BSc (Hons) Computer Science

An Evaluation of Machine Learning algorithms and Data Analysis in UFC Fight prediction.

Joshua Coyle
B00400951

4h April 2025

Supervisor: Sean Sturley

Declaration

This dissertation is submitted in partial fulfillment of the requirements for the degree of [Programme of Study] (Honours) in the University of the West of Scotland.

I declare that this dissertation embodies the results of my own work and that it has been composed by myself. Following normal academic conventions, I have made due acknowledgement to the work of others.

Name: JOSHUA COYLE

Signature: Joshua Coyle

Date: 02/04/2025

COMPUTING HONOURS PROJECT SPECIFICATION FORM

(Electronic copy available on the Aula Computing Hons Project Site)

Project Title: Can Machine Learning models improve safety when cutting weight for an MMA fight.

Student: Joshua Coyle **Banner ID:** B00400951

Programme of Study: BSc (Hons) in Computing Science

Supervisor: Sean Sturley

Moderator: Raman Singh

Outline of Project:

The aim of my project is to do an in-depth analysis of the implications of using machine learning in MMA fight prediction, specifically the Ultimate Fighting Championship (UFC). The UFC is a Mixed Martial Arts organization where fighters from all over the world compete to find who is the best above all. Using machine learning and data analysis, this project aims to explore different machine learning models to find which is best suited for a fight prediction algorithm.

MMA is a dynamic sport with a wide range of factors that influence the outcome of a fight. Using data such as fighter attributes and statistics can enable for training camp optimisation through data analysis and fight predictions. As the sports betting industry grows in popularity, the demand for leveraging data science and machine learning for accurate predictions is profound.

This project will look at studies where data science and machine learning has been used previously in sporting environments. The UFC prediction model will primarily use a dataset which holds every fighters' statistics to make predictions and compare features. Using different techniques, the paper will find the optimal machine learning algorithm which will then be tested through a deep model evaluation.

A Passable Project will:

- Review literature on machine learning in sports.
- Develop a machine learning model which can make predictions using fighter statistics.
- Make use of libraries within the chosen Python environment to delve into data analytics.

A First Class Project will:

- Display a deep analysis of machine learning methods, explaining the maths behind each one and the practicality relating to the chosen subject. This will provide a decision which one is most optimal for the practical design.
- A thorough analysis of the predictions made by the machine learning model, comparing accuracy metrics to real life examples, to get a functional gauge on how well the model performs.
- Justify the project, using real world implications.
- Conclude the paper with recommendations for further development and an evaluation of the performance relating to the different metrics of accuracy.

Resources Required: *(hardware/software/other)*

Visual Studio Code
Python Programming Language
Machine Learning Libraries for Python
Datasets to train and test the machine learning model
GitHub to display code

Marking Scheme:

		Marks	
e.g.	Introduction	15%	
	Literature Review	20%	
	Methodology	20%	
	Development of model	10%	
	Evaluation	20%	..
	Conclusion	10%	
	Critical Self-Appraisal	5%	

AGREED:

Student	**Supervisor**	**Moderator**
Name: Joshua Coyle	**Name:** Sean Sturley	**Name:** Raman Singh

IMPORTANT:

(i) ***By agreeing to this form all parties are confirming that the proposed Hons Project will include the student undertaking practical work of some sort using computing technology / IT, most frequently achieved by the creation of an artefact as the focus for covering all or part of an implementation life-cycle.***

(ii) ***By agreeing to this form all parties are confirming that any potential ethical issues have been considered and if human participants are involved in the proposed Hons Project then ethical approval will be sought through approved mechanisms of the School of CEPS Ethics Committee.***

Table of Contents

1.0 Abstract 8
2.0 Introduction 9
2.1 History of the UFC 9
2.2 Role of Machine Learning 9
2.3 Predicting UFC Fights 9
2.4 Key Factors in Declaring a Victor in a UFC Bout 10
2.4.1 Scoring a UFC Fight 10
2.4.2 Betting Odds as predictive tools 10
2.5 Justification 11
2.6 Objectives 12
4.0 Literature Review 13
4.1 Machine Learning 13
4.2 Machine Learning Models 13
4.2.1 Supervised Learning 13
4.2.2 Unsupervised Learning 13
4.3 Overview of proposed supervised learning models 14
4.3.1 Logistic Regression 14
4.3.2 Decision Trees 15
4.3.3 Artificial Neural Network 16
4.3.4 Support Vector Machine 17
4.4 Machine Learning in sports 19
4.4.1 Profiling of young boxers using unsupervised machine learning 19
4.4.2 Machine Learning Predictions for Football Match Outcomes 19
4.5 Difficulties in predicting UFC bouts 21
4.5.1 Uncertainty 22
5.0 Methodology 23
5.1 Introduction 23
5.2 Datasets 23
5.2.1 Rationale for datasets 25
5.3 Development 26
5.3.1 Libraries 26
5.3.2 Data Pre-Processing 26
5.4 Model Selection 30
5.5 Feature Engineering 32
6.0 Model Evaluation 35

7.0 Conclusion and Discussion ... 49
7.1 Introduction ... 49
7.1 Model Accuracy ... 49
7.2 Potential Improvements ... 49
7.3 Conclusion ... 50
8.0 Critical Self Appraisal ... 51
9.0 References ... 52
10.0 Appendix ... 55

1.0 Abstract

With the rise of MMA as a sport and the UFC as an organization, predicting UFC bouts is all the talk leading up to a large UFC card. Just like other sports, MMA plays a role in the betting industry. Predicting UFC fights can be important to determine which fighter is the favourite to win prior to the match. This dissertation analyses the use of machine learning methods to make predictions on UFC bouts using a dataset consisting of fighter's statistics such as significant strikes, takedowns, fight records and many more. This paper explores the different machine learning models such as Random forests, Support Vector Machines (SVM), Logistic Regression and Artificial Neural Networks. Each model will have an in-depth analysis with advantages and disadvantages for each relating to the dataset.

Comparing the different algorithms based on two datasets, one containing UFC fighter statistics (4,000 records) and the other containing previous UFC fights (6,000 records), it was found using a k-fold cross validation that a logistic regression model was most suitable to make the predictions. Using scoring techniques which are commonly used in the UFC to judge fights, each feature was given a weight respective to how much they influence the outcome of a fight.

Once the data was pre-processed and the model was trained, the results were compared to fifty recent UFC bouts. The fights would be simulated through the model where the predictions made could be compared to the actual outcome of each fight. The logistic regression algorithm returned an accuracy rate of 70%, closely reflecting its F1 score of 0.7.

The model returns a confidence score which represents the probability that the predicted winner would win. The model's performance was further validated using 18 fights from recent UFC cards, their opening betting odds for each fighter were used to establish an accuracy rating on the model's confidence score. Comparing the model's confidence to the opening odds, the results showed an average difference to be 11.3%, giving a similar prediction to the experts within the industry.

2.0 Introduction

The code of this project can be found at:
https://github.com/joshcoyle02/UFC-Fight-Predictor

2.1 History of the UFC

The UFC (Ultimate Fighting Championship) is the largest recognised mixed martial arts organization in the world (*History of UFC | UFC*, 2018), bringing athletes from across the globe that specialise in different forms of martial arts such as wrestling, jiu-jitsu, muay Thai, boxing and karate. The term Mixed martial arts (MMA) is exactly what it sounds like, a mixture of some if not all martial arts blended into the one sport. Although the sport of mixed martial arts was conceived in recent times, the inspiration that lead to its creation can be traced back to ancient Greece (*A brief history of Mixed Martial Arts.*, 2023)**.** As stated previously, MMA is a collection of all martial arts which lead to the founding of the UFC in 1993 by Art Davie, Rorion Gracie, John Milius, Bob Meyrowitz and David Isaacs (Nag, 2023). It was during this time that UFC 1 took place, which would signify the beginning of a future titan in the sporting world. On November 12th 1993, eight thousand fans would gather in the McNichols Sports Arena in Denver, Colorado to witness the first ever UFC event (Benson, 2019). UFC 1 consisted of eight fighters all of which had expertise in one specific area of martial arts who would compete in a tournament with only two rules: no biting and no eye gouging. It wasn't until January 2001 that Dana White, who is the current president of the UFC today, would purchase the company for $2 million alongside his two business partners Loronzo and Frank Fertitta.

2.2 Role of Machine Learning

Machine learning is a subset of Artificial Intelligence which uses algorithms to discover patterns and make predictions on a specific dataset. Unlike typical computer programming where instructions must be hard coded in for each component to work, Machine learning can learn from the dataset and become more "intelligent" over time.

2.3 Predicting UFC Fights

Due to the chaotic nature of MMA fights, it is tough to predict who the winner will be no matter what the statistics say. Although statistics play a huge role in match making and predicting which fighter will prevail, anything can happen within the cage due to the vast array of paths to victory. This has made machine learning very successful in sports like football and basketball, where an algorithm can be given a dataset which represents a team's statistics over the course of a season to then make predictions for future games and player performances.

2.4 Key Factors in Declaring a Victor in a UFC Bout

2.4.1 Scoring a UFC Fight

MMA is a savage sport with many athletes from all over the world who excel in one area of martial arts or be proficient in all. Due to the immense collection of different martial arts involved in MMA, there are many paths to victory. Most UFC fights are three rounds lasting five minutes each with main event and championship matches having five rounds. The UFC uses a 10 point round scoring system. Each round is scored independently in favour of one fighter, for example if Fighter A wins round one, then the score for the round would be 10-9 to fighter A if the opponent still fought competitively. If a fighter wins a round decisively, the score is sometimes 10-8 and if a fighter dominates the round, it can be scored 10-7 in rare cases. Once the fight has ended and all rounds have concluded, the score cards are added up to declare the winner. This means that if fighter A wins all three rounds by 10-9, the final scorecards will read 30-27 in favour of fighter A by unanimous decision.

2.4.2 Betting Odds as predictive tools

Betting odds use expert insight and statistical analysis of fighters to determine probabilities of a fighter winning. Due to the UFC being an American company, the odds are displayed in their native format. To understand betting odds and how they can help evaluate a machine learning model, it is best to look at an example:

In a recent match-up within the UFC, Alex Pereira faced Magomed Ankalaev. Using the opening odds, which represents the probabilities of each fighter winning before the bout takes place, we have -162 Ankalaev and +126 Pereira (*How To Bet On UFC Fights | Ultimate Guide to UFC & MMA Betting | OddsChecker*, no date). When looking at the betting odds, (-) signifies the favourite and (+) signifies the underdog. Using this method, it is a good way to compare the accuracy of the machine learning model in comparison to expert opinion. In the American betting market, if a $100 was placed on Ankaleav ro win with -162 odds, the bet would return $162 if he was the victor. Using these odds, the probability can be calculated to return a percentage which aligns more with machine learning practices, something that will be expanded upon within the evaluation chapter in this paper.

2.5 Justification

In 2023 it was reported by the American Gaming Association that the total spent on sports betting reached $119.8 billion in the United States which is an increase of 22.9% from the previous year (Fisher, 2024).

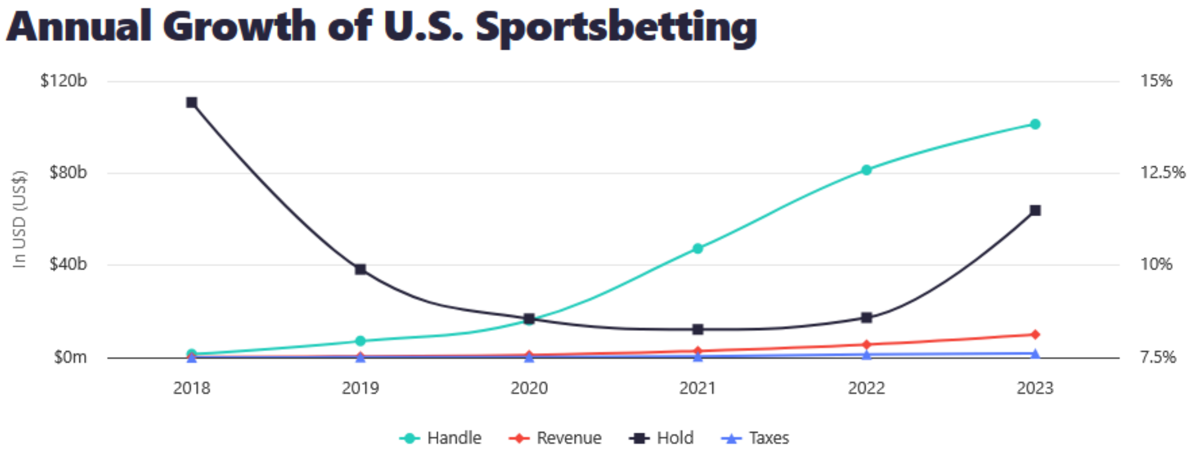


Figure 1. Growth in sports betting plotted to a graph

The surge in sports betting conveys growing financial stakes and interest in sports prediction. With the exponential growth in gambling, bettors are increasingly reliant on statistics to make accurate predictions on upcoming events. Knowing how arbitrary UFC outcomes can be at times, accurate statistical predictions can improve the

Accurate predictions have a massive impact on fighter strategies, performance and matchmaking. Using data analytics can give an insight into fighters' statistics which can help training methods. Fighters' will be able to identify their strengths and weaknesses on a statistical level, with visualisations that represent the disparity in comparison to their opponents. This allows for training camps to be optimised for specific fights, which could result in immense improvement within specific areas of a fighter's skill set that may have otherwise been overlooked.

