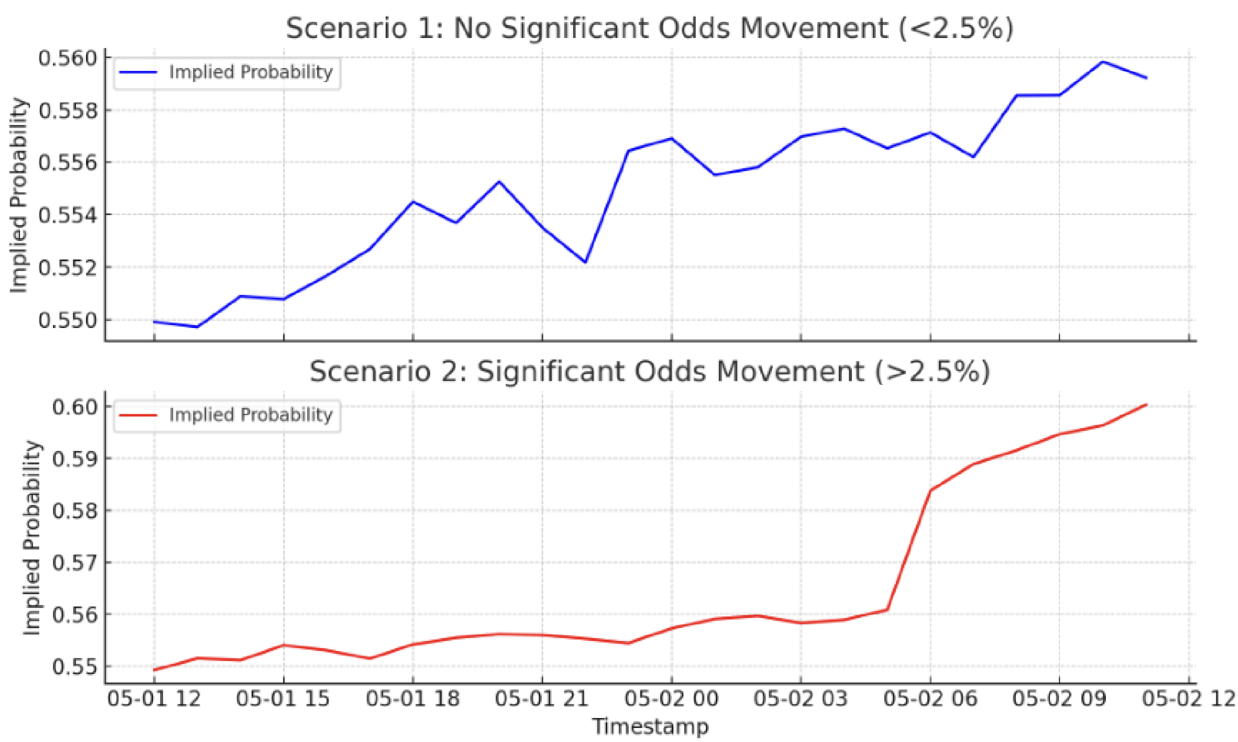
UFC Sports Betting Report

https://github.com/lucas-meadows/Final-Project

Scope:
In this report, we are exploring, given pre-fight market behaviour and fighter characteristics, can we detect whether a fighter's odds moved significantly before their bout?
We have defined a significant odds movement as any change in the average market price (the middle of the bid/ask spread, averaged across all markets) that is over 2.5%. IN our data, for ease of computation, we converted these bid/ask spread averages into implied probabilities, giving us a feature that was standardized between 0 and 1. While this may seem like a relatively low threshold of significance, large betting markets are relatively efficient markets, where typical volatility is relatively low, thus fluctuations over 2.5% seems a strong threshold. In our processed dataset we found over half of all such markets did not see any price fluctuations larger than 2.5%, which validates this claim.

Below is an example of the two sides of the binary classifier. You can see on the second market how there is one instance where there is a large, sudden, shift in the market of over 2.5%. This is the behaviour we are attempting to classify in this problem.



Audience:

Such a report could be useful to a variety of groups. This retrospective classification can aid sportsbooks in understanding the conditions under which a line moved (or did not) and whether it was predictive, helping them anticipate volatility and manage their market exposure. While quantitative bettors might benefit from this at the other end, and use it to identify patterns and market conditions which may help them "beat the market". This investigation primarily adds value to both parties as a classification tool that can help identify trends and spot individual patterns that are predictable, through which could be utilised to make money, either through betting intelligently, or hedging themselves.

It is important to clarify that this model is not intended to say exactly why odds move, it does not establish a causal relationship. Instead it serves to signal to a user the conditions under which market shifts have tended to occur in the past, based on observable patterns in the data. This type of analysis can inform such groups in their ongoing practices by highlighting structural patterns that may serve as the foundation for developing real-time tools that provide actionable signals in the future.

