UFC Betting Prediction Engine

INFO 7390 – Advances in Data Science/Architecture

Siddhartha Vale & Karan Kathayat

Abstract

In this project, we look at the competitive accuracies decided by betting companies like Bet365, Odd shark and Bovada. They have a certain system of measuring the accuracy of wins, unfortunately for those who have bet of fights and lost a lot of money know that their accuracies are quite low and have a great risk factor. As the brief of this project, we set out to find a dataset that we could use to measure the accuracies and improve them by using complex neural network algorithms. After extensive research and data exploration of a sample dataset, we saw that these established betting companies have a rather low 50 – 55% accuracy. As our mission, we aimed to find a dataset and use algorithms that would surpass those odds.

Table of Contents

Introduction

The main concept behind the project is creating a more accurate system for predicting fight wins to minimize risks for future gamblers. Betting companies use a simple method to predict a win which leads to low accuracy in the making of a decision. The first dataset we sourced from Kaggle had many errors, and is a relatively old dataset, with barely any edits made on it since its creation. Our understanding for research done is that betting companies use datasets similar to this for their calculations. After applying 7 models (Logistic Regression, Random Forest, SVM, Decision Tree, Perceptron, Naïve Bayes & KNeighbor) we notice that the accuracies attained from these models are too low to make wise decisions on the victor of a fight. For this we used a different data set, one with no missing values, a new set of data inputs and one with less incoherent noise.

With the new dataset we applied three models, all being forms of supervised learning and with one which is a neural network model. All algorithms showed a great improvement on the accuracies found from the previous data set. Two of the models tried and tested were similar to those tried in dataset 1 (Logistic Regression & Random Forrest) to see the exact difference in accuracy from the first and second dataset, while the third was Multi-layer Perceptron, a proven model in improving accuracy measurements for any classification or regression data.

Research

Accuracy measuring plays an important part of our daily lives, everything from the food we eat, car we buy or the bet we gamble on has had a tremendous amount of research and multiple market surveys done to get them, in order to make them worth it in the consumers eyes, or profitable when it comes to gambling. Sports bets are decided by companies after measuring the skill of all the athletes or their number of wins and losses. Some companies look only at the number of wins or loses a player has had, while others look at the overall player ability.

With the first dataset we used from Kaggle [1], we noticed some work done of measuring the accuracies for the fighters and their fights. The code written by Bart Gortat, was well described with many models being used, however with a lack of data exploration, data cleaning and was amounts of incoherent data, it was difficult to see if this would be a suitable dataset for us to use. So we used this opportunity to tweak the existing machine learning algorithm and applying it to another dataset based on a real-world context.

Gortat used many supervised learning models that best measured the accuracy of the data that was available. Here are some examples of the accuracies seen through his work.

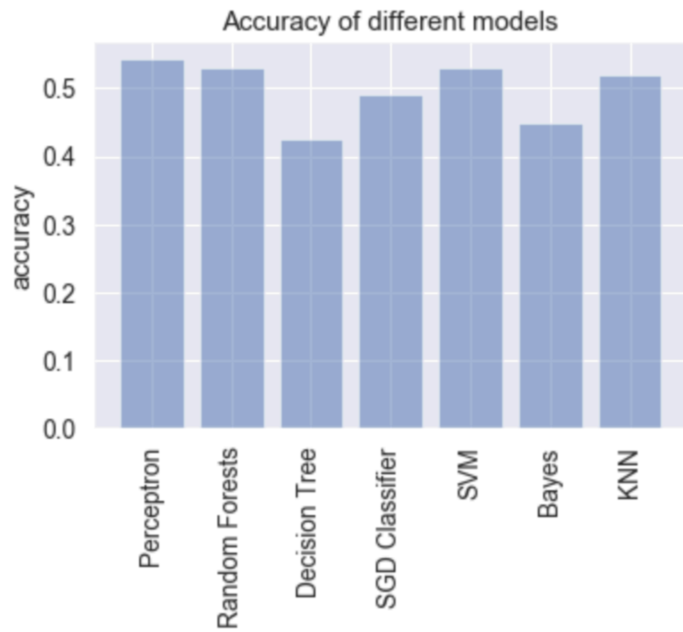


FIGURE 1: ACCURACY COMPARISON OF ALL 7 MODELS FOR KAGGLE DATASET

Figure 1 shows us that the highest accuracy for this data set without any changes made to it is 53% to the perceptron model. Although these accuracies change after the required steps are taken to minimize outliers, non-numeric, and incoherent data.

