STOCK PRICE FORECASTING USING BAYESIAN NETWORK

by
Parisa Tabassum (ID: 14301009)
Mita Halder (ID: 14301012)

Supervisor
Dr. Mahbub Alam Majumdar

A thesis submitted to the faculty of
BRAC University
in partial fulfillment of the requirements for the degree of

Bachelor of Science
in
Computer Science and Engineering

Department of Computer Science and Engineering
BRAC University

August, 2018

STATEMENT OF THESIS AUTHORSHIP

We hereby declare that this is our own work to the best of our knowledge and belief and we understand that it does not contain any material that has been published or written by another researcher, nor does it contain any material which, to a substantial extent, has been accepted for any award of any other degree or diploma at BRAC University or any other educational institution, except where due acknowledgements made in the thesis.

Signatures of the Authors:

____

Parisa Tabassum   ID: 14301009

____

Mita Halder   ID: 14301012

STATEMENT OF THESIS APPROVAL

The thesis of Parisa Tabassum and Mita Halder has been approved by the following supervisor.

Signature of the Supervisor:

____________________________ ____________________________

Dr. Mahbub Alam Majumdar Date Approved

Professor

Department of Computer Science and Engineering

BRAC University

ACKNOWLEDGEMENTS

First and foremost, we express our solemn to almighty Allah, the merciful, who has given us the ability, strength and opportunity to perform this thesis work.

We are grateful to our supervisor, Dr. Mahbub Alam Majumdar, whose contribution in stimulating suggestions and encouragement, helped us to coordinate our research especially in coming up with innovative ideas and moving forward to the next level. We take this opportunity to express a deep sense of gratitude to him for his aspiring guidance, constant supervision, and friendly advice during the research work, which were extremely valuable, both theoretically and practically.

We want to thank our parents for their immense support and believe throughout, to reach our goal for this research. We thank all our friends inside the Department of Computer Science and Engineering at BRAC University.

Lastly, we would like to express our gratitude to Department of Computer Science and Engineering, BRAC University and our teachers for their reluctance support throughout our struggle in completing our thesis on time.

TABLE OF CONTENTS

ABSTRACT

In a financially volatile market, as the stock market, it is important to have a very precise prediction of a future trend. Because of the financial crisis and scoring profits, it is mandatory to have a secure prediction of the values of the stocks. Predicting a non-linear signal requires advanced algorithms of machine learning. The literature contains the stock price prediction algorithm by using Bayesian network. The network is determined from the daily stock price. The prediction error is evaluated from the daily stock price and its prediction. The present algorithm is applied for predicting Google, Procter & Gamble and General Motors stock price. The results of this study show that the algorithm is capable of predicting future stock price more accurately than a lot of another machine learning algorithm available so far.

KEYWORDS

Stock Market, Bayesian Network, Ward Method, K2 Algorithm.

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

Stock markets are trading institutions [24] where stocks (equity) and other financial instruments such as bonds are offered for trade. For stocks, the market generally operates a 'willing-buyer, willing-seller' trade, where buyers' and sellers' prices are matched for a fit. If there is no match, then no trade takes place and waits for a future match or expires. In most stock 3 exchanges, the common and easily accessible market is the equity market (stocks), where the entry investment can be as low as USD1. The equity market is therefore more active, having many players and hence a segment worthy of further study. The performance of stock markets is measured on a daily basis by some key indicators such as 'share index', which is a measure of the performance of some stocks picked from the different sectors of the market. Such an index is important in not only gauging the performance of trades in the stock exchange but also the economic performance of the particular country as a whole. Shareholders however do not directly execute the trade, nor is there any meeting between buyers and sellers for negotiations. Shareholders trade by giving instructions to their Stockbrokers, who in turn execute the orders. Stockbrokers usually also advise clients on where to trade. In their advisory role, some Stockbrokers base their advice on the fundamentals of the various stocks or undertake technical analysis. However, none of these predictive methods have assurance of profit as they usually just indicate a future trend and a likely up or down price movement and not the real expected future stock price. Stockbrokers need to be empowered, through better predictive tools, to enable them have some capability to provide the best advice to their clients. A predictive tool that Stockbrokers can use to guide on exact price movements, as a basis of investment, is therefore desirable. This can be an artificial intelligence (AI) system based on neural networks. Due to the importance of stock markets, investment is usually guided by some form of prediction. However, predicting the stock market is not a trivial task. To start with, there is need to model the trend of the stock prices, which is nonlinear. For this reason, we decided to develop a model to predict stock price. We developed the model using Bayesian networks. For this process, we used three different company's stock prices. At first, we transformed stock price process in to discrete values set by ward method. Then we determined the Bayesian network from the discrete values' set. At last we predicted the stock price using the network.

1.2 BAYESIAN NETWORK

Bayesian Networks [23] are a type of Probabilistic Graphical Model (PGM). A PGM is a model where a graph expresses the conditional dependence structure between random variables. It uses a set of assumptions concerning the generation of some sample data, and similar data from a larger population. Bayesian Networks [25] are a type of PGM that are used to build models from data or expert opinion. They are also commonly referred to as Bayes nets, Belief networks and sometimes Causal networks.

Bayesian probability is an interpretation of the concept of probability, in which, instead of frequency or propensity of some phenomenon, probability is interpreted as reasonable expectation representing a state of knowledge or as quantification of a personal belief.

Bayesian networks are probabilistic and graphical. They are built from probability distributions and also use the laws of probability for prediction and anomaly detection, for reasoning and diagnostics, decision making under uncertainty and time series prediction. Bayesian Networks use Directed Acyclic Graph (DAG) to analyze data and their correlation.

A Bayesian network is a graph which is made up of Nodes and directed Links between them. In Bayesian networks, each node represents a variable such as someone's height, age or gender or multiple variables. Nodes with more than one variable are known as multi-variable nodes. A variable might be discrete or might be continuous. A discrete variable is one with a set of mutually exclusive states such as Gender = {Female, Male}. A continuous variable is a variable that has an infinite number of possible values such as someone's age. Continuous distributions can depend on each other and can also depend on one or more discrete variables. It is capable of using Latent variables which can model hidden relationships. Links are added between nodes to indicate that one node directly influences the other. When a link does not exist between two nodes, this does not mean that they are completely independent, as they may be connected via other nodes. They may however become dependent or independent depending on the evidence that is set on other nodes. Although links in a Bayesian network are directed, information can flow both ways.

1.3 APPLICATIONS OF BAYESIAN NETWORK

Bayesian Networks can be used [23] for a wide range of tasks. It is frequently used in prediction. Time series prediction is a great example of BN Prediction. It is popular for its usefulness in medical sector. It is used in anomaly detection, diagnostics and medicine prescribing. It can be used for the technology behind biomonitoring, document classification, information retrieval, semantic search, image processing, spam filter, turbo code, system biology etc. Reasoning and decision making under uncertainties are great use of Bayesian networks. It can also be used to develop gene regulatory networks. Automated insight technologies are developed using Bayesian networks. Overtime as more data is used by the algorithm the performance of the system improves. When the network has access to better data we see dramatic breakthroughs. So really great improvements can happen when we apply Bayesian networks to as many fields as possible.

1.4 STOCK MARKET WORK PROCESS

When we buy a stock, we’re buying a piece of the company. When a company needs to raise money, it issues shares. This is done through an initial public offering (IPO), in which the price of shares is set based how much the company is estimated to be worth [10], and how many shares are being issued. The company gets to keep the money raised to grow its business, while the shares (also called stocks) continue to trade on an exchange, such as the New York Stock Exchange (NYSE).

