

Final Project

Yifei Yang, Yiming Yuan

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.

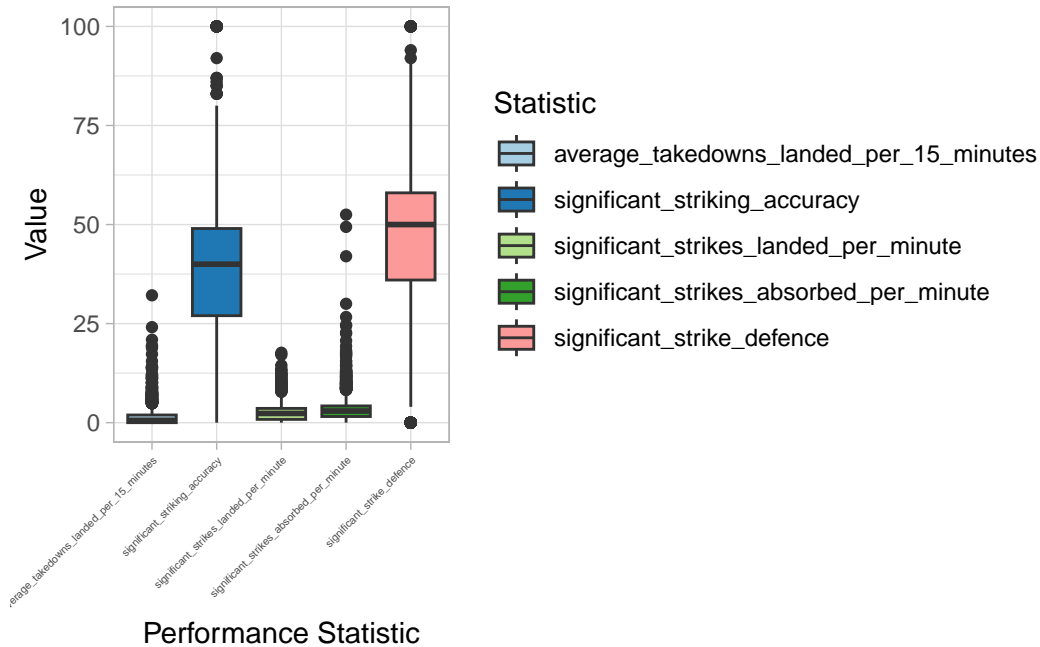


Figure 1. Box Plots of Fighting Statistics

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 win_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.

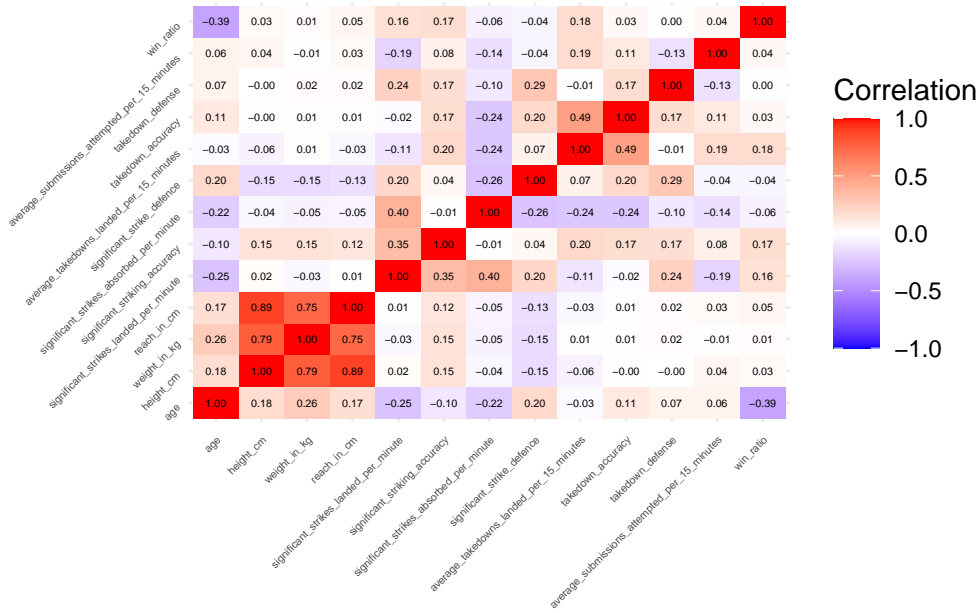


Figure 2. Heatmap of Specified Variables

Ordinal regression model selection: The data were fit by the ordinal logistic regression model based on model fit parameters (AIC: 3586.268) to explore the relationships between fighters' characteristics. 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.

According to **Table1**, the confusion matrix shows the distribution of actual versus predicted group memberships, which indicates that the test accuracy of this ordinal model is approximately 71%. Thus, the ordinal model can make more accurate predictions and it is feasible to find more influential predictor by ordinal model.

Table1. Confusion Matrix and Accuracy from the Ordinal Model

		Predicted		
Actual		Low	Medium	High
	Low	2	47	7
	Medium	0	251	620
	High	0	191	1858
Accuracy:		0.7093413978495		

Results

Table2. Characteristics of UFC Fighters with Ordinal Model

	Coefficients	OR	SD.Error	t.value
age	-0.1151	0.8912	0.0067	-17.2641
stanceOrthodox	-0.3938	0.6745	0.8161	-0.4825
stanceSideways	4.3460	77.1721	4.4600	0.9744
stanceSouthpaw	-0.1067	0.8988	0.8213	-0.1299
stanceSwitch	-0.5320	0.5874	0.8386	-0.6344
average_takedowns_landed_per_15_minutes	0.1554	1.1681	0.0318	4.8864
significant_striking_accuracy	0.0117	1.0118	0.0035	3.3222
significant_strikes_landed_per_minute	0.0676	1.0699	0.0342	1.9768
significant_strikes_absorbed_per_minute	-0.0528	0.9485	0.0181	-2.9248
significant_strike_defence	-0.0007	0.9993	0.0033	-0.2175
Low Medium	-8.6075	0.0002	0.9195	-9.3612
Medium High	-5.0338	0.0065	0.9050	-5.5622

Results from our ordinal regression model are presented in **Table2**. Age reported a negative coefficient (-0.115), indicating that for each one-unit increase in age, a fighter has approximately 0.8912 times the odds of being in a higher category of win ratio compared to younger fighters, assuming all other variables are held constant. Average takedowns landed per 15 minutes has a positive coefficient (0.155), suggesting that a fighter will have approximately 1.1681 times the odds of being in a higher category of win ratio, if he increases one unit in average takedowns landed per 15 minutes. Significant strike defence shows a negligible negative coefficient (-0.0007179), which is not evident to prove that it has significant effect on win ratio groups.

The Sideways stance has a significantly positive coefficient (coef = 4.3460381), while the coefficients for other stances are all negative. However, since the absolute t-values of all stance types are less than 1.96, there is no strong evidence to reject the null hypothesis that the stance does not have significantly effect on win ratio groups.

Therefore, the factors which have statistically significant effect on win ratio groups are age, average_takedowns_landed_per_15_minutes, significant_striking_accuracy, significant_strikes_landed_per_minute, and significant_strikes_absorbed_per_minute.

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 simplify the prediction task and potentially lose granularity in the data, which affects the model's ability to accurately predict the win ratios.