National College of Ireland

BSc in Computing

Data Analytics

2021-2022

Gavin Walsh

X17364783

X17364783@student.ncirl.ie

Predictive Analysis of UFC Fights

Technical Report

Contents

Executive Summary

Within this report, all the major talking points in regards to this project are discussed which ranges from the reasons why predicting UFC fights was chosen and an insight into the background of the project. The technologies section will discuss the different technologies that were used to enhance this project such as RStudio acting as the main hub for the project as well as the R Language being used to incorporate the coding. There will be a data section in which will contain a detailed description of the data which will discuss how the data was discovered and also how the data was transformed for it to be in the best state possible. The main attributes involved in this data set will also be discussed and why it worked out and enhanced the project. Following this will be the methodology section in which the KDD Methodology will be discussed and will be detailed with each step of the method followed and explained closely. Next, the report will go in depth with the predictive analysis approach to the project and reasons why certain algorithms were chosen and singled out will be explained such as using a Neural Network. Hyperparameters will be a big talking point of the analysis section with how important they are to creating a good model. Following the analysis discussion, the results that had been gathered from the analysis of the data will be considered and these results will be displayed. Once the results that were retrieved are visible, following it up will be the evaluation of the results and how the results can be interpreted. Finally, the last point will be how the project could be improved with more time and knowledge gained with the subject.

1.0 Introduction

1.1. Background

Mixed Martial Arts has a lot of variables to take into account and can be quite unpredictable due to this. There are many paths to victory such as a Knockout, Submission or Decision. There are many disciplines included with Mixed Martial Arts, whether it be Boxing, Kickboxing, Muay Thai, Jiu-Jitsu, Judo or even Wrestling. Usually, fighters fight other fighters who are close to them in the rankings which means they are at a similar skill level and thus makes it even harder to predict at times.

A literature review was conducted regarding the area of sports outcome prediction with the main focus being sports similar to MMA such as tennis to find some predictive analysis papers to get a better understanding. One of the first papers found from the research was predictive analysis in regards to Tennis [1]. Within this paper, different types of machine learning algorithms were discovered that can be used in projects like this current project. Although this was a larger scale paper, this was a good indicator into how these projects work and how this should move forward as a project.

Another paper found within this realm was a paper that also predicted UFC fights [2]. Within this paper, different machine learning algorithms were carried out to gather accuracies on the data set used. Perceptron, SVM, Random Forest, Decision Tree are some of the algorithms used here. The results that were gathered here ranged from between 50% and 61% with the exception for the Naïve Bayes model which only produced 21%. Choosing some of these algorithms and beating their accuracy would only enhance this project.

Predictions are a huge part of the MMA community which is visible by the rise of an application called Verdict MMA or the website which can be reached at VerdictMMA.com [3]. This is a platform where users can predict the outcome of fights in which they can earn points for predicting correctly and move up the ranks. This platform has become popular and is even used on some MMA live broadcasts! It would be interesting to use some logic and algorithms to predict these fights.

1.2. Aims

The main aims of the project were

- Locate and collect a sizable dataset of MMA Fights that includes usable attributes to be used as the cornerstone of the fight predictive analysis.
- Build a predictive model that is trained on a subset of the original dataset of MMA fights to measure and discuss the performance of the model up against the original dataset.
- Compare different model using the cornerstone that was collected.
- Use parameter tuning along with other machine learning methods to increase the accuracy of the model to the highest that can be achieved.
- Gain knowledge on which parameters have the most value when it comes to fight predictions.
- Record the results gathered from the built models and display them on graphs.
- Use research papers and the accuracies within them to compare with the accuracy that was gathered within my model.

1.3. Technology

RStudio

R studio is an open-source integrated development environment (IDE) for R Language. Used as the main hub of the project as within RStudio, the data was able to be cleaned, manipulated the data and also stored the processed data here to be viewed. It is the main home of the R programming language in which it was used with this project. Contains packages that were used here such as caret, neuralnet, h2o and randomforest. Allowed the results to be viewed and gathered within RStudio.

R Language

The R language is widely used among statisticians and data miners for developing statistical software and data analysis. R is the programming language that was used throughout this project and incorporated the machine learning algorithms that were chosen to build the models and to produce results for the models through predict functions. Also allowed for different features such as grid search for hyperparameter tuning to be implemented.

Excel

Used to edit data set + change the file from csv to excel file. Used to hold figures and create graphs for the results section of the report which included accuracies and hyperparameter tuning for both the Neural Network model and also the Random Forest model.

1.4. Structure

Introduction

Within the introduction section, there will be a discussion about the background of the project and the context behind it. Research that was conducted on similar ideas will be viewed within this section too. Following this, there will be an aims section in which the main aims of the project will be displayed which can range from finding the right dataset to creating the model for the project. The technologies section will follow up the aims section in which the technologies used within the project will be listed and discussed. Lastly, the current section, structure, will be visible. This section focuses on explaining the structure of the report and what to expect in the future sections!

Data

The data section will hold high importance. As a main driving factor behind the project. Getting the right data with the right variables is needed. This will be discussed in the data section! There will also be an open discussion on where the data came from along with if there were other datasets that could have been used. The number of rows and also variables within the data set will also be touched on here.

Methodology

This section will discuss the process on how the bridge was made between going from the data to getting the results and analysing. Within the methodology section, there will be an explanation on why the KDD method was followed along with an in-depth guide through these steps for this project which will include the data selection, data pre-processing, data transformation and data models.

Analysis

The analysis approaches will be discussed here along with the reasoning behind choosing this approach. An explanation for choosing certain variables within the modelling phase will be conducted here also. Hyperparameters will be a driving force in this section and how it came about using each different hyperparameter.

Results

This section will be used to hold all the key results of the analysis that was conducted previously. Tables, charts and figures can all be used here to show the results that were gathered. There will be an explanation behind these results too along with a reasoning on why they are important to the project.

Conclusion

The conclusion made of the final results will be present here. Discussion on which model performed the best along. There will be a comparison of results with a paper from the literature review to view how the results compare to each other. Strengths and weaknesses of the project was also discussed here.

Future Work

This is where information on what could be done with extra time will be discovered. Different models and techniques were discussed here to improve the project in the future!

References

The references section of the project will only be used to hold the references that were used throughout the rest of the report.

Appendices

Within the appendices section, extra figures related to the project that were not shown previously are viewable here. Along with this, the project proposal and project plan will be available to be seen here. Lastly, the reflective journals will be available at the end of this section.

2.0 Data

After thorough research through avenues such as Google Datasets and Kaggle, the most suitable dataset for this project was discovered through Kaggle. The dataset “ufc_master” contains all the relevant data that was needed for this project. Within this dataset contains MMA fight data ranging from 2021 all the way back to 2010. There are 4896 different entries of MMA fights while there are 119 variables within these entries which vary from dates to names to winners to fighter stances. Other options were explored such as scraping data from the UFC website but after further review, they in fact do not allow this which meant that avenue was blocked. Similar to the UFC website, another website by the name of Sherdog which contains fight data was also explored but they also did not allow web scraping of their data. These were the main sources of MMA data following research. Although it would have been ideal to gather some more data to go along with the dataset that was already retrieved, it just wasn’t possible but the data from Kaggle contains some really good information and 4896 different fights to analyse so it is way more than enough to use.

Variable Name	Description
R_fighter	Name of red corner fighter
B_fighter	Name of blue corner fighter
Winner	Winner of the fight
weight_class	Weight class of fighters
gender	Gender of fighters
no_of_rounds	Number of rounds scheduled for fight
B_current_lose_streak	Current losing streak of blue fighter until this fight
B_current_win_streak	Current winning streak of blue fighter until this fight
B_draw	Number of draws for blue fighter
B_avg_SIG_STR_landed	Average number of significant strikes landed for blue fighter
B_avg_SIG_STR_pct	Average percentage of significant strikes landed by blue fighter
B_avg_SUB_ATT	Average number of submissions attempted by blue fighter
B_avg_TD_landed	Average number of takedowns landed by blue fighter
B_avg_TD_pct	Average takedown percentage for blue fighter
B_longest_win_streak	Longest win streak obtained by blue fighter
B_losses	Number of losses for blue fighter
B_total_rounds_fought	Number of rounds fought within the UFC
B_total_title_bouts	Number of title fights within the UFC

B_win_by_Decision_Majority	Number of wins via majority decision for blue fighter
B_win_by_Decision_Split	Number of wins via split decisions for blue fighter
B_win_by_Decision_Unanimous	Number of unanimous decision wins by blue fighter
B_win_by_KO/TKO	Number of wins by KO for blue fighter
B_win_by_Submission	Number of submission wins for blue fighter
B_win_by_TKO_Doctor_Stoppage	Number of wins by doctor intervention for blue fighter
B_wins	Number of total wins for blue fighter
B_Stance	Stance used by blue fighter
B_Height_cms	Height of blue fighter
B_Reach_cms	Reach of blue fighter
B_Weight_lbs	Weight of blue fighter
R_current_lose_streak	losing streak of red fighter until current fight
R_current_win_streak	Winning streak of red fighter until current fight
R_draw	Number of draws for red fighter
R_odds	Odds for red fighter to win
B_odds	Odds for blue fighter to win
R_ev	Further odds for red fighter
B_ev	Further odds for blue fighter

date	Date of fight
Location	Location of fight
Country	Country of the fight
Title_bout	If the fight was a title fight or not
R_avg_SIG_STR_landed	Average significant strikes landed by red fighter
R_avg_SIG_STR_pct	Average significant strikes percentage for red fighter
R_avg_SUB_ATT	Average submission attempts by red fighter
R_avg_TD_landed	Average takedowns landed for red fighter
R_avg_TD_pct	Average takedown landed percentage for red fighter
R_longest_win_streak	Longest win streak obtained by red fighter
R_losses	Number of losses for red fighter
R_total_rounds_fought	Number of rounds fought by red fighter
R_total_title_bouts	Number of title fights for red fighter
R_win_by_Decision_Majority	Number of majority decision wins by red fighter
R_win_by_Decision_Split	Number of split decision wins for red fighter
R_win_by_Decision_Unanimous	Number of unanimous decision wins by red fighter
R_win_by_KO/TKO	Number of KO wins by red fighter
R_win_by_Submission	Number of submission wins for red fighter

