

Msc Data Science Project Proposal

Edward Gill *

April 2019

Abstract

To be decided closer to the completion date

Date: 16 April 2019 Supervisor: George Magoulas M.Sc. Data Science project proposal, Department of Computer Science and Information Systems, Birkbeck College, University of London, 2019 This proposal is substantially the result of my own work, expressed in my own words, except where explicitly indicated in the text. I give my permission for it to be submitted to the JISC Plagiarism Detection Service. The proposal may be freely copied and distributed provided the source is explicitly acknowledged.

*adding thanks

Contents

1	Introduction	4
2	Background	5
3	Aims and Objectives	6
4	Pitfalls of Machine Learning in Trading	8
4.1	Heteroskedasticity in Financial Markets	11
5	Building Blocks of Currency Trading Strategies	14
5.1	Drivers of Currency Markets	14
5.1.1	Economic Fundamentals	14
5.1.2	Price momentum	14
5.1.3	Investor Positioning	15
5.1.4	Currency Valuation	15
5.2	Using a Multi Factor Model For Currency Prediction	15
5.3	The Myth of Machine Learning in Finance?	17
6	Machine Learning Model Overview	17
6.1	Artificial Neural Networks	17
6.2	Support Vector Machines	20
6.3	Ensemble Methods: Random Forests	21
7	Proposed Methodology	23
7.1	Performance Evaluation	28
7.2	Assessing Model Performance	30
7.3	Deflated Sharpe Ratio	31
8	Proposed Work plan	32

1 Introduction

One is stuck by the velocity of technological change we have encountered since the turn of the millennium, the world has embarked on a technological revolution which has brought the availability of immense processing power to the masses. Computational power so large that government sponsored computer science programs of the early nighties could only dream of, in fact, the processing power of most smartphones are larger than the supercomputers of yesteryear [?].

Moore's Law, the well known law which states that the number of transistors in a dense integrated circuit doubles about every two years, has held for over fifty years [1].

The rise in computing power has opened a veritable Pandora's box of of highly complex and computationally intensive models which can now be applied to any data set, whether appropriate or not. While machine learning models have lead to rapid and impressive use cases across a host of different fields of research, this paper will focus on one oft researched area, global financial markets and specifically currency markets. The allure of applying machine learning methods to finance is well entrenched, where researchers dream of finding the holy grail of uncovering a repeatable pattern with in the data which leads to guarantees of consistent alpha generation (returns unexplained by conventional market beta factors [2]).

A cursory search of Google, another revolution of big data processing and clustering algorithms, uncovers a plethora of academic papers which profess of having found patterns in the data that could only have been uncovered by complex models with non-linear capabilities. While big data and machine learning models can and do give fantastic results in the area of biological science or as recommender systems for movies or advertisements, applying these same methods to financial markets is a different and more challenging task due to the underlying unpredictability of finance.

This proposal will firstly outline some background research already undertaken alongside some of the more common pitfalls that can impact model performance and

model generalization. The rest of this proposal dives into the data to be used, the model building framework, the main pitfalls which make applying complex algorithms in finance susceptible to out of sample failure [3] and also a brief analysis of machine learning models to be used. [?].

2 Background

The task of uncovering repeatable patterns in financial data remains incredibly precarious, financial markets are highly non stationary, exhibit leptokurtic distributions (fat tails) and can shift regimes at highly unpredictable speeds. 1 Shows the rolling volatility of EURUSD (with daily snapshots created at 8am and 4pm) which shows that the series volatility not constant and even the volatility across different daily time snapshots can deviate significantly.

Much has been written on the complexity of financial markets and how researchers

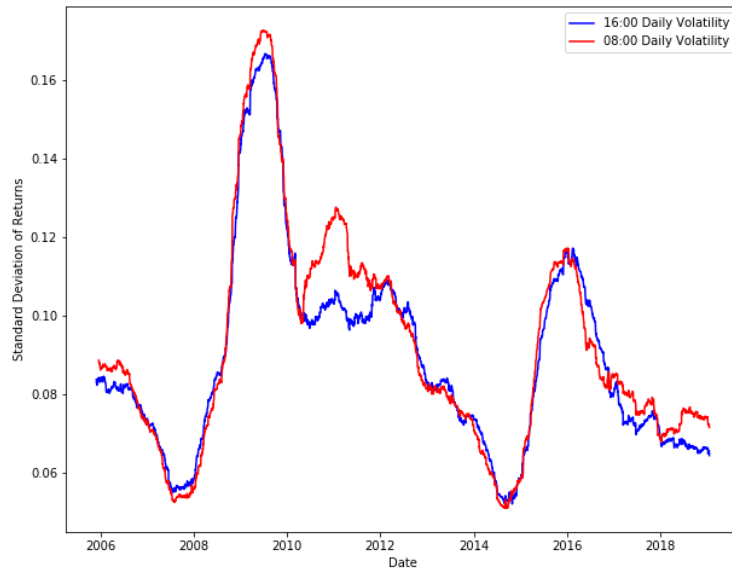


Figure 1: EURUSD Rolling Annualised Volatility Since 2002

must overcome such dynamic systems [4].

While the explosion of increased computational power has helped apply machine learning in very useful tasks such as speech recognition and image processing, the applicability of these methods in finance are, on the face of it at least, questionable [5].

Most papers exploring the financial field make use of swathes of price based technical indicator data and this can lead to issues around overfitting and dimensionality, as we can perhaps find spurious correlations in the data [3]. In the era of big data, finance is one area which has some of the smallest data sets available, as one important factor is in short supply, time. Financial markets have relatively little history compared to other fields of study. Most accurate data capture goes back to the early nineties and unless the domain is high frequency tick data, then there is generally very little data to test with when compared to some non financial data sets, for example, movie recommendation algorithms have many millions of users each with their own individual movie met data attributes [6].

3 Aims and Objectives

The main aim of the project is to understand how machine learning methods can be applied to assist in currency forecasting and the classification of macroeconomic regimes in which a particular model (or combination of models) may perform best.

Thus the project will try to outline best practises when applying machine learning techniques to currency trading, this may take the form of an outright prediction model or a set of models which also forecast shifts in market regimes. This will allow us to combine information from various sources to provide the highest probability of out of sample trading success.

The main objectives in order to achieve the stated aim of the project are shown below.

1. *Data Retrieval, Cleaning and Standardisation*

- One of the most important aspects of a trading model is to ensure that the underlying data is of high quality and doesn't contain erroneous values. The proposed model will use hourly spot exchange rates for G10 currencies ¹, various economic data from the Federal Reserve Economic Database, the International Monetary Fund and the Organisation for Economic Cooperation and Development (OECD) will be used to try to uncover exploitable patterns in the data.
- This data will also need to be clean and mapped to correct date scales so that the model only uses data available at the time of trade signal generation. The raw data itself will need to be standardised to allow the model to understand the true relationships between the features and the output. Standardising the data by its historical standard deviation is one technique which could be used.

