REPUBLIC OF TURKEY

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Information Technologies

PREDICTING FLUCTUATION OF TRENDS IN STOCK MARKET USING DEEP LEARNING

Iman Mohammed Taher MOHAMMED

Master`s Thesis

Supervisor

Asst. Prof. Dr. Ayca Kurnaz TURKBEN

Istanbul, 2022

PREDICTING FLUCTUATION OF TRENDS IN STOCK MARKET USING DEEP LEARNING

Iman Mohammed Taher MOHAMMED

Information Technologies

Master`s Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled PREDICTING FLUCTUATION OF TRENDS IN STOCK MARKET USING DEEP LEARNING prepared by IMAN MOHAMMED TAHER MOHAMMED and submitted on 11/08/2022 has been **accepted unanimously** for the degree of Master of Science in Information Technologies

____
Asst. Prof. Dr. Ayca Kurnaz TURKBEN

Supervisor

Thesis Defense Jury Members:

Asst. Prof. Dr. Ayca Kurnaz TURKBEN	Faculty of Engineering and Architecture, Altinbas University	___________________
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Faculty of Engineering and Architecture, Altinbas University	___________________
Asst. Prof. Dr. Serdar KARGIN	Faculty of Engineering and Architecture, Beykent University	___________________

I hereby declare that this thesis meets all format and submission requirements of a Master`s Thesis.

Submission date of the thesis to the Graduate Education Institute: ____/____/____

I hereby declare that all information presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Iman MOHAMMED

Signature

DEDICATION

First, I dedicate my dissertation to my beloved father, Mohammed Taher though he is no longer in this world, he was my inspiration to pursue my master's degree. thank you very much my father I will make sure to put your name in all my prayers.

This work is also dedicated to my mother who has always loved me unconditionally.

This dissertation is dedicated to my husband who has been a constant source of support and encouragement during the challenge of my master's degree and during my life.

To my beloved kids: Mariam, and Hamza who I want them to be proud of their mother.

Lastly, I dedicated this to all my family and friends who encourage and support me.

ACKNOWLEDGEMENT

I would like to express my gratitude to my supervisor Asst. Prof. Dr. Ayca KURNAZ for all support during my study. I would also like to thank my friends and family who supported me into the study.

ABSTRACT

PREDICTING FLUCTUATION OF TRENDS IN STOCK MARKET USING DEEP LEARNING

Mohammed, Iman

M.Sc., Information Technologies, Altınbaş University,

Supervisor: Asst. Prof. Dr. Ayca Kurnaz TURKBEN

Date: 08/2022

Pages: 62

Recent studies indicate that variations in the value of the stock market are difficult to anticipate due to the large number of unknowns and variables that affect its value on any given day. This includes the current market conditions, investor opinion toward a certain firm, and political developments. Hurriedly and arbitrarily, the pricing marketplaces are selected when setting the stock price. Due to the fact that it is not rare for the stock market to be dynamic and disorderly (due to several reasons), the stock market's direction is categorized as a random process, with more shifting possibilities in short time frames Therefore, an attempt to carry out a price forecast in the stock market can bring great benefits to investors, by increasing the level of information about the financial market, minimizing exposure to financial risk. In this sense, a computational technique called Artificial Neural Networks (ANNs) can be applied. The Artificial Neural Network (ANN) simulates on computers the functioning of the human brain in a simplified way.

Keywords: ANN, Stock Market, Ml, Dl

TABLE OF CONTENTS

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

GDP	:	Gross Domestic Product
AI	:	Artificial Intelligent
SVM	:	Support Vector Machine
NN	:	Neural Network
ANN	:	Artificial Neural Network
ML	:	Machine Learning
SDDE	:	Software Development Project Estimation
BPNN	:	Backpropagation Neural Networks
DCF	:	Discounted Cash-Flow
IDX	:	Indonesian Stock Index
JKSA	:	Jakarta Stock Exchange
MLP	:	Multi-layer Perceptron
KNN	:	K Nearest Neighbors
RF	:	Randon Forest
ROI	:	Return On Investment
MAE	:	Mean Absolute Error
RMSE	:	Root Mean Square Error
MAPE	:	Mean Absolute Percentage Error

LSTM	:	Long Short Time Memory

1 INTRODUCTION

1.1 BACKGROUND

Recent studies indicate that variations in the value of the stock market are difficult to anticipate due to the large number of unknowns and variables that affect its value on any given day. This includes the current market conditions, investor opinion toward a certain firm, and political developments. Hurriedly and arbitrarily, the pricing marketplaces are selected when setting the stock price. Due to the fact that it is not rare for the stock market to be dynamic and disorderly (due to several reasons), the stock market's direction is categorized as a random process, with more shifting possibilities in short time frames. Throughout the course of time, some features tend to progress in a linear fashion. Due to their disorderliness and unpredictability, investments in securities are subject to a high degree of risk. Gaining a grasp of the probable future price variations might be of great assistance in avoiding this problem [1]

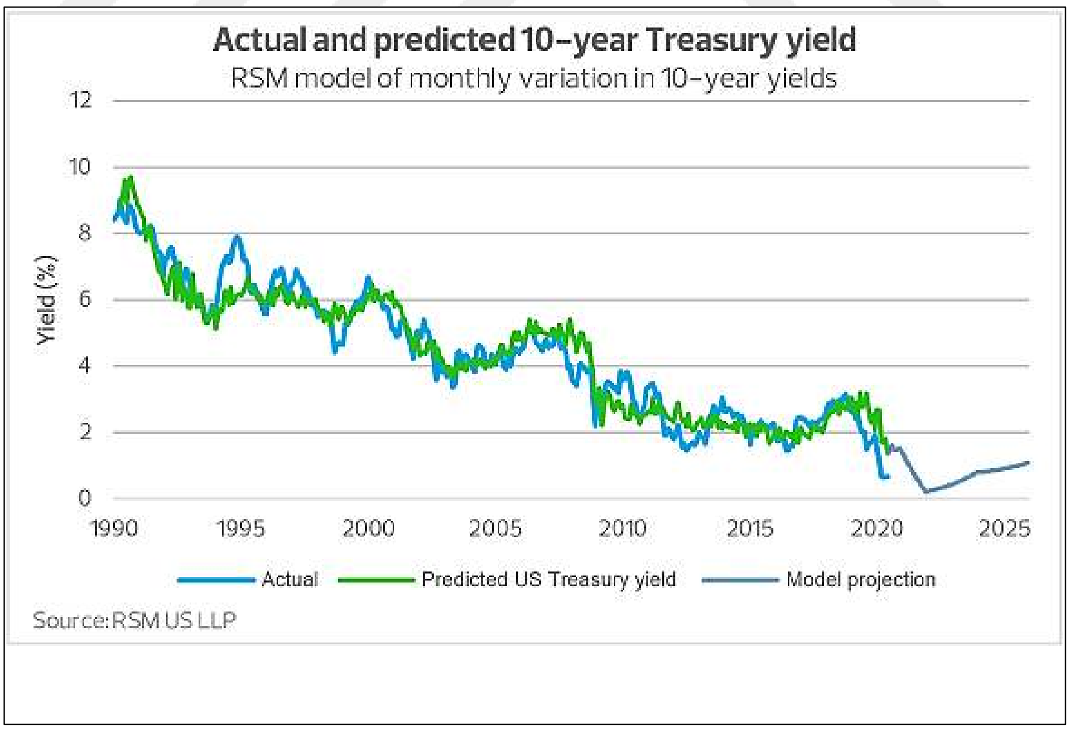

Figure 1.1: Actual vs predicted stock market within the next 10 years [1]

1.2 PLANNING THE PROBLEM

This work aims to predict the share price direction of the global stock market, as an alternative to closing prices, since it can help to minimize the investment risks of operations.

For this, this study uses three machine learning algorithms - random Forest (RF), artificial neural networks (ANN) and support vector machines (SVM). In addition, the variables that most influence the trends in the daily share prices companies are defined statistical parameters are stipulated to compare the efficiency of the proposed algorithms. In this way, it is possible to analyze through measurements (accuracy, precision $recall$, specificity, among others) the best models developed [2].

This degree thesis stems from a personal interest in the analysis of financial markets and the development of automatic trading systems. This aspect, together with the curiosity towards the topic of artificial intelligence, laid the foundations for the birth of this research thesis. The advances that world technology is experiencing today allow results that were unimaginable before: from self-driving cars to voice assistants now present in all newly marketed electronic devices, there are no longer the applications of artificial intelligence in everyday life. On the other hand, today's society is used to being surrounded by every possible form of technology. However, what we almost never ask ourselves is: “which tool allows us to obtain these results?”. The answer in most cases is one: machine learning [3].

1.3 THESIS JUSTIFICATION

Globalization is a process of deepening economic, social, cultural and political integration. This economic level has made the application of financial resources a complex task for investors. This is due to the amount of information coming from various parts of the world. In general, every company or individual investor, when applying their resources, thinks about at least two aspects: return and risk. With an adequate level of information and knowledge of the financial market, it is possible, for a certain level of return, to reduce exposure to risk. Therefore, an attempt to carry out a price forecast in the stock market can bring great benefits to investors, by increasing the level of information about the financial market, minimizing exposure to financial risk. In this sense, a computational technique called Artificial Neural Networks (ANNs) can be applied. The Artificial Neural Network (ANN) simulates on computers the functioning of the human brain in a simplified way. It has the ability to recognize patterns, identify regularities, deal with noisy, incomplete or inaccurate data and preview non-linear systems, which makes its application interesting in the financial market [4]. The use of ANNs can help investors in the choice of assets as it offers the possibility of predicting the behavior of stock prices in the future and thus subsidizing the decisions to buy and/or sell securities. In this context, this article aims to develop an artificial neural network capable of predicting the price accordingly and, with that, explore the ability of neural networks the research problem is the question: how the following artificial neural networks can be applied in the stock price prediction process?

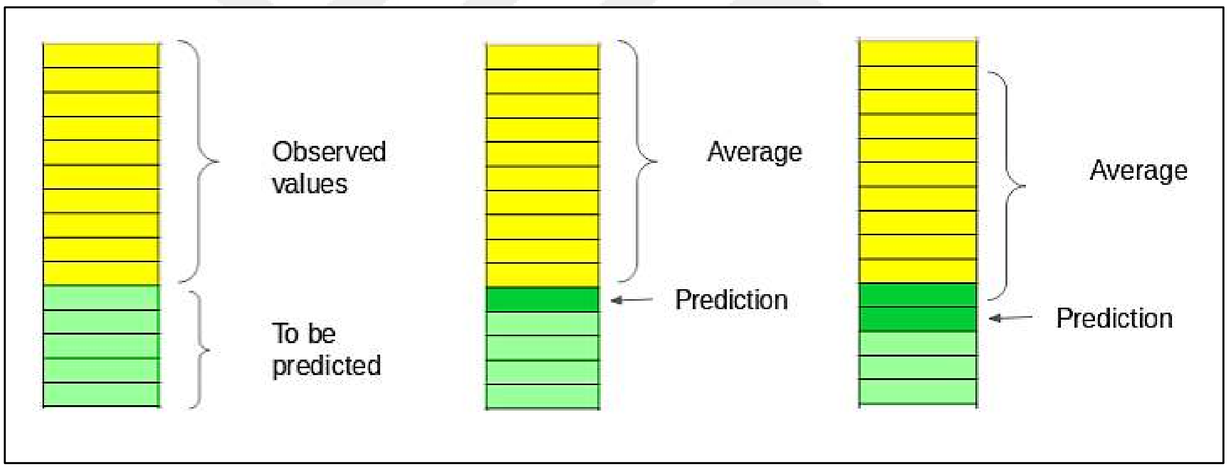

Figure 1.2: Dataset variables in the prediction process [4]

1.4 THESIS OUTLINE

This dissertation is divided into 6 chapters, as follows:

Chapter 1 establishes the target theme of the research, presents the objectives to be achieved with the work, which involves the research of machine learning algorithms applied to the forecast of the stock market. It also presents the challenges, which include the comparison of different learning techniques and the evaluation of statistical parameters of the results. Therefore, it guides on the way in which this written work was structured to facilitate the understanding of the whole.

Chapter 2 provides an overview of technical and fundamental analysis, in addition to describing in detail the works present in the literature related to the research topic, emphasizing studies that employ the same algorithms and techniques presented here. There is also a discussion of the results obtained by the main references that guided the development of this study.

Chapter 3 specifies the algorithms used: RF, SVM and RNA. It addresses its mathematical formulations, describing the functioning, advantages and particular aspects of each machine learning model. Furthermore, the chapter lists the software tools that allowed the development of artificial intelligence models.

In chapter 4, the methodology used to develop the stock forecasting algorithms is presented, describing the pre-processing steps, techniques for dividing the training and test samples and the extraction of attributes from the historical price series. In addition, the statistical parameters that will be used to evaluate the results described in chapter 5 are described.

Chapter 5 explains in detail all the results achieved, dealing with the efficiency of each proposed prediction model. The data are presented in a structured way in tables and graphs, allowing the comparison from different perspectives of the particularities of each algorithm and forecast horizon used.

Chapter 6 presents the conclusion, the contributions made by this research and proposals for possible future related works. The appendices contain the codes of the forecasting algorithms developed for each model and forecast horizon proposed.

2 LITERATURE REVIEW

2.1 CHAPTER OVERVIEW

On the basis of a review of the previous research carried out in the field of computer science, the objective of this chapter is to investigate potential new directions for research into stock market forecasting (MF). A scientific method known as a systematic literature review is used to evaluate and characterize research that makes use of similar methods or takes place under similar circumstances. This method is then used to compare experiments in each group to identify key findings as well as those that require additional independent investigation. These findings will serve as fodder for further study into the use of machine learning algorithms to predict the overall stock market values and trends of the future for experts in the fields of computer science and finance who are familiar with the aforementioned fields. Researchers were able to gain fresh information about algorithms in a range of information technology and scientific domains by using a procedure known as literature review. The following are examples of three papers that made use of this methodology and received a significant number of citations.

2.2 PREVIOUS WORKS

Jasic and Wood (2019):

For the purpose of predicting daily stock market index returns, the author created an artificial neural network using data from a variety of international stock markets. Efforts are being made to assist traders in making a profit. Short-term stock market index returns are predicted with the use of untransformed data inputs, which are fed into univariate neural networks [5].

Enke and Thawornwong (2018):

Using machine learning information gathering methods, investigate the relationships between a broad variety of financial and economic indices. By computing the information gain for each model variable, it is possible to build a ranking of the model variables. When constructing forecasting models, a threshold is defined in order to guarantee that only the most significant variables are included in the models [6].

Wen et al. (2019):

By doing a thorough literature review, we discovered and researched machine learning methods for estimating software development labor (SDEE). Software development project estimation (SDEE) is a technique for estimating the time and resources required to accomplish a software development project. The articles included in this collection were initially published between 1991 and 2019 and were subjected to a rigorous review process to ensure that they were relevant and of high quality. Each system was assessed on four dimensions: machine learning, model estimate accuracy, model components, and estimation context. Machine learning was the most important factor. The results of this literature review technique are used to provide recommendations to researchers and to provide assistance to practitioners. In a distinct information technology context [7],

Malhotra (2018):

conducted a review of the literature on the use of machine learning models to predict software failures A method known as software fault prediction may be used to identify faulty modules or classes in the code early in the software development lifecycle, allowing for faster remediation (SFP). For this evaluation, we looked at papers that appeared between 1991 and 2013. A rating and assessment were given to each article based on its expected performance and comparative analyses with other statistics and machine learning methodologies. According to a survey of seven different kinds of studies, further research is needed in order to produce insights that may be applied to a broad variety of software defects in general. A set of guidelines and recommendations were developed as a consequence of the study by a group of researchers and practitioners. In fields of research where it is difficult to predict the future, extensive literature assessments have been employed to great effect [8].

Chavan and Patil (2020):

By examining nine previously published works that describe different model input parameters, we may further our understanding of artificial neural network (ANN)-based stock market prediction. They are seeking for the input parameters that have the biggest influence on the correctness of the model, and they will find them. According to their research, microeconomic elements are largely used to anticipate stock market index values, and that most machine learning algorithms depend on technical characteristics rather than fundamental variables to make individual stock price forecasts, rather than the other way around. As an extra bonus, hybridized parameters beat single-input variable techniques in terms of overall performance [9].

