A Minor Project Final Report on

**Stock Market Prediction and Analysis using Machine Learning**

Submitted in Partial Fulfillment of the Requirements for

the degree of **Bachelors of IT Engineering**

under Pokhara University

Submitted By:

**Arjun Saud, 161505**

**Atul Kayastha, 161506**

**Sakar Bajimaya, 161528**

Under the Supervision of

**Er. Madan Kadariya**

Date: 26th Dec. 2019

**ACKNOWLEDGEMENT**

We would like to express our sincere thanks to Nepal College of Information Technology for providing us an opportunity to perform a minor project for the partial fulfillment of the requirements for the degree of B.E in Information Technology.

It’s our privilege to express our sense of gratitude and respect to our respected supervisor **Er. Madan Kadariya** for his unparalleled knowledge of moral fiber and judgment along with his know-how-based method which was an immense support for the completion of our project. We would also like to express our thanks to all the staff under Department of Information Technology for helping us in the successful completion of this project.

Lastly, we would like to thank our parents and beloved friends who helped and motivated us for the completion of the project.

**Group Members:**

Arjun Saud

Atul Kayastha

Sakar Bajimaya

# **ABSTRACT**

Stock Price forecasting 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 shares in the share market. Many methods like technical analysis, fundamental analysis, time series analysis and statistical analysis, etc. are all used to attempt to predict the price in the share market but none of these methods are proved as a consistently acceptable prediction tool.

In this project we attempt to implement an Artificial Neural Network approach to predict stock market prices. Artificial Neural networks are very effectively implemented in forecasting stock prices, returns, and stock modeling, and the most frequent methodology is the Backpropagation algorithm. This project is for Nepalese users as the prediction is done on the listed companies of Nepal Stock Exchange Ltd (NEPSE). We outline the design of the Neural Network model with its salient features and customizable parameters. We select a certain group of parameters with relatively significant impact on the share price of a company. With the help of statistical analysis, the relation between the selected factors and share price is formulated which can help in forecasting accurate results. Although, share market can never be predicted, due to its vague domain, this project aims at applying Artificial Neural Network in forecasting the stock prices.

**Keywords:** Neural Networks, Artificial Neural Network, NEPSE

Table of Contents

[**ABSTRACT**](#_r9dewn7p66w3) **3**

[**INTRODUCTION**](#_pa679zft0bz9) **6**

[PROBLEM STATEMENT](#_bdx42f8h6dt2) 8

[OBJECTIVES](#_n1vdmkptjhw) 8

[SIGNIFICANCE OF THE STUDY](#_as97i78ve1l2) 9

[**LITERATURE REVIEW**](#_s7vwa5sykbnu) **10**

[**METHODOLOGY**](#_r9a7ms553fz) **14**

[**REQUIREMENT ANALYSIS AND FEASIBILITY STUDY**](#_bpjc5jhr6bmy) **15**

[Feasibility Study](#_fefh0zbw2r6c) 15

[Requirement Analysis](#_6tmbt43gu3fw) 16

[Non-functional Analysis](#_oooxmi60jcmp) 16

[**SYSTEM DESIGN AND ARCHITECTURE**](#_w0x1ttpacy3y) **17**

[Use Case Diagram](#_joy8rqgok3h) 17

[Use Case Index](#_z155ea9nkco9) 17

[Use Case Description](#_lz7na576g899) 18

[SYSTEM FLOW DIAGRAM](#_peeou4slx4c7) 19

[SEQUENCE DIAGRAM](#_k5583kx2sq9z) 20

[IMPLEMENTATION STEPS](#_cez8ifikegui) 21

[Understanding the Objective](#_fdrpjpff2ktl) 21

[Data Collection](#_lzrg3l6k32ip) 21

[Data Pre-Processing: Data Wrangling](#_gnsl8dqk5to6) 21

[Data Processing: Data Training](#_8jxxsgq9ty5s) 21

[Forecasting Results](#_fzoftl85vbco) 21

[Plot Visualizations](#_el40aoo3u5tc) 22

[View & Analyze the results](#_l58w5aqxac3o) 22

# **INTRODUCTION**

Today we live and breathe data. Forecasting the stock exchange data is an important financial subject which involves an assumption that the fundamental information publicly available in the past has some predictive relationships to the future stock returns. Stock market forecasting contains uncovering the market trends, planning investment tactics, identifying the best time to purchase the stocks and which stocks to purchase. A stock exchange or equity business sector is a non-direct, non-parametric framework that is difficult to model with any sensible exactness. It is the mix of speculators who need to purchase or offer or hold a share at a specific time. Prediction will continue to be an exciting locale of research, making scientists in the analytics field always desiring to enhance the existing forecasting models. The motivation is that companies and individuals are empowered to make investment decisions to develop a viable system about their future endeavors.

