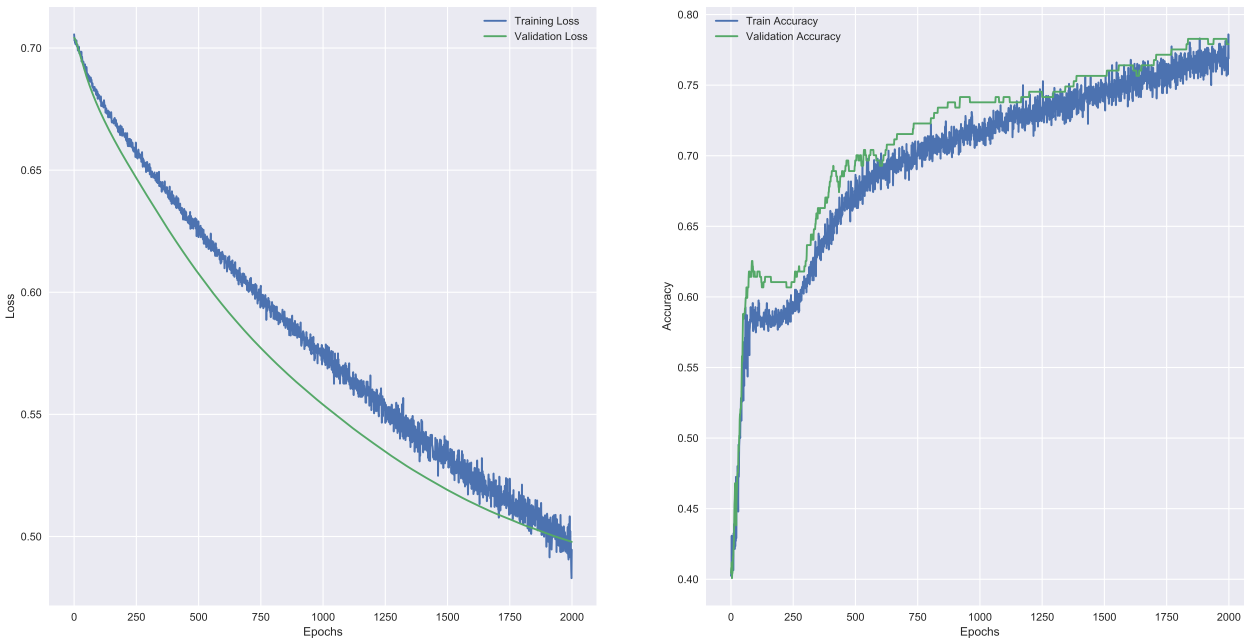

Figure 4: Loss (left) and accuracy (right) curves of a model attempting to predict the outright winner of a game. This model is tuned differently to try and increase the prediction accuracy of the outright winner.

5.3 Comparison of Datasets

The results show that the independent dataset always performed better than the dependent dataset. In this analysis, it is important to note that NFL games are played once a week. Therefore, for the independent dataset, it is known that the previous games used will always be the games directly before the game taking place. A previous game count of 4 implies that the oldest game used will be from 4 weeks ago. On the other hand, the dependent dataset requires a search in the past for similar location games. An example of when this may be problematic is when, for instance, a team plays 4 games at home, then 3 on the road If the same previous game count of 4 is used and the next game is at game, the the oldest statistics used would have been from 8 weeks, or nearly 2 months, prior to the game being tested. The results help support the idea that games played well in the past are less likely to influence the against the spread prediction than games played more recently. However, the independent dataset did not perform considerably better than the dependent dataset, and so this claim may require additional supporting evidence from future tests to verify its integrity.

5.4 Interpretability Results

The results of the interpretability analysis can help inform which team statistics are most important for against the spread prediction. Firstly, influencing turnovers, passing touchdowns, and rushing touchdowns resulted in much more profitable models. Both passing and rushing touchdowns are direct representations of the amount of points scored for a team in a game. It makes sense that the model focuses on these areas because it understands that if a team has historically been able to score more touchdowns than other team, they are more likely to score more points in future games and therefore cover more spreads too.