Traders and investors continue to buy and sell the stock of the company on the exchange, although the company itself no longer receives any money from this type of trading. The company only receives money from the IPO.

Traders and investors continue to trade a company's stock after the IPO because the perceived value of company changes over time. Investors can make or lose money depending on whether their perceptions are in agreement with "the market." The market is the vast array of investors and traders who buy and sell the stock, pushing the price up or down.

Trying to predict which stock will rise or fall, and when, is very difficult. Over time stocks as a whole tend to rise, which is why many investors choose to buy a basket of stocks in various sectors (this is called diversification) and hold them for the long-term. Investors who use this approach do not concern themselves with moment-to-moment fluctuations in stock prices. The ultimate goal of buying shares is to make money by buying stocks in companies you expect to do well, those whose perceived value (in the form of the share price) will rise.

Mature and established companies may also pay a dividend to shareholders. A dividend is a cut of the company's profit, which the company sends to shareholders as long as the company continues to pay the dividend. Aside from the dividend, the share price will continue to fluctuate. The losses and gains associated with the share price are independent of the dividend. Dividends can be large or small – or nonexistent (many stocks don't pay them). Investors seeking regular income from their stock market investments tend to favor buying stocks that pay high dividends.

When you buy shares of a company, you own a piece of that business and therefore have a vote in how it is run. While there are different classes of shares (a company can issue shares more than once), typically owning shares gives you voting rights equal to the number of shares you own. Shareholders as a whole, based on their individual votes, select a board of directors and can vote on major decisions the company is making.

For every stock transaction, there must be a buyer and a seller. When you buy 100 shares of stock (called a "lot") someone else must sell it to you. Either buyers or sellers can be more aggressive than the other, pushing the price up or down.

When the price of a stock goes down, sellers are more aggressive because they are willing to sell at a lower and lower price. The buyers are also timid and only willing to buy at lower at lower prices. The price will continue to fall until the price reaches a point where buyers step in and become more aggressive and willing to buy at higher prices, pushing the price back up.

Investors don't all have the same agenda, which leads traders to sell stocks at different times. One investor may hold stock that has grown significantly in price and sells to lock in that profit and extract the cash. Another trader may have bought at a higher price than the stock now sells for, putting the trader in a losing position. That trader may sell to keep the loss from getting bigger. Investors and traders may also sell because they believe stock is going to go down, based on their research, and want to take their money out before it does.

How many shares change hands in a day is called volume. Many stocks on major exchanges, such as the NYSE or NASDAQ, have millions of shares issued. That means potentially thousands of investors in a stock may decide to buy or sell on any particular day. A stock that has lots of daily volume is attractive to investors because the volume means they can easily buy or sell their shares whenever they please.

When volume is inadequate, or no one is actively trading a stock, it's still usually possible to dispose of a small number of shares because the exchanges mandate certain traders (firms) to provide volume. These traders are commonly referred to as market makers and act as buyers and sellers of last resort when there are no buyers or sellers. They don't have to stop a stock from rising or falling though, which is why most traders and investors still choose to trade stocks with lots of volume, and thus not rely on these "market makers," which are now mostly electronic and automated. There are still people on the floor of the NYSE. Those men and women in the blue jackets trade stocks for their firms and also help facilitate orders from the public.

Stocks are issued by companies to raise cash, and the stock then continues to trade on an exchange. Overall stocks have risen over the long-term, which makes owning shares attractive. There are also additional perks such as dividends (income), profit potential and voting rights. Share prices also fall, though, which is why investors typically choose to invest in a wide array of stocks, only risking a small percentage of their capital on each one. Shares can be bought or sold at any time, assuming there is enough volume available to complete the transaction, which means investors can cut losses or take profits whenever they wish.

1.5 STOCK PRICE PREDICTION

Stock market prediction [24] is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. It is a popular and important topic in financial and academic studies. Share Market is an untidy place for predicting since there are no significant rules to estimate or predict the price of share in the share market. Prediction methodologies fall into two broad categories which can (and often do) overlap. They are fundamental analysis and technical analysis. Fundamental analysis is Analyzing whether the company's current stock prices reflects its future revenue-generating potential. Technical analysis is the type of analysis which predicts future stock price movements on the basis of historical patterns. Several factors can affect stock price of a company including company news, company performance, industry performance, investor sentiment, interest rates, economic outlook, inflation, deflation, economic shocks, political shocks, changes in economic policy etc.

1.6 THESIS ORIENTATION

The subsequent sections of the paper have been organized as follows. Chapter 2 features the related work and existing approaches based on our proposed method. Chapter 3 introduces the dataset used in this paper, as well as the proposed approach for handling the data. Chapter 4 explains the algorithms we used to determine the network along with the prediction procedure. Finally, Chapter 5 contains the experimental results and it concludes and summarizes the report along with discussing our future plan.

CHAPTER 2

LITERATURE REVIEW

Stock Price Forecasting is getting more and more attention as an active area of research recently. Theoretical and experimental challenges in this field is peaking researchers' interests and its potential to yield significant profit is motivating companies to fund the researches. Since its creation in March 8, 1817, stock market has been a constant interest of scholars. Hasbrouck, Sofianos and Sosebee's [1] paper on New York Stock Exchange Systems and Trading Procedures gives a detailed insight on New York Stock Exchange systems, trading rules and procedures. It is successful in fulfilling its objective to provide researchers with a detailed institutional framework for studying quote and transaction data generated by U.S. securities trading.

Yue Xu [2] combined the time series analysis technique with information from the Google trend website and the Yahoo finance website to predict weekly changes in stock price. Important news or events related to a specific stock over a five-year span were recorded and the weekly Google Trend index values on this stock were used to measure the magnitude of these events. Results obtained revealed correlation between the changes in weekly stock prices and the values of important news or events from the Google trend website.

Forecasting stock price is a very complicated task. Most stockbroker use technical, fundamental or time series analysis in trying to predict stock prices. However, these strategies do not result in reliable outcome as they guide on trends and not the most likely price. It is required to use enhanced methods to predict the most accurate result. Researchers have used different methods and different sets of input variables to predict stock prices over the past few decades.

Adebiyi and Adewumi [3] built stock price predictive model using the autoregressive integrated moving average (ARIMA) model. Results showed that the ARIMA model has a strong potential for short-term prediction.

Xing, Sun, Wang, & Yu. [24] introduce a kind of method based on Hidden Markov Model to forecast stock price trend. Which is different from the existing stock prediction, they attempted to find the hidden relationship existing between the stock prices and corresponds to the Hidden Markov Model. The experimental result showed that, this method can get pretty accurate result, particularly effective in short period prediction.

Marček [4] described the basic notion of fuzzy linear regression models based on fuzzy parameter extension principle. The paper presented the autoregressive (AR) model which used the fuzzy parameters extension principle and Fuzzy Neural Network (FNN) principle for estimating and predicting stock prices. The results presented that initial results from using FNN architecture were better than basic Artificial Neural Network (ANN) architecture for daily frequencies.

Hegazy, Soliman and Salam [5] implemented a machine learning model to predict stock market price. The algorithm integrated Particle swarm optimization (PSO) and least square support vector machine (LS-SVM). The PSO algorithm selected best free parameters combination from the study of stocks historical data and technical indicators for LS-SVM to avoid over-fitting and local minima problems. The proposed model was applied and evaluated using thirteen benchmark financials datasets and compared with artificial neural network with Levenberg-Marquardt (LM) algorithm. The obtained results revealed that the proposed model has better prediction accuracy and the potential of PSO algorithm in optimizing LS-SVM.

