JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Prediction of Stock Market Data using Various Data Mining Tools

[1]Atish Tangawade, [2]Aniket A. Muley

[1] Research Scholar, [2]Associate Professor

[1,2]Department of Statistics, School of Mathematical Sciences, SRTMU Nanded (India)

Abstract: In this study, an identification of stock market data is performed and future prediction is performed. Here, secondary data is taken for the analysis of TCS Company dataset during 2004-2020. Further, different Time series prediction models are used as well as compared and among that most preferred model is identified on the basis of various performance measures. ARIMA(1,1,0) is found to be suitable for the dataset.

***Keywords* – Stock Market, ARIMA, TCS, Prediction, SVM.**

I. INTRODUCTION

The financial market is a dynamic and composite system where people can buy and sell currencies, stocks, equities and derivatives over virtual platforms supported by brokers. Nowadays, the stock market is attracting more and more people's notice with its high challenging risks and high return over. A stock exchange market depicts savings and investments that are advantageous to increase the effectiveness of the national economy. The future stock returns have some predictive relationships with the publicly available information of present and historical stock market indices. The data mining and its tool has played a vital role in exploring the data from different ware houses. Using data mining tools and analytical technologies we do a quantifiable amount of research to explore new approach for the investment decisions. The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening new business or the need of high salary career.

The inconsistency in SM occurs due to the factors such as economic condition, political activities and trader's prediction. From the fluctuating market, traders always try to make transactions over a small time frame to achieve frequent profit. The key of stock traders to get maximum profits with less risk is conceivable by establishing an accurate trading decision making tool with respect to time. The future price is predicted with better accuracy by concerning the patterns of historical data (price and volume) [1].

Stock markets are affected by many factors causing the uncertainty and high volatility in the market. Although humans can take orders and submit them to the market, automated trading systems (ATS) that are operated by the implementation of computer programs can perform better and with higher momentum in submitting orders than any human. However, to evaluate and control the performance of ATSs, the implementation of risk strategies and safety measures applied based on human judgments are required. Many factors are incorporated and considered when developing an ATS, for instance, trading strategy to be adopted, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that enable the prediction of the future stock value, and specific news related to the stock being analyzed.

Time-series prediction is a common technique widely used in many real-world applications viz., weather forecasting and financial market prediction. It uses the continuous data in a period of time to predict the result in the next time unit. Many time-series prediction algorithms have shown their effectiveness in practice. The most common algorithms now are based on its special type - Long-short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Stock market is a typical area that presents time-series data and many researchers study on it and proposed various models. ARIMA is a statistical model which is known to be efficient for time series forecasting especially for short-term prediction. In this study, we propose a model for forecasting the stock market trends based on the technical analysis using historical stock market data and ARIMA model. This model will automate the process of direction of future stock price indices and provides assistance for financial specialists to choose the better timing for purchasing and/or selling of stocks. We have also used LSTM and SVM models to predict the stock price. The results are shown in terms of visualizations using Python programming language. The obtained results reveal that the ARIMA model has a strong potential for short-term prediction of stock market trends.

Objectives

1. To perform descriptive study of stocks of TCS company.
2. To understand nature of stock prices in last two decades.
3. To identify and discover useful information from the dataset.
4. To predict the stock prices in order to make more informed and accurate investment decisions.
5. Best time span for predicting the future price of the share.

6. To understand the behavioral patterns as well as seek to develop a new hybrid forecasting approach based on ARIMA for estimating VWAP ratios.

II. RIVIEW OF LITERATURE

Angadi and Kulkarni [2] made an attempt to develop a prediction model for forecasting the stock market trends based on the technical analysis using historical time series stock market data and data mining techniques. The experimental results obtained demonstrated the potential of ARIMA model to predict the stock price indices on short -term basis which are helpful to investors in the stock market to make profitable investment decisions whether to buy/sell/hold a share. Their obtained results showed that, ARIMA model can compete reasonably well with emerging forecasting techniques in short-term prediction. Devi et.al. [3] selected top 4 NSE – Nifty Midcap50 companies having max Midcap value for analysis. They collected stock data for the past five years and The Box Jenkins methodology issued to identify the model. After that, they trained using ARIMA model with different parameters. The test criterions like Akaike Information Criterion Bayesian Information Criterion (AICBIC) were applied to predict the accuracy of the model. The performance of the trained model is analyzed and it also tested to find the trend and the market behavior for future forecast. The MAPE, PMAD and % Error accuracy is applied to determine the discrimination between the actual historical data and the forecast data.