Reducing turnovers also increases the amount of profitable models, which suggest that prior success in taking the ball away from an opponent has a large affect on a team's ability to cover the spread. A team that gives up possession of the ball not only loses an opportunity at points, but also gives this opportunity right back to the opposing team. The models make the connection that keeping possession of the ball will result in a team scoring more points. By the logic discussed previously for the passing and rushing touchdowns, teams are then more likely to cover the spread.

Another very interesting result is that increasing the amount of yards for a team actually has detrimental impact on the amount of profitable models. It is important to recognize that this inverse relationship still implies that the model does look at total yards as an important statistic, but believes that having more yards is worse for covering the spread. It is hard to reason what the actual result should be. Having more yards often means that an offense is performing well. However, more yards does not directly translate into more points, as a team may have multiple 80 yard drives but fail to score on each one. Therefore, this inverse relationship suggests that yardage does play a factor, but does not necessarily imply that more yards are better.

Lastly, influencing the current wins statistic resulted in the amount of profitable models being relatively unchanged. While some models improved, others got worse. The changes in the amount of profitable models was so marginal that it suggests that the amount of wins coming into a game is less important than the other statistics. A team may be undefeated and have a game against a team who has yet to win. However, this is often accounted for when defining the spread value, where an undefeated team would have a much larger spread margin to make up when playing a bad team versus when playing an average team. This may help inform bettors to avoid believing that successful teams are more likely to beat substantially worse teams when it comes to against the spread betting. This also suggests that the value of the Vegas line is very well placed as it helps account for discrepancies between the skills of two teams.

6 Future Work

Plenty of continuing experiments could be run to further investigate how to improve the profitability of models. In this paper, the previous game count only went up to a maximum of 4. Additional tests could be performed by increasing this threshold to see if a correlation between previous games and prediction results can be found. For example, data could be used in a season-to-date format, where the data is averaged over all past games in a single season. Another experiment could incorporate additional features added to the input vectors to see if other statistics will aid in prediction. One potential addition could include encoding information about the starting quarterback. This may come in the form of a feature which holds the amount of cumulative games played by the quarterback up to that game. This allows the model to understand when a rookie or a veteran player is playing, which may help improve results. In order increase the robustness of these models, a new data architecture could be use that avoids taking averages over past games. When taking averages, important information pertaining to the order of previous games is lost. In this format, the input vector would be extended such that it contained all the information of past games separately. For example, a previous game count of 3 would result in an input vector concatenating the statistics of all 3 previous games into one vector that is now 3 times in length. This allows for the statistics from these past games to be passed to the network directly. Interpretability analysis could then be performed in hopes of further investigating if particular past games are more important for prediction than others.

7 Conclusion

The ability for machine learning models to make against the spread predictions continues to be a difficult problem to solve. The results demonstrate that even with advanced modeling techniques, Vegas bookmakers will often continue to come out on top in these types of bets. It can be concluded that betters should prefer to focus on games in a team's direct history, as team's prior performance in a certain location is likely not as relevant. Turnovers, rushing touchdowns, and passing touchdowns seem to be potential partial indicators for helping inform on against the spread predictions, but more experimentation and analysis must be performed to validate these results before being useful for a bettor. Also, while total yardage is often seen as an important statistic for the performance of a team, it is unfair to assume that gaining more yards will result in teams doing better against the spread. Lastly, while the method of choosing the spread is based on public action, these results illustrate how well Vegas sets the line to make this problem as difficult as possible to solve.

References

[1] *Gaming Revenue Information*. 2020. URL: https://gaming.nv.gov/index.aspx?page=149.

[2] "Americans Will Wager $6 Billion on Super Bowl LIII". In: *American Gaming Association* (Jan. 2019). URL: https://www.americangaming.org/resources/super-bowl-liii-wagering-estimates/.

