

# UFC Judge Scoring Analysis

*DSCI 522 group 315*

*2020/1/23 (updated: 2020-01-25)*

## Contents

Summary . . . . .	1
Introduction . . . . .	1
Methods . . . . .	1
Results & Discussion . . . . .	2
References . . . . .	6

## Summary

In this study, we aim to assess whether the judging in the Ultimate Fighting Championship (UFC) is consistent with the official rules and guidance. To answer this question, we need to identify the key predictors for winning determined by the judges based on the performance and evaluate whether these predictors are consistent with the priorities of round assessments according to the official rules. We built a regression model using the logistic regression algorithm to assign weights to the features and used recursive feature elimination (RFE) approach to identify the strong predictors. Among the selected 11 features by RFE, 7 features are related to Striking/Grappling performance which should be considered as the top factor in judgment based on the UFC official rules. Our final logistic regression model using these selected features performed well on validation data set with accuracy score of 0.83. It correctly predicted 368 out of 446 test cases and incorrectly predicted 78 cases with 40 being false positive and 38 false negative. Our results showed that the judges generally comply to the UFC rules and put weights on some additional factors.

## Introduction

The Ultimate Fighting Championship (UFC) is the world's premier mixed martial arts (MMA) organization. The events are to identify the world's best martial artists. Since it is a relatively new sport, the rules are not well established and need improvement. The results of the fights are often controversy when the winning is needed to be decided by judges based on the competitors' performance(instead of ending by submission and technical knockout). The offical rules providing some general guidance in judging the winning and the criteria with priority from high to low are effective Striking/Grappling, effective Striking/Grappling and cage/ring Control(The ABC MMA Rules Committee 2017). As more and more controvercy of the judging decisions raised, a systematic analysis of the UFC data would be very informative to evaluate the overall quanlity of UFC judging system.

In this project, we are trying to identify the key predictors for winning in previous UFC events and examine whether these key predicitors are in line with the UFC offical guidance. This analysis is very significant because it may serve as a quanlity control approach for the UFC judging system and also help to improve the rules and training strategies for the judges in the future.

## Methods

### Data

The original data was obtained from Kaggle user Rajeev Warrier (Rajeev Warrier 2019). The data has also been downloaded and uploaded to a GitHub repo to avoid issues for users who do not have kaggle accounts.Each row in the dataset represents statistics from an UFC event, including the performance features and winners (Red or Blue). The data was pre-processed by only selecting the features related to fight performance and for the each feature, the ratio of Blue fighter versus the Red fighter was calculated. The target was computed as whether the Blue figher wins or not.

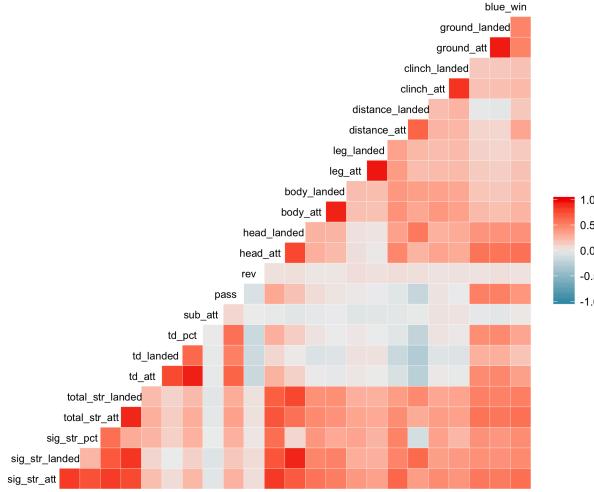


Figure 1: Figure 1. Correlation matrix for features and target.