R_win_by_TKO_Doctor_Stoppage	Number of doctor intervention wins for red fighter
R_wins	Total wins for red fighter
R_Stance	Fighting Stance of red fighter
R_Height_cms	Height of red fighter
R_Reach_cms	Reach of red fighter
R_Weight_lbs	Weight of red fighter
R_age	Age of red fighter
B_age	Age of blue fighter
lose_streak_dif	Difference between losing streak of fighters
win_streak_dif	Difference between win streak of fighters
longest_win_streak_dif	Difference between longest win streaks of fighters
win_dif	Win differential between fighters
loss_dif	Loss differential between fighters
total_round_dif	Total rounds fought differential between fighters
total_title_bout_dif	Number of title fights differential between fighters
ko_dif	Number of Kos differential between fighters
sub_dif	Number of submissions differential between fighters
height_dif	Height difference between fighters

reach_dif	Reach difference between fighters
age_dif	Age difference between fighters
sig_str_dif	Significant strike differential between fighters
avg_sub_att_dif	Average submission attempts differential
avg_td_dif	Average takedown differential
empty_arena	If the arena was empty or not due to covid
constant_1	Constant variable
B_match_weightclass_rank	Rank within weight class of fight for blue fighter
R_match_weightclass_rank	Rank within weight class of fight for red fighter
R_Women's Flyweight_rank	Red fighters rank at women’s flyweight
R_Women's Featherweight_rank	Red fighters ran at women’s featherweight
R_Women's Strawweight_rank	Red fighters rank at women’s strawweight
R_Women's Bantamweight_rank	Red fighters rank at women’s bantamweight
R_Heavyweight_rank	Red fighters rank at heavyweight
R_Light Heavyweight_rank	Red fighters rank at light heavyweight
R_Middleweight_rank	Red fighters rank at middleweight
R_Welterweight_rank	Red fighters rank at welterweight
R_Lightweight_rank	Red fighters rank at lightweight
R_Featherweight_rank	Red fighters rank at featherweight

R_Bantamweight_rank	Red fighters rank at bantamweight
R_Flyweight_rank	Red fighters rank at flyweight
R_Pound-for-Pound_rank	Red fighter rank on the pound for pound rankings
B_Women's Flyweight_rank	Blue fighters rank at women’s flyweight
B_Women's Featherweight_rank	Blue fighters rank at women’s featherweight
B_Women's Strawweight_rank	Blue fighters rank at women’s strawweight
B_Women's Bantamweight_rank	Blue fighters rank at women’s bantamweight
B_Heavyweight_rank	Blue fighters rank at heavyweight
B_Light Heavyweight_rank	Blue fighters rank at light heavyweight
B_Middleweight_rank	Blue fighters rank at middleweight
B_Welterweight_rank	Blue fighters rank at welterweight
B_Lightweight_rank	Blue fighters rank at lightweight
B_Featherweight_rank	Blue fighters rank at featherweight
B_Bantamweight_rank	Blue fighters rank at bantamweight
B_Flyweight_rank	Blue fighters rank at flyweight
B_Pound-for-Pound_rank	Blues rank on the pound for pound rankings
Better_rank	Better rank of both fighters
Finish	How the fight ended
Finish_details	Details on how it ended

Finish_round	Round the fight was ended
Finish_round_time	The time within the round the fight was finished
Total_fight_time_secs	How long the fight was
R_dec_odds	Odds for a red fighter decision win
B_dec_odds	Odds for a blue fighter decision win
R_sub_odds	Odds for a red fighter submission win
B_sub_odds	Odds for a blue fighter submission win
R_ko_odds	Odds for a red fighter ko win
B_ko_odds	Odds for a blue fighter ko win

Fig 1 – Description of Variables within dataset

Once the dataset was retrieved, the file type was changed from csv to an excel file type to be read in RStudio. The dataset was then imported into RStudio. Within RStudio, the data began to be pre-processed so that it was in a better state to be used as the cornerstone of the project. The variables that are present within this dataset were seen as very useful and impactful with variables such as current_win_streak, reach_dif, age etc. these attributes have not been seen used in researched projects around the proximity of this project. These variables will make the models stand out compared to the others. Key attribute values were changed in order to enhance the models down the line. Studying the dataset further, it was deemed important that the string data within this dataset should be changed and was completed in the following section.

Upon further research down the line, more variables were removed due to them being deemed as not needed. Some of these variables included were the ranks. These were removed as most of them contained NA values and it wouldn't make sense changing them to actual figures due to not knowing exactly where they rank in their divisions so it wouldn't make sense if all the NA values were turned to 0's. other variables that were removed came in the form of if there was an empty arena, the date, the location etc. some of these would have been valuable if the project was based around a sole fighter to see if they perform better in a certain location or a certain time period.

3.0 Methodology

The methodology that was chosen here was the Knowledge Discovery in Databases also known as the KDD method. This method was chosen over the CRISP-DM method as the KDD method is focused more on the research compared to the CRISP-DM method. This method contains different steps to follow to complete it. These steps are the following: Data Selection, Data Pre-processing, Data Transformation, Data Mining, Analysis, Results, Evaluation. These steps are the driving force of the method and must be followed.

3.0.1. Data Selection

The selection of data has already been carried out by choosing to use the data that was retrieved from Kaggle which was the "ufc_master" data set which contains 4896 fights and 119 variables. The selected data was changed to an excel file and then imported into RStudio so that the pre-processing of the data could begin.

3.0.2. Data Pre-Processing

This section is dedicated to handling issues that are within the data set. After some data exploration of the data set, it was found that 50 different variables were deemed unnecessary to the project. These were variables such as the date and the location of the fight. these would have been valuable if the project was based around a sole fighter to see if they perform better in a certain location or a certain time period. Other variables that were removed were variables such as the rankings in the division of each fighter. As previously stated, these were deemed as not needed due to the fact that the majority of the entries here were filled with NA values as only the top 15 fighters in every division get ranked which means there would be so much more NA entries than entries that contain values and adding 0's to the NA values wouldn't make sense as it would skew the data too much compared to if there was only one or two NA entries that could just be filled with a 0.

This was the case for the rest of the variables that didn't have many NA values. 0's was added in to replace the NA values so that the features could be deemed as usable. This was the case for variables such as 'R_avg_SIG_STR_landed'.

Following the data pre-processing, the data set is now left with the same number of rows which is 4896 but now contains 69 variables down from 119. Although all of these variables left might not be used in the final product, these are the variables that are the most useful for the implementation of data mining.

3.0.3. Data Transformation

Now that the pre-processing of the data was completed, the transformation of the data commenced to improve the data into the best state possible. This started with changing the stance variables from string data to numerical data. Each type of fighting stance was given a specific number such as "Orthodox = 0". This was then followed by changing the main variable within the dataset, which was the 'Winner' variable to either 0 for the red fighter winning and 1 for the blue fighter winning. This variable was also changed to numeric. Other variables such as 'Gender' which was deemed usable was also changed to numeric and given numeric values. Next, the weight class variable was given numeric values for all 12 of the weight classes just in case they were used in the models. That concluded the changing string variables to numeric values.

Following this, normalisation was then introduced into the dataset. This decision was made due to normalisation having benefits like creating a more compact, concise set of values for a variable, it can help the run time of models and also can provide better results on the backend of the models. The variables in which normalisation was used with were the age variables for both the red and blue fighters, the current win and losing streaks for each set of fighters and also the wins and losses between the red and blue fighters. To carry this procedure out, the max value of each variable was found in which was used to divide into each value of the variable. This means that the highest value will equal to 1 and everything below it will be between 0-1.

Once the data set was in a good state to move forward, the data was partitioned into an 80/10/10 split. These splits were the Training Data, Validation Data and Test Data. The Training Data for building the models, Validation Data for validating the outcomes of the models and then the Test Data used at the very end to verify the results.

3.0.4. Data Mining Models

Following research undertaken on machine learning algorithms within the space of 1v1 sports such as MMA, Boxing and tennis, a lot of the algorithms are used quite frequently. The algorithms that were chosen here were the Neural Network algorithm and also the Random Forest algorithm. Although one algorithm could be used alone, having an extra algorithm to compare the results enhances the project.

Neural Networks

A Neural Network is a machine learning algorithm that is a series of algorithms that can view underlying relationships in a provided dataset by processes that are quite similar to how the brain functions [4]. The main structure of a neural network is it has 3 main components. The input layer, the hidden layer and the output layer. This algorithm is very useful for predictive analysis and is also excellent for using numeric data! This is why this model was chosen. To create this model, the 'neuralnet' package was used in R to create the model. To fully create the model, the variable to be the predictor was chosen along with the variables that will be used to get the prediction on the back end. The data set must also be chosen along with the hyperparameters that were to be used! The hyperparameters and variables will be discussed in the analysis section! The diagram below shows how the Neural Networks come together visually [5]. Overall, this was a very useful choice of algorithm due to its nature of working well with predcitve analysis with numeric data.

