**Bayesian Analysis of NFL Sports Betting Outcomes**

**By Xiaofan Jiao**

**Abstract**

This study employs Bayesian methods alongside logistic regression and random forest models to predict win probabilities in sports betting, focusing on team matchups, game-level conditions, and potential biases. By analyzing historical win/loss records and game-specific features, the Bayesian approach iteratively updates prior beliefs using season-to-date results. In comparison, logistic regression and random forest models provide parametric and non-parametric benchmarks, respectively. Rigorous data preprocessing includes the imputation of missing values through multi-layered strategies based on temporal and spatial trends. Results aim to highlight model calibration, predictive power, and actionable insights for identifying mispriced betting odds. The study’s findings underscore the potential for leveraging probabilistic models in sports betting to uncover exploitable inefficiencies.

1. **Introduction**

Sports betting has evolved into a complex market where accurate win probability predictions can yield significant advantages. Central to this is the development of predictive models that consider both historical team performance and game-specific conditions such as weather and spread magnitude. Bayesian models offer a robust framework for incorporating prior information and updating it iteratively as new data becomes available, making them particularly well-suited for dynamic domains like sports betting.

The primary goal of this study is to estimate win probabilities for team matchups, evaluate model calibration, and identify scenarios where betting odds diverge from model predictions. The dataset was downloaded from Kaggle (Crabtree, n.d.).Special attention is given to biases such as home advantage and weather effects, as well as the role of spread magnitude in shaping win probabilities. The study’s contribution lies in comparing the performance of Bayesian, logistic regression, and random forest models in the context of sports betting, providing insights into their relative strengths and limitations.

1. **Data Processing**

To prepare the dataset for analysis, a structured, multi-layered approach was applied to address missing values and standardize variables, ensuring data quality and consistency for predictive modeling. The dataset was initially filtered to include only events occurring on or before September 2024, restricting the analysis to relevant historical data that aligns with the study's objectives.

A significant challenge was the presence of missing values in weather-related variables (`weather\_temperature`, `weather\_wind\_mph`, `weather\_humidity`). To address this, a hierarchical imputation strategy was implemented. Missing weather data were imputed based on three levels of granularity: median values for the same stadium, month, and year; median values for the same stadium and month (irrespective of the year); and median values for the same month and year across all stadiums. This approach ensured that imputation leveraged the most specific context available, preserving the spatial and temporal trends inherent in the data.

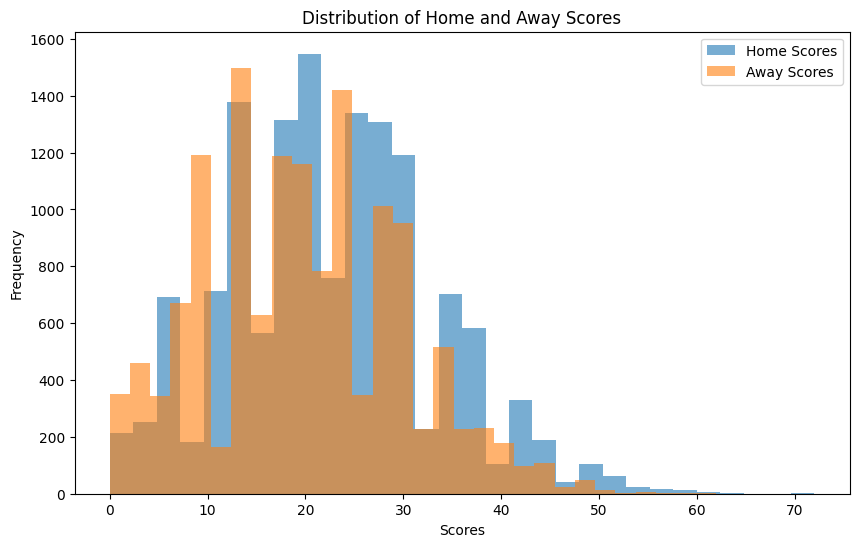
For game-specific variables, such as `spread\_favorite` and `over\_under\_line`, missing values were handled using matchup-based imputation. Median values were calculated for each unique combination of home and away teams, enabling imputation based on historical trends for the same matchups. In cases where matchup data were unavailable, missing values remained unfilled at this stage, allowing for further imputation using dataset-wide trends.

