INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

IJARIIT

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 3 - V7I3-1986)
Available online at: https://www.ijariit.com

Stock Forecasting and Analysis using LSTM and Prophet

Sidharth S.
sidharthsnl83@gmail.com
Marian Engineering College, Kazhakkoottam, Kerala

Sambhu S. S.
sambhusubramannian@gmail.com
Marian Engineering College, Kazhakkoottam, Kerala

Vishnu Jayakumar
vishnujk95@gmail.com
Marian Engineering College, Kazhakkoottam, Kerala

Buddha Prabodh J. S.
buddhaprabodh224@gmail.com
Marian Engineering College, Kazhakkoottam, Kerala

ABSTRACT

Trading is tough. Part-time folks trading from their home are up against professionals. The professionals are really great at taking the money. Where there is potential reward, there is potential risk. The results equity curve might only show the reward side of the equation, but risk is always there. So it is a wise idea to implement Artificial intelligence and Machine learning technologies into this market which can disrupt the current scenario. In this study we use LSTM and PROPHET to predict the future stock prices. The forecast models are then compared to the actual data to measure their performance against each other. The result shows that Prophet generally outperforms LSTM.

Keywords: *Stock Market, Stock Forecasting, LSTM, Prophet.*

1. INTRODUCTION

When companies reach a certain state of growth, a huge amount of extra capital will be required to further expansion. Because of which the companies will do initial public offering (IOP), it is the process of offering the shares of a private corporation/company to the public in a new stock issuance [1]. This allows the company to generate capital from public investors also allows the public investors to participate in the offering.

The place where the shares of the companies who have done their IPO is listed is called as the stock market. Here people can buy or sell the shares of publically listed companies. A trader can trade in stock market only through a registered intermediary known as the stock broker. The buying and selling of shares take place through electronic medium [2].

The price of these listed shares may vary according to the status, strategies and the revenue and sales report of the company. These causes rise and fall in the share prices. This is an opportunity for traders to generate better profits.

Stock market is so important for an economy since it is considered as one of the indicator of the status of the economy of the country. If the participation on the stock market is increased, then it can benefit the participating traders as well as the whole economy. But to be successful in the market, one needs to be an expert in predicting the future stock prices. Here comes the problem that 90 percent of the new traders coming in to the market face a big loss because they might not be as good in predicting the future stock prices. Here we can make use of machine learning techniques to predict the future stock prices so that it can be used as a reference for the traders while taking trade decisions.

Here in this work, we predict the future stock prices using neural network model LSTM and additive regressive model Prophet. The models are compared using several performance parameters and the best one is used in making a platform using python. Python streamlit as well as html is used for the making of the platform.

1.1. Existing System

Algorithmic trading is used by traders to execute trades with the help of technology. Algorithmic trading uses a computer program that follows a defines set of instructions to place a trade [3]. In algorithm trading the decisions are made by just using the conditions that are prescribed in the algorithm. There is no innovative use of modern machine learning techniques or any kind of Neural networks involve. Some of the disadvantage of algorithm trading are listed below

- Algorithm trading is not always accurate and it can trigger many orders which in turn can lead to dramatic rise or fall in the stock prices.
- Since it requires both programming as well as market knowledge it can be done only those people who have the required knowledge and expertise.
- Another limitation is algorithmic trading is not universal and one algorithm cannot be applied to every situation and because of that constant monitoring is required [4].

2. DATASET PROPERTY

There exist different types of data. Like cross-sectional data, panel data and times series for some. Understanding what type of data is needed to solve the problem and what type of data can be obtained from available sources is important for formulating the right methodology for analysis.

Cross-sectional data is obtained by taking observations from multiple individuals at the same point in time, time does not play any significant role in these types of data [5]. Data observed over multiple entities over multiple points in time we get panel data (longitudinal data). Whereas time series data is made up of quantitative observations on one or more measurable characteristics of an individual entity and taken at multiple points in time. Time series data is typically characterized by several interesting internal structures such as trend, seasonality, stationarity, autocorrelation, and so on [5].

In this paper, we look at the closing stock prices of several corporations throughout time intervals.

2.1. Time series data

Time series data is a collection of observations obtained through repeated mesurements over time [6].When plotted on a graph, one of the axis will be always time.

Time series data have many implication in our daily life. As our world get extreamely instrumented, sensors and systems are constantly producing streams of time series data. Those data has numerous applications across various industries.

