BETTING ON CERTAINTY: ARBITRAGE ON THE PREMIER LEAGUE

DATA 221: Introduction to Machine Learning
Alec Chen & Emily Tong

I. INTRODUCTION: SOURCING, DATA CLEANING, AND LITERATURE SEARCH

The goal of this project was to produce a set of tools to help in *Arbitrage Betting*, a process of ensuring profits through exploiting uneven pricing between odds-providers (e.g. DraftKings, Bet365)[1]. In particular, we chose Soccer, as it is commonplace for books to provide *draw* outcome money-lines, which allows for a 'true' certainty[2]. Our primary approach was to predict possible arbitrage profit margins, both in categorical, and regressive forms, on a given game given a variety of factors, which differed model-to-model. Whereas substantial attention has been paid to determining arbitrage-friendly betting markets, fewer forays have been conducted into predicting Arbitrage opportunities based on *temporal preconditions*; is it possible to predict based on previous data when books are more likely to price outcomes badly relative to one another?

Whereas betting is typically defined on its risk, Arbitrage Betting is defined on its utter lack thereof. Sportsbooks derive their odds through calculating their own internal predictions for game outcomes, on top of a healthy margin, which ensures the book is more likely to make money than lose it. However, methods with which different books set their odds are anything but uniform. Thus, certain opportunities to *arbitrage* odds arise; in rare cases, certain permutations of books' listed odds will yield implied win percentages where, relative to other sportsbooks, a given sportsbook will assume a given outcome is overly likely. In these special scenarios, the smart bettor will stake money on all three outcomes, guaranteeing that their losses will be entirely mitigated by their wins.

$$Total\ Arbitrage\ Percentage = \frac{100}{Home\ Odds} + \frac{100}{Draw\ Odds} + \frac{100}{Away\ Odds}$$

$$Profit = \frac{Stake}{Total\ Arb.\,\%} - Stake$$

Relevant Formulae; cr: Franck, Verbeek, Nüsche: "Inter-market Arbitrage in Betting"

Using this definition, we flagged games with arbitrage opportunities as was done in analogous previous studies, such as a 2019 analysis of Super Rugby data from oddsportal.com[3]. This intra-market "*Super Rugby"* study found abundant arbitrage opportunities, with an average return of 1.94%. This was far lower than what was found in another 2013 study of inter-market arbitrage with sampled odds from European soccer leagues[4]. Beyond these yield differences, the studies were methodologically different; the rugby study was focused on intra-market discrepancies, and the soccer study involved multiple markets.

Super Rugby's identification of copious arbitrage opportunities, despite their single-market approach brought into question for us whether one really needed an abundance of odds across diverse bookmakers to actually identify arbitrage opportunities. However, the *Super Rugby* authors themselves acknowledged in their concluding remarks, that extrapolating their results to betting in other sports might not be that useful; the nature of the sport, its scoring, and average bettor may contribute enough difference such that a more sport-by-sport *ad hoc* approach might prove more effective[5].

---

[1]Franck, Verbeek, Nuesch. "Inter-market Arbitrage in Betting", *Economica*, 318 (2012); 300-325, https://onlinelibrary.wiley.com/doi/full/10.1111/ecca.12009.
[2] Insofar as betting in most sports does not allow for staking on "draw" outcomes, e.g. Basketball, Football etc.
[3] Matthew Buckle and Chun-Sung Huang. "The Efficiency of Sport Betting Markets: An Analysis Using Arbitrage Trading within Super Rugby." *International Journal of Sport Finance*, 3 (2018); 279+,,
https://go.gale.com/ps/i.do?id=GALE%7CA568257897&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=15586235&p=AONE&sw=w&userGroupName=anon%7Ee4d79edc.
[4] Franck, 2013
[5]Buckle and Huang, "The Efficiency of Sport Betting Markets", pg. 279+

Following a thorough search of usable datasets, we landed upon the website "*football-data.co.uk*," with relevant odds data spanning three decades: 1992-2023[6]. This website provides not only historical football results and betting odds, but also live scores, and game data (shots taken, save percentage, et cetera). Alec, responsible for data formatting and selection, opted to limit our usage to 2008-2023, as many statistics were gathered only over certain time intervals. Additionally, both labeling and formatting were very inconsistent between years, rendering many columns impractical to parse through. Nevertheless, this temporal scope was wider than what was used in previous literature on the same dataset[7], though we opted to focus exclusively on the Premier League as opposed to all European leagues.

