Project Proposal
On
A Survey on Time Series Analysis based Stock Price Prediction Using ARIMA and FBProphet

Nupur Singh — Pradeep Raj Thapaliya — Sachin Janwalkar

Abstract

Stock market or equity market have a profound impact in today's economy. A rise or fall in the share price has an important role in determining the investor's gain.Stock price prediction has been a big challenge since many years due to the unpredictable nature of the market with around 200 transactions per second, making it hard to predict the stock price. Stock market also known as share market or equity market is a cluster of sellers and buyers of stock. Big and private companies sell a part of their company to investors in order to generate crowdfunding. Since the stock market is quite sensitive to economic sentiments and rapid changes, the main objective of the stock market is to develop new techniques to forecast higher profits for stock prices. This paper aims to examine time series analysis methods using two different frameworks ARIMA and FBProphet ,by which the future stock prices can be accurately predicted.

1   Goals and Methodology

The proposed method is a model independent approach. Here we are trying to fit the data in two different models built based on ARIMA and FBProphet. As part of this project we will empirically study the challenges and the results in both the techniques and also try to produce the survey results.

1.1   Goals and Hypothesis

1.1.1   Research Questions

1. How effective Time Series analysis is in terms of stock price trend prediction.

2. Which among ARIMA and FBProphet is better for time series analysis and under which circumstances.

3. A qualitative and quantitative analysis of ARIMA and FBProphet

   In this paper we compare the performance of standard Autoregressive Integrated Moving Average (ARIMA) models [27], which we take as a benchmark, to Prophet – a

forecasting tool developed by Facebook and based on a Generative Additive Model (GAM)

1.2 Methodology

We start by importing data from yahoo finance , from which we can fetch stock market data of various companies like 'Amazon', 'Apple' etc within the given timestamp the data set consists of 7 attributes. Once we have scrapped the most recent stock data we can proceed with data exploratory analysis:

1. Data Importing we will be importing data in R

2. Data Cleaning

3. Data Analysis

   We will be using diverse set of packages, functions, plots and graphical methods to analyse our data set

4. Implementation:

   We consider ARIMA models and compare them to Prophet, a scalable forecasting tool by Facebook based on a generalized additive model. The performance of both the models for forecasting stock prices is compared to obtain a good benchmark for the problem under study. A brief description of both the models is available in the late section.

2 Metrics

Since stock prices prediction is essentially a regression problem, the RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) will be our current model evaluation metrics. Both are useful measures of forecast accuracy.

$$MAPE \;=\; \frac{1}{N} \;\times\; \sum_{t=1}^{N} \left| \frac{A_t \;-\; F_t}{A_t} \right|$$

$$RMSE \;=\; \sqrt{\frac{1}{N} \;\times\; \sum_{t=1}^{N} (A_t \;-\; F_t)^2}$$

, where N = the number of time points, At = the actual / true stock price, Ft = the predicted / forecast value.

RMSE gives the differences between predicted and true values, whereas MAPE measures this difference relative to the true values. For example, a MAPE value of 12percent indicates that the mean difference between the predicted stock price and the actual stock price is 12

3 Project Outline

3.1 Literature Reviews and Related Work

1. Ch.Raga Madhuri; Mukesh Chinta; V V N V Phani Kumar(2020 7th International Conference on Smart Structures and Systems (ICSSS)), Stock Market Prediction for Time-series Forecasting using Prophet upon ARIMA

   In this parer, various techniques are utilized to enhance the historical pattern of stock marketing trade and tell future results accordingly. The stock market is distinguished as vigorous, unforeseeable, and non-linear.

2. Anusha Garlapati; Doredla Radha Krishna; Kavya Garlapati(2021 6th International Conference for Convergence in Technology (I2CT)),Stock Price Prediction Using Facebook Prophet and Arima Models

   It has discussed stock market trends and analyzed different patterns of data, and done analysis for future forecasting of stock prices. For this commentary and foresight models like ARIMA and FACEBOOK PROPHET are used. To construct these models, data is deduced on Stock price predictions from 2012–2020. This analysis has further prospective for investigation in the future.

   MAPE is used as a parameter that shows that the models are adequate in forecasting retail valuation.

   This empirical inquiry designated that the ARIMA (2,1,2) model is best for predicting Stock prices.

   FACEBOOK PROPHET here is utilized to exhibit future forecasting of stock prices.

3.2 Data Sources and Reference Data

We pull the yahoo finance dataset using the script. yfinance library offers a reliable, threaded, and Pythonic way to download historical market data from Yahoo Finance [1]. We are using the Apple, Google, and Tesla Stock Dataset for performing the stock price prediction and forecasting. The dataset source used for this purpose is from Yahoo Finance. We are using more than 5k daily stock trading for each company. Dataset has each tuple with Date, Open, Close, Low, High, Adj Close, and Volume fields.