	Date	Open	High	Low	Close	Adj Close	Volume
1595	2021-05-04T00:00:00+05:30	131.1900	131.4900	126.7000	127.8500	127.6332	137564700
1596	2021-05-05T00:00:00+05:30	129.2000	130.4500	127.9700	128.1000	127.8828	84000900
1597	2021-05-06T00:00:00+05:30	127.8900	129.7500	127.1300	129.7400	129.5200	78128300
1598	2021-05-07T00:00:00+05:30	130.8500	131.2600	129.4800	130.2100	130.2100	78892700
1599	2021-05-10T00:00:00+05:30	129.4100	129.5400	126.8100	126.8500	126.8500	87808400

Fig.1. Raw dataset

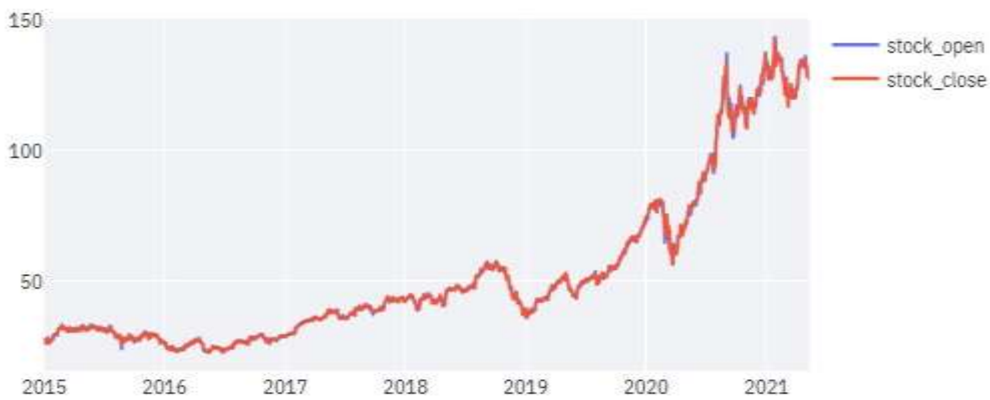

Fig.2. Plotted graph

There are mainly two kinds of time series data [6].They are:

*i.*Measurements gathered at regular time intervals (metrics)

*ii.*Measurements gathered at irregular time intervals (events)

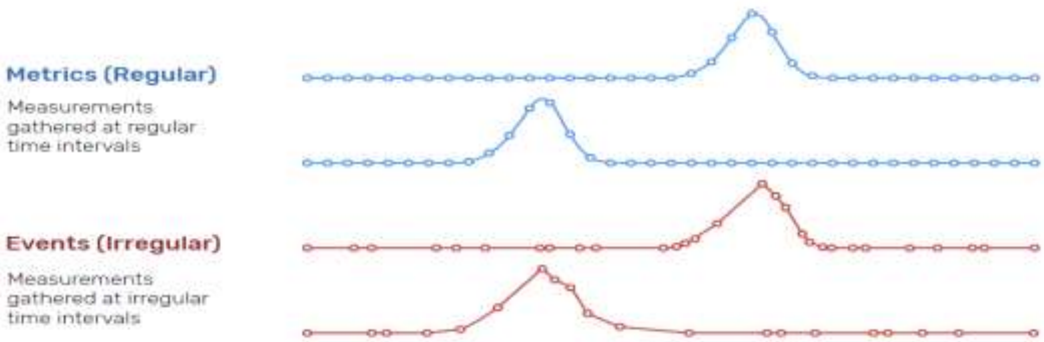

Fig.3. Two kinds of time series data

3. FORECASTING MODEL

3.1. Forecasting using LSTM

LSTM is a special type of RNN. These networks are adept at recognizing long-term dependencies. Hochreiter and Schmidhuber first announced it in 1997. These networks are clearly built to avoid the long-term reliance problem, yet they are accustomed to recalling information for lengthy periods of time. The LSTM cell is depicted in Fig 3 shows a pictorial representation of LSTM cell [7].

In comparison to other neural networks, LSTM has a unique structure [8]. Unlike traditional RNNs, which have a single neural network layer with a feedback loop, LSTMs have a memory block or cells instead of a single neural network layer. Each cell or block contains three gates, and the flow of data information via the cells is regulated by the cell state.

