

Data Analysis, Visualization and Prediction of Stock Market Prices of K-Electric

Noman Islam¹, Misbah Afzal², Muhammad Aarsal Wali², Hamza Shakeel²

¹Karachi Institute of Economics and Technology, Karachi, Pakistan

²NED University of Engineering and Technology, Karachi, Pakistan

Corresponding author: Noman Islam (e-mail: noman.islam@gmail.com).

Received: 15/05/2022, Revised: 20/08/2022, Accepted: 25/09/2022

Abstract— Predicting stock price is a trend yet very challenging task. It is because the stock prices depend upon several internal and external factors. Stock price prediction can be very useful for financial sectors and the government and help in informed decision-making. This paper analyzes the stock market prices of K-Electric Karachi. It is found that the stock prices of K-electric depend on the stock prices of the refinery sector. The paper analyzes the stock price data of the two sectors. Also, the paper compares the stock price prediction based on moving average, auto-regressive integrated moving average (ARIMA), convolutional neural network and long short-term memory (LSTM) model. It is found that ARIMA outperforms the other algorithms. A set of experiments were conducted to test the performance of algorithms. The algorithms were analyzed based on different metrics such as root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

Index-Terms — Stock price prediction, ARIMA, CNN, LSTM

I. INTRODUCTION

A stock market is a place where stocks/shares of a company are traded. Stock prices estimate a country's future economy and contribute to a company's growth [11]. Uncertainty is the crucial factor that makes stock market prediction a formidable task, among many other factors. According to the efficient market hypothesis, it is impossible to model the stock price prediction problem [15]. However, few frameworks have successfully predicted stock prices with good accuracy.

Finance and business keep track of stock indexes and prices to gain a fair idea of investment. Getting updates from social and technical resources is considered best practice because the risk factor is greater. Among various factors contributing to the prediction of stock prices are market anticipation, confidence in the company, national policies and domestic factors [10, 11].

However, it is only efficiently possible with machine learning, and big data techniques as the volume of data are large with its variable and unusual nature [10]. Stock's historical values play a major part in future stock prediction. But unusual events are again a big challenge here. Stock prices can be predicted using technical approaches such as statistics, machine learning, deep learning, or time series analysis.

Karachi Electric (KEL) is one of the largest electricity-producing companies in Pakistan. It covers the whole Karachi region. In this paper, KEL's stock prices are compared with a company in

the refinery sector. It is observed that the stock price of the electric sector is directly linked with the stock prices of the refinery sector. Further, the stock prices is a sequential data dependent on past trends. Hence, time series analysis can be performed using auto-regressive integrated moving average (ARIMA), long short-term memory (LSTM) and convolutional neural network. Therefore, this paper analyzes various time series models for stock price prediction.

II. LITERATURE REVIEW

In literature, several algorithms are proposed for stock price prediction. There are two different approaches to analyzing and predicting stock market trends. One of the approaches is qualitative and based on several external factors like the name/repute of the company, socio-political and economic factors, and stock market news circulating in the news, social media posts/blogs by the economic analyst. The second type is the technical analysis method based on historical stock prices like opening, closing, and volume traded to predict the future stock price. [14] hypothesized that stock price prediction also requires understanding the market style. Considering the market style while predicting stock prices can yield good accuracy. A stock prediction model based on market-style clustering has been proposed.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One important approach was using news from media to predict stock prices. In this direction, [1] proposed two stock market price prediction techniques. The first technique is based on news from social media. News headlines were extracted from the web and sentiment analysis was performed with positive, negative and neutral labels. Labels were calculated by taking the difference of current day opening price and last day closing price. If the value was zero or negative, it was marked as 0; if it was positive, label was given as 1. The second technique was based on technical analysis using deep learning. A deep learning model was used to predict the stock prices on the next day. Data was collected comprising historical values, normalized using the min-max normalization technique. LSTM was used with a step size of 120 days with four hidden layers. Each of the layer comprises 75 nodes. LSTM can extract features from time series data and model long-term dependencies. After evaluation, the system got 97% accuracy.