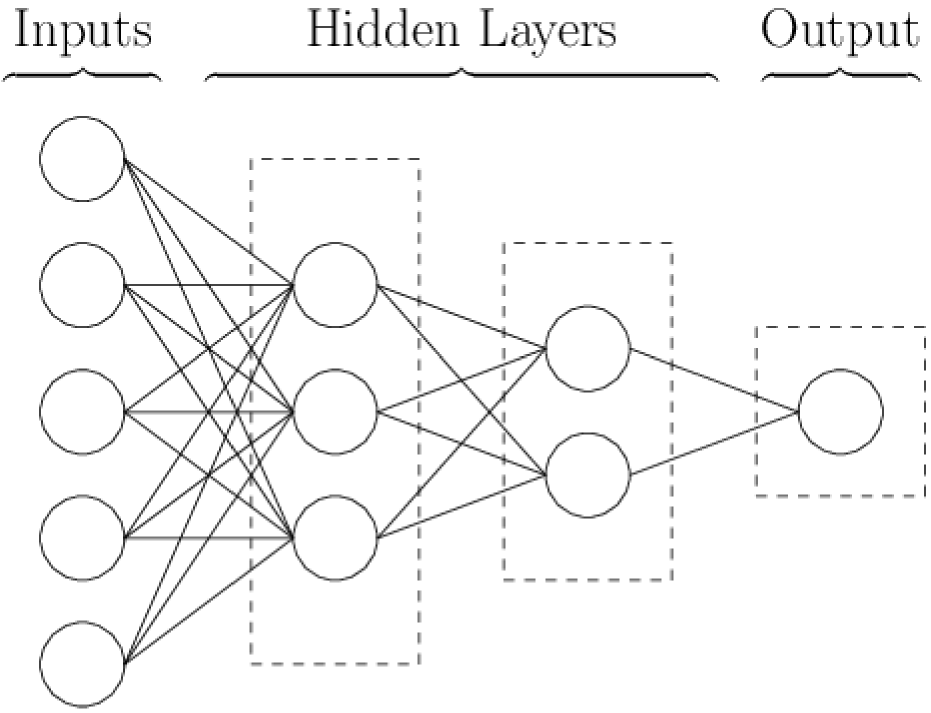


Fig 2 – Neural Network diagram

Random Forest

The Random Forest algorithm was the second algorithm incorporated into the project. Random Forest's consist of many decision trees. It randomly chooses variables to use within the building of the trees and gathers a certain number of trees together and finding the average prediction of the trees, hence called a forest [6]. This model is quite different on the surface to a Neural Network which aids the results due to giving more variety. The 'randomforest' package was used to build this model. As with the NN, a variable was chosen as the predictor along with the variables that will be used to predict the outcome. Hyperparameters are included here too. The diagram below shows how the Random Forest model runs visually [6]. Good choice of algorithm due to its effect predictive nature while also bringing a different style than a Neural Network.

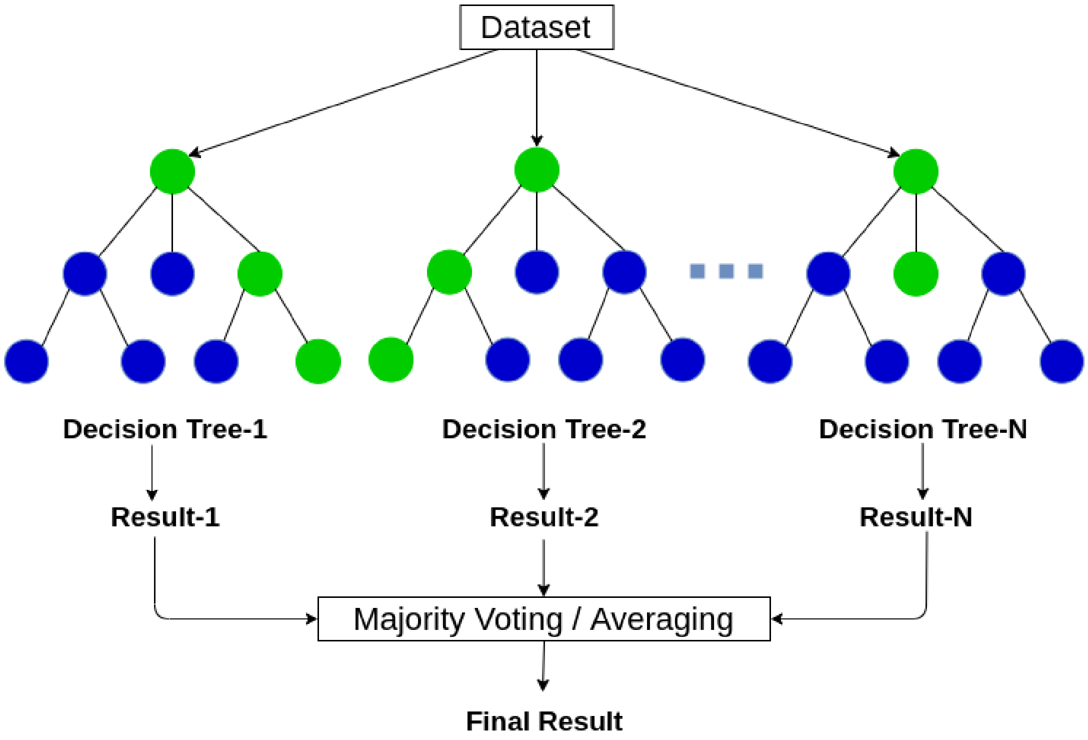


Fig 3 – Random Forest diagram

4.0 Analysis

For the analysis of the data, the results will be influenced on what happened in this section. Choosing the variable to be predicted from the models occurred. This was the 'Winner' variable as the main point of the models is to predict the winner of a fight. The 'Winner' variable was used within both the Neural Network and also the Random Forest.

To find out which variables are to be included in the models to predict the outcome, a correlation matrix was used to see which variables will have the most impact on the 'Winner' variable.

From studying each relationship in terms of correlation with all the other variables and the 'Winner' variable, the variables that were chosen to be used were the following : avg_td_dif + avg_sub_att_dif + reach_dif + total_title_bout_dif + longest_win_streak_dif + win_streak_dif + lose_streak_dif + R_age + R_win_by_Decision_Split + B_win_by_Decision_Split + R_current_lose_streak + R_current_win_streak + B_current_lose_streak + B_current_win_streak. These were the variables which showed the best correlation with the Winner and therefore were included into the models.

	Winner
Winner	1.00
avg_td_dif	0.12
avg_sub_att_dif	0.07
reach_dif	0.09
total_title_bout_dif	0.10
longest_win_streak_dif	0.08
win_streak_dif	0.15
lose_streak_dif	0.06
R_age	0.07
R_win_by_Decision_Split	0.09
R_current_win_streak	0.10
B_current_lose_streak	0.06
B_current_win_streak	0.07

Fig 4 – Correlation Matrix of variables chosen

Now that the features have been chosen for the models to get the prediction that's needed and also provided the best features to get the best results, hyperparameters were then implemented to get the results to the best place possible. Hyperparameter tuning is very important as if it's not done correctly, it can make your results worse instead of enhancing them. Hyperparameters for Neural Networks and Random Forests are not the same so they will be split up for this section.

4.0.1. Neural Network

For the Neural Network hyperparameters, the first one implemented was declaring the training data set as the data used to build the model. Following this, the stepmax was introduced and given a really high value so it could pass through the algorithm without going over the maximum number of steps. The last straight forward hyperparameter was the rep parameter in which was given a value of 1 repetition. Next the learning rate was introduced. The learning rate was given 3 different values which were 0.1, 0.01 and 0.001. these values were put into the model and the results were recorded to see which learning rate was the right choice for the finished model. For this model, 2 layers were used as a hyperparameter to gather the following results.

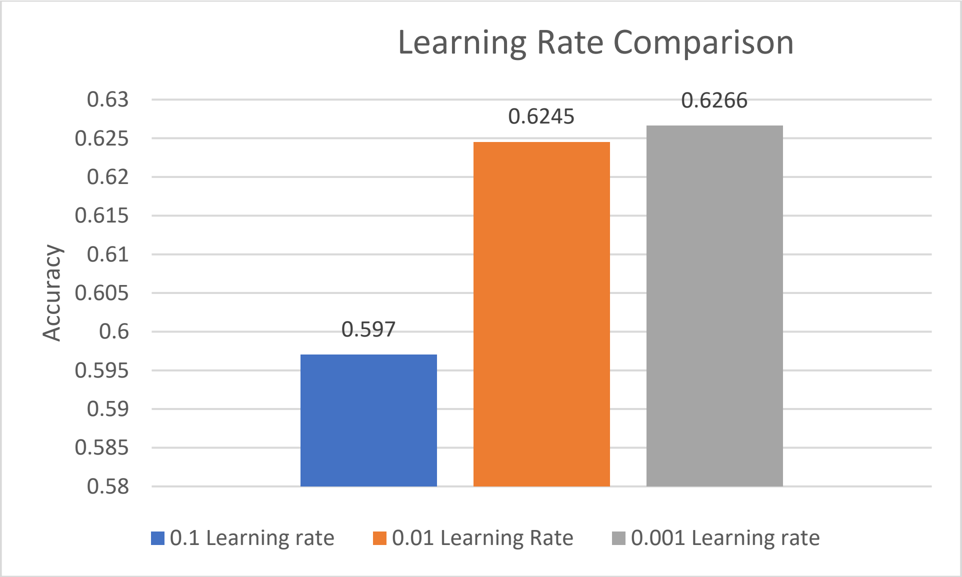


Fig 5 – Learning Rate Comparison

From the graph above, the results show that a learning rate of 0.1 performed the worst of the 3. 0.01 and 0.001 are close but ultimately 0.001 just edges the 0.01 learning rate which means that a learning rate of 0.001 was included into the Neural Network model. These figures are gathered based on the validation data set and not the final test set.