Chong, Han and Park (2020):

Deep learning networks may be used for stock market research and forecasting, among other things. Deep learning networks may be utilized for high-frequency stock market forecasting since they do not rely on past predictor information in order to create forecasts. Deep learning algorithms for stock market research and forecasting offer both advantages and limitations in terms of accuracy. With high-frequency intraday stock returns as input, the researchers investigate the influence of three unsupervised feature extraction methodologies on the network's overall ability to anticipate future market behavior: principal component analysis, autoencoder, and the constrained Boltzmann machine [10].

Lee (2019):

In order to anticipate the trend of stock markets, a support vector machine (SVM) using a hybrid feature selection technique was built. A new hybrid feature selection approach called F-score and Supported Sequential Forward Search makes it easier to select the ideal feature subset from the original feature set by combining filter techniques with wrapper methods in a new hybrid feature selection approach called Supported Sequential Forward Search (F SSFS). In order to evaluate the accuracy of this SVM-based model integrated with F SSFS in terms of prediction accuracy, backpropagation neural networks (BPNN) and feature selection methodologies such as information gain and correlation-based feature selection are used [11].

Yeh, Huang and Lee (2019):

troubleshoot support vector regression stock market value forecasting issues caused by hyperparameters in kernel functions. At this point in the procedure, the hyperparameter's value has already been determined. It is possible to combine the benefits of multiple hyperparameter settings and boost overall system performance in their system. As a result of using both sequential minimum optimization and gradient projection, they create a two-stage multiple-kernel learning algorithm [12].

Table 2.1: Summary of literature review.

Author	year	Findings
Jasic and Wood	2019	The author constructed an artificial neural network utilizing data from many foreign stock markets to anticipate daily stock market index results. Traders are being helped to make a profit.
Enke	2018	Investigate the correlations between a wide range of financial and economic variables using machine learning. The information gain for each model variable may be used to rank the variables.
Wen	2019	The articles in this collection were published between 1991 and 2019. We assessed each system's estimating context, model components, and estimate accuracy.
Malhotra	2018	Early detection of incorrect modules or classes in the code allows for speedier correction (SFP). For this review, we looked at articles from 1991 to 2013.

Table 2.1: Summary of literature review ”tables continued”

Chavan and Patil	2020	We can learn more about ANN-based stock market prediction by studying nine previously published papers that detail distinct model input parameters.
Chong, Han and Park	2020	The researchers use high-frequency intraday stock returns to test three unsupervised feature extraction methods on the network's ability to forecast future market behavior: PCA, autoencoder, and the constrained Boltzmann machine
Lee	2019	. Supported Sequential Forward Search is a revolutionary hybrid feature selection process that combines filter and wrapper methods to identify the optimal feature subset from the original feature set.
Yeh, Huang and Le	2019	Kernel function hyperparameters cause support vector regression stock market forecasting issues. The hyperparameter's value is known. There are many ways to increase system performance. In two stages, sequential minimum optimization and gradient projection are used to train numerous kernels.

2.3 CONCLUSION OF LITERATURE REVIEW

Investing in the stock market is a challenging process that requires much research. Machine learning techniques have been increasingly used to test if they can enhance market forecasting over conventional methods [13].

This chapter’s goal is to identify future machine learning stock market prediction research pathways based on prior research. Researchers conduct systematic literature reviews to find relevant peer-reviewed journal publications over the last two decades and logically group research that employ comparable approaches and circumstances. Studies using artificial neural networks, support vector machines, genetic algorithms, and other AI technologies fall into four categories: Each category is investigated to identify what similarities and differences exist, what is limited, and where further research is needed. Conclusions and prospective research directions are discussed finally.

3 MATERIALS AND METHODS

3.1 CHAPTER OVERVIEW

In this chapter we aim at explaining two concepts:

a. Experimental analysis: Where the different approaches adopted to respond to the points defined above are exposed. The goal is to use real historical data to experiment with machine learning algorithms and see if we get conclusive results.

b. Recommendation of a method: At the end of this thesis, we must show that machine learning techniques are useful for forecasting stock market movements, and consequently recommend a method that can be deployed in practice or demonstrate their weakness and their inability to beat the market and therefore reject the hypothesis of the possibility of predicting financial market movements using these techniques.

3.2 ELEMNTS OF THE PROPOSED MODEL

Predictability of short-term financial market movements: This is the main assumption. We devote a large part of this chapter to discussing it. This question divides a large number of researchers and financial market specialists. We believe that there are pockets of informational inefficiency that can be attributed to several factors that we will discuss later. If we accept that the movements of stocks are purely random, there is no reason to try to get into forecasting these series. For this work, we hypothesize that the trajectory of financial securities provides enough information to predict the trend of these series, up to a certain level, for financial data of short period (minute, hour today) and for the type of titles to use. Our results at the end will have to conclude us if this assumption holds the road [14].

Transaction costs: The second assumption that we will use for simplification purposes is that on transaction costs. First, it is assumed that there are no costs associated with transactions. Our goal at the beginning is to test the predictive power of the models to be used and not to implement algorithms to test their profitability. A comparison between the strategies to be tested will be made at the end, where we will integrate these costs to be able to quantify the profitability generated by these algorithms. An estimate of these costs per transaction is used for this purpose [14].

Exchange mechanisms and the microstructure of financial markets: Only liquid stock market assets are covered by this work, and it is assumed that it is possible to trade them at any time with the historical closing price. This assumption does not take into account the microstructure of the markets, where there is generally a difference between the expected price of a transaction and the price at which the transaction is actually executed ('slippage'). This phenomenon is accentuated during periods of high volatility when market orders are used, or also when there is not enough demand to maintain the expected price [14]. To deal with the profitability bias that will be generated in our theoretical results and to adapt to the reality of the market, we will apply an adjustment factor at the end to penalize the calculated profitability. Example: to open long positions, a cost factor will be integrated by adding to the prices used, an average of the difference between the 'Bid' price and the 'Ask' price of the asset used. The same is done to close these positions but deducting this factor from the closing price [15].

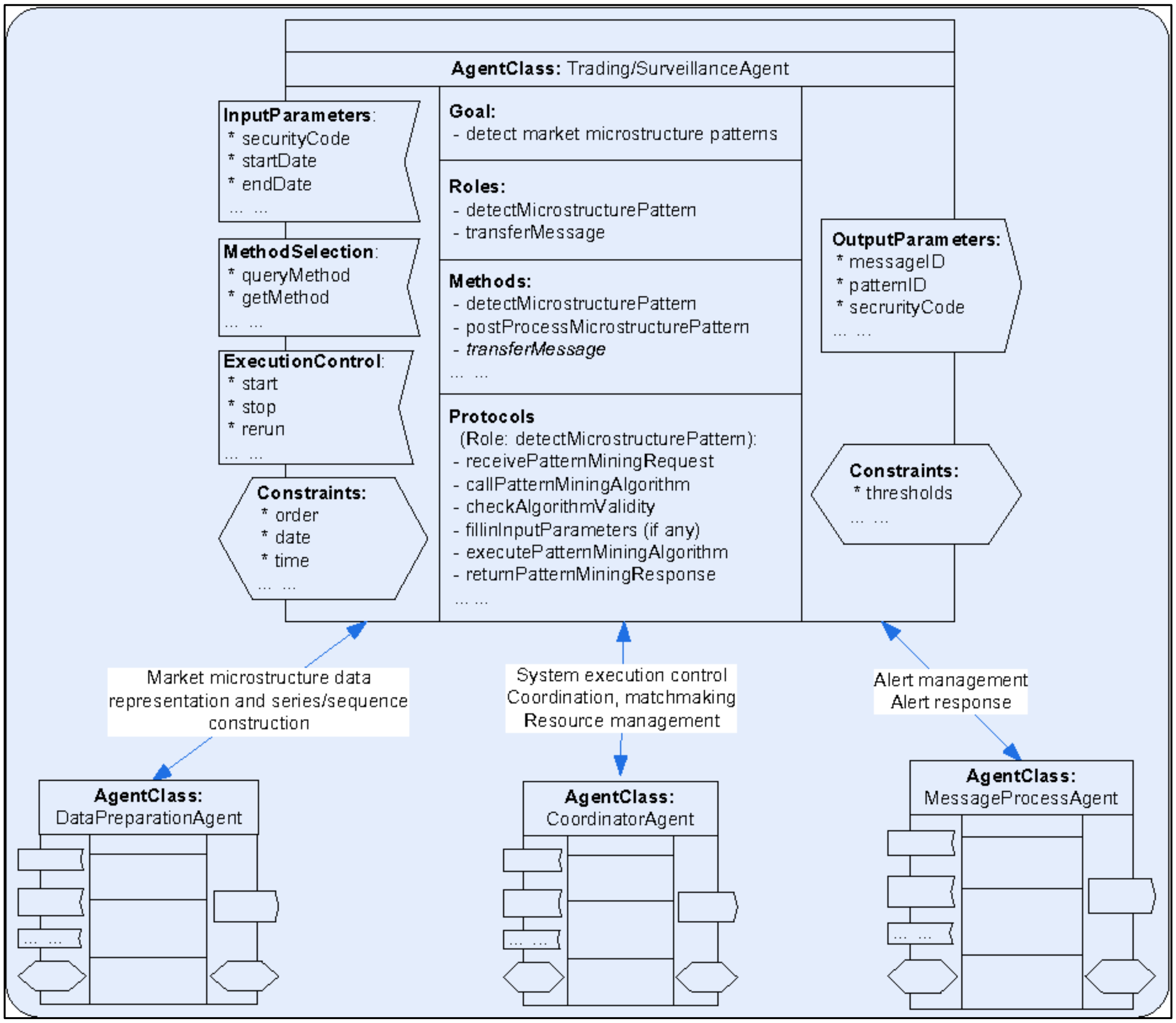

Figure 3.1: Exchange mechanisms and the microstructure of financial markets [16]

Short sales: We will assume here that it is possible to carry out short sales at any time. In other words, we do not want to limit ourselves to just predicting a bullish movement to buy the security and then sell it. But also, when the price tends to have a downward movement, we can start by borrowing it to sell it and buy it back afterwards, at a price that we think is lower than that of the sale. In practice, short selling is not permitted for certain securities or during certain crisis periods.

Profit threshold: The last assumption is related to the expected level of return for each transaction. In other words, one does not want to enter positions if one does not expect the price to vary by a certain percentage (threshold) during a predetermined time interval [16]. Example: a variation of one cent upwards for a stock worth $100 is not at all interesting to consider it as a bullish movement. We will use this logic in defining the trend that we will use as the target variable to train our models. This question will be discussed in more detail in the section on methodology and experimental analysis [17].

3.3 STEPS OF THE PROPOSED METHOD

The data used are financial series of futures contracts representing transactions on stock market indices and commodities that have high liquidity. Some challenges were encountered during the analysis and prediction of these financial series and deserve to be discussed, namely:

Data processing: cleaning, aggregation, balancing of data, transformation, imputation of missing data, creation and attribute selection [18].

Financial data time window: this is the periodicity of data aggregation. It may vary depending on the objective of the predictive modeling, from a few milliseconds or minutes to a few days. In order to have more flexibility depending on the choice of this parameter, we have created a function that allows to aggregate the data according to this parameter. Example: if the window is equal to 5 minutes, the data used in the training of our models and the prediction of the results will be of a granularity of 5 minutes instead of one minute.

The modeling and learning process: several points need to be considered when building predictive models. Among other things, overfitting and underfitting, conceptual drifts, selection of the parameters of the predictive models to use, model bias and the interpretability of the results.

Evaluation of predictive models: in order to quantify the validity and accuracy of predictions, several metrics and evaluation techniques can be used. The choice of which model to use may be based on several points that will be discussed later in Chapter 4. All of these points will be discussed in detail in the section on experimental analysis.

3.4 EFFICIENCY OF FINANCIAL MARKETS

3.4.1 Definition of Market Efficiency

The notion of efficient markets is a pillar of conventional finance theory. Investors cannot benefit from market imbalances due to the fact that the values of financial assets are decided by taking into account all of the available information. The literature on finance invests considerable time and effort to exploring this subject [19].

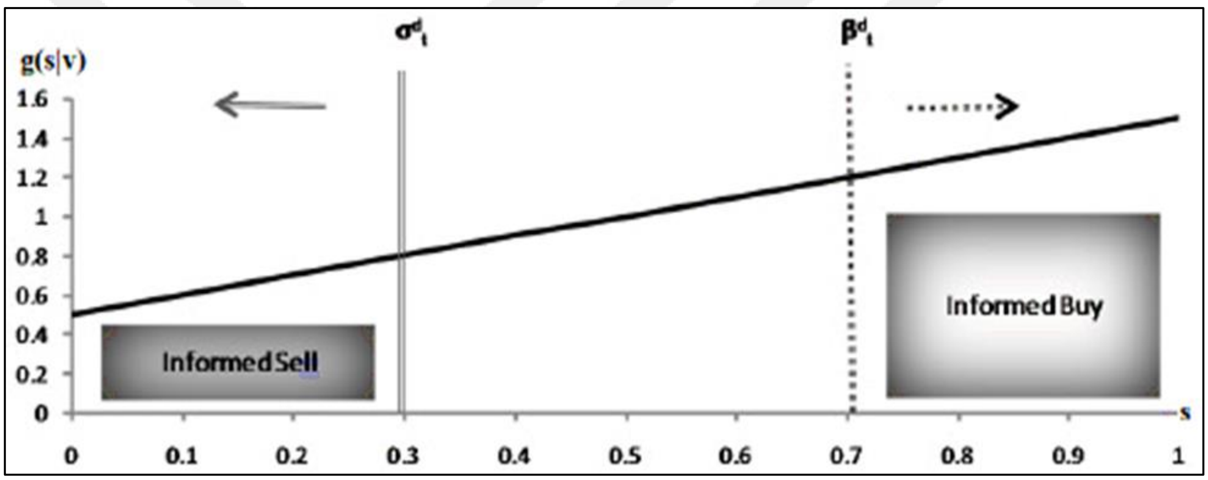

Figure 3.2: Market imbalance, leading to the realization of profits [20]

On the other hand, many people believe that the values of financial assets are prone to chaotic fluctuations and that no investment strategy can overcome management in an efficient market. Investing for the long run in a passive index. In spite of the fact that many say that financial series are predictable and that it is conceivable "to beat the market for a level of proportionate risk 3," the general view is that the route that stock prices follow consists of a series of random steps. This occurrence directly influenced the development of the "random walk theory of stock price changes." According to this theory, a speculator's mathematical expectation is considered to be equal to zero [22]. Fama conducted an empirical study on market efficiency in 1965 and, as a consequence of his results, brought the concept of an efficient market to the public for the very first time. According to him, the behavior of a random walk is the model that most closely corresponds to the stock market's asset prices. As soon as new information becomes accessible, a product's price will be adjusted to reflect the new conditions. In the framework of a formal economic argument, Samuelson is credited with having originally argued that "Prices of publicly traded equities fluctuate randomly." In order to improve his odds of winning, he adopts the martingale strategy rather than the random walk [23]. The difficulties of measuring efficiency in a single (global) form were noted, and it was recommended that it be divided into three distinct forms based on the kind of information that the market uses to represent current values of securities. Below are descriptions of the three distinct efficiency categories:

In its weakest form, this term indicates that present prices include all historical price information.

All relevant public information, such as current pricing, earnings reports, and merger and acquisition operations.

As a result, current prices and titles are determined from all relevant data, including information only known to insiders. If markets are efficient, anomalous gains are not possible, and reasonable behavior should be anticipated [25].

Due to the validity of this assumption, the market will never underestimate or overestimate the worth of the object being traded, allowing for the establishment of equilibrium. Due to this assumption, it is futile to try to evaluate or predict fluctuations in financial stability. The only factor that matters is the degree of risk taken relative to the expected reward. To put it another way, if an investor wants to raise their earnings, they must also increase their risk tolerance [26].

