

UFC Odds Project

(COMP3125 Individual Project)

Daniel LeBlanc
Wentworth 2027

I. INTRODUCTION

The relationship between sportsbook betting odds and UFC fight outcomes offers a great opportunity to evaluate how real-world markets can predict performance in mixed martial arts. As the UFC grows globally at an exponential rate, its large volume of fight statistics, combined with its extensive history of betting data, provides a unique dataset for examining whether favorites win at rates consistent with their betting probabilities and how factors such as weight class or fighter attributes influence their predictability. Prior research on combat sports shows that betting markets are often efficient but exhibit biases, such as overestimating underdog fighters, leaving open questions about the reliability and fairness of their probabilities. This report integrates statistical data with betting odds to analyze market accuracy and examine the dynamics underlying fight outcomes, setting the stage for a deep evaluation of how expectations align with reality.

II. DATASETS

A. Source of dataset

The datasets used in this analysis were obtained from publicly available Kaggle repositories, each created and maintained by members of the MMA and UFC analytics community. Both datasets contain documentation from the authors describing how the data was collected, sorted, and displayed using official UFC statistics and event records. Kaggle's version history and community validation further support the reliability of these sources.

B. Character of the datasets

The project uses 2 datasets: a comprehensive UFC fight statistics dataset and a separate betting odds dataset. The UFC fight dataset contains 151 columns of fighter metrics, fight details, and performance statistics, all stored in CSV format with thousands of rows representing individual fights. The odds dataset contains the moneyline values and matchup statistics for each fight. After importing, both datasets were cleaned by removing duplicates, standardizing column names, and converting weight classes into numeric values. The datasets were merged using fighter pair values to align fight statistics to their corresponding betting odds. These preparations ensure that the combined dataset is consistent, analyzable, and ready for modeling.

III. METHODOLOGY

In this project, two analytical approaches were used: a statistical model to look into predictive accuracy and visual methods to show patterns in the data.

A. Logistic Regression

Logistic Regression was used to model the probability that the betting favorite wins based on the probability from moneyline odds. This method assumes a linear relationship between predictions and odds, and it was chosen because it directly evaluates how well the

betting markets reflect real outcomes. The model was implemented using `LogisticRegression()` from scikit-learn after creating variables such as implied probability and the binary outcome (`favorite_won`).

B. Visualization

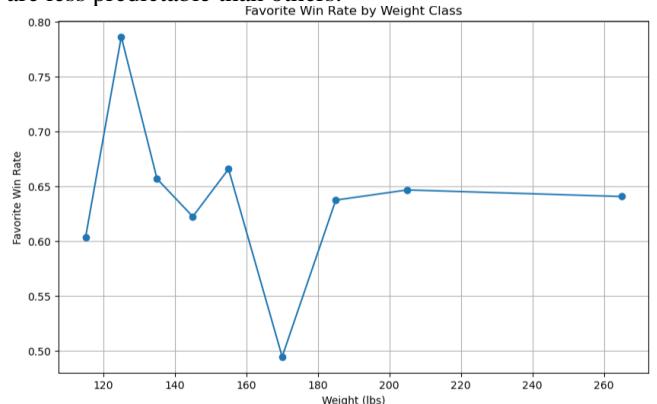
Visualizations were created using Python, creating scatterplots and distribution plots to inspect trends and data behavior. These visuals help identify patterns and inconsistencies that the models may not show. All visuals were created using Matplotlib, with adjustments made for readability and axis naming.

IV. RESULTS

This section presents the analysis findings using the visualizations created from the merged dataset. Each subsection addresses a question, presents one visualization, explains its meaning, and interprets the results.