While manually searching through column data, Alec found that almost all betting-related columns were either missing large time spans of data, or were only gathered for a short period of time. Faced with no consistency between listed books year-on-year (likely due to certain books actually joining and leaving the markets at different times), he opted to simply search for the highest decimal odds for each of the three outcomes, rather than dealing with specific books; to find a given game's *best* arbitrage opportunity, one need only look at the highest odds. Another set of very consistent data were the game statistics (Game Attendance, Shots on Goal, Fouls etc), so it could be included. Following this severe column pruning, there was only one row which registered as NaN; the culprit turned out to be an entire empty row in the 2013 dataset. After this, he converted date, results, and other non-numeric data into float values, i.e. the "Date" column was converted to an Ordinal format. To build a row of training data for a given game, the game's statistics and odds were dropped, and data from the most recent past 5 games that each *respective team participated in* was appended to the end. While it is possible that there could be cross-contamination (in the case of the neural network) between training and testing data, where the neural network encodes information about specific games, the size renders this more unlikely.

The inclusion of 13 books in the Premier League Football Data was promising firstly for its variety of bookmakers (Bet365, Blue Square, Bet&Win, Gamebookers, Ladbrokes…), which allowed for an intermarket approach, similar to the Franck Soccer study. Additionally, the inclusion of match statistics paved a route towards insights along different data axes. For instance, perhaps the number of yellow and red cards received or the number of goals scored by the top scorers could prove useful in predicting arbitrage opportunities.

Past studies on inefficiencies in the sports betting market, as identified through arbitrage opportunities, utilized bivariate Poisson distribution models, Bayesian network models, Random Forest, Boosting, and Support Vector Machines[8]. In this study, we opted to create Logistic Regression and Neural Network models to predict arbitrage.

II. Exploring The Data

Before explicitly modeling, Emily spent some time exploring the data, separated by odds and 'soccer data' in order to find some interesting trends and phenomena.

Taking an average of all books for every game yielded 2.8 decimal odds for the home team winning, 4.01 for a draw, and 4.76 for away. Since reciprocal odds can be defined as the implied probability of winning, assuming fair odds, home teams are in aggregate favored to win before a game begins since it has the lowest average odds. Looking at the average odds within seasons, this trend holds year-on-year; away odds are always highest; home always lowest. *(See Fig. 1)*

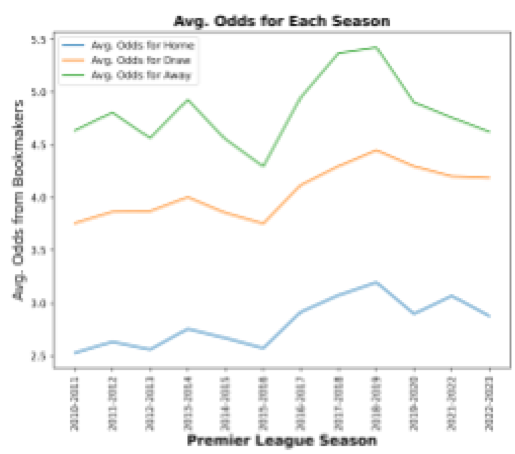

Fig. 1: The average odds across all bookmakers over time for the home team winning, a draw, or the away team winning.

[6] Click *here* for all CSVs, and *here* for the website. Data Sources: BBC, ESPN Soccer, XScores, and Sports.com (deprecated, 2002).
"Betting odds for weekend games are collected Friday afternoons, and on Tuesday afternoons for midweek games."
Click *here* for a full Data Dictionary of columns and their respective labels.

[7] Constantinou, Anthony Costa, and Normal Elliot Fenton. "Profiting from arbitrage and odds biases of the European football gambling market." *The Journal of Gambling Business and Economics* 7, no. 2 (2013): 41-70; Johannes Stubinger and Julian Knoll. "Beat the Bookmaker - Winning Football Bets with Machine Learning (Best Application Paper)". *International Conference on Innovative Techniques and Applications of Artificial Intelligence* (2018): 219-223.

[8] Ibid.

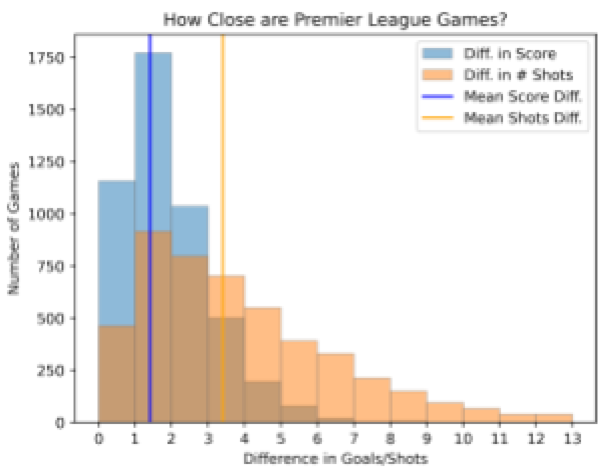

Fig. 2: Visualization of the distribution of difference in goals in all Premier League games, as well as the distribution of the difference in shots on goal in all games.