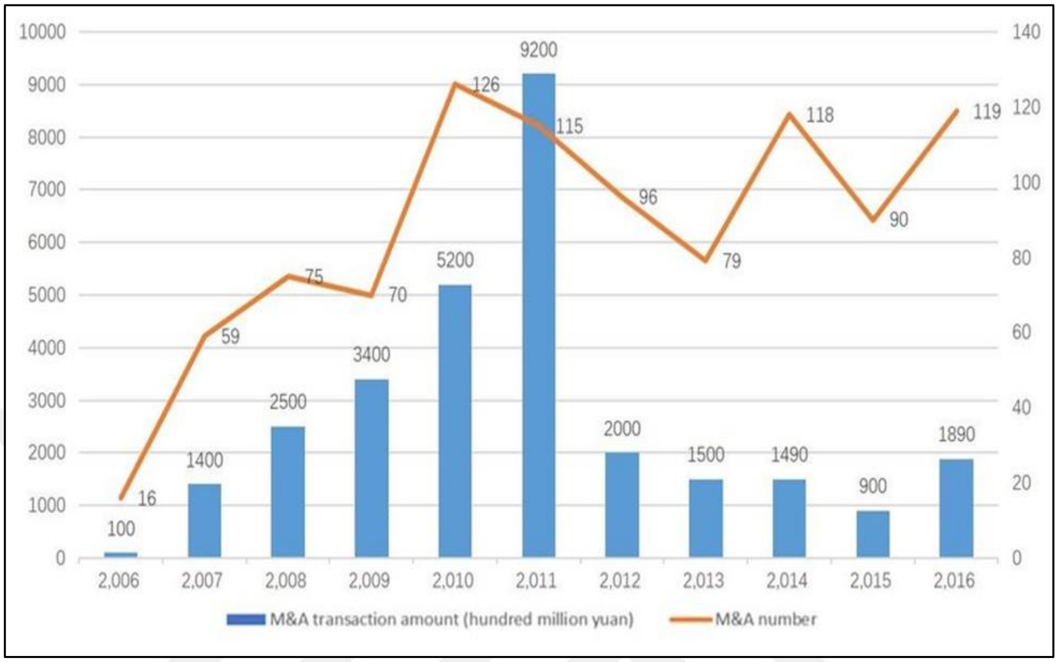

Figure 3.3: Amount and number of M&A transactions in China by 2021 [26]

3.4.2 Definition of a Random Walk

Random stroll describes a process in which there is no visible structure to the order in which phases are accomplished. The only thing that can determine the march's path is its present condition; its history has no impact whatsoever. According to this view, the questioned behavior is entirely unconscious [28]. A random stroll of n steps, for instance, would be the result of tossing a coin n times. Assume that each coin toss results in a one-point bonus for landing on heads and a one-point deduction for landing on tails, as seen in the accompanying figure. Due of this, it is impossible to predict the outcome of the (n+1)th throw solely on the results of the preceding n throws. As a direct consequence, the outcomes of each throw are seen as independent of one another. The total amount of steps for these treks is 10 million.

$$E[S] = E\left[\sum_{i=0}^{n} Xi\right] = \sum_{i=0}^{n} E[Xi] = \sum_{i=0}^{n} P[Xi].Xi = 0, \quad 0 < i < \infty \tag{3.1}$$

Indeed, each time we redo the experiment, we obtain a new random walk with a different pace. By doing the exercise 1000 times, we obtained the set of random walks by considering that all the walks start at the same starting point, and which is zero. We see that the distribution of the 1000 steps is centered around zero, which is the mathematical expectation of all the steps generated. This can be proven using the weak law of large numbers theorem. In other words, if the number of times a random experiment is redone is so great, for a balanced random walk with zero as its origin, the cumulative gain that will be obtained tends towards zero. In general, if we want to calculate X, the number of heads obtained by making n independent throws. It is a random variable that follows a binomial distribution of parameters (n: # of trials and p: the probability of having a success).

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{3.2}$$

Since the price of a security is already affected by variations in its price, adjusting the price of a security does not rely on the prior variation in its price. A toss of a coin might be used to depict the future of the stock market. If we suppose that stock market prices follow a random walk, then it is impossible to predict the stock market's price n steps from now (units of time: minutes, hours, days, etc.). Before the efficient market theory was proposed, the random walk hypothesis had been explored for almost 150 years [31]. A result of the efficient market hypothesis, but not a prerequisite for the theory's validity.

3.4.3 Critique of the Efficient Market Hypothesis

In the previous section we tried to relate the arguments behind the thought that financial markets are efficient and purely random. In the next section we will give our own arguments about this hypothesis. The price series of financial assets are not completely governed by chance. If the movements of stock market assets are completely random [32]. So why are fund managers or traders of high frequency financial securities among the highest paid professionals in the world. Why even investment banks invest colossal sums for innovation and the creation of winning strategies? Several factors cast doubt on the theory of the efficiency of financial markets. In addition, the possibility that the relevant information is not used appropriately is not taken into account in the hypothesis of the random walk of stock price movements. Moreover, the exceptional and recurring performance achieved by certain investors or certain speculators is a perfect counterexample to cast doubt on the validity of this hypothesis [33].

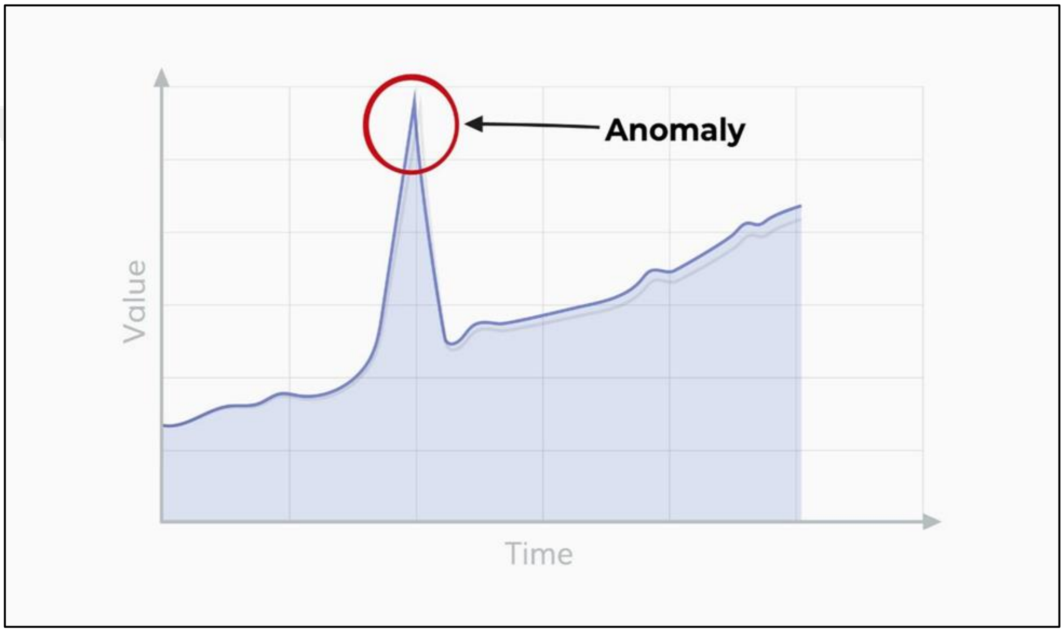

Figure 3.4: Stock Market anomaly [35]

These anomalies are attributed to the strategy adopted. For example, famed investor Warren Buffet attributes his success to thorough fundamental analysis of the companies he invests in. He buys the shares of companies that are undervalued and have great long-term growth potential. Over a period of 52 years, Buffet generated profits 155 times higher than the S&P 500 stock market index with an average annual rate of return of 20.8% During 48 years of activity, it has made profits 46 times (compared to 37 times for the S&P 500 index) and it has managed to beat the market 39 times. Mathematically, this is only probable in 4 times out of 1 billion the performance is so amazing that when we interpose the annual return of the stock of the company owned by Buffet "Berkshire Hathaway" with the return of the S&P, the latter looks like a flat line Another example is that of the mathematician James Simon who founded the Renaissance investment fund and who uses advanced quantitative and computational techniques to perform algorithmic trading. The fund has generated an average annual return of 35% since its inception in 1988. Between 1994 and 2014 the fund generated an average annual return of 71% [37]. The fund employs quantitative researchers who manipulate massive amounts of data using highly efficient and scalable technologies for computing and running highly complex quantitative algorithms.

3.4.4 The Existence of Anomalies

A return is said to be abnormal if it is not explained by the general trend of the market or even the expected return on the security. Regardless of the form of efficiency considered, irregularities have been observed in global stock markets persistently, we cite a few in the following.

The January Effect: Stocks that underperformed in the fourth quarter of the previous year tend to outperform in January. This may be justified by the capital gains tax offset of losses for underperforming stocks that investors seek to exit at year-end. Selling pressure is increased such that the stock becomes undervalued the day of the week effect: Research has shown that financial stocks tend to move more on Fridays than on Mondays, and that there is a bias towards positive performance on Fridays. It's not a huge gap, but it is a persistent gap Mergers and acquisitions: Several studies have shown that, in the short term, the securities for which a merger has taken place, are characterized by superior performance compared to shares of similar companies. This abnormal return is all the more pronounced if the merger was carried out by a public offer financed in money, and not in shares the shares of small firms perform better than those of large companies: This can be explained by the life cycle of these companies [38]. The growth opportunities of a small company are much greater than those of a large one. A supporter of classical financial thinking can justify this anomaly by the risk that a small company represents compared to a large one. In fact, he can say that it is normal to expect to have more returns by investing in a small company, because the risk is also higher.

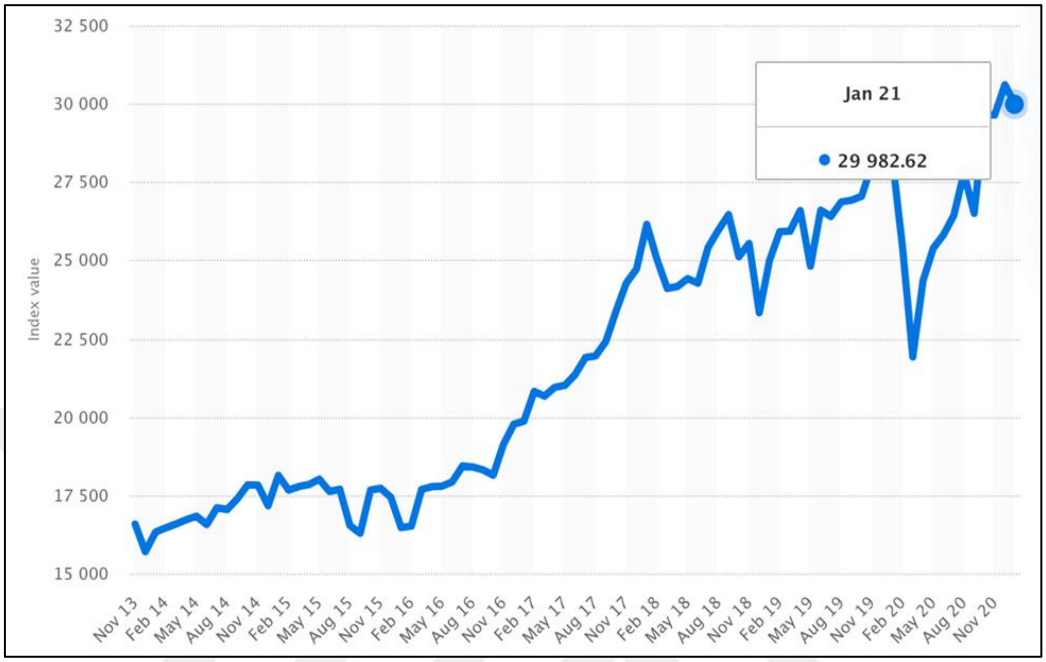

Figure 3.5: January effect in stock market [40]

3.4.5 Yield Autocorrelation Tests

At the same time as the discovery of anomalies, progress in statistical modelling, in particular statistical tests of autocorrelation, have made it possible to test and question the validity of the hypothesis of the random walk of stock market movements. These tests demonstrated the existence of positive autocorrelations in price movements. That is to say, when the price of a security moves in one direction, there is a high probability that it will continue in the same direction. In other words, the direction or trend of the price depends on its historical movements, which is contradictory with the random walk hypothesis. The first test that was developed was that of BOX and Pierce This test consists of examining the residuals (ε) obtained by comparing a series of data to a series of random numbers. If there is an autocorrelation between these residuals, then we can conclude that the series does not follow a random walk. Another approach that is most commonly used to test this hypothesis is based on the Dickey-Fuller test where one seeks to test whether a unit root is present in the tested series All these anomalies and experimental studies come to the same conclusion, that stock prices are auto-correlated and therefore, it is possible to predict the direction of the price of a security based on its historical data [42]. In the next section, we will see the techniques that can be used to make these predictions.

3.4.6 Methods for Forecasting Financial Assets

The price of a stock may be predicted using one of these three techniques. In this context, there are three distinct types of analysis: basic, technical or visual, and quantitative. Although the objective of these evaluations is to provide the same information, they are built on separate foundations and might possibly be amalgamated to increase the profitability of stock trading. Basic analysis adherents examine both microeconomic and macroeconomic aspects in order to give an explanation for price changes. To understand the long-term pricing trend, one must consider both macroeconomic and microeconomic factors, such as inflation and unemployment, as well as sales, debt levels, and profitability [43]. The following table contains these items. Some people have trust in technical analysis because they believe it can give enough information about a company's previous performance, such as its return on investment (ROI), stock price, and trading volume, to predict the company's future course. This argument is used to support their claim that financial time series are an effective indicator of investors' expectations and behaviour. In these series, the tension between supply and demand leads in the production of bullish or bearish patterns. In quantitative approaches, mathematical modeling, econometrics, and other sophisticated computer tools are used to assess and predict the movement of financial series and to execute strategies. All of these components are addressed by quantitative trading: strategy identification and backtesting, order optimization, and risk management [44]. Aside from academic journals, there are not a great deal of publications that emphasize quantitative research methodologies. Consequently, people who are in possession of successful techniques or procedures have little incentive to share them with the rest of the world.

3.4.7 Fundamental Analysis

The great majority of long-term investors begin with fundamental analysis when seeking to assess the value of a firm, a specific stock, a futures contract, or the market as a whole. The objective of this kind of study is to compare the market value of a financial asset to its intrinsic worth (also known as its theoretical value). This comparison may assist investors in evaluating whether or not to invest in the asset in question if the asset is affordable and its underlying value is expected to improve in the near future [46]. However, if the current market value of a security exceeds its real value, that security is regarded overpriced and its value is expected to continue to drop. Utilizing a vast array of financial and stock market data, fundamental analysis provides a framework for evaluating possible investments. Fundamentalists often use one of the four main types of fundamental analysis:

Economic analysis begins with an examination of the global economic environment, which includes the Gross Domestic Product (GDP), unemployment rates, interest rates, and industrial output. The vigor of the economy and its expanding potential are the primary topics of this study.

Before doing a sector study, it is necessary to analyze the economics and competitive dynamics of a particular market niche. This may be accomplished by analyzing a variety of criteria, including market share distribution, sector development, industry rivals, and entry barriers. Because of these findings, we now have a deeper understanding of the subject and can predict its future developments with more precision.

During this phase, accounting records (such as a balance sheet, income statement, and cash flow statement) and financial ratios are analyzed in order to determine the firm's worth (example: profitability, indebtedness)

The price earnings ratio, book value per share, earnings per share, and other stock market ratios are used to establish if the market value of a firm is overvalued or underpriced by comparing it to its theoretical value and the industry in which it works. Accounting experts' preferred techniques for performing valuations [47] The Initial Example: Dividend Revenue Following Discounting It was established by Gordon and Shapiro and is based on the assumption that the price of a share equals the present value of its dividends, which are expected to grow eternally at a rate represented by g.

$$P0 = \sum_{k=0}^{\infty} D0 \frac{(1+g)^t}{(1+r)^t} \tag{3.3}$$

$$P0 = D0 \frac{(1+g)}{(r-g)} = \frac{D1}{(r-g)}$$

3.4.8 The Discounted Cash-Flow (DCF) Model

In our approach, we focus on the cash flows that remain after all costs have been subtracted. These cash flows are referred to as "free cash flows," and they are used to account for future spending on both existing operations and assets. This strategy lowers amounts according to a weighted average cost of capital, which defines how much lenders and investors may expect to make on their investments. Despite the fact that it is not one of the most often used accounting valuation models, many experts and practitioners consider it to be one of the most significant. He also realized that its applicability in the actual world was uncertain [48]. The preliminary investigation's findings Numerous studies have shown the accuracy with which fundamental analysis can anticipate stock prices. The objective of these research was to determine if previous profitability and cash flow may be used in a number of ways to predict future intrinsic value. Financial ratios and the cost of equity may be used to produce accurate estimates of the stock's intrinsic value, which is defined by the value of the company's shares, according to research conducted between 1988 and 2000. The availability of fundamental information may have a substantial effect on a company's stock price. The firms' fundamental analysis is based on the quarterly financial reports obtained from these reports. The graph in Figure demonstrates that Microsoft's share price fluctuated significantly between July 2017 and September 2018. The red and green arrows depict the price movement of the action on the day when the company's results were released. This movement occurred on the day the results were announced.

