Quest Journals
Journal of Research in Applied Mathematics
Volume 8 ~ Issue 10 (2022) pp: 05-16
ISSN(Online) : 2394-0743 ISSN (Print): 2394-0735
www.questjournals.org

Research Paper

Evaluation of forecasting methods for Stock Market Data Stream

Ankitha K, Sukhesh K N, Vinitha Serrao, Satyanarayana
Department of Statistics Mangalore University, Mangalagangothri, India.

Abstract
In recent years, knowledge discovery in Data Stream plays an important role in many real-time applications. Data Streams are real-time, continuous, ever-changing with time, massive and unlimited. The construction of a suitable predictive model to handle the change in data according to time possess a great challenge for researchers. One of the applications of Data Stream is stock market price. In a global market, financial return on investments and movement of market indicators are filled with uncertainties and high volatility. Computation of Value at Risk measure the amount of potential loss that could happen in an investment over a specified time period and helps investors to choose stock with minimum risk. Stock market prediction forecasts the future value of the company stock. Future stock price prediction leads to significant profit. This research considers 5 stocks of BSE stock market (India) to identify a set of best portfolios that an investor can maintain for risk reduction and high profitability. This study uses support vector regression (SVR) to predict stock prices and prediction performance is compared with the traditional ARIMA-GARCH and ARIMA-ANN Hybrid models. The results suggest that SVR has better predictive power (comparable with ARIMA-ANN Hybrid) compared to all other models.

Keywords: *Data Stream, Stock price, Value at risk, SVR, Hybrid model.*

***R**eceived 22 Sep., 2022; **R**evised 03 Oct., 2022; **A**ccepted 05 Oct., 2022 © The author(s) 2022.*
P**ublished with open access at **www.questjournals.org

I. Introduction

Data Stream is data that is observed continuously generated by various sources. Such type of data should be processed incrementally by using streamprocessing techniques without having access to all of the data. The construction of a suitable predictive model to handle the change in data according to time possess great challenges for researchers.There are two evolution in data
1)Concept drifting: Concept drifting happens whenever class labels changes due to changes in time.
2)Conceptevolution: Concept evolution occurs when one or more new class labels emerge on class label set.

Concept drift mayinclude in the data that means the properties of the stream may change over time. Data streams are huge in volume, endless, infinite and agile information.Nowadays, the data streamis widely used in telecommunications, financial securities, retail trade, banking, weather forecasting and other fields. The stock price data are real-time, time dependent, continuous, massive and unlimited. Continuous variation in the stock market lays the user into the situation of confusion.
There are some reasons which separate data streams from traditional data mining:

- Data elements in the stream arrive on-line.
- The system cannot control the order in which data elements arrive.
- Data streams size is potentially boundless.

The stock market is a place where the stocks, shares and securities of various companies are traded and it is also known as the equity market.In the financial term stock market are the ordered market in which stockbrokers and traders can buy and sell shares, bonds and stocks.The stock exchange is controlled by the Commission to ensure that everyone follows the trading rules and regulations of various companies which are registered in the exchanges. These companies reveal financial information. They also reveal that whether the company is in a profitable position or not. According to the financial position of the company, the investors decide whether they invest in a particular company or not.

Benefits of the Stock Trading:

- The Stock price will vary over a period of time. Shareholders can sell their shares at a profit when the stock price goes up.
- Companies pay dividends to their shareholders(fixed dividend to preferred shares), which is a source of income for investors.
- Shareholders have some measure of control over the company through voting.
- Compared to all other investments (real estate, art or jewellery) shares can be bought and sold more easily.
- Tax rate for dividend income and capital gains are less.

The ability to forecast the future based on past data is an important leverage to push the organization forward. Time Series forecasting, the forecast of a time-ordered variable, playsan important role under this scenario, where the goal is to predict the behaviour of complex systems by looking at past data.Financial time series analysisgives a comprehensive and systematic introduction to financial econometric models and their application to modelling and prediction. Financial markets play a vital role in any country's economy. Monetary policies are generally based on stock exchange indices, foreign exchange rates, inflation rates, interest rates, etc. Financial time series contain an element of uncertainty known as volatility.

II. Review of Literature:

According to Gama (2010) main reasons which separate data stream from traditional data mining are: a) Data streams size ispotentially boundless b) data stream arrive on-line c) limitations in memory space d) the system cannot control or determine how data stream elements arrive. Taiwo andDaramola(2019) provided a systematic review of big data streaming analytics to understand tools and technologies to analyse and solve key issues in big data stream analysis.

Rockafellar and Stanislav (1999) introduced a new approach to optimizing a portfolio to reduce the risk of high losses. Value-at-Risk (VaR) has a role in the approach, but the emphasis is on Conditional Value-At-risk (CVaR) which is also known as Mean Excess loss. This is a suitable technique for investment companies, brokerage firms, mutual funds, and any business that evaluates risks.Shweta Tiwari and Alka Gulati (2011) studied Prediction of the Stock Market from Stream Data Time Series. They proposed a treebased data mining algorithm that takes market's behaviour and interest as input and gives a time-series pattern in the dynamic stock market. The results shows that decision tree and neural networks for classification have equivalent properties.Aparna et al (2016) applied prediction models for the Indian stock market to predict the stock market trend. Models were built for both dailyand monthly predictions. The results shows that the decision boosted tree performs better than the support vector machine. Ghani and Rahim (2019) conducted a case study on Malaysia Natural Rubber Price Modelling and forecasting using ARMA-GARCH. Predicting Stock Price direction using Support Vector Machines is carried out by Saahil Madge (2015). They used closing price of 34 technology stocks and calculated price volatility for individual stocks along with momentum. This model predicts whether the stock price in the future is lower or higher than it is on the current day. Nadeem Akhtar (2011) in their study "Statistical Data Analysis of Continuous Streams using Stream DSMS" used the Data Stream Management System tool STREAM to model and analyse Road Traffic analysis and Habitat Monitoring analysis. This result shows that STREAM is more interactive and graphically shows the relation model of a project. STREAM model is comparatively more robust than other models. Sunil et al (2019) applied the ARIMA-ANN Hybrid model to improve the accuracy of forecasting Jasmine prices. They compared the accuracy of the hybrid model with ELM, MLP, SARIMA and NNETAR(Neural Network Auto Regressive model) and concluded thatHybrid model improves the accuracy of forecasting.