Considering the game itself: relative to other popular betting sports like Basketball or American Football, Soccer is very low scoring. In Soccer, games are often close, with the average score differential being 1.4 goals, and the average shots on goal differential being 3.4. This makes the outcome of games even more uncertain, since scores are relatively discrete, outcomes could be— interpreted as being more 'random,' as allowing a single goal could differentiate between a win and a loss. This highly 'risky' betting environment accentuates the applicability of Arbitrage, as we might extrapolate that more random outcomes could result in larger pricing differentials. Furthermore, in comparing soccer players and teams, there are many more team-based differentiating X-factors, such as team synergy (think the Argentina-Saudi Arabia World Cup upset, or Japan-Germany). Therefore, directly predicting game outcomes might prove difficult. The upside though, is that it is likely *also* very hard for books, so given accessible statistics, predicting book pricing may prove more feasible. In addition, the goal of understanding the relationship between team performance and the presence of arbitrage opportunities is also constrained by the limitations of statistical variables as proxies for the abstract factors they represent. *(See Fig. 2)*

	Interval of Profit Margin	Number of Games
0	[-inf, -2%]	3110
1	[-2, -0%]	1719
2	[+0%, +2%]	529
3	[+2%, +inf]	51

Figure 3: Class Distribution; multi-class classification

Perhaps most importantly, however, there were only 580 opportunities for arbitrage of non-zero profit, out of 5409 game instances. This heavy class imbalance is important to acknowledge, as it renders raw accuracy somewhat less useful. Additionally, despite having thousands of rows of training data, it might still not be enough, especially as there were only around *50* opportunities of the 580 where profit was above 2%. Due to this, for many of the models, we opted to measure performance with *Balanced Accuracy*, a metric which returns the *average accuracy per class, weighting each accuracy equally despite frequency differences.* Therefore, we set a Balanced Score of ***25%*** as a threshold for the *four-category classification* and ***50%*** for *the binary classification. (See Fig. 3)*

III. NAÏVE MODELING APPROACH: BINARY LOGISTIC REGRESSION

Emily applied Logistic Regression to tackle the problem as both a binary and mutli-class classification problem. (Binary defined as positive or negative yields; See Fig. 3 for multi-class definitions). Splitting the data into equally-sized training and testing sets with a reproducible random state, Emily created multiple Logistic Regression models using various input features to tackle the problem as both types of problems.

INPUT & OUTPUT FOR BINARY LOGISTIC REGRESSION

X	*Using 5 Prev. Games' Data:* *Current Date, Teams Playing; Betting Odds & Game Statistics for each teams' previous five games (10 previous games total, all X data standard scaled)* *Past Season's Standings*: Everything in 'Using 5 Prev. Games' Data' as well as number of games played, wins, draws, losses, goals for, goals against, goal difference, and points from the season before (all X data standard scaled)
y	*'1' corresponds to arbitrage opportunity (Profit Margin > 0), '0' to no arbitrage opportunity*

The initial Binary Classification Model using game statistics for each team's previous five games yielded an accuracy of approximately *89.36%*. Emily also produced a Receiver Operating Characteristic Curve (ROC Curve) for the model, shown in blue below. The Area Under Curve (AUC) below the blue line is 0.609. An AUC of 0.50 would represent a 'pure random' scenario, as depicted by the gray line. The additional area between the blue and gray lines shows the 'better than random' performance of the model. However, she found a balanced accuracy score of 50.4%; given this result, the model is likely *not meaningfully classifying at all*. Emily theorized that because the data was incredibly imbalanced, it may be useful to lower the probability threshold to reflect the actual distribution of different *y* results. However, in recreating this model, one should adjust this value to that which reflects their idiosyncratic preference for false arbitrage opportunities versus false non-arbitrage instances.

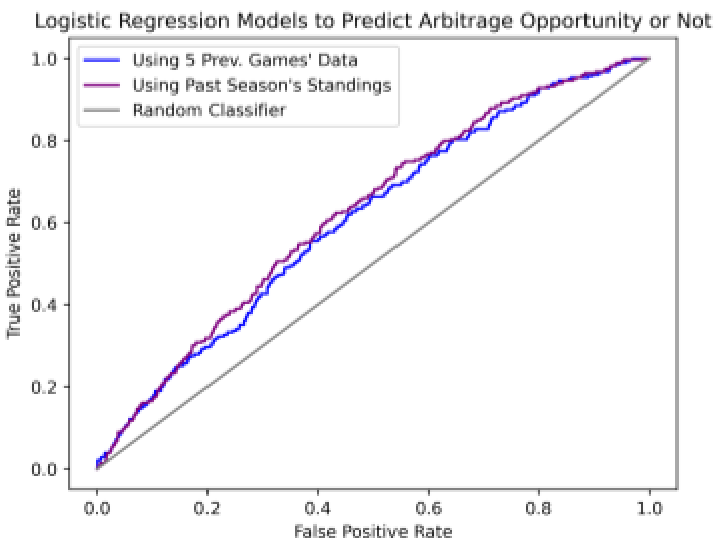

Fig. 4: Receiver Operating Characteristic Curves for two Logistic Regression models tackling the binary classification problem of predicting whether or not there will be a positive profit margin.