[3] *How To Bet On NFL Games*. URL: https://www.oddsshark.com/nfl/how-to-bet.

[4] Roger C. Vergin and John J. Sosik. "No place like home: an examination of the home field advantage in gambling strategies in NFL football". In: *Journal of Economics and Business* 51.1 (Jan. 1999), pp. 21–31. DOI: 10.1016/s0148-6195(98)00025-3.

[5] Rory P. Bunker and Fadi Thabtah. "A machine learning framework for sport result prediction". In: *Applied Computing and Informatics* 15.1 (2019), pp. 27–33. ISSN: 2210-8327. DOI: https://doi.org/10.1016/j.aci.2017.09.005. URL: http://www.sciencedirect.com/science/article/pii/S2210832717301485.

[6] Roger C. Vergin. "Overreaction in the NFL point spread market". In: *Applied Financial Economics* 11.5 (2001), pp. 497–509. DOI: 10.1080/096031001752236780. URL: https://doi.org/10.1080/096031001752236780.

[7] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. "Supervised machine learning: A review of classification techniques". In: *Emerging artificial intelligence applications in computer engineering* 160.1 (2007), pp. 3–24.

[8] Michael C Purucker. "Neural network quarterbacking". In: *IEEE Potentials* 15.3 (Aug. 1996), pp. 9–15. DOI: 10.1109/45.535226.

[9] Joshua Kahn. "Neural Network Prediction of NFL Football Games". In: *World Wide Web Electronic Publication* (Jan. 2003).

[10] John A. David et al. "NFL Prediction using Committees of Artificial Neural Networks". In: *Journal of Quantitative Analysis in Sports* 7.2 (2011). DOI: 10.2202/1559-0410.1327. URL: https://doi.org/10.2202/1559-0410.1327.

[11] Gregory J Matthews. "Improved paired comparison models for NFL point spreads by data transformation". MA thesis. Worcester Polytechnic Institute, May 2005.

A Appendix

A.1 Test Accuracies

Table 9: Testing accuracies using previous games independent of location

Model Number	Training Years	Testing Years	Prev. Games	Testing Accuracy	Underdogs Chosen	Favourites Chosen	Seed
#402	2001-2005	2006	1	0.5344	177	85	0
#403	2002-2006	2007	1	0.4440	168	91	0
#404	2003-2007	2008	1	0.5287	168	93	0
#405	2004-2008	2009	1	0.5444	146	113	0
#406	2005-2009	2010	1	0.4580	150	112	0
#407	2006-2010	2011	1	0.5292	228	29	0
#408	2007-2011	2012	1	0.5057	207	56	0
#409	2008-2012	2013	1	0.5058	177	80	0
#410	2009-2013	2014	1	0.4981	240	23	0
#411	2010-2014	2015	1	0.4922	184	74	0
#412	2011-2015	2016	1	0.4867	187	76	0
#413	2012-2016	2017	1	0.4535	168	90	0
#414	2013-2017	2018	1	0.5135	152	107	0
#415	2014-2018	2019	1	0.4786	155	102	0
#416	2001-2005	2006	2	0.5344	127	135	0
#417	2002-2006	2007	2	0.4595	148	111	0
#418	2003-2007	2008	2	0.5096	157	104	0
#419	2004-2008	2009	2	0.5367	158	101	0
#420	2005-2009	2010	2	0.4389	145	117	0
#421	2006-2010	2011	2	0.5292	212	45	0
#422	2007-2011	2012	2	0.5133	215	48	0