III. Methodology

3.1 Value at risk(VaR)

Value at risk is mainly concerned with market risk, but the concept is also applicable to other types of risk such as credit risk, operational risk. From the viewpoint of a financial institution, VaR can be defined as the maximal loss of a financial position during a given time period for a given probability. In this view, one treats VaR as a measure of loss associated with a rare (or extraordinary) event under normal market conditions. Alternatively, from the viewpoint of a regulatory committee, VaR can be defined as the minimal loss under extraordinary market circumstances. Both definitions will lead to the same VaR measure, even though the concepts appear to be different. It can measure risk and is an important consideration when firms make tradingdecisions.

3.1.1 Parametric approach

The parametric method VaR is popular and simplebecause it is based onmean and standard deviation of the portfolio. The assumption of Parametric VaR is that the returns from their portfolios are normally

distributed. This helps the manager to use the calculated standard deviation to compute standard normal z score to determine risk position. This method is suitable for risk measurement problems where the distributions are known and reliable. This method is unreliable when the sample size is too small.

3.1.2 Non-Parametric approach

Historical VaR is a better methodology to use when distribution of return series cannot be determined. This calculation is much easier than Parametric VaR.This method works based onthe ranking ofpast historical returns from lowest to highest(at a specific confidence level). For example, if 100 past returns are available then 95% confidence VaR is computed by looking at 5th data point in the ranked series. This means that 95% of the time VaR will notbe worse than the computed amount.

3.1.3 Conditional value at risk (CVaR)

CVaR is calculated by taking a weighted average between the VaR and losses exceeding the value at risk. Conditional Value at Risk is the extension of VaR. The VaR model allows the managers to limit the likelihood of incurring losses caused by certain types of risk. In VaR model,tail end of the distribution of loss is not typically assessed. If losses occur, the amount of the losses will be substantial in value. Suppose we want 5% CVaR, first arrange these return values in increasing order. Then calculate value N*(5/100)=r (say).Then find rth smallest return value and find the average return values of the first r smallest returns. Let average of first r smallest returns are "m" and T be the threshold value. Suppose we invested 'S' amounts, then CVaR is given by

$$\text{CVaR}\ (0.05, T) = -m*S$$

3.2 TIME SERIES ANALYSIS

A Time Series (TS) is a sequence of observations which are equally spaced in discrete time intervals. In this analysis,different statistical methods are used to extract the required information from the data. A basic assumption in time series analysis is that some aspects of past patterns will remain same in the future. Start by plotting the time series data and looking for non-stationary components such as trend, seasonality etc. Eliminate these components using various methods, in order to get stationary data. After the analysis of the observed time series, predict the required future values.

3.2.1 TESTING FOR THE PRESENCE OF TREND COMPONENT

- **Mann-Kendall trend test:**

The Mann-Kendall trend test is a nonparametric test used to identify a trend in a series, even if a seasonal component exists. The hypothesis for this test is as follows:

H_0: There is no trend in the data.

H_1: There is monotonic trend in the data.

The Mann-Kendall test is based on the calculation of Kendall's tau measure of association between two samples, which itself is based on the ranks of the samples. Here the main assumption is that observations are independent which means that observations are not serially correlated over time.

3.2.2 TIME SERIES MODEL

Autoregressive Moving Average Process (ARMA) model:

Definition: Let $\{X_t, t\in I\}$ be a time series process, $\{\varepsilon_t\ ,\ t\in I\}$ be a white noise process with mean zero and variance σ^2 and I be the Index Set.Then $\{X_t\}$ is said to be follow$X_t \sim ARMA\ (p, q)$, if it has the following representation

$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_p X_{t-p} + \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \cdots - \alpha_q \varepsilon_{t-q}$$

Autoregressive Integrated Moving Average (ARIMA) Model:

This model is one of the most popular and frequently used stochastic time series model. This model assumes that time series is linear and follows a particular known statistical distribution, such as the normal distribution. In this model a non-stationary time series is made stationary by applying finite differencing of the data points.

Let $\{X_t, t \in I\}$ denote a non-stationary time series, non-stationary due to trend component. Let $\{\varepsilon_t, t=\pm1,\pm2, ...\}$ be a sequence of white noise. Then $\{X_t, t\in I\}$ is said to follow Autoregressive Integrated Moving Average process if it is has the following representation

$$\emptyset(B)(1-B)^d X_t = \theta(B)\varepsilon_t$$

$\emptyset(B)=1-\beta_1 B-\ldots\ldots\ldots-\beta_p B^p$

$$\theta(B)=1-\alpha_1 B-\ldots\ldots\ldots-\alpha_q B^q$$

Where $\alpha_1,\alpha_2,\ldots\ldots,\alpha_q$ are MA parameters and $\beta_1,\beta_2,\ldots\ldots,\beta_p$ are AR parameters. Where d is the order of difference required to make given time series to stationary time series. This model is known as Box-Jenkins model.

3.3 Financial time series model

The properties of stochastic volatilities can be studied by analysing the associated log return series, which we refer to as financial time series. The volatilities are measured in terms of conditional variances of such series. The basic idea behind the volatility study is that the log-returns series is either uncorrelated or with minor order serial correlations, and it is a dependent series. To study the dynamic structure of stochastic volatilities, several conditional heteroscedastic models are introduced.

3.3.1 ARCH and GARCH models

These models are considered as conditional heteroscedastic models. Volatility is an important factor in options trading. Volatility means that conditional variance of the underlying asset return. It is important in risk management.

3.3.2 GARCH model:

ARCH model is simple, and it requires many parameters to describe the volatility process of an asset return. Bollerslev (1986) proposed a useful extension known as the generalized ARCH or the GARCH model. The term GARCH is referred to both ARCH and GARCH models. A GARCH (p,q) model is defined as follows

$$y_t = \sigma_t \varepsilon_t and \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

Where $\{\varepsilon_t\}$ is a sequence of white noise,$\alpha_0>0$, $\alpha_i \geq 0, \beta_j \geq 0$ and α_i's are referred as ARCH parameters and β_j's are referred to as GARCH parameters respectively. Error is assumed to follow the standard normal or a standardized Student-t or a generalized error distribution.

Basic assumptions of GARCHmodel,

a) Kurtosis value of Close price/ return series should be greater than 3
b) The datashould satisfy Heteroscedasticity
c) The sum of coefficients of GARCH models should be less than one.
d) The squared residual series are uncorrelated.

The null hypothesis Ljung-Box states that there are ARCH or GARCH errors. Rejecting the null hypothesis means that there are no such errors in the conditional variance. Testing for GARCH effect and diagnostic checking are same as ARCH model.

