Bachelor's thesis

# Machine learning in combat sports

***Mikoláš Bartoš***

Department of Computer Science
Supervisor: Ing. Matej Uhrín

10$^{\text{th}}$ August, 2021

# Acknowledgements

I want to express my gratitude to my supervisor Ing. Matej Uhrín for his guidance and patience throughout the preparation of this work. My thank you goes also to my dear friends and family, whose unconditional support has been tremendously helpful.

Finally, I would like to thank my school and faculty for the very opportunity to be writing this work.

# Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as school work under the provisions of Article 60(1) of the Act.

In Prague on 10$^{\text{th}}$ August, 2021 ......................

Czech Technical University in Prague
Faculty of Electrical Engineering


## Citation of this thesis

Bartoš, Mikoláš. *Machine learning in combat sports*. Bachelor's thesis. Czech Technical University in Prague, Faculty of Electrical Engineering, 2021.

# Abstrakt

Tato bakalářská práce zkoumá příležitosti pro generování profitu na sázkařských trzích pro smíšená bojová umění (MMA) za pomoci strojového učení. Oficiální data z Ultimate Fighting Championship (UFC) byla shromážděna za účelem předpovídání výsledků budoucích zápasů pomocí navržené neuronové sítě. Experimentální ztrátové funkce byly použity ke snížení korelace s předpověďmi bookmakera s cílem využít výhodnější pozice sázkaře v porovnání s bookmakerem. Kellyho kritérium a jeho alternativní podoba pro více souběžně konaných zápasů byly následně aplikovány jako kritérium pro alokaci finančních zdrojů na sázkařské příležitosti. Pro zhodnocení navrženého modelu společně s dvěma strategiemi sázení byla využita metoda bootstrap, která zvyšuje pravděpodobnost, že dosažený profit není výsledkem náhody.

**Klíčová slova** bojové sporty, smíšená bojová umění, sázkařské trhy, strojové učení, neuronové sítě, sportovní analýza, strategie sázení

# Abstract

This work examines the opportunities for profit generation on the mixed martial arts (MMA) betting market using machine learning. Official data from the Ultimate Fighting Championship (UFC) was acquired and processed to be used by the proposed neural network model to predict fight outcomes. Experimental loss functions decreasing correlation with bookmaker's estimates were used in the training process to exploit the discussed advantage a bettor holds over a bookmaker. The Kelly Criterion and its alternative for simultaneous games were then applied as wealth allocation policies on historical odds. The model and the two betting strategies were assessed using the bootstrap method to rule out any randomness of the achieved betting returns.

**Keywords** combat sports, mixed martial arts, betting markets, machine learning, neural networks, sport analysis, betting strategies

# Contents

# List of Figures

# List of Tables

Chapter 1

# Combat sports

To defeat an opponent using allowed techniques in a one-on-one fight is the ultimate objective of combat sports. Rules of individual combat sports, though, differ a lot. Differences are found mainly in the techniques contestants are allowed (or forbidden) to execute, but even definitions of victory vary across different sports.

## 1.1 History

Thousands of years old cave paintings depicting men taking part in wrestling activities from different corners of the world serve as proof, that forms of mock combat were a worldwide phenomenon long before globalization took place.

In terms of modern sport, boxing was the first combat sport to achieve mass recognition. Amateur boxing has been part of the Olympic Games ever since their re-introduction in 1904 and professional boxing was undoubtedly one of the most popular sports in the 20$^{th}$ century, producing celebrities like Muhammad Ali or Mike Tyson. Many more combat sports emerged during the 20$^{th}$ century though, usually built on the foundations of various martial arts or their combinations.

And it was the blatant differences across different combat sports and martial arts that gave birth to the arguably most popular combat sport of the 21$^{st}$ century which will be the main point of focus of this thesis.

## 1.2 Mixed Martial Arts

The name of the wealthiest promotion - The Ultimate Fighting Championship - tells a lot about the nature of Mixed Martial Arts (hereafter referred to as MMA). It was born to find the superior fighting style. The initial lack of rules lured masses of fans, but eventually proved controversial and resulted in political backlash.

In the search for lawfulness and thus business profitability, efforts were made to standardize MMA in the early 21st century. Despite there still being countries not recognizing MMA as a sport or even banning it, these now represent minority and their numbers are further diminishing. Globally, though, most big markets already recognize MMA as a legitimate sport rather than a no-rules spectacle. This transition allowed sports betting companies to start booking the fights and, consequently, MMA to become dubbed the world's fastest-growing sport.

### 1.2.1 Fighting promotions

As in professional boxing, MMA events are organized by promotions. The aforementioned Ultimate Fighting Championship (UFC) is by far the most successful and popular promotion worldwide. Examples of other promotions are Absolute Championship Berkut based in Russia or the Czech-Slovakian Oktagon. The rules might differ slightly across promotions, all must comply with the Unified Rules of MMA [1], though.

### 1.2.2 Techniques

Techniques used in MMA can be categorized into two categories: striking and grappling. Striking techniques include kicks, knee strikes, punches, and elbow strikes. Grappling techniques include clinch holds, pinning holds, submission holds, sweeps, throws, and takedowns.

### 1.2.3 Bout outcome

A fight usually results in a victory achieved by:

- judges' decision after an allotted amount of time has elapsed or
- knockout/technical knockout or
- submission.

Judges' decision can result in a draw, this is, however, very unusual. Disqualification or 'No contest' decision can occur in case of breach of rules.

### 1.2.4 Rounds

Traditionally, MMA matches are separated into 5-minute rounds. The majority of fights consist of three rounds, title bouts being an exception and lasting five rounds. There is a 1-minute break between rounds.

### 1.2.5 Combat area

All MMA contests take place in either a cage or a ring that meet certain requirements.

### 1.2.6 Attire

All fighters are required to fight with gloves and a mouthguard for protection. Male fighters must wear shorts, a protective genital cap, and be barechested. Female fighters ought to wear short shorts and a snug-fitting top such as sports bras.

### 1.2.7 Weight Divisions

To ensure fairness and an entertaining spectacle, only fighters of similar weight can fight each other. Promotions define their weight divisions with strict boundaries and every match is scheduled for a fixed weight division.

### 1.2.8 Events

On the professional level, fights are not stand-alone events. Rather than that, about a dozen of fights are usually grouped to form a fight night.

## 1.3 Betting

Betting has accompanied sport ever since the ancient Olympic Games [2] and the link is just as tight when it comes to combat sports.

### 1.3.1 Betting opportunities

Due to the ever-increasing regularity of sports events and accessibility of online gambling services, the number of opportunities bettors can place their bets on keeps growing. In combat sports, two types of betting opportunities can be distinguished: betting on a winner and propositional bets.

Betting on the winner of a fight is the most traditional form of betting. Noticeably in MMA, a bet cannot be placed on a draw even though it is a possible outcome.

Propositional bets allow betting on a fight in other ways. Bets can be placed on whether the fight would end in a submission, knockout, or judges' decision, other opportunities offer odds on the length of the fight, etc. These opportunities vary a lot across different bookmakers and they also depend on the eminence of a fight.

### 1.3.2 Odds

Betting odds reflect the bookmakers' estimation of how likely is an event to happen. Depending on available information and market movements, odds can change in time. From this perspective, we distinguish between opening odds (first odds issued by the bookmaker on an opportunity) and closing odds (odds just before the start of the event).

### 1.3.3 In-play betting

Online gambling gave birth to a whole new form of betting. In-play betting refers to wagering on an event that has started but not yet finished. Here, gamblers have the option to continue to bet once an event has started, and adapt their bets depending on how the event is progressing [3].

In UFC, the popularity of in-play betting is still very low with only 8% of bets placed in-play, compared to 70% in tennis [4]. This, however, is likely to change thanks to the UFC Event Centre which is the industry's first sports betting product created specifically for a major sports brand offering live statistics and live betting opportunities [5].

Chapter 2

# Existing research

With the surge in popularity and with a substantial amount of publicly available data, MMA has recently enjoyed an increasing volume of analysis.

## 2.1 Sport analysis

Most of the works have focused on describing the sport in terms of underlying numbers, trying to objectively find trends, patterns, implicit rules, or any sort of regularity in otherwise very subjective and unpredictable sport. The ultimate goal of these studies is to find implications with regards to improving training design, training efficiency, or in-fight strategies.

In [6], authors tried to determine performance indicators (and their combinations) contributing the most to the bout outcome. Rate-dependent data (relative to fight length) proved superior when analyzing data for an outcome. Their findings, that landed ground strikes, grappling activity, and striking accuracy are the decisive indicators, are, however, in sharp contrast with the results of a different work [7] which identifies striking while keeping distance as the ”best approach to increase the athletes' chance of performing well in a bout.” The authors contemplate the difference potentially being caused by the usage of different statistical methods or the usage of parametric data in a non-parametric setting. In other research, [8] provides insights into statistical differences between individual weight classes, frequencies of actions in individual rounds of a fight, or the ratio of high and low-intensity effort.

## 2.2 Outcome prediction

Another way to use data, statistics and machine learning methods is to try to predict future outcomes, in our context meaning prediction of the winner of a bout yet to occur. The main dissimilarity between these two approaches stems from the different information used to train classifiers (training data). The already presented studies' training data consisted of data collected exclusively within the given fight. This approach yields valuable information about how much each variable contributes to the final outcome. However, it cannot be used for predictive purposes as the necessary data is only available after the fight has finished.

Arguably due to the unpredictability and novelty of MMA, the research conducted on predicting outcomes is relatively sparse in this domain.

