**BA305 Project Report**

****

**UFC Combat Strategies**

**Recommendation Model**

Team B3:

Avi Dhruva

Brynjolfur Brynjolfsson

Haoyang (Steven) Zhang

Mohammed Alshubaibi

Zhiyu (Eric) Chen

**1. Introduction**

**1.1 Dataset Selection**

For our project, we chose to develop a recommendation algorithm aimed at helping MMA coaches analyze matchups within specific weight classes. Initially, we considered various topics, including an analysis of weather impacts on Uber pricing and an exploration of the relationship between the expected time of arrival of Uber Eats deliveries and tipping patterns. However, we ultimately selected the MMA coach recommendation model. This decision was driven by our interest in the subject and the availability of a comprehensive dataset that offered insights into the nuances of MMA fighting strategies. This abundance of data availability and personal interest convinced us that focusing on MMA would yield the most engaging and informative outcomes for our project.

**1.2 Selected Dataset Overview**

Our dataset consists of 6,012 entries, each representing a UFC fight spanning from 1994 to 2021. This dataset includes 144 variables, encompassing both categorical and numerical data types. Categorical variables include fighter names, fight dates, and fight outcomes, while numerical variables capture detailed metrics such as the number of significant strikes attempted, takedowns, and control time measured in seconds. Notably, the numerical data did not require standardization as they consist of direct counts or time measurements, which naturally align well for analytical purposes.

Given the comprehensive nature of our data and its recent coverage up to 2021, we believe it provides a holistic foundation for our recommendation algorithm. Although MMA styles evolve, the extensive 27-year timeframe of our dataset ensures its relevance and utility in understanding current trends and dynamics in the sport. This reassures us that our dataset is exceptionally well-suited for developing insightful recommendations for MMA coaches.

**1.3 Project Goals & Plan**

We initially planned to develop a model that could predict the likelihood of winning a fight based on factors such as fighter weight, fighting styles, and strike rates. However, we realized that while this model was relatively straightforward to build, it offered limited insights to the UFC community. In order to find more valuable insights, we pivoted towards a more ambitious goal: creating a recommendation model that advises fighters on the optimal fighting style to maximize their chances of victory in each weight class.

This recommendation model has the goal of providing substantial value to UFC fighters and their coaches, providing strategic guidance on which aspects of fighting to emphasize during pre-fight preparations. Moreover, the model's predictions could also serve enthusiasts and bettors, offering them a tool to make more informed decisions and potentially enhancing their success in betting outcomes.

Ultimately, our project aims to equip coaches with a deeper understanding of the most effective fighting techniques, enabling them to fine-tune their training approaches and better prepare their fighters for success in the octagon.

**2. Data Processing - (**[**Dataset**](https://www.kaggle.com/datasets/rajeevw/ufcdata)**)**

**2.1 Removing Irrelevant Features**

After selecting the UFC dataset for our analysis, we initiated the data-cleaning process. Our first step involved eliminating certain weight classes, specifically the catchweight and women's featherweight classes, due to insufficient data points. The catchweight class contained 24 observations, and the women’s featherweight class had 8 observations. We determined that such limited observations would not yield significant insights, prompting their exclusion from our analysis.

In addition, there are certain variables that can be irrelevant to our modeling algorithm. For instance, the original dataset also includes statistics of their previous opponents. However, these features might not affect the result of the game as it is not related or determined by the fighter or coaching team.

**2.2 Cleaning Data Entries**

Once we had chosen the relevant variables and discarded the unnecessary ones, we proceeded with the data-cleaning process. Initially, we modified our dataset to address issues with data formatting, particularly where numerical values were mistakenly stored as strings, such as '1.1' instead of 1.1. To rectify this, we removed quotation marks from the data. Subsequently, we looked through the dataset for missing or NaN values, eliminating these rows to enhance data integrity. After doing so, we were left with 3788 observations for our dataset.

**2.3 Splitting fighters**

After cleaning our dataset, we prepared for the next stages of our data analysis. To organize our data, we divided it into two distinct groups according to fighter colors: one for the red fighter and another for the blue fighter. This approach ensured that the data for each fighter was meticulously separated. Statistics about the red fighter were exclusively grouped together, and likewise for the blue fighter. This categorization allowed us to organize observations for each side, enhancing the clarity and readability of the data.

Initially, the data for red and blue fighters were distributed across different columns. Through careful splitting and rearranging, we successfully consolidated all fighter data into rows under a single column. This reorganization not only simplified the data presentation but also facilitated a more intuitive analysis process. This also left us with our fighters dataset which included 1334 observations, which we would use for clustering.