Figure 3.6: Microsoft stock market since January 2017 up to April

Next the release of the results, there is a significant increase in stock price, which continues the following day (usually after the conclusion of trading or just before it). This trend often happens after the market closes for the day or just before. It is not straightforward to see that the price at the start of the next day is significantly different from the price at the end of the day when the findings were released. On October 26, 2017, Microsoft reported earnings of $0.84 per share. According to industry experts' forecasts, the company should have made $0.72. Surprise! This is a 17 percent increase over the previous year. The closing price of that day's trading session was $78.76 per share (just before the earnings report was released) [49]. The next day, the stock started trading at $84.37, a rise of more than 7 percent. When there is good news, or even news that exceeds analyst estimates, investors react swiftly to purchase the stock; nevertheless, when there is bad news, or news that falls short of expert projections, investors move swiftly to sell. The price of stocks fluctuates in response to announcements on macroeconomic issues. Inflation, monetary policy, and interest rates are the key subjects of debate in these remarks. These changes will have an impact on the whole market. In proportion to how closely a security is tied to the market, the effect will be amplified. Despite this, macroeconomic updates and statements are issued on a regular basis. Therefore, using them in intraday price calculations is unnecessary. On the other side, announcements may significantly enhance the volatility of the market. This suggests that evaluating the fundamentals of a financial asset alone will not enough to estimate the price movement of that asset in the short term [50]

3.5 TECHNICAL ANALYSIS

Technical analysis, sometimes referred to as chart analysis, evaluates a company's historical data, in contrast to fundamental analysis, which focuses largely on the company's fundamentals. The goal of technical analysis is to determine the ideal timing to purchase or sell a particular investment. This may be accomplished by identifying the price's trend direction. Relevant are just price and volume changes that happened in the past. Technical analysis is a technique that seeks to predict the future value of stocks by evaluating charts of current stock prices. According to AT, stock prices represent all already available information, and they fluctuate when new patterns emerge. The concept behind AT. As long as the established pattern persists, it is more likely that future prices will adhere to the pattern than that they would deviate from it. The mismatch between supply and demand for goods is the fundamental factor influencing both price rises and declines. Although the research of a financial report's underlying numbers might be arduous and time-consuming, technical analysts focus mostly on price and volume fluctuations. Utilizing prior price and volume data, these statistical models provide signals. These indications may then be used to anticipate the future direction of the market. In addition to stock markets, technical analysis incorporates futures markets, currency markets, interest rate markets, commodity markets, and any other financial market with a data flow [51]. In technical analysis, indicators generated from stock market data, such as price and volume, are used. It is possible to aggregate granular data based on the required frequency of analysis. In contrast to fundamental analysis, which is dependent on long-term data and can only be used for long-term projections, technical analysis depends on short-term data due to its availability and granularity. Combining fundamental research with technical analysis is a valid approach for determining the worth of long-term investments.

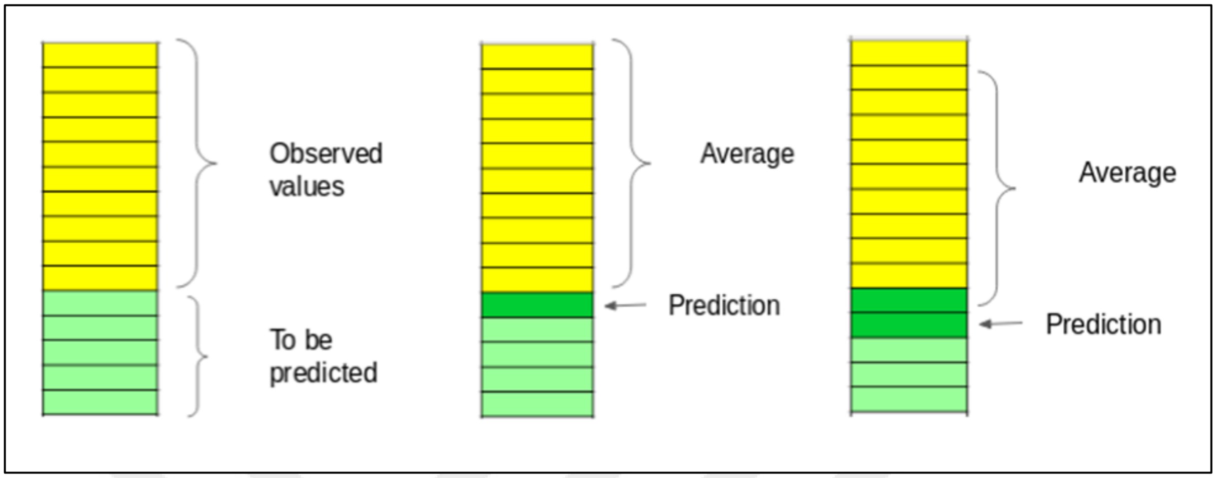

Figure 3.7: Technical analysis scheme for stock market.

3.5.1 Seasonality

In the context of time series, "seasonality" (abbreviated "St") refers to the periodic effect that appears at preset intervals. Seasonality may occur at any time of day or week, as well as daily, weekly, monthly, and annually. There is also the possibility of seasonality repeating itself.

Due to taxes, investors are urged to sell their losses at the end of the year so they may purchase them in January and complete the year with only profitable lines in their accounts [53].

In addition to long-term seasonality, there is also something known as short-term seasonality. For the "Turn of the Month" effect, it has been shown that returns are higher on the three days prior and the three days after the end and beginning of the month, respectively, and that yields are, on average, lower on Mondays than on other days of the week.

Due to the nature of the project, it is more essential to address the intraday seasonality of this project. In fact, research investigating the seasonality of the equities, futures, and exchange rate markets during the same trading session have been done. Seasonality may be explained by the microstructure of the market, the dissemination of news, as well as the behavior and psychology of market players. On intraday stock market charts, it is possible to gain a good image of the seasonality of transaction volumes with no effort [54].

To demonstrate this, we may split the trading day into the following three distinct time periods:

The morning sessions run from 9:30 a.m. to 12:00 p.m., the lunch sessions run from 12:00 p.m. to 2:00 p.m., and the afternoon sessions run from 2:00 p.m. to 4:00 p.m.

3.5.2 Irregular Fluctuations

This portion of the time series cannot be explained by seasonality or trend alone. These events are characterized by variation that is both random and weak in strength. There is also the option of alluding to threats, sounds, or waste. Other unforeseen outcomes, such as a stock market crash, terrorist attacks, or political crises, are possible [56]. Even though they may not occur often, these events have the ability to drastically alter the course of a financial series. In general, they are difficult to predict and need the use of specialized models for successful forecasting. Stationarity is an essential attribute to possess while studying time series data. It is important for these series to be stationary in order to be modeled. In other words, they display no discernible patterns and are unaffected by seasonality. Analysts of time series often operate under the assumption that stationarity is unreliable. Unfortunately, this is not always the case under turbulent financial circumstances. Nevertheless, by using the idea of differentiation, a time series may be turned into one that is only weakly stable.

3.5.3 Decomposition of a Financial Time Series

After seeing the main components of a time series, it will be wise to study the behavior of a time series by isolating trend and seasonality. These two operations are called series relaxation and seasonal adjustment. Once these components are eliminated, we obtain the random series ϵt of irregular fluctuations (residuals).

The seasonality shows a recurring decline during the first months of the year and a rise at the end of the year. Accentuated downward irregular fluctuations are observed at the end of 2008, the end of 2011 and in the middle of 2013. It should be noted that the models of decomposition and smoothing (ex: moving averages) can be used in the prediction of the trend . Although we have explored this track, the objective of our project is to test machine learning methods. This is why we will limit ourselves to the statistical analysis of the behavior of a financial time series and not to their use in the prediction [58].

3.5.4 Data for the Prediction of Financial Series

There is an abundance of available financial data that may be utilized to predict the movement of a financial asset. The section on forecasting approaches includes both transaction data and basic financial data relevant to market, industry, and firm circumstances. In an attempt to anticipate the price movement of an investment on the stock market, analysts study price and volume-related technical data [60].

3.6 OVERVIEW OF ALGORITHMIC

Trading In the previous sections we have seen forecasting techniques that can be used for the prediction of price trends in financial assets. Indeed, it is possible to automate the trading of financial assets by creating a system that is based on rules defined a priori. When these rules are satisfied, the system sends a buy or sell order depending on the signal that is generated.

a. using fundamental analysis: if the theoretical value of the security calculated, using fundamental ratios or fundamental valuation models, is higher than its market value, then the security is undervalued and could represent a good potential for growth. growth in the future. It should be noted that this kind of rules can be used for the long term or for the selection of securities on a fundamental basis.

b. using technical analysis: technical indicators as seen in the previous chapter can be used to generate market entry and exit signals.

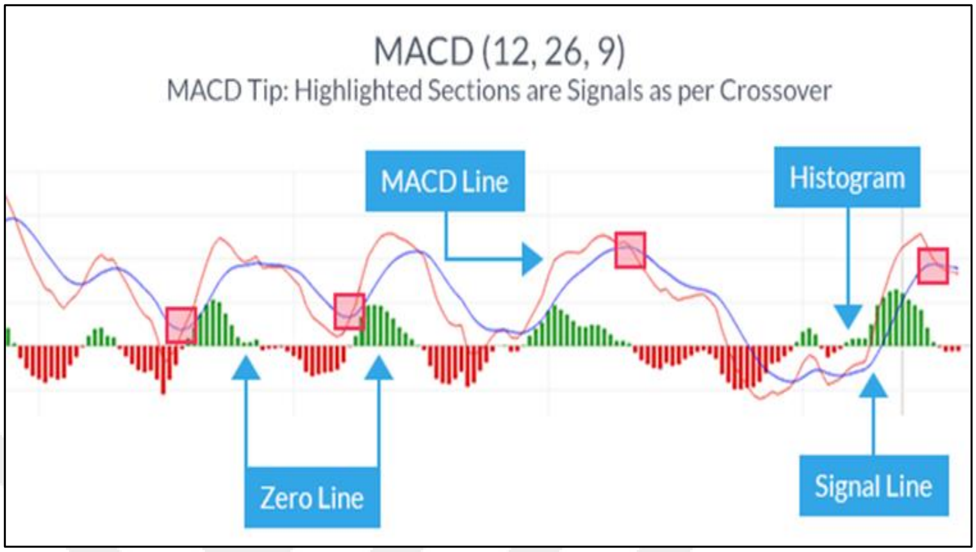

Figure 3.8: Stock market technical indicators.

The Bollinger Bands are an intriguing illustration in this context. In order for such a system to be effective, it must be able to accept the changing characteristics of financial time series and develop winning rules for each and every asset or market (e.g., volatility, non-stationarity, conceptual drift). The most major constraint of a rule-based system is the limiting number of variables and rules that it can properly handle [61]. Verifying the performance and efficacy of a large number of elements and regulations that may be evaluated without the use of modern quantitative methods is exceedingly challenging. This necessitates a system that is dynamically scalable and capable of learning and adapting to changes in investment strategy when new data is introduced. This chapter will discuss the principles of machine learning and the algorithms that may be employed to get insights from acquired data. This article will continue with a discussion of the most popular algorithms, including decision trees, vector support machines, neural networks, and ensemble approaches.

3.6.1 Types of Automatic Learning

The subject of artificial intelligence known as machine learning focuses on the process of building computer systems that can self-learn based on data and prior experiences, or by interacting with their "environment" (ML). Machine learning's ability to "learn" and "adapt" to new data without needing a priori programming is one of the most crucial and effective aspects of this science. There are several methods for machines to automatically learn, and the approach utilized depends on the unique issue at hand and the available data. Figure 3.9 depicts some of the machine learning methods that are used most often nowadays [62].

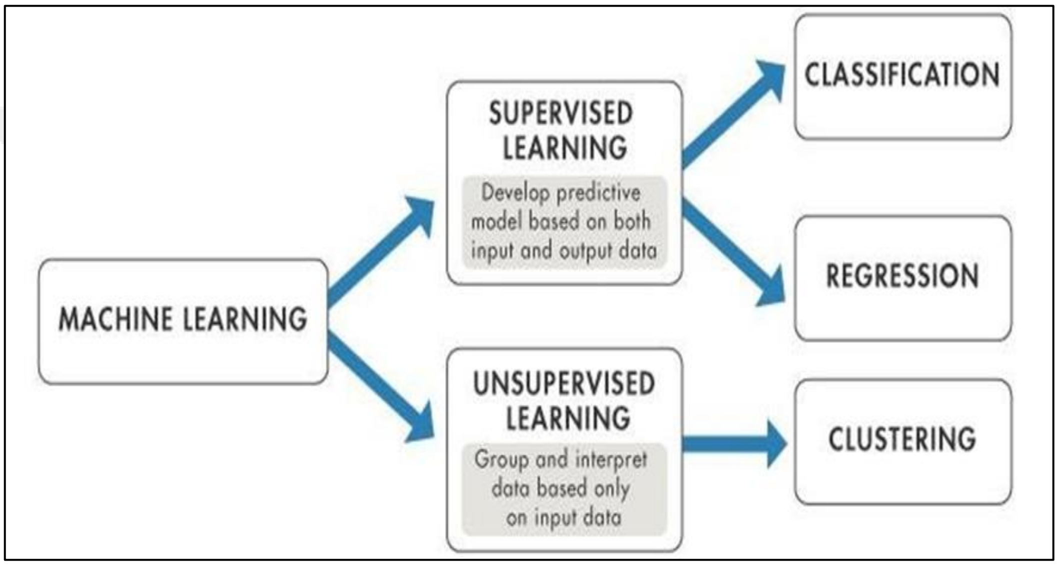

Figure 3.9: A Summary of the most common types of machine learning [62]

3.7 SUPERVISED LEARNING

Using a database of real-world events that have been researched and evaluated to guarantee their applicability as instructional examples, the technique is refined. The primary purpose is to perform an examination of correlations between the variables that act as inputs and the outputs, while the secondary objective is to obtain information about inputs that are related with outputs for which no information is available.

$$\text{J} = \sum_{i=1}^{n} \frac{(mX_i + c - Y_i)^2}{n} \tag{3.4}$$

mX denotes the total number of instances, mY represents the total number of examples, yn represents the total number of examples, etc. X is a collection of the aforementioned characteristics (discrete or continuous) There are two unique types of Y output values: continuous and discrete. In the context of supervised learning, two distinct kinds of tasks exist [64]:

The anticipated target variable has discrete values ranging from -1 to I-1. This is equivalent to classifying each individual thing (or label). This is the scenario that we should be aware of if we are attempting to predict the future movement of an asset (high, neutral, low). This specific study included six numerical indicators and one emotional indicator (categorical variable). The date will just serve as an index in our modeling table; it will not be included into the model itself.

3.7.1 Unsupervised Learning

For this type of learning, the learning database does not contain a target variable (as we have seen in supervised learning). There is only one set of data collected as input. The algorithm must discover the structure on its own based on the data. This technique is used to partition data into groups of homogeneous elements. Distance is often the most used as a measure of similarity between groups

3.7.2 Semi-supervised Learning

This is a hybrid technique that blends supervised and unsupervised learning by employing data from the same labeled and unlabeled observations-containing dataset. The procedure of data tagging may be costly and time-consuming; hence, this choice is better. Additionally, there is the risk of human bias in the labeled data. In this situation, semi-supervised learning is an excellent option so long as there are just a few label options. In addition, including a substantial amount of unlabeled data into the training process has the potential to increase the model's performance while decreasing the amount of time and money required to build it [65].

3.7.3 Reinforcement Learning

Learning takes place without supervision, by interaction with the environment (principle of trial/error) and by observing the results of the actions taken. Each action in the sequence is associated with a reward. The goal is to determine the optimal behavioral strategy to maximize the total reward. For this, a simple return of the results is necessary to learn how the machine should act. This is called the reinforcement signal. It can be very beneficial for high frequency financial forecasting where the environment is dynamic and as a result it is difficult to find or automate effective strategies manually [66].

