Comparison of Machine Learning Techniques for the Prediction of UFC Fights

Owen Brooks
Drexel University
ECEC 487 - Pattern Recognition
Instructor: Dr. Andrew Cohen

1 Abstract
With the growing popularity of both sports and sports betting, machine learning is increasingly being used in an attempt to predict the outcome of popular sporting events. While also growing in popularity, the UFC, a mixed-martial-arts league, has not received the same focus from the machine learning community. As the sport has recently become more mainstream, detailed and extensive datasets are now available to the public. In this work, we attempt to compare and contrast different models based on how effective they are in predicting a one-on-one fight. Our experimental design will be based on approaches that have achieved good results in other sports. The algorithms tested include Xtreme Gradient Boosting, Gradient Boosting Trees. Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Multiple Layer Perceptron (MLP). In addition to this, the proper metric on which to evaluate and how to apply a prediction model to sports betting will be discussed.

2 Introduction
Sporting contests have historically drawn great interest not only from viewers but from those looking to predict the outcome. Betting on events is extremely popular which is reflected in a global sports betting market that is valued at around 150 billion USD. Viewing this through the lens of pattern recognition, a model that is able to successfully predict a sporting contest would have great value in both bettings directly and to the companies that run the sportsbooks. Various works have been published in this area. The vast majority of these covers the application of machine learning algorithms to popular sports such as football, basketball or horse racing [2]. In this work, these algorithms will be used to predict UFC fights. The UFC is a professional league for mixed-martial-arts or MMA, a fighting sport.

In a UFC fight, two fighters meet head to head in an octagon. The fights are typically three, five-minute rounds with the exception of title fights which are extended to 5 rounds. The fight can be ended early by a knockout, a submission or a technical knockout where the referee is forced to stop the fight on behalf of one fighter. If a fight does not end early the decision is based on how a committee of judges scores the fight.

We choose the UFC or Ultimate Fighting Championship for several reasons. Firstly my partner, Michael Manley, and I are members of the Drexel Wrestling team. As wrestling is one of the disciplines included in MMA, we have subject knowledge that is helpful in feature engineering

[2]. Additionally, in relation to sports leagues such as the NFL, for football, the UFC is still in its infancy having been started in 1993. It is however very popular with UFC 229 on October 6th, 2018 having 2.4 million pay-per-view buys. The popularity of the sport is important, as we have found it often relates closely to the availability of data. In order to use a test-train procedure on modern classification models, a large dataset with many features allows for more accurate predictions [2]. The UFC collects and releases detailed statistics on fights held, recording details such as the frequency and type of each attack taken by the fighters. They also maintain data on their fighter's physical attributes such as height and weight. This data is available online and can be collected using web scraping. These factors all make UFC a good candidate for the application of machine learning and pattern recognition.

3 Related Work

Dwivedy and Yadav [1] performed one of the few comparisons of machine learning algorithms for UFC prediction. They claimed appropriate data was only available after 2013 and based their finds on a dataset of 1700 fights. Using this they compared results of the train-test approach on Decision Trees, SVG, KNN, Perceptron, SVM, Random Forest, and Bayes. Random forest and SVM, with an RBF kernel, were shown to perform the best with 59.8% and 61.1% accuracy respectively. In an alternate approach, Singh [4] used only Random Forest to classify a dataset of 11,886 fight, with 1,390 of these being fights in the UFC. This experiment, using cross-validation with 10 folds, yielded an AUC of 0.69. Singh also correctly asserted that a model with good accuracy does not directly correlate to a profitable model.

Unlike the UFC, soccer has an abundance of work related to this subject. Chang and Huang [3] tested a multi-layer perceptron in the prediction of the winners of the 2006 World Cup. The dataset was divided based on the stages that occur throughout the tournament. Their process was interesting in that their training data included only the stages previous to the stage they were trying to predict. This experiment resulted in 76.9% accuracy. However, this came after a determination that games ending in a tie could not be predicted and were removed from the dataset.