Non-stationary time series show varying degrees of volatility in some periods, and these volatility depend on the past of the sequence. Some other methods like Markov chain and recurrent neural network provide us different approach to model the temporal dependence of non-stationary time series, for instance in financial field. As a special financial time series application, stock market prediction is a classic problem which has been studied extensively using various machine learning tools and techniques. Tay and Cao [4] and Kim [5] studied various parameters (the upper bound and the kernel parameter) of SVM to select optimal values for the best prediction performance. In addition to this, the work of Hassan and Nath [6] and Gupta and Dhingra [7] shows that Hidden Markov Model (HMM) is widely used in financial time series price forecasting and is better than ANN and ARIMA method in price predicting accuracy. However, the hyper-parameter settings (such as the number of hidden state) of HMM is easily influenced by people and there is no authoritative principle to guide the selection of parameters.

Radaideh et al. [8] studied the use of decision tree classifier on the historical prices of the stocks to create decision rules that give buy or sell recommendations in the stock market. Such proposed model can be a helpful tool for the investors to take the right decision regarding their stocks based on the analysis of the historical prices of stocks in order to extract any predictive information from that historical data. Their results for the proposed model were not perfect because many factors including but not limited to political events, general economic conditions, and investors' expectations influence stock market.

Enke and wong [9] proposed 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.

Olaniyi et al. [10] dealt with regression analysis as a data mining technique and developed tool for exploiting especially time series data in financial institution. A prediction system has been built that uses data mining technique to produce periodically forecasts about stock market prices. Our technique complement proven numeric forecasting method using regression analysis with technology taking as input the financial information obtained from the daily activity summary (equities) published by Nigerian Stock Exchange. Some other research used the techniques of technical analysis [11], in which trading rules were developed based on the historical data of stock trading price and volume. Technical analysis as illustrated in [12] and [13] refers to the various methods that aim to predict future price movements using past stock prices and volume information. It is based on the assumption that history repeats itself and that future market directions can be determined by examining historical price data. Thus, it is assumed that price trends and patterns exist that can be identified and utilized for profit. Most of the techniques used in technical analysis are highly subjective in nature and have been shown not to be statistically valid.

Osman Hegazy et al.[13] proposed train Least Square- Support Vector Machine (LS-SVM) Flower Pollination Algorithm (FPA), Bat algorithm (BA), Modified Cuckoo Search (MCS), Artificial Bee Colony (ABC), and Particle Swarm Optimization (PSO) to optimize and predict the stock prices. These algorithms automatically select best free parameters combination for LS-SVM. Author collected Six financial technical indicators derived from stock historical data as inputs to proposed models. The Researchers shows that the proposed models have quick convergence rate at early stages of the iterations. They achieved better accuracy than compared methods in price and trend prediction. They also overcame over fitting and local minima problems found in ANN and standard LS-SVM.

Jack Y. Yang et al. [14] used PSO based selective neural network ensemble (PSOSEN) algorithm used for the Nasdaq-100 index of Nasdaq Stock Market SM and the S&P CNX NIFTY stock index analysis. In the algorithm, each neural network is obtained by bagging and is trained by PSO algorithm, and then the networks selected according to the pre-set threshold are combined. Their experimental results show that the improved algorithm is effective and outperforms GA based selective ensemble (GASEN) algorithm for the stock index forecasting problems. Author presents a PSOSEN algorithm for stock index forecasting of two well-known stock indices namely Nasdaq-100 index of NasdaqSM and the S&P CNX NIFTY stock index. They also compared the test results of the GASEN algorithm and the PSOSEN algorithm on the two data sets. It is shown from the test results that the PSOSEN algorithm is more accurate when compared with the GASEN algorithm. Moreover, the improved algorithm is faster than the GASEN algorithm. Since selective neural network ensemble is shown better in many fields, more rigorous testing on more complex problems will be performed in future works.

Some researchers have proposed decision making tools based on artificial intelligence (AI) and deep learning techniques to enhance the accuracy of the SM forecasting. Bustos et al. [16] have contributed a brief survey of different methods implemented to predict stock market price. Artificial neural network (ANN)-based model is preferred to forecast SM [17–20], because of its benefits such as organic learning, nonlinear data processing, fault tolerance and self-overhaul. Feed forward neural network (FFNN) is a simple form of ANN, which has very less computational time to solve simple problems with less complexity [21]. The effectiveness of deep feed forward neural network to forecast stock indices price is analyzed by a comparative analysis with ANN.

Multilayer Perceptron (MLP) is a persuasive ANN to be used for regression and is efficient enough to predict SM price. Back propagation neural network (BPNN) is also implemented to solve SM index forecasting problem [14, 15].