Finally, any remaining missing values were filled using overall medians calculated across the dataset. Weather variables and game-specific features were standardized back to numeric data types after imputation, ensuring compatibility with subsequent modeling stages.

1. **Exploratory Data Analysis (EDA)**

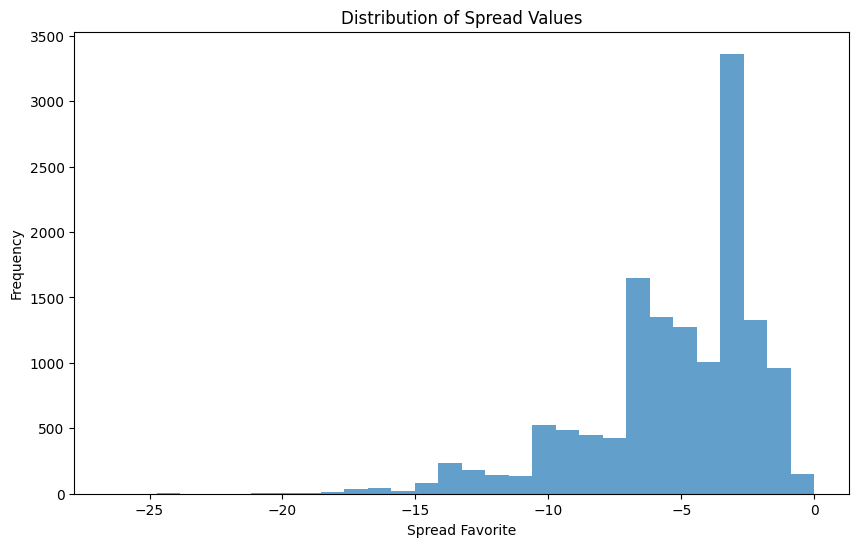
3.1 Distribution of Scores (Home vs. Away Teams)

The distribution of home and away team scores reveals key patterns in scoring dynamics. Home teams typically exhibit a slightly higher frequency of scores in the mid-range (20-30 points), reflecting a potential home-field advantage. Away scores, while following a similar distribution, show a marginally lower frequency in higher score ranges. This finding aligns with the hypothesis that home-field conditions positively impact team performance, likely due to factors such as crowd support, familiarity with the field, and reduced travel fatigue.



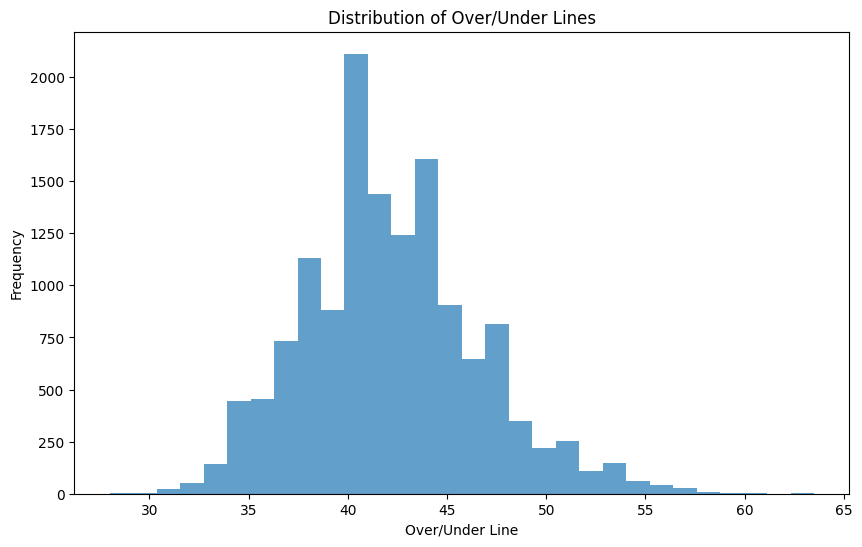
3.2 Spread Favorite Distribution

The spread favorite, representing the anticipated winning margin of the favorite team, exhibits a skewed distribution towards negative values. This indicates that most games involve a projected favorite with a winning margin of less than 10 points. Extreme negative values (e.g., -20 to -25) are rare, reflecting matchups where one team is overwhelmingly favored. This distribution is crucial for calibrating predictive models, as extreme values could introduce bias if not appropriately accounted for.