2.6 Objectives

- This paper will analyse different machine learning techniques that are most suitable to predict UFC matches. Due to the result of a UFC match being either win or lose, a classification model will be most optimal for a machine learning algorithm of this nature as the output will always be one of these. Despite there being the occasional draw in the UFC, it is very rare and hard to account for.

- Once each suggested classification model has been analysed, this paper will give sufficient reasoning to choose a specific one to build the algorithm. Each model will be compared with advantages and disadvantages which relate to the proposed idea.

- Once a machine learning algorithm has been selected, the methodology chapter will explain in detail how the model was created, which data has been used and how it was prepared for the predictions.

4.0 Literature Review

4.1 Machine Learning

The term "Machine Learning" can be traced back to the 1950's. Originally, Alan Turing, A mathematician and computer scientist from England came forward with the idea of computers being able to become more efficient at tasks that they are assigned to without the need for extra code through human intervention (Nilsen, 2018).

An algorithm that qualifies as machine learning can adapt and improve results through repetition using similar data to the testing dataset. A supervised learning model will be fed features (input variables) and labels (output variables) which will then run the data through the model while refining its parameters to increase accuracy of the prediction. Using the training data, the model will repeat this process for multiple iterations. Once the model has ran its course, its performance can be assessed to calculate the accuracy and error. Through repetitive training, a machine learning model can give accurate predictions based on patterns it has learned through analysing the data.

4.2 Machine Learning Models

There are multiple different types of machine learning algorithms, but this paper will analyse the most suitable one for the proposed dataset.

4.2.1 Supervised Learning

Supervised Machine Learning is a type of machine learning where the dataset that the model uses to train, and test has labels (R. China, 2023). A supervised learning algorithm has a dataset which contains known values for independent (input) and dependent (output) variables. Each independent value has a corresponding dependent value, meaning the model can analyse the data to recognise patterns and structure within the dataset. In the book "Machine Learning in Radiation Oncology" published in 2015, The authors El Naqa and M.J. Murphy describe an analogy that allows for simple understanding of supervised learning. Naqa and Murphy detailed a hypothetical dataset where the algorithm would be given features such as colour, size and smell which would be used to predict whether the output value was either an apple or an orange (El Naqa and Murphy, 2015). In essence, supervised learning attempts to mimic how humans make decisions. For example, referring to the orange and apples analogy described by Naqa and Murphy, as humans we identify something as an orange based on attributes like its colour, taste, texture and smell. We know this because of previous encounters with an orange, meaning we can categorize the features to the orange, which in machine learning terms is the labelled target output, based on the fact we know what an orange is already.

4.2.2 Unsupervised Learning

Unsupervised Learning is a type of machine learning algorithm that uses raw data with no labels to try and map data into clusters and find relationships or patterns (Ghahramani, 2004). Instead of giving a discrete value for a target variable, unsupervised learning simply groups similar datapoints together to visualize. Unsupervised learning can help spot outliers using a cluster graph. It is for this reason that unsupervised learning is used frequently in finance companies to detect fraudulent activity.

4.3 Overview of proposed supervised learning models

4.3.1 Logistic Regression

Logistic regression is a supervised machine learning algorithm that calculates the probability of the output variable belonging to a certain class. Logistic regression is a classification algorithm, meaning it classifies the output based on its probability to belong in one class or another (Kleinbaum and Klein, 2010). For example, using the orange and apples analogy the model would ingest input variables such as the taste, colour and texture to then calculate the probability that the output belongs to either class apples or oranges.

The equation for linear regression is:

$$\boldsymbol{z = a + b_1X_1 + b_2X_2 + \dots b_kX_k}$$

(Tuzsuz, no date)

z is the log odds, which the function is trying to calculate.

a represents the bias, which is the constant value of z when all the independent variables are equal to zero.

b_i represents the coefficients, which are also referred to as the **weights**. The coefficients determine how much each independent variable (X_i) contributes to the outcome.

X_i is the input features (independent variables) which are used to predict. Each instance of X_i represents a different feature. For example, relating to the prediction of UFC bouts, X_1 can be significant strikes landed and X_2 is the number of knockdowns.

Logistic regression uses the sigmoid function to calculate probabilities. The sigmoid function takes the linear input data and transforms it into a value between 0 and 1 which represents the probability.

The sigmoid function is typically presented with the formula:

$$\boldsymbol{\sigma(x) = \frac{1}{1 + e^{-x}}}$$

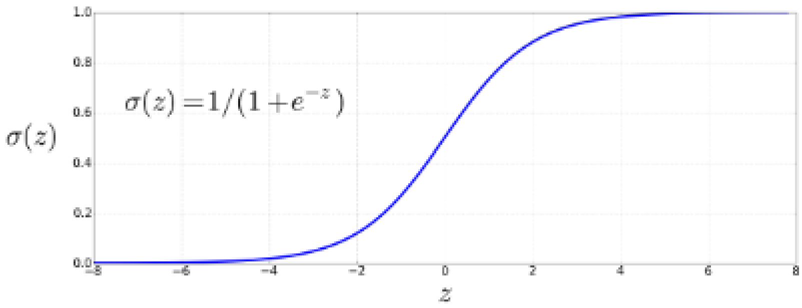
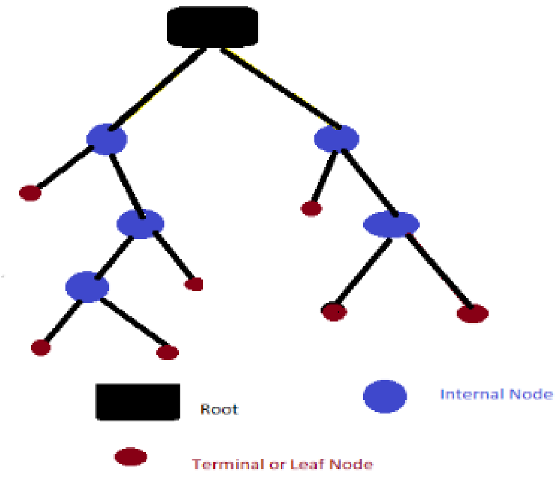


Figure 2. Sigmoid Function

4.3.2 Decision Trees

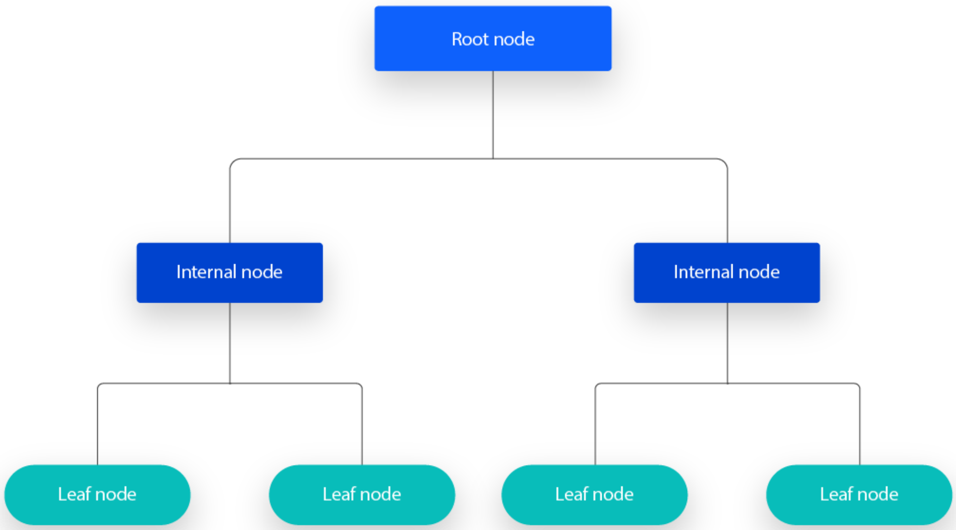
Decision Trees are a supervised learning method that involves splitting the data into different branches based on feature values to make predictions. The inspiration for decision trees came from a regular tree, where the structure is made up from branches and nodes where the nodes represent decisions and the branches represent outcomes (Ali *et al.*, 2012).



(Ali *et al.*, 2012).

Figure 3. Decision tree diagram

A decision tree starts with a root node, which is where the algorithm begins. The branches which extend from the root node lead to internal nodes which are also known as decision nodes where the data can be split further based on the feature values into leaf nodes. (IBM, 2021).



(IBM, 2021)

Figure 4. Decision tree structure

Entropy

The entropy of a decision tree is the measurement of impurity within the dataset. The value of the entropy represents the randomness of the data and is always a value between 0 and 1. The closer the value of entropy is to 0 the better, meaning the node is “purer”. If the value of entropy is 1, this means that the data within the node is split evenly. This means the decision tree cannot make an accurate binary classification prediction as the data is split 50/50 (Jijo, Abdulazeez, 2021).
The formula for entropy:

$$Entropy\ (p) = -\sum_{I=1}^{N} p_i\ log_2 p_i$$

In relation to the UFC fight predictor model, an example of a node with high entropy could be that both fighters have ten fights, each with five wins and five losses. Using this logic, the entropy would be 1, as the wins statistic would be even at a 50/50 split, with neither fighter holding an advantage over the other in the ‘wins’ statistic.

4.3.3 Artificial Neural Network

An artificial neural network (ANN) is a supervised machine learning model. Neural networks attempt to mimic the human brain, with the model consisting of neurons that transmit information to each other (Wang, 2003). These networks are composed of multiple layers which include an input, output and one or more hidden layers.

(Nagyfi, 2018)

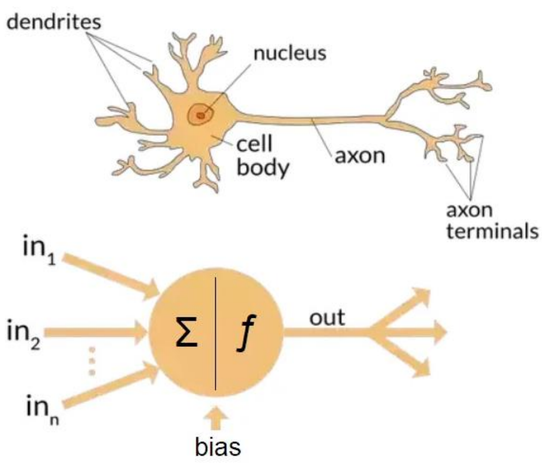


Figure 5. Biological neuron and Artificial Neuron

The diagram depicts an artificial neuron compared to a biological neuron. The artificial neuron takes an input just like the dendrite within the biological neuron, then passes it through the node which represents the neuron’s soma. Within the node the sum of the inputs is added together along with the bias term which is then passed through an activation function. In a biological neuron, if the signals generated from the dendrites meet the threshold, the neuron sends an electrical impulse through the axon to communicate with other neurons. Similarly, the activation function determines whether the weighted sum of the inputs including the bias should be fired onto the next neuron.

(Tóth, 2018)

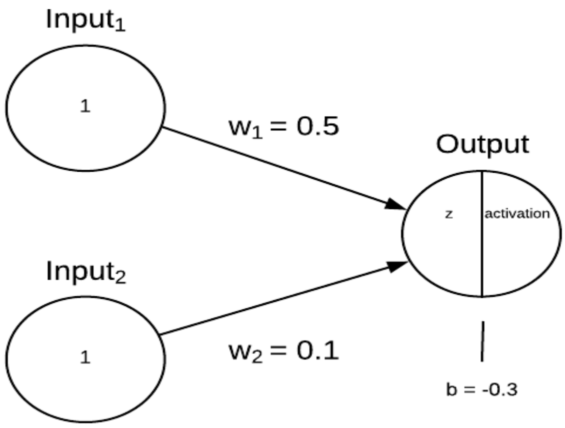


Figure 6. Artificial Neuron Structure

To visualise this concept, the image above depicts forward propagation, which is the method of flowing data through a neural network. The input one and two both represent the raw data being used to make the prediction. $\mathbf{Input_1}$ represents the significant strikes for fighter and $\mathbf{Input_2}$ represents the takedown defence. Due to the significant strikes being more important when predicting the outcome of a fight, it has a higher weight value ($\mathbf{w_1}$) than takedown defence ($\mathbf{w_2}$). Both of these are then multiplied by their weights and added together along with the bias: (Tóth, 2018)

$$\boldsymbol{Z = input_1 * w_1 + input_2 * w_2 + b}$$
$$\boldsymbol{Z = (1 * 0.5) + (1 * 0.1) - 0.3 = 0.3}$$

The result is then passed through an activation function (Sigmoid, tanh, ReLU, Softmax) to introduce non-linearity to ensure complex patterns are modelled sufficiently. The value of Z will then be passed onto the next hidden layer.