In this case, however, Emily aimed to maximize *balanced accuracy*. To that end, she chose a threshold that reflected the distribution of arbitrage opportunities in the dataset. 10.72% of observations in the dataset were labeled as true instances of arbitrage opportunities. A probability threshold of 0.107 led to an increased balanced accuracy of 58.04%. In other words, decreasing the probability necessary to conclude a given sample *was* an arbitrage opportunity to reflect the actual distribution of results maximizing our balanced accuracy, which aligned with our goals.

Additionally, Emily incorporated more input features for a second Logistic Regression model also created to predict whether or not there was an arbitrage opportunity. At the time of each game that was played, there would not only have been game statistics from the previous five games that were played by both the home and away teams, but also standings from the previous season (if the teams were in the league previously). Using statistics from ESPN, Emily merged each match with relevant data for the previous year's seasons as a way of introducing more aggregated performance statistics. Using these features in addition to that from the previous five games, the Logistic Regression model had an ever-so-slightly higher balanced accuracy of 59.11% using the same custom probability threshold of .107. This model's performance (purple line in Fig. 4) can also be compared to the previous model with less features' performance through its Area Under Curve (AUC) of 0.6248, which is slightly higher than the AUC of the model only using game statistics from the previous five games. With only this small difference, it is not certain whether adding more data is actually useful in informing the presence of arbitrage opportunities.

IV. NAÏVE MODELING APPROACH: MULTICLASS LOGISTIC REGRESSION

INPUT & OUTPUT FOR MULTICLASS LOGISTIC REGRESSION

X	*Current Date, Teams Playing; Betting Odds & Game Statistics for each teams' previous five games (10 previous games total)*
y	*'0' a profit margin within [-∞, -2%], '1' in [-2, -0%], '2' [+0%, +2%], '3' in [+2%, +∞]* [9]

[9] See *Fig. 3* for a table of the categorical encoding of the profit margin intervals.

When faced with a multi-class classification problem, however, a Logistic Regression model fitted on the same data is much worse at predicting the interval the profit margin would be in. With this more difficult problem, the Logistic Regression Model accuracy drops to 62.48%. The balanced accuracy score, however, is only 30.57%; this metric deals with imbalanced datasets such as this one, whereas in a majority of games the profit is negative because there is no arbitrage opportunity and bookmakers craft odds so that the expected value is negative. However, it must be noted that for this four-class set, **a baseline 'random' score is 25%**, so 30.57%, contextualized, is still a significant result.

As was done with the binary classification problem, Emily integrated data from the previous season for each team into an additional model. This model also had marginally higher accuracies and balanced accuracies of 63.22% and 31.62%. But regardless of whether the model had additional information to use in predictions, neither model was able to correctly classify any observations with the true label of having positive profit margins: in this multi-class problem, the model never recognized any arbitrage opportunities that existed. The test set had 279 instances of positive profit margins out of 2705 total instances, but the model consistently incorrectly classified everything with a positive profit margin as belonging to one of the categories with a negative profit margin; this is shown below in the bottom two rows of each confusion matrix: all the actual instance in class 2 and 3 were incorrectly classified as class 0 or 1. The **only logistic regression models that had true positive classifications** of actual arbitrage opportunities as such **were the two binary classification models using probability thresholds set to the true proportion of arbitrage opportunities present in the dataset as a whole** (.1072).

