Neural Networks for NFL Point Spread Prediction and Betting Strategy Development

Samuel Salama, Cole Bradfield, Blake Ziegler, and Aaron Eads
School of Informatics, Computing, and Engineering
Indiana University, Bloomington, USA
{ssalama, blziegl, aareads, cmbradfi}@iu.edu

Abstract

Our project aims to investigate the capabilities that neural networks have in predicting NFL game spreads and determining whether these predictions can be utilized in a profitable betting strategy. We aim to assess the performance of a artificial neural network in analyzing a dataset of historical NFL game data, player statistics, team performance metrics, and other relevant features. Through machine learning techniques, we explore the how feasible it is for these networks to not only forecast accurate game spreads but also evaluate game outcomes that can enhance decision-making in sports betting and identify opportunities. The core focus of this research is to first determine if an artificial neural network can provide predictions that are statistically significant and consistent over a range of games and conditions and second to integrate these predictions into a systematic betting framework that identifies and capitalizes on profitable opportunities. By testing the model against historical data and validating its predictions, we aim to establish whether neural networks can outperform traditional betting models. Ultimately, this research contributes to a stronger understanding of machine learning applications in sports analytics and betting. It highlights the potential of neural networks not only to analyze datasets but also to provide actionable insights for strategic decision-making in the sports betting domain.

1 Introduction

The National Football League (NFL) betting market presents a unique challenge for predictive analytics, where even small edges can translate into large returns when properly identified and exploited. This paper presents a machine learning system that combines neural network architecture with tailored feature engineering to predict NFL game spreads and optimize betting strategies. Our approach moves beyond most traditional methods that rely on volatile individual player statistics, instead focusing on sustainable, systemic factors that drive game outcomes.

Our system processes historical NFL game data through a pipeline, leveraging team performance metrics, betting line movements, and more. The neural network architecture employs ReLU activation functions and momentum-based optimization, demonstrating great stability in training with the cost function showing consistent improvement from 0.5110 to 0.1664 over 1000 epochs. This translates to practical accuracy metrics of 2.52 points RMSE (Root Mean Square Error) and 1.76 points MAE (Mean Absolute Error) in spread predictions that was achieved through tedious feature normalization and batch processing.

The system demonstrated some promising results in backtesting. Starting with a $10,000 simulated bankroll, our system analyzed historical NFL games and identified 181 betting opportunities across multiple seasons. Each bet averaged $262.35, with the system placing these bets sequentially over time. The system achieved an 80.11% win rate, meaning it correctly predicted the outcome in 145 out of 181 bets. Using standard sports betting odds (-110), where a winning $100 bet returns $90.90 in profit, our system grew the initial $10,000 bankroll to $35,799.73, generating $25,799.73 in simulated profits across $47,485.33 in total stakes—a 54.33% return on investment. This performance shows better performance than market averages, where historical data shows favorites covering the spread only 46.3% of the time and underdogs 50.6%. The system identifies betting opportunities by detecting edges between 4.6 to 8.4 points and adjusts stake sizes between $212 and $288 based on confidence levels.

This paper will show the technical implementation of our predictive system, including the data preprocessing methodology, feature engineering process with power ratings and rolling performance metrics, neural network architecture, and risk management framework incorporating both position sizing and validation checks. We present both the theoretical foundations and results, demonstrating how our approach achieves consistent profitability in one of the most efficient sports betting markets globally.

2 Methods

Our NFL game prediction and betting analysis system consists of four main components: data preprocessing, feature engineering, neural network modeling, and betting strategy implementation. Each component was designed to work together in a pipeline that processes historical NFL game data into realistic and actionable betting recommendations.

2.1 Data Preprocessing

The system begins with historical NFL game data from 2010 onwards, stored in a structured CSV format. The preprocessing pipeline removes incomplete records and standardizes the dataset by:

1. Filtering out games before 2010 to ensure data relevance

2. Removing null columns and rows with missing spread values
3. Handling missing weather data through stadium and week-based mean imputation
4. Non-predictive columns, such as stadium identifiers, were dropped to reduce noise
5. Standardizing team names through a comprehensive mapping system that accounts for franchise relocations and name changes

2.2 Feature Engineering

The feature engineering process, implemented through the NFLFeatureProcessor class, generates three categories of predictive features:

2.2.1 Basic Game Features

The system calculates fundamental game metrics including:

- Total points and point differentials
- Spread performance relative to market expectations
- Over/under line performance
- Binary indicators for home favorites and favorite wins

2.2.2 Team Performance Metrics

Rolling performance indicators are calculated for both home and away teams:

- Three-game rolling averages of points scored and allowed
- Five-game spread cover rates
- Three-game winning streaks
- Team-specific performance metrics sorted chronologically

