NFL SPORTS BETTING NEURAL NETWORKS

Rheyan Ampoyo, Louis Cocks
Authors contributed equally to this work
Affiliation
University of Virginia
Charlottesville

ABSTRACT

The growing popularity of sports betting has spurred interest in data-driven models capable of accurately estimating the probability of game outcomes. This project proposes the use of neural networks to predict the probabilities of football betting results, such as match winners, point spreads, and over/under totals. Using large-scale historical data, including team performance metrics, player statistics, injury reports, and weather conditions, the model aims to uncover complex nonlinear relationships that traditional statistical models often overlook. So, how can neural network architecture will be optimized through techniques such as feature selection, hyperparameter tuning, adaptive learning, and dropout regularization to balance predictive accuracy and generalization? The architecture of Fusion Transformers handle these steps well with added flexibility for multi-leg parlay encoder caching and interpretable results. With these results evaluated against benchmark models such as logistic regression and random forests using metrics such as Expected Calibration Error (ECE), Parlay Calibration Error (PCE), and AUC-ROC score. This project aims to showcase how deep learning enhances decision-making in sports analytics, specifically by generating probabilistic predictions that more effectively capture uncertainty in football outcomes providing both theoretical insights and practical applications for analysts, bettors and researchers.

1 Motivation

Sports betting, particularly in football, has evolved into a data-rich industry where analytics plays a critical role in evaluating outcomes and odds. Despite this, many existing models used by analysts and bookmakers still rely on simplified statistical approaches that often fail to capture the complex, nonlinear dynamics of team performance, player interactions, and situational factors influencing game results. As the volume and granularity of sports data continues to increase, spanning player metrics, play-by-play statistics, injuries, and even environmental conditions there is a growing opportunity to apply machine learning (ML) techniques that can uncover deeper patterns within the data.

Neural networks, known for their ability to model high-dimensional and nonlinear relationships, present a promising solution for estimating the likelihood of football betting outcomes such as match results, point spreads, and totals. This research aims to design and evaluate a neural network framework that predicting probabilistic outcomes based on historical and contextual game data. This project will investigate how Temporal Fusion Transformers (TFT) and adaptable feature sets affect predictive model performance, accounting for gaps in the use of the Kelly criterion for odds risk management, computationally efficient modeling, and multi-leg parlay scenarios for model testing. Ultimately, this study seeks to contribute to the field of sports analytics by demonstrating how deep learning can improve the predictability and reliability of betting forecasts, offering practical and valuable insights.

2 Dataset

Data is pulled from two main public resources listed below and will be sourced through R and python package calls: nflreadr & nntrn:

- nflreadr

  - Package contains play by play and player specific data.
    * Columns examples consist of Game_id (chr), season_type(chr), Week(int) and game_date(chr)
    * CSV formatted for games.csv, pbp.csv, and more.
- nntrn
  - robust Github repository with api parameters for game by game data; team data; league data; officials data and more.

3 Literature Review

3.1 Systematic Review of Machine Learning in Sports Betting: Techniques, Challenges, and Future Directions

(Link)

In [Galekwa et al., 2024], researchers examined the growing role of ML in predicting sports outcomes and optimizing betting strategies. Following PRISMA guidelines, the authors screened hundreds of studies which excluded non-English, theoretical, or non-empirical work to identify research that applies ML methods to real betting data. Across sports, decision trees and neural networks consistently outperformed traditional statistical models along with some niche neural networks, particularly when enhanced with contextual features such as: team form, head-to-head history, player statistics, and betting odds. Researchers reviewed in this meta-analysis ([Walsh and Joshi, 2024]) found that optimizing models for calibration rather than accuracy yielded up to 69.86% higher betting returns, emphasizing probability reliability over simple prediction. In American football (Section 4.5), researchers leveraged play-by-play data to capture situational dynamics, while in [Deng and Zhong, 2020] the deep neural network achieves greater accuracy than 95% in soccer prediction, the potential of deep learning. Overall, the review highlights key challenges such as data limitations in context, non-stationarity, and over-reliance on accuracy metrics and concludes that future success in sports betting ML will depend on calibrated, context-rich models that balance predictive precision with real-world profitability.

3.1.1 Identify Gaps

A key limitation of the existing studies is the absence of binary classification metrics and modeling approaches directly applicable to multi-leg betting scenarios. Since this project aims to develop an application capable of accurately predicting the outcome of individual legs within a parlay, there is a need for models that emphasize binary or quantile prediction results, such as win-or-loss probabilities rather than continuous performance outcomes like expected yards or completion percentage. Addressing this gap will allow for a more direct alignment between predictive modeling and real-world betting decision-making.