Taking a sport agnostic approach, Bunker and Thabtah [2] attempted to clearly lay out the steps needed to predict sporting contests using Artificial Neural Nets (ANN). After examining the work of those using ANN's in various sports, their goal was to standardize the process and create what they called the Sports Result Prediction CRISP-DM framework. This involves the following steps: domain understanding, data understanding, data preparation, modeling, model evaluation, and deployment.

5 Experimental Results

I. Dataset

As mentioned above, UFC data is published by the organization online. We were able to obtain a web scraped dataset of 5061 fights, from kaggle.com. The data initially contained 160 features however, through my partner's feature selection process our final data set had only 28 features.

This can be simplified into 14 core features on each fighter going into the contest (Fig. 1). The process followed the steps in the Sports Result Prediction CRISP-DM framework [2]. This eliminated some of the features that were not impactful or noisy. It also increased the accuracy of our models. The fighters were labeled as red or blue, with the features using the same convention.

- Reach in centimeters
- Stance
- Age
- Average ground strikes landed
- Average takedown attempted
- Average takedown landed
- Average total head strikes percentage
- Average opposition distance landed
- Average opposition head strikes landed
- Average opposition takedown percentage
- Average opposition total strikes attempted
- Average opposition total strikes landed
- Weight/height ratio
- Wins by Split Decision

Figure 1: Core 14 Features. Each feature shown has a red and blue variant for each fighter.

II. Model Selection

Classification models were primarily selected using the literature review process [2]. The dataset was not linearly separable which was also a factor in our selection criteria. Through this, we selected four commonly used algorithms for sports prediction: K-Nearest Neighbors (KNN), Multi-Layer-Perceptron (MLP), Support Vector Machine (SVM), and Random Forest. The following sections will detail the parameter tuning and the results of these algorithms. Through further research, we discovered that many of the online data science competitions are being won using decision-tree based, ensemble models [5]. My partner, Michael Manley, tested two such models: Gradient Boosting Trees and Xtreme Gradient Boosting (XGBoost). The results of these algorithms will be compared with those in the previously mentioned list.

III. K-Nearest Neighbors

The first classification model tested was KNN, which uses the classifications of the closest K fights to make its final predictions. Using cross-validation, the data was divided into 5 samples [2]. The train test procedure was then done on these samples using K values ranging from 10 to 40 (Fig. 2). Using a combination of the mean and median accuracy of all 5 samples the optimum K value was found to be 25.

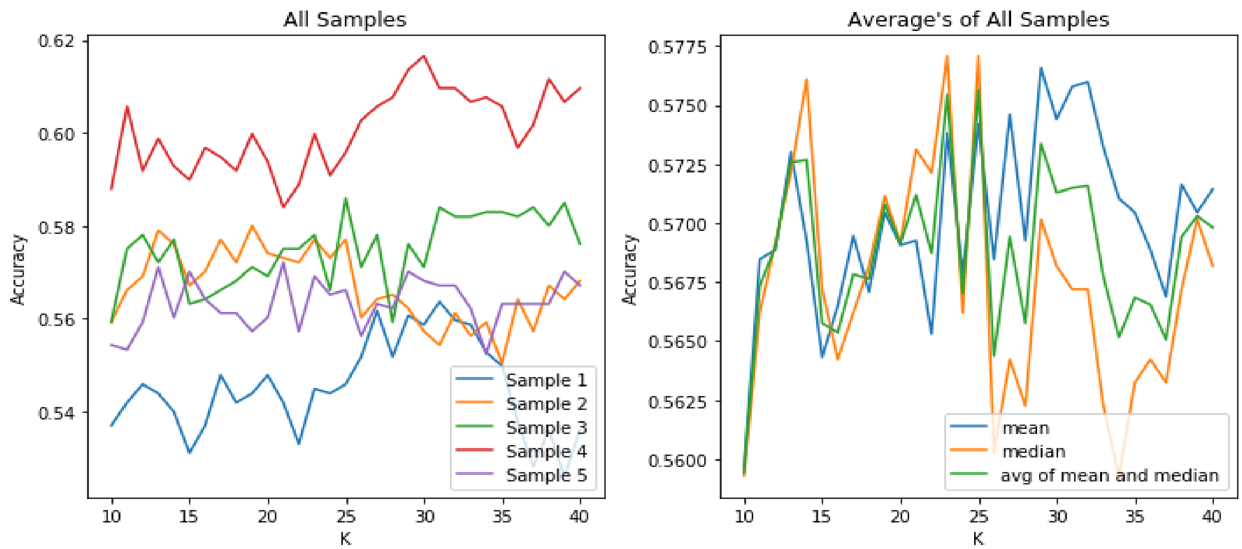

Figure 2: Results of Cross-Validation Testing to Find Optimum K Value

Using the previously found K value, the model achieved an accuracy of 59.82% and AUC of .64 (Fig. 3).

