VILNIUS UNIVERSITY

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

KUNAL BANERJEE

VERTES PROGNOZAVIMAS NAUDOJANT "DEEP LEARNING"	**FORCASTING OF STOCK VALUES USING DEEP LEARNING**

MASTER THESIS

Master Student: **Kunal Banerjee**

Academic Supervisor________________
Assoc.Prof. Dr.Diemante Teresiene

Work submission date: 05.01.2021
Registration No.__________________

2021

CONTENT

LIST OF FIGURES

LIST OF TABLES

Introduction :

With the globlalised and changing world, accurate and serene solutions of every problem becomes the basic need of the global populations. In the recent years, the Great Financial Distress (2008), changed the mentality of the people and they want to be warned beforehanded of such dismail distress. Every investor want to gain returns from the market with a low expose of risk associated them. So in Global Investment market, accurate prediction of the financial market in the future is obligatory now a days. Nevertheless standard methods prevail in predicting the true value of the stock such as Fundamental Analysis, which evaluate the past performance and credibility of the company and Technical Analysis which predict th future value based on the past value of the stocks. But with the advent of digitalization prediction of stock market moved towards different realms. The most appropriate methods involving in the prediction of stocks in Artificial Neural Networks (ANNs) which uses the back propagation of error algorithms.

Deep Learning is a subset of machine learning which help many researchers in their respective research to predict the outcomes. Hochreiter`s group (2014) used it to detect off target and toxic effects of environmental chemicals in nutrients. Similarly it also helps in predicting of the future stock price but the main problem lies within the complexity of the layers and training methods, which is the main aspect of this research paper.

Relevancy of the topic:

Stocks are volatile and the stock market had been studied over time to understand the trend or behavior pattern. The main motto behind it is to beat the market and maximize the profit. Most of the investors are unaware of the basic knowledge of trading but still invest as speculators. So there is always a need of a model which can accurately predict the value. The more accurate the model can predict, the more profit can be earned. This motivates people to study the market and also motivates me to do this research.

Research Problem:

How Deep Learning algorithm helps to predict the future stock price using the past values of the stocks?

The Goal of the Research:

The main goal of this research is to study the previous scientific articles and journals on deep learning and to analyze their model. This will help the author to develop a unique model using python code which can easily extract the future value of a stock using the past time series data and to attain the predicted value accurately. Moreover another main goal is to analyze the accuracy of the result through statistical instruments.

The Objective of the Research:

1. Study and analyze the previous scientific papers and journals and find out the methods that had been used to forecast the value of stocks.
2. Study the coding techniques in python as the layers of the model will be trained though python.
3. With the help of previous papers and journals, proper methodology will be implemented and appropriate coding will be done.
4. Examine the results of the research with actual close value of the to get the accuracy rate of the model.

Research Methodology:

To support this research, the theoretical value is taken from S&P stock exchange, it is a stock market index that measure 500 large companies in the United States. This index is market value weighted index whose components are weighted according to the total market value of the outstanding shares of each company. The time series values of different companies have been taken for twenty years and coding is been done in python and neural networks have been trained to put forward and validate and eventually reveal the true values of the stock.

Structure of the Research Paper:

The research paper has been divided into four parts. The first part is the literature review and theoretical approach, the second part is the Methodology which concentrates on the key variables, established hypothesis, sampling, research strategy, the method of data collection and extraction, the construction of the model, the third part is the analysis of the empirical results. It elaborates the results which have been received during the research work. The last part is the conclusion and further innovation in the same fields, where future researchers can work upon in the same field.

Limitation of the Research:

One of the biggest limitations is gathering the past data of each company to project the future value. In this volatile market, the data drastically change and gathering big data i.e. 20 years of historical value of each stocks of the company was a huge work. Moreover another big problem is technical as of there were no software available to the author to code in python. This technical problem has been eliminated by using online base for coding in python in Google collaborator, but still for processing such high volume of data, better processor needed to calculate the values faster and accurate.

There are many articles and journals about this topic but the main scientific journals which gave the idea of up taking such research are of authors Fank.C.Park and Eunsuk Chong (2017) where they implemented LSTM model to predict the trends of the KOSPI market exchange, which is q major stock exchange in South Korea. The model is endorsed with encoders and Boltzmann machine, which enable them to evaluate the trends to a high accuracy.

1. Theoretical Issues of Deep Learning in Stock Valuation Process:

This part of the thesis deals with the basic structure of the layers of the Deep Learning, its origin and defines the basic framework. Moreover it also gives insight knowledge of the previous scientific works by other authors and relates it for further study about the topic. The major concern of this part is the groundwork and preliminaries related to the subject and provide general overview about the topic of research.

Predicting the time series value of stocks is not a easy job for the analysts as stocks of a company always depends on various factors or in other words interdependent on many factors of the market; fundamental analysis and technical analysis have to be done and nowadays the market sentiments such as market news also may an imporatant role as it affect the stock value. Previously, there are various asset pricing models to estimate the return for one time period ahead and many different autoregressive models, such as ARIMA, vector autoregressive models, random walks, exponential smoothings and moving average for long time forecasting. But still there are anomaly persisting in the global market to predict the accurate value of the stock.

Recent development of technology, with advancement in computers, access to the Big Data, and creation and computational of complex algorithm opened the doors of Artificial Intelligence. Even though it is in the early phase of development , still it is gaining popularity in the professional world of traders, investment firms, portfolio managers, and bankers. With the discovery of more accurate programming such as machine language and complex networks such as deep learning, it is gaining much more attention now a days. In Financial sectors, it changes the landscape of the investment management, results in the optimisation of portfolio, mitigate risk and evaluate the different investment oppurtunities worldwide.

According to Steve Culp, Senior managing director of Accenture Finance and Risk services (2017), Artificial Intelligence will become the disruptive technology in the field of banking industry, as it reconstruct the traditional operative models and processes.According to IDC Research (2016), a survey had been done on companies executive and industry experts, which shows that AI will grow with compound annual growth rate (CAGR) of 55.1 % over the year 2016-2020. There are many companies which are implementing Artificial Intelligence, for example in portfolio management (Schwab Intelligent Portfolio(2017)), in algorithm trading (Renaissance Technologies(2017); Walnut Algorithms(2017)), loan and insurance underwritings, fraud detention and other new analysis.

Globally there are drastic changes in the field of Finance, and hence it is very apparent that there will be superior solutions in the performance of the financial instruments in contrast to traditional methods with the development and acceptance of AI in the financial sector. This study is dedicated to the research of future prediction stock exchanges of **S&P 500 index** from United States with supervised training technique of huge amount of historical data and in conjunction with the complex computational power of Deep Learning, which is a subset of Machine Learning, a part of Artificial Intelligence (AI). It is inspired be the structure and function of the brain called Artificial neutral networks, and concerned with the algorithms.

1.1 The origin of Deep Learning:

In the year 1943, Walter Pitts and Warren McCulloch built the first Thresholded Logic Unit, (TLU) which was based on neural networks of the human brain and can learn the AND and OR logic functions. This was an attempt to mimic the way biological neurons worked, through combination of mathematics and algorithms. Since then deep learning has evolved steadily. In the year 1957, Frank Rosenblatt picked up their idea with the develpoment of the perceptron, (it is a linear model binary classifier with a simple input output relationship), which took the concept of logical extent. Later in the decade AI researcher, Marvin Minsky discovered that the perceptrons are not sensitive to small changes and can easily be fallacy. Later Minsky and Seymour Papert claimed that the perceptrons and fundamentally flawed as they are incapable to handle certain problems non linear functions.

Henry J. Kelley in the year 1960 developed the basics of a continuous back propagation model and Stuart Dreyfus came up with the simpler version based on the chain rule in the year 1962. In 1986, AI researcher, Geoff Hilton, along with David Rumelhart and Ronald Williams published “ Learning representations by back propagating errors“ , where it had been detailed about the hidden layers of neurons to get around the problems faced by the percpeptrons. With sufficient data set and computing power they can compute the nonlinear functions something known as the “ Universal Approximation Theorem“. The approach works by backpropagating errors from higher layers of the networks to the lower layers, with enough data to train those layers and sufficient computing powers to calculate the interconnections.

In 1965, the earliest efforts of developing Deep Learning algorithms started by Alexey Grigoryevich Ivakhnenko and Valentin Grigor`evich Lapa.Kunijiko Fukushima designed neural networks with multiple pooling and convolution layer in the year 1979, called Neocognitron which used a hierarchical, multilayered design. This networks resembles the modern versions, but were trained with a reinforcement strategy of recurring activation in multiple layers.

In 1970, Back Propagation, which is the use of errors in training Deep Learning models evolved using a FORTAN code by Seppo Linnainmaa. However the concept was not applied to neural networks untill the year 1985. In the year 1989, Yann LeCun demonstrated the first practical application of back propagation at Bell Labs. He combined comvolutional neural networks with back propagation onto read handwritten digits. In 1995, Vladimir Vapnik and Dana Cortes developed the support vector machine, a system for mapping and recognizing similar data. Long short term memory or LSTM was developed in 1997 by Juergen Schmidhuber and Sepp Hochreiter for recurrent neutral networks.

Researcher Yann Lecun developed LeNet-5 at AT&T Bell Labs in 1998, which recognise handwritten images on checks using an iteration of this approach known as Convolution Neural Networks (CNNs). Eventually to get a deep network with many layers of trained neural nets and can understand the complex situations large data set and complex algorithms are needed. Hence here we need a machine learning, which allows the computer to learn using large data set instead of hard coded rules.

Around the year 2000, “features“ was discovered, which means lessons formed in the lower layers were not being learned by the upper layers, which was called Vanishing Gradient Problem. The source of this problem turned out to be certain activation functions. It was solved by layer to layer pre-training and the development of LSTM. However by 2011, the speed of GPU(Graphics Processing Unit) increased significantly and made it possible for CNN to train without layer to layer pre-training.

Currently, the processing of Big Data and the evolution of Artificial Intelligence are both dependent on Deep Learning, as there are unlimited scope in Deep Learning and in future it will evolve in enormous size.

1.2 The Basic Mechanism:

First of all, we have to understand what we meant by “learning“. In general it is defined as the process of gaining knowledge, experience or being taught. But to teach a machine, we need different structural descriptions or models which contains information about the raw data. Structural descriptions can take many forms for example, Decision trees, Linear regression and Neutral Network weights; in which there are different ways of applying the rules.

Linear algebra is the bedrock of machine learning and deep learning. Linear algebra, statistics (descriptive statistics and inferential statistics), calculus and with other basic mathematical geometry provide the mathematical underpinnings to solve the equations which is used to build the models. Fundamentally machine learning is based on algorithmic techniques to minimize the error in the equation through optimization (finding the value of an arguement that minimizes or maximizes a function) and solving systems of linear equations.

Deep Learning is a subset of machine learning method which allows to train the AI to predit output, given within the set of input. The training can be either supervised training, unsupervised training or reinforcement. The roots of the deep learning is in its neural networks which has been inspired by the biological neural networks of the brain. According to researchers, there are more than 500 trillion connections between neurons in the human brain. In contrast to it, the artificial neural networks dont even come close to such a drastic number.

Neural networks are the computational model which carry weights, which are the primary means of long term information storage and with updatation of such weights the neural networks learns new information. The behaviour of neural networks is shaped by its network architecture, which can be defined as the

- Number of neurons
- Numbers of layers
- Types of connection between layers

The evolution of artificial neuron was started with the development of single layer perceptron to multilayer perceptron. The artificial neuron of the multilayer perceptron is similar but it adds flexibility in the type of activation layer. These multilayer perceptron can solve the XOR logic function which was unsolved by the single layer perceptron. the difference have been shown in the figures below.

The word Deep in the deep learning refers to more than one hidden layers. Each connection between the neurons are associated with a weights, which dictates the importance of input value.There are large number of parameter and layers in one of four fundamental network architechture:

1. Unsupervised pretrained networks (UPNs)
2. Convolutional neural networks (CNNs)
3. Recurrent neural networks (RNNs)
4. Recursive neural networks.

Unsupervised pretrained networks (UPNs)

They have basically three layers : an input layer, a hidden layer (encoding) and a decoding layer. This network is trained in back propagation, which set a target value to be equal to the input values. Internally, it has a hidden layer that describes the code represent the inputs.

Convolutional neural networks (CNNs)

The CNN or ConvNet commonly applied to analyzing visual imaginary. The name indicates that the network employs a mathematical operation called convolution which is a special kind of linear operation. They are inspired by the biological visual cortex, and are regularized version of multilayer perceptrons which means each layer in one layer is fully connected to all neurons in the next layer.

Recurrent neural networks (RNNs)

Recurrent neural networks add additional weights to the network to create cycles in the network graph in an effort to maintain an internal state. Here layered topology of a multilayer perceptron is preserved, but every element has a weighted connection to every other element in the architecture and has a single feedback connection to itself.

Recursive neural networks.

It is created by applying same set of weights recursively over a structured input, and produce structured prediction over variable size input structures, or scalar prediction on it, by traversing a given structue in topological order. It is very useful in natural language processing, mainly phrases and sentences continuous representations based on word embedding.

1.3 Review of Major Preliminaries :

As aforesaid there were many researchers and scientists who researched about the movement of the stock variables. The most important work that has been done by Dimitrios Maditinos and Prodromos Chatzoglou (2012) compares the uses of artificial intelligence with the traditional and classic methods. Different methods and models are discussed which can be helpful for the prediction of the future real world complex problems. According to Wong *et al.* (1995) the most frequent areas of NN applications are production/operations(53.5%) finance (25.4%). According to Chakraborty *et. al.*, (1992); Theriou and Tsirigotis , (2000) the future behaviour of real world time series using NN can be predicted because it can learn nonlinear relationship between the input and output. If there can be the integration of knowledge and NN, complicated real time problems can be solved easily (Giles and Omlin, 1993). Another method is to represent prior knowledge in the form of error managements for the better training of neural networks (Abu-Mostafa,1993).

In the past few years, there are many surveys on the use of ANN to analyse the traditional classification and identification of problems in accounting and finance. Wong *et al,.*(1997) surveyed 203 articles from 1988 to 1995. The classification was done on the basis of year of publication, application area, journal and various other decision characteristics means of development. O`Leary (1998) analysed 15 articles that applied ANNs to predict corpoate failure bankrupcy. Similarly Zhang *et al.*, (1998) surveyed 21 articles that addressed modelling issues when ANNs are applied for forecasting and additional 11 articles that compares relative performance of ANNs.

The potential benefit of machine language in financial decision making was, first explored from a research perspective in Hawley *et al.* (1990) with a focus on applied neural networks as an aide to financial decision making. Echoing its future benefit to banking, a number of ealy studies also appeared in the Journal of Banking and Finance (1990), which explored the potential for ML to improve lending decisions and credit risk management. while Varetto (1998) built on his study by applying genetic learning algorithms to the same topic. These recent applications also include,the understanding of default recovery rates (Cheng and Cirillo, 2018); learning optimum hedging rates (Nian *et al.* , 2018)

Shaikh A.Hamid and Zahid Iqbal (2004) also compared the volatility forecasts from neural networks with implied volatility of S&P Index Futures options using the Barone-Adesi and Whaley American future option pricing model. The authors conclude that neural networks can predict the volatility specially when there is a ready availability of large mass of data. Similarly, Paolo Giudici (2018), tries to encourage the growth and development of financial technologies. He focused on the emerging topic of the financial technology; big data analytics, artificial itelligence and blockchain technology and their main application in banking, asset management and in paymwnt systems. The author was concerned and also focused on the risk involved with the projects of developing the technologies and to mitigate the involved risks. The author took the reference of Frontiers in Artificial Intelligence and share with the community the best practices to measure fintech risks. According to the author a framework have to be developed that can help both fintechs and supervisors.

The authors Luigi Troiano, Elena Mejuta and Prevesh Kriplani (2017) describe the trend predictions and mainly focused on Restricted Boltzmann Machines (RBM), rather than Auto-Encoders(AE), which is an alternative means of performing feature reduction. they investigate the comparison of the application of both RBM and AE and attempted to outline the architectual and input space charateristics that can affect the quality of prediction. The data had been taken from S&P 500 index from 01 Jan 2007 to 01 Jan 2017, with multiple technical indicators such as Absolute Price Oscillator (APO), MESA Adaptive Moving Average (MAMA) etc and to avoid a bias in result, the final comparison is done between RBM and AE through 10-fold cross validation.

Leonardo dos Santos Pinheiro and Mark Dras (2017) also used deep learning and evaluate RNN with character-level language model which had been pretrained for both intraday and interday stock market forecasting. The data had been taken from Standard and Poor`s 500 index, both for individual companies and overall index. Dai and Le(2015), proposed a technique to improve the training of RNNs with alanguage model. In this work this approach outerperformed training the same model from random initialization and achieved state of the art in several benchmaks. These were motivations for character-level language models, (Kim *et al.*,(2016)) and (Radford *et al*,.(2017)) showed the capacity of learning high level representations despite their simplicity. In this works the authors propose an automated trading system that, given the release of news information about the company automatically predict the changes of the stock prics. This system is trained to predict both changes in the stock price of the company as well as corresponding stock index too, here in this case S&P 500 index. It also consider the intraday changes and having an hour of window after the release of the news, and for changes between the closing price of the current trading session and the closing price of next day session. This model is pretrained by a character level language model.

Ritika Singh and Shashi Srivastava (2016), In the scientific paper from Indian School of mines(IIT), Dhanbad, evaluated on Google stock price multimedia data (chart) from NASDAQ. The objective of the study is to demonstrate that deep learning can improve the forecasting accuracy. They compare 2-Directional 2- Dimensional Principal Component Analysis (2D)^2PCA + Deep Neural Network (DNN) method with state of the art method 2-Directional 2- Dimensional Principal Component Analysis (2D)^2PCA + Radial Basis Functional Neural Network (RBFNN). The result was also compared with Recurrent Neural Network (RNN) and it is found that the accuracy for hit rate is improved by 15.6%.

Ariel Navon and Yosi Keller (2017) evaluate the prediction of time series data of stocks from NASDAQ. Their approach was to input the raw data to the deep learning network and predict the temporal trends of the stocks. They derive the probabilistic output of the NN which had been applied to the raw data and optimizes the average return from the result of the research. The performance of the model compares favorably with contemporary benchmarks of two years backtesting. Similaly Sedya Kalyoncu et al. (2020) develop a robust stock prediction model which can predict the future trend of the stocks.fo the research they consider the Turkish companies and applied LSTM model to gather the prediction of the companies stocks. In the accuracy test, they used RMSE and MSE and evaluate that LSTM can predict stock trend to a greater extent.

Moreover, Ryo Akita, Akira Yoshihara and Takashi Matsubara (2016) evaluate the stock prediction using both the quantitative and qualitative data using Paragraph Vector and Long Short Term Memory (LSTM). According to the authors, investors make financial decisions which depends on various factors such as consumer price index, price earnings ratios, and miscellaneous events reported in social media. They took Tokyo Stock Exchange, for their research and fifty companies have been taken to imply the research method. The qualitative information had been taken by Paragraph Vector and the quantitative information through LSTM. They conclude that the textual method is better in evaluating the stock value than only taking the numeric past data trends.

Kaustubh Khare and Omkar Darekar (2017) contribute a considerable measure to the unpredictability of the security market. They consider 10 unique stocks from New York Stock Exchange, develop a short term deep learning model to evaluate the short term fluctuation of the aforesaid stocks. This paper discusses about two t sorts of Artificial Neural Networks, one Feed Forward Neural Networks and another Recurrent Neural Networks. The review uncovers that Feed Forwards Multilayer Perceptron perform superior to Long Short-Term Memory, at predicting the short - term prices of a stock.

Moving further, Wei Bao, Jun Yue and Yulei Rao (2017), proposed model which consists of three parts in their research articles. It is a neutral network consisting of multiple single layer autoencoders in which the output feature of each layer is wired to the input of successive layer,(Chen Y. *et al* ,(2014)). The two methods which are incorporated to help increase predictive accuracy. LSTM is a type of recurrent neutral network, with feedback attached to some layers of the networks. Kai Chen, et al (2015) presented paper which is modeled and predicted China stock returns using LSTM. The historical data of China Stock Market were taken for review and transformed into 30-days-long sequences with 10 learning features and 3-day earning rate. They compared with random prediction method, the LSTM model improved the accuracy of stock returns prediction from 14.3% to 27.2%. This effort demonstrated the power of LSTM in stock market prediction in China.

Salvatore Carta and Andrea Corriega (2020), develop a model based on deep learning and deep reinforcement learning to predict the stock value. The data had been extracted ffrom S&P 500 index and several meta learning decisions were infused to prepare a robust model which can process the data fast and can generate stock signals through different iterations. The proposed model showed a great outcome as they evaluate the prediction to a greater extent and revealed that it replicates the convectional buy-hold strategy of the market. In a similar experiment K.P.Soman et al. (2017), used RNN and CNN sliding window model and fused it with LSTM to predict the future trends of the price of the market. They considered the NSE exchange and apply the sliding window model approach to predict the short fluctuations in the exchange. The outcome was appropriate and tested through percentage error method, which revel that LSTM models are appropriate for the prediction of the future trends of the stock markets.

The data was taken as six stock indices to test the prediction ability of the model: CSI 300 index from China, Nifty 50 index from India, Hang Seng index from Hong Kong, Nikkei 225 index fromTokyo, S&P 500 index and DJIA index from New York. Technically WSAEs-LSTM model was applied to every indices to forcast the movement of each stock and to observe the prediction of stock moving trends. For each stock index, three types of variables are used,first is the historical stock trading data, second is the technical indicators of stock trading and third is the macro economy variables.

The predictive method (Chan et al.,(2016)), is applied to get the predicted stock index, then evaluating in two dimentions: the predictive accuracy and profitability. The accuracy of the predictive value is evaluated by using three measurements, mean absolute percentage error (MAPE), correlation coefficient (R) and Theil`s inequality coefficient (Theil U). To capture the performance of WSAEs-LSTM other three models have been used, which are WLSTM , LSTM and also conventional RNN.

Moreover Lavrenko *et al.,(*2000) Combined the trends of stock exchange and financial news articles, and predict the trends using the content of news articles.Schumaker and Chen (2009), compared several different textual representations : Bag of Words, Noun phrases, and Named entities for stock price prediction. They showed that support vector machines (SVM) with proper noun features is superior in predicting the trnd of stock prices. Hagenau *et al,.(*2012), predicted the difference of open and close prices of a stock with the data from DGAP and EuroAdhoc whci are corporate announcements of Germany and UK respectively, by using bigram(two adjacent words and two word combination as features) and conducted selection according to Chi-Square statistics with respect to each brand of stock prices. Ding *et al,.*(2014), employed Open IE (Information Extraction) techniques to extract the actor and object of events from titles of news articles and predicted the S&P 500 index, by using deep neutral network model as classifier and achieved better than SVM model.

Guanhao Feng, Jingyu He and Nicholas G.Polson (2018) uses multi layer deep learning searches for nonlinear factors fo pedicting returns, such as rectified linear units (ReLU) or long-shrt term memory for time series effects. Deep Learning is capable of extracting nonlinear factors and provides a powerful alternatives to features selection and shrinkage methods, Feng *et al.,(*2017). In the research works of Kozak et al.,(2017); Gu and Xiu (2018) showed the machine learning based predictors in emperical finance. Feng *et al.,(*2018), predicts cross sectional returns with deep learning in a portfolio context.Improved methods are also suggested by Campbell and Thompson(2007), Rapach *et al.*,(2010) and Harvey *et al,*.(2016).

To conclude a huge number of models has been generated till date and those proposed models are not comparable because of the extraction of data either structured or unstructured. However in every model, which uses the LSTM, showed a greater efficiency in predicting the stock values. Many researchers have been successful in predicting the movement of the variables of the stock market with their own perceived theory and knowledge based models. Most of the model proved that they better and can generate more profitable approach than the traditional approach. Moreover these models are using various techniques and approaches to guarantee the accuracy of their predictions. However there is no fixed model for this purpose because there are many attributes involved which affect the accuracy of their predictions. But it has been proved that there is a commendable future of such prediction models in the future, and with time it will always be a scope for more research.

2. Methodology:

There are generally three types of training of data in deep learning; supervised, unsupervised and reinforcement. In this paper the author will use a supervised training technique to set the historical data from the particular index to the neuron layers, through coding in python. Python gives a general coding where the layers can be increased according to the users and appropriate tools are inbuild which help in training of historical data.

Here the historical data of different stocks from various sectors have been extracted from the S&P500 Index, and training will be provided to the layers of neurons in the deep learning model. It generally follows three steps which are generally not distinct:

1. Training : A training set of correct behaviour is analysed and some representation of the newly knowledge is stored which is often is some form of rules.
2. Validation : The rules are checked and if necessary then extra training will be given. Sometimes additional test data are also used.
3. Application: the rules are used in responding some new situations.

After such completion of training the neurons of the model of deep learning will have the capacity to predict the stocks for the future because predefined rules have been set in the neurons as activation key.

2.1 The Basic Architecture:

The prediction of the market value is of great importance to help in maximizing the profit of the stock while keeping the risk of such stocks at low levels. Recurrent neural networks (RNN) have proved one of the most powerful tool or models for processing the time series sequencing data. Within this the LSTM (Long Short-term memory) is one of the most successful RNNs architectures. LSTM has feedback connections, which is not like feed forward neural networks. It is generally composed of a cell, an input gate, an output gate and a forget gate.

The LSTM is used to classify, process and make predictions based on the time series data, since there can be lags of unknown duration between important events within the time series. It deals with the vanishing gradient problem, which is encountered while training the ANN with gradient based learning methods and back propagation. In this method each of neural network`s weights receive an update proportional to the partial derivatives of the error functions with respect to the current weight in each iteration of training. The input gate controls the extent to which a new value flows into the cell and the forget gate controls the extent to which a value remains in the cell, whereas the output gate controls the the extent to which the value in the cell is used to compute with the activation functions.

2.2 Established Hypothesis:

The hypothesis which has been taken in this research is that: **If** the historical stock prices are taken as big data and **Then** put it in deep learning neural networks, can it be possible to predict the accurate closing price of the stocks through the python programming.

If the hypothesis agrees, then what will be the accuracy rate of the predicting stock values? The whole research is been done taking this factors in mind. The future closing price can be detected by the coding done in python but to show the accuracy, the author will use regression analysis.

2.3 The steps of analysis:

There are various types of neural networks that can be developed by the combination of different factors like network topology, training method etc. In this research paper the author considered the use of LSTM (Long Short-Term Memory) with the activation of Rectified Linear unit (ReLU).
There are several stages in the process of methodology which can be analyzed in a sequential series.

2.3.1 Raw Data Collection:

The historical data of nearly 90 companies have been collected from the Yahoo Finance of S&P 500 index, by using the stock ticker name. Yahoo finance facilitates to download the historical data between any two given dates. The companies have been taken from different sectors of the economy. The closing price of every company have been recorded daily with the range **01.01.2000 to 10.12.2020**. This data will be used to predict the future stock prices of the companies which have been considered. To evaluate the proposed system, the author has trained and tested the technique over the publically available stocks. The collected data contains daily stock trends over the period of 20 years, which can be termed as the Big Data. The attributes of the data set has been given in Table 1.

Features	Description
Date	Corresponding Date of Stock values
Open	Opening price of stock on Particular date
High	Highest selling stock value of the day
Low	The lowest value of the selling price of the stock on a given date
Close	Contains closing value of the stock on the given date
Volume	The number of shares traded on the given day
Adjusting Close	The closing price of the stock after paying dividends to the investors

Table 1 : Yahoo Finance Data Set
Source: Author`s Worksheet

2.3.2 Data Processing & Extraction and Training Neural Networks:

Out of the huge dataset, the data have been filtered and particularly 90 companies have been considered for the research paper. The filtration has been done keeping in mind the numerical availability of the data and with reference of the other input within that time frame. After the dataset have been transformed into a clean data set, it has been divided into training data, with which the neurons will be trained accordingly for the full period of experiments and the other part of the data have been kept for the testing or validation purpose, so it is called the validation set or test set.

The data is fed to the neural networks and trained for prediction with random weights and biases. The training that the author done in such a way that the system minimizes the error and improve the performance of the model. There are many training algorithms but the most important is the back propagation algorithm. In this paper multilayer feed forward network have been established with back propagation algorithm.

Back propagation is important because hidden layers are having no training target value and they must be trained based on the errors from the previous layers. The output layer is the only layer with has a target value to compare. As the errors are back propagated through the nodes the connection of the weights are changed accordingly until the errors in the weights are sufficiently small to be accepted.

The architecture of the model includes an input layer followed by 3 hidden LSTM layers with ReLU activation and a dense output layer with linear activation function. Here a deep learning library "Keras "have been used with python coding because it's a high level in nature but very simple to use. Within the frame two modules have been created, the training and the test module. The testing data has been kept as 5-10 percent of the dataset.

The flowchart of the problem has been defined accordingly

- Importing the functions from the library available in the python.
- Getting the stock Quotes.
- Get the number of Rows and Column in the dataset.
- Visualizing the closing price history.
- Creating a new data frame with only the 'close column'.
- Scaling the data.
- Creating the training data set.
- Converting the x_train and y_train to numpy array.
- Reshaping and building the data into three dimensional to apply LSTM.
- Compile the model & Train the model
- Creating a testing data set & Reshaping the data.
- Get the model predicted price values & Get the RMSE value
- Plot the Data.
- Show the valid and predicted price.
- Get the stock Quote.

