Introduction

As economic uncertainty grows in the United States, an increasing number of Americans have turned to sports betting as a potential source of supplemental income. Many casual bettors believe their knowledge of basketball gives them an edge in predicting outcomes, leading them to place parlay bets, a high-risk, high-reward wager that combines multiple selections into a single ticket. However, the inherent volatility of sports makes long-term profitability difficult, as even well-informed bets can fail due to unpredictable factors such as injuries, officiating, or simple variance.

This project explores whether machine learning can provide a more disciplined approach to parlay betting by analyzing historical data, player performance, and situational trends. Rather than relying on intuition or anecdotal insights, the goal is to develop a data-driven model that identifies statistically favorable parlay combinations. By leveraging predictive algorithms, we aim to determine whether a systematic approach can outperform traditional betting strategies—or at least minimize the irrational decision-making that often leads to losses.

To achieve long-term profitability in sports betting, the fundamental challenge is not simply predicting game outcomes accurately but doing so more effectively than the sportsbooks' own models. Bookmakers employ sophisticated algorithms and vast datasets to set odds that inherently favor the house, ensuring a built-in margin known as the "vig" or "juice." For bettors to gain an edge, their predictive models must identify mispriced lines—instances where the true probability of an outcome exceeds the implied probability reflected in the odds. In the context of parlay betting, where the house edge compounds with each additional leg, this task becomes even more difficult. A successful machine learning approach must not only assess individual game probabilities with precision but also optimize parlay combinations in a way that exploits discrepancies between the sportsbook's lines and the model's expected value. By systematically identifying and capitalizing on these inefficiencies, the goal is to shift the odds in the bettor's favor, turning a profit over time despite the bookmakers' structural advantage.

Related Work

Recent advancements in machine learning have enabled data-driven approaches to sports betting. Several studies have explored predictive modeling for basketball outcomes and betting strategies. Below is a list of key contributions in this field:

[1] C. Walsh and A. Joshi, "Machine learning for sports betting: Should model selection be based on accuracy or calibration?," Machine Learning with Applications, vol. 16, p. 100539, 2024, doi: 10.1016/j.mlwa.2024.100539.

[2] B. Loeffelholz, E. Bednar, and K. Bauer, "Predicting NBA games using neural networks," J. Quant. Anal. Sports, vol. 5, no. 1, pp. 1-7, Jan. 2009, doi: 10.2202/1559-0410.1156.

[3] Horvat T, Job J. The use of machine learning in sport outcome prediction: A review. WIREs Data Mining Knowl Discov. 2020; 10:e1380. https://doi.org/10.1002/widm.1380

[4] S. Jain and H. Kaur, "Machine learning approaches to predict basketball game outcome," 2017 3rd International Conference on Advances in Computing,Communication & Automation (ICACCA) (Fall), Dehradun, India, 2017, pp. 1-7, doi: 10.1109/ICACCAF.2017.8344688.

[5] R. M. Galekwa, J. M. Tshimula, E. G. Tajeuna, and K. Kyandoghere, "A systematic review of machine learning in sports betting: Techniques, challenges, and future directions," arXiv, preprint arXiv:2410.21484, 2024. [Online]. Available: https://arxiv.org/abs/2410.21484

Methodology

Data Collection
The dataset will be compiled from multiple sources, including the NBA API which provides game statistics, player performance, and team rankings, sportsbook odds data scraped from betting platforms (e.g., DraftKings, FanDuel) to analyze market trends, and player tracking data such as player efficiency rating (PER), win shares, and plus-minus.

Feature Selection
Key features for the model include:
- Team Performance Metrics (win/loss record, home/away performance, recent form)
- Player Statistics (points per game, assists, rebounds, injuries)
- Situational Factors (back-to-back games, rest days, rivalry games)
- Betting Market Data (odds movements, public betting percentages)

Model Development
The following ML techniques will be evaluated:

Supervised Learning Models
- Logistic Regression (baseline)
- Random Forest (for feature importance)
- XGBoost (gradient boosting for improved accuracy)

Reinforcement Learning Approach
- A Q-learning model to optimize bet selection based on evolving odds.

Ensemble Methods
- Stacking multiple models to improve prediction robustness.

Evaluation Metrics
Model performance will be assessed using:
- Accuracy & Precision – To measure prediction correctness.

- Return on Investment (ROI) – To evaluate profitability compared to random betting.
- Sharpe Ratio – To assess risk-adjusted returns.

Implementation

The project will be implemented in Python using libraries such as:

- Scikit-learn (for traditional ML models)
- TensorFlow/Keras (for neural networks)
- Pandas & NumPy (for data processing)

By combining these methodologies, the project aims to develop a robust ML-based system for optimizing basketball parlay bets.