**Predicting NFL Game Outcomes: Data-Driven Insights for Game Predictions**

By: Juston Suell  
Course: DSC680 – Applied Data Science – Bellevue University

**Business Problem**

NFL games are inherently unpredictable, making accurate forecasting a challenge for analysts, coaches, and fantasy football enthusiasts. This project aims to leverage historical performance data, turnovers, and offensive efficiency to develop a machine learning model that predicts game results. By analyzing past data, the model identifies the most influential factors in determining the final score, helping users make more informed decisions.

**Background/History**

The field of sports analytics has grown exponentially in recent years, with data-driven insights becoming a cornerstone of team strategy and performance enhancement. Traditionally, predictions relied on expert opinions and historical trends, but modern techniques like machine learning offer a systematic approach to game analysis. By examining past NFL games, this project aims to uncover key patterns and improve prediction accuracy through statistical modeling.

**Data Explanation**

The data for this project was sourced from Pro-Football-Reference, which provides comprehensive statistics on team performance, player metrics, and game outcomes. The dataset was curated to focus on key performance indicators, including passing yards, rushing yards, and turnovers, which are strong predictors of game results.

**Methods**

The predictive models used in this project included Logistic Regression as a baseline, Random Forest for decision-tree-based pattern recognition, and XGBoost for its superior predictive capability. These models were trained on a dataset that included key performance variables such as passing efficiency, rushing balance, and turnover impact. The dataset was divided into training (80%) and testing (20%) subsets, with XGBoost emerging as the most effective model based on accuracy, precision, and recall.

**Analysis and Insights**

The analysis revealed several important findings:

1. Turnovers had the strongest impact on game outcomes. Teams with fewer turnovers significantly increased their probability of winning.
2. Passing efficiency correlated strongly with wins. Teams with higher passing yards generally performed better, highlighting the importance of quarterback play.
3. Rushing yards played a crucial role in maintaining a balanced offense. While not as dominant as passing efficiency, teams with stronger rushing attacks had a competitive advantage.

To support these insights, visualizations provided deeper context. Correlation heatmaps for the Cowboys and Ravens illustrated relationships between run-pass balance, total offense, and turnover impact. The Cowboys' heatmap showed a moderate negative correlation (-0.42) between total offense and turnover impact, while the Ravens' heatmap indicated a stronger negative correlation (-0.66). Additionally, a scatter plot of run-pass balance versus total offense demonstrated how offensive balance impacts yardage, with the Cowboys clustering around lower balance values and the Ravens showing a broader strategic range.

A screenshot of a graph

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

A diagram of a game

AI-generated content may be incorrect.

**Conclusion**

This project highlights the value of predictive modeling in NFL game analysis. By leveraging historical data and machine learning, the model identifies key factors that contribute to a team's success, providing actionable insights for analysts, coaches, and fantasy football players. These findings can help teams optimize strategies and improve decision-making both before and during games.

**Assumptions, Limitations, Challenges**

The model assumes that historical performance trends, such as the influence of turnovers and passing efficiency, will remain consistent in future games. However, unpredictable factors like mid-game injuries, coaching decisions, and referee calls were not included in the analysis. Additionally, data inconsistencies across seasons required careful feature engineering to ensure accuracy. While the model offers valuable insights, its reliance on historical data limits its ability to adapt to real-time game dynamics. Incorporating live player stats and mid-game adjustments could further refine predictions.

**Future Uses/Applications**

This approach has broad potential beyond football. It could be expanded to predict outcomes in basketball, baseball, or esports, where statistical analysis of team and player performance plays a critical role. Additionally, integrating real-time performance tracking could improve prediction accuracy for in-game decision-making, fantasy sports, and sports betting.

**Recommendations**

To improve game outcomes, teams should focus on reducing turnovers, as they are consistently linked to game success and play a critical role in determining results. Maintaining a balanced offensive strategy that optimizes both passing and rushing plays can significantly enhance overall efficiency. Additionally, teams should refine their offensive strategies by analyzing opponent weaknesses, optimizing play-calling based on defensive tendencies, and leveraging game data to enhance decision-making. Prioritizing these areas can provide teams with a strategic advantage in both preparation and execution.

**Ethical Assessment**

This project adheres to ethical standards by using publicly available data and avoiding proprietary information. The model’s limitations are clearly communicated to prevent misinterpretation, particularly in gambling-related contexts. Transparency and responsible use are essential as predictive analytics in sports continues to grow.

**Appendix**

Figures:

1. Cowboys Feature Correlation Heatmap – Highlights relationships between metrics like run-pass balance, total offense, and turnover impact.

A screenshot of a graph

AI-generated content may be incorrect.

1. Ravens Feature Correlation Heatmap – Shows similar correlations for the Ravens, with slightly different trends.

A screenshot of a graph

AI-generated content may be incorrect.

1. Scatter Plot of Run-Pass Balance vs. Total Offense – Demonstrates the impact of offensive strategy on total yardage.

A diagram of a game

AI-generated content may be incorrect.