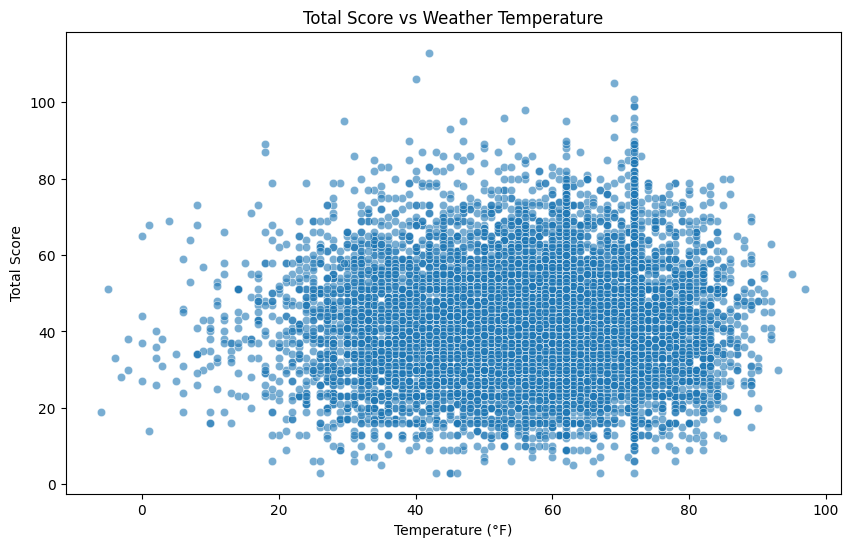
3.3 Over/Under Line Distribution

The over/under line, reflecting the expected combined score for both teams, displayed a bell-shaped distribution centered around 40-45 points. This suggests that the majority of games are anticipated to be moderately high-scoring. Outliers on both ends highlight games with either exceptionally strong defensive performances (lower scores) or explosive offensive matchups (higher scores).



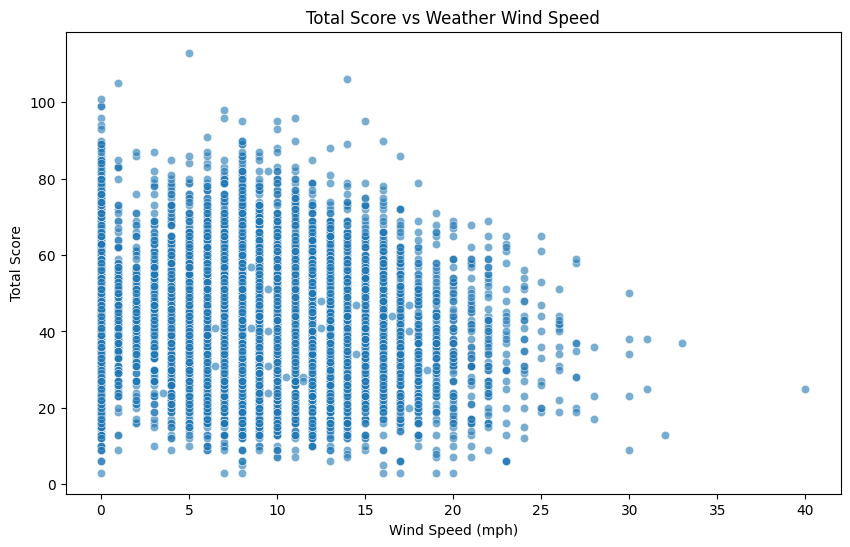
3.4 Relationship Between Total Score and Weather Temperature

The scatterplot between total scores and weather temperature indicates no strong linear relationship, suggesting that scoring is not heavily influenced by temperature alone. However, a denser clustering around mid-range temperatures (40°F to 70°F) might hint at optimal playing conditions that facilitate normal scoring patterns. Extreme temperatures (both low and high) do not appear to significantly alter scoring distributions, though further statistical analysis could confirm this observation.