In [9], the author uses logistic regression for building a prediction model using cumulative career statistics for each fighter prior to the examined fight. He compares the performance of models trained on data including either basic count variables (e.g. total strikes landed) or second-level variables (e.g. striking ratio). Separately, the former outperformed the latter, but the best predictive accuracy was achieved by combining the most significant variables from both categories. Such model was also superior to simple prediction models based on random chance or winning percentage.

In [10], multiple learning algorithm for UFC fight prediction are compared. The Support Vector Machine (SVM) method proved to be the most resilient and, paradoxically, achieved the best performance when trained on raw data. Worth noting, the authors believe the used dataset has much room for improvement that could, possibly, have a positive impact on the performance of the whole model.

Even though these works show promising results, they do not try to fully optimize the models for either predictive accuracy or profitability. Some methods, for example, neural networks, multi-level models, or the introduction of more advanced features have all been untapped in the context of MMA, as has been the application of machine learning methods on MMA betting markets.

More profound research of this type has been carried out in other sport disciplines.

## 2.3 Other sports

Apart from other combat sports, such as boxing, which to our knowledge is yet to be analyzed in any way, drawing comparisons to other disciplines is fairly difficult due to the one-on-one nature of MMA and the fact, that fighters usually take months-long breaks between individual bouts. Among popular sports, the highest degree of similarity with MMA can be found in tennis, table tennis, and badminton.

Compared to previous state-of-the-art tennis prediction models, improvement in profitability on tennis betting markets has been achieved with an artificial neural network-based model in [11]. The model owes its success partly to weighting historical matches during feature extraction. More relevant matches such as those played recently or on a given surface carry more weight than other matches.

Instead of maximizing predictive accuracy, authors in [12] introduce and confirm a hypothesis that ”correlation of outcome predictions with the bookmaker’s predictions is detrimental for the bettor, and that suppressing such correlation will result in models allowing for higher profits”. Using convolutional neural networks and adapting modern portfolio theory they arrive at a model systematically generating cumulative profits in experiments on NBA data. Models trained with decorrelation loss function yielded higher profits, despite lower prediction accuracy.

An extensive overview of the use of machine learning in sport outcome prediction is provided in [13], where over a hundred papers were analyzed, yet no piece on the topic of combat sports was included in the final work.

Chapter **3**

# Problem definition

This work aims to maximize profit on the MMA betting market with a focus on long-term robustness. This includes finding a predictive model and implementing a betting strategy.

## 3.1 Constraints

Since propositional bet offers vary a lot among individual fights, we limit ourselves to wagering on an outright winner.

We also leave out any in-play betting opportunities as this would require collection and processing of real-time data which is beyond the scope of this work.

## 3.2 Measurement of profit

Return on investment will be used throughout this work as a measure of profitability. It uses the following simple formula:

$$ROI = \frac{W^{cur} - W^0}{W^0} \tag{3.1}$$

where

$W^{cur}$ ... Wealth in the moment of measurement
$W^0$ ... Initial wealth

Chapter 4

# Proposed solution

This chapter describes the proposed solution to achieving the goal set in Chapter 3. It explains in detail the data and the methods used, plus the reasoning for their selection.

Models were implemented in Python using the PyTorch library [14], data were manipulated using the Pandas library [15]. The entire codebase, plus information on how to navigate within the project, can be found in a public repository [16].

## 4.1 Data

Historical data about UFC fights were acquired from publicly available online sources. Three types of data were obtained: fighters' information and fights stats from the official UFC website ufcstats.com, and betting odds from bestfightodds.com.

### 4.1.1 Fighter details

The fighter detail pages at ufcstats.com provide information on all fighters who have ever taken a fight in the UFC. It contains personal information such as date of birth, height, weight, reach and stance and also career statistics such as the average number of significant strikes landed per minute.

The provided career statistics, however, hold no value to us. That is because the values are calculated relative to the day the data were acquired, which would represent future information in the training process of a classifier. Thus, these data were discarded and will not appear further in this work.

A similar argument could be raised regarding the fighters' personal information, as, apart from the date of birth, all other attributes are potentially volatile. This issue may be comfortably ignored for the height and reach of a fighter since no fighter younger than 19 years or older than 47 years has ever taken part in a UFC bout and this range provides certainty of little-to-no change in these aspects. As per weight, we choose to discard this information as well thanks to the fact that fighters often change weight divisions throughout their careers.

The stance attribute describes which foot the fighter favours as his front foot when facing an opponent. Three categories of stance exist - orthodox (right foot preference), southpaw (left foot preference), and switch (no preference). With the assumption, that any change in a stance within a career would result in the fighter falling into the switch category, we decide to keep using this piece of information.

### 4.1.2 Fights data

This dataset contains information on individual fights. It includes basic information about the bouts such as names of participants, weight division, result, and length of a fight, but importantly also statistics recorded during the fight.

In total 28 measured statistics were obtained for each fight. Generally speaking, they describe the number of strikes landed or attempted by each fighter, the part of the opponent's body they targeted, takedown and submission attempts, knockdowns, reversals, and for how long they controlled the opponent.

| R_TOTAL_STR. | R_TD | R_SUB_ATT | R_CTRL | R_HEAD | R_BODY | R_DISTANCE | R_GROUND |
|---|---|---|---|---|---|---|---|
| 9 of 14 | 0 of 2 | 0 | 2:12:00 | 6 of 11 | 1 of 1 | 9 of 14 | 0 of 0 |

Table 4.1: Example of the statistics collected within a fight (does not include all collected variables)

### 4.1.3 Odds data

Closing odds on the outright winner for 4240 UFC fights that took place between 21$^{st}$ March, 2010 and 14$^{th}$ March, 2020 were obtained from bestfightodds.com. Hereafter, we only take into consideration these 4240 fights we managed to obtain odds data for and the contestants who took part in them.

## 4.2 Feature extraction

### 4.2.1 MMA bout representation

Before explaining how data were manipulated to produce features, let us first discuss the representation of data in the learning and prediction process.

To train a predictive model using supervised machine learning methods, a set of labeled examples is needed. In the case of MMA bout prediction, one example corresponds to one historical bout. Each entry in a training set consists of two parts:

1. A vector of features (X), in our case representing the characteristics of the bout and the contestants

2. A label (y), in our case representing the bout outcome.

A trained model can then be used to compute result probability estimation for future bout using a vector of the same features describing the upcoming match.

### 4.2.2 Bout outcome representation

In a UFC bout, fighters are distinguished by the colour of their corner. The higher ranked contestant fights out of a red corner, their opponent out of a blue corner. Hereafter, I will refer to the former as a 'Red fighter' and the latter as a 'Blue fighter'. Since we are restricted to betting on an outright winner, our representation of bout outcome is equivalent to a representation of bout winner, which only allows two values thus making it a binary representation. The label value can be accordingly defined as:

$$y = \begin{cases} 1, & \text{if Red fighter won} \\ 0, & \text{if Blue fighter won} \end{cases}$$

All matches that did not end in a victory of either contestant were removed from the training dataset.

### 4.2.3 Missing values

Thanks to different sources of data, inconsistencies in the fighters' names occurred across individual datasets resulting in us not being able to match some entries. Furthermore, some fighters' personal information is also missing. This reduces the number of fights with all information available to 3712.

### 4.2.4 Historical averaging

To be able to effectively predict an outcome of a fight, information on both fighters is needed. As described in 4.1.1, some values are easily accessible before the bout commences, however, for the vast majority of performance indicators, values must be estimated based on past performances. Historical averaging represents a simple method of doing so.

In Table 4.2, the process of obtaining historical averages is illustrated on the Khabib Nurmagomedov vs. Pat Healy fight from 21$^{\text{st}}$ September, 2013. In this example, the number of significant strikes landed by Nurmagomedov is estimated by finding all his previous bouts, extracting the desired statistics from each respective bout, and computing an arithmetic mean from these values.

| Fight | Significant strikes landed | Fight length (s) |
|---|---|---|
| vs. Trujillo | 23 | 900 |
| vs. Tavares | 22 | 115 |
| vs. Tibau | 25 | 900 |
| vs. Shalorus | 35 | 848 |
| Per bout average | 26.25 | |
| Per minute average | 2.28013 | |

Table 4.2: Illustration of historical averaging

Using simple arithmetic mean has its flaws that stem mainly from the nature of MMA. Looking at the example, the higher the estimated value of significant strikes landed for Nurmagomedov, the better we expect he had done in his past fights. Yet in one of the most glorified bouts in UFC history, the winner landed mere 5 significant strikes as Connor McGregor achieved victory by knockout within 13 seconds of the first round. MMA bouts can last anywhere between a couple of seconds and 25 minutes, therefore adjusting estimates to bout length should provide more accurate estimates.

For that reason, instead of using naive 'per bout' averaging, we decide to obtain per minute averages which can be defined as:

$$\boldsymbol{x} = \frac{\sum_{i=1}^{n} x_i}{\frac{\sum_{i=1}^{n} t_i}{60}} \tag{4.1}$$

where:

$\boldsymbol{x}$ ... target statistic estimate
$x_i$ ... statistic recorded in $i$th past fight
$t_i$ ... length of the $i$th past fight in seconds

### 4.2.5 Weighted averaging

Even when accounting for the length of a fight, we still assume that each of the fighter's past bouts contributes equally to their current state, which is a rather naive approach, considering the wide range of factors influencing the significance of a past bout in the assessment of fighters current abilities. This problem can be addressed by implementing weighted averages, giving higher weights to bouts with higher relevance.

