



# Machine Learning problem : Predicting UFC's matches results

A promotional poster for a UFC Fight Night event. It features two fighters: on the left, a man with a beard and short hair (Ortega), and on the right, a man with a short buzz cut (The Korean Zombie). The background is dark and moody. The text is centered and includes the UFC logo, the event name, the fighters' names, the date, and the venue.

**UFC**  
**FIGHT NIGHT**

FEATHERWEIGHT BOUT  
#2 **ORTEGA** *vs* **THE KOREAN ZOMBIE** #6

#UFCBUSAN **DEC 21 SAT** 사직실내체육관  
**SAJIK ARENA**



# Introduction

- The UFC is currently the biggest MMA (Mixed Martial Arts) organisation in the world where the best fighters are competing against each other. It started to be popular in the 90s.
- Around 500 fighters distributed in several weights division (from flyweight to heavyweight)
- Around 400 fights organized per year
- UFC Bettings have become plebiscite by an increasing number of viewers. Some big events where fights are very publicized can gather a huge amount of bettings.



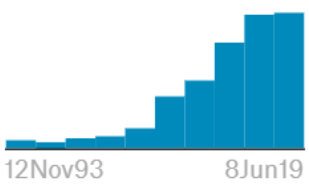
# Our goal

- Predict the winner at each UFC event (the one with the highest probability) using a Machine Learning model : a classification algorithm
- Collect all data available, preprocess it and perform some feature engineering in order to merge it as a fittable dataset for our model.
- Compare different classification algorithms and keep the best one
- Optimize it as far as we can (hyperparameter tuning, bias/variance trade-off)





# The dataset

data.csv (5.09 MB)							
				20 of 100 columns	Views		
	A R_fighter	A B_fighter	A Referee	date	A location	A Winner	
	1334 unique values	1774 unique values	Herb Dean 14% John McCarthy 12% Other (189) 74%		Las Vegas, Neva... 24% London, England, ... 2% Other (155) 74%	Red Blue Other (1)	
1	Henry Cejudo	Marlon Moraes	Marc Goddard	2019-06-08	Chicago, Illinois, USA	Red	
2	Valentina Shevchenko	Jessica Eye	Robert Madrigal	2019-06-08	Chicago, Illinois, USA	Red	
3	Tony Ferguson	Donald Cerrone	Dan Miragliotta	2019-06-08	Chicago, Illinois, USA	Red	
4	Jimmie Rivera	Petr Yan	Kevin MacDonald	2019-06-08	Chicago, Illinois, USA	Blue	
5	Tai Tuivasa	Blagoy Ivanov	Dan Miragliotta	2019-06-08	Chicago, Illinois, USA	Blue	
6	Tatiana Suarez	Nina Ansaroff	Robert Madrigal	2019-06-08	Chicago, Illinois, USA	Red	
7	Aljamain Sterling	Pedro Munhoz	Marc Goddard	2019-06-08	Chicago, Illinois, USA	Red	

5144 rows  
(matches) and  
145 columns  
(variables)



# The dataset

- The dataset is the list of every UFC fight in the history of the organization
- Every rows contain informations about both fighters : fight details (variables) and the winner (target variable)
- Fighters are represented by « red » or « blue » corner.
- Some of the variables of the dataset describe fighters statistics. **Those statistics are computed and averaged on all previous fights until the one that is happening (very important !).** That's why our model must be updated after each UFC event to really be efficient.



# My plan

Data preprocessing (including cleaning if needed)



Feature Engineering



Split data



Build several classification models (ML and DL)



Optimize them (hyperparameters with cross-validation)



Compare accuracies and keep the best model using precision, recall and F-1 score evaluation indicators