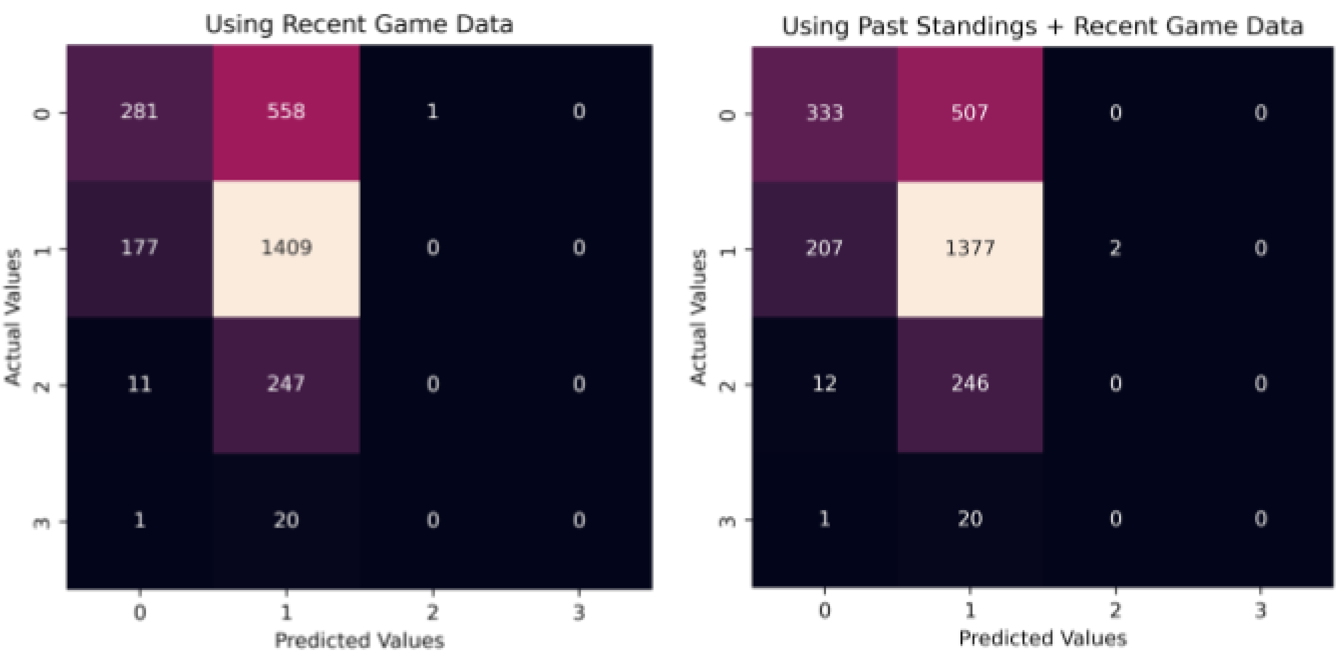

Fig. 5: Confusion Matrix for each of the two Multi-Class models. The left is the model using only data from the previous five games to predict on, and the right is the model using that recent game data as well as the past season's standings for the relevant teams.

V. DEEP LEARNING APPROACH: MULTICLASS NEURAL NETWORK

INPUT & OUTPUT FOR MULTICLASS NEURAL NETWORK

X	*Current Date, Teams Playing; Betting Odds & Game Statistics for each teams' previous five games (10 previous games total); data was scaled using the scikit-learn 'Standard Scaler' package.*
y	*Possible Profits were categorized into one of four categories (as denoted in fig. 3). They were then subsequently One Hot Encoded, into a 4 x length array.*

When examining the results of our Logistic Regression models Alec theorized that certain predictors might vary non-linearly, calling for a revised approach. The solution that he landed upon was a basic classification Feed Forward Neural Network. Unlike the Logistic Regression, the hyperparameters of FFNNs, in addition to randomized parameter 'starting positions,' vary so widely that reproductions might achieve results that vary widely from those achieved here. The primary non-standard aspect of this implementation however, was the loss function. Whereas models (label-smoothed or otherwise) fitted on generally balanced data use the standard "Categorical Cross-Entropy" metric,[10] Alec elected to implement Focal Crossentropy with a gamma value of 2.0, at the recommendation of the Meta article which proposed it[11]. The strength of Focal Crossentropy is the role of a gamma parameter which allows for additional weighting for classes which are underrepresented in the data: in this case, the rows which actually *were* valid Arbitrage opportunities. In terms of model structure, hard guidelines were few and far between; the nearest specificity available was "the optimal size of the hidden layer is usually between the size of the input and size of the output layers," and that including hidden layers above two generally yields sharply diminishing returns[12]. After some tinkering, pruning, and trial-and-error, the final structure was defined as in Fig. 5.

