STOCK MARKET PREDICTION USING MACHINE LEARING TECHNIQUES

Balamurugan*, Keerthana GR**, Narmatha A**,Pushparani R**
Assistant Professor*, Student**
Department of Computer Science and Engineering
Adhiyamaan College of Engineering
Hosur (TN), India

]

Abstract

To late quite a while securities exchange Examine may be challenging to settle on dependable predictions this paper proposes a approach towards prediction about securities exchange patterns utilizing machine Taking in model. Those mogul need created a number about stock analyses to what's to come prediction. Demonstrating and foreseeing about equity future price, dependent upon those current monetary majority of the data and news, will be from claiming gigantic utilization of the moguls. The new models might a chance to be produced In light of days gone by preparing datasets. An prominent Also generally utilized factual strategy for time arrangement determining may be the ARIMA model. Our primary theory might have been that Eventually Tom's perusing applying machine Taking in Furthermore preparation it on the previous data, it will be could reasonably be expected should foresee the shutting rate of the stock cost. These strategies would used to figure if the cost of a stock later on will be higher over its value for a provided for day, In light of recorded information same time giving a in-depth seeing of the models constantly utilized.

Keywords:Stockmarketprediction,Data Mining, ARIMA model, Auto regression

I.INTRODUCTION

Over the most recent couple of years forecast of financial exchange is picking up the consideration of speculators. Information mining is the procedure to distinguish the examples and build up relationship through information examination. Concentrate the information from the substantial informational index. The securities exchange development is effectively foresee the financial specialists and furthermore can return enough benefits out of market utilizing forecast. Stock value expectation is somewhat a dangerous activity. Gathering the information from the UCI site and it very well may be gotten to by any one. Via preparing the past dataset and to foresee the more precise outcome. Relapse is a factual estimation used to decide the relationship betweenone free factor under arrangement of other evolving variable. Relapse investigation is the nonexclusive term for a few stastical test for assessing the connection between interim dimension needy and free factors. We are thinking about the connection between one free factor and one ward variable utilizing basic direct relapse. The connection between one ward variable and more than one autonomous variable utilizing numerous relapses.

An Auto Regression and Integrated Moving Average (ARIMA) is a machine learning strategy. This

models hand-off vigorously on autocorrelation designs in the information. ARIMA technique determining is not the same as most strategies since it doesn't accept a specific example in the verifiable information of the arrangement .

II. OBJECTIVE

Our main aim of this project is to build a predictive model which will predict the market closing price ahead of time with better precision.

To predict the stock price movements with average accuracy of around 80%.

III. LITERATURE SURVEY

Stock Market Value Prediction Using Neural Networks:[1] In this paper, two kinds of neural networks, a feed forward multilayer Perceptron (MLP) and an Elman recurrent network, are used to predict a company's stock value based on its stock share value history. Both the networks are trained using backpropagation algorithm. The obtained results shows that amount of error using MLP neural network is less comparatively and MLP predicts close to the real one in comparison to the other methods.

Stock Price Prediction Using K-Nearest Neighbor (KNN) Algorithm:[2]In this paper, the author has applied k-nearest neighbor algorithm and non-linear regression approach in order to predict stock price. A prediction process for five listed companies on the Jordanian Stock Market was carried out, and is considered to be the first of its type implemented in Jordan as a case study using real data and market circumstances. Consequently, a robust model was constructed for the purpose set out. The data was extracted from five major listed companies on the Jordanian stock exchange, the sample data was used for training data set (about 200 records for each company). The author adopted a prediction algorithm tool of KNN with k=5 to perform such tests on the training data sets. According to the results, KNN algorithm was stable and robust with small error ratio, so the results were rational and reasonable. In addition, depending on the actual stock prices data; the prediction results were close to actual prices. Stock Value Prediction using Artificial Neural Networks.2013:[4] In this paper ,a stock price prediction model using multi-layer feed forward Artificial Neural Network (ANN) is used. In this model, the author has used back propagation algorithm containing one input layer, one hidden and one output layer. It is repeatedly processed and process continues until it results maximum accuracy As the closing price of any stock already covers other attributes of the company, the author have used historical stock prices (closing) for training the network. The model was generat Indian Stock Market Predictor System:[6] In this paper the author has combined both the statistical numeric data and sentiments of the stock on the internet to predict future prices in the stockMarket,there are two kinds of data to be extracted historical data about a firm and news articles regarding that company. Gathering data from internet is solely based on the (SOR) Subject of Reference (e.g.ICICI bank).Some web mining techniques (ex. crawler) are used to gather all web pages. The system uses a predictor based on Neural Network. The training of the Neural Network is done first. After training, the system is fed with historical stock prices and postings or text of the news articles about a particular firm as inputs. The neural network with 1 input layer, 2 hidden layers and 1 output layer have been used. The model is predicting for the day i using theprevious four days values for the company a and the fifth input is the Sentiment value of i-1 day. Our system is based on real time data thus it can be used anytime. The system takes historical data from yahoo finance as