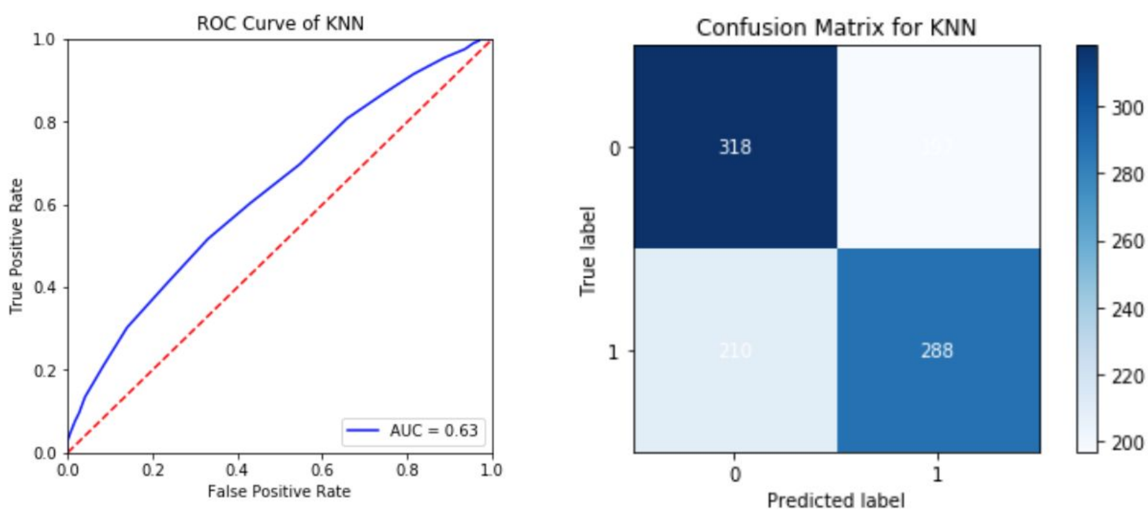

Figure 3: ROC Curve and Confusion Mat. of KNN

IV. Multi-Layer-Perceptron

The next model tested, MLP, is seemingly the most commonly used in sports prediction [2]. It involves a number of connected layers, with each layer having a certain number of neurons. These neurons are then mapped to a single output value - the model's prediction. In order to find the best MLP shape, the following parameters were held constant: the learning rate was .01, the activation function of hidden layers was relu, the activation function of the output layer was sigmoid, the batch size was 100, and the epochs was 100. The results of 9 different shapes are shown in Fig. 4.

Shape	Accuracy	Loss (MSE)
(10)	.6110	.6655
(20)	.6120	.6609
(40)	.6021	.6775
(10,10)	.6130	.6535
(20,20)	.6199	.6510
(10,10,10)	.6080	.6604
(20,20,20)	.6031	.6860
(10,20)	.6130	.6672
(20,40)	.6189	.6743

Figure 4: Testing of MLP Shape. For example, (10,10,10) is an input layer, three hidden layers each with 10 neurons, and an output layer.

Using the best combination of accuracy and loss, the shape of an input layer with 27 neurons, 2 hidden layers with 20 neurons each, and one output neuron were chosen. This model achieved an accuracy of 62% and an AUC of .68 (Fig. 5).