Stock price prediction is a heated topic in prediction study of financial area. The stock market is essentially a non-linear, non-parametric system that is extremely hard to model with any reasonable accuracy. Investors have been trying to find a way to predict stock prices and to find the right stocks and right timing to buy or sell. Most of the techniques used in technical analysis are highly subjective in nature and have been shown not to be statistically valid. Recently, data mining techniques and artificial intelligence techniques like decision trees, rough set approach, and artificial neural networks have been applied to this area. Data mining refers to extracting or mining knowledge from large data stores or sets. Some of its functionalities are the discovery of concept or class descriptions, associations and correlations, classification, prediction, clustering, trend analysis, outlier and deviation analysis, and similarity analysis. Data classification can be done in many different methods; one of those methods is the classification by using Decision Tree. It is a graphical representation of all possible outcomes and the paths by which they may be reached.

The use of ANN in business environments has been increasing over the last few years. Excellent algorithm has been applied to predict stock price or index. Interest in neural networks has led to a considerable surge in research activities in the past decade. Artificial neural network models are based on the neural structure of the brain. The brain learns from experience and so do artificial neural networks. As a useful analytical tool, ANN is widely applied in analyzing the business data stored in database or data warehouse. Identifying customer behavior patterns and predicting stock price are emerging areas of neural network research and its application. Most of the companies have created new methods of evaluating financial data and investment decisions. Artificial Neural Networks are being used by most companies for improved forecasting capabilities in an analysis of the stock market. So, artificial neural network suits better than other models in predicting the stock market.

The idea of forecasting using neural network is to find an approximation of mapping between the input and output data through training. The trained neural network is then used to predict the values for the future This research work presents the use of artificial neural network as a forecasting tool for predicting the stock market price.

Mostly the approaches are in terms of fundamental approach and technical approach. For the long-term valuation fundamental approach is used. Every stock is having its own value that does not depend on the price of the stock that is known as Intrinsic value. The proposed model works through phases of data collection, feature processing, fuzzy logic mapping and stock value calculation. Fuzzy logic is used to map the quality as well as quantity valuation factors. The IF THEN rules are applied on the linguistic variable. The fuzzy model outcomes the stock value which is used to provide stock worth. The stock value is calculated by Dividend discount model. Accuracy of the system is 0.77. The results offer the backbone for the value and not the price.

Another method is Data Mining. Decision making process for business can be risky. Corporate decision makers have to make decisions to protect company’s benefit and lower the risk. In order to evaluate data mining approach on forecasting, a tool, called IFF, was developed for evaluating and simulating forecasts. Specifically, data mining techniques and simulation’s ability to predict, evaluate and validate Port Industry forecasts is tested. Accuracy is calculated with data mining methods. Finally, the probability of user's and simulation model confidentiality is calculated. The results of the research indicate that data mining approach on forecasting and Monte Carlo method have the capability to forecast on Port industry and, if properly analyzed, can give accurate results for forecasts. We study a multivariate Markov chain model for categorical data sequences to fuzzy time series.

The proposed method gets a higher average forecasting accuracy rate than some of the existing methods on temperature prediction.

## 

## **PROBLEM STATEMENT**

Stock market is very vast and difficult to understand. It is considered too uncertain to be predictable due to huge fluctuation of the market. Stock market prediction task is interesting as well as divides researchers and academics into two groups, those who believe that we can devise mechanisms to predict the market and those who believe that the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction. Investing in a good stock but at a bad time can have disastrous results, while investing in a stock at the right time can bear profits. Financial investors of today are facing this problem of trading as they do not properly understand as to which stocks to buy or which stocks to sell in order to get optimum results. So, the proposed project will reduce the problem with suitable accuracy faced in such a real time scenario.

## 

## **OBJECTIVES**

* The main objective of this study is to study about different methodology and get a stock market prediction tool to obtain more accurate stock prediction price and to evaluate them with some performance measures.
* This study can be used to reduce the error proportion in predicting the future stock prices. It increases the chances for the investors to predict the prices more accurately by reducing error percentage and thus gain benefits in share markets.
* After getting the idea about different methods of stock market forecasting techniques we can understand that by using which methods we will get more accurate results. Then, we will be able to reduce the amount of error by which investors can invest their valuable money in the stock market at the right time.

## **SIGNIFICANCE OF THE STUDY**

* The project will be useful for investors to invest in the stock market based on the various factors.
* The project target is to create a web application that analyses previous stock data of companies and implement these values in data mining algorithm to determine the value that particular stock will have in the near future with suitable accuracy. These predicted and analyzed data can be observed by individual to know the financial status of companies and their comparisons.
* The main feature of this project is to generate an approximate forecasting output and create a general idea of future values based on the previous data by generating a pattern.

