|  |
| --- |
| **Logo, company name  Description automatically generated** |
| Faculty of Computers and Information Technology (FCIT) |
|  |
| CSC 624 |
|  |
| Developing a Neural Network Model for Predicting Stock Market: The Case of the Saudi Market |
| Author:  Latifah Albalawi, SID: 421009767 |
| Rabda Alenzi, SID: 421009760  Mona Alasmari, SID: 421009753  Manal Alswat, SID: 421009761 |
| Supervisor: Dr. Mohammad Hijji |
| Submitted in partial fulfillment of the requirements for the Degree of MSc in Data Science |
| Academic Year: 2020/21 |

**Declaration of Originality**

Office Stamp

This project consists of all my work and has not been copied in part or in full from any other source, except where duly acknowledged. As such, all use of previously published works (from books, journals, magazines, the internet, etc.) has been acknowledged within the main report with a corresponding entry in the References list.

I agree that an electronic copy of this report may be stored and used for plagiarism prevention and detection.

I understand that cheating and plagiarism constitute a breach of University Regulations and will be dealt with accordingly.

Signed: Date:

**Copyright**

The copyright of this project and report belongs to the University of Tabuk.

# Abstract

Having the adequate decisions related to the most profitable stock market is complex due to several factors, one of these factors is the price of the stocks during the past years. Artificial Neural Network (ANN) is widely used in the aspect of prediction for decisions making. In the case of the Stock market in Saudi Arabia, there is lack of tools and res with related to predicting Stothe ck market using the Radial Basis Function Neural Network (RBFNN). The aim of this project is, to provide inwith visitors a viable system to help them find promising opportunities for success in market Saudi Market exchange. The adopted method is based on the RBFNN, as it is effective in small to medium datasets, moreover, it is used for high non-linearity datasets. The performance metrics are further improved by optimizing the hyperparameters in the RBFNN

**Ta****ble of Content**

[Abstract 2](#_Toc90421219)

[1 Introduction 9](#_Toc90421220)

[1.1 Background of the Project 9](#_Toc90421221)

[1.2 Project Aim 9](#_Toc90421222)

[1.3 Project Objectives 9](#_Toc90421223)

[Report Structure 10](#_Toc90421224)

[2 Investigation 11](#_Toc90421226)

[2.1 Introduction 11](#_Toc90421227)

[2.2 Background to the Stock Market in Saudi Arabia 11](#_Toc90421228)

[2.2.1 History of the Saudi Stock Market 11](#_Toc90421229)

[2.2.2 Terms Used in the Saudi Stock Market 12](#_Toc90421230)

[2.2.3 Companies in the Saudi Stock Market 12](#_Toc90421231)

[2.2.4 Challenges in the Saudi Stock Market 13](#_Toc90421232)

[2.3 Background to ANN Models for Predicting the Stock Market 13](#_Toc90421233)

[2.3.1 What is ANN? 13](#_Toc90421234)

[2.3.2 Radial Basis Function Neural Network 15](#_Toc90421235)

[2.3.3 Related Works to ANN in Predicting the Stock Market 17](#_Toc90421236)

[2.3.4 Background to Algorithms used in ANN in Predicting the Stock Market 18](#_Toc90421237)

[2.4 Summary 19](#_Toc90421238)

[3 Methodology 20](#_Toc90421239)

[3.1 Introduction 20](#_Toc90421240)

[3.2 Phase 1: Determining the General Idea of ​​the project 21](#_Toc90421241)

[3.3 Phase 2: Literature Review on the Project Domain 21](#_Toc90421242)

[3.4 Phase 3: Requirement Analysis 22](#_Toc90421248)

[4 The Designed RBFNN Model 25](#_Toc90421259)

[5 Implementation & Testing 28](#_Toc90421262)

[6 Results 29](#_Toc90421263)

[7 Conclusions 31](#_Toc90421269)

[7.1 Achievements 31](#_Toc90421270)

[7.2 Future Work 31](#_Toc90421271)

[Bibliography and References 32](#_Toc90421272)

[Appendix A – Code 36](#_Toc90421273)

[Appendix B – Dataset 36](#_Toc90421274)

[Appendix C – Questionnaire 36](#_Toc90421275)

**Additional Materials on the Accompanying CD**

* One copy of the project on a CD

**List of Tables**

**FIGURE TITLE PAGE**

Table 1: References in the project 22

[Table 2: Project Schedule 30](#_Toc90418689)

[Table 3: Risk Management 30](#_Toc90418690)

**List of Figures**

**FIGURE TITLE PAGE**

[Figure 1: ANN architecture (NN 2021) 14](#_Toc90418697)

[Figure 2: The Radial Basis Function Neural Network Architecture 15](#_Toc90418698)

[Figure 3: The adopted project methodology 20](file:////Users/the_full_moon/Library/Containers/com.apple.mail/Data/Library/Mail%20Downloads/4B64520C-E337-497D-BEC1-7E70E26D7414/MSc%20Project_v3%20-111.docx#_Toc90418699)

[Figure 4: Search Domain 21](#_Toc90418700)

[Figure 5: Radial Basis Function Neural Network 22](#_Toc90418701)

**List of Abbreviations**

CMA Capital Market Authority

ANN Artificial Neural Network

G20 Group of Twenty

AI Artificial Intelligence

STC Saudi Telecom Company

SABlC Saudi Basic Industries Corporation

LSSVM Least-squares support-vector machines

RBFNN Radial Basis Function Neural Network Model

RNN Recurrent Neural Network

POS Particle Swarm Optimization DAN2 Dynamic Artificial Neural Networks ML Multi-Layer Perceptron

GARCH Generalized Autoregressive Conditional Heteroskedasticity

TWSE Taiwan Stock Exchange

SSE Shanghai Stock Exchange

KOSPI Korea Composite Stock Price Index

Nikkei225 Nihon Keizai Shimbun

**Acknowledgements**

We would like to thank our parents for their encouragement and support in achieving our postgraduate study. Additionally, we would like to thank our supervisor, Dr. Mohammad Hijji, for his support and guidance, which ultimately allowed us to accomplish this project.

# Introduction

This chapter aims to provide a background on the project and objectives that were developed to reach the project goal and will provide a report on each chapter.

## 1.1 Background of the Project

The stock exchange in Saudi Arabia is a cornerstone for the kingdom’s economy and a major player in the Middle East and around the world. Therefore, accurate predictions for the different share prices can help investors profit from trading in that important sector. A predictor model with the proper performance assessment strategy can be considered the main component in any Intelligent Recommendation system in that regard. The proposed methodology is based on a technique that proved to be useful in similar datasets; it showed improved performance metrics compared to other comparable machine learning techniques **(Khan, Alin, et al. 2011, Olatunji, Al-Ahmadi, et al. 2011)**.

Saudi Arabia's financial system has grown in recent years, becoming stronger and more efficient. The quality of services provided by financial institutions has improved significantly as part of Vision 2030's financial developments, and the Saudi market's capabilities contribute significantly to economic growth **(Alghamedi and Misfer 2012)**.

Previous research has indicated that the capital market performs well in terms of volume and liquidity, and as a result, the Saudi stock market is regarded as one of the major financial markets in the developed world. Furthermore, Saudi Arabia's economy is heavily reliant on oil as a significant source of revenue, and the stock market is predicted to be volatile when oil prices fluctuate **‏(Alghamedi and Misfer 2012)**.

Saudi Arabia is also a member of the Group of Twenty (G20), an international economic cooperation group comprised of leaders from 19 countries. Together, the G20 members represent about 80% of the global economy, two-thirds of the world's population, and three-quarters of global trade. Moreover, Saudi Arabia is currently compiling a package of economic reforms and opening to the world. Finally, the Saudi stock market suffers from severe fluctuations during different periods due to its exposure to external crises, such as the COVID-19 pandemic if a pandemic or a similar crisis were to cause the economy to decline, the economist would need to rein in excessive volatility to ensure financial and macroeconomic stability **(Alghamedi and Misfer 2012)**.

## 1.3 Project Aim

This project aims to:

“Develop an artificial neural network (ANN) model to predict the trends of the stock market in Saudi Arabia.”

## 1.4 Project Objectives

To reach this aim, the following objectives have been developed:

1. Conduct a theoretical review of the available literature on the subject of the stock market in Saudi Arabia and develop an ANN model.
2. Determine the suitable parameters for proper accuracy in the developed ANN model for the case of the Stock market in Saudi Arabia.
3. Develop an effective ANN model for predicting the trends of the stock market in Saudi Arabia.
4. Reduce the prediction time by using adequate artificial intelligence (AI) algorithms.
5. Support end-users in decision-making related to the most profitable stock market in Saudi Arabia.
6. Evaluate the developed ANN model to predict the trends of the stock market in Saudi Arabia based on the set parameters.

