Predicting NFL Game Outcomes Using ELO Ratings and Machine Learning Models

Abstract

This project employs a comprehensive method to predicting NFL game outcomes that includes both the ELO rating system and machine learning algorithms. The dataset includes 2,510 regular season games played between 2009 and 2018. We use feature engineering techniques to compute post-hoc characteristics such as scores, turnovers, and other performance indicators. Two machine learning models, Random Forest and Recurrent Neural Network (RNN), are used to forecast game outcomes based on the averages of the previous ten games. In addition, an ELO rating system is created using a Neo4j graph database to dynamically track team performance over time. The final model predictions are tested against real game results, yielding a prediction accuracy of 63% over the last 500 games.

Introduction to the Problem Statement

Predicting NFL game outcomes is difficult due to the sport's intricacy and the multiple elements that influence game results. The objective of this project is to forecast game outcomes using both classic ELO rating systems and current machine learning approaches. The ELO rating system allows for dynamic tracking of team performance over time by altering ratings depending on game results, whereas machine learning models can find hidden trends in historical data that traditional approaches may not capture.

Play-by-Play Structure of NFL Games

An NFL game (Contributors to Wikimedia projects, 2024) is divided into four quarters, each lasting 15 minutes, with the possibility of overtime if the game remains tied at the end of regulation. Throughout the game, teams rotate between offense and defense, with the goal of advancing the ball into the other end zone to earn points. Each play in an NFL game is made up of numerous important elements that are captured in play-by-play statistics. These variables include:

- Down: The number of attempts a team has to advance the ball 10 yards.
- Yards Gained: The number of yards a team advances the ball during a play.
- Play Type: Whether the play was a pass, run, or special teams play (e.g., punt or field goal attempt).
- Turnovers: Whether the ball was lost to the opposing team through an interception or fumble.
- Time of Possession: The amount of time a team controls the ball during a drive.
- Penalties: Infractions committed by players that result in yardage penalties or loss of downs.

In our context of predicting game outcomes, these variables are aggregated over entire games rather than individual plays. For example:

- Turnover Margin: The difference between turnovers committed and turnovers forced by a team can be a strong predictor of game outcomes.

- Time of Possession: Teams that control the ball for longer periods tend to have more opportunities to score.
- Total Yards Gained: Summing all yards gained by a team throughout the game gives an indication of offensive performance.

By aggregating these play-by-play variables into game-level metrics, we can create features that will be employed in both our ELO rating calculations and machine learning algorithms. These characteristics influence overall team performance and are essential to making accurate predictions about future games.

Relevance to Our Prediction Task

In this project, we are not interested in forecasting individual game plays, but rather with predicting the overall outcome (win/loss/tie). Understanding how individual plays contribute to overall team performance is critical for feature engineering. For instance:

1. Turnovers: Teams with fewer turnovers are generally more likely to win because they maintain possession and scoring opportunities.
2. Time of Possession: Teams that control possession for longer periods often dominate games by keeping their defense rested and their offense on the field.
3. Yards Gained: A team's ability to gain yards consistently is directly related to its ability to score points.

By converting these play-level variables into game-level metrics, we generate features that may be employed in both ELO computations and machine learning models. These features allow us to quantify team performance, which has a direct impact on our ability to predict future results.

Literature Review

The ELO rating system in this project is mostly based on the well-known ELO methodology, which was first devised to rank chess players (*Elo Rating System - Chess Terms*, n.d.) (Regan & Haworth, 2011). The ELO system automatically adjusts team ratings depending on game results, giving a simple yet effective method for tracking relative team strengths over time. In this project, we use the ELO rating system to forecast NFL game outcomes by modifying team ratings after each game depending on the actual vs.expected results.

The ELO system has been utilized in a variety of sports, including football, where it is used to assess team performance over time. The basic principle of ELO is that when a team wins a game, its rating rises and falls, with the size of the change determined by the difference in ratings between the two teams. This strategy enables dynamic adjustments in squad strength as new games are played.

In our implementation, team ratings are stored and updated using a Neo4j graph database. Each NFL team is represented as a node, and each game as a connection between two nodes (teams). The relation saves relevant game metrics like scores and dates. Following each game, both teams' ELO ratings are adjusted based on the actual and expected outcomes determined using the following formula:

$$E_A = \frac{1}{1 + 10^{\frac{\left(R_B - R_A\right)}{400}}}$$

where:

- E_A is the expected probability of team A winning
- R_A and R_B are the current ratings of teams A and B respectively

The rating update formula used is:

$$R'_A = R_A + K(S_A - E_A)$$

where:

- K is the k-factor (set to 20 in our implementation)
- S_A is the actual score (1 for win, 0.5 for tie, 0 for loss)
- E_A is the expected score

The Neo4j-based approach efficiently stores and retrieves team ratings, allowing for dynamic updates after each game. This approach was inspired by comparable implementations in the sports analytics literature, which used graph databases to describe team-game relationships (Bunker & Thabtah, 2019).

