A Comparative Analysis of the performance of the LSTM Network and ARIMA Model for the forecasting of a non-stationary financial time-series

Amitesh Anand Pandey

Lotus Valley International School, Gurugram, India

---***--

Abstract- *Forecasting the stock-market is an age-old requirement in an investor's tool-kit for successful investments. The stock market is affected by various factors that include (but are not limited to) sentiments, demand, supply, policy, political climate, among others. This makes predicting the dynamics of the stock market an incredibly challenging task. It is a combination of both random and deterministic phenomena. For this task, a variety of deep-learning and stochastic models have been developed. This paper attempts to compare two of the most used predictive models: LSTM-RNN and ARIMA over data collected from State Bank of India's historical stock price.*

Key Words: *Stock-Market, Prediction, Forecasting, Deep-Learning, ARIMA, LSTM*

1. INTRODUCTION

It has been believed for quite some time now that the efficient-market hypothesis (EMH) is not entirely correct [1]. Stock prices of most (if not all) public organisations and companies fluctuate regularly. These fluctuations can be caused by various factors outside the control of the company's performance like Public Sector Undertakings, GDP trends of the country, Foreign Exchange Rate, public policies and Oil Prices [2]. Due to this, neural networks have proved to be an excellent method to model and analyse financial time-series, as they are able to learn and process various factors by themselves [3]. On the other hand, stochastic models like ARIMA (Auto-Regression Integrated Moving Average) are also a viable option for time-series that display stationarity (in that, they are independent of the time at which they are observed). For forecasting financial time-series using LSTM-RNN (Recurrent Neural Network), this paper utilises the Keras library. For proper implementation of the ARIMA model, we use the work-flow prescribed by M. Angadi, A. Kulkarni, in [4].

2. RELEVANT LITERATURE

A variety of papers on individual forecasting models like LSTM and ARIMA already exist.

Authors of [5] provide evidence for the practicality and effectiveness of the LSTM model compared to SVMs and Backpropagation in forecasting time-series data.

In [6], the authors conclusively assert the accuracy and implications of using the ARIMA model for stock price prediction. This is a useful work that renders the absolute importance of stochastic processes in the market.

In [7], the authors provide a detailed account of the multitude of machine-learning models ranging from SVMs to LSTMs that can be useful in our endeavour.

3. METHODOLIGIES AND WORK-PROCESS

Even though a variety of APIs give the feature of importing historical stock prices and data in a CSV format, this paper utilises Yahoo Finance's historical data collection for reliable results. This paper uses comma-separated value (CSV) data of State Bank of India's stock price from 13-03-2015 to today for both the ARIMA and LSTM model. This data consists of any stock's seven primary characteristics that include Volume, Opening Price and Closing Price, among others. For simplicity, and to also limit the scope of our models to one-dimensional space, we will train our models on only the closing price of SBI's stock.

3.1 Preprocessing of Data

As in any dataset, our dataset consists of a variety of abnormalities which may result in Linear Algebra errors for ARIMA and exploding gradients for LSTM. For this, we drop all the null-values from our dataset and convert it into a simple array of the stock's closing prices. This technique is widely used and can also be seen in [8]. Other than cleaning the data for proper training, to make the process efficient, we also normalise, denoise and transform the data. This also helps us avoid erroneous output.

3.2 Check for Stationarity

While stationarity is not necessary for our LSTM model, it is a requirement (and a limitation) of the ARIMA model. To check for stationarity, we use the wide-spread Dickey-Fuller Test. On running the significance test on our data, we find that the data is indeed non-stationary.

```
Results of dickey fuller test
Test Statistics              -2.036841
p-value                       0.270657
No. of lags used              7.000000
Number of observations used  1454.000000
critical value (1%)          -3.434855
critical value (5%)          -2.863530
critical value (10%)         -2.567829
dtype: float64
```

