Final Project

Yifei Yang, Yiming Yuan

Read in the data

Introduction and data

**Questions: In what ways can MMA fighters improve their performance?**

Around 300 million individuals worldwide identify themselves as fans of Mixed Martial Arts (MMA), with its popularity peaking in nations such as the United States, the United Kingdom, Brazil, Singapore, and China. The Ultimate Fighting Championship (UFC) is the premier organization in the MMA world. In the dynamic landscape of MMA, quantifying fighter performance is essential for both athletes and analysts. Our objective is to reveal the significance and implications of that factors that contribute to the fighters' performance.

Today's data comes from UFC fighter statistics collected in 2024, encompassing various dimensions of fighter performance results and parameters. **wins**, **losses** and **draws** represent the number of a fighter's victories, dogfalls and failures throughout their career. **height_cm**, **weight_in_kg** and **reach_in_cm** quantifies fighters physical attributes. **stance** is a categorical variable including Orthodox/Southpaw/Switch, which highlights fighters' preferred combat orientation. Performance metrics such as **significant_strikes_landed_per_minute** and **significant_striking_accuracy** record fighters' precision, timing, and offensive capabilities, while **significant_strikes_absorbed_per_minute** and **significant_strike_defence** quantify their defensive skills. **takedown_accuracy** and **average_takedowns_landed_per_15_minutes** assess fighters' proficiency in executing takedowns, while **takedown_defense** shows their ability to counter opponents' takedown attempts. Lastly, **average_submissions_attempted_per_15_minutes** measures fighters' inclination towards submission-based tactics, reflecting their grappling proficiency.

We add two new variables: **win_ratio** is calculated as the ratio of wins to the sum of wins, draws, and losses, offering a measure of a fighter's success rate; and **age** is derived from fighters' date_of_birth statistics, providing insight into the fighters' maturity and experience within the competitive landscape.

Figure 1. visualizes the distribution of fighters' striking and takedowns metrics.

Figure 1. Box Plots of Fighting Statistics

sources:

https://www.kaggle.com/datasets/aaronfriasr/ufc-fighters-statistics?resource=download
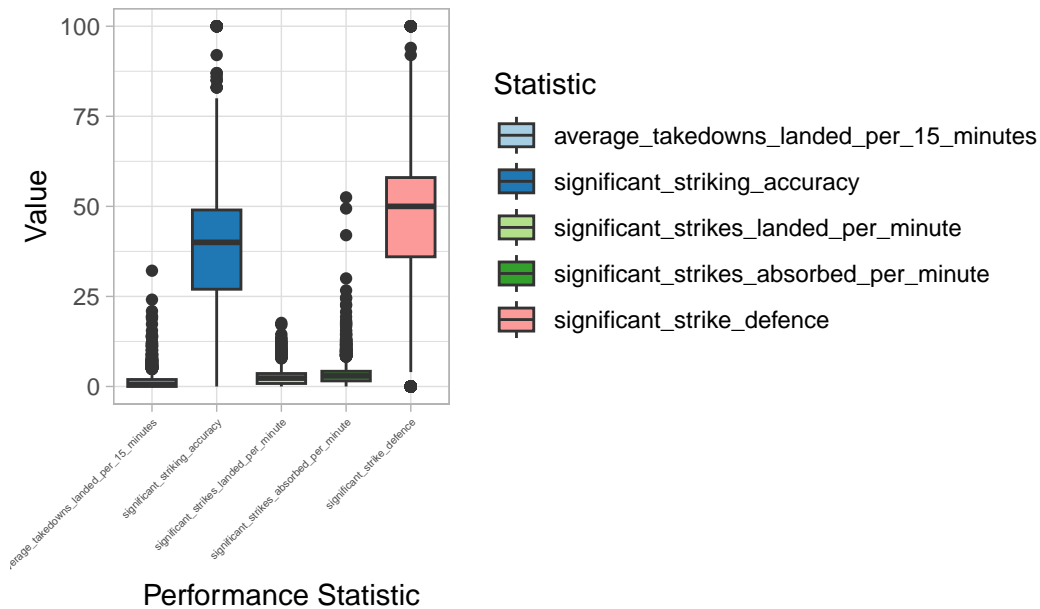
https://www.euronews.com/business/2023/09/27/the-booming-billion-dollar-business-of-combat-sports

Methodology

Predictor selection: We utilized a heatmap to visually explore and identify the variables most closely associated with fighters' win_ratio. The heatmap allowed us to observe the strength and patterns of correlation between various predictors and the wins_ratio. Hence, we selected 7 predictors that demonstrated the highest correlation coefficients with the win_ratio for inclusion in the model, and they are Age,Stance,Average_takedowns_landed_per_15_minutes, Significant_striking_accuracy, Significant_strikes_landed_per_minute, Significant_strikes_absorbed_per_minute, Significant_strike_defence.