3.4 Support Vector Machines (SVM):

Support vector regression (SVR) is used in the field of Timeseries forecasting, with better outcomes. For instance, Drucker et al. (1997), Muller et al. (1997), and Cao and Tay (2001) suggest that SVR is a promising method for Timeseries forecasting, and it offers several advantages: a smaller number of free parameters, better forecast ability, and faster training.

In SVR, the idea is to map the data events X into a p-dimensional feature space, through a nonlinear mapping, so that it is possible to fit a linear regression model to the data points in this space. Other attractive properties of SVR are related to the use of kernel functions, which make them applicable both to linear and nonlinear forecasting problems, and the absence of local minima in the error surface, due to the convexity of the fitness function and its constraints.The training dataset T is represented by

$$T=\{(x_1,y_1),(x_2,y_2),\ldots\ldots\ldots,(x_m,y_m)\}$$

Where $x \in X$ are the training inputs and $y \in Y$ are the training expected outputs.

Consider a nonlinear function,

$$f(x) = w^T \emptyset(x_i) + b \ldots\ldots\ldots\ldots\ldots\ldots..(1)$$

Where w is the weight vector, b is the bias and$\emptyset(x_i)$ is the high dimensional feature space linearly mapped from input space x. The objective is to fit the training dataset T by finding a function $f(x)$ that has the smallest possible deviation ε from the targets y_i.Equation (1) can be rewritten into a constrained convex optimization problem as follows:

Minimize $\frac{1}{2} w^T w$

Subject to $\begin{cases} y_i - w^T \Phi(x_i) - b \leq \varepsilon \\ w^T \Phi(x_i) + b - y_i \leq \varepsilon \end{cases}$ ……..(2)

The aim of the objective function represented in equation (2) is to minimize w while satisfying the other constraints. One assumption is that $f(x)$ exists, i.e. the convex optimization problem is feasible. This assumption is not always true, therefore, one might want to trade off errors by the flatness of the estimate. Vapnik reformulated equation (2) as

$$\text{Minimize} \quad \frac{1}{2}w^T w + C\sum_{i=1}^{m}(\xi_i^+, \xi_i^-)$$

$$\text{Subject to} \quad \begin{cases} y_i - w^T\Phi(x_i) - \mathrm{b} \le \varepsilon + \xi_i^+ \\ w^T\Phi(x_i) + \mathrm{b} - y_i \le \varepsilon + \xi_i^- \\ \xi_i^+,\ \xi_i^- \ge 0 \end{cases} \quad \text{.....(3)}$$

Where C<0 is a prespecified constant that is responsible for regularization and represents the weight of the loss function and ξ_i^+, ξ_i^-be the slack variables. The first term of the objective function $w^T w$ is the regularized term, whereas the second term $C\sum_{i=1}^{m}(\xi_i^+, \xi_i^-)$ is called the empirical term and measures the ε-insensitive loss function.

To solve Equation (3), Lagrangian multipliers $(\alpha_i^+, \alpha_i^-, \eta_i^+, \eta_i^-)$ can be used to eliminate some of the primal variables. The final equation that translates the dual optimization problem of SVR is

$$\text{Minimize} \quad \frac{1}{2}\sum_{i=1}^{m}K(x_i, x_j)(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-) + \varepsilon\sum_{i=1}^{m}(\alpha_i^+ + \alpha_i^-) - \sum_{i=1}^{m}(\alpha_i^+ - \alpha_i^-)$$

$$\text{Subject to} \quad \begin{cases} \sum_{i=1}^{m}(\alpha_i^+ - \alpha_i^-) = 0 \\ \alpha_i^+, \alpha_i^- \in [0, C] \end{cases} \quad \text{.....} \quad (4)$$

Where $K(x_i, x_j)$ is the kernel function, the above formulation allows the extension of SVR to nonlinear functions, as the kernel function allows nonlinear function approximations while maintain the simplicity and computational efficiency of linear SVR.The performance and good generalization of SVR depend on three training parameters:(i)The kernel function (ii) C (the regularization parameter) (iii) ε (the insensitive zone).

3.5 ARIMA – ANN Hybrid model

Hybrid model is a combination between linear and nonlinear models that usually used for increasing the forecast accuracy. In general, the mathematical form of combination between linear and nonlinear models is as follows:$X_t = L_t + N_t + \mathcal{E}_t$, Where L_t is a linear component andN_t is a nonlinear component of the model. In this paper, Neural Network (NN) is used for modelling the nonlinear component. Estimation of this hybrid model is done in two steps. In first step, the linear component is modelled to get the residuals and then apply a nonlinear model to this residual for handling the nonlinear component. In this paper, ARIMA model is used for handling the linear component. Assumee_t is residual at period t from the first linear model or ARIMA model. i.e. $e_t = X_t - \widehat{L_t}$where $\widehat{L_t}$is the forecast of linear model at period t. Then, NN is applied for modelling e_t as follows:

$$e_t = \mathrm{f}(e_{t-1}, e_{t-2}, \ldots, e_{t-p}) + v_t = \widehat{N_t} + v_t$$

Where f(.) is a nonlinear function from the NN model. Hence, the forecast value of the ARIMA-ANN Hybrid model is as follows

$$\widehat{X_t} = \widehat{L_t} + \widehat{N_t}$$

IV. Empirical Study

The daily information about open, close, low and high price for five stocks viz., Reliance, HDFC Bank, Infosys, TCS and HindustanUnilever (HUL) were gathered for 10 years from the year 2012 to 2021 from BSE-Bombay Stock Exchange. Closing price is used for analysis.The daily data from January 2012 to December 2020 are used for training the model and data from January 2021 to December 2021 are used as test data. R studio software was used for the analysis. The Closing price represents the most up-to-date valuation of a security until trading commences again on the next trading day.

Objectives: Examining the changes in the stock market, Computation value at risk to decide the most appropriate stock for investments, and examine the suitability of the well-known models in order to forecast the stock price.

Table 1: Descriptive Statistics for the Stock price of various company

	Reliance	HDFC	Infosys	TCS	HUL
Minimum	676.2	406.1	526.7	918.3	265.9
Maximum	2731.5	2564.9	4350.8	3954.8	2812.1
Mean	1155.6	1266.1	1787.2	2157.5	1142.8
Variance	177460.4	325427.7	914317.1	406284.1	468732.7
Skewness	1.530476	0.44707	0.5911712	-0.0082398	0.6202346
Kurtosis	4.542561	2.064458	1.988723	2.63519	2.061761

From Table 1,observed that Skewness is less than zero for TCS Company, so most of the values are concentrated on the right of the mean. This shows that stock price of TCS is negatively skewedwhereas stock price of Reliance, HDFC, Infosys, HUL Companies shows Positive skewness. Also observed that the kurtosis is greater than 3 for Reliance Company, so thatprice distribution is leptokurtic and for remaining companies price distribution is platykurtic.