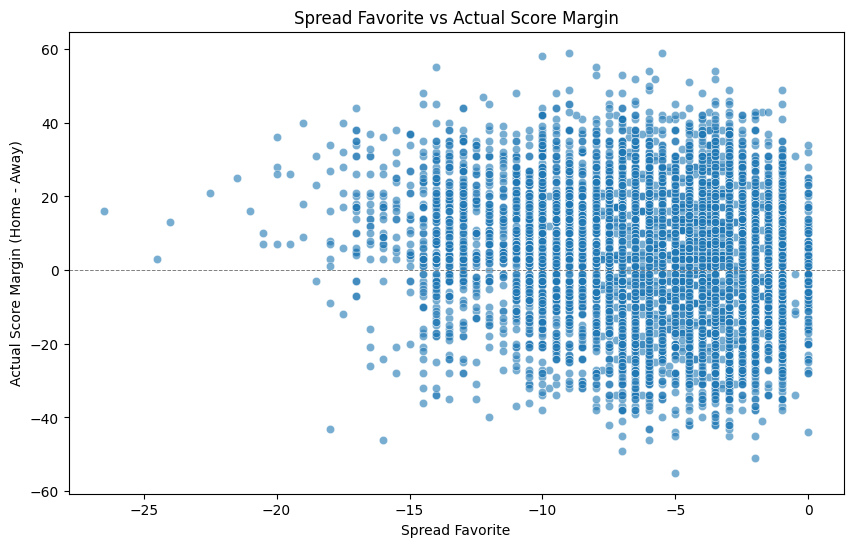
3.5 Relationship Between Total Score and Weather Wind Speed

Total scores tend to decrease slightly as wind speed increases, particularly beyond 15 mph. This relationship aligns with expectations, as high wind speeds can disrupt passing plays and field goal attempts, reducing scoring opportunities. The dispersion of scores remains consistent at lower wind speeds, suggesting that moderate wind conditions do not impose significant restrictions on gameplay.



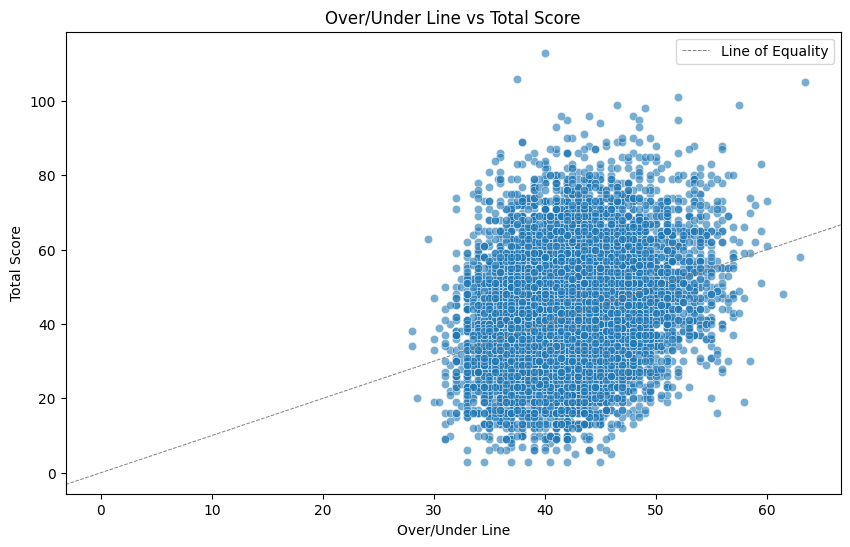
3.6 Spread Favorite vs. Actual Score Margin

The relationship between spread favorite and actual score margin reveals a clear trend: as the spread favorite becomes more negative (indicating a stronger anticipated favorite), the actual score margin tends to reflect larger victories for the favorite. However, there is considerable variance, particularly in games with spreads near zero, where outcomes are less predictable. This finding underscores the importance of accounting for uncertainty in predictive models.