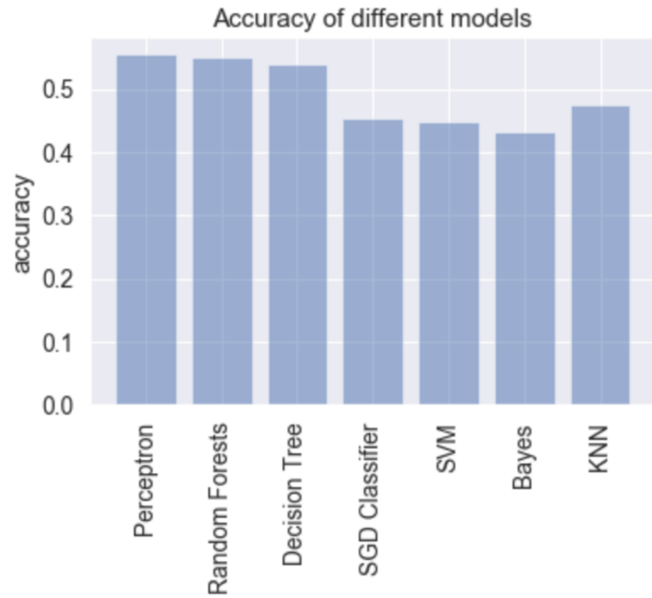


FIGURE 2: ACCURACY MEASUREMENT AFTER CHANGES MADE IN DATASET

Figure 2, much like figure 1 leaves perceptron in the same place with the highest accuracy, while giving Random Forest and Decision Tree a higher accuracy than SVM after the changes are made.

After carrying out some of our own EDA, we noticed that Rounds 4 & 5 were not as important in the analysis of accuracy as most fights did not have a round four or five, and those that did, had no effect on the decision.

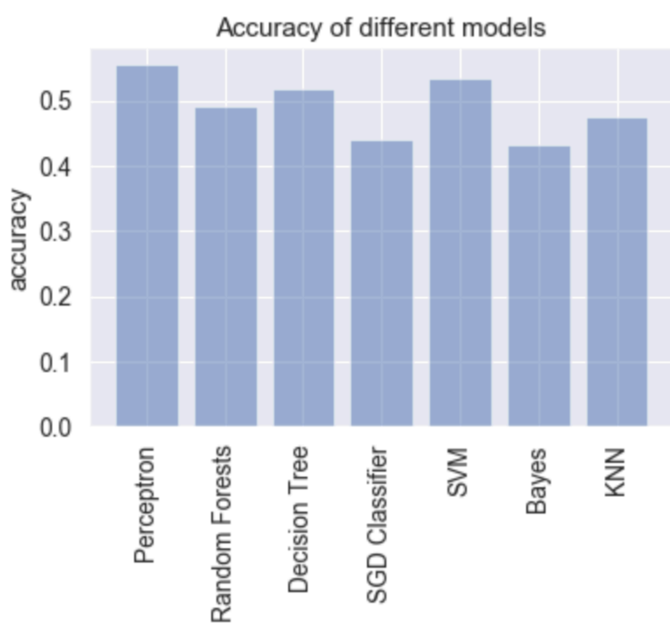


FIGURE 3: ACCURACY ANALYSIS AFTER DROPPING ROUNDS 4 & 5

After dropping rounds four & five we notice that we get a similar model to what we noticed from the first one, again guaranteeing our initial hypothesis of rounds 4 & 5 not affecting the results of the whole model. In this scenario the Perceptron increases in accuracy, but only by one percent which is not enough for us to choose this dataset.

```
(count    1235.000000
 mean        0.958704
 std         1.637015
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         1.000000
 max        17.000000
 Name: B__Round1_Grappl
 count    1235.000000
 mean        1.217004
 std         1.676862
 min         0.000000
 25%         0.000000
 50%         1.000000
 75%         2.000000
 max        17.000000
 Name: R__Round1_Grappl
```

FIGURE 6: GAPPLING TAKEDOWNS ROUND 1

```
(count    1235.000000
 mean        0.710121
 std         1.293364
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         1.000000
 max         8.000000
 Name: B__Round2_Grapp
 count    1235.000000
 mean        0.882591
 std         1.396330
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         1.000000
 max         9.000000
 Name: R__Round2_Grapp
```

FIGURE 5: GAPPLING TAKEDOWNS ROUND

```
(count    1235.000000
 mean        0.675304
 std         1.284112
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         1.000000
 max        13.000000
 Name: B__Round3_Grap
 count    1235.000000
 mean        0.806478
 std         1.322645
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         1.000000
 max        11.000000
 Name: R__Round3_Grap
```

FIGURE 4: GAPPLING TAKEDOWNS ROUND 3

In the grappling we see that in round one the grapples are highest for red when compared to round three. Also, blue has a lower grapple mean than red, which show that red is more skilled in grappling, because they are more experienced fighters.

```
(count    1235.000000
 mean        6.037247
 std         8.050906
 min         0.000000
 25%         0.000000
 50%         3.000000
 75%         9.000000
 max        56.000000
 Name: B__Round1_Strik
 count    1235.000000
 mean        7.625101
 std         8.808029
 min         0.000000
 25%         1.000000
 50%         5.000000
 75%        11.000000
 max        61.000000
 Name: R__Round1_Strik
```