4.3.4 Support Vector Machine

A support vector machine (SVM) is a type of supervised machine learning that is commonly used for classification and regression techniques (Jakkula, no date). SVMS's, like other machine learning models, aim to maximise prediction accuracy while maintaining generalisation to avoid overfitting the data. Similar to artificial neural networks, SVMs are regularly used in pattern recognition for image classification and speech recognition. SVMs use kernels to handle different types of classification and regression tasks, enabling them to model both linear and non-linear relationships in a dataset.

.

Support Vector Machines use different kernels depending on the data. There are four that will be examined within the scope of this paper (Tiwari, 2025):

- Linear Kernel
- Polynomial Kernel
- Radial Basis Function Kernel (RBF)
- Sigmoid Kernel

Linear Kernel

The linear kernel is the simplest kernel used in support vector machines. A linear kernel is used when the data is linearly separable, meaning the decision boundary can be mapped by a straight line to divide the classes in a classification task.

The formula for the linear kernel: $\boldsymbol{K(x, y) = x.y}$

Polynomial Kernel

The polynomial kernel is used when data becomes more complex. To account for more nonlinear data, where the datapoints cannot be separated by a linear decision boundary, it maps the input features to a higher dimensional space where separation is possible.

The formula for the polynomial kernel: $\boldsymbol{K(x, y) = (x.y + c)^{d}}$

Radial Basis Function Kernel

An RBF kernel works similarly to a polynomial kernel, where it excels at mapping high dimensional data. Despite the similarities, an RBF kernel is used where the data is messy and has no structure to it.

The formula for the RBF kernel: $\boldsymbol{K(x, y) = e\text{^}(-\gamma\ ||x - y||^2)}$

Sigmoid Kernel

The sigmoid kernel shares similarities to a neural network in the way it processes data. Homogeneous to a neural network, the sigmoid kernel uses an activation function.

The formula for the sigmoid kernel: $\boldsymbol{K(x, y) = tanh(\gamma\ .x\text{^}T\ y + r)}$

Based on the dataset used in the prediction model, the RBF kernel would be best suited to handle the data. Due to the unpredictability of the sport, the RBF kernel would be efficient at mapping outliers within the dataset and handling the different statistics and their implications on predictions.

4.4 Machine Learning in sports

4.4.1 Profiling of young boxers using unsupervised machine learning

In August 2021, a study was conducted to analyse the attributes of young boxers to profile their physical performance (Ćenanović and Kevrić, 2023). The study utilises unsupervised machine learning models, specifically k-Medoids clustering, to categorize the athletes based on several variables. The research involved an analysis of young boxers with at least one year of training experience using attributes such as:

- Rear hand punch force
- Punch Velocity
- Maximal isometric handgrip strength
- Lower-limb muscle power

The goal of the study was to profile the young boxers using unsupervised machine learning algorithms. The research sought to classify the boxers based on the physical and functional characteristics. The results of the classification would allow the athletes to have an insight into different performance profiles, which can help optimise training regimes by identifying strengths and weaknesses.

Within the study, there were two groups of boxers that were split between higher body mass and handgrip strength and lesser body mass. It was found that one group, having higher body mass, was physically stronger but less explosive and the other group having higher power and explosiveness being lighter and not as strong.

Applying these methodologies to a UFC fight predictor can be useful in match making. To gain an insight into fighters' statistics can make for competitive fights, making the sport more balanced.

4.4.2 Machine Learning Predictions for Football Match Outcomes

Research conducted by Hengzhi Chen, a student from Bashu Secondary School in Chongqing China, looked at using machine learning models to predict the outcome of a football match. The study explored various machine learning algorithms and analysed how player statistics could be used as a metric to predict match results (Chen, 2019).
Chen uses data directly from the official FIFA website, which gives 34 different statistics for each player.

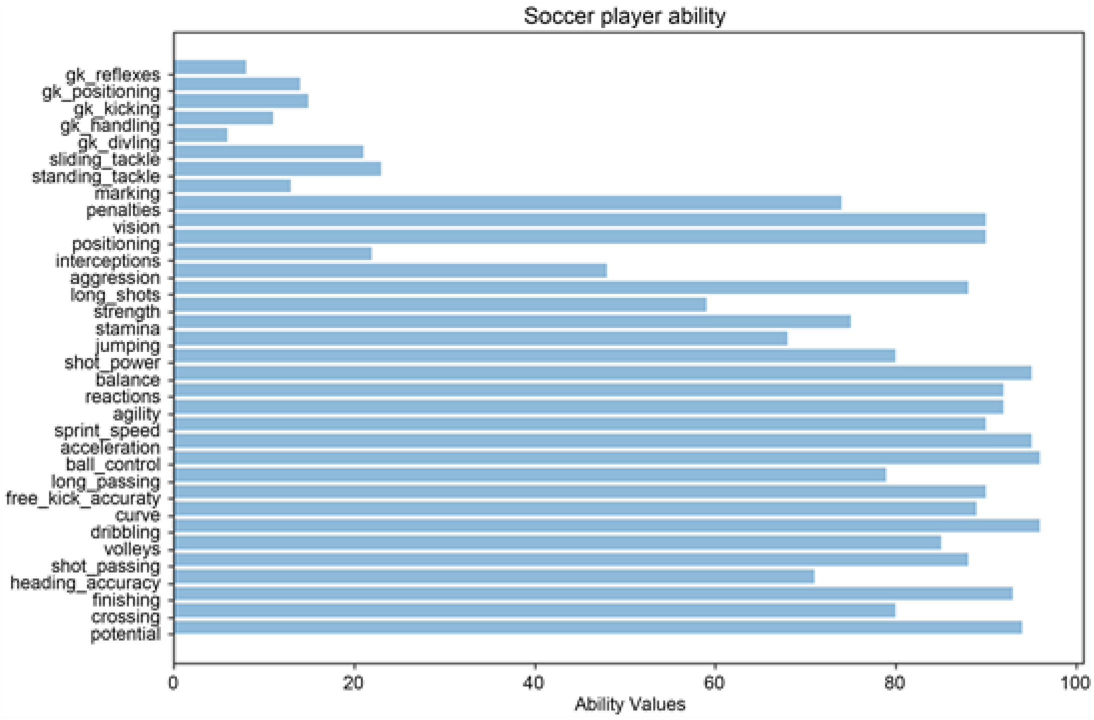


Figure 7. Soccer Player skills plotted to a graph

Chen uses 22 different players and their statistics, as in a game of football there are 2 teams both with 11 players. An interesting point that is highlighted in this paper is the difficulty of considering unpredictable factors that may affect the accuracy of a prediction. In the paper, Chen discusses the challenges in accounting for unforeseeable factors that impact a team's chance of winning a match, such as weather and coaching. To tackle these challenges, the algorithms would require additional data which may not always be available. This can be a common problem with complex machine learning algorithms that set out to make complicated predictions. Machine learning uncertainty can usually be classified as one of two types, Aleatoric uncertainty and Epistemic Uncertainty (Hüllermeier and Waegeman, 2021). Aleatoric uncertainty is caused by inherent randomness or noise within the data. The randomness of data is impossible to eliminate even when additional data is added. No matter how much data is added, certain factors are simply unpredictable. Chen highlights this in his research paper on football match prediction. Chen acknowledges the importance of weather and though it does play a role in the outcome of a football game, to predict it accurately is near impossible as the exact conditions will always be unpredictable.

Once the data for the football players has been collected, it is then trained on three different machine learning models to find which is most accurate. Chen begins by inputting the data of 22 players in two teams to predict the outcome of the match. The output can either be 'win', 'loss' and 'draw'. The study uses a Support Vector Machine (SVM) with a polynomial kernel which is later changed to linear and gives a correct rate of 0.542. The data is then fed through a random tree algorithm, where the optimal number of trees is found to be between 400 and 800. The random tree returns a correct rate of 0.545. Finally, Chen trains the data on a convolutional neural network with 4 layers, two convolution layers, one pooling layer and one full connection layer. After adjusting the configuration of the neural network, it returns a correct rate of 0.574, the highest out of the three algorithms.

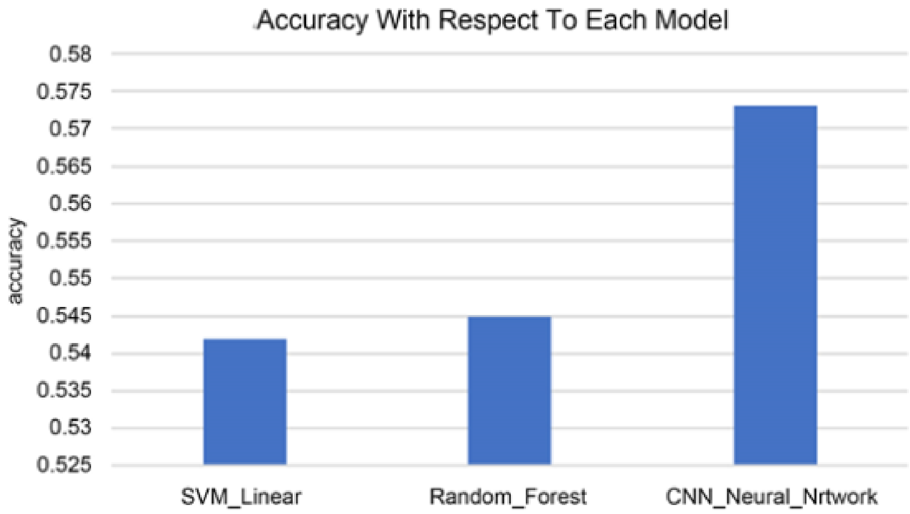


Figure 8. Accuracy Comparison of algorithms

Taking these results into account, the average accuracy score for the prediction models is 55.3% with the convolutional neural network achieving a score of 57.4%. As Chen highlights in the conclusion of the study, BBC football analyst, Mark Lawerson has an accuracy score of 52% when predicting football match outcomes. This highlights the potential for machine learning to aid sports predictions as even the best brains in football struggle to reach the success rate of a machine learning algorithm.

4.5 Difficulties in predicting UFC bouts

Predicting the outcome of a mixed martial arts bout is a difficult task due to the extensive ability of the athletes and the many ways for them to win. Even though the statistics help when trying to decide which fighter will be victorious, the other fighter always has a chance to win if they land a big shot or a flash submission (*Why MMA is the Hardest Sport to Predict - Martial Nerd*, no date)**.** A paper published in January 2022 reviews the impact of psychological prowess in mixed martial arts. The study, written by Sydney Cooper and Marc Lochbaum discusses ideas such as motivation, self-confidence and mental resilience and how they can be pivotal in a fighters' success (Cooper and Lochbaum, 2022). Within the study, the pair found that there are several psychological factors that can impact a fighters' mental state leading to an uncertain performance. External factors such as a large audience and financial security seems to have an impact -on a fighters' ability to perform under pressure. Due to the myriads of athletes who currently fight for the UFC, it is impossible to know if crowd presence has an impact on everyone. According to an article written by Forbes following the recent UFC event, UFC 311, The athletes who are at the top of their division consistently make good money, with Islam Makhachev, the lightweight champion ending the night with a purse of $200,000 for one fight (Reinsmith, 2025). The same cannot be said for fighters who are further down in the rankings for their division however, with the standard pay without a win bonus is around $12,000 - $24,000. Although this may seem like a lot of money, the fighters under the UFC banner only fight around three times a year meaning, the take home pay after coaching fees and taxes can be low considering the risk involved.

4.5.1 Uncertainty

Aleatoric Uncertainty

Leading on from Hengzhi Chen's acknowledgment of uncertainty in machine learning, it is important to recognise the impact that epistemic and aleatoric factors can have on prediction accuracy (Hüllermeier and Waegeman, 2021). Similarly to the football match prediction study, and how limiting factors like weather caused aleatoric uncertainty, the same can be said for predicting UFC bouts. An example of aleatoric uncertainty could be a fighter getting injured. Such an event is unpredictable and cannot be accounted for in statistical models, however it can significantly alter the outcome of a fight despite what the statistics may suggest. Due to there being no way of accounting for aleatoric uncertainty, it is important to acknowledge the potentiality of such an occurrence happening without attempting to incorporate them into the model.