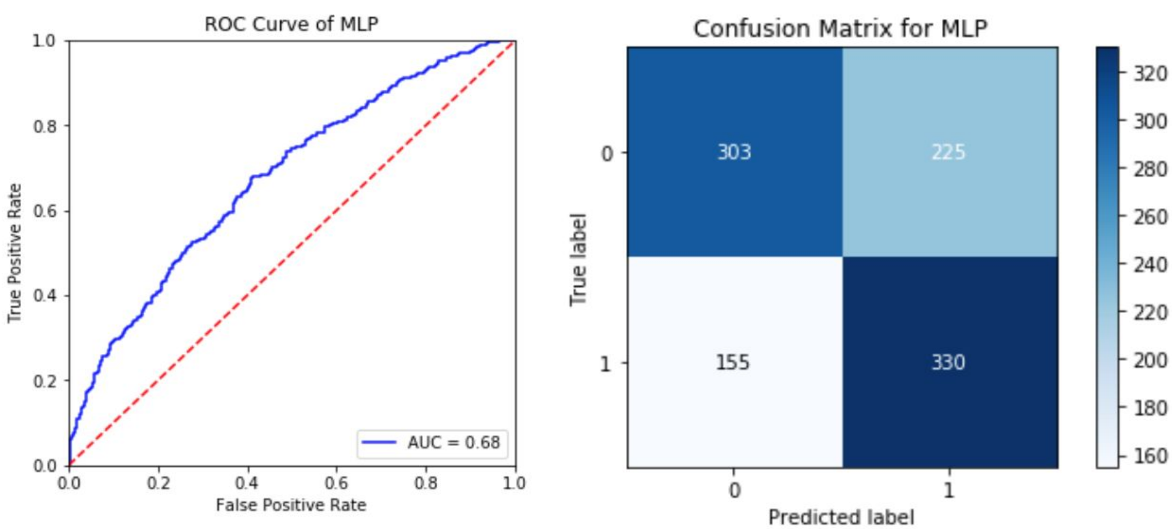

Figure 5: ROC Curve and Confusion Mat. of KNN

V. Support Vector Machine

The next model used was SVM, where the features are mapped into a hyperplane in order to create the prediction. A polynomial, radial basis function (RBF) and sigmoid kernel were all tested, with the RBF achieving the best results (Fig. 6).

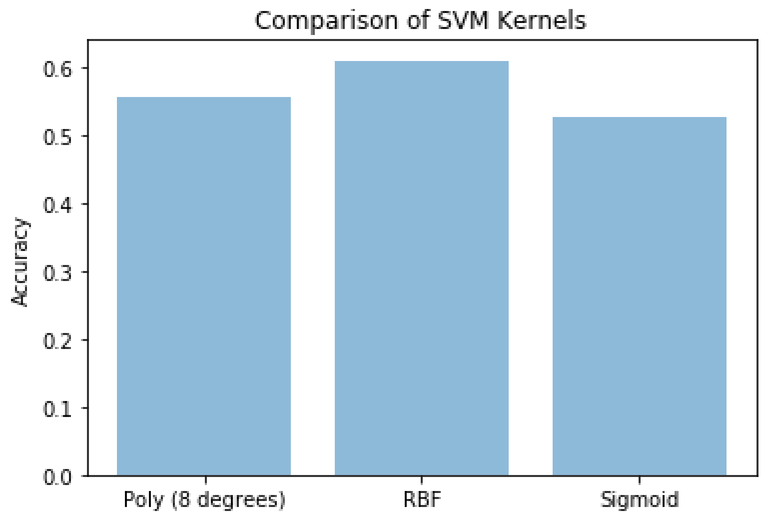

Figure 6: Comparison of SVM Kernels

Using the RBF kernel, the model achieved an accuracy of 63.33% and AUC of .70 (Fig. 7).