Analysis:

To analysis the efficiency of the model, to common attributes were taken one the Root Mean Square Error (RMSE) is used. It is calculated as in equation (a)

$$rmse = \sqrt{} \sum_{1}^{n}(Pi - Oi)^2. \qquad \text{(a)}$$

Where P= Predicted Value

O= Observed Value

It is the standard deviation of the prediction errors. Residuals are a measure of how far the points are from regression line data. RMSE measure the spread of the residuals. This has been calculated within the coded program. Another attribute was taken R^2, which is calculated through the excel as in equation (b)

$$R^2 = 1 - (\frac{RSS}{TSS}) \qquad \text{(b)}$$

Where RSS= Sum of Square of Residuals

TSS= Total Sum of Squares

Whereas R-squared is a relative measure of fit, RMSE is an absolute measure of fit.

3. The Usage of Deep Learning in Different Sectors:

In this part of the thesis the final analysis and the results have been evaluated. A stock market sector comprises of many group of stocks that have a lot of common features in common with each other, usually because they are in similar industries. There are 11 different stock market sectors, according to the most commonly used classification system: the Global Industry Classification Standard (GICS). Out of which, nine sectors have been choosen and ten companies from every sector have been selected according to inception date of the company and availability of the data in the yahoo finance

As per the methodology,every time series stock values have been put in the coding model and the predicted price have been taken out as on 10th december 2020. While selecting the companies one of the important impact was the availability of the stock price, as in the model big data have been taken to validate the model. Out of twenty years of time series data of stock value, 60% have been put to test the model and rest is to validate and predict the outcome. Moreover to illustrate the actual outcome of every company, particular graph have been shown in the appendixes which can be referered in the Ref. column of the table.

3.1 Sector- Communication Services:

In the S& P stock exchange, this sector is very much new and comprises of telecommunication services (both wireless and fixed landline) and media and entertainment companies. Out of numerous companies 10 companies have been selected as per the availability of the big data, as shown in Table 2. As we can see the RMSE value is quite low in most of the companies except The Walt Disney company, (Rmse = 16.92) the reason can be the data point might have a high separation value which the model cannot predict properly and hence there is a huge difference in the actual close price and predicted close price.

But if we observe the Bar graph in figure 1 and the plot diagram in the predicted close in figure 2 we can observe that the R^2 = 0.9976, which is very close to 1, illustrate that the model can predict to an great efficiency.

	Security	Ticker Symbol	Sector	Sub Industry	RMSE	Actual Close	Predicted Close	Ref	Headquater, Location
1	Discovery Inc (Class A)	DISCA	Communication Services	Broadcasting	0.356	28.79	27.86	Graph 24	Silver Spring, Maryland
2	Discovery Inc (Class C)	DISCK	Communication Services	Broadcasting	0.477	25.48	25.42	Graph 25	Silver Spring, Maryland
3	Dish Network	DISH	Communication Services	Cable & Satellite	0.988	36.20	36.77	Graph 26	Meridian, Colorado
4	Interpublic Group	IPG	Communication Services	Advertising	0.233	23.63	23.45	Graph 47	New York
5	Netflix Inc.	NFLX	Communication Services	Movies & Entertainment	1.95	501.08	493.95	Graph 48	Los Gatos, California
6	Verizon Communication	VZ	Communication Services	Integrated Telecommunication Services	0.193	60.50	61.55	Graph 49	New York
7	The Walt Disney Company	DIS	Communication Services	Movies & Entertainment	16.92	154.69	176.02	Graph 50	Burbank, California
8	Omnicom Group	OMC	Communication Services	Advertising	1.21	63.95	65.15	Graph 51	New York
9	AT & T Inc	T	Communication Services	Integrated Telecommunication Services	0.307	30.69	31.10	Graph 73	Dellas, Texas
10	Comcast Corp.	CMCSA	Communication Services	Cable & Satellite	0.442	50.48	51.73	Graph 74	Philadelphia, Pennsylvania

***Table 2*: Final Result of 10 Companies of sector " Communication Services"**

***Source*: Author`s worksheet**

After observing the prediction outcome (Table 2) it is clear that the RMSE value ranges from (0. 3 to 1.95),except The Walt Disney Company which is having high value of 16.92, nevertheless the R^2 is nearly 1 and referring to plot diagram, (figure 2) all values are coincide with the exact predictions. Even to understand the prediction of such companies, actual graphs are provided (Annex 2)

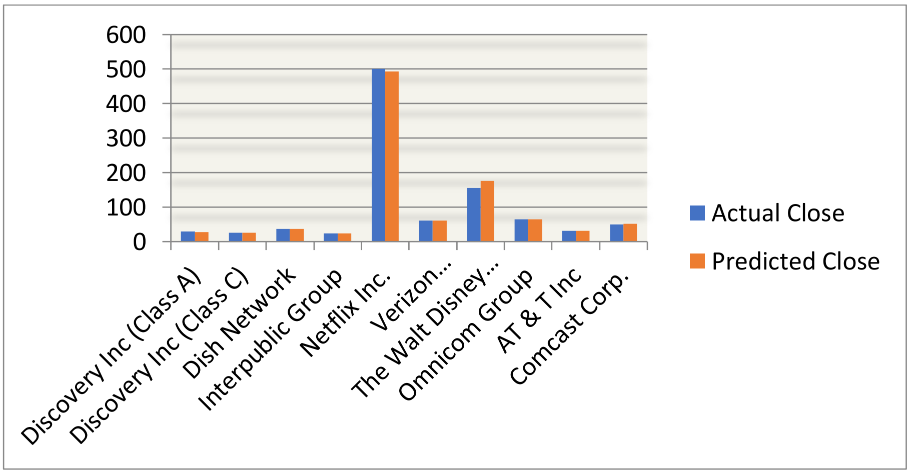


***Figure 1*: Bar Graph showing the prices of "Communication Services"**

***Source*: Author`s worksheet**

This graph (figure 1) is showing the accuracy of the Actual and Predicted stock value which had been extracted through the process, in x axis the name of the companies are provided for better understanding of the values calculated.

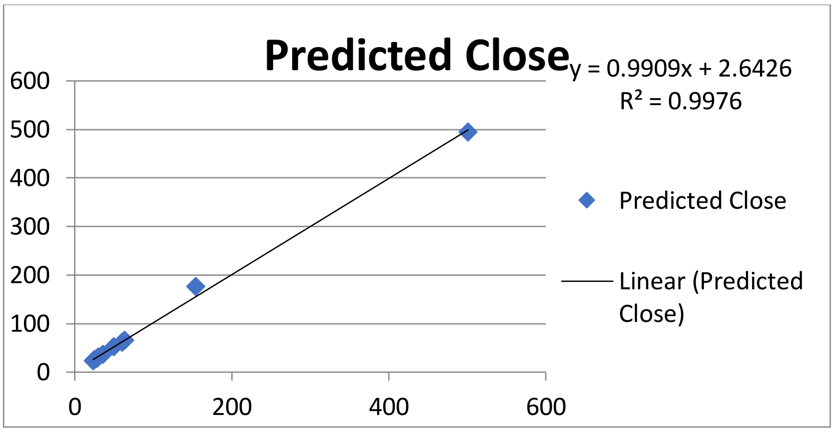


***Figure 2*: Plot Diagram of predicted close“Communication Sector“**

***Source*: Author`s worksheet**

After considering the result (Table 2) and applying the R^2method in the Real Estate, (R^2 =0.9976) it is very much observed that most of the predicted close are coinciding with the regression line and the model is having high accuracy in this sector.

3.2 Sector- Financials:

In the S&P stock exchange, this sector comprises of business which handles money, mainly the banks and financial intermediataries such as insurance, brokerage house mortgage related companies, consumer finance and many more. one of the important sector is Financials,the seleted companies are shown in Table 3. As we can see the RMSE values in all the companies are low ranging from 0.067 to 1.68 and there is a close predicted price in most of the companies.

	Security	Ticker Symbol	Sector	Sub Industry	RMSE	Actual Close	Predicted Close	Ref	Headquater, Location
1	Aflac Inc.	AFL	Financials	Life & Health Insurance	0.437	45.34	45.68	Graph 7	Columbus, Georgia
2	Allstate Corp.	ALL	Financials	Property & Casualty Insurance	0.255	105.41	103.65	Graph 11	Northfield Township, Illinois
3	American Express	AXP	Financials	Consumer Finance	0.067	121.83	122	Graph 15	New York
4	Bank Of America Corp.	BAC	Financials	Diversified Banks	1.19	29.11	30.06	Graph 21	Charlotte, North Carolina
5	Cboe Global Market	CBOE	Financials	Financial Exchanges & Data	1.22	86.94	93.16	Graph 32	Chicago, Illinois
6	Fifth Third Bancorp	FITB	Financials	Regional Banks	0.377	27.78	26.76	Graph 52	Cincinnati, Ohio
7	Globe Life Inc.	GL	Financials	Life & Health Insurance	1.068	94.87	93.21	Graph 53	McKinney, Texas
8	Hartford Financial Svc. Gp	HIG	Financials	Property & Casualty Insurance	1.038	47.40	45.60	Graph 54	Hartford, Connecticut

9	JPMorgan Chase & Co.	JPM	Financials	Diversified Banks	1.68	120.26	123.17	Graph 55	New York
10	KeyCorp.	KEY	Financials	Regional Banks	0.638	16.08	16.50	Graph 56	Cleveland, Ohio

***Table 3*: Final Result of 10 Companies of sector " Financials"**

***Source*: Author`s worksheet**

After observing the prediction outcome (Table 3) it is clear that the RMSE value ranges from (0.06 to 1.68), and referring to plot diagram, (figure 4) all values are coincide with the exact predictions, with slight difference in some rare cases. Even to understand the prediction of such companies, actual graphs are provided (Annex 2)

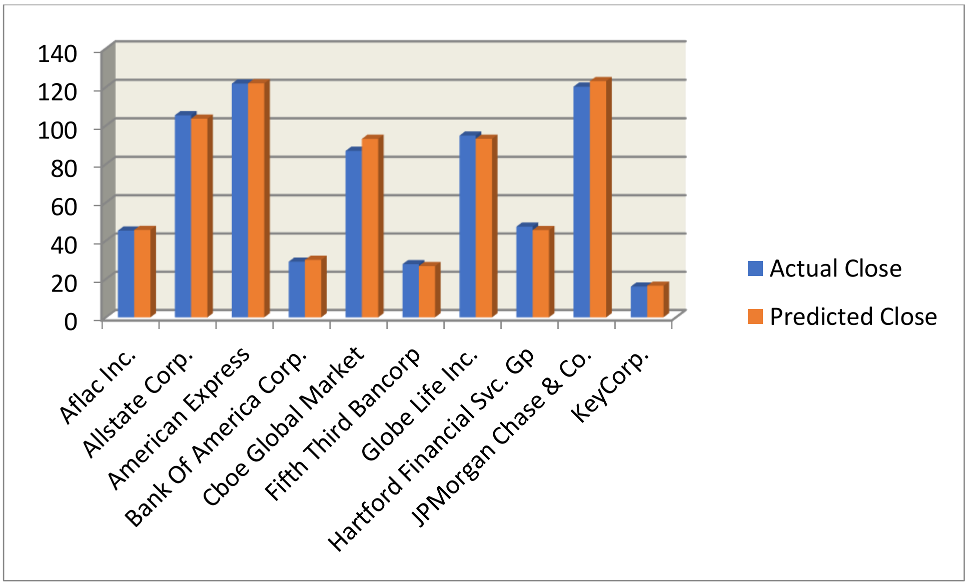


***Figure 3*: Bar Graph showing the prices of "Financials"**

***Source*: Author`s worksheet**

This graph (figure 3) is showing the accuracy of the Actual and Predicted stock value which had been extracted through the process, in x axis the name of the companies are provided for better understanding of the values calculated.

If we take a close view of the figure 2, mostly every company`s actual close collides with the predicted close. So in this sector the model is very accurate in predicting the close price of the selected companies. Even the R^2= 0.9964, showing its strong correlation.

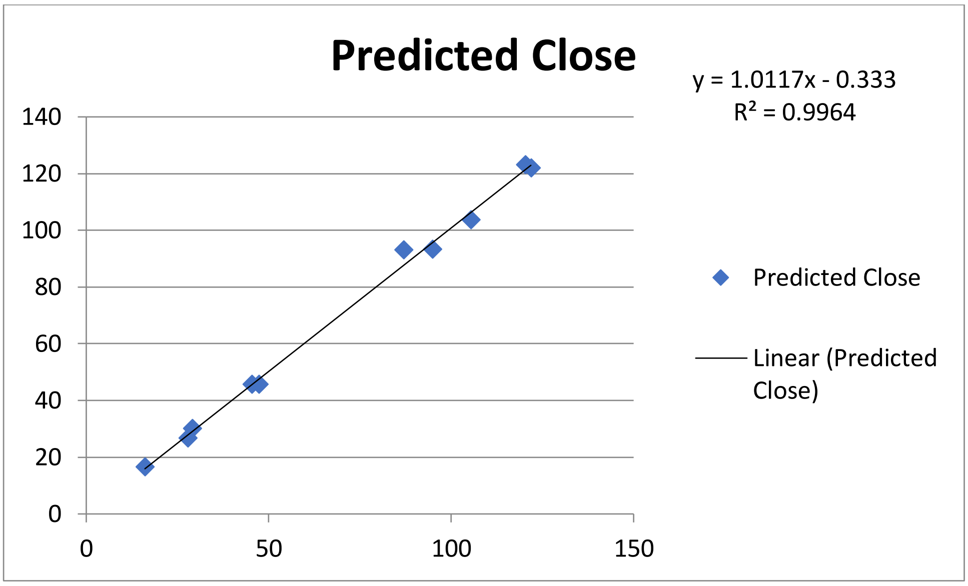


***Figure 4*: Plot Diagram of predicted close"Financials"**

***Source*: Author`s worksheet**

After considering the result (Table 3) and applying the R^2method in the Real Estate, (R^2 =0.9964) it is very much observed that most of the predicted close are coinciding with the regression line and the model is having high accuracy in this sector.

3.3 Sector- Energy:

The next sector is Energy which comprises mainly of Oils and Natural gas companies. It also comprises of production of coal and ethanol, but this sector does not include any renewable energy companies, which actually comprises in utilities sectors.

	Security	Ticker Symbol	Sector	Sub Industry	RMSE	Actual Close	Predicted Close	Ref	Headquater, Location
1	Baker Hughes Co	BKR	Energy	Oil & Gas Equipment & Services	0.751	23.12	22.45	Graph 29	Houston, Texas
2	Cabot Oil & Gas	COG	Energy	Oil & Gas Exploration & Production	0.209	17.65	16.6	Graph 30	Houston, Texas
3	Apache Corporation	APA	Energy	Oil & Gas Exploration & Production	4.26	16.50	18.30	Graph 39	Houston, Texas
4	Chevron Corp	CVX	Energy	Integrated Oil & Gas	2.87	93.34	92.59	Graph 40	San Ramon, California
5	Concho Resources	CXO	Energy	Oil & Gas Exploration & Production	4.89	63.95	57.46	Graph 41	Midland, Texas
6	ConocoPhillips	COP	Energy	Oil & Gas Exploration & Production	0.482	44.00	41.9	Graph 42	Houston, Texas
7	Devon Energy	DVN	Energy	Oil & Gas Exploration & Production	0.066	16.57	16.35	Graph 43	Oklahoma City, Oklahoma
8	EOG Resources	EOG	Energy	Oil & Gas Exploration & Production	2.48	54.68	51.05	Graph 44	Houston, Texas
9	Exxon Mobil Corp.	XOM	Energy	Integrated Oil & Gas	0.891	44.00	41.35	Graph 45	Irving, Texas
10	Halliburton Co.	HAL	Energy	Oil & Gas Equipment & Services	0.841	19.99	19.65	Graph 46	Houston, Texas

***Table 4*: Final Result of 10 Companies of sector " Energy"**

***Source*: Author`s worksheet**

The seleted companies are shown in (Table 4). As we can see the RMSE values in all the companies are ranging from (0.209 to 4.89) and the prediction value is still close with a normal difference of 2 dollars close price.Even to understand the prediction of such companies, actual graphs are provided (Annex 2)

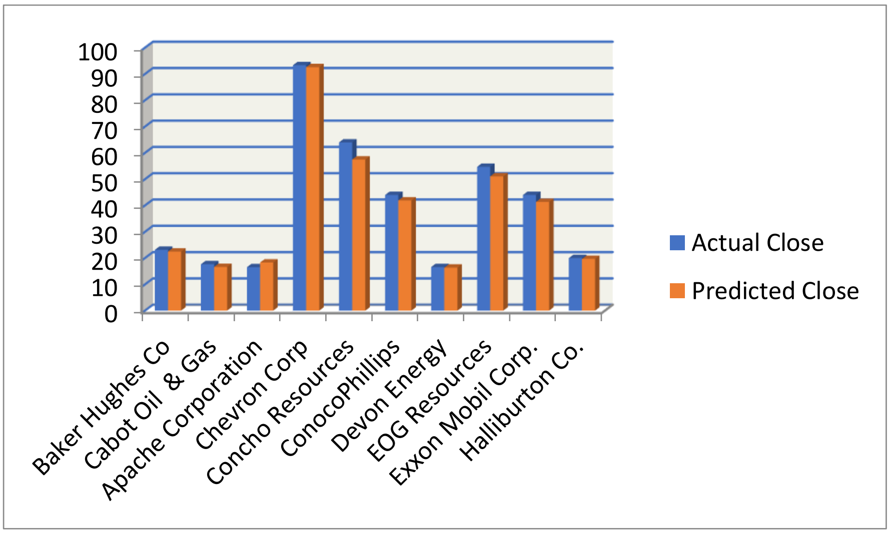


***Figure 5*: Bar Graph showing the prices of "Energy"**

***Source*: Author`s worksheet**

This graph (figure 5) is showing the accuracy of the Actual and Predicted stock value which had been extracted through the process, in x axis the name of the companies are provided for better understanding of the values calculated.

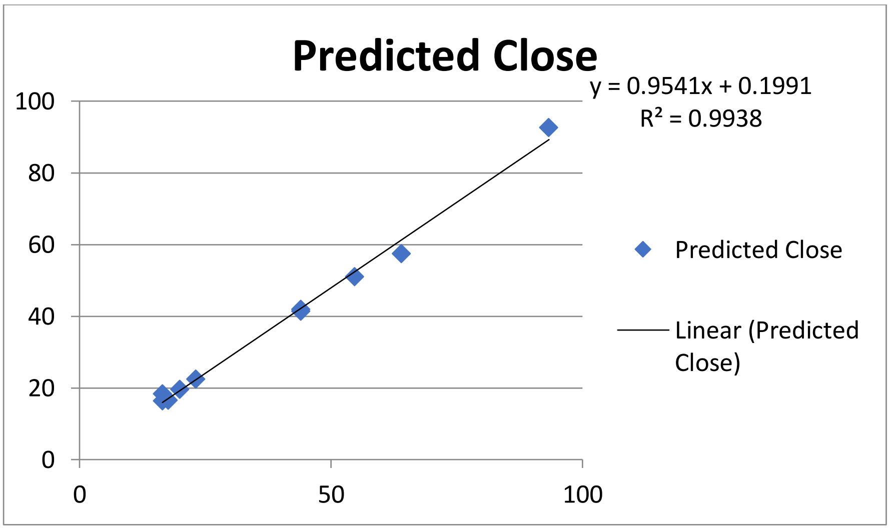


***Figure 6:* Plot Diagram of predicted close"Energy"**

***Source*: Author`s worksheet**

After considering the result (Table 4) and applying the R^2method in the Real Estate, (R^2 =0.9938) it is very much observed that most of the predicted close are coinciding with the regression line except Concho Resources, having high RMSE value.

3.4 Sector- Healthcare:

It mainly has two components, one include companies that develop the pharmaceuticals and the other is the other is suppliers of healthcare and equipment services. The seleted companies are shown in Table 5. As we can see the RMSE values in all the companies are relatively high ranging from 0.92 to 6.711.

	Security	**Ticker Symbol**	**Sector**	**Sub Industry**	**RMSE**	**Actual Close**	**Predicted Close**	**Ref**	**Headquater , Location**
1	Abbott Laborities	ABT	Health Care	Health Care Equipment	1.033	106.51	102.4	Graph 2	North Chicago, Illinois
2	AbbVie Inc.	ABBV	Health Care	Pharmaceuticals	1.169	107.48	104.6	Graph 3	North Chicago, Illinois
3	Abiomed Inc.	ABMD	Health Care	Health Care Equipment	6.711	267.92	264.16	Graph 4	Danvers, Massachuset ts
4	Amgen Inc.	AMGN	Health Care	Biotechnology	2.58	228.16	226.43	Graph 16	Thousand Oaks, California
5	Johnson & Johnson	JNJ	Health Care	Pharmaceuticals	4.31	152.25	145.45	Graph 57	New Brunswick, NJ
6	Lilly (Eli) & Co	LLY	Health Care	Pharmaceuticals	2.34	161.00	148.2	Graph 58	Indianapolis , Indiana
7	Medtronic Plc.	MDT	Health Care	Health Care Equipment	0.134	112.68	112.81	Graph 59	Dublin, Ireland
8	Merck & Co.	MRK	Health Care	Pharmaceuticals	0.92	82.98	81.95	Graph 60	Kenilworth, NJ
9	PerkinElmer	PKI	Health Care	Health Care Equipment	3.89	145.91	134.07	Graph 61	Waltham, Massachuset ts
10	Pfizer Inc.	PFE	Health Care	Pharmaceuticals	0.89	41.73	40.08	Graph 62	New York

***Table 5*: Final Result of 10 Companies of sector “ Health Care“**

***Source*: Author`s worksheet**

After observing the prediction outcome (Table 5) it is clear that the RMSE value ranges from (0.13 to 6.71), and referring to plot diagram, (figure 8) but still most of the values are coincide with the exact predictions, and having R^2 nearly to 1. Even to understand the prediction of such companies, actual graphs are provided (Annex 2)

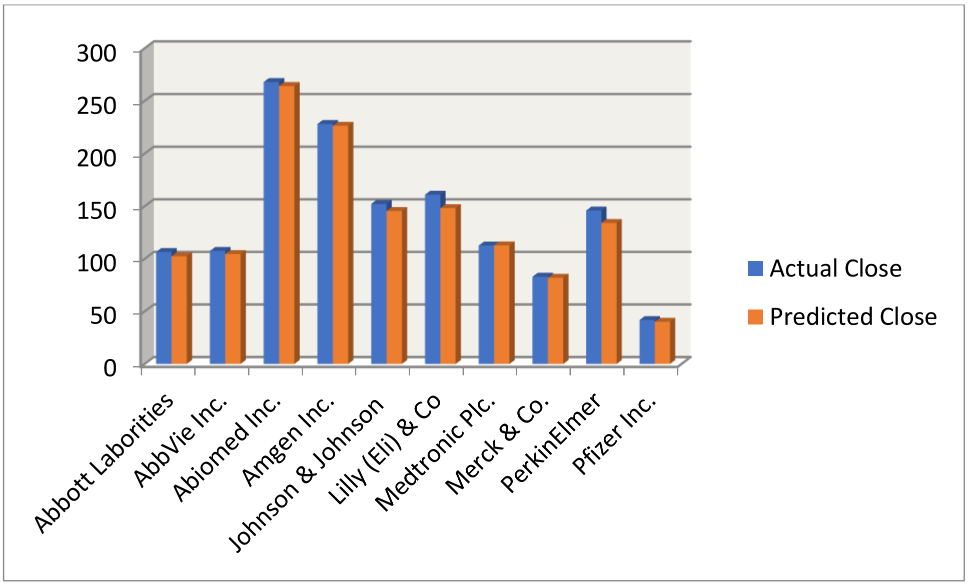


Figure 7: **Bar Grapg showing the prices of "Health Care"**

Source**: Author`s worksheet**

This graph (figure 7) is showing the accuracy of the Actual and Predicted stock value which had been extracted through the process, in x axis the name of the companies are provided for better understanding of the values calculated.

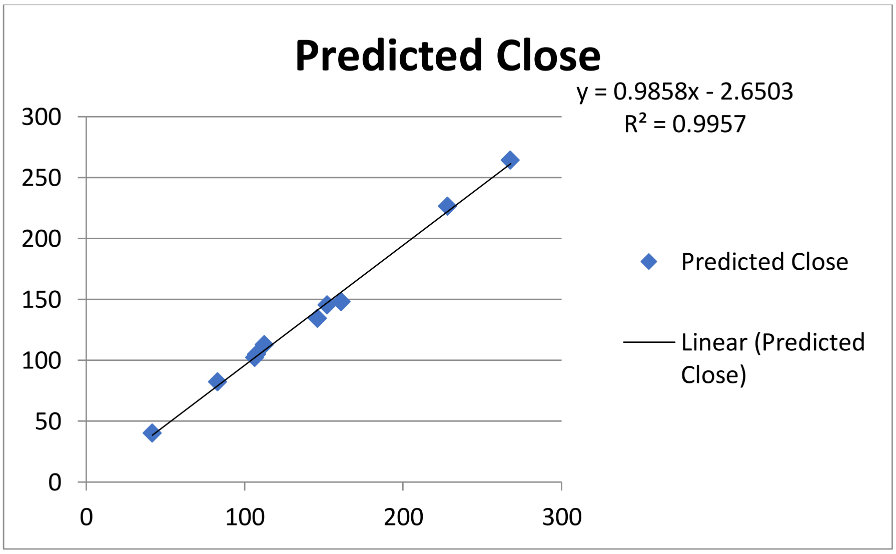


Figure 8**: Plot Diagram of predicted close"Healthcare"**

Source**: Author`s worksheet**

After considering the result (Table 5) and applying the R^2method in the Real Estate, (R^2 =0.9957) it is very much observed that most of the predicted close are coinciding with the regression line and the model is having high accuracy in this sector.

3.5 Sector- Industrials:

The next sector that has been choosen for the analysis is Industrials, which comprises of different companies which involves the uses of heavy equipments, such as railways, airlines , logistic companies and even companies like building products, electrical equipments are in this sector and hence it is the back bone of every index having huge volume of trading per day,mostly the big companies have been seleted, shown in Table 6. As we can see the RMSE values in all the companies are low and predicting the accurate close price for most of the selected companies.

	Security	Ticker Symbol	Sector	Sub Industry	RMSE	Actual Close	Predicted Close	Ref	Headquater, Location
1	3M Company	MMM	Industrials	Industrial Conglomerates	0.664	173.49	173.25	Graph 1	St. Paul, Minnesota
2	Alaska Air Group Inc	ALK	Industrials	Airlines	2.95	51.29	49.55	Graph 9	Seattle, Washington
3	American Airlines Group	AAL	Industrials	Airlines	0.424	17.99	17.01	Graph 13	Fort Worth,Texas
4	Equifax Inc.	EFX	Industrials	Research & Consulting Services	1.93	183.38	170.47	Graph 28	Atlanta, Georgia
5	Catepillar Inc.	CAT	Industrials	Construction Machinery & Heavy Trucks	0.898	178.85	177.05	Graph 31	Deerfield, Illinois
6	General Dynamics	GD	Industrials	Aerospace & Defense	1.43	150.36	151.50	Graph 63	Falls Church, Virginia
7	General Electric	GE	Industrials	Industrial Conglomerates	0.022	11.32	10.94	Graph 64	Boston, Massachusett s
8	Honeywell Int`l Inc.	HON	Industrials	Industrial Conglomerates	3.49	211.97	198.23	Graph 65	Morristown, NJ
9	Illinois Tool Works	ITW	Industrials	Industrial Machinery	4.52	202.55	212.32	Graph 66	Glenview, Illinois
10	Cummins Inc.	CMI	Industrials	Industrial Machinery	2.23	219.55	216.49	Graph 75	Columbus, Indiana

***Table 6*: Final Result of 10 Companies of sector " Industrials"**

***Source*: Author`s worksheet**

After observing the prediction outcome (Table 6) it is clear that the RMSE value ranges from (0.02 to 4.52), and referring to plot diagram, (figure 10) most of the values are coincide with the exact predictions, except two which are having high RMSE.Even to understand the prediction of such companies, actual graphs are provided (Annex 2)

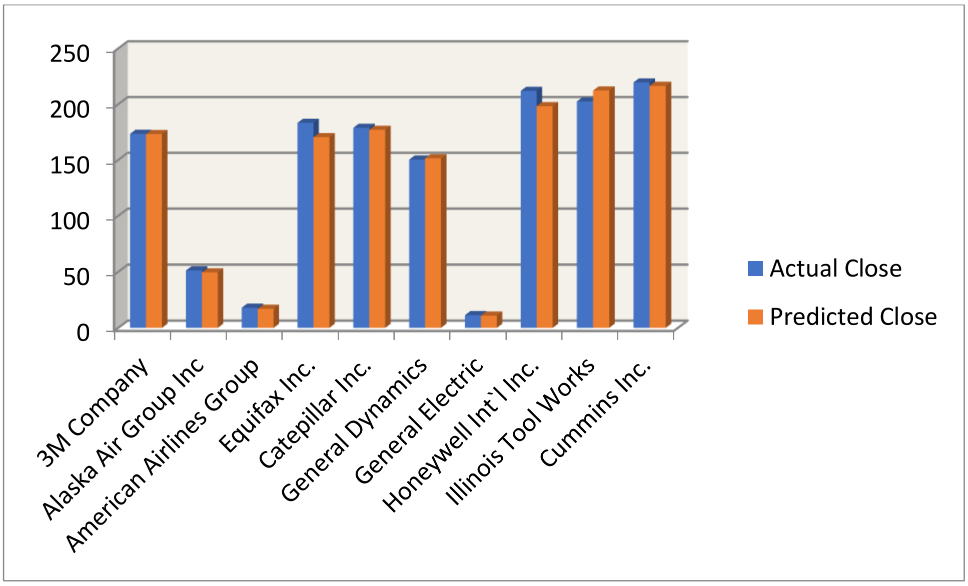


***Figure 9*: Bar Grapg showing the prices of "Industrials"**

***Source*: Author`s worksheet**

This graph (figure 9) is showing the accuracy of the Actual and Predicted stock value which had been extracted through the process, in x axis the name of the companies are provided for better understanding of the values calculated.