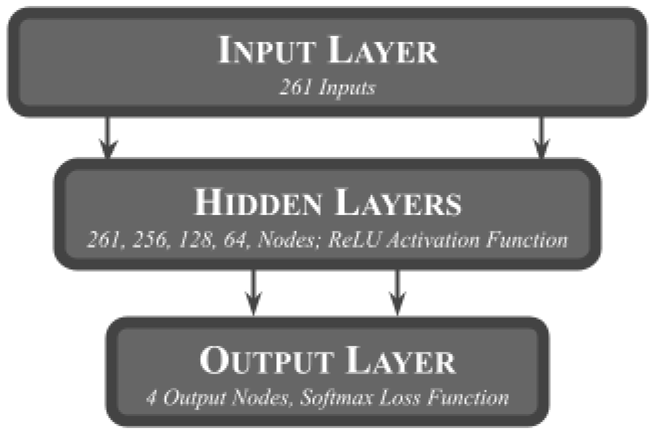

Figure 5: Classification FFNN Structure

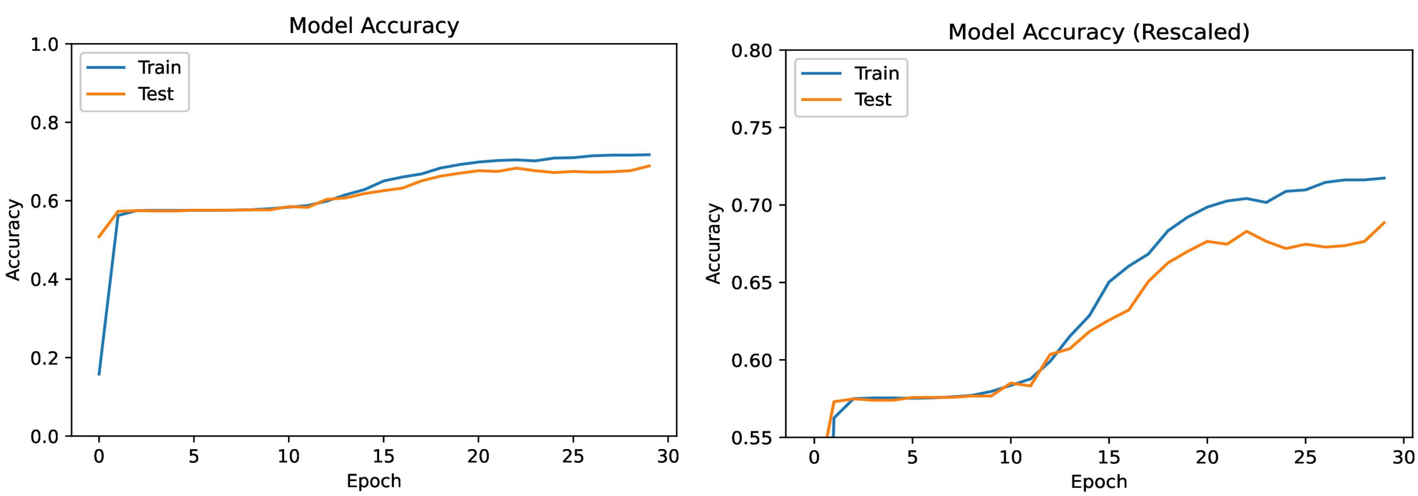

Figure 6: Unbalanced Accuracy Scores; Rescaled (Right)

[10] Elliot Gordon-Rodriguez, et al. "Uses and Abuses of the Cross-Entropy Loss: Case Studies in Modern Deep Learning". *Proceedings of Machine Learning Research*, vol. 137; 1-10. http://proceedings.mlr.press/v137/gordon-rodriguez20a/gordon-rodriguez20a.pdf.

[11] Tsung-Yi Lin et al. "Focal Loss for Dense Object Detection". *International Conference on Computer Vision* (2017). https://research.facebook.com/file/1009710453184631/focallossdensedetection.pdf

[12] Warren S. Sarle, FAQ posts for "comp.ai.neural-nets", 2002, ftp://ftp.sas.com/pub/neural/FAQ.html.zip.

Before examining model performance, *it is important to state that better results are almost certainly possible with a more robust, complex neural net. The conclusions drawn here pertain to the specific model architecture implemented as is outlined above.* Examining the performance, we only see a maximum 68.85% unbalanced score, which is an improvement over the Logistic Regression trained on the same data, but not significant. However, the model achieved a *Balanced Score of **67.99%***, compared to the 30.57% of the analogous Logistic Regression Model. To reiterate, the baseline 'random' score is *25%* (looking at the graph, this seems to be roughly the accuracy that the model began at). This means that despite being *as accurate* if not slightly *more accurate* than the other model, the NN Classifier was capable of doing so while being *agnostic to category imbalances.* It even performed better by 10% than the *Binary Logistic Classifier*. This essentially confirms the hypothesis that these data are highly non-linearly related.

The Accuracy x Epoch data, indicates that overtraining is certainly possible, and these results were achieved after much trial-and-error surrounding batch sizing and n-epoch setting. Alec found that while the model was *capable of* 100% accuracy in the training set, this came at the drastic expense of the test set being almost entirely unpredictable. In general, the more epochs the model was trained on, the more uncorrelated the test/training accuracies became, pointing towards very standard overfitting. In addition, larger batches seemed to yield better results, possibly because the likelihood of arbitrage opportunities *being in* a given batch was higher.