**2.4 Addressing Old Records**

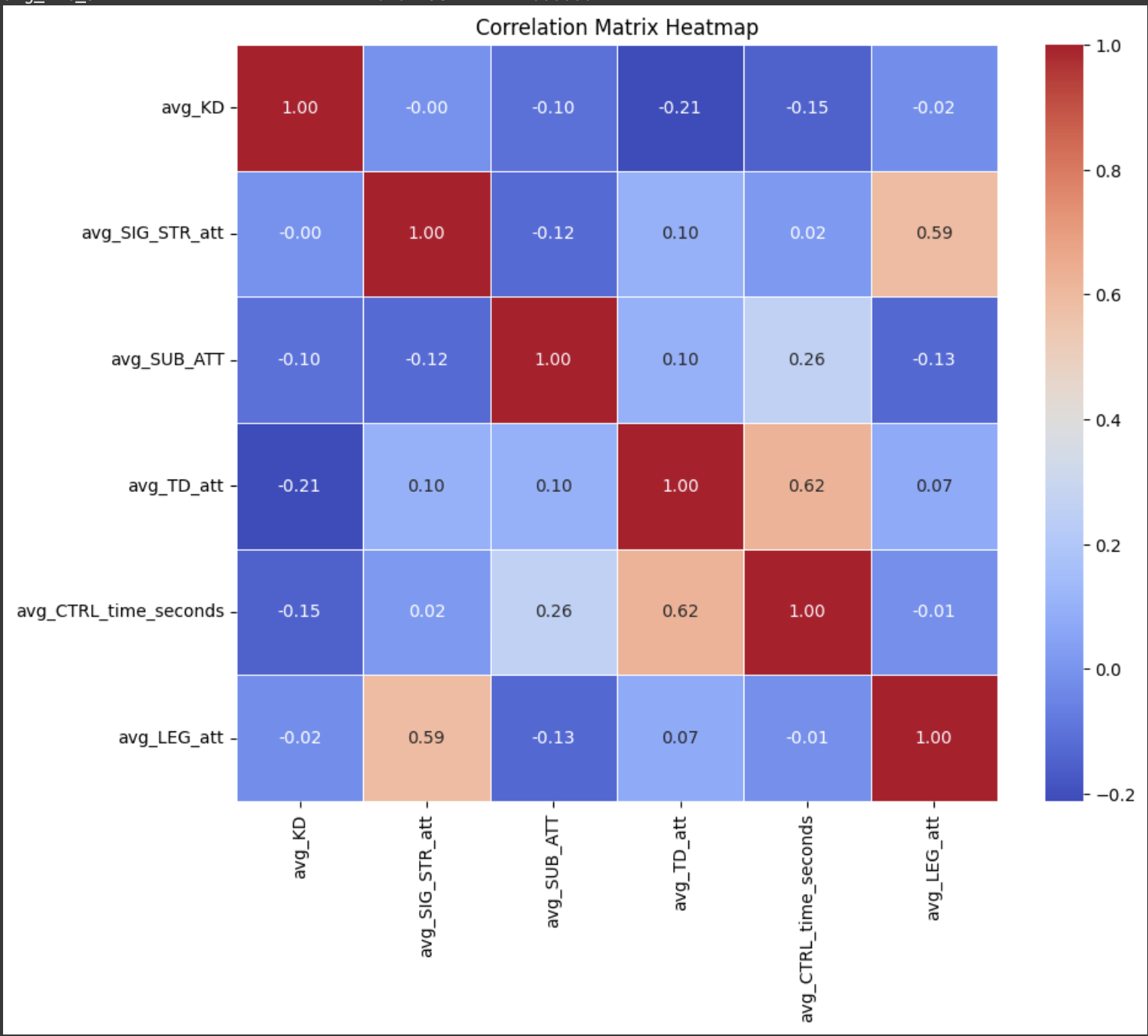
After reorganizing our dataset to focus on individual fighters rather than specific fights, we encountered the issue of duplicate entries, as many fighters appear multiple times due to participating in numerous bouts. This led us to contemplate the best approach for selecting representative data points for each fighter to use in our clustering analysis.

We debated between using either the first recorded fight or the most recent fight for each fighter. After considering our options, we concluded that these options might not significantly impact our analysis. This is because fighters in the UFC typically train in their chosen fighting styles throughout their fighting career, and these fundamental styles do not drastically change over time. However, it is important to note that while their core style remains consistent, fighters often enhance their repertoire by integrating techniques from other fighting styles into their existing skill set.

**3. Methodology**

**3.1: How to categorize?**

As we moved forward with our analysis, we were tasked with selecting which variables to use for clustering from a list of over 144 features. Our goal was to choose variables that could distinctly represent the two predominant fighting styles in MMA: grappling and striking. Leveraging our prior knowledge of MMA, we identified several key indicators for each style. For strikers, we selected significant strikes attempted, leg kicks attempted, and knockdowns attempted; for grapplers, we chose submissions attempted, takedowns attempted, and control time in seconds. To ensure that our assumptions about these variables being indicative of their respective fighting styles were correct, we conducted a correlation analysis. We ran a correlation matrix for these six variables and confirmed that indeed, the variables within each group were significantly correlated, affirming their relevance for distinguishing between the different fighting techniques in our clustering model.