2. *Feature Extraction*

- Feature extraction for financial time series prediction is a very challenging task and given the inherent noise in the data, features which have an intuitive justification for predicting future prices is important as given the limited number of data points, there is a higher risk of over fitting if the training data results alone drive the feature selection process.

3. *Model Training and Architecture Selection*

- The project itself will test the efficacy of a range of algorithms in trying to predict currency movements, this project will also make use of non price data in order to assist the model in understanding potential regime changes in the trading environment which can impact the models learned patterns

4. *Model Validation and Strategy Back testing*

¹Hourly FX Spot Data Provided by Citi Foreign Exchange

- Once a model architecture has been selected, it is important to understand not only classification rates but also how the model would perform in a live market setting, which can be approximated by creating a back testing framework which includes the impact of transaction costs for each signal the model generates.

5. *Testing on Fully Withheld Data*

- While the test data allows us to understand how the model performs, there will also be an opportunity to analyse how the selected model will work on new unseen data and can provide a sense of how the model will generalise in future trading.

4 Pitfalls of Machine Learning in Trading

The problem to overcome in applying ML to the financial trading is the level of signal to noise in the data. To try to differentiate the problems ML practitioners face in the financial domain vs others, lets look at an example. Take the case of image processing, if a neural network is fed a matrix of pixels which form the image of cats and dogs, and the neural network trains the model based on the historical patterns it observes. The well trained neural network will no doubt be able to identify cats and dogs with high classification rates, however, if we provide an image of a rhino, the model likely classifies the rhino as a large dog due to the historical patterns that have been learned.

Similarly with the Alpha Go project [7] at the company Deep Mind, the developers created an excellent model which would calculate all possible combinations of moves to find the most optimal route to victory. This model convincingly beat the worlds best human Go players.

That extremely well built model is beholden to the past patterns it has learned

over millions of iterations playing within the well defined rules of the game of Go. In finance, which one could assume is somewhat akin to playing the ancient game of Go (or more closely related would possibly be poker) where you can learn all the historical patterns and probabilities surrounding the asset you are trading. This model will likely work very well even in the testing period. However, in finance the rules can change rapidly, causing structural breaks to occur in past relationships. Alpha Go does not need to worry about the impact of a 280 character tweet from Donald Trump entirely changing the rules of the game.

For example, on the day of the 2016 US presidential election, equity markets in the US initially crashed, the market was horrified to see a president who openly criticized western allies, spoke of protectionism and of building walls against major trading partners.

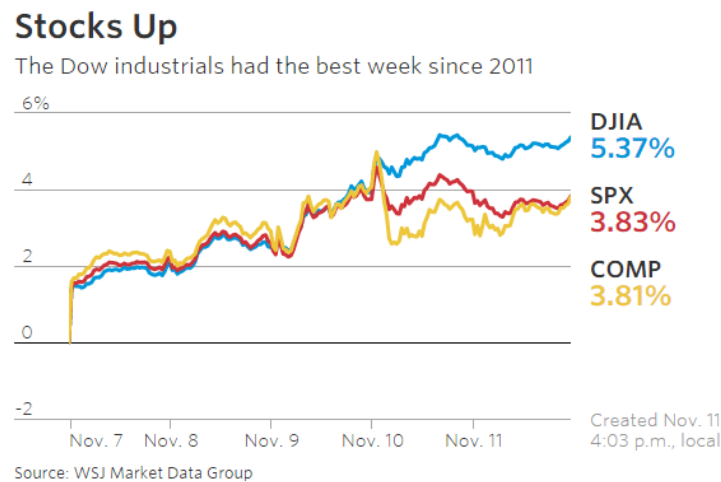


Figure 2: US Equity Market Performance During US Election

However, post his initial speech (which was much less controversial and business friendly) markets fully reversed their losses and ended up 5% higher ². Thus even models that had predicted a Trump win (and subsequent market crash) and positioned as such based on the historical correlations in the lead up to election day would have incurred severe losses. Currency markets were also impacted , with EURUSD rising

nearly 3% as the election result was announced, before subsequently falling by a larger amount after Trump's conciliatory speech 3.

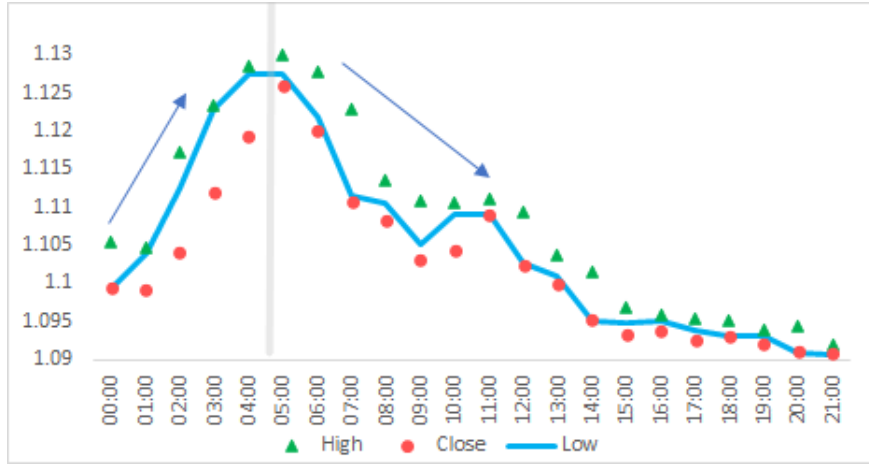


Figure 3: EURUSD Price Action US Election Day 2016 Before and After Trump's Initial Victory Speech

Correlations and patterns in markets can be extremely fickle, financial markets are incredibly complex and dynamic [8]. Why is this so one may ask, well part of the issue lies in the efficiency of informational flow across markets currently [can you give a quick look at how quickly trends have built up or the speed of trends standardized by the underlying move, ie. can you see if we build up the same % move but in much shorter times?], barriers to entry are much smaller and there has likely been a compression in the informational edge that smart money investors (such as hedge funds) previously have had. The market price itself is a reflection of all these moving parts and thus even if a robust pattern has been found, it is likely to be arbitrated as more market actors step in to trade away any mis-pricing [9].

One issue that is particularly pertinent in the financial domain is the issue of data dimensionality and the underlying size of each individual data set, even if there was fifty years of daily financial data, this is less than 20,000 data observations (assuming a 252 trading day calendar), thus while in other fields of research there may be millions of data points for categorization and training, this is not the case for finance. This

means that one must think carefully before adding features to the model. There are thousands of potential predictors of a financial asset (in fact [10] used 180 predictors when trying to forecast financial market crisis events), and this means that due to the number of possible variations, the probability of finding a strategy by luck remains high, [11] show a trading strategy which would have passed even the most stringent evaluation methods in both training and testing which selected its stocks based on alphabetical ordering.