#423	2008-2012	2013	2	0.5058	179	78	0
#424	2009-2013	2014	2	0.5057	234	29	0
#425	2010-2014	2015	2	0.4884	187	71	0
#426	2011-2015	2016	2	0.5171	211	52	0
#427	2012-2016	2017	2	0.4903	176	81	0
#428	2013-2017	2018	2	0.5328	185	74	0
#429	2014-2018	2019	2	0.4864	177	80	0
#430	2001-2005	2006	3	0.5802	153	109	0
#431	2002-2006	2007	3	0.4595	136	123	0
#432	2003-2007	2008	3	0.5019	157	104	0
#433	2004-2008	2009	3	0.5637	157	102	0
#434	2005-2009	2010	3	0.4313	141	121	0
#435	2006-2010	2011	3	0.5175	213	44	0
#436	2007-2011	2012	3	0.5019	216	47	0
#437	2008-2012	2013	3	0.5019	188	69	0
#438	2009-2013	2014	3	0.5133	238	25	0
#439	2010-2014	2015	3	0.5039	187	71	0
#440	2011-2015	2016	3	0.5171	213	50	0
#441	2012-2016	2017	3	0.4648	171	85	0
#442	2013-2017	2018	3	0.5483	145	114	0
#443	2014-2018	2019	3	0.5136	198	59	0
#444	2001-2005	2006	4	0.5191	127	135	0
#445	2002-2006	2007	4	0.4517	154	105	0
#446	2003-2007	2008	4	0.4981	160	101	0
#447	2004-2008	2009	4	0.5598	164	95	0
#448	2005-2009	2010	4	0.4618	151	111	0
#449	2006-2010	2011	4	0.5292	214	43	0

#450	2007-2011	2012	4	0.5133	223	40	0
#451	2008-2012	2013	4	0.5058	175	82	0
#452	2009-2013	2014	4	0.5133	232	31	0
#453	2010-2014	2015	4	0.5349	177	81	0
#454	2011-2015	2016	4	0.5095	211	52	0
#455	2012-2016	2017	4	0.5137	151	104	0
#456	2013-2017	2018	4	0.5405	161	98	0
#457	2014-2018	2019	4	0.4903	170	87	0
#571	2001-2005	2006	1	0.4580	5	257	1
#572	2002-2006	2007	1	0.5058	16	243	1
#573	2003-2007	2008	1	0.4828	2	259	1
#574	2004-2008	2009	1	0.4942	7	252	1
#575	2005-2009	2010	1	0.4924	7	255	1
#576	2006-2010	2011	1	0.4630	5	252	1
#577	2007-2011	2012	1	0.4601	5	258	1
#578	2008-2012	2013	1	0.5214	19	238	1
#579	2009-2013	2014	1	0.5057	12	251	1
#580	2010-2014	2015	1	0.4767	28	230	1
#581	2011-2015	2016	1	0.5057	16	247	1
#582	2012-2016	2017	1	0.5349	9	249	1
#583	2013-2017	2018	1	0.4710	13	246	1
#584	2014-2018	2019	1	0.4319	37	220	1
#585	2001-2005	2006	2	0.4389	6	256	1
#586	2002-2006	2007	2	0.5290	8	251	1
#587	2003-2007	2008	2	0.4866	5	256	1
#588	2004-2008	2009	2	0.4942	9	250	1
#589	2005-2009	2010	2	0.4885	6	256	1

#590	2006-2010	2011	2	0.4591	16	241	1
#591	2007-2011	2012	2	0.4753	7	256	1
#592	2008-2012	2013	2	0.5214	17	240	1
#593	2009-2013	2014	2	0.5019	11	252	1
#594	2010-2014	2015	2	0.4961	23	235	1
#595	2011-2015	2016	2	0.5057	16	247	1
#596	2012-2016	2017	2	0.5331	15	242	1
#597	2013-2017	2018	2	0.4556	13	246	1
#598	2014-2018	2019	2	0.4280	24	233	1
#599	2001-2005	2006	3	0.4466	4	258	1
#600	2002-2006	2007	3	0.5290	28	231	1
#601	2003-2007	2008	3	0.4789	5	256	1
#602	2004-2008	2009	3	0.4749	12	247	1
#603	2005-2009	2010	3	0.4924	7	255	1
#604	2006-2010	2011	3	0.4436	20	237	1
#605	2007-2011	2012	3	0.4753	7	256	1
#606	2008-2012	2013	3	0.5292	17	240	1
#607	2009-2013	2014	3	0.4943	7	256	1
#608	2010-2014	2015	3	0.4845	22	236	1
#609	2011-2015	2016	3	0.5095	13	250	1
#610	2012-2016	2017	3	0.5156	20	236	1
#611	2013-2017	2018	3	0.4556	17	242	1
#612	2014-2018	2019	3	0.4280	30	227	1
#613	2001-2005	2006	4	0.4351	1	261	1
#614	2002-2006	2007	4	0.5097	11	248	1
#615	2003-2007	2008	4	0.4789	5	256	1
#616	2004-2008	2009	4	0.4903	8	251	1