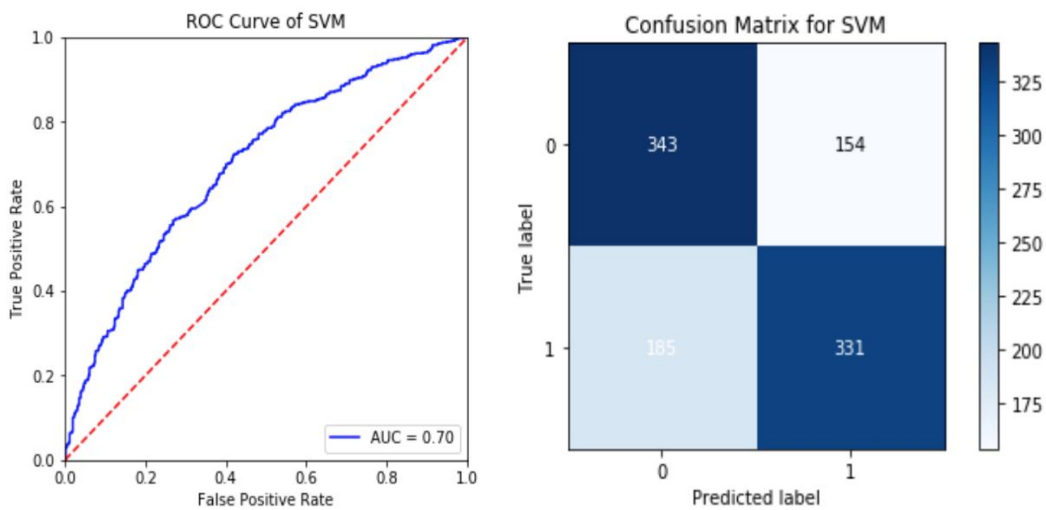

Figure 7: ROC Curve and Confusion Mat. of SVM

VI. Random Forest

The last model tested was Random Forest. This is an ensemble algorithm, meaning it combines the decisions of multiple decision trees in making a prediction. For parameter tuning, the value for the number of trees in the random forest was varied from 15 to 50. After the initial increase from 15 to 20 trees, the accuracy began to fluctuate (Fig. 8). The optimum value in the range tested turned out to be 22 trees.

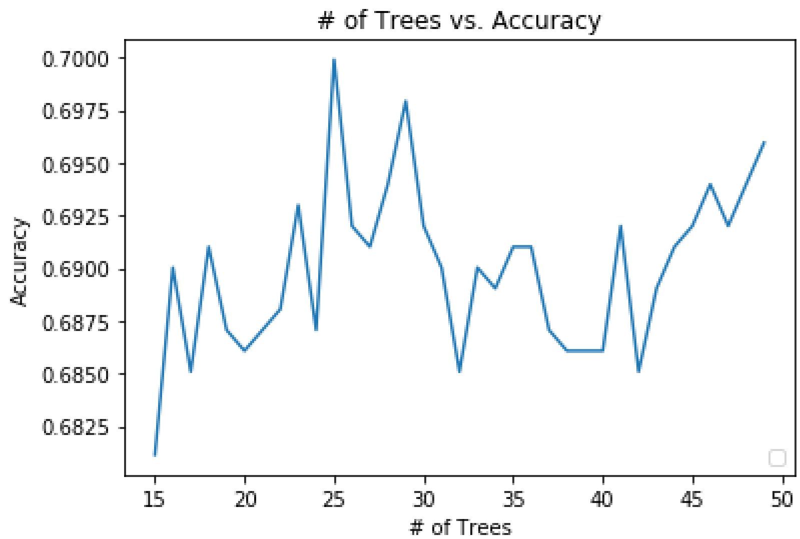

Figure 8: The Effects of the Number of Trees in Random Forest on Accuracy

With the number of trees set to 22, the Random Forest model achieved an accuracy of 68.29% and an AUC of .76 (Fig. 9).

