Matthew Olivo

CPSC 392

What are the main predictors behind a win and a draw?

Changed to: What are the main predictors behind the number of fights a fighter has won?

Changed: Logistic Regression to Linear Regression

Changed D: how many wins added when certain moves are thrown

1. I will be using a logistic regression model to answer this question because the outcome variable I am trying to predict is categorical. I will be utilizing a Train Test Split as my method of cross validation and use lasso to find the best variables to use.
2. A logistic regression model is a simple yet effective way to view which features have the largest impact. I am using lasso in order to not only reduce overfitting but to help in feature selection. Train test split cross validation is very practical and more efficient than other validation methods.
3. The coefficients of the best features in our models will have the highest value.
4. I plan to make a grouped scatterplot which shows which types of moves or strikes were most often associated with a win. Additionally, I will make a boxplot which displays when a certain percentage/amount of each move or strike is thrown, what the likelihood of winning is.

Does age play a role in a fighter’s chances of winning?

1. For this question, I will first need to look at the results of my previously mentioned logistic regression model and see if the age variable in any way is significant. My previously mentioned lasso model will tell me if the age is significant as well.
2. These are already implemented models that will provide me with answers to multiple different questions. lasso regression will tell me if age is one of the best predictors of a win, and my logistic regression will tell me how good of a predictor it is of a win.
3. Will take note of age’s coefficients in logistic and lasso regression.
4. I will plot a scatterplot showing the relationship between wins and fighter’s age. Next I will make boxplots by age group displaying win percentage spread across the group.

Difference in fighting styles and chances of a win?

1. This question will best be solved by a KMeans clustering algorithm with closer to 7 or 8 clusters. Variables will be standardized, and I will use silhouette scores.
2. I am using this many clusters as there are said to be about that many prevalent fighting styles in the UFC today. Predictor variables will need to be standardized as all are not operating on the same scale. Silhouette scores will be used to ensure there is cohesion and separation amongst the clusters. I am admittedly unsure of my decision to use a KMeans algorithm over a HAC because I do enjoy the idea of being able to use a dendrogram.
3. Identify groups given certain fighting style choices and compare those choices with optimal choices.
4. I will display key variables graphed against each other in a scatterplot to show obvious separations between certain clusters. If switch to HAC, a dendrogram displaying cluster separation and cohesion.