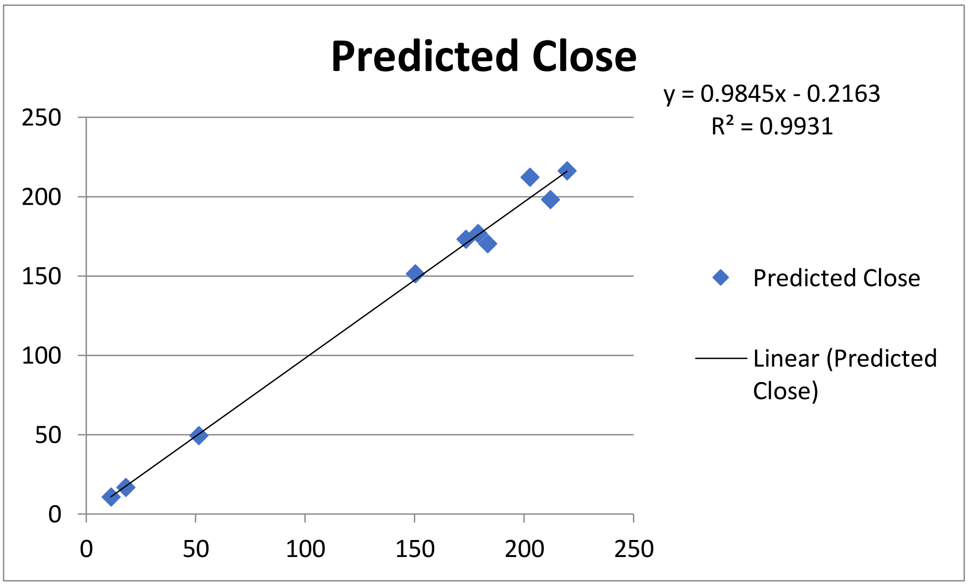


***Figure 10*: Plot Diagram of predicted close"Industrials"**

***Source*: Author`s worksheet**

After considering the result (Table 6) and applying the R^2method in the Real Estate, (R^2 =0.9931) it is very much observed that most of the predicted close are coinciding with the regression line except Honeywell Int Inc and Illinois Tool Works, because of the data availability in the extraction.

3.6 Sector- Information Technology:

The next sector which is very dynamic as it provides technological innovation, like software development, hardware and technological equipments as shown in Table 7. As we can see the RMSE values in all the companies are pretty high and predicting close price has a large difference with the actual close price for most of the selected companies.

	Security	Ticker Symbol	Sector	Sub Industry	RMSE	Actual Close	Predicted Close	Ref	Headquater, Location
1	Accenture plc	CAN	Information Technology	IT Consulting & Other Services	6.99	246.16	253.79	Graph 5	Dublin, Ireland
2	Adobe Inc.	ADBE	Information Technology	Application Software	14.46	476.86	517.41	Graph 6	San Jose, California
3	Amphenol Corp.	APH	Information Technology	Electronic Components	7.92	131.41	115.33	Graph 17	Wallingford, Connecticut
4	Analog Devices Inc.	ADI	Information Technology	Semiconductors	10.62	141.27	160.13	Graph 18	Norwood,Massachusetts
5	ANSYS	ANSS	Information Technology	Application Software	9.64	334.04	306.44	Graph 19	Canonsburg, Pennsylvania
6	KLA Corporation	KLAC	Information Technology	Semiconductor Equipment	3.65	255.97	244.87	Graph 86	Milpitas, California
7	Mastercard Inc.	MA	Information Technology	Data Processing & Outsourced Services	10.25	331.51	349.78	Graph 87	Harrison, New York
8	Micron Technology	MCHP	Information Technology	Semiconductors	0.75	139.91	139.88	Graph 88	Boise
9	Microsoft Technology	MSFT	Information Technology	System Softwares	4.60	210.52	225.12	Graph 89	Redmond
10	Motorola Solution Inc	MSI	Information Technology	Communication Equipments	22.57	171.86	135.19	Graph 90	Chicago

***Table 7*: Final Result of 10 Companies of sector " Information Technology"**

***Source*: Author`s worksheet**

After observing the prediction outcome (Table 7) it is clear that the RMSE value ranges from (0.75 to 22.57), which is very high and referring to plot diagram, (figure 12) most of values are not coincide with the exact predictions. Even to understand the prediction of such companies, actual graphs are provided (Annex 2)

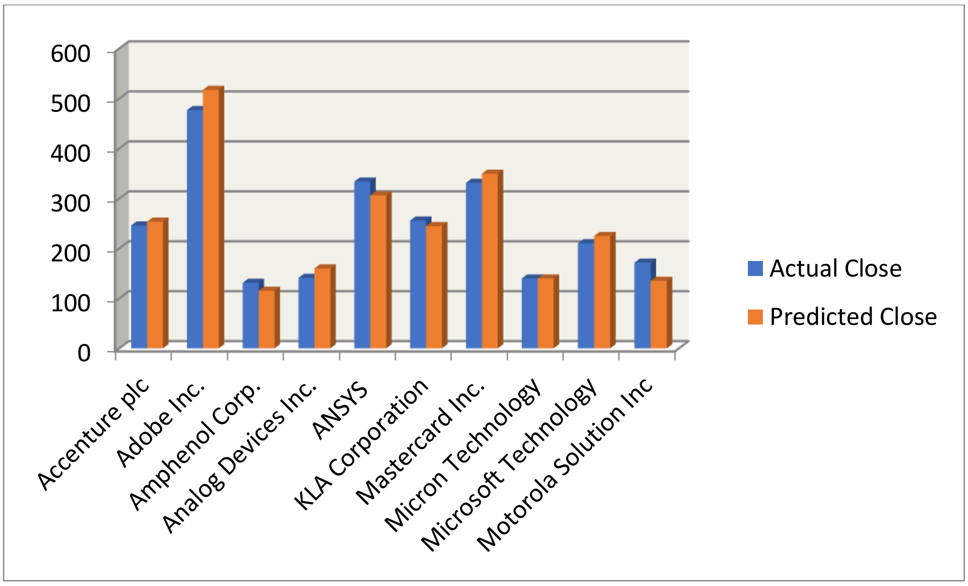


Figure 11**: Bar Graph showing the prices of "Information Technology"**

Source**: Author`s worksheet**

This graph (figure 11) is showing the accuracy of the Actual and Predicted stock value which had been extracted through the process, in x axis the name of the companies are provided for better understanding of the values calculated.

There is a huge difference in most of the predicted value with R^2=0.97.

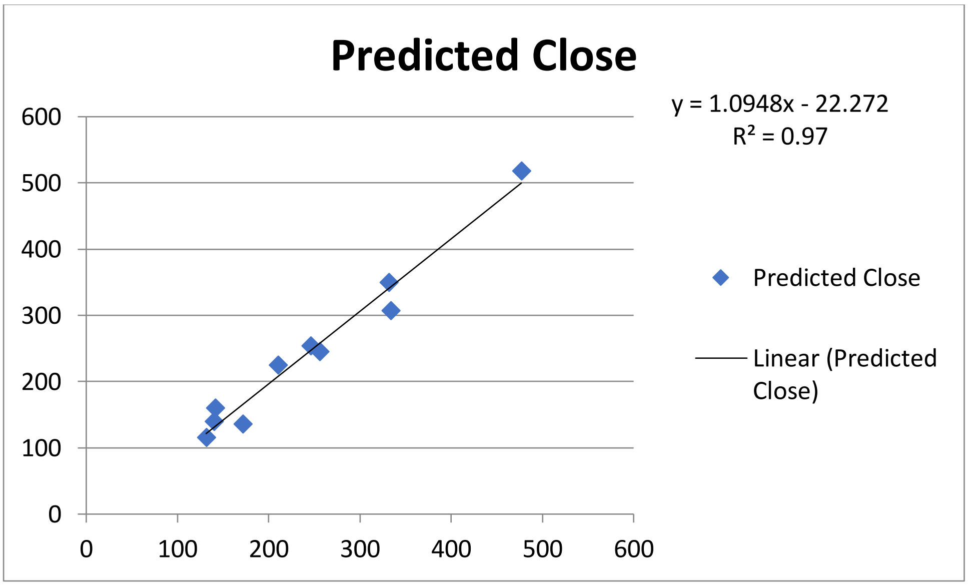


Figure 12**: Plot Diagram of predicted close"Information Technology"**

Source**: Author`s worksheet**

After considering the result (Table 7) and applying the R^2method in the Real Estate, (R^2 =0.97) it is very much observed that most of the predicted close are not coinciding with the regression line and hence the model is not appropriate in this Information Sector. Even here the root mean square error value is high comparing with other sector`s root mean square error value.

3.7 Sector- Materials:

The next sector selected for analysis is materials to have a diversified view of the industry,the seleted companies are shown in Table 8. It comprises of different companies which provide goods for manufacturing and other applications, like forest products, chemicals etc. The RMSE values in all the companies are pretty low except Linde Inc (9.28) and Air product and Chemicals Inc. (5.03).

	Security	Ticker Symbol	Sector	Sub Industry	RMSE	Actual Close	Predicted Close	Ref	Headquater , Location
1	Air Product & Chemicals Inc.	APD	Materials	Industrial Gases	5.03	268.82	256.93	Graph 8	Allentown, Pennsylvani a
2	Albemarle Corp.	ALB	Materials	Specialty Chemicals	1.38	141.42	140.27	Graph 10	Charlotte,No rth Carolina
3	Amcor Plc	AMCR	Materials	Paper Packaging	0.029	11.53	11.51	Graph 12	Warmley, Bristol
4	Ball Corp.	BLL	Materials	Metal & Glass Containers	2.63	93.26	95.36	Graph 20	Broomfield, Colorado
5	Eastman Chemical	EMN	Materials	Diversified Chemicals	1.32	102.98	100.25	Graph 67	Kingsport, Tennessee
6	Ecolab Inc.	ECL	Materials	Specialty Chemicals	1.28	223.08	220.24	Graph 68	St. Paul, Minnesota
7	Freeport McMoRan Inc.	FCX	Materials	Copper	0.458	24.86	25.32	Graph 69	Phoenix,Ari zona
8	International Paper	IP	Materials	Paper Packaging	1.3	48.63	50.39	Graph 70	Memphis, Tennessee
9	International Flavours & Fragrances	IFF	Materials	Specialty Chemicals	0.528	112.58	115.07	Graph 71	New York
10	Linde Plc.	LIN	Materials	Industrial Gases	9.28	250.94	232.22	Graph 72	Guildford, Surrey, UK

***Table 8*: Final Result of 10 Companies of sector " Materials"**

***Source*: Author`s worksheet**

After observing the prediction outcome (Table 8) it is clear that the RMSE value ranges from (0.02 to 9.28), and referring to plot diagram, (figure 14) all values are still coincide with the exact predictions. Even to understand the prediction of such companies, actual graphs are provided (Annex 2)

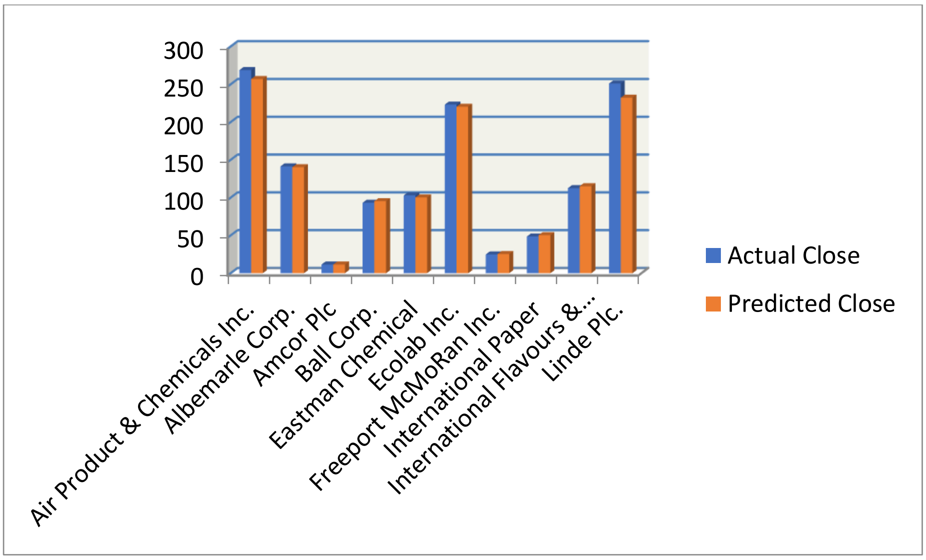


***Figure 13*: Bar Graph showing the prices of "Materials"**

***Source*: Author`s worksheet**

This graph (figure 13) is showing the accuracy of the Actual and Predicted stock value which had been extracted through the process, in x axis the name of the companies are provided for better understanding of the values calculated.

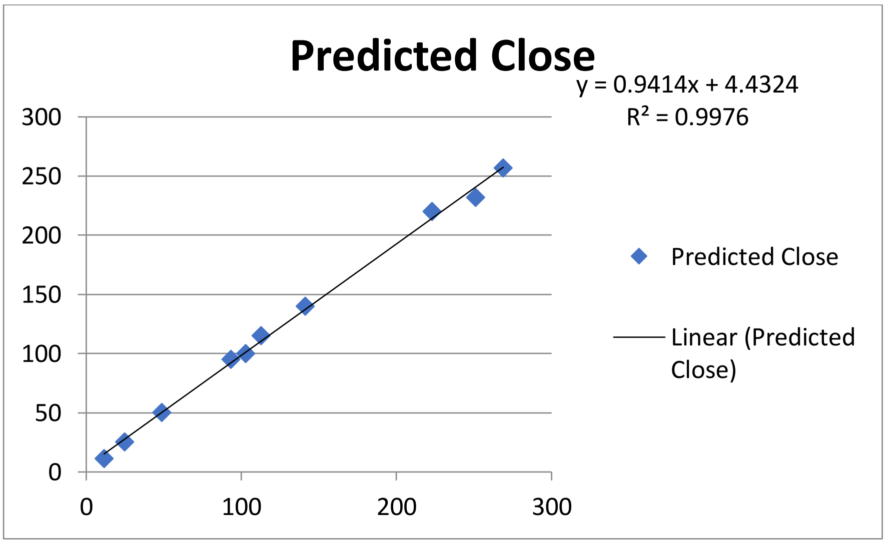


***Figure 14*: Plot Diagram of predicted close"Materials"**

***Source*: Author`s worksheet**

After considering the result (Table 8) and applying the R^2method in the Real Estate, (R^2 =0.9976) it is very much observed that most of the predicted close are coinciding with the regression line, however the RMSE value of some companies are high.

3.8 Sector- Utilities:

The next sector selected for analysis is Utilities to have a diversified view of the industry,as it provides basic amenities for the society,the seleted companies are shown in Table 9. Typically people used to hold utilities for long term gains. Here we can see that the RMSE of every company are low and the predicted price is very close to the actual price

	Security	Ticker Symbol	Sector	Sub Industry	RMSE	Actual Close	Predicted Close	Ref	Headquater, Location
1	American Electric Power	AEP	Utilities	Electric Utilities	1.98	82.48	81.24	Graph 14	Columbus,Ohio
2	CMS Energy	CMS	Utilities	Multi-Utilities	1.01	59.43	58.87	Graph 22	Jackson, Michigan
3	Consolidated Edison	ED	Utilities	Electric Utilities	0.375	73.34	75.31	Graph 23	New York
4	Edison Int`l	EIX	Utilities	Electric Utilities	3.45	62.88	59.33	Graph 27	Rosemead, California
5	NextEra Energy	NEE	Utilities	Multi-Utilities	0.363	73.27	72.34	Graph 33	Juno Beach, Florida
6	NiSource Inc.	NI	Utilities	Multi-Utilities	0.032	22.57	23.85	Graph 34	Merrillville, Indiana
7	PPL Corp.	PPL	Utilities	Electric Utilities	0.429	28.12	28.42	Graph 35	Allentown, Pennsylvania
8	Pinnacle West Capital	PNW	Utilities	Multi-Utilities	1.035	80.73	81.19	Graph 36	Phoenix,Arizona
9	Sempra Energy	SRE	Utilities	Multi-Utilities	2.28	128.72	126.67	Graph 37	San Diego, California
10	Xcel Energy Inc.	XEL	Utilities	Multi-Utilities	0.82	65.18	64.51	Graph 38	Minneapolis, Minnesota

***Table 9*: Final Result of 10 Companies of sector " Utilities"**

***Source*: Author`s worksheet**

After observing the prediction outcome (Table 9) it is clear that the RMSE value ranges from (0.03 to 12.28), and referring to plot diagram, (figure 16) all values are coincide with the exact predictions. Even to understand the prediction of such companies, actual graphs are provided (Annex 2)

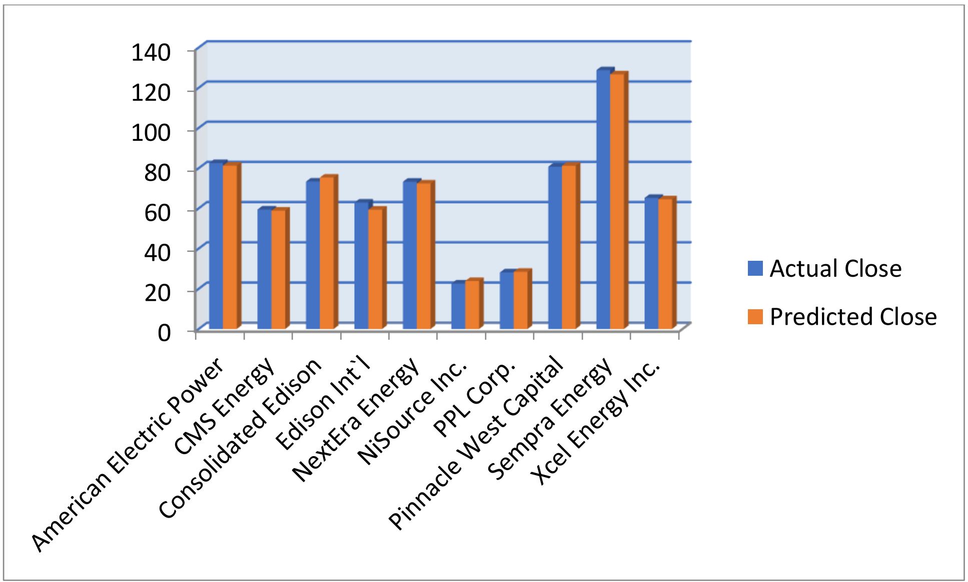


***Figure 15*: Bar Graph showing the prices of "Utilities"**

***Source*: Author`s worksheet**

This graph (figure 15) is showing the accuracy of the Actual and Predicted stock value which had been extracted through the process, in x axis the name of the companies are provided for better understanding of the values calculated.

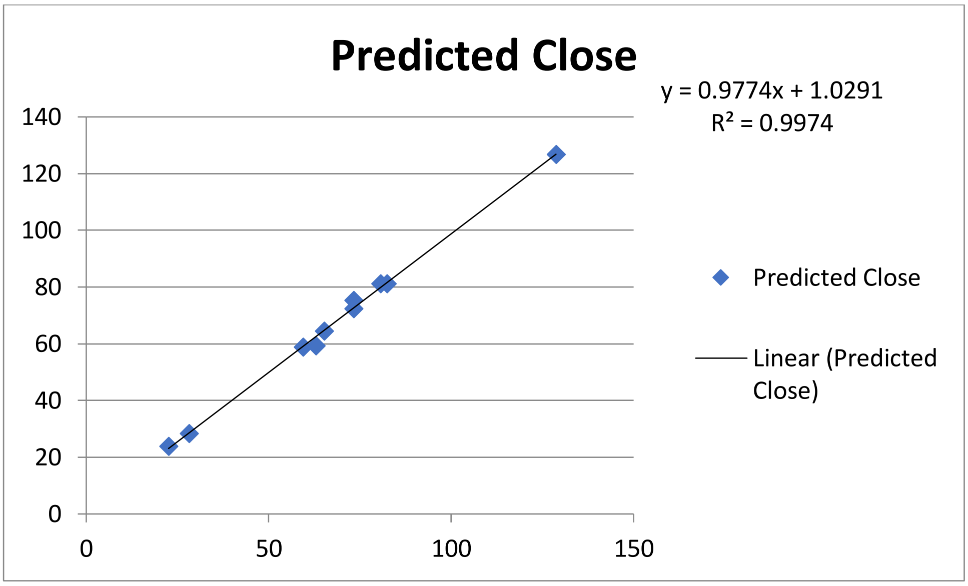


***Figure 16*: Plot Diagram of predicted close"Utilities"**
***Source*: Author`s worksheet**

After considering the result (Table 9) and applying the R^2method in the Utilities, (R^2 =0.9974) it is very much observed that most of the predicted close are coinciding with the regression line, hence the accuracy of the model is nearly perfect.

3.9 Sector- Real Estates:

The next sector selected for analysis is Real Estates with diversified REIT`s shown in Table 10. The RMSE of every selected company are very low except the Crown Castle International Corp and Digital Realty Trust, which are pretty high because of the data extracted was less than the period.

	Security	**Ticker Symbol**	**Sector**	**Sub Industry**	**RMSE**	**Actual Close**	**Predicted Close**	**Ref**	**Headquater, Location**
1	Apartmemt Ivestment & Management	AIV	Real Estate	Residential REITs	0.012	4.38	4.42	Graph 76	Denver, Colorado
2	AvalonBay Communities	AVB	Real Estate	Residential REITs	0.503	162.52	169.57	Graph 77	Arlington, Virginia
3	Boston Properties	BXP	Real Estate	Office REITs	1.21	101.90	101.12	Graph 78	Boston, Massachusetts
4	Crown Castle International Corp.	CCI	Real Estate	Specialized REITs	9.42	157.22	147.34	Graph 79	Huston, Texas
5	Digital Realty Trust Inc.	DLR	Real Estate	Specialized REITs	5.96	129.66	127.13	Graph 80	San Fransisco, Californai
6	Equity Residential	EQR	Real Estate	Residential REITs	1.14	59.79	61.22	Graph 81	Chicago, Illinois

7	Healthpeak Preperties	PEAK	Real Estate	Health Care REITs	0.05	29.54	30.85	Graph 82	Long Beach, California
8	Host Hotels & Resorts	HST	Real Estate	Hotel & Resort REITs	0.09	14.83	15.07	Graph 83	Bethesda, Maryland
9	Iron Mountain Inc.	IRM	Real Estate	Specialized REITs	0.6	30.29	30.92	Graph 84	Boston, Massachusetts
10	Kim Realty	KIM	Real Estate	Retail REITs	0.03	14.82	15.06	Graph 85	New Hyde Park, New york

***Table 10*: Final Result of 10 Companies of sector " Real Estates"**

***Source*: Author`s worksheet**

After observing the prediction outcome (Table 10) it is clear that the RMSE value ranges from (0.03 to 9.42), and referring to plot diagram, (figure 18) some values are not showing the exact predictions. Even to understand the prediction of such companies, actual graphs are provided (Annex 2)

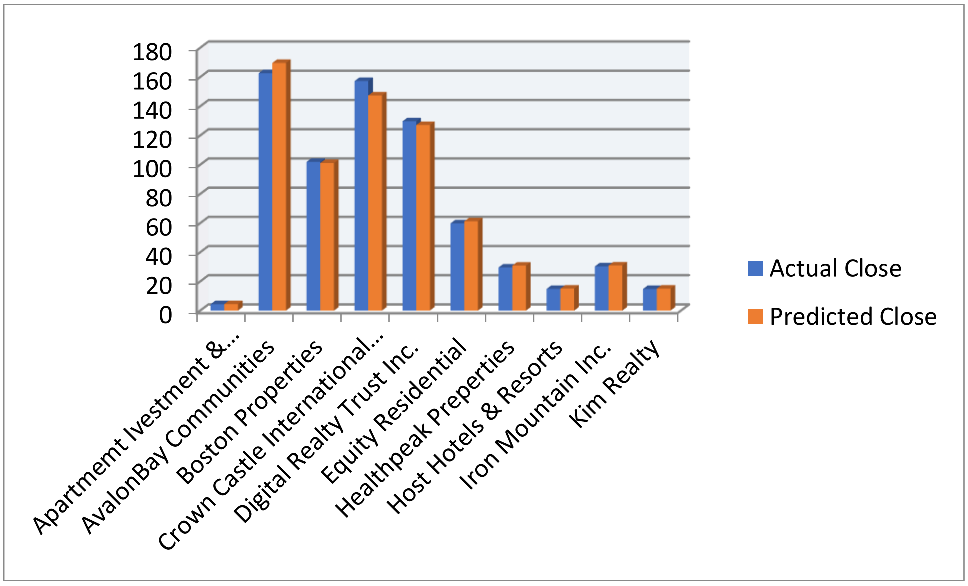


***Figure 17:* Bar Graph showing the prices of "Real Estate"**

***Source*: Author`s worksheet**

This graph (figure 17) is showing the accuracy of the Actual and Predicted stock value which had been extracted through the process, in x axis the name of the companies are provided for better understanding of the values calculated.

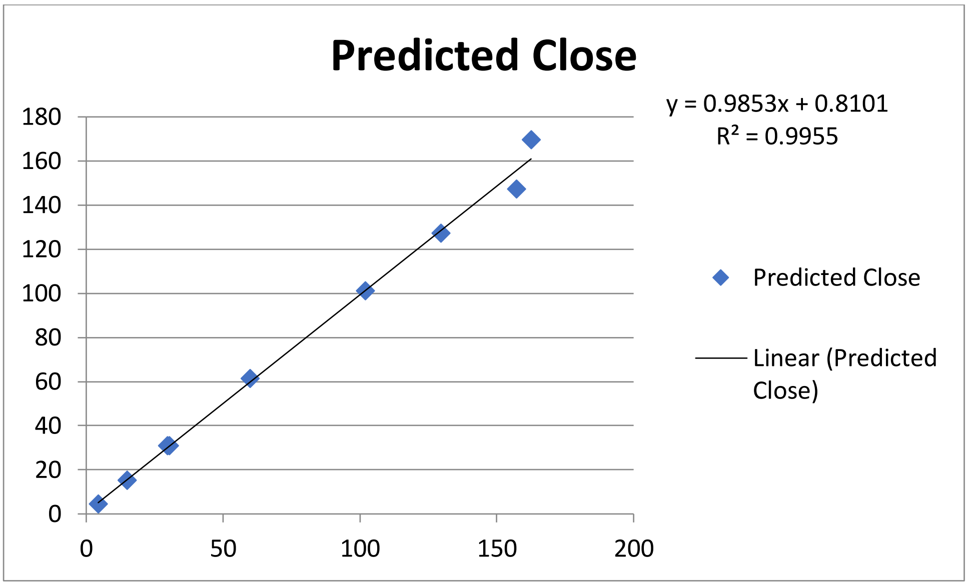


Figure 18**: Plot Diagram of predicted close"Real Estate"**

Source**: Author`s worksheet**

After considering the result (Table 10) and applying the R^2method in the Real Estate, (R^2 =0.9955) it is very much observed that most of the predicted close are coinciding with the regression line except two companies Crown Castle International Corp and Digital Realty, which shows slightly different prediction because of the volatility of the data.

3.10 Research limitations and Further Innovation:

One of the biggest limitations of this research is to gather the past data of each company to project the future value. In this volatile market, the data drastically change and gathering big data i.e. 20 years of historical value of each stocks of the company was a huge work. Moreover Bloomberg terminal was useful is such procurement of data but the extraction becomes very messy by the author. To some extent this problem was solved by using online source of data from yahoo finance. During coding the direct extraction of this big data is done by using the ticker name of the stock and taking the online availability of the data.

Another problem which the author faced is technical as of there were no software available to the author to code in python. This technical problem has been eliminated by using online base for coding in python in google colaboratory, which gives a base for the beginners to code in any language.

With the dynamic world, it is not only the numerical aspect but also the social information which affect the stock price, so to be very precise in predicting the stock price we should also take in account the social media analysis or in other words qualitative factors which is beyond the analysis of this research paper. There is Paragraph Vector technique or craw let technique which can capture the effect of social information and predict the effect of such information on the price of the stock. So there are various possibilities available for further innovation in this fields of study.

Moreover, the next generation can use quantum computing which is a super computing model and can handle enormous amount of data per second and can predict accurate stock prices and even bonds or indices all over the world. But this computing is still on its first stage, and it needs high language and coding techniques to train the neural networks. It is still in the first stage of research and needs few more years to commercialize which can easily eliminate the errors and other limitations in predicting the value of future stock price.