[6] proposed the best fit model for 10 years of data stock exchange. The major objective was to determine a pattern of the stock market trend for historical data. Data was run through three different models. The first model was autoregressive integrated moving average (ARIMA), a widely used technique for time series data. It makes an association between data points and considers the difference in stock prices. Second model is simple moving average which is a variation of simple average i.e., succeeding values are also considered to calculate final predicted value. Third model is the holt-winters method, which considers historical data and seasonality. Features included opening price, closing price, high and low volumes. Predictions were made on day, month and year basis. ARIMA was used to make these predictions. As a concluding point, a low prediction error was observed for a simple moving average model. The second-best performance was given by holt-winters method. However, ARIMA couldn't perform well on 10 years of data.

[7] combined the effect of machine learning and sentiment analysis to propose a more robust approach for stock prediction. The proposed approach comprises both qualitative and quantitative analysis. Three modules were implemented i.e. machine learning, sentiment analysis and fuzzy logic. Each of the models was then ensembled into one recommender model. The extracted features were the opening price of the particular day, the lowest and highest prices, simple moving average estimates, exponential moving average estimates of opening/closing prices, and the exponential moving average of lowest and highest prices.

Time series analysis is important to forecast and predict the stock market price and trend. The time series analysis is useful in identifying trends, seasons, cycles and variations of stock and stock investors interested in short-term or long-term investment decision-making. Time series have many forecasting methods i.e. exponential smoothing, regression, and threshold but the authors in [2] proposed ARIMA algorithm for the stock market

flow and variation. For performance measurement and accuracy, RMSE and MAPE value was used.

The time series data is characterized by its dependence on time. The data changes and recurs seasonally with respect to time. The three main attributes of a time series data are i) linear time and cycle time ii) time points and time intervals, iii) sequential time, branch time and multi-angle time. The visualization of time series can be expressed using a calendar view and it is helpful to analyze the massive data flow w.r.t to time intervals[3].

Data visualization in various approaches is considered for prediction. [4] represented the data in a graphical and pictorial form which is easily understandable and effective for a large dataset. It's a powerful technique for analyzing and interpreting large and complex data. Common visualization techniques used were line graphs, bar charts, scatter plots and pie charts. These techniques can be more appropriate using different colors or sizes.

[5] presents a cross-reference to exchange-based stock trend (CREST) prediction approach. The objective is to analyze the Wipro limited stock registered in two stock exchanges: the national stock exchange (NSE) and the Bombay stock exchange (BSE). It was hypothesized that variations of stocks in one exchange could be used for predicting the trend in other exchanges based on LSTM. Various metrics such as RMSE, directional accuracy, precision, recall and F1-measure were used to measure the performance.

To predict the future closing price, artificial neural networks (ANN) and random forest (RF) were used [6]. ANN is one of the techniques from machine/ deep learning based on a network of perceptrons (neurons) that exchanges data with each other and learn fundamental trends/patterns from data using backpropagation and then generalize from data for future prediction. Random forest (RF) is an ensemble machine learning technique that uses multiple decision trees to conclude. RF can be used for regression as well as classification tasks. RMSE, MAPE and mean bias error (MBE) was used to evaluate the model performance of both models, and it was found that ANN performed better than Random Forest.

A convolutional neural network (CNN) and bi-directional LSTM has been proposed in [11]. In [12], CNN has been used to model stock prices. Various machine learning and LSTM based regression models have been evaluated in [13]. [15] analyzed the stock prices of Indian companies using various statistical, machine learning and deep learning approaches.

To conclude this section, deep learning has been intensively used for stock price prediction. Interested readers are directed to [10] for a survey on techniques based on deep learning. These techniques include deep neural networks, convolutional neural networks, long short-term memory models, hybrid attention networks, self-paced learning and wavelet. However, it is found

that the work requires consideration and one-on-one analysis of specific markets. Hence, this paper proposes a stock price prediction model for a market of a third-world country i.e. Karachi stock exchange (KSE).

III. METHODS

This research aims to analyze the stock market data, forecast using different models, and compare the results. The model will give a minimum error with the best model. Following are the steps for a proposed approach:

- *Data Collection:* We used KSE stock market data to analyze, visualize and do the forecasting
- *Data Preparation:* Analysis has been performed on the last 8 year stock data. Splitting the data into train and test data.
- *Model fitting:* ARIMA, Exponential Moving Average, CNN and LSTM were fit
- *Analysis:* Check the residuals and accuracy of each model
- *Forecast:* Do the forecasting using best model

A. DATA VISUALIZATION

We have compared stock market trends with single stock data to see an impact. After visualization, we found that KEL stock is not 80-90% of the times moving up and down according to overall stock market trend. We visualized the same sector stock and found that they are almost showing similar kind of up and down movement.