#617	2005-2009	2010	4	0.4885	4	258	1
#618	2006-2010	2011	4	0.4514	16	241	1
#619	2007-2011	2012	4	0.4677	11	252	1
#620	2008-2012	2013	4	0.5097	18	239	1
#621	2009-2013	2014	4	0.4905	10	253	1
#622	2010-2014	2015	4	0.4922	22	236	1
#623	2011-2015	2016	4	0.5323	17	246	1
#624	2012-2016	2017	4	0.5216	15	240	1
#625	2013-2017	2018	4	0.4402	15	244	1
#626	2014-2018	2019	4	0.4319	23	234	1
#683	2001-2005	2006	1	0.5267	137	125	2
#684	2002-2006	2007	1	0.4710	211	48	2
#685	2003-2007	2008	1	0.4904	144	117	2
#686	2004-2008	2009	1	0.5097	127	132	2
#687	2005-2009	2010	1	0.5267	172	90	2
#688	2006-2010	2011	1	0.5175	221	36	2
#689	2007-2011	2012	1	0.5247	212	51	2
#690	2008-2012	2013	1	0.4825	207	50	2
#691	2009-2013	2014	1	0.5285	196	67	2
#692	2010-2014	2015	1	0.5000	198	60	2
#693	2011-2015	2016	1	0.4829	208	55	2
#694	2012-2016	2017	1	0.4884	219	39	2
#695	2013-2017	2018	1	0.5135	132	127	2
#696	2014-2018	2019	1	0.5409	163	94	2
#697	2001-2005	2006	2	0.5229	206	56	2
#698	2002-2006	2007	2	0.4595	190	69	2
#699	2003-2007	2008	2	0.5402	139	122	2

#700	2004-2008	2009	2	0.4903	128	131	2
#701	2005-2009	2010	2	0.5534	181	81	2
#702	2006-2010	2011	2	0.5253	223	34	2
#703	2007-2011	2012	2	0.5209	223	40	2
#704	2008-2012	2013	2	0.4864	210	47	2
#705	2009-2013	2014	2	0.5285	210	53	2
#706	2010-2014	2015	2	0.4961	211	47	2
#707	2011-2015	2016	2	0.5095	217	46	2
#708	2012-2016	2017	2	0.5058	216	41	2
#709	2013-2017	2018	2	0.5521	148	111	2
#710	2014-2018	2019	2	0.5447	174	83	2
#711	2001-2005	2006	3	0.5115	199	63	2
#712	2002-2006	2007	3	0.4247	191	68	2
#713	2003-2007	2008	3	0.5096	127	134	2
#714	2004-2008	2009	3	0.5174	141	118	2
#715	2005-2009	2010	3	0.5153	179	83	2
#716	2006-2010	2011	3	0.5175	229	28	2
#717	2007-2011	2012	3	0.5171	228	35	2
#718	2008-2012	2013	3	0.4903	203	54	2
#719	2009-2013	2014	3	0.4981	204	59	2
#720	2010-2014	2015	3	0.5000	198	60	2
#721	2011-2015	2016	3	0.5133	220	43	2
#722	2012-2016	2017	3	0.5039	209	47	2
#723	2013-2017	2018	3	0.5521	118	141	2
#724	2014-2018	2019	3	0.5253	159	98	2
#725	2001-2005	2006	4	0.4656	181	81	2
#726	2002-2006	2007	4	0.5251	153	106	2