## 1.5 Report Structure

The structure of this report consists of four chapters that are outlined as follows:

* **Chapter 2: Investigation**

In this chapter, we will speak about the history and emergence of the Saudi stock market, as it presents the most important companies trading in the Saudi stock market; the most important terms used and the challenges; and will discuss the background of the ANN algorithm and Radial Basis Function Neural Network (RBFNN). This chapter will also mention previous studies that use an ANN algorithm to predict the stock market.

* **Chapter 3: Methodology**

In this chapter, we will speak about the methodology focused on the use of Machine Learning to predict the Saudi stock market.

* **Chapter 4:** **The Designed RBFNN Model**

In this chapter, we will illustrate our RBF Network architecture and its advantages over the other Neural Networks.

* **Chapter 5: Implementation & Testing**

In this chapter, the implementation will be described via the language and tools used for the implementation, algorithms, and libraries in the implementation, and the work of cleaning data.

* **Chapter 6: Results**

In this chapter, we will discuss the results that we have found and also compare between the different machine learning model and our main model which is the RBF Network.

* **Chapter 7: Conclusions**

In this chapter, the achievements made to accomplish the project objectives and future proposals related to the development of the work will be presented also, we will describe the limitations that we faced during our work. Is addition to the project management.

# Investigation

## 2.1 Introduction

This chapter aims to provide a background for the stock market in Saudi Arabia and the related ANN and RBFNN models, and will include some related work to ANN models in predicting the stock market.

## 2.2 Background to the Stock Market in Saudi Arabia

### 2.21 History of the Saudi Stock Market

The Saudi stock market is one of the largest financial markets in developed countries. Oil is the main source of the Saudi economy, and the fluctuations of the stock market are likely to be based on fluctuations in oil prices **(Alshammari, Ismail, et al. 2020)**.

The stock market is comprised of the investors who own shares of a company. The share price depends on the company and rises when the company is doing well. The stock also rises when the economy is high **(Market 2018)**.

Saudi Arabia is also a de facto member of the Organization of Petroleum Exporting Countries, as it plays an important role in the oil market, owning billions of barrels of oil. However, in 2016, Saudi Arabia took measures to reduce its dependence on oil as a major activity for the economy, and developed strategies to reduce its dependence and diversify non-oil sources of income; this is referred to as Vision 2030 **(Khashan 2017)**.

The Saudi stock market, known as Tadawul, is a Saudi company that deals with trading and depositing securities **(Alghamedi, A. (2012))(Alshogeathri 2011)**. The Tadawul offers values, Islamic securities, trade exchanged assets and shared assets. As of now, the Tadawul has almost 200 organizations recorded for trading. The Tadawul All Share Index is the significant financial exchange list that tracks the exhibition of all organizations recorded on the Saudi stock exchange **(Alshogeathri 2011, Alghamedi and Misfer 2012).**

In this section, we will talk about the emergence of the Saudi financial market; it began as an informal market in the fifties and continued until the eighties, when the government of the Kingdom of Saudi Arabia stepped in and established the Capital Market Authority (CMA) in 2003, where it was entrusted with the tasks of developing the financial market **(Alshogeathri 2011).** The Saudi stock market witnessed a development, as authority was transferred to the CMA, where it became the regulator of shares in Saudi Arabia, and trading at that time was limited to Saudis only; trading was eventually allowed and opened in the Gulf Cooperation Council in 2008, and the Market Authority also provided approval for Saudi shares to foreign investors to participate in trading. Then, the Saudi CMA was allowed to open the way for foreign institutions to invest in 2015. Several measures have been taken by the CMA Pot allow foreign investors to own up to 49% of the shares **(Lerner, Leamon et al. 2017).**

By the end of 2017, Nasdaq and Tadawul consented to an arrangement to change Tadawul's post-exchange innovation infrastructure. Now complete, it permits the Saudi bourse to present new resource classes, like subsidiaries, to the market. The connection between securities exchange improvement and financial development has for some time been a huge subject of discussion. The dominant parts of exact examinations on this theme cantered on progressed showcases and created developing business sectors, and no significant review exists for business sectors in oil-based economies. To reach the results, we have to use various blended techniques, such as consolidating quantitative and subjective techniques, to improve the review's legitimacy and unwavering quality **(Alghamedi and Misfer 2012, Lerner, Leamon et al. 2017)**.

### 2.2.2 Terms Used in the Saudi Stock Market

In the Saudi stock market, there are some terms used related to financial, policy, or platform for trading. In this section, the definitions and the differences between all terms will be explained as follows:

* ***CMA:*** It is an organized body that monitors the Saudi financial market and makes regulations and laws that aim to protect investors **(ksa 2021)**.
* ***Stocks:*** A financial value of the shareholder’s share in the company’s capital. It is considered a document that establishes the right of the shareholders in the company. The decrease and rise in share prices depend on the company’s strength in the financial trading market **(terms 2021).**
* ***Money bills:*** A legal document of the investor's ownership of stocks in the company **(terms 2021)**.
* ***Money Market:*** The place where securities are traded **(Market 2017)**.
* ***Tadawul:*** A platform for the stock trading market in Saudi Arabia, where the value of shares for companies in the stock market is estimated at 8.9 trillion Saudi riyals **(ksa 2021, tadawul 2021)**.
* ***Portfolio:*** It considers all the securities owned by the investor and completes this portfolio based on the purchase of shares and added investments **(Terms 2020)**.

### 2.2.3 Companies in the Saudi Stock Market

Saudi Arabia is one of the largest economies in the Gulf, as Saudi Arabia includes the largest energy and electricity companies in the Middle East **(KSA 2015).** Saudi Arabia also includes the second largest petrochemical company, Saudi Basic Industries Corporation (SABIC), which is one of the largest oil companies in the world. SABIC is the best company to distribute profits in the Saudi stock market, as it achieved profits for the first quarter of 2021 with a rate of 24% in revenues **(KSA 2015, market 2021)**.

Saudi banks, such as Al-Ahly Bank and Al-Rajhi Bank, are among the largest government investments in project financing; the former has a long history in the Saudi Stock Market, whereas the latter provides banking services that comply with Islamic law and provides investment services in the Investment Fund **(KSA 2015, stocks 2021)**.

Telecom is one of the most important telecommunications companies with diversified services. The first national company is the Saudi Telecom Company (STC), where 30% of the company's shares were sold to the private sector, where 70% shares are distributed to the Saudi Investment Fund, 16.16% are free shares, and 7% are owned by the General Insurance Corporation. Social and 76.7% pension institution It is considered one of the most important and reliable Saudi investments, and it has proven its effectiveness during the Covid-19 pandemic **(KSA 2015, STC 2021).**

There are also other companies that are considered the best in the Saudi stock market, such as the Kingdom Holding Company, which is the largest investment company in the world, and Samba, SAP, Almarai, and the Saudi Electricity Company **(companies 2016)**.

### 2.2.4 Challenges in the Saudi Stock Market

The Kingdom of Saudi Arabia has faced many challenges in the stock market and the economy in general in 2020, as the Saudi financial market was exposed to fluctuations that negatively affected the performance of the stock market from several factors, such as the COVID-19 pandemic and the judiciary in Syria and Yemen, where the oil price fell. However, with the developments that we are witnessing today in the country and the decisions of the Saudi government to mitigate these repercussions **(challenges 2021)**, diversifying the sources of economic income in the Kingdom of Saudi Arabia has the right to restore the Saudi stock market, the total output, and the decline of the Saudi economy by 3.7 % during the year 2020 for the oil and non-oil sectors, despite the decline in the economy in Saudi Arabia. Although, the stock market deviated from the expectations of the International Monetary Fund by 4.5% **(economy 2020)**.

One of the challenges facing researchers and investors is the prediction of stock market prices, as there is no full awareness of the fluctuations of the financial market **(Sharma and Kaushik 2018)**.

## 2.3 Background to ANN Models for Predicting the Stock Market

### 2.3.1 What is ANN?

Nowadays, the use of automated learning algorithms has become increasingly popular due to their ease of use and high results, and there are many algorithms that have proven effective, such as the neural network algorithm, as many researchers prefer to use the ANN algorithm because it provides high-accuracy results. A neural network is known as an ANN or an SNN. ANN is a supervised learning algorithm that shows us the basic relationships in a dataset through a model inspired by how the brain works. It solves the problems that occur in the field of AI, the field of machine learning, and the field of deep learning. ANN is a very powerful and flexible model that works well with unstructured inputs, such as audio, video, and other. ANN can achieve high performance, and networks are typically organized in layers. There are three types of neural network layers: the first layer is the Input layer containing Input variables, the second layer is the hidden layer containing any layer between the Input and Output layers, and the third layer is Output layer containing generated predictions for classification or regression, as shown in Figure 3 below **(NN 2021)**. ANN has two main phases: the feedforward phase computes values from inputs to outputs, and the backpropagation phase adjusts the weights to minimize the loss function **(algorithm 2021) (Wanjawa and Muchemi 2014, ANNML 2020)**.