We also visualized the refinery and electric sector and found that they are moving opposite direction to each other almost 50% of the time. Time series decomposition algorithm is used to identify the trend and season of particular stock data. That can help investor to take the decision accordingly.

B. ARIMA MODEL

Box and Jenkins and ARIMA models were popularized by the ARIMA model and are often referred to as Box-Jenkins models. We have applied ARIMA model on 10 stock data and found that best ARIMA (p,d,q) model is (0,1,1) depending on AIC value. Forecasting is done on 95% confidence interval.

C. LONG-SHORT TERM MEMORY (LSTM) MODEL

LSTM is a powerful time series model that can be used to model sequential data. A sequential data can be used to learn the trends in data such that data is dependent on past data. LSTM can be used to model the data various time steps ahead into the future. An LSTM cell maintains states controlled through various gates to make predictions for long- and short-term data.

- *Cell state:* It represents the internal memory of the cell that can model both long-term and short-term dependencies.
- *Hidden State:* It stores the last output generated by the LSTM network. The state information is calculated based on the current input. The hidden state and cell state is used to predict future data such as stock market prices and temperature forecasting. Moreover, the hidden state is modelled to use long-term, short-term, or both to make the next prediction.
- *Input gate:* It controls how information from current input alters the cell state.
- *Forget Gate:* It controls how information from current input and previous cell state flows into the current cell state.
- *Output Gate:* It regulates how information from current cell state is written into the hidden state. LSTM can pick either short-term or long-term memories based on the output gate.

D. MULTIVARIATE TIME SERIES CNN MODEL

Convolutional neural network (CNN) model are used for image classification and designed to process the pixel data. That's why using it for predicting stock price is unusual. As our data of the stock pertains to 1D time series data, we adopted the "conv1d" function in the model that accepts a stock data and other properties such as the number of filters, and stride as input. We have applied the multivariate CNN Models that takes multiple parallel input time series and a single output time series. In our case we have four input series, which are open, high, low, and volume of the stocks and closing indicator is used as an output series. We split the data into samples and restructured it to three-dimensional, expected by a CNN model as an input. The first dimension denotes the samples, the second dimension is the number of time steps per sample and the last dimension accounts for the number of time series steps or features (i.e. 4). We have used only one convolutional layer of 32 filters with a kernel size of 3. The average pooling layer is applied to reduce the feature maps before the internal representation is flattened to one long vector. We choose the rectified linear unit (ReLU) as an activation function, and ADAM optimizer is used in the model with 100 epochs.

IV. RESULTS

The dataset and stock prices of KEL and NRL have been extracted from the Karachi stock exchange (KSE) [9]. The moving average of closing prices of refinery companies is shown in Fig 1. The figure highlights the moving average values based on last 10, 20 and 50 days compared to the actual closing price. It also highlights that the refinery companies have greater revenue in the third quarter(FY-Q3) of each year.

