



# Métodos de previsão de tendência de mercados de ações

**MIGUEL ALEXANDRE BORGES DA SILVA**

Outubro de 2021

# Stock Market Tendency Prediction Methods

**Miguel Silva**

**A dissertation submitted in partial fulfillment of  
the requirements for the degree of Master of Science,  
Specialisation Area of Computer Systems**

**Supervisor: Dr. Carlos Ferreira**

**Evaluation Committee:**

President:

Members:

Porto, October 15, 2021



# Dedictory

I would like to thank my supervisor for giving me support throughout the making of this thesis, and my family for helping me when times were tough and motivating me to continue.



# Abstract

Investing in stock market is not an easy task, it could be a high-risk investment. Stock price depends on several factors that can be internal factors, such as, the company itself which is down-grade his worth or external factor such as politics, or even the financial environment of the country of that specific market. Investors need careful decision-making skills to maintain profit in the long run. Usually they use techniques that can help them to make a better decision, or at least facilitate that “job”. In the past years, many machine learning models have been investigated for this purpose. Being an area, which is very influenced by uncontrollable factors, it’s not easy to get a perfect solution. This thesis demonstrates a possible methodology to produce a good price prediction of the stock market and make it available to the general public.

**Keywords:** Stock Markets, Machine Learning, Neural Networks



# Resumo

Investir no mercado de ações não é uma tarefa fácil, pode tornar-se um investimento de algo risco. O preço de stock depende de vários fatores que podem ser internos, como por exemplo, uma empresa que reduza o seu valor, ou externos como política, ambiente financeiro do país. Investidores precisam de adquirir experiência em planeamento de modo a manter ou aumentar o lucro. Normalmente, estes usam técnicas que lhes facilita a tomar uma melhor decisão. Desde á uns anos atrás que machine learning tem sido estudado, mas como é uma área que tem fatores incontroláveis, faz com que não seja possível obter uma solução perfeita. Esta Tese demonstra uma possível metodologia que produz uma boa previsão de preço do mercado de ações e o tornar disponível para o publico.



# Contents

|                        |             |
|------------------------|-------------|
| <b>List of Figures</b> | <b>xiii</b> |
|------------------------|-------------|

|                       |           |
|-----------------------|-----------|
| <b>List of Tables</b> | <b>xv</b> |
|-----------------------|-----------|

|                                                       |          |
|-------------------------------------------------------|----------|
| <b>1 Introduction</b>                                 | <b>1</b> |
| 1.1 Context . . . . .                                 | 1        |
| 1.2 Problem . . . . .                                 | 1        |
| 1.3 Objectives . . . . .                              | 1        |
| 1.4 Approach . . . . .                                | 2        |
| 1.5 Document Structure . . . . .                      | 2        |
| <b>2 State of the art</b>                             | <b>3</b> |
| 2.1 Artificial Intelligence . . . . .                 | 3        |
| 2.1.1 Applied AI . . . . .                            | 3        |
| 2.1.2 Generalised AI . . . . .                        | 3        |
| 2.1.3 Machine Learning . . . . .                      | 4        |
| 2.2 Machine Learning Concepts . . . . .               | 4        |
| 2.2.1 Supervised Learning . . . . .                   | 4        |
| 2.2.1.1 Classification Algorithms . . . . .           | 5        |
| 2.2.1.2 Logistic Regression . . . . .                 | 5        |
| 2.2.1.3 Naïve Bayes Classifier . . . . .              | 6        |
| 2.2.1.4 K-Nearest neighbor . . . . .                  | 6        |
| 2.2.1.5 Decision Tree . . . . .                       | 7        |
| 2.2.1.6 ARIMA . . . . .                               | 8        |
| 2.2.1.7 Random Forest . . . . .                       | 9        |
| 2.2.1.8 Support Vector Machines . . . . .             | 10       |
| 2.2.1.9 Regression Algorithms . . . . .               | 10       |
| 2.2.1.10 Simple Linear Regression . . . . .           | 11       |
| 2.2.1.11 Lasso Regression . . . . .                   | 11       |
| 2.2.2 Unsupervised Learning . . . . .                 | 12       |
| 2.2.2.1 Clustering . . . . .                          | 13       |
| 2.2.3 Reinforced Learning . . . . .                   | 14       |
| 2.2.4 CRISP-DM . . . . .                              | 15       |
| 2.2.5 LSTM . . . . .                                  | 16       |
| 2.3 Related work . . . . .                            | 18       |
| 2.3.1 Neural Networks . . . . .                       | 18       |
| 2.3.1.1 Theory of Levenberg- Marquardt . . . . .      | 19       |
| 2.3.1.2 Theory of scaled conjugate gradient . . . . . | 19       |
| 2.3.1.3 Theory of Bayesian regularization . . . . .   | 20       |
| 2.3.1.4 Case Study . . . . .                          | 20       |
| 2.3.1.5 Results . . . . .                             | 20       |

|          |                                       |           |
|----------|---------------------------------------|-----------|
| 2.3.1.6  | Conclusion . . . . .                  | 23        |
| 2.3.1.7  | Case Study 2 . . . . .                | 24        |
| 2.3.1.8  | Results . . . . .                     | 24        |
| 2.3.1.9  | Conclusion . . . . .                  | 25        |
| 2.3.2    | Natural Language Processing . . . . . | 26        |
| 2.3.2.1  | Case Study . . . . .                  | 26        |
| 2.3.2.2  | Results . . . . .                     | 26        |
| 2.3.2.3  | Conclusion . . . . .                  | 27        |
| 2.4      | Technologies . . . . .                | 27        |
| 2.4.1    | Programming Languages . . . . .       | 27        |
| 2.4.1.1  | Python . . . . .                      | 27        |
| 2.4.1.2  | R . . . . .                           | 28        |
| 2.4.1.3  | PHP . . . . .                         | 29        |
| 2.4.1.4  | Conclusion . . . . .                  | 29        |
| 2.4.2    | Development Tools . . . . .           | 30        |
| 2.4.2.1  | Jupyter Notebook App . . . . .        | 30        |
| 2.4.2.2  | Spyder . . . . .                      | 30        |
| 2.4.2.3  | Visual Studio Code . . . . .          | 31        |
| <b>3</b> | <b>Analysis and Design</b>            | <b>33</b> |
| 3.1      | Problem Description . . . . .         | 33        |
| 3.2      | Objectives . . . . .                  | 33        |
| 3.3      | Investigation . . . . .               | 33        |
| 3.4      | Data Sources . . . . .                | 33        |
| 3.4.1    | Datasets . . . . .                    | 33        |
| 3.4.1.1  | Yahoo! Finance . . . . .              | 33        |
| 3.5      | Domain Model . . . . .                | 34        |
| 3.6      | Use Cases . . . . .                   | 35        |
| 3.6.1    | Register . . . . .                    | 36        |
| 3.6.2    | Login . . . . .                       | 37        |
| 3.6.3    | Add Stock Market . . . . .            | 37        |
| 3.6.4    | Remove Stock Market . . . . .         | 37        |
| 3.6.5    | Read Prediction Results . . . . .     | 38        |
| 3.6.6    | Predict Price Tendency . . . . .      | 38        |
| 3.7      | Non-Function Requirements . . . . .   | 39        |
| 3.7.1    | Functionality . . . . .               | 39        |
| 3.7.2    | Reliability . . . . .                 | 39        |
| 3.7.3    | Performance . . . . .                 | 39        |
| 3.7.4    | Supportability . . . . .              | 39        |
| 3.8      | Component Model . . . . .             | 39        |
| <b>4</b> | <b>Value Analysis</b>                 | <b>41</b> |
| 4.1      | Value . . . . .                       | 41        |
| 4.2      | Perceived Value . . . . .             | 41        |
| 4.3      | CANVAS . . . . .                      | 41        |
| 4.4      | QFD . . . . .                         | 41        |
| 4.5      | Functional Requirements . . . . .     | 41        |
| <b>5</b> | <b>Case Study</b>                     | <b>43</b> |

|          |                                    |           |
|----------|------------------------------------|-----------|
| 5.1      | Datasets . . . . .                 | 43        |
| 5.1.1    | Datasets by sector . . . . .       | 43        |
| 5.1.1.1  | Energy . . . . .                   | 43        |
| 5.1.1.2  | Healthcare . . . . .               | 44        |
| 5.1.1.3  | Technology . . . . .               | 45        |
| 5.1.2    | Datasets by Country . . . . .      | 46        |
| 5.1.2.1  | Australia . . . . .                | 46        |
| 5.1.2.2  | Germany . . . . .                  | 47        |
| 5.1.2.3  | United States of America . . . . . | 48        |
| 5.2      | Prediction Models . . . . .        | 49        |
| 5.2.1    | ARIMA . . . . .                    | 49        |
| 5.2.1.1  | Model Parameters . . . . .         | 49        |
| 5.2.2    | Random Forest . . . . .            | 50        |
| 5.2.2.1  | Model Parameters . . . . .         | 50        |
| 5.2.3    | Neural Network . . . . .           | 52        |
| 5.2.3.1  | Model Parameters . . . . .         | 52        |
| 5.3      | Evaluation Methodology . . . . .   | 52        |
| 5.3.1    | Metrics . . . . .                  | 52        |
| 5.3.1.1  | Root-Mean-Square Error . . . . .   | 52        |
| 5.3.1.2  | $R^2$ Score . . . . .              | 53        |
| 5.4      | Results . . . . .                  | 53        |
| 5.4.1    | Energy . . . . .                   | 53        |
| 5.4.1.1  | ARIMA . . . . .                    | 53        |
| 5.4.1.2  | Random Forest . . . . .            | 54        |
| 5.4.1.3  | Neural Network . . . . .           | 56        |
| 5.4.2    | Healthcare . . . . .               | 57        |
| 5.4.2.1  | ARIMA . . . . .                    | 57        |
| 5.4.2.2  | Random Forest . . . . .            | 58        |
| 5.4.2.3  | Neural Network . . . . .           | 59        |
| 5.4.3    | Technology . . . . .               | 61        |
| 5.4.3.1  | ARIMA . . . . .                    | 61        |
| 5.4.3.2  | Random Forest . . . . .            | 62        |
| 5.4.3.3  | Neural Network . . . . .           | 63        |
| 5.4.4    | Australia . . . . .                | 64        |
| 5.4.4.1  | ARIMA . . . . .                    | 64        |
| 5.4.4.2  | Random Forest . . . . .            | 65        |
| 5.4.4.3  | Neural Network . . . . .           | 66        |
| 5.4.5    | Germany . . . . .                  | 67        |
| 5.4.5.1  | ARIMA . . . . .                    | 67        |
| 5.4.5.2  | Random Forest . . . . .            | 68        |
| 5.4.5.3  | Neural Network . . . . .           | 69        |
| 5.4.6    | United States of America . . . . . | 71        |
| 5.4.6.1  | ARIMA . . . . .                    | 71        |
| 5.4.6.2  | Random Forest . . . . .            | 72        |
| 5.4.6.3  | Neural Network . . . . .           | 73        |
| <b>6</b> | <b>Case Study Result Analysis</b>  | <b>75</b> |
| <b>7</b> | <b>Implementation</b>              | <b>77</b> |

|          |                                    |           |
|----------|------------------------------------|-----------|
| 7.1      | Web Application . . . . .          | 77        |
| 7.1.1    | Development . . . . .              | 77        |
| 7.1.2    | Web Pages . . . . .                | 77        |
| 7.1.2.1  | Login Page . . . . .               | 77        |
| 7.1.2.2  | Register Page . . . . .            | 78        |
| 7.1.2.3  | Portfolio Page . . . . .           | 78        |
| 7.1.2.4  | Market List Page . . . . .         | 79        |
| 7.2      | Machine Learning . . . . .         | 80        |
| <b>8</b> | <b>Conclusion</b>                  | <b>81</b> |
|          | <b>Bibliography</b>                | <b>83</b> |
| <b>A</b> | <b>Canvas</b>                      | <b>85</b> |
| <b>B</b> | <b>QFD</b>                         | <b>87</b> |
| <b>C</b> | <b>Full Error Results</b>          | <b>91</b> |
| C.1      | Energy . . . . .                   | 92        |
| C.1.1    | ARIMA . . . . .                    | 92        |
| C.1.2    | Random Forest . . . . .            | 93        |
| C.1.3    | Neural Network . . . . .           | 94        |
| C.2      | Healthcare . . . . .               | 95        |
| C.2.1    | ARIMA . . . . .                    | 95        |
| C.2.2    | Random Forest . . . . .            | 96        |
| C.2.3    | Neural Network . . . . .           | 97        |
| C.3      | Technology . . . . .               | 98        |
| C.3.1    | ARIMA . . . . .                    | 98        |
| C.3.2    | Random Forest . . . . .            | 99        |
| C.3.3    | Neural Network . . . . .           | 100       |
| C.4      | Australia . . . . .                | 101       |
| C.4.1    | ARIMA . . . . .                    | 101       |
| C.4.2    | Random Forest . . . . .            | 101       |
| C.4.3    | Neural Network . . . . .           | 102       |
| C.5      | Germany . . . . .                  | 102       |
| C.5.1    | ARIMA . . . . .                    | 102       |
| C.5.2    | Random Forest . . . . .            | 103       |
| C.5.3    | Neural Network . . . . .           | 103       |
| C.6      | United States of America . . . . . | 104       |
| C.6.1    | ARIMA . . . . .                    | 104       |
| C.6.2    | Random Forest . . . . .            | 104       |
| C.6.3    | Neural Network . . . . .           | 105       |

# List of Figures

|      |                                                                                                      |    |
|------|------------------------------------------------------------------------------------------------------|----|
| 2.1  | Supervised Learning . . . . .                                                                        | 5  |
| 2.2  | K-Nearest Neighbor Method . . . . .                                                                  | 7  |
| 2.3  | Decision Tree Method . . . . .                                                                       | 8  |
| 2.4  | Support Vector Machine Method . . . . .                                                              | 10 |
| 2.5  | Simple Linear Regression Method . . . . .                                                            | 11 |
| 2.6  | Lasso Regression Method . . . . .                                                                    | 12 |
| 2.7  | Unsupervised Learning . . . . .                                                                      | 13 |
| 2.8  | Clustering Method . . . . .                                                                          | 14 |
| 2.9  | Reinforced Learning . . . . .                                                                        | 15 |
| 2.10 | CRISP-DM Method . . . . .                                                                            | 16 |
| 2.11 | LSTM Method . . . . .                                                                                | 18 |
| 2.12 | Neural Network . . . . .                                                                             | 19 |
| 2.13 | Real and predicted NASDAQ index values for four prior days from 28 January to 18 June 2015 . . . . . | 22 |
| 2.14 | Real and predicted NASDAQ index values for nine prior days from 28 January to 18 June 2015 . . . . . | 23 |
| 2.15 | Result of the training for the NKE stocks with different dataset times . . . . .                     | 24 |
| 2.16 | Result of training for NKE and GOOGL stocks with different number of epochs . . . . .                | 25 |
| 2.17 | Python logo . . . . .                                                                                | 28 |
| 2.18 | R logo . . . . .                                                                                     | 28 |
| 2.19 | PHP logo . . . . .                                                                                   | 29 |
| 2.20 | Jupyter logo . . . . .                                                                               | 30 |
| 2.21 | Spyder logo . . . . .                                                                                | 30 |
| 2.22 | Visual Studio Code logo . . . . .                                                                    | 31 |
| 3.1  | Domain Model . . . . .                                                                               | 35 |
| 3.2  | Use Cases . . . . .                                                                                  | 36 |
| 3.3  | Component Model . . . . .                                                                            | 39 |
| 5.1  | Example of a Partial Correlation Plot . . . . .                                                      | 49 |
| 5.2  | Example of a Auto Correlation Plot . . . . .                                                         | 50 |
| 7.1  | Login page . . . . .                                                                                 | 77 |
| 7.2  | Register page . . . . .                                                                              | 78 |
| 7.3  | Portfolio page . . . . .                                                                             | 79 |
| 7.4  | Market list page . . . . .                                                                           | 79 |



# List of Tables

|      |                                                                                |    |
|------|--------------------------------------------------------------------------------|----|
| 2.1  | Trained models and their $R^2$ score . . . . .                                 | 20 |
| 2.2  | Various model structures for four working days and their $R^2$ score . . . . . | 21 |
| 2.3  | Various model structures for nine working days and their $R^2$ score . . . . . | 22 |
| 2.4  | LSTM model summary . . . . .                                                   | 24 |
| 2.5  | The value of loss for GOOGL and NKE for different number of epochs . . . . .   | 26 |
| 2.6  | Accuracy of used algorithms . . . . .                                          | 27 |
| 3.1  | Example of stock market historical data provided by Yahoo! Finance . . . . .   | 34 |
| 5.1  | Various Stock Markets on the Energy Sector . . . . .                           | 44 |
| 5.2  | Various Stock Markets on the Healthcare Sector . . . . .                       | 45 |
| 5.3  | Various Stock Markets on the Technology Sector . . . . .                       | 46 |
| 5.4  | Various Stock Markets in Australia . . . . .                                   | 47 |
| 5.5  | Various Stock Markets in Germany . . . . .                                     | 48 |
| 5.6  | Various Stock Markets in the USA . . . . .                                     | 48 |
| 5.7  | Correlation plot of dataset features . . . . .                                 | 50 |
| 5.8  | Best Errors for each iteration and historical data . . . . .                   | 51 |
| 5.9  | Average baseline error for various historical data . . . . .                   | 51 |
| 5.10 | ARIMA error results for the energy sector . . . . .                            | 53 |
| 5.11 | ARIMA time results for the energy sector . . . . .                             | 54 |
| 5.12 | Random Forest error results for the energy sector . . . . .                    | 54 |
| 5.13 | Random Forest time results for the energy sector . . . . .                     | 55 |
| 5.14 | Neural network error results for the energy sector . . . . .                   | 56 |
| 5.15 | Neural network time results for the energy sector . . . . .                    | 56 |
| 5.16 | ARIMA error results for the healthcare sector . . . . .                        | 57 |
| 5.17 | ARIMA time results for the healthcare sector . . . . .                         | 58 |
| 5.18 | Random Forest error results for the healthcare sector . . . . .                | 58 |
| 5.19 | Random Forest time results for the healthcare sector . . . . .                 | 59 |
| 5.20 | Neural network error results for the healthcare sector . . . . .               | 59 |
| 5.21 | Neural network time results for the healthcare sector . . . . .                | 60 |
| 5.22 | ARIMA error results for the technology sector . . . . .                        | 61 |
| 5.23 | ARIMA time results for the technology sector . . . . .                         | 61 |
| 5.24 | Random Forest error results for the technology sector . . . . .                | 62 |
| 5.25 | Random Forest time results for the technology sector . . . . .                 | 62 |
| 5.26 | Neural network error results for the technology sector . . . . .               | 63 |
| 5.27 | Neural network time results for the technology sector . . . . .                | 63 |
| 5.28 | ARIMA error results for Australia . . . . .                                    | 64 |
| 5.29 | ARIMA time results for Australia . . . . .                                     | 65 |
| 5.30 | Random Forest error results for Australia . . . . .                            | 65 |
| 5.31 | Random Forest time results for Australia . . . . .                             | 66 |
| 5.32 | Neural network error results for Australia . . . . .                           | 66 |

|      |                                                                         |    |
|------|-------------------------------------------------------------------------|----|
| 5.33 | Neural network time results for Australia . . . . .                     | 67 |
| 5.34 | ARIMA error results for Germany . . . . .                               | 67 |
| 5.35 | ARIMA time results for Germany . . . . .                                | 68 |
| 5.36 | Random Forest error results for Germany . . . . .                       | 68 |
| 5.37 | Random Forest time results for Germany . . . . .                        | 69 |
| 5.38 | Neural network error results for Germany . . . . .                      | 69 |
| 5.39 | Neural network time results for Germany . . . . .                       | 70 |
| 5.40 | ARIMA error results for the United States of America . . . . .          | 71 |
| 5.41 | ARIMA time results for the United States of America . . . . .           | 71 |
| 5.42 | Random Forest error results for the United States of America . . . . .  | 72 |
| 5.43 | Random Forest time results for the United States of America . . . . .   | 72 |
| 5.44 | Neural network error results for the United States of America . . . . . | 73 |
| 5.45 | Neural network time results for the United States of America . . . . .  | 73 |
| 6.1  | Best error results for each prediction method . . . . .                 | 75 |

# Chapter 1

## Introduction

### 1.1 Context

A stock market is a platform for trading of a company's stocks at an agreed price. It's a sector that demands a certain amount of knowledge on market trends because supply and demand are the determinants of share prices. With the constant growing influence of the stock markets, people started becoming more interested in predicting price tendencies. As technology improves, stock traders are moving towards to use Artificial Intelligence rather than statistical analysis, which helps to take immediate investment decisions. One of the main aims of a decision support system is to be able to predict the stock price tendency, enabling people to invest at the right time. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instruments traded on an exchange.

### 1.2 Problem

The stock market is an attractive area for those who intend to obtain extra income, but due to its volatility and complexity, people are susceptible to a great degree of risk if they don't have the necessary knowledge or enough time to spend learning. At the moment there are some platforms that try to predict prices of stock markets by various means although they're very uncertain and prone to subjective parameters that cannot be measured through statistics.

### 1.3 Objectives

This project aims to help those with or without prior experience on stock markers and all its related subjects. By providing this assistance, the user will be able to make better choices on their selection of shares, and also be able to maintain a constant surveillance of possible opportunities. It seems to be very difficult to replace the professionalism of an experienced trader for predicting the stock price, but because of the availability of a remarkable amount of data and technological advancements we can now be able to formulate an appropriate algorithm for prediction whose results can be used to help investors make profit. These algorithms' accuracy is directly proportional to gains made by using them.

## 1.4 Approach

The evolution of technology brought new possibilities in the prediction systems area. Nowadays are used in almost every activity sector. New models are being made and tested. In stock market area, making predictions it's not easy due the consequences and conditions involved. There are already some solutions like Yahoo Finance and Google Finance which are used as support for investors, helping them to make some decisions about what they should do to the stock of a company. With this project, I intend to make a study of automatic learning methodologies that could be useful to predict prices of the stock market. This study will use historical data of actions from various companies, indexes and related news that are published on available media. There will also be developed an web application that allows the user to obtain information about certain stock markets and predict the price tendency and variation.

## 1.5 Document Structure

This thesis is structured like so:

In the **Introduction** chapter, it's mentioned the problem this thesis tries to discuss and the approach to a possible solution.

In the **State of the art** chapter, we find the state of the art of prediction methods and also explanation of the most notable machine learning terms along with some case studies.

In the **Analysis and Design** chapter, it's displayed the analysis and design for the project to be made.

In the **Value Analysis** chapter, it's portrayed some business value terms, the value proposition canvas, the quality function deployment and the requirements to develop the proposed solution.

In the **Case Study** chapter, it's shown the process done to obtain the results, how the model's training was done as well the parameters needed to optimize it, and the results obtained.

In the **Case Study Result Analysis** chapter, I discuss the results shown in the previous chapter, explaining how they relate to each other and why some were better than others.

In the **Implementation** chapter, It's shown how the project was made.

In the **Conclusion** chapter, I mention the conclusion I derived from developing the proposed solution.

## Chapter 2

# State of the art

### 2.1 Artificial Intelligence

Artificial intelligence is the idea of programming machines to be capable of thinking like humans, as a means to facilitate the human workload. AI is split between two branches, applied AI, which uses principles of simulating human thought to carry out one specific task, and generalised AI, which seeks to develop machines capable to turn their hands to any task, much like a person.

#### 2.1.1 Applied AI

Applied AI enhances software applications and puts advanced machine learning to use, providing high levels of accuracy and adaptation over time.

##### Advantages

- Rapid decision-making

Applied AI enables "human-like" judgment to reduce error and predict outcomes, achieving end-to-end process automation and enhancing smart device ecosystems.

- Computer capability with a human touch

Applied AI bridges the gap between digital and physical worlds while striving to limit the impact of model error, social ethics and human bias in the process.

- Efficiency

Applied AI improves efficiency and throughput, saving time and money.

- Automation

Applied AI frees up employees from manual processes by increasing automation.

- Increased revenue

Applied AI increases profitability by identifying and helping solve complex business issues through its machine learning and adaptive capabilities.

#### 2.1.2 Generalised AI

Generalised AI, also known by strong AI, is a theoretical form of AI used to describe a certain mindset of AI development. It aims to create intelligent machines that are indistinguishable

from the human mind, for that a AI machine would have to learn through input and experiences, constantly progressing and advancing its abilities over time, although it only exists today as a theoretical concept versus a tangible reality, research is constantly being done.

### **Advantages**

- Multi-tasking

Generalised AI can perform a variety of functions, eventually teaching itself to solve for new problems.

- Independent

Generalised AI doesn't require human interference to define the parameters of its learning algorithms and to provide the relevant training data to ensure accuracy.

### **2.1.3 Machine Learning**

Machine learning is a branch of artificial intelligence that uses statistical techniques or algorithms to enable a computer to become better at what it does.

Machine learning discovers patterns and structures, finds unusual data points, predicts values and categories, and solves a variety of problems. It also allows companies to consume very large amounts of data from a variety of sources and make sense out of it.

## **2.2 Machine Learning Concepts**

Depending on the problem and data available, various methods can be used to create a prediction, these methods differ on the way the solution is approached and how the data is used.

### **2.2.1 Supervised Learning**

Supervised learning is based on training a data sample from a data source with correct classification already assigned (Sathya and Abraham 2013). Such techniques are utilized in feed forward or Multi Layer Perceptron models, diagram in Figure 2.1.

These MLP have three distinctive characteristics:

- One or more layers of hidden neurons that are not part of the input or output layers of the network that enable the network to learn and solve any complex problems.
- The non-linearity reflected in the neuronal activity is variant.
- The interconnection model of the network exhibits a high degree of connectivity.

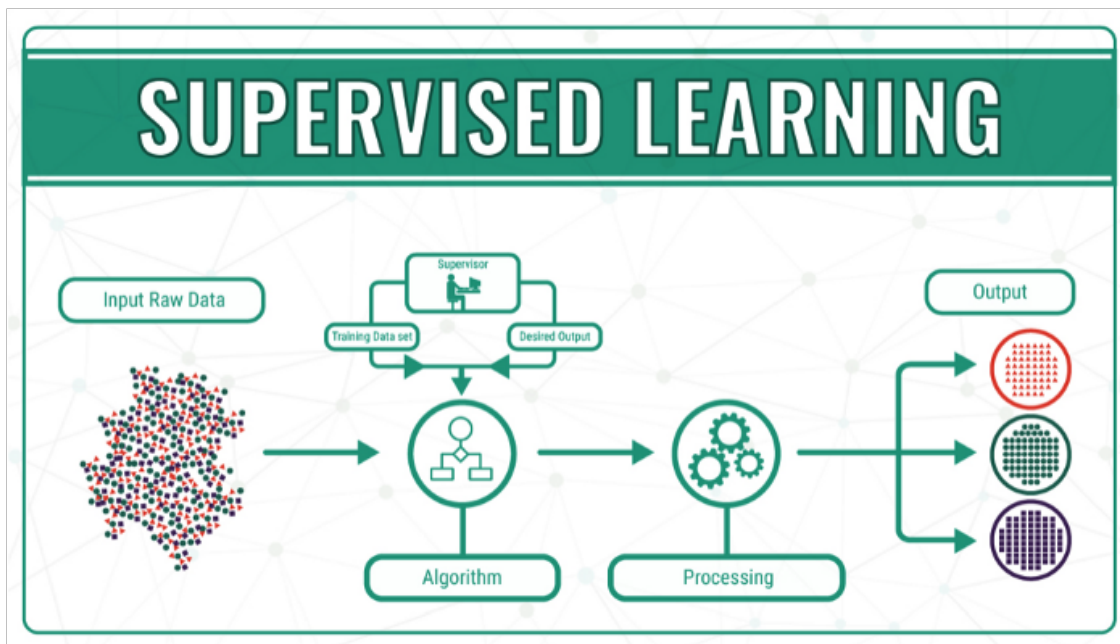


Figure 2.1: Supervised Learning

### 2.2.1.1 Classification Algorithms

Classification is one of the most frequently encountered decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to the object in question (*Classification Algorithms in Machine Learning: How They Work* 2020).

### 2.2.1.2 Logistic Regression

Logistic regression is a calculation used to predict a binary outcome. Independent variables are analyzed to determine the outcome with the results falling into one of two categories. These variables can be categorical or numeric, but the dependent variable is always categorical.

$$P(Y = 1|X) \text{ or } P(Y = 0|X)$$

It calculates the probability of dependent variable  $Y$ , given independent variable  $X$ .

#### Advantages

- Simple to implement.
- Effective.
- Feature scaling not needed.
- Hyper-parameters tuning not required.

#### Disadvantages

- Poor performance on non-linear data.
- Poor performance with irrelevant and highly correlated features.

- Can be easily outperformed by other algorithms.
- High reliance on proper presentation of data.

### 2.2.1.3 Naive Bayes Classifier

Naive Bayes calculates the possibility of whether a data point belongs within a certain category or not.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

#### Advantages

- Very fast for real time predictions.
- Scalable with large datasets.
- Insensitive to irrelevant features.
- Multi class prediction.
- Good performance with high dimensional data.

#### Disadvantages

- Independence of features impacts negatively.
- Probability outputs are not very accurate.
- Training data is required to accurately represent population.

### 2.2.1.4 K-Nearest neighbor

K-nearest neighbors (K-NN) is a pattern recognition algorithm that used training datasets to find the k closest relatives in future examples.

When K-NN is used in classification, the calculations are done to place data within the category of its nearest neighbor. If  $k = 1$ , then it would be placed in class nearest 1. K is classified by a plurality poll of its neighbors, example in Figure 2.2.

#### Advantages

- Simple to understand.
- No assumption about data.
- Constantly evolving.
- Can solve multi-class problems.

#### Disadvantages

- Slow for large datasets.
- Worse for datasets with a large number of features.
- Has to scale data.
- Does not work well on imbalanced data.
- Sensitive to outliers.

- Can't deal with missing values.