In addition to utilizing ELO to rate teams, we use machine learning models like Random Forests and Recurrent Neural Networks (RNNs) to forecast game results using historical data. These models have been demonstrated to perform well in sports prediction tasks by identifying complicated patterns in data that traditional statistical methods may miss.

Thus, the majority of our implementation is based on well-established approaches in both ELO-based ranking systems and machine learning models for sports prediction.

Objectives of the Study

The goal of this project is to predict NFL game outcomes using both old ELO rating methods and current machine learning techniques. The specific objectives are listed below:

1. Implement an ELO Rating System:

   Develop an ELO rating system using a Neo4j graph database to track team performance dynamically over time. The system will update team ratings after each game based on the actual result versus the expected result, allowing for dynamic adjustments in team strength.

2. Predict NFL Game Outcomes Using Machine Learning:

   Employ machine learning models to predict NFL game outcomes based on historical data. The models will use features derived from the last 10 games' averages, including metrics such as scores, turnovers, and time of possession.

3. Evaluate Prediction Accuracy:

   Evaluate the accuracy of the predictions made by both the ELO rating system and machine learning models. The accuracy will be measured by comparing predicted outcomes with actual results from the last 500 games in the dataset.

4. Compare Feature Sets:

   Compare the effectiveness of different feature sets (top 10 features vs top 25 features) in improving model performance. This comparison will help determine which features are most important.

Data Collection

The dataset used in this study consists of NFL regular-season games played between 2009 and 2018. The data was collected from publicly available NFL play-by-play datasets, which provide detailed information about each game, including team performance metrics and game outcomes. The dataset contains a total of 2,510 games, with each row representing a single play in the game and its associated characteristics.

Data Description

Each record in the dataset is a play which includes the following key variables:

- Game ID (game_id): A unique identifier for each game.
- Home Team (home_team): The name of the team playing at home.
- Away Team (away_team): The name of the team playing away.
- Home Score (score_home): The total points scored by the home team.
- Away Score (score_away): The total points scored by the away team.
- Game Date (game_date): The date when the game was played.
- Turnovers: The number of times each team lost possession of the ball due to fumbles or interceptions.
- Passing Completions: The number of successful pass completions made by each team.
- First Downs Due to Penalties: The number of first downs awarded to each team due to penalties committed by the opposing team.
- Time of Possession: The amount of time each team controlled the ball during a drive. etc.

There are 255 characteristics play-by-play in the raw data of which 49 important characteristics per team were derived from the play-by-play data for each game.

Data Preprocessing: Calculating Features for Last 10 Games Averages from Raw Data

The process of calculating features for the last 10 games averages from the raw NFL play-by-play data involves several steps, from data cleaning and transformation to feature engineering. This section details the entire process, including how some games were dropped due to insufficient historical data (i.e., fewer than 10 prior games), and why the last 10 games averages were chosen as features for machine learning (ML) and deep learning (DL) models.

Step 1: Loading Raw Data

The raw data used in this study consists of detailed play-by-play information for each NFL game played between 2009 and 2018. Each row in the dataset represents a single play within a game, with variables such as:

- Game ID (game_id): A unique identifier for each game.
- Home Team (home_team): The team playing at home.
- Away Team (away_team): The team playing away.
- Play Type: Whether the play was a pass, run, or special teams play.
- Yards Gained: The number of yards gained on each play.
- Down: The down number during the play (1st, 2nd, 3rd, or 4th down).
- Turnovers: Whether the ball was lost to the opposing team through an interception or fumble.
- Time of Possession: How long a team controlled the ball during a drive.

These variables are aggregated at the game level to create summary statistics that reflect overall team performance in each game.

Step 2: Aggregating Play-Level Data into Game-Level Metrics

The play-by-play data can be grouped into various categories such as offensive metrics, defensive metrics, and special teams metrics. These categories help us understand different aspects of team performance. For example:

- Offensive Metrics: Include total yards gained, passing completions, and time of possession.
- Defensive Metrics: Include turnovers forced and sacks recorded.
- Special Teams Metrics: Include field goals made and punt returns.

By aggregating these play-by-play metrics into game-level statistics as done by (Bosch, 2018), we can create summary features that reflect overall team performance in each game. For example:

- Total Yards Gained: Summing all yards gained by a team during the game gives an indication of offensive performance.
- Turnover Margin: The difference between turnovers committed and turnovers forced by a team can be a strong predictor of game outcomes.
- Time of Possession: Teams that control possession for longer periods often dominate games by keeping their defense rested and their offense on the field.

These aggregated metrics are then used to calculate post-priori characteristics.

Step 3: Calculating Post-Priori Characteristics

Once the raw data is aggregated into game-level metrics, we calculate post-priori characteristics that summarize team performance over their last 10 games. These characteristics include:

- Average Points Scored (Last 10 Games): The average points scored by a team over their last 10 games.
- Average Points Allowed (Last 10 Games): The average points allowed by a team over their last 10 games.
- Turnover Margin (Last 10 Games): The difference between turnovers committed and turnovers forced by a team over their last 10 games.
- Time of Possession (Last 10 Games): The average time of possession for a team over their last 10 games.