Heatmap of Specified Variables
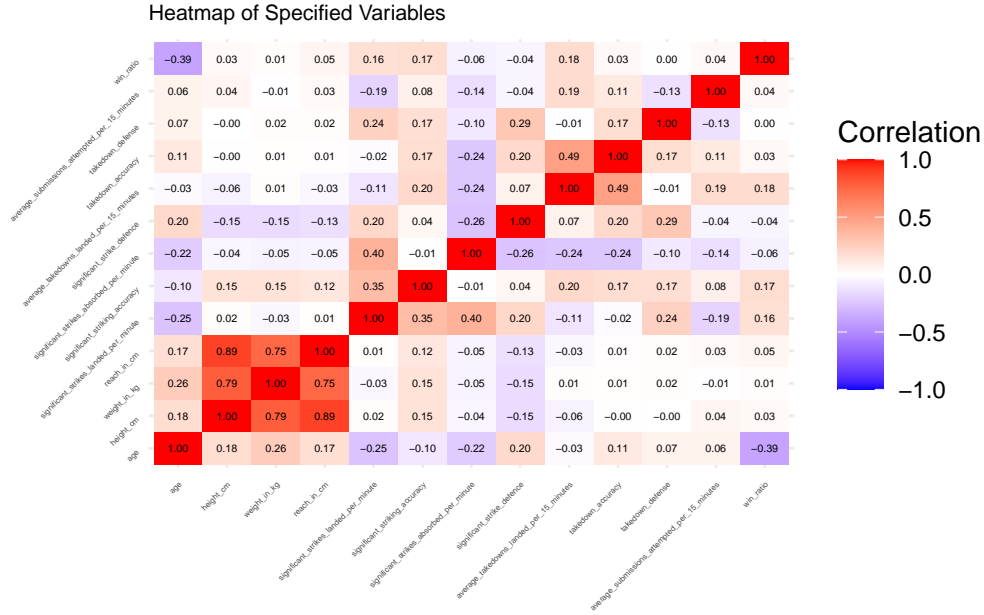
Ordinal regression model selection: We first utilize a linear model with with interaction terms and and observed an R-squared value of 0.29, indicating suboptimal predictive performance. Then we fit an ordinal model and test its accuracy. Specifically, the response variable win_ratio is categorized into 0-0.33, 0.33-0.67, and 0.67-1, and labelled as "Low", "Medium", "High" respectively. Our analysis explore the relationships between fighters' win_ratio groups (Low, Medium, and High) and a set of predictors.

By creating a confusion matrix that shows the distribution of actual versus predicted group memberships, we find that the test accuracy of this ordinal model is 0.71. Thus, the ordinal model can make more accurate predictions and it is feasible to find more influential predictor by ordinal model.

```
        Predicted
Actual    Low Medium High
  Low       2     47     7
  Medium    0    251   620
  High      0    191  1858

[1] "Accuracy:  0.709341397849462"
```

Results

Continuous Variable:

- Age shows a negative coefficient (-0.115), indicating that as fighters age, their likelihood of being in a higher win ratio group decreases significantly.

- Average takedowns landed per 15 minutes has a positive coefficient (0.155), suggesting that higher rates of takedowns landed are associated with being in a higher win ratio group.

- Significant striking accuracy has a positive effect (0.012) on the win ratio group, indicating that more accurate strikers tend to be in higher win ratio groups. However, it has a small effect.

- Significant strikes landed per minute also positively affects the win ratio group (coefficient = 0.068), suggesting that fighters who land more strikes per minute tend to have better win ratios.

- Significant strikes absorbed per minute has a negative coefficient (-0.053), indicating that fighters absorbing more strikes per minute are likely to belong to lower win ratio groups.