Lastly for the Neural Network, the number of hidden layers had to be determined, just like the learning rate, there were 3 different values that were put into the model and the results were recorded to see which number of layers performed the best. The 3 different values used were 1,2 and 3 hidden layers. As this was done following the learning rate gathering, a learning rate of 0.001 was used to build the models with the different number of layers.

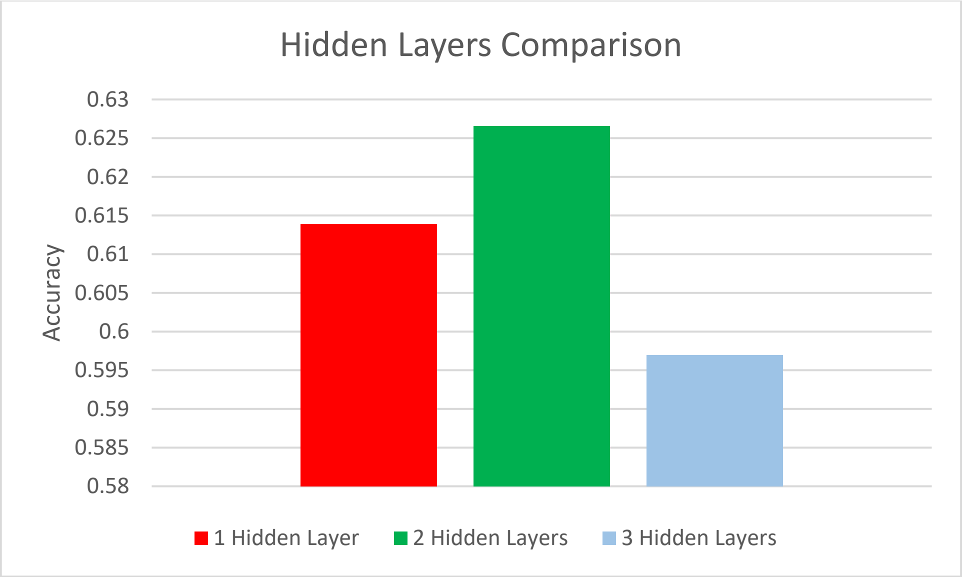


Fig 6 – Hidden Layer Comparison

From the results provided above, the value of 2 hidden layers was a big performer in comparison to the other 2, especially 3 hidden layers which came in just below 0.60. 1 hidden layer didn’t perform too badly being just over 1% behind 2 hidden layers. This means that a value of 2 was given to the hidden hyperparameter in the model building for the Neural Network.

Now that the variables and hyperparameters have been chosen and tuned to the best possible stage, the model for the Neural Network was built and ready to give results.

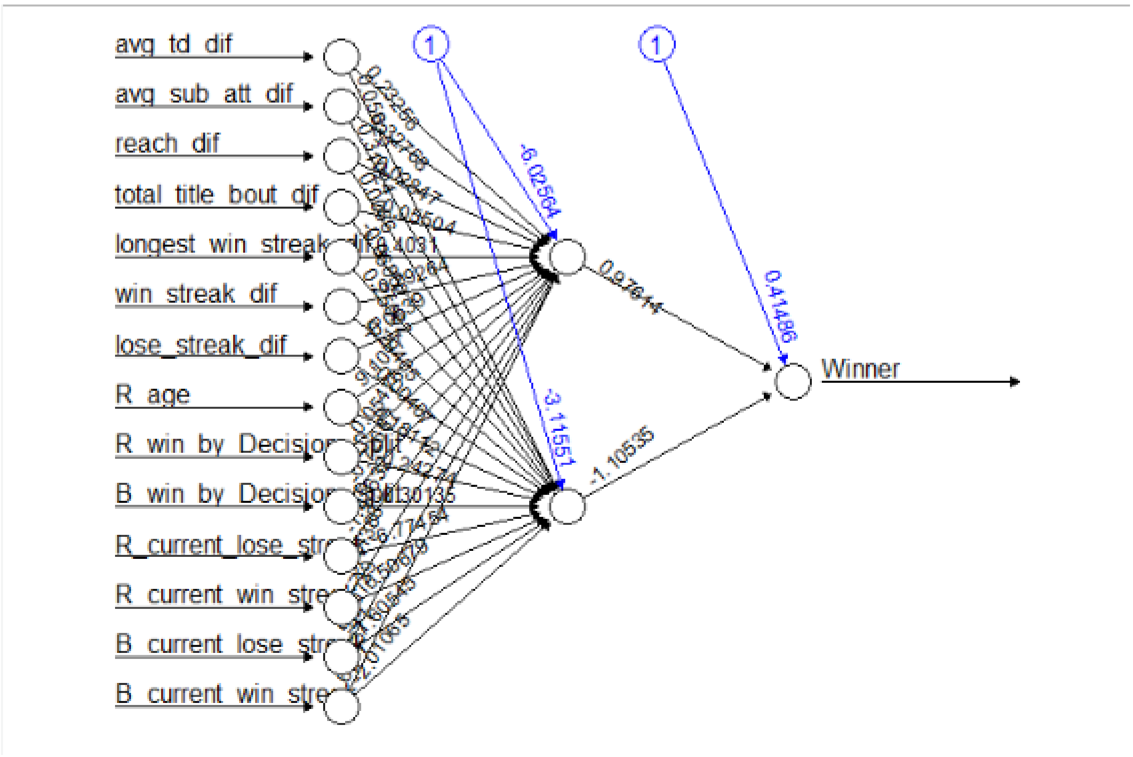


Fig 7 – Plot for Neural Network

4.0.2. Random Forest

The Random Forest model was then worked on. The same variables were used within this model too so that there is consistency across the board. The 'Winner' variable was also chosen again as the predictor. In regard to hyperparameters, learning rate was the only one to include within this model. The learning rate for the Random Forest was retrieved in a different fashion in comparison to the Neural Network. Grid Search was implemented for the Random Forest to gather the best learning rate. This was implemented by using the 'h2o' package! To find the best performing learning rate for the random forest, the residual deviance was used to figure it out. Whichever learning rate had the lowest figure would be the one to be used.

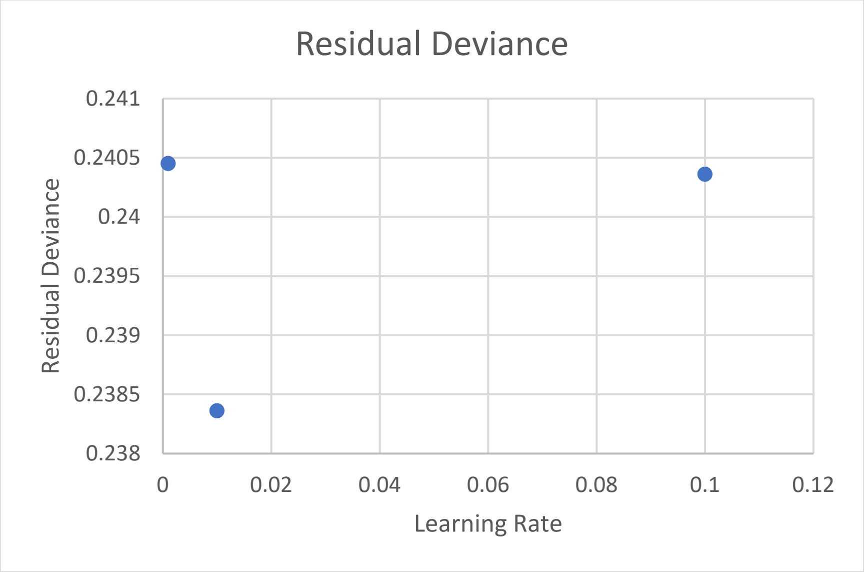


Fig 8 – Residual Deviance for the Learning Rate – Random Forest

The results for this were very close to each other but a learning rate of 0.001 was the worst performer just above 0.24045 and 0.1 right behind it at 0.24036. the best performer was 0.01 which came in at 0.23836. this means that a learning rate of 0.01 was used as a hyperparameter of the Random Forest model.

As the predictor, variables and hyperparameters were chosen to be included into the building of the Random Forest model, the model was then built to completion ready to produce results.

5.0 Results

To gather the results, the percentage coming out of the model predictor was rounded up or down depending on which side of 50% it landed on. This was done to match the 'Winner' variable which was 0 for the red corner fighter and 1 for the red corner fighter.

Once the models were built and the prediction method was carried out to gather the results, the accuracy of the model was then found as seen below for both the Neural Network and the Random Forest models.