# **LITERATURE REVIEW**

Over the past two decades many important changes have taken place in the environment of financial markets. The development of powerful communication and trading facilities has enlarged the scope of selection for investors. Forecasting stock return is an important financial subject that has attracted researcher’s attention for many years. It involves an assumption that fundamental information publicly available in the past has some predictive relationships to the future stock returns. In order to be able to extract such relationships from the available data, data mining techniques are new techniques that can be used to extract the knowledge from this data. For that reason, several researchers have focused on technical analysis and using advanced math and science. Extensive attention has been dedicated to the field of artificial intelligence and data mining techniques.

Some models have been proposed and implemented using the above mentioned techniques, the authors of Tsang, P.M., Kwok,P., Choy, S.O., Kwan, R., Ng, S.C., Mak, J., Tsang, J., Koong, K., and Wong, T. made an empirical study on building a stock buying/selling alert system using back propagation neural networks (BPNN), their NN was codenamed NN5. The system was trained and tested with past price data from Hong Kong and Shanghai Banking Corporation Holdings over the period from January 2004 to December 2005. The empirical results showed that the implemented system was able to predict short-term price movement directions with accuracy about 74%.

The research by Wu, M.C., Lin, S.Y., and Lin, C.H., used decision tree technique to build on the work of Lin. where Lin tried to modify the filter rule that is to buy when the stock price rises k% above its past local low and sell when it falls k% from its past local high. The proposed modification to the filter rule was by combining three decision variables associated with fundamental analysis. An empirical test, using the stocks of electronics companies in Taiwan, showed Lin’s method outperformed the filter rule. According to Wu, M.C., Lin, S.Y., and Lin, C.H.,, in Lin’s work, the criteria for clustering trading points involved only the past information; the future information was not considered at all. The research by Wu, M.C., Lin,S.Y., and Lin, C.H., aimed to improve the filter rule and Lin’s study by considering both the past and the future information in clustering the trading points. The researchers used the data of Taiwan stock market and that of NASDAQ to carry out empirical tests. Test results showed that the proposed method outperformed both Lin’s method and the filter rule in the two stock markets.

The model of Wang, J.L., Chan, S.H. (2006) “Stock market trading rule discovery using two-layer bias decision tree”, applied the concept of serial topology and designed a new decision system, namely the two layers bias decision tree, for stock price prediction.

The methodology developed by the authors differs from other studies in two respects;

1. First, to reduce the classification error, the decision model was modified into a bias decision model.
2. Second, a two-layer bias decision tree is used to improve purchasing accuracy.

The empirical results indicated that the presented decision model produced excellent purchasing accuracy, and it significantly outperformed than random purchase.

The authors Enke, D., Thawornwong, S. presented an approach that used data mining methods and neural networks for forecasting stock market returns. An attempt has been made in this study to investigate the predictive power of financial and economic variables by adopting the variable relevance analysis technique in machine learning for data mining. The authors examined the effectiveness of the neural network models used for level estimation and classification. The results showed that the trading strategies guided by the neural network classification models generate higher profits under the same risk exposure than those suggested by other strategies.

The research by Cao, Q., Leggio, K.B., and Schniederjans, M.J., was basically a comparison between the work of Fama and French’s model and the artificial neural networks in order to try to predict the stock prices in the Chinese market. The purpose of this study is to demonstrate the accuracy of ANN in predicting stock price movement for firms traded on the Shanghai Stock Exchange. In order to demonstrate the accuracy of ANN, the authors made a comparative analysis between Fama and French’s model and the predictive power of the univariate and multivariate neural network models. The results from this study indicated that artificial neural networks offer an opportunity for investors to improve their predictive power in selecting stocks, and more importantly, a simple univariate model appears to be more successful at predicting returns than a multivariate model.

Al-Haddad et al., presented a study that aimed to provide evidence of whether or not the corporate governance & performance indicators of the Jordanian industrial companies listed at the Amman Stock Exchange (ASE) are affected by variables that were proposed and to provide the important indicators of the relationship of corporate governance & firms’ performance that can be used by the Jordanian industrial firms to solve the agency problem. The study random sample consists of (44) Jordanian industrial firms. The study founds a positive direct relationship between corporate governance and corporate performance.

Hajizadeh et al. provided an overview of application of data mining techniques such as decision tree, neural network, association rules, and factor analysis and in stock markets. Prediction stock price or financial markets has been one of the biggest challenges to the AI community. Various technical, fundamental, and statistical indicators have been proposed and used with varying results.