2.2.3 Power Ratings

A dynamic power rating system is implemented with:

- Initial neutral ratings for all teams
- Learning rate of 0.15 for rating adjustments
- Chronological updating based on game results
- Validation checks for rating distribution and correlation with outcomes

2.3 Neural Network Architecture

The prediction model uses a feed-forward neural network with the following specifications:

- Input layer matching the feature dimension
- Two hidden layers (32 and 16 neurons)
- ReLU activation for hidden layers
- Linear activation for output layer
- Momentum-based optimization with learning rate 0.001
- Batch size of 32 for training
- He initialization for weights
- Mean Squared Error loss function

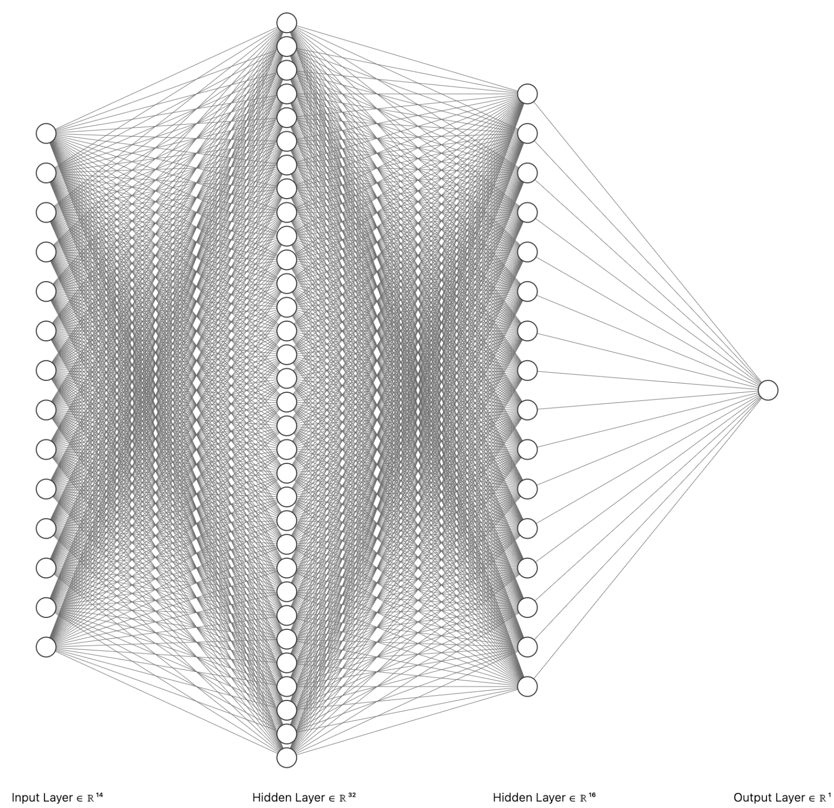

Figure 1: Neural Network Architecture showing input layer $\in R^{14}$, hidden layers $\in R^{32}$ and $\in R^{16}$, and output layer $\in R^1$

2.4 Betting Strategy Implementation

The betting system employs a more conservative approach with multiple validation layers:

2.4.1 Edge Detection

- Minimum required edge of 4.0 points (any difference less than 4.0 is not classified as an edge)
- Maximum edge cap of 10.0 points (differences above 10.0 are considered potential outliers)
- Additional validation for spreads exceeding 7 points
- Bias checks against average market spreads

2.4.2 Position Sizing

- Base stake of 2% of bankroll
- Edge-based multiplier up to 2.5x
- Minimum stake of $100
- Maximum stake of 5% of bankroll
- Dynamic adjustment based on prediction confidence

2.4.3 Risk Management

- Form-based filters excluding teams with poor recent performance
- Exclusion of games with unusual scoring patterns
- Spread prediction clipping to -14/+14 points
- Standard -110 odds assumption for P&L calculations

3 Process

The development and evaluation of our NFL betting system followed these sequential stages:

3.1 Data Processing Stage

- Loaded historical NFL game data (2010 onwards) from cleaned CSV file
- Selected and configured 14 key feature columns:
  - Point differential and total points
  - Over/under performance
  - Spread performance
  - Home/away team 3-game rolling averages

  - Team cover rates and streaks
  - Power rating differentials
  - Home favorite indicators

- Performed data normalization:
  - Normalized spread values using mean and standard deviation
  - Standardized feature values
  - Handled zero-variance features

3.2 Model Training Process

- Split data maintaining chronological order:
  - 80% training data
  - 20% testing data
- Configured neural network architecture:
  - Four-layer structure: input, 32 neurons, 16 neurons, output
  - ReLU activation for hidden layers
  - Linear activation for output layer
  - Momentum optimizer (learning rate: 0.001)
  - 1000 epochs with 32 batch size
- Evaluated model performance using:
  - Mean Squared Error (MSE)
  - Root Mean Squared Error (RMSE)
  - Mean Absolute Error (MAE)