VI. Deep Learning Approach: Regression Neural Network

Input & Output For Regression Neural Network

X	*Current Date, Teams Playing; Betting Odds & Game Statistics for each teams' previous five games (10 previous games total); data was scaled using the scikit-learn 'Standard Scaler' package.*
y	*Possible Profits were directly listed as a positive or negative decimal figure, then scaled using the scikit-learn 'Standard Scaler' package.*

While the Deep Learning model for Classification was highly successful, the metrics of "accuracy" between different categories seemed somewhat unmeaningful (a gross miscategorization from < -2% to > 2% was weighted the same as an 'adjacent' miscategorization). To truly understand the model's performance and capabilities, a 'distance' based metric would be a better performance quantifier. So, while the "Regression Neural Network" was technically a different model than the classifier, its internal node structure was maintained the same (fig. 3). The difference was that instead of categorical cross-entropy, loss would be determined from *distance* instead, specifically, Mean Absolute Error. Alec chose MAE over Mean Squared Error primarily because MSE generally smooths outliers, whereas outliers in the Arbitrage setting are exactly the target prediction. Reformatting the categorical *y* data was trivial insofar as removing the code to categorize arbitrage outcomes, as well as applying the Standard Scaler package.

Examining these result data, we see a much 'smoother' training process, likely because the error was better defined compared to the categorical error not differentiating *how wrong* a result was. Unlike with the other classification tasks, we lack some context as to what the Error actually *means*, especially because the y data was normalized. However, we can assume that at around .77 was the baseline 'random' performance, and that 0 would mean perfect prediction. By this standard, our val_loss, or test set data was predicted 28.57% on a scale from random to perfect. All we can observe is that there was certainly meaningful improvement above our baseline, and the error divergence can again be seen between training and testing sets as more epochs are performed.

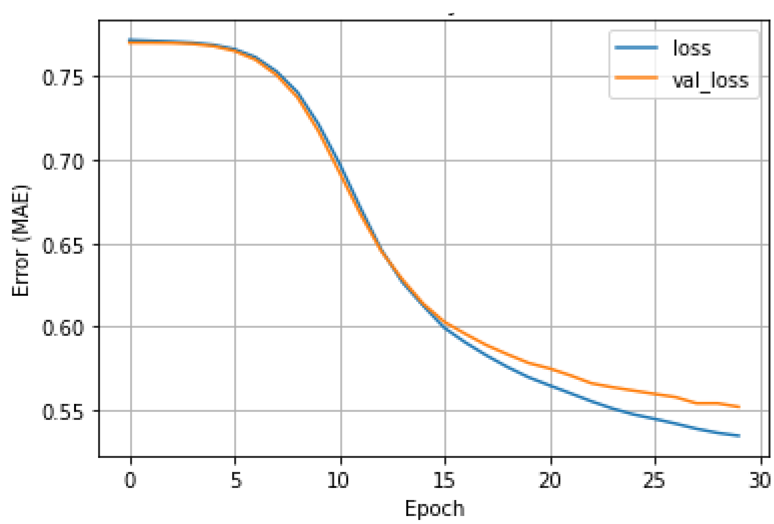

Figure 7: Loss (MAE)

VII. Conclusion & Final Model Selection

This study suffers somewhat in so-called "Final Model Selection," as only two of the four models can be substantively compared, being trained on the same tasks. To that end, the deep learning approach certainly yielded better results, both balanced and otherwise; it even yielded better balanced and unbalanced scores than the Binary Classification task, which was profoundly less precise and 'easier' than the four-category one. Where it becomes particularly tricky is when comparing the Multiclass and Regression Neural Networks. Really, the only comparability between the two is which would be more useful in trying to predict arbitrage. Here, clearly the Regression approach gives more information, and is hypothetically more 'accurate' having a better, more descriptive error function, but less credence in prioritizing arbitrage opportunities. So, a "Final Model Selection" would have to settle squarely between the two Deep Learning approaches.

The primary methodological concern here would likely be: because each row of training data contains information from previous games, would it be possible that included information could contaminate training for another row through the models encoding, or 'remembering' information. If so, the training and test sets would no-longer be properly siloed off. However, this is unlikely, given the relative simplicity of our models; the most complex model: the neural network only has four hidden layers, and seems hardly likely to be capable of 'remembering' specific game statistics.

Generally, this study has proven that it is indeed possible to meaningfully predict 'above-random' arbitrage opportunities, purely from historical data. So, the direction going forward is twofold:

1. What other exotic data could improve accuracy and what hyperparameters in the deep-learning approach could be tweaked to further prevent overfitting, improve accuracy, and further prioritize rarer arbitrage opportunities?
2. An advantage of these models is that they do not meaningfully use very soccer-specific techniques. So, if the Premier League itself is inhospitable to Arbitrage, in what other contexts would it be more useful to apply similar techniques?