#727	2003-2007	2008	4	0.5249	129	132	2
#728	2004-2008	2009	4	0.5290	150	109	2
#729	2005-2009	2010	4	0.5115	168	94	2
#730	2006-2010	2011	4	0.5292	220	37	2
#731	2007-2011	2012	4	0.5171	224	39	2
#732	2008-2012	2013	4	0.5058	203	54	2
#733	2009-2013	2014	4	0.4677	196	67	2
#734	2010-2014	2015	4	0.5388	206	52	2
#735	2011-2015	2016	4	0.5133	216	47	2
#736	2012-2016	2017	4	0.4902	209	46	2
#737	2013-2017	2018	4	0.5367	132	127	2
#738	2014-2018	2019	4	0.5214	172	85	2

Table 10: Testing accuracies using previous games dependent on location

Model Number	Training Years	Testing Years	Prev. Games	Testing Accuracy	Underdogs Chosen	Favourites Chosen	Seed
#346	2001-2005	2006	1	0.5038	121	141	0
#347	2002-2006	2007	1	0.4324	129	130	0
#348	2003-2007	2008	1	0.5249	155	106	0
#349	2004-2008	2009	1	0.5135	162	97	0
#350	2005-2009	2010	1	0.4847	133	129	0
#351	2006-2010	2011	1	0.5058	224	33	0
#352	2007-2011	2012	1	0.5057	213	50	0
#353	2008-2012	2013	1	0.5136	181	76	0
#354	2009-2013	2014	1	0.4791	229	34	0
#355	2010-2014	2015	1	0.4884	181	77	0
#356	2011-2015	2016	1	0.4905	204	59	0
#357	2012-2016	2017	1	0.4825	159	98	0
#358	2013-2017	2018	1	0.5290	166	93	0
#359	2014-2018	2019	1	0.5331	167	90	0
#360	2001-2005	2006	2	0.4695	122	140	0
#361	2002-2006	2007	2	0.4595	150	109	0
#362	2003-2007	2008	2	0.4904	162	99	0
#363	2004-2008	2009	2	0.5212	184	75	0
#364	2005-2009	2010	2	0.4962	172	90	0
#365	2006-2010	2011	2	0.5136	230	27	0
#366	2007-2011	2012	2	0.4981	217	46	0
#367	2008-2012	2013	2	0.5019	182	75	0
#368	2009-2013	2014	2	0.4981	232	31	0
#369	2010-2014	2015	2	0.5233	172	86	0

#370	2011-2015	2016	2	0.5057	202	61	0
#371	2012-2016	2017	2	0.5255	156	99	0
#372	2013-2017	2018	2	0.5328	157	102	0
#373	2014-2018	2019	2	0.5136	178	79	0
#374	2001-2005	2006	3	0.4771	122	140	0
#375	2002-2006	2007	3	0.4402	131	128	0
#376	2003-2007	2008	3	0.5441	164	97	0
#377	2004-2008	2009	3	0.4981	180	79	0
#378	2005-2009	2010	3	0.5000	167	95	0
#379	2006-2010	2011	3	0.4981	236	21	0
#380	2007-2011	2012	3	0.5133	211	52	0
#381	2008-2012	2013	3	0.4903	185	72	0
#382	2009-2013	2014	3	0.5171	239	24	0
#383	2010-2014	2015	3	0.5388	174	84	0
#384	2011-2015	2016	3	0.5057	212	51	0
#385	2012-2016	2017	3	0.4901	149	104	0
#386	2013-2017	2018	3	0.5097	173	86	0
#387	2014-2018	2019	3	0.5214	190	67	0
#388	2001-2005	2006	4	0.5038	171	91	0
#389	2002-2006	2007	4	0.4402	147	112	0
#390	2003-2007	2008	4	0.5441	150	111	0
#391	2004-2008	2009	4	0.4556	179	80	0
#392	2005-2009	2010	4	0.5191	174	88	0
#393	2006-2010	2011	4	0.4669	226	31	0
#394	2007-2011	2012	4	0.5323	220	43	0
#395	2008-2012	2013	4	0.4903	185	72	0
#396	2009-2013	2014	4	0.5133	236	27	0