FIGURE 9: BODY STRIKING ROUND 1

```
(count    1235.000000
 mean        3.450202
 std         5.389504
 min         0.000000
 25%         0.000000
 50%         1.000000
 75%         5.000000
 max        47.000000
 Name: B__Round3_Strik
 count    1235.000000
 mean        4.198381
 std         5.903080
 min         0.000000
 25%         0.000000
 50%         2.000000
 75%         6.000000
 max        48.000000
 Name: R__Round3_Strik
```

FIGURE 8: BODY STRIKING ROUND 2

```
(count    1235.000000
 mean        3.450202
 std         5.389504
 min         0.000000
 25%         0.000000
 50%         1.000000
 75%         5.000000
 max        47.000000
 Name: B__Round3_Strik
 count    1235.000000
 mean        4.198381
 std         5.903080
 min         0.000000
 25%         0.000000
 50%         2.000000
 75%         6.000000
 max        48.000000
 Name: R__Round3_Strik
```

FIGURE 7: BODY STRIKING ROUND 3

In the body striking we see that in round one the strikes are highest for red when compared to round three. Also, blue has a lower strike rate than red, which show that red is more skilled in body striking, because they are more experienced fighters.

```
(count    1235.000000
 mean        9.783806
 std        15.640389
 min         0.000000
 25%         0.000000
 50%         4.000000
 75%        13.000000
 max       134.000000
 Name: B__Round1_Strik
 count    1235.000000
 mean       11.572470
 std        14.287913
 min         0.000000
 25%         1.000000
 50%         6.000000
 75%        17.000000
 max       103.000000
 Name: R__Round1_Strik
```

FIGURE 12: STRIKE CLINCH ROUND 1

```
(count    1235.000000
 mean        6.757895
 std        11.100346
 min         0.000000
 25%         0.000000
 50%         2.000000
 75%         9.000000
 max        76.000000
 Name: B__Round2_Strik
 count    1235.000000
 mean        8.479352
 std        12.359690
 min         0.000000
 25%         0.000000
 50%         3.000000
 75%        12.000000
 max        86.000000
 Name: R__Round2_Strik
```

FIGURE 11:: STRIKE CLINCH ROUND 2

```
(count    1235.000000
 mean        5.744130
 std        10.326719
 min         0.000000
 25%         0.000000
 50%         1.000000
 75%         7.000000
 max        79.000000
 Name: B__Round3_Strik
 count    1235.000000
 mean        6.435628
 std        10.181323
 min         0.000000
 25%         0.000000
 50%         2.000000
 75%         9.000000
 max        71.000000
 Name: R__Round3_Strik
```

FIGURE 10: STRIKE CLINCH ROUND 3

In the clinch striking we see that in round one the strikes are highest for red when compared to round three. Also, blue has a lower strike rate than red, which show that red is more skilled in clinch striking, because they are more experienced fighters.

```
(count    1235.000000
 mean        0.162753
 std         0.656773
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         0.000000
 max         7.000000
 Name: B__Round1_Strik
 count    1235.000000
 mean        0.264777
 std         0.919982
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         0.000000
 max        10.000000
 Name: R__Round1_Strik
```

FIGURE 15: HEAD KICKS ROUND 1

```
(count    1235.000000
 mean        0.133603
 std         0.549757
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         0.000000
 max         6.000000
 Name: B__Round2_Strik
 count    1235.000000
 mean        0.228340
 std         0.756178
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         0.000000
 max         6.000000
 Name: R__Round2_Strik
```

FIGURE 14: HEAD KICKS ROUND 2

```
(count    1235.000000
 mean        0.131984
 std         0.474202
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         0.000000
 max         4.000000
 Name: B__Round3_Strik
 count    1235.000000
 mean        0.150607
 std         0.510813
 min         0.000000
 25%         0.000000
 50%         0.000000
 75%         0.000000
 max         4.000000
 Name: R__Round3_Strik
```

FIGURE 13: HEAD KICKS ROUND 3

In the head kicks we see that in round one the kicks are highest for red when compared to round three. Also, blue has a lower kicks rate than red, which show that red is more skilled in head kicks, because they are more experienced fighters.

```
(count    1235.000000
 mean       11.036437
 std        14.199133
 min         0.000000
 25%         0.000000
 50%         6.000000
 75%        16.000000
 max        97.000000
 Name: B__Round1_Strik
 count    1235.000000
 mean       13.676113
 std        14.959222
 min         0.000000
 25%         3.000000
 50%         9.000000
 75%        19.000000
 max       131.000000
 Name: R__Round1_Strik
```

FIGURE 18: HEAD STRIKES ROUND 1

```
(count    1235.000000
 mean        9.978947
 std        14.187773
 min         0.000000
 25%         0.000000
 50%         4.000000
 75%        15.000000
 max        91.000000
 Name: B__Round2_Strik
 count    1235.000000
 mean       12.562753
 std        15.641079
 min         0.000000
 25%         1.000000
 50%         7.000000
 75%        18.000000
 max       103.000000
 Name: R__Round2_Strik
```

FIGURE 17: HEAD STRIKES ROUND 2

```
(count    1235.000000
 mean        9.043725
 std        14.111895
 min         0.000000
 25%         0.000000
 50%         2.000000
 75%        12.000000
 max        89.000000
 Name: B__Round3_Strik
 count    1235.000000
 mean       10.564372
 std        14.926878
 min         0.000000
 25%         0.000000
 50%         5.000000
 75%        15.000000
 max       120.000000
 Name: R__Round3_Strik
```

FIGURE 16: HEAD STRIKES ROUND 3

In the head strikes we see that in round one the strikes are highest for red when compared to round three. Also, blue has a lower strike rate than red, which show that red is more skilled in head strikes, because they are more experienced fighters.

