



PREDICTING UFC FIGHTS

In this project, we create a machine learning categorizing model to predict the outcomes of UFC fights using individual fighter statistics. In testing, our overall error rate was 25%.

Authors
Sophia Manodori
Greta Anesko



INTRODUCTION

The Ultimate Fighting Championship is a multi billion dollar corporation that hosts fighting events around the world. International mixed martial arts athletes meet at these weekly events to fight for the division titles. Each match between two fighters has either 3 or 5 minute rounds and ends via submission, knockout, technical knockout (called by the referee), or by judge's decision.

Our dataset was from kaggle and included data from 7226 fights since 1996. It contains all individual fighter statistics as well as the in fight statistics. We primarily focused on using fighter statistics to predict fight outcomes. Fighter statistics include physical statistics such as height, weight, and stance, and career statistics such as strikes accuracy and takedown accuracy.

METHODOLOGY

We performed variable selection using a logistic regression model to identify the most significant predictors. We then split the data into a training set (the first 80% of the data) and the testing set (the most recent 20%). We trained logistic regression, LDA, QDA and Naive Bayes models on the training set and tested them on the test set and compared the error rates of each model between classes and overall error rates.

RESULTS

Of all the models tested, the LDA model performed the best, with a 24% error rate on the testing data. This makes sense since most of the variables seemed to be normally distributed and there was not a noticeable grouping correlation. The testing error rate for predicting Red fighter's wins was 22%, while the error for predicting Blue fighter's wins was 43%. 61% of wins in the test dataset are Red fighter wins, with similar proportions in the training and testing sets, which may account for some of the increased accuracy for Red wins.

PREDICTORS

To select predictors we evaluated only variables that were differences between the two fighter's statistics (the Red fighter's statistic minus the Blue fighter's statistic). This ensured that the models were making their predictions based on equal amounts of data from each fighter. The variables we identified as significant predictors through logistic regression were the differences between fighters in:

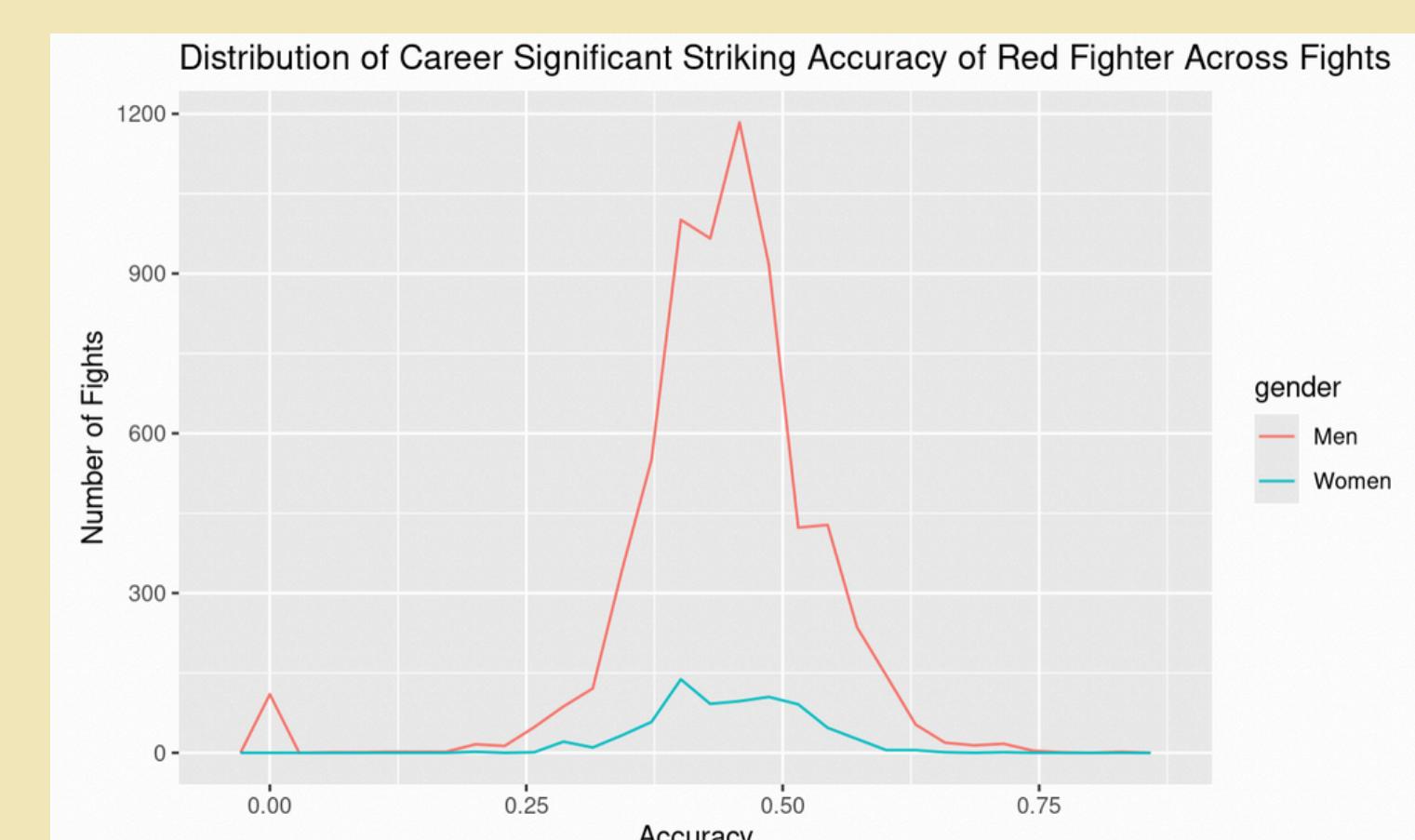
- Number of wins
- Number of losses
- Weight
- Age
- Strikes landed per minute
- Strikes absorbed per minute
- Significant striking rate
- Takedown defense rate
- Average number of takedowns per 15 minutes

PREDICTIONS AND FUTURE TESTING

To test our model on upcoming fighting events, we compiled the fighter statistics for the UFC Fight Night on April 28th and ran the model on the fights. The model correctly predicted the outcome of 9 out of 10 fights. Our model is optimal for sports betting as it offers specific probabilities for each fighter, it making it optimal for placing parlay bets since the risk can be minimized by choosing the highest probability fighters. It also occasionally will differ from the UFC favorite, highlighting ideal candidates for underdog bets.