Soni surveyed somnolescent literature in the domain of machine learning techniques and artificial intelligence used to predict stock market movements. Artificial Neural Networks (ANNs) are identified to be the dominant machine learning technique in stock market prediction area.

El-Baky et al.proposed a new approach for fast forecasting of stock market prices. The proposed approach uses new high-speed time delay neural networks (HSTDNNs). The authors used the MATLAB tool to simulate results to confirm the theoretical computations of the approach.

V. Vamitha, M. Jeyanthi, S. Rajaram and T. Revathi’s research about Multivariate Markov Chain also gave a new approach in the stock market prediction systems. Since 1993 researchers proposed many methods for forecasting enrollments, Temperature prediction, stock price etc. in time variant and time invariant first order, higher order, two factor and dual variables.

In this paper, we propose a model to temperature predictions from correlated categorical data sequence obtained from similar source. We study a multivariate Markov chain model for categorical data sequences to fuzzy time series. The proposed method gets higher average forecasting accuracy rate than some of the existing methods on temperature prediction.

Anass Nahil proposed a new method on stock market prediction which will help many investors to invest their money in the right time by which they will get more benefits in the near future. Their proposed method was about support vector machine (SVM). It is a popular tool in time series forecasting for the capital investment industry. This machine learning technique which is based on a discriminative classiﬁer algorithm, forecasts more accurately the ﬁnancial data. By examining the stock price of 5 Moroccan banks, the experiment shows that the SVM can perform better when we add the global evolution of the market to the independent variables.To express the global evolution of the market, three indices of the Casablanca Stock Exchange are used: MASI, MADEX and Banks Sector Index.

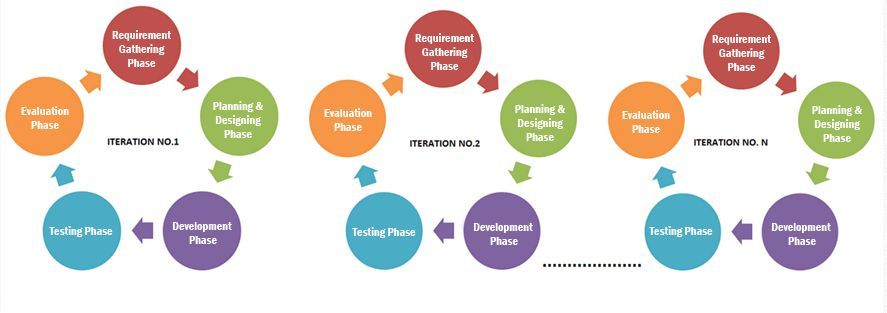
Narendra Pahuja, Abhishek Oturkar, Kailash Sharma, Jatin Shrivastava, Dimple Bohra’s ARIMA model made a huge change in stock market prediction. Over the years it is observed that stock market data is nonlinear, chaotic & dynamic.

This paper is going to present a predictive model for prices of the stocks with the help of ARIMA model. The stock data which is published by the Nepal Stock Exchange (NEPSE) has been used with the model developed for the prediction of stock price. From the results which are obtained, we come to the conclusion that for short-term prediction the ARIMA model has a great potential & also it shows competence with the already present methods for stock price prediction.

# **METHODOLOGY**

The proposed method for developing the system is based on agile model of SDLC. In the agile methodology after every development iteration, we are able to see the result and understand whether it is up to our expectations or not. This is one of the advantages of agile software development life cycle model. Extreme programming is one of the practical use of the agile model.

* Firstly, data is collected and sorted for relevancy from various sources.
* Secondly, analysis is carried out on the collected data by examining the current market direction, tracking the industry group and specific companies after which the data is represented and scored accordingly.
* At last, a Neural Network is designed by a suitable algorithm yielding best accuracy is chosen to predict the stock value.



1. Requirement Gathering Phase: Our requirements being the stock prices value of Nepalese companies of a long time span (4-5 years), we will be using Web Scraping, by building a web crawler for [www.nepalstock.com](http://www.nepalstock.com) . Building data frames and consecutive Data-sets for a limited number of companies.
2. Planning and Designing Phase: The scraped data from NEPSE will be stored in our local drive, which can be accessed automatically from our system as open CSV’s. A neural-network preferably based on suitableMachine Learning , Artificial Neural Network algorithm is researched and appropriate planning is done for data storing.
3. Development Phase: The retrieved datasets are then used to train our neural-network model. Various external open data sets are also used in this process. Datasets are separated and cleaned as well to remove missing data portions.
4. The main system is developed which displays the current stock prices of an individual/user selected company.
5. It displays the predicted value of closed prices for the next day.
6. It compares its output with the predicted values in the market for calculating errors and accuracy.
7. Testing Phase: The Neural Network model is tested with the warehoused datasets of the individual companies.