Figure 1: Heat map for correlation matrix of clustering variables

**3.2: Clustering**

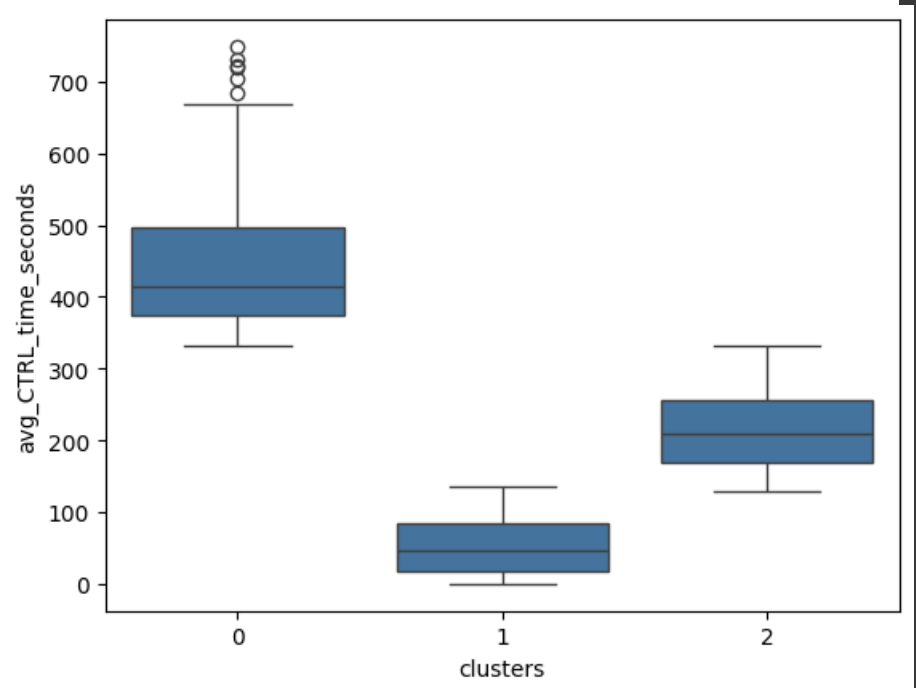
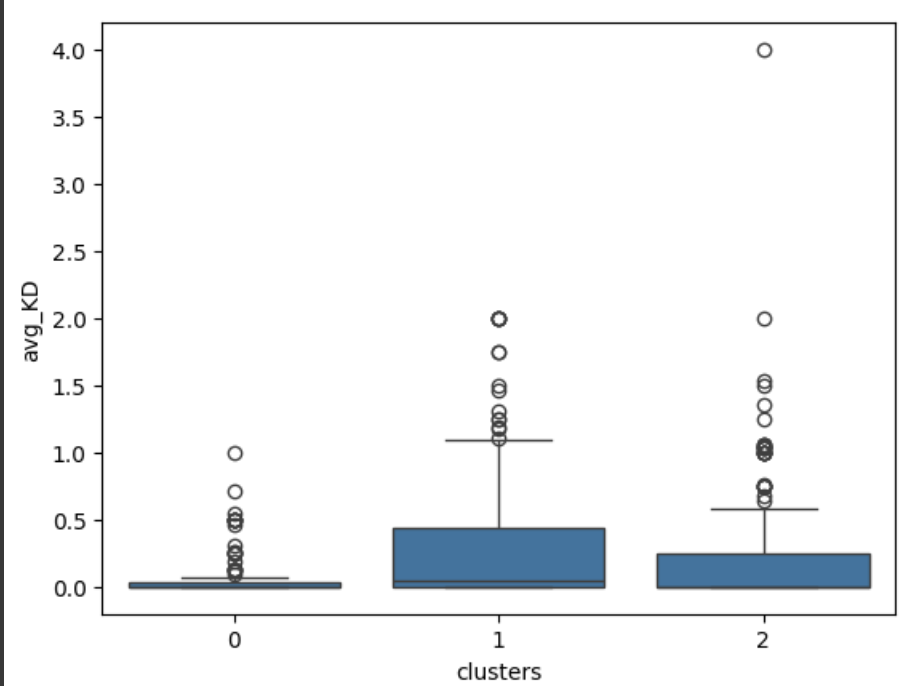
After selecting the relevant variables, we conducted a cluster analysis, which were segmented into three distinct groups: grapplers, strikers, and balanced fighters. This analysis aimed to categorize each fighter into a specific group with precision, utilizing the chosen variables to ensure accurate classifications.

**3.3: Implications of clustering and assigning fighting styles**

Adopting this approach had certain limitations. The restricted number of clusters meant that some fighters had to be categorized into broader groups that might not perfectly align with their precise fighting styles. Consequently, the options for categorizing fighters into the specific groups of grappler, striker, or balanced were somewhat limited, potentially oversimplifying the diversity of their techniques but still representative of the sport.

**3.4: Visualizing our Clusters**

To visualize our clusters, we chose boxplots for their clarity and simplicity. We created a boxplot for each of the six variables used to cluster the fighters. Our analysis identified three distinct clusters: Cluster 0, predominantly strikers with high striking stats and low grappling stats; Cluster 1, primarily grapplers with the opposite pattern; and Cluster 2, which showed a balanced mix of both attributes. The accompanying figures highlight these differences, showing "Average Knockdowns" for strikers and "Average Control Time" for grapplers, effectively demonstrating the clusters' characteristics.