These post-priori characteristics are simply the averages of the game-level metrics calculated earlier. They provide a more comprehensive view of team performance over time rather than focusing on individual games. By averaging performance metrics over the last 10 games, we capture trends in recent form while smoothing out anomalies from individual games.

Step 4: Dropping Games with Insufficient Historical Data

One challenge in calculating last 10 games averages is that some teams may not have played enough prior games to calculate these averages. For example, teams playing in the first few weeks of the 2009 season do not have sufficient historical data (i.e., fewer than 10 previous games). As a result, these games are dropped from the dataset to ensure that all remaining records have valid last 10 games averages. This step ensures that our machine learning models are trained on complete data without missing values for key features.

Step 5: Feature Selection and Engineering

After calculating post-priori characteristics, we select the top features based on their correlation with game outcomes. These features include:

- Points Scored (Last 10 Games): Teams that consistently score more points tend to win more games.
- Points Allowed (Last 10 Games): Teams that allow fewer points tend to have stronger defenses and win more games.
- Turnover Margin (Last 10 Games): Teams with positive turnover margins often have more scoring opportunities and are more likely to win.
- Time of Possession (Last 10 Games): Teams that control possession for longer periods often dominate games by keeping their defense rested and their offense on the field.

These features are used as inputs for both ELO calculations and machine learning models.

Why Choose Last 10 Games Averages?

The decision to use last 10 games averages as features is supported by research from Owen and Galle (2017), who demonstrated that recent performance is a strong predictor of future outcomes in sports analytics. By focusing on the most recent games, we capture current team form and momentum, which are critical factors in predicting future success. Owen and Galle argue that "recent performance metrics provide more accurate predictions than season-long averages because they reflect short-term trends such as player injuries or tactical adjustments" (Owen & Galle, n.d.).

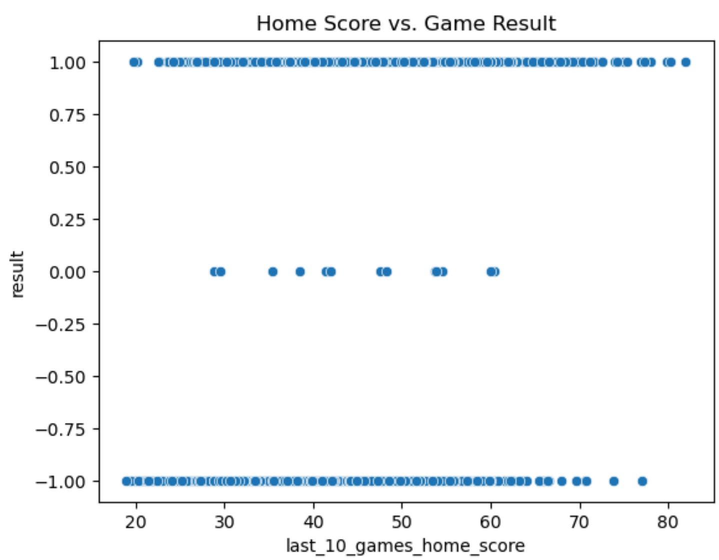
In our case, using the last 10 games averages allows us to account for fluctuations in team performance throughout the season while still providing enough historical context to make reliable predictions.

Exploratory Data Analysis (EDA)

The Exploratory Data Analysis (EDA) phase plays a crucial role in understanding the relationships between various features in the dataset and how they influence game outcomes. In this project, several visualizations were created to explore the distribution of key features, their correlation with game results, and their predictive power.

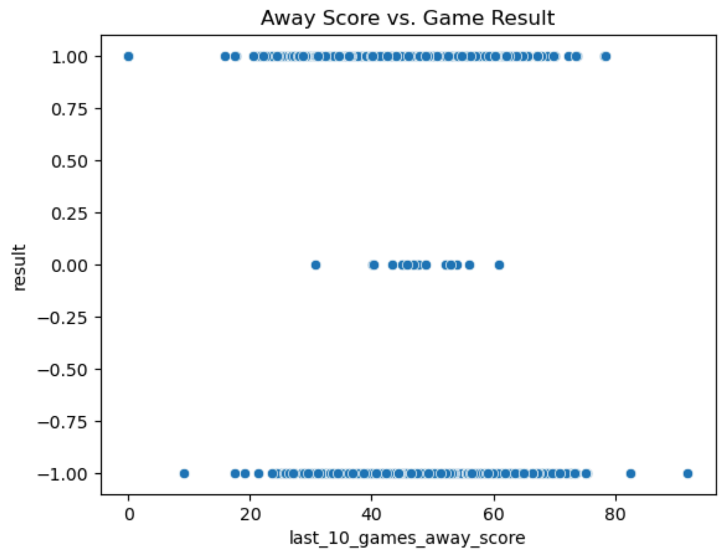
1. **Scatter Plots: Home and Away Scores vs. Game Result**
   - **Home Score vs. Game Result:** The scatter plot shows the relationship between the home team's score over their last 10 games (last_10_games_home_score) and the game result (result), where 1 represents a home win, 0 represents a tie, and -1 represents an away win.