3.8 SUPERVISED LEARNING ALGORITHMS

3.8.1 The K Nearest Neighbors Method

The statistical technique known as KNN, which stands for "k-nearest neighbors," may be used to both classification and regression studies. This nonparametric strategy is often the method of choice for doing nonlinear financial forecasts. This preference may be explained by the following two key factors:

Utilizing a variety of tests, it has been shown that the KNN technique has a substantial predictive capacity. first, the method's reduced algorithmic complexity relative to alternative global techniques, such as neural networks and genetic algorithms

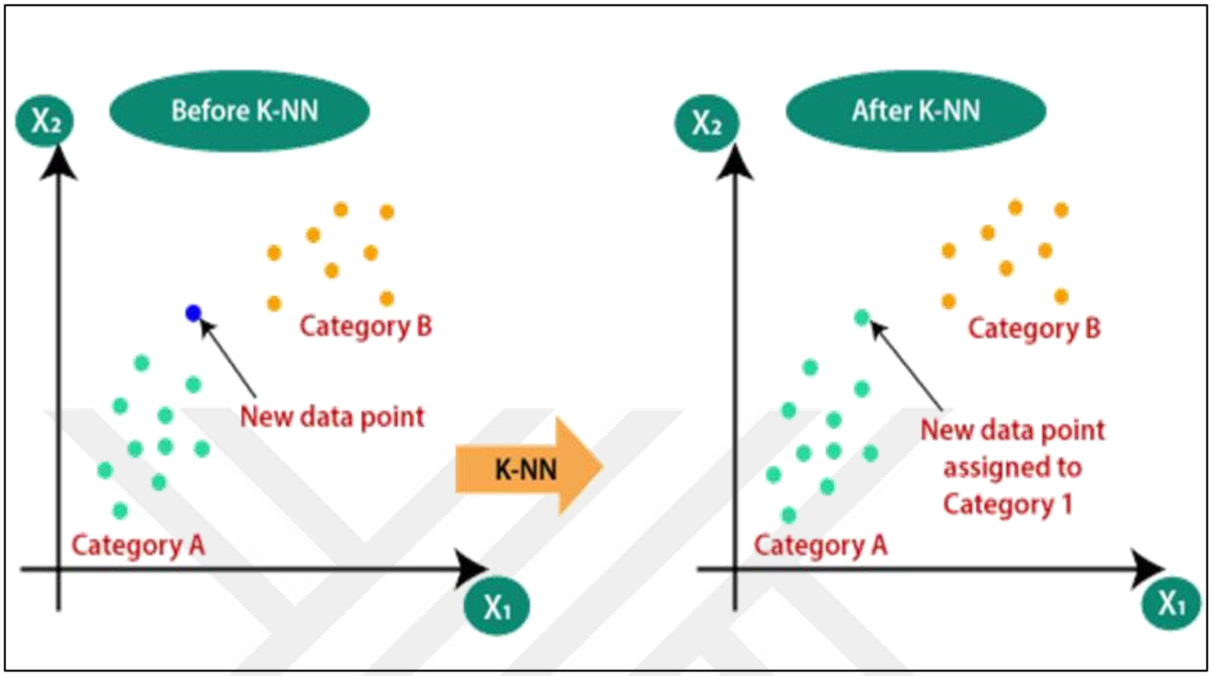

Figure 3.10: Dynamic Time warping in time series forecasting.

3.8.2 Logistic Regression

Logistic regression is a predictive model that seeks to find associations between a vector of random explanatory variables, of a numerical or categorical nature, (x1, x2…,xk), and a binary or multinomial categorical dependent variable Y. This is simply a non-linear transformation of linear regression, where one seeks to predict a class instead of a continuous numerical value. To do this, we use a logistic function also called 'Sigmoid', to return the probabilities that are used to separate the classes to be predicted. Example: if the probability generated by a logistic regression, to predict the trend of a security, is greater than 0.5, then the predicted trend will be bullish, if not, it is bearish. This model produces a logistic curve, which is limited to values between 0 and 1 [67].

3.8.3 Support Vector Machines

The Support Vector Machine is a tool for solving classification and regression problems (SVM). On the other hand, this function's primary aim is categorization. If we are given an Ndimensional space, for instance, we will look for a margin hyperplane that appropriately divides the data and is as far away from all of the observations as is realistically feasible. The goal is to reduce the likelihood of forming inaccurate generalizations. The separation between the separator plane and one of the training instances is minimal [68].

3.8.4 Neural Networks

The artificial neural network (ANN) contains a set of highly interconnected artificial neurons inspired by the biological neurons of the brain. The goal is to imitate certain functions of the human brain, such as memorization by association, learning for example, etc.

A. Network of biological neurons

The brain is composed of a large number of biological neurons, each composed of a cell body, dendrites and an axon (see Figure 3.11). These neurons are interconnected and thus create a larger neural system. Connections, called synapses, are made between the dendrite of one neuron and the axon of another. When a neuron receives a signal via the axon, it activates or not this signal, depending on its strength, and transmits it via its dendrites to all the connected neurons [69]

B. Artificial neural network

A component of artificial neural networks consists of input and output layers as well as (in most instances) one or more hidden layers of units that convert the input into something the output layer can utilize (ANN). It is vital to choose proper weights for the connections between an ANN's input, masked, and output nodes in order to limit the overall number of errors it generates. The computations are accomplished by adding the weighted sum of the inputs, where the weighting is a function of the connection quality).

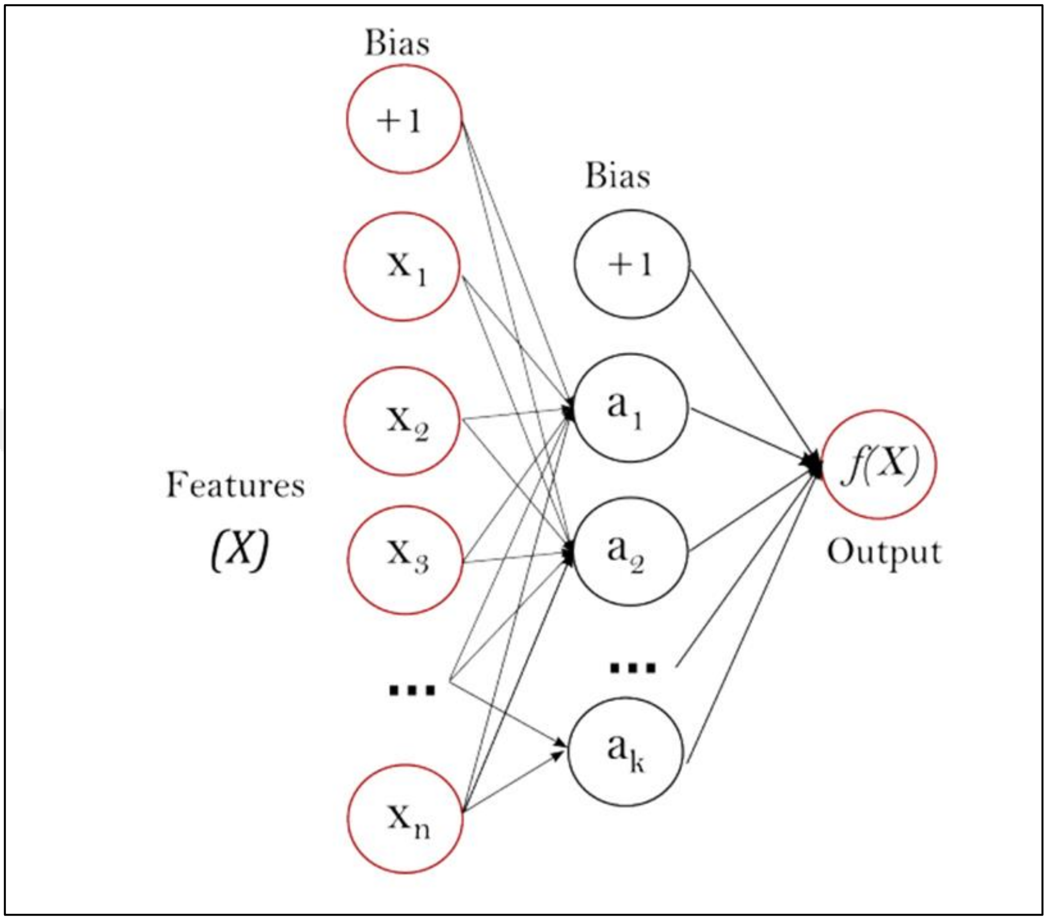

Figure 3.11: ANN structure.

This approach is defined by [70] as an unsupervised, multi-layer perceptron methodology that may be trained on any given data set. This is the case regardless of the number of variables in the input data or the number of output dimensions. The purpose of this study is to categorize the data; hence, the model must be capable of training a nonlinear approximation function for classification as well as regression. The non-linear layers existing between the input and output layers are referred to as "hidden layers," and this term is used to characterize these layers.

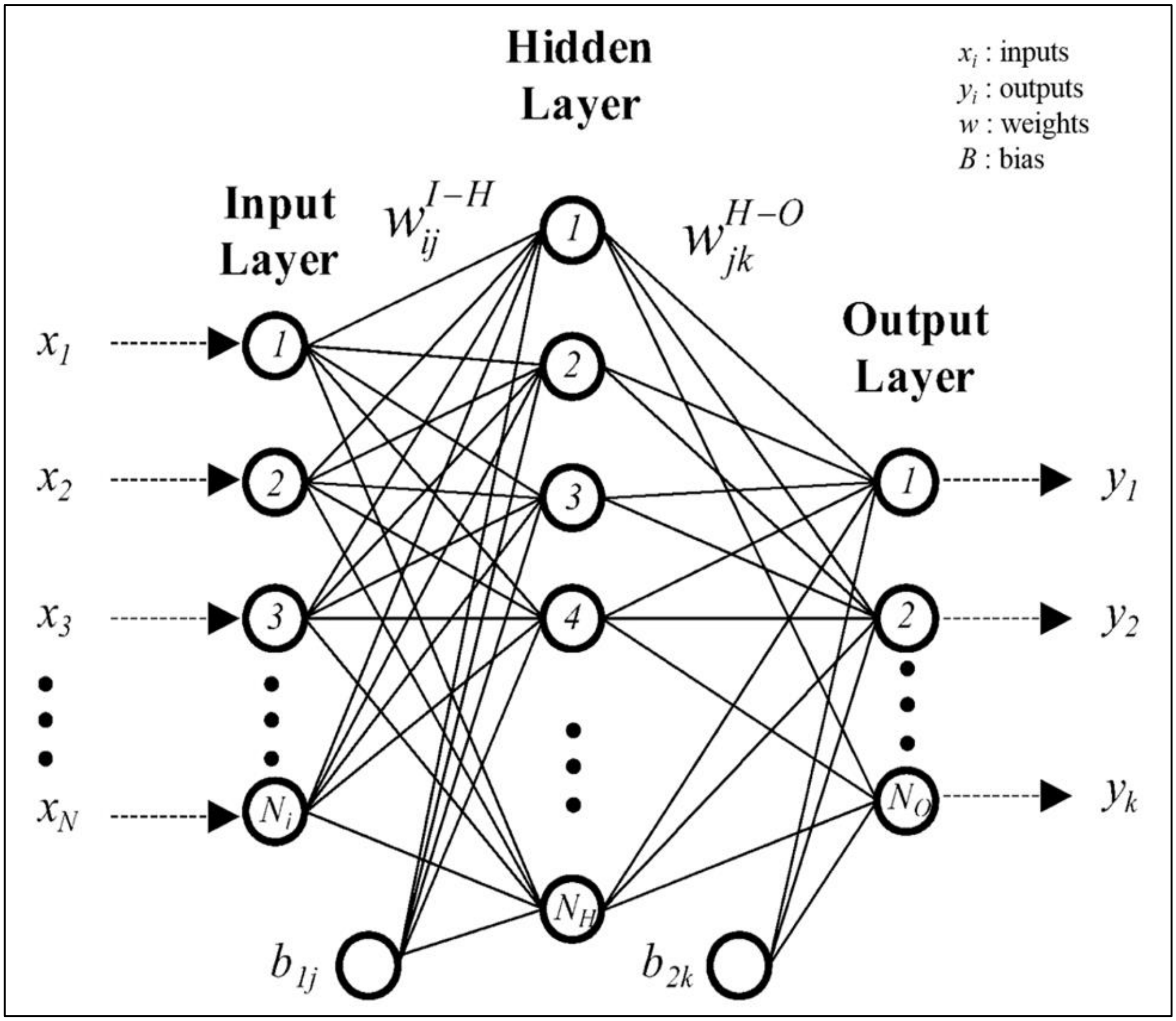

Figure 3.12: Weights learning function in MLP-ANN

The MLP is illustrated from Figure: Each circle in the input layer contains what is called a neuron {xi |x1, x2, ..., xm} and each circle in the hidden layer represents a neuron in the hidden layer and transforms the values of the previous layer based on a sum with weights w1x1 + w2x2 + ... + wmxm, followed by a non-linear activation function Logistic regression can be defined as a technique to calculate the parameters of a logistic model, ie , a model that uses logistic functions to model the value of a dependent binary variable. Thus, this technique is widely used for classification problems whose classes to be predicted can assume 2 possible values, such as a decision to buy or not buy shares in the financial market, scope delimited by the project. illustrates the steps for predicting the output by the algorithm. As in MLP Logistic Regression applies weights to input variables to find a calculated sum of them. Then, instead of going through hidden layers, that sum goes through a attaching function that returns an associated probability. This probability, finally, passes through a step function that classifies the output based on a threshold [71]. In the logistic model, the hit logarithm (referring to outputs classified as "1") is a linear combination of one or more binary or continuous independent variables, which are the model's inputs. The probability corresponding to the value "1", can vary continuously between 0 and 1 and the function that translates this probability to the final class is called a logistic function.

4 PROPOSED METHOD

4.1 CHAPTER OVERVIEW

The overall design of the system help with analysis will allow the development of new strategies. Below are the development stages of a new strategy:

a. Development of the strategy

   The analyst developer will have a development environment and a data access library available. He can base his strategy on simple mathematical formulas or rely on the results of other strategies. For each strategy, the developer analyst will choose the data that will be executed on his strategy. He may choose to execute it on a single or several actions and on part or all of the prices.

b. Loading the strategy into the analysis system

   When loading the strategy, a copy will be saved in order to keep a history. For each strategy, all the parameters associated with it will also be saved.

c. Executing the strategy

   When a strategy is executed, the parameters and data are provided to it as they occur. Once executed, its results are saved. These results are the different buy/sell decisions that it will have deducted as it is executed.

d. Generation of a strategy performance report

   The generation of the report will take place once all the parameters have been passed on the strategy. It will provide the details on the execution specifying the strategy which provided the best result according to the parameters.