Red Fighter	Blue Fighter	P(Red)	P(Blue)
Sandhagen	Figueiredo	0.5451256875	0.4548743125
Ridder	Nickal	0.8677388848	0.1322611152
Ponzinibbio	Rodriguez	0.6552771791	0.3447228209
Jackson	Marcos	0.433604851	0.566395149
Smotherman	Sidey	0.6732768667	0.3267231333
Stephens	Jones	0.2301701167	0.7698298833
Santos	Tate	0.6162587221	0.3857612779
Loder	Bekoev	0.6281413677	0.3718586323
Rodriguez	Robertson	0.697331031	0.302668969
Bolanos	Le	0.6051956489	0.3948043511
Petersen	Mayes	0.8461694534	0.1538305466
Miller	Petrović	0.3633011558	0.6366988442

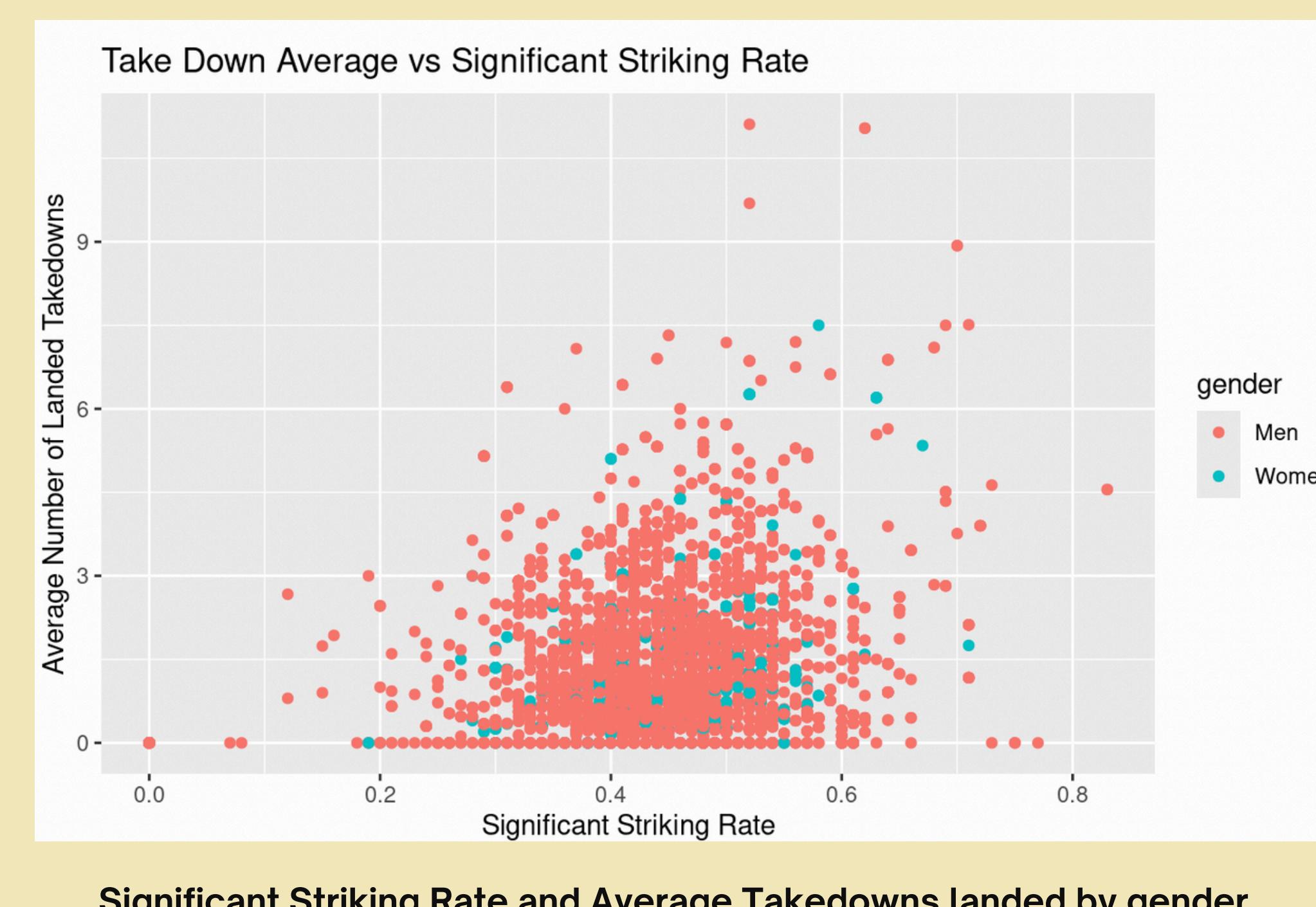
Fight Predictions for May 2nd Fight Night (winners are shaded, fights with odds differing from the UFC official predictions are shown in blue)



Distribution of Striking Accuracy

CONCLUSION

Overall our model was incredibly successful. In the future we plan on creating a model with variables selected for women's fights, and also consider using neural networks to create a model that identifies the most likely method of winning.



Significant Striking Rate and Average Takedowns landed by gender