4.1 Value at risk

Table 2: Value at risk and Conditional Value at risk (5% level of significance)

Company	Value at risk		Conditional Value at risk
	NonParametric	Parametric	
Reliance	26617.07	30219.40	41586.86
HDFC	22753.88	28573.55	44109.91
Infosys	27697.16	40641.59	48833.84
TCS	27041.07	34060.97	41120.57
HUL	**20591.57**	**23533.71**	**29260.78**

From Table 2,based on Non-Parametric VaR, observed that for an investment of Rs 10 lakh,the value at risk is Rs. 20591.57 for alpha=0.05.This implies that there is 95% chance of loss not exceeding Rs. 20591.57.Table also shows that HUL followed by HDFC and Reliance companies arethe top 3 companies in terms of minimum Value at Risk.Therefore only these 3 Company's close prices are considered to build forecasting models.

4.2 Time series analysis on stock market price

Figure 1: Time profile of closing price of Reliance Company

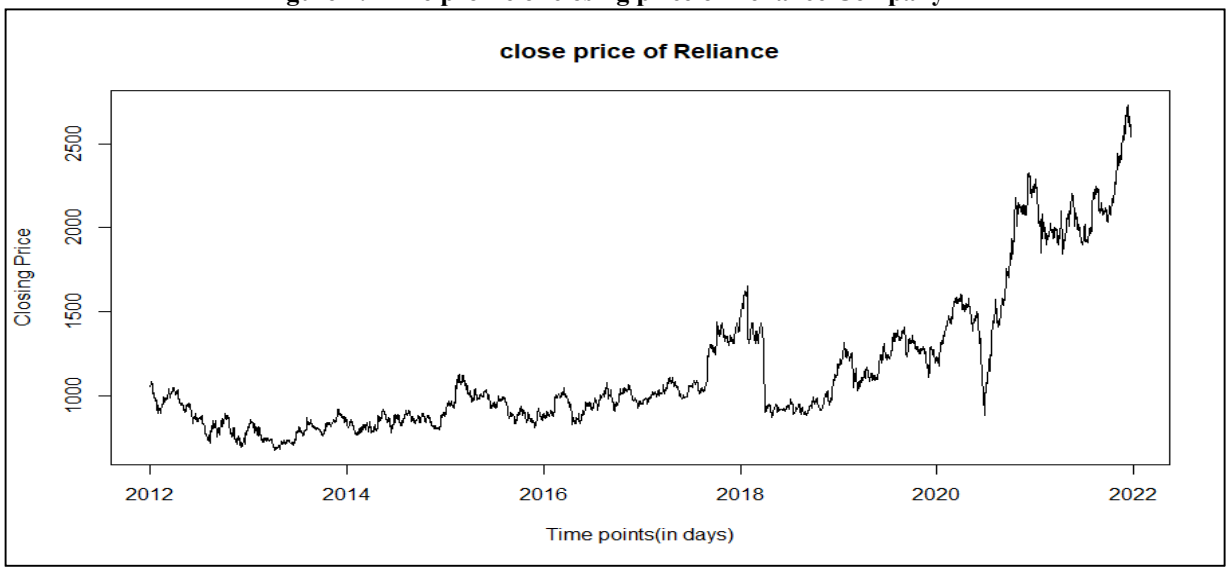

Figure1, indicates that there is a trend component(Upward trend) in the data. Hence observed series is non stationary.

Test for Trend:

Mann-Kendall Trend Test:

H_0 : There is no monotonic trend in the data.

H_1: There is monotonic trend in the data.

tau = 0.736 **2-sided p-value <=2.22e-16**

The p-value of the Mann-Kendall Trend Test is less than significance level 0.05. Therefore, reject H_0 and conclude that there is a monotonic trend in the closing price of the reliance company.

Making the series stationary using variate difference method:

Variance of original series is 186582.5

Variance of first differenced series is 545.3002

Variance of second differenced series is 1021.793

Observed that the variance of the first differenced series is less than the variance of second differenced series. Therefore, by variate difference method, the first differenced series is stationary (d=1).

Figure 2: Plot of ACF and PACF of Stationary Series

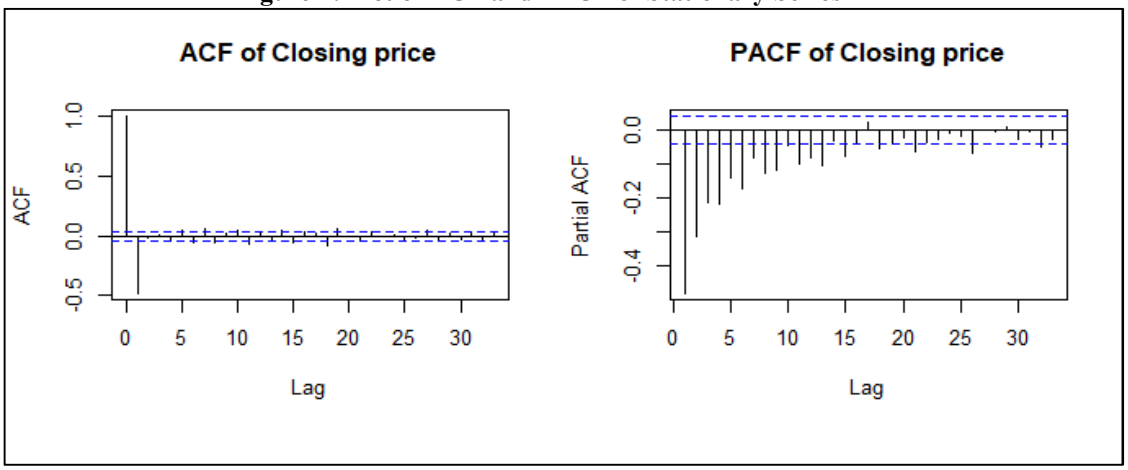

Figure 2, represents the ACF and PACF for stationary series. In the ACF plot,there is one autocorrelation that lies outside the 2σ limits and hence the maximum possible order of q is 1.Similarly in PACF plot, there are 13 partial autocorrelation lies outside the 2σ limits and hence the maximum possible value of p is 13.