Data:

We incorporated two data sets to inform this model. Firstly, we imported a dataset from Kaggle which compiled statistics for UFC fighters. Due to the lack of timestamps on this data, we only included "static" features, those that did not update, and therefore would not risk data leakage to older fights. These variables were the following:

- "Id" - fighter ID
- "Name" - fighter's name, comprehensible data, helps merge data sets
- "Division" - weight class
- "Gender"
- "Status" - active or not
- "Place of Birth"
- "Age"
- "Octagon Debut" - when they first fought in the UFC
- "Height"
- "Reach" - arm span
- "Leg reach" - hip bone to heel

These features were constant no matter the time stamp of the fight.

Our other dataset involved polling data from Odds.API, where we compiled data from all betting markets for UFC fights over a one month period. This data was snapshots of each market in a moment in time, from which, we had to engineer variables to represent key market characteristics, without including data such as individual price jumps, which would cause data leakage within our model. These features we produced were:

- 'start_prob_fighter_A' - opening probability for fighter A
- 'start_prob_fighter_B' - opening probability for fighter B
- 'avg_spread_across_books_fighter_A' - average bid/ask spread across sportsbooks
- 'avg_num_books_fighter_A' - number of sportsbooks which held this market
- 'time_to_fight_hrs_fighter_A' - how long the market was live for before the fight

Our feature engineering was isolated towards our second dataset, which compiled the market snapshots into time series data. We produced numerous features which were used in our models, primarily the average bid/ask spread across all sports books, alongside isolating the starting probabilities for both fighters, and calculating the time to the fight by measuring the length of each time series.

Prior to any data processing and feature engineering, these datasets were particularly large. The Kaggle dataset had 2990 instances, and 4 columns which consisted of over 35 dictionary pairs. Our market data was also large, consisting of over 1000 market snapshots, these were compiled into over 340 rows of data, with 10 columns. The size of the data was reduced greatly due to incompletions, and unfortunately our data size could not be expanded due to the costliness of betting market data. We did, however, test and account for this in our problem, and our choice of model accounts for the limited size of the dataset.
In future, it would be beneficial to collect data with greater time frequency, as the tick rate we were subjected to was 30 minutes, whereas betting markets often operate at tick rates in the range of milliseconds. Alternatively, it would have been beneficial to be able to pull data over a greater range of dates, this would encompass more betting markets due to different fights occurring at different times.

Exploratory Data Analysis:
To check the completion and quality of the dataset we first had to engineer our desired features, and compile the market snapshots to check whether there was ample data. In compiling the market snapshots, I could then produce the features of each, then checking that each one was able to be computed and each column complete. Only when merging the datasets, and dropping any market data rows with incomplete fighter data, could we confirm whether there was an existing problem to be analyzed, adn whether there was sufficient data to approach the models.
We found that there was an approximately 50% split between rows with target variable 0 and target variable 1, showing that there was a problem to be investigated.
The shape of our data was 89 observations for 36 features, meaning it was feature rich but was weak in the number of observations. We were unable to gather more data (the Odds API is very expensive for large amounts of data), however, this was sufficient to run some algorithms. The limited number of instances meant our model complexity was limited, algorithms like decision trees, random forests or XG boost would be at risk of majorly overfitting the data.
We tested this fear by utilising a Random Forest Model, for which the results gave a precision of 1.00 for the classification of large price shifts, and a recall of 1.00 for instances without significant price shifts. This demonstrates the model being very cautious of classifier instances as 1 (large line movements), results such as these are tell-tale signs of data overfitting, and confirm our earlier concerns. As a result, it made sense for us to focus on lower complexity models, thus we explored our retroactive classification problem using (less complex) Logistic Regression models.
More general checks included verifying whether our processed data made sense. adding the odds for fighters A and B added to 1, which meant the data made intuitive sense. Probability changes were always better -1 and 1, likewise we checked that ages were within a threshold of 18 to 45 years, and height was between 48 and 84 inches, to ensure that all values were within a range that made sense.

Results:

Being informed by our EDA to steer clear of overly complex models, we began by applying a standard logistic regression model in order to classify whether a fight's odds had experienced a significant movement of more than 2.5% in the lead up to the bout. The model, with the results demonstrated below,m yielded overall accuracy of 70% with a notable performance imbalance between the two classes. The model favoured predicting large movement with a higher recall of 0.82, yet lower precision of 0.64. This suggests that it captured most true positives but at the cost of some false alarms (slight overfitting). This performance indicated that there was some general pattern recognition but the model was faltering due to its small dataset size.

Standard Logistic Regression

Class	Precision	Recall	F1-Score	Support
0 (No Movement)	0.78	0.58	0.67	12
1 (Movement)	0.64	0.82	0.72	11
Overall Avg.	0.71	0.7	0.69	23

To address this issue, we introduced a cross-validation to better estimate the out-of-sample performance, and ensure that the model is truly generalized and was not reflecting a singular train-test split. We saw meaningful improvements across all of the key metrics. Accuracy increased to 74%, with both precision and recall being more balanced across both classes. This is suggestive of an improvement in the stability of the models predictions, and thus less reliant on the individual patterns in a single train/test split.