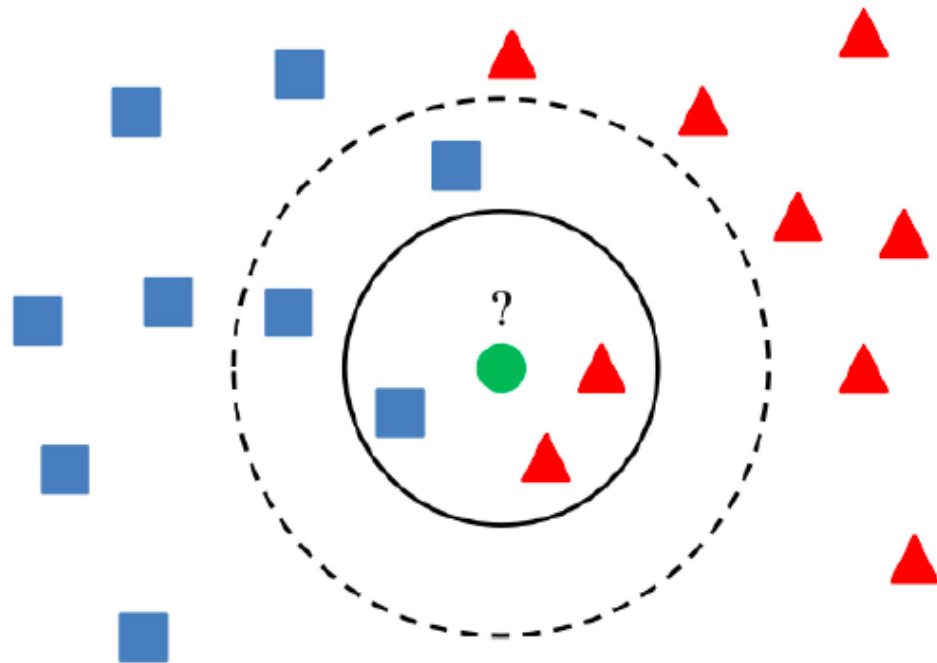


Figure 2.2: K-Nearest Neighbor Method

#### 2.2.1.5 Decision Tree

A decision tree is great for classification problems, as it's able to order classes on a precise level. It works like a flow chart, separating data points into two similar categories at a time from the "tree trunk" to "branches" to "leaves", where the categories become more finitely similar. This creates categories within categories, allowing for organic classification with limited human supervision, shown example in Figure 2.3.

##### Advantages

- Normalization or scaling of data not needed.
- Can handle missing values.
- Easy to explain.
- Easy visualization.
- Automatic feature selection.

##### Disadvantages

- Prone to overfitting.
- Sensitive to data changes.
- Higher time required to train.

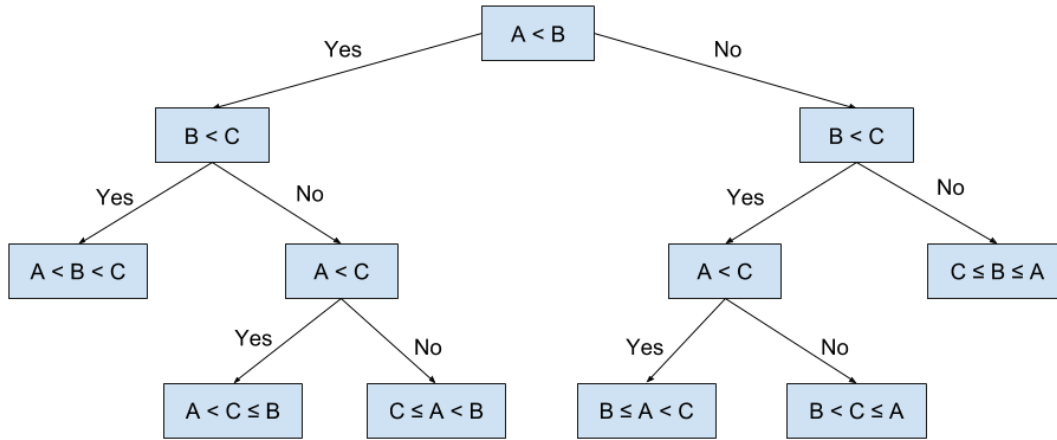


Figure 2.3: Decision Tree Method

### 2.2.1.6 ARIMA

Auto Regressive Integrated Moving Average (ARIMA) is based on an Auto Regressive Moving Average (ARMA) but fitted to time series data to better understand the data or to predict points in the future. The model uses three integers greater than or equal to zero:

- Auto Regressive Part (AR)

Dependent relationship between an observation and some number of lagged observations.

- Integrated Part (I)

The differencing of raw observations in order to make the time series stationary

- Moving Average Part (MA)

Dependency between an observation and a residual error from a moving average applied to lagged observations.

An ARIMA process has the following formula:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \epsilon_t$$

Where  $L$  is the lag operator, the  $\phi_i$  are the parameters of the auto regressive part of the model, the  $\theta_i$  are the parameters of the moving average part, the  $\epsilon_i$  are error terms and  $d$  is a positive integer that controls the level of differencing.

### 2.2.1.7 Random Forest

The random forest is an expansion of a decision tree, in that, firstly it's constructed some axis real world decision trees with training data, then fit your new data within one of the trees as a "random forest".

It averages the data to connect it to the nearest tree on the data scale. Random forest models are helpful as they remedy for the decision tree's problem of "forcing" data points within a category unnecessarily.

The algorithm is characterized by these parameters:

- Max Features

Represents the maximum number of features the algorithm is allowed to try in an individual tree. By increasing this value, the accuracy of the model increases as the number of considered options for each node is higher. This value greatly impacts the speed of the model's training, so striking a balance is necessary for each occasion.

- N Estimators

Number of trees to be created before taking the maximum voting or averages of predictions. Like the number of max features, the higher the number of n estimators the better the accuracy but lower the speed.

- Min Sample Leaf

Sample size to be considered in each leaf of the decision tree (the last node of each branch). This value also represents the complexity on each leaf, meaning that a smaller value can introduce noise in the training data.

#### Advantages

- Able to deal with highly correlated data.
- Reduced error.
- Good performance on imbalanced datasets.
- Good handling of huge amount of data.
- Overcomes missing data.
- Little impact of outliers.
- No problem of overfitting.
- Useful to extract feature importance.

#### Disadvantages

- Features need to have some predictive potential.
- Predictions of the trees need to be uncorrelated.
- Difficult to measure optimal parameters.

### 2.2.1.8 Support Vector Machines

A support vector machine (SVM) uses algorithms to train and classify data within degrees of polarity, taking it to a degree beyond X/Y prediction.

The SVM assigns a hyper-plane that best separates the results by drawing a line between the clusters and thus declaring each cluster as positive and negative. The more complex the data, the more accurate the predictor will become, shown in Figure 2.4.

#### Advantages

- Optimized for datasets with higher dimensions.
- One of the best algorithms when classes are separable.
- Outliers have less impact.
- Suited for extreme case binary classification.

#### Disadvantages

- Requires a large amount of time to process larger datasets.
- Poor performance with overlapped classes.
- Selecting appropriate hyper parameters is important.
- Selecting the appropriate kernel function is difficult.

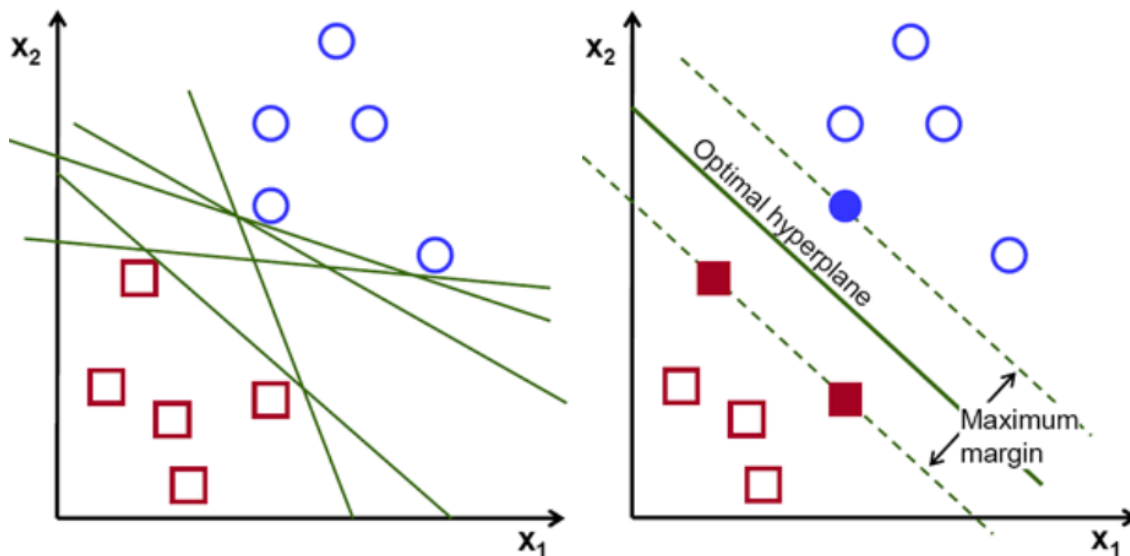


Figure 2.4: Support Vector Machine Method

### 2.2.1.9 Regression Algorithms

Regression outputs for probability as a continuous number value between 0 and 1. It predicts quantities or the probability that something will occur, like property values in a particular location or the effects an economic crisis may have on the stock market (Bhatia 2021).

### 2.2.1.10 Simple Linear Regression

Simple linear regression is a statistical method that enables users to summarise and study relationships between two continuous variables. Linear regression is a linear model wherein a model that assumes a linear relationship between the input variables and the single output variable, example depiction in Figure 2.5.

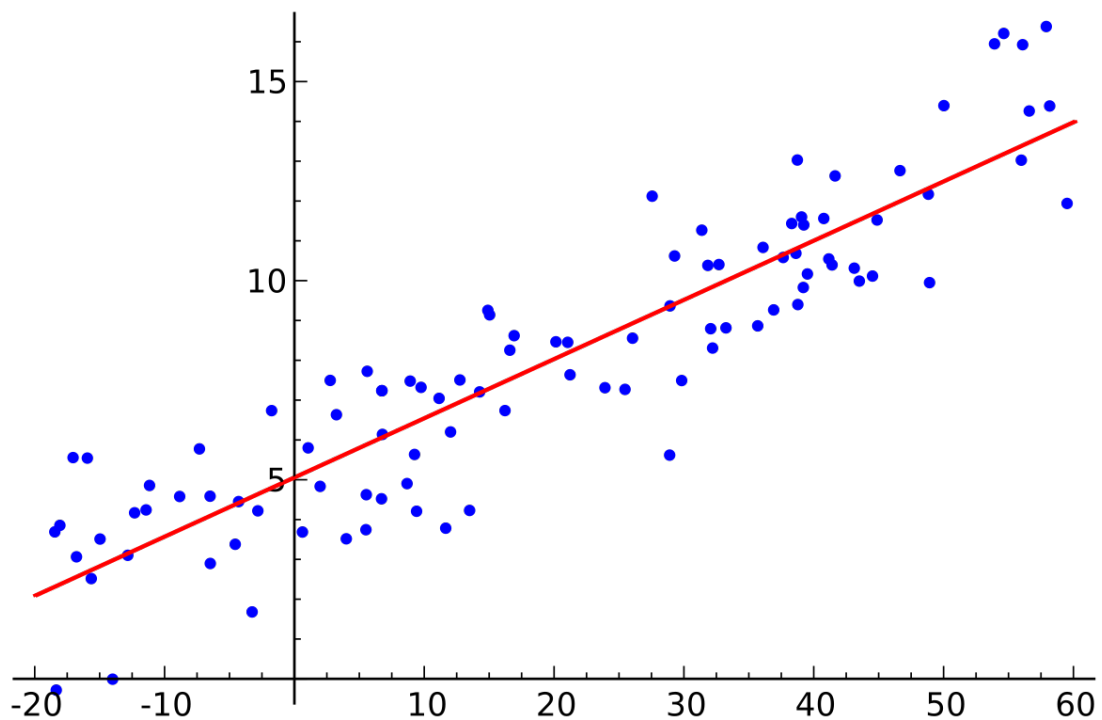


Figure 2.5: Simple Linear Regression Method

### 2.2.1.11 Lasso Regression

LASSO stands for Least Absolute Selection Shrinkage Operator wherein shrinkage is defined as a constraint of parameters. The goal of lasso regression is to obtain the subset of predictors that minimize prediction error for a quantitative response variable. The algorithm operates by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink towards a zero.

Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model. Variables with non-zero regression coefficients are most strongly associated with the response variable. Explanatory variables can be either quantitative or categorical. This lasso regression analysis is a shrinkage and variable selection method and it helps analysts to determine which of the predictors are the most important, example in Figure 2.6.

#### Advantages

- Good at selecting features.
- Avoids overfitting.

#### Disadvantages

- Selected features are highly biased.
- Algorithm selects arbitrary one feature from a set of correlated features.
- For different data, the feature selected can be very different.
- Prediction performance is worse than other algorithms.

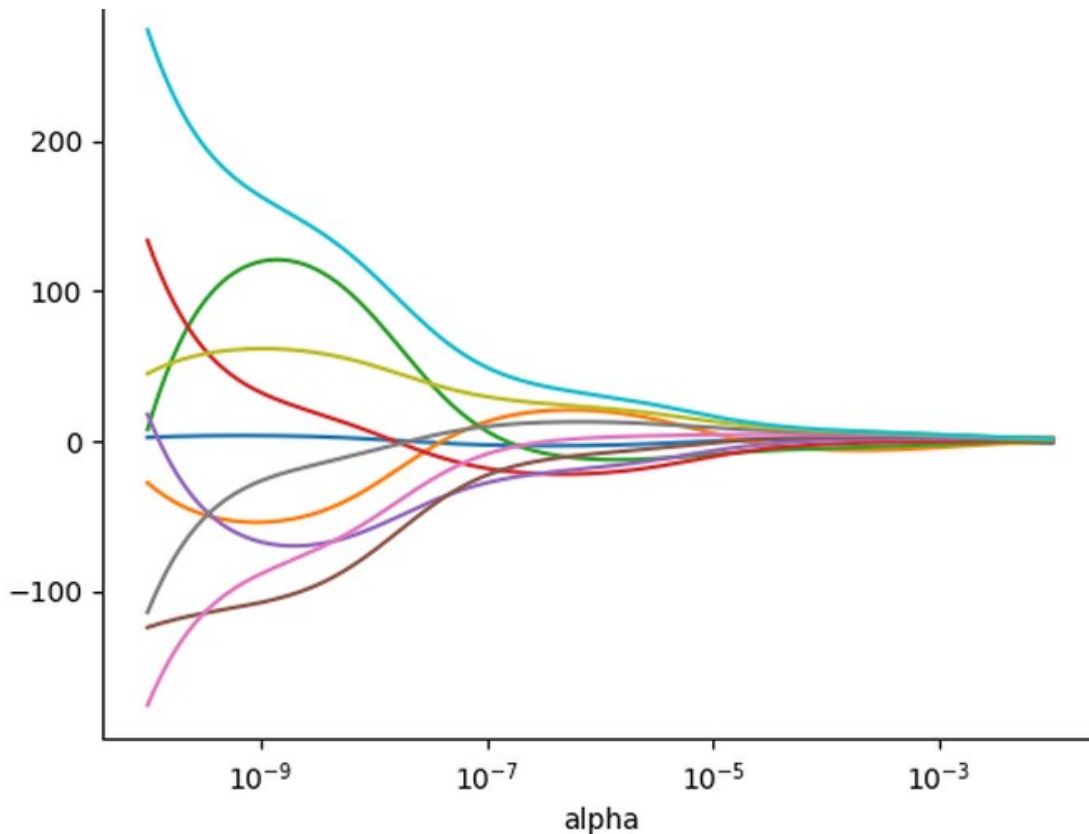


Figure 2.6: Lasso Regression Method

### 2.2.2 Unsupervised Learning

Self-Organizing neural networks learn using unsupervised learning algorithms to identify hidden patterns in unlabelled input data. This unsupervised refers to the ability to learn and organize information without providing an error signal to evaluate the potential solution. The lack of direction for the learning algorithm can sometimes be advantageous, since it lets the algorithm to look back for patterns that have not been previously considered, diagram in Figure 2.7.

The main characteristics are:

- It transforms an incoming signal pattern of arbitrary dimension into one or 2 dimensional map and performs this transformation adaptively.
- The network represents a feed-forward structure with a single computational layer consisting of neurons arranged in rows and columns.
- At each stage of representation, each input signal is kept in its proper context.

- Neurons dealing with closely related pieces of information are close together and they communicate through synaptic connections.

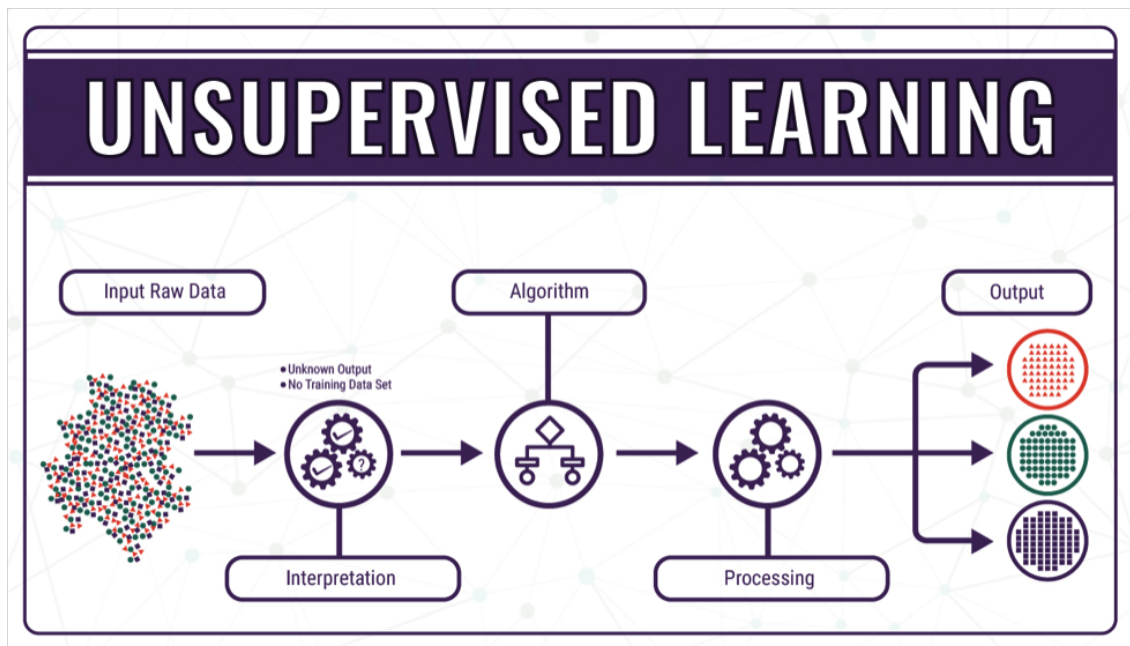


Figure 2.7: Unsupervised Learning

### 2.2.2.1 Clustering

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same group are more similar to other data points in the same group and dissimilar to the data points in other groups (*Clustering in Machine Learning* 2020).

Clustering is used to determine the intrinsic grouping among the unlabeled data present. This algorithm must make some assumptions which constitute the similarity of points and each assumption makes different and equally valid clusters, shown example in Figure 2.8. Some of the methods used are:

- **Density Based:** This method considers the clusters as the dense region having some similarity and difference from the lower dense region of the space. This method has good accuracy and the ability to merge two clusters.
- **Hierarchical Based:** The clusters formed in this method forms a tree shaped structure based on the hierarchy. New clusters are formed using the previously formed one.
- **Partitioning:** This method partitions the objects into k clusters and each partition forms one cluster. This method is used to optimize an objective criterion similarity function such as when the distance is a major parameter.
- **Grid Based:** In this method the data space is formulated into a finite number of cells that form a grid like structure. All the clustering operation done on these grids are fast and independent of the number of data objects.

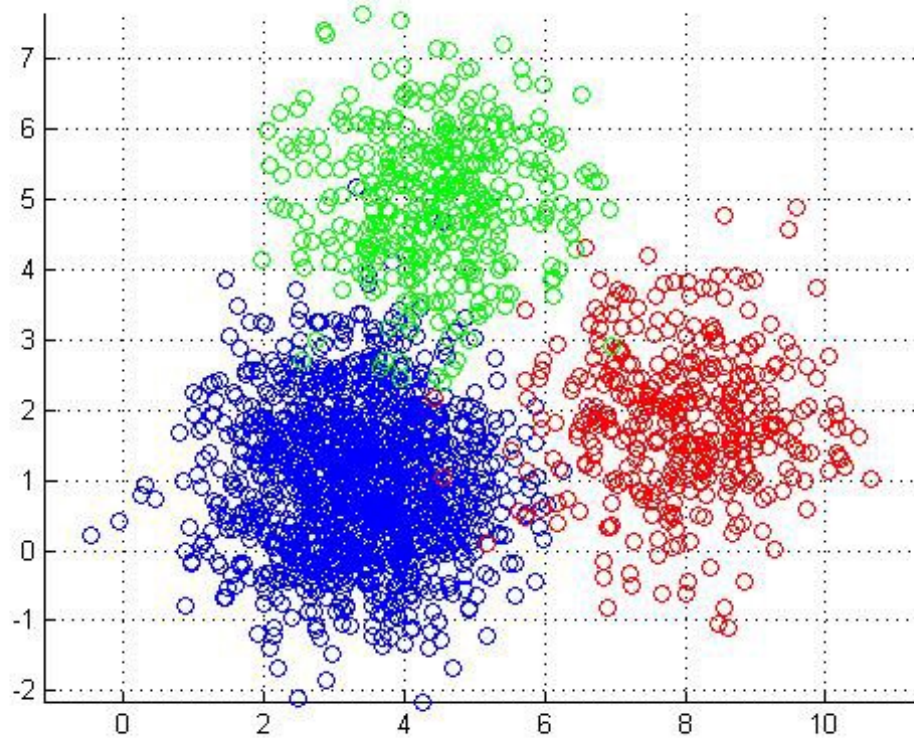


Figure 2.8: Clustering Method

### 2.2.3 Reinforced Learning

Reinforced learning is explained mostly as "trial and error". In reinforcement learning, a machine or computer program chooses the optimal path or next step in a process based on previously learned information. Machines learn with maximum reward reinforcement for correct choices and penalties for mistakes (*Top Machine Learning Algorithms Explained: How Do They Work?* 2020).

Like unsupervised learning, reinforcement models don't learn from labeled data, but instead of finding patterns, they learn by trial and error, diagram in Figure 2.9.

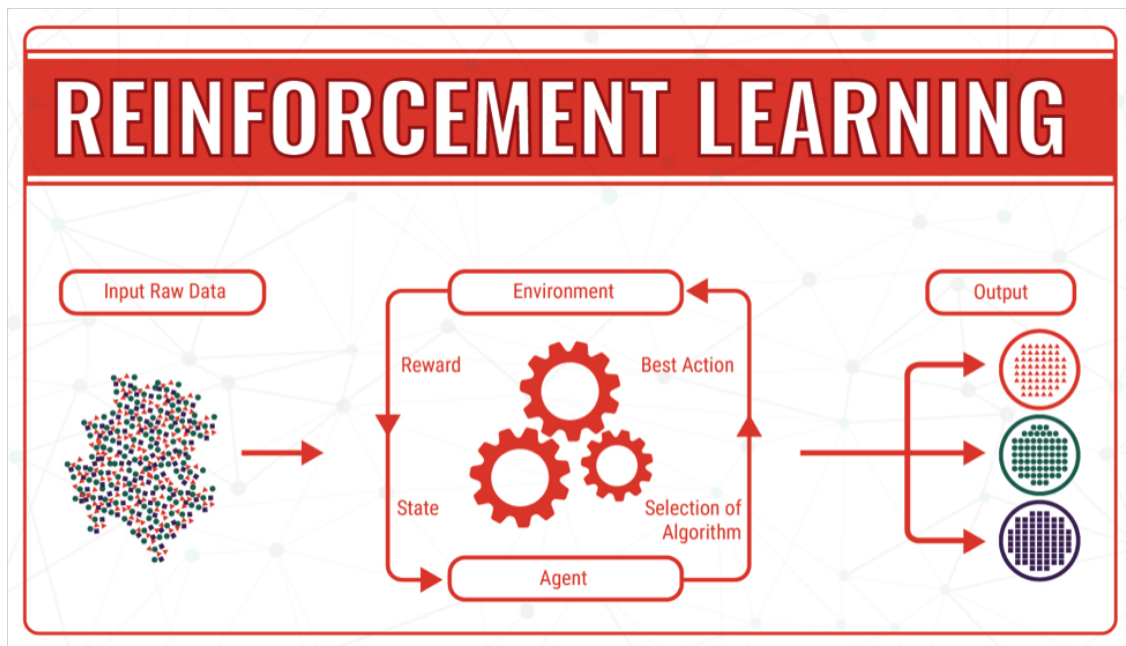


Figure 2.9: Reinforced Learning

### 2.2.4 CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) is a comprehensive data mining methodology and process model that provides anyone with a complete blueprint for conducting a data mining project, depiction in Figure 2.10.

CRISP-DM divides the life cycle of a data mining project into six phases:

#### 1. Business Understanding

- Understanding project objectives and requirements;
- Data mining problem definition;

#### 2. Data Understanding

- Initial data collection and familiarization;
- Identify data quality issues;
- Initial obvious results;

#### 3. Data Preparation

- Record and attribute selection;
- Data cleaning;

#### 4. Modeling

- Run the data mining tools;

#### 5. Evaluation

- Determine if results meet business objectives;

- Identify business issues that should have been addressed earlier;

## 6. Deployment

- Put the resulting models into practice;
- Set up for continuous mining of the data;

## Advantages

- Cost effective.
- Encourages best practices.
- Uniform framework.
- Industry standard.

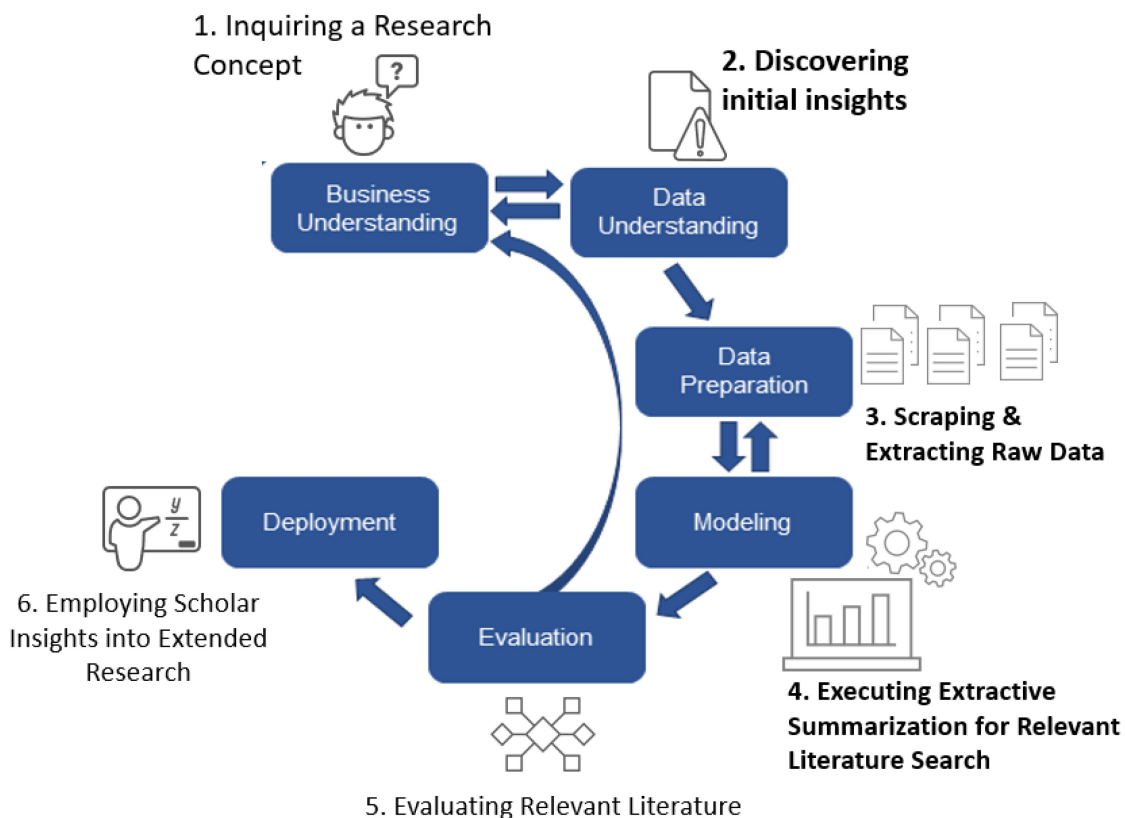


Figure 2.10: CRISP-DM Method

## 2.2.5 LSTM

Long Short-Term Memory (LSTM) networks are a modified version of a recurrent neural networks, which makes it easier to remember past data in memory. LSTM is well-suited to classify, process and predict time series given time lags of unknown duration. It trains the model by using back-propagation (Mittal 2019), example diagram in Figure 2.11.

1. **Input gate**: Discover which value from input should be used to modify the memory. Sigmoid function decides which values to let through and tanh function gives weightage

to the values which are passed deciding their level of importance.

$$i_t = \sigma(W_i.[h_t - 1, x_t] + b_i)$$

$$C_t = \tanh(W_C.[h_t - 1, x_t] + b_C)$$

2. **Forget gate:** Discover what details to be discarded from the block. It's decided by the sigmoid function. it look at the previous state and the current input and outputs a number between 0 and 1.

$$f_t = \sigma(W_f.[h_t - 1, x_t] + b_f)$$

3. **Output gate:** The input and the memory of the block is used to decide the output. Sigmoid function decides which values to let through and tanh function gives weightage to the values which are passed deciding their level of importance and then it's multiplied with the sigmoid output.

$$o_t = \sigma(W_o.[h_t - 1, x_t] + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

### Advantages

- Optimal for:
  - Language modelling.
  - Image processing.
  - Speech and handwriting recognition.
  - Music generation.
  - Language translation.