III. METHODOLOGY

This study deals with the TCS company dataset for analyzing dataset with data mining tools. Here, we have considered secondary dataset of TCS. This dataset were taken from website: (https://www.kaggle.com/datasets)

Collection of Data:

In this study, to analyze stock prices of TCS Technologies data following variables are under consideration:

- The Open and Close columns indicate the opening and closing price of the stocks on a particular day.
- The High and Low columns provide the highest and the lowest price for the stock on a particular day, respectively.
- The Volume column tells us the total volume of stocks traded on a particular day.
- The volume weighted average price (VWAP) is a trading benchmark used by traders that gives the average price a security has traded at throughout the day, based on both volume and price. It is important because it provides traders with insight into both the trend and value of a security. The Turnover column refers to the total value of stocks traded during a specific period of time. The time period may be annually, quarterly, monthly or daily

The following methods are used to predict the VWAP ratios of the gathered datasets. The detailed information of the tools is as follows:

Light GBM

It is a gradient boosting framework that makes use of tree based learning algorithms that is considered to be a very powerful algorithm when it comes to computation. It is considered to be a fast processing algorithm.

Dummy Regressor

The Dummy Regressor is a kind of Regressor that gives prediction based on simple strategies without paying any attention to the input Data. It always predicts the R^2 score as 0 for both the mean and median, since it is always predicting a constant without having an insight of the output. (In general, best R^2 score is 1 and Constant R^2 score is 0). The Linear Regression Model seems to fit a little better than the Dummy Regressor in terms of "mean squared error", "median absolute error" and "R^2 score".

ARIMA

Auto Regressive Integrated Moving Average (ARIMA) model is among one of the more popular and widely used statistical methods for time-series forecasting. It is a class of statistical algorithms that captures the standard temporal dependencies that is unique to a time series data. The ARIMA model uses statistical analyses in combination with accurately collected historical data points to predict future trends and business needs. The ARIMA model is typically denoted with the parameters (p, d, q), which can be assigned different values to modify the model and apply it in different ways.

Support vector machine (SVM)

SVM is one popular algorithm used for many classification and regression problems. Primarily, it is used for Classification problems in data mining and machine learning. It is one of the supervised learning models that are based on the concept of hyperplane as decision boundaries.

Long Short Term Memory (LSTM)

It is a model or architecture that extends the memory of recurrent neural networks. LSTM introduces long-term memory into recurrent neural networks.

Performance Measures

In this study, we have used the following performance measures:

- **MAE (Mean absolute error)** represents the average of the absolute difference between the actual and predicted values in the dataset. It measures the average of the residuals in the dataset.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_{i_{pred}} - Y_{i_{act}}|$$

Where, $Y_{i_{pred}}$ is predicted value of i^{th} observation, $Y_{i_{act}}$ is actual value of i^{th} observation, n is sample size.

- **RMSE:** It is the square root of the average of squared differences between actual values and predicted values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{i_{pred}} - Y_{i_{act}})^2}$$

- **MAPE:** The mean absolute percentage error (MAPE) is the mean or average of the absolute percentage errors of forecasts. Error is defined as actual or observed value minus the forecasted value. Percentage errors are summed without regard to sign to compute MAPE. The smaller the MAPE the better the forecast.

$$MAPE = \frac{1}{n}\left(\sum_{i=1}^{n}\frac{|Y_{i_{pred}} - Y_{i_{act}}|}{|Y_{i_{act}}|}\right) \times 100$$

- **Confusion Matrix:** A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.
  Let's now define the most basic terms, which are whole numbers (not rates):
  True positives (TP): These are cases in which we predicted yes when it is actually yes.
  True negatives (TN): We predicted no, and they it is actually yes (Also known as a "Type I error.")
  False positives (FP): We predicted yes, but it is actually no
  False negatives (FN): We predicted no, but they actually yes (Also known as a "Type II error.")

This is a list of rates that are often computed from a confusion matrix for a binary classifier:

$$Accuracy = \frac{(TP + TN)}{N}$$

$$Missclassifiacation\ Rate = \frac{(FP + FN)}{N}$$

Which is equivalent to 1 minus Accuracy, also known as "Error Rate"

True Positive Rate/Sensitivity/Recall: $True\ Positive\ Rate = \frac{TP}{actual\ yes}$

False Positive Rate: $False\ Positive\ Rate = \frac{TP}{actual\ no}$

True Negative Rate/Specificity: $True\ Negative\ Rate = (1 - FPR) = \frac{TN}{actual\ no}$

Precision: $Precision = \frac{TP}{predicted\ yes}$

Prevalence: $Prevalance = \frac{acual\ yes}{N}$

F Score: This is a weighted average of the true positive rate (recall) and precision.