Epistemic Uncertainty

Epistemic uncertainty is caused by limitations in the dataset and the models' understanding of the dynamics of the prediction topic. Dissimilar to aleatoric uncertainty, epistemic uncertainty can be reduced by implementing more data and refining the algorithms. Epistemic uncertainty does affect the accuracy of the UFC fight predictor models. Although there is a vast number of fighters within the UFC who all have their respective statistics, uncertainty can arise when a fighter only has a few fights, as the data will be limited compared to other more experienced fighters. This makes predictions vary in accuracy depending on the fighters. A fighter with many bouts which return a detailed spectrum of statistics will give the model plenty of data to feed from, unlike a fighter with less fights which results in fewer features. Throughout the testing of the model, it is optimal to use fighters with experience in the UFC to give the best results.

5.0 Methodology

5.1 Introduction

This section details the methods used to develop and compare the proposed machine learning models to find which is most optimal for predicting UFC matches. The methods use data analytics and machine learning techniques within a python environment to compare fighters' statistics gathered from performances in previous UFC bouts. The objective is to design a model that can closely predict UFC fights and allow for fight analysis using visualisations.

5.2 Datasets

To ensure accuracy when performing predictions, there must be an adequate amount of data for each fighter. The proposed prediction models require two different datasets, one for UFC fighter statistics and another for previous UFC matchups and the outcome. The data that is being used are both sourced from Kaggle, a reliable open-source depository for machine learning algorithms and datasets. The first dataset contains over four thousand different UFC athletes along with their statistics that are relevant to predicting a winner for a bout. The second dataset holds previous UFC bouts that happened in the past, with both the fighters and the winner.

Dataset 1 (Fighter Statistics) Example Data

Feature Name	Description	Example
Name	First name and Surname of the UFC fighter.	Alex Pereria
Nickname	The given nickname of the UFC fighter.	“Poatan”
Wins	The total wins in the fighters’ professional MMA career.	8
Losses	The total losses in the fighters’ professional MMA career.	3
Draws	The total draws in the fighters’ professional MMA career.	1
Height (cm)	The fighters’ height in centimetres	185cm
Weight (kg)	The fighters’ weight in kilograms.	70.3kg
Reach (cm)	The fighters’ reach in centimetres.	175.2cm
Stance	The fighters’ preferred fighting stance.	Southpaw
Significant Striking Accuracy	The percentage of significant strikes that land and are effective.	42%
Significant Strikes landed every minute.	The average significant strikes the fighter lands each minute.	3.3
Significant Strikes Absorbed every minute	The average of significant strikes the fighter receives every minute.	6.22
Significant Strike defence	The percentage of significant strikes that the fighter successfully defends.	65%
Average takedowns landed per 15 minutes	The average takedowns the fighter completes within a 15-minute period.	10.23
Takedown Accuracy	The percentage of takedowns that are successful.	75%
Takedown Defence	The percentage of takedowns that the fighter successfully defends.	48%
Average Submissions attempted per 15 minutes	The average number of submission attempts made by the fighter every 15 minutes.	11.7

Dataset 2 (Previous UFC Bouts) Example Data

Feature Name	Description	Example
R_fighter	The R_fighter column represents the name of the red corner fighter within the UFC bout.	Dan Ige
B_fighter	The B_fighter column represents the name of the blue corner fighter within the UFC bout.	Peter Yan
Winner	The winner column represents which fighter won the bout.	Blue

5.2.1 Rationale for datasets

Fighter's Statistics Dataset

The fighter's statistics dataset was chosen as it has an adequate number of features and fighters. There are fifteen features in the dataset that are factors in a fighter's performance inside the cage. There are also over four thousand fighters with their own respective statistics which is enough data to train and test a machine learning model.

Apart from features such as name and nickname which are simply there to identify the fighter, each other variable has a crucial role in predicting the outcome of a UFC bout. The dataset holds expansive information including physical attributes that give insight into the ability of each fighter which is essential when declaring the victor of a fight due to the nature of the sport.

This data was sourced from Kaggle which is a widely recognised platform for open-source datasets. This gives credibility to the data being used as Kaggle is a reliable tool used by many in developing data analytics models.

Previous UFC bouts dataset

The previous UFC bouts dataset was used due to it containing over six thousand different UFC fights. This dataset was also acquired from Kaggle, making it a trustworthy source of data.

5.3 Development

This section will detail the steps taken to develop and test the machine learning algorithm. Four of the chosen models will be given the dataset to train and test to see which is the most accurate. Once the models have been compared, the best one will be used in a final accuracy examination where it will be tested in comparison to fighter matchups that have already happened in the UFC to see if the predictions match the real-life outcomes.

5.3.1 Libraries

The implementation utilised several of Python's machine learning libraries:

1. **Scikit-learn**
   - Classification algorithms (Logistic Regression, Random Forest Classifier, SVC, Neural Network)
   - Data preprocessing tools (Standard Scaler)
   - Validation tools (Stratified K-Fold, Cross validation score)
   - Performance evaluation metrics (F1 score, accuracy score)

2. **Pandas**
   - Enables data manipulation and feature engineering
   - Presents data in DataFrames for statistics
   - Handles cleaning the data and missing values
   - Allows merging of datasets.

3. **Matplotlib**
   - Allows for data visualisations
   - Helps identifies patterns in data and fighter statistics comparison

4. **NumPy**
   - Gives statistical calculations for the analysis of features

Each library plays an essential role in the development of the fight predictor model.

5.3.2 Data Pre-Processing

Once the dataset is obtained from Kaggle, it must be pre-processed to optimise results. To do this, several steps must be taken to transform the raw data into a format that is sufficient for training and testing on a machine learning model.

Sample Data taken from ufc-fighters-statistics.csv

name	nickname	wins	losses	draws	height_cm	weight_in_	reach_in_	stance	date_of_bi
Robert Drysdale		7	0	0	190.5	92.99		Orthodox	########

significan	significan	significan	significan	average_t	takedown	takedown	average_submissions_attempted_per_15_minutes
0	0	0	0	7.32	100	0	21.9

Handling Missing Values

Within a dataset as large as the UFC fighters' statistics one, it is expected that there will be missing values that have to be dealt with before training the model. Due to the vast number of data entries that range from active athletes to fighters long retired, there are some statistics unaccounted for. After analysing the data, there are four columns which have consistent missing values: height_in_cm, reach_in_cm, stance and date_of_birth.

In the 'reach_in_cm' column in the statistics dataset, there are 1,924 missing values. Due to this number being so high, it was logical to remove the column entirely from the dataset. Nearly half of the fighters were missing their reach statistic, meaning it would only serve to complicate the prediction model.

Due to the date of birth column being in a date format, this had to be changed for the model to use it as a feature. To change the date of birth to a valid feature, a new column was created in the data set which would hold the fighters' age. To do this, the year in which each fighter was born was subtracted from the current year to calculate their age.

Fighter age calculation

Age is a very significant factor in a fighters' performance in the UFC. As a fighter begins to get older, it's natural that their athleticism and their ability begins to deteriorate. Fighters over the age of 45 were also removed as they are presumed to be retired and no longer actively competing under the company.

Using Matplotlib, we can visualize the correlation between age and a decline in a fighters' career.

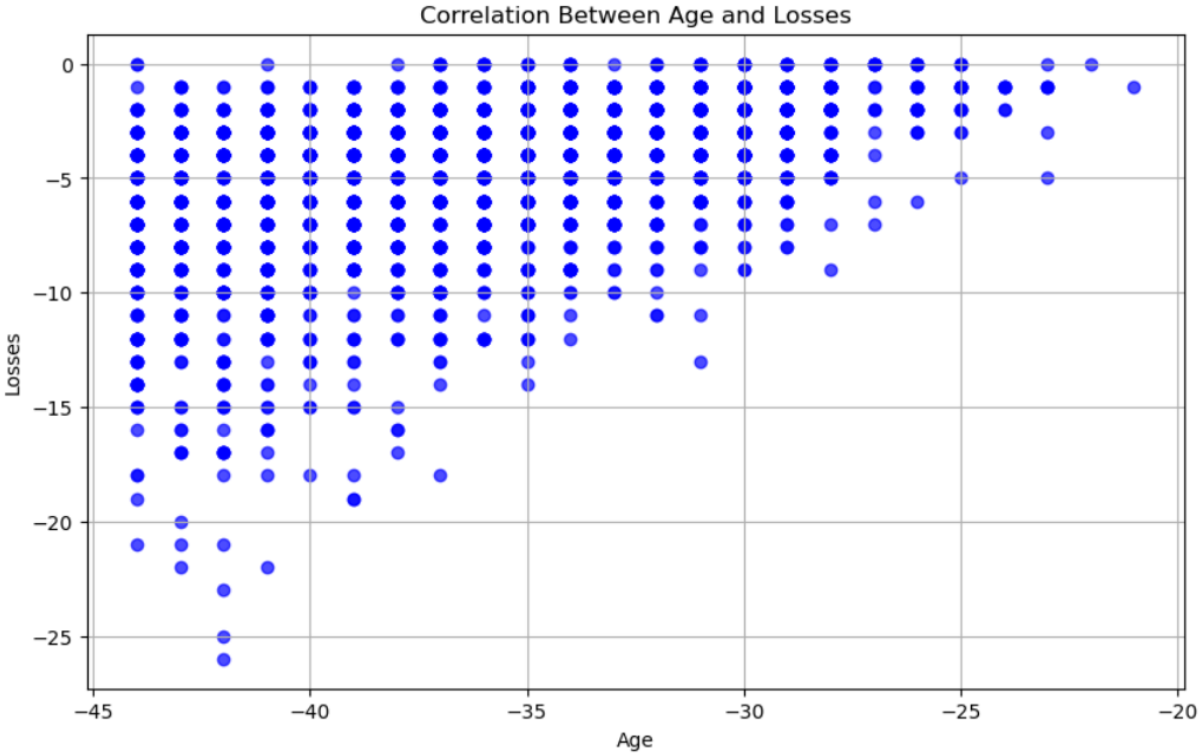


Figure 9. Age statistic plotted against losses statistic

Based on the relationship between age and losses, it seems that the longer a UFC Fighters' career goes on, they're losses increase. A research paper written by Christopher Kirk, from Burnley College in England, sought to investigate the correlation between physical attributes and the implications of them within an MMA bout (Kirk, 2016). The study analysed 287 professional MMA fighters from different weight classes and whether age, height and reach influenced how the fighters lost a bout. The results of the study found a strong correlation between a fighters' age and their chances of winning. A fighter who is older is more likely to lose a fight due to factors such as athleticism and their ability to take damage decreasing. It was found that older fighters are more likely to lose by knockout or technical knockout over anything else, supporting the fact that age can negatively impact a fighters' chance of winning.

Cleaning the UFC fights dataset

The UFC fights dataset is essential for training the prediction model. This data gives the model previous UFC fights to use as a benchmark and holds the target column, winner' where the value can either be 'red' or 'blue' depending on which corner won the fight. To train the model with the previous fights data frame, the type for the winner column must be changed to an integer value to align with the other data and help the algorithms learn more effectively. To do this, the values were changed to a binary format, with red being swapped out for '1' and blue being replaced by '0'. This ensures that the machine learning models can use the previous fights as a feature when making predictions.

Combining the datasets

With there being two datasets, one containing the fighter's statistics and one containing previous UFC matches, combining them with one dataset gives an expansive view of each fight and the fighters attributes and stats. Combing the datasets allows the model to use the 'winner' column as a target variable to train.

5.4 Model Selection

The data was trained on the following machine learning algorithms to find which is the best fit for the final prediction model:

- Random Forest Classifier
- Support Vector Machine (SVM)
- Logistic Regression
- Neural Network

To figure out which model is best for the data, each was passed through a K-fold cross validation using Stratified K-Fold with 6 splits. K-Fold cross validation offers a deeper insight into data, by splitting it into 5 different folds which are used for training while reserving the last fold for testing.

To compare the performance of each model on the training data, the results of the k-fold cross validation are represented in the graph below.