Tiwari, Bharadwaj and Gupta [22] proposed use of Data analytics to be used in assist with investors for making right financial prediction so that right decision on investment can be taken by Investors. Two platforms were used for operation: Python and R. various techniques like Arima, Holt winters, Neural networks (Feed forward and Multi-layer perceptron), linear regression and time series are implemented to forecast the opening index price performance in R. While in python Multi-layer perceptron and support vector regression were implemented for forecasting Nifty 50 stock price and also sentiment analysis of the stock was done using recent tweets on Twitter. Nifty 50 (ANSEI) stock indices was considered as a data input for methods which are implemented. 9 years of data was used. The accuracy was calculated using 2-3 years of forecast results of R and 2 months of forecast results of Python after comparing with the actual price of the stocks.

Mittal and Goel [6] applied sentiment analysis and machine learning principles to find the correlation between 'public sentiment' and 'market sentiment'. Twitter data was used to predict public mood and the predicted mood and previous days' Dow Jones Industrial Average(DJIA) values were used to predict the stock market movements. They propose a new cross validation method for financial data in order to test the results and obtained 75.56% accuracy using Self Organizing Fuzzy Neural Networks (SOFNN) on the Twitter feeds and DJIA values from the period June 2009 to December 2009.

Wamkaya and Lawrence [7] proposed the use of Artificial Neural Network that is feedforward multi-layer perceptron with error backpropagation. They developed a model of configuration 5:21:21:1 with 80% training data in 130,000 cycles. The research developed a prototype and tests it on 2008 - 2012 data from stock markets where prediction results show mean absolute percentage error (MAPE) between 0.71% and 2.77%.

Dunne [27] analyzed existing and new methods of stock market prediction. He took three different approaches at the problem: Fundamental analysis, Technical Analysis, and the application of Machine Learning. He found evidence in support of the weak form of the Efficient Market Hypothesis, that the historic price does not contain useful information but out of sample data may be predictive. He showed that Fundamental Analysis and

Machine Learning could be used to guide an investor’s decisions. He demonstrate a common flaw in Technical Analysis methodology and showed that it produces limited useful information. Based on his findings, algorithmic trading programs were developed and simulated using Quantopian.

Selvin, Vinayakumar, Gopalakrishnan, Menonans and Soman K.P [8] experimented on a model independent approach. Instead of fitting the data to a specific model they identified the latent dynamics existing in the data using deep learning architectures. They used three different deep learning architectures for the price prediction of NSE listed companies and compared their performance. The three models were Neural Network (NN), Long Short-Term Memory networks (LSTM) and Convolutional Neural Network (CNN). The performance of the models was quantified using percentage error. The maximum value of error percentage was obtained for each model. The result showed that CNN was giving more accurate results than the other two models.

Kita, Zuo, Harada and Mizuno [9] developed stock price prediction algorithm using Bayesian network. The algorithm used the network twice. First, the network was determined from the daily stock price and then it was applied for predicting the daily stock price which was already observed. The prediction error was evaluated from the daily stock price and its prediction. Second, the network was determined again from both the daily stock price and the daily prediction error and then it was applied for the future stock price prediction. Numerical results showed that the maximum prediction error of the present algorithm was 30%.

CHAPTER 3

DATA HANDLING

3.1 DATA COLLECTION

We collected end of day closing data of Google, Procter & Gamble and General Motors. Closing price is the trading price of a security at the end of the trading day. The New York Stock Exchange has the most famous closing bell. At 4 p.m. Eastern time, the closing bell rings. At that time, the last trading price for each security on the exchange becomes the closing price. The closing price does not necessarily mean the end of all trading on that security for the day. In fact, it simply means the floor of the exchange is closed. After-hours markets remain open, as do other exchanges in other countries and time zones, which provides opportunity for the price to change. A negative closing in the stock market occurs when a company's stock ended the trading day at a lower price than it opened with that day. We selected one multinational technology company, one multinational consumer goods corporation and one multinational design and manufacture corporation to test our system on different types of company and ensure its credibility. We collected these data from yahoo finance as Microsoft Excel document.

3.2 READ DATA USING PYTHON

We used Xlrd package to read data. Xlrd is a Module that allows Python to read data from MS Excel files. It is a library that also allows formatting information from Excel files, whether they are .xls or .xlsx files. The package itself is pure Python with no dependencies on modules or packages outside the standard Python distribution. Xlrd will safely and reliably ignore any of these if present in the file: Charts, Macros, Pictures, any other embedded object, VBA modules Formulas, Comments, Hyperlinks, Auto filters, Advanced Filters, Pivot tables, Conditional formatting and Data validation.

3.3 TRANSFORM DATA INTO DISCRTE VALUES

3.3.1 CLUSTER ANALYSIS

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group known as cluster are more similar to each other than to those in other clusters. It is a process of partitioning a set of data or objects into a set of meaningful sub-classes, called clusters. It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics. It can be used either as a stand-alone tool to get insight into data distribution or as a preprocessing step for other algorithms.

A cluster is a subset of objects which are similar. A subset of objects such that the distance between any two objects in the cluster is less than the distance between any object in the cluster and any object not located inside it. It is a connected region of a multidimensional space containing a relatively high density of objects.

In our thesis, clustering helped us understand the natural grouping or structure in the data set. We found structure in the data by isolating group of examples that are similar. Our clustering method produced high quality clusters in which the intra-class similarity is high and the inter-class similarity is low.

3.3.2 WARD METHOD

Ward method [26] is an approach for performing cluster analysis. Basically, it looks at cluster analysis as an analysis of variance problem, instead of using distance metrics or measures of association.

This method involves an agglomerative clustering algorithm. It will start out at the leaves and work its way to the trunk, so to speak. It looks for groups of leaves that form into branches, the branches into limbs and eventually into the trunk. Ward's method starts out with n clusters of size 1 and continues until all the observations are included into one cluster [18].

Ward method determines the clusters to minimize the Euclid distances from samples to the cluster centers. Euclid distance is ordinary straight-line distance between two points in Euclidean space.

In previous studies, the Ward method and the uniform clustering were compared and then, the results show that the accuracy of Ward method was better than that of the uniform clustering. Therefore, Ward method is employed as the clustering algorithm.

The notation z, C_i and c_i denote the sample, the cluster and its center, respectively. The estimator is given as,

$$D(Ci, Cj) = E(Ci \cup Cj) - E(Ci) - E(Cj) \quad \dots \dots \dots (3.1)$$

$$E(Ci) = \sum_{z \in ci} d(z, ci)^2 \quad \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (3.2)$$

where the notation d(z, c_i) denotes the Euclid distance between z and c_i.

CHAPTER 4

NETWORK DETERMINATION AND PREDICTION ALGORITHM

4.1 CONDITIONAL PROBABILITY

The conditional probability of an event B is the probability that the event will occur given the knowledge that an event A has already occurred. This probability is written P(B|A), notation for the probability of B given A. In the case where events A and B are independent where event A has no effect on the probability of event B, the conditional probability of event B given event A is simply the probability of event B, that is P(B).

If events A and B are not independent, then the probability of the intersection of A and B (the probability that both events occur) is defined by [19]

P(A and B) = P(A)P(B|A).

From this definition, the conditional probability P(B|A) is easily obtained by dividing by P(A):

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} \quad \dots \dots \dots (4.1)$$

When the probabilistic variable x_0 depends on the probabilistic variable x_i, the relation of the variables is represented as,

$$x_0 \rightarrow x_1 \quad \dots \dots \dots \dots (4.2)$$

where the node x_i and x_0 are named as the parent and the child nodes, respectively.