## Analysis

The logistic regression algorithm was used to build a regression model to assign weights to all the features in the pre-processed data. The features included in the final model were selected using recursive feature elimination (RFE) approach with cross validation. A final regression model were built using these selected features and used to predict on the validation data set. The R and Python programming languages (R Core Team 2019; Van Rossum and Drake 2009) and the following R and Python packages were used to perform the analysis: docopt (de Jonge 2018), janitor (Sam Firke 2020), tidyverse (Wickham 2017), GGally (al 2018), docopt (Keleshev 2014), os (Van Rossum and Drake 2009), requests (Reitz 2019), Pandas (McKinney 2010), numpy (Oliphant 2006), altair (VanderPlas et al. 2018), matplotlib (Hunter 2007), selenium (Muthukadan 2019), scikit-learn (Buitinck et al. 2013), chromedriver-binary (Kaiser 2019). The scripts for generating the analysis and the report together with all the relevant information can be found [here](#).

## Results & Discussion

To explore the relationships between different features and between features and target, we plotted the pair-wise correlation matrix (Figure 1). The graph showed that features `sig_str_att`, `head_att`, `total_str_att` are highly correlated with the target. It also indicated there are some interaction between the features.

We categorized the features into four groups and in each group we explored the realtionships between features and target (Figure 2). In general, the results showed that all the distributions of features between winning and losing were overlapped for some extend in different groups. However, each group has some features with significant difference in the means, which indicating these features may be strong predictors for the target. The result is consistent with the feature-target correlation analysis.

We chose logistic regression model to assign weights to all the features and used recursive feature elimination (RFE) with cross validation approach to identify the most relevant features and ranked them according to the weigths (Table 1). Features with higher weights are stronger predictors for winning. Among the 11 features, the top first feature and another 6 features are indicators of Striking/Grappling performance, which should be the first prioritiy in judging criteria according to the UFC offical rules (The ABC MMA Rules Committee 2017).This indicates the rules were generally followed by the judges. Four features does not belong to the