     Key Observation: There is a clear pattern where higher home scores are associated with home wins (result = 1). Conversely, lower home scores are more likely to result in away wins (result = -1). Ties tend to occur within a narrower range of scores.

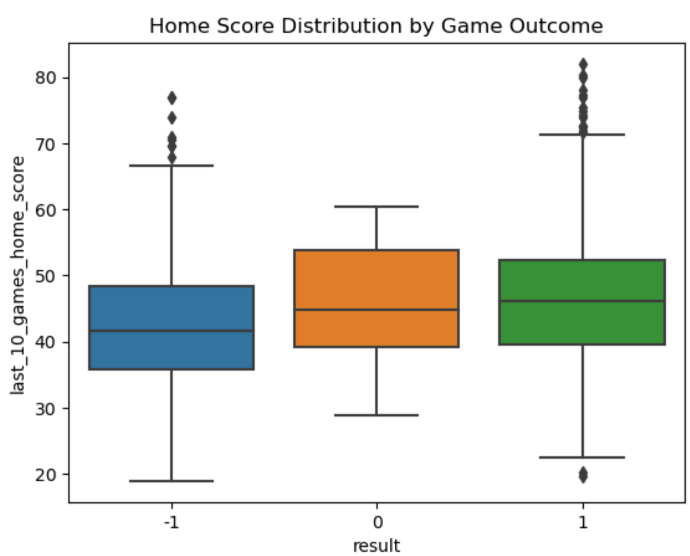


   - **Away Score vs. Game Result:** Similarly, the scatter plot for away scores (last_10_games_away_score) shows that higher away scores are associated with away wins (result = -1), while lower away scores are more likely to result in home wins (result = 1).

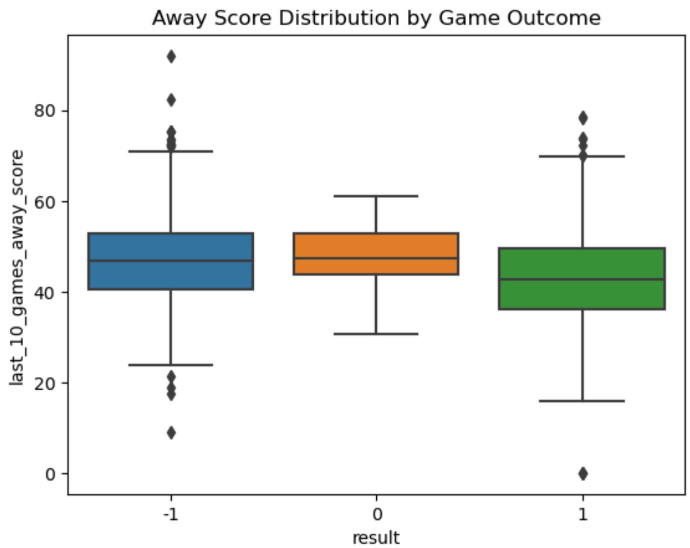
     Key Observation: The spread of away scores is broader compared to home scores, suggesting that away teams experience greater variability in performance. However, higher away scores still correlate with better outcomes for the away team.



2. **Box Plots: Score Distribution by Game Outcome**
   - Home Score Distribution by Game Outcome: The box plot compares the distribution of home team scores across different game outcomes (home win, tie, away win).
     - Key Observation: Home teams that win games tend to have higher median scores compared to those that lose or tie. The interquartile range (IQR) for home wins is also wider, indicating that winning home teams can score a broader range of points.
     - Outliers: Several outliers are observed for home wins where teams scored significantly more than the typical range (above 70 points). These outliers represent exceptionally high-scoring games.



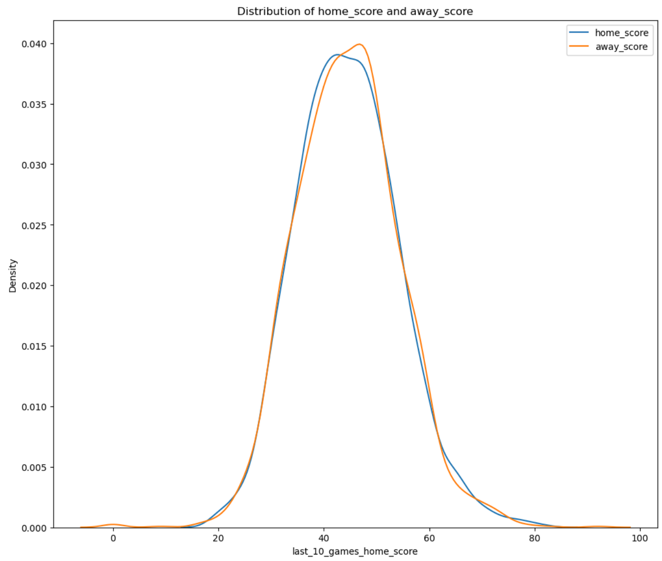
- Away Score Distribution by Game Outcome: The box plot for away team scores shows a similar pattern, where winning away teams tend to score more points compared to losing or tying teams.
    - Key Observation: Away teams that win generally have higher median scores compared to those that lose or tie. However, the IQR for away wins is narrower than that for home wins, suggesting that winning away teams tend to perform within a more consistent range of scores.
    - Outliers: There are several outliers where losing away teams scored very few points (below 20),



indicating poor offensive performance in those games.

