Machine Learning problem : Predicting UFC's matches results

Introduction

- The UFC is curently 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.
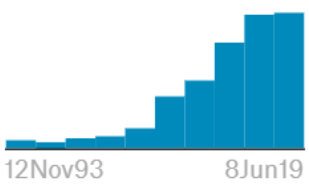
Our goal

- 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

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