4.1 Heteroskedasticity in Financial Markets

Complex systems that continuously evolve make applying ML techniques in a way that results in a robust model very difficult [11]. Financial markets have been shown to exhibit heteroskedasticity in their returns [?] and as such creating a robust systematic trading strategy relies on more than just a simple data input/signal output type of architecture.

In trading systems, classification of regimes and subsequent buy or sell signals must also be compared vs the expected value of the that trade. A model with an 80% classification success rate will under perform if it mis-classifies the the 20% of samples which result in large moves in the asset. The corollary of that is by only targeting classification of outliers you reduce your sample set and thereby reduce the statistical validity of the results. There is rarely a free lunch in financial markets and as such there will be no free lunch in using machine learning methods to predict financial times series. For example 4, shows that if an investor were to miss out on the best 25 days of the Standard and Poors 500 companies equity index (S&P 500) over the last fifty years, the reduction in returns would have been very severe [?] difference in returns is stark. Given that 25 trading days represents only 0.005% of the total trading days , returns are significantly impacted. This shows the importance of being correctly positioned for outlier events and also the fact that one model can be severely impacted by one off or black swan type events [?].

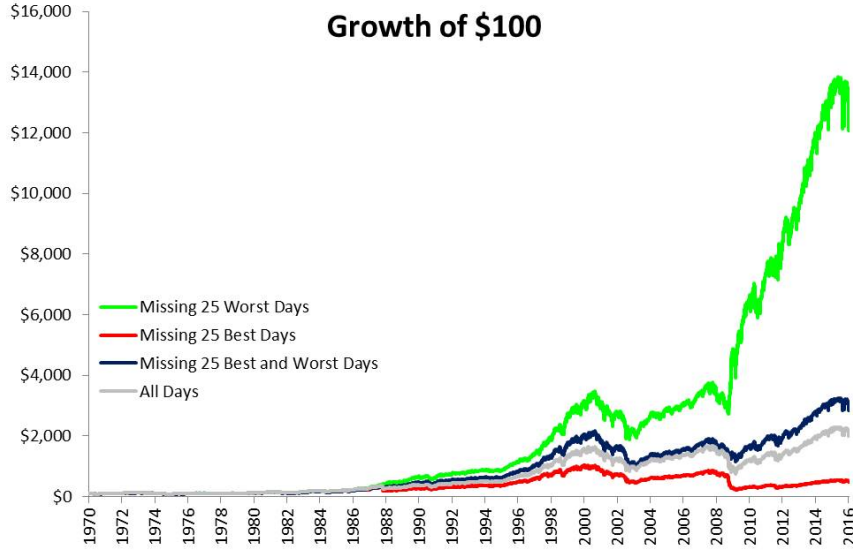


Figure 4: Growth of \$100 Invested Since 1975, source : Market Watch

Previous research has centered on price based data possibly as it is easily available and also because it is one dataset that is likely to have higher frequency data available [12] [13] [14]. This is perhaps due to the innate link between financial markets and the behavior of sometimes non rational human investors. In fact, even when analysing the daily changes in EURUSD , we can see that distribution of future returns can shift quickly, and the regime a model has learned can end up shifting so that the model under performs. For example, 5 splits daily distribution of EURUSD percentage returns by two different regimes, one in 2008 vs a second regime in 2014, each market regime results in a shift in the distribution of the currency and thus when training a model, one must be cognizant of the impact the wider market regime can have on a currency.

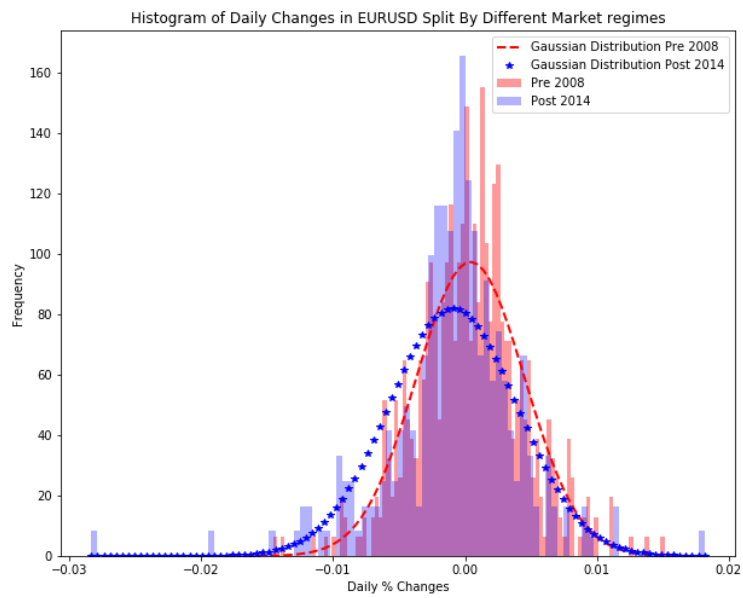


Figure 5: Distribution of Currency Returns Shifts According to The Market Regime

5 Building Blocks of Currency Trading Strategies

One area which seems under researched (from the preliminary research performed), is to use lower frequency economic data to assist in a model's prediction accuracy. While price based metrics offer the easiest route to higher frequency data, they also lack the ability to capture the fundamental economic factors that also play a role in currency direction [?].

5.1 Drivers of Currency Markets

Firstly, before we analyze feature selection methods, it is also important to try to understand what consists of a good trading signal, whether discretionary or systematic in its generation. Below we showcase factors that can impact currency specific direction.

5.1.1 Economic Fundamentals

One of the most important factors that sets the broad direction of a currency is how the country is performing in economic terms relative to peers. [15] show that measuring relative economic momentum between different countries has the ability to forecast currency direction over medium term horizons. That paper showed statistically significant results that economic fundamentals have an important role in relative capital flows and subsequent currency direction.

5.1.2 Price momentum

Price momentum here refers to the historical trend in the currency, research does suggest there exists exploitable momentum within the currency space [16]. The reasoning for the existence of momentum in currencies is likely due to decision delays amongst different investor segments who build up positions in currencies over time. Auto-correlation in currencies may also exist as information is priced slowly over time

(such as economic releases) and thus as better economic data is published, this adds further confirmation to a trading signal and thus conviction in the trade increases.

5.1.3 Investor Positioning

Currency positioning relates to the magnitude of the investments communities holdings of a currency. Positioning and investor flow information can have informational content that impacts the models performance [17]. For instance, if the finalised model triggered a buy recommendation on a currency, but if the wider market had already built up a sizable long position in that currency, then the likelihood of further buyers coming in to drive the price higher is diminished. While positioning is very hard to empirically measure, one can use certain price metrics to gauge proxy positioning and thus avoid trades where the chance of currency reversion is high.

