



## UFC

Starting in 1993 as a professional mixed martial arts (MMA) organization, UFC® has revolutionized the fight business and today stands as a premium global sports brand, media content company and the largest Pay-Per-View (PPV) event provider in the world.

The goal is to find “the Ultimate Fighting Champion” by staging a single-night tournament featuring the best athletes skilled in the various disciplines of all martial arts, including karate, jiu-jitsu, boxing, kickboxing, grappling, wrestling, sumo and other combat sports.

## CONTENT

Each row is a compilation of both fighter stats. Fighters are represented by 'red' and 'blue' (for red and blue corner). So for instance, red fighter has the compiled average stats of all the fights except the current one. The stats include damage done by the red fighter on the opponent and the damage done by the opponent on the fighter (represented by 'opp' in the columns) in all the fights this particular red fighter has had, except this one as it has not occurred yet (in the data). Same information exists for blue fighter.

The target variable is 'Winner' which is the only column that tells you what happened. Column definitions can be found in appendix 1.

## PROBLEM STATEMENT

A Training Dataset comprising of the various details of the match has been provided. You need to train your model using the given data set. You may use any or all variables from the dataset to predict the outcome of the match.

Also attached is a test dataset that **DOES NOT** contain the winner of the match. You need to use your model to predict the outcome of the matches in the test dataset. You may use R/Python or any programming language of your choice.

For submission, kindly send a zip file containing:

- the testing dataset ALONG with the winner predictions.
- Code of your model
- A short report provide detailing the logical flow along with the reasoning behind steps performed and insights derived from the data.

Name of the file should be in the format: TeamName\_CollegeName.zip

Submissions shall be evaluated on Mean Absolute Error between your predicted WinnerName and the observed WinnerName in the **test dataset (real outcome of the match)**. The observed WinnerName shall not be shared with the participants.

Depth of analysis and Logical Flow shall also have weightage in the results.

Please send the write-up and necessary files for analysis(excel sheets/code files) to **datascience.club@iimbg.ac.in** by **17th November 2019 23:59:59**

For any clarifications, contact:

Vipul Sehgal

[vipuls2020@iimbg.ac.in](mailto:vipuls2020@iimbg.ac.in)

ph: +918281400554



## Appendix 1: Column definitions:

Variable	Meaning (R_ and B_ prefix signifies red and blue corner fighter stats respectively)
_opp_	containing columns is the average of damage done by the opponent on the fighter
KD	is number of knockdowns
SIG_STR	is no. of significant strikes 'landed of attempted'
SIG_STR_pct	is significant strikes percentage
TOTAL_STR	is total strikes 'landed of attempted'
TD	is no. of takedowns
TD_pct	is takedown percentages
SUB_ATT	is no. of submission attempts
PASS	is no. times the guard was passed?
HEAD	is no. of significant strikes to the head 'landed of attempted'
BODY	is no. of significant strikes to the body 'landed of attempted'
CLINCH	is no. of significant strikes in the clinch 'landed of attempted'
GROUND	is no. of significant strikes on the ground 'landed of attempted'
win_by	is method of win
last_round	is last round of the fight (ex. if it was a KO in 1st, then this will be 1)
last_round_time	is when the fight ended in the last round
Format	is the format of the fight (3 rounds, 5 rounds etc.)
Referee	is the name of the Ref
date	is the date of the fight
location	is the location in which the event took place
Fight_type	is which weight class and whether it's a title bout or not
WinnerName	is the name of the winner of the fight
WinnerCorner	is the name of the winning corner of the fight
Stance	is the stance of the fighter (orthodox, southpaw, etc.)
Height_cms	is the height in centimeter
Reach_cms	is the reach of the fighter (arm span) in centimeter
Weight_lbs	is the weight of the fighter in pounds (lbs)
age	is the age of the fighter
title_bout	Boolean value of whether it is title fight or not
weight_class	is which weight class the fight is in (Bantamweight, heavyweight, Women's flyweight, etc.)
no_of_rounds	is the number of rounds the fight was scheduled for
current_lose_streak	is the count of current concurrent losses of the fighter
current_win_streak	is the count of current concurrent wins of the fighter
draw	is the number of draws in the fighter's ufc career
wins	is the number of wins in the fighter's ufc career
losses	is the number of losses in the fighter's ufc career
total_rounds_fought	is the average of total rounds fought by the fighter
total_time_fought(seconds)	is the count of total time spent fighting in seconds
total_title_bouts	is the total number of title bouts taken part in by the fighter
win_by_Decision_Majority	is the number of wins by majority judges decision in the fighter's ufc career



win_by_Decision_Split	is the number of wins by split judges decision in the fighter's ufc career
win_by_Decision_Unanimous	is the number of wins by unanimous judges decision in the fighter's ufc career
win_by_KO/TKO	is the number of wins by knockout in the fighter's ufc career
win_by_Submission	is the number of wins by submission in the fighter's ufc career
win_by_TKO_Doctor_Stoppage	is the number of wins by doctor stoppage in the fighter's ufc career
Sex	Gender of the fighters