Figures 2 & 3: Visualizing our clusters using boxplots

**3.5: Creating our model**

To guide coaches in developing winning strategies, we aim to advise the optimal fighting strategy that the athletes can utilize to maximize the likelihood of winning. Originally, the accuracy score was considered as the performance measure that evaluates our models. However, as we developed our models and tested their performance, we discovered that the accuracy was low across various models due to the high complexity and randomness due to the nature of the sport. Our model only achieves an accuracy of around 65%, which did not satisfy our internal criteria. Given that prediction errors are inevitable, it's essential to select a new performance measure that offers a more accurate assessment.

In practical scenarios, such as being part of a fighter's coaching team, it's crucial to minimize the risk of losing when predicting a win, which affects our credibility and is measured by precision. Simultaneously, we aim to reduce instances where we unexpectedly win after predicting a loss, reflecting recall. While surprising wins may seem beneficial, they indicate that fighters and coaches are expending unnecessary resources on a contest they are likely to win. Therefore, the F1 score, which balances both precision and recall, is the most suitable performance measure for our needs.

We need to train models specifically for each weight class rather than using the entire dataset, as fighters in different categories likely exhibit distinct characteristics that influence their chances of winning.

**3.5.1: Decision tree**

After selecting the F1 score as the performance measure and segmenting the dataset by weight class, we trained a decision tree model for each class intending to optimize performance. The decision tree approach was chosen because it allows coaches to easily understand which strategies could increase their chances of winning, thereby facilitating the development of more effective training plans. To determine the optimal tree depth, we implemented a loop that tested various depths until the model became too complex for practical interpretation. We then recorded the tree depth that yielded the best performance on the testing dataset for each respective dataset.

Using 4 known weight classes as examples, we found that decision trees with shallow depths and lower complexity perform better. These simpler models are also easier to interpret, enabling coaches to identify the best winning strategies.