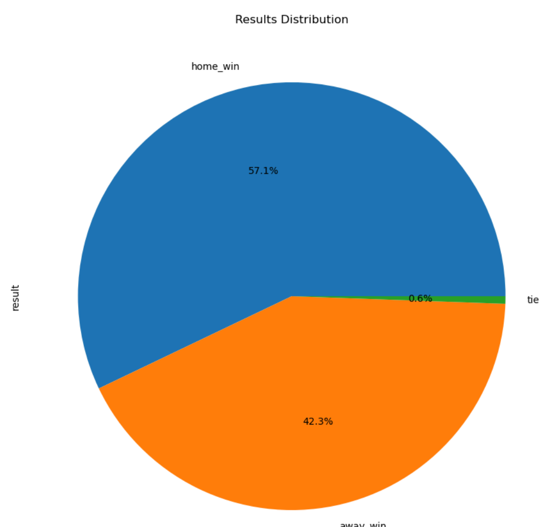
3. Distribution of Home Score and Away Score

- The KDE plot compares the distribution of scores for home teams (last_10_games_home_score) and away teams (last_10_games_away_score).
- Key Observations:
    - The distributions for both home and away scores are very similar, with slight differences indicating that home teams tend to score slightly more than away teams.
    - Both distributions are centered around a mean score of approximately 40 points, with a slight skew toward higher scores for home teams.
    - The tails of the distributions indicate that extreme high or low scores are relatively rare but can occur for both home and away teams.



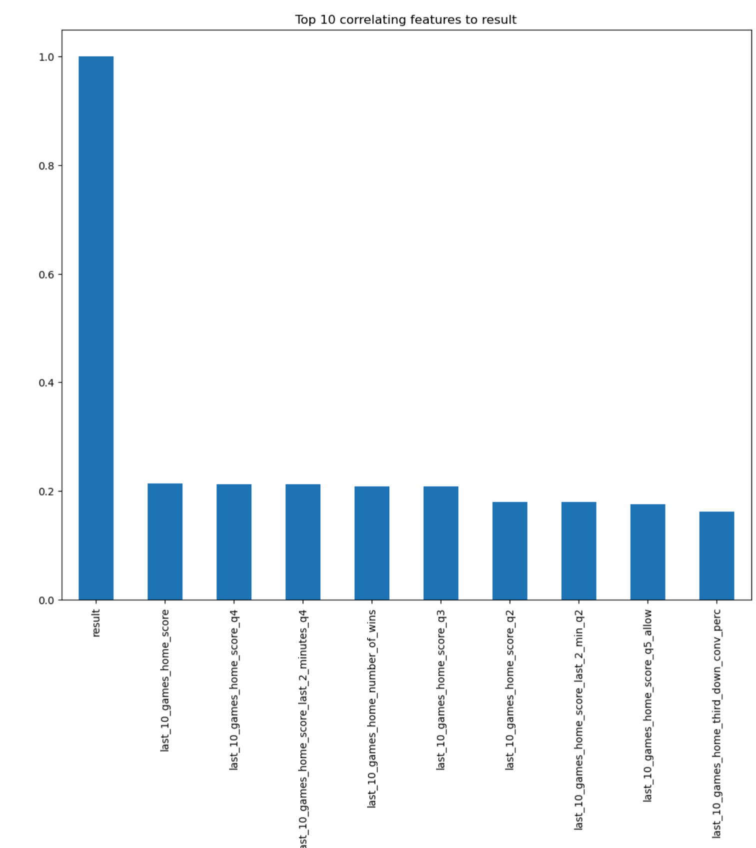
4. Results Distribution

- The pie chart shows the distribution of game outcomes: home win, away win, and tie.
- Key Observations:
  - Home Wins Dominate: Home wins account for 57.1% of all games, which supports the hypothesis of a home-field advantage in NFL games.
  - Away Wins: Away wins make up 42.3% of all games, showing that while home teams have an advantage, away teams still win a significant portion of games.
  - Ties Are Rare: Ties account for only 0.6% of all games, reflecting how uncommon tied outcomes are in NFL regular-season games.



5. Top 10 Correlating Features to Result

- The bar chart shows the top 10 features that have the highest correlation with the game result (win, loss, or tie).
- Key Observations:
    - The last_10_games_home_score has a strong correlation with the game result, which is expected since teams that score more points are more likely to win.
    - Quarter-Specific Scores: Features like last_10_games_home_score_q4 (home team's score in the 4th quarter) and last_10_games_home_score_last_2_minutes_q4 (home team's score in the last two minutes of the 4th quarter) are also highly correlated with game outcomes. This suggests that performance in critical moments of the game, especially late-game scoring, plays a significant role in determining the outcome.
    - Wins and Momentum: The feature last_10_games_home_number_of_wins also shows a strong correlation, indicating that teams with recent wins tend to maintain momentum and perform better in subsequent games.
    - Defensive Metrics: Features like last_10_games_home_score_q3_allow (points allowed by the home team in the 3rd quarter) and last_10_games_home_third_down_conv_perc (third-down conversion percentage) show moderate correlations. This suggests that defensive strength and efficiency on critical downs are important factors for winning games.