ROC Curve: This is a commonly used graph that summarizes the performance of a classifier over all possible thresholds. It is generated by plotting the True Positive Rate (y-axis) against the False Positive Rate (x-axis) as varying the threshold for assigning observations to a given class.

- **AIC (Akaike Information Criteria):** Akaike Information Criteria (AIC) is an evaluation of a continual in addition to the corresponding interval among the undetermined, accurate, and justified probability of the facts. The AIC calculation is done with the following formula:

$$AIC = 2k - 2\ln(L(\hat{\theta}))$$

- **BIC (Bayesian Information Criteria):** Bayesian Information Criteria (BIC) is an evaluation of the purpose of the possibility, following the model is accurate, under a particular Bayesian structure. The BIC calculation is done with the following formula:

$$BIC = kln(n) - 2\ln(L(\hat{\theta}))$$

θ = the set (vector) of model parameters

$L(\hat{\theta})$ = the likelihood of the fitted model given the data when evaluated at the maximum likelihood estimate of θ

k = the number of estimated parameters in the fitted model

Following flow chart highlights work-flow of our study:

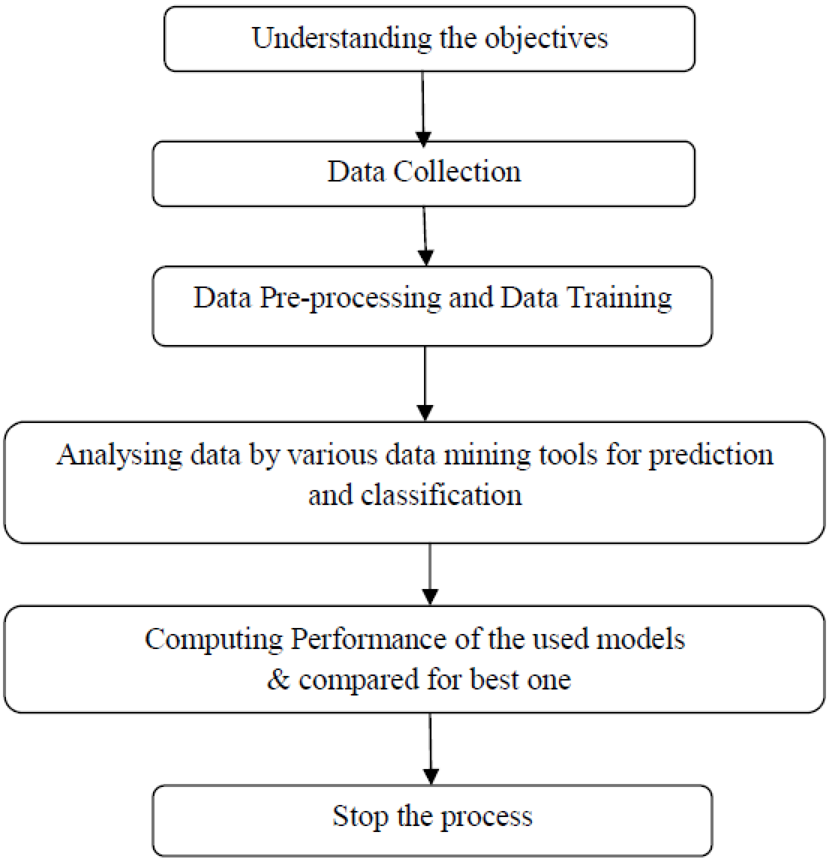

Fig. 1 Flow chart

Fig. 1 represents the flow chart of our study. It explores our study direction for prediction and classification of the stock market data. The Light-GBM, Dummy Regressor and ARIMA methods were used for prediction whereas LSTM and SVM methods used for classification. Further various performance measures are checked and compared to select best predictive model.

IV. Result and Discussion

In this study, the results obtained through the data mining tools are summarized below:

Time Series Analysis:

- **Visualization of Data:**

Fig. 2 (a) TCS Full Dataset

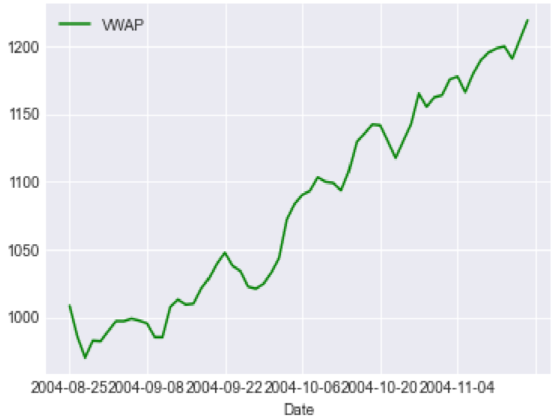

Fig. 2(b) TCS First 60 values

Seasonal decomposition:

We can decompose a time series into trend, seasonal and remainder components. The series can be decomposed as an additive or multiplicative combination of the base level, trend, seasonal index and the residual.