With regards to 4.1, we define weighted per minute average as:

$$\boldsymbol{x} = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\frac{\sum_{i=1}^{n} w_i \cdot t_i}{60}} \tag{4.2}$$

where:

$w_i \ldots$ weight of $i$th past fight, $w_i \in (0, 1)$

The most evident factor that influences the relevance of a fight is the amount of time elapsed since it took place as we can expect that fighter's recent bout reflects their current abilities more accurately than an older bout. This can be reflected by assigning weights using different functions. We opt to compute the weights by an exponential function:

$$w_i = \delta^{\Delta t} \tag{4.3}$$

where

$\delta \quad \ldots$ discount rate, $\delta \in (0, 1)$
$\Delta t \ldots$ time elapsed since the $i$th fight

Weights could be potentially fine-tuned to reflect things such as past head-to-head matches or fights with common opponents. Another potentially huge factor might be the styles of previous opponents as before each match, one of the most discussed topics in the MMA community is the difference in styles of each fighter and how they can cope against a given style. If these debates are justified, past bouts against fighters with similar attributes to those the contestant is about the face might be assigned higher weights.

### 4.2.6 Form-related features

It is hard to deny that psychology has a great effect on all performers, MMA fighters included. Fighters' mental states could therefore offer a very important indicator of the outcome of an upcoming fight. Unfortunately, such information is difficult to represent and even more difficult to capture without access to fighters.

But what we can do with our data, is to take into account the recent and overall form of a fighter, assuming that past performances have an effect on future performances. For this reason, we added three features to our existing set of features - winning streak, losing streak, and winning percentage.

### 4.2.7 Debuts

An issue that arises with historical averaging is the representation of fighters making their UFC debuts. Despite the fact that the fighter entering UFC for the first time could be very experienced with tens of wins to his name, his UFC record is nonexistent, making it impossible to extract features from our bout database. We decide to leave out all fights including at least one fighter making their debut.

## 4.3 Model selection

Any supervised machine learning method could be used for the problem of predicting MMA bouts [10]. A technique that is yet to be explored in the context of combat sports, despite its global popularity in both research and industry, is the artificial neural network.

### 4.3.1 Artificial neural networks

Inspired by the functioning of the human brain, an artificial neural network is based around the concept of neurons that combine inputs to an output which is then passed to other neurons. Neurons are represented by nodes which are typically organized into layers. The output of a neuron in a layer is first computed using its input values and weights and then fed forward to all neurons in the following layer. The output is calculated using a nonlinear function:

$$f(x) = \sigma(\boldsymbol{w}^T \boldsymbol{x} + b) \tag{4.4}$$

where

$\sigma$ ... nonlinear activation function
$\boldsymbol{w}$ ... vector of neuron's weights
$\boldsymbol{x}$ ... vector of input values
$b$ ... bias

### 4.3.2 Network architecture

Regarding the network structure of our choice, single hidden layer neural networks with any continuous bounded nonlinear activation functions can form decision regions with arbitrary shapes [17] and approximate any continuous function [18], therefore we stick to such architecture in this work.

Our output layer, producing the probability of Red fighters' win, employs the Sigmoid activation function which is useful for the representation of probability as it only produces a positive outcome in the range from 0 to 1:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{4.5}$$

The hidden layer could be used with a plethora of activation functions. We decide to implement a novel activation function dubbed either Swish or Sigmoid Linear Unit (SiLu), which has the potential to outperform the much recognized ReLu function [19]. Its definition uses the Sigmoid function defined in 4.5:

$$SiLu(x) = x \cdot \sigma(x) \tag{4.6}$$

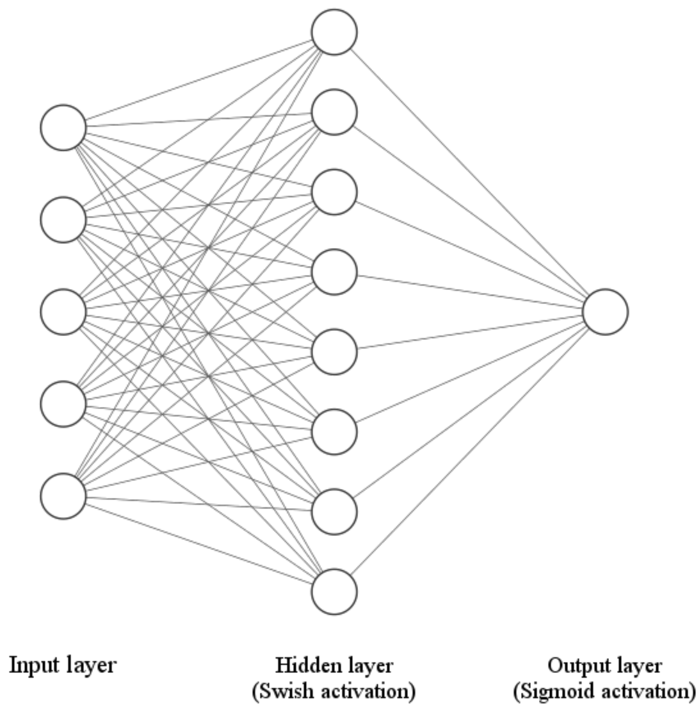


Figure 4.1: Artificial neural network architecture

### 4.3.3 Training algorithm

Neural networks can be trained in a variety of ways, but the most common approach is called back-propagation. In the training process, the model computes and propagates values through the network to the output layer which is called the forward pass. The quality of the output values in comparison with the ground truth target values is then evaluated using a chosen criterion - a loss function. The results are propagated back (hence back propagation) through the network in order to update the model's weights [20].

The network iterates over many cycles until reaching an acceptable loss function value. Most commonly, Gradient-based methods, which are known to converge to a local minimum, are utilized for minimization of the loss function. The weights update in iteration t+1 can be described as [21]:

$$\boldsymbol{w}_{t+1} = w_t - \mu \cdot \nabla J(\boldsymbol{w}_t) \tag{4.7}$$

where

$\mu$ ... learning rate
$\nabla J(\boldsymbol{w}_t)$ ... gradient of the loss function

In our implementation, we stick with the trusted algorithm of stochastic gradient descent (SGD) with momentum. The bouts are propagated through the network in batches of a certain size, the propagation of all training samples (in 1 or more batches) is labeled as an epoch.

### 4.3.4 Loss functions

Traditionally, binary cross-entropy (BCE) loss function is used in binary classification problems. However, the choice of a loss function is typically regarded as an empirical problem and for all convex loss functions, the sign of the minimum of the expected risk coincides with the Bayes optimal solution [22].

$$BCE = y \cdot log(x) + (1 - y) \cdot log(1 - x) \tag{4.8}$$

where

$x$ ... output of a neural network
$y$ ... ground truth target value

### 4.3.5 Regularization

Prevention of overfitting is one of the biggest challenges when searching for a classifier. Dropout represents a simple method addressing the problem for neural networks with little extra computational effort. The idea behind dropout is to simulate training of multiple neural networks at once, resulting in a superior classifier that is less prone to overfitting. This is achieved by randomly (with a predefined probability) dropping out units and their connections from a network during training. A network with $n$ nodes can then be seen as collection of $2^n$ different possible reduced neural networks [23].

Another cheap way to improve generalization is early stopping. In the optimal scenario, once overfitting begins to happen in the training process, validation loss begins to rise which might be seen as the moment to cease the training loop. In practice, validation loss might be reaching multiple local minima before severe overfitting emerges [24]. We utilize this fact by letting the training loop run over a fixed number of epochs and saving the model that achieved the all-time lowest validation loss.

### 4.3.6 Hyperparameter optimization

Based on experiments, observations, and suggestions in other works, some parameters were fixed and others' values were restricted to a reasonable range to reduce the computational complexity of hyperparameter optimization.

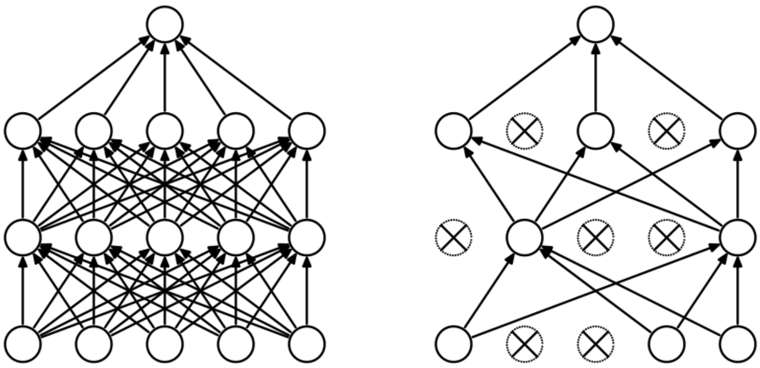

Figure 4.2: The left figure shows a standard neural network with two hidden layers. The right figure represents an example of a neural network reduced by applying dropout during the training process. Image retrieved from [23].

**Discount rate for time-discounting of past bouts**

After training a neural network with fixed parameters on different datasets using different discount rates, we measured the performance of the neural network on the validation dataset. The different discount rates yielded very similar results with no major trend or relationship apparent. We decide to fix the discount rate at $\delta = 0.4$ which marginally achieved the best validation loss, but we assume that the discount rate should not have drastic effects on the classifier.