### Disadvantages

- Fails to resolve vanishing gradients.
- Requires a lot of resources and time.
- Has a limit for long term information.
- Easily affected by random weights.
- Prone to overfitting.

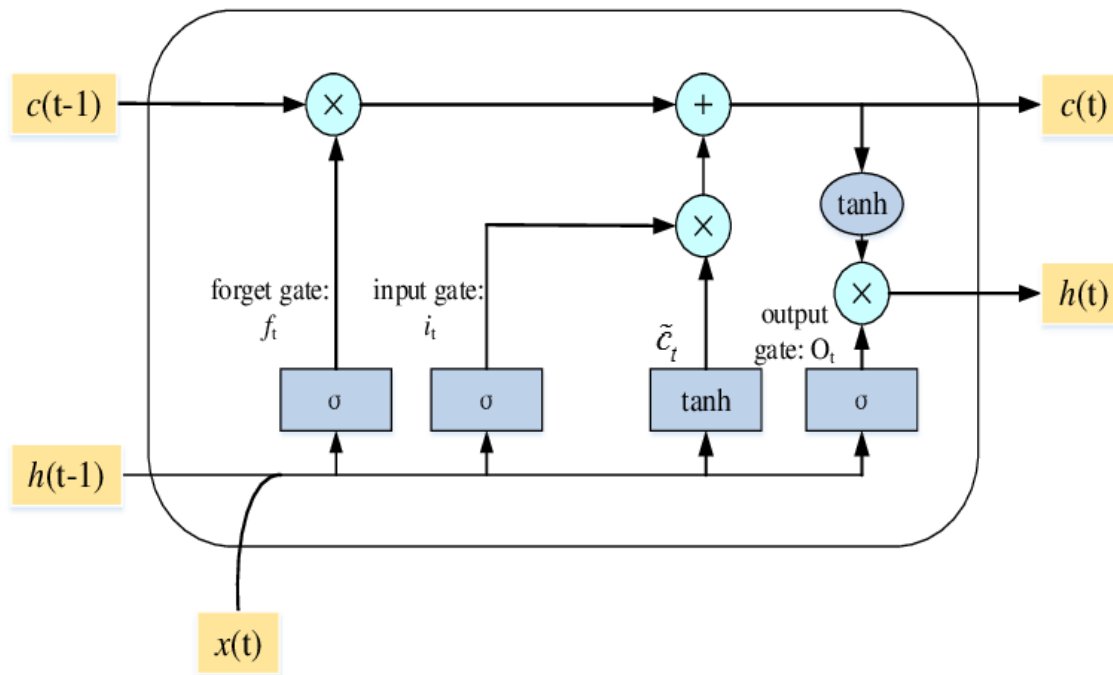


Figure 2.11: LSTM Method

## 2.3 Related work

### 2.3.1 Neural Networks

A Neural Network is a bio-inspired system with several single processing elements, called neurons. The neurons are connected each other by joint mechanism which is consisted of a set of assigned weights. Multi layer perceptron (MLP) is a common approach in regression-type problems. A MLP network has three layers: input layer, output layer, and hidden layer. Neuron takes the values of inputs parameters, sums them up according to the assigned weights, and adds a bias. By applying the transfer function, the value of the outputs would be determined. The number of neurons in input layer corresponded to the number of input parameters. The architecture of a typical MLP is presented in Figure 2.12 (Hedayati, Moghaddam, and Esfandaryi 2016).

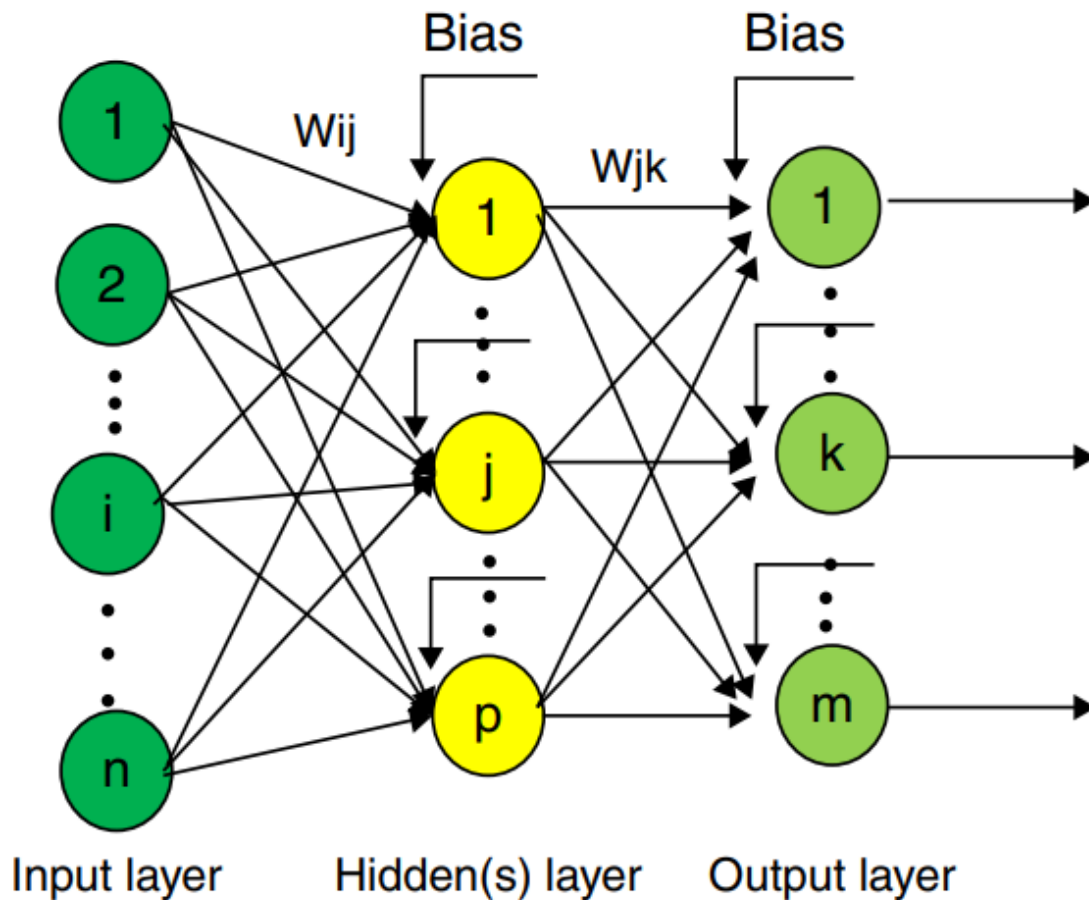


Figure 2.12: Neural Network

There are several learning techniques such as scaled conjugate gradient (SCG), Theory of Levenberg-Marquardt (LM), Theory of Bayesian regularization (BR), etc. that are used for training and developing the constructed models

#### 2.3.1.1 Theory of Levenberg- Marquardt

Developed to approximate the second-order training speed to avoid the computation of the Hessian matrix and used for solving a non-linear least square problem. One of the significant merits of this approach is that it performs similarly to gradient search and Newton method for large values, merges the best attributes of the steepest-descent algorithm and the Gauss-Newton technique.

#### 2.3.1.2 Theory of scaled conjugate gradient

Used with the intent to reduce the performance function by doing the search along the conjugate gradient direction to find the step size due to the nature of most back-propagation algorithms modifying the step size in each iteration causing an impact on the speed of the system. A key advantage of this algorithm is that it does not line search at each iteration as compared to all other methods, lowering the need for more computation-ally expensive hardware.

### 2.3.1.3 Theory of Bayesian regularization

Bayesian regularized artificial neural networks eliminate or reduce the requirement for lengthy cross-validation. Hence perform more robustly than standard back-propagation. Is a mathematical process that converts a non-linear regression into a structured statistical problem in the manner of ridge regression. A key advantage of this algorithm is that it considers the probabilistic nature of the weights in the network related to the given data set.

### 2.3.1.4 Case Study

In 2015 a study was made to predict the NASDAQ index. The methodology used in this study considered the short-term historical stock prices as well as the day of week as inputs (Hedayati, Moghaddam, and Esfandiyari 2016).

Daily stock exchange rates of NASDAQ from January 28, 2015 to 18 June, 2015 are used to develop a robust model. First 70 days (January 28 to March 7) are selected as training dataset and the last 29 days are used for testing the model prediction ability. For constructing the model, training, and testing procedure MATLAB software R2010a was used. The performance of ANNs was evaluated using the determination coefficient ( $R^2$ ) and the mean square error (MSE) of the modeled output [1].

### 2.3.1.5 Results

The study divided the experiments in two types, first using four prior working days and then nine prior working days. For each type, they did 7 experiments with the same neural network model, with 3 hidden layers with 20-40-20 neurons respectably and using different training and transfer functions, see Table 2.1. From the results, we can notice that the  $R^2$  score is very good throughout the training and tests done, meaning that we could theoretically use any of the training and transfer functions for our purpose that we would get good results. Overall, training with the Levenberg- Marquardt function gave the best results.

Table 2.1: Trained models and their  $R^2$  score

| No. | Training function | Transfer function | $R^2$  |        |            |        |
|-----|-------------------|-------------------|--------|--------|------------|--------|
|     |                   |                   | Train  | Test   | Validation | Total  |
| 1   | LM                | TANSIG            | 0.9925 | 0.9869 | 0.8864     | 0.974  |
| 2   | LM                | PURELIN           | 0.9457 | 0.9675 | 0.9027     | 0.9395 |
| 3   | LM                | LOGSIG            | 0.9989 | 0.9698 | 0.7339     | 0.9475 |
| 4   | OSS               | LOGSIG            | 0.9166 | 0.9133 | 0.8669     | 0.9069 |
| 5   | OSS               | PURELIN           | 0.7016 | 0.8824 | 0.8230     | 0.7675 |
| 6   | OSS               | TANSIG            | 0.9386 | 0.8917 | 0.9408     | 0.9267 |
| 7   | GDA               | LOGSIG            | 0.9016 | 0.8308 | 0.8497     | 0.8649 |

After determining the optimal training function, they had to determine the optimal number of hidden layers and neurons to use. The method they used was to randomly divide data retrieved from 99 days, leaving 60% to train, 20% to validate and the rest for the testing.

They compared the model validation'  $R^2$  score of various structures and have chosen the best one on both types of experiments, refer to Table 2.2 and Table 2.3.

#### Four prior working days

Table 2.2: Various model structures for four working days and their  $R^2$  score

| No. | Structure      | $R^2$  |         |            |        |
|-----|----------------|--------|---------|------------|--------|
|     |                | Train  | Test    | Validation | Total  |
| 1   | 2              | 0.8177 | 0.9616  | 0.9493     | 0.8692 |
| 2   | 5              | 0.9250 | 0.9188  | 0.9605     | 0.9264 |
| 3   | 5-5            | 0.9229 | 0.9724  | 0.8631     | 0.9212 |
| 4   | 5-10           | 0.2185 | -0.0165 | -0.2065    | 0.0838 |
| 5   | 10-10          | 0.9534 | 0.9602  | 0.6811     | 0.9344 |
| 6   | 10-20          | 0.9059 | 0.9758  | 0.9108     | 0.9263 |
| 7   | 40-40          | 0.9003 | 0.9639  | 0.9616     | 0.9264 |
| 8   | 50-100         | 0.9576 | 0.9324  | 0.9393     | 0.9483 |
| 9   | 100-200        | 0.9390 | 0.9466  | 0.9533     | 0.9393 |
| 10  | 200-300        | 0.9267 | 0.9642  | 0.8822     | 0.9276 |
| 11  | 20-40-20       | 0.9386 | 0.8917  | 0.9408     | 0.9267 |
| 12  | 20-50-20       | 0.9403 | 0.9417  | 0.9077     | 0.9374 |
| 13  | 50-100-50      | 0.6837 | 0.8108  | 0.7785     | 0.7326 |
| 14  | 20-40-40-20    | 0.8990 | 0.8445  | 0.8093     | 0.8739 |
| 15  | 10-20-20-10    | 0.8977 | 0.9602  | 0.9015     | 0.9109 |
| 16  | 10-20-20-20-10 | 0.9304 | 0.9341  | 0.9456     | 0.9329 |

As seen on Table 2.2,  $R^2$  score had desirable values when the number of hidden layers was 2 and the number of neurons in hidden layers were 40. It is worthwhile noting that any changes in number of neurons would influence the model proficiency.

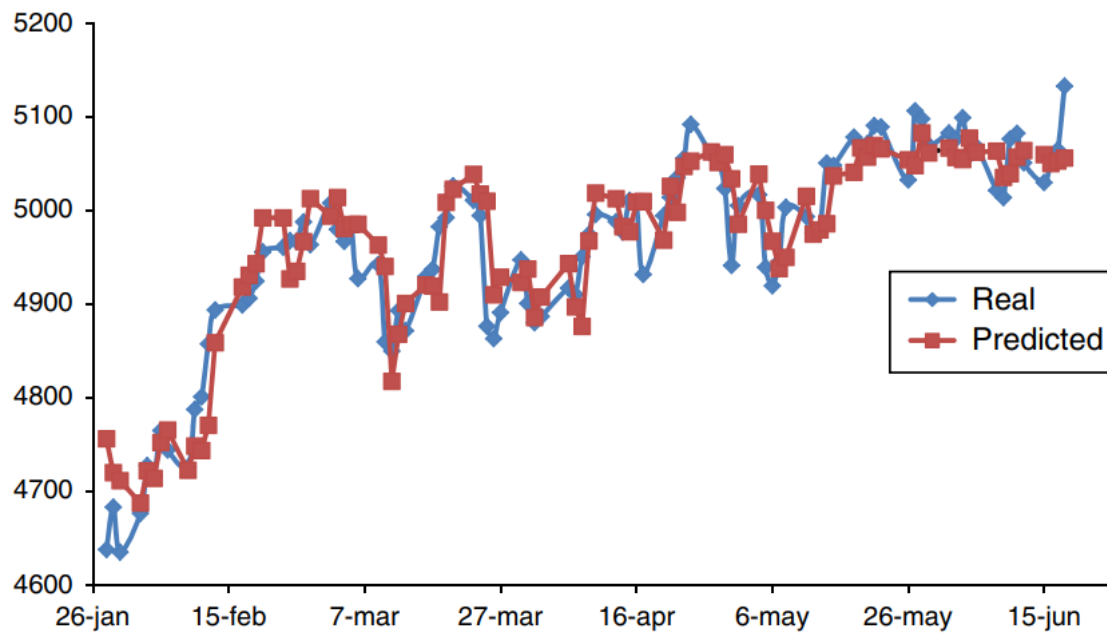


Figure 2.13: Real and predicted NASDAQ index values for four prior days from 28 January to 18 June 2015

Figure 2.13 shows the real and predicted NASDAQ index values for four prior days in 99 days.

### Nine prior working days

Table 2.3: Various model structures for nine working days and their  $R^2$  score

| No. | Structure      | $R^2$  |        |            |        |
|-----|----------------|--------|--------|------------|--------|
|     |                | Train  | Test   | Validation | Total  |
| 1   | 5              | 0.8195 | 0.8480 | 0.8425     | 0.8274 |
| 2   | 10             | 0.8025 | 0.7113 | 0.8608     | 0.7882 |
| 3   | 5-5            | 0.8437 | 0.7762 | 0.8518     | 0.8280 |
| 4   | 5-10           | 0.8127 | 0.8554 | 0.8262     | 0.8180 |
| 5   | 20-20          | 0.8344 | 0.9116 | 0.7524     | 0.8389 |
| 6   | 20-30          | 0.8859 | 0.8690 | 0.8756     | 0.8707 |
| 7   | 50-100         | 0.8335 | 0.9028 | 0.8300     | 0.8292 |
| 8   | 200-300        | 0.8473 | 0.7899 | 0.8938     | 0.8506 |
| 9   | 300-400        | 0.8476 | 0.8685 | 0.7988     | 0.8435 |
| 10  | 20-50-20       | 0.8648 | 0.7937 | 0.8809     | 0.8372 |
| 11  | 20-40-20       | 0.9318 | 0.8827 | 0.9622     | 0.9262 |
| 12  | 50-100-50      | 0.8697 | 0.8367 | 0.8388     | 0.8552 |
| 13  | 20-40-40-20    | 0.7761 | 0.9205 | 0.8601     | 0.8188 |
| 14  | 20-50-50-20    | 0.8359 | 0.8462 | 0.8977     | 0.8443 |
| 15  | 10-20-20-20-10 | 0.8578 | 0.8348 | 0.6704     | 0.8107 |

In Table 2.3, it is found that a network with three hidden layers and 20-40-20 neurons in hidden layers was the optimized network.

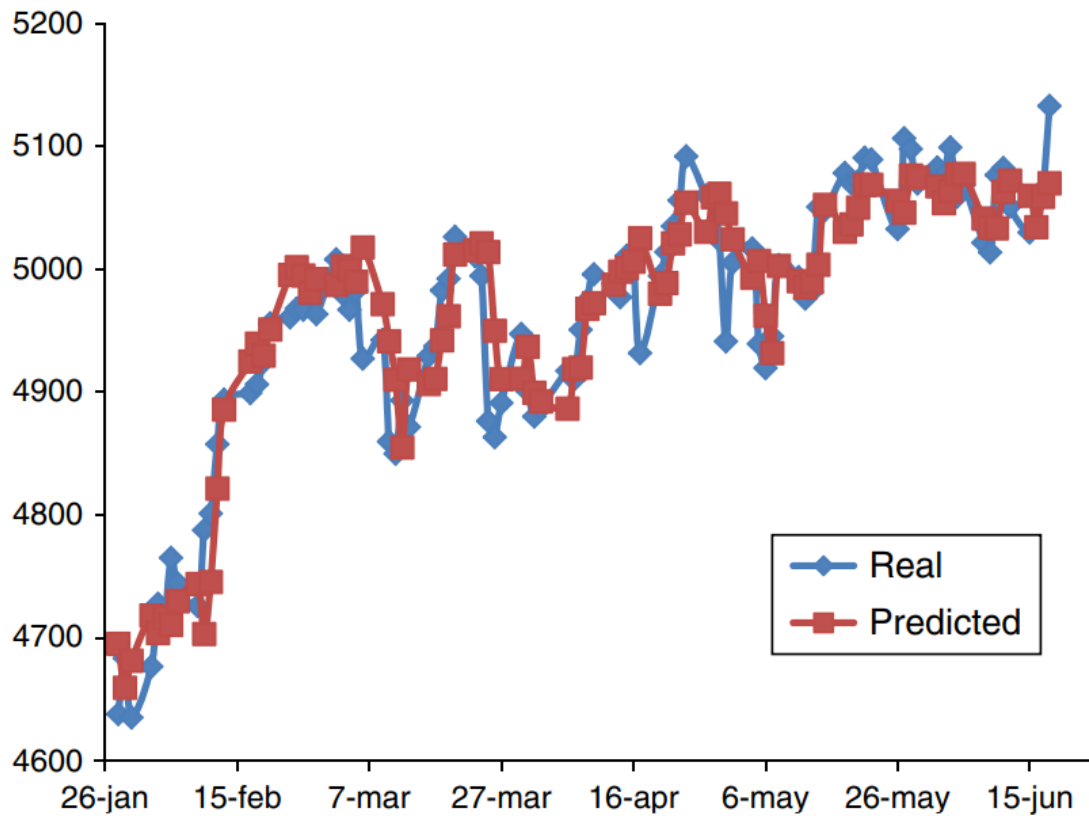


Figure 2.14: Real and predicted NASDAQ index values for nine prior days from 28 January to 18 June 2015

Figure 2.14 shows the real and predicted NASDAQ index values for nine prior days in 99 days. Accordingly, there is no distinct difference between the prediction ability of the four and nine prior working days as input parameters.

### 2.3.1.6 Conclusion

The model uses the values of NASDAQ exchange rate of last four and nine working days as well as the day of week as the input parameters. For four prior working days, applying OSS training method and TANGSIG transfer function in a network with 20-40-20 neurons in hidden layers resulted in an optimized trained network with  $R^2$  values of 0.9408 for validation dataset. For this dataset, the maximum  $R^2$  values for the networks with OSS training method and TANGSIG transfer function would be obtained when the number of hidden layers was 2 and the number of neurons was 40-40. For nine prior working days a network with 20-40-20 neurons in hidden layers, OSS training method and LOGSIG transfer function are the optimized network with validation  $R^2$  of 0.9622. The model outputs show that there is no distinct difference between the prediction ability of the four and nine prior working days as input parameters.

### 2.3.1.7 Case Study 2

In 2020, an article was written with the intent to report an attempt of predicting the stock market using LSTM. It used data of daily opening prices of two stocks in the New York Exchange NYSE (GOOGL and NKE) extracted from Yahoo! Finance. The GOOGL's data covered the period going from 19/8/2004 to 19/12/2019 and the NKE's data was from 4/1/2010 to 19/12/2019 (Moghar and Hamiche 2020).

The model was build using the LSTM method, it used 80% of the data for the training process and the remaining for the testing process. They used the mean squared error to optimize the model and trained with various epochs, refer to 2.4.

Table 2.4: LSTM model summary

| Layer (type)                         | Output Shape   | Parameters |
|--------------------------------------|----------------|------------|
| lstm_1 (LSTM)                        | (None, 50, 96) | 37632      |
| dropout_1 (Dropout)                  | (None, 50, 96) | 0          |
| lstm_2 (LSTM)                        | (None, 50, 96) | 74112      |
| dropout_2 (Dropout)                  | (None, 50, 96) | 0          |
| lstm_3 (LSTM)                        | (None, 50, 96) | 74112      |
| dropout_3 (Dropout)                  | (None, 50, 96) | 0          |
| lstm_4 (LSTM)                        | (None, 96)     | 74112      |
| dropout_4 (Dropout)                  | (None, 96)     | 0          |
| dense_1 (Dense)                      | (None, 1)      | 97         |
| Total number of parameters : 260,065 |                |            |
| Trainable parameters : 260,065       |                |            |
| Non-trainable parameters :0          |                |            |

### 2.3.1.8 Results

After training the neural network the result of the testing has shown different results, the number of epochs as well the length of the data have both significant impact on the testing result.

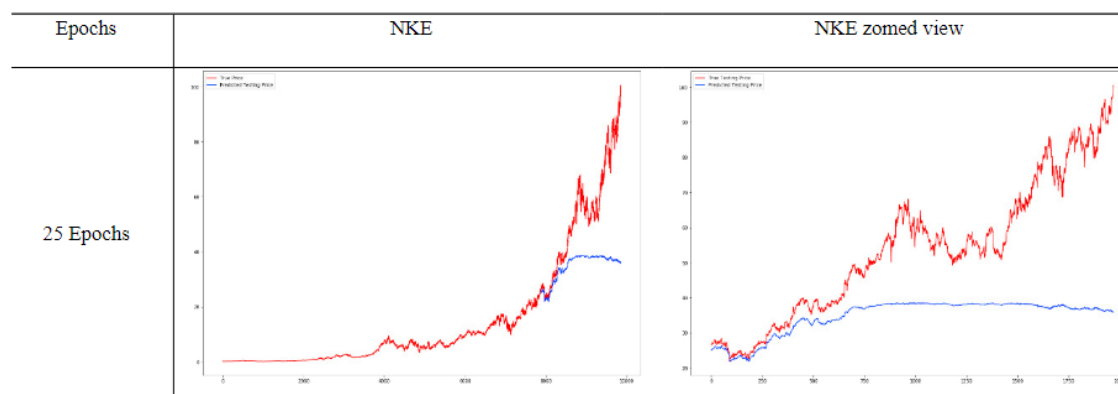


Figure 2.15: Result of the training for the NKE stocks with different dataset times

After observing the data, we can see that at first the data was less volatile and have lower values, in the figure 2.15, the red lines represent the real market values and the blue lines represent the predicted price value, after the NKE start having bigger values, the asset becomes more volatile, then the nature of this asset changed. In the case is better to avoid

this type of change. The model has lost trace of opening prices around 600 to 700 days of testing which conforms the change in data. The result for the dataset for different number of epochs is given by the figure 2.16

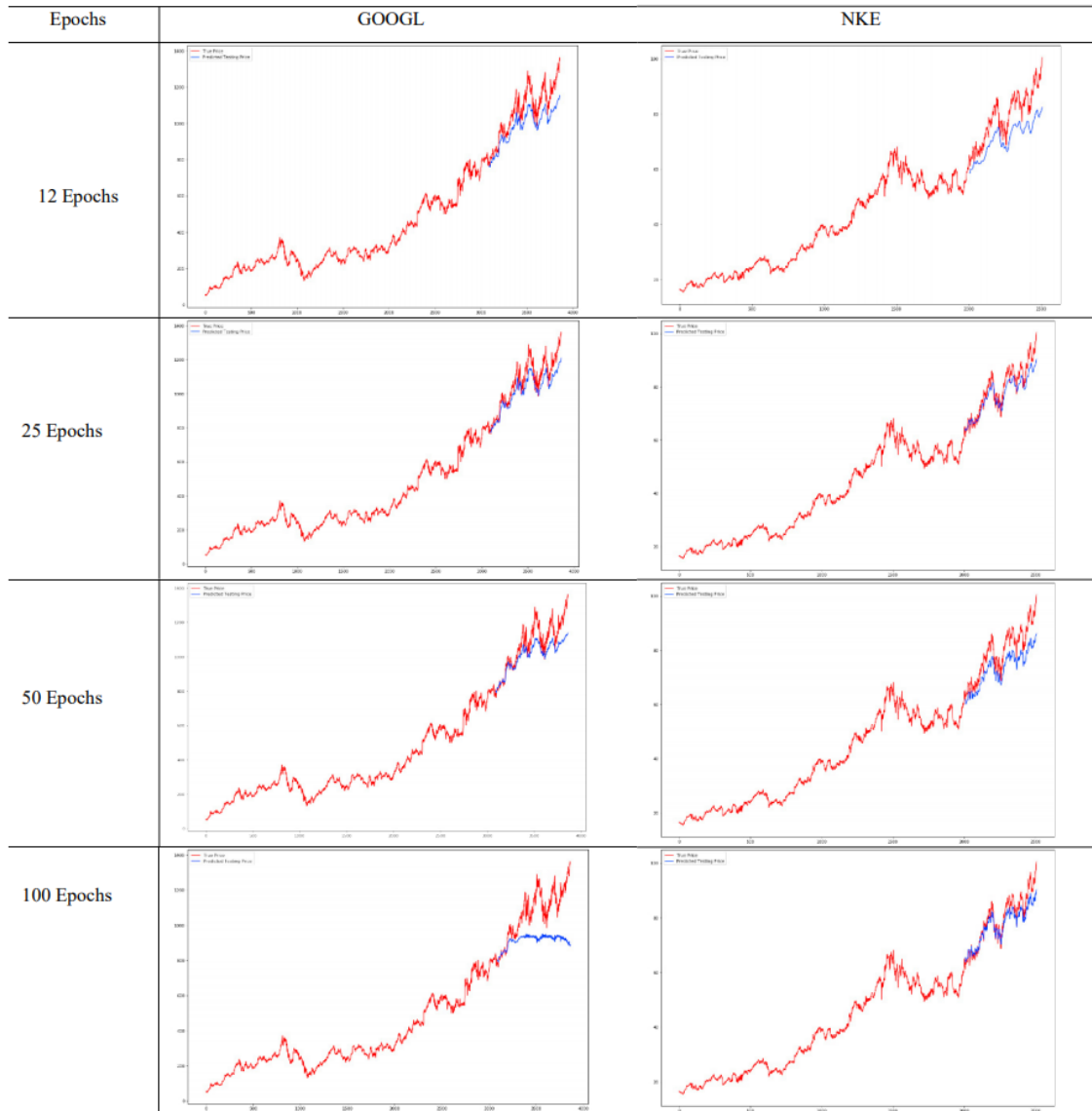


Figure 2.16: Result of training for NKE and GOOGL stocks with different number of epochs

For different datasets we can observe that training the less data and more epochs can improve our testing result and at the same time allows us to have a better forecast and prediction values. The following table 2.5 shows the precision of the training and testing for all the epochs for both NKE and GOOGL asset prices.

### 2.3.1.9 Conclusion

This case study proposes RNN based on LSTM built to forecast future values for both GOOGL and NKE assets, the result of the model has shown some promising results. The

testing result conforms that the model is capable of tracing the evolution of opening prices for both assets

Table 2.5: The value of loss for GOOGL and NKE for different number of epochs

|            | GOOGL                 |          | NKE                   |          |
|------------|-----------------------|----------|-----------------------|----------|
|            | processing Time / sec | Loss     | processing Time / sec | Loss     |
| 12 epochs  | 264                   | 0.0011   | 132                   | 0.0019   |
| 25 epochs  | 550                   | 0.001    | 275                   | 0.0016   |
| 50 epochs  | 1100                  | 6.57E-04 | 550                   | 0.001    |
| 100 epochs | 2200                  | 4.97E-04 | 1100                  | 8.74E-04 |

### 2.3.2 Natural Language Processing

Natural Language Processing (NLP) refers to the manipulation of natural languages, such as text, by software. Some of the most common applications of NLP include speech recognition, chat-bots, auto-correct, virtual assistants, and sentiment analysis (Velay and Daniel 2018).

#### 2.3.2.1 Case Study

On June 2018, a study was made using Natural Language Processing to predict the DJIA stock market index. They extracted indicator vectors using NLP techniques, such as Linear Discriminant Analysis, Decision Trees, Naive Bayes, etc, which allowed them to train different algorithms to predict the trend.

Using news headlines as the source for the training process, they started by removing stop words and named entities as these would prevent the making of a complete autonomous system due to the necessity of adjacent context analysis. Then they converted the remaining words to vectors of fixed length to compose the dataset.

#### 2.3.2.2 Results

After testing most of the combinations of embedding and algorithms, they found that results rarely improved on luck. Deep learning algorithms especially had difficulty figuring out the link between the headlines and the index trend.

These predicted single values for whatever input was provided, generally around 0.5, which shows very low confidence in the selected class. The best results they found were using linear regression on bag-of-words vectorized data, which had a 57% accuracy. The results using this configuration can be found in Table 2.6. They believe this result is due to the algorithms not properly extracting information from the large vectors. The headlines are also very context dependent, with recurrent actors for periods of times, before becoming irrelevant. This means that their models should be trained on short time frames, unless it learns information that will not help predict future trends. This limitation also implies that the system is unable to react to breaking news due to not knowing enough about its context to correctly predict a trend.