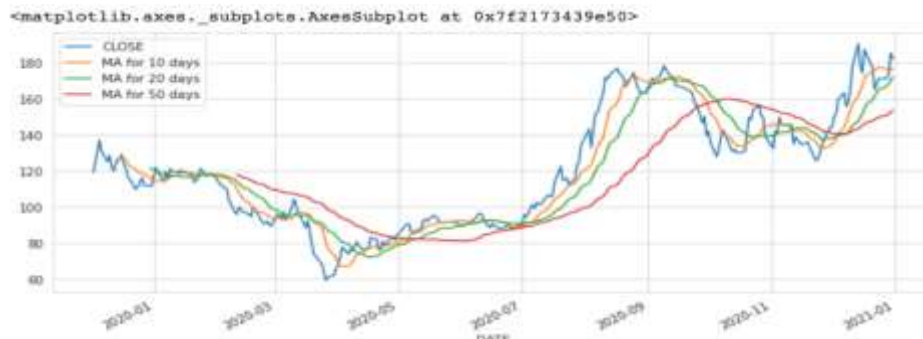


FIGURE 1: Moving average of refinery

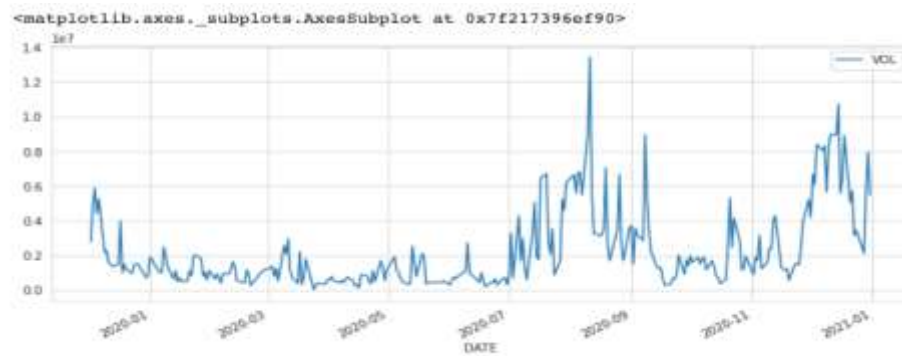


FIGURE 2: Share prices for the refinery company

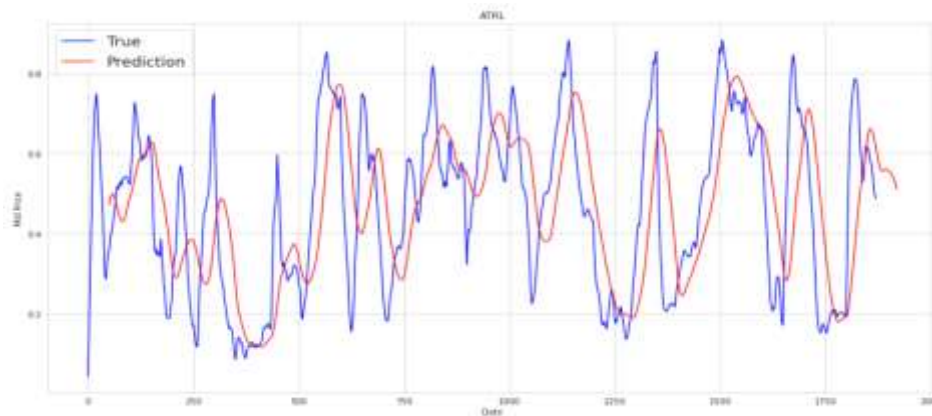


FIGURE 3: Simple average prediction for stock price

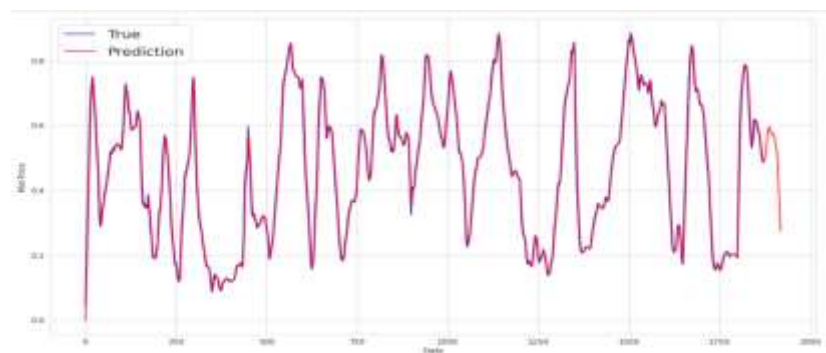


FIGURE 4: Moving average prediction for stock prices

From Fig. 2, it is found that the number of shares for a refinery company are most sold from July onwards i.e. the start of fiscal year(FY). From Fig. 3 we found that a simple average cannot predict correct values for stocks because it cannot accommodate historical data i.e. historical values have a greater impact on the next day, month and year's stock price. There is a difference between actual and predicted prices.

Figure 4 shows the exponential moving average that produces best stock market prediction results. The predicted values are very close to the actual price. Figure 5 shows the result obtained using LSTM. It produces the second-best results and considers all 10 years of historical data.