3.7 Over/Under Line vs. Total Score

The scatterplot of over/under line against total score demonstrates a positive correlation, as expected. Games with higher over/under lines generally result in higher total scores, though significant variance exists around the line of equality. This variance highlights instances where games deviate from betting expectations, offering potential opportunities for identifying mispriced bets.



1. **Methodology**

This section outlines the implementation of three predictive models—Bayesian hierarchical, logistic regression, and random forest models—used to estimate win probabilities for team matchups based on historical game outcomes and predictive features. A comparative analysis of these models was conducted to evaluate their performance and identify biases in their predictions.

4.1 Bayesian Win Probability Model

The Bayesian win probability model estimates the likelihood of a home team winning a game by incorporating team-specific strengths and league-wide variability through a hierarchical structure. This approach captures differences in team performance while explicitly modeling uncertainty, making it particularly effective for sports betting predictions. The dataset was prepared with an outcome variable for each game (1 for a home win, 0 for an away win), and teams were mapped to unique indices to enable the model to differentiate between them and attribute performance differences to team-specific factors.

The model is structured as a hierarchical Bayesian framework, integrating league-wide trends and team-specific variability. A global parameter represents the league-wide average team strength, while captures variability in team strengths around this average. Each team’s strength is modeled as a normal distribution centered on , allowing robust estimation even for teams with limited data. Game-level win probabilities are calculated using a logistic (sigmoid) function of the difference between the home and away teams’ strengths, ensuring stronger teams are more likely to win while preserving uncertainty in individual outcomes. Game results are modeled using a Bernoulli distribution, linking observed outcomes to the predicted probabilities.

Posterior distributions of model parameters were estimated using Markov Chain Monte Carlo (MCMC) sampling, with 3,000 samples and 2,000 tuning iterations to ensure convergence. The hierarchical structure enabled the model to incorporate prior knowledge, improving predictions for teams with sparse data. Posterior distributions quantified uncertainty, providing nuanced insights compared to deterministic models. Additionally, the model’s ability to update dynamically with new data makes it particularly valuable for capturing evolving team strengths and adapting to dynamic sports environments like betting markets.

4.2 Logistic Regression Model

The logistic regression model was implemented as a parametric baseline for predicting win probabilities, relying on linear relationships between predictors and outcomes. This approach provided interpretable results but was limited in its ability to capture non-linear interactions inherent in complex sports data. Despite these limitations, the model served as a straightforward and effective benchmark for comparison against more sophisticated methods. The model utilized both team-specific and game-level features to predict outcomes. Team-specific factors included the historical win rates of the home and away teams, providing a measure of each team’s past performance. Game-level features included the spread magnitude, home/away status, and weather conditions such as temperature, wind speed, and humidity. Additionally, the over/under line, which reflects the total expected points in a game, was included as a predictor to capture broader game dynamics.

4.3 Random Forest Model

The random forest model served as a non-parametric alternative for predicting win probabilities, leveraging an ensemble of decision trees to uncover complex, non-linear interactions between predictors. This flexibility made the model well-suited for capturing intricate patterns in sports data that might be overlooked by more restrictive parametric models like logistic regression. The model utilized the same set of predictors as the logistic regression model, incorporating both team-specific and game-level features. Team-specific factors, such as historical win rates for home and away teams, were combined with game-level details, including spread magnitude, home/away status, weather conditions (temperature, wind speed, and humidity), and the over/under line. This feature set allowed the model to analyze a comprehensive range of variables influencing game outcomes.