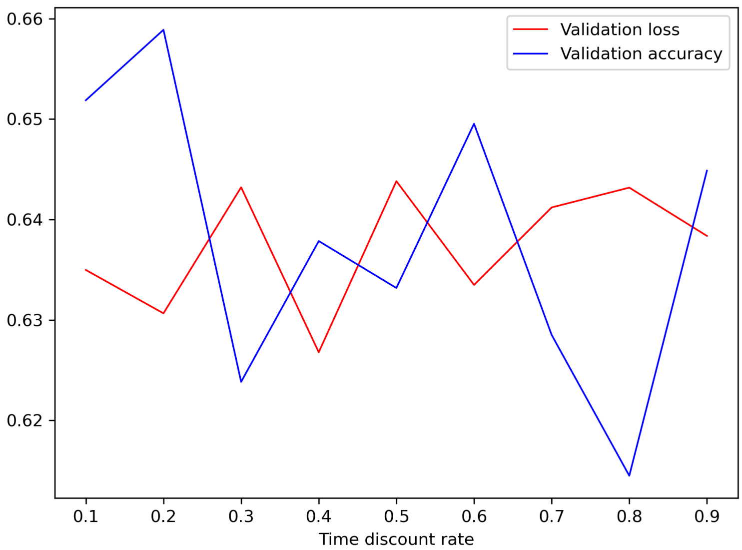


Figure 4.3: Results of a neural network trained using datasets with different discount rates for time-discounting

**Learning rate of the SGD optimizer**
With a relatively small training dataset, we can afford to set the learning rate low, which is said to significantly improve generalization on complex problems [25], without suffering from too long training times. Using a trial-error approach we arrived at the learning rate $\mu = 0.00008$.

The rest of the hyperparameters, including the number of nodes in a hidden layer, the momentum of the SGD optimizer, dropout rate, and batch size, were optimized using a hyperparameter optimization framework Optuna [26]. Computational resources for finding the optimal hyperparameters were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development, and Innovations Infrastructures.

## 4.4 Accuracy versus profit

Forecasting accurately and forecasting profitably are not equivalent. None of the two necessary implications, which are often implicitly present in (sports) forecasting studies, are valid. Both implications fail due to more complex relationships caused by the presence of a bookmaker, whose error distribution in comparison with bettor's error distribution vastly influences the betting returns. In some cases, profitable strategies might also be a result of randomness, but chances of arriving at such strategy decrease with an increasing number of bets placed [27].

More often than not, maximization of accuracy is the default approach when training a predictive model and that is well justifiable in many applications, where reaching the highest possible accuracy is the ultimate goal of the model. This might include models used by individual sportsmen or sports teams to analyze their own or opponent's game, ranking systems, bookmakers' models, or, specifically in the combat sports field, models for creating equal and therefore likely entertaining match-ups. However, such an approach can lead to sub-optimal results on the betting markets.

By aiming for a model that optimizes predictive accuracy, a bettor gives away one of the biggest advantages he holds over a bookmaker. While bookmaker has no option but to try to predict the real probabilities of uncertain events, a bettor has the freedom to choose the opportunities he finds profitable given the bookmakers' offered odds. In other words, the bettor does not need to predict the exact probability of any event, it is enough for him to identify opportunities where the bookmaker undervalued the probability and therefore offers higher odds. This task is substantially easier than the one bookmakers face [28].

This concept is rather theoretical and indeed, success on betting markets can be achieved with a model trained for maximal accuracy, but even in practice, it is becoming increasingly difficult to beat a bookmaker through a model with higher predictive accuracy as betting companies employ teams of data analysts and likely possess richer and more granular data from third-party data providers. Moreover, betting companies' odds are not a direct reflection of bookmakers' probability estimates since they incorporate a profit margin, which offers them space for error and, in the long run, guarantees them profit.

### 4.4.1 Decorrelation

As shown in [27], in situations where the bettor's predictive model achieves equal or lower predictive accuracy than the bookmaker's model, betting returns decrease with increasing error correlation between the two models. To put it another way, if a bettor is unable to beat a bookmaker in predictive accuracy (which is very likely as described in previous paragraphs), then having similar estimates as bookmaker results in lower profit. Therefore, it might be desirable for a bettor to intentionally decrease the correlation between his model's and bookmaker's probability estimates. I will further refer to this concept as decorrelation.

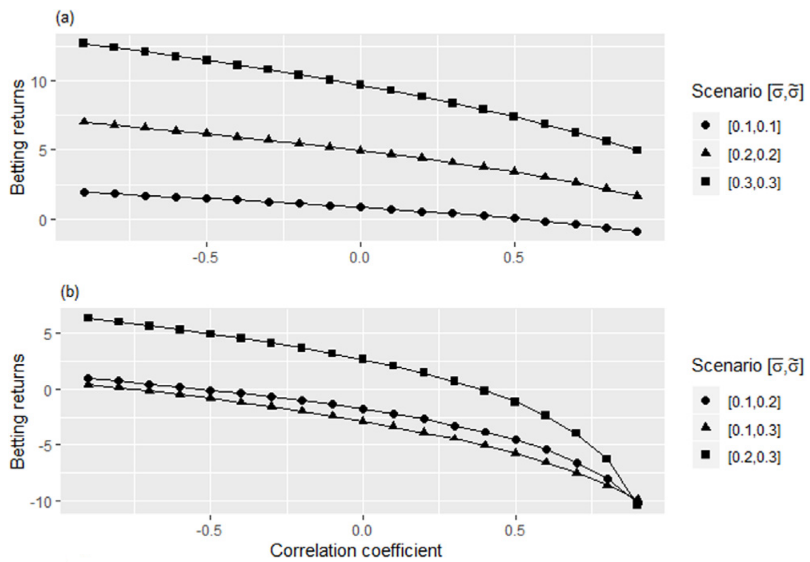


Figure 4.4: Effect of error correlation on betting returns. Numbers in the scenario descriptions refer to the bookmaker and bettor errors respectively. Image retrieved from [27].

In the context of machine learning, decorrelation can be encouraged in several ways. One approach could lie in playing with the weights of training examples based on the bookmaker's odds by highlighting the importance of opportunities, where the underdog (in terms of odds) succeeds, with higher weights. While the actual calculation of weights may differ as explored in [12], they all cause high-odds outcomes to contribute more to the training error.

This, in theory, forces the model to identify opportunities where bookmaker underestimates the underdog's chances.

Another, more subtle, neural network-specific solution lies in slightly changing the objective of the training process by adjusting the loss function. The idea is, that instead of only measuring our estimates' distance from the ground-truth, distance from the bookmaker's estimates is calculated too, and during the training process, the error is minimized while maximizing the distance from the bookmaker. Generally, we can define such loss function as:

$$Loss = D(M \parallel T) - c \cdot D(M \parallel B) \tag{4.9}$$

where

$D(M \parallel T)$ ... distance between model's estimates and ground-truth
$D(M \parallel B)$ ... distance between model's and bookmaker's estimates
$c$ ... decorrelation constant defining the significance of the decorrelation term

The balance between minimizing prediction error and maximizing distance from the bookmaker depends heavily on the decorrelation constant $c$ of the decorrelation term in 4.9. The optimal value of $c$ subsequently depends on the distance function we use in the decorrelation term as can be seen in 4.5. Numerous functions are available for computation of both distances in 4.9, however, we restrict ourselves to these four:

1. Binary cross-entropy (BCE)

   As established in 4.3.4, binary cross-entropy is the default function for the calculation of the distance between the model's estimates and ground truth.

2. Mean squared error (MSE)

   MSE is often used for the computation of the model's average error in comparison with ground truth. It was also suggested in [12] as a possible distance measure for the decorrelation term in 4.9.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(p_i - \hat{p}_i)^2$$

3. Kullback-Leibler divergence

   The Kullback-Leibler divergence [29] is a measure of how far distribution Q is from distribution P. It is not technically a statistical distance function due to its asymmetric nature, but in practice, it serves very similar purposes. For the two probability distributions on the same probability space X, it can be defined as:

$$D_{KL}(Q \parallel P) = \sum_{x \in X}^{N} P(x) \cdot log\left(\frac{P(x)}{Q(x)}\right)$$

4. Jensen-Shannon divergence

   The Jensen-Shannon divergence is another method of measuring the distance between two probability distributions. It is based on the Kullback-Leibler divergence but is tweaked to be symmetric and to always produce a finite value.

$$D_{JS}(Q \parallel P) = \frac{1}{2} D_{KL}(Q \parallel M) + \frac{1}{2} D_{KL}(P \parallel M)$$

   where

$$M = \frac{1}{2}(Q + P)$$

To reduce the complexity of the task at hand, we fix the function used for the calculation of the distance between the model's estimates and ground truth to be binary cross-entropy thus the number of variations of the two distance functions drops from 16 to 4.

The figure 4.5 shows, how individual distance functions used in the decorrelation term differ in sensitivity to the decorrelation constant $c$. For this reason, the value of $c$ ought to be optimized separately for each distance function. However, it is very unclear what the optimum is. Results in 4.4 suggest that the lower the correlation the higher the betting returns, yet this relationship is likely to be more complex in a real-life scenario. Usually, hyper parameters are selected based on the highest predictive accuracy or lowest value of loss function. None of these, though, are reasonable in our scenario where predictive accuracy is not the main objective and the loss function is directly influenced by the value of $c$. Consequently, betting returns on the validation set will be our criterion for selecting optimal hyper parameters' values.

| Decorrelation term | Optimal value of $c$ | Betting returns on validation set |
|---|---|---|
| BCE | 0.4 | 1.014 |
| MSE | 2.6 | 0.623 |
| KL | 0.5 | 1.198 |
| JS | 1.9 | 0.851 |

Table 4.3: Results of optimization of decorrelation constant $c$ for loss functions using different distance measures.