If the node x_0 has multiple parent nodes, the class of the parent nodes is defined as the class,

$$Pa(x_0) = \{x_1, x_2, \dots\dots\dots, x_M\} \quad \dots \dots \dots (4.3)$$

where the notation x_i and M denote the parent nodes of the node x_0 and the total number of the parent nodes, respectively.

The dependency of the child node x_0 to the class of parent nodes $Pa(x_0)$ is quantified by the conditional probability $P(x_0|Pa(x_0))$[17] which is given as,

$$P\left(x0\middle|Pa(x0)\right)\frac{P(x0)P(Pa(x0)|x0)}{P(Pa(x0))} \quad \dots \dots \dots (4.4)$$

Where,

$$P\big(Pa(x0)\big) = \prod_{i=1}^{M} P(xi)P(Pa(x0)|xi) \quad \dots \dots \dots (4.5)$$

4.2 K2 METRIC

Several scores such as BD Metric and K2Metric can be used for estimating the Bayesian network. In this study, K2Metric is used for evaluating the network validity [11].

The K2 metric is a well-known evaluation measure or scoring function for learning Bayesian networks from data [12]. K2 algorithm is used to learn the topology of a Bayes Net. It is derived by assuming uniform prior distributions on the values of an attribute for each possible instantiation of its parent attributes. This algorithm heuristically searches for the most probable belief–network structure given a database of cases. K2 algorithm takes the out clusters of Ward algorithm as inputs. K2 treats the clusters as nodes. The output for each node is a printout of the parents of the node. With this parent and child node knowledge the network topology is built.

K2Metric is defined as follows,

$$K2 = \prod_{j=1}^{M} \frac{(L-1)!}{(Nj + L - 1)!} \prod_{k=1}^{L} Njk! \qquad \dots \dots \dots (4.6)$$

And

$$Nj = \sum_{k=1}^{L} Njk \qquad \dots \dots \dots (4.7)$$

where the parameter N, M and L denote total number of nodes and total numbers of states for x_0 and $Pa(x_0)$, respectively.

The notation N_{jk} denotes the number of samples of the state $x_0 = x^k$ on the condition

$Pa(x_0) = Y^j$

4.3 K2 ALGORITHM

Set a parent node class $Pa(x_0)$ as empty class Ø .

Estimate K2Metric S_{best} of the network composed of the node x_0 and the class $Pa(x_0)$.

1) Set i = 1.

2) Add x_i to $Pa(x_0)$.

3) Estimate K2Metric of the network composed of the node x_0 and the class $Pa(x_0)$.

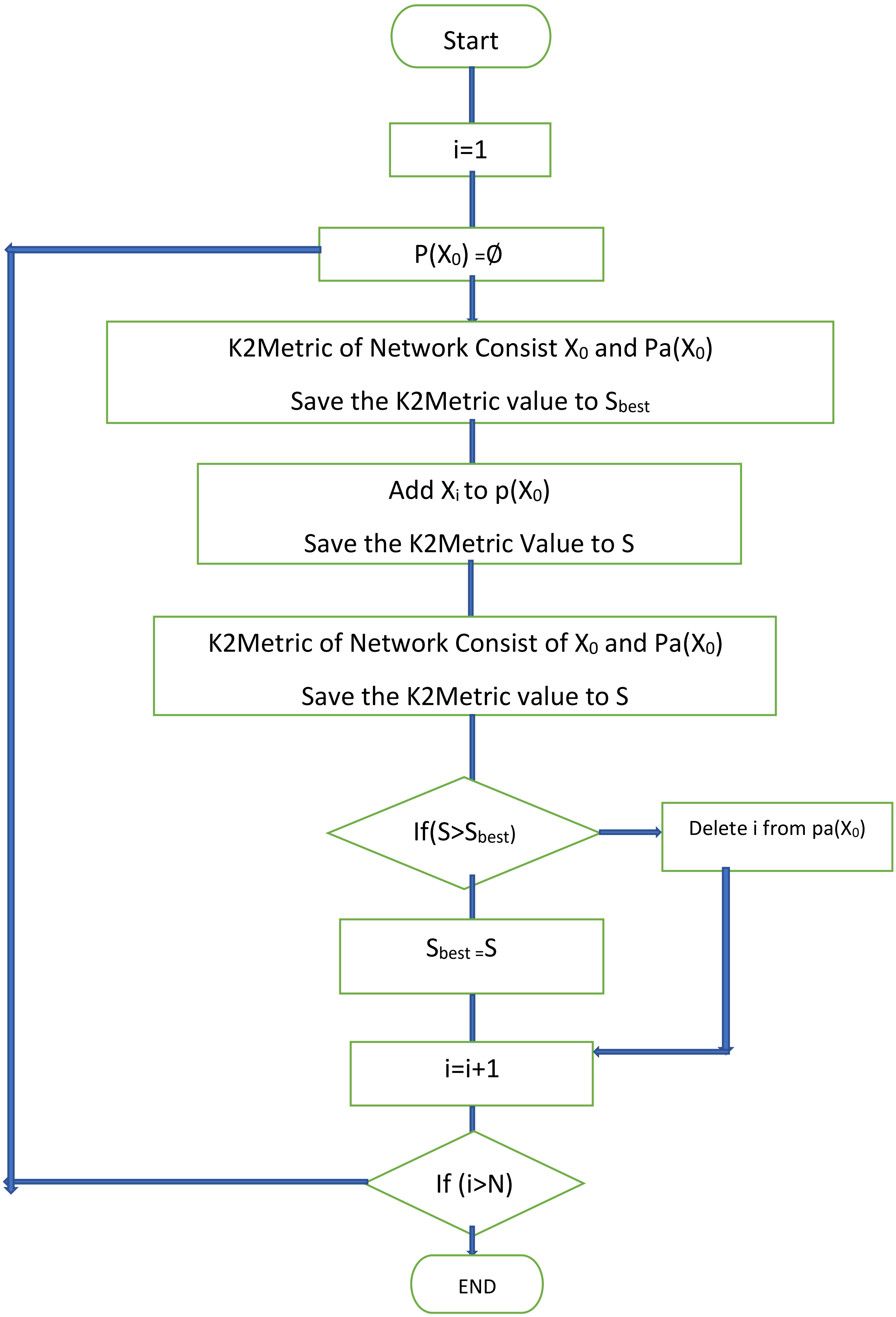
4) If $S \le S_{best}$, remove x_i from the class $Pa(x_0)$.

5) If $S > S_{best}$, set $S_{best} = S$.

6) Set i = i+1.

7) If $i \le N$, goto step 4.

8) Set as the Bayesian network B, the network composed of the node x_0 and the class $Pa(x_0)$.


Start
i=1
P(X0) =Ø
K2Metric of Network Consist X0 and Pa(X0)
Save the K2Metric value to Sbest
Add Xi to p(X0)
Save the K2Metric Value to S
K2Metric of Network Consist of X0 and Pa(X0)
Save the K2Metric value to S
If(S>Sbest)
Delete i from pa(X0)
Sbest =S
i=i+1
If (i>N)
END

4.4 PROBABILISTIC REASONING

The marginal distribution of a subset of a collection of random variables is the probability distribution of the variables contained in the subset. It gives the probabilities of various values of the variables in the subset without reference to the values of the other variables. This contrasts with a conditional distribution, which gives the probabilities contingent upon the values of the other variables.

When the evidence ***e*** of the random variable is given, the probability $P(x_i|e)$ is estimated by the marginalization algorithm[15] with the conditional probability table.

