STOCK MARKET PREDICTION DURING COVID-19 USING LSTM

Sectoral Analysis: Healthcare, Real Estate and IT-India

1.Ashutosh Singh, 2.Bhramari Verma, 3.Prof. Ajay Kr. Srivastava
[1]Student, [2]Student, [3]Professor
Department Of Information Technology,
Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM), Lucknow, India
1 mrashu.singh21@gmail.com , 2 bhramariverma@gmail.com , 3 ajaykrsrivastava.it@srmcem.ac.in

Abstract—

Air pollution is one of the major environmental issues that affect the health and well-being of people living in cities around the world. For this purpose, it is important that the Air Quality Index (AQI) is predicted with maximum accuracy. This research aims to propose a system for the prediction of the Air Quality Index using machine learning algorithms. Various algorithms are used for the prediction of the Air Quality Index. These include the use of the Random Forest algorithm, Support Vector Machine (SVM), Decision Tree, and Linear Regression. The system has been implemented using Python and a web application based on the Django framework. The experimental results show that the use of machine learning algorithms for the prediction of the Air Quality Index is quite effective.

Index Terms—

LSTM, COVID-19, stock market prediction, Indian stock market, sectoral analysis, time series forecasting

I. INTRODUCTION

In late 2019, the COVID-19 outbreak in Wuhan quickly turned into a worldwide crisis. It was not just a crisis for public health; it was also a crisis for the world economy. As governments froze all movement, the global supply chain simply buckled, leading to a near-total paralysis of international markets. Large economies such as China and America saw their growth slow down, and their stock markets also slowed down with huge price declines and fluctuations. India was also affected by the economic impact of the pandemic. While Healthcare and IT industries recorded positive growth because everybody needed medicine and technology, the Real Estate market was on the verge of dying. This is why we are unable to talk about the overall market anymore because we need to assess each market to know who is actually surviving.

This era of massive volatility has led to a major weakness of traditional forecasting models, which revealed they were not designed to account for the dynamic long-term dependencies of a global crash event. To fix this, our research utilizes Long Short-Term Memory (LSTM) Networks. What makes this model unique is its ability to remember patterns and filter out the market noise, which usually leads to the failure of traditional statistical models. Though it is known how the pandemic affected the market, little research has been done on this subject at a sectoral level using deep learning models. This research aims to use this model on all of India's major sectors, in order to determine what truly makes or breaks a sector. The goal of this research is, at the end of it all, for financial professionals to be able to measure risk in a world where economic laws are no longer applicable.

II. LITERATURE REVIEW

COVID-19 forced a massive, near-instantaneous recalibration of how we study market economics. Scholars witnessed a fundamental fracturing of how global markets function. If the early papers were a simple damage report, the more recent literature offers a narrative of total stability breakdown. The biological spread of the virus and the sudden death of market liquidity became one and the same. It was a massive fear contagion. The response of the equity markets was to break away from their conventional behaviors, which in turn resulted in a wave of abnormal returns, thereby wiping out even the most robust regional diversification strategies. [1],[2]. Lumping the entire market together is a massive mistake, as studies by He et al. [3] and Mazur et al. [4] clearly show. When the pandemic hit, we observed tech and healthcare gaining a massive boost but real estate and hospitality nearly died. This blow was experienced by India instantly [5].

The current research largely employs classical econometric models such as OLS, GARCH, and event study toolkits [6,7,8]. However, the nature of financial data during global crash is far too volatile for standard math, and full of hidden dependencies which the classical models cannot calculate. To fix this, computational finance is pivoting toward the fluid, adaptive power of deep learning [9]. In this regard, the Long Short-Term Memory (LSTM) model stands out as a much deeper and more reliable forecasting. Utilizing the architecture pioneered by Hochreiter and Schmidhuber [10], LSTMs bypass the vanishing gradient problem that kills older sequential models. While the ARIMA model is restricted in its approach, the LSTMs have the advantage of remembering long-term sequences, thus tracing the non-linear movements of the current financial market [11].