the input and then uses the back-propagation algorithm to predict the future values of the stock using concept of artificial neural networks Our system provides features like comparing two or more stocks and helping the user to select best available stock from the selected once. System provides detail analysis regarding including the graph of a particular stock. The system also has a feature of sending alert messages to the user regarding the information about the stock. David Enke and Suraphan Thawornwong [1] employed machine learning techniques to assess the predictive relationship of versatile financial and economic variables by using neural network models for effective forecast of future values. Mayankkumar B Patel et al. used artificial neural network (ANN) to make prediction of stock price for the companies listed under National Stock Exchange (NSE). Dase R.K and Pawar D.D. used neural network to predict the stock rate as it possesses the ability to extract utile information from a large dataset. Halbert white used neural network and learning methods to model nonlinear regularities in asset price movement and reported his findings. Debashish Das et al. applied a combination of neural network and data mining to make a reliable prediction of stock market because neural network is capable to extract utile information from a large dataset and data mining has ability to forecast future trends. Sneha Soni did survey on different machine learning techniques that were used to make stock market prediction and identified Artificial Neural Network as a dominant technique to forecast stock market.

IV. EXISTING SYSTEM

In existing system, the author has applied k-nearest neighbor algorithm and non-linear regression approach in order to predict stock price. A prediction process for five listed companies on the Jordanian Stock Market was carried out, and is considered to be the first of its type implemented in Jordan as a case study using real data and market circumstances. Consequently, a robust model was constructed for the purpose set out. According to the results, KNN algorithm was stable and robust with small error ratio, so the results were rational and reasonable. In addition, depending on the actual stock prices data; the prediction results were close to actual prices.

Disadvantage:

• Concealed bits of knowledge are not Considered amid the forecast

• Low Accuracy

• Tedious

PROPOSED SYSTEM

Complex relationships between inputs and outputs may not always allow us to find patterns. Neural Networks is gaining much attention these days because of its capability of solving such problems. It has robust ability to discover relationship in the input data set without a priori assumption of the knowledge of relation between the input and the output data. It can be used to build model that identify unknown hidden patterns in data which can be further used for prediction purposes.

Advantage:

• Ready to foresee by considering concealed bits of knowledge.

• High Accuracy.

• Less tedious.

• User friendly

V.SYSTEM ARCHITECTURE

A System architecture or systems architecture is the conceptual model that defines the structure, behavior, and the more views of a system .An architecture description is a formal and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

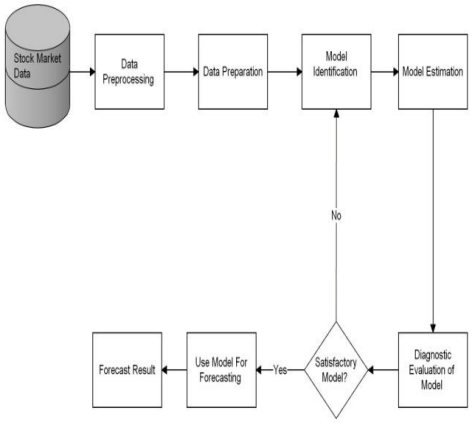


Figure 1.1 Stock Market Prediction

Step 1: Collecting Datas from UCI and stored in the Database

Step 2: To Remove the Unwanted Datas.

Step 3: Arranging the datas in a sequence manner.

Step 4: ARIMA model can be performed in this process.

Step5: To estimate the ARIMA model to process the next step.

Step 6: The diagnostic evaluation has been processed the evaluation of the model.

Step 7: Satisfactory model is used to fix the accurate value and it can be processed if the accuracy is correct it will process the step4 otherwise it can be process the step 8.

Step 8: Displays the accurate value of the predicted results in this model.