#397	2010-2014	2015	4	0.4651	161	97	0
#398	2011-2015	2016	4	0.5133	208	55	0
#399	2012-2016	2017	4	0.4940	151	100	0
#400	2013-2017	2018	4	0.5097	161	98	0
#401	2014-2018	2019	4	0.5214	176	81	0
#515	2001-2005	2006	1	0.4389	8	254	1
#516	2002-2006	2007	1	0.5367	10	249	1
#517	2003-2007	2008	1	0.4943	5	256	1
#518	2004-2008	2009	1	0.4903	4	255	1
#519	2005-2009	2010	1	0.4771	9	253	1
#520	2006-2010	2011	1	0.4864	13	244	1
#521	2007-2011	2012	1	0.4829	9	254	1
#522	2008-2012	2013	1	0.5136	19	238	1
#523	2009-2013	2014	1	0.4943	15	248	1
#524	2010-2014	2015	1	0.5039	27	231	1
#525	2011-2015	2016	1	0.5019	19	244	1
#526	2012-2016	2017	1	0.5331	14	243	1
#527	2013-2017	2018	1	0.4556	11	248	1
#528	2014-2018	2019	1	0.4280	22	235	1
#529	2001-2005	2006	2	0.4313	6	256	1
#530	2002-2006	2007	2	0.5290	10	249	1
#531	2003-2007	2008	2	0.4866	5	256	1
#532	2004-2008	2009	2	0.4788	13	246	1
#533	2005-2009	2010	2	0.4695	11	251	1
#534	2006-2010	2011	2	0.4669	20	237	1
#535	2007-2011	2012	2	0.4677	11	252	1
#536	2008-2012	2013	2	0.5175	18	239	1

#537	2009-2013	2014	2	0.4829	12	251	1
#538	2010-2014	2015	2	0.4922	28	230	1
#539	2011-2015	2016	2	0.5057	14	249	1
#540	2012-2016	2017	2	0.5176	18	237	1
#541	2013-2017	2018	2	0.4672	10	249	1
#542	2014-2018	2019	2	0.4436	26	231	1
#543	2001-2005	2006	3	0.4313	8	254	1
#544	2002-2006	2007	3	0.5212	10	249	1
#545	2003-2007	2008	3	0.4828	2	259	1
#546	2004-2008	2009	3	0.4672	10	249	1
#547	2005-2009	2010	3	0.4771	9	253	1
#548	2006-2010	2011	3	0.4514	16	241	1
#549	2007-2011	2012	3	0.4563	12	251	1
#550	2008-2012	2013	3	0.5175	20	237	1
#551	2009-2013	2014	3	0.4981	18	245	1
#552	2010-2014	2015	3	0.4806	21	237	1
#553	2011-2015	2016	3	0.5019	15	248	1
#554	2012-2016	2017	3	0.5336	22	231	1
#555	2013-2017	2018	3	0.4595	8	251	1
#556	2014-2018	2019	3	0.4514	32	225	1
#557	2001-2005	2006	4	0.4389	4	258	1
#558	2002-2006	2007	4	0.5367	22	237	1
#559	2003-2007	2008	4	0.4866	1	260	1
#560	2004-2008	2009	4	0.4942	15	244	1
#561	2005-2009	2010	4	0.4809	6	256	1
#562	2006-2010	2011	4	0.4514	20	237	1
#563	2007-2011	2012	4	0.4525	15	248	1