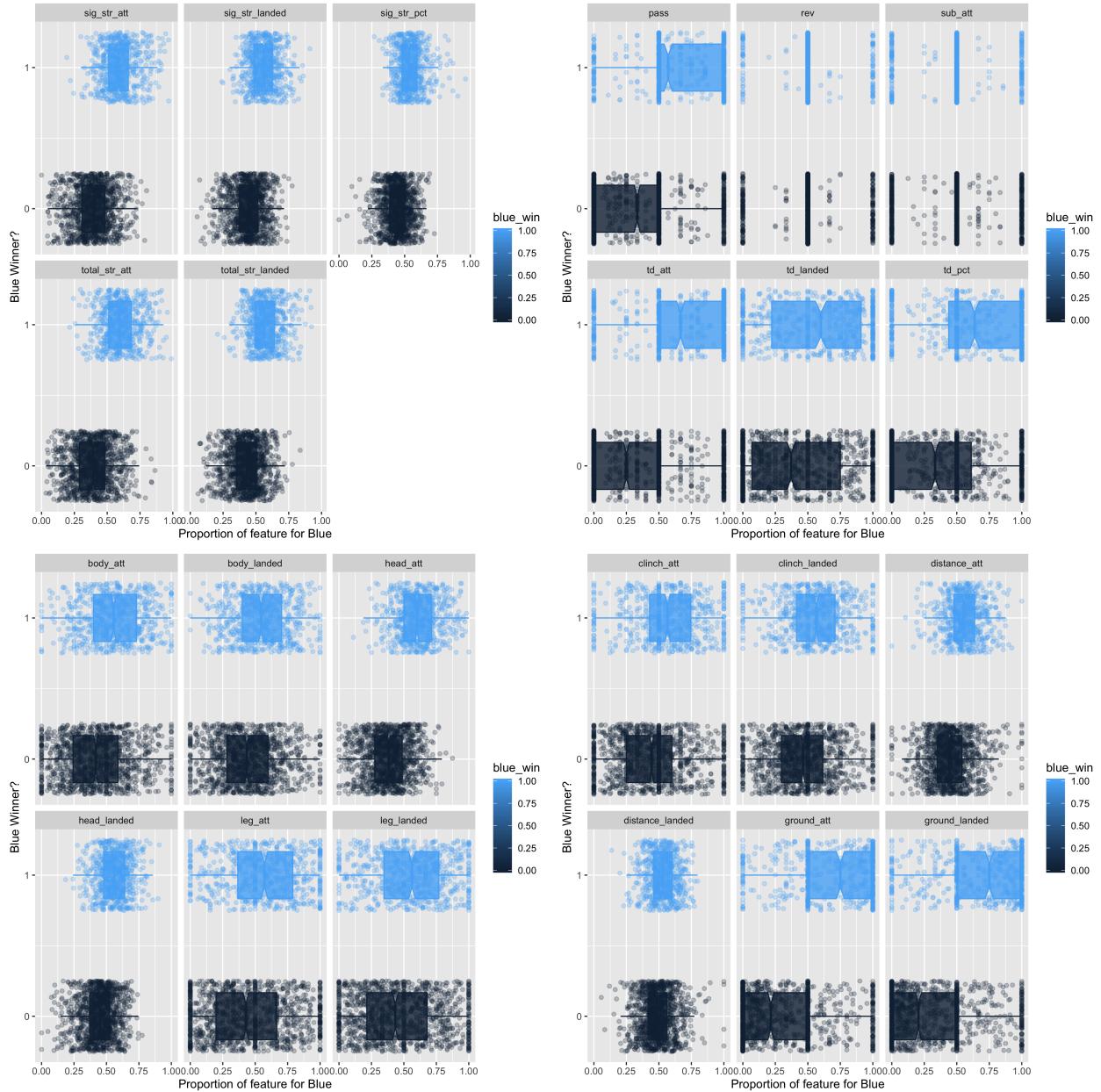


Figure 2: Comparison of the distributions of the predictors between winning and losing in different groups. Top left: striking features, top right: ground features, bottom left: attacks-to features, bottom right: attacks-from features.

Table 1: Table 1. REF selected features and their weights.

Features	Weights
sig_str_att	-3.4293230
head_att	-3.2237744
total_str_att	-2.8965256
td_att	-2.3422950
distance_att	-1.4891628
distance_landed	1.3455524
pass	-1.2626370
total_str_landed	-1.1250887
td_pct	1.1130846
td_landed	0.7163905
ground_att	-0.6729079



Figure 3: Figure 3. The train and validation error for including different numbers of features in the model .

Strking/Grappling group, such as the top second feature “head\_att”, suggesting there are some other factors which the judges put higher weights on.

*Note: The features’ full name and explanation can be found here.*

To validate our feature selection, we calculate the train and test errors using recursive feature elimination (RFE) with the “n\_features\_to\_select” hyperparameter ranging from 1 to 22 (Figure 3). It is observed that when the “n\_features\_to\_select” hyperparameter is between 8 and 11, the train and validation errors are relatively small, confirming the results from RFE with cross validation.

We built our final logistic regression model using the 11 features selected by RFE with cross validation and compared it with a logistic regression model using all the features. The accuracy on the train and validation data sets are reasonable for both models. And the accuracy for the model with feature selection is slightly better than the model without feature selection.

The confusion matrix showed similar accuracy, where our model correctly predicted 368 out of 446 validation cases and incorrectly predicted 78 cases with 40 being false positive and 38 false negative cases (Figure 4).

In sum, we identified the key predictors for UFC winning. Most of these strong predictors are indicators of Strking/Grappling performance, which is consistent with the top criteria in UFC judgment. However, several other predictors are also heavily weighted by the judges. Although our model performed well on the validation data set, there is still room to improve and optimize the model to achieve better accuracy.

Table 2: Table 2. Accuray of model performance on train and validation data sets.