6. Additional Insights from EDA

Based on these visualizations, several insights can be drawn:

- Score Disparity and Outcome: Both scatter plots and box plots confirm that higher team scores (whether home or away) are strongly correlated with winning outcomes. This reinforces the importance of offensive performance as a key predictor of game results.
- Variability in Away Team Performance: Away teams exhibit greater variability in both their scores and outcomes compared to home teams. This suggests that while some away teams can perform exceptionally well and win games, others struggle significantly.
- Critical Performance Moments: The scatter plots indicate that there is a threshold score above which teams are much more likely to win. For example, home teams scoring above 50 points rarely lose games

Hypotheses for the Study

Based on the visualizations and insights provided from the Exploratory Data Analysis (EDA) phase, several hypotheses can be formulated for the overall study. These hypotheses can either be verified through further statistical analysis or commented upon based on the trends observed in the data.

1. **Home-Field Advantage Hypothesis**
    - Statement: Teams playing at home are more likely to win than teams playing away.
    - Rationale: The scatter plots and box plots for home scores vs. game result indicate that higher home scores are strongly associated with home wins (result = 1). Additionally, the pie chart showing the distribution of game outcomes reveals that 57.1% of all games result in home wins, supporting the hypothesis of a home-field advantage.
2. **Offensive Performance Hypothesis**
    - Statement: Teams that score more points are more likely to win, regardless of whether they are playing at home or away.
    - Rationale: Both scatter plots (home and away scores vs. game result) confirm that higher team scores—whether by the home or away team—are strongly correlated with winning outcomes. This reinforces the importance of offensive performance as a key predictor of game results.
3. **Variability in Away Team Performance Hypothesis**
    - Statement: Away teams exhibit greater variability in performance compared to home teams.
    - Rationale: The scatter plot for away scores shows a broader spread compared to home scores, suggesting that away teams experience greater variability in performance. Some away teams perform exceptionally well and win games, while others struggle significantly.
4. **Defensive Strength Hypothesis**

- Statement: Teams that allow fewer points (stronger defenses) are more likely to win.
- Rationale: Features related to defensive performance, such as last_10_games_home_score_q3_allow (points allowed by the home team in the 3rd quarter), show moderate correlations with game outcomes. This suggests that defensive strength plays an important role in determining success, especially in critical moments like third downs or late-game situations.

5. **Critical Performance Moments Hypothesis**
    - Statement: Teams that perform well in critical moments (e.g., 4th quarter or last 2 minutes) are more likely to win.

- Rationale: Features like last_10_games_home_score_q4 (home team's score in the 4th quarter) and last_10_games_home_score_last_2_minutes_q4 (home team's score in the last two minutes of the 4th quarter) are highly correlated with game outcomes. This suggests that performance in critical moments of the game, especially late-game scoring, plays a significant role in determining the outcome.

6. **Momentum Hypothesis**

- Statement: Teams with recent wins tend to maintain momentum and perform better in subsequent games.
- Rationale: The feature last_10_games_home_number_of_wins shows a strong correlation with game outcomes, indicating that teams with recent wins tend to continue performing well. This supports the idea that momentum from previous victories can influence future success.

Machine Learning (ML) and Deep Learning (DL) Models

In this section, we will discuss the machine learning (ML) models and deep learning (DL) models used in the project, focusing on their architecture, hyperparameters, and results. Additionally, the ELO rating system is also discussed as a traditional method for predicting NFL game outcomes. Each approach has its strengths and weaknesses, and they were evaluated based on their ability to predict game outcomes accurately.

1. Random Forest Classifier (Machine Learning Model)

Architecture and Hyperparameters:

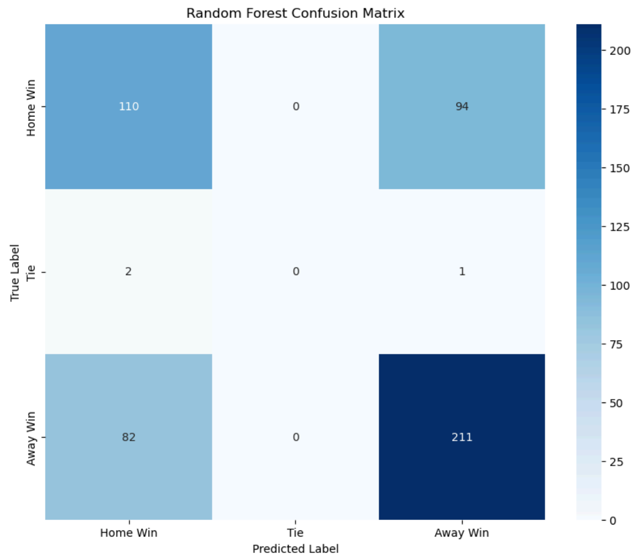
- The Random Forest model was tuned using a grid search with cross-validation to optimize hyperparameters. The best parameters were:
    - n_estimators: 200
    - max_depth: 10
    - min_samples_split: 5
    - min_samples_leaf: 2
- These parameters improved the model's performance on the test set.