Despite a surge in pandemic market studies and machine learning, the two rarely overlap in the Indian context. While most market studies have stuck to the conventional formulas to track the market crash, others have resorted to the use of AI to make predictions about the market prices without focusing on the struggles of the industries. This is a major void in the market. The present study aims to fill this void by applying the LSTM framework to the analysis of the Healthcare, IT, and Real Estate industries, which form the backbone of the Indian economy.

III. METHODOLOGY

The proposed system utilizes a structured time series forecasting pipeline with a Long Short-Term Memory (LSTM) network for forecasting the stock prices. Our approach progresses from data cleaning, sequence building, to training the model and verifying the accuracy of the results. This is to ensure the LSTM model is able to capture the changing dependencies that characterize the financial crisis.

A. Data Collection

In order to develop a reliable model for prediction, we obtained sector-wise data (Healthcare, IT, and Real Estate) from the National Stock Exchange website. We archived this information under a daily market log, but specifically extracted closing prices for our model. We have done this because it is generally accepted that closing price is the most accurate anchor for what a stock is actually worth at the end of the day. Pandas is then used to parse the raw data into an easily understandable form. This is important so that the LSTM model can process the input correctly.

B. Data Processing

Before training the model, preprocessing was done on the data, keeping in mind consistency and stability during the learning process. We processed the closing prices into the numerical values required for the LSTM. Thereafter, Min-Max normalization was used to normalize the values within a range from 0 to 1. As the data was sequential in nature, considering the domain of stock markets, the data was split into training sets and testing sets with a ratio of 65:35 based on their occurrence order.

C. Time-Series Sequence Generation

We used a sliding window technique to train the model on how to identify trends over a period of time with a 150 days look-back period. Hence, the entire previous window of 150 days was taken into account before making a single prediction. The data was transformed to a three-dimensional data set (samples, time steps, features), where the number of features is one (closing price).

D. LSTM Model Architecture

For the predictive aspect of our research, we have developed a stacked Long Short-Term Memory (LSTM) architecture within the TensorFlow and Keras frameworks. Unlike traditional Recurrent Neural Networks (RNNs) employed in the processing of sequential information, LSTMs are particularly developed to avoid the vanishing gradient problem and, therefore, are capable of holding their state over long periods of time, which is an important feature to handle the volatile nature of the Indian market.

E. Model Training

We ran the training for 200 epochs with a batch size of 64. The aim here is to keep the Mean Squared Error (MSE) as low as possible. With the help of the learning rate of 0.01, the system will be fine-tuning its weight values. This was done to increase the degree of precision for stock prediction.

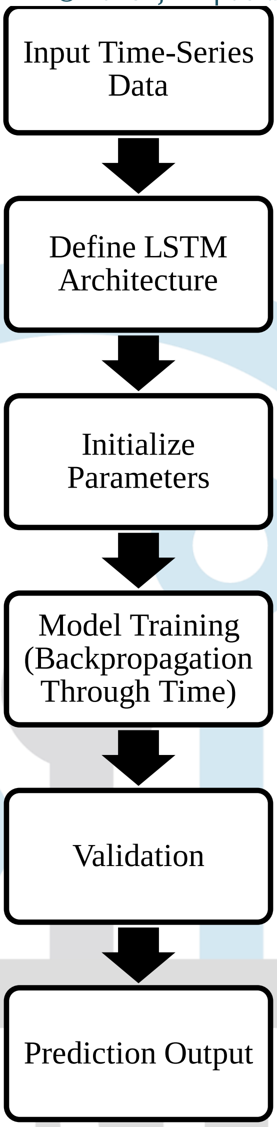

Fig. 1. Workflow illustrating the training process of the proposed LSTM model for stock market prediction

F. Performance Evaluation

In order to understand how effectively the model performed in the real world, we have made use of the Root Mean Square Error (RMSE) measure. This is because the model is focused on reducing the variance with respect to the output during the training phase. Hence, RMSE is a more realistic measure since it calculates the error based on the stock prices. This makes it an invaluable tool in understanding the extent of errors made by the model in its output in a way that is both statistically accurate and easy to understand.