The results of optimization on the validation set 4.4.1 back up our hypotheses that different values of $c$ are required for each distance function. On

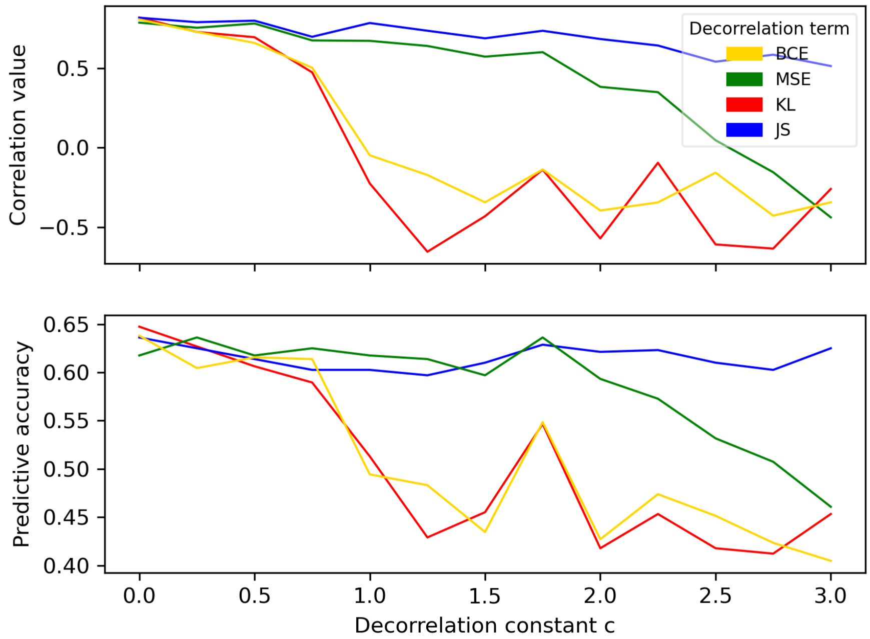


Figure 4.5: Effect of the decorrelation constant on predictive accuracy and correlation of model's and bookmaker's probability estimates.

the other hand, with such small margins in the betting returns and the small sample size of the validation set, it would be naive to jump to conclusions on the ordering of individual functions. We will further compare all four versions of our custom loss functions in Chapter 5.

## 4.5 Betting strategies

With the predictive model complete, the last step needed before placing bets is to determine the amount of money we want to bet on any single opportunity. Multiple approaches to optimal wealth allocation are being used by gamblers and investors trying to exploit different markets.

### 4.5.1 Reinvestment

All strategies can be executed in one of two ways - using a flat stake or reinvesting the obtained winnings. In the case of flat stake, all accumulated profit goes directly into the bettor's pocket and is not further used for betting.

Therefore, the fraction of the bettor's wealth he decides to allocate to an opportunity will always be a fraction of his initial budget.

In contrast, a bettor adopting the reinvestment strategy uses his achieved profit to extend his betting budget. When the bettor possesses a profitable model, then reinvestment obviously leads to much higher profits than flat stake as the amounts of money he is placing on opportunities grows with the growing bank.

Although there are scenarios, where constant bets are justified, reinvestment is the more sensible and realistic strategy in the sports betting context. Online betting services also limit the minimal figures one can withdraw from the betting account, hence withdrawing money after every winning bet is practically unrealizable.

All results on betting markets in this work were achieved employing the reinvestment strategy.

### 4.5.2 Kelly Criterion

One of the most common and significant wealth allocation strategies is the Kelly Criterion which optimizes wealth growth rate using logarithmic utility function [30]. The goal of maximization of average logarithmic growth rate for a single binary betting opportunity can be defined as:

$$\underset{\boldsymbol{b}}{\text{maximize}}\, \mathbb{E}[G(b)] = p \cdot \log(1 + (o-1) \cdot b) + q \cdot \log(1-b) \tag{4.10}$$

where:

$p$ . . . win probability
$o$ . . . offered odds
$b$ . . . fraction of our wealth we decide to bet
$q$ . . . loss probability

If we differentiate $\mathbb{E}[G(b)]$ with respect to $b$ and set the derivative equal to zero, we get the maximum and the optimal strategy:

$$b_{max} = \frac{p \cdot (o-1) - q}{o-1} \tag{4.11}$$

For games where the win and loss probabilities are known, the Kelly betting strategy guarantees optimal wealth growth with the zero probability of going bankrupt. In cases where probability distributions are not known, these guarantees, as might be expected, vanish. Therefore using such a strategy in cases where bettor only possesses probability estimates (as we do) is very likely to lead to overbetting and bankruptcy.

To mitigate the risks and avoid overbetting, we implement the fractional Kelly strategy, using a static constant $\omega \in (0, 1]$ to adjust the fraction of wealth we are betting. The constant $\omega$ will hereafter be referred to as Kelly fraction.

$$b = \omega \cdot \frac{p \cdot (o-1) - q}{o-1} \tag{4.12}$$

The value of $\omega$ affects the potential betting returns and the variance a bettor will experience. The higher the value, the more volatile wealth trajectories can be expected and the bigger is the risk of going bankrupt as can be seen in Figure 4.5.2.

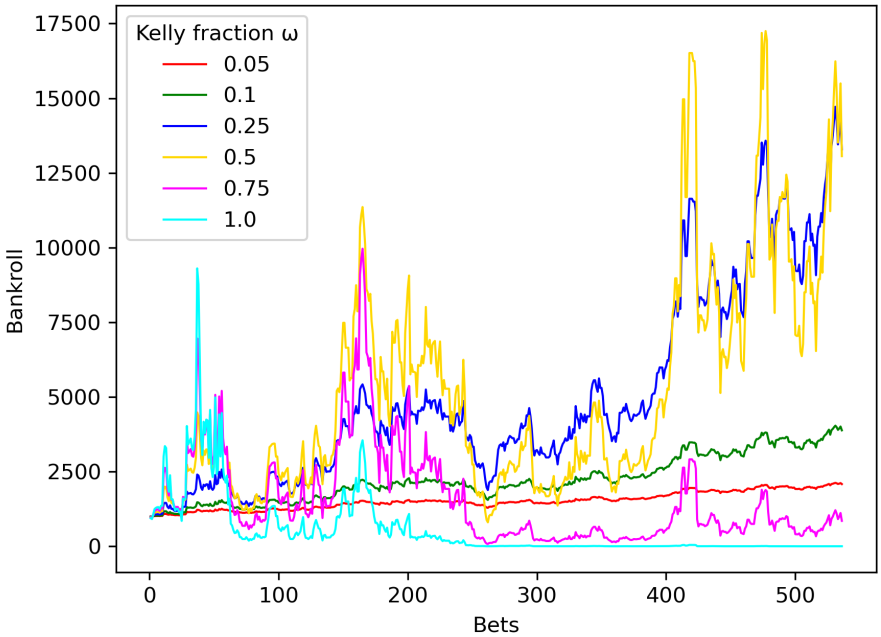


Figure 4.6: Wealth trajectories achieved with the same predictive model but different Kelly fractions used for determining the wealth the bettor places on each opportunity.

In this example, the riskiest strategy achieves the highest profit after the first approximately 50 bets but ends up bankrupt after a series of losses making the bettor unable to place any more bets. The second most aggressive strategy meets a similar fate as these two approaches are the only ones not to generate profit. This is of huge importance with regards to our goal which is long-term profit as specified in 3, hence the effects of the value of $\omega$ will be further examined after defining our evaluation framework.

### 4.5.3 Kelly strategy for simultaneous fights

As mentioned in 1.2.8, MMA fights are usually grouped into events called fight nights, where multiple fights take place one after another with about 30-minute breaks between them. Indeed, treating each fight as a unique binary opportunity in a sequential manner as described above is viable, yet it might prove advantageous to handle bouts within one fight night as simultaneous opportunities.

The main difference between these two approaches can be described using a simple example. Imagine yourself as a bettor with information on two upcoming fights. In the first one, you estimate the probability of the red fighter winning slightly higher than the bookmaker, in the second one, you are almost certain that the blue fighter will take the win, despite being the underdog in terms of odds (due to, for example, private information about an injury of the favourite). If you treat the opportunities separately in chronological order, the Kelly criterion will suggest allocating let us say 10% and 40% of wealth respectively. But would it not be better to allocate more money on the second opportunity and ignore the first one when you know there is a much higher chance of success and a larger payout?

What happens is that the sequential Kelly strategy does not take into account the opportunities of future fights, even though all information about these fights is already known at the moment of making the decision as this process will, in practice, take place just before the start of a fight night, and neither odds nor the possessed prediction model is likely to change during the duration of a single fight night.

To define the Kelly strategy for simultaneous games, we first need to generalize the Kelly criterion definition to include more than one opportunity:

$$\begin{aligned} \underset{\boldsymbol{b}}{\text{maximize}} \quad & \mathbb{E}[\log(\mathbf{R} \cdot \mathbf{b})] \\ \text{subject to} \quad & \sum_{i=1}^{K} b_i = 1.0, \; b_i \geq 0 \end{aligned} \tag{4.13}$$

The idea is obviously the same - we are trying to maximize the growth rate using the logarithmic utility function. What is different is the representation of opportunities as we need to accommodate multiple assets at once. For this reason, we introduce the return matrix $\mathbf{R}$, where each column stands for a single asset (a bet we can place). We also add to our return matrix a risk-free asset allowing our strategy to put money aside, hence the total number of opportunities for a fight night with $n$ fights is going to be $2n + 1$.

