Which is a better predictor for the outcome of UFC fights - physical attributes or career data?

Talhaa Hussain

University of Exeter

2023

- The UFC is the biggest mixed martial arts promotion in the world, and has been credited with popularising the sport globally. Created in 1993, the UFC has promoted over 200 events, and generated over 1 billion USD in 2021 [1].
- Fights take place between two competitors, with the higher ranked fighter (or champion) fighting out of the red corner and the other fighter representing the blue corner.
- As with all sports, there's a lot of speculating and predicting surrounding the outcome of a match-up, for various purposes such as sports analytics and gambling.
- This project aims to explore which factors are better for predicting the outcomes of fights - physical attributes of fighters, or their career achievements and statistics.

The Dataset

Background

- The dataset is a list of every single fight to take place in the organisation since its inception in 1993, all the way up to 2021.
- Over 100 fields (features) and 6'000 records (instances) are provided, with metrics on both both fighters, the date, location and context included.
- The dataset is from kaggle.com, and is highly commended. The original data was scraped from ufcstats.com, a popular website for tracking and recording data about fighters, bouts and events in the UFC.
- The dataset can be accessed at https://www.kaggle.com/datasets/rajeevw/ufcdata.

Classification

Machine Learning

The three possible types of machine learning technique that could've been used were...

Regression

- Used to predict **continuous** values, such as numerical data.

Classification

- Involves a set number of **classes**, and aims to classify data into these classes.

Clustering

- Organises data into **groups**, based on patterns in the data.

A classification approach was used, because the goal of the ML algorithm is to be able to predict outcomes of fights as either "red wins" or "blue wins".

Multi Layer Perceptron Neural Networks

Machine Learning

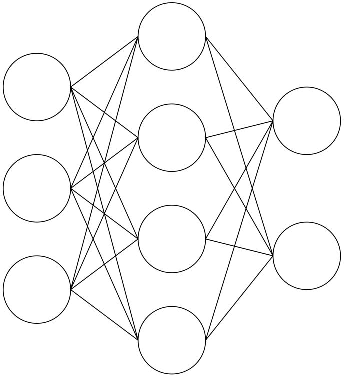

Figure 1: A MLP Neural Network. Inputs from previous layers feed to the next layer.

- Multi Layer Perceptrons (MLPs) are a type of neural network, a supervised learning technique, built up from layers of perceptrons.
    - Supervised learning techniques train on labelled data to classify new inputs.
- Perceptrons in a layer "feed forward" to the next layer, providing input to the perceptrons in that layer.
- MLPs take features as inputs and classify the given item into a certain class based on these features.

Multi Layer Perceptron Neural Networks (cont.)

Machine Learning

MLPs were selected because of their ability to implicitly find relationships between variables and their well-documented effectiveness. MLPs are widely considered one of the best supervised classification machine learning techniques, and work effectively with non-linear and complex relationships [2].

This is particularly useful for this problem, as mixed-martial arts is a sport that is dependent on several factors, some clearly visible and others much more hidden or implicit. MLPs have the ability to extract these hidden variables and effectively harness and utilise them to draw conclusions and make classifications.

Considerations

- Prior to 21/03/2010, the red/blue system was not used - all winning fighters prior to this are assigned as "red" in the dataset. After careful consideration, it is believed that this should have no negative effect. This will be confirmed later with testing.
- BMI is not included in the dataset, but may be a desirable feature to have. This will be computed using height and weight of fighters from other fields - feature engineering.

Encoding and Normalisation

- Most fields are already numerical. For those that aren't, a simple integer encoding is used. For example, for fight results: 0 = Red wins; 1 = Blue wins.
- Data is normalised to avoid one feature taking dominance when training the MLP. This is done by scaling each column, so that the maximum value is 10.0 and the minimum is 1.0.
- Records with NULL fields are eliminated.

Data Cleaning and Pre-processing (cont.)

- Draws are not accounted for, since they are extremely rare and make up a negligible part of the dataset. Furthermore, most contemporary and professional analytical tools don't predict draws between fighters.
- Weight classes and sex are eliminated from the dataset. Weight classes are redundant when the fighter's weight is available as a separate feature, and both fighters always fight at the same weight class, against someone of the same sex.
- The dataset is split into two parts, one part for use by the MLP considering physical attributes, the other for use by the MLP concerned with career data.
    - Physical attribute features include both fighters' weights, heights, BMIs (engineered), reach and stance (hand preference).
    - Career data features include both fighters' current winning/losing streaks, their wins and losses (career totals), and their career longest winning streak.