Model	**Accuracy**	**Range**	**F1 Score**	**Range**
Random Forest	0.6912	±0.0124	0.6712	±0.0208
SVM	0.7019	±0.0187	0.6943	±0.0194
Logistic Regression	0.7026	±0.0185	0.6963	±0.0190
Neural Network	0.6567	±0.0159	0.6508	±0.0255

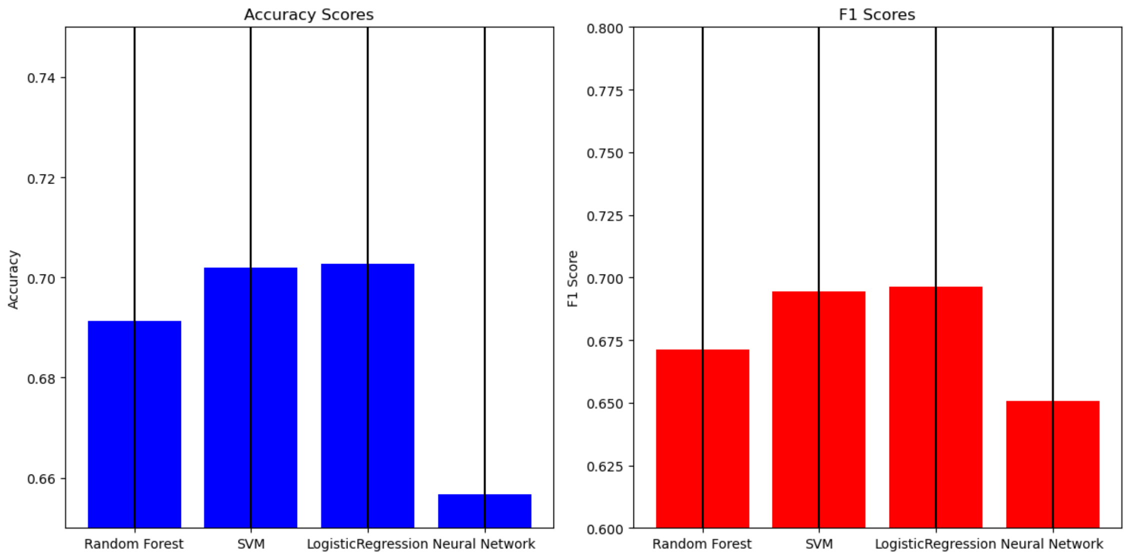


Figure 10. F1 and Accuracy scores of each algorithm

Based off the F1 and accuracy scores for each model from the K-fold cross validation, Logistic Regression seems to be the most optimal for the predictor model, despite it being nearly identical to the support vector machine results. Logistic Regression has consistency in its performance which is evident by its low variance of ±0.0185 for the accuracy and ±0.0194 for the F1 score.

For the prediction model to work as intended, the program must take two fighters names through user input and fetch their respective statistics from the fighter's statistics data frame.

Gathering the fighter's statistics

```
def fighter_features(fighter_name, fighters_df):
    fighter_features = fighters_df[fighters_df['name'] == fighter_name].drop(columns = ['name'])
    if fighter_features.empty:
        print(f"Fighter '{fighter_name}' not found in the dataset.")
    else:
        return fighter_features.iloc[0]
```

The 'fighter_features' function takes the user input fighter name and aligns it with the corresponding statistics. Once the match has been found, the name column is removed as the model requires numerical data to make the prediction. If the fighter's name is not in the fighters_df data frame, the code prints an error message.

As most machine learning models expect a single row of data as an input, each fighter's statistics were merged into one to help the model analyse the difference between them accurately.

Gathering the fighter's statistics

```
def combine_features(fighter_A, fighter_B, fighters_df):
    features_A = fighter_features(fighter_A, fighters_df)
    features_B = fighter_features(fighter_B, fighters_df)

    combined_features = pd.concat([features_A.add_suffix('_A'), features_B.add_suffix('_B')]).to_frame().T
```

Using the fighters' features gathered from the previous function, two new variables are created to store them. Using the statistics for fighter A and fighter B, a new single row pandas data frame is created called 'combined_features'. This new data frame concatenates the statistics for each fighter, adding either A or B as a tag so the features are distinct from each other.

5.5 Feature Engineering

The fighter's statistics dataset holds many different features that have a varying impact on the chances of winning a fight. It is important to consider the difference between defensive and offensive statistics in a UFC match. For example, takedown defence is an important skill that can prevent an opponent from advancing the score cards through takedowns and wrestling. Despite how useful good takedown defence is, successfully defending a takedown doesn't return a fighter any points on the judge's scorecards. Using this logic, there are some statistics which will have a larger weight on the outcome of a fight. For instance, a statistic like "significant strikes landed per minute" has a larger impact on the outcome of the fight as it's counted on the score cards and can inflict damage on the opponent.

To accurately establish the weights of each feature being used for the predictions, this study utilizes coefficient magnitude analysis. Once the data set had been trained on the logistic regression model, each features coefficient was plotted to a graph to determine which were most important in a fighter's chances of winning.

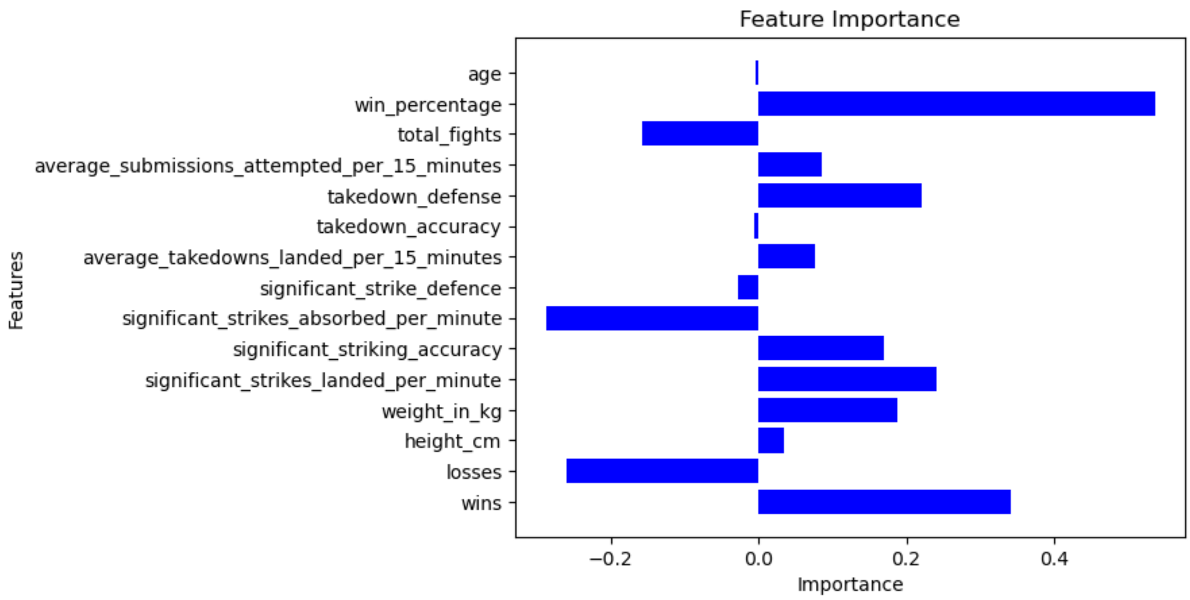


Figure 11. Features and their coefficients

Feature	**Coefficient**
wins	0.3421
losses	-0.2604
height_cm	0.0347
weight_in_kg	0.1870
significant_strikes_landed_per_minute	0.2409
significant_striking_accuracy	0.1699
significant_strikes_absorbed_per_minute	-0.2875
significant_strike_defence	-0.0279
average_takedowns_landed_per_15_minutes	0.0760
takedown_accuracy	-0.0067
takedown_defense	0.2211
average_submissions_attempted_per_15_minutes	0.0850
total_fights	-0.1568
win_percentage	0.5363
age	-0.0037

Using these coefficients as a gauge on how important certain statistics are, we can highlight fighters' chances of winning based on their values for each feature.

Each statistic which directly impacts the outcome of a fight has its own assigned weight value. According to an article written by elite sports on the UFC judging criteria, points are awarded based on damage done by significant strikes and wrestling takedowns the most (*In-Depth Guide - How are UFC Fights Scored?*, no date). Knowing this, the weights are adjusted to make the model asses fights similarly to how a judge would.

Key Findings

- Offensive Dominance
  - Positive coefficients for significant strikes landed per minute and takedown accuracy reinforce the idea of damage winning fights.
- Defensive Struggles
  - Takedown defence having a negative weight seems unorthodox, however this means that the model has found that defensive specialists lose decisions more than offensive specialists.
- Total Fights and Damage
  - As the total fights feature has a negative coefficient, it can be assumed that this is due to factors such as age and damage over years of competition. Despite the belief that experience is a positive thing, at which sometimes it can be, a fighter who has been through many tough fights has been found to have a lesser chance of winning.
- Age
  - Complimentary to earlier findings in the methodology, age plays a large impact on a fighter's chances of winning.

Now that the weight of each statistic can be perceived, it allows for an in-depth analysis of potential match ups in the UFC. Using fighters' features, we can visualise each fighter on a graph to get a deeper understanding of their statistical differences. Using a potential match-up, we can see the possibilities for data analytics within MMA. Jon Jones vs. Tom Aspinall is a highly anticipated fight that will unify the heavyweight championship. Inputting both fighters into the model, we can visualise their statistics in comparison to each other, which can help athletes identify weaknesses they should work on and potential advantages they may have over their opponent.

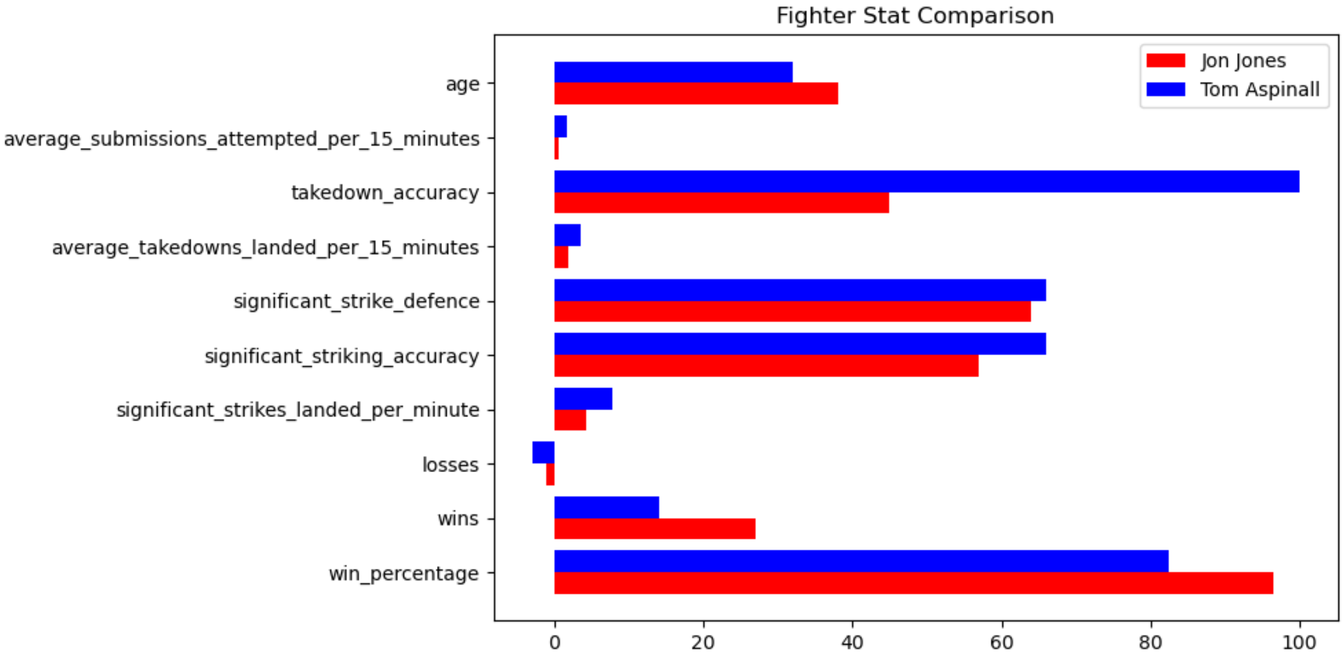


Figure 12. Fighter statistics comparison

6.0 Model Evaluation

Using the F1 score as a metric of accuracy is a good way to determine how well the model can make predictions, especially on imbalanced datasets, however, it is best to compare the results to real world outcomes. To evaluate the general accuracy, previous UFC cards can be fed through the model to see if it gives similar results to the actual fights. This method will give a broader insight into the reliability of the model and its overall accuracy in comparison to the ground truth.

To make the predictions as close to real life as possible, it is recommended to use recent UFC bouts as the statistics dataset is regularly updated, meaning older cards from previous years may have fights were fighters underperformed in relation to their current statistics. The UFC cards being used are all sourced from the official UFC website, where the fighters who fought and the results can be found.