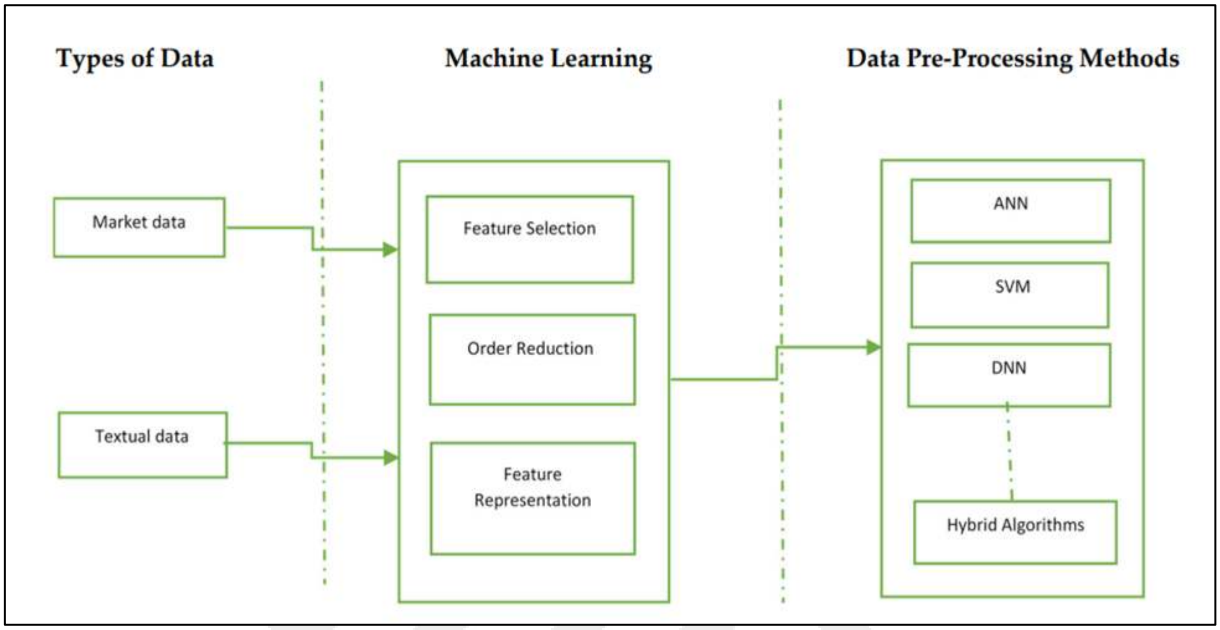

Figure 4.1: Stock price model with a non-linear rise weekly

The loading and execution of a strategy will be done directly by the development environment. Thus, no additional technical knowledge will be necessary for the developer. When the developer develops a strategy, he specifies the time range over which he wants his strategy to be executed, the time gap (second, minute, hour, etc.) and the type of underlying. Thus, when executing the strategy, the analysis module will first retrieve the parameters of the strategy. It will load the underlying from the database (currencies, stocks, bonds, etc.), then prepare the strategy execution environment. Then it will save a copy of the policy in case another developer wants to use it later. It will save the settings, then start the actual execution of the strategy. For each signal provided by the strategy, the analysis support module will save the signal (Buy/Sell), the price, the underlying (ex: EUR/USD currency, BNP share, etc.), the spread of time (e.g.: minute, day, month, etc.). Then it will associate this backup with the policy and settings. Thus, the analysis assistance module will have an entire copy of the strategy and will be able to rerun it at any time or update it. Once executed and the results saved, the analysis assistance module will generate a report in PDF format with the various data necessary to evaluate the strategy. The developer will then be able to find out how to optimize his strategy or submit it to his managers for production. This report will contain data concerning the overall execution of the strategy (execution time, number of executions in relation to the parameter), details concerning the best result obtained (the graph and with which broker it was the most profitable, and the total percentage), then the other results classified by performance.

4.2 SYSTEM MODELLING

The neural network module will be the most complex module to implement. It requires a good understanding of the subject of neural networks and in particular the technical aspects of neural networks. However, market prediction systems being very poorly documented in general, we will mention any technical decisions that could be purely arbitrary. As described in the functional specifications, this module will have a neural network. The type of this neural network will be supervised learning and back propagation. faced with two problems concerning the stock market price: - Recognize the historical stock market price that is closest to the current price. - Generate a probable sequence. Here is a simple stock market price representing a drop followed by a price increase illustrated by the following figure: Monday the share is worth 40, Tuesday it is worth 20, Wednesday it is worth 40, Thursday it is worth 60 and finally Friday it is worth 80. If we put a correspondence table as for character recognition of the type: Monday=40, Tuesday=20, Wednesday=40, Thursday=60, Friday=80 It will therefore be possible from this correspondence table to find in the stock price history the model corresponding to this stock price. The following week, the stock price is represented by the following figure:

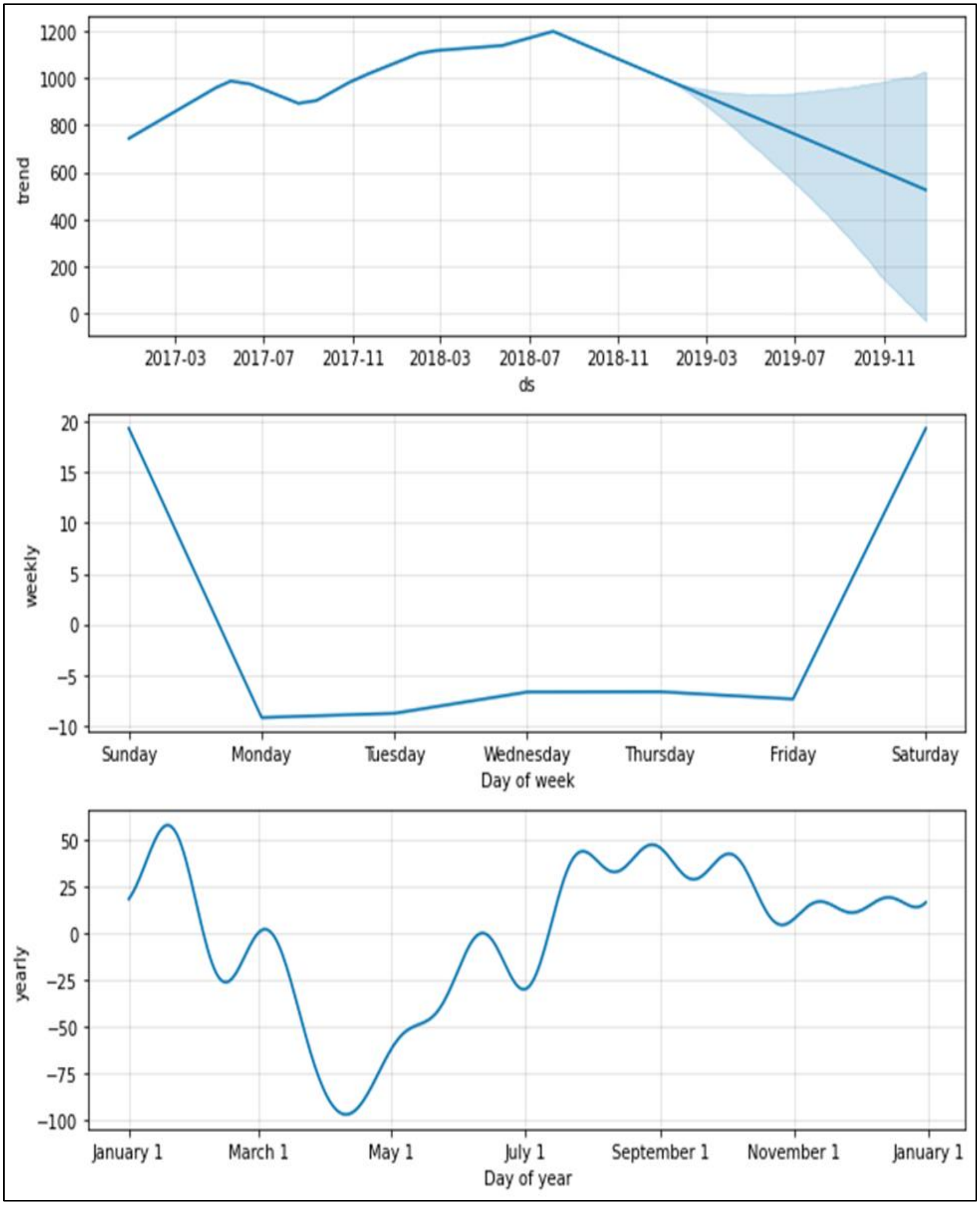

Figure 4.2: Stock price model with a non-linear rise weekly.

4.3 PROPOSED ALGORITHM

The models used for training and forecasting the data set are defined below.

4.3.1 MLP (Multi-Layer Perceptron)

It is possible to learn the function f(): Rm - Ro from an input data set with m dimensions (variables in the training set) and from an input data set with m dimensions (variables in the training set) (variables in output). The purpose of this study is to categorize the data; hence, the model must be capable of training a nonlinear approximation function for classification as well as regression. Non-linear layers that exist between the input layer and the output layer are considered hidden layers. There may be one or more of these layers. The below illustration shows the MLP: The circles in the input layer indicate neurons in the hidden layer. Each circle in the input layer represents one of these neurons in the hidden layer and affects the values of the layer below it using a sum of weights consisting of w1x1 + w2x2 +... + wmxm, followed by a nonlinear activation function: An MLP technique that employs a hidden layer between the input and output layers to produce the model's parameters may be compared to a logistic model, which use logistic functions to represent the value of a dependent binary variable. Because it incorporates logistic functions, a logistic model is categorized as an MLP technique. Due to the limitations imposed by the project's scope, this approach is often used for classification issues in which the classes to be predicted may take on two distinct values. This would include the choice to purchase shares on the financial market rather than pass up the chance to do so. demonstrates the method the algorithm employs to predict the outcome. Logical regression assigns weights to each variable while computing the sum of the input variables. The total is then sent by an attachment function, which, rather than traversing any hidden layers, gives a linked probability. The output is ultimately categorized using a step function that does so based on the probability [23]. The logistic model takes as inputs one or more independent binary or continuous variables, and the hit logarithm refers to outputs categorized as "1." If the chance of "1" is interpreted as a logistic function, then it has the capacity to fluctuate between 0 and 1 at any given moment.

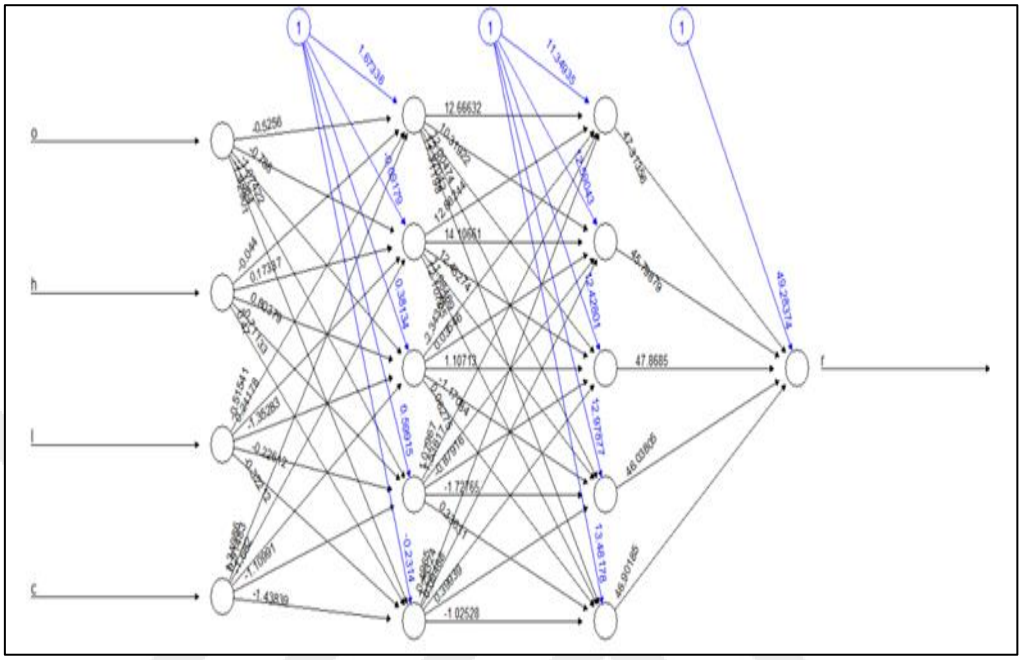

Figure 4.3: ANN structure for the prediction model.

The number of neurons contained in the input and output layer will be decided according to the investment strategy that we will adopt. Thus, for an investment strategy lasting several days, the number of input prices must be sufficient for the neural network to find the optimal model. We specify the configuration of these layers in the paragraph on learning. The hidden layer is going to be the core of the neural network. It is this layer that will perform most of the calculations. It will be composed in the same way as the other layers of a set of neurons. It will take data from the input layer as input and send these results to the output layer. The number of neurons contained in this layer is not fixed and does not depend on external criteria such as the investment strategy. This number of neurons is evaluated without however justifying it between half of the neurons of the input layer and its double. To successfully evaluate the number of neurons needed for the hidden layer, we will take the example given In their example they distinguish the course model from the ideal course. The price pattern is a set of strategic points of the ideal price. The following figure illustrates our example. In the two graphs A and B the ideal course is represented by dotted lines, the model by circles and the prediction by a continuous line. In the neural network that generated the prediction of graph A (left), the hidden layer contains 5 neurons. The hidden network layer of graph B (right) has 20 of them. We can notice that on graph B, the prediction is in perfect agreement with the model represented by the circles. But it is very far from the ideal. This effect is called overtraining. This case happens when the training models contain a lot of noise, and the neural network tries to reproduce this noise as much as possible rather than giving an approximation.

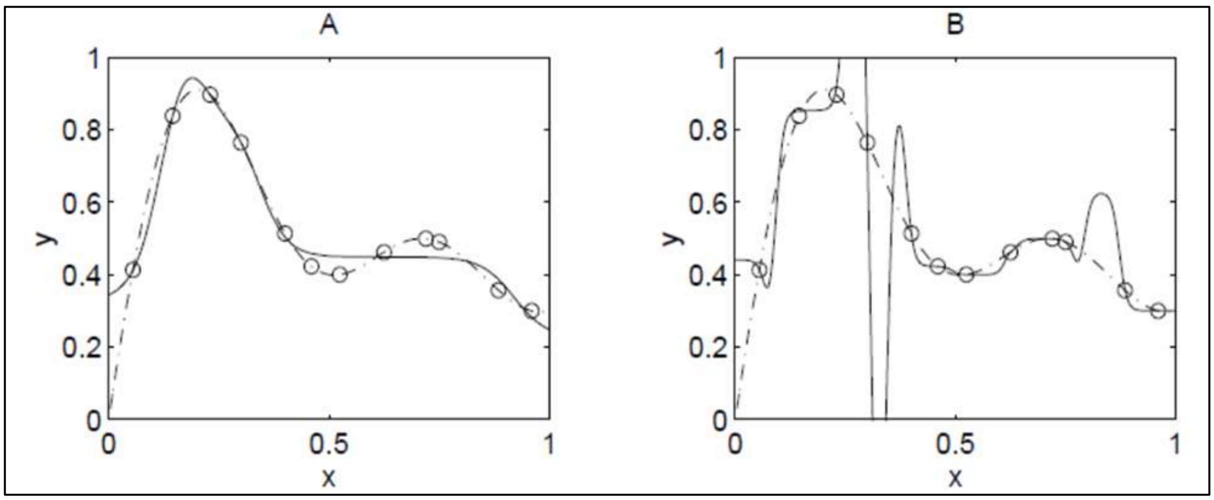

Figure 4.4: Where a is the actual values and B are the predicted values using ANN

Some neural networks do not realize the sum of the products but the product of the products such as “Sigma-pi” networks. This type of network is interesting for High Order Neural network type typologies which, should be the subject of in-depth study. The second function of the neuron is the activation function. Its role is to decide whether it should transmit the result of the summation function. This function corresponds to the nerve impulse of a biological neuron (or action potential). This action potential of a biological neuron will transmit the sum of the "graded potentials" of the synapses (summation function) if this sum exceeds a threshold of excitability of the neuron. Below this threshold, this sum is not transmitted. The activation function will act in the same way. As we saw in the functional specifications chapter, several activation functions exist. The most common is the sigmoid function

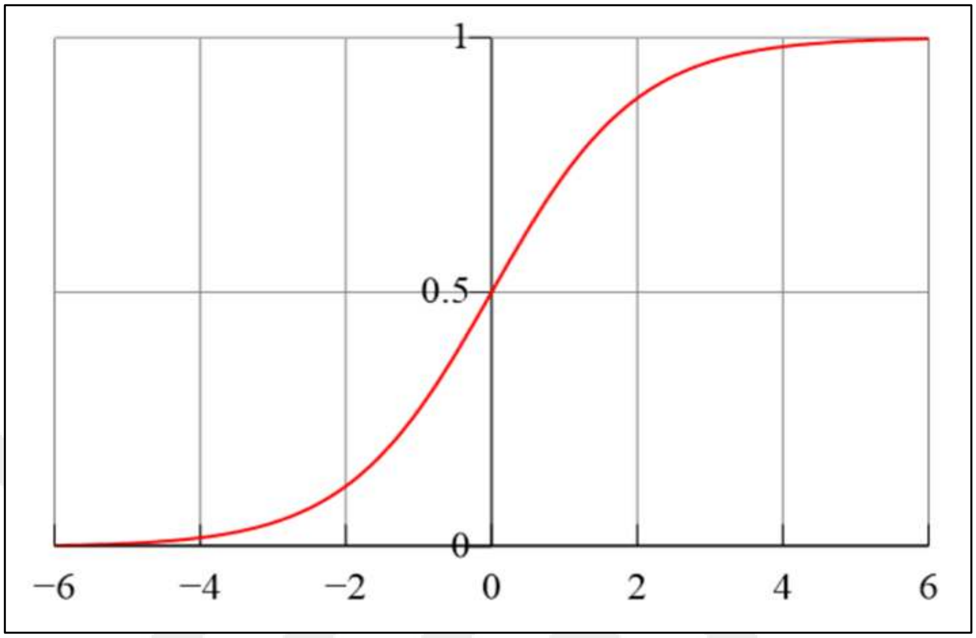

Figure 4.5: Applying sigmoidal function in the ANN prediction

The goal of the system we are developing will be to achieve maximum profit over a very short period. It is over this period of time that we will have to establish our paradigms in order to train the neural network. suggest some principles to define these paradigms:

i. Clean the data.

ii. Scale and seasonally adjust data.

iii. Use an appropriate method to choose a good starting point.

iv. Use a specialized method to avoid the local optimum.