There are various types of ANNs, such as RBFNN, which is considered an important type in the ANN because it learns fast and its ideas are comparable to the theory of approximation function**(John A. Bullinaria 2004).** ANN techniques are divided into several categories: a classification neural network is trained in classifying a set of data into a specific category, prediction ANN is trained to produce expected outputs from specific inputs, clustering ANN is trained in identifying a unique feature of data and classifying it into different categories, and association ANN is trained in remembering a certain pattern **(AI 2021, ANN 2021).**

Thus, the ANN algorithm has many different advantages, including the ability to be used in classification and aggregation, and it also has the ability to process in a balanced manner, as it does more than one important job at a time. It also has the ability to train a machine quickly and make decisions, and stores information on the entire ANN algorithm. Therefore, if some information is lost, it would not affect the performance of the algorithm **(Wu and Lu 1993).**

Diagram

Description automatically generated

Figure 1: ANN architecture (NN 2021).

### 2.3.2 Related Works to ANN in Predicting the Stock Market

**{Liang, 2021 #47}** proposed a system to use AI techniques to predict the stock market and analyse stock market data using an RBFNN because it is a new and effective neural network and has a high computing speed and a high ability to generalize; therefore, it uses a non-linear planning function, which is very suitable for forecasting in the stock market. By comparing data processing between the RBFNN and the feedforward neural network, it was proved that the radial network is much better based on the proven results. Also, it provides better predictions for stock prices when compared to 40 other prediction models **.**

**{Alotaibi, 2018 #45}** proposed a system wherein ANNs used the reverse propagation algorithm to predict the movement of the Saudi stock market. As the dataset was chosen from real data from the Saudi stock exchange (TADAWUL stock market exchange), where was used 2 futures in the dataset, which is the average price and the closing price, and they were trained and then the data was tested, where it proved that the ANN using the back propagation algorithm has a high capacity and accuracy for predictions in the stock market**.**

**{Vrbka, 2017 #17**}The capacity to precisely anticipate share value advancement is critical for financial backers as far as boosting their riches goes. By utilizing high accuracy anticipating frameworks, financial backers can create incredible gains.

**{Loayza, 2021 #18}** The considerations for this model refer to the methodological analysis from the point of view of the Peruvian financial environment, as well as the framework conceptualization of neural networks, such as architecture, back propagation algorithms, and analysis parametric of neural networks**.**

**{Padhiary, 2011 #19 }** Stock value forecasting is a significant field of examination in finance because of the supposition that the market is effectively anticipated. At that point, the financial backers might get the greatest returns. The financial exchange or value market is a public market where an enormous measure of capital is put and exchanged everywhere. Numerous analysts guarantee that the market is dynamic, non-direct, convoluted, and turbulent in nature.

**{Fallatah, 2011 #95}** stated that accuracy was measured through the correlation coefficient, as this refers to how close the expected values ​​are in the direction of the actual price values, so that if they recorded low values ​​in the (RMSE error) as the model will become highly predictable, as the result was 99.9%**.**

**{Sharma, 2021 #20}** proposed that the irregular conduct of financial exchanges makes estimating difficult, and new ways to deal with gauging models continues to be looked for, such as customary measurable procedures, like autoregressive coordinated moving normal, for example**.**

**{Nti, 2021 #22}** suggested that a few Petabytes of information are created consistently from various sources, which influence the financial exchange. A reasonable and proficient combination of these information sources (factors) provide insight to offer better pre-phrasing precision on the securities exchange.

**{Madhu, 2021 #23**} explained that different AI strategies have been planned and created to manage the issue of foreseeing the future pattern of choice cost, such as SVM.

**{Alturki, 2020 #24}** The recurrent neural network RNN was proposed to test the weak shape of the Saudi stock market and produce a trading signal to predict the next day's trading signal for many stocks in the Saudi market. The results showed that RNN achieves trading with an accuracy of 55%.

**{Olatunji, 2013 #25}** mentioned that the outcomes from the exact work demonstrated that the proposed ANN model predicts the following day shutting value financial exchange esteem with an exceptionally low RMSE down to 1.817.

### 2.3.4 Background to Algorithms used in ANN in Predicting the Stock Market

The movement of a stock index is the consequence of numerous conceivable elements, such as a company's growth and profit-making capacity, local economy, social and political situations, and the global economic position, making the stock index, trend, and market predictions a difficult assignment for academics.

To avoiding financial risk and maximize returns, accurate projections are essential. Fundamental and technical stock analysis are the two types of stock analyses. The first is an examination of a stock's intrinsic value based on basic variables, such as a company's development and profit-making capacity, the growth of its industry group, and the overall economic trend. The second is a mathematical study based on historical stock index records. The most basic type of analysis is to generate a prediction based on the trend of stock movement in a graph. Complex statistical methods and machine learning algorithms are used in more advanced investigations **(Analysis 2021)**.

To estimate the Dow Jones index, Bollen et al. (**Dickinson, Ganger et al. 2015)** analyzed data from Twitter. Guresen et al. **(Guresen, Kayakutlu et al. 2011)** utilised four models to predict the NASDAQ index: ANN Multilayer Perceptron (MLP), Dynamic Architecture for Artificial Neural Network (DAN2), GARCH-MLP, and GARCH-DAN2. To predict SSE, TWSE, KOSPI, and Nikkei225, Wang et al. used Elman recurrent neural networks paired with a stochastic time effective function. ANN has been used to anticipate both established and fledgling stock markets. Kara et al. **(Wang, Wang et al. 2016)**, for example, employed ANN and SVM to predict the movement of the Turkish ISE 100 index by importing many technical indicators, and discovered that ANN's predictions were correct **(Kara, Boyacioglu et al. 2011**)

When the indicators were imported into four models—ANN, SVM, Random Forest, and Nave-Bayes classifier models—Patel et al. proposed preparing a trend to determine the data of technical indicators prior to importing them into models and found that it provided improved prediction results than a conventional import procedure when the indicators were imported into CNX Nifty and S&P Bombay Stock Exchange markets (**Patel, Shah et al. 2015)**. Manish and Thenmozhi employed ANN, SVM, logit, and Random Forest to forecast the daily direction of the S&P CNX NIFTY Index and discovered that SVM beat the others. In all the studies mentioned above, either ANN or SVM came out on top **(Manish and Thenmozhi 2005)**.

## 2.4 Summary

In this chapter, we talked about the development and establishment of the Saudi stock market, as the Saudi market is one of the most important and largest markets in the developed world, and we got acquainted with the most important terms of the Saudi stock market. The best companies affiliated with Saudi Arabia, such as the energy and electricity companies, SABIC, Saudi banks, Telecom and others, as well as the challenges that the Kingdom of Saudi Arabia faced in the Saudi stock market, impact of the country's economy. Additionally, the ANN algorithm is considered one of the most important algorithms and is preferred by many researchers for its accuracy and ease of use, and because of how neural networks and RBFNNs work with regard to the previous research was that used pertaining to the ANN algorithm and to predict the stock market; this ultimately proved that it is an accurate algorithm that produces accurate results.

# Methodology

## 3.1 Introduction

This chapter aims to demonstrate the adopted methodology that was designed based on the "Waterfall methodology," as shown in Figure 3 below **(Model 20211).**

Figure 3: The adopted project methodology (Model 20211).

## 3.2 Determining the General Idea of ​​the project

This phase refers to the determination of the initial lines of the project and chooses an accurate idea of ​​​​the forecast and the field of forecasting. According to Vision 2030, the economy is a key player in the success of the country. Therefore, stock market predictions and how to incorporate machine learning techniques in these predictions is necessary to build an efficient predictive model **(Vasudevan 2021**).

## 3.3 Literature Review on the Project Domain

The research idea began by reviewing the Saudi stock market and, stock market research in general, to understand the different techniques used in such an important sector. Similar use cases have been planned the research methodology very carefully. The research was conducted through research related to the project through the two scopes of the desired research (ANN prediction of the stock market).

Figure 4: Domains of the conducted literature review.

The scientific papers and research specialized in the field of stock market prediction through machine learning algorithms were searched and nearly 10 research articles related to what was searched were found. Table 1 below provides an overview of the used references in this project.