5.1.4 Currency Valuation

Uncovering the true fundamental fair value of a currency has been the focus of international academic research for many years citeRogoff1996, and it remains an elusive and rather ambiguous task as currency drivers can undergo structural shifts. Purchasing Power Parity currency levels , which rely on the Bassala Samuelson effect and the law of one price [18] is one method of trying to anchor the value of a currency to one fundamental price level. While currencies can deviate from fair value, there is some evidence that developed market² currencies tend to mean revert in the long term [?].

5.2 Using a Multi Factor Model For Currency Prediction

While the above drivers of currency markets are not exhaustive, it can provide a framework for constructing a systematic trading strategy. Currency drivers can shift over

²Developed Market Currencies consist of USD, EUR, JPY, CHF, GBP, AUD, NZD, NOK , SEK, CAD

the short and medium term and while this can all be somewhat captured by price action, it is also prudent to try to understand the overall economic picture as that has been shown to provide some indication of future currency trends [19].

The view behind this idea is that as economies of various currencies outperform peers, then those economies attract foreign investor capital and thus undergo currency appreciation. Thus it is important to focus on not only price action but also the underlying fundamentals.

Therein lies the main problem when combining machine learning and finance, that other areas such as the biological or physical sciences do not encounter, is that at the core of each price, is a human or human created trading algorithm which have collectively decided on a price level on which to transact.

Every price of a trade in financial markets consist of the dreams, fears and greed of human emotion. Not only this, but market investors are influenced by past patterns and incorporate new information (including newly published research) into future decisions, thus altering the environment in which the historical pattern has been learned. There is evidence that anomaly (excess) profits published in research papers tend to decay substantially in the period after publication, which suggests any new findings are arbitrated away quickly as the new information is incorporated into the market [9].

While early academic research espoused the efficient market hypothesis, there are certain times where the market is not fully rational [20]. Thus opportunities do exist, but finding repeatable opportunities requires more than just analysing price patterns in data. Given the randomness of markets, even poor strategies can perform well for long periods of times due to sheer luck. This creates new challenges for practitioners of machine learning in finance, in fact, new research has also strated to focus on protocols and best practices around data mining in finance [11].

5.3 The Myth of Machine Learning in Finance?

Given the inherent challenges when creating systematic strategies in markets, one may wonder if creating a trading strategy based on machine learning methods is a fruitless endeavor.

Due to the unique issues related to applying machine learning in the financial domain, this paper proposes to test various models with the goal of finding a robust methodology for extracting features which also have the ability to generalise into the future. Once we have the results and performance of each model, we can then compare and contrast how each model performs both in the training and validation phases.

To ensure we can test the selected model fully on new data, the last one year of test data will be truncated from the data until all validation and testing phases have been completed, this allows us to see how the finalised model would perform on fully unseen data even after initial testing and calibration phases across all the models selected.

6 Machine Learning Model Overview

Below we provide a brief overview of various models we will test in our analysis

6.1 Artificial Neural Networks

Artificial Neural Networks (ANN) were conceptually first written about in the early 1940s [21] in the early 1940s when neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper on how neurons could work in theory, Donal Webb then expanded on this in his paper "The Organization of Behavior" while the first computerised neural networks were developed by IBM in models called "ADALINE" and "MADALINE. At their core, ANNs are merely a series of large matrix multiplications where weights and biases, which represent the strength of connections between nodes, are optimized to provide the best approximation to the desired

output. The popularity of ANNs have grown as enhanced computational power led to the lowering of barriers to entry in the field. One of the more popular networks used in practice is the mutlilayer perceptron, initially introduced by Frank Rosenblatt in 1958 [22]. A perceptron is the simplest form of a layered network and consists of a single neuron with an adjustable weight. The simplest form of the perceptron can divide the hyperplane into a 2-dimensional space which is then used to classify the given inputs. The hyper plane is defined by the linearly separable function below.

$$\sum_{i=1}^n x_i w_i - \theta = 0 \quad (1)$$

Multi layered perceptrons can be defined as

$$NeuralNet_j^l = \sum_{i=1}^{N_l} w_{ij}^{l-1} y_i^{l-1} \quad (2)$$

where $NeuralNet_j^l$ is the l th layer of the j th neuron which gives us the weighted sum of its inputs. The weights of the connections to the next layer $l = 1$ are used to calculate the weighted sum of inputs for the next layer in the network.

$$Y_j^l = ActivationFunction(NeuralNet_j^l) \quad (3)$$

This weighted sum of inputs is then passed through to an activation function which maps the output of the hidden layer to a space defined by the activation function, in this step we can introduce non linearity into the model itself by using a non linear activation function such as the sigmoid function described below.

$$y = \frac{1}{1 + e^{-x}} \quad (4)$$

This maps any series of values onto a plain of $[-1,1]$ and the output depends on the values produced by the node which feeds the activation function.

While there are a vast array of model training techniques , one method to train an ANN is to use backpropagation, where updates to the networks weights are back-propagated through each layer. The model has an error function which is used to

grade the models performance, which could be the the squared difference between the models output and the desired output (mean squared error).

The universal approximation theorem [?] of the MLP model in neural networks states that ANNs can approximate any function and thus in theory, should be able to identify any patterns within the data that another method would also identify.

The temptation to introduce multiple layers with thousands of connections in order to find a solution is ever present, however (and especially y in the realm of finance) this increases the risk of over-fitting the model or worse selecting a model based on sheer luck, which fails when tested on live unseen data.

ANNs have been used extensively in academic financial journals, [23] made use of feedforward multilayer perceptrons to prediction the direction of the EURUSD currency rate, while [14] applied probabilistic neural networks and Least Square Support Vector Machine models to assist in the prediction of Chinese equities. Literature on using ANNs to forecast currencies has a long history as even at the turn of the millenium where three ANN based forecasting models using standard backpropagation, scaled conjugate gradient and backpropagation with Baysian regularization where implemented in order predict movements in various Australian dollar currency crosses [?]. [24] tested a variety of ANN models and Support Vector Machines (SVM) to try to predict the price of equities traded on the Istanbul Stock Exchange , they uncovered model classification accuracy consistently around 70% on test data. The all of the research cited make use of various technical indicators which are derived directly from the price series, for example, [24] make use of ten price based technical indicators to uncover patterns within equities traded on the Turkish stock exchange. While on average most prior research suggested above 70% classification rates were in general attainable, there has also been some research to suggest that ANNs often show inconsistent performance on noisy data [25], [26] [27] , which highlights the challenges in applying complex models as the risk of overfitting the data remains an ever present threat.