- Significant strike defence shows a negligible negative coefficient, suggesting minimal impact on win ratio groups.

Categorical variable:

Sideways (coef = 4.3460381): This stance has a significantly positive coefficient, indicating that fighters adopting a Sideways stance have a higher likelihood of being in a higher win ratio group compared to the reference category. However, the statistical significance is marginal (t-value close to 1), thereby it may not be robust across different samples.

Additionally, the coefficients for other stances are all negative, while the t-value indicates that their effect is not statistically significant.

4. *Linear model for comparison*

```
Call:
lm(formula = win_ratio ~ age + stance + average_takedowns_landed_per_15_minutes +
    significant_striking_accuracy + significant_strikes_landed_per_minute +
    significant_strikes_absorbed_per_minute + significant_strike_defence +
    age * average_takedowns_landed_per_15_minutes + age * significant_striking_accuracy +
    age * significant_strikes_absorbed_per_minute + age * significant_strike_defence,
    data = ufc)

Residuals:
     Min       1Q   Median       3Q      Max
-0.72177 -0.06487  0.00156  0.07423  0.66218
```

```
Coefficients:
                                                      Estimate Std. Error t value
(Intercept)                                          1.130e+00  6.512e-02  17.358
age                                                 -1.218e-02  8.557e-04 -14.230
stanceOrthodox                                       5.375e-03  4.994e-02   0.108
stanceSideways                                       1.853e-01  1.315e-01   1.410
stanceSouthpaw                                       9.690e-03  5.017e-02   0.193
stanceSwitch                                         1.104e-02  5.080e-02   0.217
average_takedowns_landed_per_15_minutes             -3.342e-03  6.007e-03  -0.556
significant_striking_accuracy                       -1.792e-03  7.956e-04  -2.253
significant_strikes_landed_per_minute                7.883e-03  1.656e-03   4.760
significant_strikes_absorbed_per_minute              5.719e-03  4.378e-03   1.306
significant_strike_defence                          -1.763e-03  7.565e-04  -2.331
age:average_takedowns_landed_per_15_minutes          3.698e-04  1.615e-04   2.289
age:significant_striking_accuracy                    6.309e-05  1.885e-05   3.347
age:significant_strikes_absorbed_per_minute         -3.119e-04  1.141e-04  -2.734
age:significant_strike_defence                       4.203e-05  1.773e-05   2.370
                                                     Pr(>|t|)
(Intercept)                                          < 2e-16 ***
age                                                  < 2e-16 ***
stanceOrthodox                                       0.914298
stanceSideways                                       0.158761
stanceSouthpaw                                       0.846868
stanceSwitch                                         0.827888
average_takedowns_landed_per_15_minutes             0.578025
significant_striking_accuracy                        0.024348 *
significant_strikes_landed_per_minute                2.03e-06 ***
significant_strikes_absorbed_per_minute              0.191499
significant_strike_defence                           0.019831 *
age:average_takedowns_landed_per_15_minutes          0.022133 *
age:significant_striking_accuracy                    0.000826 ***
age:significant_strikes_absorbed_per_minute          0.006295 **
age:significant_strike_defence                       0.017840 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1215 on 2961 degrees of freedom
  (1135 observations deleted due to missingness)
Multiple R-squared:  0.2995,    Adjusted R-squared:  0.2962
F-statistic: 90.41 on 14 and 2961 DF,  p-value: < 2.2e-16
```

We also fitted a linear model with with interaction terms and and observed an R-squared value of 0.29, indicating suboptimal predictive performance.

Discussion

In terms of how fighters improve their performance, we can conclude that they should put more effort into improving their rates of takedowns landed, striking accuracy, striking numbers, and strikes absorbed, because these factors are predicted to be more closely associated with wins ratio. In contrast, which stance they choose and the significant strike defence do not contribute much to improve their wins ratio.

As for limitations, the ordinal model may oversimplify the complex dynamics of MMA fights and fail to capture all relevant predictors. To improve future analyses, we could consider incorporating additional variables such as fighters' training regimens, injury history, and psychological factors. Additionally, since we only categorize the wins ratio into three groups, this may simplif the prediction task and potentially lose granularity in the data, which affects the model's ability to accurately predict the win ratios.