#564	2008-2012	2013	4	0.5097	26	231	1
#565	2009-2013	2014	4	0.5133	16	247	1
#566	2010-2014	2015	4	0.4961	25	233	1
#567	2011-2015	2016	4	0.5133	20	243	1
#568	2012-2016	2017	4	0.5538	24	227	1
#569	2013-2017	2018	4	0.4479	7	252	1
#570	2014-2018	2019	4	0.4514	22	235	1
#627	2001-2005	2006	1	0.4656	187	75	2
#628	2002-2006	2007	1	0.4942	181	78	2
#629	2003-2007	2008	1	0.4943	147	114	2
#630	2004-2008	2009	1	0.5521	136	123	2
#631	2005-2009	2010	1	0.4733	170	92	2
#632	2006-2010	2011	1	0.5097	225	32	2
#633	2007-2011	2012	1	0.5437	217	46	2
#634	2008-2012	2013	1	0.4981	201	56	2
#635	2009-2013	2014	1	0.4867	203	60	2
#636	2010-2014	2015	1	0.5426	199	59	2
#637	2011-2015	2016	1	0.5285	224	39	2
#638	2012-2016	2017	1	0.4942	216	41	2
#639	2013-2017	2018	1	0.4749	136	123	2
#640	2014-2018	2019	1	0.5136	168	89	2
#641	2001-2005	2006	2	0.4389	152	110	2
#642	2002-2006	2007	2	0.5174	149	110	2
#643	2003-2007	2008	2	0.4981	142	119	2
#644	2004-2008	2009	2	0.5328	149	110	2
#645	2005-2009	2010	2	0.4962	176	86	2
#646	2006-2010	2011	2	0.5097	227	30	2

#647	2007-2011	2012	2	0.4905	217	46	2
#648	2008-2012	2013	2	0.4942	192	65	2
#649	2009-2013	2014	2	0.5285	202	61	2
#650	2010-2014	2015	2	0.5504	203	55	2
#651	2011-2015	2016	2	0.5551	217	46	2
#652	2012-2016	2017	2	0.5059	207	48	2
#653	2013-2017	2018	2	0.4672	108	151	2
#654	2014-2018	2019	2	0.4825	154	103	2
#655	2001-2005	2006	3	0.4695	158	104	2
#656	2002-2006	2007	3	0.4942	183	76	2
#657	2003-2007	2008	3	0.5249	143	118	2
#658	2004-2008	2009	3	0.5174	147	112	2
#659	2005-2009	2010	3	0.5305	175	87	2
#660	2006-2010	2011	3	0.5370	230	27	2
#661	2007-2011	2012	3	0.4867	222	41	2
#662	2008-2012	2013	3	0.5019	198	59	2
#663	2009-2013	2014	3	0.5019	203	60	2
#664	2010-2014	2015	3	0.5155	206	52	2
#665	2011-2015	2016	3	0.5019	217	46	2
#666	2012-2016	2017	3	0.5020	210	43	2
#667	2013-2017	2018	3	0.5405	93	166	2
#668	2014-2018	2019	3	0.5292	162	95	2
#669	2001-2005	2006	4	0.4542	142	120	2
#670	2002-2006	2007	4	0.4981	134	125	2
#671	2003-2007	2008	4	0.5249	155	106	2
#672	2004-2008	2009	4	0.5058	146	113	2
#673	2005-2009	2010	4	0.5420	184	78	2

#674	2006-2010	2011	4	0.5136	234	23	2
#675	2007-2011	2012	4	0.4753	229	34	2
#676	2008-2012	2013	4	0.5097	214	43	2
#677	2009-2013	2014	4	0.4715	203	60	2
#678	2010-2014	2015	4	0.5155	200	58	2
#679	2011-2015	2016	4	0.5323	217	46	2
#680	2012-2016	2017	4	0.4781	205	46	2
#681	2013-2017	2018	4	0.5637	103	156	2
#682	2014-2018	2019	4	0.5370	160	97	2