ARIMA -GARCH Hybrid Model fitting

Based on the orders of p and q we fitted the different ARIMAmodels with different combinations of (p,d,q). The residuals obtained from the ARIMA are used to predict the volatility. To fit the GARCH model, ACF and PACF of squared residual are examined. Based on the orders of s and m, different GARCH models with different combinations of(m, s) are constructed. The final model is the one with maximum Ljung-Boxtest p-value and minimum AIC.Most appropriate model for the observed series is ARIMA (12,1,1) and GARCH (1,2) model with minimum AIC and highest p-value.

Figure 3 Forecast from ARIMA-GARCH model

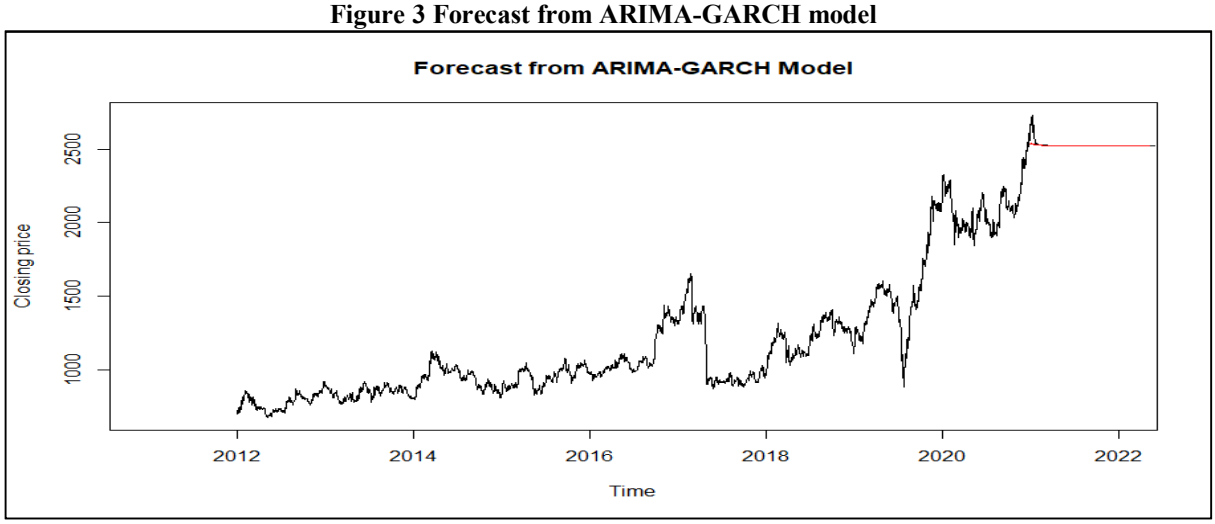

Figure 3, represents the forecasted values for next 365 days using ARIMA-GARCH model

Figure 4 Forecast from SVM Model

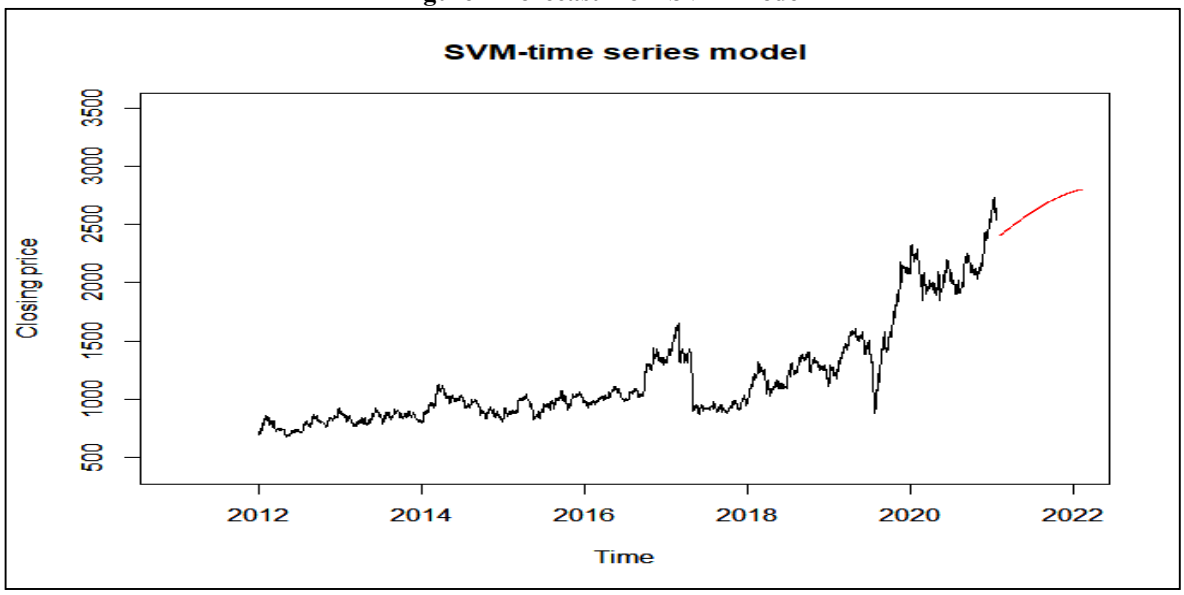

Figure 4, represents the forecasted values for next 365 days using SVM model.