Model	***Accuracy***
Neural Network	***0.6148***
Random Forest	***0.6004***

Fig 9 – Final Results for Model Prediction

As can be seen above, the Neural Network model produced an accuracy of 61.48% while the Random Forest produced an accuracy of 60.04% when using the Test Data. This means that the Neural Network was the better performer in terms of predicting the winner of a UFC fight compared to the Random Forest model. Although the accuracy of the Random Forest model is not too far away from the Neural Network model at a 0.0144/1.44%. Now we will view the performance of the Test Data, which wasn't used until the gathering of the final accuracy above, up against the Validation Data, which was used during the building stages of the model to gather accuracies for hyperparameters and variables.

Model	***Data Set***	***Accuracy***
Neural Network	**Test Data**	**0.6148**
	Validation Data	**0.616**
Random Forest	**Test Data**	**0.6004**
	Validation Data	**0.5865**

Fig 10 – Results of the models for prediction with Test and Validation Data

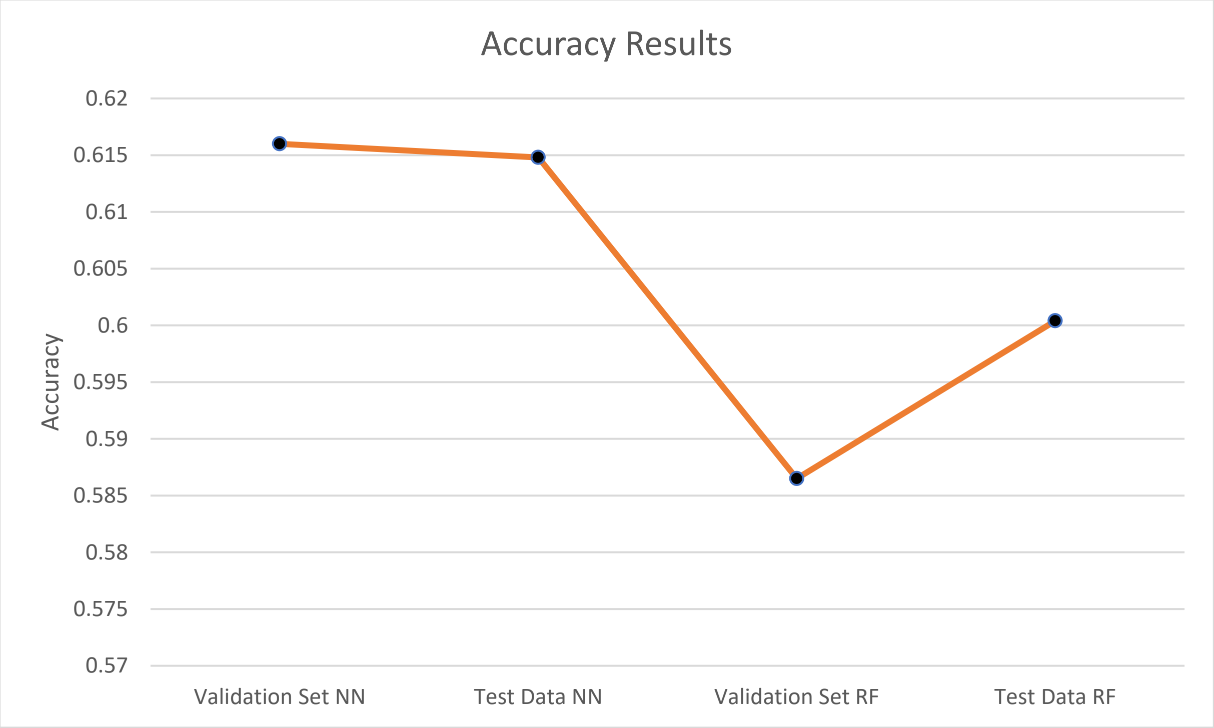


Fig 11 – Line Graph for Results with Test and Validation Data

For the Neural Network, the Validation Data outperformed the Test Data with an accuracy of 61.60% vs 61.48%. Although it's only a measly .12% more than the Test Data, it still just slightly outperforms the Test Data. It is still a really good performance for the test Data as usually the Validation Data outperforms the Test Data by a wider margin.

For the Random Forest model, the Test Data actually outperforms the Validation Data by a nice margin too. The Validation Data accuracy came in at 58.65% while the Test Data produced an accuracy of 60.04%! this was surprising as like stated above, there is normally a wide margin difference in favour of the Validation Data compared to the Test Data.

Overall, with these 4 accuracies, the Random Forest model using the Validation Set is lacking compared to the rest which are all within the range of 60% and above!

6.0 Conclusions

In conclusion, the Neural Network model came away as the strongest model that was built between the Neural Network and the random Forest models. The gap between these 2 models was not too big but Neural Network still performed best. The hyperparameter tuning definitely enhanced the results of these models to give higher accuracies in terms of the UFC fight winners. With how unpredictable the sport of Mixed Martial Arts can be at most times, an accuracy of just over 60% isn't unreasonable at all! Although these results are good, during the building stage testing hyperparameters with the validation set, a higher accuracy was gathered at 62.66%.

A major strength this project can take away is that it outperformed the research paper within the literature review [2] earlier in this report which focused on predicting UFC fights also.

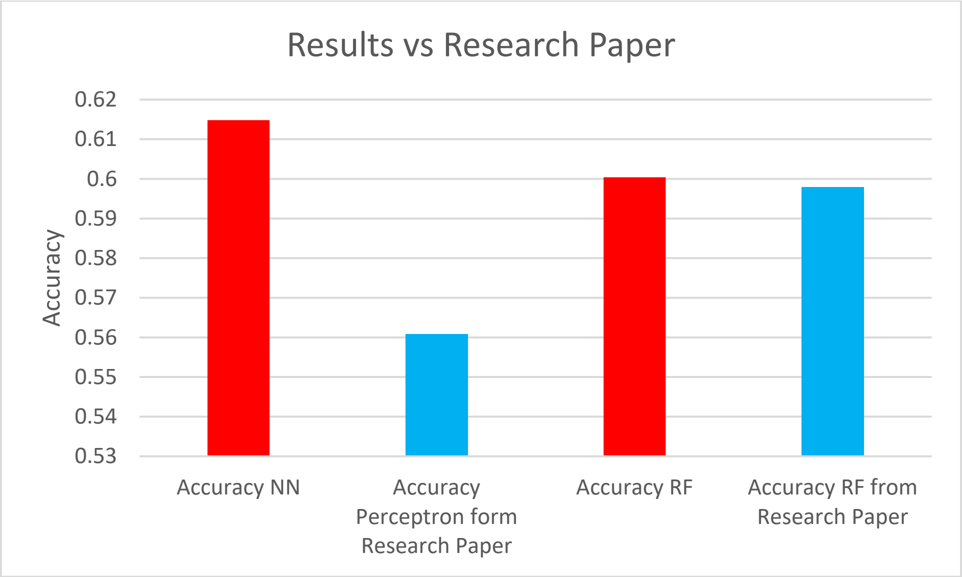


Fig 12 – Results vs Results from Research Paper

With the results gathered from this project in red and results from the research paper in blue [2]. Perceptron was the closest model to a Neural Network model in which the accuracy gathered from here outperforms the Perceptron model with the perceptron model coming in at 56.08% going up against the 61.48% of the Neural Network model. This gives a gap of 5.4% in favour of the Neural Network model which is a good performance. For the 2 Random Forest models, our model slightly outperforms the model from the research paper with the accuracies at 60.04% and 59.79% respectively. A slight gap of .25% between the 2 but this means that the models built here outperform both these models from the research paper which is a strength of this project!

The outperformance of the models may be due to differing variables that were used to build the models. In the research paper, the features chosen were completely different than the features that were chosen here. This may have been the main factor behind these models performing better than the research paper models!

What this research paper has over this project though is the volume of algorithms compared to 2 here. There is multiple more within that paper that gives more variety than here. This can be seen as a limitation of this project.

Overall, both models that were developed and built for the prediction of the UFC fights in question managed to produce reasonable accuracies to move forward with and possibly predict fights in the future especially using the Neural |Network model that was created.

7.0 Further Development or Research

With additional time to gather more knowledge on the subject and gather more resources, the first point of action would be to include more machine learning algorithms to compare against the Neural Network and the Random Forest model that were created here. This would give the chance to gather an even higher top accuracy!

Another thing would be to somehow gather extra fight data since there was no useful data that was available to be used by the public. This would enhance the whole project by having extra data to work with.

Extra hyperparameter tuning would also be a high priority to work on especially for the Random Forest to try make it even better than the state it is currently in. it would also be beneficial learning about hyperparameters for other machine learning algorithms that could potentially be used in the future!

Adding even more variety to different sections would also help gathering more knowledge regarding this subject. For example, with the Random Forest model, a normal grid search was implemented to find the learning rate to be used within the model. Expanding the knowledge here by using a random grid search instead of just the normal grid search would help learn more techniques to use.

Finally, gathering data for an upcoming fight card and all the fighters on it could be quite interesting to test on the models created. Predicting the winners before the fights and comparing it to the actual results of the fights to gain an accuracy would be intriguing.

8.0 References

[1] - Wilkens, S., 2021. *Sports prediction and betting models in the machine learning age: The case of tennis,* s.l.: Journal of Sports Analytics,.

[2] - Hitkul, K. Y. N. a. D. M., 2018. A Comparative Study of Machine Learning Algorithms for Prior Prediction of UFC Fights. *Harmony Search and Nature Inspired Optimization Algorithms Advances in Intelligent Systems and Computing,* pp. 67-76.