6.2 Support Vector Machines

One model which prior research suggests is useful for uncovering repeatable patterns in financial data are Support Vector Machines (SVM) and Support Vector Regression (SVR). SVMs were initially introduced by Cortes and Vapniks [?] which led to the development of non linear models which could achieve binary classifications of data while SVRs output real-valued predictions. In similar fashion to an ANN, the SVM tries to classify an n-dimensional feature vector according to where that data lies on an n-dimensional space.

The model makes use of the kernel trick [?] to transform the original data into a higher dimensional space such that we can then separate the data points. The function updates the parameters of the function of the kernel function such that it can more accurately classify the data points. The error correction model in the case of a SVM is a maximisation problem where the model tries to maximise the distance between all the separating hyper planes it has constructed. The maximisation problem can take the form of the following

$$Maximise \sum_{i=1}^N w_i - \sum_{i=1}^N \sum_{j=1}^N w_i w_j * y_i y_j * K(x_i, x_j) \quad (5)$$

where w_i denotes the coefficients of the i^{th} observation ($i = 1, 2, \dots, N$), y_i is the output of observation i and x_i, x_j are vectors which undergo dimensional transformation using the selected kernel. [28] made use of SVMs to find patterns in Turkish equities achieving classification success rates of around 60% when tested on hold out data. Both polynomial and radial basis SVMs were trained on a host of price based technical data with the radial SVM outperforming the polynomial SVM. The polynomial SVM also had a much higher variability in classification success rates and thus the radial basis may be best suited to predicting financial time series. The results of the radial SVM were significant with a test statistic of 3 and p-value of 1%.

Similarly, [29] tested a selection of four different machine learning methods (ANN,

SVM, Random Forest and Naive Bayes) using ten technical indicators. They also transform the predictor space into trend deterministic states, which map the original continuous values into discrete values based on the trend of the continuous valued indicator (downward trend = -1, upward trend = 1). They find the models performance is improved when trend deterministic data is used instead of continuous data. They postulate that the models perform better on the trend deterministic data as this transformation inputs the underlying trend of the indicator which then allows the model to learn the relationship between the underlying predictor trends and the actual output trend. SVM classification rates using this approach (for both radial and polynomial kernels) approached nearly 90% on average. [12] also tested the perform of an SVM model (radial kernel) to predict Nikkei futures prices³ and achieved classification rates of 73% on test data. The authors also found that combining model forecasts increased out-of-sample test results. This is a architecture that will be tested in the project as well. Overall, prior research has found significant results for using SVMs as method to predict future time series such as equities ([30], [13], [25]) and currency movements [31].

6.3 Ensemble Methods: Random Forests

Ensemble methods are a technique in machine learning which uses the outputs of multiple models to form a consensus around the prediction, while there are many different techniques one can use, this paper proposes to explore whether ensemble methods can be used to reinforce the accuracy of the model on unseen data. One approach is to use tree based methods [?] , which involve segmenting the predictor space into a number of regions. Figure 6 looks at the prior 12 hour (x-axis) and 24 hour (y-axis) return of EURUSD with the subsequent 24 hour ahead move in the currency colour coded from red (negative return) to green (positive return). Even in such a simple split of the predictor space, we see some evidence that more subsequent

³Nikkei 225 is one of the main Japanese Stock Market Indices

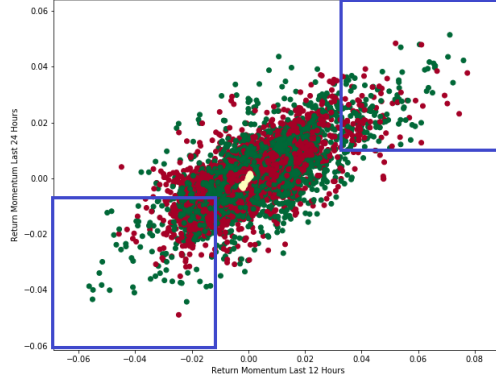


Figure 6: Tree Based Methods Segment the Predictor Space into Numerous Regions

positive days occur in the bottom left of the chart (when prior 12 and 24 hour moves in the currency have been negative). Tree based methods use this type of inference to make predictions on future data points. The goal of the decision tree is to find multiple regions which then minimise the squared error of predictions

$$\sum_{j=1}^J \sum_{i=1}^{R_j} (y_i - y_{R_j}^*)^2 \quad (6)$$

where $y_{R_j}^*$ is the mean response for the training observation within the j^{th} region. The methods tested in the project will revolve around Random Forests (introduced by Leo Breiman [?]) which averages the prediction across hundreds of single decision trees. Research has shown that Random forests can generalise better than single decision trees as they maintain the high variance while reducing the bias of the model [?]. Random forests use bootstrapping to take multiple samples from the training data and also randomly samples the data and a set of feature vectors to create a diverse array of decision trees to assist with the generalisation of the model. This method helps to create decision trees which are uncorrelated to each other and thus have lower probability of overfitting the data. [?] showed that the it was possible to achieve classification rates of over 80% for various stock price predictions when using the random forest algorithm and input features which consisted of technical based indicators. The used a random select of three feature vectors (out of a total of ten) to create each

single decision tree in the random forest model and tested the results from a 20 trees to 200 trees using a majority vote for the final classification.

[26] performed extensive tests to compare the performance of support vector machines to random forests and found that on average support vector machines outperformed random forests when predicting the direction of the Indian stock exchange. Three predictors were considered for each node while the number of trees for each model ranged from 200 to 1500, however there was no significant effect on the reduction of error rate for the out-of-sample data based on the number of trees. The RF model achieved classification rates of over 67% on out of sample data.

[?] found that RF methods outperformed SVMs and logistic regression techniques when predicting stock market crisis events, they also note that RF models are relatively robust to overfitting due to each forest only being exposed to a subset of the available feature vectors. The allure of using an ensemble method when using decision trees (DT) is that those models on an individual basis have low bias but high variance. However combining these models in an ensemble fashion can help to reduce the probability of overfitting, while increasing the generalisation of the model on unseen new data.

7 Proposed Methodology

The following section will outline the methodology to achieve the main projects aim and objectives that were outlined in section 2. One area which looks to be under researched (in the view of the author) is to combine models to handle different aspects of the trade signal generation process. This has the potential to be perform well, as one model alone is unlikely to generalise going forward on its own, especially when applied to the the domain of trading, where structural breaks can occur.

Neural Networks, by the nature of their construction need a large training set to understand the patterns within the data, thus this type of model could assist in cap-

turing the non-linearities in financial data.

However, as we have mentioned previously, patterns in financial data remain exceedingly complex and are subject to change very quickly. These changes can be a product of the current market regime. For example, regimes of strong investor risk appetite where higher yielding currencies perform well [32] has generally coincided in environments where underlying fundamental data matters little in terms of currency direction. Thus for a model to be successful, it must be capable of capturing changes in signal dynamics across different market regimes. This can be achieved by feeding features which also allow us to capture risk appetite, overall economic performance or mean reverting price patterns.