Model	Score
Training Error - No Feature Selection	0.8433128
Training Model - Selected Features	0.8461108
Testing Error - No Feature Selection	0.8206278
Testing Error - Selected Features	0.8251121

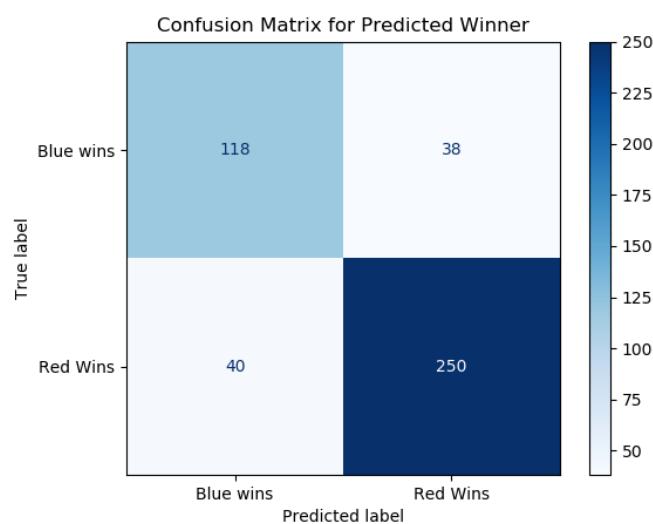


Figure 4: Figure 4. Confusion matrix of model performance on test data on validation data set.

However, the results may also reflect the inherent heterogeneity with the data set, suggesting there are big variations in the subjective judgement on different cases.

## References

- al, Barret Schloerke et. 2018. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=GGally>.
- Buitinck, Lars, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, et al. 2013. “API Design for Machine Learning Software: Experiences from the Scikit-Learn Project.” In *ECML Pkdd Workshop: Languages for Data Mining and Machine Learning*, 108–22.
- de Jonge, Edwin. 2018. *Docopt: Command-Line Interface Specification Language*. <https://CRAN.R-project.org/package=docopt>.
- Hunter, J. D. 2007. “Matplotlib: A 2D Graphics Environment.” *Computing in Science & Engineering* 9 (3): 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- Kaiser, Daniel. 2019. *Chromedriver-Binary 80.0.3987.16.0*. <https://pypi.org/project/chromedriver-binary/>.
- Keleshev, Vladimir. 2014. *Docopt: Command-Line Interface Description Language*. <https://github.com/docopt/docopt>.
- McKinney, Wes. 2010. “Data Structures for Statistical Computing in Python.” In *Proceedings of the 9th Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman, 51–56.
- Muthukadan, Baiju. 2019. *Selenium 3.141.0*. <https://selenium-python.readthedocs.io>.
- Oliphant, Travis E. 2006. *A Guide to Numpy*. Vol. 1. Trelgol Publishing USA.
- Rajeev Warrier. 2019. “UFC-Fight Historical Data from 1993 to 2019”. <https://www.kaggle.com/rajeevw/ufcdata>.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Reitz, Kenneth. 2019. *Requests: HTTP for Humans*. <https://pypi.org/project/requests/>.
- Sam Firke, Chris Haid, Bill Denney. 2020. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- The ABC MMA Rules Committee. 2017. “MMA Judging Criteria/Scoring-Approved August 2, 2016”. [http://www.abcboking.com/wp-content/uploads/2016/08/judging\\_criteriascoring\\_rev0816.pdf](http://www.abcboking.com/wp-content/uploads/2016/08/judging_criteriascoring_rev0816.pdf).
- VanderPlas, Jacob, Brian Granger, Jeffrey Heer, Dominik Moritz, Kanit Wongsuphasawat, Arvind Satyanarayan, Eitan Lees, Ilia Timofeev, Ben Welsh, and Scott Sievert. 2018. “Altair: Interactive Statistical Visualizations for Python.” *Journal of Open Source Software*, December. <https://doi.org/10.21105/joss.01057>.
- Van Rossum, Guido, and Fred L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Wickham, Hadley. 2017. *Tidyverse: Easily Install and Load the ‘Tidyverse’*. <https://CRAN.R-project.org/package=tidyverse>.