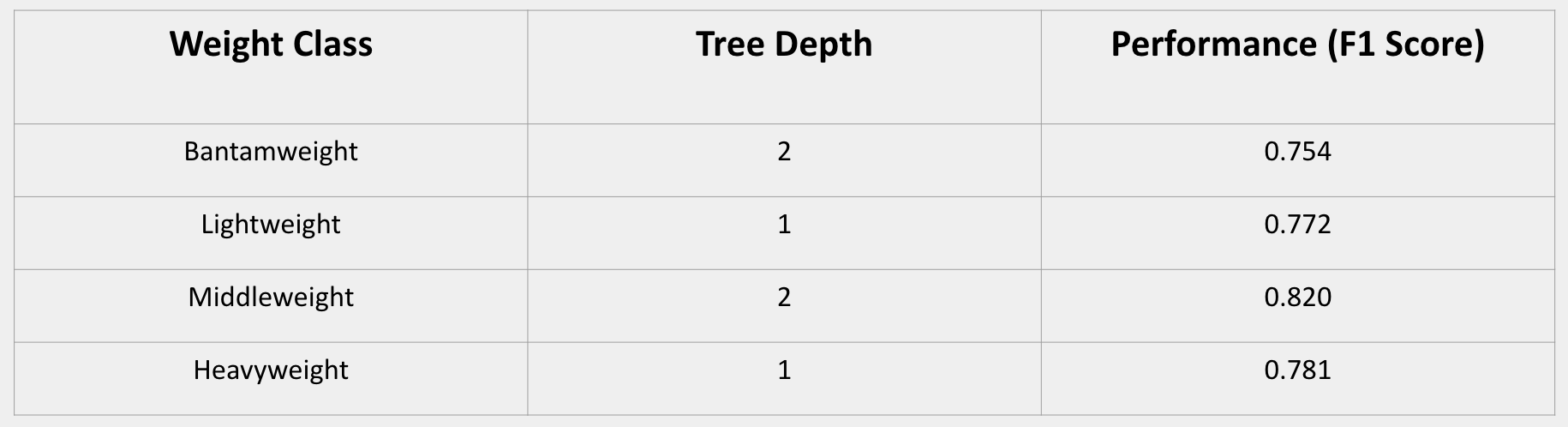
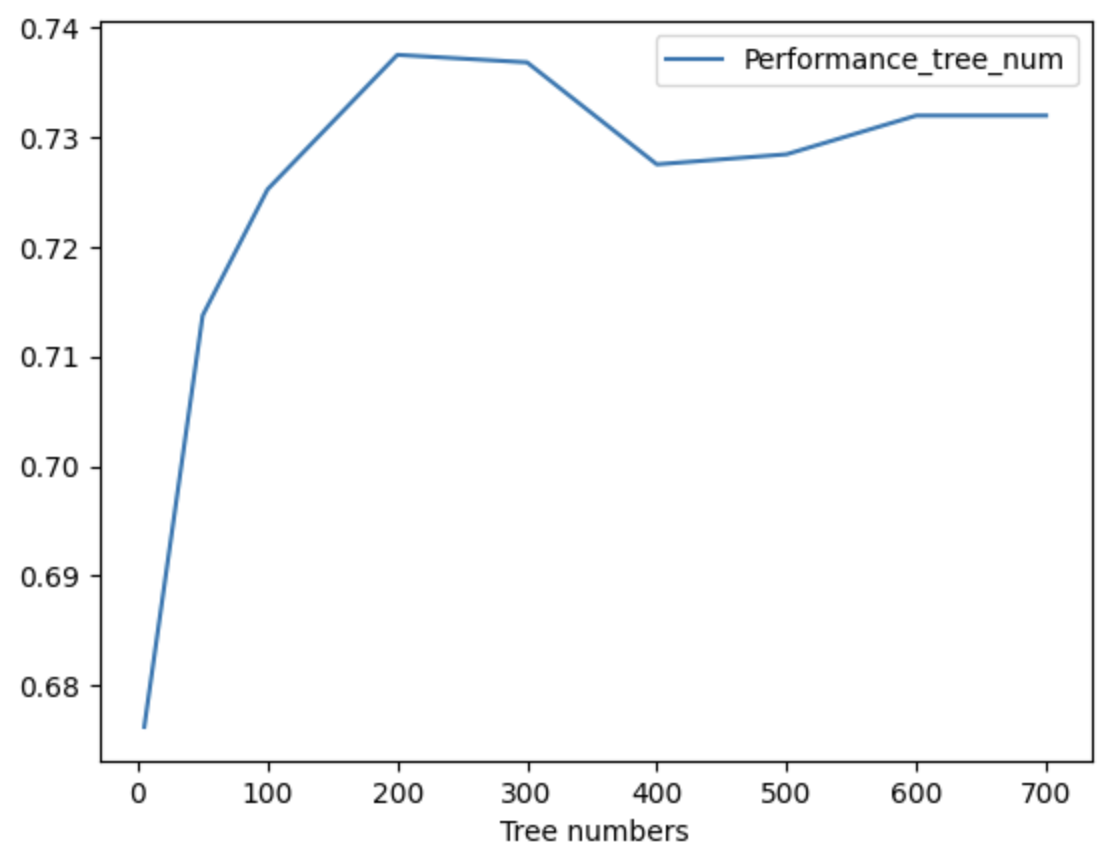
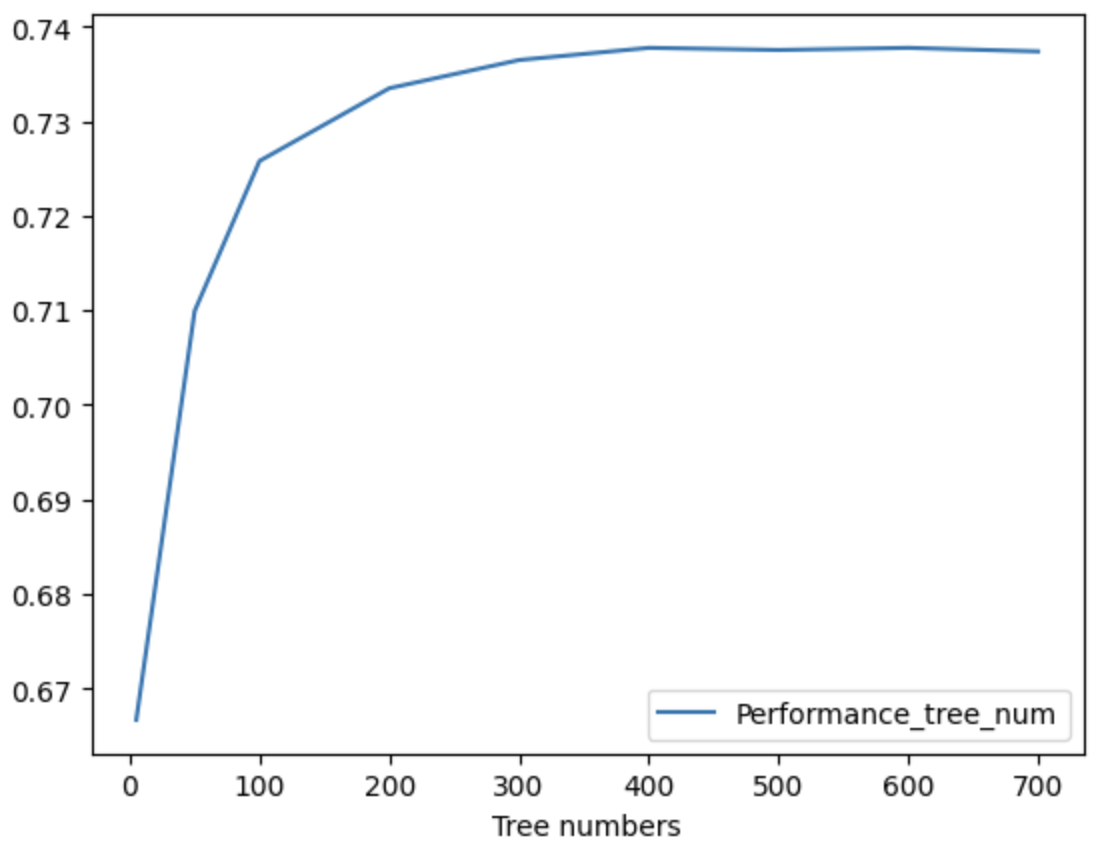


Table 1: Results from our original decision tree

**3.5.2: Random forest**

After developing our decision tree model, we explored whether performance could be enhanced with alternative algorithms. We opted to experiment with a random forest model. Initially, we varied the random forest parameters to find the optimal range for our needs. We tested different numbers of trees and the maximum number of features in each tree to fine-tune our model's accuracy.