Implementation is done on Python 3, using numpy, pandas and scikit-learn.

- The dataset CSV is imported as a pandas dataframe, which is cleaned and normalised as specified before. This is then converted to a numpy ndarray, and split into training and test data.
- A MLP classifier is initialised through scikit-learn, and trained using the training data. It is then tested using the test data, and gives results as an accuracy percentage and a confusion matrix, to give insights into performance.

This is done twice in two separate Python programs; once for physical features and once for career features. These results are then compared to evaluate their successes and shortcomings. Continual effort is made to ensure that neither program has an unfair advantage; however, complete fairness cannot be guaranteed.

- **Physical attributes** yielded the following results for ten successive runs: 64.7%, 63.2%, 63.5%, 63.2%, 62.2%, 63.8%, 64.2%, 64.7%, 63.3%, 65.0%. This gives a mean of 63.73%.
  - The average confusion matrix was $\begin{bmatrix} 935.1 & 58.3 \\ 424 & 58.6 \end{bmatrix}$.
- **Career statistics** yielded the following results for ten successive runs: 68.4%, 66.4%, 67.5%, 67.3%, 67.1%, 67.5%, 68.4%, 67.5%, 68.0%, 65.1%. This gives a mean of 67.32%.
  - The average confusion matrix was $\begin{bmatrix} 669.8 & 105.5 \\ 334.2 & 107 \end{bmatrix}$.

- At a glance, career statistics seemed to be much better features for predicting the outcome of bouts, scoring +3.59% on average across 10 attempts.
- However, career statistics also yielded a lot more false positives on average, by a landslide amount of almost double across 10 successive runs.
- Physical attributes actually performed significantly better than career statistics regarding true positives found on average, beating the latter by approximately +39.6%.
- Both models reported false positives, true negatives and false negatives somewhat proportionally on average.

Evaluating Results (cont.)

While it may **seem** like career statistics are better for predicting fight outcomes, there are a few things to consider.

- Both feature sets performed quite poorly overall at classification, with neither method ever scoring above 70%. This suggests that these features are somewhat ineffective/incomplete with regards to bout prediction, at least in their current forms.
- The significantly stronger score on true positives for the physical attribute feature set indicates that, had the dataset been larger, the outcomes between the feature sets may have been closer, and physical attributes may have even outscored career statistics for classification.
- Both feature sets struggled massively with reporting false negatives (instances where blue won the bout). This was initially assumed to be due to the skew of historical data (with red winning most bouts), but even after removing the pre-2010 dataset and retesting this remained the case, suggesting that this was not the issue.

Limitations of the Data and Results

Data Analysis

There were some clear limitations of this experiment that should be acknowledged. Some of these trade-offs are unavoidable when translating theory into practice.

- There haven't been that many UFC bouts; the dataset was quite small initially, and was shrank further by the elimination of NULL values and data cleaning. This means that training may not have happened to an optimal standard, leading to subpar scores.
- The ruleset of the UFC regarding fighter colours has changed over time, and this may have impacted statistics and learning.
- The classification techniques do not take different types of win into account, nor do they account for a fighter's combat style. The classification technique does not account for ties/draws or no contests.
- Within career statistics or physical attributes, there are some properties of fighters that are not publicly available or accessible, which may explain low accuracy scores.
- The classification does not account for meta-game - the idea that fighting has evolved over time, with certain techniques becoming more dominant and certain techniques regressing.

Conclusions

- It cannot be concluded from these results whether physical attributes or career statistics are **definitively** better for predicting the outcome of a fight. Both models have their advantages and disadvantages in different ways.
- While the models performance of the models are underwhelming, they have a low variance - classification accuracy is consistent. If some techniques can be applied to improve accuracy, both models have some potential for predicting fairly well.
- It **can** be concluded that both feature sets are somewhat incomplete; additional information is needed in order to make an accurate outcome estimate.

References

[1] (), [Online]. Available: https://www.mmafighting.com/2023/2/28/23619576/ufc-generated-record-revenues-in-2022-including-the-best-sponsorship-year-ever.

[2] K Hemalatha and K. U. Rani, "Advancements in multi-layer perceptron training to improve classification accuracy," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 5, no. 6, pp. 353–357, 2017.