We have therefore adapted them to our needs by carrying out these steps:

a. Define a period of time.

b. Split the stock price history into a sample of a duration equal to the previously defined period ("choose a good starting point").

c. Remove the noise from these courses ("clean the data").

d. Train the network with these paradigms ("avoid the local optimum").

We need to define a period of time, in order to find the models over this period. This period of time can be defined in two ways: by strategy (we only want to invest in the very short term to reduce the risk) or by carrying out tests using the neural network (for example a period of 10 minutes may contain more similar behavior than a 2 second period). We will initially choose to choose this period strategically, because the number of tests to be carried out is relatively large). For the second point, we break down the history of stock market prices by period to create models. The following graph illustrates the price breakdown over a period of twenty minutes:

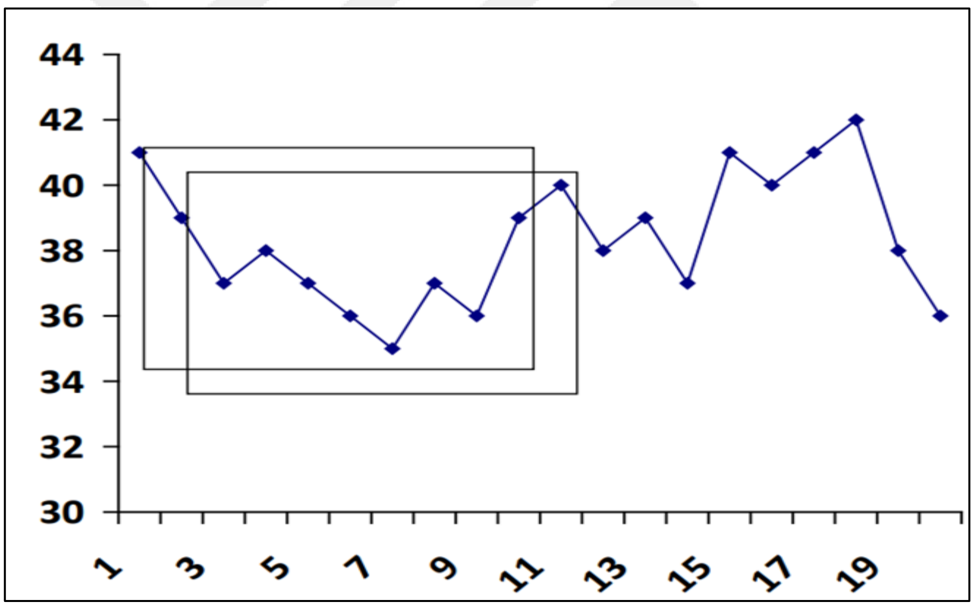

Figure 4.6: Time frame chosen for the analysis.

4.3.2 Validation Techniques

This section describes the validation techniques used for the project. Sliding Windows As the problem is a time series, the data set was divided into training and testing in such a way that, necessarily, the test set was composed of data obtained chronologically after the training data. In this way, it was possible to obtain forecasts closer to the real scenario, in which the model input obeys a temporal order. In order to obtain the best possible use of days used for training and forecasting the data set, The technique of training and prediction sliding windows was used. In this strategy, an N number of days is used to form the training set and an N0 number for the test set to perform the validation. In the next iteration, the same number of days is used for both windows, with the difference that the training and forecasting window has the beginning and end of the previous beginning and end. can be used to illustrate the explanation.

Expansion of Sliding Windows In order to find the optimal number of training data needed for the model, an iteration was performed such that, at the end of the cycle of sliding windows of N days of training, another cycle was started with N + 1 days for the training set. The Python programming language was used to implement this logic Sliding windows with 4 days used for the training set and 1 day for the test set

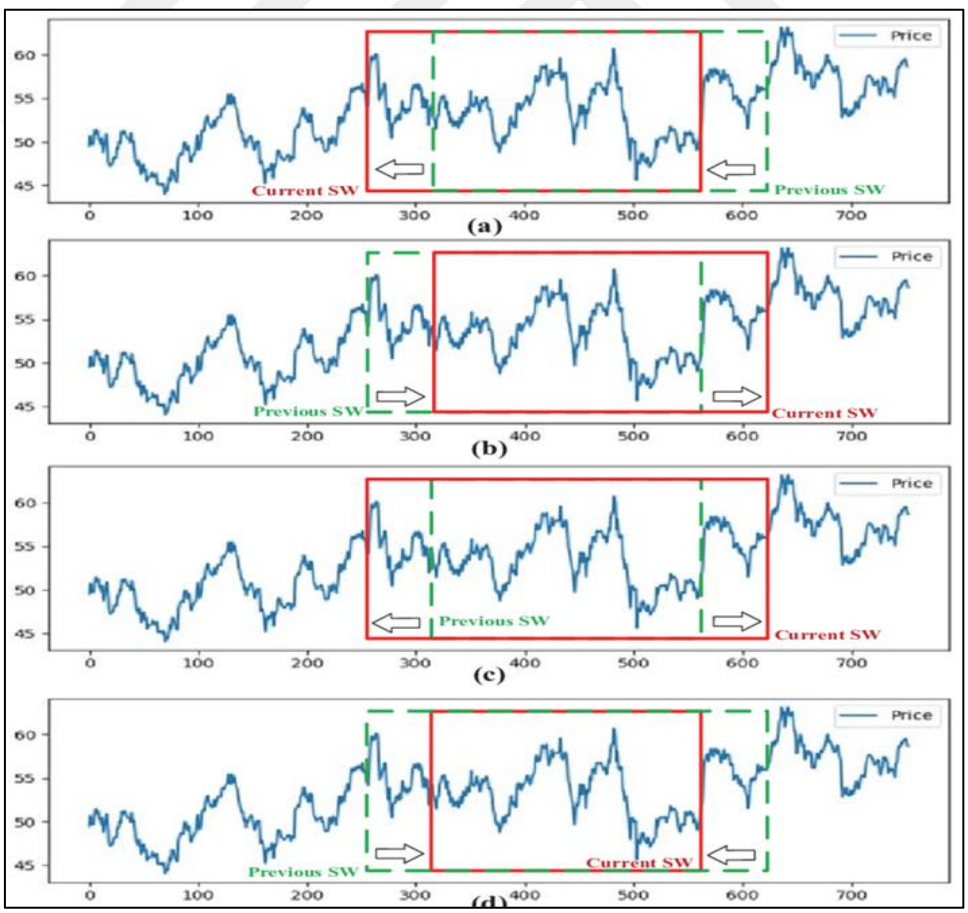

Figure 4.7: Sliding window for the predicted fluctuation timeline.

4.3.3 Validation Parameters

For each prediction made by the model, it is created a confusion matrix (5.4) before extracting the parameters used to measure the performance of the model: Confusion matrix With the confusion matrix, you can more directly visualize the predicted values in relation to the true values. Thus, the number of true positives and true negatives in relation to the rest of the set is of interest for the performance of the model. A true positive is when a class classified as "1" in the test suite actually has the true value "1". A false positive, in turn, refers Analogous thinking can be applied to true negatives and false negatives. For the domain of work, the obvious interest is a model that is able to obtain more hits than mistakes to make a profit and, in addition, identifies the greatest possible number of purchase opportunities. In this way, the confusion matrix can be reused to generate the validation parameters below. Recall (Sensitivity) Recall is defined as the number of values correctly classified as true out of the total number of true in the test set. It can be calculated by the formula below:

$$\begin{aligned}
Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
Precision &= \frac{TP}{TP + FP} \\
Recall &= \frac{TP}{TP + FN} \\
F1\text{-}score &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\end{aligned} \tag{4.1}$$

Where R is recall, FN is the number of samples classified as false negatives, and TP is the total positive values of the test set. Precision is defined as the number of values correctly classified as true out of the total number of predictions classified as true in the test set. It can be calculated by the formula below:

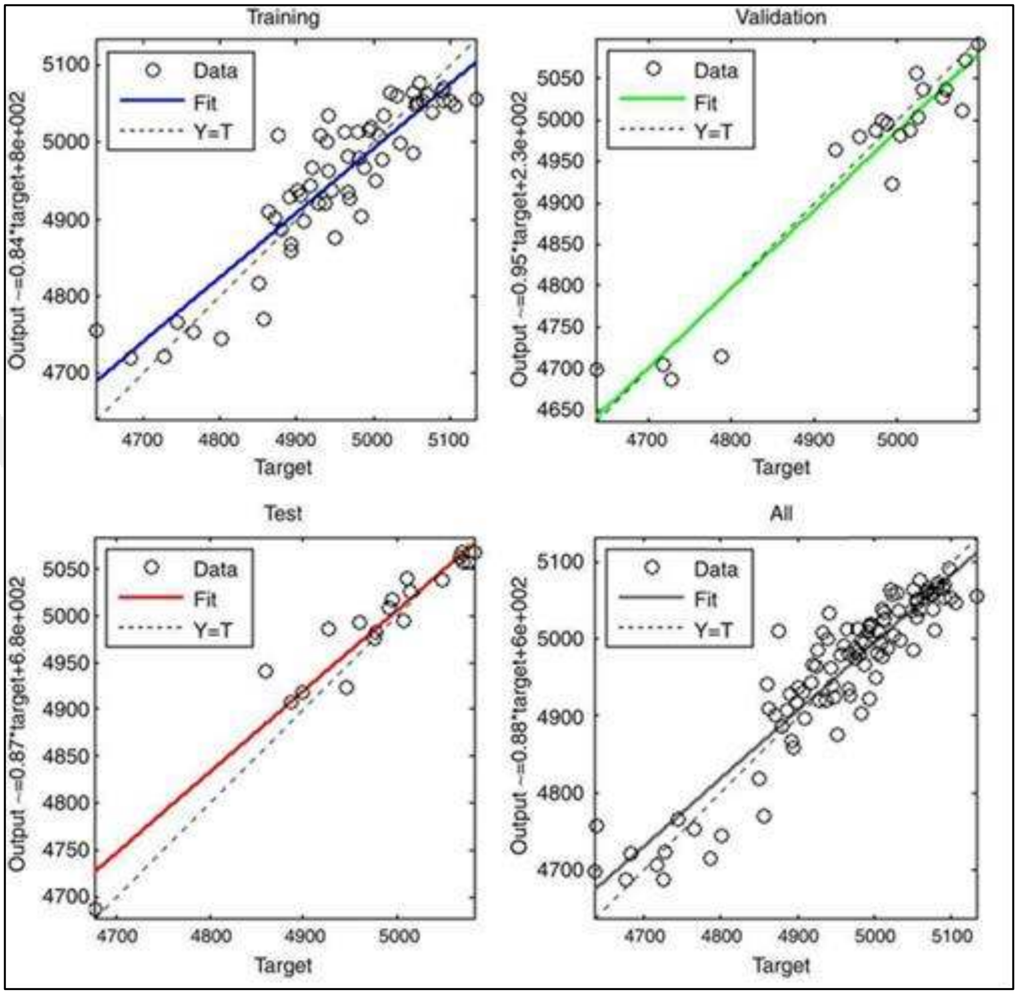

Figure 4.8: Accuracies and Recall in the graph fitting applied using ANN

Where P is precision and VP and FP are classified as samples classified as true positives and false positives, respectively. For each cycle of a given training window, the accuracies and reallocations obtained are averaged in order to obtain the best combination of training days. The higher the average of the accuracies, the higher the hit rate obtained for future samples. If the reallocation is high, it means that the model is identifying most of the buying opportunities. At this point, it is possible to choose the strategy to be followed: maintain a safer model, which is less sensitive when it comes to entering the market, but with greater chances of success; maintain a more aggressive model, which is more sensitive to identify opportunities, but less accurate in its prediction. Extending the precision analysis applied in the strategy defined in section 4.2, it is possible to immediately identify a minimum of safe precision, based on the upper x and lower y limits, that the model must reach for the maximum loss scenario. A maximum loss scenario, as discussed in the previous chapter, is characterized when all forecasts classified as false positive reach the lower limit y, causing a loss the number of securities traded. Defining that, in order to make a profit, the volume of true positives times the profit must be greater than the volume of false positives times the loss, we have the following relationship:

$$\begin{aligned} Specificity &= \frac{Number\ of\ true\ negatives}{(Number\ of\ true\ negatives\ +\ number\ of\ false\ positives)} \\ &= \frac{Number\ of\ true\ negatives}{Total\ number\ of\ individuals\ without\ the\ illness} \end{aligned} \tag{4.2}$$

Where V is the volume of securities traded, P is the accuracy of hits and x and y refer to the upper and lower limits, respectively. Establishing that the relationship and y occurs by equation can be reduced to: In short, from the above relationship, a minimum precision can be defined that never causes false positives from the x and y limits. The scenario stands out that, even with a precision lower than the safe minimum, it is still possible to obtain profit from the model if there is a portion of false positives that do not exceed the lower margin y. The two squares surrounding the classes each represent a pattern of 10-minute classes spaced one minute apart. It is these two models (and the following models shifted by one minute each time. not represented on the graph) which will be provided to the neural network for its learning. The third step is data cleaning. As we have seen, we are going to have to remove the noise from the patterns that we have extracted from the stock price history. For this, we will use a simple technique of extracting the points that seem interesting to us to achieve our objectives. These points are going to be the extremities of a rise or fall if they were large enough to be profitable for us. Taking the second model of the previous graph, we denote each point with a letter, Performance evaluation allows to evaluate and validate a prediction of the neural network. This evaluation will allow us to know the level of reliability of the predictions of the neural network. It will go through the calculation of a difference between the result and the actual price. Validation will be done by assessing this difference. Our focus will be on valuation methods dedicated to predicting stock market prices. These methods allow not only to evaluate the deviation, but to calculate the directional direction of the deviation. Thus, a difference between a falling stock price and a rising stock price prediction will be greater than if the two prices, the actual price and the predicted price, go in the same direction. the methods generally used for performance evaluation are:

i. The effective value or quadratic mean.

ii. Correlation coefficient.

iii. Percentage of absolute maximum.

iv. Percentage of absolute mean.

These methods give a good estimate of the difference between two series of values. In our case, we need to include a directional sense. To illustrate our problem, the following charts each have an actual price and a prediction, we now move on to the test phase which will allow us to appreciate the learning phase of the neural network. By providing it with the dataset displayed previously (stock market price of January 2021), we will be able to present the predictions: The first prediction includes as input to the neural network the first 20 stock market prices (in percentage difference with the previous price). Ideally, the prediction should give us the next 10 days of price (always in percentage) This graph allows us to see that the prediction is relatively far from the ideal price. Using the performance indicators detailed in the Validation and performance chapter, we have a directional symmetry of 22 out of 100 (which is very low) and an average absolute error (MSE) of -0.0057. On a second example, still providing the data for the month of January 2012 shifted by one day compared to the previous example, we obtain the following graph

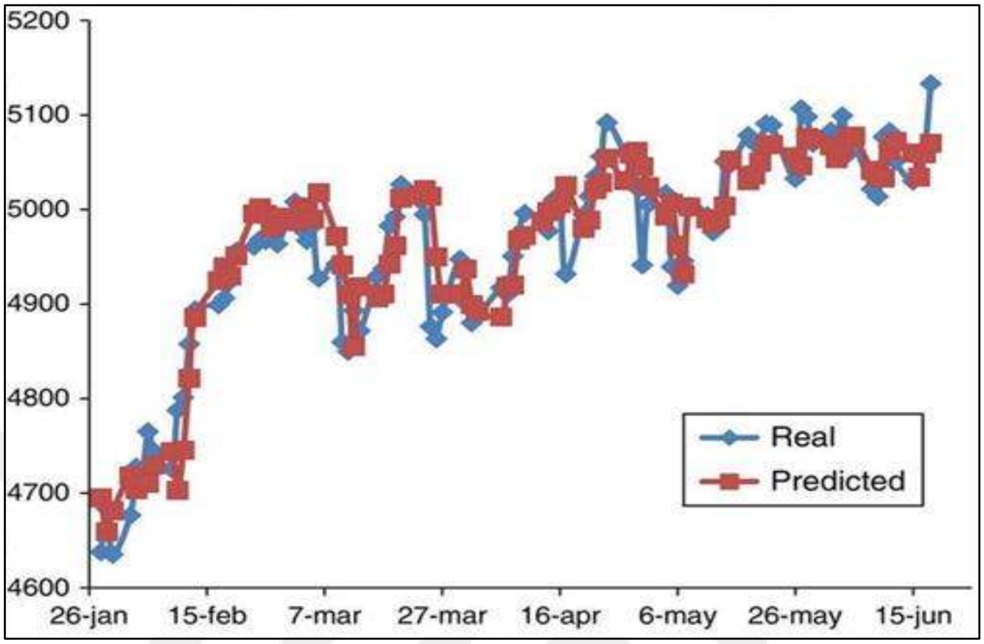

Figure 4.9: Real vs predicted values in the market fluctuations.