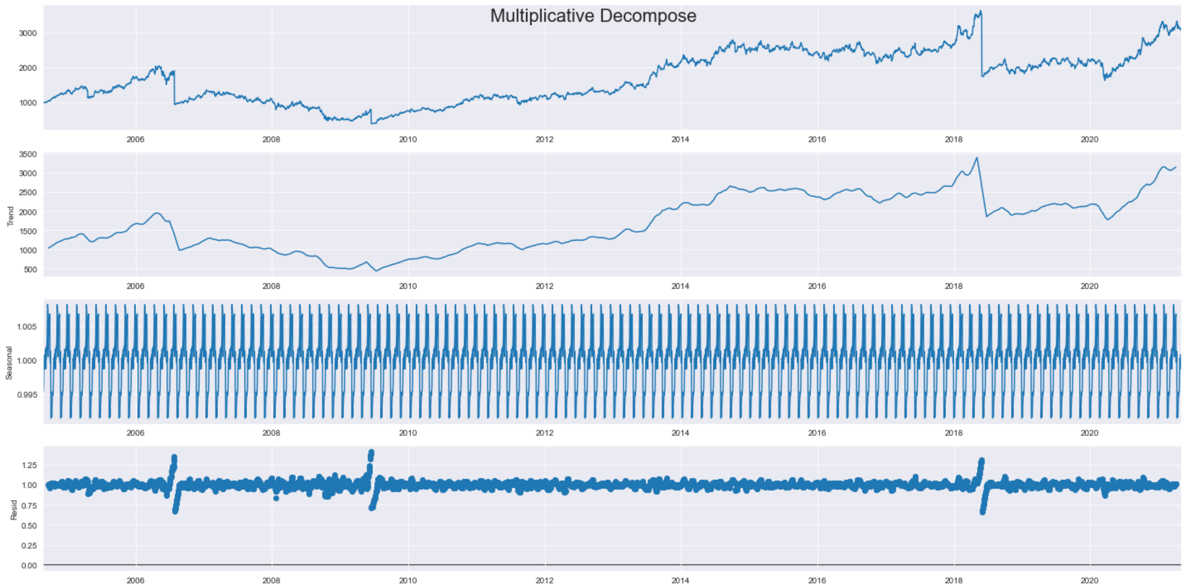

Fig. 3(a) Multiplicative decomposition of TCS Data

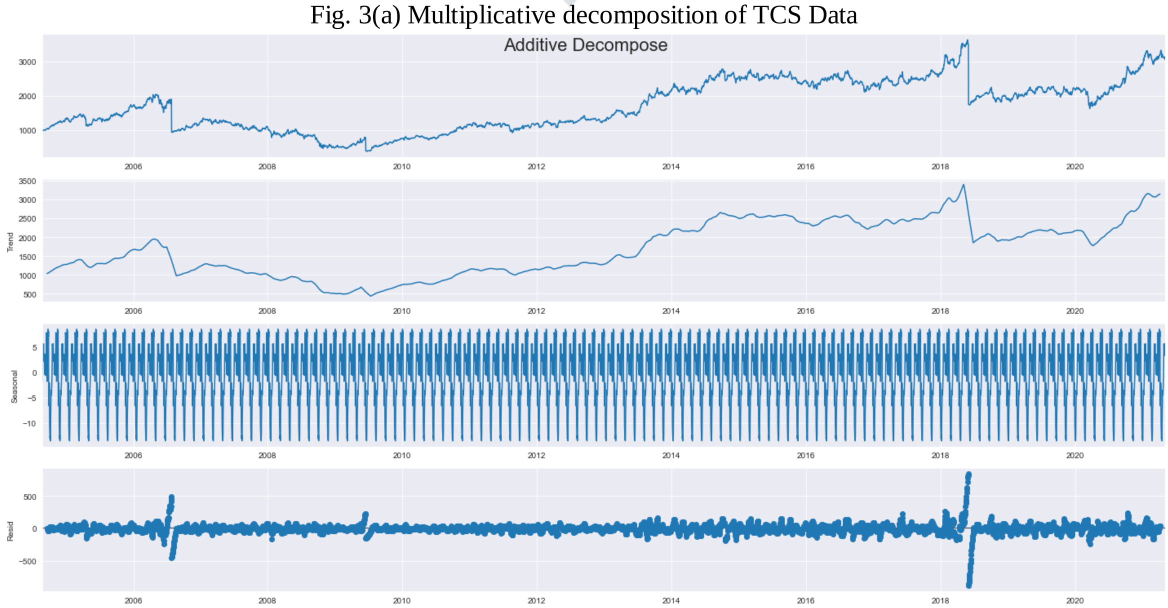

Fig. 3(b) additive decomposition of TCS Data

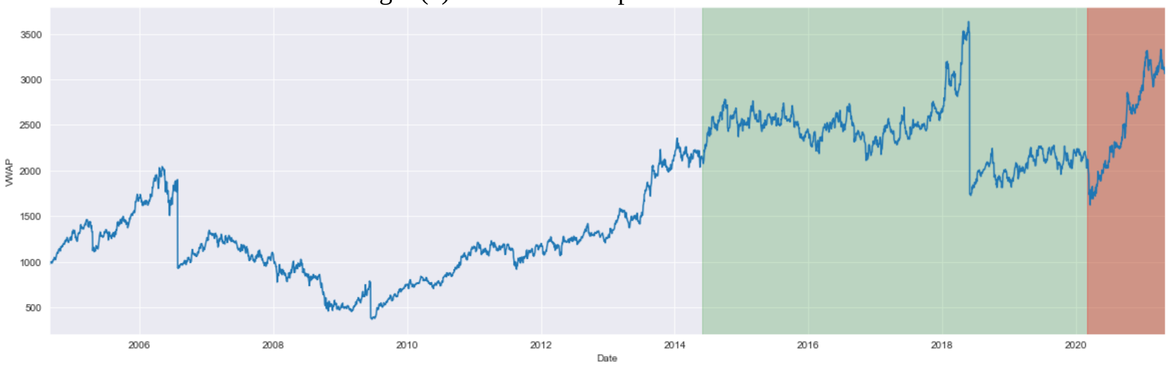

Fig 4 Time span classification of TCS Data

Fig. 3(a-b) illustrates that, the representation of TCS data from year 2000 to 2021. In Fig. 4 the green shaded portion explains the growth during Modi government establishment since 2014. The red shaded portion explores Covid -19 pandemic periods. Further, we observed that, TCS prices are increased till 2002 then it suddenly falls down. Afterwards there is upward trend in TCS stock prices till 2016, but in year 2017 prices are falls down and shows stationary behavior upto 2020.Although, TCS prices are exponentially increased in Covid-19 pandemic periods. While, VWAP ratio of WIPRO stock is reached to pick in starting of year 2002 and then there is continuous decrement in VWAP ration till 2005. Afterwards, it shows stationary behavior from year 2005 to year 2020. But, in Covid-19 pandemic period, VWAP ratio is slightly increased. For, HCL technologies stocks, figures shows that, VWAP ratios are negative exponentially decreased from year 2001 to 2004, then VWAP ratios are increased till 2008. In year 2009, VWAP ratios are slightly decreased and show stationary behavior till 2012 but then there is high increment till 2015. But, due to some reasons, VWAP ratio for HCL technology suddenly falls down then it shows stationary behavior till 2019. But in year 2020, VWAP ratio is falls down and shows slight increment in Covid-19 pandemic periods.

Stationarity

In the most intuitive sense, stationarity means that the statistical properties of a process generating a time series do not change over time. It does not mean that the series does not change over time, just that the way it changes does not itself change over time. The algebraic equivalent is thus a linear function, perhaps, and not a constant one; the value of a linear function changes as x grows, but the way it changes remains constant — it has a constant slope; one value that captures that rate of change.

Table 1 Comparative Model apporach

	Multiplicative	Additive
ADF test statistic	-1.558160	-1.558160
p-value	0.504570	0.504570
lags used	1.000000	1.000000
observations	6091.000000	6091.000000
critical value ({key})	-2.567023	-2.567023

Table (1) represents the comparative stationarity result, it is observed that by using ADF test TCS company's data shows stationarity.

Fig. 5 Dummy Model prediction results

Fig. 6 LightGBM prediction results

Fig.5 represents the dummy models prediction results and Fig. 6 represents the LightGBM models prediction results of TCS dataset.