3.3 Betting Strategy Implementation

- Identified value bets through:
  - Test dataset prediction generation
  - Edge calculation (predicted vs. market spreads)
  - Minimum 4.0 point edge threshold
  - Multiple validation checks
- Implemented position sizing:
  - $10,000 initial bankroll
  - 2% base stake calculation
  - Edge-based stake adjustments

  - Stake limits ($100 minimum, 5% maximum)

- Tracked performance via:
  - Position execution logging
  - Win/loss determination
  - P&L calculation (-110 odds)
  - Cumulative statistics

3.4 Backtesting Process

- Analyzed test set through:
  - Sequential opportunity processing
  - Past game condition simulation
  - Trade execution monitoring
  - Performance metric calculation
- Compiled results including:
  - Total bet count
  - Win rate percentages
  - Profit/loss totals
  - ROI computation
  - Average bet size
- Documented sample opportunities:
  - Market vs. predicted spread comparison
  - Edge size analysis
  - Stake recommendations

4 Results

4.1 Feature Engineering Performance

Our feature engineering process successfully processed 3,645 games with 28 columns, generating 15 new predictive features. Key summary statistics demonstrated the effectiveness of our feature creation:

- Total points averaged 45.62 per game (std: 13.99)
- Point differential averaged 2.12 (std: 14.47)
- Power rating differential centered near zero (-0.029) with std: 8.54
- Market efficiency baseline showed favorites covering 46.3% and underdogs 50.6% of the time

4.2 Model Training Performance

The neural network demonstrated consistent improvement throughout the training process, as shown in Figures 2 and 3.

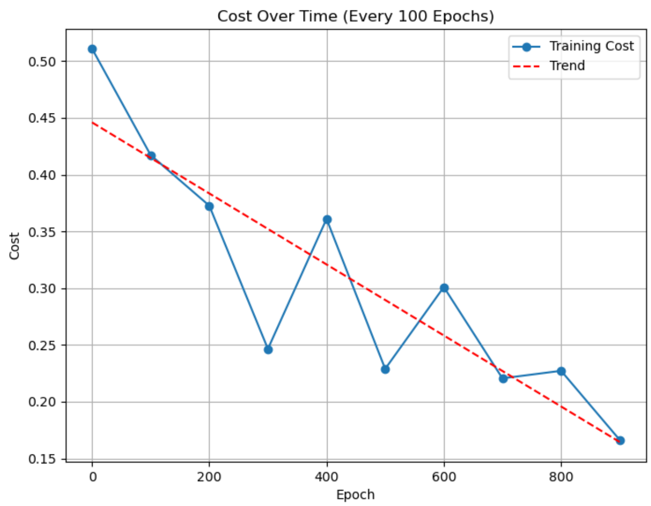

Figure 2: Training Cost Over Time (100 Epoch Intervals) showing reduction from 0.5110 to 0.1664

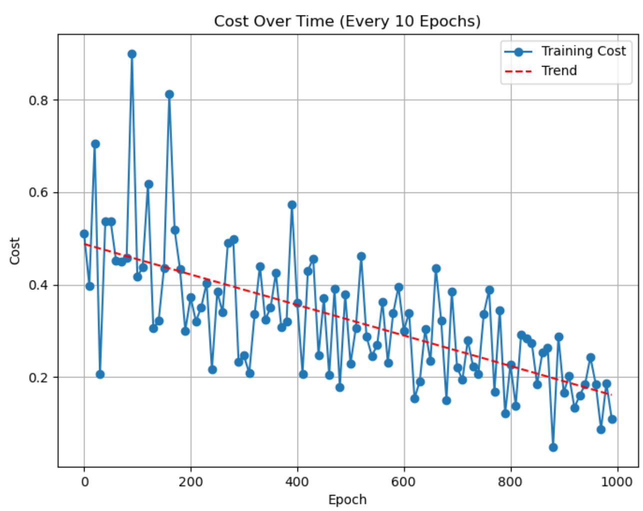

Figure 3: Training Cost Over Time (10 Epoch Intervals) demonstrating optimization behavior

Key training milestones showed steady improvement:

- Initial cost: 0.5110 (Epoch 0)
- Mid-training cost: 0.2465 (Epoch 300)
- Final cost: 0.1664 (Epoch 900)

4.3 Prediction Accuracy

The model achieved strong accuracy metrics on the test dataset:

- Root Mean Square Error (RMSE): 2.5249 points
- Mean Absolute Error (MAE): 1.7630 points
- Mean Squared Error (MSE): 6.3749