Table 1: References in the project

|  |  |
| --- | --- |
| Years | From 2011 to 2021 |
| Journal Papers | 8 |
| Conferences Papers | 13 |
| Books | 3 |
| Ph.D. thesis | 1 |
| MSc report | 1 |

## 3.4 Requirement Analysis

Based on the chosen research trends, and the selection of a sequential procedure methodology, the Saudi open-source data platform was used and a dataset was selected from it via the website **(dataset).**

The dataset was chosen for a group of companies that were selected according to their successive net income over a period of three years (2017–2019) were divided into four sections, where a period of every three months is a separate period from which the net income is determined; the figures are set in the millions. After dataset selection, a predictive model using RBFNN is implemented, validated, and tested using the collected set of data.

### –––3.4.1 Study of the Behaviour Patterns of the Data

Before teaching the machine and applying the algorithm, it is necessary to study the behaviour of the shares (rise or decrease) of the income of the joint-stock company and note the direction of the curve during the existing division.

## 3.5 Network Architecture Overview

Multi-layers are made up of numerous layers that are connected to each other. Each unit's factors are independent of the factors of the other units in the layer in a fully connected layer, implying that each unit has its own set of weights. The model will be designed using the necessary Python code snippets**(Multi-layers 1997, Kumar and Yadav 2011, MLP 2020).**

Diagram

Description automatically generated

Figure 5: RBFNN (He, Yan et al. 2019).

## 3.6 Choosing the RBFNN over the Other NN

To forecast stock prices with accuracy, we need to use a technique that works with small and medium numbers sample sizes with high variance. It may, like RBFNN, be suitable for our case study. According to the literature, RBFNN with the Gaussian activation function may be the best solution **(Madhiarasan and Deepa 2016).**

Before we dive into training and testing, we need to understand how RBFNN works and discuss its advantages.

Like any neural network, the RBF network has some hidden layers which consist of hidden neurons, the special thing in RBF architecture is the activation function of these neurons is a Gaussian function. The hidden layer generates a signal corresponding to an input vector in the input layer, and corresponding to this signal, the network generates a response.

**In-brief,** Radial basis function (RBF) networks have the advantages of easy design, good generalization, strong tolerance to input noise, and online learning ability. The properties of RBF networks make them very suitable to design flexible control systems.

RBF doesn't suffer from local minima in the same way as multi-layer perceptron. RBF networks are also good at modeling nonlinear data.

### 

### 4. The Designed RBFNN Model

### 4.1 Radial Basis Function Neural Network history

The theory of function approximation is inspired by the concept of RBF networks. Multi-Layer Perceptron (MLP) networks can learn to approximate functions using a hidden layer of sigmoidal units. RBF networks, on the other hand, take a slightly different approach, as they have several suitable characteristics as shown in Figure 2 below **(RBFN 2020)**. First, they are made up of feedforward networks with two layers. Second, a collection of radial basis functions is implemented by the concealed nodes, such as Gaussian functions. Third, as in an MLP, the output nodes implement linear summation functions. Fourth, the network is trained in two stages: first, the "weights" from the input to the hidden layer are determined, and then the weights from the hidden to the output layer are determined. Fifth, the training/learning process is quite quick. Finally, interpolation is a strong suit for the networks **(Zhao, Liang et al. 2018).**

### 4.2 Radial Basis Function Neural Network Architecture

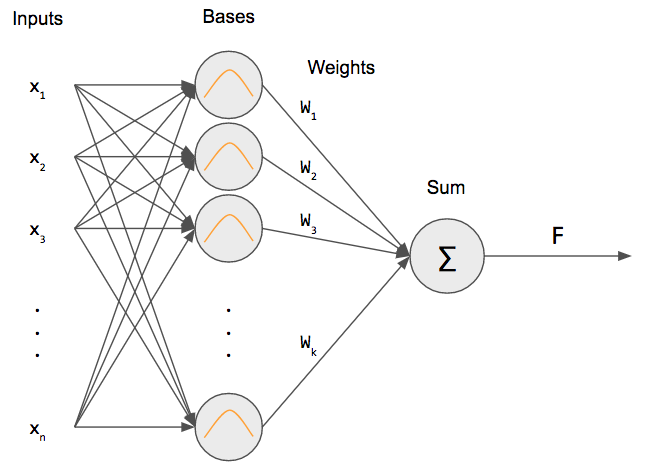


Figure 2: The Radial Basis Function Neural Network Architecture (RBFN 2020).

Moreover, regression or curve fitting refers to the general challenge of detecting an underlying function from a set of noisy training data using MLP networks. Exact interpolation is a specific instance in which the output function passes through all the data points perfectly.

### 4.3 Demonstrate the flow of data through RBFNN

To precisely interpolate a sequence of N data points in a multi-dimensional space, all the D dimensional input vectors must be mapped onto the target outputs . The goal is to find a function f(x) **(Zhao, Liang et al. 2018).**

The radial basis function method provides a collection of N basis functions, one for each data point q, with the form where is a non-linear function whose form will be explored momentarily. Thus, the qth function is determined by the distance between and , which is commonly assumed to be Euclidean. The mapping's output is then interpreted as a linear combination of the basis functions, i.e. **(Zemouri, Racoceanu et al. 2003, RBF 2004)**.

The goal is to discover the "weights" that allow the function to traverse the data points.

Combining the preceding equations yields the following weight equations:

The training data determines the distances between data points p and q, so the Gaussian and Inverse Multi-Quadric Functions are localized in the sense that:

Even the Linear Function is non-linear in the components of in two or more dimensions. This leads to the piecewise-linear interpolating function in one dimension, which is the simplest form of precise interpolation **(RBNN 2015).** There are good reasons to prefer localized basis functions for neural network mappings. Gaussian basis functions will be the focus of our study since, in addition to being localized, they have several other valuable analytic qualities. We can also see how to set their widths and naturally construct function approximations **(RBF 2004, RBNN 2015).**

To improve the predictive models’ performance, we start with the RBF networks' basic structure, which performs accurate interpolation, then improve on it in several ways **(RBNN 2015, Zhao, Liang et al. 2018)**:

1. The number M of hidden units (basis functions) does not have to equal the number N of training data points. In general, it is preferable to have M significantly less than N.
2. The training data input vectors do not have to be defined as the centers of the radial basis functions. Instead, a training algorithm can determine them.
3. The width parameter does not have to be the same for all the radial basis functions. A training algorithm can also be used to determine these.
4. Like an MLP, bias settings can be added into the linear sum of activations at the output layer. These will make up for the disparity between the average value of the basis function activations over the entire dataset and the actual value **(Zhao, Liang et al. 2018).**

## 

# 5. Implementation & Testing

### 5.1 Tools and Libraries

There are several tools used in the implementation phase, which are as follows:

* ***Built in libraries in Python Scikit Learn and NumPy:***

There serval **libraries** used, which are:

1. **Scikit**: One of the offices for machine learning specialized in Python. Its contents includes algorithms and methods used in the field of prediction, in addition to its use in the stage of data filtering and the evaluation of prototypes **(scikit-learn 2017).**
2. **NumPy**: The mother library of Python in data science; the rest of the libraries depend on it mainly for its simplicity, with its wonderful ability to deal with arrays and lists, and it has an amazing ability to quickly read the output **(Numpy 2017).**
3. **Pandas:** It is one of the most popular packages that used in the field of data science and machine learning. It is built depending on NumPy. It help us to load, save, manipulate, and analyse our dataset.
4. **Matplotlib:** Matplotlib is a comprehensive library for creating static and interactive visualizations in Python. Matplotlib helps to make the trends in data visible and easy to conclude.
5. **Seaborn:** is a Python data visualization library based on matplotlib. It has the same functionality od Matplotlib but with more advanced interface. We can say the seaborn is the upgraded version of matplotlib in user interface.
6. **TensorFlow:** It is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.
7. **Keras:** It is an API designed for human beings, not machines. Keras follows best practices for reducing cognitive load: it offers consistent & simple APIs, it minimizes the number of user actions required for common use cases, and it provides clear & actionable error messages. It also has extensive documentation and developer guides.

import pandas as pd

import numpy as np

import matplotlib.pylab as plt

%matplotlib inline

from pandas import DataFrame

from pandas import concat

from matplotlib.pylab import rcParams

from collections import defaultdict

from scipy.spatial import distance

from sklearn.cluster import KMeans

from sklearn.model\_selection import train\_test\_split , cross\_val\_score , KFold

from sklearn.metrics import explained\_variance\_score , mean\_absolute\_error , mean\_squared\_error , r2\_score

from sklearn.neural\_network import MLPRegressor

from keras.utils import np\_utils

from keras.models import Sequential

from keras.layers import Dense, Dropout, Activation,LSTM

from tensorflow.keras.optimizers import SGD

from keras.wrappers.scikit\_learn import KerasRegressor

from sklearn.preprocessing import StandardScaler

from sklearn.pipeline import Pipeline

from keras import optimizers.