As pointed out by [33], financial model performance can differ vastly prior to publication than after it ⁴, whether due to over fitting or post publication anomaly decay [9], this means we must be extra vigilante when assessing model performance.

The Python programming language (Python 3) [34] will be used as the main application to achieve the objectives, with the python packages of Numpy [35], Scipy [36], Sci-Kit Learn [37], Matplotlib [38] and Pandas [39] all providing very useful data manipulation functions to help store, clean and process data. For more complex ANN learning, the machine learning module Keras [40] will be used in the training and testing of those more complex trading models.

- *Objective 1: Data Capture, Cleaning and Standardisation*

- The initial step will be to source intraday hourly price data for a selection of currencies⁵ and related economic and investor flow data. Figure 7 shows the performance of EURUSD over the last sixteen years formed of hourly snapshots⁶

⁴Research suggested a 26% fall in performance out-of-sample and 58% decline post publication

⁵Analysis will involve developed market currencies of EUR, USD, JPY, GBP, SEK, NOK, AUD, NZD, CAD, CHF

⁶circa 90,000 observations

- There is evidence of the existence of momentum in currencies, and using higher frequency data can allow us to explore different time frames which have the most consistent momentum [41].
- Economic data can be obtained from a variety of sources, with the Federal Reserve Economic Database (FRED)⁷, IMF and OECD all providing vast treasures of low frequency economic data to create feature vectors. Relative economic growth (as measured by momentum across indicators such as consumer spending, GDP growth, industrial production, business confidence) has been shown to assist in the prediction of currency movements [15].
- In order to manipulate the data, Pandas will be used to store the data in pandas dataframes, while Sci-Kit Learn has some very useful functions such as StandardScaler which can standardise data using various different techniques (z-scoring, min-max etc)
- Standarising the data remains an important initial step that has been used across a wide variety of prior research. Subtracting the mean and dividing by the standard deviation (z-scoring) [42], min-max standardisation ([24], [26]) and other techniques such as mapping functions to scale feature vectors onto a pre-defined scale [28] [14] [29] all help to achieve the overall goal of scaling, which is to ensure that larger or more volatile features do not overwhelm features of smaller magnitude and less volatility.
- Any errors or missing data will need to be rectified and as such the data will be tested for outliers which are impossible (i.e. negative currency values) or do not fall within appropriate bands (.i.e 100% percent jumps in currency values between hourly time frames).
- Cleaning of the data will be broken down into outlier detection (to capture any erroneous inputs) and then also cross checking the data vs known events

⁷Successfully used in [10] to assist with forecasting stock market crises

in time (such as Brexit, election dates) to see if the observed outliers match the expected time frames. This provides some clarity on the time snapshot of the data as well. The time series will be transformed to a GMT date timescale so that there are no mis-matched values and to ensure the signal lags have been appropriately adhered to.

- Sci-Kit Learn also provides functions to handle unbalanced datasets and anomaly detection which will be used to assist in cleaning the raw data [?]

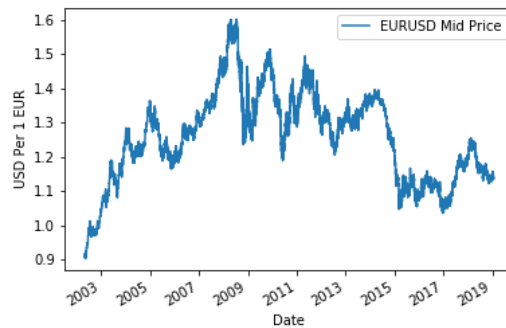


Figure 7: EURUSD Price Action

- *Objective 3: Feature Selection*

- When working with financial time series the feature selection process should be driven by economic intuition as this can lessen the probability of over-fitting [11].
- As patterns and correlations in prices can change quite quickly, it is important not to over-fit the model on the data. A transformation of the data is also needed, with some findings suggesting that performing principal component analysis to create lower dimensional features as a robust method for creating predictive models [43]. The majority of previous research used price-based technical indicators [28] [14] [29] [24] such as moving average crossovers, rate of price change, average true range (ATR) and exponential

moving averages to try to predict future currency direction. While technical indicators can be useful to forecast currencies, it is also important to measure non price based information such as economic momentum and valuation. The proposed feature set is outlined in 1.

Table 1: Model Features

Indicator	Description	Type
Short Term Momentum	5,10, 21, 55 period Moving Average	Technical
Long Term Momentum	100, 155, 200, 255 Period Moving Average	Technical
Mean Reversion	Standard Deviation from 55, 100, 200 Period Moving Average	Technical
Volatility	55, 100, 200 Period Rolling Volatility	Technical
Relative Economic Momentum	1 Year Economic Momentum	Economic
Global Economic Momentum	1 Year Global Economic Momentum	Economic
Purchasing Power Parity	Deviation From Fair Value	Fundamental Valuation

- *Objective 4: Model Training and Fine Tuning*
 - Training the model such that it can understand and capture the relationship between the features and the classification target vector (the target vector will be the future direction of the asset price across various time periods, both regression and classification models will be tested). Prior research has generally used binary classification of the future asset price move when developing models([44], [?]) [10].
 - Model training will also allow us to tune parameters to help to understand which combination of models (ANN, RNN, PNN ,SVM, RF) are best suited in capturing the nuances in the data. This stage of the process will also allow experimentation with various models to see the differences in performance. Various Python packages such as Sci-Kit Learn and Keras will be used in order to create and test the classification success rates of each type of model. [?] made use of Keras (on top of Googles Tensorflow) to implement Recurrent Neural Networks (RNN) while using Sci-Kit Learn for logistic regression, SVM and Random Forest methods.

- This is also a time where testing various combinations of models may work, one model may create a high frequency currency signal while a separate model identifies the regime where this model has historically performed the best. This stacked type of model architecture can assist in model generalisation and boost overall predictive performance ([10], [45]).
- *Objective 5: Model Validation and Strategy Back testing*
 - In order to validate the model, the selected model and its structure will need to be tested in a similar setting to real world trading. This requires the construction of a backtesting engine which can simulate the models performance taking into account transaction costs and the funding costs associated with trading currencies.
 - This will provide a much more accurate analysis of the selected trading strategy performance. While classification success rates will be important in gauging model accuracy, high classification rates may not lead to positive strategy performance if large currency moves are mis-classified. Further analysis of model validation techniques are discussed in detail in the next section.

7.1 Performance Evaluation

Analysing and outlining model performance will be one of the key objectives to be satisfied in the final stages of the project. In light of the findings from ([11], [5]), which highlights in detail the pitfalls inherent in systematic trading, the below outlines the approach to reviewing and evaluating model performance to standards generally set in financial institutions.