Methodology

1. Finding current work and accuracies to base a project on
2. EDA and Data analysis of comparison dataset
3. Finding a new dataset
4. New data cleaning & EDA
5. Data Visualization
6. Model Implementation
7. Summary & Comparisons

Results

The first process was finding a dataset that matches our needs and is relatively new. On a sport such as UFC, this is quite difficult as the sport is relatively new with its introduction being made in November 1993 [1]. As compared to another famous betting sport, horse racing which begun in 1665 and had betting for it start in the early 1800's [2].

The dataset we found, is one that has zero missing inputs and is a dataset that takes what was perfect about the Kaggle dataset and eliminates everything that is of no importance. With a relatively new dataset, we begun with the Exploratory Data Analysis (EDA).

Exploratory Data Analysis

We began by explaining what the columns mean, being a sport with many complex naming conventions the data came with short but difficult to understand abbreviations, this was the first step to understanding the data we were working with.

Acronyms Explained

- Label - This is the response variable. Either Favourite or Underdog will win
- REACH - Fighter's reach
- SLPM - Significant Strikes Landed per Minute
- STRA. - Significant Striking Accuracy
- SAPM - Significant Strikes Absorbed per Minute
- STRD - Significant Strike Defence (the % of opponents strikes that did not land)
- TD - Average Takedowns Landed per 15 minutes
- TDA - Takedown Accuracy
- TDD - Takedown Defense (the % of opponents TD attempts that did not land)
- SUBA - Average Submissions Attempted per 15 minutes
- Odds - Fighter's decimal odds spread for that specific matchup

FIGURE 19: DATASET COLUMN ABBREVIATIONS EXPLAINED

With the dataset being checked for any missing data points and being found as 100% error free with no missing values, we moved to the next step of Data visualization, where our inferences were made.

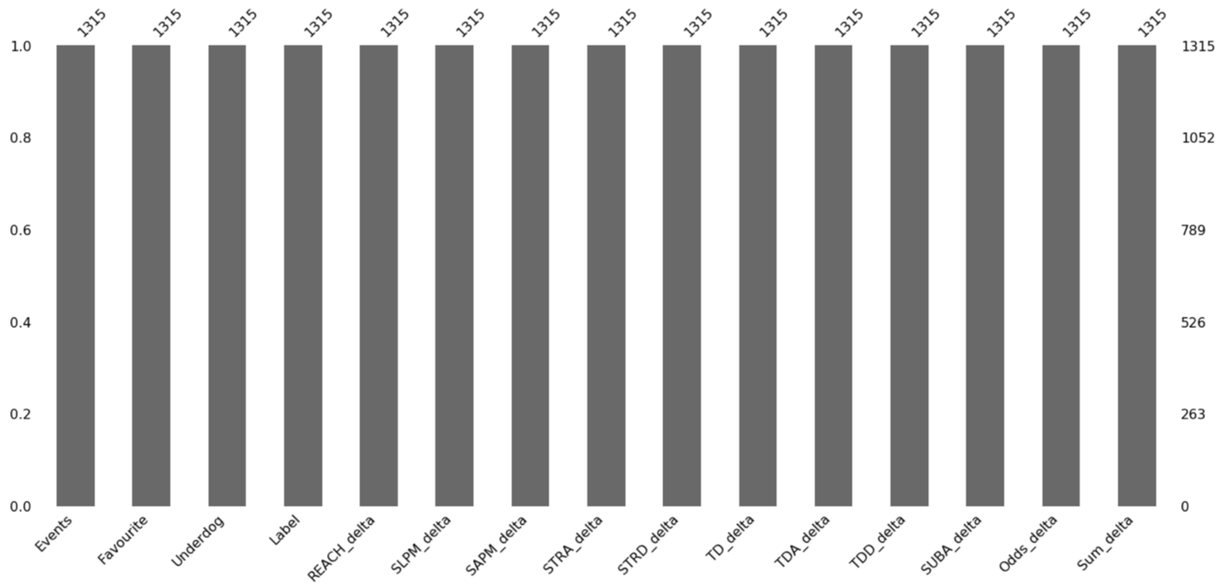


FIGURE 20: GRAPH TO SHOW IF ANY DATA INPUTS ARE MISSING

Data Visualization

With the data not missing any values the next part was to check the distribution of our dataset and evaluate the success rate for the Favorite or the Underdog (favorite being the fighter with greater experience and the underdog being the newer boxer) which is similar to red corner or blue corner for the previous dataset respectively.