As indicated by the two graphs below, which are the patterns of how performance changes when the number of trees used increases for two different train-test splits, it seems like the optimal test set performance is reached between 200 and 400.



Figures 4 & 5: Random forest model performance based on number of trees

Regarding the number of features in each tree, we observed that including more features tends to degrade performance, possibly introducing noise that the model interprets as significant. Therefore, we chose to use a limited number of features for the final random forest model to enhance its effectiveness.

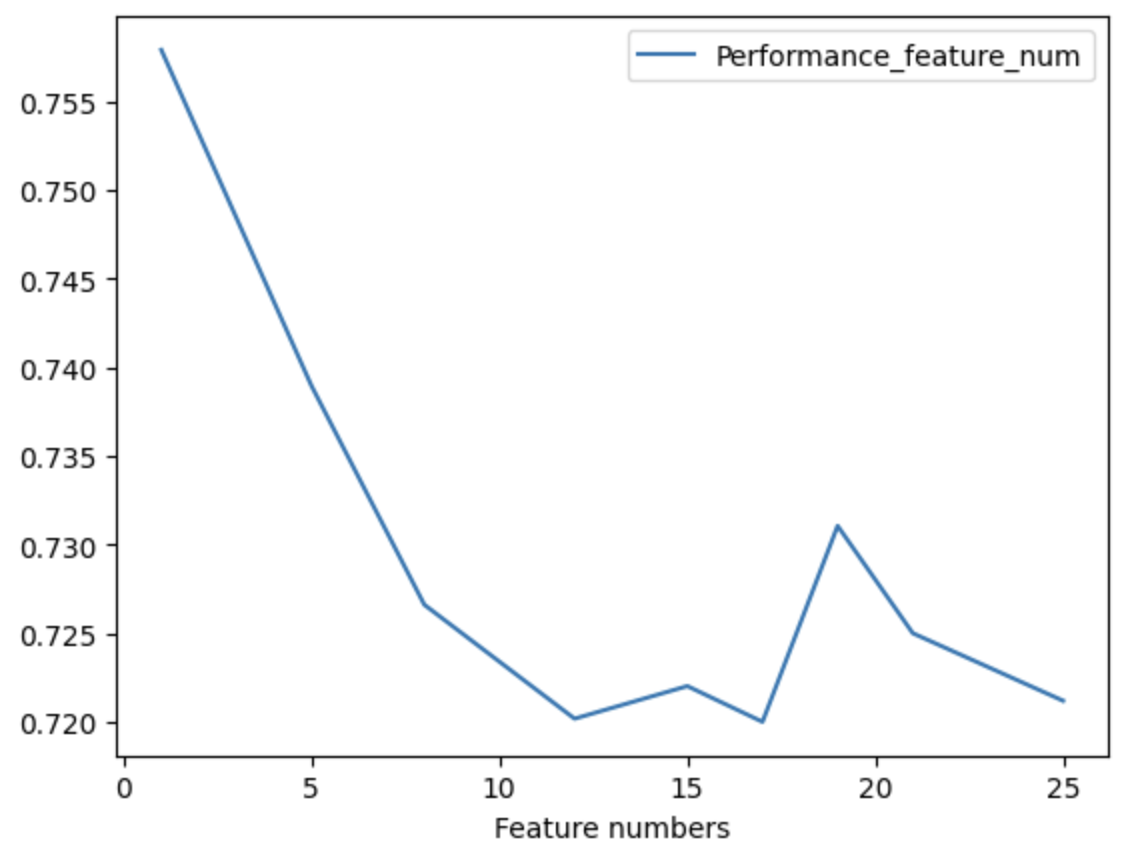


Figure 6: Random forest model performance based on the number of variables used

Once the parameter ranges were established, we trained the final random forest model using a grid search method. This involved creating loops to iterate through every possible combination of parameter values. Below, we present the results and performances of the decision tree algorithm for four common weight classes.