- *Model Validation*

Once the model has been trained, then we begin the process of testing how

the model performs on unseen data. This allows the model to score itself on previous iterations to understand the non-linear relationships within the data.

- *Results Across All Model Iterations*

Keeping track of the number of models being tested is very important as it can help define the statistical likelihood of finding a strategy that is a true positive. Even given 20 randomly selected strategies, there is a high probability that one strategy will beat the threshold needed for a statistical significance [11]. Thus, the number of trials ran must be considered so we can deflate the statistical significance of the strategy performance as a function of the number of trials [46].

- *Cross Validation*

Splitting the dataset into various training and test sets will be very important in being able to assess model performance. The most common cross validation techniques are K-fold cross validation and Leave-One-Out cross validation. This project will use a training set, validation set and finally a test dataset which will be held out until the final model has been selected. This will provide a useful gauge of out of sample performance⁸.

- *Problem of Dimensionality*

In finance, the lack of sufficient data (price) vs the number of possible predictors does lead to the issue of dimensionality, where we can add further features but not necessarily add further data points to be tested. Thus the model proposed here will remain cognizant of the number of features being used and will try to reduce the probability of finding false positive trading strategies. The dimensionality of the current data set will involve one target vector (i.e. future price move) and 17-20 predictor columns as outlined in 1.

- *Out-of-Sample Testing on Unseen Data*

⁸In depth study of cross validation techniques is outlined in [?]

As stated by [11], selection of model inputs is inherently biased (especially when done by domain experts) as the experts use heuristics to choose what they find the most economically intuitive as predictors, thus even out of sample test results are not fully out of sample. To override some of these concerns, this paper proposes to omit one years worth of data from any analysis until the model and related architecture has been fully completed and the selected model or average of model outcomes have been finalized. Then the chosen model will be blind tested on the final year of data that has been held back to give the best approximation of how the chosen model might have performed.

- *Simplicity Trumps Complexity*

The goal of this project is to try to create a model capable of generalising into the future, therefore it is important to choose a model which remains the simplest in terms of construction and feature sets. All models should be compared against a simple linear model as a benchmark, to test whether the value added by the increased complexity and reduced interpretability is substantial enough to overcome the drawbacks.

7.2 Assessing Model Performance

In order to measure the prediction accuracy of the model the mean squared error (MSE) is used. It looks at the average distance of the predicted variable from the actual value and has been used extensively in prior research [?] [47] [] .

$$MSE = \frac{1}{N} \sum_{i=1}^n (y_i - F(x_i))^2 \quad (7)$$

where y_i is the actual value and $F(x_i)$ is the value predicted by the function when the i_{th} observation from is passed to the function. Generally speaking common practice would be to choose the model which has the lowest test MSE.

Another method of model valuation is to look at the precision and recall of the model [?], whose values can be calculated from a confusion matrix, which shows the number

of correct and incorrect classifications per class. ?? shows the confusion matrix for a two class classification problem.

Table 2: Confusion Matrix			
Actual (row)	Predicted (col)	Buy	Sell
Buy		True Positive	False Negative
Sell		False Positive	True Negative

The confusion matrix allows us to calculate the precision and recall of the model, the formulas for which are shown below.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (8)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegatives} \quad (9)$$

The precision tells us how well the model has performed in terms of the number of positive signals (buy signals in the case of a trading model) that were actually triggered, while recall of the model depicts how accurate the model has been as a function of the total number of positives in the data set itself. This is an important metric to analyze when predicting a financial time series as certain periods can be heavily biased towards one class or another, for example in the case of a financial asset which has a strong trend will by design have a higher occurrence of one class vs another.

7.3 Deflated Sharpe Ratio

Given the inherent issues of overfitting in Finance ([11] [3]), it is also important to analyse the overall Sharpe Ratio [48] which measures the overall return of a trading model relative to its risk (or volatility of returns). High Sharpe ratios can be obtained by models with high classification rates or through pure overfitting or selection bias, thus it is important to deflate the Sharpe Ratio of every strategy tested to take into

account the number of models tests as well as the skew of each strategies distribution [49]. The Deflated Sharpe Ratio is a test statistic which can outline the probability of having found a false positive. It is defined below as

$$\widehat{DSR} = \left[\frac{\widehat{SR} - \widehat{SR}_0 \sqrt{T-1}}{\sqrt{1 - \gamma_3 \widehat{SR} + \frac{\gamma_4-1}{4} * \widehat{SR}^2}} \right] \quad (10)$$

where $\widehat{SR} = \sqrt{V[\widehat{SR}_n] \left((1-\gamma) Z^{-1} \left[1 - \frac{1}{N} \right] + \gamma Z - 1 \left[1 - \frac{1}{N} \exp(-1) \right] \right)}$ and \widehat{SR} is the estimated Sharpe Ratio, T is the sample length, γ_3 is the skewness of the returns distribution and γ_4 is the kurtosis of the selected strategy, $V[\widehat{SR}_n]$ is the variance across the number of strategy Sharpe Ratios tested, N is the number of independent trials of strategies run (.i.e. the number of models tested) and Z is the cumulative function of the Gaussian distribution⁹.

8 Proposed Work plan

The next page will outline the necessary steps to undertake the work needed. In total, this will require just over 300 hours of work, from the end of April until the deadline date in September 2019.

⁹Further information and theoretical proofs of the Deflated Sharpe is available from [49]

Table 3: Proposed Work Plan

Objective	Task	Description	Start Date	Completion Date
1	Data Retrieval	Source data from various public and institutional databases	April 20 th	April 26 th
2	Data Cleaning	Screen data for outliers	April 27 th	May 5 th
3	Truncation	Split data into appropriate training and testing sets	May 5 th	May 6 th
4	Choose Target	Choose variable to be classified	May 7 th	May 15 th
5	Feature Selection	Select intuitive features for the model	June 14 th	June 30 th
6	Model Preprocessing	Choose the features appropriate for each model	July 1 st	July 3 rd
7	Standardization	Transform and Center data	July 4 th	July 8 th
8	Training	Run Various Training Iterations	July 12 th	July 28 th
9	Model Validation	Cross Validation Across Models	July 28 th	August 10 th
10	Backtesting	Run trading signals through backtester engine	August 12 th	August 20 th
11	Bench-marking	Compare final model to linear classifier	August 21 st	August 25 th
12	Blind Testing	Run Finalized model on Blind data	August 26 th	August 26 th
13	Performance Evaluation	Analyse Performance Statistics	August 27 th	September 3 rd

9 Conclusion

Add conclusion once paper is more finalized.