3.2 Exploiting Sports-Betting Market Using Machine Learning

(Link)

[Hubáček et al., 2019] explores how convolutional neural networks can be used to predict sports outcomes while intentionally reducing correlation with bookmaker odds. Rather than replicating market-implied probabilities, the authors design a model that identifies informational inefficiencies by de-correlating its predictions from those embedded in betting lines. Their approach operates on a two-team data pipeline, using only the features directly relevant to the competing teams in a given matchup such as form, ranking, and performance metrics, thus emphasizing game-specific rather than league-wide information. To improve profitability, the study introduces methods that 'elaborate various measures to remodel the learned model from the bookmaker's model (approximately estimated from the odds assigned), while maintaining adequate prediction accuracy.' The main models are two types of neural network tests including a standard (deep) feed-forward architecture with 4 dense layers, and a second with 1 convolutional layer followed by 3 dense layers. Confidence-thresholding is added as an enhancement in conjunction with models utilizing logistic sigmoid output which was the utilized activation function. The findings demonstrate that this de-correlation strategy can enhance profit margins by uncovering value bets overlooked by the market, showing that model independence from bookmaker behavior can be as critical as predictive accuracy.

3.2.1 Identify Gaps

There was a trade-off between the increased level of de-correlation which dropped accuracy but increased profits and vice versa. Given this, some overlooked aspects of the bookmaker model include public betting bias exploitation by which bookmakers can adjust markets and incentives for geographical location, inherent biases, and knowledge of 'whales' (extremely large bets). A portfolio theory gap includes the missing incorporation of Kelly Criterion for bet sizing which is notably a successful strategy addressed in the meta-analysis (from [Matej et al., 2021]). A final gap is a lack of realistic market simulation where closing odds with unlimited liquidity overestimates real-world profitability; artificially inflating the direct impact of this analysis on future methodologies for real-world implementation.

3.3 Sports Betting: an application of neural networks and modern portfolio theory to the English Premier League

(Link)

By integrating machine learning with quantitative finance principles to construct optimized betting portfolios, [Jiménez et al., 2023] employ Kullback–Leibler (KL) divergence to measure disparity between model-predicted probabilities and bookmaker-implied odds, allowing the system to identify wagers where the market miss-prices true outcome likelihoods. To allocate wagers efficiently, they apply concepts from modern portfolio theory constructing an efficient frontier of betting combinations that minimize the variance for a given expected return. Building on this foundation, the study borrows from modern portfolio theory, defining an efficient frontier of betting portfolios that minimize variance for a given expected return analogous to selecting the "best" number of legs in a parlay. The model further utilizes the Kelly Criterion to determine optimal stake sizing, maximizing long-term bankroll growth while controlling for risk exposure. This combination of probabilistic learning, information theory, and portfolio optimization represents a shift from traditional single-bet prediction toward system-level betting strategies that explicitly balance risk and reward.

3.3.1 Identify Gaps

While this framework provides a rigorous mathematical basis for optimizing betting portfolios, it assumes that model-predicted probabilities are perfectly calibrated and that market conditions remain static. These assumptions limit real-world applicability, particularly in fast-changing environments like NFL games where player injuries, weather, and line movements continuously alter the betting landscape. Additionally, the framework focuses on optimizing portfolio weights rather than improving the underlying predictive accuracy of individual bets or legs—an essential component for applications like the one proposed in this project. For a sports-betting model that dynamically predicts and selects parlay legs, bridging this gap requires integrating calibrated neural network predictions with portfolio optimization, allowing the system to adapt to shifting odds and evolving game contexts while still applying Kelly-style bankroll management.

4 Methodology

This paper covers a customizable approach as taking the best raw techniques seen above while acknowledging the gaps in study, such is a deterministic listing of the best criterion and techniques under the implementation of a recently developed model, [Lim et al., 2021] Temporal Fusion Transformers (TFT). Portfolio theory is often used in the best models [Galekwa et al., 2024] & [Jiménez et al., 2023] that incorporate odds data; however Kelly criterion is seldom mentioned. It will be used to diminish hidden assumptions in the variable selection network when adjusting for different parlay legs in a manner of risk management. We will also de-correlate any bookmaker data from the datasets to support [Hubáček et al., 2019] findings. Another covered gap is widespread real-world applicability in that we have a computationally efficient model with the implementation of multi-leg parlays (not singular bets) as a pivotal point in architecture of the model.