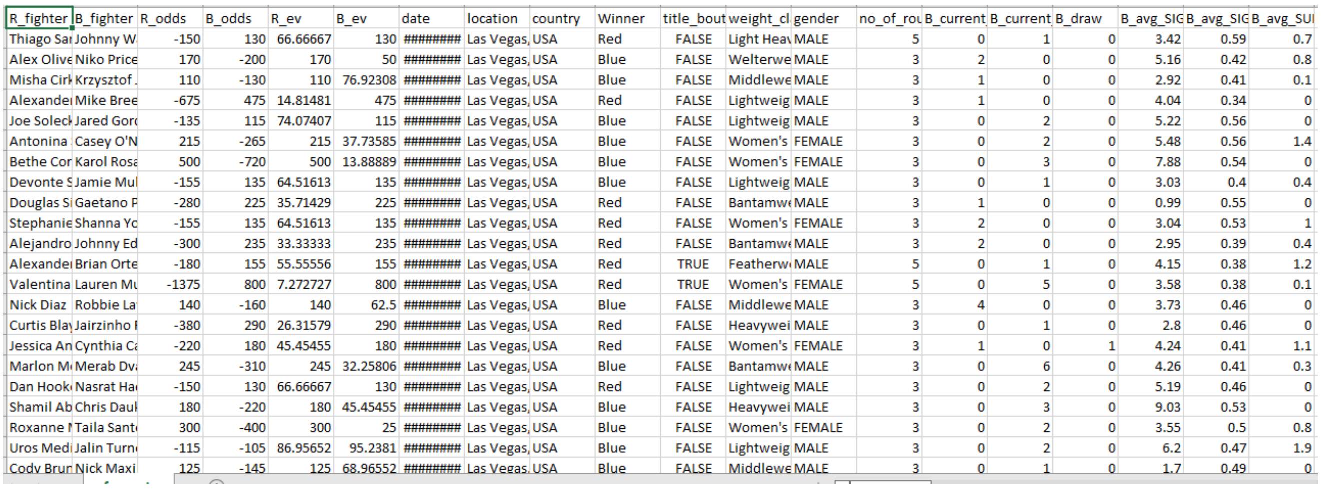
[3] - Anon., 2022. *VerdictMMA.* [Online]
Available at: https://verdictmma.com/

[4] - Education, I. C., 2020. *Neural Networks.* [Online]
Available at: https://www.ibm.com/cloud/learn/neural-networks

[5] - Cinelli, L. &. C. G. &. L. M., 2018. Vessel Classification through Convolutional Neural Networks using Passive Sonar Spectrogram Images.

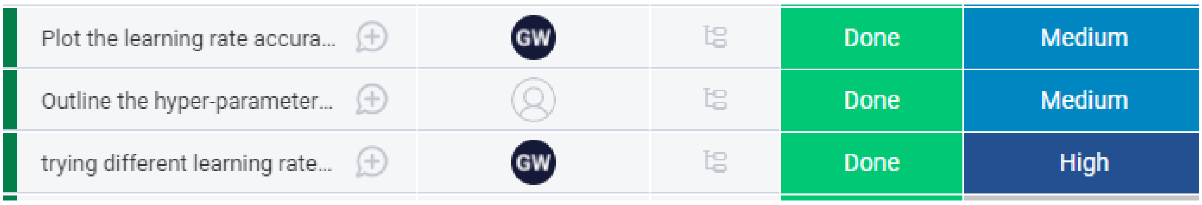
[6] - R, S. E., 2021. *Understanding Random Forest.* [Online]
Available at: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/#:~:text=Random%20forest%20is%20a%20Supervised,average%20in%20case%20of%20regression.

9.0 Appendices



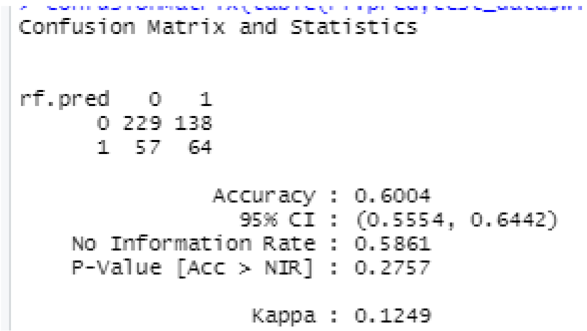
R_fighter	B_fighter	R_odds	B_odds	R_ev	B_ev	date	location	country	Winner	title_bout	weight_cl	gender	no_of_rou	B_current	B_current	B_draw	B_avg_SIG	B_avg_SIG	B_avg_SU
Thiago Sa	Johnny W	-150	130	66.66667	130	########	Las Vegas,	USA	Red	FALSE	Light Heav	MALE	5	0	1	0	3.42	0.59	0.7
Alex Olive	Niko Price	170	-200	170	50	########	Las Vegas,	USA	Blue	FALSE	Welterwe	MALE	3	2	0	0	5.16	0.42	0.8
Misha Cirk	Krzysztof	110	-130	110	76.92308	########	Las Vegas,	USA	Blue	FALSE	Middlewe	MALE	3	1	0	0	2.92	0.41	0.1
Alexande	Mike Bree	-675	475	14.81481	475	########	Las Vegas,	USA	Red	FALSE	Lightweig	MALE	3	1	0	0	4.04	0.34	0
Joe Soleck	Jared Gor	-135	115	74.07407	115	########	Las Vegas,	USA	Blue	FALSE	Lightweig	MALE	3	0	2	0	5.22	0.56	0
Antonina	Casey O'N	215	-265	215	37.73585	########	Las Vegas,	USA	Blue	FALSE	Women's	FEMALE	3	0	2	0	5.48	0.56	1.4
Bethe Cor	Karol Rosa	500	-720	500	13.88889	########	Las Vegas,	USA	Blue	FALSE	Women's	FEMALE	3	0	3	0	7.88	0.54	0
Devonte S	Jamie Mul	-155	135	64.51613	135	########	Las Vegas,	USA	Blue	FALSE	Lightweig	MALE	3	0	1	0	3.03	0.4	0.4
Douglas Si	Gaetano P	-280	225	35.71429	225	########	Las Vegas,	USA	Red	FALSE	Bantamwe	MALE	3	1	0	0	0.99	0.55	0
Stephanie	Shanna Yo	-155	135	64.51613	135	########	Las Vegas,	USA	Red	FALSE	Women's	FEMALE	3	2	0	0	3.04	0.53	1
Alejandro	Johnny Ed	-300	235	33.33333	235	########	Las Vegas,	USA	Red	FALSE	Bantamwe	MALE	3	2	0	0	2.95	0.39	0.4
Alexande	Brian Orte	-180	155	55.55556	155	########	Las Vegas,	USA	Red	TRUE	Featherw	MALE	5	0	1	0	4.15	0.38	1.2
Valentina	Lauren Mu	-1375	800	7.272727	800	########	Las Vegas,	USA	Red	TRUE	Women's	FEMALE	5	0	5	0	3.58	0.38	0.1
Nick Diaz	Robbie La	140	-160	140	62.5	########	Las Vegas,	USA	Blue	FALSE	Middlewe	MALE	3	4	0	0	3.73	0.46	0
Curtis Blay	Jairzinho R	-380	290	26.31579	290	########	Las Vegas,	USA	Red	FALSE	Heavywei	MALE	3	0	1	0	2.8	0.46	0
Jessica An	Cynthia Ca	-220	180	45.45455	180	########	Las Vegas,	USA	Red	FALSE	Women's	FEMALE	3	1	0	1	4.24	0.41	1.1
Marlon Mo	Merab Dva	245	-310	245	32.25806	########	Las Vegas,	USA	Blue	FALSE	Bantamwe	MALE	3	0	6	0	4.26	0.41	0.3
Dan Hook	Nasrat Ha	-150	130	66.66667	130	########	Las Vegas,	USA	Red	FALSE	Lightweig	MALE	3	0	2	0	5.19	0.46	0
Shamil Ab	Chris Daul	180	-220	180	45.45455	########	Las Vegas,	USA	Blue	FALSE	Heavywei	MALE	3	0	3	0	9.03	0.53	0
Roxanne M	Taila Sant	300	-400	300	25	########	Las Vegas,	USA	Blue	FALSE	Women's	FEMALE	3	0	2	0	3.55	0.5	0.8
Uros Medi	Jalin Turn	-115	-105	86.95652	95.2381	########	Las Vegas,	USA	Blue	FALSE	Lightweig	MALE	3	0	2	0	6.2	0.47	1.9
Cody Brun	Nick Maxi	125	-145	125	68.96552	########	Las Vegas,	USA	Blue	FALSE	Middlewe	MALE	3	0	1	0	1.7	0.49	0

Fig 13 – Original UFC Dataset layout in Excel



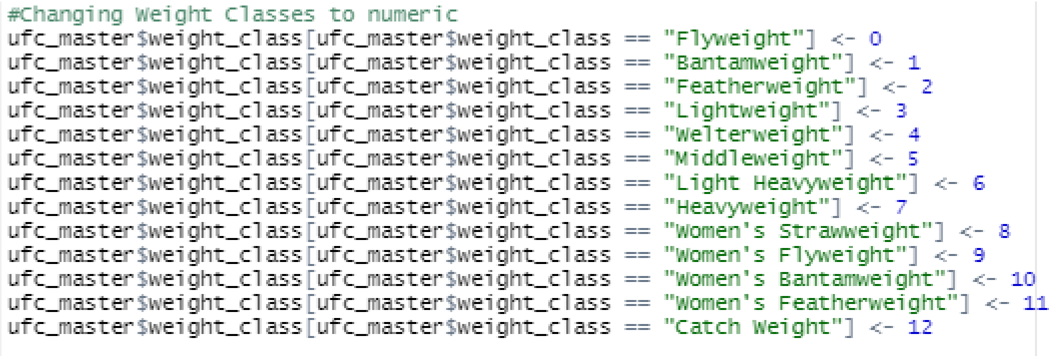
Plot the learning rate accura...	GW	Done	Medium
Outline the hyper-parameter...		Done	Medium
trying different learning rate...	GW	Done	High