# **REQUIREMENT ANALYSIS AND FEASIBILITY STUDY**

## 

## **Feasibility Study**

Simply put, the Stock Market cannot be accurately predicted. The future, like any complex problem, has far too many variables to be predicted. The stock market is a place where buyers and sellers converge. When there are more buyers than sellers, the price increases. When there are more sellers than buyers, the price decreases. So, there is a factor which causes people to buy and sell. It has more to do with emotion than logic. Because emotion is unpredictable, stock market movements will be unpredictable. It’s futile to try to predict where markets are going. They are designed to be unpredictable.

There are some fundamental financial indicators by which a company’s stock value can be estimated. Some of the indicators and factors are: Price-to-Earning (P/E) Ratio, Price-to-Earning Growth (PEG) Ratio, Price-to-Sales (P/S) Ratio, Price/Cash Flow (P/CF) Ratio, Price-to-Book Value (P/BV) Ratio and Debt-to-Equity Ratio. Some of the parameters are available and accessible on the web but all of them aren’t. So, we are confined to use the variables that are available to us.

The proposed system will not always produce accurate results since it does not account for the human behaviors. Factors like the change in the company's leadership, internal matters, strikes, protests, natural disasters, change in the authority cannot be taken into account for relating it to the change in Stock market by the machine.

The objective of the system is to give an approximate idea of where the stock market might be headed. It does not give a long-term forecasting of stock value. There are way too many reasons to acknowledge for the long-term output of a current stock. Many things and parameters may affect it on the way due to which long term forecasting is just not feasible.

## **Requirement Analysis**

After the extensive analysis of the problems in the system, we are familiarized with the requirement that the current system needs. The requirement that the system needs is categorized into the functional and non-functional requirements. These requirements are listed below: Functional Requirements Functional requirements are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

* The system should be able to generate an approximate share price.
* The system should collect accurate data from the NEPSE website in a consistent manner.

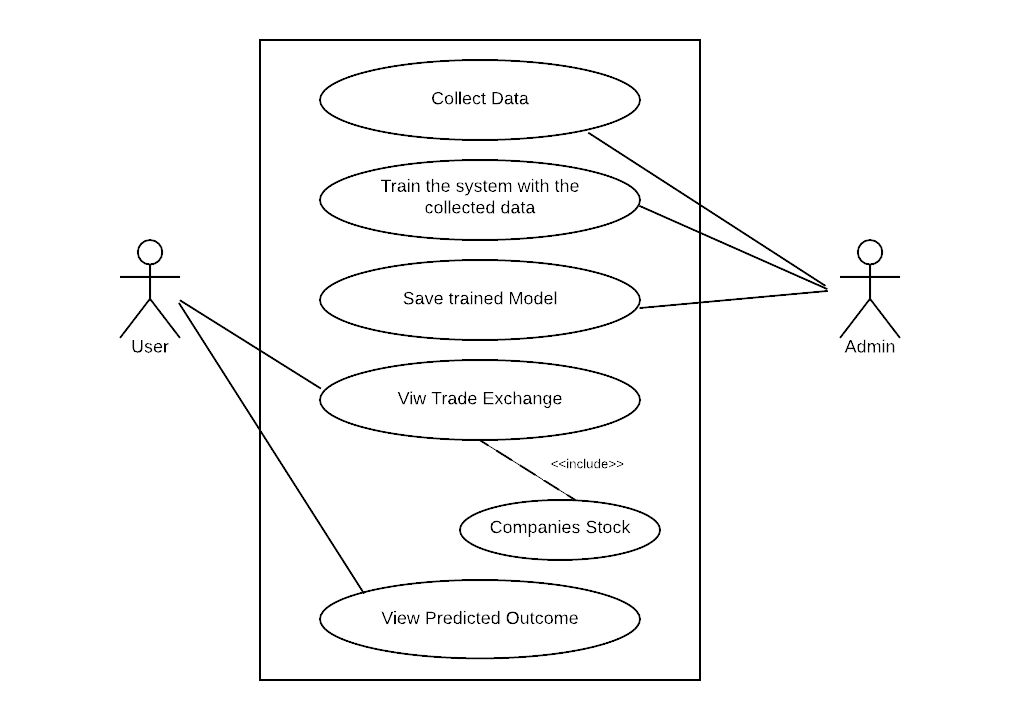
## **Non-functional Analysis**

Non-functional requirement is a description of features, characteristics and attributes of the system as well as any constraints that may limit the boundaries of the proposed system. The non-functional requirements are essentially based on the performance, information, economy, control and security efficiency and services. Based on these the non-functional requirements are as follows:

* The system should provide better accuracy.
* The system should have a simple interface for users to use.
* To perform efficiently in a short amount of time.