Table 2.6: Accuracy of used algorithms

| <b>Algorithm</b>             | <b>Accuracy</b> |
|------------------------------|-----------------|
| <b>Logistic Regression</b>   | <b>0.57</b>     |
| Linear Discriminant Analysis | 0.51            |
| K-Nearest Neighbors          | 0.46            |
| Decision Tree Classifier     | 0.49            |
| Support Vector Machine       | 0.53            |
| Random Forrest               | 0.50            |
| Extreme Gradient Boosting    | 0.52            |
| Naive Bayes                  | 0.53            |
| LSTM                         | 0.55            |
| MLP                          | 0.53            |

### 2.3.2.3 Conclusion

In this case study, it was analyzed several techniques used in predicting market trends using news articles. Different encoding techniques using word embedding were detailed. It was suggested the use of classification algorithms in order to process the news articles and predict the future trends of the DJIA index.

The study resulted in the finding that, over a long period of time, there was little to no correlation between news sentiment and the DJIA index trends. The results of the different word embedding techniques to predict the trend were found out to be just slightly better than random.

The following suggestions might improve the results obtained, but there are still flaws in the idea that a model trained on past news could predict future trends. These flaws can not be resolved until future Machine Learning models are able to learn broad context of unrelated events which humans rely on the process new information, in particular news articles.

## 2.4 Technologies

### 2.4.1 Programming Languages

#### 2.4.1.1 Python

Python was released in 1989 with a philosophy that emphasizes code readability and efficiency. It is an object-oriented programming language, which means it groups data and code into objects that can interact with and modify one another. Java, C++, and Scala are other examples. This sophisticated approach allows data scientists to execute tasks with better stability, modularity, and code readability (Science 2020), logo displayed in Figure 2.17.



Figure 2.17: Python logo

**Advantages**

- Supports all kinds of different data formats;
- Able to easily create datasets and analyse them in-depth;
- Lots of reference and documentation available;
- Pandas can easily scan through the data and clean it;
- Numpy and scikit-learn are very powerful computational Python libraries;
- Matplotlib is able to generate basic to advanced graphs and charts of the data;

**Disadvantages**

- Very dependant of third-party libraries to do most tasks;

**2.4.1.2 R**

R was developed in 1992 and was the preferred programming language of most data scientists for years. It is a procedural language which works by breaking down a programming task into a series of steps, procedures, and subroutines. This is a plus when it comes to building data models because it makes it relatively easy to understand how complex operations are carried out; however, it is often at the expense of performance and code readability, logo displayed in Figure 2.18.



Figure 2.18: R logo

**Advantages**

- Can import data from Excel, CSV and text files;
- Mainly built for statistical and numerical analysis of large data sets;
- Basic functionality encompasses the basics of analysis, optimization, machine learning, etc;

**Disadvantages**

- Package dependant for data collection tasks;
- Relies on third-party libraries for more intensive work;

### 2.4.1.3 PHP

The name PHP stands for Hypertext Preprocessor and denotes a server-side scripting language, which suggests that applications written thereon run on web servers and don't depend upon the online browser. Syntax of PHP language is analogous to C language. It is created by Rasmus Lerdorf and it appeared in 1995. PHP is being widely utilized in developing web applications and become one among main languages for developers to make new applications (*Advantages and Disadvantages of PHP* 2020), logo displayed in Figure 2.19.

Leading social networking sites like Facebook and reputed organizations like Harvard University are both supported PHP which makes PHP popular and increases its credibility.



Figure 2.19: PHP logo

#### Advantages

- Free and open source;
- Platform independent;
- Faster rate of loading over slow internet and speed;
- Easy to learn;
- Progressively stable;
- Powerful library support;
- Very flexible;

#### Disadvantages

- Less secure;
- Not suitable for big content-based web applications;
- Can be daunting due to the large amount of frameworks;
- Basic PHP lacks debugging tools;

### 2.4.1.4 Conclusion

For Data Science tasks both programming languages, Python and R, are suitable but, Python offers more capability for machine learning and deep learning projects due to its community-driven libraries that are constantly update and well documented.

For the development of the web application, PHP will be used as a result of the project being a simple user interface to show the neural network outcome.

## 2.4.2 Development Tools

### 2.4.2.1 Jupyter Notebook App

Jupyter Notebook App produce documents, which contain both computer code and rich text elements. These documents are both human-readable documents containing the analysis description and the results as well as executable documents which can be run to perform data analysis, logo displayed in Figure 2.20.



Figure 2.20: Jupyter logo

#### Advantages

- Visually intuitive organization of code.
- Static display of function outputs.
- Easy replication of notebooks.
- Data exploration and visualization

#### Disadvantages:

- No testing integration.
- Can't integrate with other IDE's.
- Difficult python code styling.

### 2.4.2.2 Spyder

Spyder is a free and open source scientific environment written in Python for Python and designed by and for scientists, engineers and data analysts. It features a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection and beautiful visualization capabilities of a scientific package, logo displayed in Figure 2.21.



Figure 2.21: Spyder logo

#### Advantages

- Well formatted comments.
- Best suited for data analysis and reporting.
- Attractive interface and UI.

**Disadvantages**

- Not suitable for collaborative work.
- Difficult to setup.
- Only works with Python.

**2.4.2.3 Visual Studio Code**

Visual Studio Code is a lightweight but powerful source code editor which runs on your desktop and is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other language and runtimes, logo displayed in Figure 2.22.



Figure 2.22: Visual Studio Code logo

**Advantages**

- Manages any npm-based applications.
- Can customize the entire environment.
- Easy integration with Git.
- Easy to learn.
- Ideal for simple projects.

**Disadvantages**

- Debugging is slow.
- Requires a lot of allocated RAM memory.
- Extension management can get complicated.



## Chapter 3

# Analysis and Design

### 3.1 Problem Description

The stock market is an attractive area for those who intend to obtain extra income, but due to its volatility and complexity, people are susceptible to a great degree of risk if they don't have the necessary knowledge or enough time to spend learning. At the moment there are some platforms that try to predict prices of stock markets by various means although they're very uncertain and prone to subjective parameters that cannot be measured through statistics.

### 3.2 Objectives

This project aims to help those with or without prior experience on stock markers and all its related subjects. By providing this assistance, the user will be able to make better choices on their selection of shares, and also be able to maintain a constant surveillance of possible opportunities. It seems to be very difficult to replace the professionalism of an experienced trader for predicting the stock price, but because of the availability of a remarkable amount of data and technological advancements we can now be able to formulate an appropriate algorithm for prediction whose results can be used to help investors make profit. These algorithms' accuracy is directly proportional to gains made by using them.

### 3.3 Investigation

This project will make use of historical data from various stock markets and with that create machine learning models capable of producing predictions about the future prices of those stock markets, by doing so, it will be able to make a correct assessment and help investors in their decisions.

### 3.4 Data Sources

#### 3.4.1 Datasets

##### 3.4.1.1 Yahoo! Finance

Yahoo! Finance is a media property that is part of Yahoo!'s network. It provides financial news, data and commentary including stock quotes, press releases, financial reports, and original content. It also offers some online tools for personal finance management. In

addition to posting partner content from a wide range of other web sites, it posts original stories by its team of staff journalists.

Neural networks will use datasets provided by Yahoo! Finance, which contain daily historical data of stock prices, example shown in Table 3.1.

Table 3.1: Example of stock market historical data provided by Yahoo! Finance

| Date         | Open   | High   | Low    | Close* | Adj Close** | Volume     |
|--------------|--------|--------|--------|--------|-------------|------------|
| Jan 29, 2021 | 235.99 | 238.02 | 231.35 | 231.96 | 231.43      | 42,468,900 |
| Jan 28, 2021 | 235.61 | 242.64 | 235.09 | 238.93 | 238.38      | 49,111,200 |
| Jan 27, 2021 | 238.00 | 240.44 | 230.14 | 232.90 | 232.36      | 69,870,600 |
| Jan 26, 2021 | 231.86 | 234.18 | 230.08 | 232.33 | 231.80      | 49,169,600 |
| Jan 25, 2021 | 229.12 | 229.78 | 224.22 | 229.53 | 229.00      | 33,152,100 |
| Jan 22, 2021 | 227.08 | 230.07 | 225.80 | 225.95 | 225.43      | 30,172,700 |
| Jan 21, 2021 | 224.70 | 226.30 | 222.42 | 224.97 | 224.45      | 30,749,600 |
| Jan 20, 2021 | 217.70 | 225.79 | 217.29 | 224.34 | 223.82      | 37,777,300 |
| Jan 19, 2021 | 213.75 | 216.98 | 212.63 | 216.44 | 215.94      | 30,480,900 |
| Jan 15, 2021 | 213.52 | 214.51 | 212.03 | 212.65 | 212.16      | 31,746,500 |
| Jan 14, 2021 | 215.91 | 217.46 | 212.74 | 213.02 | 212.53      | 29,480,800 |
| Jan 13, 2021 | 214.02 | 216.76 | 213.93 | 216.34 | 215.84      | 20,087,100 |
| Jan 12, 2021 | 216.50 | 217.10 | 213.32 | 214.93 | 214.44      | 23,249,300 |
| Jan 11, 2021 | 218.47 | 218.91 | 216.73 | 217.49 | 216.99      | 23,047,000 |
| Jan 08, 2021 | 218.68 | 220.58 | 217.03 | 219.62 | 219.12      | 22,956,200 |
| Jan 07, 2021 | 214.04 | 219.34 | 213.71 | 218.29 | 217.79      | 27,694,500 |

For the basis of training the various models and comparing algorithms, information was divided in various datasets, by sector and country

### 3.5 Domain Model

The domain model has the following concepts:

- **User**

The credentials of a user of the application.

- **Investor**

A user of the application that has a portfolio.

- **Portfolio**

A portfolio that belongs to an investor and consists of stock markets.

- **Stock Market**

A stock market is the shares of a company, it belongs to a market place and it can belong to a portfolio.

- **Market Place**

A market place is a group of stock markets and has a time zone.

- **Time Zone**

A time zone belongs to a market place.

- **Stock Market Prediction**

A prediction made by the machine learning model to each stock market.

Domain model diagram in Figure 3.1.

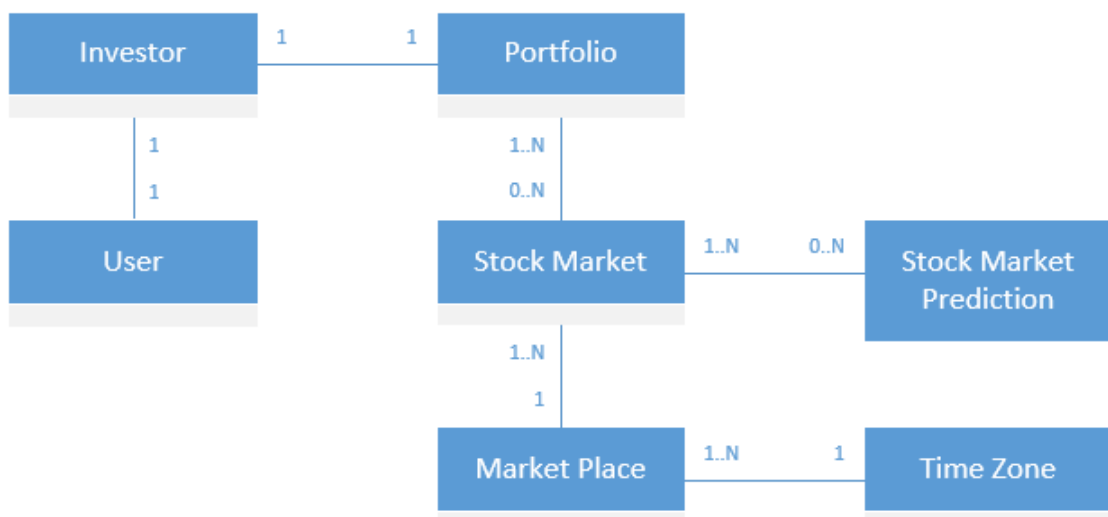


Figure 3.1: Domain Model

## 3.6 Use Cases

Diagram depicting the actions users can perform in the web application, Figure 3.2.

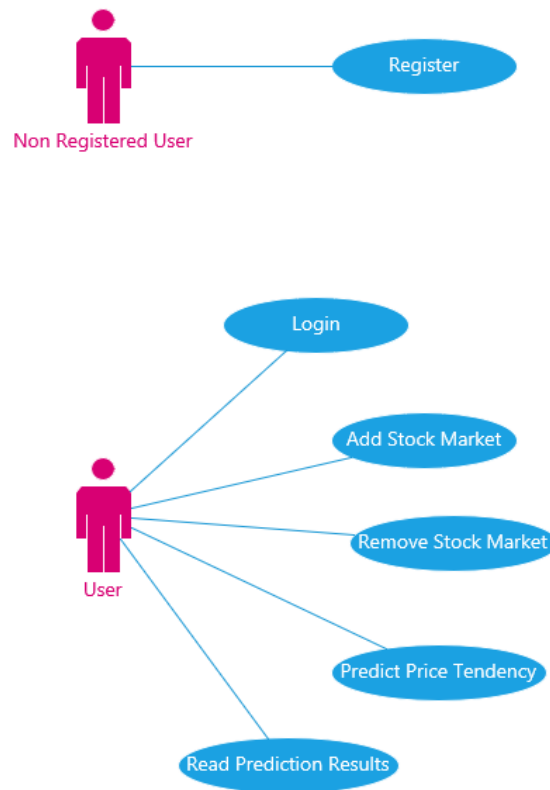


Figure 3.2: Use Cases

### 3.6.1 Register

- **Description:**
  - A unregistered user can register in the web application so that he can make use of the other functionalities.
- **Pre-conditions:**
  - None
- **Post-condition:**
  - Non registered user is registered in the data base.
- **Basic Path:**
  1. The non registered user intends to register.
  2. The system prompts the user with a form to fill in.
  3. The user fills in his information.
  4. The system registers the user in the data base.

### 3.6.2 Login

- **Description:**
  - A user inserts their credential to access the functionalities of the web application.
- **Pre-conditions:**
  - The user is registered.
- **Post-condition:**
  - User gains access to the functionalities of the web application.
- **Basic Path:**
  1. The user intends to Login.
  2. The system prompts the user with a form to fill in.
  3. The user fills in his credentials.
  4. The system redirects the user to the main page.

### 3.6.3 Add Stock Market

- **Description:**
  - A user can add a stock market to his portfolio.
- **Pre-conditions:**
  - The user is registered.
  - There's at least one stock market in the database.
- **Post-condition:**
  - The stock market is registered in the user's portfolio.
- **Basic Path:**
  1. The User intends to add a stock market to his portfolio.
  2. The system prompts the user with a list of stock markets.
  3. The user chooses one or more from the list provided.
  4. The system registers the stock markers chosen by the user in his portfolio.

### 3.6.4 Remove Stock Market

- **Description:**
  - A user can remove a stock market from his portfolio.
- **Pre-conditions:**
  - The user is registered.
  - There's at least one stock market in the user's portfolio.
- **Post-condition:**

- The stock market is removed from the user's portfolio.

- **Basic Path:**

1. The User intends to remove a stock market from his portfolio.
2. The system prompts the user with the list of stock markets from his portfolio.
3. The user chooses one or more from the list provided.
4. The system removes the stock markers chosen by the user from his portfolio.

### 3.6.5 Read Prediction Results

- **Description:**

- A user can read the results of the predictions made for the stock markets in his portfolio.

- **Pre-conditions:**

- The user is registered.
- There's at least one stock market in the user's portfolio.
- There's at least one prediction result of a stock market present in the user portfolio.

- **Post-condition:**

- The user reads the prediction results.

- **Basic Path:**

1. The User intends to read the prediction results.
2. The system shows the registered results to the user.
3. The user chooses one or more from the list provided.
4. The system removes the stock markers chosen by the user from his portfolio.

### 3.6.6 Predict Price Tendency

- **Description:**

- A user wishes to calculate the prediction of the stock markets in his portfolio.

- **Pre-conditions:**

- The user is registered.
- There's at least one stock market in the user's portfolio.

- **Post-condition:**

- A prediction is created for each stock markets in the user's portfolio .

- **Basic Path:**

1. The User intends to calculate a prediction.

2. The system shows the result of the requested prediction.

## 3.7 Non-Function Requirements

Using the FURPS model, the following non-function requirements were considered.

### 3.7.1 Functionality

The web app has to permit the access through smartphones' web browsers.

### 3.7.2 Reliability

The server hosting the web app should be able to be available 24/7 and reduce its proneness to energy failure.

### 3.7.3 Performance

Due to the nature of the stock markets, the server hosting the web app has to withstand the amount of users connected and deliver the complete content with a reasonable speed.

### 3.7.4 Supportability

Additional functionalities may be added to the web app due to user request or by necessity.

## 3.8 Component Model

Diagram that shows the components that the project is composed by, Figure 3.3

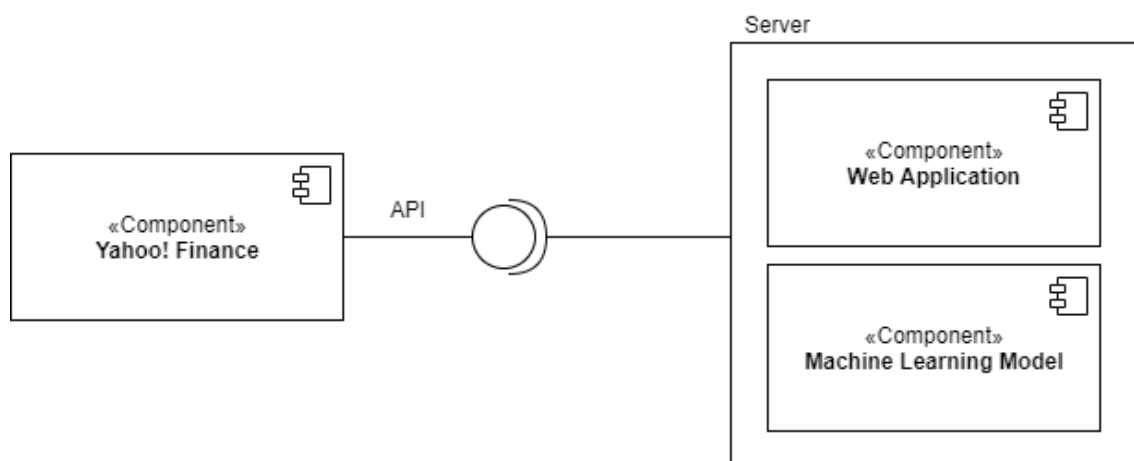


Figure 3.3: Component Model



## **Chapter 4**

# **Value Analysis**

### **4.1 Value**

Value refers to the worth in monetary terms of the technical, economic, service and social benefits a customer or company receives in exchange for the price it pays for a market offering. The customer action is traditionally a purchase, but could be a sign-up, a vote or a visit, while the cost refers to anything a customer must forfeit in order to receive the desired benefit, such as money, data, time and knowledge (Kopp 2020).

### **4.2 Perceived Value**

Perceived value is the customer' evaluation of the merits of a product or service, and its ability to meet their needs and expectation, especially in comparison with its peers. It comes down to the price the public is willing to pay for a good or service. Even a snap decision is a store involves an analysis of a product's ability to fulfill a need and provide satisfaction compared to other products under different brand names. The pricing of products takes perceived value into consideration,. In some cases, the price of a product or service may have more to do with its emotional appeal than with the actual cost of production.

### **4.3 CANVAS**

Appendix A

### **4.4 QFD**

Appendix B

### **4.5 Functional Requirements**



## Chapter 5

# Case Study

### 5.1 Datasets

#### 5.1.1 Datasets by sector

These sectors were chosen based of their prevalence on the market and due to their innate differences in price volatility and strength.

##### 5.1.1.1 Energy

The energy market is very volatile since its value not only depends on their competitors but also on their primary resources needed to meet the demands, many companies seek to expand their market by providing various ways to supply energy for consumer use. As such, it's interesting to see if all these factors are significant enough to prove a challenge to predicting models to provide a value close to reality, Table 5.1.

Table 5.1: Various Stock Markets on the Energy Sector

| Symbol         | Name                                    | Country        | Currency | Exchange | Time Zone        |
|----------------|-----------------------------------------|----------------|----------|----------|------------------|
| <b>NNA</b>     | Navios Maritime Acquisition Corporation | Monaco         | USD      | NYQ      | America/New_York |
| <b>BNXR</b>    | Brinx Resources Ltd.                    | United States  | USD      | PNK      | America/New_York |
| <b>GKP.L</b>   | Gulf Keystone Petroleum Limited         | Bermuda        | GBp      | LSE      | Europe/London    |
| <b>CTP.AX</b>  | Central Petroleum Limited               | Australia      | AUD      | ASX      | Australia/Sydney |
| <b>BWO.OL</b>  | BW Offshore Limited                     | Bermuda        | NOK      | OSL      | Europe/Oslo      |
| <b>IPL.TO</b>  | Inter Pipeline Ltd.                     | Canada         | CAD      | TOR      | America/Toronto  |
| <b>GXE.TO</b>  | Gear Energy Ltd.                        | Canada         | CAD      | TOR      | America/Toronto  |
| <b>GEI.TO</b>  | Gibson Energy Inc.                      | Canada         | CAD      | TOR      | America/Toronto  |
| <b>ESN.TO</b>  | Essential Energy Services Ltd.          | Canada         | CAD      | TOR      | America/Toronto  |
| <b>BLY.AX</b>  | Boart Longyear Limited                  | United States  | AUD      | ASX      | Australia/Sydney |
| <b>POU.TO</b>  | Paramount Resources Ltd.                | Canada         | CAD      | TOR      | America/Toronto  |
| <b>VET.TO</b>  | Vermilion Energy Inc.                   | Canada         | CAD      | TOR      | America/Toronto  |
| <b>AM</b>      | Antero Midstream Corporation            | United States  | USD      | NYQ      | America/New_York |
| <b>AKSO.OL</b> | Aker Solutions ASA                      | Norway         | NOK      | OSL      | Europe/Oslo      |
| <b>AKER.OL</b> | Aker ASA                                | Norway         | NOK      | OSL      | Europe/Oslo      |
| <b>REX</b>     | REX American Resources Corporation      | United States  | USD      | NYQ      | America/New_York |
| <b>LNEGY</b>   | Lundin Energy AB (publ)                 | Sweden         | USD      | PNK      | America/New_York |
| <b>IOX.OL</b>  | Interoil Exploration and Production ASA | Norway         | NOK      | OSL      | Europe/Oslo      |
| <b>KEL.TO</b>  | Kelt Exploration Ltd.                   | Canada         | CAD      | TOR      | America/Toronto  |
| <b>BP.L</b>    | BP p.l.c.                               | United Kingdom | GBp      | LSE      | Europe/London    |
| <b>AR</b>      | Antero Resources Corporation            | United States  | USD      | NYQ      | America/New_York |
| <b>HE.TO</b>   | Hanwei Energy Services Corp.            | Canada         | CAD      | TOR      | America/Toronto  |
| <b>SCL.TO</b>  | Shawcor Ltd.                            | Canada         | CAD      | TOR      | America/Toronto  |
| <b>APA</b>     | APA Corporation                         | United States  | USD      | NMS      | America/New_York |
| <b>BPT</b>     | BP Prudhoe Bay Royalty Trust            | United States  | USD      | NYQ      | America/New_York |
| <b>MMR.AX</b>  | MEC Resources Limited                   | Australia      | AUD      | ASX      | Australia/Sydney |
| <b>BBLS</b>    | Petrolia Energy Corporation             | United States  | USD      | PNK      | America/New_York |
| <b>OXY</b>     | Occidental Petroleum Corporation        | United States  | USD      | NYQ      | America/New_York |
| <b>ORG.AX</b>  | Origin Energy Limited                   | Australia      | AUD      | ASX      | Australia/Sydney |
| <b>HWO.TO</b>  | High Arctic Energy Services Inc         | Canada         | CAD      | TOR      | America/Toronto  |

### 5.1.1.2 Healthcare

Healthcare is characterized by the ability to prevent, diagnose, and treat physical and mental illnesses. As such, the market fluctuates when these services are required. It's a relative stable market but events like a global health crisis is sure to impact the status quo. By its stable nature, this sector was chosen to provide a bigger variety of cases to the training models, Table 5.2.

Table 5.2: Various Stock Markets on the Healthcare Sector

| Symbol | Name                                       | Country        | Currency | Exchange | Time Zone        |
|--------|--------------------------------------------|----------------|----------|----------|------------------|
| BRKR   | Brucker Corporation                        | United States  | USD      | NMS      | America/New_York |
| CI     | Cigna Corporation                          | United States  | USD      | NYQ      | America/New_York |
| BMJ    | Bristol-Myers Squibb Company               | United States  | USD      | NYQ      | America/New_York |
| ASNB   | EKIMAS Corporation                         | United States  | USD      | PNK      | America/New_York |
| IMV.TO | IMV Inc.                                   | Canada         | CAD      | TOR      | America/Toronto  |
| ABBV   | AbbVie Inc.                                | United States  | USD      | NYQ      | America/New_York |
| IPD.AX | ImpediMed Limited                          | Australia      | AUD      | ASX      | Australia/Sydney |
| BMRN   | BioMarin Pharmaceutical Inc.               | United States  | USD      | NMS      | America/New_York |
| AGEN   | Agenus Inc.                                | United States  | USD      | NCM      | America/New_York |
| ALKS   | Alkermes plc                               | Ireland        | USD      | NMS      | America/New_York |
| ADPT   | Adaptive Biotechnologies Corporation       | United States  | USD      | NMS      | America/New_York |
| RDHL   | RedHill Biopharma Ltd.                     | Israel         | USD      | NMS      | America/New_York |
| NVS    | Novartis AG                                | Switzerland    | USD      | NYQ      | America/New_York |
| UBQU   | Ubiquitech Software Corp.                  | United States  | USD      | PNK      | America/New_York |
| TRIB   | Trinity Biotech plc                        | Ireland        | USD      | NMS      | America/New_York |
| MCK    | McKesson Corporation                       | United States  | USD      | NYQ      | America/New_York |
| EARS   | Auris Medical Holding Ltd.                 | Bermuda        | USD      | NMS      | America/New_York |
| GNGBY  | Getinge AB                                 | Sweden         | USD      | PNK      | America/New_York |
| GUD.TO | Knight Therapeutics Inc.                   | Canada         | CAD      | TOR      | America/Toronto  |
| MYHI   | Mountain High Acquisitions Corp.           | United States  | USD      | PNK      | America/New_York |
| AEZS   | Aeterna Zentaris Inc.                      | United States  | USD      | NCM      | America/New_York |
| PBH    | Prestige Consumer Healthcare Inc.          | United States  | USD      | NYQ      | America/New_York |
| ABC    | AmerisourceBergen Corporation              | United States  | USD      | NYQ      | America/New_York |
| ABC.L  | Abcam plc                                  | United Kingdom | GBp      | LSE      | Europe/London    |
| BNTC   | Benitec Biopharma Inc.                     | United States  | USD      | NMS      | America/New_York |
| LH     | Laboratory Corporation of America Holdings | United States  | USD      | NYQ      | America/New_York |
| AGTC   | Applied Genetic Technologies Corporation   | United States  | USD      | NMS      | America/New_York |
| ARRAY  | Accuray Incorporated                       | United States  | USD      | NMS      | America/New_York |
| EW     | Edwards Lifesciences Corporation           | United States  | USD      | NYQ      | America/New_York |
| BLRX   | BioLineRx Ltd.                             | Israel         | USD      | NCM      | America/New_York |

### 5.1.1.3 Technology

Technology encapsulates a wide range of classifications, mostly anything can be called technology or at least depend on it. It's a very volatile market because the sheer number of ways its value can be influenced. It's possible that the prediction models trained with these markets are susceptible to less reliable results, Table 5.3.

Table 5.3: Various Stock Markets on the Technology Sector

| Symbol    | Name                                           | Country        | Currency | Exchange | Time Zone        |
|-----------|------------------------------------------------|----------------|----------|----------|------------------|
| PACT.ST   | Proact IT Group AB (publ)                      | Sweden         | SEK      | STO      | Europe/Stockholm |
| NCR       | NCR Corporation                                | United States  | USD      | NYQ      | America/New_York |
| BCOV      | Brightcove Inc.                                | United States  | USD      | NMS      | America/New_York |
| TEAM      | Atlassian Corporation Plc                      | Australia      | USD      | NMS      | America/New_York |
| IVX.DE    | InVision Aktiengesellschaft                    | Germany        | EUR      | GER      | Europe/Berlin    |
| SAIC      | Science Applications International Corporation | United States  | USD      | NYQ      | America/New_York |
| APH       | Amphenol Corporation                           | United States  | USD      | NYQ      | America/New_York |
| EVS.BR    | EVS Broadcast Equipment SA                     | Belgium        | EUR      | BRU      | Europe/Brussels  |
| XRX       | Xerox Holdings Corporation                     | United States  | USD      | NYQ      | America/New_York |
| HMNY      | Helios and Matheson Analytics Inc.             | United States  | USD      | PNK      | America/New_York |
| ADBE      | Adobe Inc.                                     | United States  | USD      | NMS      | America/New_York |
| CHKP      | Check Point Software Technologies Ltd.         | Israel         | USD      | NMS      | America/New_York |
| GLW       | Coming Incorporated                            | United States  | USD      | NYQ      | America/New_York |
| ENA.V     | Enablence Technologies Inc.                    | Canada         | CAD      | VAN      | America/Toronto  |
| LRCX      | Lam Research Corporation                       | United States  | USD      | NMS      | America/New_York |
| BGO.L     | Bango plc                                      | United Kingdom | GBp      | LSE      | Europe/London    |
| ORCL      | Oracle Corporation                             | United States  | USD      | NYQ      | America/New_York |
| HXGBY     | Hexagon AB (publ)                              | Sweden         | USD      | PNK      | America/New_York |
| DDD       | 3D Systems Corporation                         | United States  | USD      | NYQ      | America/New_York |
| TFECY     | Thin Film Electronics ASA                      | Norway         | USD      | PNK      | America/New_York |
| CIEN      | Ciena Corporation                              | United States  | USD      | NYQ      | America/New_York |
| ACN       | Accenture plc                                  | Ireland        | USD      | NYQ      | America/New_York |
| MSFT      | Microsoft Corporation                          | United States  | USD      | NMS      | America/New_York |
| IAIC      | Information Analysis Incorporated              | United States  | USD      | PNK      | America/New_York |
| IBM       | International Business Machines Corporation    | United States  | USD      | NYQ      | America/New_York |
| MELE.BR   | Melexis NV                                     | Belgium        | EUR      | BRU      | Europe/Brussels  |
| BELFA     | Bel Fuse Inc.                                  | United States  | USD      | NMS      | America/New_York |
| XRF.AX    | XRF Scientific Limited                         | Australia      | AUD      | ASX      | Australia/Sydney |
| DYSL      | Dynasil Corporation of America                 | United States  | USD      | PNK      | America/New_York |
| HEXA-B.ST | Hexagon AB (publ)                              | Sweden         | SEK      | STO      | Europe/Stockholm |

## 5.1.2 Datasets by Country

These chosen criteria for these countries were the number of available information and the variety of sectors their markets participate on.