Cross-Validated Logistic Regression

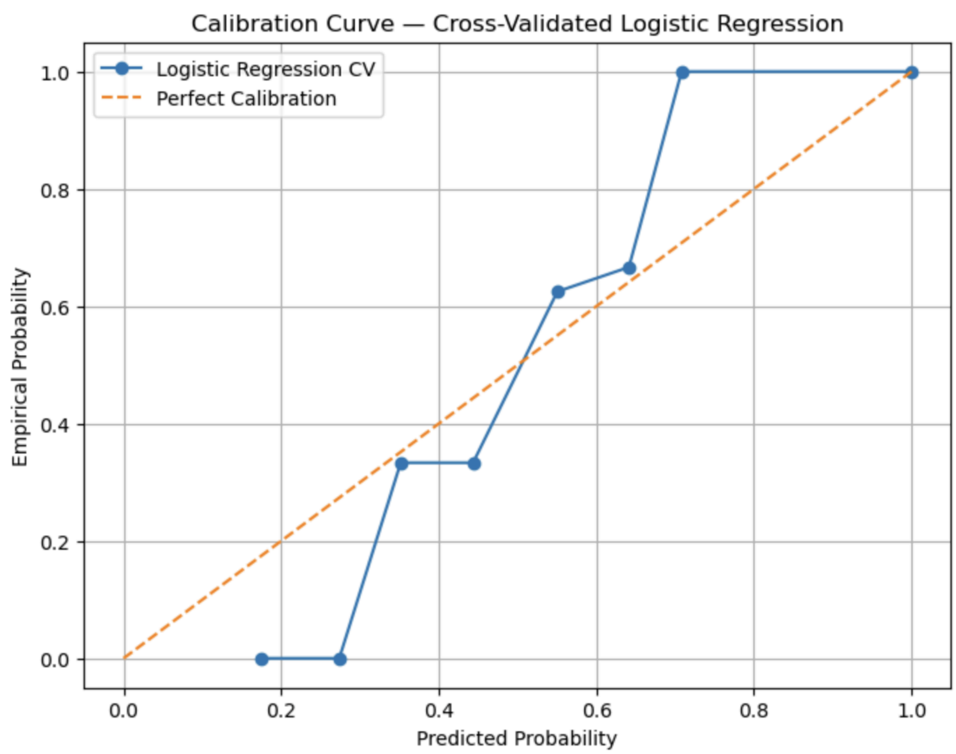
Class	Precision	Recall	F1-Score	Support
0 (No Movement)	0.8	0.67	0.73	12
1 (Movement)	0.69	0.82	0.75	11
Overall Avg.	0.75	0.74	0.74	23

Calibration Curve:

We produced a calibration curve for our best performing model, the cross-validated logistic regression. Calibration is referring to the relationship between predicted and empirical probability. It does not just measure how strong the model is, but how accurate its model is. A perfectly calibrated model predicting 60% confidence in class 1 (large line movements) should be correct 60% of the time across all predictions at that level.

In our cases, our model performs well on high confidence predictions, of 80% and up. When it assigns high probabilities to movement, the movement does indeed concur nearly 100% of the time. IN the middle of this, between 40% and 80%, the model performs alright. However, as we approach the low end of the curve the model visibly underperforms. In low confidence predictions, the model falls short and there are in fact more movements that occur than expected.

From a practical standpoint, this is a highly favorable result. IN the context of betting markets, where models are used to signal high-confidence market opportunities to traders and sports books alike, we care most about the reliability of the high-confidence predictions. A model that can inform us of a high probability of a movement enables confident action when it matters most. This model being less confident at the lower end is not critical from an operational standpoint, as these low confidence predictions would be filtered out in its practical use regardless. Ultimately, the model is conservative in low confidence settings but assertive when it has to be, which aligns well with the scope of the project prioritizing financial or wagering applications.



Conclusion:

This project set out to investigate whether large pre-fight movements in UFC betting markets could be retrospectively classified based on pre-fight behaviour and static fighter characteristics. Rather than attempting to predict future outcomes, which risked major data leakage due to the structure of our market snapshots, we instead framed our question as a classification problem applied after the fight's closure. This allowed us to explore whether we could find predictive patterns in the structure of markets, and identify those that experienced large market movements versus those that did not.

Through careful data processing and feature engineering, we produced a strong set of data for which we carefully went through a model selection process, given its features. We found that logistic regression, particularly with cross-validation, provided us with a stable and interpretable framework for this classification task. We found that our final model achieved an overall accuracy of 74%, a promising result. More importantly, it demonstrated good calibration at high thresholds, which, in a betting context, established its use for actionable insights amongst algorithmic bettors and sportsbooks alike.

While we cannot make causal claims about the features that drive these significant line movements, our results do demonstrate that meaningful patterns exist within the structure of the markets where large shifts occur. The classifier is able to distinguish between these cases and highlight which markets are more likely to exhibit significant line movements. This offers practical utility to analysts, traders and sportsbook alike, not by explaining why odds move, but allowing them to recognize the conditions under which movement is more or less likely to occur.