Fig 14 – Example of Kanban Board used to hold tasks needed to complete for project



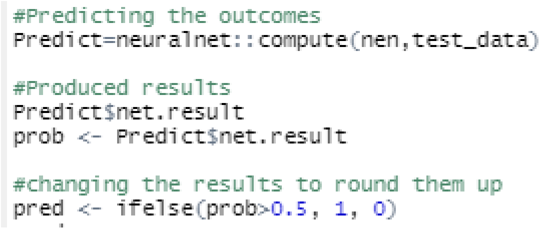
```
Confusion Matrix and Statistics


rf.pred   0   1
      0 229 138
      1  57  64

               Accuracy : 0.6004
                 95% CI : (0.5554, 0.6442)
    No Information Rate : 0.5861
    P-Value [Acc > NIR] : 0.2757

                  Kappa : 0.1249
```

Fig 15 - Example of Confusion Matrix for Random Forest model



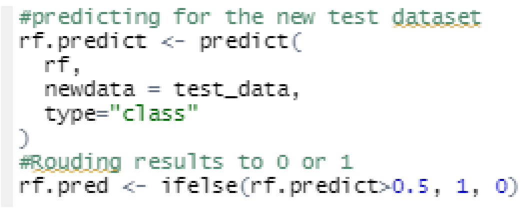
```
#Changing Weight Classes to numeric
ufc_master$weight_class[ufc_master$weight_class == "Flyweight"] <- 0
ufc_master$weight_class[ufc_master$weight_class == "Bantamweight"] <- 1
ufc_master$weight_class[ufc_master$weight_class == "Featherweight"] <- 2
ufc_master$weight_class[ufc_master$weight_class == "Lightweight"] <- 3
ufc_master$weight_class[ufc_master$weight_class == "Welterweight"] <- 4
ufc_master$weight_class[ufc_master$weight_class == "Middleweight"] <- 5
ufc_master$weight_class[ufc_master$weight_class == "Light Heavyweight"] <- 6
ufc_master$weight_class[ufc_master$weight_class == "Heavyweight"] <- 7
ufc_master$weight_class[ufc_master$weight_class == "Women's Strawweight"] <- 8
ufc_master$weight_class[ufc_master$weight_class == "Women's Flyweight"] <- 9
ufc_master$weight_class[ufc_master$weight_class == "Women's Bantamweight"] <- 10
ufc_master$weight_class[ufc_master$weight_class == "Women's Featherweight"] <- 11
ufc_master$weight_class[ufc_master$weight_class == "Catch Weight"] <- 12
```

Fig 16 – Example of how data was transformed



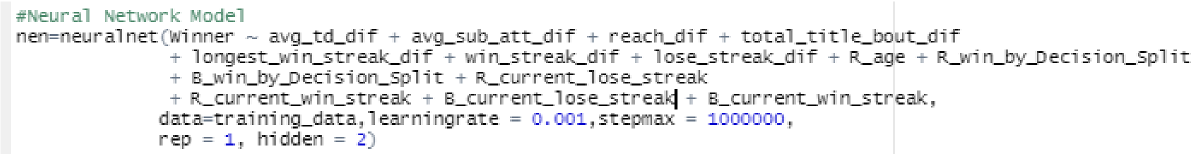
```
#Predicting the outcomes
Predict=neuralnet::compute(nen,test_data)

#Produced results
Predict$net.result
prob <- Predict$net.result

#changing the results to round them up
pred <- ifelse(prob>0.5, 1, 0)
```

Fig 17 – Predicting/Rounding Up NN results



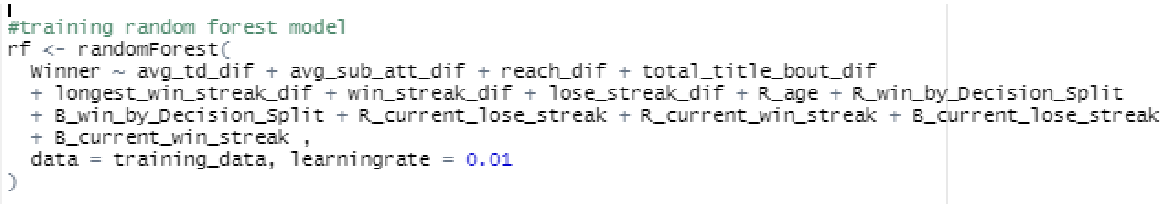
```
#predicting for the new test dataset
rf.predict <- predict(
  rf,
  newdata = test_data,
  type="class"
)
#Rouding results to 0 or 1
rf.pred <- ifelse(rf.predict>0.5, 1, 0)
```

Fig 18 – Predicting/Rounding Up RF results



```
#Neural Network Model
nen=neuralnet(Winner ~ avg_td_dif + avg_sub_att_dif + reach_dif + total_title_bout_dif
               + longest_win_streak_dif + win_streak_dif + lose_streak_dif + R_age + R_win_by_Decision_Split
               + B_win_by_Decision_Split + R_current_lose_streak
               + R_current_win_streak + B_current_lose_streak + B_current_win_streak,
              data=training_data,learningrate = 0.001,stepmax = 1000000,
              rep = 1, hidden = 2)
```

Fig 19 – Neural Network Model



```
#training random forest model
rf <- randomForest(
  Winner ~ avg_td_dif + avg_sub_att_dif + reach_dif + total_title_bout_dif
  + longest_win_streak_dif + win_streak_dif + lose_streak_dif + R_age + R_win_by_Decision_Split
  + B_win_by_Decision_Split + R_current_lose_streak + R_current_win_streak + B_current_lose_streak
  + B_current_win_streak ,
  data = training_data, learningrate = 0.01
)
```

Fig 20 – Random Forest Model

9.1. Project Proposal

9.1.1. Objectives

My main objective for this project would be to analyse mixed martial arts fights data and use that data to be able to predict the outcomes of fights. With UFC (Ultimate Fighting Championship) being the main hub of mixed martial arts fight in the world, I will be using data from UFC fights. For this project, I will need to find some detailed datasets on previous UFC fights to base the predictions on. These datasets should aid me in answering the questions such as:

- Who will win a certain fight?
- How will the fight end?
- Does weight make a difference?
- Does age come into play?

A side objective for this project would be that I want to enhance my knowledge in this field and I am sure this will definitely contribute to that! As a lot of this will be new to me, it will be challenging but it will also be very beneficial for me. I will be learning on the go with the help of other modules too which can be implemented into my project.

9.1.2. Background

As an avid fan of mixed martial arts, when I was brainstorming ideas for this project, I wanted to choose something that I was interested in and thought that this was a very good path to choose as it will keep me fully motivated to keep working on something I enjoy. Just looking at the UFC, there is a lot of variety present. There are 12 different weight classes, many different fighting styles and many different outcomes such as a knockout, a submission, draw or else a decision that can happen in a fight. I think this variety aids the project as there are many ways to angle this idea. Along with this being a passion of mine, predictions are actually a big deal within the MMA community ranging from the martial arts media to current fighters predicting how their peers' fights will end. It is one of the fun aspects within the build-up of a fight night. Everyone in the community has an opinion. I've always thought what would be the best way to predict outcomes of fights and this seems like the best way! To tackle this project, I need to do extensive research on sports analytics, especially in the field of 1v1 sports similar to mixed martial arts to help me get a better understanding from others in regards to my project.

9.1.3. State of the Art

In regards to similar work that has been carried out by others; I am currently in the process of researching this to see what is out there that is somewhat comparable to what I'm setting out to achieve with this project.

One I found that was based around the UFC was someone who was analysing certain strikes and moves and how often they are thrown or landed in different rounds. For example, it would be analysing how often takedowns are attempted in certain rounds. This information could be used to predict takedowns in the future. There doesn't seem to be many projects centred around this as I don't think the sport is as popular as different sports such as football for example.

While researching, I also saw another one which was a predictor but it had little to no detail. What makes my idea stand out and makes it different is that I want to aim to go into more details than others that have been done and use the data to get the best results possible!

9.1.4. Data

In terms of the data that is required for working towards my objectives with this project, I need to find some good datasets that contain many, many fights that have been previously contested within the UFC.

To achieve this, I will do my research within many of the top dataset websites such as Kaggle and Google Datasets to seek out the data that is required to make my project as successful as possible. I hope to find datasets to contain all the relevant information that I will need. For example, some of the data that may be required for me will be things such as which fighter won the fight, the weights of the fighters competing, how the fight ended, the age of fighters etc.

if the data is spread out between different datasets, I can clean it up and take all the relevant data from each of the datasets and combine them together to create subsets that are useful for my project. This is going to be one of the most important aspects of my project as without good and useful data, the project basically crumbles. I need to make sure I put a lot of time into the data research and gathering aspect of the project to make sure that I have the best data available to help me succeed. This should take up a good amount of my time within my project.

9.1.5. Methodology & Analysis

After doing some research on other projects when it comes to predictions, the methodology that I will aim to follow will be the KDD (Knowledge Discovery in Databases) method. This is the method of transforming your presented database and finding patterns that can be useful in your analysis. I think this method will help my project as its more focused on the research side in comparison to the CRISP-DM method.