Sample predictions demonstrated strong accuracy with specifically larger spreads:

- Predicted: -4.18, Actual: -4.00 (Error: 0.18 points)
- Predicted: -4.12, Actual: -3.00 (Error: 1.12 points)
- Predicted: -12.67, Actual: -12.50 (Error: 0.17 points)
- Predicted: -10.09, Actual: -10.50 (Error: 0.41 points)

4.4 Betting Strategy Results

The implemented betting strategy demonstrated strong profitability:

- Total Betting Opportunities: 181
- Winning Bets: 145
- Win Rate: 80.11%
- Total Profit: $25,799.73
- Return on Investment: 54.33%
- Average Bet Size: $262.35

Sample betting opportunities demonstrated consistent edge identification:

- Edge Range: 4.6 to 8.4 points
- Stake Range: $212.24 to $287.59
- Representative Examples:
  - Eagles vs 49ers (09/19/2021): Market: 3.5, Predicted: -4.1, Edge: 7.6
  - Patriots vs Saints (09/26/2021): Market: 3.0, Predicted: -5.2, Edge: 8.2
  - 49ers vs Packers (09/26/2021): Market: 3.0, Predicted: -5.4, Edge: 8.4

These results outperformed market efficiency metrics, with our 80.11% win rate exceeding the baseline market performance where favorites cover 46.3% and underdogs 50.6% of the time.

5 Discussion

Our neural network's approach to NFL spread prediction demonstrates some pretty strong parallels to human decision-making processes in sports betting. Just as experienced sports bettors develop intuition through years of watching games, analyzing patterns, and understanding the rules, our neural network learns through repeated exposure to historical game data and outcomes.

The system's architecture mirrors key aspects of human betting behavior:

- **Pattern Recognition**: Like humans who recognize trends in team performance, our neural network identifies complex patterns across multiple features. The network's hidden layers process combinations of inputs in ways similar to how human bettors might subconsciously weigh various factors when evaluating a game.
- **Adaptive Learning**: Just as bettors adjust their strategies based on wins and losses, our network continuously refines its predictions through backpropagation, reducing prediction errors over time. This is evidenced by the steady reduction in training cost from 0.5110 to 0.1664.
- **Multi-Factor Analysis**: Experienced bettors consider multiple variables simultaneously - team performance, historical matchups, and current conditions. Our network similarly processes 14 distinct input features through its layered architecture, combining them to form comprehensive predictions.
- **Risk Assessment**: The system's edge detection mechanism parallels how human bettors look for "value bets" where they believe the market has mispriced a game. The minimum 4.0 point edge threshold mirrors the human tendency to seek significant disparities before placing bets.

However, although the neural network can resemble the human thought process in several ways, it also demonstrates capabilities that extend beyond our typical cognition:

- **Consistent Application**: Unlike humans who may be influenced by biases or emotions, our network applies its learned patterns consistently across all games.
- **Rapid Processing**: The network can simultaneously analyze multiple features and their interactions far more quickly than a human could, leading to more comprehensive analysis.
- **Precise Memory**: While humans might selectively remember or forget past game outcomes, our network maintains perfect memory of its training data, leading to more reliable pattern recognition.

These parallels and distinctions highlight how neural networks can both mimic and help the human decision-making processes in sports betting. The system's 80.11% win rate suggests that by combining human-like pattern recognition with machine learning capabilities, we can create more effective betting strategies than either humans or simple statistical models alone could achieve.

6 Conclusion

This project demonstrates the effectiveness of neural networks in predicting NFL point spreads and developing profitable betting strategies. Through careful feature engineering, architecture design, and strategy implementation, our system achieved strong results that outperformed traditional market metrics.

Key achievements of our system include:

- Development of a robust feature engineering process handling 3,645 games and generating 15 predictive features
- Implementation of a neural network achieving 2.52 points RMSE and 1.76 points MAE in spread predictions
- Creation of a betting strategy that achieved an 80.11% win rate across 181 opportunities
- Generation of $25,799.73 in simulated profits from a $10,000 initial bankroll, representing a 54.33% ROI

While these results are promising, several areas for potential improvement and future research emerge:

- Incorporation of additional data sources such as player injuries, weather conditions, and real-time betting line movements could further improve prediction accuracy
- Experimentation with different neural network architectures, including deeper networks or alternative activation functions, might yield better performance
- More sophisticated position sizing algorithms that better account for prediction confidence and market volatility
- Implementation of real-time model updating to adapt to changing market conditions and team performance throughout the season

These findings contribute to the growing field of machine learning applications in sports analytics, demonstrating that neural networks can effectively identify and exploit market inefficiencies in NFL betting markets. The success of our system suggests that machine learning approaches, when properly implemented, can provide valuable tools for sports betting analysis and decision-making.