Results:

- With hyperparameter tuning, the Random Forest model achieved a test accuracy of 64.2% using the top 25 features.
- The confusion matrix for the Random Forest model revealed that:
    - Home wins were predicted with high precision.
    - Away wins were moderately well-predicted.
    - Ties remained challenging to predict due to their rarity in the dataset.

Key Observations:

- High precision for home win predictions.
- Moderate accuracy for away win predictions.
- Poor performance in predicting ties due to their low frequency.



2. Recurrent Neural Network (RNN) - Deep Learning Model

Architecture and Hyperparameters

The Recurrent Neural Network (RNN) is a type of neural network designed to handle sequential data by maintaining a hidden state that captures information from previous time steps. In this project, an RNN with three layers was used to predict game outcomes based on sequential data from the last 10 games. A similar implementation for football by (Pettersson & Nyquist, 2017) was a reference for us.

- Input Layer: The input layer consists of features derived from the last 10 games' averages for both home and away teams.
- Hidden Layers: The RNN consists of three hidden layers, each with 25 units. The activation function used is tanh, which helps capture non-linear relationships between features while maintaining smooth gradients during backpropagation.
- Output Layer: The output layer consists of three units representing the three possible outcomes: home win (0), tie (1), and away win (2). A softmax activation function is applied to convert raw scores into probabilities for each class.

- The RNN model was tuned using a grid search over the following hyperparameters:
    - hidden_size: [25, 50]
    - num_layers: [1, 2, 3]
    - dropout: [0.0, 0.2, 0.4]
    - learning_rate: [0.001, 0.01]

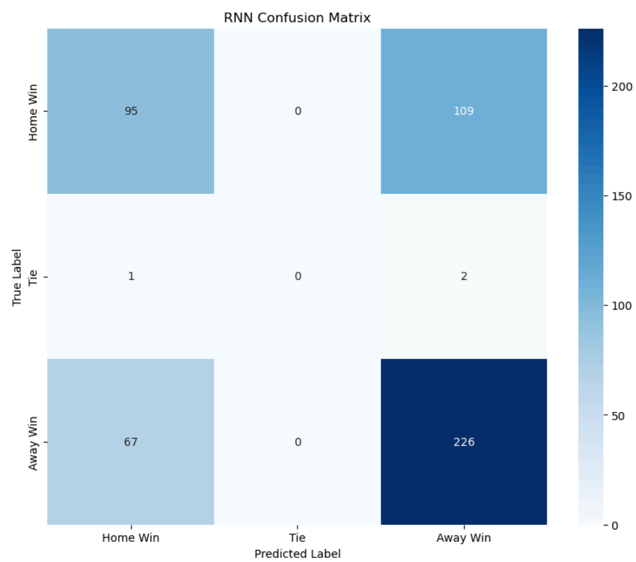
- The best configuration was:
    - hidden_size: 25
    - num_layers: 2
    - dropout: 0.2
    - learning_rate: 0.001

Results

- After hyperparameter tuning, the RNN achieved a test accuracy of 64.2%, matching the performance of the tuned Random Forest model.
- The confusion matrix for the RNN showed similar trends to the Random Forest model:
    - Strong performance in predicting home wins.
    - Moderate performance for away wins.
    - Limited success in predicting ties.

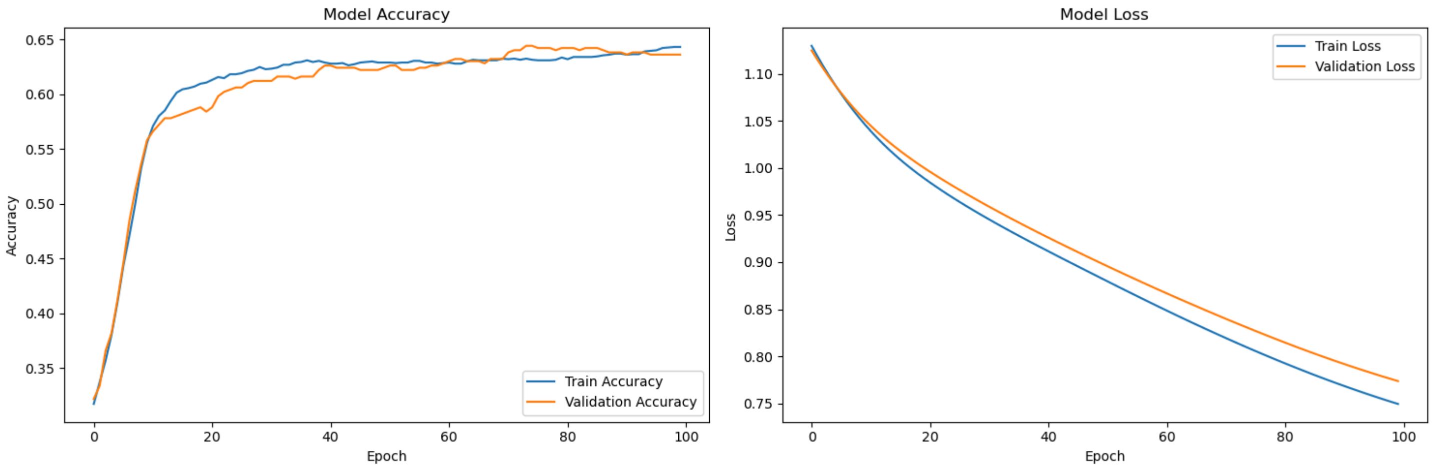
Key Observations:

- Similar trends to Random Forest, with strong performance for home wins but difficulty in predicting ties.



