UFC Fight Prediction

Helar Jaadla, Henrik Lepson, Tanel Orumaa

TARTU ÜLIKOOL · UNIVERSITAS TARTUENSIS · 1632

IT Akadeemia toetab Skype™

Background information

UFC (Ultimate Fighting Championship) is the most important and well known MMA (mixed martial arts) promotion in the world. UFC events take place every week or every other week all year round and they attract a lot of spectators. Not everyone simply watches the fights, but there are people who bet on the outcomes or participate in prediction competitions. Statistical methods and data analysis has become popular in the sports during the last decades and it has proven its effectiveness. This raises the question: can we use data about fighters physical attributes and past fights for predicting fight outcomes?

Goals

The main goals of the project were following:

- Identifying the attributes that affect the fight outcome the most
- Creating and training a classification model that uses the most important attributes to predict outcomes of future UFC fights
- Cluster analysis on fighting styles of different fighters

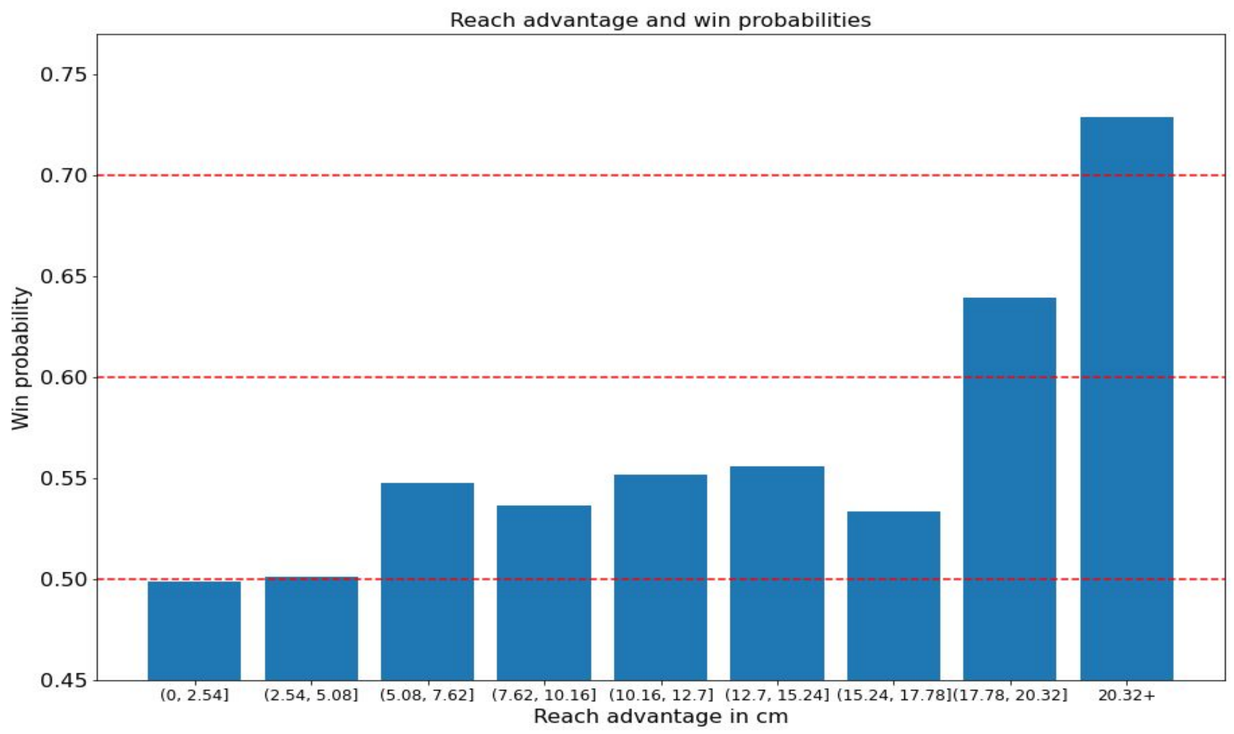
Data

The main dataset used for this project was the UFC Ultimate dataset, which is available on Kaggle. The dataset contains various information, such as the names of the fighters, their previous records, their physical attributes, average strikings stats and so on. The quality of the dataset was quite good. There were minor issues with missing values and values that were wrong, but they were easy to fix using the data that is available on the UFC website.

The first dataset is not well suited for cluster analysis and because of that a second small dataset was created using data about UFC fighters that can be found on the UFC website. This dataset contains information about 18 well known UFC fighters and their fighting tendencies (striking stats, takedowns landed etc).