### 5.2 Load the Dataset

#Reading the data

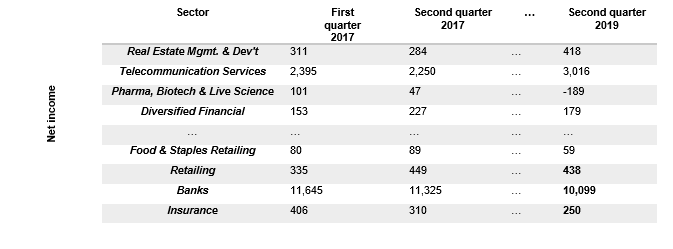
dataframe=pd.read\_excel('net-income.xlsx' ,header=None)

dataframe=dataframe.iloc[:,1:]

dataframe

Load our dataset using read\_csv function from pandas library. Select all the records but we dropped the first column as it only contains duplicated index so it isn’t useful. Display a sample of the dataset.

1. DATASET OF NET INCOME BY SECTOR(MAIN MARKET) FROM FIRST QUARTER2017-SECOND QUARTER2019



When there is a dataset that can be analyzed and subject to the possibility of learning and testing through algorithms.

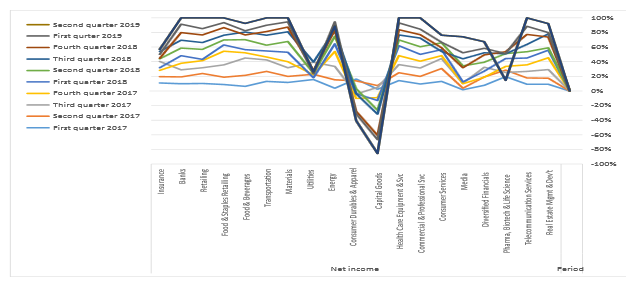
We must be sure that our data is clean and in the form that helps the models to reach the best results and capture the most important or relative information in it.

Having clean data will increase overall productivity which will help in taking the right decision.

### 5.3 Data Visualization

Data visualization allows us to capture insights from our data. It helps us to recognize new patterns and errors in the data. Making sense of these patterns helps us pay attention to areas that indicate red flags or progress. This process, in turn, drives the business ahead.

We will depend mainly on Matplotlib and Seaborn in data visualization as we mentioned before.



**Fig.6. The autocorrelation function of the Saudi stock market index**.

### 5.4 Date Pre-processing

Good data preparation helps us in reaching efficient analysis, limits errors and inaccuracies that can occur to data during processing, and makes all processed data more accessible to users.

from sklearn.preprocessing import LabelEncoder

LE = LabelEncoder()

data['type'] = LE.fit\_transform(data['type'])

in the previous code, we imported the label encoder to use in encoding our type column which means we will convert it from text form (string) to numeric form (numbers) to be passed to the model as the model can only deal with numbers.

### 5.5 Split the data

We need to split a dataset into train and test sets to evaluate how well our machine learning model performs. The train set is used to fit the model, the statistics of the train set are known. The second set is called the test data set, this set is solely used for predictions.

Training the proposed model with predefined hyperparameters the 80/20 *method* is often used for medium and small datasets, or 30/70 for large datasets. However, the project used the 80/20 *method, which is as follows*:

train\_data=data.iloc[:152,:]

test\_data=data.iloc[152:,:]

X\_train=train\_data.iloc[:,:2]

Y\_train=train\_data.iloc[:,2:]

X\_test=test\_data.iloc[:,:2]

Y\_test=test\_data.iloc[:,2:]

1. 80% of the dataset for training which is the samples from 0 to 152.
2. 20% of the dataset for testing which is the samples from 152 to the end of data.
3. In addition to our use of the *n* *fold* cross validation method, in order to adjust hyperparameters without losing data (**BookAI 2020)**.

Moreover, in order to have an effective model, the hyperparameters approach is used, where after training the proposed model with the predefined hyperparameters, we can have advanced tuning of these hyperparameters using an advanced metaheuristic optimization technique, such as PSO or any of its variants.

After splitting the data into training and testing, it is time to select the input features and the output column for both training and testing datasets.

### 5.6 Model Building

The first step here is to choose the best model that can fit our data well and give the most accurate result. However unfortunately there is a magic model. So, we will implement 3 models and compare the results to find out and compare the score of RBF Network and the other models.

We choose to implement three different models that are different in architecture and complexity level.

The three models are:

RBF Network which is our main model.

ANN is a simple neural network.

Linear Regression is one of the most straightforward and simplest models.

### 5.6.1 Building the RBFNN Model

# Gaussian function

def RBFunction(r, x,c ):

return(np.exp(- (distance.euclidean(x, c))\*\*2/r\*\*2))

# Computing radial/spread

def radial (x,c ):

r=[]

for i in range (c.shape[0]):

k=[]

for j in range(x.shape[0]) :

k.append(distance.euclidean(x.iloc[j,:], c[i]))

m=(sum(k)/len(k))

r.append(m)

return(r)

# compute the centers using Kmeans

def centers (data,n):

kmeans = KMeans(n\_clusters=n)

a = kmeans.fit(data)

return(kmeans.cluster\_centers\_)

#Preparing the RBF network

def RBF (centers,radials, X):

s=[[]for k in range (centers.shape[0])]

for i in range (centers.shape[0]):

for j in range (X.shape[0]):

X=pd.DataFrame(X)

s[i] .append(RBFunction(radials[i] ,X.iloc[j,:],centers[i]))

s=pd.DataFrame(s)

return(s.transpose())

In the above code, we have defined 4 functions which are RBFnction, radial, centers, and RBF.

The RBFunction function is responsible for applying the gaussian transformation.

The radial function contains 2 for loops (inner and outer) that are responsible of iterate and computing the radial over spread.

The centers function is responsible for computing the centers using the K-means algorithm.

RBF function is responsible for preparing the RBF Network. It also contains two loops. Finally, it returns a data frame.

def MAD\_MSD(y,pred):

madv=[]

mapev=[]

aux1=y.reset\_index(drop=True)

aux1=pd.Series(aux1)

aux2=pd.Series(pred)

for i in aux1.index:

if aux1[i]==0:

mapev.append(0)

else:

mapev.append(abs(aux1[i]-aux2[i])/aux1[i])

for i in aux1.index:

if aux1[i]!=0:

madv.append(mapev[i]\*aux1[i])

else:

madv.append(0)

mad=np.mean(madv)

#MSD:Mean squared deviation

msdv=np.square(madv)

msd=np.mean(msdv)

return(mad,msd)

def SMAPE(y,pred):

smapev=[]

smape=0

aux1=y.reset\_index(drop=True)

aux1=pd.Series(aux1)

aux2=pd.Series(pred)

for i in aux1.index:

if aux1[i]==0:

smapev.append(0)

else:

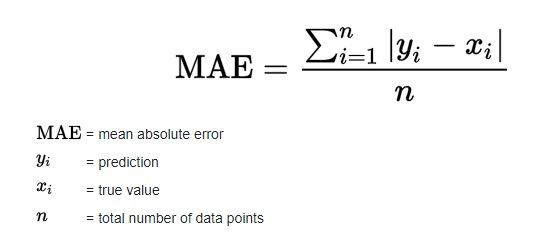
smapev.append(2\*(abs(aux1[i]-aux2[i])/((abs(aux1[i])+abs(aux2[i])))))

smape = (np.mean(smapev))\*100

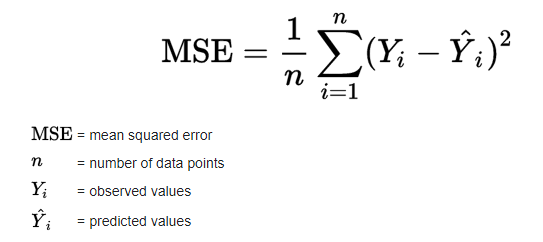
return(smape)

The above code is mainly used in the evaluation step as it creates the mean absolute error or deviation (MAE), mean squared error(MSE), and squared mean absolute error(RMAE).