Conclusion & Recommendation :

1. The first and important thing was to read previous scientific articles which help to evaluate the past trends and works that had been done uptill now on the relevant topic as well as to to evaluate the theories that had been put forward ragarding the topic. It is clear from the research of the previous literature that there had been comandable work done on the evaluation and prediction of stocks. Moreover this is not a static topic to discuss and immense work had been done till now and it will persevere in the future also. It is generally the collaboration of different software codes that are inbuild on the system which help in computing and training the data to predict the accurate result. With the ongoing development and upgradation of the computational system, the prediction of stocks in the future will be more uncomplicated. With the advancement of layers in the Deep Learning, and the enhancement of the types the training system, will help to predict the stock with greater accuracy. There are many prediction done till date with a percentage of accuracy depending on the model and the training data. Generally if large amount of time series data have been taken for training then the prediction will be much accurate depending on the model used. Through the Long Short Term Memory (LSTM) networks in python, the prediction is much accurate.
2. The main hurdle which had been resolved is to learn the basic about python and the coding technique. It always require a platform, and different hardware in the computer which can support this high level machine language and can implement the process, which the author doesn't have previously but with the help of Google collaborator it becomes easy to achieve this process of research.
3. The next step which was very much critical, to develop the proper methodology, the data extraction and the analysis process. With the information which the author gathered from previous articles helped a lot. The first process was to gather the big data, which was a huge task and the first attempt taken from Bloomberg terminal. But the data was messy and proper filtration couldn't be done. Eventually with the help of Google Collaborator, secondary data can be gathered through yahoo finance, and even the data can be filtered as shown in Table 1 above. The training of the neurons are done through the coding process (Annex 1), and proper model is achieved which can evaluate every company`s stock through the ticker name.(Annex 2).
4. Last but not the least dealing with the analysis of the research, evaluate that the model which is generated is giving quite accurate analysis in every sector of the GICS. The root mean

square error value of most of the companies is low which indicate that the predicted value is very close to the actual close price of the stock. Moreover the R^2 every company shows the value nearly to one which also indicates the perfectness of the model. But there are some predictions which shows greater difference, sector “Information and Technologies” because of the volatility of the data extracted. Moreover some of the companies also show slight variation in the predicted values of the stocks because of the unavailability of the data and the model cannot be trained properly on such data. This shortcoming of the model is due to the technical problem and the extraction problem. These shortcomings can be eliminated by using high computational hardware which can easily extract a huge amount of data which is actually called big data and train the model fast. To predict the future price trend of a stock a huge data should be needed to train the model then only a perfect outcome can be expected, because neurons have to learn it or validate itself with a huge number of data.

One of the best ways to understand and evaluate the trend of the stock prices, both quantitative and qualitative information should be gathered and implemented in the model. One of the major limitations of this research is that the author only focuses on the quantitative information and implements it in the model. As the financial world is dynamic, social information also affect the trend of the stock which should also be studied for better prediction of the stock prices. This model is unique on its own but there are lots of scope to study further with this topic implementing the craw let of Paragraph Vector technique in the coding to capture the qualitative aspect of the share market.
Moreover for further study of such model, high level machine processor should be needed for better processing and fast calculation of the big data, because these all predictions depend on the past trends of the share, so huge data have to be extracted and to implement in the model so that the model can properly validate itself.

References:

1. Abu-Mostafa, Y.,(1993),"A model for learning from hints",in Hanson, S. et al.(eds), Advances in Neural information Processing Systems, 5, San Mateo, CA: Morgan Kaufmann.
2. Alkhatib,K., Najadat,H., Hmeidi,I., and Shatnawi, M.K.A.(2003). Stock price prediction using k-nearest neighbor (kNN) algorithm. International Journal of Business Humanities and technology. 3(3):32-44.
3. Altman, E.I., Marco,G., and Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). Journal of Banking & Finance, 18(3):505-529.
4. Andrew.M.Dai, and Quoc.V.Le.,(2015), Semi supervised sequence learning. Advances in Neural Information Processing Systems.
5. Ankita. Ryo, Yoshihara. Akira., Matsubara. Takashi. and Uehara. Kuniaki.,(2016), " Deep Learning for Stock Prediction using numerical and Textual information"
6. Aziz. Saqib., Dowling. Michael., Hammami. Helmi. and Piepenbrink. Anke., **"**Machine Learning in finance: A topic modeling approach".
7. Bao. Wei. Yue. Jun and Rao. Yulei,(2017)," A deep learnig framework for financial time series using stacked autoencoders and long- short term memory".
8. Blei., D.M., Ng, A.Y., and Jordan, M.I.(2003),"Latent dirichlet allocation, Journal of Machine Learning Research.",3:993-1022.
9. Blei., D.M.,(2012), probabilistic topic models. Communication of the ACM,55(4):77-84.
10. Campbell.J.Y and Thompson.S.B,(2007). Predicting excess stock returns out of sample: an anything beat the historical average? The Review of Financial Studies 21(4),1509-1531.
11. Chakraborty, K., Mehrotra, K.,Mohan, C. and Ranka, S., (1992), " Forecasting the behavior of multivariate time series using neural networks.", Neural Networks, 5: 961-70.
12. Chan.PMJ. and Mehralizadeh. M.,(2016)," Forecasting East Asian Indices Futures via a Novel Hybrid of Wavelet-PCA Denoising and Artificial Neural Network Models.
13. Chen.Y., Lin.Z., Zhao.X., Wang.G. and Gu.Y.,(2014),"Deep Learning-Based Classification of Hyperspectral Data.". IEEE J Sel Top Appl Earth Observ Remote Sens.,7(6):2094-107.
14. Cheng, D. and Cirillo.P.(2018), A reinforced urn process modeling of recovery rates and recovery times.Journal of Banking & Finance,96:1-17.
15. Ding.X., Zhang.Y., Liu.T. and Duan.J.,(2014),"Using structured events to predict stock price movement: An empirical investigation," Proceedings of the 2014 conference on Empirical Methods in Natural Language Processing (EMNLP-14)., pp.1415-1425.

16. Ding.X., Zhang.Y., Liu.T. and Duan.J., (2015), " Deep Learning for Event Driven Stock Prediction". Proceedings of the 24^{th} International Joint Conference on Artificial Intelligence(IJCAI-15).
17. Feng. G., He.J., and Polson.G. Nicholas.,(2018)," Deep Learning for predicting Asset Returns".
18. Feng.G., Giglio.S., and Xie.D.,(2017). Deep Learning Alpha. Technical report, the university of Chicaho.
19. Giles, C. and Omlin, C., (1993), "Rule refinement with recurrent neural networks", Proceedings of International Conference on Neural Networks, San Francisco,pp.801-6.
20. Giudici. P., (2018), "Fintech Risk Management: A Research Challenge for Artificial Intelligence in Finance.
21. Gu, Shihao, K.B. and Xiu.D.,(2018). Empirical asset pricing via machine learning. https://ssrn.com/abstract=3159577
22. Hagenau.M., Liebmann.M., Hedwig.M., and Neumann.D.,(2012)."Automated news reading: Stock price prediction based on financial news using context-specific features," in System Sciences. Proceedings of the 45^{th} Annual International Conference on System Sciences (HICSS-12). IEEE.2012, pp.1040-1049.
23. Harvey,C.R., Liu.Y., and Zhu.H.,(2016). The cross-section of expected returns. The Review of Financial Studies 29(1),5-68.
24. Hawley,D.D.,Johnson, J.D., and Raina,D. (1990). Artificial neural systems: A new tool for financial decision making. Financial Analysts Journal, 46(6):63-72.
25. Kercheval,A.N. and Zhang, Y. (2015). Modelling high frequency limit order book dynamics with support vector machines. Quantitative Finance, 15(8):1315-1329.
26. Kozak.S., Nagel.S., and Santosh.S.,(2017). Shrinking the cross section. Technical report , National Bureau of Economic Research.
27. Lavrenko.V., Schmill.M., Lawrie.D., Ogilvie.P., Jenson.D. and Allan.J.,(2000)," Language models for financial news recommendation," in Proceedings of the 9^{th} International Conference on Information and Knowledge Management (CIKM-00), pp. 389-396.
28. Lee.J.W.(2001). Stock price prediction using reinforcement learning. In proceeding of the IEEE International Symposium on Inductrial Electronics,Vol1,690-695.
29. Maditinos.D., and Chatzoglou.P.,(2001). "The use of Neural Networks In Forecasting: Review of Economic Sciences,6,TEI of Epirus,pp.161-176.
30. Mojaddady,M., Nabi.M., and Khadivi.S.(2011). Stock market prediction using twin Gaussian process regression. Technical report, Department of Computer Engineering, Amirkabir University of Technology, Tehran,Iran.

31. Nian,K., Coleman, T.F., and Li,Y.(2018). Learning minimum variance discrete hedging directly from the market. Quantitative Finance,18(7):1115-1128.
32. O`Leary DE. (1998). Using neural networks to predict corporate failure. International Journal of Intelligent Systems in Accounting, Finance and Management 7:187-197.
33. Pinheiro, L.D.S., and Dras. M., (2017), “Stock Market Prediction With Deep Learning: A Character based Neural Language Model for Event based Trading“.
34. Radford.A., Jozefowicz.R and Sutskever.I,(2017),” Learning to generate reviews and discovering sentiment.
35. Renault,T.(2017). Intraday online investor sentiment and return patterns in the US stock market. Journal of Banking & Finance,84: 25-40.
36. Rapach,D.E., Strauss.J.K and Zhou.G,(2010). Out of sample equity premium prediction: Combination forecasts andlinks to the real economy. The Review of Financial Studies 23(2),821-862.
37. S.H.X.Cai and X.Lin,(2012),“ Feature extraction using restricted boltzman machine for stock price prediction,“ IEEE International Conference on Computer Science and Automation Engineering (CSAE), vol.3, pp. 80-83
38. Schumaker.R.P and Chen.H.,(2009),” Textual analysis of stock market prediction using breaking financial news: The AZFin text system,” ACM Transactions on Information Systems (TOIS-09), vol.27, no.2, pp. 12:1-12:19.
39. Singh. Ritika. and Srivastava. Shashi.,(2016),” Stock Prediction using Deep Learning“.
40. Theriou, N.G. and Tsirigotis, G. (2000), “The Construction of an Anticipatory Model for the Strategic Management Decision Making Process at the Firm level”, International Journal of Computing Anticipatory Systems, 9: 127-142.
41. Troiano.L., Mejuto.E., and Kriplani.P.,(2017),“ On the Reduction using Deep Learning for Trend Prediction in Finance“
42. Varetto, F.(1998). Genetic algorithms applications in the analysis of insolvency risk. Journal of Banking & Finance,22(10-11):1421-1439.
43. Wong, B.K, Bodnovich, T.A., Selvi, Y.,(1995). A bibliography of neural networks business application research: 1988-September 1994. Expert Systems 12(3),253-262.
44. Wong, B.K, Bodnovich, T.A., Selvi, Y.,(1997). Neural n etwork applications in business. A review and analysis of the literature (1988-95). Decision Support Systems 19:301-320.
45. Yoon Kim, Yacine Jernite, David. S. and Rush.M.A., (2016),” Character-aware neural language models”

46. Zavadskaya Alexandra (2017). Artificial Intelligence in Finance: Forecasting Stock Market Returns Using Artificial Neural Networks.

47. Zhang,G., Patuwo, E.B., Hu, M.Y., (1997), " Wavelet neural networks for function learning", IEEE Transactions on Signal Processing,43(6):1485-1497.

48. Zhao.R., Deng.Y.,Dredze.M., Verma.A., Rosenberg.D., and Stent.A,(2019),"Visual attention Model for Cross sectional stock return pediction and End to end Multimodal Market representation learning".

49. Zhang,G., Patuwo, E.B., Hu, M.Y., (1997),"Forecasting with Artificial Neural Networks: the state of the art", International journal of Forecasting 14:35-62.

50. Maad M., Mejiwel., Adam.Esen., (2019),"Overview of Neural Networks"

51. Anshul.KJ., Manik S and Manu.SG., "Algorithm Building and Learning Programming Languages Using a New Educational Paradigm"

52. Sanders. ID and Langford. S., "Students' Perceptions of Python as a First Programming Language at Wits".

53. Fangohr. H., "A comparison of C, MATLAB, and python as teaching languages in engineering"

54. Mark Lutz. Python Programming book

55. Xing. Cai**.** (2005) "On the Performance of the Python Programming Language for Serial and Parallel Scientific Computations"

56. M. Zhaparov. , R. Suliyev, (2013) **"**Python to learn programming"

57. Akshansh. Sharma, Firoj. Khan, Deepak. Sharma, Dr. Sunil. Gupta (2019), "Python: The Programme Of the Future"

58. Aaquib.Javed, Monika.Jaman, and M.Monir.Uddin (2019), "An Analysis on Python Programming Language Demand and its recent trend in Bangladesh"

VERTES PROGNOZAVIMAS NAUDOJANT "DEEP LEARNING"

Kunal Banerjee
Magistro baigiamasis darbas
Finansų ir bankininkystės programa
Ekonomikos ir verslo administravimo fakultetas, Vilniaus Universitetas
Darbo vadovė **Assoc.Prof. Dr.Diemante Teresiene**

SANTRAUKA

97 puslapis, 18 paveikslų, 10 lentelių, 90 grafikų, 58 literatūros sąrašas.

Akcijų prognozavimas vaidina vertinamąjį vaidmenį pasaulinėje investicijų rinkoje, nes visi siekia maksimaliai padidinti savo pelną. Kadangi akcijos yra labai nepastovios ir nestacionarios, labai sunku prognozuoti akcijų laiko eilučių tendenciją bet kurioje akcijų rinkoje. Be to, konkrečios bendrovės akcijų vertė priklauso nuo daugelio vidinių ir išorinių veiksnių, tokių kaip pagrindiniai veiksniai, techniniai veiksniai ir rinkos nuotaikos. Kiekvienas analitikas ir investuotojai bando išsiaiškinti tinkamą vertę. Šiais laikais interneto ar kitų socialinių tinklų naujienų nagrinėjimas kartu su natūralios kalbos apdorojimo pažanga ir teksto gavybos metodais leido tam tikru mastu nuspėti atsargas. Be to, esami dirbtinio neurono tinklo (ANN) metodai nepateikia laukiamų rezultatų investuotojams ir analitikams.

Norėdami išspręsti šį iššūkį, šiame dokumente kalbama apie atsargų vertės prognozavimą per „Deep Learning". Iš Jungtinių Valstijų akcijų rinkos indekso „S&P 500" buvo išgauti didžiuliai duomenys ir, prižiūrint mokymo techniką, sluoksnių neuronai buvo išmokyti naudojant pitono programavimą, kad gautų tinkamą būsimų akcijų prognozę. Straipsnio tikslas yra parodyti, kad giluminis mokymasis gali pagerinti akcijų rinkos prognozavimą artimiausioje ateityje.

Magistro darbo kalba: anglų

ANNEXES

Annex 1: The coding of the prediction model :

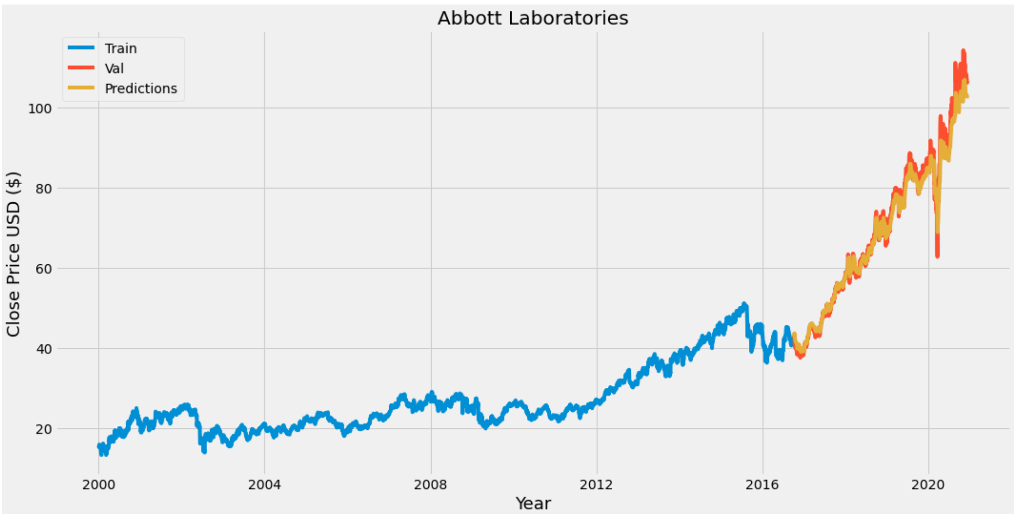
```
import math
import pandas_datareader as web
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense,LSTM
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
#get the stock Quote
df1=web.DataReader('Company_Ticke',data_source='yahoo',start='2000-01-01',end='2020-12-10')
#show the data
df1
#Get the number of rows and columns in the data set
df1.shape
#visualize the closing price history
plt.figure(figsize=(16,8))
plt.title('closing price history')
plt.plot(df1['Close'])
plt.xlabel('date',fontsize=18)
plt.ylabel('close price USD ($)',fontsize=18)
plt.show()
#create a new data frame with only the 'close column'
data1=df1.filter(['Close'])
#convert the data frame to a numpy array
dataset1=data1.values
#get the number of rows to train the model on
training_data_len1= math.ceil(len(dataset1)*.8)
training_data_len1
#scale the data
Scaler = MinMaxScaler(feature_range=(0,1))
Scaled_data=Scaler.fit_transform(dataset1)
Scaled_data
#create the training data set
#create the scaled training data set
train_data = Scaled_data[0:training_data_len1 , :]
#split the data into x_train and y_train data set
x_train = []
y_train = []

for i in range(60, len(train_data)):
  x_train.append(train_data[i-60:i, 0])
  y_train.append(train_data[i, 0])
if i <=60:
  print(x_train)
  print(y_train)
  print()
# convert the x_train and y_train to numpy array
```

```
x_train, y_train = np.array(x_train), np.array(y_train)
# Reshape the data into three dimentional to apply LSTM
x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
x_train.shape
# build the LSTM model
model = Sequential()
model.add(LSTM(50, return_sequences=True, input_shape= (x_train.shape[1], 1)))
model.add(LSTM(50, return_sequences= False))
model.add(Dense(25))
model.add(Dense(1))
#compile the model
model.compile(optimizer='adam', loss='mean_squared_error')
#Train the model
model.fit(x_train, y_train, batch_size=1, epochs=1)
#create a testing data set
#create a new array containing scaled values from index 1543 to 2003
test_data = Scaled_data [training_data_len1 - 60: , :]
#create the data set x_test and y_test
x_test = []
y_test = dataset1[training_data_len1:, :]
for i in range(60, len(test_data)):
  x_test.append(test_data[i-60:i, 0])
# convert the data into numpy array
x_test = np.array(x_test)
# Reshape the data
x_test = np.reshape(x_test, (x_test.shape[0], x_test.shape[1], 1))
# Get the model predicted price values
predictions = model.predict(x_test)
predictions = Scaler.inverse_transform(predictions)
# Get the Root Mean Squared Error (RMSE)
rmse = np.sqrt( np.mean( predictions - y_test)**2)
rmse
# Plot the Data
train = data1[:training_data_len1]
valid = data1[training_data_len1:]
valid['predictions'] = predictions
# visualize the data
plt.figure(figsize=(16,8))
plt.title('Motorola Solution Inc. ')
plt.xlabel('Year', fontsize=18)
plt.ylabel('Close Price USD ($)', fontsize=18)
plt.plot(train['Close'])
plt.plot(valid[['Close', 'predictions']])
plt.legend(['Train', 'Val', 'Predictions'], loc= ' lower right')
plt
#get the stock Quote
Company_Ticke=web.DataReader('company',data_source='yahoo',start='2000-01-01',end='2020-11-09')
# Create a New Data frame
new_df1 = Company_Ticke.filter(['Close'])
# Get the last 4 year closing price values and convert the dataframe to an array
last_4_years = new_df1[-1440:].values
# Scale the data to be values between 0 to 1
last_4_year_scaled = Scaler.transform(last_4_years)
# Create an empty list
X_test = []
# Append the last 4 years
```

```
X_test.append(last_4_year_scaled)
# convert the X_test data to numpy array
X_test = np.array(X_test)
# Reshape the data
X_test = np.reshape(X_test, (X_test.shape[0], X_test.shape[1], 1))
# Get the predicted scaled data
pred_price = model.predict(X_test)
# Undo the scaling
pred_price = Scaler.inverse_transform(pred_price)
print (pred_price)
#get the stock Quote
Company_Ticker=web.DataReader('HFC',data_source='yahoo',start='2020-11-09',end='2020-11-09')
print(Ticker_Name['Close'])
```

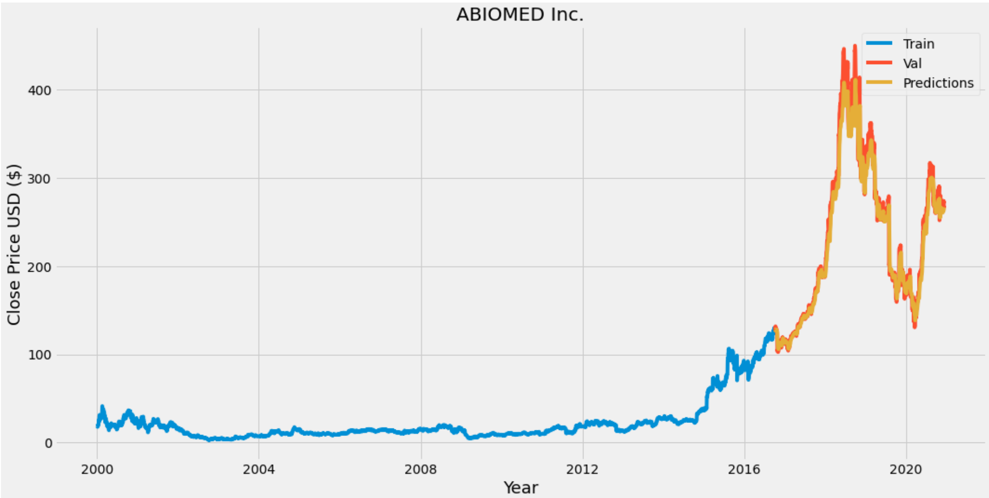
Annex 2: The Actual Prediction Analysis of the Model :

Graph 1 : Time Series values of the share of 3M company
Source: Author`s Worksheet

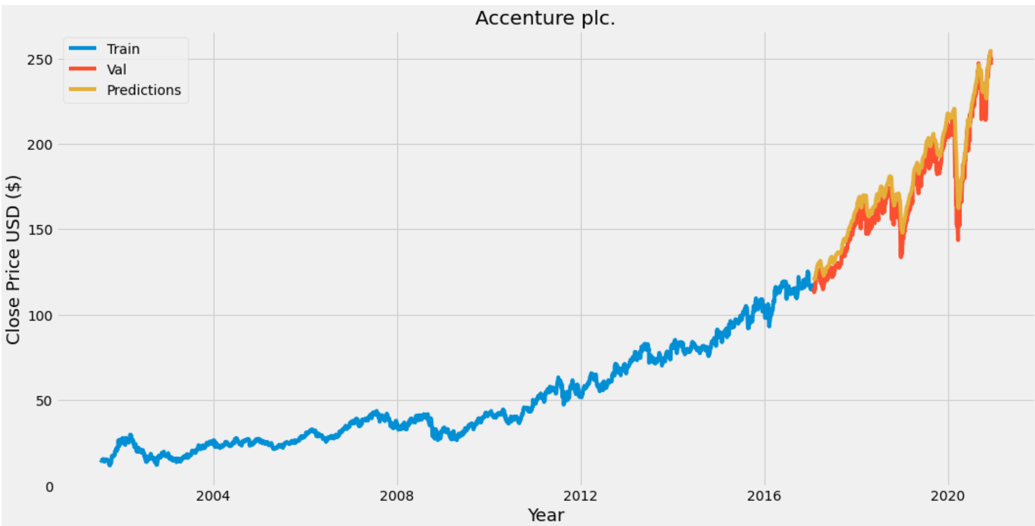


Graph 2 : Time Series values of the share of Abbott Laboratories
Source: Author`s Worksheet

Graph 3 : Time Series values of the share of AbbVie Inc.
Source: Author`s Worksheet

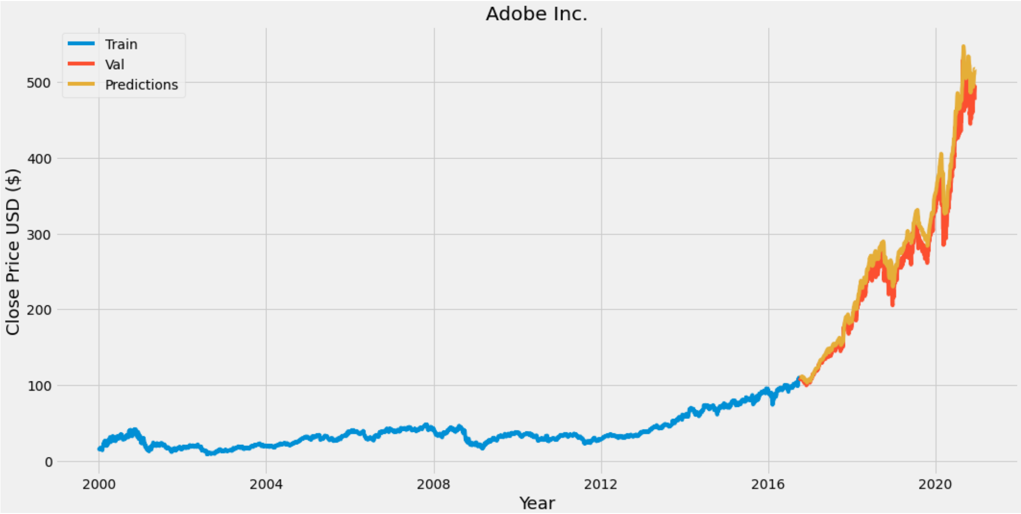


Graph 4 : Time Series values of the share of Abiomed Inc.
Source: Author`s Worksheet



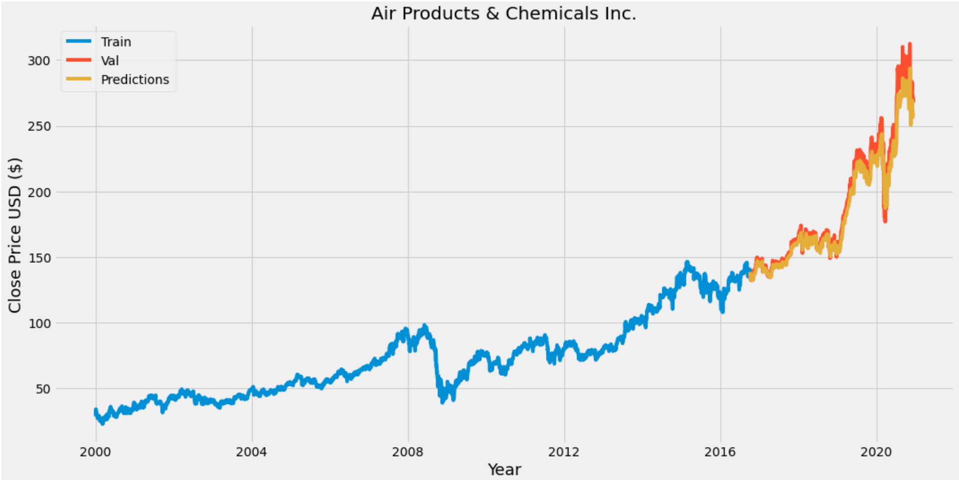
Graph 5 : Time Series values of the share of Accenture Plc.