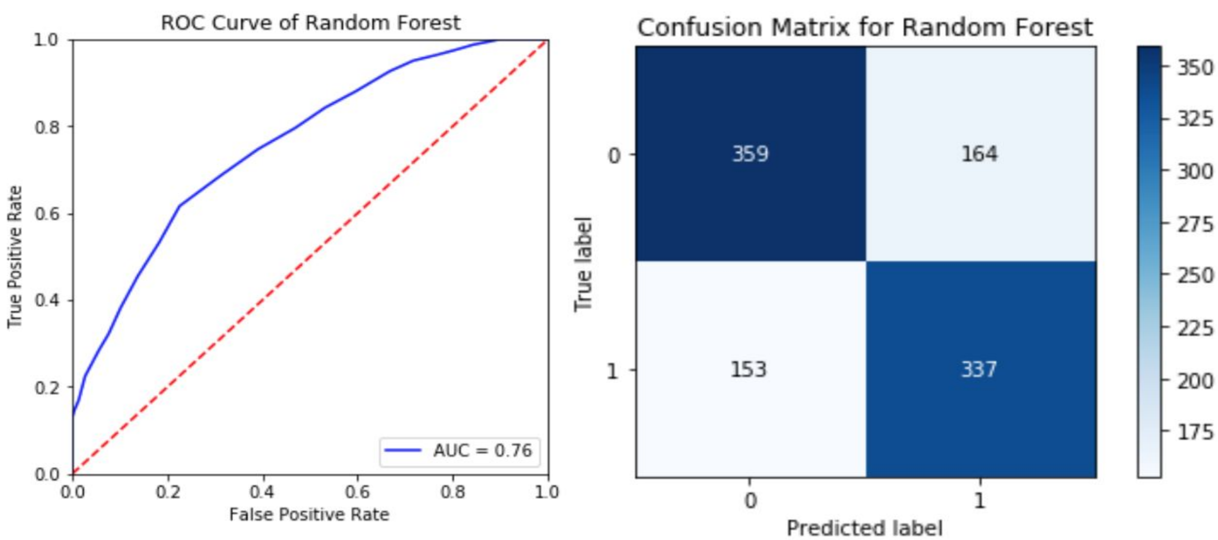

Figure 9: ROC Curve and Confusion Mat. of Random Forest

VII. Combined Results

The results of the four models mentioned above, as well as the two created by my partner, are shown in Fig. 10.

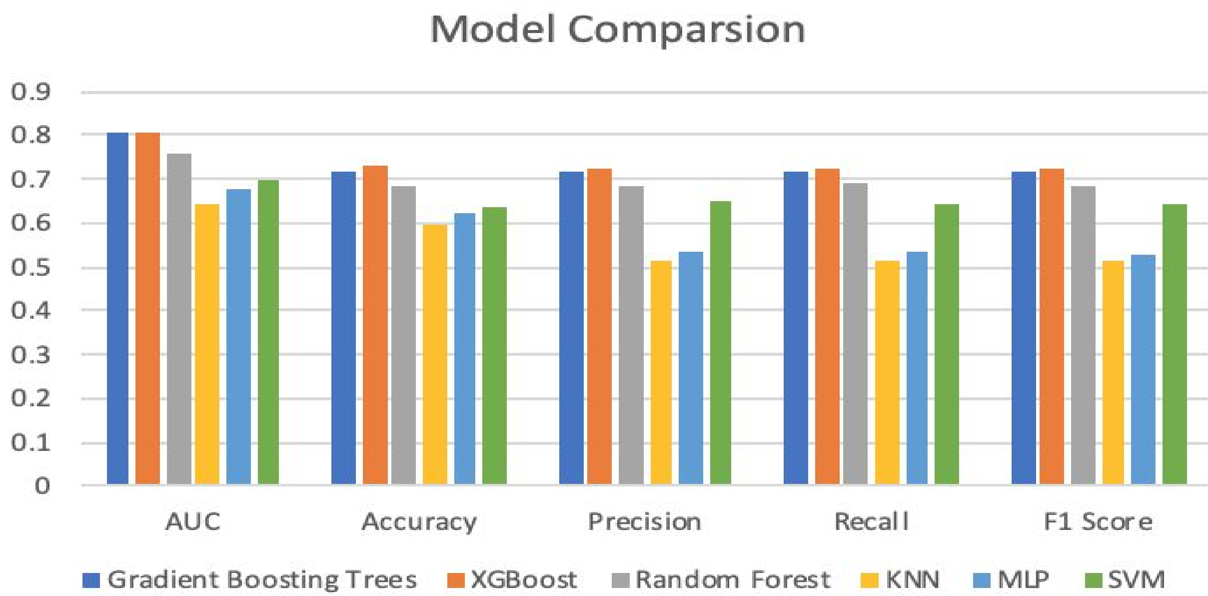

Figure 10: Comparision of all 6 Models

6 Discussion

It is immediately clear that the ensemble-based algorithms performed the best in predicting the outcome of the fights. The three Gradient Boosting Trees, XGBoost, and Random Forrest outperformed the others in all metrics. This is because these models have a higher level of complexity and combine the decisions of multiple decision trees. Similar to its results in various online data science competitions [5], XGBoost performed the best out of all the models. Further parameter tuning is needed to fully maximize the potential performance of the Random Forrest model. While the optimal number of trees was tested, it would be beneficial to go back and investigate the feature importance with Random Forrest. This could bring accuracy and AUC levels closer to the other two ensemble learning techniques.