UFC Card	**Date**
UFC Fight Night: Edwards vs. Brady	March 22, 2025
UFC 313: Pereira vs. Ankalaev	March 8, 2025
UFC 312: Du Plessis vs. Strickland 2	Feb 8, 2025
UFC Fight Night: Adesanya vs. Imavov	Feb 1, 2025
UFC 311: Makhachev vs, Moicano	Jan 18, 2025
UFC 309: Jones vs. Miocic	Nov 9, 2024

Due to the UFC matchmaking, it is rare for a fighter to be significantly mismatched statistically as the organisation prioritises fair match ups to make the most exciting fights. Most fights at this level are relatively close on paper, making exciting match ups to make the most of each fighters' skill set. Three fights have been taken from the six recent UFC cards and the fighters' statistics have been ran through the model to predict the winner. The results will display who the predicted winner is according to the prediction model, along with the actual winner. The best way to gauge the accuracy of the model is to compare the confidence metric to the opening odds for each fighter. The opening odds have been sourced from MMA odds breaker, where a vast archive of previous MMA fight odds is stored (*MMAOddsBreaker*, 2025) The fights gathered are sourced from the official UFC website. The odds gathered for the favourite from MMA Odds Breaker have been converted to a percentage using the formula:

$$Probability = \frac{|Odds|}{|Odds| + 100} \times 100$$

UFC Fight Night: Edwards vs. Brady

Fight 1 – Leon Edwards (A) vs. Sean Brady (B)
Betting Odds - +132 Edwards, -156 Brady (~60.9%, Brady)
Predicted Result – Sean Brady (**B**), Confidence: 59.42%
Actual Result – Sean Brady (**B**)

```
Model F1 Score: 0.7108
The predicted winner is: Sean Brady
Confidence: 59.42%
```

Figure 13. Brady vs. Edwards Results

What was supposed to be a close fight according to the opening betting odds turned out to be a dominant performance for Sean Brady. Brady would go on to dominate Leon Edwards for three rounds straight before securing a submission by guillotine in the fourth. Before the fight was stopped, the judges' scorecards had Brady winning all three rounds 27-30. Both the opening odds and the prediction model had anticipated a close fight, however Brady having a strong performance allowed him to out class Edwards with relative ease.

Fight 2 – Carlos Ulberg (A) vs. Jan Blachowicz (B)
Betting Odds – +295 Blachowicz, -375 Ulberg (~78.9%, Ulberg)
Predicted Result – Carlos Ulberg (**A**), Confidence: 59.16%
Actual Result – Carlos Ulberg (**A**)

```
Model F1 Score: 0.6814
The predicted winner is: Carlos Ulberg
Confidence: 59.16%
```

Figure 14. Ulberg vs. Blachowicz Results

The opening odds for the Ulberg vs. Blachowicz fight had Carlos Ulberg as the large favourite. The fight would span over three rounds where Blachowicz would win the first and Ulberg coming back to win rounds two and three. The judges score cards read 28-29 in favour of Ulberg by unanimous decision. The prediction model reflected these results accurately with the confidence in Ulberg's victory being 59.16%, a little shy of the dominant performance that the opening odds predicted.

Fight 3 – Kevin Holland (A) vs. Gunnar Nelson (B)
Betting Odds - + 114 Holland, -135 Nelson (~57.4%, Nelson)
Predicted Results – Gunnar Nelson (**B**)**,** Confidence: 67.19%
Actual Result – Kevin Holland (**A**)

```
Model F1 Score: 0.7102
The predicted winner is: Gunnar Nelson
Confidence: 67.19%
```

Figure 15. Holland vs. Nelson Results

Both the prediction model and the opening odds had Nelson winning a close fight, however Kevin Holland would prove to be the victor by winning two out of the three rounds and earning a unanimous decision over Gunnar Nelson with the judges' score cards reading 29-28 in favour of Holland. Despite the performance that Holland displayed, the prediction model gave similar results to the opening odds.

UFC 313: Pereira vs. Ankalaev

Fight 1 – Alex Pereira (A) vs. Magomed Ankalaev (B)
Betting Odds - +126 Pereira, -162 Ankalaev (~61.8%, Ankalaev)
Predicted Results – Magomed Ankalaev (**B**) Confidence: 54.37%
Actual Result – Magomed Ankalaev (**B**)

```
Model F1 Score: 0.6928
The predicted winner is: Magomed Ankalaev
Confidence: 54.37%
```

Figure 16. Pereira vs. Ankalaev Results

Alex Pereira vs Magomed Ankalaev was a championship bout for the light-heavyweight title. Before the fight, the opening odds displayed how evenly matched the pair were, with the prediction model supporting this. Ankalaev would win the fight by unanimous decision, with two judges scoring the fight 47-48 for Magomed Ankalaev and one judge scoring it 46-49 Ankalaev meaning he won three out of 5 rounds on two of the judges' scorecards and four out of five on the third judges' scorecards.

Fight 2 – Justin Gaethje (A) vs. Rafael Fiziev (B)
Betting Odds - +125 Gathje, -145 Fiziev (~59.18%, Fiziev)
Predicted Results – Justin Gaethje (**A**) Confidence: 52.30%
Actual Result – Justin Gaethje (**A**)

```
Model F1 Score: 0.7130
The predicted winner is: Justin Gaethje
Confidence: 52.30%
```

Figure 17. Gaethje vs. Fiziev Results

Justin Gaethje and Rafael Fiziev would compete in a three-round fight in the co-main event at UFC 313 before Pereira and Ankalaev. According to the opening odds, Fiziev was the favourite to win despite it being projected to be a close fight. Justin Gathje would win the fight by unanimous decision with all the judges' scoring it 29-28 for Gaethje, meaning he won two out of the three rounds. The prediction model reflects this by having Gaethje winning with a confidence score of 52.30%.

Fight 3 – Ignacio Bahamondes (A) vs. Jalin Turner (B)
Betting Odds - +180 Bahamondes, - 210 Turner (~67.74%, Turner)
Predicted Results – Jalin Turner (B) Confidence: 83.94%
Actual Result – Ignacio Bahamondes (**A**)

```
Model F1 Score: 0.7138
The predicted winner is: Jalin Turner
Confidence: 83.94%
```

Figure 18. Bahamondes vs. Turner Results

Ignacio Bahamondes vs. Jalin Turner was assumed to be a fight that would go in favor of Jalin Turner, with the opening odds and the prediction model reflecting this. Despite the odds, Bahamondes would submit Turner in the first round by triangle choke, declaring an upset victory. The prediction model displayed a similar outcome to the opening odds, however there are times when a fight can completely go in an unexpected direction due to the nature of the sport.

UFC 312: Du Plessis vs. Strickland 2

Fight 1 – Dricus du Plessis (A) vs. Sean Strickland (B)
Betting Odds - +125 Strickland, -145 Du Plessis (~59.1%, Du Plessis)
Predicted Results – Dricus Du Plessis (A) Confidence: 52.61%
Actual Result – Dricus Du Plessis (**A**)

```
Model F1 Score: 0.6900
The predicted winner is: Dricus Du Plessis
Confidence: 52.61%
```

Figure 19. Du Plessis vs. Strickland Results

Dricus Du Plessis and Sean Strickland fought a highly anticipated rematch at UFC 312, with the opening odds declaring Du Plessis, the victor of their previous bout as the favorite. The prediction model returned similar confidence in a Du Plessis victory, presenting the fight as a close one like their previous bout which ended in a narrow victory for Du Plessis. Despite the competitiveness of their previous fight, Du Plessis would win a one-sided unanimous decision with two of the judges' scorecards reading 50-45 and the other judge scoring it 49-46 all in favor of Du Plessis.

Fight 2 – Zhang Weili (A) vs. Tatiana Suarez (B)
Betting Odds - +210 Tatiana Suarez, -250 Zhang Weili (~71.4%, Zhang)
Predicted Results – Zhang Weili (**A**) Confidence: 92.80%
Actual Result – Zhang Weili (**A**)

```
Model F1 Score: 0.7006
The predicted winner is: Zhang Weili
Confidence: 92.80%
```

Figure 20. Weili vs. Suarez Results

Zhang Weili vs. Tatiana Suarez had opening odds heavily favoring Weili to win. Weili has been a dominant champion in the UFC women's strawweight division for some time, with both the model and odds reflecting this. Weili would in a unanimous decision, scoring 49-46 on all three judges' score cards. Due to the dominant career of Zhang Weili, the prediction model had a score of 92.80% confidence that she would successfully defend her belt against Suarez.

Fight 3 – Jimmy Crute (A) vs. Rodolfo Bellato (B)
Betting Odds - +240 Crute, -300 Bellato (~75.00%, Bellato)
Predicted Results – Rodolfo Bellato (**B**) Confidence: 85.15%
Actual Result – Majority Draw

```
Model F1 Score: 0.6911
The predicted winner is: Rodolfo Bellato
Confidence: 85.15%
```

Figure 21. Crute vs. Bellato Results

Despite Rodolfo Bellato being a large favorite in the opening odds and the prediction model, the fight would end in a majority draw, with one judge scoring the contest 29-27 in favor of Jimmy Crute and the other two scoring it 28-28 in favor of a draw. A draw is relatively rare in the UFC, with only 2.5% of fights ending in a draw in the light-heavyweight division in which this fight was sanctioned in. (‘UFC Fight Outcomes by Weight Class’, 2025)

UFC Fight Night: Adesanya vs. Imavov

Fight 1 – Nassourdine Imavov (A) vs. Israel Adesanya (B)
Betting Odds - +164 Imavov, -215 Adesanya (~68.25%, Adesanya)
Predicted Results – Israel Adesanya (**B**) Confidence: 50.61%
Actual Result – Nassourdine Imavov (**A**)

```
Model F1 Score: 0.7188
The predicted winner is: Israel Adesanya
Confidence: 50.61%
```

Figure 22. Imavov vs. Adesanya Results

Israel Adesanya, a former champion, reigned for years in the middleweight division before losing his belt in a prior fight. The opening odds had Adesanya winning. The prediction model also had Adesanya winning, however, the model only gave a confidence score of 50.61% meaning that the fight was very close statistically. Although Adesanya would win the first round with a 10-9 score in all three judges scorecards, Imavov would knock Adesanya out in the second round, declaring an upset over the former champion.

Fight 2 – Sergei Pavlovich (A) vs. Jairzinho Rozenstruik (B)
Betting Odds - +180 Rozenstruik, -210 Pavlovich (~67.74%, Pavlovich)
Predicted Results – Sergei Pavlovich (**A**) Confidence: 89.37%
Actual Result – Sergei Pavlovich (**A**)

```
Model F1 Score: 0.7216
The predicted winner is: Sergei Pavlovich
Confidence: 89.37%
```

Figure 23. Pavlovich vs. Rozenstruik Results

Sergei Pavlovich and Jairzinho Rozenstruik would compete in a three-round fight. Reflecting the opening odds and the model's confidence prediction, Pavlovich would win a dominant unanimous decision, scoring 30-27 on all three judges' scorecards, winning all three rounds decisively.

Fight 3 - Vinicius Oliveira (A) vs Said Nurmagomedov (B)
Betting Odds - +180 Oliveria, -210 Nurmagomedov (~67.74%, Nurmagomedov)
Predicted Results – Vinicius Oliveria (**A**) Confidence: 86.90%
Actual Result – Vinicius Oliveria (**A**)

```
Model F1 Score: 0.7116
The predicted winner is: Vinicius Oliveira
Confidence: 86.90%
```

Figure 24. Oliveria vs. Nurmagomedov Results

Despite the opening odds reflecting a victory for Said Nurmagomedov, the prediction model had a high confidence score that Vinicius Oliveira would win the fight. After the three-round fight, the judges' scorecards all read 29-28 in favor of Oliveira. Although the model predicted the fight outcome accurately, it displayed a very high confidence score for what would turn out to be a close fight, with Oliveira narrowly earning a unanimous decision.

UFC 311: Makhachev vs. Moicano

Fight 1 – Islam Makhachev (A) vs. Renato Moicano (B)
Betting Odds - +700 Moicano, -1100 Makhachev (~91.67% Makhachev)
Predicted Results – Islam Makhachev (**A**) Confidence: 83.34%
Actual Result – Islam Makhachev (**A**)

```
Model F1 Score: 0.6918
The predicted winner is: Islam Makhachev
Confidence: 83.34%
```

Figure 25. Makhachev vs. Moicano Results

Islam Makhachev was originally scheduled to fight the number one contender in the lightweight division, Arman Tsarukyan. Due to an injury just days before the fight was scheduled, Tsarukyan was forced to pull out and Renato Moicano, who is ranked tenth in the lightweight division, stepped in on short notice to face the champion. As the opening odds suggest, the fight was predicted to be heavily one sided in favor of Makhachev. The models confidence would reflect this, with a score of 83.34%, like the opening odds. Makhachev defeated Moicano in the first round by submission.