# **SYSTEM DESIGN AND ARCHITECTURE**

## **Use Case Diagram**



## **Use Case Index**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Use Case ID** | **Use Case Name** | **Primary Actor** | **Scope** | **Complexity** | **Priority** |
| 1. | Collect Data | admin | in | high | 1 |
| 2. | Train System with collected data | admin | in | high | 1 |
| 3. | Save Trained Model | admin | in | high | 1 |
| 4. | View Trade Exchange | user | in | medium | 2 |
| 5. | Companies Stock | user | in | medium | 2 |
| 6. | View Predicted outcomes | user | in | high | 1 |

### Use Case Description

**Use Case ID:1**

Use Case Name: Collect Data

Description: Every required data will be available in Nepal Stock Exchange. Admin will be able to collect the data for the system.

**Use Case ID:2**

Use Case Name: Train the system with the collected data

Description: Prediction result will be handled and generated by admin. The system will be built, through which the result of prediction and system performance will be analyzed.

**Use Case ID: 3**

Use Case Name: Save trained model

Description: The trained model is then saved for further processing. Besides, with the change of market and technology regular update of system is required. Beside there the predicted result of stock exchange and their actual price will be updated by admin on a regular basis.

**Use Case ID: 4**

Use Case Name: View Trade Exchange

Description: Company trading which is held at NEPSE can be viewed by user.

**Use Case ID: 5**

Use Case Name: Companies Stock

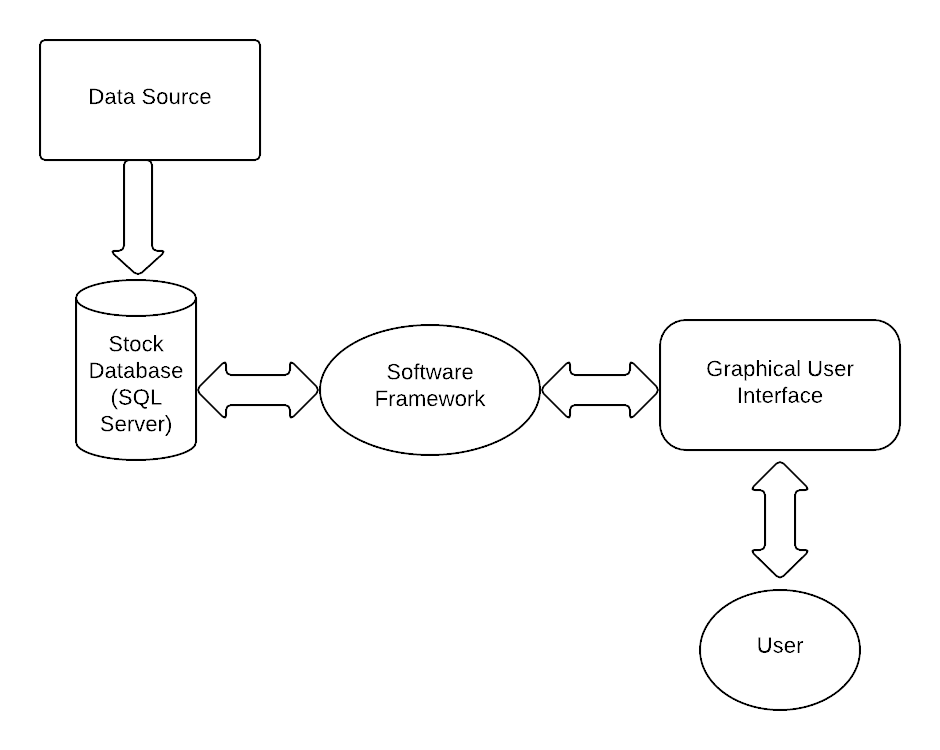
Description: It is an extended feature of view traded exchange. This includes the stock value of particular company.

