Which is a Better Predictor for the Outcome of UFC Fights - Physical Attributes or Career Data?

Talhaa Hussain

1 Background

1.1 MMA and the UFC

MMA is a sport that has gained a lot of popularity over the last few years, with many more watchers and participants. 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. With the massive pull of the UFC, the ability to predict the outcome of fights has become more and more sought after. This project aims to compare two potential ways of predicting fight outcomes. One of these involves looking at both fighters' career statistics; the other involves looking at the physical attributes of fighters. This project hopes to compare the two approaches and determine which is better in what circumstances.

1.2 The Dataset

The dataset used in this project is adapted from kaggle. The dataset itself is a list of every single fight to take place in the UFC from 1993 to 2021. Over 100 fields and 6'000 records are provided, with metrics on both both fighters, the date, location and context included. The dataset can be accessed at https://www.kaggle.com/datasets/rajeevw/ufcdata.

2 Machine Learning

2.1 Classification

There are three types of machine learning technique - regression, classification and clustering. Regression techniques are used to predict continuous values, such as numerical data; classification techniques are used to classify data into classes, and clustering techniques are used to organise data into groups, based on patterns in the data. In this project, a classification technique was chosen. This is because the goal of the ML algorithm is to be able to predict outcomes of fights as either "red wins" or "blue wins"; or, to be able to classify fights into one of two classes. It makes very little sense to use regression or clustering in this instance.

2.2 Multi Layer Perceptrons

Multi layer perceptrons (MLPs) are a type of neural network. They are a supervised learning technique that are built up from layers of perceptrons. Perceptrons are the individual units behind decision-making, inspired by the neurons of the human/animal brain. 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.

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 [1].

3 Experiment and Method

3.1 Data Cleaning and Pre-processing

Before using inputting the data into the machine learning algorithm, I eliminated any items with NULL fields, normalised all feature columns (by scaling to make sure that all values fall between 1.0 and 10.0 inclusive), and performed encoding (making sure all value were numerically represented). Data is normalised to avoid one feature taking dominance when training the MLP. For encoding, a simple integer encoding is used - for example, for fight results: 0 = Red wins; 1 = Blue wins. I also needed to engineer a desired feature that wasn't originally in the dataset - BMI. This was calculated using the height and weight of fighters (pulled from other fields). An important consideration I had to make was that 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. At first, I was concerned that there may be skewing because of this, however, after careful consideration, I believed that this should have no negative effect. This is because the relationship between winners and their winning features remains unchanged. This will be confirmed later with testing.

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.

3.2 Implementation

The implementation is done on Python 3, using numpy, pandas and scikit-learn. Two seperate Python programs are produced, one for physical features and one for career features. Their results are then compared to evaluate their successes and shortcomings. The Python files are available inside this submission.

4 Data Analysis

4.1 Results

The full results from ten successive runs can be seen in Figure 1, on the following page.

4.2 Evaluation

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. Both models reported a lot of false negatives. 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.

4.3 Limitations of the Data and Results

There were some clear limitations of this experiment that should be acknowledged. The classification techniques do not take different types of win into account, nor do they account for a fighter's combat style. There haven't been that many UFC bouts, and the classification does not account for meta-game - the idea that fighting has evolved over time.

5 Conclusion

While we haven't reached a conclusion on which is better for predicting the outcome of UFC fights (physical attributes or career historical data), we have gleaned some new knowledge from this experiment. While the models performance of the models are underwhelming, they have a low variance - classification accuracy is consistent. It can also be concluded that both feature sets are somewhat incomplete; additional information is needed in order to make an accurate outcome estimate.