Figure 5 Forecast from ARIMA-ANN Hybrid Model

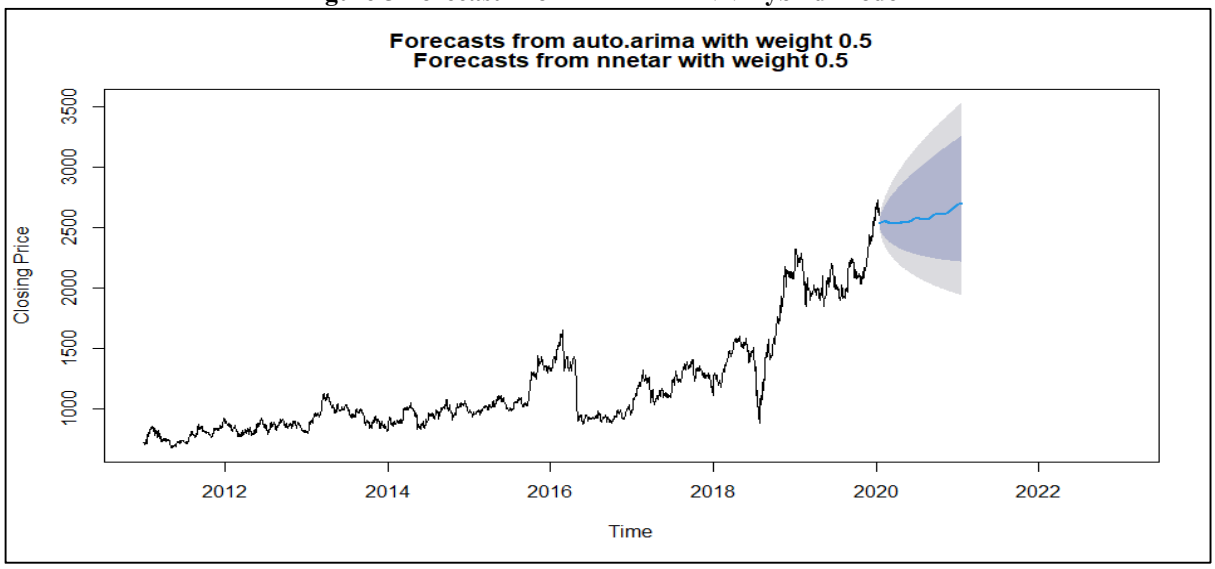

Figure 5, represents the forecasted values for next 365 days using ARIMA-ANN Hybrid model.

Table 4: Forecasting performance of different model based on Accuracy measures

Variable	SVM	ARIMA-ANN	ARIMA-GARCH
RMSE	**23.20465**	23.28777	149.0838
MAE	**15.23096**	15.24088	99.74084
MAPE	**1.278448**	1.28084	8.043128

Based on the value of Root Mean Square Error (RMSE), Mean Absolute Error(MAE),Mean Absolute Percentage Error (MAPE), SVM model and Hybrid model performance are comparable and performs better than ARIMA-GARCH. Finally SVM model is used to forecast the stock price of Reliance Company from October 2021 -October 2022.

Figure 6 Fitted Values along with Original Values for Reliance Company using SVR model

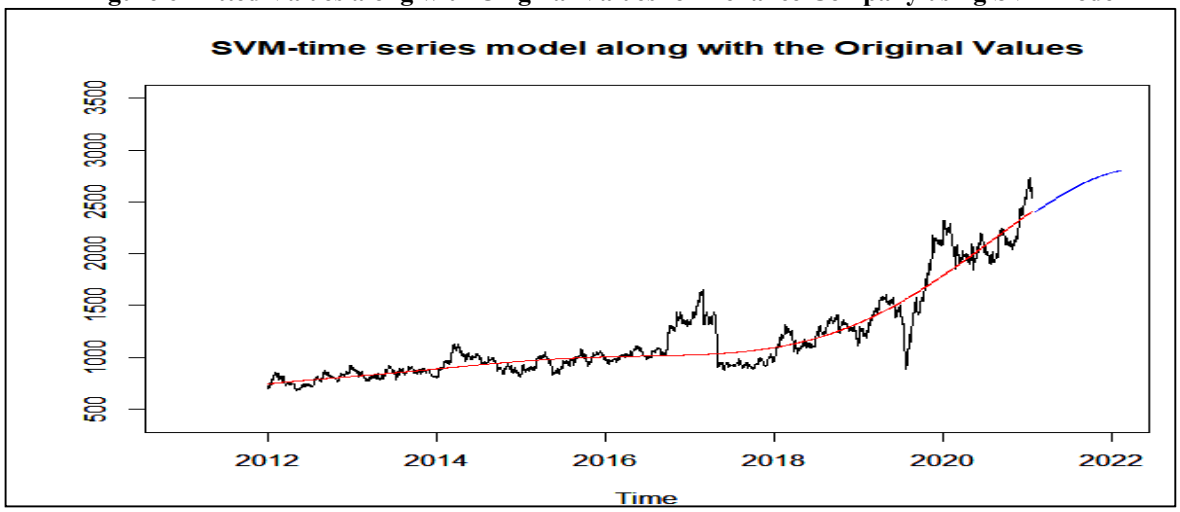

Figure 6, represents the plot of predicted values along with the actual values.

Time series analysis on stock market price of HDFC Bank:

Figure 7: Time profile of closing price of HDFC Bank

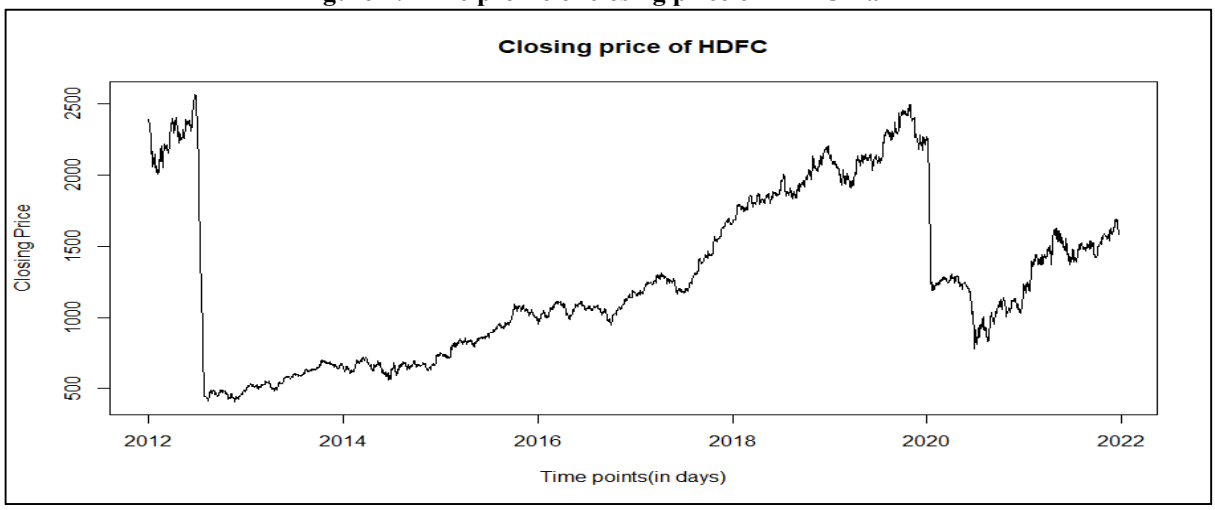

Figure7, indicates that there is a trend component (Upward trend) in the data. Hence observed series is non stationary.