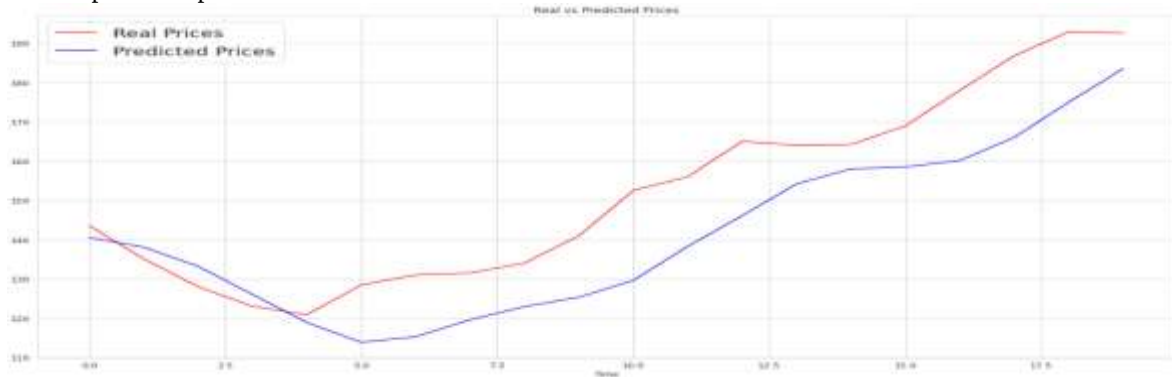


FIGURE 5: Prediction using LSTM



FIGURE 6: Decomposition to find stock trend

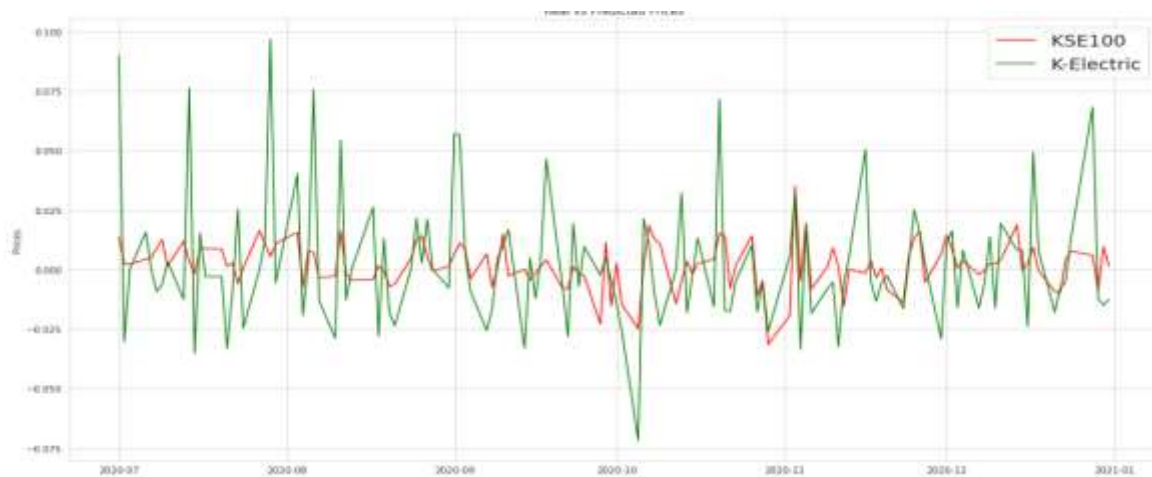


FIGURE 7: KEL stock price w.r.t to stock market trend

Figure 6 performs the decomposition to find the stock trend. Now, let's analyze KEL stock price w.r.t to the stock market trend of other companies listed in KSE100. It is shown in Fig. 7.

Figure 8 shows the sector Vs. sector comparison. KEL is moving opposite to refinery sector NRL. As can be seen, the refinery sector trend is almost opposite to the KEL trend.