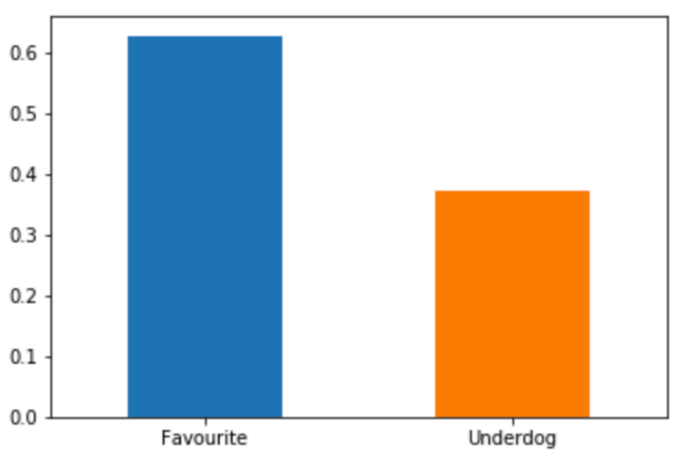


FIGURE 21: VISUAL DISPLAY OF WINS FOR FAVORITE VS UNDERDOG

Form this we can summaries that there is a 60% chance that the favorite will win and 40% chance the underdog takes the victory. From this simple visualization we can see that our accuracy has already increased in comparison to the work previously done. Although to achieve greater accuracies, more visualization is required more a deeper understanding on the dataset. Each fight have a score of strikes, takedowns and reach and this can be used to see which has the greatest impact of the decision of a fight

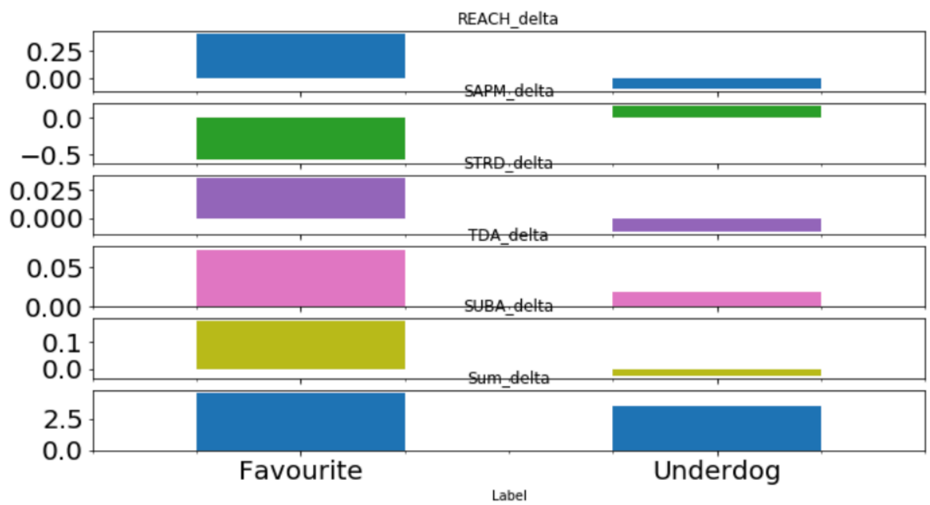


FIGURE 22: ABILITY DIFFERENCE BETWEEN FAVORITE & UNDERDOG PART 1

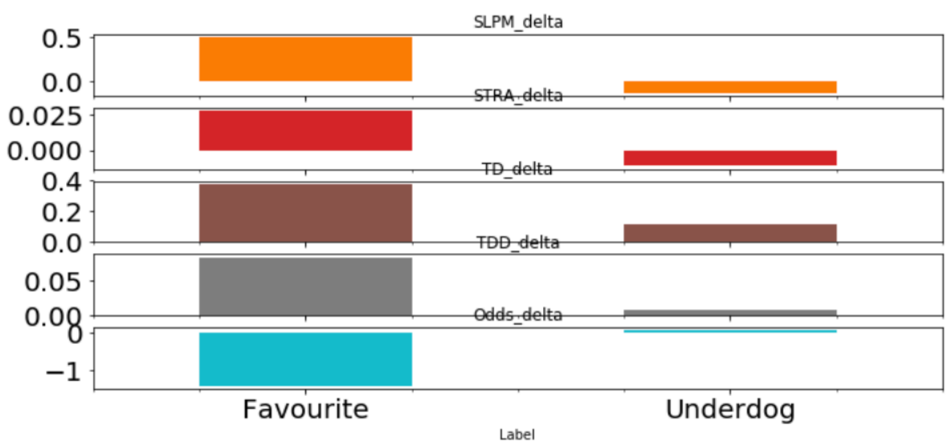


FIGURE 23: ABILITY DIFFERENCE BETWEEN FAVORITE & UNDERDOG PART 2

From these results we can infer that the favorite is the likely winner as they have more experience and greater skill in hitting their competitor with strikes and takedowns. While the underdogs do well in strike defenses, reach and takedown accuracy, their poor skill in the rest of the skills cause them to be weaker. This could be due to exhaustion. Exhaustion is a major deciding factor in a fighter change to win, as a fighter with more energy is likely to stand for longer and move better. Although exhaustion is not a factor we could measure as that can change a lot for each fighter, depending on activities that lead up to the fight, a common inference from the Kaggle dataset and this one is that the underdogs seem to always have less energy than the favorites.