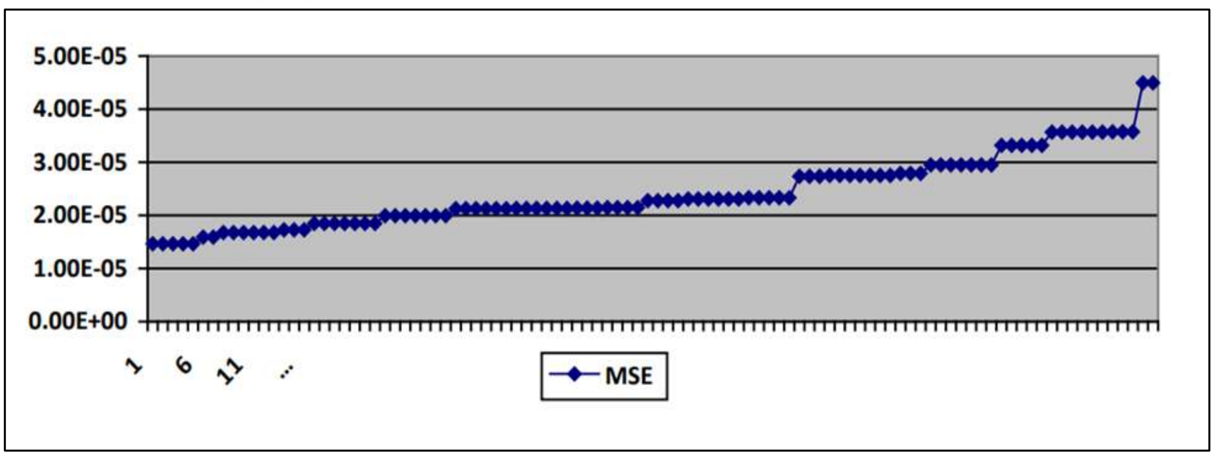

Figure 4.10: Error rate alongside epochs.

4.4 CONCLUSION

With the objective of purely validating the application of machine learning models, the training used the Logistic Regression and MultiLayer Perceptron models to generate the results obtained.

Due to the complexity associated with calculating the MLP, the training sweep for the day cycles totaled approximately 4 to 5 hours. Regarding the calibration parameters for the learning models, the standards defined by the MATLAB r2021a package [21] were used for both models. The present work aimed to apply machine learning techniques to generate price direction prediction models. From the validation using historical tests of the time series, it was possible to verify that decision making supported by artificial intelligence algorithms can be much more effective than a simple traditional technical analysis.

5 CONCLUSIONS

5.1 FINAL REMARKS

The development of this research resulted in the exposition of four machine learning algorithms aimed at predicting the financial market. In addition, different data manipulation and validation techniques were presented and discussed. The work sought to discuss the advantages and disadvantages of the classification models used. The technique of temporal division of the data proved to be more plausible for application in a real system, despite the low results compared to the other technique used.

The specification of our trading system and the development of it, allows us to have a good view of the complexity of the front-office systems that financial institutions can use. By having developed a strategy of tests and neural network, we prove the good design of our system to integrate heterogeneous modules having different functioning. We are certain that we can continue to evolve the needs and specifications of the system in order to make them complete and fully operational. This thesis and in particular the prototype was made possible thanks to contributors from the financial field, external to the project, who were able to help me clarify important gray areas. trading and their calculation methods. I also worked at Natixis, still as a Business Analyst, where we assessed the value of companies, which helped me a lot in specifying needs, particularly strategies. This project was carried out in a context and with personal means while having the objective of marketing a final solution. The different market segments that we have discussed and more specifically the methods that we could use to sell our product, would allow us to enter the financial software market by starting small and studying the personalized needs of customers.

5.2 MAIN CONTRIBUTIONS

The main contribution of this work consists in proposing a method for classifying the trend of assets using different algorithms. The research sought to reproduce results already present in the literature but applied to the Brazilian market. The results obtained show that the high hit rates found in the bibliography are probably not reproducible on a daily basis. This fact is due to the validation method used, which randomly separates training and testing. The SVM-ANN function proved to be able to circumvent the problem of financial predictions, reinforcing the statement in relation to its application in other contexts, in addition to image recognition.

5.3 FUTURE WORK

For future work, the idea is to test new machine learning algorithms, such as LSTM Networks (commonly used for time series), convolutional networks and hybrid models that combine these aforementioned algorithms. In the future, forecast models may also include other metrics, such as forecast price (regression), risk level and probability of occurrence. Furthermore, the development of a portfolio management system using genetic algorithms and deep learning will also be a complement to this research.

REFERENCES

[1] M.R. Hassan and B. Nath, Stock market forecasting using hidden markov model: A new approach, in ISDA'05. Proceedings. 5th International Conference on Intelligent Systems Design and Applications, 2015.

[2] N. Jegadeesh and S. Titman, returns to buying winners and selling losers: Implications for stock market efficiency, The Journal of Finance, vol. 48, no. 1, pp. 65- 91, 1993

[3] P.M. Dechow, A.P. Hutton, L. Meulbroek and R.G. Sloan, Short-sellers, fundamental analysis, and stock returns, Journal of Financial Economics, vol. 61, no. 1, pp. 77-106, 2021.

[4] J.S. Abarbanell and B.J. Bushee, Fundamental analysis, future earnings and stock prices, Journal of Accounting Research, vol. 35, no. 1, pp. 1-24, 1997.

[5] D.P. Brown and R.H. Jennings, On technical analysis, Review of Financial Studies, vol. 2, no. 4, p. 527, 1989.

[6] L. Blume, D. Easley and M. O'hara, Market statistics and technical analysis: The role of volume, The Journal of Finance, vol. 49, no. 1, pp. 153-181, 1994.

[7] Okaz. [Online. Available: https://www.okazinvest.com

[8] H. White, Economic prediction using neural networks: the case of IBM daily stock returns, in IEEE International Conference on Neural Networks, 1988. 10. E. Tsang, J. Li and J.M. Butler, EDDIE beats the bookies, Software - Practice and Experience, vol. 28, no. 10, pp. 1033-1043, 1998.

[9] D. S. Barr and G. Mani, using neural nets to manage investments., AI Expert, vol. 9, no. 2, pp. 16-21, 1994

[10] R. Lawrence, Using neural networks to forecast stock market prices, Course Project, University of Manitoba, 1997.

[11] T. S. Quah, Using neural network for DJIA stock selection, Engineering Letters, vol. 15, no. 1, pp. 126-133, September 2007.

[12] S. Kullback and R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics, vol. 22, no. 1, pp. 79-86, 1951

[13] T. S. Quah, Using neural network for DJIA stock selection, Engineering Letters, vol. 15, no. 1, pp. 126-133, September 2007.

[14] B. Egeli, M. Ozturan and B. Badur, Stock market prediction using artificial neural networks, in International Conference on Business, Hawaii, 2003.

[15] P. Ti˘no, C. Schittenkopf and G. Dorffner, Financial volatility trading using recurrent neural networks., IEEE Transactions on Neural Networks, vol. 12, no. 4, pp. 865- 874, 2001.

[16] J. S. Jang, ANFIS: Adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man and Cybernetics, vol. 23, no. 3, pp. 665-685, 1993.

[17] O. Castillo and P. Melin, Simulation and forecasting complex financial time series using neural networks and fuzzy logic, in IEEE International Conference on Systems Management and Cybernetics, 2001.

[18] Y. F. Wang, predicting stock price using fuzzy gray prediction system, Expert systems with applications, vol. 22, no. 1, pp. 33-38, 2002.

[19] Y. Yoon and G. Swales, Predicting stock price performance: A neural network approach, Kauai, HI, USA, 1991.

[20] T. Kimoto, K. Asakawa, M. Yoda and M. Takeoka, Stock market prediction system with modular neural networks, in International Joint Conference on Neural Networks, San Diego, CA, 1990.

[21] M. Reaz, S. Z. Islam, M. A. Ali and M. S. Sulaiman, FPGA realization of backpropagation for stock market prediction, in Proceedings of the 9th International Conference on Neural Information Processing, Singapore, 2002.

[22] Z. Liao and J. Wang, Forecasting model of global stock index by stochastic time effective neural network, Expert Systems with Applications, vol. 37, no. 1, pp. 834-841, 2010.

[23] X. Wang, P. K. Phua and W. Lin, Stock market prediction using neural networks: Does trading volume help in shortterm prediction? in Proceedings of the International Joint Conference on Neural Networks, 2003.

[24] M. Sugeno and T. Yasukawa, A fuzzy logic-based approach to qualitative modeling, IEEE Transactions on fuzzy systems, vol. 1, no. 1, pp. 7-31, 1993.

[25] Y. Hiemstra, A stock market forecasting support system based on fuzzy logic, in Proceedings of the TwentySeventh Hawaii International Conference on System Sciences, Wailea, HI, 1994.

[26] Wang, Y. Zhang, S. Belkasim and R. Sunderraman, Real time fuzzy personalized web stock information agent, in Proceedings of the 2nd International Workshop on Intelligent Systems Design and Applications, 2002.

[27] Esfahanipour and W. Aghamiri, Adapted neuro-fuzzy inference system on indirect approach TSK fuzzy rule base for stock market analysis, Expert Systems with Applications, vol. 37, no. 7, pp. 4741-4748, 2010.

[28] P.C. Chang and C.H. Liu, A TSK type fuzzy rule-based system for stock price prediction, Expert Systems with applications, vol. 34, no. 1, pp. 135-144, 2008

[29] M. E. El-telbany, The Egyptian stock market return prediction: A genetic programming approach, in International Conference on Electrical, Electronic and Computer Engineering, Kunming, 2005.

[30] D. Zhang and X. Zhang, Study on forecasting the stock market trend based on stochastic analysis method, vol. 4, no. 6, 2009.

[31] S.I. Amari and A. Cichocki, Adaptive blind signal processing-neural approaches, Proceedings of the IEEE, vol. 86, no. 10, pp. 2026-2048, 1998.

[32] S. Amari, T.P. Chen and A. Cichocki, Stability analysis of adaptive blind source separation, Neural Networks, vol. 10, no. 8, pp. 1345-1351, November 1997

[33] S. Kullback and R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics, vol. 22, no. 1, pp. 79-86, 1951.

[34] J. Goldberger, S. Gordon, and H. Greenspan, an efficient image similarity measure based on approximations of KLdivergence between two gaussian mixtures, vol. 1, 2003.

[35] J. Quint, On the equivalence of weighted finite-state transducers, Barcelona, Spain, 2004.

[36] J. D. Lafferty and C. Zhai., Document language models, query models, and risk minimization for information retrieval, in Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, 2001.

[37] A. Ismail and T. Nabil, A hybrid approach for time series forecasting using artificial neural network and kullback-leibler divergence, vol. 15, no. 2, 2004.

[38] W. Wilder, New concepts in technical trading systems, First, Ed., Trend research, 1978.

[39] C. Grosan, A. Abraham, V. Ramos and S.Y. Han, Stock market prediction using multi expression programming, in portuguese conference on Artificial intelligence, 2005.

[40] R.A. Johnson and D.W. Wichern, Applied multivariate statistical analysis, 5 ed., Prentice Hall Upper Saddle River, NJ, 2002.

[41] Cabitza, F., Locoro, A., & Banfi, G. (2018). Machine learning in orthopedics: a literature review. Frontiers in Bioengineering and Biotechnology, 6, 75.

[42] Chavan, P. S., & Patil, S. T. (2013). Parameters for stock market prediction. International Journal of Computer Technology and Applications, 4(2), 337.

[43] Chiu, D. Y., & Chen, P. J. (2009). Dynamically exploring internal mechanism of stock market by fuzzy-based support vector machines with high dimension input space and genetic algorithm. Expert Systems with Applications, 36(2), 1240-1248.

[44] Lee, K. H., & Jo, G. S. (1999). Expert system for predicting stock market timing using a candlestick chart. Expert systems with applications, 16(4), 357-364.

[45] Lee, M. C. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. Expert Systems with Applications, 36(8), 10896-10904.

[46] Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., ... & Deng, X. (2016). Empirical analysis: stock market prediction via extreme learning machine. Neural Computing and Applications, 27(1), 67-78.

[47] Liao, Z., & Wang, J. (2010). Forecasting model of global stock index by stochastic time effective neural network. Expert Systems with Applications, 37(1), 834-841.

[48] C. Grosan, A. Abraham, V. Ramos and S.Y. Han, Stock market prediction using multi expression programming, in portuguese conference on Artificial intelligence, 2005.

[49] R.A. Johnson and D.W. Wichern, Applied multivariate statistical analysis, 5 ed., Prentice Hall Upper Saddle River, NJ, 2002.

[50] Cabitza, F., Locoro, A., & Banfi, G. (2018). Machine learning in orthopedics: a literature review. Frontiers in Bioengineering and Biotechnology, 6, 75.

[51] Chavan, P. S., & Patil, S. T. (2013). Parameters for stock market prediction. International Journal of Computer Technology and Applications, 4(2), 337.

[52] Chiu, D. Y., & Chen, P. J. (2009). Dynamically exploring internal mechanism of stock market by fuzzy-based support vector machines with high dimension input space and genetic algorithm. Expert Systems with Applications, 36(2), 1240-1248.

[53] Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187-205.

[54] Dai, W., Wu, J. Y., & Lu, C. J. (2012). Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. Expert systems with applications, 39(4), 4444-4452.

[55] Das, S. P., & Padhy, S. (2012). Support vector machines for prediction of futures prices in Indian stock market. International Journal of Computer Applications, 41(3).

[56] Dash, R., & Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. The Journal of Finance and Data Science, 2(1), 42-57. Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. Expert Systems with applications, 29(4), 927- 940.

[57] Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. MIS quarterly, 213-236.

[58] Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187-205.

[59] Dai, W., Wu, J. Y., & Lu, C. J. (2012). Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. Expert systems with applications, 39(4), 4444-4452.

[60] Das, S. P., & Padhy, S. (2012). Support vector machines for prediction of futures prices in Indian stock market. International Journal of Computer Applications, 41(3).

[61] Dash, R., & Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. The Journal of Finance and Data Science, 2(1), 42-57. Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. Expert Systems with applications, 29(4), 927- 940.

[62] Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. MIS quarterly, 213-236.

[63] Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. Expert Systems with Applications, 38(8), 10389-10397.\

[64] Holland, J. H. (1992). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.

[65] MIT press. Jasic, T., & Wood, D. (2004). The profitability of daily stock market indices trades based on neural network predictions: Case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965–1999. Applied Financial Economics, 14(4), 285-297.

[66] Kim, H. J., & Shin, K. S. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. Applied Soft Computing, 7(2), 569-576.

[67] Kim, K. J., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert systems with Applications, 19(2), 125-132.

[68] Kim, K. J., & Lee, W. B. (2004). Stock market prediction using artificial neural networks with optimal feature transformation. Neural computing & applications, 13(3), 255-260.

[69] Kim, M. J., Min, S. H., & Han, I. (2006). An evolutionary approach to the combination of multiple classifiers to predict a stock price index. Expert Systems with Applications, 31(2), 241-247.

[70] Kumar, L., Pandey, A., Srivastava, S., & Darbari, M. (2011). A hybrid machine learning system for stock market forecasting. Journal of International Technology and Information Management, 20(1), 3.