Table 5: Forecasting performance of different model based on Accuracy measures

Variable	SVM	ARIMA-ANN	ARIMA-GARCH
RMSE	17.9596	**17.53694**	203.7955
MAE	11.95596	**11.80114**	134.408
MAPE	1.024097	**1.00755**	11.26318

Based on the value of RMSE, MAE, MAPE, ARIMA-ANN Hybrid model and SVM model performance are comparable and performs better than ARIMA-GARCH. Finally, ARIMA-ANNHybrid model is used to forecast the stock price of Reliance Company from October 2021 -October 2022.

Figure 8 Forecast from ARIMA-ANN Hybrid model

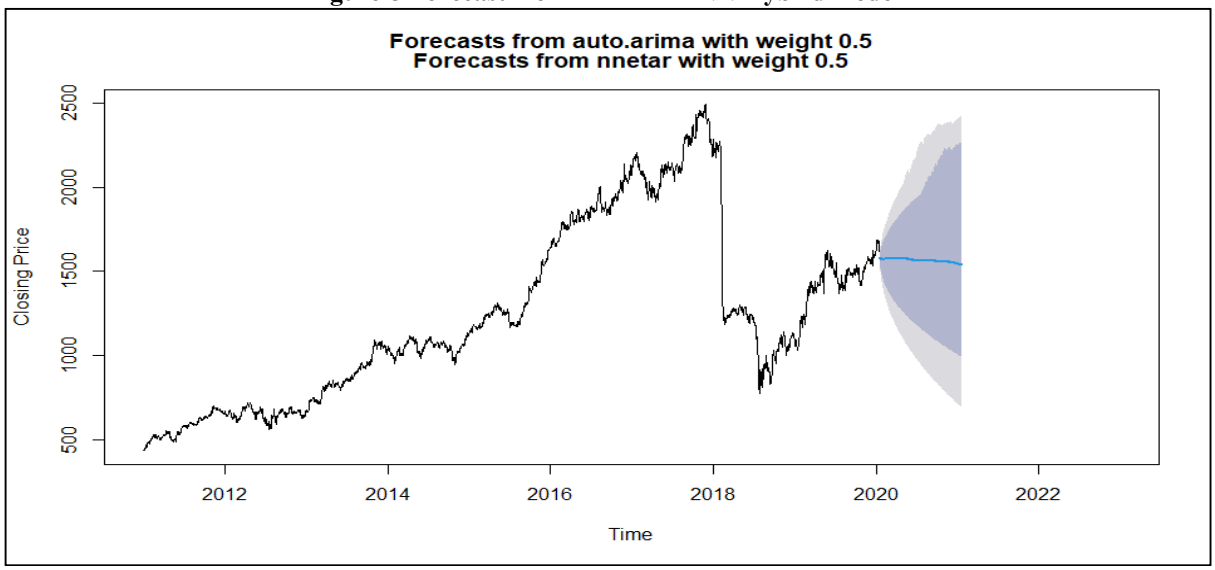

Figure 8,represents the forecasted values for next 365 days using ARIMA-ANN Hybrid model.

Time series analysis on stock market price of HUL:

Figure 9: Time profile of closing price of HUL

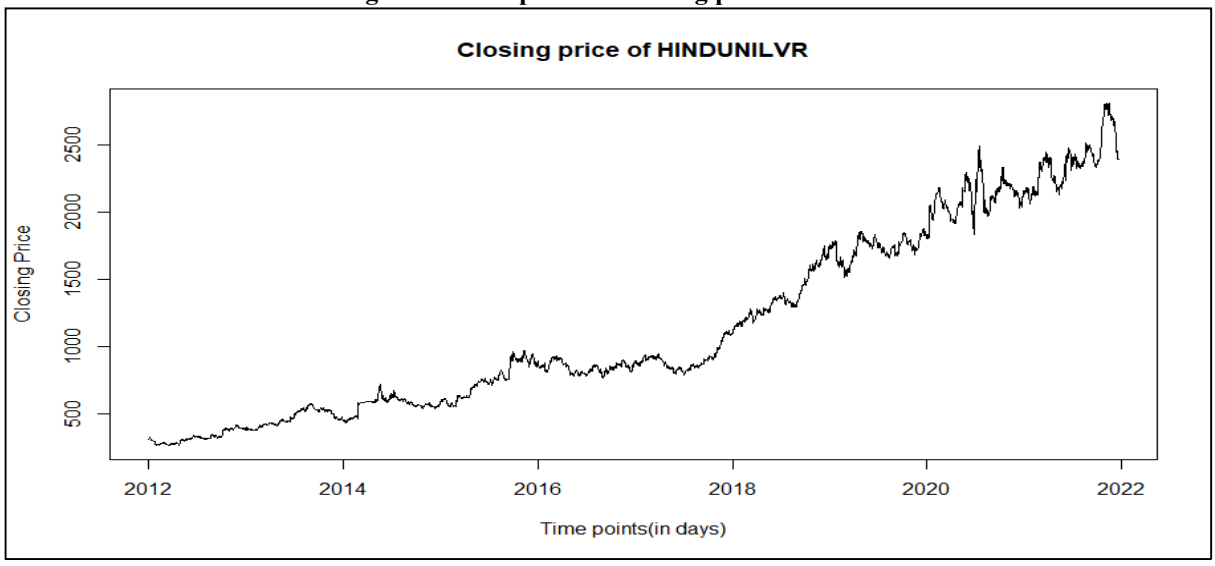

Figure9, indicates that there is a trend component (Upward trend) in the data. Hence observed series is non stationary

Table 6: Forecasting performance of different model based on Accuracy measures

Variable	SVM	ARIMA-ANN	ARIMA-GARCH
RMSE	**20.77815**	21.00564	82.15524
MAE	**12.58526**	12.66731	65.48801
MAPE	**1.018056**	1.02444	6.276301

Based on the value of RMSE, MAE, MAPE, SVM model and Hybrid model performance are comparable and performs better than ARIMA-GARCH. Finally,SVM is used to forecast the stock price of Hindustan UnileverCompany from October 2021 -October 2022