Fig -1: Results of the Dickey-Fuller Test

3.3 Seasonal Decomposition

Since our data was not appropriate in its original form for the ARIMA model, we perform a multiplicative seasonal decomposition for the best results. This coerces our data to become non-seasonal.
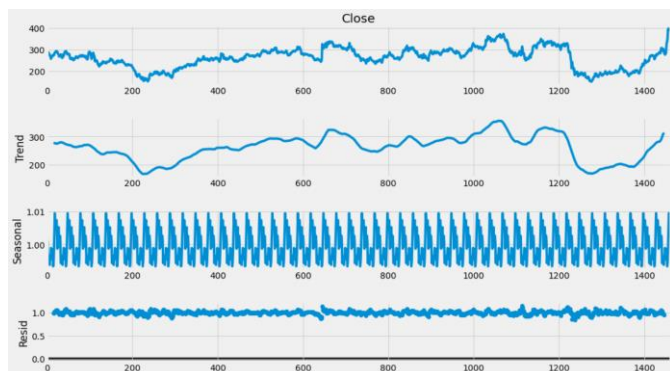

Fig -2: Results of Seasonal Decomposition

3.4 Feature Scaling of Data

Large values of the stock-price (as in the case of SBI's) may cause inefficiency and even instability in our LSTM model's weights. Due to this, we scale the data-values on a scale of 0 to 1 inclusive. This helps the model render our nodes quickly, making the network overall efficient.

3.5 Training the ARIMA Model

We use the ADFTest for an optimal d-value, out of the three required values: p(number of lag observations), d(differencing-degree) and q(width of the moving-average window) values for an ARIMA model [9]. Apart from that, we sample the time-series over a frequency m = 1. This was experimentally concluded to be the most satisfying output. With our aforementioned seasonal decomposition's help, we are allowed to put the seasonality boolean as False.

```
[[[0.70999994]
  [0.69113633]
  [0.76340903]
  ...
  [0.54522722]
  [0.55931816]
  [0.58659088]]

 [[0.69113633]
  [0.76340903]
  [0.7638636 ]
  ...
  [0.55931816]
  [0.58659088]
  [0.61749992]]

 [[0.76340903]
  [0.7638636 ]
  [0.74818173]
  ...
  [0.58659088]
  [0.61749992]
  [0.61568176]]]
```

Fig -3: Scaled Data Array

3.6 Evaluating the ARIMA Model

We used the SMAPE (Symmetric Mean Absolute Percentage Error) metric for the evaluation of our model. The reason for this is that it is widely considered to be a more accurate representation of the accuracy of stochastic models like ours than MSE, RMSE etc. [10] The error in our ARIMA model is 45%, which shows how inaccurate its results are. This is because of various reasons, including the non-stationarity of our data.

$$\text{SMAPE} = \frac{1}{n} \sum_{t=1}^{n} \frac{|F_t - A_t|}{(A_t + F_t)/2}$$

Fig -4: Expression for SMAPE

3.7 Training the LSTM Model

The LSTM model works on 3 basic gates. These are namely the Input, Output and Forget gates. The input gate serves the purpose of adding input values to the cell-state. This is carried out by the input gate in a 3-step process that involves the regulation of values using a sigmoid (or ReLu) function, the creation of a bounded array that ranges from -1 to 1 using the hyperbolic tangent function, and the addition of the regulation created in the first step to the array. We have used 100 epochs, with a batch-size of 32, to train the model. The model utilises common Mean-Squared-Error as the loss function, and Adam-Algorithm as our optimizer. We use

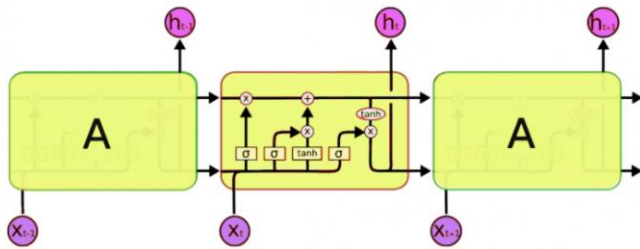Adam because it is known to be effective with non-stationary data[11].