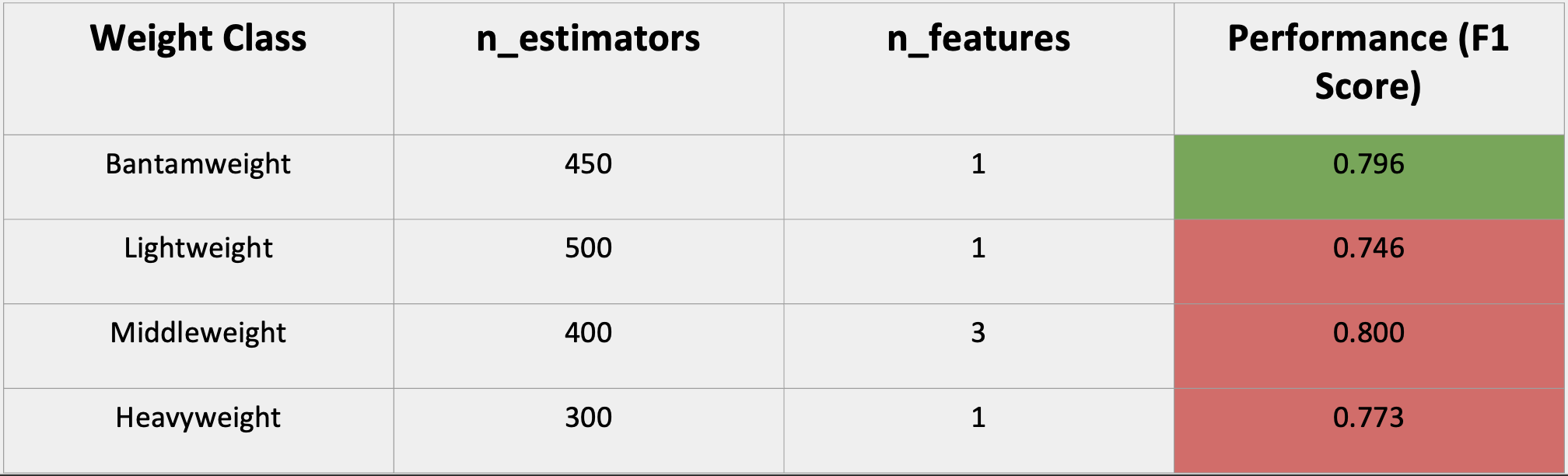


Table 2: Performance of the random forest model for different weight classes based on recommended features and estimators

The color green highlights the improvement in performance from the original single decision tree whereas the color red highlights the decrease in performance. As the table shows, a random forest model does not improve the model performance as we expected.

**3.5.3: KNN**

In order to investigate additional modeling possibilities, we implemented the KNN (k-nearest neighbors) algorithm. The premise is that if two fighters with distinct styles and skills compete multiple times without altering their tactics, the outcomes should be consistent and predictable. By analyzing fights with similar statistics and conditions, we aim to enhance the accuracy of our predictions. After testing various numbers of neighbors for prediction, we present the results for the common weight classes below.

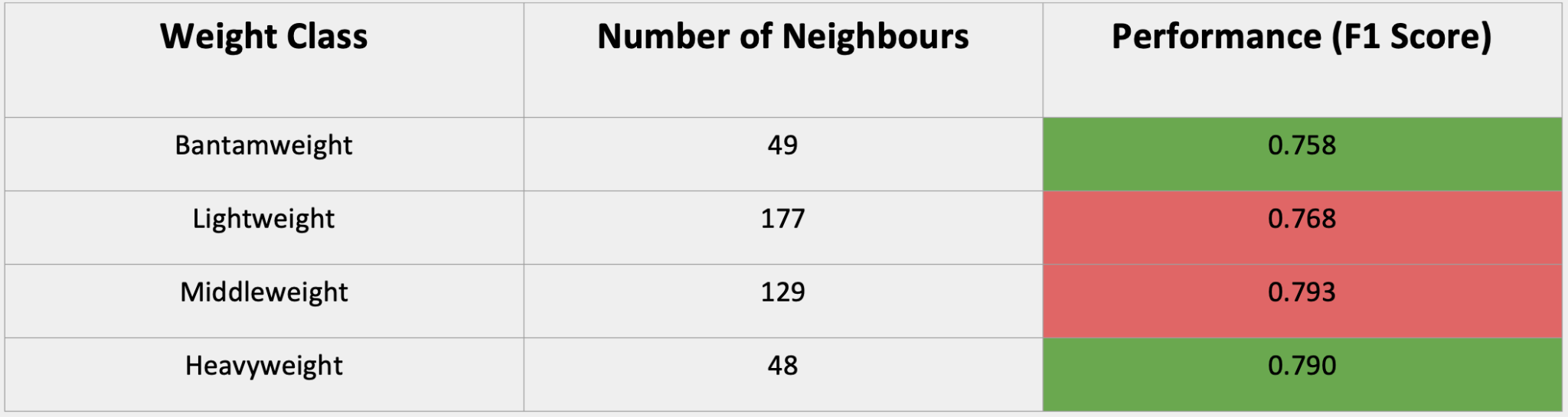


Table 3: Compares the performance of the K-NN model to our decision tree model for different weight classes

As indicated by the performance score of these models, we observe improvements in performance for Bantamweight and Heavyweight, however, performance for Lightweight and Middleweight decreased.

**4. Potential for improvements and alternative use cases**

**4.1 Potential for improvement within our dataset**

After reviewing our results, we identified potential improvements based on data availability. First, we could enhance our fight outcome predictions by increasing our dataset size. This could be achieved either by extending our data collection period, assuming UFC consistently publishes similar statistics, or by including data from other organizations like One FC, Bellator, or Eagle FC. This approach would offer insights into more recent trends, given the rapid evolution in martial arts techniques.

Additionally, we noted a bias in our dataset, with red fighters winning more often than blue, which could skew results. We also observed an imbalance in our dataset, favoring striking over grappling variables. Addressing these issues would involve sourcing datasets with balanced representations of both fighting styles, ensuring a more comprehensive analysis.

**4.2 Alternative use cases**

Our model has potential applications beyond coaching, including sports betting and video games. For sports bettors, the model can enhance their ability to accurately predict fight outcomes, potentially increasing their success rate. Betting companies could use the algorithm to refine the odds they offer for each fight, improving the betting experience.