The following breakdown describes our TFT model directly. First is the variable selection blocks that, without supervision extract static covariates, past inputs and known future inputs. Second, accessibility for usage of this model comes from computational efficiency given by pruning of irrelevant features, and caching of encoders for multi-leg parlays. Third, the static LSTM encoder/decoder handles variable-length sequences like total games played, given bye weeks. Fourth, the masked interpretable multi-head attention provides attention patterns on different focal points such as emphasis on recent games, historic rematch data, and new factors like rookies and bench players, and more. Fifth, gated residual networks are supportive of preventing gradient vanishing/exploding, adaptive learning for variable comparisons in games, and regularization to prevent overfitting, all of which improve model success. Finally, the sixth reason surrounds the uncertainty distribution, or lack thereof in most models, where now we have a linear quantile spread to further support multi-leg parlays in understanding why risk management and focused calibration of expected calibration error (ECE) [Walsh and Joshi, 2024] success over singular accuracy.

5 Experiments

To evaluate the effectiveness of the TFT model several metrics will be compared against benchmarks including ECE, ROC-AUC, Prediction Correlation to the bookmakers, and Log-Loss for cross-entropy. Along with singular metrics there exists parlay & financial related metrics such as Independence Check (Chi-Square) for independence, Kelly Criterion Betting ROI and Sharpe Ratio for risk-adjusted returns, and Parlay Calibration Error (PCE) for joint probabilities where N games with true outcomes $\mathbf{y}_{true} = (y_1, \ldots, y_N) \in \{0,1\}^N$ and predicted probabilities $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_N) \in [0,1]^N$, we define the parlay calibration error for L-leg parlays as:

$$\epsilon_{PCE}(L, M) = \frac{1}{M} \sum_{m=1}^{M} \left| \hat{p}_{parlay}^{(m)} - y_{parlay}^{(m)} \right| \tag{1}$$

where for each random parlay sample $m \in \{1, \ldots, M\}$:

$$\begin{aligned}
\mathcal{I}^{(m)} &\sim Uniform(\{I \subset \{1, \ldots, N\} : |I| = L\}) \\
\hat{p}_{parlay}^{(m)} &= \prod_{i \in \mathcal{I}^{(m)}} \hat{p}_i \\
y_{parlay}^{(m)} &= \begin{cases} 1 & \text{if } y_i = 1 \text{ for all } i \in \mathcal{I}^{(m)} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Consequently, a threshold does not exist for predictive acceptance of a leg, rather there is a quantile range for all modeled legs. Finally Prediction Correlation will be used for correlation checking and for TFT, quantile uncertainty distribution verifying each qunatiles' accuracy is critical; therefore, Prediction Interval Coverage will be calculated. A successful model will show Expected Calibrated Error (ECE)/ Parlay Calibration Error < 0.05, profitability (Kelly ROI) at least positive > 0, and de-correlation (correlation < 0.9).

References

[Deng and Zhong, 2020] Deng, W. and Zhong, E. (2020). Analysis and prediction of soccer games: An application to the kaggle european soccer database. *Insight - Statistics*, 3:1.

[Galekwa et al., 2024] Galekwa, R. M., Tshimula, J. M., Tajeuna, E. G., and Kyandoghere, K. (2024). A systematic review of machine learning in sports betting: Techniques, challenges, and future directions.

[Hubáček et al., 2019] Hubáček, O., Šourek, G., and Železný, F. (2019). Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, 35(2):783–796.

[Jiménez et al., 2023] Jiménez, V., Alberto, R., Ontiveros, L., Manuel, J., and Possani, E. (2023). Sports betting: an application of neural networks and modern portfolio theory to the english premier league. Papers 2307.13807, arXiv.org.

[Lim et al., 2021] Lim, B., Arık, S. , Loeff, N., and Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764.

[Matej et al., 2021] Matej, U., Šír, G., Hubáček, O., and Filip, (2021). Optimal sports betting strategies in practice: an experimental review. *IMA Journal of Management Mathematics*, 32.

[Walsh and Joshi, 2024] Walsh, C. and Joshi, A. (2024). Machine learning for sports betting: Should model selection be based on accuracy or calibration? *Machine Learning with Applications*, 16:100539.