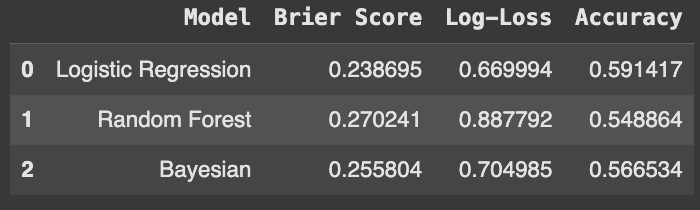
Training the random forest involved using historical game outcomes to construct an ensemble of decision trees, each trained on random subsets of data. By aggregating predictions across these trees, the model produced probabilistic win predictions for the home team. This ensemble approach enabled the random forest to capture intricate dependencies between variables, offering a flexible and powerful tool for win probability estimation. Despite its strengths, the random forest model had limitations, particularly in terms of interpretability. Unlike the logistic regression and Bayesian models, which offer more transparency in how predictions are derived, the random forest’s ensemble nature makes it challenging to isolate the contribution of individual predictors. This trade-off between complexity and interpretability is a key consideration in applying the model, particularly for scenarios where understanding predictor influence is critical. Nonetheless, its ability to handle non-linear relationships and detect complex patterns makes it a valuable addition to predictive modeling in sports analytics.

Together, these models provide a comprehensive framework for predicting win probabilities in sports betting. The Bayesian model offers robust probabilistic reasoning, logistic regression provides simplicity and interpretability, and random forests capture complex patterns, highlighting the complementary strengths of each approach.

1. **Results Interpretation**

5.1 Model Performance Comparison

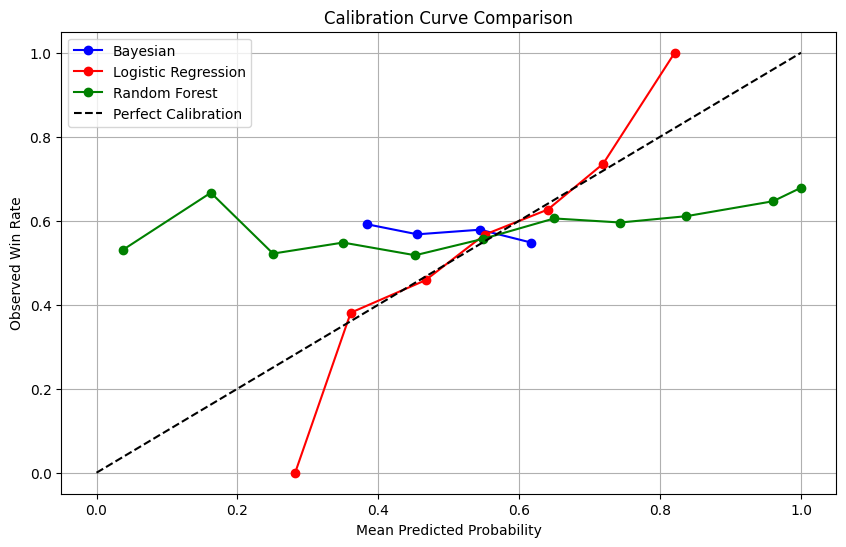
The models were evaluated using three metrics: Brier score (calibration), log-loss (probabilistic accuracy), and accuracy (binary outcome correctness). The results are summarized in the following table:



* Logistic Regression: Achieved the best overall calibration (lowest Brier score) and probabilistic accuracy (lowest log-loss). However, its accuracy was only slightly better than random guessing.
* Bayesian Model: Struck a balance between calibration and predictive accuracy, outperforming the random forest in both metrics while providing interpretable outputs.
* Random Forest: Exhibited higher error rates and weaker calibration, suggesting challenges in capturing the probabilistic nature of outcomes in a highly variable sports environment.

5.2 Calibration Analysis

A calibration curve was plotted to evaluate how well predicted probabilities aligned with observed win rates:

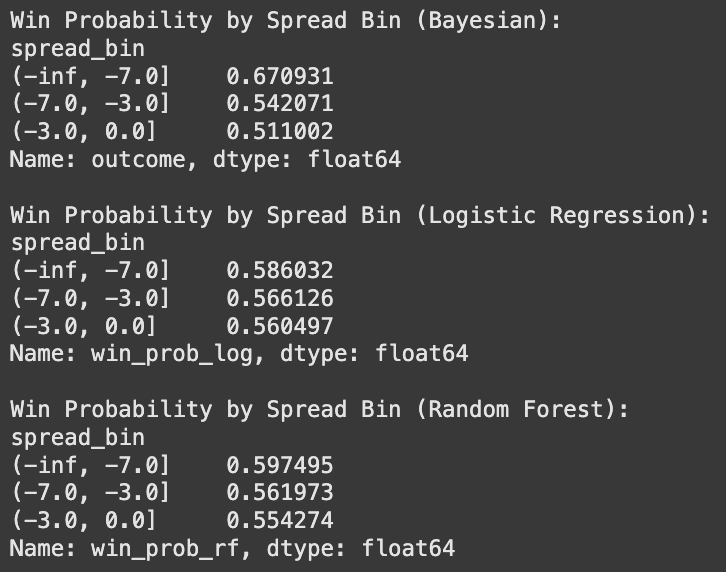


* The Bayesian model demonstrated strong calibration near the 50% win probability range, reflecting its strength in accounting for team-specific uncertainty.
* Logistic regression showed underconfidence at extreme probabilities (near 0 or 1), where it tended to overestimate uncertainty.
* Random forest predictions were less consistent, with notable deviations from the calibration line across probability ranges.

5.3 Identifying Potential Biases

To investigate potential biases, average win probabilities were calculated for different ranges of spread values:

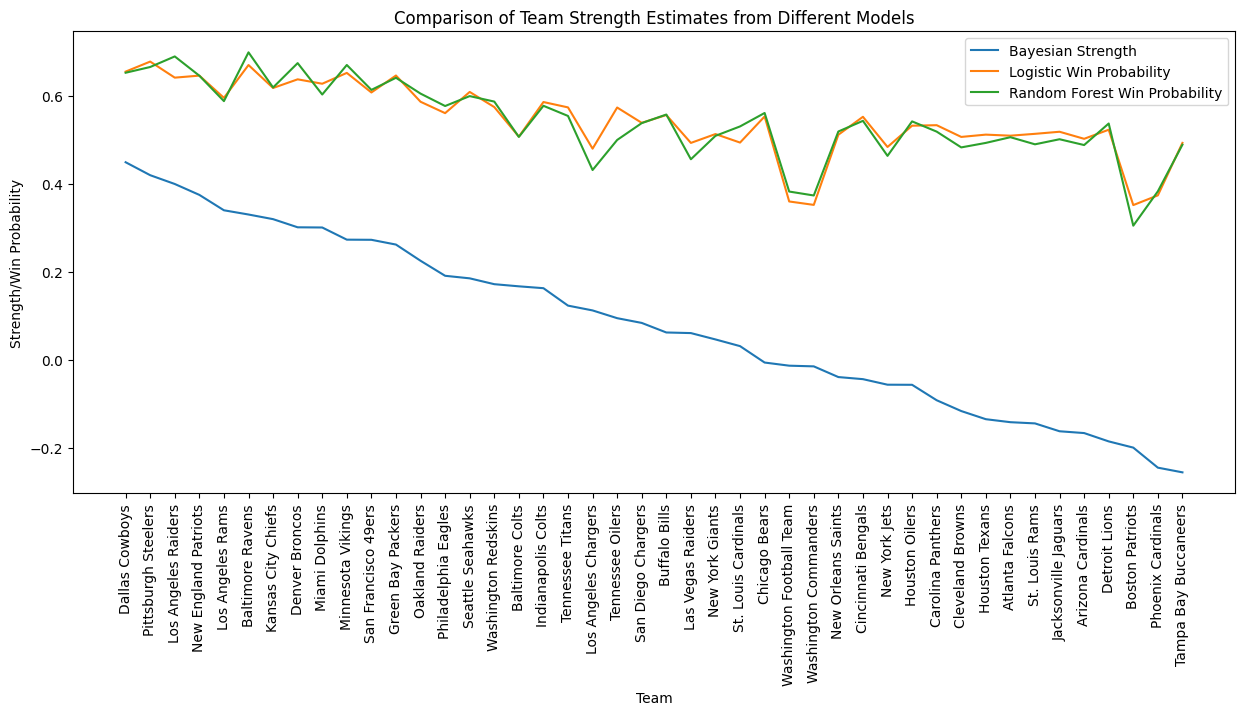
* Bayesian Model: Captured clear trends, with win probabilities decreasing as the spread became less favorable for the home team (e.g., smaller or positive spreads).
* Logistic Regression and Random Forest: Displayed similar trends but were less sensitive to spread magnitude changes, particularly for extreme spread ranges.