AUTOREGRESSIVE INTEGERATED MOVING AVERAGE MODEL

An ARIMA model is a class of statistical models for analyzing and forecasting time series data.It explicitly provides a simple powerful method for making skillful time series forecasting. The ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average and it is a generalization of the simpler Autoregressive Moving Average and adds the notion of integration.

- **AR:** Autoregression. A model that uses the dependent relationship between an observation and somenumber of lagged observations.
- **I** :Integrated. The use of differencing of raw observations (e.g. subtracting an observation from the previous time observation) in order to make the time series stationary.
- **MA**: Moving Average. A model that can be used to find the dependency between an observation and a residual error from a moving average model applied to lagged observations.
- Each of these components are explicitly specified in the model as a parameter. The notation are used by ARIMA model identity the values for(p,d,q) the values are applied byARIMA model being used.

The parameters of the ARIMA model are given below:

- **p**: How much number is include by ARIMA model, also called the lag order.
- **d**: How much time differenced in raw observations are also called the degree of differencing.
- **q**: The moving average window size, also called the order of moving average.

Linear regression model is constructed to specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary, i.e. to remove trend and seasonal structures that negatively affect the regression model.A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way ARIMA model can be configured function of an ARIMA model, and even a simple AR, I, or MA model.ARIMA model can adapted by process that generated the observations is an ARIMA

process. This process helps to motivate the need to confirm the assumptions of the model in the raw observations and in the residual errors of forecasts from the model. The next step let's take a look at how we can use the ARIMA model in Python. We will also start with loading an simple univariation of the time series.

III AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL

ARIMA stands for Autoregressive Integrated Moving Average. ARIMA is also known as Box Jenkins approach. Box Jenkins claimed that non stationary data can be made by differencing the series, Y_t. The general model for Y_t is written as,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} \ldots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots \theta_q \epsilon_{t-q}$$

Where Y_t is the different time series of values ϕ and θ are unknown parameters and ϵ are independent identically distributed error terms with zero mean. Y_t is expressed in terms of its past values and then the current and past values of their error terms.

The ARIMA model combines three basic methods:

- Auto Regression (AR):In auto-regression, the values of a given time series data are regressed on their own lagged a value, which is indicated by the "p" value in the ARIMA model.
- Differencing (I for Integrated)to this involves differencing the time series data to remove the trend and convert a non-stationary time series to a stationary one. This is represented by the "d" value in the ARIMA model. If d = 1 and it looks at the difference between two time series entries and if d = 2 it looks at the differences of the differences obtained at d=1 and so forth.
- Moving Average (MA):The moving average is the nature of the ARIMA model is represented by the "q" value which is the number of lagged values of the error term.

- Plot(dataset)while(graph is non stationary=true){smoothen graph to make it stationary}
- ACF/PACF(stationary graph)
- Estimate all possible model parameters
- Calculate AIC value of all model parameters.
- Plot (residual of model parameters)
    - If (residual graph with no lag){forecast data set with those parameters}
    - Else choose other estimated model parameters and repeat step 3.The following model is called Autoregressive Integrated Moving Average or ARIMA (p,d,q) of Y_t. We will follow the steps that can be enumerated below to build our model.

Step1: TESTING AND ENSURING STATIONARITY

The model of a time series with the Box Jenkins approach the series has to be stationary. A stationary time series defines the time series without trend one having a constant mean and variance over time which makes it easy for predicting values.

3.1 TESTING FOR STATIONARITY – We test for stationarity using the Augmented Dickey-Fuller unit root test. The p-value resulting from the ADF test has to be less than 0.05 or 5% for a time series to be stationary. If the p-value is greater than 0.05 or 5%, you conclude that the time series has a unit root which means that it is a non-stationary process.

3.2 DIFFRENCING: To convert a non-stationary process to a stationary process, we apply the differencing method. Differencing time series means finding the differences between consecutive values of a time series data. The different values form a new time series dataset which can be tested to uncover new correlations or other interesting statistical properties. We can apply the different methods consecutively more than once giving rise to the "first differences", "second order differences" etc. We can apply the appropriate differencing order (d) to make a time series stationary before we can proceed to the next step.

Step 2: IDENTIFICATION OF P AND Q

In this step we identifying the appropriate order of Autoregressive (AR) and Moving average (MA) processes by using the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF). The "Starting out with Time Series" for an explanation of ACF and PACF functions.