Fight 2 - Merab Dvalishvili (A) vs. Umar Nurmagomedov (B)
Betting Odds - +142 Dvalishvili, -170 Nurmagomedov (~62.96%, Nurmagomedov)
Predicted Results – Umar Nurmagomedov (**B**) Confidence: 88.35%
Actual Result – Merab Dvalishvili (**A**)

```
Model F1 Score: 0.7037
The predicted winner is: Umar Nurmagomedov
Confidence: 88.35%
```

Figure 26. Dvalishvili vs. Nurmagomedov Results

Umar Nurmagomedov was the favourite going into this fight, being undefeated and having some of the best coaches in the world. Despite the opening odds and the model's prediction, Merab Dvalishvili would win a close fight by unanimous decision, scoring 48-47 on two of the judges' scorecards and a controversial 49-46 on the other scorecard.

Fight 3 - Jamahal Hill (A) vs. Jiri Prochazka (B)
Betting Odds - +110 Prochazka, -130 Hill (~56.56%, Hill)
Predicted Results – Jiri Prochazka (**B**) Confidence: 58.05%
Actual Result – Jiri Prochazka (**B**)

```
Model F1 Score: 0.6788
The predicted winner is: Jiri Prochazka
Confidence: 58.05%
```

Figure 27. Hill vs. Prochazka Results

Jamahl Hill vs. Jiri Prochazka was to be a competitive fight with both athletes being evenly matched statistically. The prediction model returned a confidence score of 56.56% in favor of Prochazka, Dissimilar to the opening odds which listed Hill as the favorite. Prochazka would win round one with Hill winning round two before Jiri Prochazka won in round three by way of knockout. The model made a very accurate prediction, as the bout was close before the knockout with both fighters winning a round.

UFC 309: Jones vs. Miocic

Fight 1 – Jon Jones (A) vs. Stipe Miocic (B)
Betting Odds - +300 Miocic, -400 Jones (~80.00%, Jones)
Predicted Results – Jon Jones (**A**) Confidence: 81.90%
Actual Result – Jon Jones (**A**)

```
Model F1 Score: 0.7039
The predicted winner is: Jon Jones
Confidence: 81.90%
```

Figure 28. Jones vs. Miocic Results

Jon Jones, the heavyweight champion, is regarded as one of the greatest fighters of all time within MMA. Due to Jones being so dominant throughout his career, with his only loss being by a disqualification due to an illegal elbow, the odds for this fight were heavily in his favor. The prediction model made a near perfect match to the opening odds, with a confidence score of 81.90%, just 1.9% difference from the opening odds. The fight would be stopped in the third round with Jones defeating Miocic by technical knockout.

Fight 2 - Charles Oliveira (A) vs. Michael Chandler (B)
Betting Odds - +164 Chandler, -198 Oliveira (~66.44%, Oiliveira)
Predicted Results – Charles Oliveira (**A**) Confidence: 74.24%
Actual Result – Charles Oliveira (**A**)

```
Model F1 Score: 0.7143
The predicted winner is: Charles Oliveira
Confidence: 74.24%
```

Figure 29. Oliveira vs. Chandler Results

Charles Oliveira vs. Michael Chandler was a highly anticipated rematch between the two lightweight contenders. Due to Oliveira defeating Chandler in a previous fight, the odds were in favor of an Oliveira victory, which the prediction model reinforced. The fight would last the entire five rounds, with Oliveira defeating Michael Chandler by unanimous decision, scoring 49-46 on two of the judges' score cards and 49-45 on the other scorecard, declaring a dominant performance which confirms the confidence in the prediction model.

Fight 1 - Bo Nickal (A) vs. Paul Craig (B)
Betting Odds - +1000 Craig, -2000 Nickal (~95.24%, Nickal)
Predicted Results – Bo Nickal (**A**) Confidence: 97.31%
Actual Result – Bo Nickal (**A**)

```
Model F1 Score: 0.6946
The predicted winner is: Bo Nickal
Confidence: 97.31%
```

Figure 30. Nickal vs. Craig Results

Bo Nickal, an undefeated prospect, was matched against UFC veteran Paul Craig. Despite Paul Craig's accolades within the sport, the odds were heavily stacked against him with the prediction model matching this almost perfectly. The fight would go as expected, with Bo Nickal dominating the three round fight to score 30-27 on all three judges' scorecards.

Opening Odds Comparison

The UFC Fight prediction model has a good accuracy rate when predicting UFC fights. To measure this, out of eighteen of the recent UFC bouts used to compare the models' predictions, the average difference from the predictor model's confidence score opening odds can be calculated. This is done by adding the differences up then dividing by the amount of correctly predicted fights.

Fight	Models Confidence	Opening Odds	Difference
Leon Edwards vs. Sean Brady	59.42%	60.9%	1.48%
Carlos Ulberg vs. Jan Blachowicz	59.19%	78.9%	19.71%
Magomed Ankalaev vs. Alex Pereira	54.37%	61.8%	7.43%
Justin Gaethje vs. Rafael Fiziev	52.30%	59.18%	6.88%
Dricus Du Plessis vs. Sean Strickland	52.61%	59.1%	6.49%
Zhang Weili vs. Tatiana Suarez	92.80%	71.4%	21.4%
Jimmy Crute vs. Rodolfo Bellato	85.15%	75.00%	10.15%
Israel Adesanya vs. Nassourdine Imavov	50.61%	68.25%	17.64%
Sergei Pavlovich vs. Jairzinho Rozenstruik	89.37%	67.74%	21.63%
Islam Makhachev vs. Renato Moicano	83.34%	91.67%	8.33%
Umar Nurmagomedov vs. Merab Dvalishvili	88.35%	62.96%	25.39%
Jon Jones vs. Stipe Miocic	81.90%	80.800%	1.90%
Charles Oliveira vs. Michael Chandler	74.24%	66.44%	7.80%
Bo Nickal vs. Paul Craig	97.31%	95.24%	2.07%

To calculate the average difference between the model confidence and the opening odds, the formula is:

$$Average = \frac{x1 + x2 + x3 + x4 \ldots}{number\ of\ entries}$$

After the calculations, the average difference between the model's confidence in the prediction and the opening odds is **11.30%.**

7.0 Conclusion and Discussion

7.1 Introduction

This chapter will discuss the success of the model and potential improvements that could be introduced to make it more accurate.

7.1 Model Accuracy

As discussed previously within the model evaluation chapter, when comparing the model's confidence feature to the opening odds of recent UFC fights, the average difference between the opening odds made by industry experts and the model confidence was 11.30%. Based on this, the model seems to closely align with expert predictions and even being more accurate in some cases where the fight reflected the confidence score more than the opening odds.

To further evaluate the model accuracy in predicting the winner of UFC bouts, fifty fights were selected from recent UFC cards to be analysed by the model (see **Appendix, figure 2.1).** Like the opening odds comparison, the predictions were compared to the actual result of the fights, giving a broader look at the overall accuracy of the model. Out of the 50 fights ran through the model, 35 of them were predicted correctly. This gives a success rate of 70%, which reflects the F1 score as it is around 0.7 on average. Using this figure as a basis for declaring the overall accuracy, it can be said that the UFC fight predictor model has an accuracy rating of 70% when predicting UFC fights on a statistical level.

7.2 Potential Improvements

There are many ways to improve the accuracy of the model. Despite the addition of a second dataset which contains all previous matchups within the UFC, it is not used as a feature in the predictions. To make the model more accurate, these previous fights and their outcomes could be added to the model to determine whether two fighters have faced each other before, which could significantly impact the outcome.

Additionally, more data would serve to increase the accuracy of the model. Implementing more variables such as momentum, tracking recent win and loss streaks, could increase the accuracy of the model. Fight camp and training partners could also be added, giving the model an expansive view on the fighters' preparation and fitness leading up to the contest giving more insight into the expected performance of the athletes. Further improvements could be made by implementing fighting styles within the data. It has long been said that styles make fights, meaning the style of a fighter can sometimes be the path to victory, as certain styles are more effective at beating others. A research paper written by Isais Wellington from Southern Illinois Texas delves deep into this subject, arguing that styles not only are a potent factor in a fight, but can also be traced back to cultural and racial influences (Wellington, 2023). Although the paper focuses on boxing, the findings can be applied to mixed martial arts as well. In recent years, the sport has seen a surge of fighters coming from Dagestan, who bring a sambo style wrestling approach that has proven to be effective against explosive strikers by draining their energy over rounds of high-pace dominance on the ground (Jones, 2024).

7.3 Conclusion

In conclusion, the results of the prediction model and data analysis emphasises the potential for machine leaning practices within combat sports.

Despite the chaotic nature of MMA, the model has given a relatively high success rate. Prediction models can be very helpful within the betting industry to gain an insight into the fights at a statistical level. Using the skill statistics and physical attributes, data analysis can help athletes prepare for their next opponent by way of comparison.

To further evaluate the model's accuracy, more fights could be simulated through it instead of the fifty used in this paper. To really get an understanding of how well it can perform, using 200-300 fights could give a deeper understanding of the accuracy and help identify trends or patterns associated with certain statistics.

8.0 Critical Self Appraisal

If I were given the opportunity to start my honours project from scratch, I would implement the following changes:

During the time I was writing my interim report, I came slightly off course in balancing my workload. I had other assignments to do for different modules which led me to neglecting my honours project for a few weeks. Knowing how large of an impact this made on my time management, I would prioritise more time for research and drafting my honours project to ensure I had plenty of time.

Managing my time better would have really helped me, as in February I was involved in a car accident as a passenger. Although I suffered minor physical injuries, I did walk away with a concussion due to hitting my head as the car entered the ditch. Due to the concussion, I couldn’t focus on my dissertation or look at my screen for long without experiencing headaches. This forced me to take two weeks from work until I was at full health again, which caused me to fall behind in my schedule. If I had been better with my work management earlier in the year, I wouldn’t have been so stressed about the time I had to take off.

Another thing I regret not focusing on was implementing more features for the machine learning model. Due to the time it took to get it working and evaluate the results, I didn’t want to tamper with the foundational code I had written months earlier in fear of messing it up. Despite this though I am happy with the results and the work put in to analyse the fighter’s statistics and the different models’ performances.

Something I feel I should have explored more is the machine learning models implications on the actual project. Although I believe this paper does a good job of explaining each model and evaluating them in relation to the data, more direct connection between the UFC fight predictor model and the analysis of each algorithm could have served well.

Finally, in retrospect I would have liked to focus more on the actual models themselves. The paper does a good job of describing each model but could do with more depth in relation to the task at hand.

Despite the changes I would have made in my approach to this paper, I am happy with the final product. I believe it reflects the initial idea I had when writing my project specification well.

9.0 References

(Ali et al., 2012)
Ali, Jehad, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. 'Random Forests and Decision Trees' 9, no. 5 (2012).

(Ćenanović and Kevrić, 2023).
'Profiling the Physical Performance of Young Boxers with Unsupervised Machine Learning: A Cross-Sectional Study'. Accessed 1 April 2025. https://www.mdpi.com/2075-4663/11/7/131?trk=public_post_main-feed-card-text.

(Chen, 2019)
Chen, Hengzhi. 'Neural Network Algorithm in Predicting Football Match Outcome Based on Player Ability Index'. Advances in Physical Education 9, no. 4 (9 September 2019): 215–22. https://doi.org/10.4236/ape.2019.94015.

(Cooper and Lochbaum, 2022)
Cooper, Sydney, and Marc Lochbaum. 'A Systematic Review of the Sport Psychology Mixed Martial Arts Literature: Replication and Extension'. European Journal of Investigation in Health, Psychology and Education 12, no. 2 (February 2022): 77–90. https://doi.org/10.3390/ejihpe12020007.

(David, no date)
'Why MMA Is the Hardest Sport to Predict - Martial Nerd'. Accessed 5 February 2025. https://www.martialnerd.com/posts/mma-is-hardest-sport-to-predict.

(El Naqa and Murphy, 2015)
El Naqa, Issam, and Martin J. Murphy. 'What Is Machine Learning?' In Machine Learning in Radiation Oncology, 3–11. Springer, Cham, 2015. https://doi.org/10.1007/978-3-319-18305-3_1.

(Elite Sports, no date)
Elite Sports. 'In-Depth Guide - How Are UFC Fights Scored?' https://www.elitesports.com/blogs/news/in-depth-guide-how-are-ufc-fights-scored.

(Fisher, 2024)
Fisher, Eric. 'U.S. Sports Betting Sets Records in 2023 for Handle, Revenue'. Front Office Sports (blog), 20 February 2024. https://frontofficesports.com/u-s-sports-betting-sets-records-in-2023-for-handle-revenue/.

(Ghahramani, 2004)
Ghahramani, Zoubin. 'Unsupervised Learning'. In Advanced Lectures on Machine Learning, 72–112. Springer, Berlin, Heidelberg, 2004. https://doi.org/10.1007/978-3-540-28650-9_5.

(History of MMA and the UFC, no date)
'History of MMA and the UFC'. Accessed 13 January 2025. https://verdictmma.com/guides/[...slug].