Even continuing with the Premier League and the time span studied, one thing that may contribute to the presence of more arbitrage opportunities, and potentially also better predictions, is having more odds to work with from different bookmakers. This may require scraping multiple websites or pulling live odds using an API, which we initially attempted to do, but could not accumulate enough data in the time span for this project. Processing the existing features to create "synthetic" variables could also be beneficial: A previous study of the European Soccer betting market found that the difference between the maximum and minimum odd on each outcome (home team wins, draw, away team wins) was "positively related to the implied margin"[13]. (These researchers defined implied margin as what we considered the profit margin; it is considered implied since calculating the actual margin of bookmakers would require knowing the distribution of bets over all outcomes; the typical approach in literature is to assume a uniform distribution of bets across outcomes and that odds are determined by their true probabilities.[14])

The inclusion of synthetic features as well as more game-related statistics could also prompt an analysis of what features are most important for predictions of arbitrage opportunities and profit margins. Trying different regularization strengths for L1 regularization in the Logistic Regression models could be insightful, as well as performing principal component analysis and identifying clusters of features. However, all of this would only be applicable if it is possible to get more data with more existing arbitrage opportunities. Without lowering the probability threshold necessary to classify an observation as being an arbitrage opportunity, the Logistic Regression models barely classified any instances as being such; having more training data with existing positive profit margins could potentially allow for more instances to be classified, and also classified correctly. Delving into fixed-odds bookmakers in addition to online bookmakers could allow for more observations with arbitrage opportunities: past research into european football using five major online bookmakers and one fixed-odds bookmaker found the presence of most opportunities to be from the fixed-odds bookmaker[15]. This may be

---

[13] Nikolaos Vlastakis, George Dotsis, and Raphael N. Markellos. "How Efficient is the European Football Betting Market?". *Journal of Forecasting* 28 (2009), 426-444. 10.1002/for.1085.

[14] Vlastakis, Dotsis, and Markellos. "How Efficient is the European Betting Market?"

[15] Ibid.

because fixed-odds bookmakers are not able to update odds as easily, making them more susceptible to creating odds that would construct an arbitrage opportunity.

Also, expanding the scope of our study of Soccer outside the Premier League may be beneficial in expanding our dataset as well as improving the performance of our models. A previous examination found that about 75% of matches containing arbitrage opportunities were international competitions, and that while rare, the existing arbitrage opportunities yielded highly profitable returns[16]. From the same data source, it would also be possible to get odds data from multiple divisions and leagues aside from the Premier League, and literature has found lower divisions to contain more increased profit margins[17]

[16] Ibid.
[17] Constantinou and Fenton. "Profiting from arbitrage and odds biases of the European football gambling market."

Works Cited

Constantinou, Anthony Costa, and Normal Elliot Fenton. "Profiting from arbitrage and odds biases of the European football gambling market." *The Journal of Gambling Business and Economics* 7, no. 2 (2013): 41-70.

Elliot Gordon-Rodriguez, et al. "Uses and Abuses of the Cross-Entropy Loss: Case Studies in Modern Deep Learning". *Proceedings of Machine Learning Research*, vol. 137; 1-10. http://proceedings.mlr.press/v137/gordon-rodriguez20a/gordon-rodriguez20a.pdf.

Franck, Verbeek, Nuesch. "Inter-market Arbitrage in Betting", *Economica*, 318 (2012); 300-325, https://onlinelibrary.wiley.com/doi/full/10.1111/ecca.12009.

Johannes Stubinger and Julian Knoll. "Beat the Bookmaker - Winning Football Bets with Machine Learning (Best Application Paper)". *International Conference on Innovative Techniques and Applications of Artificial Intelligence* (2018): 219-223.

Matthew Buckle and Chun-Sung Huang. "The Efficiency of Sport Betting Markets: An Analysis Using Arbitrage Trading within Super Rugby." *International Journal of Sport Finance*, 3 (2018); 279+,, https://go.gale.com/ps/i.do?id=GALE%7CA568257897&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=15586235&p=AONE&sw=w&userGroupName=anon%7Ee4d79edc.

Nikolaos Vlastakis, George Dotsis, and Raphael N. Markellos. "How Efficient is the European Football Betting Market?". *Journal of Forecasting* 28 (2009), 426-444. 10.1002/for.1085.

Tsung-Yi Lin et al. "Focal Loss for Dense Object Detection". *International Conference on Computer Vision* (2017). https://research.facebook.com/file/1009710453184631/focallossdensedetection.pdf

Warren S. Sarle, FAQ posts for "comp.ai.neural-nets", 2002, ftp://ftp.sas.com/pub/neural/FAQ.html.zip.