IDENTIFYING THE P ORDER OF AR MODEL

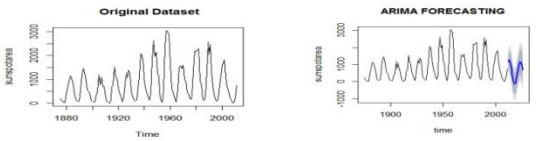
AR models that ACF will dampen exponentially and the PACF will be used to identify the order (p) of the AR model. If we have one of the significant spike at lag 1 on the PACF then we have an AR model of the order 1 i.e. AR (1). If we have significant spikes at lag using 1, 2, and 3 on the PACF and then we have an AR model of the order 3, i.e. AR (3).

IDENTIFYING THE Q ORDER OF MA MODEL

The MA models the PACF will dampen exponentially and the ACF plot will be used to identify the order of the MA process. If we have one significant spike at lag 1 on the ACF then we have to process an MA model of the order 1 i.e. MA (1). If we have significant spikes at lag 1, 2, and 3 on the ACF and we have an MA model of the order 3 i.e.

Step 3: ESTIMATION AND FORECASTING

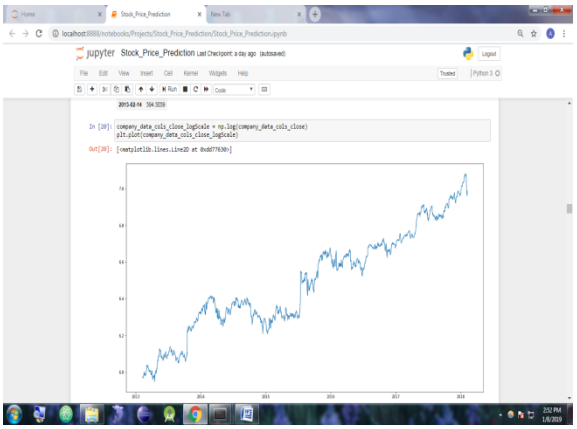
Once we have determined the parameters (p,d,q) we estimate the accuracy of the ARIMA model on a training data set and then using the fitted model to forecast the values of the test data set for forecasting function. At the end we will cross-check whether our forecasted values are in line with the actual values.



RESULT AND CONCLUSION

Predicting the direction of movements of the stock market index is important for the development of effective market trading strategies. It usually affects a financial trader's decision to buy or sell an instrument. Successful prediction of stock prices may promise attractive benefits for investors. These tasks are highly complicated and very difficult. We can use the principle of artificial neural networks to correctly predict the future value of a particular stock. This is done in steps; first we collect the past data of a stock. Then we create a neural network & train the network on the basis of the past data. Generally 80-90% of the data is taken to train the model. Then we use our model to predict the value of the stock & compare the value with the pre obtained data set.

OUTCOMES



REFERENCES

1. L.Y. Wei " The hybrid model based on the ANFIS and adaptive expectation genetic algorithm for forecasting" *Economic Model*, vol. 33, pp. 893-899, 2013.
2. J.J. Wang, JZ. Wang, Z.G. Zhang, S.P. Guo "Stock market forecasting based on the hybrid models", *Omega*, vol. 40, pp. 758-766, 2012.
3. C. Javier, E. Rosario, J.N. Francisco, J.C. Antonio, "ARIMA Model to Predict the Next Electricity Price", *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1014-1020, 2003.
4. N. Rangan, N. Titida, "ARIMA Model for Forecasting Oil Palm Price", *Proceedings of the 2nd IMT-GT Regional Conference on Mathematics Statistics and Applications*, 2006.
5. M. Khasel, M. Bijari, G.A.R Ardali, *Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and*, pp. 956-967, 2009.
6. C. Lee, C. Ho, "Short term load forecasting using lifting scheme and ARIMA model" Expert System with Applications of vol. 38, no. 5, pp. 5902-5911, 2011.
7. A. Meyler, G. Kenny, T. Quinn, "Forecasting Irish Inflation using ARIMA Models", *Central Bank of Ireland Research Department Technical Paper*, 1998.
8. JJ. Wang, JZ. Wang, Z.G. Zhang, S.P. Guo "Stock index forecasting based on the hybrid model", *Omega, vol. 40*, pp. 758-766, 2012.
9. L.C. Kyungjoo, Y. Sehwan, J. John, "Neural Network Model vs *SARIMA Model In Forecasting Korean Stock Price Index (KOSPI) Issues in Information System*, vol. 8, no. 2, pp. 372-378, 2007.