The matrix $\mathbf{R}$ will then consist of returns of all possible combinations of outcomes as each row in $\mathbf{R}$ represents one possible sequence of results. This gives us a matrix of size $2^n \times (2n + 1)$. For 2 simultaneous fights, $\mathbf{R}$ would look like this:

$$\mathbf{R} = \begin{bmatrix} o_r^1 & 0 & o_r^2 & 0 & 1 \\ o_r^1 & 0 & 0 & o_b^2 & 1 \\ 0 & o_b^1 & o_r^2 & 0 & 1 \\ 0 & o_b^1 & 0 & o_b^2 & 1 \end{bmatrix} \tag{4.14}$$

Where $o$ stands for odds, superscript is a fight identifier, and subscript represents the fighter (red or blue).

The matrix $\mathbf{b} \in \mathbb{R}^{(2n+1)\times 1}$ stands for the wealth fractions we shall place on each of the $2n+1$ available assets.

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{2n+1} \end{bmatrix} \tag{4.15}$$

The constraints on **b** defined in 4.13 are self-explanatory as one cannot possibly bet more money than he possesses.

The expectation of returns is then realized using our probability estimates for each opportunity.

$$\mathbb{E}[\log(\mathbf{R} \cdot \mathbf{b})] = \mathbf{p} \cdot \log(\mathbf{R} \cdot \mathbf{b}) \tag{4.16}$$

Reflecting the representation of **R**, the matrix $\mathbf{p} \in \mathbb{R}^{1\times 2^n}$ includes estimates of probabilities for each of the $2^n$ possible outcome sequences.

$$\mathbf{p} = \begin{bmatrix} p_1, & p_2, & \ldots, & p_{2^n} \end{bmatrix} \tag{4.17}$$

Continuing with the example of 2 simultaneous fights, the probability of the sequence defined on the first line in 4.14 would be

$$\begin{aligned} p_1 &= P(r^1) \cap P(r^2) \cap P(1) \\ &= P(r^1) \cdot P(r^2) \cdot P(1) \\ &= P(r^1) \cdot P(r^2) \end{aligned} \tag{4.18}$$

as individual fights are independent events. If we calculate the probabilities for the remaining outcome sequences in the same way, we arrive at

$$\mathbf{p} = \begin{bmatrix} P(r^1) \cdot P(r^2), & P(r^1) \cdot P(b^2), & P(b^1) \cdot P(r^2), & P(b^1) \cdot P(b^2) \end{bmatrix} \tag{4.19}$$

In practice, we implemented the Kelly strategy for simultaneous fights using the cvxpy optimization framework [31][32]. The number of simultaneous fights was fixed to 10 to mimic the real fight nights.

Similar to the original Kelly criterion for a single opportunity, overbetting is an issue when dealing with probability estimates. This can be tackled using the exact same trick of betting only fractions of the optimal value suggested by the Kelly strategy.

Chapter 5

# Results

## 5.1 Evaluation framework

### 5.1.1 Data separation

To be able to objectively evaluate and compare the performance of predictive models and betting strategies, it is appropriate to do so with predictions on data not used to train the model. This is usually done in one of two ways - using k-fold cross-validation or train-test split.

K-fold cross-validation represents an easy way of making less biased estimations of models' performances. It is performed by shuffling the dataset, splitting it into $k$ groups, and for each group training the model on remaining $(k-1)$ groups and evaluating the performance on the group itself. The performance of the model is then summarized by aggregating all $k$ individual performance indicators.

The train-test split method, on the other hand, is an even more straightforward solution where the dataset is separated into two subsets. The performance of the model is then evaluated based on the predictions and results achieved on the subset unseen by the machine learning algorithm during the learning process.

While k-fold cross-validation is the preferred choice when dealing with smaller datasets as we are, the train-test split method is the more sensible method to use in our scenario. The reasoning behind this decision is the distribution of fights in time.

Thanks to ever-increasing funding, advances in sport sciences and sports data analysis, many sports have changed significantly and the tactics used by contemporary sportsmen or sports teams are very different from those of just 10 years ago. A great example of this is football where players take much less low-value shots from long range and instead try to stay patient and create opportunities closer to the opponent's goal which present higher chances of scoring. Similarly, MMA has its trends too, and even though they are more difficult to identify, there is no doubt that current fighters utilize a different range of techniques and tactics compared to their predecessors.

Therefore, evaluation of betting results in MMA makes the most sense on the most recent data available. With that in mind, we opt for the train-test split approach and take away the most recent 20% of fights as a hold-out test set which we will evaluate performance on. The remaining 80% of fights are used in the training process of our neural network.

### 5.1.2 Bootstrap method

To increase the statistical significance of our findings and ensure that any generation of positive betting returns is not random and coincidental, we introduce a different technique that goes by the name of the bootstrap method. Generally speaking, it is a resampling method frequently used in statistics to more accurately estimate properties of observed data and the whole statistical distribution by repeatedly sampling random samples with replacement from the observed distribution realization [33].

We will use the method in a slightly different manner - to literary enlarge our test dataset and thus create significantly more ”alternative histories” on which we can test our model and betting strategies. Instead of having one sequence of $n = 536$ fights, we can retrieve hundreds or thousands of different sequences by randomly drawing samples of size $n$ with replacement. This allows us to further explore the robustness of the results achieved on betting markets.

### 5.1.3 Evaluation

With 4 proposed prediction models (trained using different decorrelation terms in the loss function) and 2 proposed betting strategies, we will now compare how they perform on the hold-out test dataset.

Each combination of model and betting strategy was evaluated using the following workflow.

1. Obtain probability estimates for fights from the test dataset
2. Draw 1000 or 100 bootstrap samples from the test dataset
3. Run a betting strategy on each of the bootstrap samples
4. Observe the achieved results

Due to different computational demands of the two proposed betting strategies, the number of bootstrap samples drawn was set at 1000 for the sequential Kelly and 100 for the simultaneous Kelly.

## 5.2 Betting returns

### 5.2.1 Sequential Kelly criterion

As shown in 5.2.1, all models were able to generate positive betting returns when conservative values of the Kelly fraction were applied in the betting strategy. Also, as expected, none of the models can sustain profitability with the riskier Kelly fraction values and the return on investment median of all models using the full Kelly strategy was very close or equal to $-1$ meaning at least half of the bootstrap sample histories ended in bankruptcy.

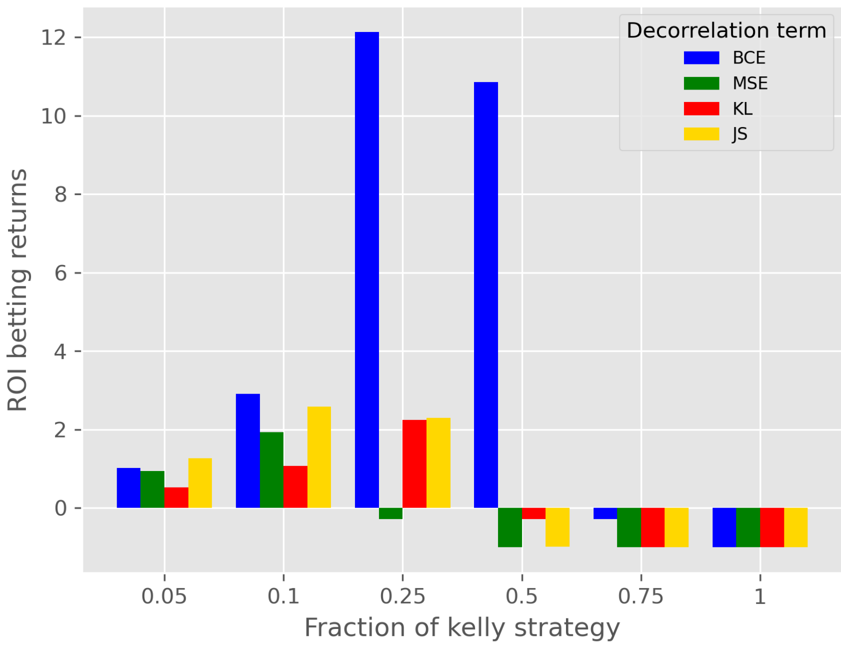


Figure 5.1: Median of betting returns of sequential Kelly betting strategy run on 1000 bootstrap samples with different loss functions and Kelly fractions.

Out of the four models, the one using binary cross-entropy as a measure of distance between model's and bookmaker's estimates in our custom loss function proves the most profitable with the ROI median of over 10 for certain Kelly fractions.

### 5.2.2 Simultaneous Kelly criterion

Similar patterns can be recognized in the results of the Kelly strategy for 10 simultaneous fights. Model using BCE in decorrelation term of loss function again outperforms the other models, but all of them successfully generate profit for the smaller Kelly fractions.

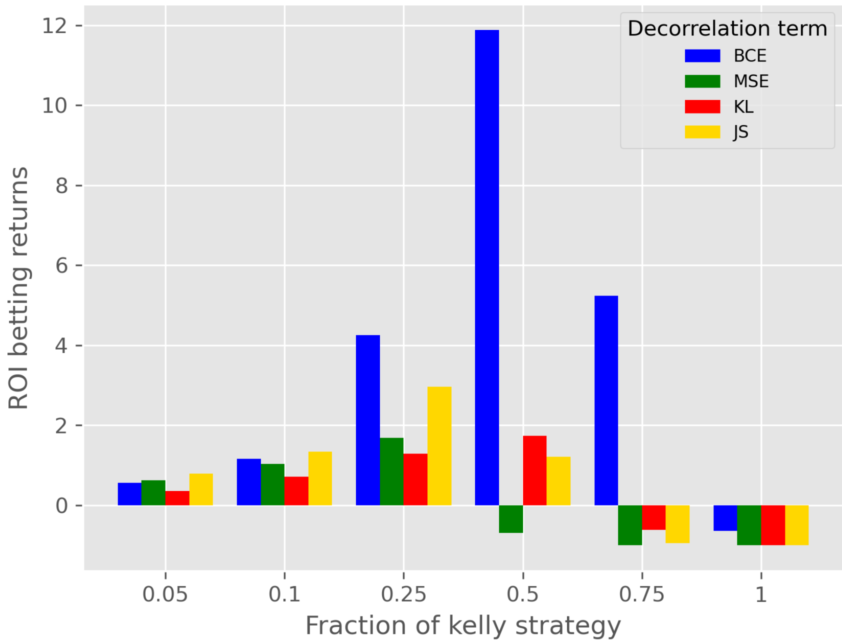


Figure 5.2: Median of betting returns of Kelly betting strategy for 10 simultaneous fights run on 100 bootstrap samples with different loss functions and Kelly fractions.