References

- [1] C. A. MacK, “Fifty years of Moore’s law,” in *IEEE Transactions on Semiconductor Manufacturing*, 2011.
- [2] R. Rebonato, “Asset Management: A Systematic Approach to Factor Investing,” *Quantitative Finance*, 2017.
- [3] M. Lopez de Prado, “The 10 Reasons Most Machine Learning Funds Fail,” *SSRN*, 2018.
- [4] B. Lebaron, “Chaos and Nonlinear Forecastability in Economics and Finance,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 1994.
- [5] D. H. Bailey, J. Borwein, M. Lopez de Prado, and Q. J. Zhu, “Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance,” *SSRN*, 2013.
- [6] I. Portugal, P. Alencar, and D. Cowan, “The use of machine learning algorithms in recommender systems: A systematic review,” 2018.
- [7] D. Silver and D. Hassabis, “AlphaGo: Mastering the ancient game of Go with Machine Learning,” 2016.
- [8] S. Camargo, S. M. Queirós, and C. Anteneodo, “Bridging stylized facts in finance and data non-stationarities,” *European Physical Journal B*, 2013.

- [9] S. M. Bartram, L. Djuranovik, and A. Garratt, “Currency Anomalies,” *SSRN Electronic Journal*, 2019.
- [10] S. P. Chatzis, V. Siakoulis, A. Petropoulos, E. Stavroulakis, and N. Vlachogiannakis, “Forecasting stock market crisis events using deep and statistical machine learning techniques,” *Expert Systems with Applications*, 2018.
- [11] R. D. Arnott, C. R. Harvey, and H. Markowitz, “A Backtesting Protocol in the Era of Machine Learning,” *SSRN Electronic Journal*, 2018.
- [12] W. Huang, Y. Nakamori, and S. Y. Wang, “Forecasting stock market movement direction with support vector machine,” *Computers and Operations Research*, 2005.
- [13] S. Shen, H. Jiang, and T. Zhang, “Stock market forecasting using machine learning algorithms,” 2012.
- [14] S. Wang and W. Shang, “Forecasting direction of China security index 300 movement with least squares support vector machine,” in *Procedia Computer Science*, 2014.
- [15] M. Dahlquist and H. Hasseltoft, “Economic Momentum and Currency Returns.” 2015.
- [16] T. J. Moskowitz, Y. H. Ooi, and L. H. Pedersen, “Time series momentum,” *Journal of Financial Economics*, 2012.
- [17] L. Menkhoff, L. Sarno, M. Schmeling, and A. Schrimpf, “Currency momentum strategies,” *Journal of Financial Economics*, 2012.
- [18] F. Hassan, “The price of development: The Penn–Balassa–Samuelson effect revisited,” *Journal of International Economics*, 2016.

- [19] D. Rime, L. Sarno, and E. Sojli, “Exchange rate forecasting, order flow and macroeconomic information,” *Journal of International Economics*, 2010.
- [20] B. Döme, “Efficiency and Return Predictability.” 2008.
- [21] B. Widrow and M. A. Lehr, “30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation,” *Proceedings of the IEEE*, 1990.
- [22] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychological Review*, 1958.
- [23] P. Czekalski, M. Niezabitowski, and R. Styblinski, “ANN for FOREX forecasting and trading,” in *Proceedings - 2015 20th International Conference on Control Systems and Computer Science, CSCS 2015*, 2015.
- [24] H. Gunduz, Y. Yaslan, and Z. Cataltepe, “Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations,” *Knowledge-Based Systems*, vol. 137, pp. 138–148, 2017.
- [25] K.-j. Kim, “Predicting Stock Price Direction using Support Vector Machines,” *Neurocomputing*, 2003.
- [26] M. Kumar and M. Thenmozhi, “Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest.” 2006.
- [27] K. J. Kim and I. Han, “Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index,” *Expert Systems with Applications*, 2000.
- [28] Y. Kara, M. Acar Boyacioglu, and Ö. K. Baykan, “Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange,” *Expert Systems with Applications*, 2011.

- [29] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques,” *Expert Systems with Applications*, 2015.
- [30] B. Gui, X. Wei, Q. Shen, J. Qi, and L. Guo, “Financial time series forecasting using support vector machine,” in *Proceedings - 2014 10th International Conference on Computational Intelligence and Security, CIS 2014*, 2015.
- [31] J. Kamruzzaman, R. Sarker, and I. Ahmad, “SVM based models for predicting foreign currency exchange rates,” 2004.
- [32] C. Christiansen, A. Rinaldo, and P. Söderlind, “The time-varying systematic risk of carry trade strategies,” *Journal of Financial and Quantitative Analysis*, 2011.
- [33] R. D. Mclean and J. Pontiff, “Does Academic Research Destroy Stock Return Predictability?,” *Journal of Finance*, 2016.
- [34] W. McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 1976.
- [35] S. Van Der Walt, S. C. Colbert, and G. Varoquaux, “The NumPy array: A structure for efficient numerical computation,” *Computing in Science and Engineering*, 2011.
- [36] D. R. Tobergte and S. Curtis, “Learning SciPy for Numerical and Scientific Computing Second Edition,” *Journal of Chemical Information and Modeling*, 2013.
- [37] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. 2017.
- [38] M. A. Wood, *Python and Matplotlib Essentials for Scientists and Engineers*. 2015.
- [39] M. I. Reiff, “Pandas,” *Journal of Developmental and Behavioral Pediatrics*, 2002.
- [40] F. Chollet, “Keras: The Python Deep Learning library,” 2015.

- [41] I. Filippou and M. P. Taylor, “Common Macro Factors and Currency Premia,” *Journal of Financial and Quantitative Analysis*, 2017.
- [42] T. Fischer and C. Krauss, “Deep learning with long short-term memory networks for financial market predictions,” *European Journal of Operational Research*, 2018.
- [43] F. Song, Z. Guo, and D. Mei, “Feature selection using principal component analysis,” in *Proceedings - 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, ICSEM 2010*, 2010.
- [44] G. Abreu, R. Neves, and N. Horta, “Currency exchange prediction using machine learning, genetic algorithms and technical analysis,” 2018.
- [45] J. Stock and M. Watson, “Forecasting with Many Predictors,” *Handbook of economic forecasting*, 2010.
- [46] D. H. Bailey and M. Lopez de Prado, “The Sharpe Ratio Efficient Frontier,” *SSRN*, 2011.
- [47] T. Nguyen, A. Khosravi, S. Nahavandi, and D. Creighton, “Neural network and interval type-2 fuzzy system for stock price forecasting,” in *IEEE International Conference on Fuzzy Systems*, 2013.
- [48] W. F. Sharpe, “The Sharpe Ratio,” *The Journal of Portfolio Management*, 2009.
- [49] D. H. Bailey and M. Lopez de Prado, “The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting and Non-Normality,” *SSRN*, 2014.