Source: Author`s Worksheet



Graph 6 : Time Series values of the share of Adobe Inc.
Source: Author`s Worksheet

Graph 7 : Time Series values of the share of Aflac Inc.
Source: Author`s Worksheet

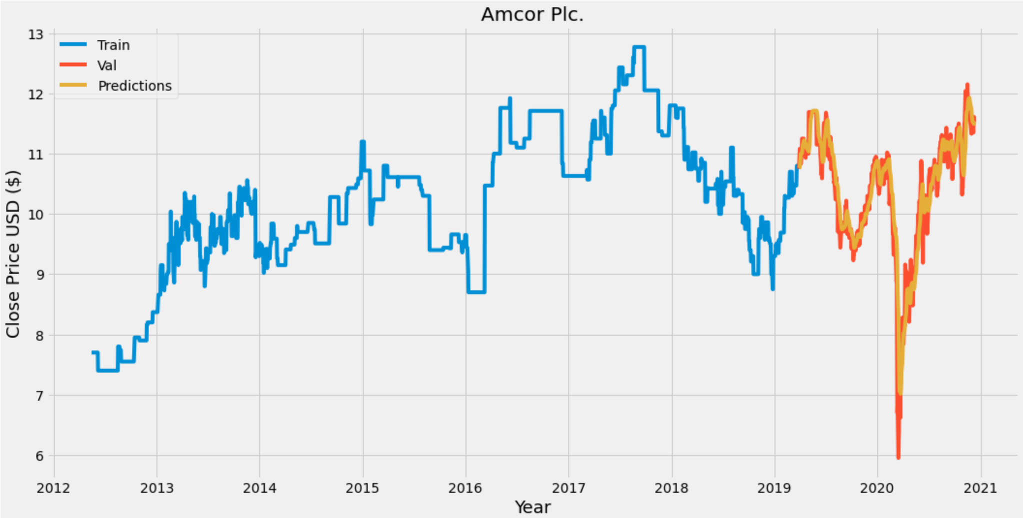


Graph 8 : Time Series values of the share of Air Products & Chemicals Inc.
Source: Author`s Worksheet

Graph 9 : Time Series values of the share of Alaska Airgroup Inc.
Source: Author`s Worksheet

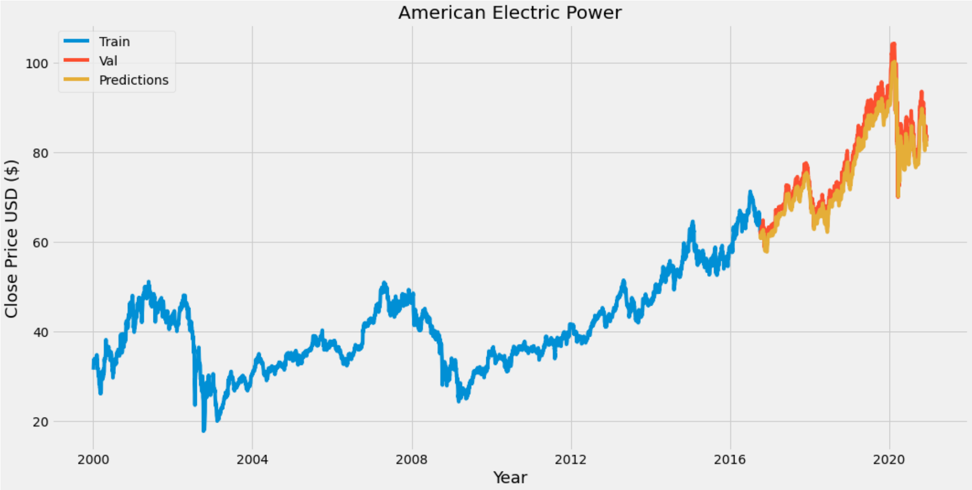
Graph 10 : Time Series values of the share of Albemarle Corp.
Source: Author`s Worksheet

Graph 11 : Time Series values of the share of Allstate Corp.
Source: Author`s Worksheet

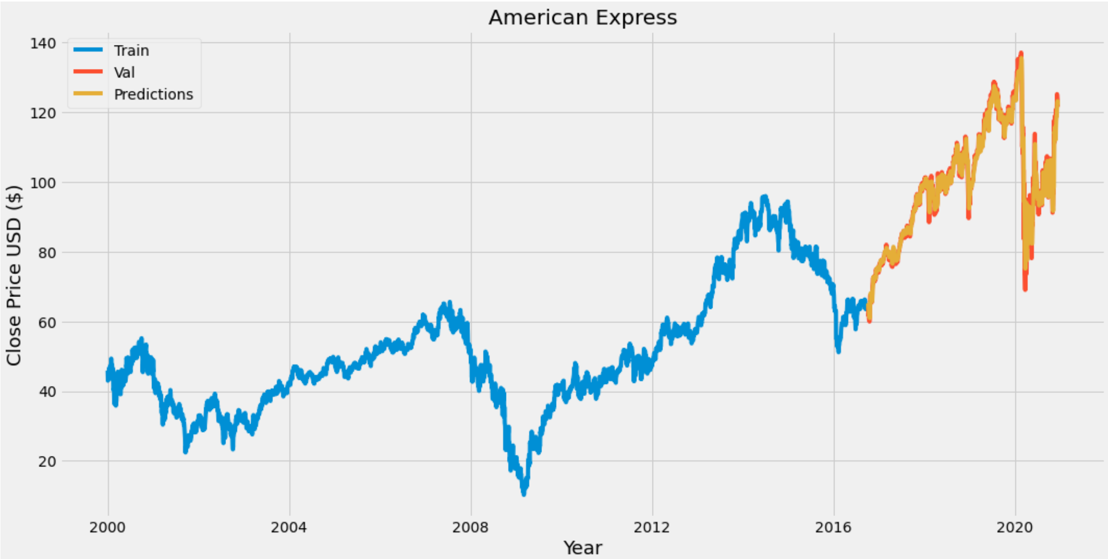


Graph 12 : Time Series values of the share of Amcor Plc.
Source: Author`s Worksheet

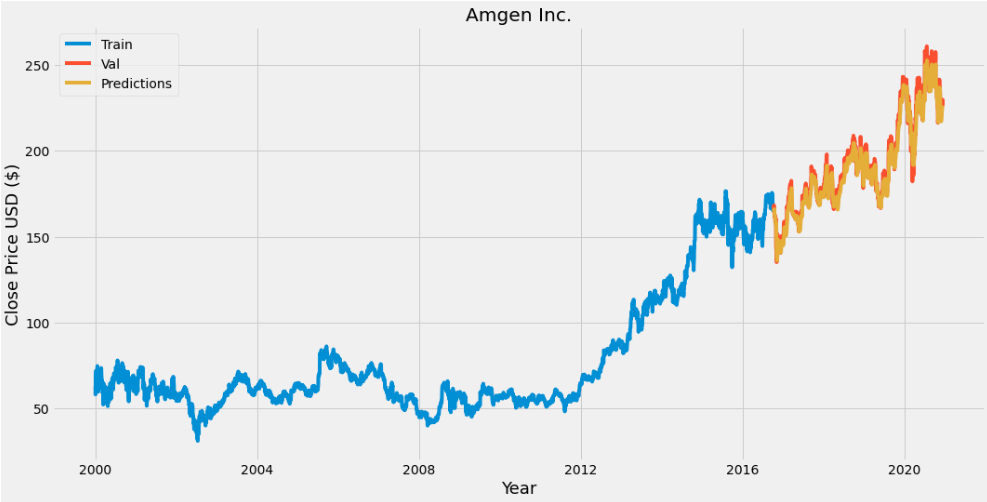
Graph 13 : Time Series values of the share of American Airlines Group.
Source: Author`s Worksheet



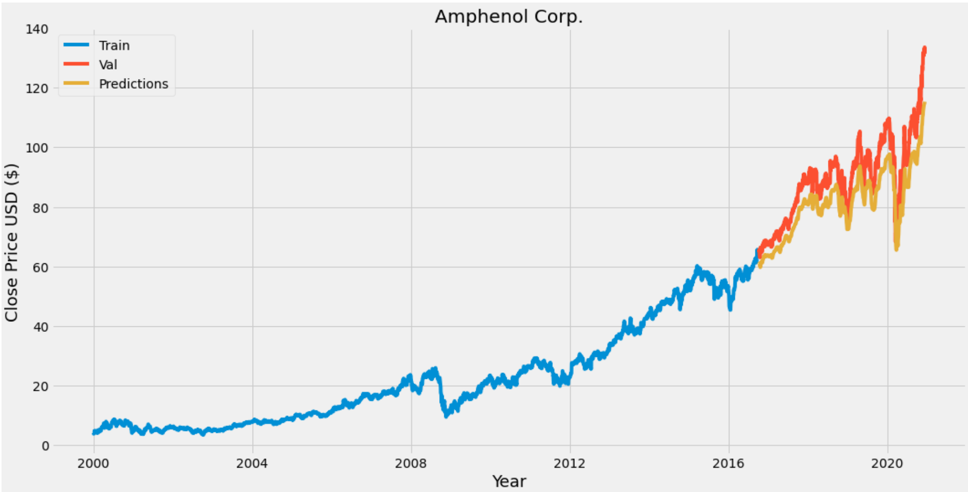
Graph 14 : Time Series values of the share of American Electric Power
Source: Author`s Worksheet



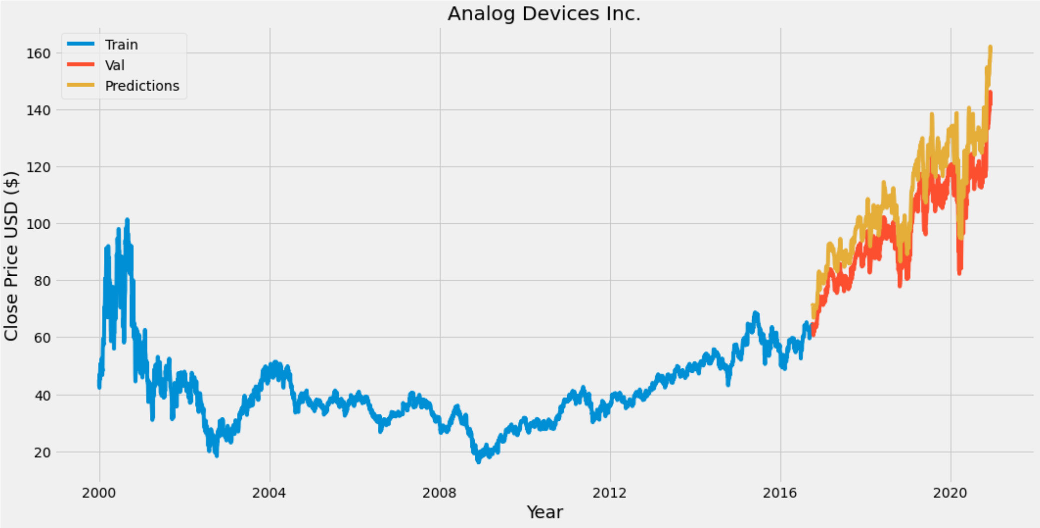
Graph 15 : Time Series values of the share of American Express
Source: Author`s Worksheet



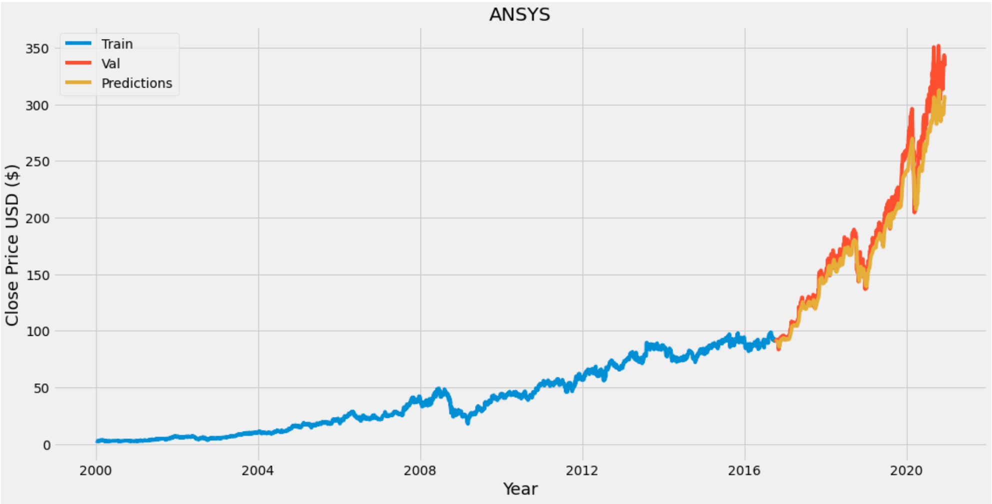
Graph 16 : Time Series values of the share of Amgen Inc.
Source: Author`s Worksheet



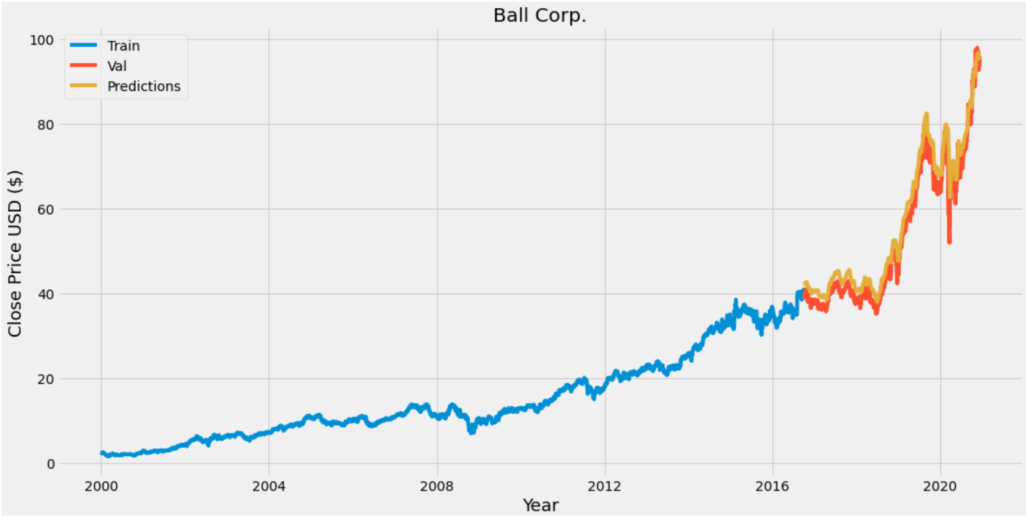
Graph 17 : Time Series values of the share of Amphenol Corp.
Source: Author`s Worksheet



Graph 18 : Time Series values of the share of Analog Devices Inc.
Source: Author`s Worksheet

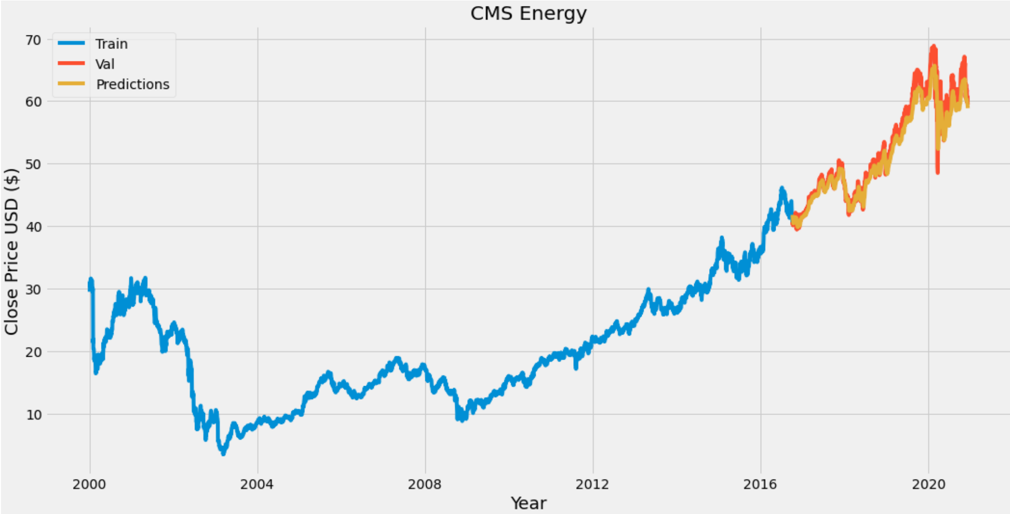


Graph 19 : Time Series values of the share of ANSYS
Source: Author`s Worksheet

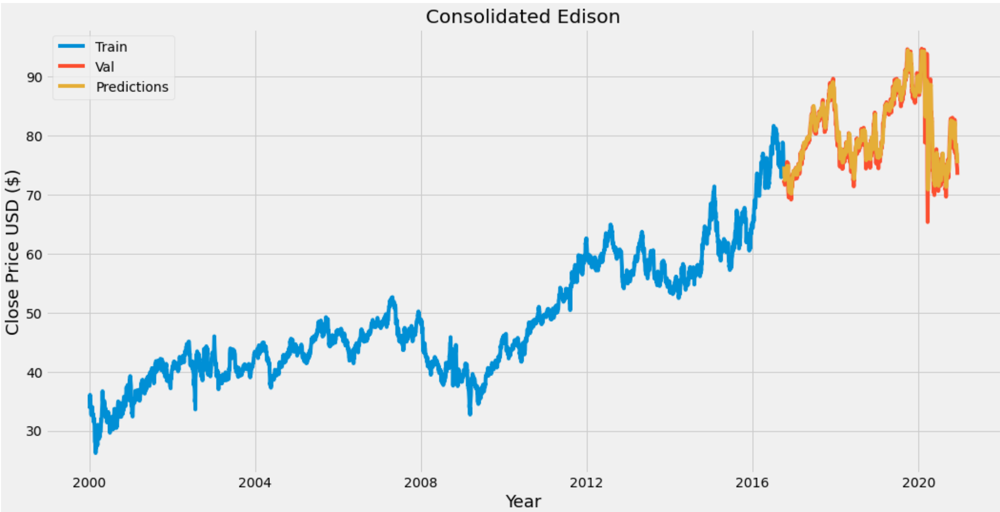


Graph 20 : Time Series values of the share of Ball Corp.
Source: Author`s Worksheet

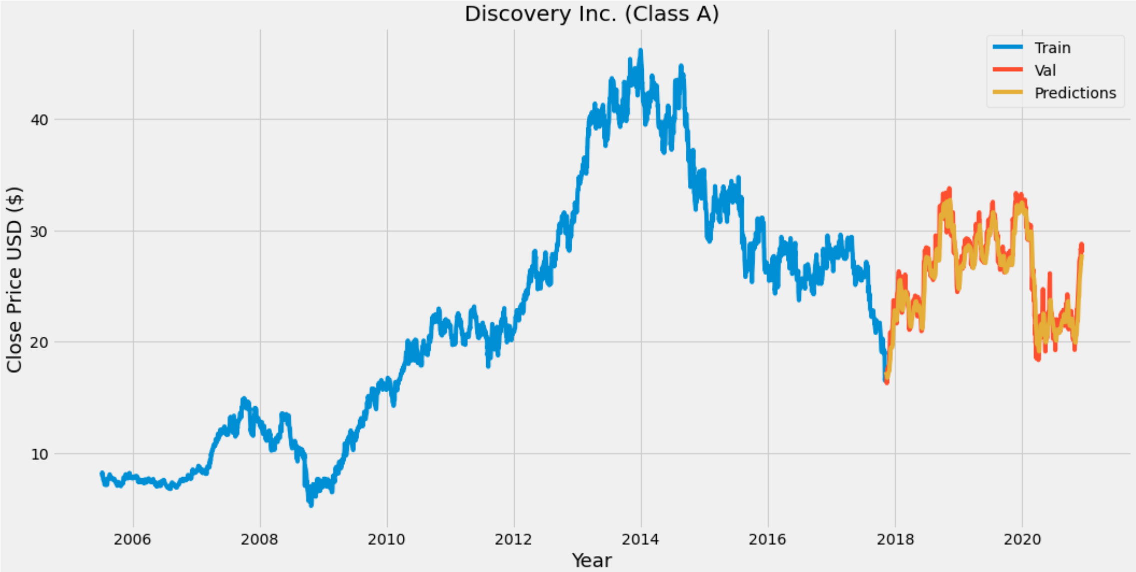
Graph 21 : Time Series values of the share of Bank OF America Corp.
Source: Author`s Worksheet



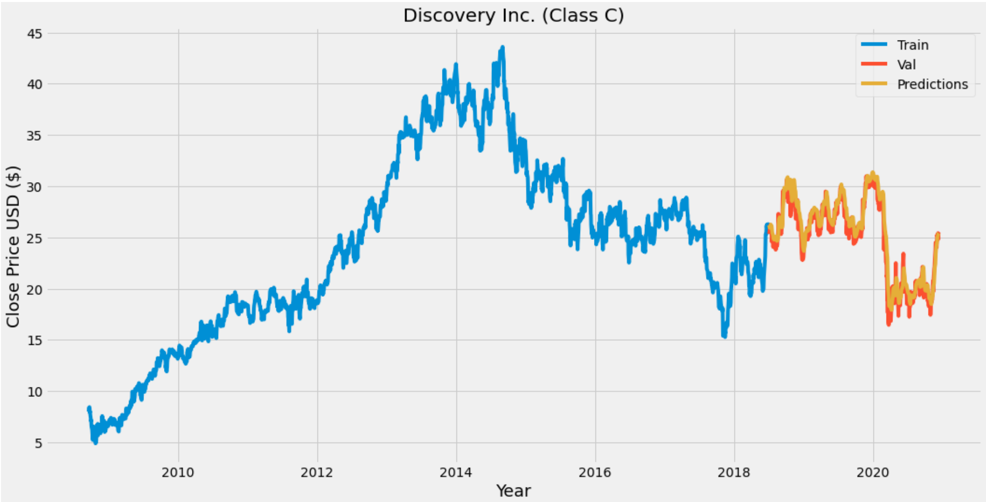
Graph 22 : Time Series values of the share of CMS Energy
Source: Author`s Worksheet



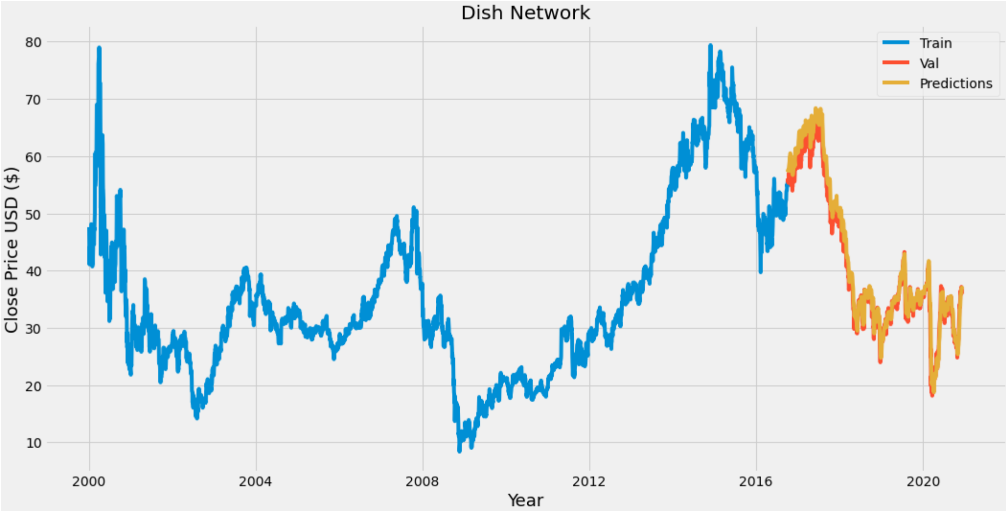
Graph 23 : Time Series values of the share of Consolidated Edison
Source: Author`s Worksheet



Graph 24 : Time Series values of the share of Discovery Inc. (Class A)
Source: Author`s Worksheet

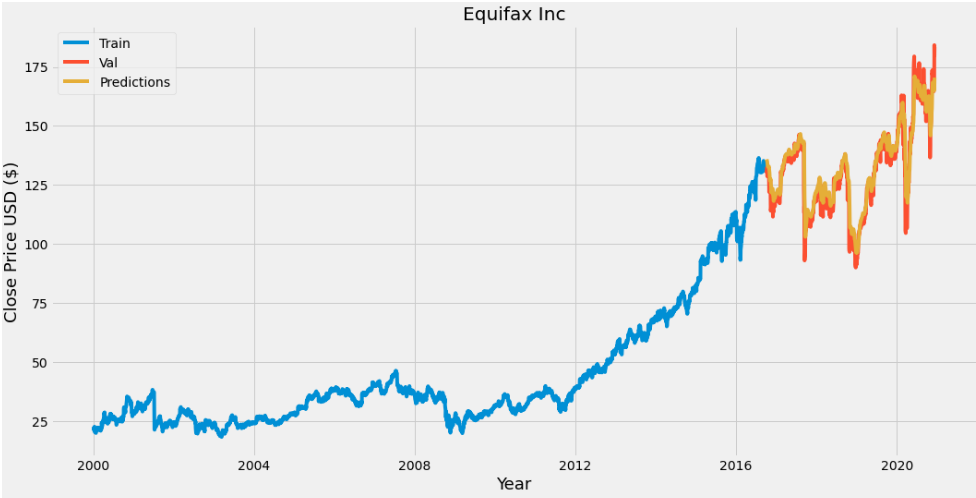


Graph 25 : Time Series values of the share of Discovery Inc. (Class C)
Source: Author`s Worksheet

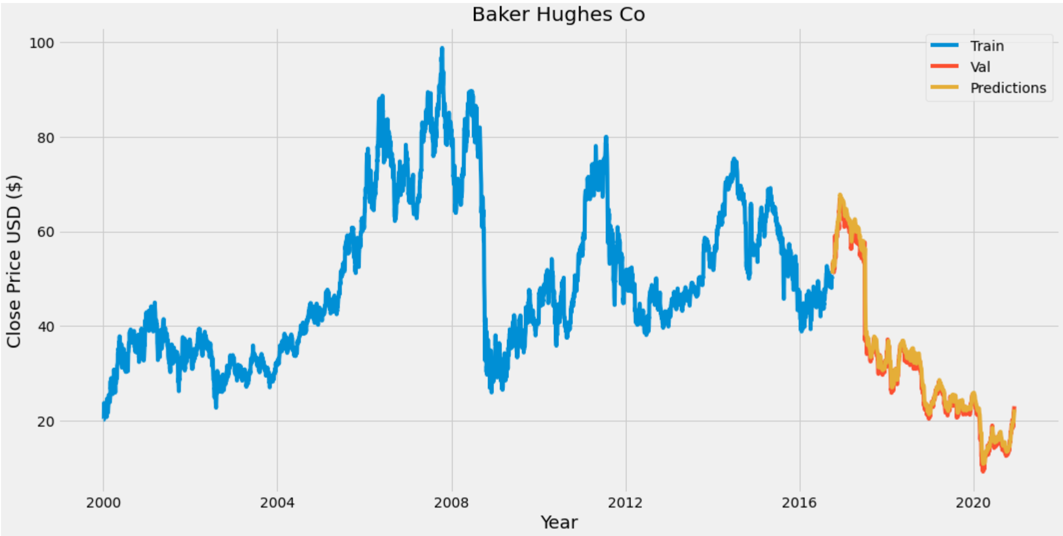


Graph 26 : Time Series values of the share of Dish Network
Source: Author`s Worksheet

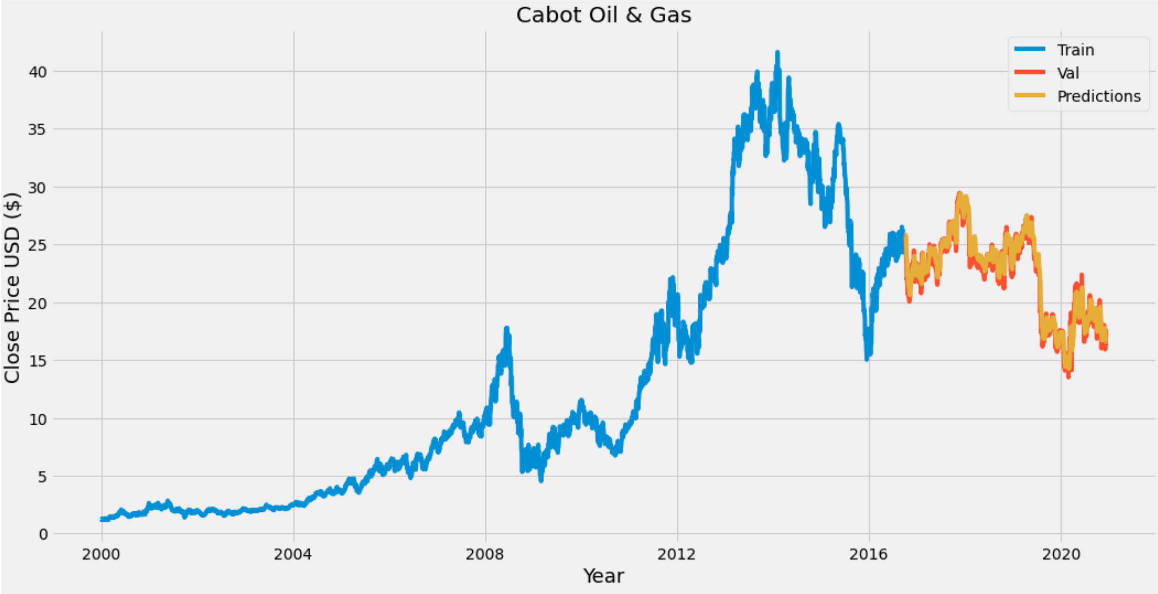
Graph 27 : Time Series values of the share of Edison Int`l
Source: Author`s Worksheet



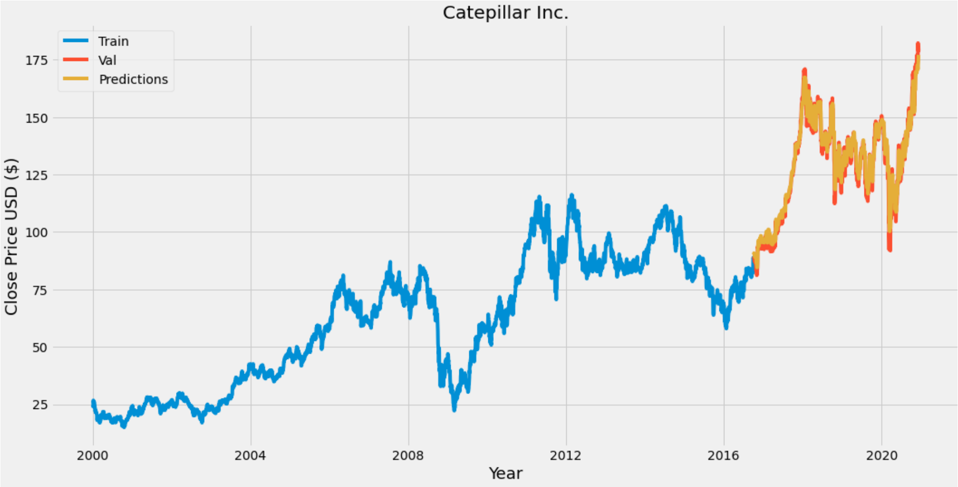
Graph 28 : Time Series values of the share of Equifax Inc
Source: Author`s Worksheet



Graph 29 : Time Series values of the share of Baker Hughes Co.
Source: Author`s Worksheet