Following table shows the all the columns, data type, and description of the dataset.

Table 1 Data columns, data type, and description

Column Name	Data Type	Description
Date	date	Date of the stock trading
Open	float64	the price at which a stock started trading
High	float64	Highest price at which a stock traded during a period
Low	float64	lowest price of the period
Close	float64	Stock value of the day closing
Adj Close	float64	more accurate measure of stocks' value of that closing of the period
Volume	int64	Total volume that is traded during period

Following snippets show the chunk of the data set that we are going to use.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2000-01-03	0.936384	1.004464	0.907924	0.999442	0.858137	535796800
1	2000-01-04	0.966518	0.987723	0.903460	0.915179	0.785788	512377600
2	2000-01-05	0.926339	0.987165	0.919643	0.928571	0.797286	778321600
3	2000-01-06	0.947545	0.955357	0.848214	0.848214	0.728290	767972800
4	2000-01-07	0.861607	0.901786	0.852679	0.888393	0.762789	460734400

Figure 1: Apple Stock Data: Snippets of the Apple stock Dataset

In this survey, we are going to explore the stock price changes over time using the various visualization tools and techniques. We will perform the following;

1. Explore the change in price of the stock over time.

2. Calculating and visualizing daily return of the stock on average.

3. Finding correlation between different stocks.

3.3 Data Processing

Many machine learning algorithms perform better or converge faster when features are on a relatively similar scale and/or close to normally distributed. We are going to perform various techniques like scaling, standardization and Normalization on the stock data. We don't have to worry about the null value and missing values because the dataset available in the yahoo finance is reliable and has no missing or null values. These terms can be explain following ways;[2]

1. Scale generally means to change the range of the values. The shape of the distribution does not change. Think about how a scale model of a building has the same proportions as the original, just smaller. That's why we say it is drawn to scale. The range is often set at 0 to 1.

2. Standardize generally means changing the values so that the distribution's standard deviation equals one. Scaling is often implied.

3. Normalize can be used to mean either of the above things (and more!).

Beside these, we will process our dataset in training, testing and validation set in such a way that it can be used for time-series analysis.

3.4 Model Selection

Here we are trying to fit the data in two different models built based on ARIMA and FBProphet. As part of this project we will empirically study the challenges and the results in both the techniques and also try to produce the survey results. For more about these models, see section 4.

4 Analysis Model

4.1 ARIMA

1. The fact of introducing ARIMA models comes from the assumption that we are not working with a non stationary dataset series. We say that time series datasets are stationary when their means, variance and autocovariance dont change during time. The mayority of economic time series are not stationary, but differencing them determined number times makes them stationary. Whit this previous operation we can apply ARIMA models to any stock price. In general we say that a temporal set Yt admits an integrated autoregressive representation with p, q and d moving average orders respectively. We denote this forecasting model by ARIMA(p, d, q).

$$Y_t = c + \phi_1 y d_{t-1} + \phi_p y d_{tp} + ... + \theta_1 e_{t-1} + \theta_q e_{tq} + e_t$$

2. In ARIMA, p denotes the number of autoregressive terms, d denotes the number of times that the set should be differenciated for making it stationary. The last parameter q denotes the number of invertible moving average terms.

3. The construction of the models is made by iterative approaches using 4 steps:

   -Identification: With the time dataset we try to incorporate a relevant research model. The objective is to find the best values reproducing the time set variable to forecast.