The marginalization algorithm gives the probability $P(x_i = X^l|e)$ as follows

$$\mathrm{P(xi\ =\ Xl}|e) = \frac{\sum_{j=1,j\neq i}^{N}\sum_{xj=X1}^{XL} P(x1, \dots, xi = xl, \dots, XN, e)}{\sum_{j=1,j\neq i}^{N}\sum_{xj=X1}^{XL} P(x1, \dots, XN, e)} \qquad \dots\ \dots\ \dots (4.8)$$

where the notation $\sum_{xj=X1}^{XL} P(xi)$ denotes the summation over all states X_1, X_2,.....,X_L of the random variable X_j.

4.5 PROCESS

The process of the prediction algorithm is summarized as follows.

Step 1:

Transform stock price return into the discrete values set by Ward method.

Step 2:

Determine the Bayesian network B from the discrete values set.

Step 3:

Predict the stock price by using the network B.

4.6 WARD METHOD CLUSTERING

In the first step we managed all our data by clustering. We used Ward Method as clustering algorithm as we can get discrete values from this algorithm and that is what we need to apply K2 algorithm.

We visualized the clusters and drew dendrogram to decide the distance between data points and numbers of clusters.

The ward algorithm has been explained in section 3.3.2

4.7 DISCRETIZATION OF STOCK PRICE RETURN

Discretization is the process of transferring continuous functions, models, variables, and equations into discrete counterparts. This process is usually carried out as a first step toward making them suitable for numerical evaluation and implementation on digital computers.

Stock price return r_t is defined as,

$$rt = (\ln Pt - \ln Pt - 1) * 100 \qquad \dots \dots \dots (4.9)$$

where the variable P_t denotes the closing stock price on the day t[14].

In the Bayesian network, random variables are specified at nodes. The use of Ward method, which is one of clustering algorithms, transforms the return into the set of discrete values. In this study, the Ward method is adopted as the clustering algorithm.

The set of the discrete values is defined as,

$$\{r^1, r^2, \dots\dots\dots, r^L\} \qquad \dots \dots \dots (4.10)$$

where the notation r^1 and L denote the discrete value and its total number, respectively.

The parameter L is determined so as to minimize AIC[16] which is defined as

$$AIC = \ln \sigma^2 + \frac{2(L+1)}{T} \qquad \dots \dots \dots (4.11)$$

where the notation σ^2 denotes the variance and the variable T is the total number of nodes. The variance σ^2 is defined as,

$$\sigma^2 = \frac{1}{T-N}\sum\nolimits_{t=N+1}^{T}(rt - rit)^2 \qquad \dots \dots \dots (4.12)$$

where N=10 and the parameter *rit* denotes the actual stock price return.

4.8 STOCK PRICE PREDICTION

Bayesian network is determined according to the K2 algorithm from the set of discrete values. The K2 algorithm needs the total order relationship of the random variables [13]. The total order relationship of the random variables is defined according to the time order

Figure 2: Total order of stock price return

Once the Bayesian network B is determined, the stock price return r_t is determined so as to maximize the conditional probability $P(r^l|B)$:

$$rt = \arg\max r^l \ \left(P(rl|B)\right) \qquad \dots \dots \dots (4.13)$$

CHAPTER 5

RESULTS ANALYSIS AND FUTURE PLAN

5.1 GOOGLE LLC

For the first example of stock average prediction we considered Google LLC. The network is determined from the daily stock price return from May 2016 to April 2017. Then the determined network is applied on new data to predict stock price of a certain date.

We determined the network from the stock price. Firstly, discrete values of the stock price clusters are calculated using Ward Method.

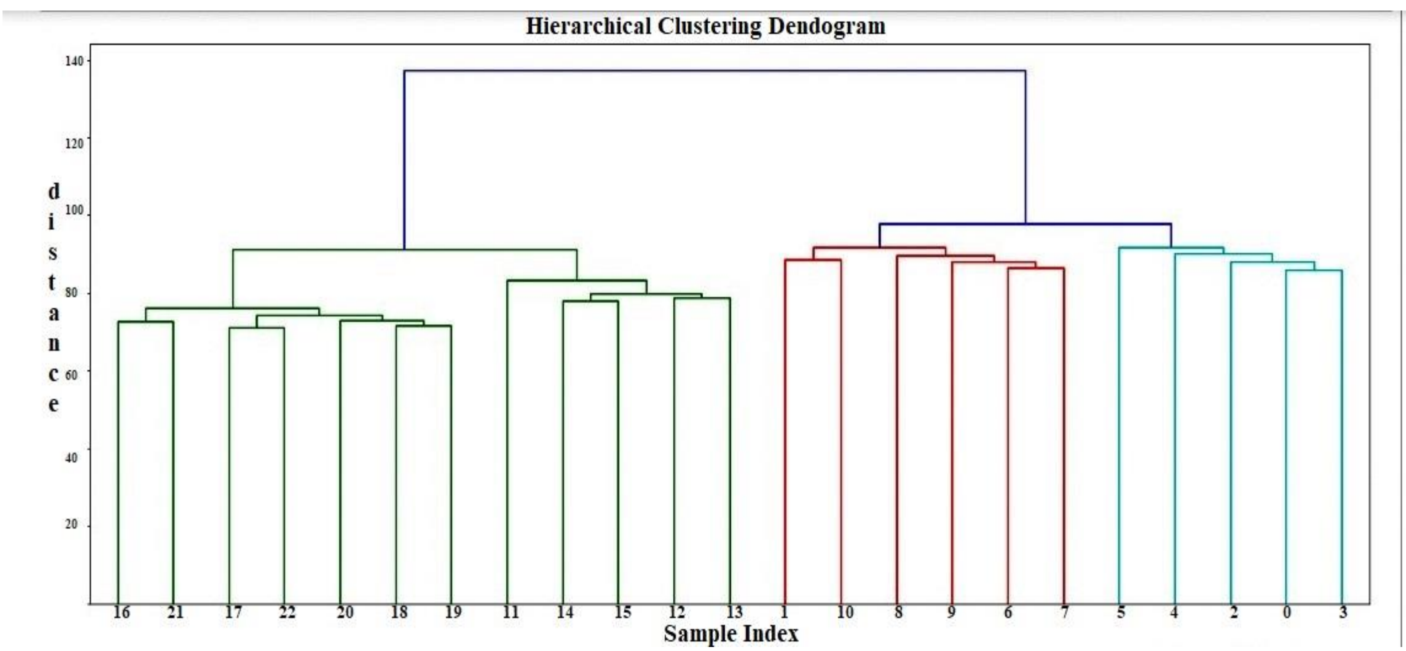


Figure 3: Dendrogram of Google LLC Stock Average

Then AIC of the network is calculated from the number of discrete values. The AIC is defined in section 4.7. The effect of the number of discrete values to the AIC of the network is compared in the table below.

Discrete Number L	AIC
2	1.62
3	1.46
4	1.27
5	1.10
6	1.06
7	1.17
8	1.35
9	1.60
10	1.65

Table 1: Discrete number versus AIC on Google LLC

We notice that the AIC is smallest at L = 6. The discrete values and the cluster parameters at L = 6 are calculated. The notation $c_l(r^l)$ means the cluster center, which is considered as the discrete value. From these values the minimum and the maximum values of the samples in the cluster are calculated. Thus, the network from stock price return has been determined. We notice that the return r_t depends on [20] the 3-days prior return r_{t-3}, 5-

days prior return r_{t-5}, 7-days prior return r_{t-7} and 10-days prior return r_{t-10}. Now we have the four nodes to apply K2 algorithm on.