**Use Case ID: 6**

Use Case Name: View Predicted Outcomes

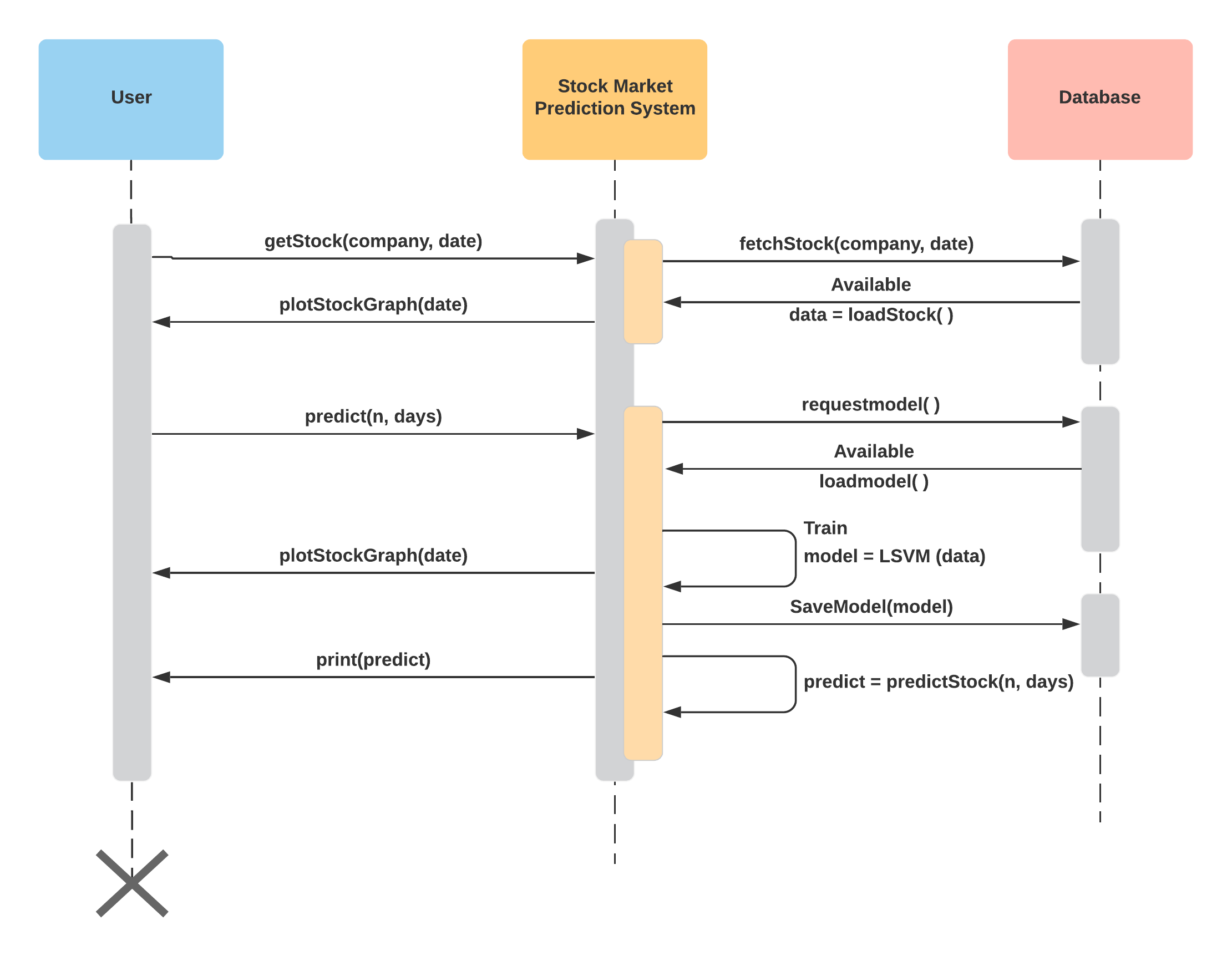
Description: This use case is the most important in the whole project. The key feature of this project is to predict the stock value of hydropower companies. Thus, this will be available in user interface and viewer can observe them.