   -Analysis and Differentiation: This step consists on studying the time set. In this study we incorporate different statistical tools like ACF and PACF tests, selecting the model parameters.
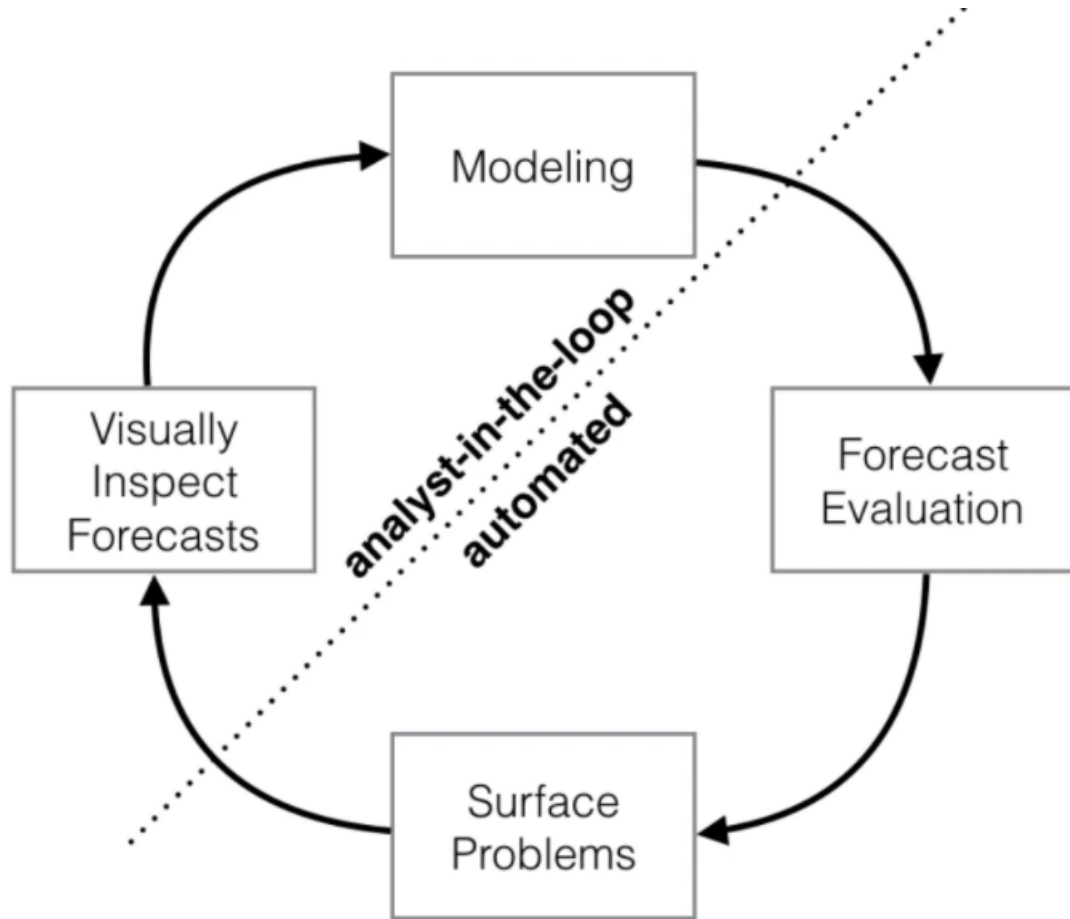
Figure 2: Prophet Forecast Flow

-Adjusting ARIMA model: We extract the determination coeficients and adjust the model.

-Prediction: Once we have selected the best model, we can make a forecasting based on probabilistic future values.

4.2 FBProphet

1. The origin of prophet comes from the application of a forecasting model into supply chain management, sales and economics. This model helps with a statistical approach in shaping business decisions.The Prophet model has been developed by Facebook's Core Data Science team and it is an open-source tool for business forecasting.

   Prophet forecast flow

   $$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

2. The Prophet model is an additive model with the components g(t) models trends, s(t) models seasonality with Fourier series, h(t) effects of holidays or large events. For this

study we are not deeply analyse the Prophet model operations because they are too dense and deep that it requires a new independent study but we did a good approach explaining .

5 Tools

5.1 Software

1. R Studio

2. Overleaf

3. Google Colab

5.2 Libraries/Packages

There are various library using for this project; FBProphet, Arima, tseries, dplyr, forecast, timeseries, ggplot2, knitr, caret etc.

5.3 Group collaborations & Version Control

Following are the collaborative tools that are going to use;

1. GitHub - To manage all the code and for the version control

2. Git - Free and Open Source technology using for update the change on code and collaboration between team members

3. Overleaf - Online tool to write the reports with clean formatted way and document the entire process

4. LaTeX - To Write the mathematical formulas. It is intended to provide a high-level, descriptive markup language that accesses the power of TeX in an easier way for writers.

5. Google Drive.

References

[1] NASDAQ. (2021, OCt 5). Microsoft Corporation (MSFT). Stock Dataset. Retrieved from https://finance.yahoo.com/quote/MSFT

[2] Jeff Hale . (2019, May 4). Scale, Standardize, or Normalize with Scikit-Learn. https://towardsdatascience.com/scale-standardize-or-normalize-with-scikit-learn-6ccc7d176a02

[3] Andrea M, Massimiliano P. (2021, Aug 21). Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices. https://arxiv.org/pdf/2107.12770.pdf

[4] https://ieeexplore.ieee.org/abstract/document/6970973?casa_token=EIhogNa2FFwAAAAA:yI9Y3Qv
QRhcKxrNiftoq9r1EZsFL9kqfGveyNZNycjK4

[5] https://medium.com/@josephabraham9996/time-series-analysis-on-stock-market-
forecasting-arima-prophet-2b60cacf604