Key findings

- Length of the arms matters. Fighters who have a reach advantage over their opponents win more often.
- Height does not matter. Taller fighters win only a little more than half of their fights (51%).
- Do not pick the rookie. Fighters making their UFC debuts win only 43% of the time.
- If a fighter has lost two or three fights in a row their chances of winning the next fight are over 70%.
- Always pick the champion (71% win rate).

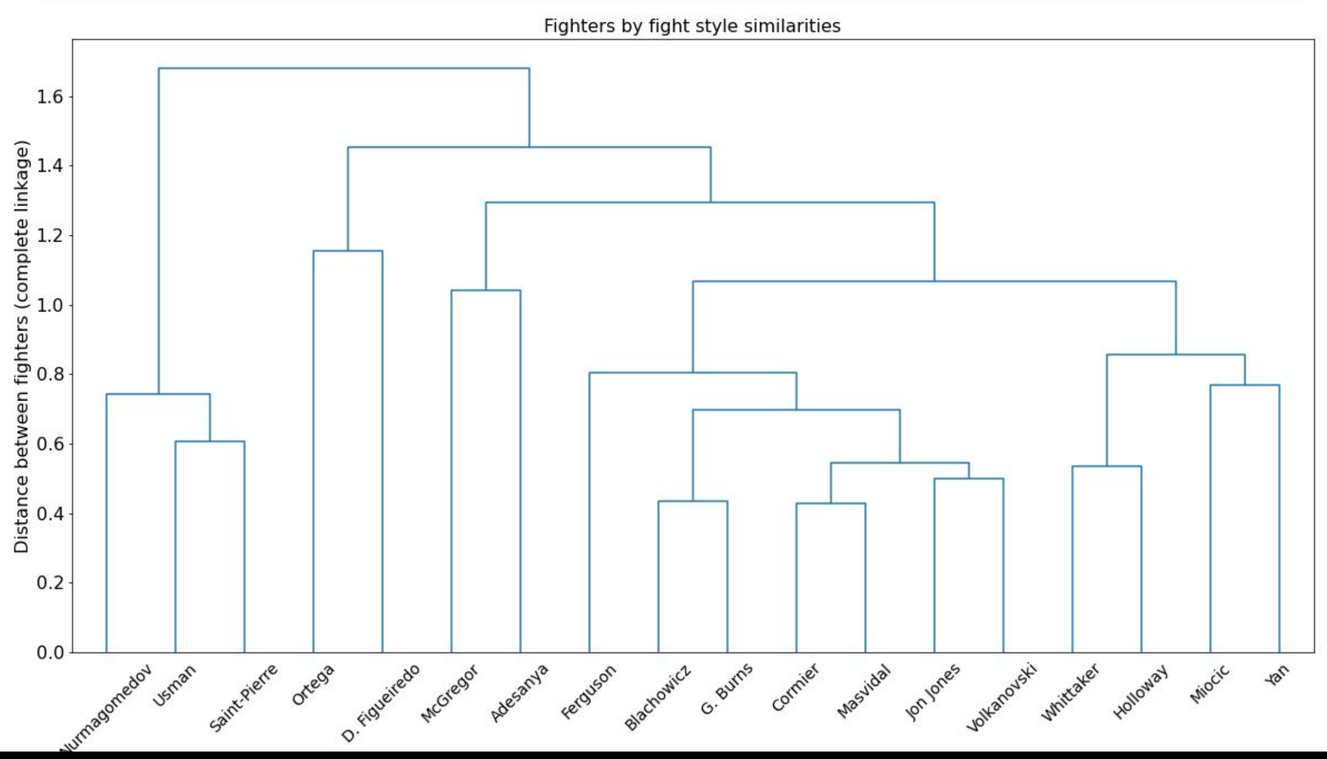


	KNN (without odds)	Random Forest (without odds)	KNN (with odds)	Random Forest (with odds)
UFC Vegas 15	60%	50%	40%	50%
UFC Vegas 16	50%	62.5%	75%	75%
UFC 256	60%	60%	70%	80%
Accuracy	57.14%	57.14%	60.71%	67.86%

With or without odds tells us whether betting odds were used for the model training.

Prediction results

The best prediction results were achieved with a Random Forest Classifier model that was able to predict fights from the last 3 UFC events (November 28, December 5 and December 12) with an accuracy of 67.9%. The expectation before the project was 70% accuracy, but because fighters are usually well matched and there is a lot of uncertainty involved in each fight it is still a good result and shows that predicting fights is possible.



Cluster analysis

Hierarchical clustering of fighters was done using complete linkage (shortest distance between cluster elements that are farthest away). Every feature in the dataset that was not already a proportion was divided by the max value of the column to avoid out of proportion effects of some of the features such as average fighting time, which can take values of up to 25 (minutes). This made the distances between fighters shorter and the end result better.

Project repository: https://github.com/Henrik895/FightPrediction