ARIMA Predictions:

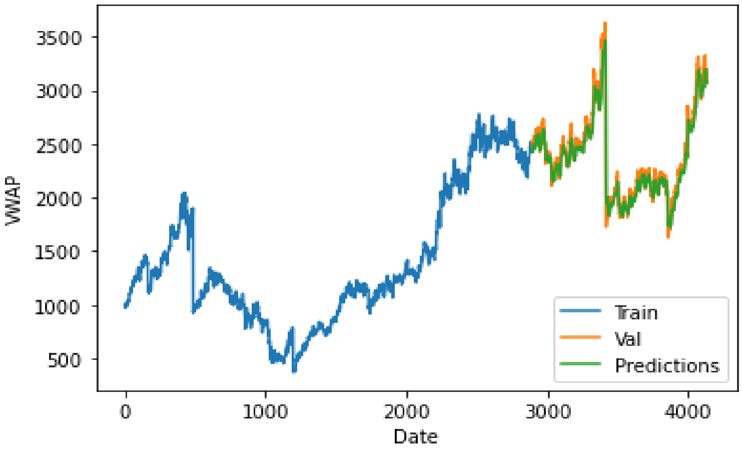

Fig. 7 ARIMA model prediction results

Comparative Study of Predictive techniques for all three datasets:

Table 2 Different models performance measures

Result	**TCS**
Dummy Regressor	
RMSE	517.5626882313153
MAE	507.0105625192726
LBGM Regressor	
RMSE	740.8829578313048
MAE	714.6186890573044
ARIMA	
RMSE	233.09264476894316
MAE	210.88550194849063

Table 2 explores that, ARIMA model gives good prediction results as compared to other ones.

Simulation result of ARIMA models by autoarima function in Python:

TCS

Performing stepwise search to minimize aic

ARIMA(2,1,2)(0,0,0)[0] intercept : AIC=51828.328, Time=3.93 sec
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=51839.372, Time=0.82 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=51822.696, Time=2.49 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=51823.002, Time=2.56 sec
ARIMA(0,1,0)(0,0,0)[0] : AIC=91482.821, Time=0.59 sec
ARIMA(2,1,0)(0,0,0)[0] intercept : AIC=51824.430, Time=3.07 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=51824.333, Time=3.11 sec
ARIMA(2,1,1)(0,0,0)[0] intercept : AIC=51845.401, Time=1.47 sec
ARIMA(1,1,0)(0,0,0)[0] : AIC=51820.803, Time=1.92 sec
ARIMA(2,1,0)(0,0,0)[0] : AIC=51822.535, Time=2.60 sec
ARIMA(1,1,1)(0,0,0)[0] : AIC=51822.431, Time=2.67 sec
ARIMA(0,1,1)(0,0,0)[0] : AIC=51821.111, Time=2.15 sec
ARIMA(2,1,1)(0,0,0)[0] : AIC=51843.523, Time=1.31 sec

Best model: ARIMA(1,1,0)(0,0,0)[0]

Here, we have performed simulation to identify the best suitable model to perform prediction of TCS. Table 3 represents the summary result of the best ARIMA models for different datasets.

Summary of the best fitted ARIMA model:

Table 3 Summary of the Best ARIMA Model

	TCS
Dep. Variable	y
Model:	SARIMAX(1, 1, 0)
No. Observations:	5242
Log Likelihood	-25907.401
AIC	51820.803
BIC	51840.496
HQIC	51827.689
Ljung-Box (L1) (Q):	0.00
Prob(Q):	0.98

Heteroskedasticity (H):	2.86
Prob(H) (two-sided):	0.00
Jarque-Bera (JB):	499767788.88
Prob(JB):	0.00
Skew:	-31.42
Kurtosis:	1514.50

Hypothesis testing and confidence interval estimation of VWAP parameter

Table 4 Hypothesis and confidence intervals for various parameters

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
VWAP	-1.898e-06	1.91e-07	-9.952	0.000	-2.27e-06	-1.52e-06
ar.L1	0.0597	0.006	9.798	0.000	0.048	0.072
sigma2	1151.0315	1.086	1060.244	0.000	1148.904	1153.159

Here we have performed Hypothesis testing and confidence interval estimation of VWAP parameter and other model parameters.

Classification using LSTM and SVM

Table 5 Classification performance using SVM

	precision	recall	f1-score	support
0	0.58	0.64	0.61	381
1	0.66	0.60	0.63	447
ACCURACY	0.62			828

Table 6 LSTM results

RMSE by LSTM	Accuracy	MAPE
104.3794	0.6183	38.1642

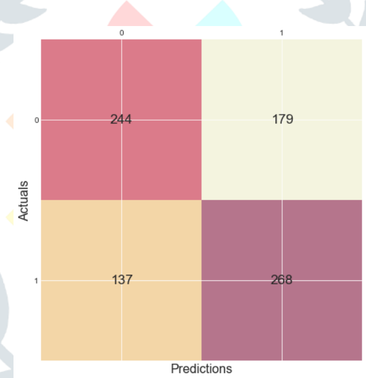

Fig. 8 Confusion Matrix

Table 5 and table 6 represent the performance measures of SVM and LSTM to compute accuracy, RMSE and MAPE of the proposed model. Fig. 8 represents the exploration of accuracy in the form of confusion matrix.