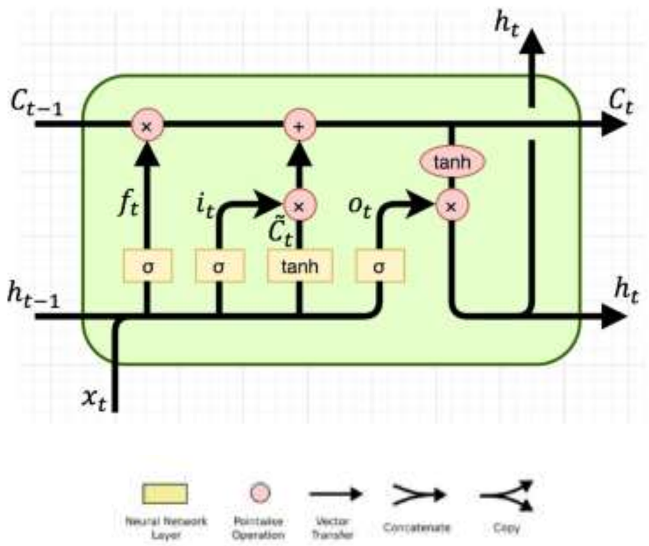

Fig.4. Structure of LSTM

Here C_{t-1}: old cells state, C_t: present cell state, h_{t-1}: Output of previous cell, h_t: Output of present cell, i_t: Input gate layer, f_t:forget gate layer, O_t : Output sigmoid gate layer.

In the figure Fig.4. horizontal line passing through the top of the diagram is known as cell state (C_{t-1}, C_t). It acts like a conveyor belt that runs over the entire network. It carries the information from the previous cell to the present and so on. The decision for storing information in cell state is taken by forget gate layer (f_t) which is also known as sigmoid layer. The output from forget gate is added to cell state using a point-wise multiplication operation. The input gate is made up of two layers: a sigmoid layer (i_t) and a tanh layer. These two are combined in the cell state by the input gate. The new values formed by the tanh layer are represented by C_t. A point-wise multiplication of sigmoid gate O_t and tanh produces output (h_t) [9].

The LSTM is capable of determining how long old information should be retained, when to remember and forget, and how to connect old memory with new input. This historical data is used to identify existing data patterns and use them to forecast future events.
The following are the steps involved in LSTM implementation:

- Importing libraries – We'll be working with a number of libraries, which we'll have to install first before importing into our environment.
- Loading the dataset – We gather the dataset from Yahoo Finance and convert it to csv format, which contains the OHLC values, which we then import into our working environment.
- Target Variable – The closing price will be our goal variable for forecasting.
- Data pre-processing – Sklearn has a preprocessing module that allows us to scale our data before fitting it into our model. Then the data is plotted.

3.2. Forecasting using PROPHET

High-quality forecasts have always been difficult to come by. As a result, there was a serious scarcity of analysts capable of producing projections with the precision needed to guide corporate decisions. The Facebook Core Data Science team created Prophet, a forecasting library for Python and R, which they open-sourced in 2017 to address this supply gap and make scaling forecasting significantly easier.

Prophet was created with the goal of "making it easier for professionals and non-experts to develop high-quality forecasts that keep up with demand." Prophet can create trustworthy and resilient forecasts with relatively little manual work (often outperforming other standard forecasting techniques), while also allowing domain knowledge to be applied via easily-interpretable parameter sets [10].
PROPHET is a time series data forecasting technique. It aspires to be able to forecast 'at scale,' which means PROPHET intends to be the automated forecasting tool that allows analysts of any background or with little to (potentially) no prior understanding of forecasting to forecast well. PROPHET "works best with time series with strong seasonal effects and multiple seasons of historical data, and is robust to outliers and fluctuations in the trend," according to Facebook. Our data is not seasonal in this example, but it

performs well. And this is where PROPHET comes in handy. Because of its autonomous nature, it can handle time series data with significant changes, so analysts don't have to worry about their data not being acceptable for forecasting with PROPHET30.

PROPHET is a simple to use programme. Analysts must prepare the dataset and generate a data frame with two columns: a 'ds' column for the date stamp (in date-time format) and a 'y' column for the forecasted measurement that must be in numerical values.Then, using the Prophet() class, analysts must generate an object with the data frame fitted into it. Analysts can then select the forecasting period they want to work with and begin forecasting. The forecasting result will have numerous columns, with the 'ds' and 'yhat' columns being the ones analysts want to look at. 'yhat' is a column in the historical data frame that contains the predicted outcome of 'y'. The variables 'ds' and 'yhat' can be plotted to show characteristics like future trend or seasonality [10].