(Hüllermeier and Waegeman, 2021)

Hüllermeier, Eyke, and Willem Waegeman. ‘Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods’. Machine Learning 110, no. 3 (1 March 2021): 457–506. https://doi.org/10.1007/s10994-021-05946-3.

(IBM, 2021)
‘What Is a Decision Tree? | IBM’, 2 November 2021. https://www.ibm.com/think/topics/decision-trees.

(Jakkula, no date)
Jakkula, Vikramaditya. ‘Tutorial on Support Vector Machine (SVM)’, 2011

(Jones, 2024)
Dagestani Muslim Fighters Dominate The MMA Scene | Fight.TV’. https://www.fight.tv/post/the-unstoppable-force-the-dominance-of-dagestani-muslim-fighters-in-mma.

(Kirk, 2016)
Kirk, Christopher. ‘The Influence of Age and Anthropometric Variables on Winning and Losing in Professional Mixed Martial Arts’. PeerJ Preprints, 22 August 2016. https://doi.org/10.7287/peerj.preprints.2380v1.

(Kleinbaum and Klein, 2010)
Kleinbaum, David G., and Mitchel Klein. Logistic Regression. Statistics for Biology and Health. New York, NY: Springer New York, 2010. https://doi.org/10.1007/978-1-4419-1742-3.

(Nag, 2023)
Nag, Sayan. ‘Who Founded the UFC? Here’s All about the Mega MMA Promotion’s Origin and Eventual Takeover’, 7 March 2023. https://www.sportskeeda.com/mma/news-who-founded-ufc-here-s-mega-mma-promotion-s-origin-eventual-takeover.

(Nagyfi, 2018)
Nagyfi, Richard. ‘The Differences between Artificial and Biological Neural Networks’. Medium, 4 September 2018. https://towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7.

(Nilsen, 2018)
Nilsen, Vanessa. ‘An Introduction to Machine Learning’, 2018.

(*How To Bet On UFC Fights | Ultimate Guide to UFC & MMA Betting | OddsChecker*, no date)
‘How To Bet On UFC Fights | Ultimate Guide to UFC & MMA Betting | OddsChecker’. Accessed 1 April 2025. https://www.oddschecker.com/us/betting-guides/betting-strategy/how-to-bet-on-ufc.

(Reinsmith, 2025)
'UFC 311 Pay: Merab Dvalishvili Earns Highest Reported Fighter Payout'. Accessed 1 April 2025. https://www.forbes.com/sites/trentreinsmith/2025/01/19/ufc-311-pay-merab-dvalishvili-earns-highest-reported-fighter-payout/.

(R. China, 2023)
'Types of Machine Learning | IBM', 1 August 2024. https://www.ibm.com/think/topics/machine-learning-types.

(Tuzsuz, no date)
Tuzsuz, Deniz. 'Sigmoid Function'. Accessed 26 January 2025. https://www.learndatasci.com/glossary/sigmoid-function/.

(Tiwari, 2025)
'Understand Kernel in Support Vector Machine — A Complete Guide | by Shriyansh Tiwari | Feb, 2025 | Medium'. https://medium.com/@st4046641/understand-kernel-in-support-vector-machine-a-complete-guide-7d99a43fda08

(*UFC Results | UFC*, no date)
'UFC Results | UFC'. Accessed 1 April 2025. https://www.ufc.com/results.

(Wang, 2003)
Wang, Sun-Chong. 'Artificial Neural Network'. In Interdisciplinary Computing in Java Programming, 81–100. Springer, Boston, MA, 2003. https://doi.org/10.1007/978-1-4615-0377-4_5.

(Wellington, 2023)
'Styles Make Fights: Coaching Boxing Style as a Racial Project: EBSCOhost'. Accessed 3 April 2025. https://web.p.ebscohost.com/ehost/detail/detail?vid=0&sid=16c04547-770b-459b-afa7-cea12cc9637f%40redis&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=175289895&db=s3h.

10.0 Appendix

Figure 1.1 k-Fold cross validation results

```
Random Forest:
Accuracy: 0.6912 ± 0.0124
F1 Score: 0.6712 ± 0.0208

SVM:
Accuracy: 0.7019 ± 0.0187
F1 Score: 0.6943 ± 0.0194

LogisticRegression:
Accuracy: 0.7026 ± 0.0185
F1 Score: 0.6963 ± 0.0190

Neural Network:
Accuracy: 0.6567 ± 0.0159
F1 Score: 0.6508 ± 0.0255
```

Figure 2.1 Matchups used for testing

	Fighter A	**Fighter B**	**Predicted Winner**	**Actual Winner**
1	Marvin Vettori	Roman Dolidze	B	B
2	Ilia Topuria	Max Holloway	B	A
3	Khamzat Chimaev	Robert Whittaker	A	A
4	Magomed Ankalaev	Aleksandar Rakic	A	A
5	Lerone Murphy	Dan Ige	A	A
6	Shara Magomedov	Armen Petrosyan	A	A
7	Anthony Hernandez	Michel Pereira	B	A
8	Rob Font	Kyler Phillips	B	A
9	Daniel Pineda	Darren Elkins	B	B
10	Tatsuro Taira	Brandon Royval	A	B
11	Chidi Njokuani	Jared Gooden	A	A
12	Grant Dawson	Rafa Garcia	A	A
13	Daniel Rodriguez	Alex Morono	A	A
14	Alex Pereira	Khalil Rountree Jr	A	A
15	Mario Bautista	Jose Aldo	A	A
16	Roman Dolidze	Kevin Holland	B	A
17	Renato Moicano	Benoit Saint Denis	B	A
18	Nassourdine Imavov	Brendan Allen	B	A
19	Bryan Battle	Kevin Jousset	B	B
20	Morgan Charriere	Gabriel Miranda	A	A
21	Alex Pereira	Jiří Procházka	B	A
22	Dan Ige	Diego Lopes	B	B

23	Anthony Smith	Roman Dolidze	A	B
24	Islam Makhachev	Dustin Poirier	A	A
25	Sean Strickland	Paulo Costa	A	A
26	Niko Price	Alex Morono	B	A
27	Jose Aldo	Jonathan Martinez	A	A
28	Anthony Smith	Vitor Petrino	B	A
29	Caio Borralho	Paul Craig	A	A
30	Alex Pereira	Jamahal Hill	A	A
31	Max Holloway	Justin Gaethje	A	A
32	Arman Tsarukyan	Charles Oliveira	B	A
33	Bo Nickal	Cody Brundage	A	A
34	Jiri Prochazka	Aleksandar Rakić	A	A
35	Aljamain Sterling	Calvin Kattar	A	A
36	Sean O'Malley	Marlon Vera	A	A
37	Dustin Poirier	Benoît Saint Denis	A	A
38	Gilbert Burns	Jack Della Maddalena	B	B
39	Curtis Blaydes	Jailton Almeida	A	A
40	Ilia Topuria	Alexander Volkanovski	B	A
41	Robert Whittaker	Paulo Costa	A	A
42	Ian Garry	Geoff Neal	A	A
43	Merab Dvalishvili	Henry Cejudo	B	A
44	Dricus du Plessis	Sean Strickland	A	A
45	Tom Aspinall	Sergei Pavlovich	B	A
46	Islam Makhachev	Alexander Volkanovski	A	A
47	Khamzat Chimaev	Kamaru Usman	A	A
48	Magomed Ankalaev	Johnny Walker	B	A
49	Sean Strickland	Israel Adesanya	B	A
50	Alexander Volkov	Tai Tuivasa	A	A

Figure 2.2 Matchup results

Fight Number	**Results**
1	Model F1 Score: 0.6975 The predicted winner is: Roman Dolidze Confidence: 59.72%
2	Model F1 Score: 0.7065 The predicted winner is: Max Holloway Confidence: 55.64%

3		Model F1 Score: 0.7065 The predicted winner is: Khamzat Chimaev Confidence: 54.88%	
4		Model F1 Score: 0.7065 The predicted winner is: Magomed Ankalaev Confidence: 65.77%	
5		Model F1 Score: 0.7065 The predicted winner is: Lerone Murphy Confidence: 57.99%	
6		Model F1 Score: 0.7065 The predicted winner is: Shara Magomedov Confidence: 74.26%	
7		Model F1 Score: 0.7065 The predicted winner is: Michel Pereira Confidence: 62.23%	
8		Model F1 Score: 0.7065 The predicted winner is: Kyler Phillips Confidence: 50.00%	
9		Model F1 Score: 0.6980 The predicted winner is: Darren Elkins Confidence: 70.35%	
10		Model F1 Score: 0.6665 The predicted winner is: Tatsuro Taira Confidence: 85.95%	
11		Model F1 Score: 0.6974 The predicted winner is: Chidi Njokuani Confidence: 67.22%	
12		Model F1 Score: 0.7354 The predicted winner is: Grant Dawson Confidence: 54.03%	
13		Model F1 Score: 0.6946 The predicted winner is: Daniel Rodriguez Confidence: 58.51%	

14		Model F1 Score: 0.6886 The predicted winner is: Alex Pereira Confidence: 76.29%
15		Model F1 Score: 0.6769 The predicted winner is: Mario Bautista Confidence: 52.95%
16		Model F1 Score: 0.6801 The predicted winner is: Kevin Holland Confidence: 56.18%
17		Model F1 Score: 0.7089 The predicted winner is: Benoit Saint Denis Confidence: 62.22%
18		Model F1 Score: 0.6993 The predicted winner is: Brendan Allen Confidence: 64.93%
19		Model F1 Score: 0.7037 The predicted winner is: Bryan Battle Confidence: 73.67%
20		Model F1 Score: 0.7037 The predicted winner is: Morgan Charriere Confidence: 82.80%
21		Model F1 Score: 0.7239 The predicted winner is: Jiri Prochazka Confidence: 81.65%
22		Model F1 Score: 0.7239 The predicted winner is: Diego Lopes Confidence: 54.29%
23		Model F1 Score: 0.7294 The predicted winner is: Anthony Smith Confidence: 66.88%
24		Model F1 Score: 0.7294 The predicted winner is: Islam Makhachev Confidence: 76.21%

25		Model F1 Score: 0.7294 The predicted winner is: Sean Strickland Confidence: 68.40%
26		Model F1 Score: 0.7058 The predicted winner is: Alex Morono Confidence: 61.76%
27		Model F1 Score: 0.7058 The predicted winner is: Jose Aldo Confidence: 52.39%
28		Model F1 Score: 0.7339 The predicted winner is: Vitor Petrino Confidence: 59.57%
29		Model F1 Score: 0.7339 The predicted winner is: Caio Borralho Confidence: 87.31%
30		Model F1 Score: 0.6955 The predicted winner is: Alex Pereira Confidence: 52.08%
31		Model F1 Score: 0.6955 The predicted winner is: Max Holloway Confidence: 52.58%
32		Model F1 Score: 0.6955 The predicted winner is: Charles Oliveira Confidence: 55.66%
33		Model F1 Score: 0.6955 The predicted winner is: Bo Nickal Confidence: 54.99%
34		Model F1 Score: 0.6955 The predicted winner is: Jiri Prochazka Confidence: 66.69%
35		Model F1 Score: 0.6955 The predicted winner is: Aljamain Sterling Confidence: 59.03%
36		Model F1 Score: 0.6955 The predicted winner is: Sean O'Malley Confidence: 52.88%

37	Model F1 Score: 0.6955 The predicted winner is: Dustin Poirier Confidence: 51.53%
38	Model F1 Score: 0.6955 The predicted winner is: Jack Della Maddalena Confidence: 67.53%
39	Model F1 Score: 0.6955 The predicted winner is: Curtis Blaydes Confidence: 54.16%
40	Model F1 Score: 0.7075 The predicted winner is: Alexander Volkanovski Confidence: 73.87%
41	Model F1 Score: 0.7075 The predicted winner is: Robert Whittaker Confidence: 59.15%
42	Model F1 Score: 0.7075 The predicted winner is: Ian Garry Confidence: 68.55%
43	Model F1 Score: 0.7075 The predicted winner is: Henry Cejudo Confidence: 59.42%
44	Model F1 Score: 0.7075 The predicted winner is: Dricus Du Plessis Confidence: 56.60%
45	The predicted winner is: Sergei Pavlovich Confidence: 65.64%
46	Model F1 Score: 0.6975 The predicted winner is: Islam Makhachev Confidence: 70.40%
47	Model F1 Score: 0.6975 The predicted winner is: Khamzat Chimaev Confidence: 74.02%
48	Model F1 Score: 0.6975 The predicted winner is: Johnny Walker Confidence: 60.34%

49		Model F1 Score: 0.6975 The predicted winner is: Israel Adesanya Confidence: 62.85%
50		Model F1 Score: 0.6975 The predicted winner is: Alexander Volkov Confidence: 60.89%