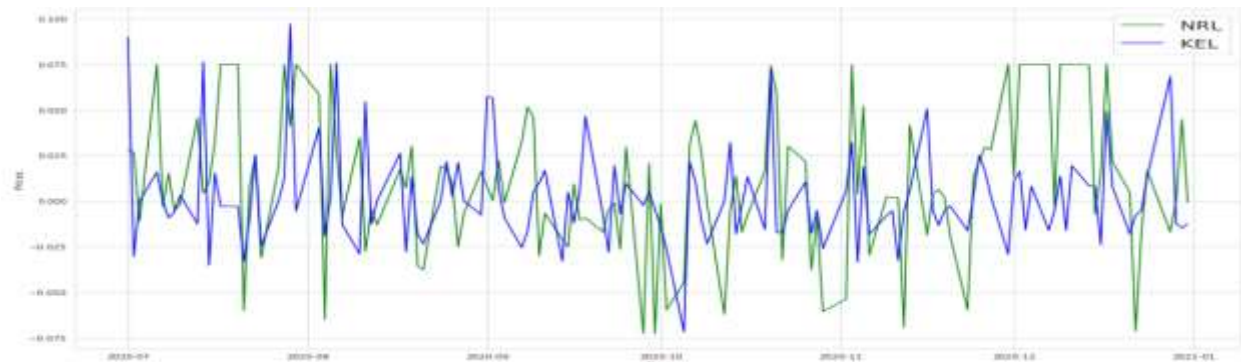


FIGURE 8: Sector wise comparison of KEL and NRL



FIGURE 9: ARIMA model for forecasting



FIGURE 10: Visualization of stock prices

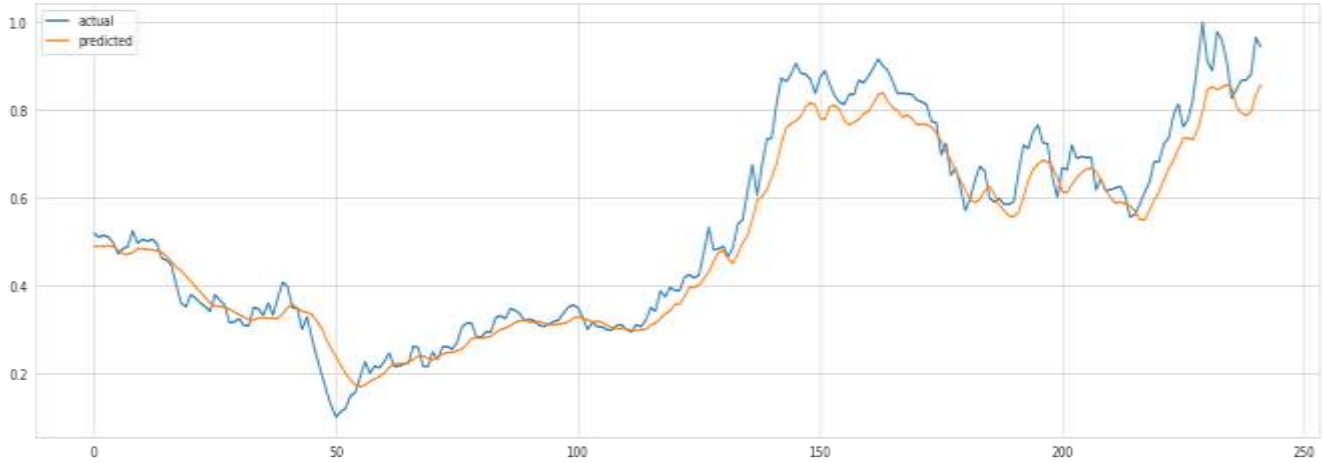


FIGURE 11: Forecasting using CNN

TABLE I
COMPARISON OF MODEL PERFORMANCE ON STOCK DATA

Stock	ARIMA			Exponential Moving Average			LSTM			CNN		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
KEL	0.98	0.85	0.26	0.017	0.018	0.052	3.47	4.66	10.91	0.0431	0.0305	0.6555
LUCK	132.12	106.37	0.17	0.020	0.020	0.05	24.56	33.00	73.58	0.0285	0.0195	0.5031
ATRL	34.94	29.39	0.24	0.01	0.011	0.029	8.10	10.88	32.65	0.0557	0.0424	0.6037
NRL	66.26	46.64	0.27	0.012	0.013	0.038	5.84	7.85	70.40	0.0273	0.0194	0.5960
TRG	22.28	16.35	0.36	0.002	0.001	0.004	12.76	17.09	37.31	0.0220	0.0162	0.7449
MEBL	21.70	18.10	0.25	0.010	0.011	0.029	10.88	14.62	85.14	0.0503	0.0438	0.6385

We then used ARIMA model for forecasting the future stock prices. The forecasting using CNN is shown in Fig. 11. The comparison of 6 stock data on ARIMA, exponential moving average, LSTM and CNN for various companies is shown in Table I.