Nevertheless, since the LSTM model has been trained on scaled values, its output is essentially in the form of abstract values between 0 and 1. In order for this output to have some real-world meaning for an economic study, it was necessary for us to re-translate it by applying an inverse Min-Max function in order for it to return to its original form in terms of Rupees. This essentially took the study from the purely mathematical world and into the real world, allowing us to see exactly how the model behaved reflected by its output in terms of actual points during the volatile changes in the stock prices of the IT, Healthcare, and Real Estate sectors.

G. Equation

Let

- n be the total number of observations
- y_i be the actual stock price at time step i
- $\hat{y}_i$ be the predicted stock price at time step i
- $(y_i - \hat{y}_i)$ be the prediction error

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

The above equation calculates the prediction error of the actual price versus predicted price of the stocks through the Root Mean Square Error (RMSE).

The above Equation (1) computes the mean of squared differences between the actual values and predicted values and then taking their square root. In doing so, this will give a high penalty to a greater error value while keeping the error in the same units as the stock price.

IV. RESULTS AND DISCUSSION

The proposed model has a high capability to grasp the temporal dependencies in the stock price values. Even after the heavy fluctuations in the stocks exchange due to the pandemic, the predictions have followed the actual stock prices quite well. This shows the power of the LSTM model in dealing with complex stock movements.

However, we did see minor offsets during extreme market spikes where the pandemic's unpredictability briefly pushed prices outside our predicted range.

Our model worked well for the most part, but struggled to track the fastest price changes. Nevertheless, the overall results were good.

Sector	Training RMSE	Testing RMSE	Observations
Healthcare	0.022879248528941378	0.029677001495567852	Low error; stable predictions
Real Estate	302.51750360965485	392.84481187556787	Higher error; volatile sector
IT	24016.764047035864	32570.21542768334	High error; large price scale

Table. 1. RMSE-Based Performance Assessment Of LSTM Model For Sectoral Stock Prediction

From Table 1, it is evident that the predicted LSTM model produces the least error in predicting the Healthcare industry. However, both the real estate and IT sectors have RMSE values that are greater, indicating that both of them are relatively volatile compared to other industries with respect to the price pattern during the current pandemic. The increased error seen in the case of the IT sector can also be associated with the high value of its price level.

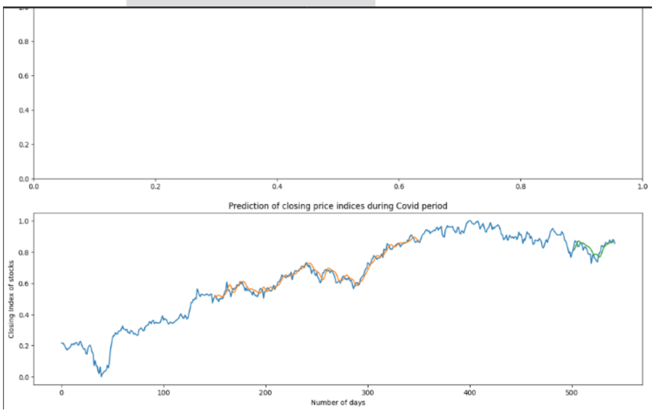

Fig. 2. Comparison of actual vs predicted stock prices for the Healthcare sector using the LSTM model

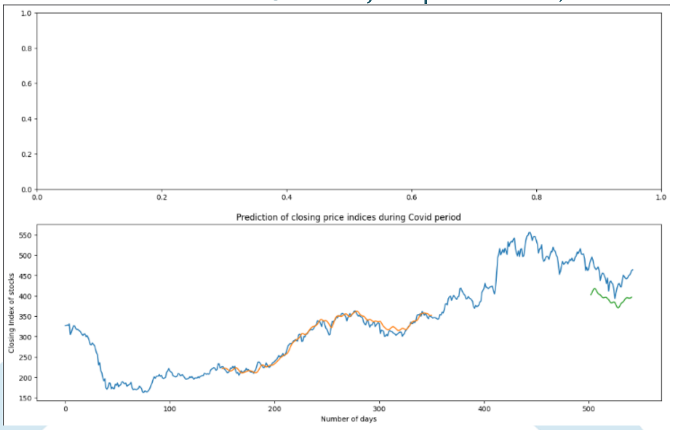

Fig. 3. Comparison of actual vs predicted stock prices for the Real-Estate sector using the LSTM model

