Paul Suh (phs92), 4741
Aditya Kompella(apk74), 5741
Kenny Liang (ql75), 5741

UFC Fight Outcome Predictions

Problem Statement: The UFC is among the world's most popular mixed martial arts organizations. In a mixed martial arts fight, two fighters, one who fights out of the red corner and another who fights out of the blue corner, are allowed to punch, kick, elbow, wrestle, and grapple their opponents. Traditionally, the red corner is given to the fighter who has a higher ranking and the blue corner is given to the fighter with a lower ranking. A standard UFC Fight has 3 rounds, each of which lasts 5 minutes. UFC Fights can end in either a win, a draw, or a loss. A fighter can win by knocking the opponent out with a strike (KO), submitting the opponent with a grappling technique (SUB), or getting a decision victory on the judges' scorecards (DEC). A fight can end in a draw if a fight lasts the full duration and the majority of the judges score the bout to be a draw. Throughout this project, we will use feature engineering, exploratory data analysis methods like PCA, and prediction models like Logistic Regression in order to accurately predict the winning fighter. We will then use the probabilities of the model to determine how much to bet and see how profitable our model is on some held-out test data.

Description of Dataset:
The Ultimate UFC Dataset is a comprehensive collection (~ 800,000 rows) of data related to the Ultimate Fighting Championship (UFC). It includes detailed information about past fights, fighter statistics, and match outcomes. Across 119 columns, some important features include statistics on Red/Blue odds, average significant strike landed, fighter ranks, and other useful data. This in-depth information on fighters and their different attributes, along with their previous histories, should be sufficient to predict if a fighter will win on average. This dataset is likely to be valuable for those interested in analyzing trends, patterns, and factors that contribute to the success or failure of fighters in a UFC event. Due to how comprehensive and large the data set is, we are confident in using it to create predictive models or to perform detailed statistical analysis on UFC fights.

The following is a link to the dataset that we are using:
https://www.kaggle.com/datasets/mdabbert/ultimate-ufc-dataset/data

Significance of Problem:
UFC fighters, gyms, and broadcasters will find many of our insights useful and interesting. For example, we could classify fighters as "strikers" or "grapplers" and see how each type of fighter fares against the other type. We could see how much stamina each fighter has based on how long their fights last and see how matching opponents with more or less stamina affects win statistics.

This information would be useful to current UFC fighters or gyms that can optimize the training of athletes. For instance, if we find that striking accuracy is a larger predictor of success, gyms might increase the amount of striking training they include in a fighter's regimen. Moreover, UFC broadcasters could utilize our results to provide better insights during live commentary.

The predominant benefactor of our work would be sports betting companies. Popular online gambling sites like DraftKings or Fanduel utilize internal statistical models as well as betting actions to set the live odds of every UFC event. It is paramount for sports betting platforms to set the odds as accurately as possible to minimize risk exposure and ensure government compliance. A statistical model that can more accurately predict the outcome of UFC events would offer significant business value to sports betting companies. We will first backtest our model on previous UFC matches and see how well our model predicts the win-loss of past UFC matches. We will then scrape live odds and statistics to see how well our model does on live matches and if our model generalizes beyond the given dataset. To know whether our model did well, we would want to see if the return from betting using what our model predicts is greater than if we bet using the live odds set by a sports betting company. Ideally, if this model performs better than the current betting odds on sites like DraftKings, companies can use this model to readjust their betting odds. A sports betting company's primary goal is to generate profits. By setting odds that accurately reflect the probabilities of different outcomes, the company aims to achieve a balanced book, where it receives roughly equal amounts of money on both sides of a bet. When the outcomes align with the predicted probabilities, the sportsbook can collect the losing bets and pay out the winning bets, ensuring a profit margin. Accurate odds help sportsbooks manage their risk and maximize profitability.

GitHub Link: https://github.com/KennyLiang2302/ufc_prediction