The mean absolute error is calculated using the following formula:



The mean squared error is calculated using the following formula:



### 5.6.2 RBFNN Model Training

nc=20

c=centers(X\_train , nc)

r=radial(X\_train,c)

inputs\_train= pd.DataFrame(RBF(c,r,X\_train))

#Test

inputs\_t= pd.DataFrame(RBF(c,r,X\_test))

model = Sequential()

model.add(Dense(1, input\_dim=nc, kernel\_initializer='normal'))

# Compile model

model.compile(loss='mean\_squared\_error', optimizer='adam')

#data preparation

inputs\_train\_arr=np.array(inputs\_train)

inputs\_t\_arr=np.array(inputs\_t)

y\_train\_arr=np.array(Y\_train)

y\_t\_arr=np.array(Y\_test)

#train the model

model.fit(inputs\_train\_arr,y\_train\_arr, epochs=1000, batch\_size=64,verbose=1)

After defining the main functions that are responsible for building the RBFNN architecture and the evaluation metrices. It is time call these functions to construct our model.

We called the centers function and passed nc=20 which represents the number of centers. After that, we called the radial function and passed the centers to it. We used the RBF to create our data frame.

We created our sequential model from Keras and added one layer to it which is a dense layer that represents the output layer with only one neuron. We compiled the model using mean squared error as our loss function and adam to be the optimizer.

Finally, we train the RBFNN for 1000 epochs with a batch size equalling 64.

### 5.6.3 Linear Regression Model

from sklearn import linear\_model

# Create linear regression object

regr = linear\_model.LinearRegression()

# Train the model using the training sets

mr=regr.fit(X\_train, Y\_train)

# Make predictions using the testing set

y\_pred = regr.predict(X\_test)

Linear Regression is one of the most straightforward models in machine learning. We imported it from SKlearn library. To make a LR model, we just take an instance from it and fitted or trained it on our train data. Finally, we make our prediction using our test sample data.

### 5.6.4 ANN Model

classifier = Sequential() #we use classification because were going to predict tje tested result

classifier.add(Dense(

units = 6,

kernel\_initializer="uniform",

activation="relu",

input\_dim = 2

))

classifier.add(Dense(

units = 6,

kernel\_initializer="uniform",

activation="relu"

))

#adding output layer (final layer), sum of units is 1 because we just have 1 output

classifier.add(Dense(

units = 1,

kernel\_initializer="uniform",

activation="sigmoid"))

classifier.compile(loss='mean\_squared\_error', optimizer='adam')

classifier.fit(

X\_train,

Y\_train,

epochs=500, batch\_size=64,verbose=1)

Building a simple artificial neural network to compare its performance and the RBFNN performance. First, we create our sequential model then we add three dense layers. The first two dense layers have the same properties which are the number of units (6), the kernel initializer (uniform), and the activation function (relu). The final dense layer consists of only one neuron and represents the output layer with a sigmoid activation function.

After that, we compile the model using mean squared error as our loss function and adam to be the optimizer. Finally, we train the RBFNN for 500 epochs with a batch size equalling 64.

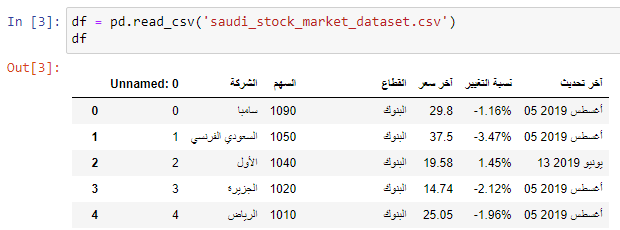
### 5.7 Testing

### 5.7.1 Real test data preparation

To accurately test and choose the best model, we got a real dataset for the Saudi Stock Market.

But we faced a problem: the new data format is different from our original(training) data format.

So we need to make some preprocessing steps to make the new data look like the original one.



First of all as we see, the type column needs to be mapped to the form of training data.

So, we will make a replacer that will replace each value in the new test data with its corresponding in the original date. Ex: (البنوك) will be mapped to (Banks)



This is our full replace dictionary.

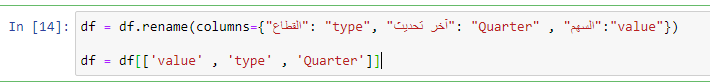
Secondly, we will extract the month for date column as our year is will be split into four quarters.



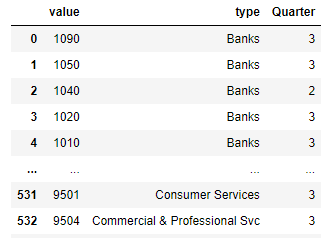
The final two steps will be:

Rename the columns.

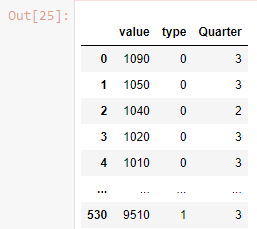
Select only relevant columns.



And this is the final look of our real testing data



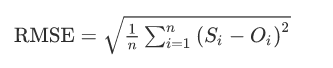
We will map the type column to numeric format as we have done in training data



### 5.7.2 Model Evaluation

As mentioned before, we applied three different models. So for the evaluation the root mean squared error (RMSE) will be our main matrix to decide the best model. The root mean squared error is defined as the square root of the mean of the square of all of the error. It is one of the most popular evaluation matrices that is used to measure the difference between the actual values of the dataset and the predicted values by the model.

RMSE formula is given as follows:



* n represents the total number of samples.
* S represents the predicted values by the model.
* represents the actual values.

For testing we have 2 types of testing data which are:

* Test sample that came from train test split
* Real test data

real\_predicted = regr.predict(real\_x\_test)

var=explained\_variance\_score(real\_y\_test, real\_predicted)

mae=mean\_absolute\_error(real\_y\_test, real\_predicted)

mse=mean\_squared\_error(real\_y\_test, real\_predicted)

rmse=np.sqrt(mse)

r2=r2\_score(real\_y\_test, real\_predicted )

metric=[mae,mse,rmse]

nn=['MAE','MSE','RMSE']

performance=pd.DataFrame([nn,metric])

This code sample is used to test the linear regression performance on the real dataset that we prepared in the previous part.

inputs\_t= pd.DataFrame(RBF(c,r,real\_x\_test))

inputs\_t\_arr=np.array(inputs\_t)

y\_t\_arr=np.array(real\_y\_test)

out=model.predict(inputs\_t\_arr)

mae=mean\_absolute\_error(real\_y\_test, out)

mse=mean\_squared\_error(real\_y\_test, out)

rmse=np.sqrt(mse)

nn=['MAE','MSE','RMSE']

metric=[mae,mse,rmse]

performance=pd.DataFrame([nn,metric])

This code sample is used to test the RBF Network performance on the real dataset that we prepared in the previous part.

out=classifier.predict(real\_x\_test)

mae=mean\_absolute\_error(real\_y\_test, out)

mse=mean\_squared\_error(real\_y\_test, out)

rmse=np.sqrt(mse)

nn=['MAE','MSE','RMSE']

metric=[mae,mse,rmse]

performance=pd.DataFrame([nn,metric])

This code sample is used to test the ANN performance on the real dataset that we prepared in the previous part.

# 6. Results

The results of the models were different in training and testing. First of all, the linear regression model managed to reach only 0.901 error on the sample test while it achieved a little higher error on real testing data which is 0.981. ANN has a good performance on the sample test with only 0.827 error but unfortunately it had a high error on the real test data which is 1.07. Finally, the RBF Network managed to have an acceptable error on test sample while it had the least error on real test data achieving only 0.802 error.

|  |  |  |
| --- | --- | --- |
| Model | Test on sample data | Test on real data |
| RBF Network | 0.841 | 0.802 |
| ANN | 0.827 | 1.07 |
| Linear Regression | 0.901 | 0.981 |

As we see from the previous table, although the linear regression is very simple and straight forward model, it managed to achieve acceptable results. The RBF Network managed to achieve the least error on the real test data. Also, we can see the ANN has a good performance on the test sample but on the real test data, it has a larger error which can be an indicator of an overfitting problem. And this proves the early mentioned advantage of RBF Network that it can handle and fit well on the small data.

We tried to increase the number of epochs to more than 1000 epoch and also tried to change the batch size such as 32 or 128 but the results did not improve much.

# 7. Conclusions

## 7.1 Introduction

In this chapter, the achievements that have been accomplished from the objectives of the project are presented, and suggestions that can be developed from the project for the future are discussed.

## 7.2 Achievements

* **Regarding Objective 1:**

**To conduct a theoretical review of the available literature on the subject of the stock market in Saudi Arabia and the development of ANN models.**

We reviewed several research articles and evaluated models related to our field of study, as well as scientific papers on neural network algorithms and the Saudi stock market, and summarized the most relevant points. In addition, we found that there is a lack of research in the area of RBFNN The number of the research that are covered in the literature review is 11.