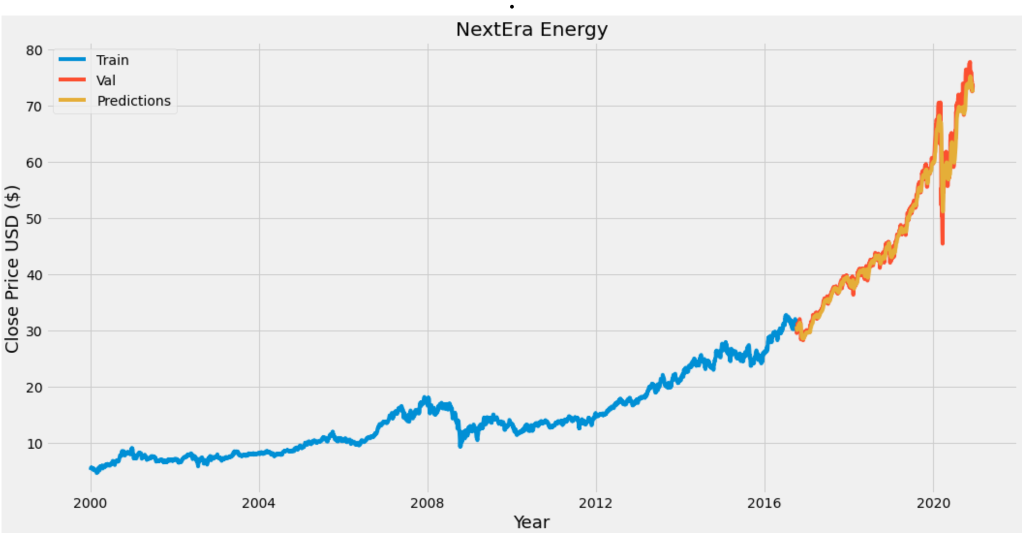


Graph 30 : Time Series values of the share of Cabot Oil & Gas
Source: Author`s Worksheet

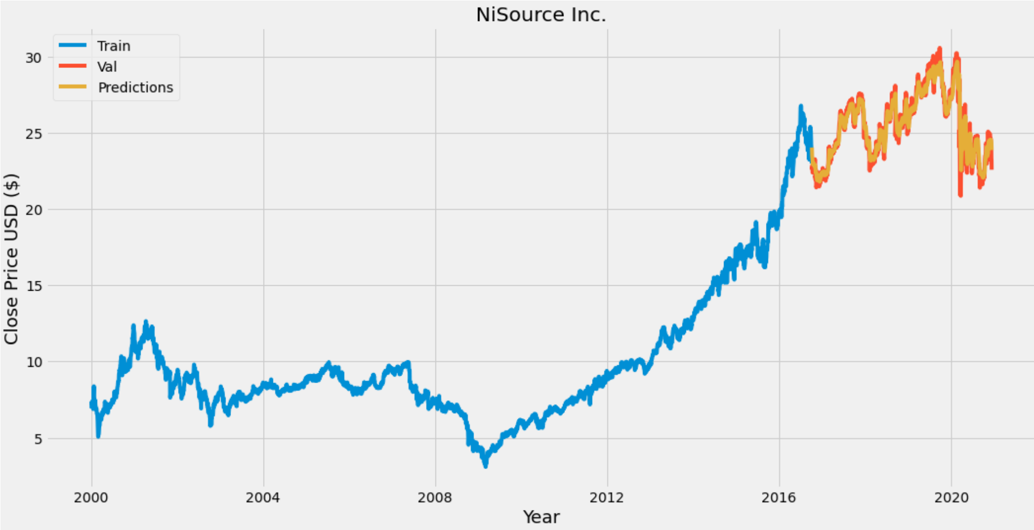


Graph 31 : Time Series values of the share of Catepillar Inc.
Source: Author`s Worksheet

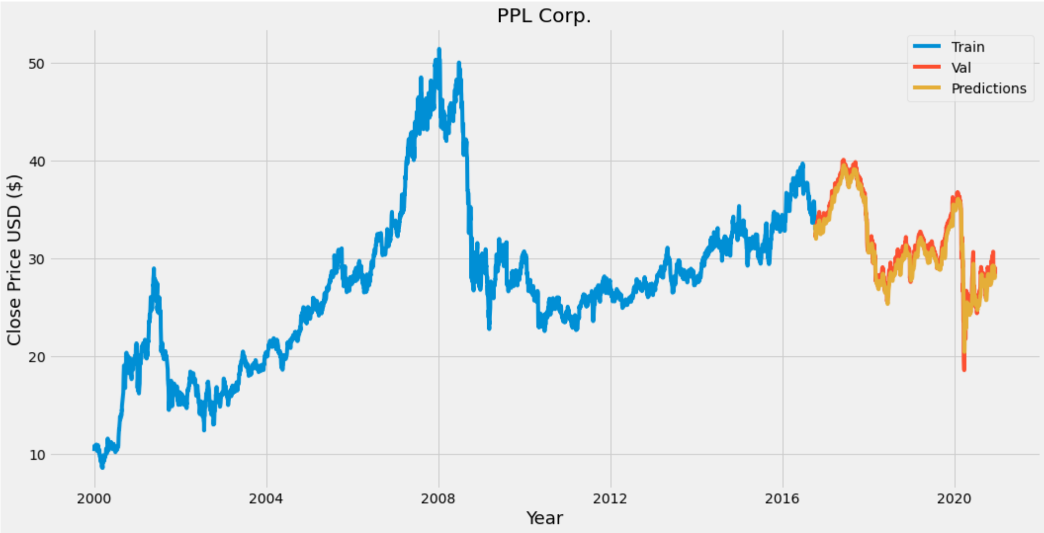
Graph 32 : Time Series values of the share of CBOE Global Market
Source: Author`s Worksheet



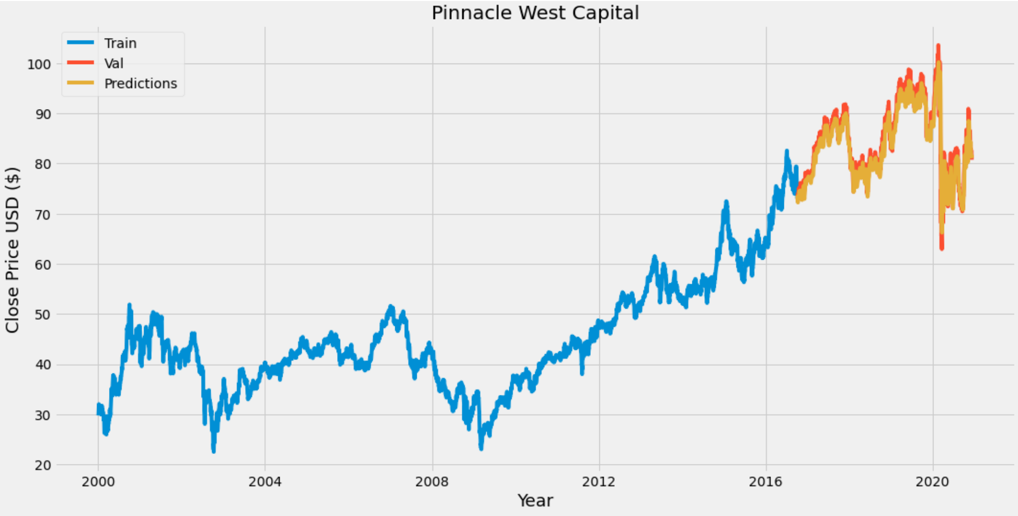
Graph 33 : Time Series values of the share of NextEra Energy
Source: Author`s Worksheet



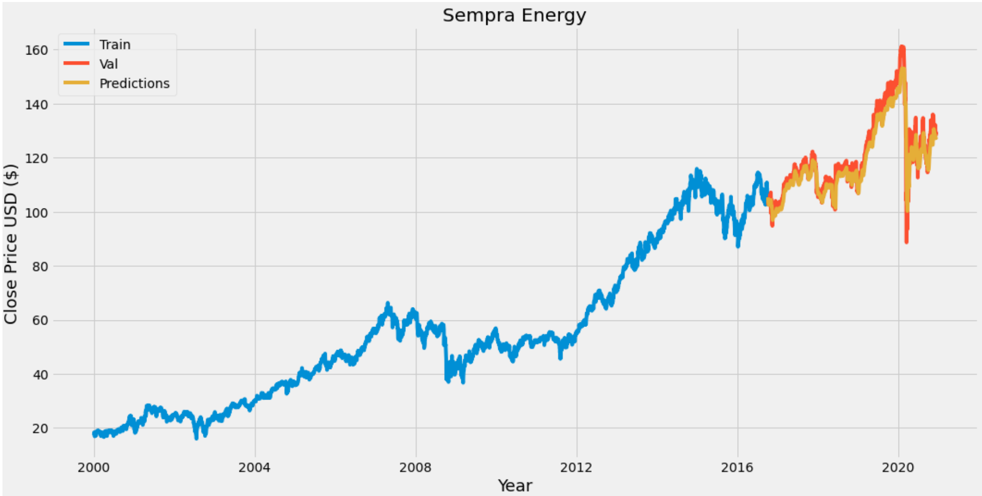
Graph 34 : Time Series values of the share of NiSource Inc.
Source: Author`s Worksheet



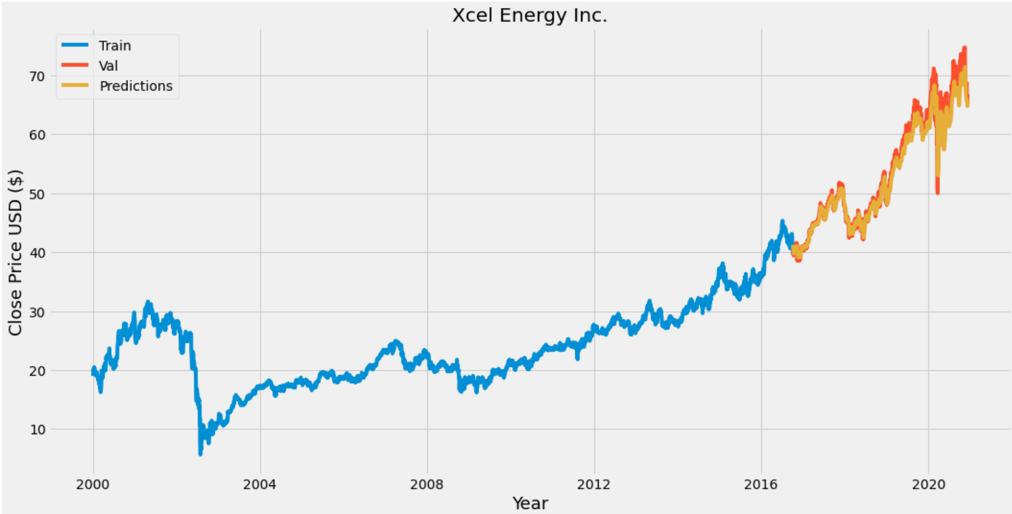
Graph 35 : Time Series values of the share of PPL Corp.
Source: Author`s Worksheet



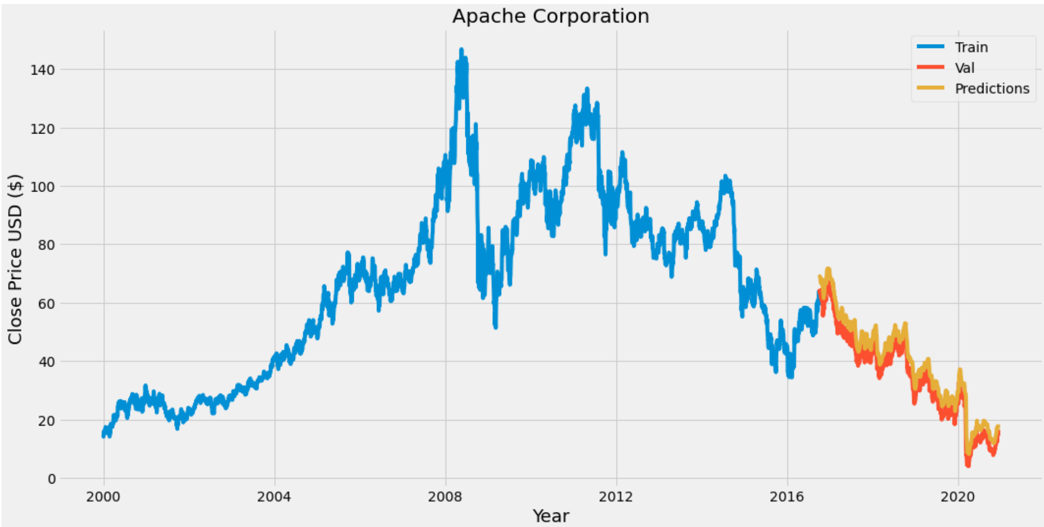
Graph 36 : Time Series values of the share of Pinnacle West Capital
Source: Author`s Worksheet



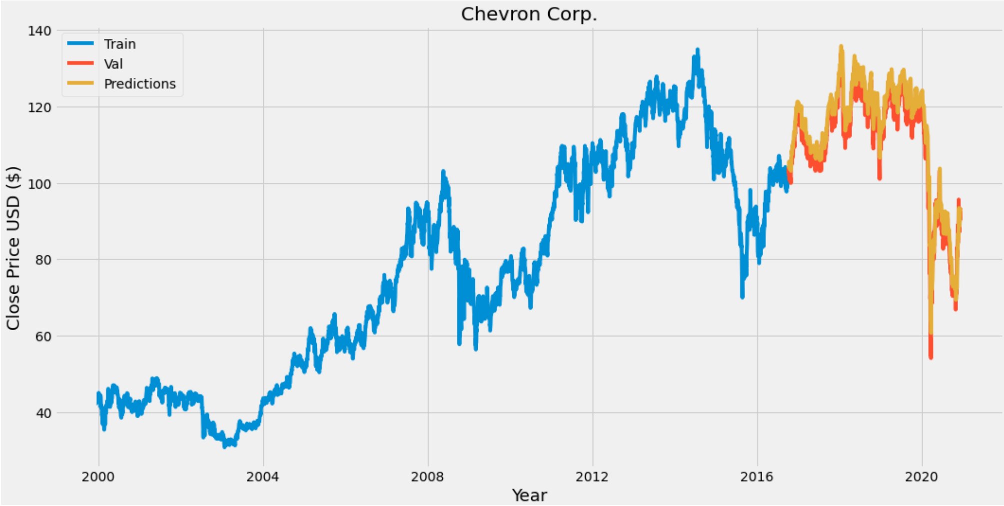
Graph 37 : Time Series values of the share of Sempra Energy
Source: Author`s Worksheet



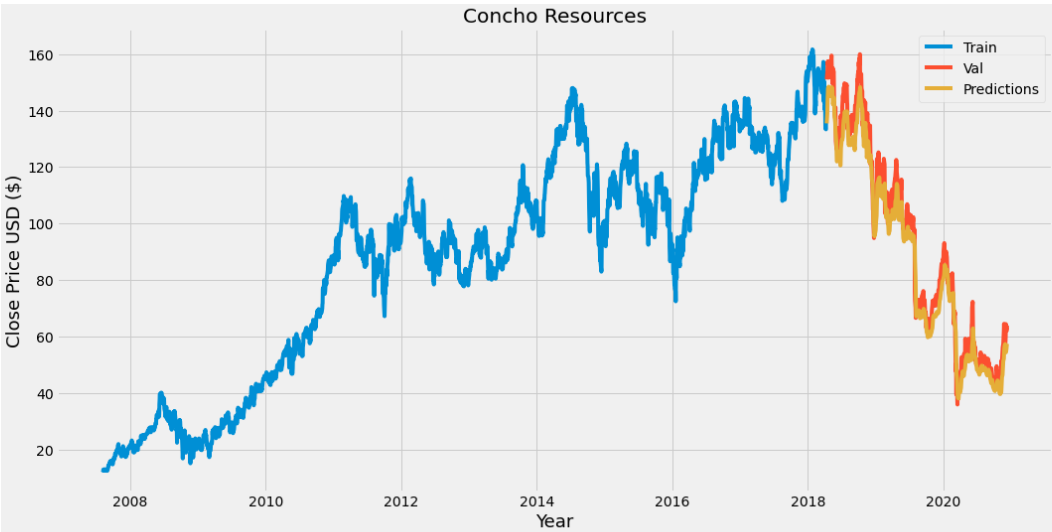
Graph 38 : Time Series values of the share of Xcel Energy Inc.
Source: Author`s Worksheet



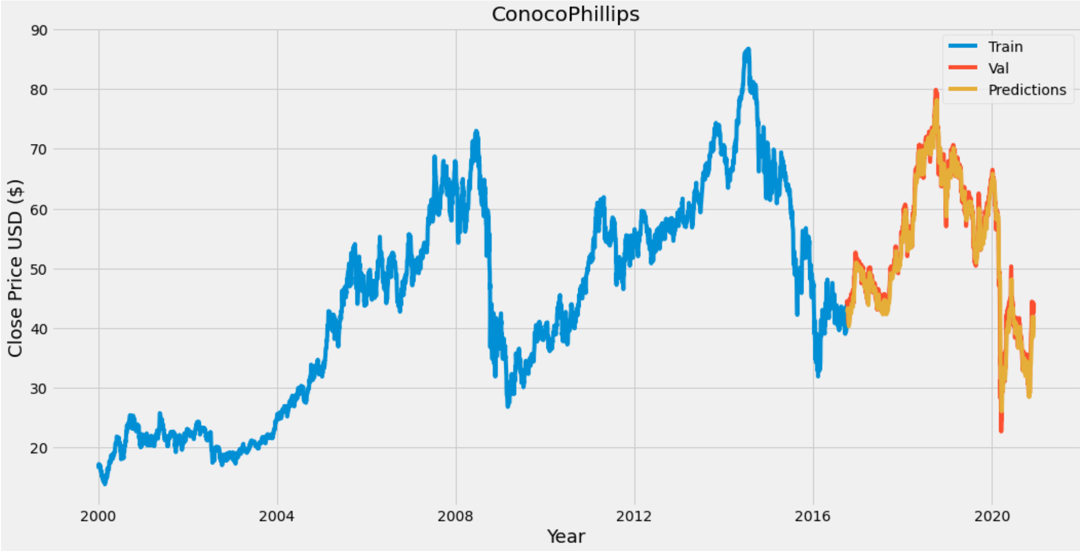
Graph 39 : Time Series values of the share of Apache Corporation.
Source: Author`s Worksheet



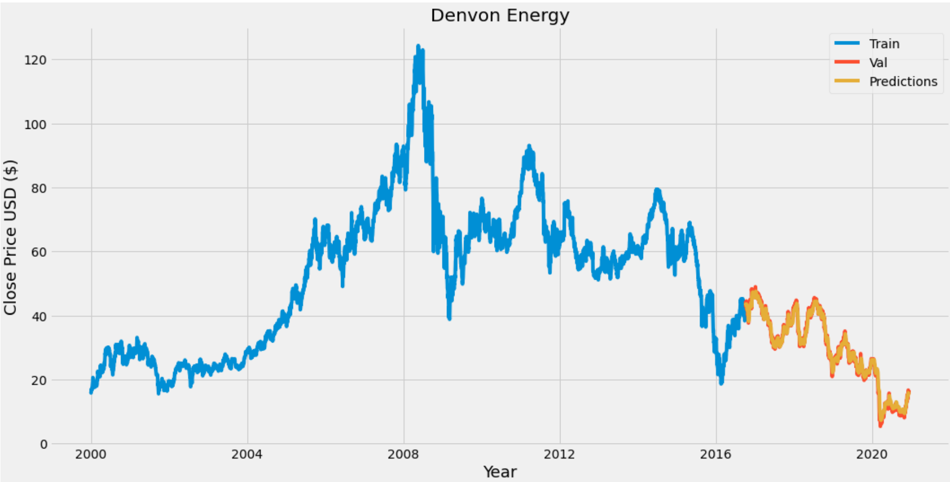
Graph 40 : Time Series values of the share of Chevron Corp.
Source: Author`s Worksheet



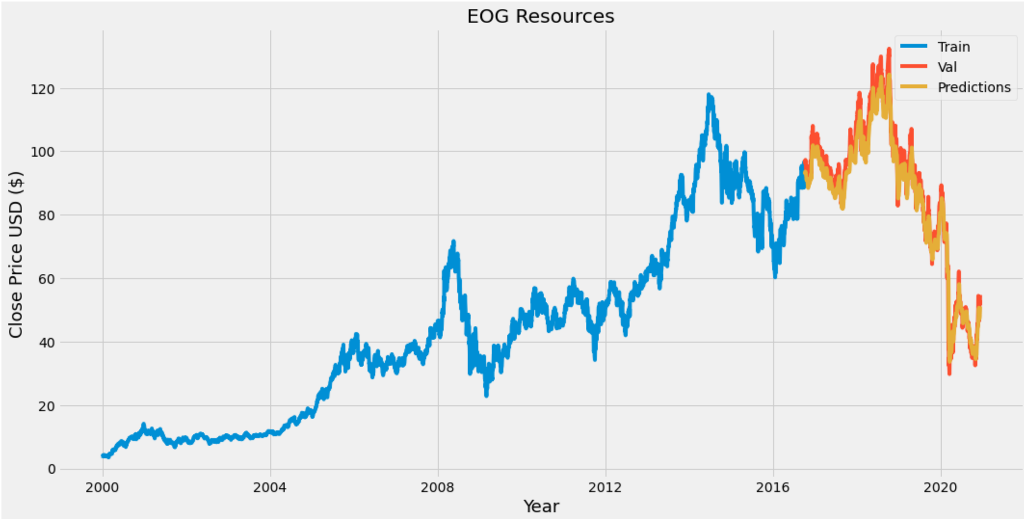
Graph 41 : Time Series values of the share of Concho Resources
Source: Author`s Worksheet



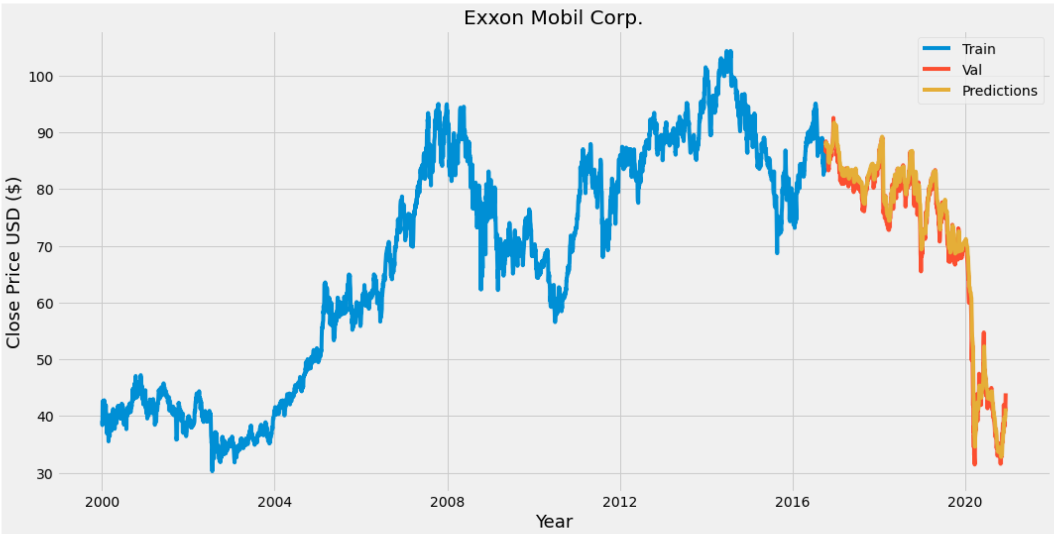
Graph 42 : Time Series values of the share of Conoco Phillips
Source: Author`s Worksheet



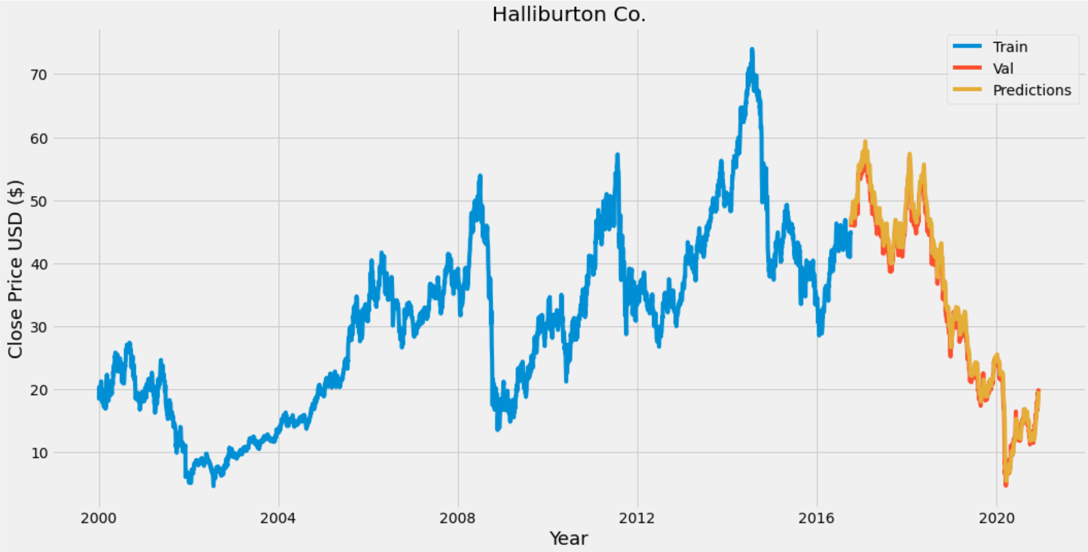
Graph 43 : Time Series values of the share of Denvon Energy
Source: Author`s Worksheet



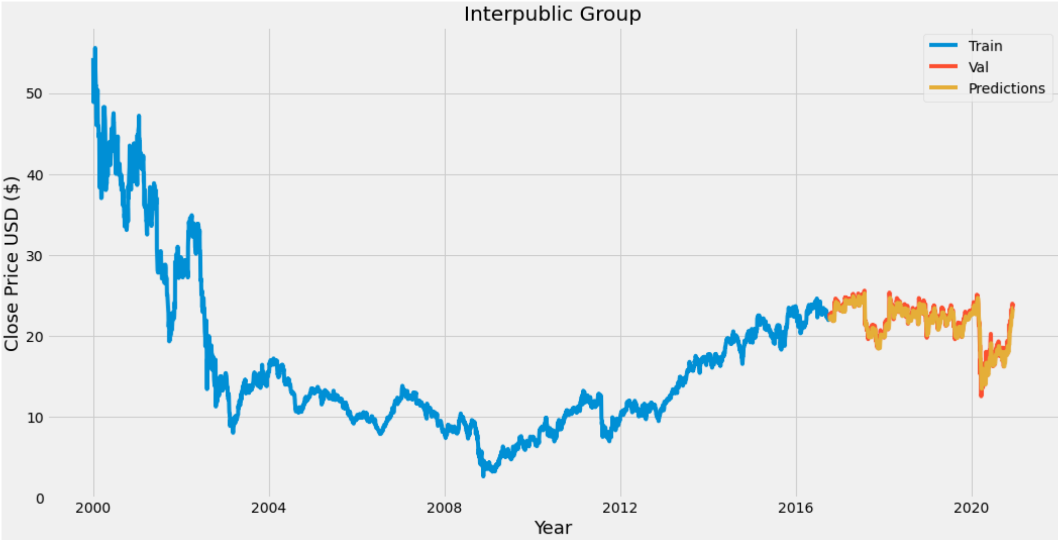
Graph 44 : Time Series values of the share of EOG Resources
Source: Author`s Worksheet



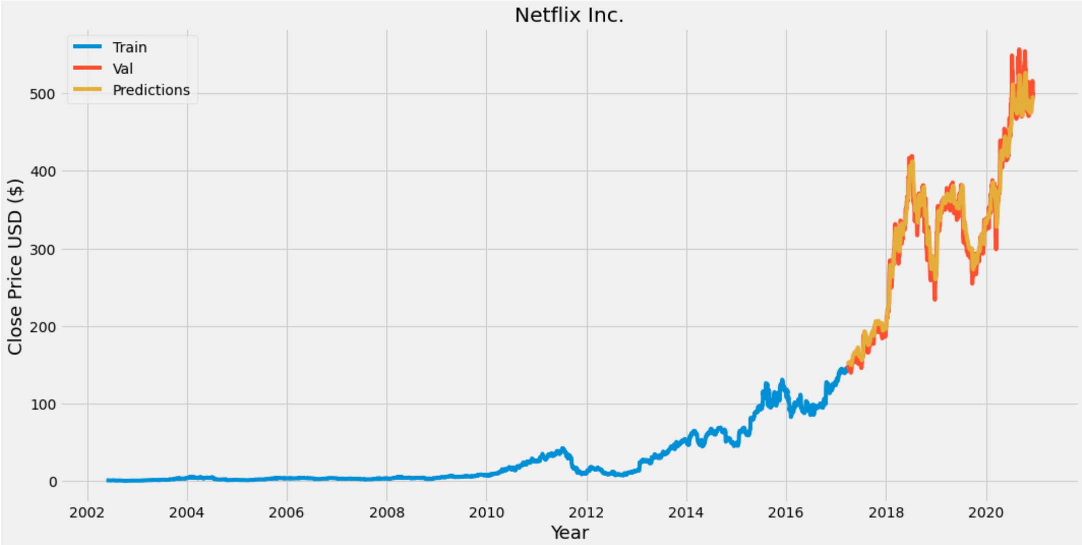
Graph 45 : Time Series values of the share of Exxon Mobil Corp.
Source: Author`s Worksheet