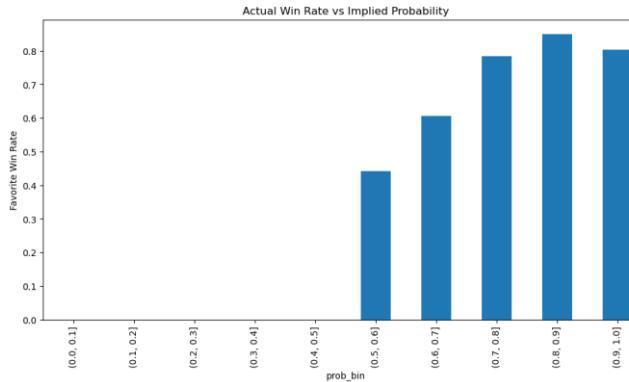
A. Favorite Win Rate by Weight Class

After converting each weight class into a numeric weight Value and calculating the fights where the betting favorite won, the graph shows clear variations between divisions. Lower weight classes, like the featherweight division (125 lbs.), did show noticeably high favorite success, while the welterweight division (170 lbs.) shows a less than 50% favorite win rate. This indicates that certain weight classes are less predictable than others.



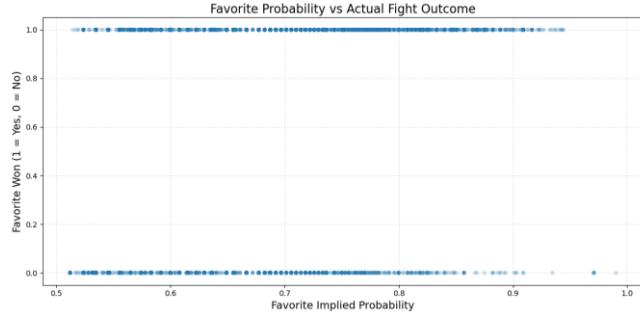
B. Favorite Accuracy vs. Implied Probability

By grouping fights into bins based on their implied probability, the results show an upward trend: as the implied probability increases, the actual favorite's win rate also increases. Lower probability bins perform inconsistently, but once the probability exceeds 60% (roughly), the win rates closely track expectations, indicating that the betting markets are generally well-calibrated and accurate.



C. Implied Probability vs. Actual Outcome

The scatter plot shows a pattern: as the favorites' implied probability increases, the distribution of the outcomes is more wins than losses. While there is some noise at lower probabilities, high probability favorites rarely lose, showing that betting odds reliably reflect fight outcomes. The tight clustering of points at the top shows the strong alignment between predicted probabilities and real fighter performance.



V. DISCUSSION

This project's results were mostly consistent, but a few limitations affected the accuracy. Some weight classes

had smaller sample sizes, which introduced noise, and the betting odds themselves may reflect maker bias rather than true probabilities. The logistic model was also limited by using only implied probability without additional fighter attributes.

Future improvements could include adding more fights, incorporating more fighter statistics, and testing more advanced models to capture more patterns. Expanding to multiple sportsbooks or analyzing the trends over time would strengthen and refine the findings.

VI. CONCLUSION

This project showed that UFC betting odds are relatively reliable indicators of fight outcomes: favorites win at similar rates across weight classes, with lighter fighters being usually more predictable. Higher implied probabilities align with higher actual win rates, and the logistic model confirms a positive relationship between odds and outcomes. These findings suggest that sportsbooks predict UFC fights with reasonable accuracy, meaning bettors can rely on odds as a realistic expectation of performance. In the real world, this helps bettors better understand risk, allows analysts to evaluate the market efficiency, and supports more informed decision-making in sports analytics and wagering.

REFERENCES

- [1] [1] UFC Stats, "Fight statistics and bout history dataset," ufcstats.com, Accessed: Jan. 2025.
- [2] [2] Kaggle, "UFC fight stats – complete historical dataset," [Kaggle.com](https://www.kaggle.com/datasets/ufc/fight-statistics-and-bout-history-dataset), Accessed: Jan. 2025.
- [3] [3] BestFightOdds, "Historical UFC betting odds and line movement," bestfightodds.com, Accessed: Jan. 2025.
- [4] [4] UFC, "Official fight results and event outcomes," ufc.com/events, Accessed: Jan. 2025.