### 5.1.2.1 Australia

Australia has a very strong financial sector as it can influence banks worldwide and provide safe and profitable investments, meanwhile the cost of living is very high due to taxes and the

amount of common products it exports are not enough to make an impact on the economy, Table 5.4.

Table 5.4: Various Stock Markets in Australia

| Symbol        | Name                        | Country   | Currency | Exchange | Time Zone        |
|---------------|-----------------------------|-----------|----------|----------|------------------|
| <b>BKW.AX</b> | Brickworks Limited          | Australia | AUD      | ASX      | Australia/Sydney |
| <b>JRV.AX</b> | Jervois Mining Limited      | Australia | AUD      | ASX      | Australia/Sydney |
| <b>CDD.AX</b> | Cardno Limited              | Australia | AUD      | ASX      | Australia/Sydney |
| <b>EME.AX</b> | Energy Metals Limited       | Australia | AUD      | ASX      | Australia/Sydney |
| <b>KOV.AX</b> | Korvest Ltd                 | Australia | AUD      | ASX      | Australia/Sydney |
| <b>MOC.AX</b> | Mortgage Choice Limited     | Australia | AUD      | ASX      | Australia/Sydney |
| <b>CTP.AX</b> | Central Petroleum Limited   | Australia | AUD      | ASX      | Australia/Sydney |
| <b>BRK.AX</b> | Brookside Energy Limited    | Australia | AUD      | ASX      | Australia/Sydney |
| <b>NHF.AX</b> | nib holdings limited        | Australia | AUD      | ASX      | Australia/Sydney |
| <b>JYC.AX</b> | Joyce Corporation Ltd       | Australia | AUD      | ASX      | Australia/Sydney |
| <b>TLS.AX</b> | Telstra Corporation Limited | Australia | AUD      | ASX      | Australia/Sydney |
| <b>WHC.AX</b> | Whitehaven Coal Limited     | Australia | AUD      | ASX      | Australia/Sydney |
| <b>IMU.AX</b> | Imugene Limited             | Australia | AUD      | ASX      | Australia/Sydney |
| <b>BMH.AX</b> | Baumart Holdings Limited    | Australia | AUD      | ASX      | Australia/Sydney |
| <b>BOL.AX</b> | Boom Logistics Limited      | Australia | AUD      | ASX      | Australia/Sydney |

### 5.1.2.2 Germany

Germany relies heavily on education, being able to provide a competent workforce capable of increasing a company's reputation and profit if applied well, but this gives way to their major economic weakness, the labor cost is very high probably due to high demands to the workforce, because of this many startups or small companies are unable to keep functioning, Table 5.5.

Table 5.5: Various Stock Markets in Germany

| Symbol  | Name                                                          | Country | Currency | Exchange | Time Zone     |
|---------|---------------------------------------------------------------|---------|----------|----------|---------------|
| DB1.DE  | Deutsche Brse AG                                              | Germany | EUR      | GER      | Europe/Berlin |
| KSB.DE  | KSB SE & Co. KGaA                                             | Germany | EUR      | GER      | Europe/Berlin |
| AAD.DE  | Amadeus FiRe AG                                               | Germany | EUR      | GER      | Europe/Berlin |
| K1R.DE  | KROMI Logistik AG                                             | Germany | EUR      | GER      | Europe/Berlin |
| DAI.DE  | Daimler AG                                                    | Germany | EUR      | GER      | Europe/Berlin |
| BMW.DE  | Bayerische Motoren Werke Aktiengesellschaft                   | Germany | EUR      | GER      | Europe/Berlin |
| BVB.DE  | Borussia Dortmund GmbH & Co. Kommanditgesellschaft auf Aktien | Germany | EUR      | GER      | Europe/Berlin |
| BAS.DE  | BASF SE                                                       | Germany | EUR      | GER      | Europe/Berlin |
| DBK.DE  | Deutsche Bank Aktiengesellschaft                              | Germany | EUR      | GER      | Europe/Berlin |
| BEI.DE  | Beiersdorf Aktiengesellschaft                                 | Germany | EUR      | GER      | Europe/Berlin |
| ADS.DE  | adidas AG                                                     | Germany | EUR      | GER      | Europe/Berlin |
| NDX1.DE | Nordex SE                                                     | Germany | EUR      | GER      | Europe/Berlin |
| BAYN.DE | Bayer Aktiengesellschaft                                      | Germany | EUR      | GER      | Europe/Berlin |
| IVX.DE  | InVision Aktiengesellschaft                                   | Germany | EUR      | GER      | Europe/Berlin |
| G1A.DE  | GEA Group Aktiengesellschaft                                  | Germany | EUR      | GER      | Europe/Berlin |

### 5.1.2.3 United States of America

The USA is a major economic superpower, due to this, it invites many influential people to start their companies there, providing a wide diversity of products that paired with people's higher buying power creates a capitalist ecosystem, that leads to a bigger inequality of wealth distribution that impairs the economic growth of the individual, Table 5.6.

Table 5.6: Various Stock Markets in the USA

| Symbol | Name                                                           | Country       | Currency | Exchange | Time Zone        |
|--------|----------------------------------------------------------------|---------------|----------|----------|------------------|
| BWEL   | JG Boswell Company                                             | United States | USD      | PNK      | America/New_York |
| VLLX   | Valley Republic Bancorp                                        | United States | USD      | PNK      | America/New_York |
| FITBI  | Fifth Third Bancorp                                            | United States | USD      | NMS      | America/New_York |
| GPACU  | Global Partner Acquisition Corp II                             | United States | USD      | NMS      | America/New_York |
| BOKF   | BOK Financial Corporation                                      | United States | USD      | NMS      | America/New_York |
| NEN    | New England Realty Associates Limited Partnership              | United States | USD      | ASE      | America/New_York |
| FGB    | First Trust Specialty Finance and Financial Opportunities Fund | United States | USD      | NYQ      | America/New_York |
| BXS    | BancorpSouth Bank                                              | United States | USD      | NYQ      | America/New_York |
| AIMC   | Altra Industrial Motion Corp.                                  | United States | USD      | NMS      | America/New_York |
| GAHC   | Global Arena Holding Inc.                                      | United States | USD      | PNK      | America/New_York |
| CCD    | Calamos Dynamic Convertible and Income Fund                    | United States | USD      | NMS      | America/New_York |
| BNXR   | Brinx Resources Ltd.                                           | United States | USD      | PNK      | America/New_York |
| BDC    | Belden Inc.                                                    | United States | USD      | NYQ      | America/New_York |
| HRL    | Hormel Foods Corporation                                       | United States | USD      | NYQ      | America/New_York |
| CI     | Cigna Corporation                                              | United States | USD      | NYQ      | America/New_York |

## 5.2 Prediction Models

For the basis of training prediction models using the proposed datasets, the following algorithms were considered.

### 5.2.1 ARIMA

#### 5.2.1.1 Model Parameters

- Auto Regression

To determine the best value for this parameter, a Partial auto correlation plot (PACF) is used. Partial correlation is the correlation between the series and its lag, PACF conveys the pure correlation between the lag and the series allowing us to know what lag is needed (Tinungki 2019).

Any auto correlation in a stationarized series can be rectified by adding enough terms. The Auto Regression parameter has to be equal to as many lags that crosses the significance limit in the PACF plot, as demonstrated in Figure 5.1

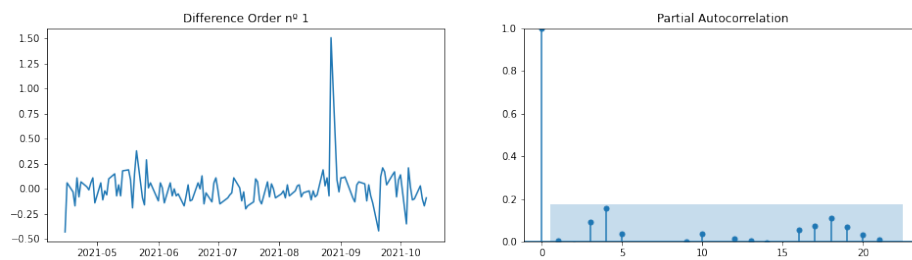


Figure 5.1: Example of a Partial Correlation Plot

- Integrated

To make use of the Integrated part of the algorithm, the data has to be stationary, meaning that the mean is consistent throughout the entire time series, using the Augmented Dickey-Fuller Test (Mushtaq 2011) it's possible to check this fact. This parameter represents the number of times the data has to be differenced to become stationary.

- Moving Average

It refers to the number of lagged forecast errors that should be considered in the model. For this a Auto correlation function (ACF) is used, and by its plot, Figure 5.2, we can determine the suitable value that should be the first closest to the significance line.

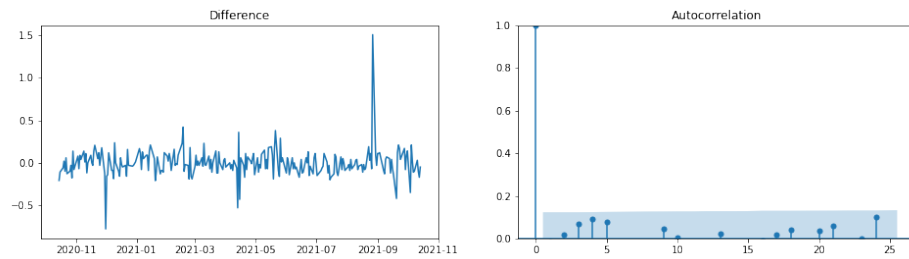


Figure 5.2: Example of a Auto Correlation Plot

- Historical Data

The amount of historical data was considered by taking both the p-value from the Augmented Dickey-Fuller Test and the number of optimal lags.

## 5.2.2 Random Forest

### 5.2.2.1 Model Parameters

- Max Features

To determine the ideal value, a correlation plot was performed. From that plot we can see that the columns "Open", "High", "Low", "Close" have really high correlation between them, this means that one value from these columns can confidently predict the others, for that reason, only one of these will be considered, shown correlation in Table 5.7.

Table 5.7: Correlation plot of dataset features

|       | Open | High | Low  | Close |
|-------|------|------|------|-------|
| Open  | 1.00 | 0.99 | 0.99 | 0.98  |
| High  | 0.99 | 1.00 | 0.98 | 0.99  |
| Low   | 0.99 | 0.98 | 1.00 | 0.99  |
| Close | 0.98 | 0.99 | 0.99 | 1.00  |

- N Estimators

This value's is very hard to optimize, since its contribution to the overall model's accuracy depends heavily on the dataset data and distribution.

Table 5.8: Best Errors for each iteration and historical data

|                      | 1 Month Error | 3 Month Error | 6 Month Error | 9 Month Error | 1 Year Error |
|----------------------|---------------|---------------|---------------|---------------|--------------|
| <b>N Estimations</b> |               |               |               |               |              |
| <b>1</b>             | 0.145373      | 0.260745      | 0.190923      | 0.228616      | 0.158123     |
| <b>2</b>             | 0.125083      | 0.269791      | 0.202969      | 0.211506      | 0.139128     |
| <b>5</b>             | 0.087662      | 0.181645      | 0.199809      | 0.215061      | 0.134321     |
| <b>10</b>            | 0.111728      | 0.175736      | 0.201663      | 0.203777      | 0.132030     |
| <b>25</b>            | 0.097876      | 0.183613      | 0.194636      | 0.190529      | 0.129503     |
| <b>50</b>            | 0.094335      | 0.167231      | 0.191903      | 0.191238      | 0.128211     |
| <b>100</b>           | 0.084560      | 0.176935      | 0.190095      | 0.197748      | 0.128225     |
| <b>250</b>           | 0.088225      | 0.179066      | 0.192652      | 0.194480      | 0.127782     |
| <b>500</b>           | 0.092211      | 0.185791      | 0.191779      | 0.195992      | 0.127958     |
| <b>1000</b>          | 0.088595      | 0.187245      | 0.192062      | 0.195555      | 0.127256     |

For this reason, this value will be kept at 100 for the training process of all models.

- Historical Data

The amount of historical data to be processed will depend on the average baseline error of the dataset, the dataset with the least average baseline error will be used.

Table 5.9: Average baseline error for various historical data

|               | 1 Month  | 3 Months | 6 Months | 9 Months | 1 Year   |
|---------------|----------|----------|----------|----------|----------|
| <b>NNA</b>    | 0.063330 | 0.109410 | 0.107810 | 0.112500 | 0.084620 |
| <b>BNXR</b>   | 0.000000 | 0.001350 | 0.001660 | 0.000580 | 0.000660 |
| <b>GKP.L</b>  | 2.898900 | 4.048090 | 3.611380 | 3.491580 | 2.590840 |
| <b>CTP.AX</b> | 0.003330 | 0.002790 | 0.003500 | 0.003430 | 0.003370 |
| <b>BWO.OL</b> | 0.783330 | 0.767700 | 0.614360 | 0.709990 | 0.642670 |
| <b>IPL.TO</b> | 0.011670 | 0.065680 | 0.066790 | 0.090420 | 0.116870 |
| <b>GXE.TO</b> | 0.016670 | 0.022940 | 0.015310 | 0.013960 | 0.016830 |
| <b>GEI.TO</b> | 0.151480 | 0.158530 | 0.255630 | 0.294380 | 0.282070 |
| <b>ESN.TO</b> | 0.006670 | 0.008820 | 0.010940 | 0.009580 | 0.010320 |
| <b>BLY.AX</b> | 0.166000 | 0.205000 | 0.235940 | 0.304890 | 0.325400 |

## 5.2.3 Neural Network

### 5.2.3.1 Model Parameters

- Epochs

The number of epochs, refers to the number of cycles the model uses to train. Each cycle uses the previous ones to try to reduce the error of the next iteration, meaning that in theory the more cycles it does, the better the model becomes, although this is not the case, there's always a threshold where the error becomes worse, so knowing the ideal number of epochs to use beforehand is very hard. For this reason model training libraries developed a way to facilitate this, called yearly stopping that makes the training stop in the instant that the error worsens.

- Sliding Window

It refers to the number of consecutive days considered to predict each subsequent value. For example, the value at date  $t$  and the value at date  $t + 1$  is used to predict the value at date  $t + 2$ . The model can train by using the current date  $t$  and previous dates  $t - 1$  as input variables to predict  $t + 1$ . For the training of the model in this project, it was used a sliding window of 5 days.

## 5.3 Evaluation Methodology

### 5.3.1 Metrics

#### 5.3.1.1 Root-Mean-Square Error

The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed. The RMSD represents the square root of the second sample moment of the differences between predicted values and observed values or the quadratic mean of these differences. These deviations are called residuals when the calculations are performed over the data sample that was used for estimation and are called errors (or prediction errors) when computed out-of-sample. The RMSD serves to aggregate the magnitudes of the errors in predictions for various data points into a single measure of predictive power. RMSD is a measure of accuracy, to compare forecasting errors of different models for a particular dataset and not between datasets, as it is scale-dependent.[1]

RMSD is always non-negative, and a value of 0 (almost never achieved in practice) would indicate a perfect fit to the data. In general, a lower RMSD is better than a higher one. However, comparisons across different types of data would be invalid because the measure is dependent on the scale of the numbers used.

RMSD is the square root of the average of squared errors. The effect of each error on RMSD is proportional to the size of the squared error; thus larger errors have a disproportionately large effect on RMSD. Consequently, RMSD is sensitive to outliers.

$$RMSD(\hat{\theta}) = \sqrt{MSE(\hat{\theta})} = \sqrt{E((\hat{\theta} - \theta)^2)}$$

### 5.3.1.2 $R^2$ Score

The coefficient of determination, denoted  $R^2$  or  $r^2$  and pronounced "R squared", is the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

It is a statistic used in the context of statistical models whose main purpose is either the prediction of future outcomes or the testing of hypotheses, on the basis of other related information. It provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}$$

## 5.4 Results

For each Sector/Zone, the results of the training with each prediction method divided by individual, sector and global parameters. Best results for each iteration shown in green.

### 5.4.1 Energy

#### 5.4.1.1 ARIMA

Table 5.10: ARIMA error results for the energy sector

|                       | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|-----------------------|-------------|--------------------|--------------------|
| <b>Symbol</b>         |             |                    |                    |
| <b>GXE.TO</b>         | 0.129200    | 0.120400           | 0.120400           |
| <b>ESN.TO</b>         | 0.061700    | 0.045500           | 0.045500           |
| <b>CTP.AX</b>         | 0.010300    | 0.010800           | 0.010800           |
| <b>BLY.AX</b>         | 1.019800    | 1.018900           | 1.018900           |
| <b>GEI.TO</b>         | 0.169800    | 0.238800           | 0.238800           |
| <b>BNXR</b>           | 0.000400    | 0.000600           | 0.000600           |
| <b>BWO.OL</b>         | 0.637200    | 1.103500           | 1.103500           |
| <b>GKP.L</b>          | 5.530300    | 9.265500           | 9.265500           |
| <b>IPL.TO</b>         | 0.176800    | 0.013700           | 0.013700           |
| <b>NNA</b>            | 0.273200    | 0.143500           | 0.143500           |
| <b>Mean</b>           | 0.800900    | 1.196100           | 1.196100           |
| <b>Std. Deviation</b> | 1.605400    | 2.718300           | 2.718300           |
| <b>Best Counter</b>   | 5           | 5                  | 0                  |

From the Table 5.10, we can notice that the individual training method was much as effective as the training method for the whole sector. Although the error overall was better when training individually.

Table 5.11: ARIMA time results for the energy sector

|                | ARIMA Time     | ARIMA Sector Time | ARIMA Global Time |
|----------------|----------------|-------------------|-------------------|
| Symbol         |                |                   |                   |
| GXE.TO         | 0:00:00.019970 | 0:00:00.009949    | 0:00:00.009972    |
| ESN.TO         | 0:00:00.009243 | 0:00:00.009972    | 0:00:00.009973    |
| CTP.AX         | 0:00:00.009996 | 0:00:00.009973    | 0:00:00.009946    |
| BLY.AX         | 0:00:00.009656 | 0:00:00.007979    | 0:00:00.009973    |
| GEI.TO         | 0:00:00.008949 | 0:00:00.009336    | 0:00:00.008948    |
| BNXR           | 0:00:00.008976 | 0:00:00.004986    | 0:00:00.005983    |
| BWO.OL         | 0:00:00.009973 | 0:00:00.008999    | 0:00:00.010046    |
| GKP.L          | 0:00:00.010972 | 0:00:00.009003    | 0:00:00.010942    |
| IPL.TO         | 0:00:00.058395 | 0:00:00.009003    | 0:00:00.009973    |
| NNA            | 0:00:00.033909 | 0:00:00.015958    | 0:00:00.019948    |
| Mean           | 0.018000       | 0.009500          | 0.010600          |
| Std. Deviation | 0.016200       | 0.002700          | 0.003600          |

The Table 5.11 shows that the ARIMA model was faster to train when using the sector's parameters.

#### 5.4.1.2 Random Forest

Table 5.12: Random Forest error results for the energy sector

|                | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|----------------|---------------------|----------------------------|----------------------------|
| Symbol         |                     |                            |                            |
| CTP.AX         | 0.003900            | 0.004400                   | 0.005300                   |
| BWO.OL         | 0.673300            | 0.983600                   | 0.983600                   |
| BLY.AX         | 0.126400            | 0.163500                   | 0.126400                   |
| IPL.TO         | 0.033600            | 0.032800                   | 0.033600                   |
| NNA            | 0.127000            | 0.173100                   | 0.173100                   |
| ESN.TO         | 0.012900            | 0.016700                   | 0.016000                   |
| GKP.L          | 3.119200            | 8.078500                   | 8.078500                   |
| GXE.TO         | 0.033500            | 0.044600                   | 0.044300                   |
| GEI.TO         | 0.191100            | 0.257100                   | 0.191100                   |
| BNXR           | 0.000200            | 0.000200                   | 0.000200                   |
| Mean           | 0.432100            | 0.975400                   | 0.965200                   |
| Std. Deviation | 0.915800            | 2.384100                   | 2.387400                   |
| Best Counter   | 9                   | 1                          | 0                          |

In the Table 5.12, we can see that the model's training using individual parameters had better accuracy.

Table 5.13: Random Forest time results for the energy sector

| Symbol         | Random Forest Time | Random Forest Sector Time | Random Forest Global Time |
|----------------|--------------------|---------------------------|---------------------------|
| CTP.AX         | 0:00:00.128628     | 0:00:00.097738            | 0:00:00.095726            |
| BWO.OL         | 0:00:00.102725     | 0:00:00.098619            | 0:00:00.097738            |
| BLY.AX         | 0:00:00.094218     | 0:00:00.110704            | 0:00:00.090749            |
| IPL.TO         | 0:00:00.092957     | 0:00:00.102726            | 0:00:00.094736            |
| NNA            | 0:00:00.102725     | 0:00:00.106222            | 0:00:00.102754            |
| ESN.TO         | 0:00:00.126452     | 0:00:00.093750            | 0:00:00.091724            |
| GKP.L          | 0:00:00.101732     | 0:00:00.095744            | 0:00:00.095743            |
| GXE.TO         | 0:00:00.096805     | 0:00:00.093749            | 0:00:00.091370            |
| GEI.TO         | 0:00:00.104721     | 0:00:00.091755            | 0:00:00.091758            |
| BNXR           | 0:00:00.094746     | 0:00:00.102628            | 0:00:00.091754            |
| Mean           | 0.104600           | 0.099400                  | 0.094400                  |
| Std. Deviation | 0.012800           | 0.006100                  | 0.003800                  |

Table 5.13 demonstrates, as a contrast to the Table 5.12, the time was less when the accuracy was also lower.

### 5.4.1.3 Neural Network

Table 5.14: Neural network error results for the energy sector

|                       | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|-----------------------|----------------------|-----------------------------|-----------------------------|
| <b>Symbol</b>         |                      |                             |                             |
| <b>BWO.OL</b>         | 0.080400             | 0.078500                    | 0.080800                    |
| <b>BLY.AX</b>         | 0.129400             | 0.067800                    | 0.072600                    |
| <b>GEI.TO</b>         | 0.070700             | 0.070000                    | 0.069800                    |
| <b>GKP.L</b>          | 0.119100             | 0.118300                    | 0.118900                    |
| <b>ESN.TO</b>         | 0.127900             | 0.122900                    | 0.129100                    |
| <b>IPL.TO</b>         | 0.109100             | 0.108700                    | 0.109200                    |
| <b>GXE.TO</b>         | 0.119800             | 0.117600                    | 0.121600                    |
| <b>NNA</b>            | 0.233700             | 0.104400                    | 0.127400                    |
| <b>BNXR</b>           | 0.254300             | 0.177300                    | 0.177600                    |
| <b>CTP.AX</b>         | 0.116100             | 0.100000                    | 0.102000                    |
| <b>Mean</b>           | 0.136000             | 0.106600                    | 0.110900                    |
| <b>Std. Deviation</b> | 0.057200             | 0.030300                    | 0.030600                    |
| <b>Best Counter</b>   | 0                    | 9                           | 1                           |

Neural Network had overwhelmingly more accuracy when training with the whole sector, as displayed in the Table 5.14

Table 5.15: Neural network time results for the energy sector

|                       | Neural Network Time | Neural Network Sector Time | Neural Network Global Time |
|-----------------------|---------------------|----------------------------|----------------------------|
| <b>Symbol</b>         |                     |                            |                            |
| <b>BWO.OL</b>         | 0:00:05.792047      | 0:00:01.086431             | 0:00:06.562652             |
| <b>BLY.AX</b>         | 0:00:05.141870      | 0:00:02.157809             | 0:00:06.328272             |
| <b>GEI.TO</b>         | 0:00:05.793229      | 0:00:02.136861             | 0:00:06.222407             |
| <b>GKP.L</b>          | 0:00:05.576512      | 0:00:00.563493             | 0:00:05.533840             |
| <b>ESN.TO</b>         | 0:00:06.775461      | 0:00:02.174112             | 0:00:05.493108             |
| <b>IPL.TO</b>         | 0:00:05.845917      | 0:00:00.922707             | 0:00:05.712312             |
| <b>GXE.TO</b>         | 0:00:06.770287      | 0:00:01.563442             | 0:00:06.779418             |
| <b>NNA</b>            | 0:00:05.237811      | 0:00:00.703120             | 0:00:05.155091             |
| <b>BNXR</b>           | 0:00:04.279974      | 0:00:00.589417             | 0:00:05.403674             |
| <b>CTP.AX</b>         | 0:00:06.073240      | 0:00:01.380576             | 0:00:06.291436             |
| <b>Mean</b>           | 5.728600            | 1.327800                   | 5.948200                   |
| <b>Std. Deviation</b> | 0.747500            | 0.654900                   | 0.554400                   |

Like Table 5.14, Table 5.15 demonstrates that the model was better when using the sector as a training basis.

## 5.4.2 Healthcare

### 5.4.2.1 ARIMA

Table 5.16: ARIMA error results for the healthcare sector

|                | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|----------------|-------------|--------------------|--------------------|
| Symbol         |             |                    |                    |
| ALKS           | 1.125700    | 1.024600           | 1.038900           |
| BRKR           | 1.008200    | 10.897200          | 1.927300           |
| IPD.AX         | 0.009400    | 0.009700           | 0.005600           |
| BMRN           | 0.853500    | 3.997600           | 0.848500           |
| CI             | 2.368600    | 11.912600          | 1.828700           |
| BMV            | 0.895900    | 3.538200           | 1.152700           |
| IMV.TO         | 0.115400    | 0.091400           | 0.153900           |
| AGEN           | 0.449200    | 0.884600           | 0.143700           |
| ASNB           | 0.007100    | 0.022100           | 0.016700           |
| ABBV           | 2.239700    | 1.425200           | 1.938700           |
| Mean           | 0.907300    | 3.380300           | 0.905500           |
| Std. Deviation | 0.801600    | 4.228100           | 0.761000           |
| Best Counter   | 3           | 3                  | 4                  |

In this Sector, the error results were overall balanced, refer to Table 5.16.

Table 5.17: ARIMA time results for the healthcare sector

|                | ARIMA Time     | ARIMA Sector Time | ARIMA Global Time |
|----------------|----------------|-------------------|-------------------|
| Symbol         |                |                   |                   |
| ALKS           | 0:00:00.069813 | 0:00:00.027925    | 0:00:00.010970    |
| BRKR           | 0:00:00.014960 | 0:00:00.026928    | 0:00:00.009934    |
| IPD.AX         | 0:00:00.011968 | 0:00:00.033910    | 0:00:00.008945    |
| BMRN           | 0:00:00.009973 | 0:00:00.027924    | 0:00:00.010971    |
| CI             | 0:00:00.010971 | 0:00:00.027953    | 0:00:00.010970    |
| BMV            | 0:00:00.028923 | 0:00:00.026930    | 0:00:00.009064    |
| IMV.TO         | 0:00:00.009973 | 0:00:00.025932    | 0:00:00.010972    |
| AGEN           | 0:00:00.008976 | 0:00:00.028497    | 0:00:00.009004    |
| ASNB           | 0:00:00.019947 | 0:00:00.021914    | 0:00:00.008969    |
| ABV            | 0:00:00.009973 | 0:00:00.025930    | 0:00:00.011001    |
| Mean           | 0.019500       | 0.027400          | 0.010100          |
| Std. Deviation | 0.018700       | 0.003000          | 0.001000          |

The time results are distributed between the individual training and the global training, as described in Table 5.17.

#### 5.4.2.2 Random Forest

Table 5.18: Random Forest error results for the healthcare sector

|                | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|----------------|---------------------|----------------------------|----------------------------|
| Symbol         |                     |                            |                            |
| BRKR           | 1.161100            | 1.832800                   | 2.243200                   |
| BMRN           | 0.633900            | 1.173500                   | 0.633900                   |
| CI             | 2.812500            | 3.877600                   | 2.806900                   |
| ASNB           | 0.004900            | 0.012400                   | 0.012200                   |
| AGEN           | 0.178800            | 0.277200                   | 0.306500                   |
| BMV            | 0.744100            | 0.677200                   | 0.744100                   |
| IPD.AX         | 0.003700            | 0.006600                   | 0.004200                   |
| ABV            | 0.828300            | 1.172700                   | 0.823800                   |
| ALKS           | 0.426300            | 0.570200                   | 0.737300                   |
| IMV.TO         | 0.104900            | 0.074800                   | 0.085400                   |
| Mean           | 0.689900            | 0.967500                   | 0.839700                   |
| Std. Deviation | 0.797900            | 1.124900                   | 0.902100                   |
| Best Counter   | 6                   | 2                          | 2                          |

In the Table 5.18, we can see the model had more accuracy using individual training method.