The main difference of the two strategies are higher betting returns achieved with Kelly fraction $\omega = 0.5$ and $\omega = 0.75$. This goes hand in hand with the notion of higher stability of the optimal Kelly for simultaneous games, as it has the opportunity to compare multiple opportunities at hand before deciding the wealth allocation. This might not result in higher betting returns than achieved by sequential Kelly (potential winnings cannot be reinvested immediately, but only after all simultaneous games have taken place), but should lower the associated risk.

## 5.3 Robustness

Despite very promising betting returns, the results presented so far do not paint the full picture. Although median values are, in our scenario, much more informative than simple mean values (ROI values range from $-1$ to $\infty$, therefore means are stretched by positive outliers), we still ought to take a closer look at the distribution of achieved returns.

### 5.3.1 Sequential Kelly criterion

A simple, yet telling indicator is the percentage of bootstrap samples we managed to achieve positive profit on. All models offer relatively reasonable ratios for low-risk Kelly fractions, but only the model using BCE in decorrelation term maintains likeable percentages for $\omega$ values other than 0.05 or 0.1.

| Decorrelation term \Kelly fraction | **0.05** | **0.1** | **0.25** | **0.5** | **0.75** | **1** |
|---|---|---|---|---|---|---|
| **BCE** | 99.2 | 98.0 | 93.3 | 78.7 | 47.7 | 19.3 |
| **MSE** | 84.3 | 78.1 | 48.3 | 5.9 | 0.2 | 0.0 |
| **KL** | 91.0 | 87.9 | 75.7 | 44.0 | 18.6 | 3.6 |
| **JS** | 92.9 | 90.3 | 71.1 | 21.2 | 1.6 | 0.0 |

Table 5.1: Percentages of simulation runs that resulted in a positive betting return for different Kelly fractions and decorrelation functions using the sequential Kelly betting strategy.

On the other end, we shall also examine the probabilities of reaching a state where betting is no longer possible. Bettors often define a risk constraint for their strategy to be reasonable such that

$$P(W^{min} \leq \alpha) \leq \beta \tag{5.1}$$

where

$W^{min}$ ... lowest value recorded in a wealth trajectory
$\alpha$ ... minimum wealth threshold
$\beta$ ... desired maximum likelihood of falling below $\alpha$

As shown in Table 5.2, the chances of going bankrupt are significant for any value of $\omega$ greater than 0.25.

### 5.3.2 Simultaneous Kelly criterion

Looking at both tables 5.3 and 5.4, we can clearly see how much safer the Kelly strategy for simultaneous games is. Still, it should be underscored that these numbers were achieved on a smaller number of bootstrap samples compared to the sequential strategy.

| Decorrelation term \Kelly fraction | **0.05** | **0.1** | **0.25** | **0.5** | **0.75** | **1** |
|---|---|---|---|---|---|---|
| **BCE** | 0.0 | 0.0 | 0.9 | 21.8 | 64.7 | 89.7 |
| **MSE** | 0.0 | 0.8 | 44.7 | 96.3 | 100.0 | 100.0 |
| **KL** | 0.0 | 0.0 | 4.3 | 52.4 | 85.3 | 98.6 |
| **JS** | 0.0 | 0.1 | 21.2 | 84.6 | 99.6 | 100.0 |

Table 5.2: Percentages of simulation runs that involved a drop of wealth below a tenth of the initial wealth using the sequential Kelly strategy. Approximation of the probabilities $P(W^{min} \leq 0.1 \cdot W^0)$ (in %).

| Decorrelation term \Kelly fraction | **0.05** | **0.1** | **0.25** | **0.5** | **0.75** | **1** |
|---|---|---|---|---|---|---|
| **BCE** | 100.0 | 100.0 | 96.0 | 89.0 | 73.0 | 45.0 |
| **MSE** | 98.0 | 86.0 | 73.0 | 35.0 | 11.0 | 0.0 |
| **KL** | 88.0 | 87.0 | 84.0 | 66.0 | 43.0 | 12.0 |
| **JS** | 97.0 | 94.0 | 86.0 | 60.0 | 25.0 | 5.0 |

Table 5.3: Percentages of simulation runs that resulted in a positive betting return for different Kelly fractions and decorrelation functions using the Kelly betting strategy for simultaneous fights.

| Decorrelation term \Kelly fraction | **0.05** | **0.1** | **0.25** | **0.5** | **0.75** | **1** |
|---|---|---|---|---|---|---|
| **BCE** | 0.0 | 0.0 | 0.0 | 5.0 | 31.0 | 66.0 |
| **MSE** | 0.0 | 0.0 | 6.0 | 57.0 | 93.0 | 100.0 |
| **KL** | 0.0 | 0.0 | 0.0 | 18.0 | 54.0 | 90.0 |
| **JS** | 0.0 | 0.0 | 2.0 | 35.0 | 76.0 | 98.0 |

Table 5.4: Percentages of simulation runs that involved a drop of wealth below a tenth of the initial wealth using the Kelly strategy for simultaneous fights. Approximation of the probabilities $P(W^{min} \leq 0.1 \cdot W^0)$ (in %).

Interestingly, if we set our risk constraint from 5.1 reasonably such that $\alpha = 0.1 \cdot W^0$ and $\beta = 0.1$, the optimal strategy in terms of betting returns will successfully pass the risk constraint. This is generally an unlikely scenario as bettors often need to make trade-offs between optimality and safety.

## 5.4 Wealth trajectories

The wealth trajectories of the optimal models for sequential and simultaneous Kelly strategy - models with BCE decorrelation term and Kelly fraction of 0.25 and 0.5 respectively - provide further insight into the distribution of betting returns achieved across different simulation runs with the initial wealth set to 1000 units, as shown in Figure 5.4.

For both strategies, the lower quartile, upper quartile and median seem to be steadily growing, which is another sign of promise. Indeed, the minimums drop below the initial wealth or even all the way to zero as some of the strategies remain profitless or hit bankruptcy, but that is just a reflection of the risks already discussed in Section 5.3.

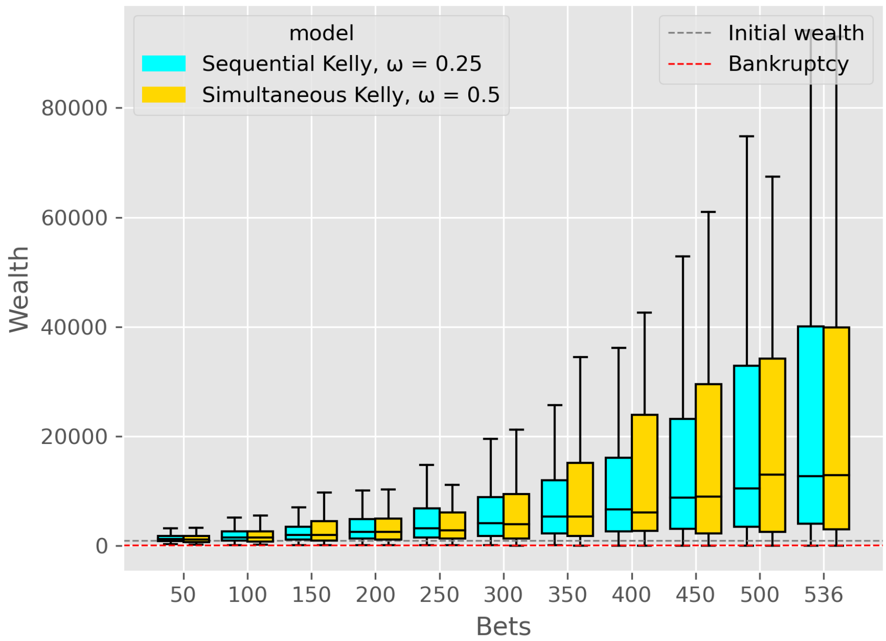


Figure 5.3: Comparison of wealth trajectories of the two most optimal combinations of a model and betting strategy.

Chapter 6

# Conclusion

In Chapter 1 we introduced the domain of combat sports with an emphasis on the sport of mixed martial arts which has been particularly trending in the last couple of years. In previous works, as explained in Chapter 2, various machine learning and analysis methods have been applied in order to describe the nature of MMA or to correctly predict the outcome of MMA fights. However, to my knowledge, there has been no attempt to apply these methods with the goal of generating profit on MMA betting markets. Thus, in this thesis, the goal was set to use machine learning and betting strategies to generate positive betting returns.

First, we obtained publicly available data on fighters, historical fights, and corresponding odds to work with. We processed these data to produce a set of features and selected an artificial neural network as the model of our choice. In the vast majority of applications, predictive accuracy is the main priority and consequently, prediction models are optimized to correctly classify as many presented examples as possible. But as we discussed in section 4.4, it is worth considering a different approach when applying the model on betting markets. Four experimental loss functions for our neural network were introduced to enforce a lower correlation between model's and bookmaker's probability estimates. To finally exploit the betting markets, we introduced two different implementations of the Kelly wagering strategy in section 4.5.