* **Regarding Objective two:**

**To determine the suitable parameters for proper accuracy in the developed ANN model for the case of the stock market in Saudi Arabia.**

Several techniques were used to reach the appropriate prediction accuracy through the model, such as, for example, through the testing phase. The N-fold cross validation method was used, which is a very powerful technique for modeling super parameter optimizations and modelling selections that try to overcome the problem of overfitting the training dataset. Furthermore, the RBFNN model was used, which gave us high capabilities in modeling non-linear data.

## 7.3 Project Management

## 7.3.1 Introduction

This chapter aims to break down the tasks necessary to complete the project on time, as well as manage the risks that may arise during the performance of the task and how they can be modified.

## 7.3.2 Project Schedule

In the table below, the project tasks and the time of each task are broken down:

Table 2: Project Schedule

|  |  |  |  |
| --- | --- | --- | --- |
| # | Task | Estimated Duration | Estimated Deadline |
|  | Define project problem | 8 days | 1 Oct 2021 |
|  | Lecture review | 13 days | 16 Oct 2021 |
|  | Define the aim and objective | 5 days | 23 Oct 2021 |
|  | Methodology | 2 weeks | 11 Nov 2021 |
|  | Requirement | 4 days | 2 Dec 2021 |
|  | Analysis | 1 week | 18 Dec 2021 |
|  | Design | 1 week | 02 Jan 2022 |
|  | Implementation | 2 weeks | 24 Jan 2022 |
|  | Testing and Training the models’ stages | 1 month | 2 Feb 2022 |

## 7.3.3 Risk Management

In the table below, the risks that may arise during the research and how to mitigate these risks are presented:

Table 3:Risk Management

|  |  |  |
| --- | --- | --- |
| **#** | **Risk** | **Mitigation** |
| 1 | Look of the appropriate dataset related to the Saudi stock market | Using another dataset and merging dataset together |
| 2 | Find the appropriate papers related to Saudi the stock market | Using other dataset related to a different case study |
| 3 | Find the suitable model related to the dataset | Developing the model to be compatible with the dataset |
| 4 | The information on the project and document is lost or damaged | Using the cloud to store the project document |

## 7.4 Limitations

1. The most common limitation of all deep learning projects is the number of observations (dataset size)

As the neural network is complex and has a huge number of parameters so it needs a large enough dataset to reach an acceptable performance.

2. Results on the real test data has a larger error than the error on the test sample data and again this could be a result of dataset size.

## 7.5 Future Work

The suggestions for some ideas that could develop the work in the future are as follows:

1. Trying to get larger dataset as it will help the model training process.

2. One of the proposed models for its accuracy in predicting stock market prices is LSSVM, working on it and comparing it with RBFNN may improve the study results.

3. Trying to tune the network's hyper-parameters to help the model generalize on unseen data better.

**Bibliography and References**

AI (2021). "30 Artificial intelligence in the stock market." Retrieved 3 Dec, 2021, from <https://www.arab-cio.org/%D8%A7%D9%84%D8%B0%D9%83%D8%A7%D8%A1-%D8%A7%D9%84%D8%A7%D8%B5%D8%B7%D9%86%D8%A7%D8%B9%D9%8A-%D9%88%D8%A3%D9%87%D9%85%D9%8A%D8%AA%D9%87-%D9%81%D9%8A-%D8%B3%D9%88%D9%82-%D8%A7%D9%84%D8%A3%D9%88%D8%B1>.

Alghamedi, A. and A. Misfer (2012). Assessing the impact of stock market development on economic growth in Saudi Arabia: An Empirical Analysis, Durham University.

algorithm, N. (2021). "1 Neural Network ". Retrieved 24 NOV, 2021, from <https://www.investopedia.com/terms/n/neuralnetwork.asp>.

Alotaibi, T., et al. (2018). "Saudi Arabia stock market prediction using neural network." International Journal on Computer Science and Engineering **9**(2): 62-70.

Alshammari, T. S., et al. (2020). "Modeling and forecasting saudi stock market volatility using wavelet methods." The Journal of Asian Finance, Economics, and Business **7**(11): 83-93.

Alshogeathri, M. A. M. (2011). Macroeconomic determinants of the stock market movements: empirical evidence from the Saudi stock market, Kansas State University.

Alturki, F. A. and A. M. Aldughaiyem (2020). "Trading Saudi Stock Market Shares using Multivariate Recurrent Neural Network with a Long Short-term Memory Layer." Machine learning **11**(9).

Analysis (2021). "Technical Analysis." Retrieved 15 Nov, 2021, from <https://www.investopedia.com/ask/answers/difference-between-fundamental-and-technical-analysis/>.

ANN (2021). "2 Artificial Neural Networks." Retrieved 1 DEC, 2021, from <https://www.xenonstack.com/blog/artificial-neural-network-applications>.

ANNML (2020). "What is a Neural Network in Machine Learning." Retrieved 14 Nov, 2021, from <https://www.tutorialspoint.com/what-is-a-neural-network-in-machine-learning>.

BookAI (2020). Artificial intelligence applications and the future of learning technology.

challenges (2021). "The fluctuations and challenges of the Saudi stock market." Retrieved 18 Nov, 2021, from <https://www.mubasher.info/news/3742097/%D8%B3%D9%88%D9%82-%D8%A7%D9%84%D8%A3%D8%B3%D9%87%D9%85-%D8%A7%D9%84%D8%B3%D8%B9%D9%88%D8%AF%D9%8A%D8%A9-%D8%A8%D8%A7%D9%84%D8%B9%D8%A7%D9%85-2020-%D8%AA%D9%82%D9%84%D8%A8%D8%A7%D8%AA-%D9%88%D8%AA%D8%AD%D8%AF%D9%8A%D8%A7%D8%AA-%D9%88%D9%86%D9%87%D8%A7%D9%8A%D8%A9-%D8%A5%D9%8A%D8%AC%D8%A7%D8%A8%D9%8A%D8%A9>.

companies, s. (2016). "The 10 largest companies in Saudi Arabia by market capitalization." Retrieved 25 Nov, 2021, from <https://makkahnewspaper.com/article/586627/%D8%A3%D8%B9%D9%85%D8%A7%D9%84/%D8%A3%D9%83%D8%A8%D8%B1-10-%D8%B4%D8%B1%D9%83%D8%A7%D8%AA-%D9%81%D9%8A-%D8%A7%D9%84%D8%B3%D8%B9%D9%88%D8%AF%D9%8A%D8%A9-%D9%85%D9%86-%D8%AD%D9%8A%D8%AB-%D8%A7%D9%84%D9%82%D9%8A%D9%85%D8%A9-%D8%A7%D9%84%D8%B3%D9%88%D9%82%D9%8A%D8%A9>.

dataset Saudi open-source data

Dickinson, B., et al. (2015). "Dimensionality reduction of distributed vector word representations and emoticon stemming for sentiment analysis." Journal of Data Analysis and Information Processing **3**(04): 153.

economy, S. (2020). "130 Saudi economy." Retrieved 13 Nov, 2021, from <https://www.aleqt.com/2020/12/30/article_2002761.html>.

Endnote. "Endnote." Retrieved 12 Nov, 2021, from <https://endnote.com/>.

Fallatah, Y. (2011). "Saudi Arabia stock prices forecasting using artificial neural networks." from <https://ieeexplore.ieee.org/document/6041425>.

Guresen, E., et al. (2011). "Using artificial neural network models in stock market index prediction." Expert Systems with Applications **38**(8): 10389-10397.

He, H., et al. (2019). "Tree height estimation of forest plantation in mountainous terrain from bare-earth points using a DoG-coupled radial basis function neural network." Remote sensing **11**(11): 1271.

John A. Bullinaria (2004). 22Radial Basis Function Networks.

Kara, Y., et al. (2011). "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange." Expert systems with Applications **38**(5): 5311-5319.

Khan, Z. H., et al. (2011). "Price prediction of share market using artificial neural network (ANN)." International Journal of Computer Applications **22**(2): 42-47.

Khashan, H. (2017). "Saudi Arabia's Flawed" Vision 2030"." Middle East Quarterly.

KSA, C. (2015). "Companies in Saudi Arabia." Retrieved 29 NOV, 2021, from <https://marcopolis.net/top-10-companies-in-saudi-arabia.htm>.

ksa, c. (2021). 3 Guide for Investment Terminology

Kumar, M. and N. Yadav (2011). "Multilayer perceptrons and radial basis function neural network methods for the solution of differential equations: a survey." Computers & Mathematics with Applications **62**(10): 3796-3811.