We applied K2 algorithm and determined the parent and the child nodes. From this we determined the network topology. We applied the results in equation 4.13 to find stock price return of a single day. We compared this result with the actual price of that day to determine the percentage of error thus determining accuracy [21] of the model.

In case of Google stock average prediction, the average and the maximum errors of the present algorithm are 3% and 27%.

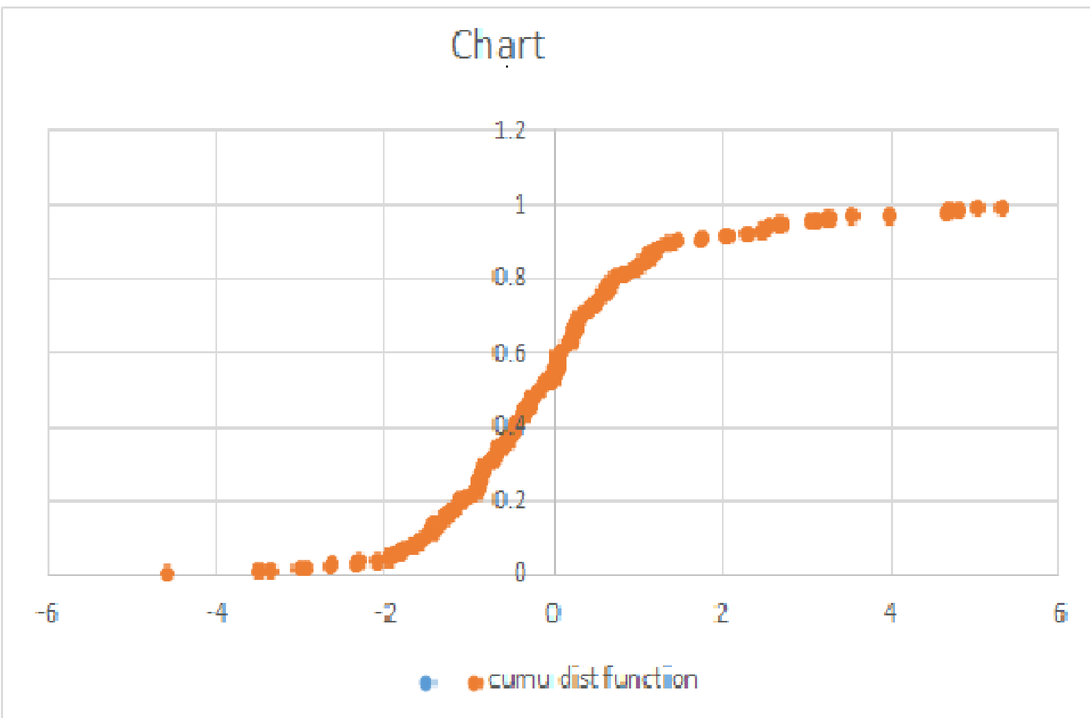


Figure 4: Google LLC Price Prediction Graph

5.2 PROCTER & GAMBLE CO.

For the first example of stock average prediction we considered Google LLC. The network is determined from the daily stock price return from May 2016 to April 2017. Then the determined network is applied on new data to predict stock price of a certain date.

We determined the network from the stock price. Firstly, discrete values of the stock price clusters are calculated using Ward Method.

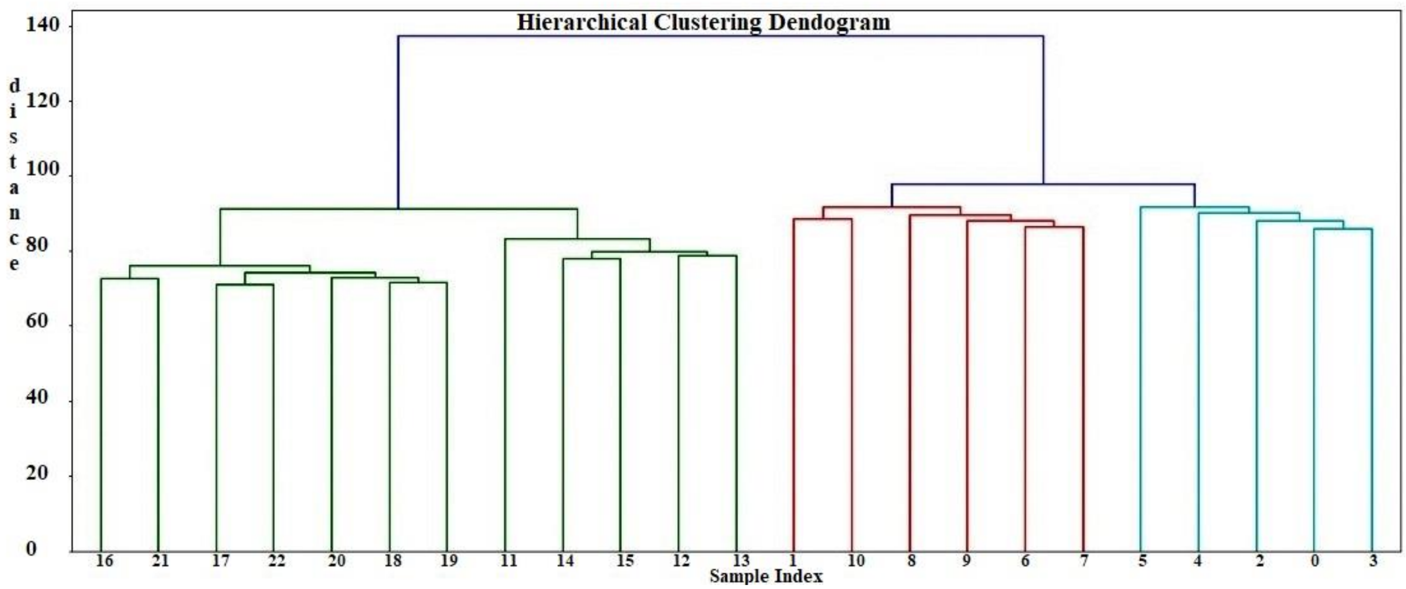


Figure 5: Dendrogram of Procter & Gamble Co. Stock Average

Then AIC of the network is calculated from the number of discrete values. The AIC is defined in section 4.7. The effect of the number of discrete values to the AIC of the network is compared in the table below.

Discrete Number L	AIC
2	1.55
3	1.41
4	1.13
5	1.11
6	1.27
7	1.43
8	1.55
9	1.58
10	1.74

Table 2: Discrete number versus AIC on Procter & Gamble Co.

We notice that the AIC is smallest at L = 3. The discrete values and the cluster parameters at L = 3 are calculated. The notation $c_l(r^l)$ means the cluster center, which is considered as the discrete value. From these values the minimum and the maximum values of the samples in the cluster are calculated. Thus, the network from stock price return has been determined. We notice that the return r_t depends on [20] the 4-days prior return r_{t-4}, 5-

days prior return r_{t-5} and 8-days prior return r_{t-8}. Now we have the four nodes to apply K2 algorithm on.

We applied K2 algorithm and determined the parent and the child nodes. From this we determined the network topology. We applied the results in equation 4.13 to find stock price return of a single day. We compared this result with the actual price of that day to determine the percentage of error thus determining accuracy of the model.

In case of Procter & Gamble stock price prediction, the average and the maximum errors of the present algorithm are 7% and 31%.