Figure 10 Forecast from SVM model

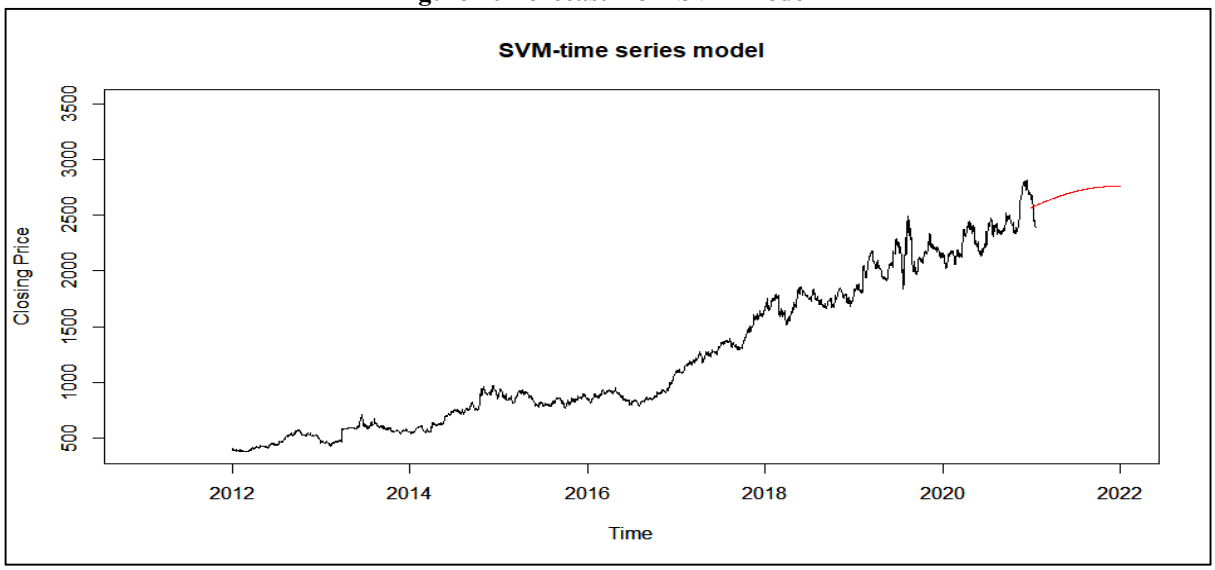

Figure 10,represents the forecasted values for next 365 days using SVM model.

V. Conclusion

Developing predictive models for the stock market is challenging, but it is an important task when building profitable financial market transaction strategies. Forecasting the stock market is one of the most effective tools for risk management and portfolio diversification. In the literature, there are several forecasting techniques available to obtain accurate forecasts which helps the investors in decision making. Previous studies shows that there is no single method that can be applied to all markets uniformly.

This study aimed to examine the predictive performance of ARIMA, artificial intelligence, support vector machines and hybrid models to find an appropriate model to forecast the stock price. We considered the daily stock market price of Reliance, HDFC, Infosys, TCS, and Hindustan Unilever. Hindustan Unilever is the first choice for the investor as it shows the minimum value at risk followed by HDFC and Reliance. From the time profile, the above 3 companies show an upward trend in the price series. The accuracy measure results suggest that the SVR (comparable with Hybrid model)time series model shows better forecasting performance in the case of Reliance and Hindustan Unilever while the Hybrid model shows better performance in HDFC.

References:

[1]. RUEY S. TSAY, Analysis of Financial Time Series Second Edition, University of Chicago Graduate School of Business
[2]. Aparna Nayak, Manohar Pai MM and Radhika M Pai (2016): Prediction models for Indian stock market, Procedia Computer Science
[3]. Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura (2018): Stock Price Prediction Using Support Vector Regression on Daily and Up to the Minute Prices, the Journal of Finance and Data Science. Volume 4, Issue 3, Pages 183-201
[4]. C. W. Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, Kevin W. Hamlen & Nikunj C. Oza (2011): "Facing the reality of data stream classification: coping with scarcity of labeled data", Knowledge and Information System, vol. 33, p. 32, http://dx.doi.org/10.1007/s10115-011- 0447-8.
[5]. Drucker, H., Chris, J., Burges, C., Kaufman, L., Smola, A. and Vapnik, V. (1997): Support Vector Regression Machines, Advances in Neural Information Processing Systems, 9, 155-161.
[6]. J. Gama, (2010): knowledge discovery from data streams, Chapman and Hall/CRC, Taylor and Francis Group
[7]. L. Cao, F. Tay (2001): Financial Forecasting Using Support Vector Machines, Neural Computing & Applications, DOI:10.1007/s005210170010
[8]. M Md Ghani and H A Rahim (2019) :Modelling and forecasting of volatility using ARMA-GARCH: Case Study on Malaysia Natural Rubber Prices , IOP Conference Series Materials Science and Engineering , 548(1):012023 DOI:10.1088/1757-899X/548/1/012023
[9]. M. Mallikarjuna and R. Prabhakara Rao (2019): Evaluation of forecasting methods from selected stock market returns, Financial Innovation**5**, Article number: 40.
[10]. Mauro castelli, Fabiana Martins Clemente, Ales Popovic, Sara Silva, and Leonardo Vannneschi (2020): A Machine Learning Approach to predict Air Quality in California, Hindwai Complexity, WILEY
[11]. Michel Alexandre, Fernando Fagundes Ferreiria, Camilo Rodrigues Neto (2018): Order book, order flow and returns:evidence from the Brazeilian stock exchange,Redeca v.5,p.24-38.
[12]. Nadeem Akhtar (2011): Statistical data analysis of continuous streams using stream DSMS, International Jouranal of Data base management system, DOI:10.5121/ijdms.2011.3206
[13]. R. Tyrrell Rockafellar and Stanislav Uryasev (2000): Optimization of Conditional Value-at-Risk, Journal of risk, volume 2
[14]. Saahil Madge and Swati Bhatt (2015): Predicting Stock Price Direction using Support Vector Machines, Independent Work Report Spring.
[15]. Shweta Tiwari, Alka Gulati (2011): Prediction of Stock Market from Stream Data Time Series Pattern using Neural Network and Decision Tree, International Journal of Electronics & Communication Technology,issue 2,volume 4.

[16]. Subhabrata Choudhury, Subhajyoti Ghosh, Arnab Bhattacharya, Kiran Jude Fernandes, Manoj Kumar Tiwari(2013): A real time clustering and SVM based price-volatility prediction for optimal trading strategy , Neurocomputing, Elsevier.
[17]. Sunil, Satyanarayana, Sachin Acharya, Arun Kumar Jogi(2019): Application Of Hybrid Model For Forecasting Prices Of Jasmine Flower In Bangalore , International Journal Of Scientific & Technology Research, Volume 8, Issue 11.
[18]. Taiwo Kolajo, Olawande Daramola and Ayodele Adebiyi1,(2019):Big data stream analysis: a systematic literature review, Journal of the Big data, Springer.