While a bar plot can show us the best outcomes between favorites and underdogs, it cannot show us the distribution leading to the result, for this limitation we create a violin plot, this shows us how the data variates between the maximum, minimum and all quartiles within the interquartile range.

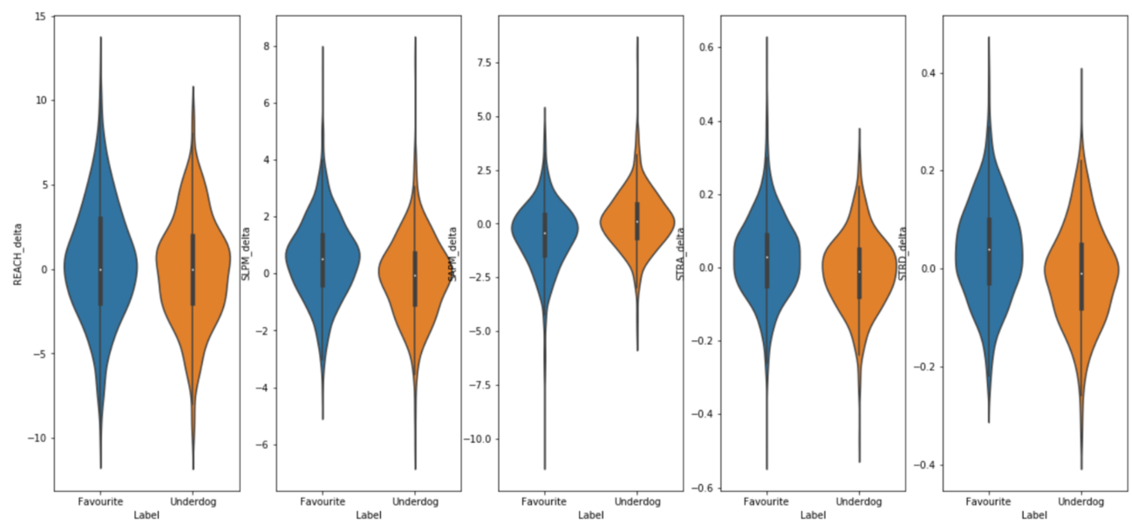


FIGURE 24: VIOLIN PLOT OF FAVORITE VS. UNDERDOG IN SKILLS PART 1

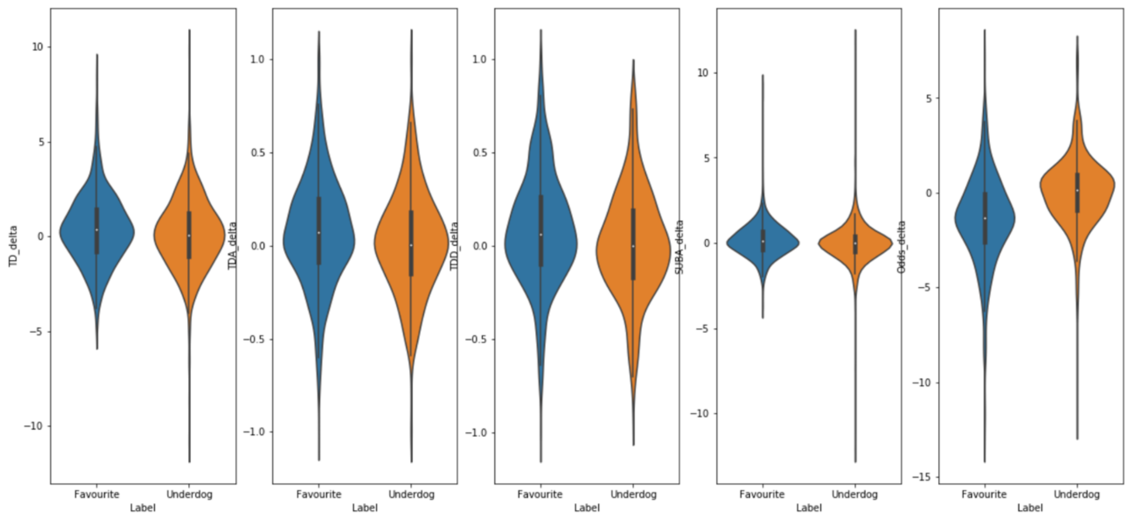


FIGURE 25: VIOLIN PLOT OF FAVORITE VS. UNDERDOG IN SKILLS PART 2

Favorite vs. Underdog

1. Reach

In reach we can see a very similar range, favorites and underdogs have an analogous interquartile range, while favorites improve on their maximum and minimum variables.

Not Important

2. Strikes Landed per minute

In SLPM we notice that there is a significant difference between the favorite and the underdog, giving the favorite fighter a serious advantage. With a small tail, the minimum of favorites is very close to the 15^{th} percentile of the underdog.