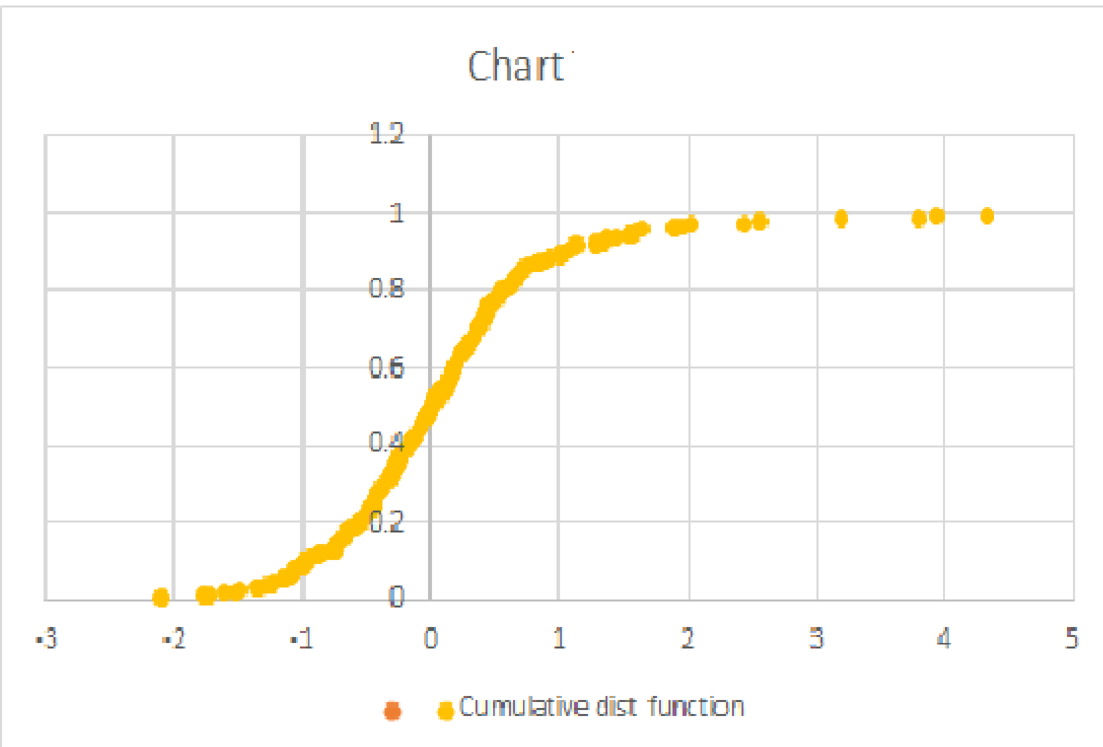


Figure 6: Procter & Gamble Price Prediction Graph

5.3 GENERAL MOTORS COMPANY

For the first example of stock average prediction we considered Google LLC. The network is determined from the daily stock price return from May 2016 to April 2017. Then the determined network is applied on new data to predict stock price of a certain date.

We determined the network from the stock price. Firstly, discrete values of the stock price clusters are calculated using Ward Method.

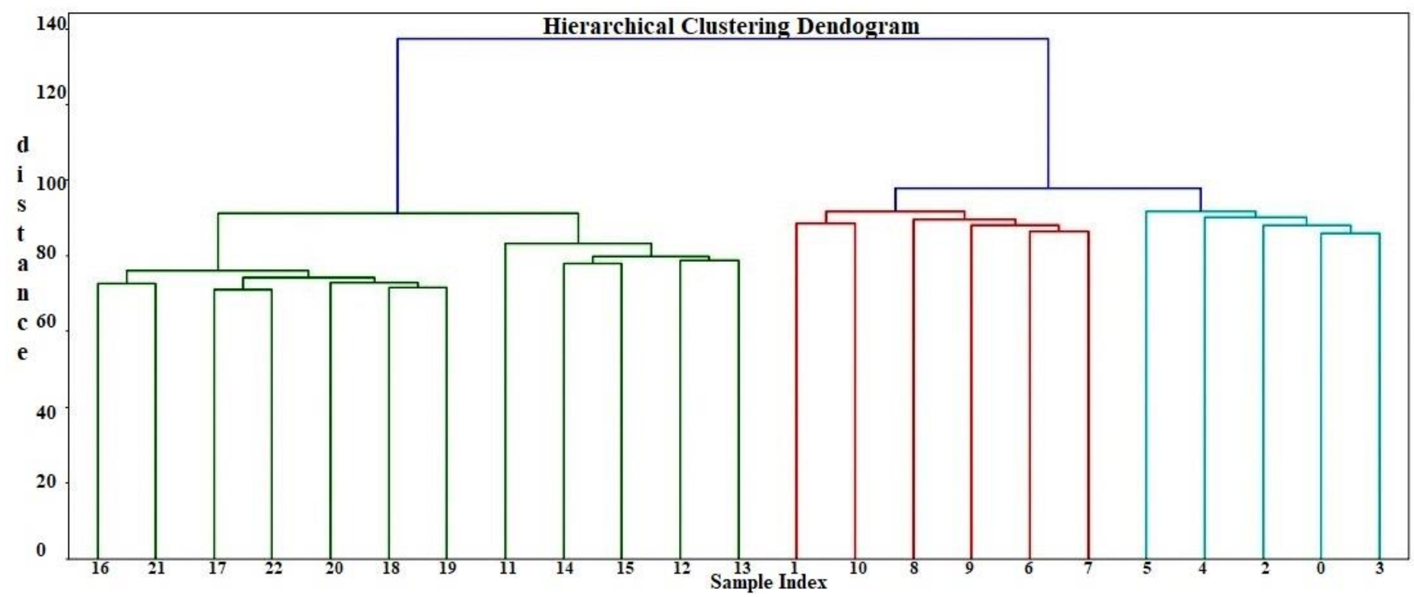


Figure 7: Dendrogram of General Motors Company Stock Average

Then AIC of the network is calculated from the number of discrete values. The AIC is defined in section 4.7. The effect of the number of discrete values to the AIC of the network is compared in the table below.

Discrete Number L	AIC
2	1.58
3	1.19
4	1.68
5	1.733
6	1.815
7	1.87
8	1.91
9	2.06
10	2.90

Table 3: Discrete number versus AIC on General Motors Company

We notice that the AIC is smallest at L = 5. The discrete values and the cluster parameters at L = 5 are calculated. The notation $c_l(r^l)$ means the cluster center, which is considered as the discrete value. From these values the minimum and the maximum values of the samples in the cluster are calculated. Thus, the network from stock price return has been determined. We notice that the return r_t depends on [20] the 3-days prior return r_{t-3}, 5-

days prior return r_{t-5}, 6-days prior return r_{t-6} , 8-days prior return r_{t-8} , and 10-days prior return r_{t-10} . Now we have the four nodes to apply K2 algorithm on.

We applied K2 algorithm and determined the parent and the child nodes. From this we determined the network topology. We applied the results in equation 4.13 to find stock price return of a single day. We compared this result with the actual price of that day to determine the percentage of error thus determining accuracy [21] of the model.

In case of General Motors stock average prediction, the average and the maximum errors of the present algorithm are 4% and 28%.