Table 5.19: Random Forest time results for the healthcare sector

|                | Random Forest Time | Random Forest Sector Time | Random Forest Global Time |
|----------------|--------------------|---------------------------|---------------------------|
| Symbol         |                    |                           |                           |
| BRKR           | 0:00:00.112698     | 0:00:00.406946            | 0:00:00.091726            |
| BMRN           | 0:00:00.105256     | 0:00:00.096741            | 0:00:00.093260            |
| CI             | 0:00:00.106714     | 0:00:00.261515            | 0:00:00.090785            |
| ASNB           | 0:00:00.106714     | 0:00:00.149892            | 0:00:00.090730            |
| AGEN           | 0:00:00.117686     | 0:00:00.092752            | 0:00:00.098734            |
| BMV            | 0:00:00.115691     | 0:00:00.117910            | 0:00:00.091727            |
| IPD.AX         | 0:00:00.139598     | 0:00:00.092763            | 0:00:00.090796            |
| ABV            | 0:00:00.096673     | 0:00:00.101728            | 0:00:00.093321            |
| ALKS           | 0:00:00.116198     | 0:00:00.090757            | 0:00:00.090787            |
| IMV.TO         | 0:00:00.100753     | 0:00:00.114684            | 0:00:00.089763            |
| Mean           | 0.111800           | 0.152600                  | 0.092200                  |
| Std. Deviation | 0.012000           | 0.103200                  | 0.002600                  |

The Table 5.19, demonstrates that the model took less time using global parameters.

### 5.4.2.3 Neural Network

Table 5.20: Neural network error results for the healthcare sector

|                | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|----------------|----------------------|-----------------------------|-----------------------------|
| Symbol         |                      |                             |                             |
| ALKS           | 0.111400             | 0.112300                    | 0.113000                    |
| BRKR           | 0.087100             | 0.086000                    | 0.086500                    |
| IMV.TO         | 0.108200             | 0.104100                    | 0.104600                    |
| ASNB           | 0.211800             | 0.218600                    | 0.218700                    |
| BMRN           | 0.125100             | 0.127700                    | 0.128400                    |
| AGEN           | 0.115500             | 0.089000                    | 0.088500                    |
| BMV            | 0.198500             | 0.103400                    | 0.107200                    |
| CI             | 0.079400             | 0.079900                    | 0.079600                    |
| IPD.AX         | 0.118100             | 0.119400                    | 0.119400                    |
| ABV            | 0.119400             | 0.124700                    | 0.095700                    |
| Mean           | 0.127400             | 0.116500                    | 0.114200                    |
| Std. Deviation | 0.041300             | 0.037400                    | 0.037800                    |
| Best Counter   | 5                    | 3                           | 2                           |

In the Table 5.20, the results are balanced.

Table 5.21: Neural network time results for the healthcare sector

| Symbol         | Neural Network Time | Neural Network Sector Time | Neural Network Global Time |
|----------------|---------------------|----------------------------|----------------------------|
| ALKS           | 0:00:06.230454      | 0:00:05.520481             | 0:00:05.661638             |
| BRKR           | 0:00:06.510451      | 0:00:05.977857             | 0:00:05.982889             |
| IMV.TO         | 0:00:05.745659      | 0:00:06.558668             | 0:00:05.743273             |
| ASNB           | 0:00:06.055534      | 0:00:06.050432             | 0:00:06.056470             |
| BMRN           | 0:00:06.578597      | 0:00:06.412959             | 0:00:06.547333             |
| AGEN           | 0:00:05.999888      | 0:00:06.146710             | 0:00:06.486980             |
| BMJ            | 0:00:05.119887      | 0:00:05.572695             | 0:00:06.602493             |
| CI             | 0:00:06.365934      | 0:00:06.412767             | 0:00:06.454945             |
| IPD.AX         | 0:00:06.738219      | 0:00:06.440378             | 0:00:07.367847             |
| ABBV           | 0:00:05.099892      | 0:00:05.204571             | 0:00:06.427923             |
| Mean           | 6.044500            | 6.029800                   | 6.333200                   |
| Std. Deviation | 0.573500            | 0.460400                   | 0.498300                   |

The Table 5.21, shows that the best times were between using individual and global parameters.

### 5.4.3 Technology

#### 5.4.3.1 ARIMA

Table 5.22: ARIMA error results for the technology sector

|                | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|----------------|-------------|--------------------|--------------------|
| Symbol         |             |                    |                    |
| HMNY           | 0.000700    | 0.000600           | 0.000500           |
| BCOV           | 0.936600    | 0.618800           | 0.545100           |
| EV.S.BR        | 0.506200    | 0.586500           | 0.542200           |
| PACT.ST        | 2.789400    | 5.126300           | 2.385500           |
| TEAM           | 10.684800   | 18.082500          | 11.066600          |
| IVX.DE         | 0.858900    | 3.450600           | 0.941100           |
| XRX            | 1.261000    | 0.543300           | 0.926700           |
| NCR            | 3.728600    | 1.955700           | 2.697400           |
| APH            | 1.909900    | 1.129700           | 1.844100           |
| SAIC           | 2.138300    | 1.528300           | 1.458600           |
| Mean           | 2.481400    | 3.302200           | 2.240800           |
| Std. Deviation | 2.931800    | 5.143700           | 3.051000           |
| Best Counter   | 3           | 3                  | 4                  |

The Table 5.22 displays the error results as balanced.

Table 5.23: ARIMA time results for the technology sector

|                | ARIMA Time     | ARIMA Sector Time | ARIMA Global Time |
|----------------|----------------|-------------------|-------------------|
| Symbol         |                |                   |                   |
| HMNY           | 0:00:00.009974 | 0:00:00.025958    | 0:00:00.009945    |
| BCOV           | 0:00:00.029920 | 0:00:00.026900    | 0:00:00.009973    |
| EV.S.BR        | 0:00:00.019947 | 0:00:00.026929    | 0:00:00.009539    |
| PACT.ST        | 0:00:00.009973 | 0:00:00.029919    | 0:00:00.009973    |
| TEAM           | 0:00:00.017952 | 0:00:00.034876    | 0:00:00.009943    |
| IVX.DE         | 0:00:00.025930 | 0:00:00.022955    | 0:00:00.009003    |
| XRX            | 0:00:00.011967 | 0:00:00.026928    | 0:00:00.008978    |
| NCR            | 0:00:00.010971 | 0:00:00.026955    | 0:00:00.009000    |
| APH            | 0:00:00.025931 | 0:00:00.025931    | 0:00:00.010004    |
| SAIC           | 0:00:00.011968 | 0:00:00.025902    | 0:00:00.009972    |
| Mean           | 0.017500       | 0.027300          | 0.009600          |
| Std. Deviation | 0.007600       | 0.003200          | 0.000500          |

The Table 5.23 shows that the model took less time training with global parameters.

### 5.4.3.2 Random Forest

Table 5.24: Random Forest error results for the technology sector

|                | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|----------------|---------------------|----------------------------|----------------------------|
| <b>Symbol</b>  |                     |                            |                            |
| BCOV           | 0.224200            | 0.226600                   | 0.325200                   |
| IVX.DE         | 0.427300            | 0.427300                   | 0.427300                   |
| EVS.BR         | 0.280000            | 0.357100                   | 0.357100                   |
| TEAM           | 5.415400            | 6.211400                   | 7.544400                   |
| APH            | 0.791200            | 1.001900                   | 1.024000                   |
| NCR            | 0.984200            | 0.802300                   | 1.034600                   |
| PACT.ST        | 2.364300            | 2.054900                   | 2.054900                   |
| SAIC           | 1.278800            | 1.444600                   | 0.717200                   |
| XRX            | 0.186700            | 0.193700                   | 0.202400                   |
| HMNY           | 0.000200            | 0.000200                   | 0.000200                   |
| Mean           | 1.195200            | 1.272000                   | 1.368700                   |
| Std. Deviation | 1.555900            | 1.753800                   | 2.133000                   |
| Best Counter   | 7                   | 2                          | 1                          |

In the Table 5.24, its presented that Random Forest had more accuracy using individual parameters.

Table 5.25: Random Forest time results for the technology sector

|                | Random Forest Time | Random Forest Sector Time | Random Forest Global Time |
|----------------|--------------------|---------------------------|---------------------------|
| <b>Symbol</b>  |                    |                           |                           |
| BCOV           | 0:00:00.100704     | 0:00:00.098735            | 0:00:00.089748            |
| IVX.DE         | 0:00:00.126662     | 0:00:00.135270            | 0:00:00.093748            |
| EVS.BR         | 0:00:00.143616     | 0:00:00.298202            | 0:00:00.095768            |
| TEAM           | 0:00:00.181484     | 0:00:00.097739            | 0:00:00.090757            |
| APH            | 0:00:00.105717     | 0:00:00.272778            | 0:00:00.092787            |
| NCR            | 0:00:00.103698     | 0:00:00.090758            | 0:00:02.583644            |
| PACT.ST        | 0:00:00.092776     | 0:00:00.103722            | 0:00:00.090757            |
| SAIC           | 0:00:00.118682     | 0:00:00.306583            | 0:00:00.100755            |
| XRX            | 0:00:00.106223     | 0:00:00.119811            | 0:00:00.092257            |
| HMNY           | 0:00:00.098736     | 0:00:00.134961            | 0:00:00.095310            |
| Mean           | 0.117800           | 0.165900                  | 0.342600                  |
| Std. Deviation | 0.027000           | 0.089100                  | 0.787400                  |

The Table 5.25 shows that the model took less time training with global parameters.

### 5.4.3.3 Neural Network

Table 5.26: Neural network error results for the technology sector

|                | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|----------------|----------------------|-----------------------------|-----------------------------|
| <b>Symbol</b>  |                      |                             |                             |
| BCOV           | 0.100300             | 0.076900                    | 0.076900                    |
| IVX.DE         | 0.135300             | 0.133200                    | 0.136200                    |
| TEAM           | 0.110200             | 0.118000                    | 0.108400                    |
| EVS.BR         | 0.099800             | 0.100500                    | 0.103900                    |
| HMNY           | 0.081200             | 0.080900                    | 0.080700                    |
| NCR            | 0.091000             | 0.090800                    | 0.090200                    |
| PACT.ST        | 0.084300             | 0.089500                    | 0.085900                    |
| SAIC           | 0.098100             | 0.097800                    | 0.096400                    |
| APH            | 0.112100             | 0.112300                    | 0.112100                    |
| XRX            | 0.082900             | 0.084100                    | 0.109200                    |
| Mean           | 0.099500             | 0.098400                    | 0.100000                    |
| Std. Deviation | 0.015700             | 0.017000                    | 0.016900                    |
| Best Counter   | 4                    | 2                           | 4                           |

The results portrayed in the Table 5.26 shows that the results were overall balanced.

Table 5.27: Neural network time results for the technology sector

|                | Neural Network Time | Neural Network Sector Time | Neural Network Global Time |
|----------------|---------------------|----------------------------|----------------------------|
| <b>Symbol</b>  |                     |                            |                            |
| BCOV           | 0:00:06.096583      | 0:00:06.607324             | 0:00:06.648563             |
| IVX.DE         | 0:00:06.119863      | 0:00:06.217402             | 0:00:05.648187             |
| TEAM           | 0:00:05.592235      | 0:00:06.615537             | 0:00:06.048221             |
| EVS.BR         | 0:00:06.063368      | 0:00:05.947455             | 0:00:06.561246             |
| HMNY           | 0:00:06.593446      | 0:00:06.913215             | 0:00:06.644848             |
| NCR            | 0:00:06.019696      | 0:00:05.857359             | 0:00:06.128235             |
| PACT.ST        | 0:00:06.728528      | 0:00:05.676050             | 0:00:06.484801             |
| SAIC           | 0:00:06.184771      | 0:00:06.644683             | 0:00:06.671658             |
| APH            | 0:00:07.183894      | 0:00:05.945534             | 0:00:05.600809             |
| XRX            | 0:00:06.660800      | 0:00:08.203829             | 0:00:04.914708             |
| Mean           | 6.324300            | 6.462800                   | 6.135100                   |
| Std. Deviation | 0.459200            | 0.737100                   | 0.590200                   |

The results portrayed in the Table 5.27 shows that the times were overall balanced.

#### 5.4.4 Australia

##### 5.4.4.1 ARIMA

Table 5.28: ARIMA error results for Australia

|                       | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|-----------------------|-------------|--------------------|--------------------|
| <b>Symbol</b>         |             |                    |                    |
| <b>CTP.AX</b>         | 0.010300    | 0.011300           | 0.010800           |
| <b>BKW.AX</b>         | 0.843500    | 0.878600           | 0.927700           |
| <b>NHF.AX</b>         | 0.195900    | 0.175800           | 0.217700           |
| <b>JRV.AX</b>         | 0.045800    | 0.047700           | 0.039700           |
| <b>CDD.AX</b>         | 0.156900    | 0.184300           | 0.177200           |
| <b>JYC.AX</b>         | 0.495800    | 0.301400           | 0.313600           |
| <b>TLS.AX</b>         | 0.069600    | 0.038600           | 0.030500           |
| <b>KOV.AX</b>         | 0.479700    | 0.533400           | 0.531600           |
| <b>EME.AX</b>         | 0.050300    | 0.016800           | 0.015100           |
| <b>BRK.AX</b>         | 0.002300    | 0.001800           | 0.001800           |
| <b>Mean</b>           | 0.235000    | 0.219000           | 0.226600           |
| <b>Std. Deviation</b> | 0.266000    | 0.271500           | 0.284600           |
| <b>Best Counter</b>   | 4           | 3                  | 3                  |

The results portrayed in the Table 5.28 shows that the results were overall balanced.

Table 5.29: ARIMA time results for Australia

|                | ARIMA Time     | ARIMA Sector Time | ARIMA Global Time |
|----------------|----------------|-------------------|-------------------|
| <b>Symbol</b>  |                |                   |                   |
| CTP.AX         | 0:00:00.008977 | 0:00:00.008975    | 0:00:00.009974    |
| BKW.AX         | 0:00:00.027952 | 0:00:00.011969    | 0:00:00.012965    |
| NHF.AX         | 0:00:00.008947 | 0:00:00.008975    | 0:00:00.010991    |
| JRV.AX         | 0:00:00.008948 | 0:00:00.009973    | 0:00:00.011967    |
| CDD.AX         | 0:00:00.008975 | 0:00:00.009973    | 0:00:00.009974    |
| JYC.AX         | 0:00:00.027950 | 0:00:00.011000    | 0:00:00.010999    |
| TLS.AX         | 0:00:00.008976 | 0:00:00.010001    | 0:00:00.010942    |
| KOV.AX         | 0:00:00.008976 | 0:00:00.008976    | 0:00:00.009974    |
| EME.AX         | 0:00:00.009973 | 0:00:00.009973    | 0:00:00.009973    |
| BRK.AX         | 0:00:00.020972 | 0:00:00.008976    | 0:00:00.009043    |
| Mean           | 0.014100       | 0.009900          | 0.010700          |
| Std. Deviation | 0.008200       | 0.001000          | 0.001100          |

The results portrayed in the Table 5.29 shows that the results were overall balanced.

#### 5.4.4.2 Random Forest

Table 5.30: Random Forest error results for Australia

|                | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|----------------|---------------------|----------------------------|----------------------------|
| <b>Symbol</b>  |                     |                            |                            |
| CTP.AX         | 0.004400            | 0.004400                   | 0.005300                   |
| BRK.AX         | 0.001700            | 0.001600                   | 0.001400                   |
| JYC.AX         | 0.067300            | 0.136900                   | 0.072100                   |
| JRV.AX         | 0.025700            | 0.025700                   | 0.018600                   |
| EME.AX         | 0.018000            | 0.046300                   | 0.041200                   |
| CDD.AX         | 0.030100            | 0.030100                   | 0.028500                   |
| BKW.AX         | 0.344500            | 0.610300                   | 0.392200                   |
| TLS.AX         | 0.031100            | 0.035300                   | 0.035900                   |
| NHF.AX         | 0.096700            | 0.111800                   | 0.086400                   |
| KOV.AX         | 0.074100            | 0.061500                   | 0.131600                   |
| Mean           | 0.069400            | 0.106400                   | 0.081300                   |
| Std. Deviation | 0.096300            | 0.173000                   | 0.110400                   |
| Best Counter   | 5                   | 1                          | 4                          |

In the Table 5.30 the results were balanced but tended towards the individual parameters.

Table 5.31: Random Forest time results for Australia

|                | Random Forest Time | Random Forest Sector Time | Random Forest Global Time |
|----------------|--------------------|---------------------------|---------------------------|
| <b>Symbol</b>  |                    |                           |                           |
| CTP.AX         | 0:00:00.103230     | 0:00:00.104232            | 0:00:00.089731            |
| BRK.AX         | 0:00:00.097738     | 0:00:00.097739            | 0:00:00.090781            |
| JYC.AX         | 0:00:00.107220     | 0:00:00.098245            | 0:00:00.093750            |
| JRV.AX         | 0:00:00.104720     | 0:00:00.131649            | 0:00:00.091070            |
| EME.AX         | 0:00:00.094086     | 0:00:00.110731            | 0:00:00.093749            |
| CDD.AX         | 0:00:00.091650     | 0:00:00.104719            | 0:00:00.088593            |
| BKW.AX         | 0:00:00.118021     | 0:00:00.186012            | 0:00:00.100271            |
| TLS.AX         | 0:00:00.111672     | 0:00:00.099732            | 0:00:00.092778            |
| NHF.AX         | 0:00:00.102753     | 0:00:00.094746            | 0:00:00.092747            |
| KOV.AX         | 0:00:00.121696     | 0:00:00.107750            | 0:00:00.091782            |
| Mean           | 0.105300           | 0.113600                  | 0.092500                  |
| Std. Deviation | 0.009700           | 0.027500                  | 0.003200                  |

The Table 5.31 shows favorable results when the model uses the global parameters.

#### 5.4.4.3 Neural Network

Table 5.32: Neural network error results for Australia

|                | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|----------------|----------------------|-----------------------------|-----------------------------|
| <b>Symbol</b>  |                      |                             |                             |
| KOV.AX         | 0.102900             | 0.104500                    | 0.103300                    |
| NHF.AX         | 0.093900             | 0.097200                    | 0.094400                    |
| TLS.AX         | 0.100700             | 0.100200                    | 0.101300                    |
| JRV.AX         | 0.102300             | 0.102100                    | 0.100800                    |
| EME.AX         | 0.117800             | 0.117100                    | 0.117100                    |
| JYC.AX         | 0.117400             | 0.121000                    | 0.115700                    |
| CDD.AX         | 0.117900             | 0.116300                    | 0.116700                    |
| BRK.AX         | 0.108400             | 0.106900                    | 0.107300                    |
| BKW.AX         | 0.091700             | 0.092500                    | 0.092000                    |
| CTP.AX         | 0.116500             | 0.102300                    | 0.104500                    |
| Mean           | 0.107000             | 0.106000                    | 0.105300                    |
| Std. Deviation | 0.009600             | 0.008800                    | 0.008500                    |
| Best Counter   | 3                    | 5                           | 2                           |

The results portrayed in the Table 5.32 shows that the results were overall balanced.

Table 5.33: Neural network time results for Australia

|                | Neural Network Time | Neural Network Sector Time | Neural Network Global Time |
|----------------|---------------------|----------------------------|----------------------------|
| Symbol         |                     |                            |                            |
| KOV.AX         | 0:00:05.339858      | 0:00:05.161321             | 0:00:06.958779             |
| NHF.AX         | 0:00:06.617261      | 0:00:06.610945             | 0:00:08.807752             |
| TLS.AX         | 0:00:05.823825      | 0:00:06.366531             | 0:00:06.962290             |
| JRV.AX         | 0:00:06.443756      | 0:00:06.322426             | 0:00:15.697090             |
| EME.AX         | 0:00:05.347409      | 0:00:05.575015             | 0:00:05.957717             |
| JYC.AX         | 0:00:06.653584      | 0:00:05.269551             | 0:00:08.789687             |
| CDD.AX         | 0:00:06.208772      | 0:00:06.453208             | 0:00:07.306893             |
| BRK.AX         | 0:00:06.884996      | 0:00:06.348552             | 0:00:08.574928             |
| BKW.AX         | 0:00:06.577633      | 0:00:05.817577             | 0:00:08.176215             |
| CTP.AX         | 0:00:05.432631      | 0:00:06.241473             | 0:00:06.377129             |
| Mean           | 6.133000            | 6.016700                   | 8.360800                   |
| Std. Deviation | 0.596600            | 0.521200                   | 2.769400                   |

The Table 5.33 show better results between individual and sector parameters.

## 5.4.5 Germany

### 5.4.5.1 ARIMA

Table 5.34: ARIMA error results for Germany

|                | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|----------------|-------------|--------------------|--------------------|
| Symbol         |             |                    |                    |
| AAD.DE         | 10.035500   | 4.132700           | 4.132700           |
| K1R.DE         | 0.377200    | 0.098700           | 0.098700           |
| BEI.DE         | 1.750100    | 1.418900           | 1.418900           |
| BAS.DE         | 1.812800    | 2.342100           | 2.342100           |
| KSB.DE         | 23.519100   | 25.805700          | 25.805700          |
| BVB.DE         | 0.051100    | 0.046900           | 0.046900           |
| DAI.DE         | 4.796700    | 4.871900           | 4.871900           |
| DB1.DE         | 2.713700    | 2.735900           | 2.735900           |
| DBK.DE         | 0.447900    | 0.400900           | 0.400900           |
| BMW.DE         | 3.263200    | 3.603000           | 3.603000           |
| Mean           | 4.876700    | 4.545700           | 4.545700           |
| Std. Deviation | 6.809900    | 7.268100           | 7.268100           |
| Best Counter   | 5           | 5                  | 0                  |

The Table 5.34 shows that the results were balanced.

Table 5.35: ARIMA time results for Germany

|                | ARIMA Time     | ARIMA Sector Time | ARIMA Global Time |
|----------------|----------------|-------------------|-------------------|
| <b>Symbol</b>  |                |                   |                   |
| AAD.DE         | 0:00:00.010970 | 0:00:00.010939    | 0:00:00.008976    |
| K1R.DE         | 0:00:00.016955 | 0:00:00.003989    | 0:00:00.004955    |
| BEI.DE         | 0:00:00.009973 | 0:00:00.011968    | 0:00:00.011969    |
| BAS.DE         | 0:00:00.029919 | 0:00:00.008976    | 0:00:00.008976    |
| KSB.DE         | 0:00:00.021942 | 0:00:00.010942    | 0:00:00.011968    |
| BVB.DE         | 0:00:00.018943 | 0:00:00.008976    | 0:00:00.009974    |
| DAI.DE         | 0:00:00.021941 | 0:00:00.009973    | 0:00:00.008976    |
| DB1.DE         | 0:00:00.013962 | 0:00:00.010000    | 0:00:00.008976    |
| DBK.DE         | 0:00:00.011990 | 0:00:00.010971    | 0:00:00.008976    |
| BMW.DE         | 0:00:00.024933 | 0:00:00.010971    | 0:00:00.009374    |
|                |                |                   |                   |
| Mean           | 0.018200       | 0.009800          | 0.009300          |
| Std. Deviation | 0.006600       | 0.002200          | 0.001900          |

The Table 5.39 show better results between the sector and global parameters.

#### 5.4.5.2 Random Forest

Table 5.36: Random Forest error results for Germany

|                | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|----------------|---------------------|----------------------------|----------------------------|
| <b>Symbol</b>  |                     |                            |                            |
| K1R.DE         | 0.082500            | 0.226400                   | 0.226400                   |
| KSB.DE         | 6.608200            | 14.234800                  | 14.234800                  |
| AAD.DE         | 3.698700            | 3.626600                   | 3.626600                   |
| DAI.DE         | 1.390800            | 1.640200                   | 1.640200                   |
| BMW.DE         | 1.499300            | 1.090100                   | 1.090100                   |
| BAS.DE         | 0.758600            | 1.541400                   | 1.541400                   |
| DB1.DE         | 1.445900            | 1.445900                   | 1.445900                   |
| DBK.DE         | 0.117600            | 0.117600                   | 0.117600                   |
| BVB.DE         | 0.212100            | 0.212100                   | 0.212100                   |
| BEI.DE         | 0.730500            | 0.973600                   | 0.973600                   |
|                |                     |                            |                            |
| Mean           | 1.654400            | 2.510900                   | 2.510900                   |
| Std. Deviation | 1.934600            | 4.026700                   | 4.026700                   |
| Best Counter   | 8                   | 2                          | 0                          |

The Table 5.36 demonstrates that the model had more accuracy when using individual parameters.

Table 5.37: Random Forest time results for Germany

|                | Random Forest Time | Random Forest Sector Time | Random Forest Global Time |
|----------------|--------------------|---------------------------|---------------------------|
| Symbol         |                    |                           |                           |
| K1R.DE         | 0:00:00.099735     | 0:00:00.094772            | 0:00:00.086795            |
| KSB.DE         | 0:00:00.101758     | 0:00:00.125813            | 0:00:00.089785            |
| AAD.DE         | 0:00:00.098735     | 0:00:00.096549            | 0:00:00.092727            |
| DAI.DE         | 0:00:00.100351     | 0:00:00.107409            | 0:00:00.092725            |
| BMW.DE         | 0:00:00.110330     | 0:00:00.119680            | 0:00:00.094746            |
| BAS.DE         | 0:00:00.105225     | 0:00:00.102725            | 0:00:00.093747            |
| DB1.DE         | 0:00:00.098735     | 0:00:00.134691            | 0:00:00.094105            |
| DBK.DE         | 0:00:00.092752     | 0:00:00.101752            | 0:00:00.089786            |
| BVB.DE         | 0:00:00.095433     | 0:00:00.116750            | 0:00:00.091753            |
| BEI.DE         | 0:00:00.094746     | 0:00:00.109422            | 0:00:00.091754            |
| Mean           | 0.099800           | 0.111000                  | 0.091800                  |
| Std. Deviation | 0.005200           | 0.013000                  | 0.002400                  |

Random Forest took less time training Germany's data, as shown in Table 5.37.

#### 5.4.5.3 Neural Network

Table 5.38: Neural network error results for Germany

|                | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|----------------|----------------------|-----------------------------|-----------------------------|
| Symbol         |                      |                             |                             |
| DAI.DE         | 0.111500             | 0.111300                    | 0.110500                    |
| BMW.DE         | 0.090800             | 0.090700                    | 0.092100                    |
| BAS.DE         | 0.084400             | 0.113500                    | 0.111600                    |
| BVB.DE         | 0.190400             | 0.088400                    | 0.088700                    |
| BEI.DE         | 0.073400             | 0.090300                    | 0.088900                    |
| AAD.DE         | 0.102400             | 0.101100                    | 0.101500                    |
| DBK.DE         | 0.109100             | 0.107500                    | 0.109400                    |
| KSB.DE         | 0.111300             | 0.110800                    | 0.112400                    |
| DB1.DE         | 0.102600             | 0.114100                    | 0.099900                    |
| K1R.DE         | 0.152200             | 0.153000                    | 0.152700                    |
| Mean           | 0.112800             | 0.108100                    | 0.106800                    |
| Std. Deviation | 0.032600             | 0.017800                    | 0.017700                    |
| Best Counter   | 3                    | 5                           | 2                           |

The Table 5.38 portrays the error results as balanced.

Table 5.39: Neural network time results for Germany

| Symbol         | Neural Network Time | Neural Network Sector Time | Neural Network Global Time |
|----------------|---------------------|----------------------------|----------------------------|
| DAI.DE         | 0:00:05.743255      | 0:00:05.979348             | 0:00:08.149559             |
| BMW.DE         | 0:00:06.752004      | 0:00:06.092361             | 0:00:09.719083             |
| BAS.DE         | 0:00:06.592633      | 0:00:04.926336             | 0:00:05.889122             |
| BVB.DE         | 0:00:04.958791      | 0:00:06.599797             | 0:00:09.575579             |
| BEI.DE         | 0:00:05.907320      | 0:00:06.295215             | 0:00:08.251275             |
| AAD.DE         | 0:00:06.507790      | 0:00:06.370868             | 0:00:09.005068             |
| DBK.DE         | 0:00:06.426664      | 0:00:06.420546             | 0:00:09.463248             |
| KSB.DE         | 0:00:07.073226      | 0:00:06.342592             | 0:00:09.436265             |
| DB1.DE         | 0:00:06.515009      | 0:00:06.719525             | 0:00:08.777530             |
| K1R.DE         | 0:00:05.600965      | 0:00:05.340583             | 0:00:06.562878             |
| Mean           | 6.207800            | 6.108700                   | 8.483000                   |
| Std. Deviation | 0.637700            | 0.565200                   | 1.313600                   |

In the Table 5.39, we can see that the model took less time training with individual and sector parameters.

## 5.4.6 United States of America

### 5.4.6.1 ARIMA

Table 5.40: ARIMA error results for the United States of America

|                       | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|-----------------------|-------------|--------------------|--------------------|
| <b>Symbol</b>         |             |                    |                    |
| <b>GPACU</b>          | 0.071300    | 0.038800           | 0.043300           |
| <b>BWEL</b>           | 8.249000    | 48.167500          | 8.071400           |
| <b>VLLX</b>           | 0.496100    | 2.758800           | 0.446900           |
| <b>BXS</b>            | 0.629500    | 1.044700           | 1.052400           |
| <b>FITBI</b>          | 0.758300    | 0.618100           | 0.728900           |
| <b>BOKF</b>           | 3.147800    | 5.670900           | 2.345700           |
| <b>AIMC</b>           | 1.303200    | 2.927300           | 1.508400           |
| <b>GAHC</b>           | 0.000900    | 0.000400           | 0.000700           |
| <b>FGB</b>            | 0.147400    | 0.062600           | 0.078300           |
| <b>NEN</b>            | 1.516200    | 0.896200           | 1.801700           |
| <b>Mean</b>           | 1.632000    | 6.218500           | 1.607800           |
| <b>Std. Deviation</b> | 2.378300    | 14.085600          | 2.284000           |
| <b>Best Counter</b>   | 2           | 5                  | 3                  |

The Table 5.40 presents the results as overall balanced.