Prophet Forecasting Model is an Additive Regressive Model

$$y(t) = g(t) + h(t) + s(t) + et$$

y(t) — Additive Regressive Model
g(t) — Trend Factor
h(t) — Holiday component
s(t) — Seasonality Component
et — Error term

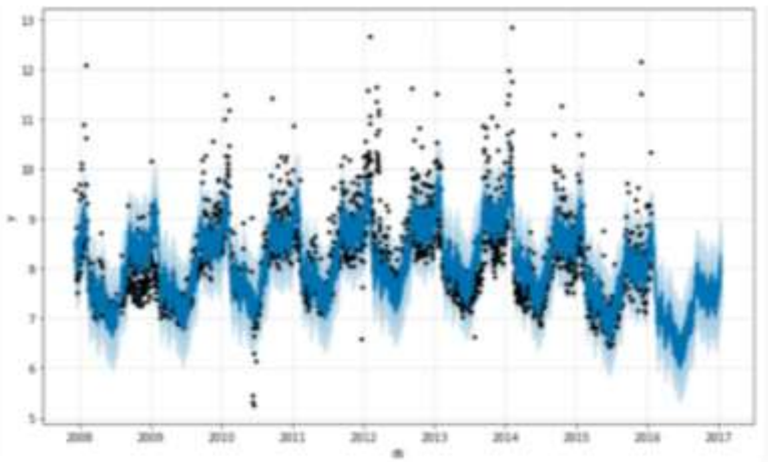

Fig.5. Example Graph using Prophet

It implements two possibles trend models and the trend factor can be adjusted using two models:
a) Logistic growth model — It handles non-linear growth with saturation.

Non-linear growth with saturation — Growth starts out nearly exponential (geometric), then slows to linear (arithmetic) as saturation sets in, and finally stops at maturity.
x_0 — X-value of sigmoid's point
L — Curve's Maximum value
k — Logistic growth rate or steepness of the curve
m = Prophet(growth='logistic')
m.fit(df)

b) Piecewise Linear Model — It's a simple linear model modification that's really handy. A single linear model may not provide a sufficient explanation or description . Breakpoints are the values of x when the slope changes . The value of breakpoints may or may not be known before the study; when unknown, it must be calculated

By default Prophet uses linear model for its growth.

We can fine-tune these parameters (trend components) in our prophet model by adjusting the breakpoints (also known as changepoints) and the total CAP (market size or capacity value) — the total CAP must be supplied for each row in a data frame and is not required to be constant. CAP can be an escalating sequence if the market size is expanding.

4.RESULTS

4.1 Output of LSTM

The forecasted data for the company "Apple" using LSTM is depicted in Fig.5.

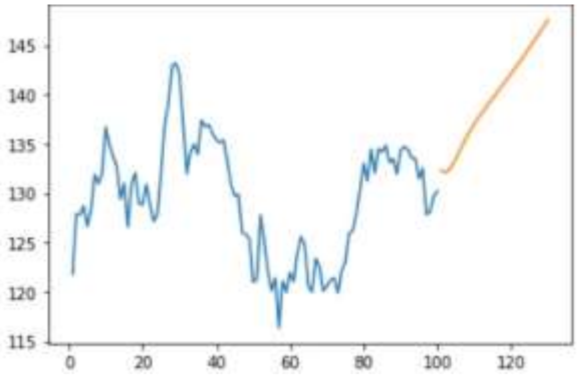

Fig.6. Forecasted Graph using LSTM

4.2 Output of PROPHET

The forecasted data for the company "Apple" using Prophet is depicted in Fig.7.

	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive_ter
1978	2022-06-03T00:00:00+05…	194.9227	181.1789	202.9944	185.8685	203.7883	-2.6130	
1979	2022-06-04T00:00:00+05…	195.0673	177.8136	200.1046	186.0044	203.9895	-5.5636	
1980	2022-06-05T00:00:00+05…	195.2120	177.9223	200.8291	186.1215	204.1985	-5.4256	
1981	2022-06-06T00:00:00+05…	195.3566	181.8084	203.9475	186.2476	204.4074	-2.0854	
1982	2022-06-07T00:00:00+05…	195.5013	182.3084	204.8927	186.3817	204.6126	-1.8277	