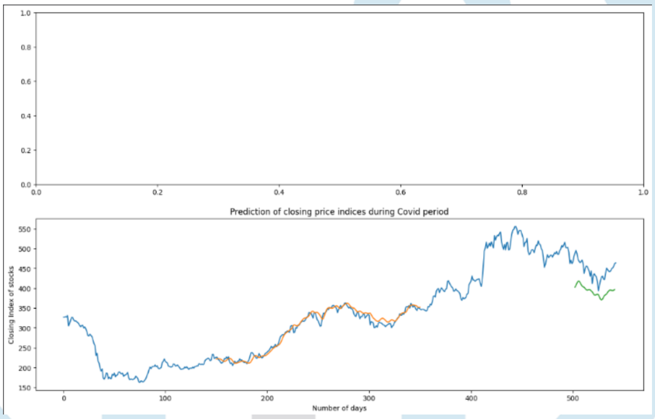

Fig. 4. Comparison of actual vs predicted stock prices for the IT sector using the LSTM model

V. CONCLUSION AND FUTURE WORK

In this present study, a stacked LSTM network was used to predict the stock market trends for India's Healthcare, IT, and Real Estate sectors during the pandemic period. The complex non-linear patterns within the data could be captured with the help of historical data related to closing prices of shares fed into the network. Our results demonstrate how our model tracks changes very quickly, even in the unprecedented times of the COVID-19 pandemic. The RMSE scores back up the model's performance which reveals that the system handled most sectors well other than the IT sector where the error rate was slightly increased. One of the major shortcomings of our current model lies in the fact that it does not consider any other data sources except for internal data into consideration. Right now, the model ignores outside factors like policy changes and market mood. Our next step is to bridge this gap by layering in more diverse information from the real world and incorporate real-world sentiments to improve accuracy.

REFERENCES

[1] C.-L. Chang and Q. Cai, "Stock return anomalies identification during the COVID-19 pandemic," *Economic Analysis and Policy*, vol. 79, pp. 168–183, 2023.

[2] N. James, M. Menzies, and G. A. Gottwald, "On financial market correlation structures and diversification benefits across and within equity sectors," *arXiv preprint arXiv:2202.10623*, 2022.

[3] Q. He, J. Liu, S. Wang, and J. Yu, "The impact of COVID-19 on stock markets," *Economic and Political Studies*, vol. 8, no. 3, pp. 275–288, 2020.

[4] M. Mazur, M. Dang, and M. Vega, "COVID-19 and the March 2020 stock market crash. Evidence from S&P1500," *Finance Research Letters*, vol. 38, p. 101690, 2021.

[5] S. Singh and Y. Neog, "Contagion effect of COVID-19 outbreak: Another recipe for disaster on Indian economy," *Public Affairs*, vol. 20, no. 4, p. e2171, 2020.

[6] S. Liu, Y. Wang, D. He, and C. Wang, “The COVID-19 pandemic and the global stock market: Evidence from an event study,” *Finance Research Letters*, vol. 36, p. 101605, 2020.

[7] S. Baig, H. A. Butt, O. Haroon, and S. A. Rizvi, “Deaths, panic, lockdowns and US equity markets: The case of COVID-19 pandemic,” *Finance Research Letters*, vol. 38, p. 101701, 2021.

[8] T. Bollerslev, “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[9] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[10] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] S. Siami-Namini, N. Tavakoli, and A. S. Namin, “A comparison of ARIMA and LSTM in forecasting time series,” in *Proc. 17th IEEE Int. Conf. Machine Learning and Applications (ICMLA)*, 2018, pp. 1394–1401