Table 5.41: ARIMA time results for the United States of America

|                       | ARIMA Time     | ARIMA Sector Time | ARIMA Global Time |
|-----------------------|----------------|-------------------|-------------------|
| <b>Symbol</b>         |                |                   |                   |
| <b>GPACU</b>          | 0:00:00.010970 | 0:00:00.028922    | 0:00:00.010001    |
| <b>BWEL</b>           | 0:00:00.019947 | 0:00:00.026928    | 0:00:00.009973    |
| <b>VLLX</b>           | 0:00:00.005984 | 0:00:00.016955    | 0:00:00.006981    |
| <b>BXS</b>            | 0:00:00.009945 | 0:00:00.027925    | 0:00:00.009974    |
| <b>FITBI</b>          | 0:00:00.010970 | 0:00:00.028923    | 0:00:00.010970    |
| <b>BOKF</b>           | 0:00:00.018950 | 0:00:00.030918    | 0:00:00.009950    |
| <b>AIMC</b>           | 0:00:00.010971 | 0:00:00.026929    | 0:00:00.009973    |
| <b>GAHC</b>           | 0:00:00.010970 | 0:00:00.027925    | 0:00:00.009973    |
| <b>FGB</b>            | 0:00:00.009005 | 0:00:00.028923    | 0:00:00.009013    |
| <b>NEN</b>            | 0:00:00.017952 | 0:00:00.021942    | 0:00:00.008976    |
| <b>Mean</b>           | 0.012600       | 0.026600          | 0.009600          |
| <b>Std. Deviation</b> | 0.004700       | 0.004100          | 0.001100          |

The Table 5.41 shows better results when trained with global parameters.

### 5.4.6.2 Random Forest

Table 5.42: Random Forest error results for the United States of America

|                       | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|-----------------------|---------------------|----------------------------|----------------------------|
| <b>Symbol</b>         |                     |                            |                            |
| <b>BOKF</b>           | 1.494500            | 1.606500                   | 1.985700                   |
| <b>FGB</b>            | 0.031700            | 0.050900                   | 0.028900                   |
| <b>VLLX</b>           | 0.136000            | 0.136000                   | 0.136000                   |
| <b>GPACU</b>          | 0.025400            | 0.031700                   | 0.033200                   |
| <b>GAHC</b>           | 0.000300            | 0.000300                   | 0.000300                   |
| <b>AIMC</b>           | 0.746100            | 0.745700                   | 1.102900                   |
| <b>BXS</b>            | 0.565100            | 0.493700                   | 0.603400                   |
| <b>BWEL</b>           | 8.627300            | 8.556000                   | 8.455700                   |
| <b>FITBI</b>          | 0.101600            | 0.337000                   | 0.282500                   |
| <b>NEN</b>            | 0.708800            | 0.636000                   | 0.636000                   |
| <b>Mean</b>           | 1.243700            | 1.259400                   | 1.326500                   |
| <b>Std. Deviation</b> | 2.502100            | 2.475200                   | 2.448100                   |
| <b>Best Counter</b>   | 5                   | 3                          | 2                          |

The Table 5.42 presents the results as overall balanced.

Table 5.43: Random Forest time results for the United States of America

|                       | Random Forest Time | Random Forest Sector Time | Random Forest Global Time |
|-----------------------|--------------------|---------------------------|---------------------------|
| <b>Symbol</b>         |                    |                           |                           |
| <b>BOKF</b>           | 0:00:00.106986     | 0:00:00.111701            | 0:00:00.094769            |
| <b>FGB</b>            | 0:00:00.116700     | 0:00:00.099241            | 0:00:00.095632            |
| <b>VLLX</b>           | 0:00:00.102241     | 0:00:00.095772            | 0:00:00.098736            |
| <b>GPACU</b>          | 0:00:00.102725     | 0:00:00.099733            | 0:00:00.092752            |
| <b>GAHC</b>           | 0:00:00.099734     | 0:00:00.108709            | 0:00:00.094255            |
| <b>AIMC</b>           | 0:00:00.096742     | 0:00:00.097247            | 0:00:00.095252            |
| <b>BXS</b>            | 0:00:00.114497     | 0:00:00.097247            | 0:00:00.092235            |
| <b>BWEL</b>           | 0:00:00.122496     | 0:00:00.106223            | 0:00:00.091729            |
| <b>FITBI</b>          | 0:00:00.102724     | 0:00:00.098179            | 0:00:00.090264            |
| <b>NEN</b>            | 0:00:00.097236     | 0:00:00.092752            | 0:00:00.094939            |
| <b>Mean</b>           | 0.106200           | 0.100700                  | 0.094100                  |
| <b>Std. Deviation</b> | 0.008800           | 0.006100                  | 0.002400                  |

The Table 5.43 shows better results when trained with global parameters.

### 5.4.6.3 Neural Network

Table 5.44: Neural network error results for the United States of America

|                | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|----------------|----------------------|-----------------------------|-----------------------------|
| <b>Symbol</b>  |                      |                             |                             |
| VLLX           | 0.162800             | 0.163400                    | 0.165100                    |
| FGB            | 0.099200             | 0.100800                    | 0.101200                    |
| FITBI          | 0.094500             | 0.094800                    | 0.094500                    |
| BWEL           | 0.111400             | 0.112800                    | 0.111000                    |
| BOKF           | 0.116800             | 0.118400                    | 0.116700                    |
| BXS            | 0.104100             | 0.103400                    | 0.103000                    |
| AIMC           | 0.108500             | 0.125600                    | 0.107800                    |
| NEN            | 0.097500             | 0.097000                    | 0.097200                    |
| GAHC           | 0.074100             | 0.075200                    | 0.073300                    |
| GPACU          | 0.129400             | 0.127300                    | 0.129900                    |
| Mean           | 0.109800             | 0.111900                    | 0.110000                    |
| Std. Deviation | 0.022500             | 0.022800                    | 0.023200                    |
| Best Counter   | 3                    | 2                           | 5                           |

The Table 5.44 presents the results as overall balanced.

Table 5.45: Neural network time results for the United States of America

|                | Neural Network Time | Neural Network Sector Time | Neural Network Global Time |
|----------------|---------------------|----------------------------|----------------------------|
| <b>Symbol</b>  |                     |                            |                            |
| VLLX           | 0:00:05.679151      | 0:00:05.096723             | 0:00:06.446260             |
| FGB            | 0:00:06.413490      | 0:00:06.289567             | 0:00:08.436127             |
| FITBI          | 0:00:06.409315      | 0:00:05.609610             | 0:00:07.703428             |
| BWEL           | 0:00:06.399490      | 0:00:06.544274             | 0:00:08.447522             |
| BOKF           | 0:00:06.824875      | 0:00:06.546273             | 0:00:08.151258             |
| BXS            | 0:00:06.421756      | 0:00:05.881659             | 0:00:08.973395             |
| AIMC           | 0:00:06.651207      | 0:00:05.944576             | 0:00:09.966595             |
| NEN            | 0:00:05.822001      | 0:00:06.292079             | 0:00:08.730575             |
| GAHC           | 0:00:06.637110      | 0:00:06.446762             | 0:00:10.177853             |
| GPACU          | 0:00:06.222567      | 0:00:05.986872             | 0:00:07.552907             |
| Mean           | 6.348100            | 6.063800                   | 8.458600                   |
| Std. Deviation | 0.358800            | 0.460500                   | 1.110800                   |

The Table 5.45 shows better results when trained with the sector parameters.

## Chapter 6

# Case Study Result Analysis

By analysing the results of the case study, I can conclude that:

- The ARIMA model depends heavily on the context of the data, since the results have shown that its best performance is balanced between all types of training.
- The Random Forest model is best suited when it's trained with individual parameters for each stock presented, as proven by all the best results being the case.
- The Neural Network had the most varied distribution between the training types. This can mean that the model is unreliable or that it adapts to the situation accordingly.

After analysing the mean of all results, the following Table 6.1 was created, demonstrating its best values shown in green

Taking these factors into account we can conclude that neural networks had more accuracy against all other methods, and it will be used in the implementation of this project.

Table 6.1: Best error results for each prediction method

|                      | ARIMA Error | Random Forest Error | Neural Network Error |
|----------------------|-------------|---------------------|----------------------|
| <b>energy</b>        | 3.441900    | 1.176500            | 0.115400             |
| <b>healthcare</b>    | 1.693100    | 1.269800            | 0.103400             |
| <b>technology</b>    | 4.071800    | 1.440900            | 0.101700             |
| <b>australia</b>     | 0.178400    | 0.056300            | 0.112700             |
| <b>germany</b>       | 4.236900    | 1.462900            | 0.104400             |
| <b>united_states</b> | 1.489900    | 1.158000            | 0.113800             |

Full results in Appendix C



## Chapter 7

# Implementation

In this chapter, it's displayed the implementation of the web application and the machine learning model.

### 7.1 Web Application

#### 7.1.1 Development

The web application was developed using the programming language PHP, using Visual Studio Code as its development tool.

#### 7.1.2 Web Pages

##### 7.1.2.1 Login Page

The first page of the website a user can encounter, a registered user has to input its credentials to access the website, Figure 7.1.

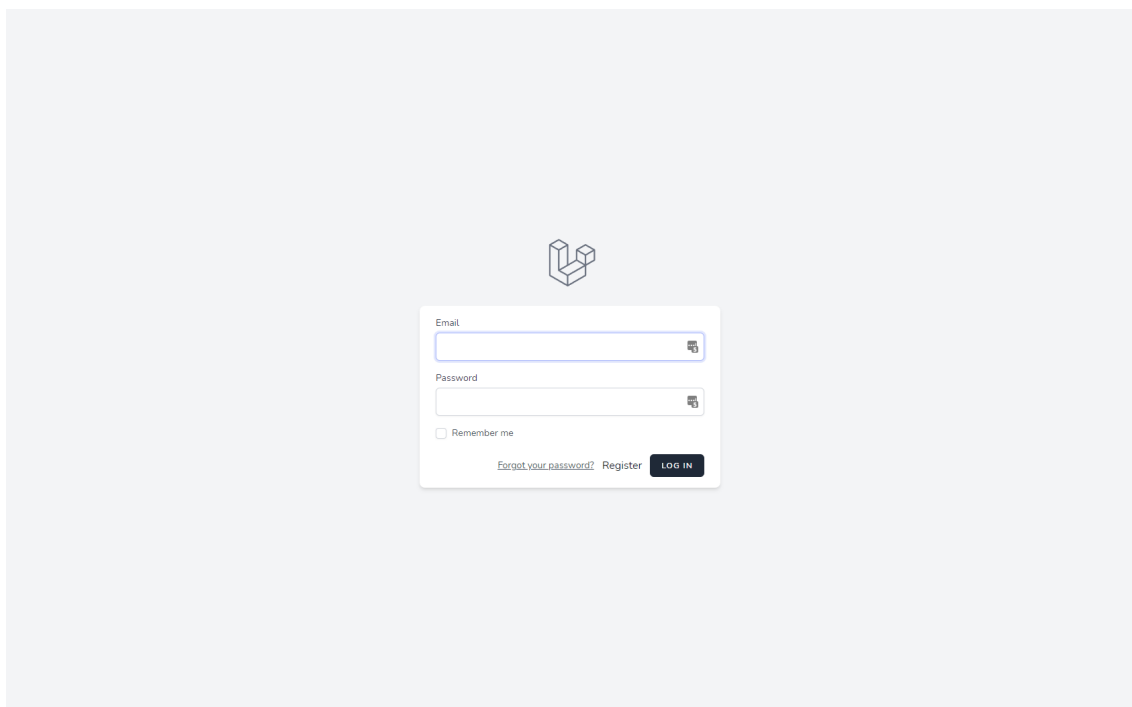


Figure 7.1: Login page

### 7.1.2.2 Register Page

The page responsible for allowing new users to register into the website, Figure 7.2.

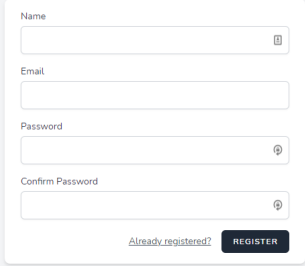
The image shows a registration form centered on a light gray background. At the top of the form is a logo consisting of three stacked cubes. Below the logo are four input fields: 'Name', 'Email', 'Password', and 'Confirm Password'. Each field has a small icon on its right side (a square for Name, an envelope for Email, and a key for both Password and Confirm Password). At the bottom of the form, there is a link that says 'Already registered?' and a dark gray button with the word 'REGISTER' in white capital letters.

Figure 7.2: Register page

### 7.1.2.3 Portfolio Page

A page depicting the overview of a investor's portfolio, a user can manage their portfolio in this page, Figure 7.3.

| Portfolio                             |                                 |       |       |          |          |       |                   |               |          |                   | Portfolio               | Lista de Mercados | Terminar Sessão |
|---------------------------------------|---------------------------------|-------|-------|----------|----------|-------|-------------------|---------------|----------|-------------------|-------------------------|-------------------|-----------------|
|                                       |                                 |       |       |          |          |       |                   |               |          |                   | Pesquisa                |                   |                 |
| Symbol                                | Name                            | Error | Score | Tendency | Variance | Tempo | Last Verification | Market Place  | Currency | Time Zone         |                         |                   |                 |
| <a href="#">VLY</a>                   | Valley National Bancorp         | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York  |                         |                   |                 |
| <a href="#">TEO</a>                   | Telecom Argentina SA            | ??    | ??    | =        | ??       | ??    | ??                | Argentina     | USD      | America/New_York  |                         |                   |                 |
| <a href="#">MAHABANK.NS</a>           | BK OF MAHARASHTRA               | ??    | ??    | =        | ??       | ??    | ??                | India         | INR      | Asia/Kolkata      |                         |                   |                 |
| <a href="#">NEM</a>                   | Newmont Corporation             | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York  |                         |                   |                 |
| <a href="#">CPL66.SA</a>              | COPEL PNB N1                    | ??    | ??    | =        | ??       | ??    | ??                | Brazil        | BRL      | America/Sao_Paulo |                         |                   |                 |
| <a href="#">MR.UN.TO</a>              | MELCOR REIT                     | ??    | ??    | =        | ??       | ??    | ??                | Canada        | CAD      | America/Toronto   |                         |                   |                 |
| <a href="#">TSL.TO</a>                | TREE ISLAND STEEL LTD           | ??    | ??    | =        | ??       | ??    | ??                | Canada        | CAD      | America/Toronto   |                         |                   |                 |
| <a href="#">EHI</a>                   | Western Asset Global High Incom | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York  |                         |                   |                 |
| <a href="#">HPE</a>                   | Hewlett Packard Enterprise Comp | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York  |                         |                   |                 |
| <a href="#">AKVA</a>                  | ARKANOVA ENERGY COMPANY         | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York  |                         |                   |                 |
|                                       |                                 |       |       |          |          |       |                   |               |          |                   |                         |                   |                 |
| A mostrar 1 até 10 de 25 linhas       |                                 |       |       |          |          |       |                   |               |          |                   | 10 registros por página |                   |                 |
| <a href="#">Ver Lista de Mercados</a> |                                 |       |       |          |          |       |                   |               |          |                   | 1 2 3 >                 |                   |                 |

Figure 7.3: Portfolio page

### 7.1.2.4 Market List Page

A list of all available stock markets a user can add to his portfolio, Figure 7.4.

| Mercados                              |                                 |       |       |          |          |       |                   |               |          |                  | Portfolio               | Lista de Mercados | Terminar Sessão |
|---------------------------------------|---------------------------------|-------|-------|----------|----------|-------|-------------------|---------------|----------|------------------|-------------------------|-------------------|-----------------|
|                                       |                                 |       |       |          |          |       |                   |               |          |                  | Pesquisa                |                   |                 |
| Symbol                                | Name                            | Error | Score | Tendency | Variance | Tempo | Last Verification | Market Place  | Currency | Time Zone        |                         |                   |                 |
| <a href="#">ALL</a>                   | Allstate Corporation (The)      | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York |                         |                   |                 |
| <a href="#">ADBE</a>                  | Adobe Inc.                      | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York |                         |                   |                 |
| <a href="#">GOOGL</a>                 | Alphabet Inc.                   | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York |                         |                   |                 |
| <a href="#">ACN</a>                   | Accenture plc                   | ??    | ??    | =        | ??       | ??    | ??                | Ireland       | USD      | America/New_York |                         |                   |                 |
| <a href="#">ABBV</a>                  | AbbVie Inc.                     | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York |                         |                   |                 |
| <a href="#">MT</a>                    | Arcelor Mittal NY Registry Shar | ??    | ??    | =        | ??       | ??    | ??                | Luxembourg    | USD      | America/New_York |                         |                   |                 |
| <a href="#">LLY</a>                   | Eli Lilly and Company           | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York |                         |                   |                 |
| <a href="#">APA</a>                   | APA Corporation                 | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York |                         |                   |                 |
| <a href="#">ADM</a>                   | Archer-Daniels-Midland Company  | ??    | ??    | =        | ??       | ??    | ??                | United States | USD      | America/New_York |                         |                   |                 |
| <a href="#">BMHAX</a>                 | BAUMART FPO                     | ??    | ??    | =        | ??       | ??    | ??                | Australia     | AUD      | Australia/Sydney |                         |                   |                 |
| A mostrar 1 até 10 de 1315 linhas     |                                 |       |       |          |          |       |                   |               |          |                  | 10 registros por página |                   |                 |
| <a href="#">Ver Lista de Mercados</a> |                                 |       |       |          |          |       |                   |               |          |                  | 1 2 3 4 5 ... 132 >     |                   |                 |

Figure 7.4: Market list page

## 7.2 Machine Learning

The machine learning part of the project was created using Python as its programming language and Jupyter as its development tool.

The prediction model chosen to train the stock market data models was neural networks, as by according to the case study results presented previously, was the best contender between the tested prediction models.

To train models using neural networks, the script fetches the data from yahoo! finance using its API, then prepares the dataset for training, and uses the Tensorflow library to create the model, train and predict values.

## Chapter 8

# Conclusion

The world of stock markets is a vast one, spanning many years of economic fluctuation, we witnessed very successful stories and hoped to get a fraction, so we pursued the idea of becoming an investor, but that is not a easy task in the slightest. The normal attitude is to learn about the necessary terms and situations, but due to the intensive influx of information, one starts to give up on the idea. That's where these projects come in, people trying to figure out a way to automate the process of investing and help those who don't have the time to dedicate.

With this thesis I concluded that although there are a lot of approaches to predict market values, such task is not easy since we can get close to the real values but not precise enough to make an impact because the amount of conditions and situations that can impact the future of a stock market are beyond unpredictable that number analysis is not able to make up for it.



# Bibliography

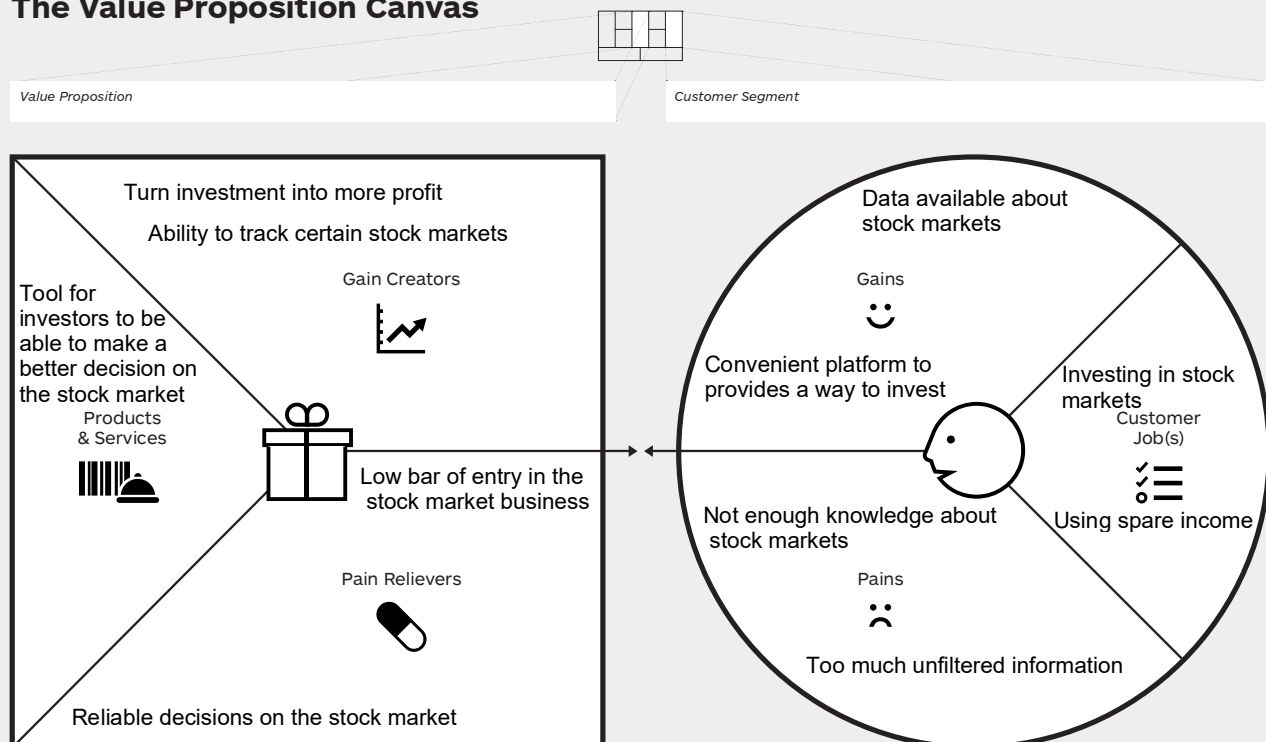
- Advantages and Disadvantages of PHP* (Nov. 2020). url: <https://www.geeksforgeeks.org/advantages-and-disadvantages-of-php/>.
- Bhatia, Richa (Feb. 2021). *Top 6 Regression Algorithms Used In Analytics & Data Mining*. url: <https://analyticsindiamag.com/top-6-regression-algorithms-used-data-mining-applications-industry/>.
- Classification Algorithms in Machine Learning: How They Work* (Aug. 2020). url: <https://monkeylearn.com/blog/classification-algorithms/>.
- Clustering in Machine Learning* (Feb. 2020). url: <https://www.geeksforgeeks.org/clustering-in-machine-learning/>.
- Hedayati, Amin, Moein Moghaddam, and Morteza Esfandyari (July 2016). "Stock market index prediction using artificial neural network:" in: *Journal of Economics, Finance and Administrative Science*. doi: 10.1016/j.jefas.2016.07.002.
- Kopp, Carol M. (Dec. 2020). *Understanding Perceived Value*. url: <https://www.investopedia.com/terms/p/perceived-value.asp>.
- Mittal, Aditi (Oct. 2019). *Understanding RNN and LSTM*. url: <https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e>.
- Moghar, Adil and Mhamed Hamiche (2020). "Stock Market Prediction Using LSTM Recurrent Neural Network". In: *Procedia Computer Science* 170. The 11th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 3rd International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops, pp. 1168–1173. issn: 1877-0509. doi: <https://doi.org/10.1016/j.procs.2020.03.049>. url: <https://www.sciencedirect.com/science/article/pii/S1877050920304865>.
- Mushtaq, Rizwan (Aug. 2011). "Augmented Dickey Fuller Test". In:
- Sathya, R. and Annamma Abraham (Feb. 2013). "Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification". In: *International Journal of Advanced Research in Artificial Intelligence* 2. doi: 10.14569/IJARAI.2013.020206.
- Science, Data-Driven (May 2020). *Python vs R for Data Science: And the winner is..* url: <https://medium.com/@datadrivenscience/python-vs-r-for-data-science-and-the-winner-is-3ebb1a968197>.
- Tinungki, G M (Feb. 2019). "The analysis of partial autocorrelation function in predicting maximum wind speed". In: 235, p. 012097. doi: 10.1088/1755-1315/235/1/012097. url: <https://doi.org/10.1088/1755-1315/235/1/012097>.
- Top Machine Learning Algorithms Explained: How Do They Work?* (Aug. 2020). url: <https://monkeylearn.com/blog/machine-learning-algorithms/>.
- Velay, Marc and Fabrice Daniel (June 2018). "Using NLP on news headlines to predict index trends". In:



## **Appendix A**

# **Canvas**

## The Value Proposition Canvas



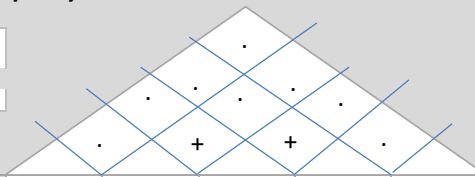
## **Appendix B**

### **QFD**

## Quality Function Deployment

Title: Stock Market Tendency  
Prediction Methods

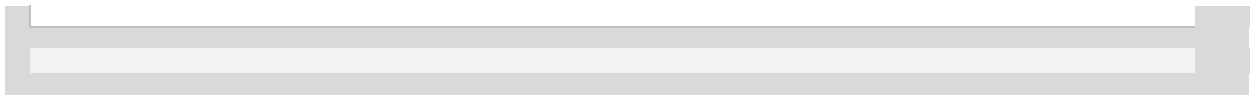
Date: 05/03/2021



|                |          |                |          |
|----------------|----------|----------------|----------|
| Correlation:   | +        | .              | -        |
|                | Positive | No correlation | Negative |
| Relationships: | 9        | 3              | 1        |
|                | Strong   | Moderate       | Weak     |

|                            |                                | Desired direction of improvement (↑,0,↓)               |                                                 |                                  |                                                    |                                |                                           |                                          |                |                |   |
|----------------------------|--------------------------------|--------------------------------------------------------|-------------------------------------------------|----------------------------------|----------------------------------------------------|--------------------------------|-------------------------------------------|------------------------------------------|----------------|----------------|---|
| 1: low, 5: high            |                                | Functional Requirements(How) →                         | Customizable Interface                          | Reliable website hosting         | Stock market recommendations                       | Accurate prediction algorithms | Extensive list of available stock markets | Competitive evaluation (1: low, 5: high) |                |                |   |
| Customer importance rating | Customer Requirements (What) ↓ | Weighted Score                                         |                                                 |                                  |                                                    |                                |                                           | Satisfaction rating                      | Yahoo! Finance | Google Finance |   |
| 1                          | 1                              | Accessible graphic user interface                      | 9                                               | 0                                | 0                                                  | 0                              | 0                                         | 9                                        |                | 9              | 3 |
| 2                          | 4                              | Get a large variety of stock markets                   | 0                                               | 0                                | 9                                                  | 0                              | 9                                         | 72                                       |                | 9              | 9 |
| 3                          | 5                              | Consistent and reliable prediction data                | 0                                               | 9                                | 3                                                  | 9                              | 0                                         | 105                                      |                | 1              | 1 |
| 4                          | 2                              | Able to add a subset of stock markets to his portfolio | 3                                               | 0                                | 9                                                  | 0                              | 9                                         | 42                                       |                | 9              | 9 |
| 5                          | 3                              | Assortement of options to choose from when predicting  | 3                                               | 0                                | 0                                                  | 3                              | 1                                         | 21                                       |                | 0              | 0 |
| 6                          |                                |                                                        |                                                 |                                  |                                                    |                                |                                           | 0                                        |                |                |   |
| 7                          |                                |                                                        |                                                 |                                  |                                                    |                                |                                           | 0                                        |                |                |   |
| 8                          |                                |                                                        |                                                 |                                  |                                                    |                                |                                           | 0                                        |                |                |   |
| 9                          |                                |                                                        |                                                 |                                  |                                                    |                                |                                           | 0                                        |                |                |   |
|                            |                                | Technical importance score                             | 24                                              | 45                               | 69                                                 | 54                             | 57                                        | 249                                      |                |                |   |
|                            |                                | Importance %                                           | 10%                                             | 18%                              | 28%                                                | 22%                            | 23%                                       | 100%                                     |                |                |   |
|                            |                                | Priorities rank                                        | 5                                               | 4                                | 1                                                  | 3                              | 2                                         |                                          |                |                |   |
|                            |                                | Current performance                                    |                                                 |                                  |                                                    |                                |                                           |                                          |                |                |   |
|                            |                                | Target                                                 | Able to customize almost everything on the page | Website 99% online and available | Create automatic connections between stock markets | 95% of accuracy over real data | Contain mostly available market places    |                                          |                |                |   |
|                            |                                | Benchmark                                              |                                                 |                                  |                                                    |                                |                                           |                                          |                |                |   |
|                            |                                | Difficulty                                             | 4                                               | 3                                | 5                                                  | 5                              | 3                                         | 1: very easy, 5: very difficult          |                |                |   |
|                            |                                | Cost and time                                          | 4                                               | 3                                | 5                                                  | 5                              | 2                                         | 1: low, 5: high                          |                |                |   |
|                            |                                | Priority to improve                                    | 10                                              | 15                               | 10                                                 | 20                             | 5                                         |                                          |                |                |   |

Comments/Conclusion:







## Appendix C

# Full Error Results

## C.1 Energy

### C.1.1 ARIMA

| Symbol  | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|---------|-------------|--------------------|--------------------|
| GXE.TO  | 0.129200    | 0.120400           | 0.120400           |
| BPT     | 0.162000    | 0.163500           | 0.163500           |
| GEI.TO  | 0.169800    | 0.238800           | 0.238800           |
| ORG.AX  | 0.349200    | 0.223800           | 0.223800           |
| ESN.TO  | 0.061700    | 0.045500           | 0.045500           |
| KEL.TO  | 0.298300    | 0.182800           | 0.182800           |
| AR      | 1.070100    | 1.429100           | 1.429100           |
| HWO.TO  | 0.112700    | 0.124700           | 0.124700           |
| IPL.TO  | 0.176800    | 0.013700           | 0.013700           |
| POU.TO  | 1.280500    | 1.199800           | 1.199800           |
| LNEGY   | 0.835500    | 0.954100           | 0.936200           |
| HE.TO   | 0.001100    | 0.000000           | 0.000000           |
| OXY     | 3.918100    | 2.916800           | 2.916800           |
| AKER.OL | 63.766600   | 68.473000          | 68.473000          |
| GKP.L   | 5.530300    | 9.265500           | 9.265500           |
| IOX.OL  | 0.070200    | 0.076400           | 0.076400           |
| BP.L    | 9.953000    | 8.209900           | 8.209900           |
| SCL.TO  | 0.259900    | 0.264900           | 0.264900           |
| VET.TO  | 0.551400    | 0.667800           | 0.667800           |
| AKSO.OL | 3.600800    | 1.446700           | 1.446700           |
| AM      | 0.789100    | 0.471500           | 0.471500           |
| BLY.AX  | 1.019800    | 1.018900           | 1.018900           |
| BBLS    | 0.006000    | 0.003600           | 0.003600           |
| BNXR    | 0.000400    | 0.000600           | 0.000600           |
| REX     | 2.870800    | 4.088600           | 4.088600           |
| CTP.AX  | 0.010300    | 0.010800           | 0.010800           |
| BWO.OL  | 0.637200    | 1.103500           | 1.103500           |
| APA     | 1.910700    | 2.301900           | 2.301900           |
| NNA     | 0.273200    | 0.143500           | 0.143500           |