In the realm of video games, our model could be integrated into games like the UFC video game to render AI-opponent behavior more realistic based on fighter statistics. Additionally, it could be used in a management simulation game where players act as managers of fighters, with the outcomes based on the probabilities calculated by our algorithm.

Another interesting application could be a website where users input their statistics and those of their friends to receive fight outcome predictions, adding a fun and interactive element to social gatherings or online forums.

**5. Conclusion**

**5.1 What We Learned**

Interestingly, the most significant lesson from our project was recognizing the limitations of our predictive models. We found that accurately predicting outcomes in sports is particularly challenging due to the high impact of immeasurable factors, such as a fighter's mood and well-being on the day of the fight.

Additionally, we discovered that simply adding more variables to our models often generated noise rather than improving predictions. This occurs because an increased number of variables can introduce more randomness into the data, negatively affecting model performance. Moreover, more complex models tend to overfit, becoming too tailored to the training data and less effective on new, unseen data. This instability highlights the importance of balancing model complexity with practical applicability in sports-related projects.

**5.2 Overcoming Obstacles**

During our project, we faced several challenges that demanded innovative solutions. One major issue was choosing the appropriate measure of accuracy. Initially, using accuracy as a metric allowed our model to overlook increases in false positives and negatives, provided there was an increase in true positives. This was problematic since our goal was to minimize both false positives and negatives rather than solely maximizing true positives. Through research, we discovered that the F1 score, which incorporates both precision and recall, effectively addresses this issue, so we opted to use it.

Another challenge was predicting and providing recommendations for matchups within the same fighting class. Relying solely on fighting style proved inadequate for coaches dealing with strikers facing other strikers. To resolve this, we included additional variables not used in our initial clustering algorithm in our subsequent models. This approach allows coaches to focus on refining specific techniques such as head strikes or clinches, offering more targeted and effective training strategies beyond general fighting style.

**5.3 Closing Thoughts**

After finalizing our project, we feel like we have been successful in achieving our goal. We have 3 models that can predict the outcome of fights with, relatively, decent accuracy. Although we originally expected to get a higher accuracy, the process of going through this project has, as we previously mentioned, taught us that other immeasurable factors can contribute to the outcome of models, especially in the context of sports.

We feel like going through this project has helped us further understand the course concepts and how we will be applying them in the future. Overall we are very happy with what we have achieved and learned throughout this project.

Appendix:

| **Feature** | **Description** |
| --- | --- |
| fighter | Weight class for each specific fighter (example: heavy-weight) |
| avg\_KD | Average number of knockouts / per fighter |
| avg\_SIG\_STR\_att | Average number of significant strikes attempted / per fighter |
| avg\_SUB\_ATT | Average number of submission attempts / per fighter |
| avg\_TD\_att | Average number of takedowns attempted / per fighter |
| avg\_CTRL\_time\_seconds | Average number of seconds spent on ground, “in control” |
| avg\_LEG\_att | Average number of legkicks attempted |

Table 4: Features of the fighters dataset for clustering

| **Feature** | **Description** |
| --- | --- |
| R\_fighter | Red fighter name |
| B\_fighter | Blue fighter name |
| Winner | Winner of the fight (blue or red) |
| weight\_class | Weight class of the fighter |
| R\_fighter\_style | Red fighters style |
| B\_fighter\_Style | Blue fighters style |
| avg\_kd\_diff | Difference in average knockdowns between fighters |
| avgsigstrpct\_diff | Difference in average significant strike percentage between fighters |
| avgtdpct\_diff | Difference in average takedown percentage between fighters |
| avgsubatt\_diff | Difference in average submissions attempted between fighters |
| avgrev\_diff | Difference in average reversals between fighters |
| avgsigstratt\_diff | Difference in average significant strikes attempted between fighters |
| avgsigstrlanded\_diff | Difference in average significant strikes landed between fighters |
| avgtotalstratt\_diff | Difference in average total strikes attempted between fighters |
| avgtotalstrlanded\_diff | Difference in average total strikes landed between fighters |
| avgheadlanded\_diff | Difference in average head strikes landed between fighters |
| avgtdlanded\_diff | Difference in average takedowns landed between fighters |
| avgbodylanded\_diff | Difference in average body strikes landed between fighters |
| avglegatt\_diff | Difference in average leg strikes attempted between fighters |
| avgleglanded\_diff | Difference in average leg strikes landed between fighters |
| avgdistanceatt\_diff | Difference in average strikes attempted at a distance between fighters |
| avgdistancelanded\_diff | Difference in average strikes landed at a distance between fighters |
| avgclicnhatt\_diff | Difference in average significant strikes attempted in a clinch between fighters |
| avgclinchlanded\_diff | Difference in average significant strikes landed in a clinch between fighters |
| avggroundatt\_diff | Difference in average strikes attempted on the ground between fighters |
| avggroundlanded\_diff | Difference in average strikes landed on the ground between fighters |
| avgctrltime\_diff | Difference in average control time in seconds between fighters |

Table 5: Explanation of the features in our model dataset