5.4 Comparison of Team Strength Estimates

The graph illustrates the estimated team strengths derived from the Bayesian model compared to the predicted win probabilities from the logistic regression and random forest models. The Bayesian model produces a broader range of team strengths, highlighting its ability to capture more nuanced distinctions between stronger and weaker teams. Conversely, the logistic regression and random forest models exhibit clustered predictions near the upper strength range, potentially reflecting limitations in their ability to account for team variability across the league.

This discrepancy suggests that the Bayesian model, with its hierarchical structure, is better equipped to estimate relative strengths when data availability varies across teams. For instance, weaker teams (e.g., Tampa Bay Buccaneers or Arizona Cardinals) are more distinctly separated in the Bayesian estimates, while the logistic regression and random forest models compress these differences. Such detailed differentiation is crucial in sports betting, where identifying underestimated or overestimated teams can lead to actionable insights. Additionally, the smoother progression in Bayesian estimates demonstrates its robustness in assigning meaningful strengths even to teams with limited data, reinforcing its value in dynamic betting markets.



1. **Application of the Bayesian Model for October 2024**

The Bayesian win probability model was applied to predict the outcomes of 60 games that occurred in October 2024. Real game outcomes were collected and compared with the model's predictions. The model correctly predicted the outcome of 34 games, achieving an accuracy of 57%. While this accuracy may seem modest, it aligns with expectations in a domain characterized by inherent uncertainty and high variability. In the context of sports betting, even small predictive advantages can provide meaningful opportunities for capitalizing on mispriced betting odds.

1. **Discussion and Conclusion**

Although the accuracy of the Bayesian model (57%) might appear low at first glance, it is important to contextualize this performance within the broader framework of sports betting. Sports betting markets are highly efficient, and even achieving a modest edge over betting odds can yield significant financial returns. The Bayesian model's predictive capability lies not solely in its raw accuracy but in its ability to identify mispriced bets by estimating true win probabilities with calibrated uncertainty.

The October 2024 analysis highlights the model's capacity to generate actionable insights, particularly in cases where betting odds deviate from predicted probabilities. This alpha—small, consistent advantages over the market—is a valuable resource for bettors. The Bayesian framework, with its focus on updating beliefs iteratively based on incoming data, is uniquely suited to dynamically adjust team strengths and improve predictions as the season progresses. Furthermore, the model's probabilistic nature enables more nuanced decision-making than binary win/loss predictions. By providing win probabilities, the model allows bettors to assess value in odds rather than focusing purely on outcomes, which is particularly valuable in betting scenarios where margins are tight.

1. **Recommendations and Future Studies**

Based on the findings of this study, several recommendations can enhance the practical utility of the Bayesian model in sports betting. A key strength of the model lies in identifying mispriced betting odds rather than solely achieving high accuracy. Bettors should focus on "value bets," where the model’s predicted probabilities significantly deviate from implied probabilities in the betting markets, such as estimating a 70% win probability when odds suggest only 50%. Future implementations should incorporate betting odds as a feature to refine predictions further and uncover systematic biases in odds setting, such as overestimation of home-field advantage. Additionally, the model’s probabilistic outputs can be weighted by confidence to prioritize high-certainty predictions, such as games with extreme spreads or significant disparities in team strength. By focusing on these value-driven strategies, bettors can maximize profitability while minimizing exposure to high-risk scenarios.

**9. Reference**

Crabtree, T. (n.d.). NFL scores and betting data [Dataset]. Kaggle. Retrieved November 29,

2024, from <https://www.kaggle.com/datasets/tobycrabtree/nfl-scores-and-betting-data?select=spreadspoke_scores.csv>