All the steps within the KDD method will aid my project. These steps are:

- Data Selection and Integration - main aspect of this project as without the right data, it can’t go ahead.
- Cleaning of Data and Pre-processing – This step will fix errors in the data such as null values, low quality data or data in wrong columns etc
- Transformation of Data – Prepares data to be fed to the data mining algorithms.
- Data Mining – This is the main step. Algorithms are used to find useful patterns from the data
- Evaluation of patterns – Once the patterns are found, you can display them in charts or graphs.

I believe if I follow this methodology and these steps I have stated, I will be able to make great progress on my project. I hope to learn more about data mining from other modules to help me with this!

I think creating charts can be key for the analysis of the project as it will display the results I have found and also it will make it easier to analyse the data when it is put on charts.

9.1.6. Technical Details

In this project I plan on implementing machine learning into my project. There are many different machine learning algorithms that can be used. For example, there is the Decision Tree algorithm and also the Random Forests algorithm. This will be used to give predictions in relation to the data provided.

With Machine Learning, I can split the datasets into 3 pieces:

- Training Data: This is used to train whichever algorithms are chosen.
- Testing Data: Views the performance of the result.
- Validation Data: This is left to the end to validate the results.

Technologies that should be used are:

RStudio - Used for statistical computing with the R language.

SPSS – This can aid me with a more in-depth statistical analysis.

Tableau – I can view my datasets within Tableau. There are tools available that also may help.

Data Mining – Used to find patterns within the data.

Kaggle, Google Datasets – To find data for this project.

9.1.7. Project Plan

Project Tasks	Task	October	November	December	January	February	March	April	May
Initial Tasks	• Brainstorming • Project Pitch • Feedback	🟦							
Research	• Project-Proposal • Similar Project Research • Dataset Research • Mid-Point Presentation		🟦	🟦	🟦				
Data Gathering	• Gathering Data needed				🟩	🟩	🟩		
Developing Data	• Cleaning data • Transform Data					🟪	🟪		

	• Represent Data					🟪	🟪		
Analysis	• Analysing Data that was gathered						🟥	🟥	
Reporting	• Results • Publishing • Presenting • Documentation • Showcase							🟧	🟧

As of right now, this is my first draft of my project plan. Although things may vary, I will try to stick to this. At this point I am currently in the research phase of the plan. I am currently analysing other projects and also analysing this type of analysis which is a 1v1 sport.

9.2. Reflective Journals

Student Name	Gavin Walsh
Student Number	X17364783
Course	BSc in Computing, specialising in Data Analytics

Month: October

What?

This month was all about getting back into the swing of things with college and trying to deal with all modules in the most efficient way possible as its going to be a very busy semester. This month for my project, I have brainstormed ideas that I could use for my project. I submitted my final idea which was in video form explaining the details of my project. I also just got assigned my supervisor for the project too. I have also begun working on my Project Proposal.

So What?

This was pivotal for my project as this is the starting point and starting base to the rest of the project. Without an idea, this project can’t go ahead. In regards to the challenges, I found it very tough to come up with an idea for this project. It took some time before something clicked for me as I was keen on building my project around the area of sport but I wasn't sure how to approach it. As far as successes go, I came to the conclusion that I wanted my project to be based around Mixed Martial Arts data. My project idea was accepted with amendments requested.

Now What?

To address this, I will have to wait to receive information about the amendments with my project idea so I can proceed with my project. I will also now have to keep steady progress on my Project Proposal so that I get it completed comfortably before the deadline.

Student Signature	Gavin Walsh

Student Name	Gavin Walsh
Student Number	X17364783
Course	BSc in Computing

Month: November

What?

A lot has happened in this month. It started off with my project being accepted with amendments. I needed to make the vision for my project a little clearer with datasets etc. I then got assigned my supervisor which is William Clifford. We have had multiple meetings discussing my project and talking about the direction of my project. I then completed my project proposal for my project. My work so far has mainly consisted of research. I found a suitable dataset that I can use for my MMA project which contains some needed attributes. Along with this, I was doing some research on other projects that are similar to this which includes 1v1 sports such as tennis. I also submitted my ethics form for my dataset.

So What? What this means for my project is that as soon as it got accepted and figured out the amendments, I could crack on with the project proposal for a rough draft of how my project should go. Following this I could then start my research process which is vital for my project. I was very happy with how things turned out this month as I got my proposal done and started some initial research along with meeting with my supervisor! Although there is still a lot of research to do, a lot has been done up to this point!
Now What? What can you do to address outstanding challenges? For my next research, I want to use avenues such as Google Scholar to view some high-level papers on projects that are similar to mine to see how it is done and see if there are any bits I can take for my own project. Along with this, I'd like to get my dataset into RStudio and become more familiar with it.

Student Signature	Gavin Walsh

Student Name	Gavin Walsh
Student Number	X17364783
Course	BSc in Computing

Month: December

What?

This month with my project started off with more in depth into research papers on projects that are similar to mine. This came in the form of 1v1 sports prediction models such as tennis. These were interesting to read and to view how these projects come together although it is a very tough read at first! Next for me was working on my dataset. I cut out some variables first, then I started to transform my data more to make it easier for me. I did this by changing some of the possibly usable string data into int data. This was variables such as winners, genders, stances and weight classes. Following this I began to look at the different algorithms to implement into my project. After researching different papers with them etc, I came to the conclusion I wanted to try implement a Neural Network into my project. I would like to better other people's results that I have seen!

So What?

December was a big month for my project as it got my dataset into a good state and also a lot of research was done! Although it was challenging to get a good grasp of the research papers that I had found. I'm now at the stage where I can work with the algorithm which will be challenging for me to implement as it will be my first time!

Now What?

To address outstanding challenges, I will do continuous research regarding Neural Networks and use it on dummy data to make sure I understand it to implement it into my project. This will be tough but it will also be good to learn something completely fresh too!

Student Signature	Gavin Walsh

Student Name	Gavin Walsh
Student Number	X17364783
Course	BSc in Computing

Month: January

What?

As we just came off the Holiday and exams period, I didn’t manage to get too much done this month. I did some more research into Neural Networks and tried to implement it on a dataset. Unfortunately I was running into errors so I still need to keep trying to implement the Neural Network.

So What?

What this means for my project is that I need to put in some extra time into the project to cover for the time missed over the break. This will help me complete my project with hopefully time to spare and hopefully not be in a complete rush at deadlines!

Now What?

Next, I will have to keep at it with implementing the Neural Network algorithm so that I get a full understanding of it and also a fully working Neural Network. Once I complete this, then I can move onto whats next.

Student Signature	Gavin Walsh

Student Name	Gavin Walsh
Student Number	X17364783
Course	BSc in Computing

Month: February

What?

This month, I began searching for other data that I could incorporate into my data set but most web sites that could be used for scraping such as the UFC website didn't allow it so I couldn't use it. I made considerable progress on my neural network as I got it implemented on a basic level and using different attributes to see how different it looks. I managed to write a draft of the first few sections of the final report so that I'm not in a rush towards the deadline of the report.

So What?

I'm content with how this month turned out as I surpassed an error with my neural network that was holding me back while I made progress on the final report too.

Now What?

Next up for me is to work on the output of my neural network. Some more problems have arisen but hopefully I can find solutions quickly so I can progress further. Also, I'd like to make more ground on the final report draft this month.

Student Signature	Gavin Walsh

Student Name	Gavin Walsh
Student Number	X17364783
Course	BSc in Computing specialising in Data Analytics

Month: March

What?

This month, I worked more on my neural network implementation. I started adding in a learning rate to enhance the results. While working on this, I noticed when I ran the neural network, it would take an extra long time to process. This may have been due to some variable figures being larger than others. Due to this, I normalised some of my data to make it easier to read. I found the max figure within the variable then divided the rest of the column by the max number. Along with this, I continued working on the first draft for the report and cleaning up some of the previous sections!

So What?

A success was making more progress on the draft. This will help in the long run as to not having to rush the final documentation. Also, the normalisation was a success as it enhanced my data set. The output of the neural network is still a challenge for me but it will hopefully be resolved soon!

Now What? What can you do to address outstanding challenges? Now, I will continue working on my neural network to make it as good as it can be. More research will be conducted for this. I will also try to get more sections of the draft completed to give me even more breathing room come the end of the project cycle!

Student Signature	Gavin Walsh

Student Name	Gavin Walsh
Student Number	X17364783
Course	BSc in Computing specialising in Data Analytics

Month: April

What? As it is approaching the finish line, a lot has to be done! This month I began to look at ways to enhance the project. At first, I tried to create a loop to show me the best learning rate hyperparameter for my neural network implementation but I couldn't solve that problem. So, I decided to instead incorporate a grid search into my project to show me the best learning rate to use. This was implemented using the h2o package. This allowed me to see which learning rate to include. More work was conducted on the report also. I have also been working on turning my Winner variable into a single node as the results that are being produced give me two different figures that split the possibilities of the winner of the fight.
So What?

There was good progress especially finding the best learning rate to produce the best results but there is also stuff to be worked on. Its good to have more of the report worked on too as it wont be the biggest rush come the final days of the project.

Now What?

What can you do to address outstanding challenges?

Next, I will continue working away to fix the node problem with my results and also may have to gather a backup plan if it doesn’t work out correctly. Some more fine tuning with the parameters can be used with also more work on the report!

Student Signature	Gavin Walsh