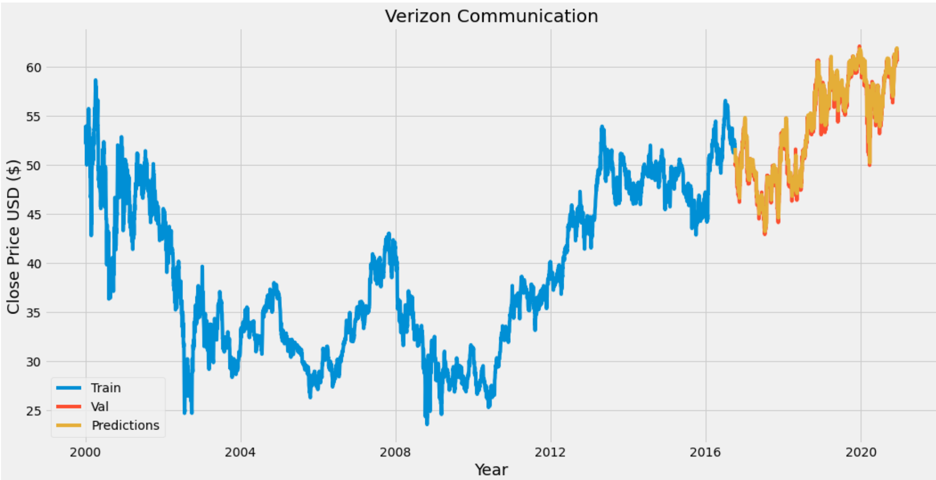
Graph 46 : Time Series values of the share of Halliburton Co.
Source: Author`s Worksheet



Graph 47 : Time Series values of the share of Interpublic Group
Source: Author`s Worksheet

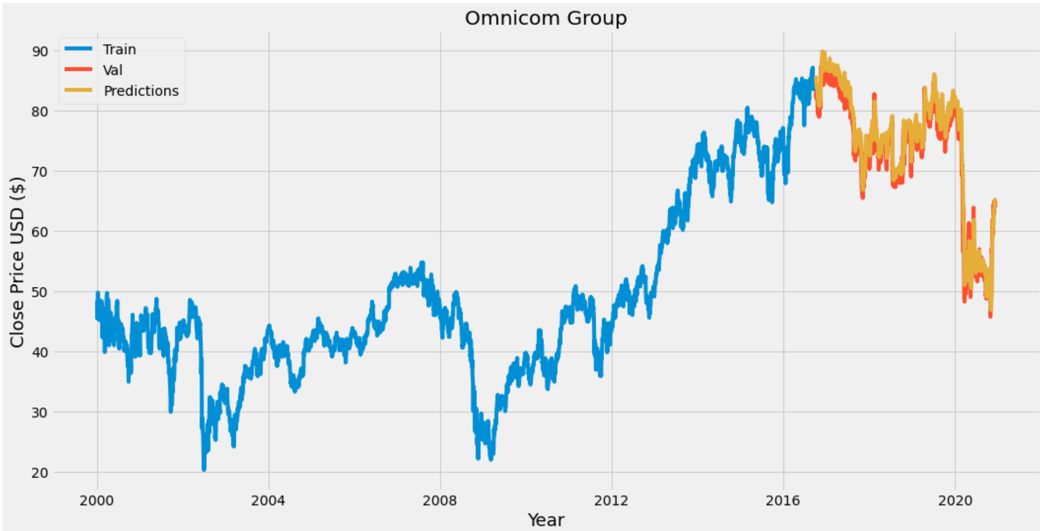


Graph 48 : Time Series values of the share of Netflix Inc.
Source: Author`s Worksheet

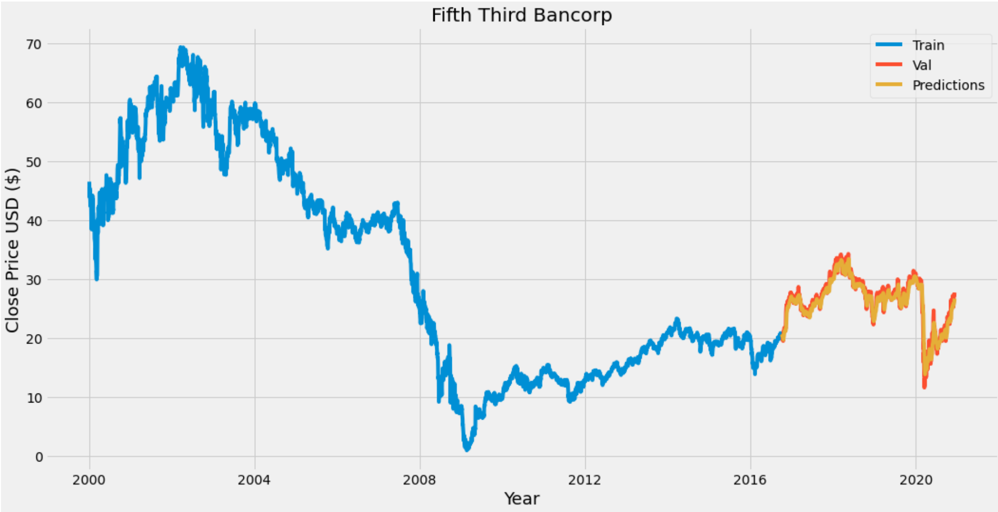


Graph 49 : Time Series values of the share of Verizon Communication
Source: Author`s Worksheet

Graph 50 : Time Series values of the share of The walt Disney Company.
Source: Author`s Worksheet



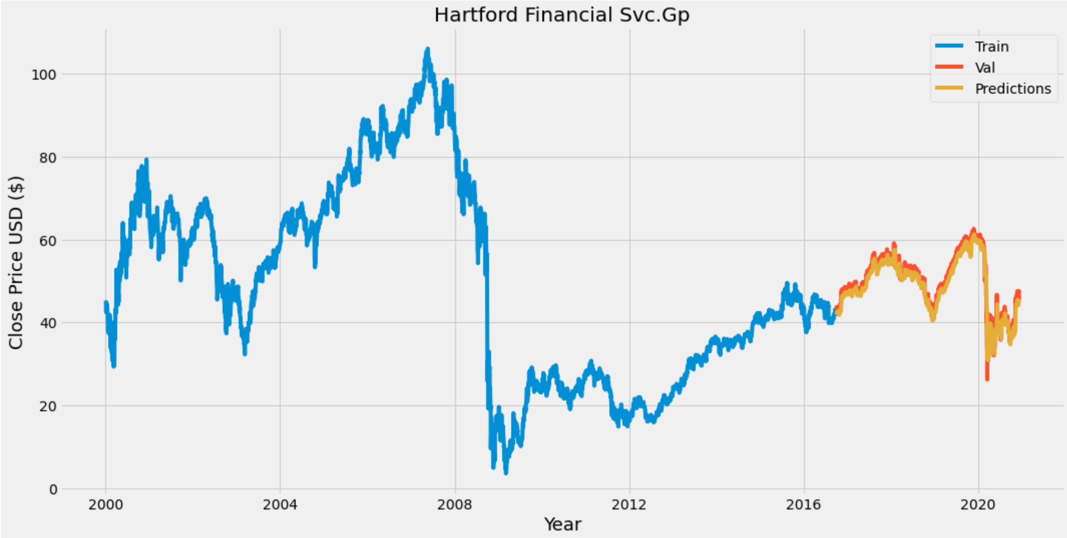
Graph 51 : Time Series values of the share of Omnicom Group.
Source: Author`s Worksheet



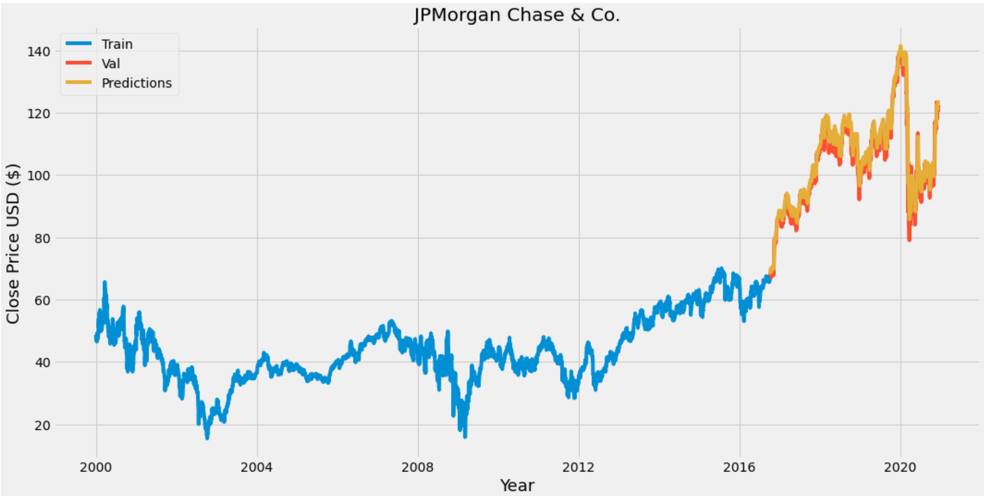
Graph 52 : Time Series values of the share of Fifth Third Bancorp
Source: Author`s Worksheet

Graph 53 : Time Series values of the share of Globe Life Inc.
Source: Author`s Worksheet

Graph 54 : Time Series values of the share of Hartford Financial Svc. Gp



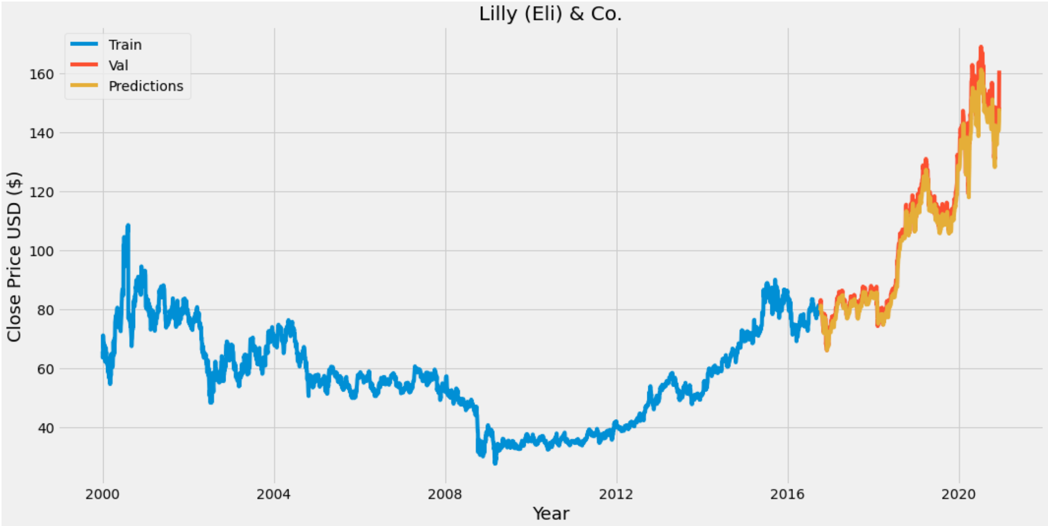
Graph 54 : Time Series values of the share of Hartford Financial Svc. Gp
Source: Author`s Worksheet



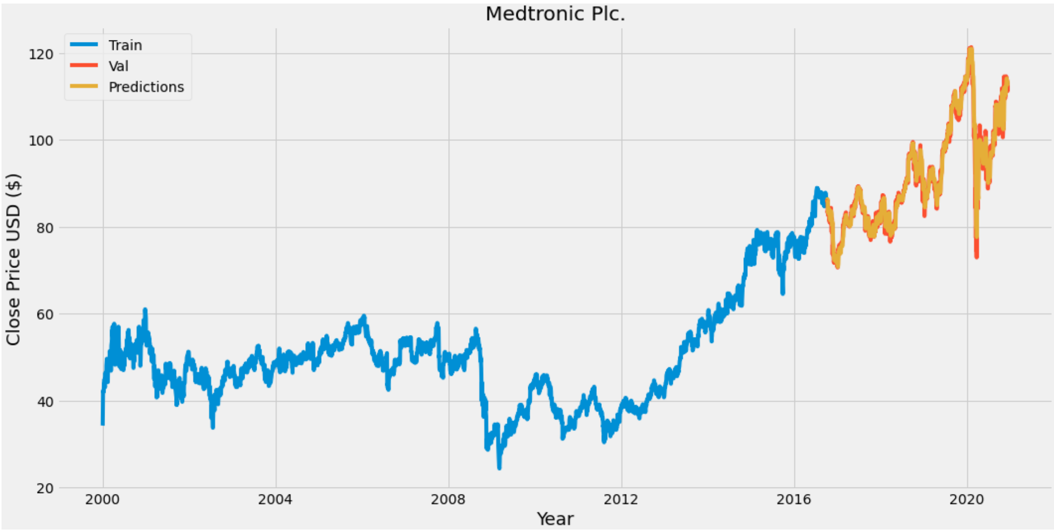
Graph 55 : Time Series values of the share of JPMorgan Chase & Co.
Source: Author`s Worksheet

Graph 56 : Time Series values of the share of KeyCorp
Source: Author`s Worksheet

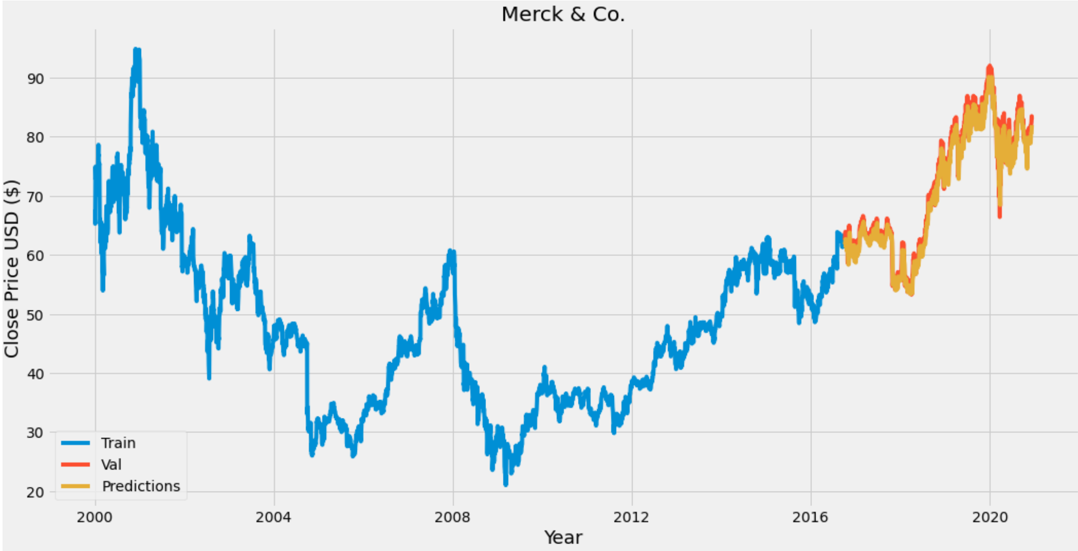
Graph 57 : Time Series values of the share of Johnson & Johnson
Source: Author`s Worksheet



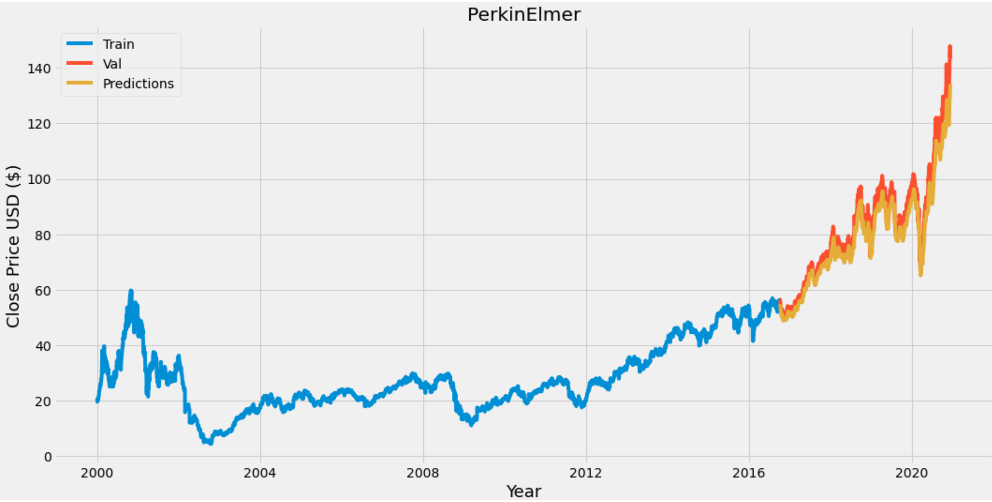
Graph 58 : Time Series values of the share of Lilly (Eli) & Co.
Source: Author`s Worksheet



Graph 59 : Time Series values of the share of Medtronic Plc.
Source: Author`s Worksheet

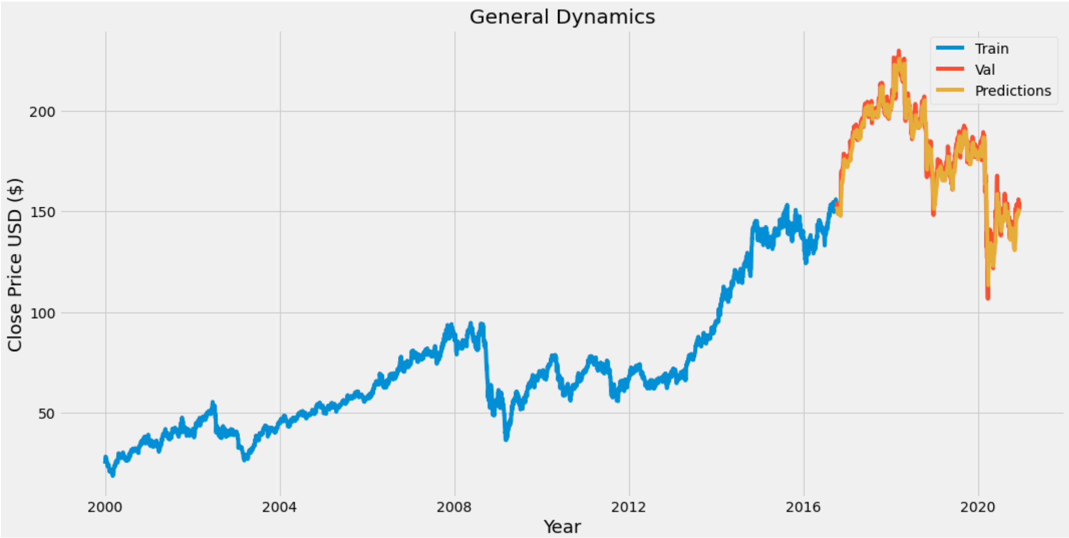


Graph 60 : Time Series values of the share of Merck & Co.
Source: Author`s Worksheet

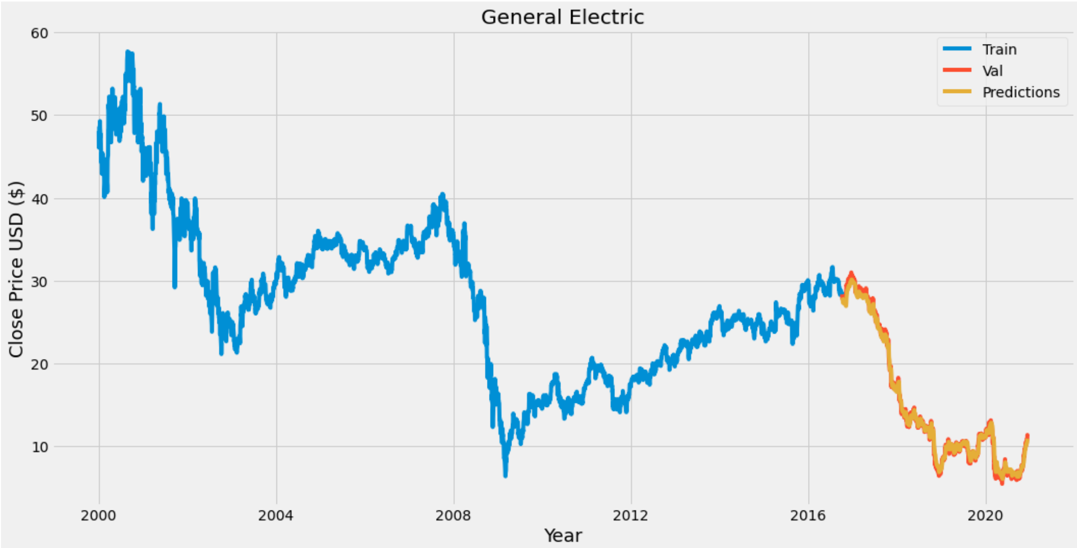


Graph 61 : Time Series values of the share of PerkinElmer
Source: Author`s Worksheet

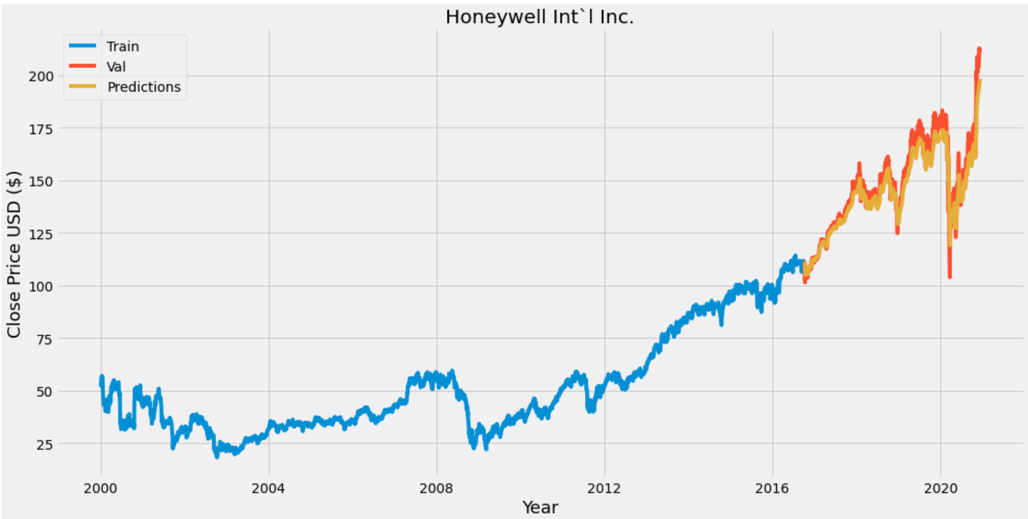
Graph 62 : Time Series values of the share of Pfizer Inc.
Source: Author`s Worksheet



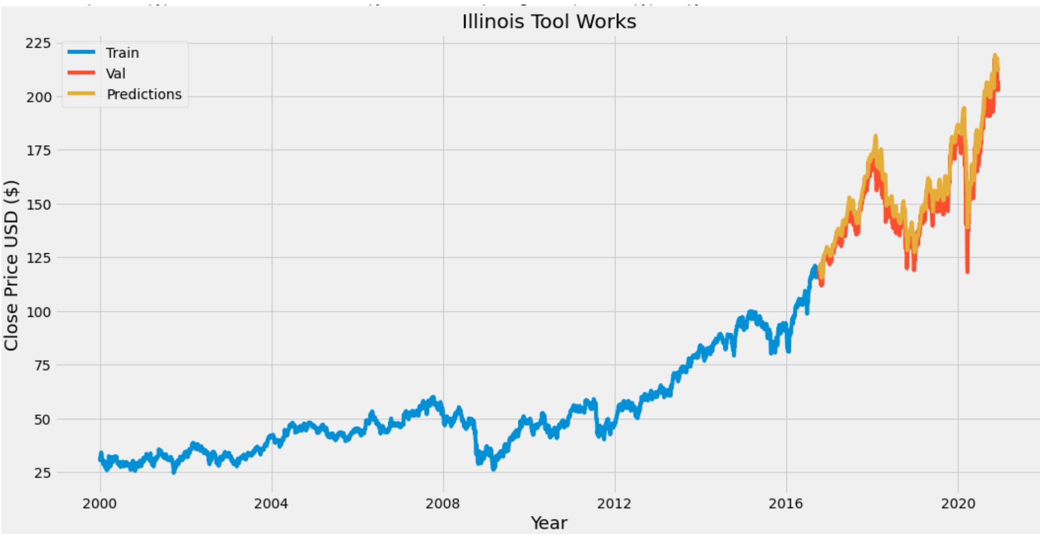
Graph 63 : Time Series values of the share of General Dynamics
Source: Author`s Worksheet



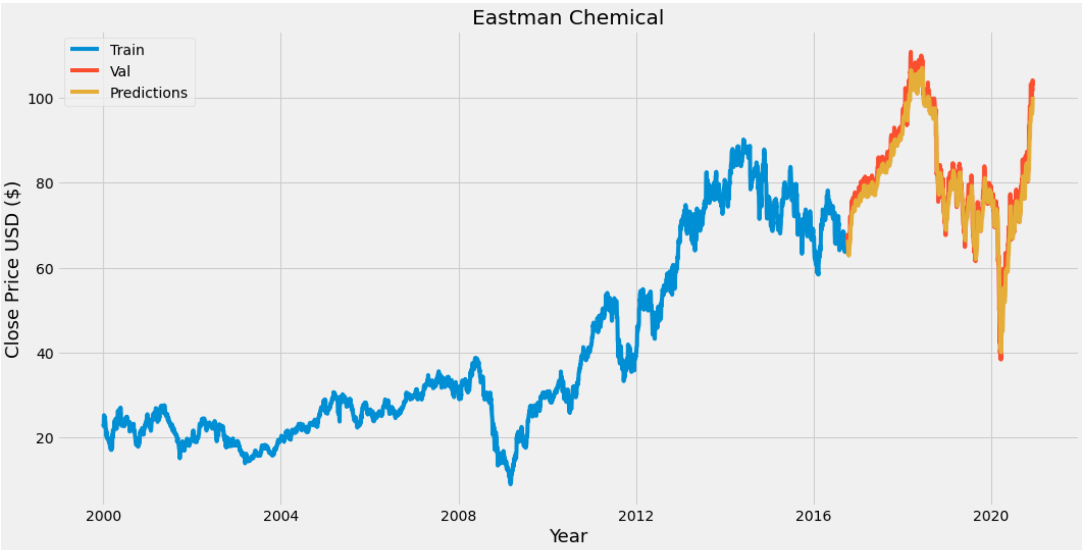
Graph 64 : Time Series values of the share of General Electric
Source: Author`s Worksheet



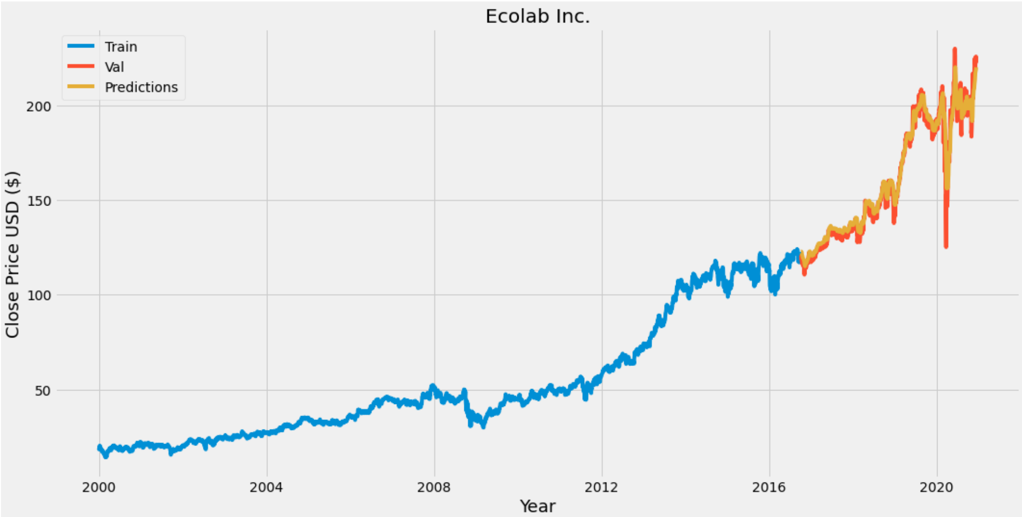
Graph 65 : Time Series values of the share of Honeywell Int`l Inc.
Source: Author`s Worksheet



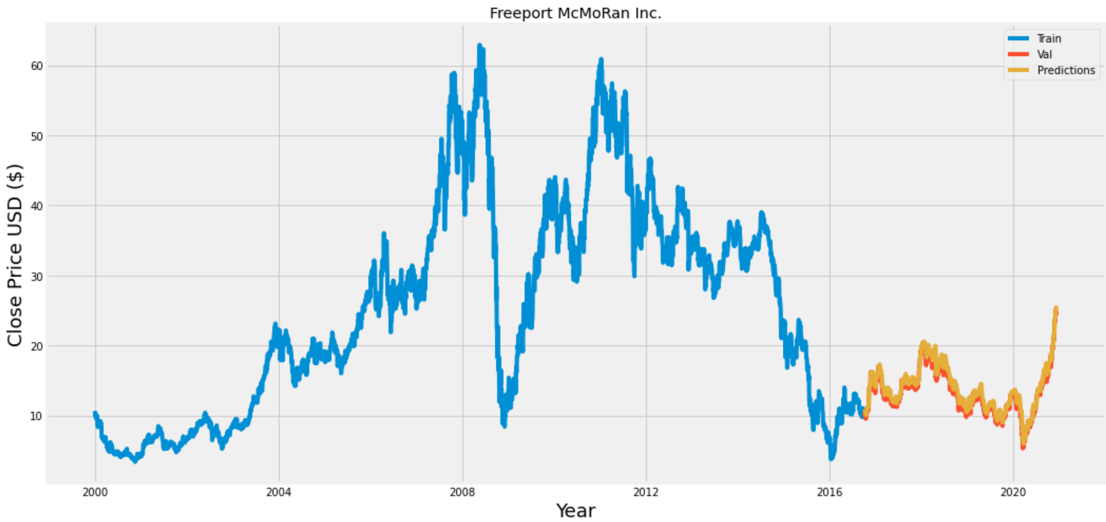
Graph 66 : Time Series values of the share of Illinois Tools Works
Source: Author`s Worksheet



Graph 67 : Time Series values of the share of Eastman Chemical
Source: Author`s Worksheet