Forecast plot for 1 years

Fig.7.1. Forecasted Data in Tabular Form

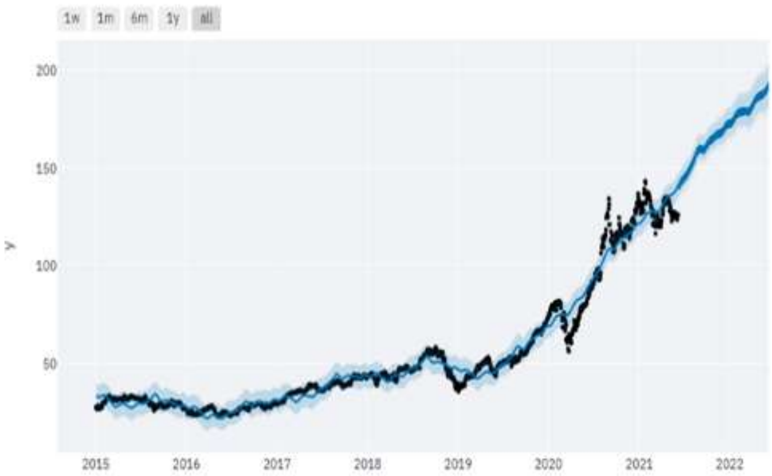

Fig.7.2. Forecasted Data in Graphical Form

5. ACCURACY ANALYSIS

Error metrics enable us to track the efficiency and accuracy through various metrics as shown below–

5.1. Mean Square Error (MSE)

The loss function for least squares regression is MSE, which is the average of the squared error. It is the sum of the square of the difference between the anticipated and actual target variables, divided by the number of data points, over all data points [11].

$$\sum_{i=1}^{n} \frac{\left(w^T x(i) - y(i)\right)^2}{n}$$

5.2. Root Mean Squared Error (RMSE)

The square root of MSE is RMSE. MSE is calculated in units equal to the target variable's square, whereas RMSE is calculated in the same units as the target variable [11]. MSE, like the squared loss function from which it comes, essentially penalises larger errors more severely due to its construction.

5.3. Mean Absolute Error (MAE)

MAE is a statistic that assesses the average magnitude of mistakes in a set of forecasts without taking into account their direction. It's the average of the absolute differences between forecast and actual observation over the test sample, where all individual deviations are given equal weight [12].

$$\mathrm{MAE} = \frac{1}{n} \sum_{j=1}^{n} |y_j - \hat{y}_j|$$

5.4. R Squared

R-squared (R^2) is a statistical measure that quantifies the percent of variation explained by an independent variable or variables in a regression model for a dependent variable. The R-squared value indicates how much the variation of one variable explains the variation of the other. So, if a model's R2 is 0.50, the model's inputs can explain nearly half of the observed variation [13].

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}$$

Table-1: Comparison based on error metrics

PARAMETERS	FB PROPHET	LSTM
MSE	41.2242	62160.9654
RMSE	6.4206	249.321
MAE	5.4989	224.7766
R Squared	-0.9860	-835458.884

From the analysis it is clearly shown that prophet is much ahead of LSTM in both accuracy as well as Prophet generates less error in comparison to LSTM. Therefore, Prophet can be labelled as a better model.

6. CONCLUSION

The project will help the amateur trading community to take better decisions while making trades. Since we have compared two different models from two different sectors machine learning, we get a clear idea about the working and accuracy of these techniques. The platform created is having several immaculate advantages. They are,

- The method can be back-tested using available historical and real-time data to check the viability of the trading strategy.
- Loss percentage can be reduced.
- Traders can take better decisions.
- Traders can test the accuracy of their forecast.

Stock market penetration in developed markets like United States are more than 52% and majority of the population is benefited by the impeccable advantages offered by the market. Where as in India, where the market penetration is less than 2%. Our project will make the process of interaction with the stock market much easier, thereby increase market penetration in India for the masses.

7. REFERENCES

[1] Intial Public Offering, https://www.investopedia.com/
[2] Hiransha M, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P (2018). NSE Stock Market Prediction Using Deep-Learning Models
[3] Basics of Algorithmic Trading: Concepts and Examples, https://www.investopedia.com
[4] Advantages and Disadvantages of Algorithmic Trading, https://www.letslearnfinance.com
[5] Types of Data, https://subscription.packtpub.com/
[6] Time Series Data, https://www.influxdata.com/what-is-time-series-data/
[7] Adil MOGHAR, Mhamed HAMICHE (2020). Stock Market Prediction Using LSTM Recurrent Neural Network
[8] Salman Ahmed, Saeed-Ul Hassan, Naif Radi Aljohani, Raheel Nawaz (2020). FLF-LSTM: A novel prediction system using Forex Loss Function
[9] Understanding LSTM Networks, http://colah.github.io/posts/2015-08-Understanding-LSTMs/
[10] Christophorus Beneditto Aditya Satrio, William Dermatan, Bellatasya Unrica Nadia, Novita Hanafiah (2020). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET
[11] Mean Squared Error and Root Mean Squared Error, https://www.oreilly.com/
[12] Mean Absolute Error, https://medium.com/
[13] R- Squared, https://www.investopedia.com/