Important

3. Strike Absorbed per minute

Unlike the prior two, in this graphic representation a lower average is better, as in this aspect every strike absorbed makes the fighter weaker and increases their exhaustion, giving their competitor a greater advantage. This is easily seen to be an advantage in the favor of the favorite.

Important

4. Strike Accuracy

STRA is an accuracy measure for the number of strikes given by one fighter to another. While the quartile range for this scenario is similar, the maximum range increases the mean for favorites far beyond the underdog.

Not as important

5. Strike Defense

STRD shows the skill of a fighter to deflect a punch or kick thrown their way. This is seen in the graph that favorites have a better defense rating, meaning that they can avoid absorbing a strike while reducing their competitions energy.

Not as important

6. Takedowns

The underdog has a far greater spread in takedowns, and with their higher maximum takedown rate, makes this another factor that could be considered as important when an underdog wins. However, underdogs have a far lower minimum that the favorites too. With the minimum difference between favorites and underdogs being five for the whole fight.

Not as important

7. Takedown Accuracy

TDA is the only scenario here that has very similar accuracies for the favorite and the underdog, so there is a low chance that this would be a factor in the win accuracy.

Not as important

8. Takedown Defense

The take down defense shows the amount of times a takedown was deflected, and not allowed to occur because of the strength of one fighter to defend themselves.

Not as important

9. Average Submissions per 15 mins

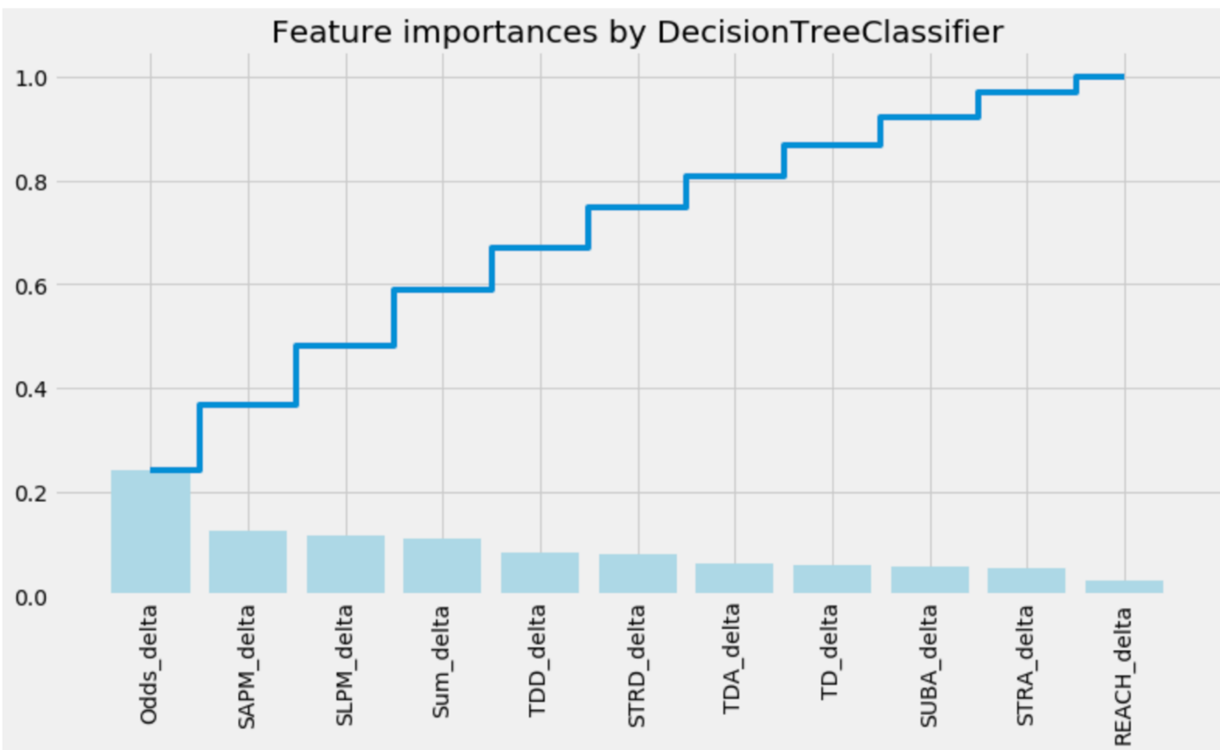
A submission, is when both fighters are on the ground and one fighter has the other in a locked position, not allowing them to move or break free.

Not as important

10. Odds

These are not a skill, rather the likelihood a competitor wins or loses. In an accuracy measuring project this is the greatest factor as it directly corelated with the accuracy of win/loss.

Most Important



Data Implementation

There Models Implemented [4]:

1. Logistic Regression

```
Logistic Regression
--------------------
Best Paramerters:
{'solver': 'newton-cg'}
Best Score:
0.6927756653992395
Best Model:
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
          intercept_scaling=1, max_iter=100, multi_class='warn',
          n_jobs=None, penalty='l2', random_state=None, solver='newton-cg',
          tol=0.0001, verbose=0, warm_start=False)

Scoring method:
make_scorer(accuracy_score)
```

2.