## C.1.2 Random Forest

| Symbol  | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|---------|---------------------|----------------------------|----------------------------|
| BLY.AX  | 0.126400            | 0.163500                   | 0.126400                   |
| OXY     | 0.893600            | 0.744900                   | 0.695500                   |
| GEI.TO  | 0.191100            | 0.257100                   | 0.191100                   |
| REX     | 1.914400            | 3.136700                   | 1.914400                   |
| KEL.TO  | 0.195300            | 0.197600                   | 0.195300                   |
| ORG.AX  | 0.078100            | 0.064200                   | 0.078100                   |
| AKSO.OL | 0.422300            | 0.460100                   | 0.460100                   |
| BBL.S   | 0.007200            | 0.007200                   | 0.007200                   |
| POU.TO  | 0.377700            | 0.324900                   | 0.306900                   |
| APA     | 0.663800            | 0.861200                   | 0.694700                   |
| AR      | 0.518500            | 0.635800                   | 0.642600                   |
| IOX.OL  | 0.033200            | 0.063400                   | 0.063400                   |
| BWO.OL  | 0.673300            | 0.983600                   | 0.983600                   |
| GKP.L   | 3.119200            | 8.078500                   | 8.078500                   |
| HWO.TO  | 0.041200            | 0.038300                   | 0.035900                   |
| NNA     | 0.127000            | 0.173100                   | 0.173100                   |
| HE.TO   | 0.000000            | 0.000000                   | 0.000000                   |
| GXE.TO  | 0.033500            | 0.044600                   | 0.044300                   |
| CTP.AX  | 0.003900            | 0.004400                   | 0.005300                   |
| BPT     | 0.122600            | 0.170600                   | 0.150300                   |
| BNXR    | 0.000200            | 0.000200                   | 0.000200                   |
| VET.TO  | 0.273500            | 0.404600                   | 0.428800                   |
| LNEGY   | 3.108900            | 3.142400                   | 3.108900                   |
| SCL.TO  | 0.193300            | 0.147700                   | 0.131300                   |
| BP.L    | 5.488800            | 5.488800                   | 5.488800                   |
| IPL.TO  | 0.033600            | 0.032800                   | 0.033600                   |
| ESN.TO  | 0.012900            | 0.016700                   | 0.016000                   |
| AKER.OL | 15.298300           | 15.298300                  | 15.298300                  |
| AM      | 0.165600            | 0.232800                   | 0.165500                   |

### C.1.3 Neural Network

|         | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|---------|----------------------|-----------------------------|-----------------------------|
| Symbol  |                      |                             |                             |
| BPT     | 0.090400             |                             | 0.083700                    |
| BNXR    | 0.254300             |                             | 0.177300                    |
| ORG.AX  | 0.149000             |                             | 0.141800                    |
| BP.L    | 0.114900             |                             | 0.114400                    |
| GKP.L   | 0.119100             |                             | 0.118300                    |
| SCL.TO  | 0.099800             |                             | 0.098600                    |
| BLY.AX  | 0.129400             |                             | 0.067800                    |
| IOX.OL  | 0.099500             |                             | 0.086300                    |
| AKER.OL | 0.112500             |                             | 0.112400                    |
| APA     | 0.120400             |                             | 0.120000                    |
| NNA     | 0.233700             |                             | 0.104400                    |
| VET.TO  | 0.117500             |                             | 0.119900                    |
| AM      | 0.119100             |                             | 0.118300                    |
| AR      | 0.113200             |                             | 0.117600                    |
| HWO.TO  | 0.133000             |                             | 0.130100                    |
| BWO.OL  | 0.080400             |                             | 0.078500                    |
| IPL.TO  | 0.109100             |                             | 0.108700                    |
| ESN.TO  | 0.127900             |                             | 0.122900                    |
| LNEGY   | 0.162600             |                             | 0.160900                    |
| HE.TO   | 0.161100             |                             | 0.151600                    |
| POU.TO  | 0.121300             |                             | 0.115900                    |
| GXE.TO  | 0.119800             |                             | 0.117600                    |
| OXY     | 0.120200             |                             | 0.118900                    |
| AKSO.OL | 0.116600             |                             | 0.117500                    |
| GEI.TO  | 0.070700             |                             | 0.070000                    |
| CTP.AX  | 0.116100             |                             | 0.100000                    |
| REX     | 0.108200             |                             | 0.085800                    |
| KEL.TO  | 0.120500             |                             | 0.118000                    |
| BBL.S   | 0.300000             |                             | 0.168400                    |

## C.2 Healthcare

### C.2.1 ARIMA

|        | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|--------|-------------|--------------------|--------------------|
| Symbol |             |                    |                    |
| ADPT   | 3.076800    | 4.317400           | 3.578000           |
| BMV    | 0.895900    | 3.538200           | 1.152700           |
| AGEN   | 0.449200    | 0.884600           | 0.143700           |
| RDHL   | 0.551000    | 3.876800           | 0.304600           |
| NVS    | 0.486700    | 3.064900           | 0.633200           |
| ABC.L  | 19.362100   | 19.455100          | 16.533000          |
| UBQU   | 0.000100    | 0.000300           | 0.000100           |
| BLRX   | 0.110100    | 0.070900           | 0.165600           |
| IMV.TO | 0.115400    | 0.091400           | 0.153900           |
| ARRAY  | 0.128400    | 0.226900           | 0.118500           |
| GNGBY  | 0.626700    | 2.104900           | 0.805700           |
| EW     | 4.964300    | 9.009000           | 5.648300           |
| ABBV   | 2.239700    | 1.425200           | 1.938700           |
| LH     | 3.317800    | 17.828600          | 4.271400           |
| ALKS   | 1.125700    | 1.024600           | 1.038900           |
| ABC    | 2.905300    | 2.896900           | 2.746800           |
| PBH    | 3.088000    | 2.502300           | 1.977400           |
| TRIB   | 0.095900    | 0.524600           | 0.096300           |
| IPD.AX | 0.009400    | 0.009700           | 0.005600           |
| ASNB   | 0.007100    | 0.022100           | 0.016700           |
| MCK    | 2.250100    | 3.090100           | 2.538500           |
| BMRN   | 0.853500    | 3.997600           | 0.848500           |
| AEZS   | 0.067100    | 0.100200           | 0.067200           |
| BRKR   | 1.008200    | 10.897200          | 1.927300           |
| GUD.TO | 0.076300    | 0.107800           | 0.036900           |
| BNTC   | 0.157600    | 0.199800           | 0.109400           |
| CI     | 2.368600    | 11.912600          | 1.828700           |
| AGTC   | 0.416700    | 0.486000           | 0.412900           |
| MYHI   | 0.001700    | 0.008400           | 0.001100           |

## C.2.2 Random Forest

|        | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|--------|---------------------|----------------------------|----------------------------|
| Symbol |                     |                            |                            |
| AEZS   | 0.025400            | 0.079500                   | 0.049300                   |
| UBQU   | 0.000100            | 0.000100                   | 0.000100                   |
| AGEN   | 0.178800            | 0.277200                   | 0.306500                   |
| IMV.TO | 0.104900            | 0.074800                   | 0.085400                   |
| BLRX   | 0.078900            | 0.079200                   | 0.108700                   |
| BMV    | 0.744100            | 0.677200                   | 0.744100                   |
| GUD.TO | 0.078900            | 0.080600                   | 0.078700                   |
| IPD.AX | 0.003700            | 0.006600                   | 0.004200                   |
| ARRAY  | 0.200600            | 0.200600                   | 0.123400                   |
| ABC    | 1.469000            | 1.441300                   | 2.066400                   |
| GNGBY  | 0.213000            | 0.213000                   | 0.277100                   |
| MYHI   | 0.001800            | 0.005200                   | 0.002500                   |
| ABC.L  | 36.937900           | 15.115600                  | 15.115600                  |
| MCK    | 2.451700            | 4.335600                   | 1.999800                   |
| RDHL   | 0.423000            | 1.035400                   | 1.046100                   |
| NVS    | 0.234900            | 0.566700                   | 0.235300                   |
| ABBV   | 0.828300            | 1.172700                   | 0.823800                   |
| PBH    | 0.833700            | 0.978400                   | 1.110500                   |
| AGTC   | 0.107100            | 0.092500                   | 0.098600                   |
| BMRN   | 0.633900            | 1.173500                   | 0.633900                   |
| ALKS   | 0.426300            | 0.570200                   | 0.737300                   |
| EW     | 1.267700            | 1.618400                   | 1.533300                   |
| BNTC   | 0.081100            | 0.082400                   | 0.091800                   |
| TRIB   | 0.103700            | 0.113500                   | 0.103700                   |
| CI     | 2.812500            | 3.877600                   | 2.806900                   |
| ADPT   | 1.160200            | 1.703300                   | 1.196400                   |
| BRKR   | 1.161100            | 1.832800                   | 2.243200                   |
| LH     | 2.691000            | 1.683200                   | 3.190600                   |
| ASNB   | 0.004900            | 0.012400                   | 0.012200                   |

## C.2.3 Neural Network

|        | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|--------|----------------------|-----------------------------|-----------------------------|
| Symbol |                      |                             |                             |
| BRKR   | 0.087100             | 0.086000                    | 0.086500                    |
| BMJ    | 0.198500             | 0.103400                    | 0.107200                    |
| BLRX   | 0.090300             | 0.091300                    | 0.089700                    |
| GUD.TO | 0.157800             | 0.155100                    | 0.154600                    |
| AGEN   | 0.115500             | 0.089000                    | 0.088500                    |
| ADPT   | 0.141000             | 0.103300                    | 0.104500                    |
| BMRN   | 0.125100             | 0.127700                    | 0.128400                    |
| ABC    | 0.118200             | 0.117100                    | 0.117300                    |
| NVS    | 0.219700             | 0.089200                    | 0.089700                    |
| CI     | 0.079400             | 0.079900                    | 0.079600                    |
| ASNB   | 0.211800             | 0.218600                    | 0.218700                    |
| PBH    | 0.104700             | 0.103700                    | 0.102000                    |
| IPD.AX | 0.118100             | 0.119400                    | 0.119400                    |
| UBQU   | 0.099000             | 0.099500                    | 0.099600                    |
| ABBV   | 0.119400             | 0.124700                    | 0.095700                    |
| AGTC   | 0.099700             | 0.096600                    | 0.095600                    |
| IMV.TO | 0.108200             | 0.104100                    | 0.104600                    |
| TRIB   | 0.088500             | 0.073700                    | 0.074100                    |
| ALKS   | 0.111400             | 0.112300                    | 0.113000                    |
| BNTC   | 0.078600             | 0.078900                    | 0.079000                    |
| ABC.L  | 0.103700             | 0.112000                    | 0.112400                    |
| AEZS   | 0.064300             | 0.064400                    | 0.064200                    |
| MCK    | 0.107800             | 0.127300                    | 0.108000                    |
| RDHL   | 0.194400             | 0.131700                    | 0.132100                    |
| MYHI   | 0.098900             | 0.087300                    | 0.086400                    |
| ARRAY  | 0.087900             | 0.088400                    | 0.088200                    |
| LH     | 0.092800             | 0.093100                    | 0.093500                    |
| GNGBY  | 0.088800             | 0.089500                    | 0.089200                    |
| EW     | 0.076200             | 0.076200                    | 0.076500                    |

## C.3 Technology

### C.3.1 ARIMA

|           | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|-----------|-------------|--------------------|--------------------|
| Symbol    |             |                    |                    |
| SAIC      | 2.138300    | 1.528300           | 1.458600           |
| TEAM      | 10.684800   | 18.082500          | 11.066600          |
| ACN       | 5.668900    | 13.680500          | 3.441000           |
| BGO.L     | 12.708000   | 21.679800          | 13.711800          |
| XRF.AX    | 0.057700    | 0.066900           | 0.056000           |
| HEXA-B.ST | 5.147600    | 11.893900          | 6.204300           |
| ADBE      | 19.358300   | 62.791900          | 9.527500           |
| MELE.BR   | 1.721200    | 6.804600           | 3.433100           |
| HXGBY     | 0.826600    | 1.728400           | 0.663900           |
| PACT.ST   | 2.789400    | 5.126300           | 2.385500           |
| NCR       | 3.728600    | 1.955700           | 2.697400           |
| IBM       | 4.209200    | 3.229600           | 3.103500           |
| EVS.BR    | 0.506200    | 0.586500           | 0.542200           |
| HMNY      | 0.000700    | 0.000600           | 0.000500           |
| BELFA     | 0.598200    | 0.989300           | 0.658100           |
| IAIC      | 0.124300    | 0.157700           | 0.219100           |
| GLW       | 0.620100    | 1.524500           | 0.460800           |
| ORCL      | 4.137100    | 2.773300           | 4.696900           |
| LRCX      | 26.430900   | 28.580300          | 38.421300          |
| CIEN      | 0.751000    | 3.126200           | 0.464500           |
| ENA.V     | 0.001200    | 0.001300           | 0.001800           |
| CHKP      | 4.732700    | 6.668100           | 3.841500           |
| DYSL      | 0.090100    | 0.093500           | 0.150000           |
| DDD       | 1.862900    | 3.760600           | 1.802400           |
| XRX       | 1.261000    | 0.543300           | 0.926700           |
| APH       | 1.909900    | 1.129700           | 1.844100           |
| BCOV      | 0.936600    | 0.618800           | 0.545100           |
| MSFT      | 7.791400    | 6.607500           | 4.817100           |
| IVX.DE    | 0.858900    | 3.450600           | 0.941100           |

### C.3.2 Random Forest

| Symbol    | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|-----------|---------------------|----------------------------|----------------------------|
| XRF.AX    | 0.014200            | 0.034800                   | 0.022600                   |
| DYSL      | 0.019600            | 0.019600                   | 0.019600                   |
| EVS.BR    | 0.280000            | 0.357100                   | 0.357100                   |
| ADBE      | 6.287400            | 9.755400                   | 8.245900                   |
| APH       | 0.791200            | 1.001900                   | 1.024000                   |
| BELFA     | 0.291700            | 0.291700                   | 0.362600                   |
| SAIC      | 1.278800            | 1.444600                   | 0.717200                   |
| ORCL      | 1.411200            | 2.203000                   | 1.736900                   |
| CIEN      | 0.737700            | 0.795000                   | 0.597500                   |
| DDD       | 0.514300            | 0.529100                   | 0.700700                   |
| IVX.DE    | 0.427300            | 0.427300                   | 0.427300                   |
| HMNY      | 0.000200            | 0.000200                   | 0.000200                   |
| ACN       | 2.008100            | 5.290500                   | 4.020900                   |
| GLW       | 0.533600            | 0.425600                   | 0.548600                   |
| IAIC      | 0.097000            | 0.097000                   | 0.102200                   |
| TEAM      | 5.415400            | 6.211400                   | 7.544400                   |
| BCOV      | 0.224200            | 0.226600                   | 0.325200                   |
| IBM       | 1.472200            | 1.787700                   | 1.758100                   |
| ENA.V     | 0.000000            | 0.000000                   | 0.000000                   |
| HEXA-B.ST | 1.751500            | 3.022100                   | 3.022100                   |
| LRCX      | 6.777800            | 6.777800                   | 10.785100                  |
| PACT.ST   | 2.364300            | 2.054900                   | 2.054900                   |
| MELE.BR   | 1.347300            | 1.737300                   | 1.737300                   |
| CHKP      | 1.237400            | 1.613500                   | 0.989100                   |
| HXGBY     | 0.160100            | 0.160100                   | 0.221500                   |
| BGO.L     | 3.072900            | 7.164300                   | 7.164300                   |
| XRX       | 0.186700            | 0.193700                   | 0.202400                   |
| MSFT      | 2.101200            | 4.600500                   | 3.373900                   |
| NCR       | 0.984200            | 0.802300                   | 1.034600                   |

### C.3.3 Neural Network

| Symbol    | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|-----------|----------------------|-----------------------------|-----------------------------|
| APH       | 0.112100             | 0.112300                    | 0.112100                    |
| BGO.L     | 0.095400             | 0.098200                    | 0.095200                    |
| HXGBY     | 0.079800             | 0.079200                    | 0.080600                    |
| ACN       | 0.095800             | 0.096200                    | 0.097100                    |
| XRF.AX    | 0.112600             | 0.110000                    | 0.109200                    |
| TEAM      | 0.110200             | 0.118000                    | 0.108400                    |
| DDD       | 0.086700             | 0.089000                    | 0.086100                    |
| LRCX      | 0.097800             | 0.095000                    | 0.097800                    |
| SAIC      | 0.098100             | 0.097800                    | 0.096400                    |
| EVS.BR    | 0.099800             | 0.100500                    | 0.103900                    |
| ORCL      | 0.108000             | 0.109200                    | 0.121800                    |
| GLW       | 0.074100             | 0.074200                    | 0.073800                    |
| IBM       | 0.082200             | 0.081800                    | 0.081500                    |
| IVX.DE    | 0.135300             | 0.133200                    | 0.136200                    |
| MSFT      | 0.097700             | 0.098300                    | 0.098900                    |
| MELE.BR   | 0.096500             | 0.100700                    | 0.096200                    |
| IAIC      | 0.129500             | 0.131100                    | 0.130500                    |
| XRX       | 0.082900             | 0.084100                    | 0.109200                    |
| BCOV      | 0.100300             | 0.076900                    | 0.076900                    |
| HMNY      | 0.081200             | 0.080900                    | 0.080700                    |
| NCR       | 0.091000             | 0.090800                    | 0.090200                    |
| ADBE      | 0.087000             | 0.071200                    | 0.070100                    |
| CHKP      | 0.113600             | 0.111800                    | 0.115000                    |
| BELFA     | 0.081900             | 0.082400                    | 0.081500                    |
| DYSL      | 0.127900             | 0.128000                    | 0.128400                    |
| ENA.V     | 0.236100             | 0.235500                    | 0.235300                    |
| CIEN      | 0.094400             | 0.094400                    | 0.093800                    |
| PACT.ST   | 0.084300             | 0.089500                    | 0.085900                    |
| HEXA-B.ST | 0.076700             | 0.079000                    | 0.077700                    |

## C.4 Australia

### C.4.1 ARIMA

|        | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|--------|-------------|--------------------|--------------------|
| Symbol |             |                    |                    |
| WHC.AX | 0.306800    | 0.252000           | 0.250100           |
| BOL.AX | 0.004200    | 0.004000           | 0.004100           |
| IMU.AX | 0.041900    | 0.051400           | 0.050300           |
| BMH.AX | 0.000000    | 0.000000           | 0.000000           |
| JRV.AX | 0.045800    | 0.047700           | 0.039700           |
| JYC.AX | 0.495800    | 0.301400           | 0.313600           |
| CTP.AX | 0.010300    | 0.011300           | 0.010800           |
| CDD.AX | 0.156900    | 0.184300           | 0.177200           |
| EME.AX | 0.050300    | 0.016800           | 0.015100           |
| NHF.AX | 0.195900    | 0.175800           | 0.217700           |
| BRK.AX | 0.002300    | 0.001800           | 0.001800           |
| BKW.AX | 0.843500    | 0.878600           | 0.927700           |
| KOV.AX | 0.479700    | 0.533400           | 0.531600           |
| TLS.AX | 0.069600    | 0.038600           | 0.030500           |

### C.4.2 Random Forest

|        | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|--------|---------------------|----------------------------|----------------------------|
| Symbol |                     |                            |                            |
| NHF.AX | 0.096700            | 0.111800                   | 0.086400                   |
| CDD.AX | 0.030100            | 0.030100                   | 0.028500                   |
| KOV.AX | 0.074100            | 0.061500                   | 0.131600                   |
| WHC.AX | 0.074700            | 0.089000                   | 0.120300                   |
| BMH.AX | 0.000000            | 0.000000                   | 0.000000                   |
| IMU.AX | 0.015300            | 0.017900                   | 0.016700                   |
| BKW.AX | 0.344500            | 0.610300                   | 0.392200                   |
| JRV.AX | 0.025700            | 0.025700                   | 0.018600                   |
| BRK.AX | 0.001700            | 0.001600                   | 0.001400                   |
| TLS.AX | 0.031100            | 0.035300                   | 0.035900                   |
| CTP.AX | 0.004400            | 0.004400                   | 0.005300                   |
| JYC.AX | 0.067300            | 0.136900                   | 0.072100                   |
| BOL.AX | 0.004800            | 0.004900                   | 0.002800                   |
| EME.AX | 0.018000            | 0.046300                   | 0.041200                   |

### C.4.3 Neural Network

|         | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|---------|----------------------|-----------------------------|-----------------------------|
| Symbol  |                      |                             |                             |
| TL.S.AX | 0.100700             | 0.100200                    | 0.101300                    |
| NHF.AX  | 0.093900             | 0.097200                    | 0.094400                    |
| BMH.AX  | 0.309600             | 0.189000                    | 0.190200                    |
| CTP.AX  | 0.116500             | 0.102300                    | 0.104500                    |
| EME.AX  | 0.117800             | 0.117100                    | 0.117100                    |
| JRV.AX  | 0.102300             | 0.102100                    | 0.100800                    |
| WHC.AX  | 0.113000             | 0.111900                    | 0.111300                    |
| BOL.AX  | 0.113200             | 0.114300                    | 0.114100                    |
| KOV.AX  | 0.102900             | 0.104500                    | 0.103300                    |
| JYC.AX  | 0.117400             | 0.121000                    | 0.115700                    |
| CDD.AX  | 0.117900             | 0.116300                    | 0.116700                    |
| IMU.AX  | 0.109300             | 0.110700                    | 0.108800                    |
| BRK.AX  | 0.108400             | 0.106900                    | 0.107300                    |
| BKW.AX  | 0.091700             | 0.092500                    | 0.092000                    |

## C.5 Germany

### C.5.1 ARIMA

|         | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|---------|-------------|--------------------|--------------------|
| Symbol  |             |                    |                    |
| K1R.DE  | 0.377200    | 0.098700           | 0.098700           |
| BAYN.DE | 0.932100    | 0.805100           | 0.805100           |
| BEI.DE  | 1.750100    | 1.418900           | 1.418900           |
| KSB.DE  | 23.519100   | 25.805700          | 25.805700          |
| BMW.DE  | 3.263200    | 3.603000           | 3.603000           |
| NDX1.DE | 1.115000    | 1.299000           | 1.299000           |
| BA S.DE | 1.812800    | 2.342100           | 2.342100           |
| G1A.DE  | 0.983800    | 0.844500           | 0.844500           |
| BVB.DE  | 0.051100    | 0.046900           | 0.046900           |
| DB1.DE  | 2.713700    | 2.735900           | 2.735900           |
| AAD.DE  | 10.035500   | 4.132700           | 4.132700           |
| ADS.DE  | 15.656200   | 14.206400          | 14.206400          |
| DBK.DE  | 0.447900    | 0.400900           | 0.400900           |
| IVX.DE  | 0.858900    | 0.941100           | 0.941100           |
| DAI.DE  | 4.796700    | 4.871900           | 4.871900           |

### C.5.2 Random Forest

| Symbol  | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|---------|---------------------|----------------------------|----------------------------|
| BAS.DE  | 0.758600            | 1.541400                   | 1.541400                   |
| BVB.DE  | 0.212100            | 0.212100                   | 0.212100                   |
| G1A.DE  | 0.445000            | 0.445000                   | 0.445000                   |
| IVX.DE  | 0.427300            | 0.427300                   | 0.427300                   |
| K1R.DE  | 0.082500            | 0.226400                   | 0.226400                   |
| BMW.DE  | 1.499300            | 1.090100                   | 1.090100                   |
| AAD.DE  | 3.698700            | 3.626600                   | 3.626600                   |
| KSB.DE  | 6.608200            | 14.234800                  | 14.234800                  |
| BEI.DE  | 0.730500            | 0.973600                   | 0.973600                   |
| ADS.DE  | 3.753900            | 3.750100                   | 3.750100                   |
| BAYN.DE | 0.457800            | 0.457800                   | 0.457800                   |
| NDX1.DE | 0.315400            | 0.521500                   | 0.521500                   |
| DBK.DE  | 0.117600            | 0.117600                   | 0.117600                   |
| DAI.DE  | 1.390800            | 1.640200                   | 1.640200                   |
| DB1.DE  | 1.445900            | 1.445900                   | 1.445900                   |

### C.5.3 Neural Network

| Symbol  | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|---------|----------------------|-----------------------------|-----------------------------|
| DB1.DE  | 0.102600             | 0.114100                    | 0.099900                    |
| BAYN.DE | 0.080800             | 0.078800                    | 0.079500                    |
| BVB.DE  | 0.190400             | 0.088400                    | 0.088700                    |
| BEI.DE  | 0.073400             | 0.090300                    | 0.088900                    |
| BAS.DE  | 0.084400             | 0.113500                    | 0.111600                    |
| DBK.DE  | 0.109100             | 0.107500                    | 0.109400                    |
| KSB.DE  | 0.111300             | 0.110800                    | 0.112400                    |
| G1A.DE  | 0.105100             | 0.099800                    | 0.101600                    |
| K1R.DE  | 0.152200             | 0.153000                    | 0.152700                    |
| DAI.DE  | 0.111500             | 0.111300                    | 0.110500                    |
| ADS.DE  | 0.104100             | 0.100500                    | 0.101000                    |
| IVX.DE  | 0.133400             | 0.142900                    | 0.134700                    |
| BMW.DE  | 0.090800             | 0.090700                    | 0.092100                    |
| NDX1.DE | 0.082200             | 0.079200                    | 0.081600                    |
| AAD.DE  | 0.102400             | 0.101100                    | 0.101500                    |

## C.6 United States of America

### C.6.1 ARIMA

|              | ARIMA Error | ARIMA Sector Error | ARIMA Global Error |
|--------------|-------------|--------------------|--------------------|
| Symbol       |             |                    |                    |
| <b>BNXR</b>  | 0.000400    | 0.004400           | 0.000600           |
| <b>FGB</b>   | 0.147400    | 0.062600           | 0.078300           |
| <b>HRL</b>   | 1.064600    | 0.775200           | 1.090900           |
| <b>AIMC</b>  | 1.303200    | 2.927300           | 1.508400           |
| <b>CCD</b>   | 0.957100    | 0.926300           | 0.905800           |
| <b>GAHC</b>  | 0.000900    | 0.000400           | 0.000700           |
| <b>BWEL</b>  | 8.249000    | 48.167500          | 8.071400           |
| <b>BDC</b>   | 2.203900    | 1.506500           | 2.444300           |
| <b>NEN</b>   | 1.516200    | 0.896200           | 1.801700           |
| <b>BOKF</b>  | 3.147800    | 5.670900           | 2.345700           |
| <b>GPACU</b> | 0.071300    | 0.038800           | 0.043300           |
| <b>VLLX</b>  | 0.496100    | 2.758800           | 0.446900           |
| <b>CI</b>    | 2.342800    | 11.912600          | 1.828700           |
| <b>BXS</b>   | 0.629500    | 1.044700           | 1.052400           |
| <b>FITBI</b> | 0.758300    | 0.618100           | 0.728900           |

### C.6.2 Random Forest

|              | Random Forest Error | Random Forest Sector Error | Random Forest Global Error |
|--------------|---------------------|----------------------------|----------------------------|
| Symbol       |                     |                            |                            |
| <b>BOKF</b>  | 1.494500            | 1.606500                   | 1.985700                   |
| <b>CI</b>    | 2.791300            | 3.878000                   | 2.807100                   |
| <b>VLLX</b>  | 0.136000            | 0.136000                   | 0.136000                   |
| <b>BXS</b>   | 0.565100            | 0.493700                   | 0.603400                   |
| <b>GPACU</b> | 0.025400            | 0.031700                   | 0.033200                   |
| <b>AIMC</b>  | 0.746100            | 0.745700                   | 1.102900                   |
| <b>GAHC</b>  | 0.000300            | 0.000300                   | 0.000300                   |
| <b>FGB</b>   | 0.031700            | 0.050900                   | 0.028900                   |
| <b>BWEL</b>  | 8.627300            | 8.556000                   | 8.455700                   |
| <b>HRL</b>   | 0.551600            | 0.541600                   | 0.484400                   |
| <b>NEN</b>   | 0.708800            | 0.636000                   | 0.636000                   |
| <b>FITBI</b> | 0.101600            | 0.337000                   | 0.282500                   |
| <b>BNXR</b>  | 0.000200            | 0.000200                   | 0.000200                   |
| <b>BDC</b>   | 1.315400            | 1.436900                   | 1.112800                   |
| <b>CCD</b>   | 0.275000            | 0.351100                   | 0.387200                   |

### C.6.3 Neural Network

|        | Neural Network Error | Neural Network Sector Error | Neural Network Global Error |
|--------|----------------------|-----------------------------|-----------------------------|
| Symbol |                      |                             |                             |
| FITBI  | 0.094500             | 0.094800                    | 0.094500                    |
| FGB    | 0.099200             | 0.100800                    | 0.101200                    |
| NEN    | 0.097500             | 0.097000                    | 0.097200                    |
| GAHC   | 0.074100             | 0.075200                    | 0.073300                    |
| BNXR   | 0.253000             | 0.178200                    | 0.179000                    |
| HRL    | 0.217800             | 0.113800                    | 0.113600                    |
| BXS    | 0.104100             | 0.103400                    | 0.103000                    |
| GPACU  | 0.129400             | 0.127300                    | 0.129900                    |
| CI     | 0.079900             | 0.079700                    | 0.079500                    |
| CCD    | 0.110000             | nan                         | nan                         |
| BOKF   | 0.116800             | 0.118400                    | 0.116700                    |
| BDC    | 0.120300             | 0.121300                    | 0.121600                    |
| BWEL   | 0.111400             | 0.112800                    | 0.111000                    |
| VLLX   | 0.162800             | 0.163400                    | 0.165100                    |
| AIMC   | 0.108500             | 0.125600                    | 0.107800                    |