V. Conclusion:

In this study, we have dealt with time series data of renowned company like TCS. Here, we have applied time series traditional approach and further compared with advance data mining tools. It is observed that, ARIMA model perform well as compared with other models. Overall, the results of the study reveals that ARIMA model is suitable for stock market dataset and gives precise results and hence in upcoming future if someone may use this method to predict the stock VWAP ratio, this method will be fruitful.

VI. References:

1. Schabacker, R. (2005). *Technical analysis and stock market profits*. Harriman House Limited.
2. Angadi, M. C., & Kulkarni, A. P. (2015). Time Series Data Analysis for Stock Market Prediction using Data Mining Techniques with R. *International Journal of Advanced Research in Computer Science*, *6*(6).
3. Devi, B. U., Sundar, D., & Alli, P. (2013). An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50. *International Journal of Data Mining & Knowledge Management Process*, *3*(1), 65.
4. Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *omega*, *29*(4), 309-317.
5. Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1-2), 307-319.
6. Hassan, M. R., & Nath, B. (2005, September). Stock market forecasting using hidden Markov model: a new approach. In *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)* (pp. 192-196). IEEE.

7. Gupta, A., & Dhingra, B. (2012, March). Stock market prediction using hidden markov models. In *2012 Students Conference on Engineering and Systems* (pp. 1-4). IEEE.
8. Al-Radaideh, Q. A., Assaf, A. A., & Alnagi, E. (2013, December). Predicting stock prices using data mining techniques. In *The International Arab Conference on Information Technology (ACIT'2013)* (pp. 1-8).
9. Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with applications*, *29*(4), 927-940.
10. Olaniyi, S. A. S., Adewole, K. S., & Jimoh, R. G. (2011). Stock trend prediction using regression analysis–a data mining approach. *ARPN Journal of Systems and Software*, *1*(4), 154-157.
11. Wu, M.C., Lin, S.Y., and Lin, C.H., (2006) "An effective application of decision tree to stock trading",Expert Systems with Applications, 31, pp. 270-274.
12. Tsang, P.M., Kwok, P., Choy, S.O., Kwan, R., Ng, S.C., Mak, J., Tsang, J., Koong, K., and Wong, T.L.(2007) "Design and implementation of NN5 for Hong Kong stock price forecasting", Engineering Applications of Artificial Intelligence, 20, pp. 453-461.
13. Murphy, J.J., (1999) Technical Analysis of the Financial Markets: a Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance.
14. JZ Wang JJ Wang ZG Zhang SP Guo 2011 Forecasting stock indices with back propagation neural network Expert Syst Appl 38 11 14346 14355 https://doi.org/10.1016/j.eswa.2011.04.222 15.
15. D Zhang S Lou 2020 The application research of neural network and BP algorithm in stock price pattern classification and prediction Futur Gener Comput Syst 115 872 879 https://doi.org/10.1016/j.future.2020.10.009
16. O Bustos A Pomares-Quimbaya 2020 Stock market movement forecast: a systematic review Expert Syst Appl 156 113464https://doi.org/10.1016/j.eswa.2020.113464
17. AH Moghaddam MH Moghaddam M Esfandyari 2016 Stock market index prediction using artificial neural network J Econ Financ Adm Sci 21 41 89 93 https://doi.org/10.1016/j.jefas.2016. 07.002
18. M Qiu Y Song F Akagi 2016 Application of artificial neural network for the prediction of stock market returns: the case of the Japanese stock market Chaos Solitons Fract 85 1 7 https://doi.org/10.1016/j.chaos.2016.01.004
19. D Vukovic Y Vyklyuk N Matsiuk M Maiti 2020 Neural network forecasting in prediction Sharpe ratio: evidence from EU debt market Physica A 542 123331 https://doi.org/10.1016/j.physa. 2019.123331
20. E Guresen G Kayakutlu TU Daim 2011 Using artificial neural network models in stock market index prediction Expert Syst Appl 38 8 10389 10397 https://doi.org/10.1016/j.eswa.2011.02. 068
21. S Gupta LP Wang 2010 Stock forecasting with feedforward neural networks and gradual data sub-sampling Aust J Intel Inf Process Syst 11 4 14 17