Results on a hold-out dataset show that all suggested models were able to generate significant profit with certain configurations of the betting strategies. The results also support the already widely accepted fact that the full Kelly strategy for wealth allocation is rather unsustainable when making decisions based on probability estimates. This uncertainty was in our case successfully overcome using the fractional Kelly approach.

Best average betting returns were achieved using binary cross-entropy function as a measure of distance in the custom loss functions, with return on investment values climbing up to around 1200% for both betting strategies. For the same model, 93.3% and 89% of simulated runs resulted in a profit which indicates respectable generalization.

## 6.1 Future work

This work provides much potential for future work and could be improved or expanded in numerous ways. Utilization of more data or findings of important performance indicators could result in higher informational efficiency, potentially leading to the ability to incorporate gambling on the in-game betting markets or propositional bets. Also worth exploring would be different betting strategies such as the Modern Portfolio Theory.

Also, the usage of custom loss functions to decrease correlation between model's and bookmaker's estimates is highly experimental. There might be a plethora of other practices that would encourage the desired behaviour of prediction models. A more rigid analysis is also required to enlighten the relationship between the distance functions used in the two terms of the decorrelation loss functions. Potentially, some combinations could be more powerful than other. A similar analysis could then be applied to the relationship of the form of a loss function and the subsequently applied betting strategy.

# Bibliography


[1] California Athletic Commission. *The Unified Rules Of Mixed Martial Arts*. URL: `https://www.dca.ca.gov/csac/forms_pubs/publications/unified_rules_2017.pdf`.

[2] Sally Gainsbury. *Gambling on the Olympics*. 2010.

[3] Elizabeth A Killick and Mark D Griffiths. "In-play sports betting: A scoping study". In: *International Journal of Mental Health and Addiction* 17.6 (2019), pp. 1456–1495.

[4] Philip Conneller. *UFC Hopes to Revolutionize Fight Betting with New Global In-Play Product*. Oct. 2019. URL: `https://www.casino.org/news/ufc-hopes-to-revolutionize-fight-betting-with-new-global-in-play-product/`.

[5] IMG ARENA. *Rush Street Interactive partners with IMG ARENA to debut UFC Event Centre*. Oct. 2020. URL: `https://www.imgarena.com/news/view/rush-street-interactive-partners-with-img-arena-to-debut-ufc-event-centre`.

[6] Lachlan P James et al. "Identifying the performance characteristics of a winning outcome in elite mixed martial arts competition". In: *Journal of science and medicine in sport* 20.3 (2017), pp. 296–301.

[7] Bianca Miarka, Ciro José Brito, and John Amtmann. "Performance probabilities and outcome effects in technical–tactical factors with bout phase changes of mixed martial arts". In: *International Journal of Performance Analysis in Sport* 17.4 (2017), pp. 510–520.

[8] Bianca Miarka et al. "Comparisons of time-motion analysis of mixed martial arts rounds by weight divisions". In: *International Journal of Performance Analysis in Sport* 15.3 (2015), pp. 1189–1201.

[9] Jeremiah Douglas Johnson. "Predicting outcomes of mixed martial arts fights with novel fight variables". PhD thesis. University of Georgia, 2012.

[10] Karmanya Aggarwal Hitkul, Neha Yadav, and Maheshwar Dwivedy. “A Comparative Study of Machine Learning Algorithms for Prior Prediction of UFC Fights”. In: *Harmony Search and Nature Inspired Optimization Algorithms Advances in Intelligent Systems and Computing* (2018), pp. 67–76.

[11] Michal Sipko and William Knottenbelt. “Machine learning for the prediction of professional tennis matches”. In: *MEng computing-final year project, Imperial College London* (2015).

[12] Ondřej Hubáček, Gustav Šourek, and Filip Železn. “Exploiting sports-betting market using machine learning”. In: *International Journal of Forecasting* 35.2 (2019), pp. 783–796.

[13] Tomislav Horvat and Josip Job. “The use of machine learning in sport outcome prediction: A review”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.5 (2020), e1380.

[14] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

[15] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: `10.25080/Majora-92bf1922-00a`.

[16] *Mikoláš Bartoš / Machine Learning in Combat Sports*. URL: `https://gitlab.fel.cvut.cz/bartomik/ml-in-combat-sports`.

[17] Guang-Bin Huang, Yan-Qiu Chen, and Haroon A Babri. “Classification ability of single hidden layer feedforward neural networks”. In: *IEEE Transactions on Neural Networks* 11.3 (2000), pp. 799–801.

[18] Kurt Hornik. “Approximation capabilities of multilayer feedforward networks”. In: *Neural networks* 4.2 (1991), pp. 251–257.

[19] Prajit Ramachandran, Barret Zoph, and Quoc V Le. “Searching for activation functions”. In: *arXiv preprint arXiv:1710.05941* (2017).

[20] P Jeatrakul and KW Wong. “Comparing the performance of different neural networks for binary classification problems”. In: *2009 Eighth International Symposium on Natural Language Processing*. IEEE. 2009, pp. 111–115.

[21] Léon Bottou. “Stochastic gradient learning in neural networks”. In: *Proceedings of Neuro-Nımes* 91.8 (1991), p. 12.

[22] Lorenzo Rosasco et al. “Are loss functions all the same?” In: *Neural computation* 16.5 (2004), pp. 1063–1076.

[23] Nitish Srivastava et al. “Dropout: a simple way to prevent neural networks from overfitting”. In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[24] Lutz Prechelt. “Early stopping-but when?” In: *Neural Networks: Tricks of the trade*. Springer, 1998, pp. 55–69.

[25] D Randall Wilson and Tony R Martinez. “The need for small learning rates on large problems”. In: *IJCNN’01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222)*. Vol. 1. IEEE. 2001, pp. 115–119.

[26] Takuya Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019.

[27] Fabian Wunderlich and Daniel Memmert. “Are betting returns a useful measure of accuracy in (sports) forecasting?” In: *International Journal of Forecasting* 36.2 (2020), pp. 713–722.

[28] Ondřej Hubáček and Gustav Šır. “Beating the market with a bad predictive model”. In: *arXiv preprint arXiv:2010.12508* (2020).

[29] Solomon Kullback and Richard A Leibler. “On information and sufficiency”. In: *The annals of mathematical statistics* 22.1 (1951), pp. 79–86.

[30] John L Kelly Jr. “A new interpretation of information rate”. In: *The Kelly capital growth investment criterion: theory and practice*. World Scientific, 2011, pp. 25–34.

[31] Steven Diamond and Stephen Boyd. “CVXPY: A Python-embedded modeling language for convex optimization”. In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.

[32] Akshay Agrawal et al. “A rewriting system for convex optimization problems”. In: *Journal of Control and Decision* 5.1 (2018), pp. 42–60.

[33] Bradley Efron. “Bootstrap methods: another look at the jackknife”. In: *Breakthroughs in statistics*. Springer, 1992, pp. 569–593.

# BACHELOR'S THESIS ASSIGNMENT

## I. Personal and study details

Student's name: **Bartoš Mikoláš** Personal ID number: **487230**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Computer Science**

Study program: **Open Informatics**

Specialisation: **Software**

## II. Bachelor's thesis details

Bachelor's thesis title in English:

**Machine learning in combat sports**

Bachelor's thesis title in Czech:

**Strojové učení v bojových sportech**

Guidelines:

Mixed martial arts is currently one of the fastest growing sports world wide. It has also been largely left out by the existing research. This thesis is focused on prediction of bout outcomes in the sport of mixed martial arts. Specifically, the aim of the thesis is to find a prediction model optimizing profitability on betting markets utilizing methods of data analysis, machine learning and portfolio optimization.

1. Review relevant literature on combat sports bout outcome prediction.
2. Prepare and process suitable historical data.
3. Seek to understand the dynamics of the sport and discuss important findings.
4. Propose multiple approaches combining predictive modeling with decision making focused on long term robustness and profitability.
5. Evaluate and compare proposed approaches using several suitable metrics.

Bibliography / sources:


1. James, Lachlan P., et al. "Identifying the performance characteristics of a winning outcome in elite mixed martial arts competition." Journal of science and medicine in sport 20.3 (2017): 296-301.
2. Johnson, Jeremiah Douglas. Predicting outcomes of mixed martial arts fights with novel fight variables. Diss. University of Georgia, 2012.
3. Hitkul, Karmanya Aggarwal, Neha Yadav, and Maheshwar Dwivedy. "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 (2018): 67-76.
4. Miarka, Bianca, Ciro José Brito, and John Amtmann. "Performance probabilities and outcome effects in technical–tactical factors with bout phase changes of mixed martial arts." International Journal of Performance Analysis in Sport 17.4 (2017): 510-520.
5. Taleb, Nassim Nicholas. "On the statistical differences between binary forecasts and real-world payoffs." International Journal of Forecasting (2020).


Name and workplace of bachelor's thesis supervisor:

**Ing. Matej Uhrín, Department of Computer Science, FEE**

Name and workplace of second bachelor's thesis supervisor or consultant:

Date of bachelor's thesis assignment: **04.03.2021** Deadline for bachelor thesis submission: **13.08.2021**

Assignment valid until: **19.02.2023**

| Ing. Matej Uhrín<br>Supervisor's signature | Head of department's signature | prof. Mgr. Petr Páta, Ph.D.<br>Dean's signature |
|---|---|---|

## III. Assignment receipt

The student acknowledges that the bachelor’s thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the bachelor’s thesis, the author must state the names of consultants and include a list of references.

.

Date of assignment receipt

Student’s signature