Lerner, J., et al. (2017). "The CMA and the Saudi stock market crash of 2006." BELLA Research Group and Saudi Capital Market Authority (CMA).

Liang, Y. (2021). 6 STOCK MARKET FORECASTING BASED ON AR T FORECASTING BASED ON ARTIFICIAL

INTELLIGENCE TECHNOLOGY

Loayza, R., et al. (2021). Application of Emerging Technologies in the development of a model of Neural Network for the financial forecast in the Stock Exchange of Lima, EasyChair.

Madhiarasan, M. and S. Deepa (2016). "New criteria for estimating the hidden layer neuron numbers for recursive radial basis function networks and its application in wind speed forecasting." Asian Journal of Information Technology **15**(21): 4377-4391.

Madhu, B., et al. (2021). "A Comparative Study of Support Vector Machine and Artificial Neural Network for Option Price Prediction." Journal of Computer and Communications **9**(5): 78-91.

Manish, K. and M. Thenmozhi (2005). Forecasting stock index movement: A comparison of support vector machines and random forest. Proceedings of ninth Indian institute of capital markets conference, Mumbai, India.

market, c. S. (2021). "A company that distributes profits in the Saudi market." Retrieved 29 NOV, 2021, from <https://saeadat.com/saudi-stocks/best-dividends-stocks-in-the-saudi-market/>.

Market, M. (2017). "What is the Money Market." Retrieved 25 NOV, 2021, from <https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/what-is-money-market/>.

Market, W. i. t. S. (2018). "stok market ksa." Retrieved 24 NOV, 2021, from Stock Market - What is the Stock Market and How it Works (corporatefinanceinstitute.com).

MLP (2020). "Multilayer receptor neural network." Retrieved 11 Dec, 2021, from <https://sciences24.com/%D8%A7%D9%84%D8%B4%D8%A8%D9%83%D8%A9-%D8%A7%D9%84%D8%B9%D8%B5%D8%A8%D9%8A%D8%A9-%D8%B0%D9%88-%D8%A7%D9%84%D9%85%D8%B3%D8%AA%D9%82%D8%A8%D9%84%D8%A7%D8%AA-%D9%85%D8%AA%D8%B9%D8%AF%D8%AF%D8%A9-%D8%A7>.

Model, W. (20211). "Software Development Life Cycle(SDLC) & All models." Retrieved 1Dec, 2021, from <https://gkmaurya.blogspot.com/2011/04/software-development-life-cyclesdlc-all.html>.

Multi-layers (1997). Destructive Water: Water-Caused Natural Disasters, Their Abatement and Control.

NN, I. (2021). "5 A Friendly Introduction to [Deep] Neural Networks." from <https://www.knime.com/blog/a-friendly-introduction-to-deep-neural-networks>.

Nti, I. K., et al. (2021). "A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction." Journal of Big Data **8**(1): 1-28.

Numpy (2017). "32 ML . Libraries0." Retrieved 10 Dec, 2021, from <https://pythonat.com/articles/%D9%85%D9%83%D8%AA%D8%A8%D8%A9-numpy-%D8%A7%D9%84%D8%AE%D8%B7%D9%88%D8%A9-%D8%A7%D9%84%D8%A3%D9%88%D9%84%D9%89-%D9%81%D9%8A-%D8%B9%D9%84%D9%85-%D8%A7%D9%84%D8%A8%D9%8A%D8%A7%D9%86%D8%A7%D8%AA-%D8%A8>.

Olatunji, S., et al. (2013). "Forecasting the Saudi Arabia stock prices based on artificial neural networks model." International journal of intelligent information systems **2**(5): 77-86.

Olatunji, S. O., et al. (2011). Saudi Arabia stock prices forecasting using artificial neural networks. Fourth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2011), IEEE.

Padhiary, P. K. and A. P. Mishra (2011). "Development of improved artificial neural network model for stock market prediction." International Journal of Engineering Science and Technology **3**(2): 1576-1581.

Patel, J., et al. (2015). "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques." Expert systems with applications **42**(1): 259-268.

RBF (2004). Radial Basis Function.

RBFN (2020). "Radial Basis Function Networks." Retrieved 15 Nov, 2021, from <https://pythonmachinelearning.pro/using-neural-networks-for-regression-radial-basis-function-networks/>.

RBNN (2015). Radial Basis Function Networks: Algorithms.

scikit-learn (2017). "31 ML . Libraries." Retrieved 10 Dec, 2021, from <https://pythonat.com/articles/%D9%85%D9%83%D8%AA%D8%A8%D8%A9-scikit-learn-%D9%84%D9%85%D8%AD%D8%A9-%D8%B3%D8%B1%D9%8A%D8%B9%D8%A9-%D8%B9%D9%86-%D9%85%D9%83%D8%AA%D8%A8%D8%A9-%D8%A7%D9%84%D8%A8%D8%A7%D9%8A%D8%AB%D9%88%D9%86>.

Sharma, D. K., et al. (2021). "Integration of genetic algorithm with artificial neural network for stock market forecasting." International Journal of System Assurance Engineering and Management: 1-14.

Sharma, S. and B. Kaushik (2018). "Quantitative analysis of stock market prediction for accurate investment decisions in future." Journal of Artificial Intelligence **11**(1): 48-54.

STC, c. (2021). "Names of telecom companies in Saudi Arabia." from <https://mhtwyat.com/%d8%a3%d8%b3%d9%85%d8%a7%d8%a1-%d8%b4%d8%b1%d9%83%d8%a7%d8%aa-%d8%a7%d9%84%d8%a7%d8%aa%d8%b5%d8%a7%d9%84%d8%a7%d8%aa-%d9%81%d9%8a-%d8%a7%d9%84%d8%b3%d8%b9%d9%88%d8%af%d9%8a%d8%a9/>.

stocks, i. (2021). "Investment stocks in the Saudi market." Retrieved 23 NOV, 2021, from <https://www.almaal.org/investment-shares-in-the-saudi-market>.

tadawul (2021). "tadawul KSA." Retrieved 22 NOV, 2021, from <https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%B3%D9%88%D9%82_%D8%A7%D9%84%D9%85%D8%A7%D9%84%D9%8A%D8%A9_%D8%A7%D9%84%D8%B3%D8%B9%D9%88%D8%AF%D9%8A%D8%A9_(%D8%AA%D8%AF%D8%A7%D9%88%D9%84>).

Terms (2020). "5 اهم المصطلحات في السوق السعودي ". Retrieved 14 Nov, 2021, from <https://fintechwall.com/investment/stock-market-terms-that-every-beginner-should-know/>.

terms, c. (2021). "terms of the capital market authority." Retrieved 24 NOV, 2021, from <https://cma.org.sa/Awareness/Pages/Glossary.aspx>.

Vasudevan, B. (2021). Effective Implementation of Neural Network Model with Tune Parameter for Stock Market Predictions. 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), IEEE.

Vrbka, J. and Z. Rowland (2017). Stock price development forecasting using neural networks. SHS Web of Conferences, EDP Sciences.

Wang, J., et al. (2016). "Financial time series prediction using elman recurrent random neural networks." Computational intelligence and neuroscience **2016**.

Wanjawa, B. W. and L. Muchemi (2014). "ANN model to predict stock prices at stock exchange markets." arXiv preprint arXiv:1502.06434.

Wu, S.-I. and R.-P. Lu (1993). Combining artificial neural networks and statistics for stock-market forecasting. Proceedings of the 1993 ACM conference on Computer science.

Zemouri, R., et al. (2003). "Recurrent radial basis function network for time-series prediction." Engineering Applications of Artificial Intelligence **16**(5-6): 453-463.

Zhao, B., et al. (2018). Short-term load forecasting based on RBF neural network. Journal of Physics: Conference Series, IOP Publishing.

Appendix A – [Code](https://github.com/lateefa788/Graduation-Project/blob/main/RFNNv3%20(1)%20updated%20(1).ipynb)

Appendix B – [Real Test Dataset](https://github.com/lateefa788/Graduation-Project/blob/main/saudi_stock_market_dataset.csv) [Training Dataset](https://github.com/lateefa788/Graduation-Project/blob/main/net-income-by-sector-main-market-from-first-quarter-2017-second-quarter-2019%20(1).xlsx)

Appendix C – Questionnaire

The designed Questionnaire for the expert in the Saudi stock market

|  |
| --- |
| 1. What is the field of companies that you would like to participate in? 2. What is your customary upper limit for the number of shares you participate in? 3. What is the methodology used when you contribute? 4. Do you follow the rise or fall of the shares of the target company? 5. What is the time range of expectancy for the model? 6. What accuracy do you hope for as a percentage? 7. What functionalities do you hope for in the model? |