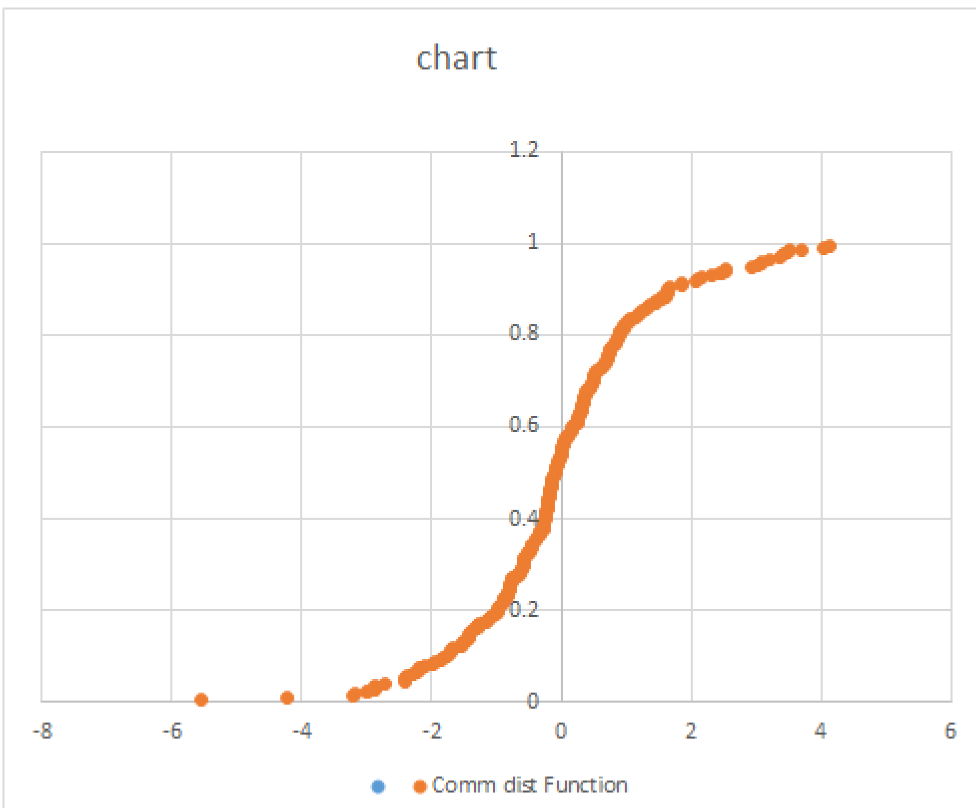


Figure 8: General Motors Price Prediction Graph

Results

Company	Min Error(%)	Max Error(%)	Average Error(%)
Google LLC	0.0236	3.4472	2.4667
Procter & Gamble	0.0458	6.0365	2.8214
General Motors Company	0.0652	7.2391	4.1636

5.4 Comparison

Model	Maximum Error(%)	Minimum Error(%)	Average Error(%)
AR	7.5091	0.0615	2.6657
MA	9.6259	0.0319	2.6859
ARMA	9.1204	0.0401	2.6739
ARCH	8.8839	0.0597	2.6992
CNN	8.8900	0.0236	2.2400
BN	7.8279	0.0451	3.1494

5.5 FUTURE PLAN

The stock price prediction algorithm using Bayesian network was presented in this study. Bayesian network can represent the stochastic dependency between random variables via an acyclic directed graph. In the previous study, the Bayesian network was determined from the time-series stock price data in order to predict the stock price. In this study, the previous algorithm is applied for predicting the stock price which was already observed. The prediction errors are estimated by the difference between the actual and predicted stock prices. Then, the new network is determined from both time-series stock price data and its prediction errors in order to predict the stock price.

Google, Procter & Gamble and General Motors stock price were considered as the numerical examples. In case of Google stock average prediction, the average and the maximum errors [21] of the present algorithm are 3% and 27%. In case of Procter & Gamble stock price prediction, the average and the maximum errors of the present algorithm are 7% and 31%. In case of General Motors stock average prediction, the average and the maximum errors of the present algorithm are 4% and 28%.

As we can see from the results that the present algorithm predicts the stock price almost accurate but it still is not accurate. To reduce error, we want to work more with this algorithm. We want to use ten years' stock data to determine the network next time. We also want to use more than three companies to reduce error.

We want to make the algorithm's overall performance better. So, we want to reduce CPU time.

In future we want to determine an algorithm which will use the network twice. First, the network will be determined from the daily stock price and then, it will be applied for predicting the daily stock price which was already observed. The prediction error will be evaluated from the daily stock price and its prediction. Second, the network will be determined again from both the daily stock price and the daily prediction error and then, it will be applied for the future stock price prediction. These are the ways we are planning to improve the network search algorithm in order to reduce error.

References:

[1] Hasbrouck, Sofianos & Sosebee. (1993). New York Stock Exchange Systems and Trading Procedures. Retrieved June 22, 2018.

[2] Xu. (n.d.). Stock Price Forecasting Using Information from Yahoo Finance and Google Trend. Retrieved June 21, 2018.

[3] Adebiyi, Ayo & Adewumi. (2014). Stock price prediction using the ARIMA model. Retrieved June 14, 2018.

[4] Marček. (2003). STOCK PRICE FORECASTING: AUTOREGRESSIVE MODELLING AND FUZZY NEURAL NETWORK. Retrieved June 13, 2018.

[5] Hegazy, Soliman & Salam. (2013). A Machine Learning Model for Stock Market Prediction. Retrieved June 22, 2018.

[6] Mittal & Goel. (2012). Stock Prediction Using Twitter Sentiment Analysis. Retrieved June 19, 2018.

[7] Wanjawa. (n.d.). ANN Model to Predict Stock Prices at Stock Exchange Markets. Retrieved June 28, 2018.

[8] Selvin & Gopalakrishnan. (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. Retrieved June 22, 2018.

[9] Kita, Zuo Harada, & Mizuno. (2012). Application of Bayesian Network to stock price prediction. Retrieved May 18, 2018.

[10] Mitchell. (2017). How the Stock Market Works. Retrieved July 10, 2018.

[11] Pearl & Russell. (2001). Handbook of Brain Theory and Neural Networks. Retrieved July 15, 2018.

[12] Ruiz. (n.d.). Illustration of the K2 Algorithm for Learning Bayes Net Structures. Retrieved May 12, 2018.

[13] COOPER, & HERSKOVITS. (1992). A Bayesian Method for the Induction of Probabilistic Networks from Data. Retrieved May 23, 2018.

[14] Dmouj. (2006). Stock price modelling: Theory and Practice. Retrieved July 01, 2018.

[15] HECKERMAN, GEIGER & CHICKERING. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. Retrieved June 3, 2018.

[16] Engle. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Retrieved May 15, 2018.

[17] Pearl. (2016). Probabilistic reasoning in intelligent systems: Networks of plausible inference. Retrieved July 8, 2018.

[18] Murtagh. (2011). Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. Retrieved May 22, 2018.

[19] ACKERMAN, FREER & ROY. (2010). ON THE COMPUTABILITY OF CONDITIONAL PROBABILITY. Retrieved July 7, 2018.

[20] Ariel. (1984). A MONTHLY EFFECT IN STOCK RETURNS. Retrieved May 25, 2018.

[21] Bonini, Zanetti & Bianchini. (2006). Target Price Accuracy in Equity Research. Retrieved June 14, 2018.

[22] Tiwari, Bharadwaj & Gupta. (2017). Stock Price Prediction Using Data Analytics. Retrieved July 7, 2018.

[23] Margaritis. (2003). Learning Bayesian Network Model Structure from Data. Retrieved May 3, 2018.

[24] Xing, Sun, Wang & Yu. (2013). The Analysis and Prediction of Stock Price. Retrieved July 1, 2018.

[25] Horný. (2014). Bayesian Networks. Retrieved May 13, 2018.

[26] Murtagh. (2011). Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. Retrieved July 3, 2018.

[27] Dunne. (2015). Stock Market Prediction. Retrieved June 4, 2018.