Training Progress

- Plots of training loss and accuracy over epochs showed steady convergence.
- Final training loss: 0.7411
- Final test loss: 0.6788



While RNNs are powerful for capturing temporal dependencies in sequential data, they did not significantly outperform simpler models like Random Forest in this case. This could be due to limited sequential complexity in the dataset or insufficient tuning of hyperparameters such as learning rate or batch size.

Feature Selection: Top Features vs. All Features

Feature Sets Evaluated

1. Top 10 Features:
    - Selected based on correlation with game outcomes.
    - Includes metrics like last_10_games_home_score, last_10_games_home_score_q4, and average_point_diff.
    - Achieved an accuracy of 61% with Random Forest.
2. Top 25 Features:
    - Includes additional metrics such as turnovers, time of possession, and defensive metrics (last_10_games_home_total_turnovers, last_10_games_away_yards_gained).
    - Achieved an accuracy of 64.2%, outperforming both top 10 and all features.

3. All Features:
    - Includes all numerical features from the dataset.
    - Achieved an accuracy of 61%, indicating that including too many features may introduce noise and reduce model performance.

Reason for Choosing Top 25 Features

The top 25 feature set struck a balance between including important predictive metrics and avoiding overfitting caused by irrelevant or redundant features. Metrics like turnovers and time of possession added significant predictive power compared to the top 10 feature set.

3. ELO Rating System

Overview

The ELO rating system (Silver, 2018) is a traditional method used to rank players or teams based on their performance in head-to-head competitions. In this project, ELO ratings were calculated for each team after every game, with ratings being updated based on actual game results versus expected results.

Calculation Process

For each game:

1. The current ELO ratings for both teams are retrieved.
2. Expected win probabilities are calculated using the following formula:

$$E_A = \frac{1}{1 + 10^{\frac{(R_B - R_A)}{400}}}$$

where:

- E_A is the expected probability of team A winning
- R_A and R_B are the current ratings of teams A and B respectively

The rating update formula used is:

$$R'_A = R_A + K(S_A - E_A)$$

where:

- K is the k-factor (set to 20 in our implementation)
- S_A is the actual score (1 for win, 0.5 for tie, 0 for loss)
- E_A is the expected score

The new ratings are stored in a Neo4j graph database, where each team is represented as a node and each game as a relationship between nodes.

Results

The ELO rating system achieved a prediction accuracy of:

- 63% on predicting outcomes for the last 500 games in the dataset.

This result demonstrates that ELO ratings can effectively capture team strength over time and provide reasonably accurate predictions for future games.

Conclusion:

Each model—Random Forest, RNN, and ELO—has its strengths:

- The Random Forest classifier performed well with limited feature engineering, achieving an accuracy of around 64%.
- The RNN model captured sequential dependencies but did not outperform simpler models.
- The ELO rating system provided dynamic insights into team performance over time and achieved a prediction accuracy of around 63%.

These results suggest that combining traditional methods like ELO with modern machine learning techniques could lead to even better predictions in future iterations of this project.

References

Bosch, P. (2018). *Predicting the winner of NFL-games using Machine and Deep Learning.* Vrije universiteit, Amsterdam.

Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics*, *15*(1), 27–33. https://doi.org/10.1016/j.aci.2017.09.005

Contributors to Wikimedia projects. (2024, October 30). *American football rules*. Wikipedia. https://en.wikipedia.org/wiki/American_football_rules

Elo rating system - Chess terms. (n.d.). Chess.Com. Retrieved November 15, 2024, from https://www.chess.com/terms/elo-rating-chess

Owen , Z., & Galle, V. (n.d.). *Predicting the NFL*. https://vgalle.github.io/files/NFLReport.pdf

Pettersson, D., & Nyquist, R. (2017). *Football Match Prediction using Deep Learning* [CHALMERS UNIVERSITY OF TECHNOLOGY Gothenburg, Sweden]. https://odr.chalmers.se/server/api/core/bitstreams/4ad5c2f4-f2bc-477d-9397-8b179b5a893f/content

Regan, K. W., & Haworth, G. M. (2011, January 1). *Intrinsic chess ratings*. Unknown. https://www.researchgate.net/publication/221606300_Intrinsic_Chess_Ratings

Silver, N. (2018, September 5). How our NFL predictions work. *FiveThirtyEight*.

https://fivethirtyeight.com/methodology/how-our-nfl-predictions-work/