As MLP is the most popular model for sports prediction [2], it is important to directly discuss the results since this was our second lowest performing model. While the difference in sports is a huge factor, our MLP performed 14.9% worse than that of Chang and Huang [3]. The parameter selection done involved the sequential tuning of the shape of the net [3]. In order to increase accuracy and AUC, sequential tuning should also be done to determine parameters like the number of epochs, learning rate, dropout layers, and activation functions. This process is the most challenging part of developing an accurate MLP model.

Comparatively, we performed better than others trying to predict UFC fight results. Our SVM achieved a 9.2% higher AUC than Singh's [4]. This was a result of both our parameter tuning and feature engineering. Additionally, our dataset being UFC focused, while Singh's used multiple fighting leagues in his prediction. In comparison to the work of Dwivedy and Yadav [1] our KNN, MLP, SVM, and Random Forest performed 9.1%, 12.3%, 3.4%, and 13.4%

respectively, using their metric of accuracy. Similar to their work our SVM and Random Forrest did perform better than KNN and MLP. Our use of the Sports Result Prediction CRISP-DM [2], in approaching the prediction problems, specifically our domain knowledge in application to feature engineering, likely led to our advantage. In addition to this, the parameter tuning done on our models helped to achieve the higher accuracy numbers.

As is, our findings are not directly applicable to sports betting. In order to bridge this gap some additional work must be done, as a model that predicts when to place a bet, must also factor in the betting odds. The term for this is cost-based learning and can be done by either factoring in the betting odds as a feature or combining the model prediction with the betting odd after the prediction. An incorrect decision to bet would cause a monetary loss and thus is worse than a decision not to bet on a fighter that ends up loosing. Thus the prediction value of the model, not the classification should be used in a cost-based learning model.

7 Conclusion

Predicting UFC fights provides an undeniably unique challenge in terms of dataset collection, feature engineering, modeling, and model evaluation. The techniques used in feature engineering are vastly different as the sport envolves two individuals, facing off, one on one. From the works of others and our experiment, it is clear to see some models perform better than others in sports prediction and more specifically UFC prediction. Ensemble-based, boosting algorithms have become commonplace in online data science competitions for a reason. They achieve the best results in predicting the outcome of a UFC fight. The ever-popular and widely used MLP model requires extensive parameter tuning in order to achieve the results of those previously mentioned.

When predicting the results of any sport using machine learning, it is important to follow a framework such as the one laid out by Bunker and Thabtah [2] in order to produce clear, reproducible results. It is advisable for others attempting a similar task to follow the steps: domain understanding, data understanding, data preparation, modeling, model evaluation, and deployment. This works well for the UFC, but could also be applied to other, lesser-known sports as well.

8 References

1. Yadav, N., Yadav, A., Bansal, J. C., Deep, K., & Kim, J. H. (2018). Harmony Search and Nature Inspired Optimization Algorithms: Theory and Applications, Ichsa 2018.

2. Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. Applied Computing and Informatics, 15(1), 27–33. doi: 10.1016/j.aci.2017.09.005

3. Huang, K.-Y., & Chang, W.-L. (2010). A neural network method for prediction of 2006 World Cup Football Game. The 2010 International Joint Conference on Neural Networks (IJCNN). doi: 10.1109/ijcnn.2010.5596458

4. Betting on UFC Fights – A Statistical Data Analysis. (2013, June 22). Retrieved December 6, 2019, from https://partyondata.com/2011/09/21/betting-on-ufc-fights-a-statistical-data-analysis/.

5. Morde, V. (2019, April 8). XGBoost Algorithm: Long May She Reign! Retrieved December 6, 2019, from https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d.