V. CONCLUSION

The historical data for the period of five years from 2011 to 2021 were taken into account for analysis. RMSE, MAE and MAPE are used to identify the best model. Initially, a simple moving average was taken for one-year stock data, which seemed to be under-performed because the major impact of historical values was missing. Secondly, the exponential moving average was calculated using 10 years of data and performed best among all models. The RMSE for all predictions was found to be approximately 0.01 on average.

This work can be extended in various directions. For instance, the generative adversarial network (GAN) and attention-based models can be tested for the prediction of stock prices [13].

FUNDING STATEMENT

The author(s) received no specific funding for this study.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest to report regarding the present study.

REFERENCES

- [1] Agarwal, Harshit, et al. "Analysis and Prediction of Stock Market Trends Using Deep Learning." *Lecture Notes in Networks and Systems*, 2020, pp. 521–31. Crossref, doi:10.1007/978-981-15-3369-3_39.
- [2] B. Uma Devi, D.Sundar and Dr. P. Alli3 "An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50" *International Journal of Data Mining & Knowledge Management Process (IJDKP)* Vol.3, No.1, January 2013, pp 65-75. DOI : 10.5121/ijdkp.2013.3106
- [3] Yujie Fang, HuiXu and Jie Jiang. "A Survey of Time Series Data Visualization Research" *IOP Conference Series: Materials Science and Engineering*, 2020, doi:10.1088/1757-899X/782/2/022013
- [4] Matthew N. O. Sadiku, Adebawale E. Shadare, Sarhan M. Musa and Cajetan M. Akujuobi "Data Visualization" *International Journal of*

- [5] Ankit Thakkar, KinjalChaudhari. "CREST: Cross-Reference to Exchange-Based Stock Trend Prediction using Long Short-Term Memory" Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2019). 10.1016/j.procs.2020.03.328.
- [6] Kulkarni, Mugdha, et al. "Time Series Data Analysis for Stock Market Prediction." SSRN Electronic Journal, 2020. Crossref, doi:10.2139/ssrn.3563111.s.
- [7] Pathak, Ashish, and Nisha P. Shetty. "Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis." *Advances in Intelligent Systems and Computing*, 2018, pp. 595–603. Crossref, doi:10.1007/978-981-10-8055-5_53.
- [8] MeharVijh, DeekshaChandola, VinayAnandTikkiwal, Arun Kumar. "Stock Closing Price Prediction using Machine Learning Techniques" Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2019). 10.1016/j.procs.2020.03.326 10.1016/j.procs.2020.03.326
- [9] <https://www.ksestocks.com/MarketSummary>
- [10] Hu, Z., Zhao, Y. and Khushi, M., 2021. A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1), p.9
- [11] Lu, W., Li, J., Wang, J. and Qin, L., 2021. A CNN-BiLSTM-AM method for stock price prediction. *Neural Computing and Applications*, 33(10), pp.4741-4753.
- [12] Mehtab, S. and Sen, J., 2020. Stock price prediction using convolutional neural networks on a multivariate timeseries. *arXiv preprint arXiv:2001.09769*.
- [13] Mehtab, S., Sen, J. and Dutta, A., 2020, October. Stock price prediction using machine learning and LSTM-based deep learning models. In *Symposium on Machine Learning and Metaheuristics Algorithms, and Applications* (pp. 88-106). Springer, Singapore.
- [14] Li, X. and Wu, P., 2021. Stock price prediction incorporating market style clustering. *Cognitive Computation*, pp.1-18.
- [15] Mehtab, S. and Sen, J., 2020. A time series analysis-based stock price prediction using machine learning and deep learning models. *International Journal of Business Forecasting and Marketing Intelligence*, 6(4), pp.272-335.