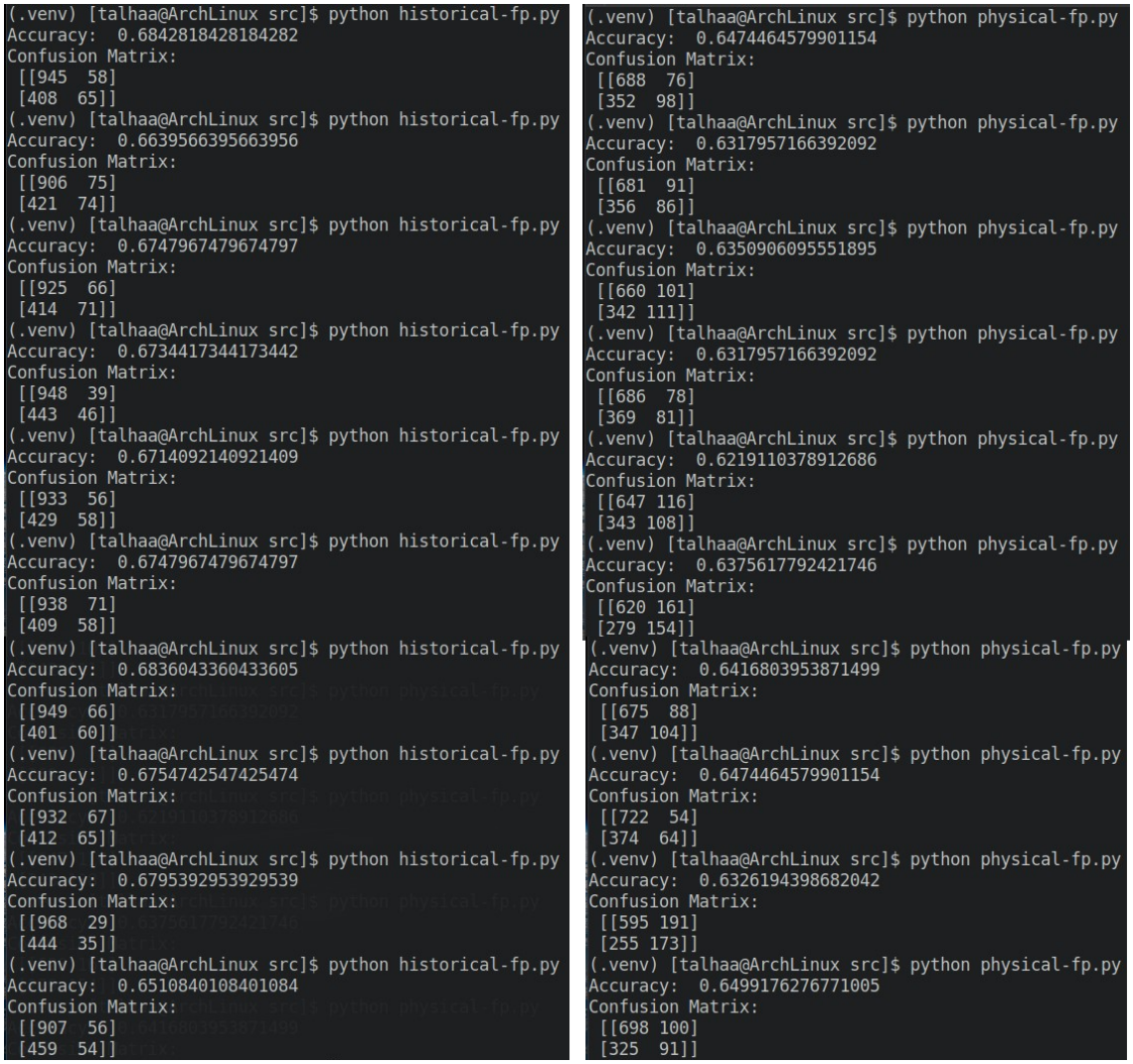

Figure 1: All results for 10 successive runs, including confusion matrices. On the right, we have physical-fp.py, for the physical feature set; on the left, we have historical-fp.py, for the opposite.

References

[1] K Hemalatha and K Usha Rani. Advancements in multi-layer perceptron training to improve classification accuracy. *International Journal on Recent and Innovation Trends in Computing and Communication*, 5(6):353–357, 2017.