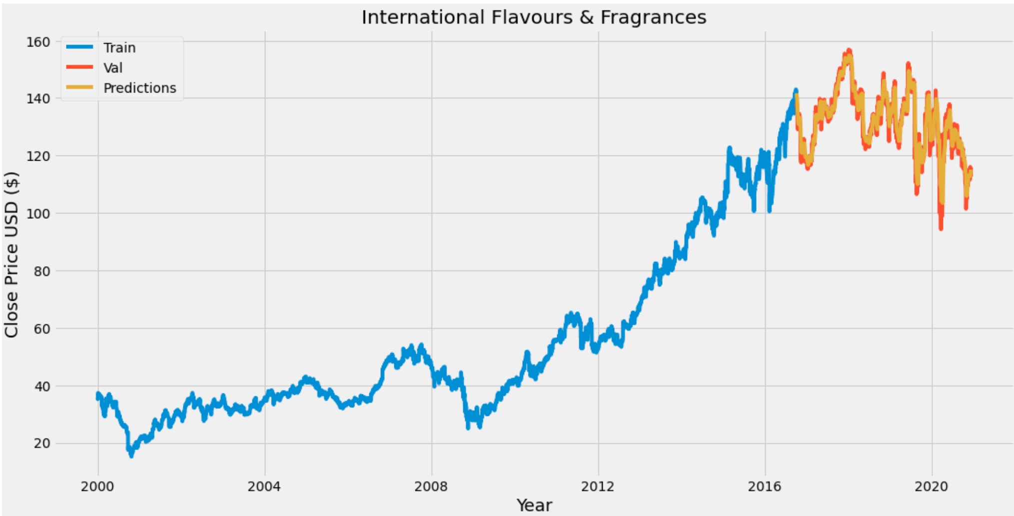


Graph 68 : Time Series values of the share of Ecolab Inc.
Source: Author`s Worksheet



Graph 69 : Time Series values of the share of Freeport McMoRan Inc.
Source: Author`s Worksheet

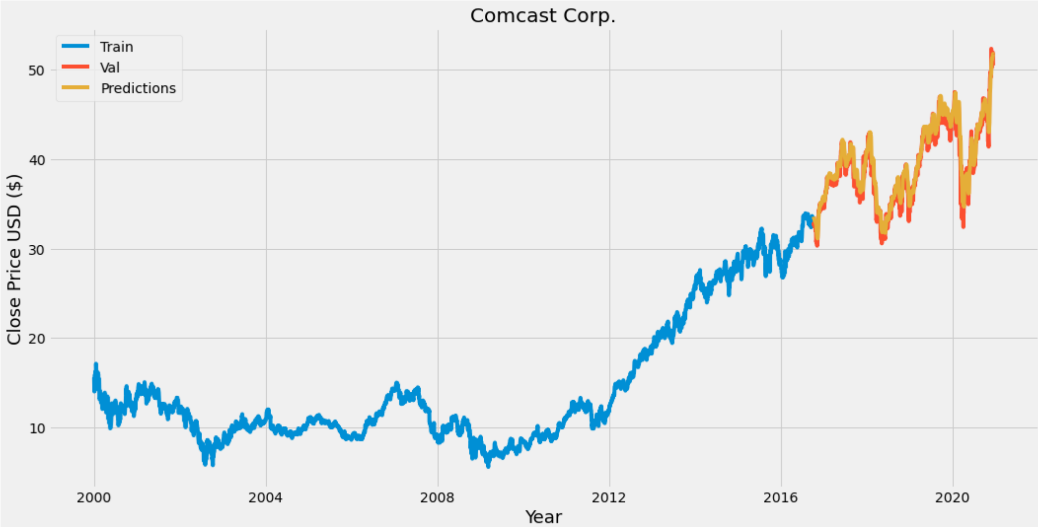
Graph 70 : Time Series values of the share of International Paper
Source: Author`s Worksheet



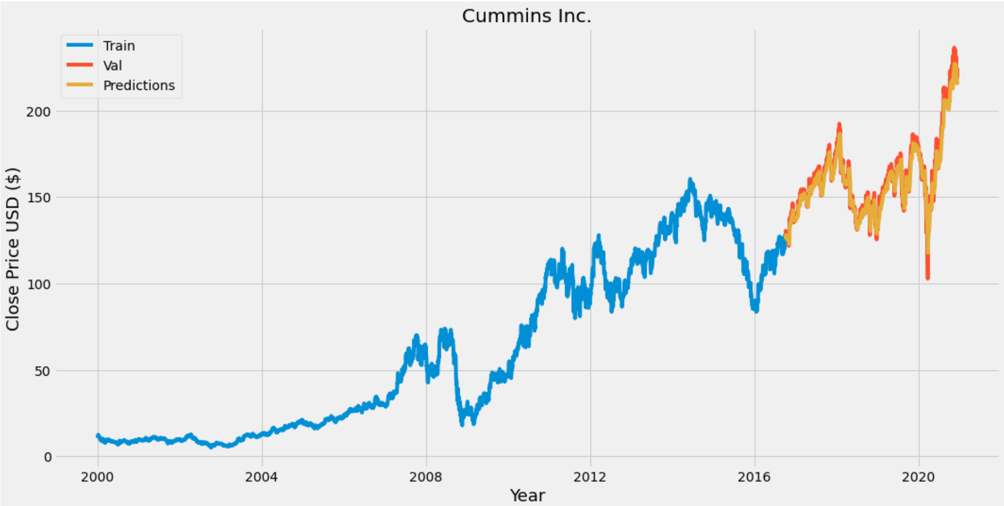
Graph 71 : Time Series values of the share of International Flavours & Fragnance
Source: Author`s Worksheet

Graph 72 : Time Series values of the share of Linde Plc.
Source: Author`s Worksheet

Graph 73 : Time Series values of the share of AT & T Inc.
Source: Author`s Worksheet

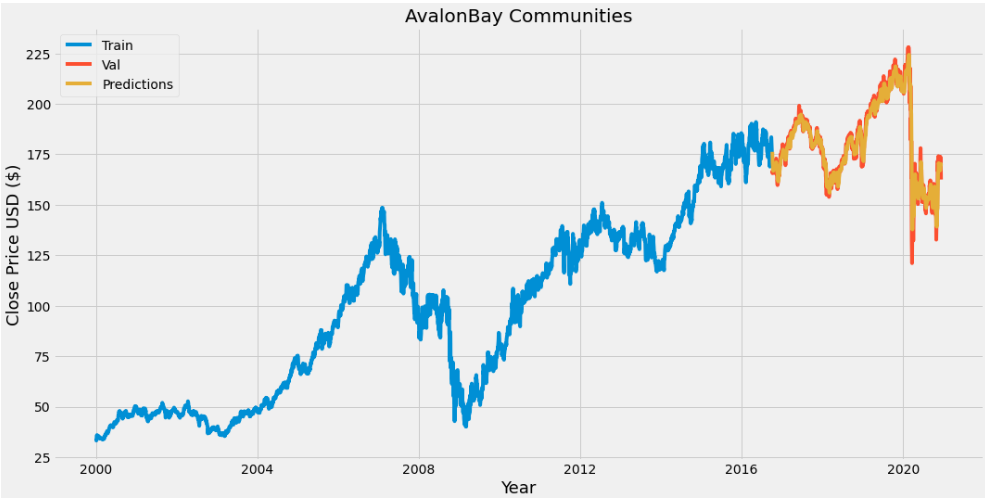


Graph 74 : Time Series values of the share of Comcast Corp.
Source: Author`s Worksheet

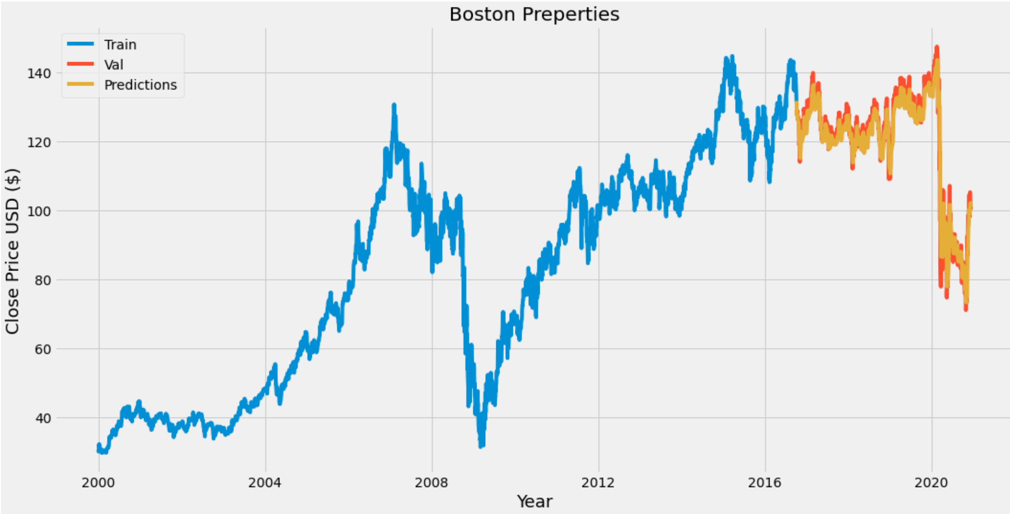


Graph 75 : Time Series values of the share of Cummins Inc.
Source: Author`s Worksheet

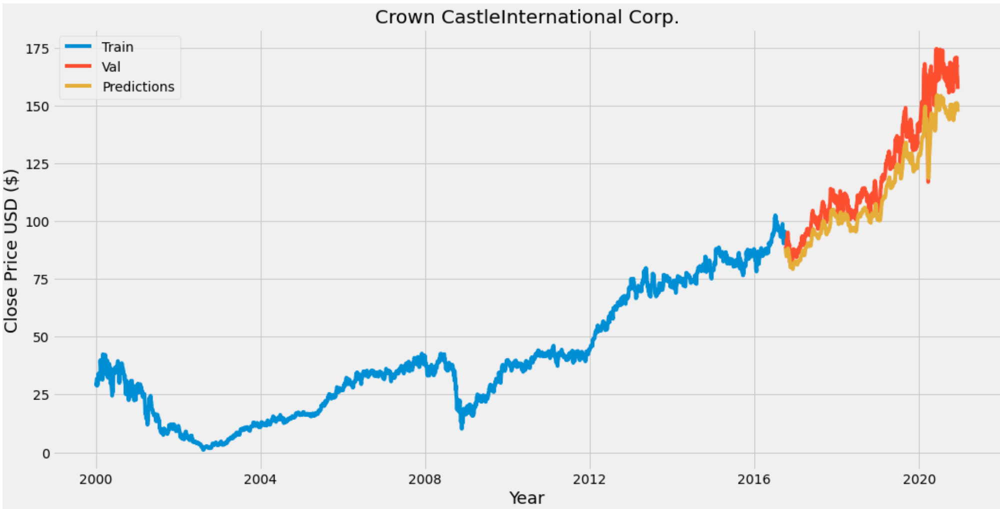
Graph 76 : Time Series values of the share of Apartment Investment & Management
Source: Author`s Worksheet



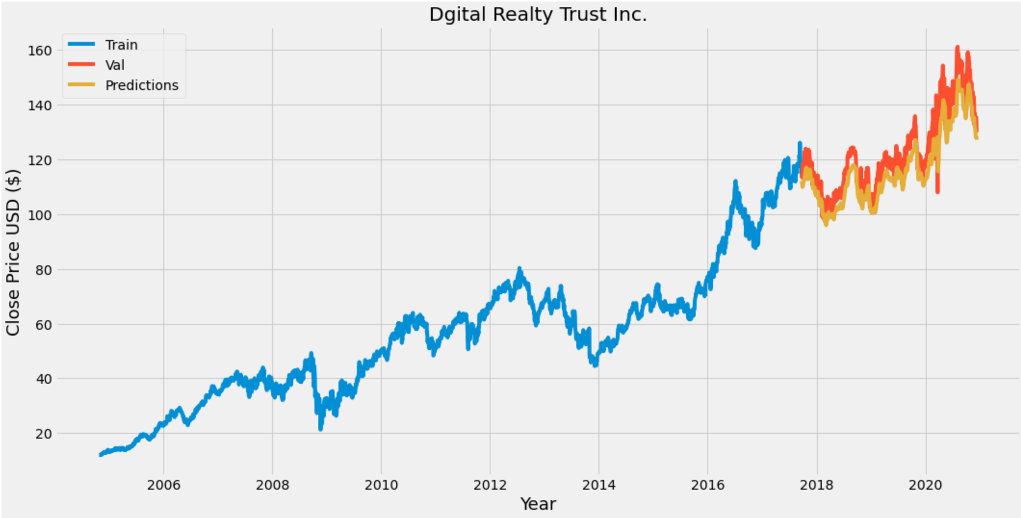
Graph 77 : Time Series values of the share of AvalonBay Communities
Source: Author`s Worksheet



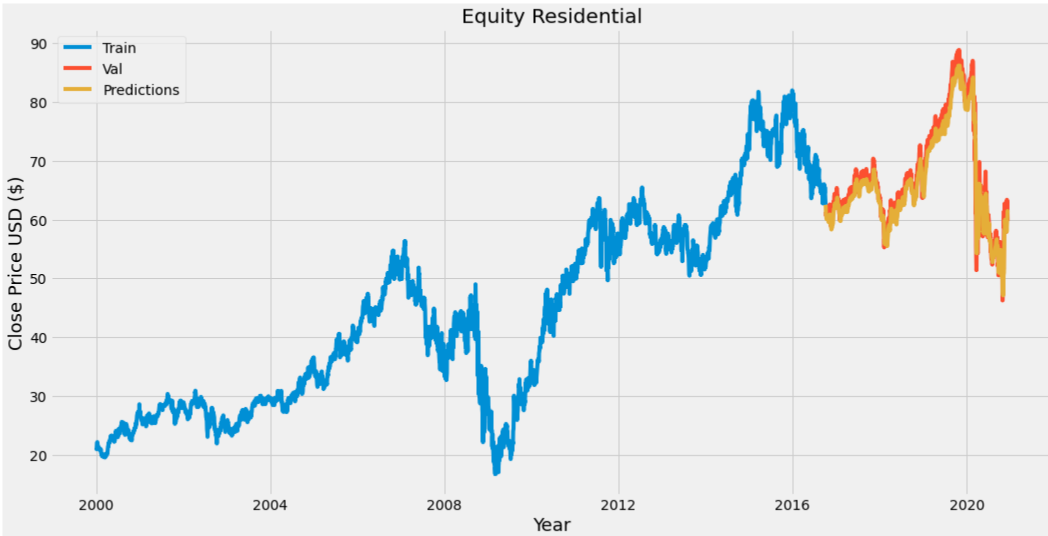
Graph 78 : Time Series values of the share of Boston Properties
Source: Author`s Worksheet



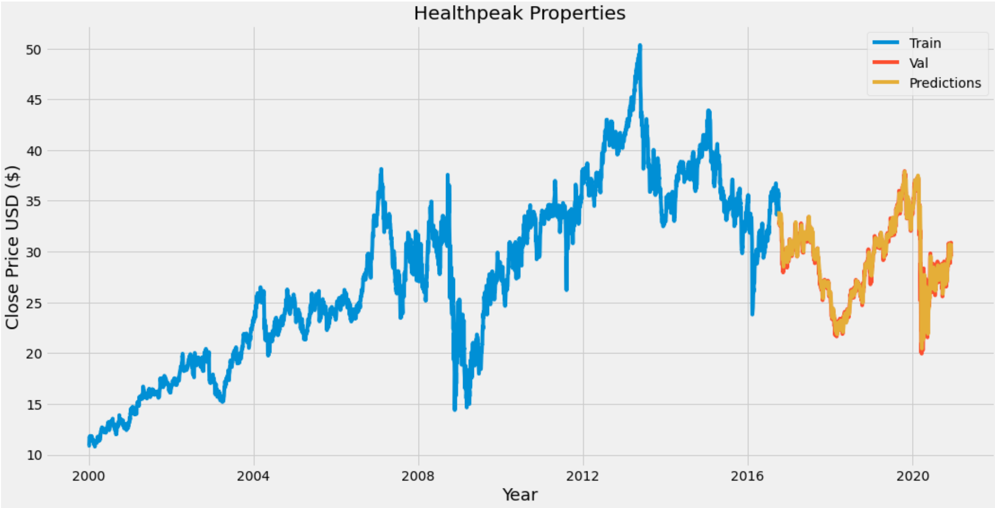
Graph 79 : Time Series values of the share of Crown Castle International Corp.
Source: Author`s Worksheet



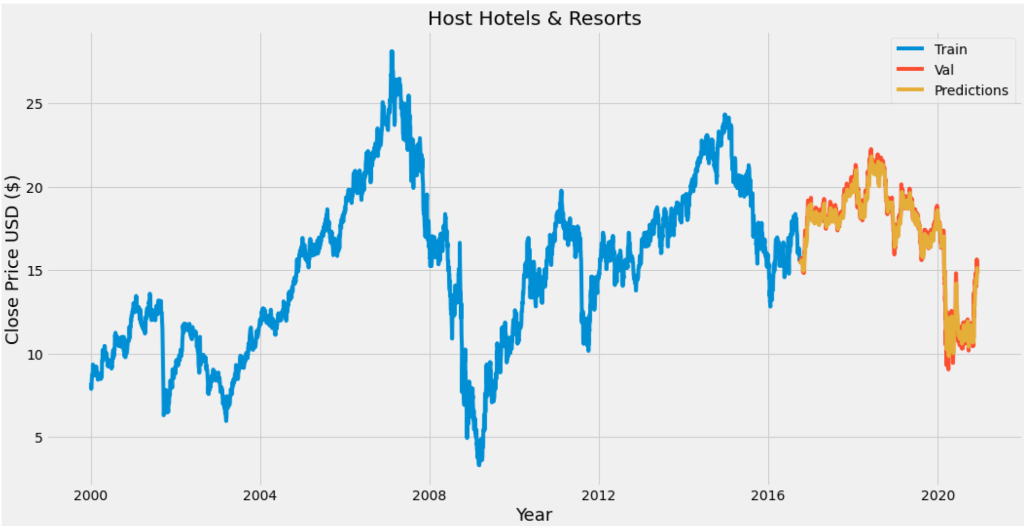
Graph 80 : Time Series values of the share of Digital Realty Trust Inc.
Source: Author`s Worksheet



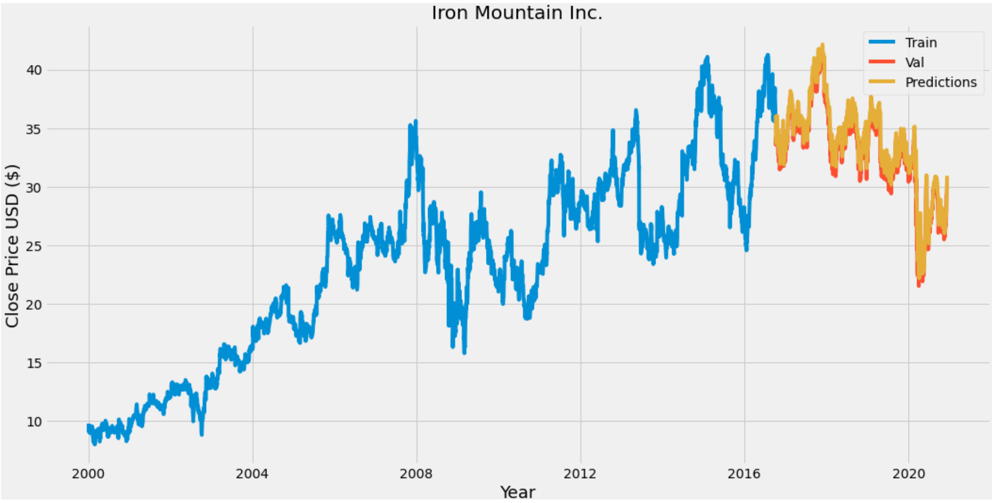
Graph 81 : Time Series values of the share of Equity Residential.
Source: Author`s Worksheet



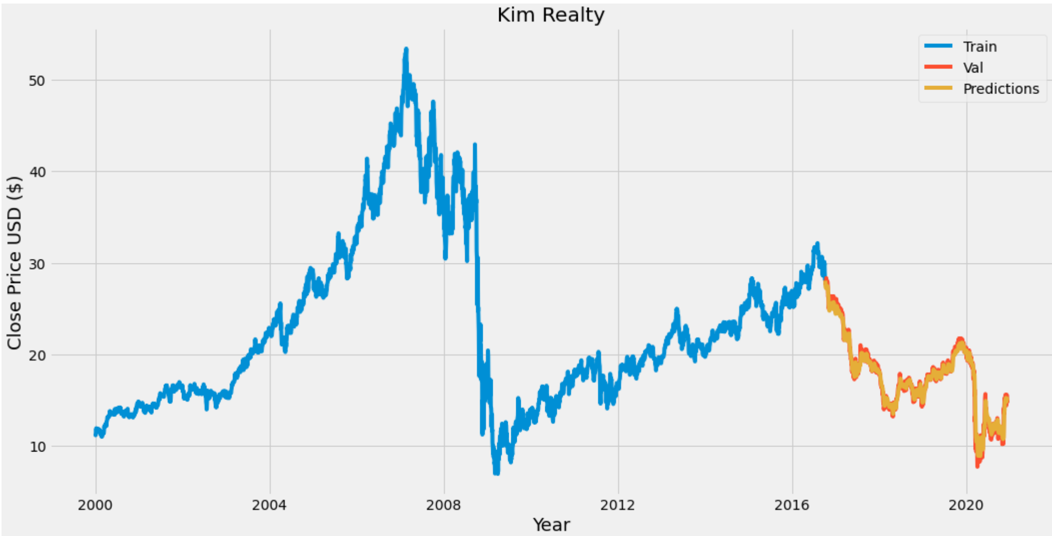
Graph 82 : Time Series values of the share of Healthpeak Properties
Source: Author`s Worksheet



Graph 83 : Time Series values of the share of Host Hotels & Resorts
Source: Author`s Worksheet

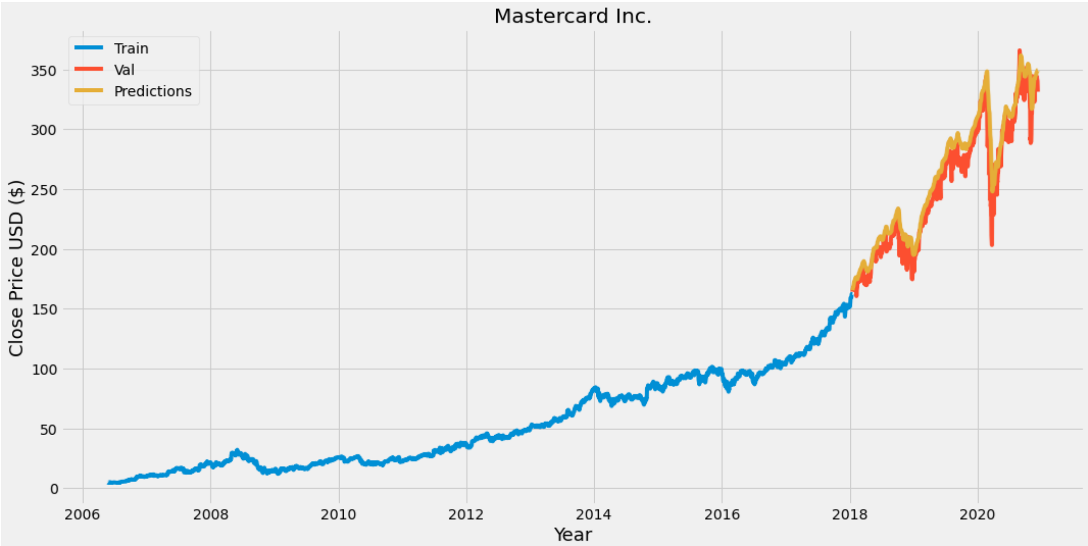


Graph 84 : Time Series values of the share of Iron Mountain Inc.
Source: Author`s Worksheet

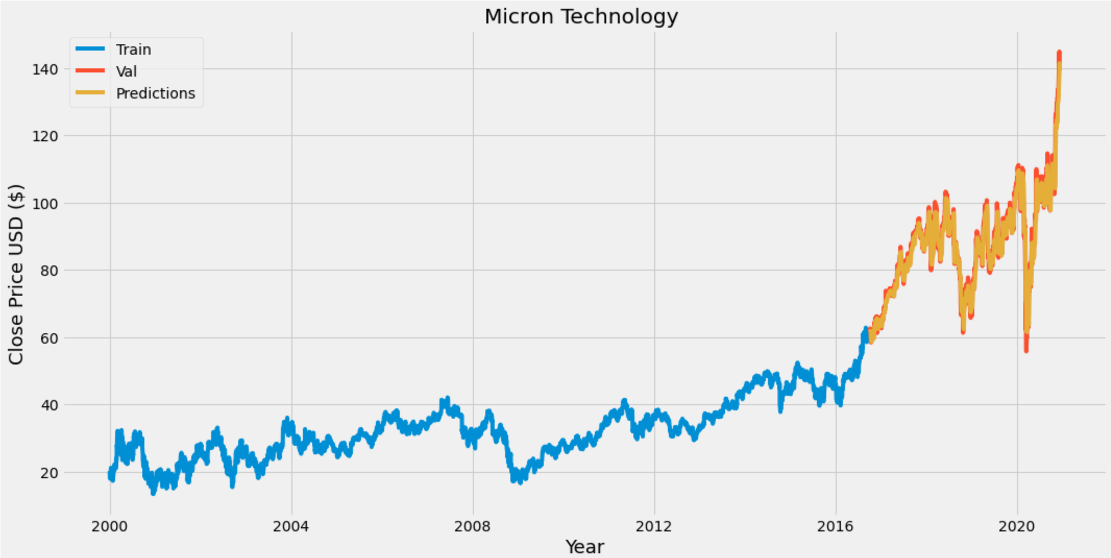


Graph 85 : Time Series values of the share of Kim Realty.
Source: Author`s Worksheet

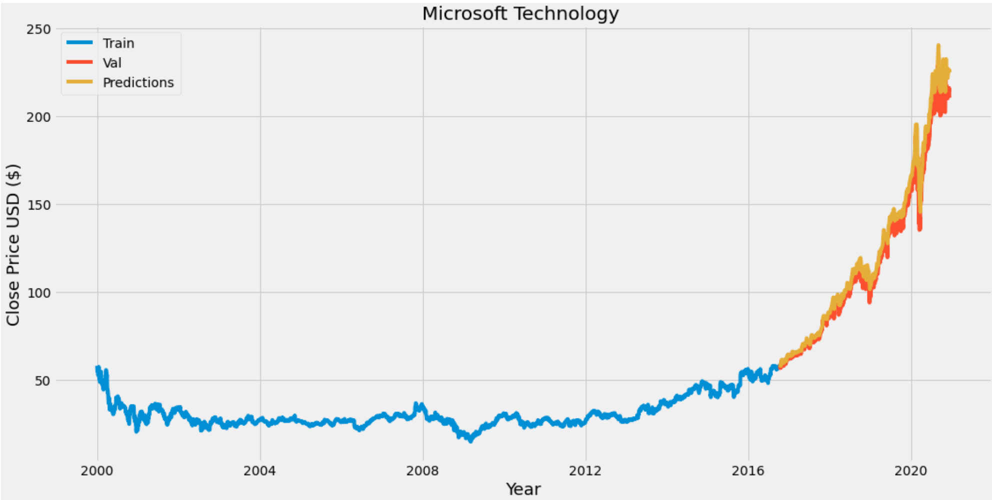
Graph 86 : Time Series values of the share of KLA Corporation.
Source: Author`s Worksheet



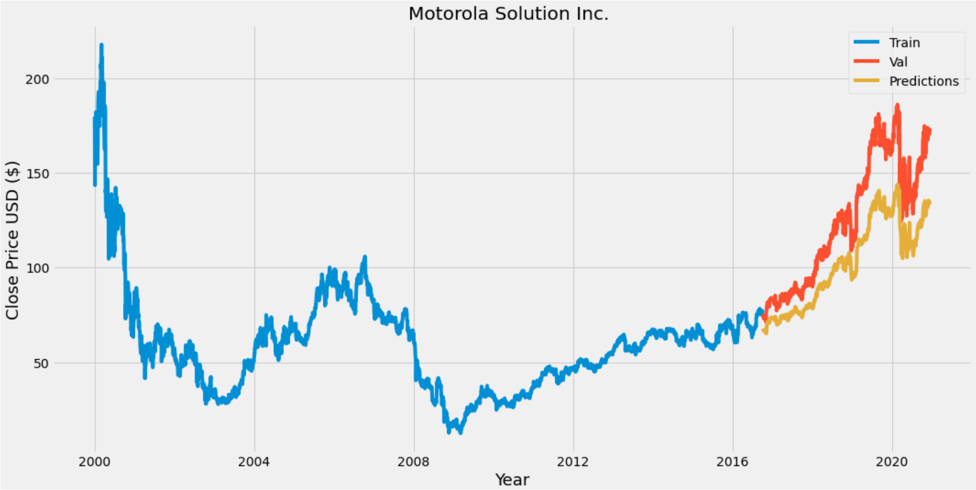
Graph 87 : Time Series values of the share of Mastercard Inc
Source: Author`s Worksheet



Graph 88 : Time Series values of the share of Micron Technology
Source: Author`s Worksheet



Graph 89 : Time Series values of the share of Microsoft Technology
Source: Author`s Worksheet



Graph 90 : Time Series values of the share of Motorola Solution Inc.
Source: Author`s Worksheet