2. Random Forest

```
RandomForestClassifier
----------------------
Best Paramerters:
{'criterion': 'gini', 'max_depth': 5, 'max_features': 'log2', 'min_samples_leaf': 8, 'min_samples_split': 2, 'n_estim
ators': 4}
Best Score:
0.6988593155893537
Best Model:
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
            max_depth=5, max_features='log2', max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=8, min_samples_split=2,
            min_weight_fraction_leaf=0.0, n_estimators=4, n_jobs=None,
            oob_score=False, random_state=1, verbose=0, warm_start=False)

Scoring method:
make_scorer(accuracy_score)
```

3. Neural Network (MLP)

```
Multi Layer Perceptron (MLP)
----------------------------
Best Paramerters:
{'activation': 'tanh', 'hidden_layer_sizes': (10, 10), 'learning_rate': 'constant', 'solver': 'adam'}
Best Score:
0.7026615969581749
Best Model:
MLPClassifier(activation='tanh', alpha=0.0001, batch_size='auto', beta_1=0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=(10, 10), learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=1, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)

Scoring method:
make_scorer(accuracy_score)
```

Discussion

In the first dataset we notice our accuracies being as low as 51% and at the highest a 60%, which is basically the natural odds of any two-sided game. Building a system for betting system on accuracies as low as that is not much different from betting on the probability of getting heads on an unbiased coin. In order to reduce the risk of a bet the accuracy by which a predication can be made must increase, as risk is inversely proportioned to the to accuracy.

In a simple context, a coin that is biased towards heads is always going to reduce the risk of a bet when the bettor chooses heads. For example, a coin tossed ten times gets seven heads and 3 tails, meaning the accuracy is 70%, so the risk for a bettor reduces drastically too.

In our dataset we must improve the accuracy of wins for favorites or underdogs to reduce the risk. According to the kernels we sourced from Kaggle and set as a basis to improve accuracies on we found the highest accuracy after data manipulation was 60%. With the new dataset we first created a probability scenario a bookie would use to decide the victor and loser, with this we got a 3% increase in accuracy from the earlier dataset, which already showed the improvements of the new data.

As further analysis we created two similar models to the Kaggle dataset and one different but common supervised learning model, from which we found that these models gave us accuracies far greater than the earlier dataset. Logistic regression & Random Forest gave us an accuracy of 69%, while Multi-layer Perceptron gave us accuracies that were far greater than our expectation, with a 10% boost from the best model and scenario of the previous dataset and 7% above a bookie score. The Neural Networks (MLP) accuracy was 70%

Inference

From our research and testing of the new dataset we can infer that many of the skills could have led to an obvious victory or were not supportive enough to reach a valid conclusion. Amongst these inferences was the appearance of favorite fighters winning on an average of 63% of the fights they were in, and that most finishes came in at round 3 which made they need for rounds 4 and 5 unnecessary. Fights that concluded by decision in round three had empty inputs for round four and five which led to those columns having a lot of missing data.

Other inferences from the data explorations we found was the sizable disparity between the mean values for the columns between favorite fighters and the underdogs. This led to the favorites having a serious advantage in the scoring system of both datasets, which reemphasized our initial assumption, that favorites have an obvious advantage in winning over the underdogs, this could be as another feature we noticed was the disparity in mean age of the favorites and underdogs. Favorites in almost all fights were older than their underdog competitors which means they were more experienced.

The number of kick and punch strikes reduced exponentially from the first round to the second, with a gradual reduce between the second and third rounds, from these statistics we took into consideration a fighter's level of exhaustion as each round passed, with underdogs having a greater exhaustion rate than the favorites in every round.

After conducting data explorations with the dendrogram and the heatmaps we saw that there were not many strong correlations between two or more columns in either dataset, so there is a low possibility to have collinearity. However, in the final dataset we noticed an order of importance in the decision for a win, where odds was given most importance followed by strikes attempted per minute, strikes landed per minute, takedown defense, strike defense, takedowns attempted, takedowns landed, submissions, strikes attempted & finally the least important was reach. From this feature importance, we can supply this to coaches to use while training their fighters.

Lastly an important inference we found was that a fighter that once fought as a favorite, did not always remain a favorite, with increase number of fights a fighter can get better as they learn how to tackle other fighters of varying skills, however it could also go another way, where the fighter loses practice or their ability overtime and becomes weaker than their competitor, in this case becoming the new underdog. This was noticed when we measured the max age for both fighters was the same, at age 46. This. Shows that the fighter that was once a favorite, within a small amount of time also played as an underdog.

References

[1] Aggarwal, Karmanya Kaggle (2013)

[2] Clyde, Is illegal in Canada Gentry III (2001)

[3] Stefano, Dan Former UFC champ helps promote Pittsburgh event, (2009)

[4] Brown, Nicholas GitHub, (2018)