Fig -5: Architecture of LSTM (*source: colah.github.io*)

3.8 Evaluating the LSTM Model

Our LSTM Model proves to be very useful in the modelling of stock prices. It has an accuracy of about 89%, which is an incredible improvement from the ARIMA model, which failed despite a confidence-interval of 95%. A variety of reasons caused this increment, and one is ARIMA's limitation itself when it comes to the analysis of non-stationary data. Furthermore, our seasonal decomposition method may have also had a vital role in the ARIMA model's poor performance.

4. Graphical Results

We provide the graphical results for the predictions made by the two models. This is for greater intuition about the polarity in the effectiveness of the two models.

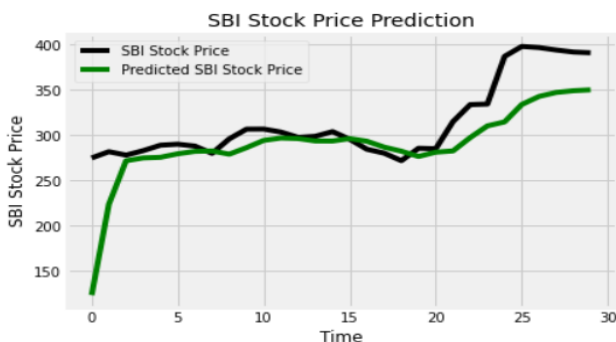
Fig -5: Prediction of the ARIMA Model (Scaled)

Fig -6: Predictions of the LSTM Model

5. CONCLUSIONS

After a detailed analysis of the error-graphs of both the models and a clear depiction through the forecast graphs, we can with reasonable certainty conclude that for non-stationary time-series, LSTM networks perform significantly better than other stochastic models like ARIMA. This gives us an insight into the further scope of Neural Networks' application in forecasting stock-prices as most if not all, trends in the financial market are non-stationary and non-seasonal.

REFERENCES

[1] Shostak, F. In defense of fundamental analysis: A critique of the efficient market hypothesis. Rev Austrian Econ 10, 27–45(1997). https://doi.org/10.1007/BF02538483

[2] Alam, S., Miah, M., & Karim, M. (2016). Analysis on Factors that Affect Stock Prices: A Study on Listed Cement Companies at Dhaka Stock Exchange. Research Journal of Finance and Accounting, 93-113

[3] H. N. Balakrishnan, A. Kathpalia, S. Saha, and N. Nagaraj, "Chaosnet: A chaos based artificial neural network architecture for classification," Chaos 29, 113125 (2019).

[4] Mahantesh C. Angadi, Amogh P. Kulkarni, "Time Series Data Analysis for Stock Market Prediction using Data Mining Techniques with R"; International Journal of Advanced Research in Computer Science

[5] D. Mahendra Reddy et al. "Stock Market Analysis using LSTM in Deep Learning"; International Journal of Engineering Research & Technology

[6] Nikhil Jain et al. "Stock Exchange Forecast Using ARIMA"; International Research Journal of Engineering and Technology

[7] V. H. Shah, "Machine learning techniques for stock prediction," Foundations of Machine Learning— Spring, vol. 1, no. 1, pp. 6–12, 2007

[8] Mrs. Nivethitha et al. "Future Stock Price Prediction using LSTM Machine Learning Algorithm." International Research Journal of Engineering and Technology

[9] Paul, J. et al. "Selection of Best ARIMA Model for Forecasting Average Daily Share Price Index of Pharmaceutical Companies in Bangladesh: A Case Study on Square Pharmaceutical Ltd." *Global Journal of Management and Business Research* 13 (2013): n. pag

[10] Sungil Kim, Heeyoung Kim, A new metric of absolute percentage error for intermittent demand forecasts, International Journal of Forecasting, Volume 32, Issue 3, 2016, Pages 669-679, ISSN 0169-2070, https://doi.org/10.1016/j.ijforecast.2015.12.003

[11] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In International Conference on Learning Representations (ICLR), 2015.