## **SYSTEM FLOW DIAGRAM**



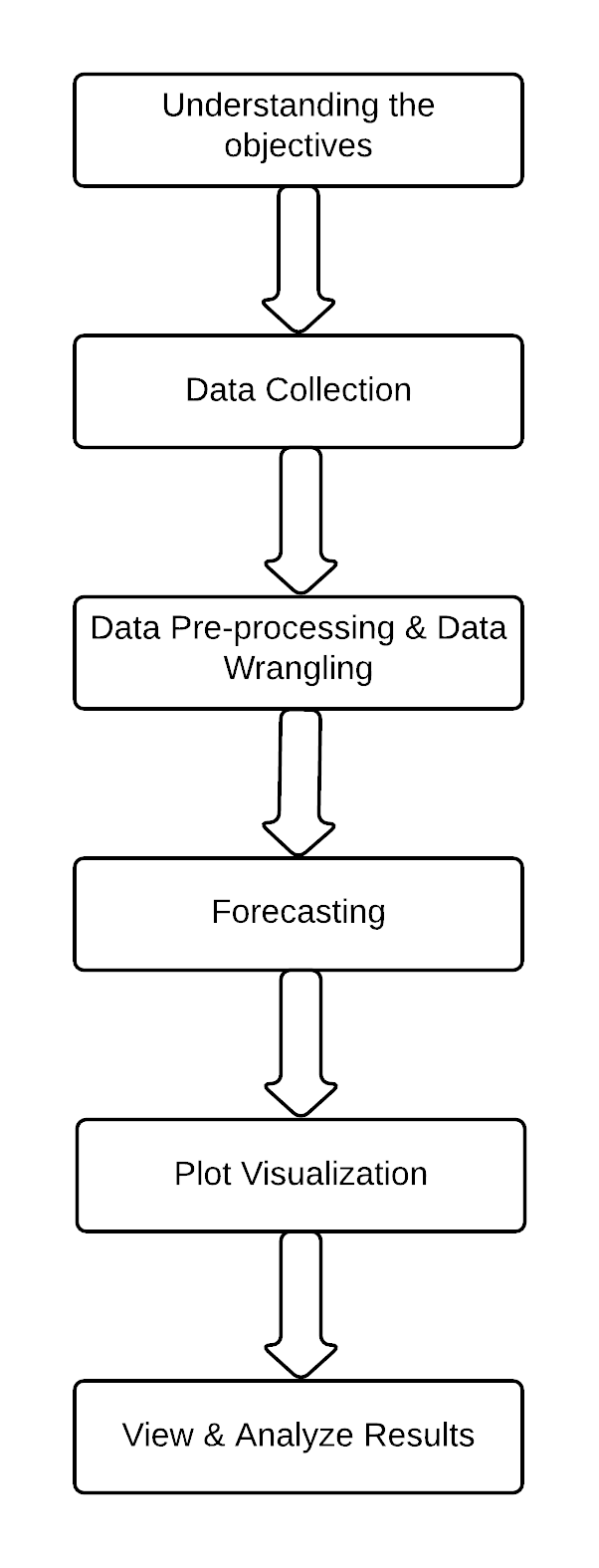
## 

## **SEQUENCE DIAGRAM**



## **IMPLEMENTATION STEPS**

System Architecture is a model that defines the behavior of a system in the conceptual model. The huge systems are decomposed into subordinate systems to provide a similar set of services. The beginning layout strategy of perceiving these sub-systems and building up a structure for sub-systems control and cooperation is called architecture design. As shown below, Fig. …. includes seven major steps to implement the system and each step is explained below:



### Understanding the Objective

The first step in developing a project is to understand the objective which involves an understanding of the intent and essentials of a system. This comprehension is used as a problem description and a preparatory system to accomplish the expectations. The objective of our project is neither to build a system that makes billions nor to waste billions too. But the objective is to develop a system that finds the direction of change of stock price indices based on the correlations between stock prices and help the investors in the stock market in taking a decision whether to buy/sell/hold a stock by providing the results in-terms of visualizations.

### Data Collection

Once the understanding of the objective is over, the next step is to collect the data. Data collection involves the understanding of initial observations of the data to identify the useful subsets from hypotheses of the hidden information. The data we use are the data obtained from Scraping from NEPSE’s website.

### Data Pre-Processing: Data Wrangling

The data pre-processing stage involves all the activities to prepare the final dataset from the preparatory raw information. The data preparation tasks can be performed several times as there is no specific order. These tasks include the selection of a record, table, attribute and cleaning of data for modeling tools.

### Data Processing: Data Training

In technical analysis investors use the auto regressive and moving average models to forecast the stock trends. Major steps involved here are **identification, parameter estimation and forecasting.** These steps are repeated until an appropriate model is identified for prediction.

### Forecasting Results

The process of making predictions of the future by relying upon the past and present data is known as forecasting. Various prediction techniques are used by the stock analysts to evaluate the future stock trends value. Prediction also offers a significant standard for organizations that have a long-term perception of actions. We use ‘forecast’ package for predicting the future stock trends based on the analysis of past trends. This ‘forecast’ package provides a number of forecasting functions for displaying the time series predictions along with exponentials and space models.

### Plot Visualizations

Data visualization is a graphical representation of the numerical data. After forecasting the stock market trends we visualize the results for short-term investment assistance in-terms offline charts, candlesticks charts, bar charts, and histograms.

### View & Analyze the results

Once after plotting the results in-terms of visualizations we can find out the correlations to get the short-term predictions. In the next section we provide some of the screen shots by which the investor can analyze and predict the future stock trends of a particular company at a specific time period. So, the investors in the stock market can use this as assistance to sell/buy/hold a share.

# **FORECASTING METHODOLOGIES**