STOCK MARKET ANALYSIS

Caption

Report by :

Aishwarya Melige Suryanarayana
Arunsai Chettipally

Stock Market Analysis : Cover Page

Group 8

Student 1 Name:

Arunsai Chettipally

Student 2 Name:

Aishwarya Melige Suryanarayana

603-930-7971

541-300-7023

Chettipally.a@northeastern.edu
Meligesuryanarayan.a@northeastern.edu

Percentage of Effort Contributed by Arunsai : 50%

Percentage of Effort Contributed by Aishwarya : 50%

Signature of Arunsai : Arunsai

Signature of Aishwarya: Aishwarya

Submission Date : 13th December 2021

ABSTRACT:

The prediction of stock market can be defined as to determine the industry stock's future value or other finances which is on the national stock exchange. The advantage of a successful prediction of a stock's future price could results insignificant profit. The efficient-market hypothesis recommends that stock costs mirror all right now accessible data and any value changes that are not founded on recently uncovered data subsequently are an unpredictable. Models like prophet are the machine learning models which are some of best time series machine learning models. Each have its own perspective to choose for a time series machine learning model and use them for the prediction. We will be using both model for prediction of stock price and compute the results and compare them for best accuracy and performance.

INTRODUCTION :

The time series forecasting has been become a trend in these recent years by the researchers. There are many interesting methods and algorithms, which have been proposed for prediction. Whether one wish to predict some stock prices or some the consumption of electricity, time plays an important factor that is widely considered for the time Series Models. For example, it would be very interesting to know when the price of the stock is going to rise up but also when it is going to rise up.A time series is simply can be defined as a series of data points which have been ordered in a time orderly manner. In time series, time is usually the independent variable and the goal is usually to make a forecast for the future.However, there are some aspects which needs to be considered when dealing with time series.

Stationarity

The Stationarity is an important characteristic of the time series. If the statistical properties do not change over the time, the time series can be referred to as stationary. It can be also depicted as that it has constant mean and variance, and the covariance is independent of time. Since we see a growing trend in the stock prices, or the volatility might increase over the time period (meaning the variance is changing).The Dickey-fuller test is used, which is a statistical test that we run to determine if a time series process is stationary or not. The dickey-fuller test is a test which test the null hypothesis that a union root is present.

- If it is, then p >0, and the process is stationary.
- Otherwise, p = 0, the null hypothesis is rejected, and the process is considered to be stationary.

Seasonality

The periodic fluctuations in the time series is referred as the seasonality. For example, the online sales of a company or website increased during some vacation or festival like Diwali before slowing down again.The seasonality can also be derived from the autocorrelation plot if it has a sinusoidal shape. The look at the period can give the length of the season.

Autocorrelation

The autocorrelation informally can be described as the similarity between observations as a function of the time lag between them. If the autocorrelation plot looks like sinusoidal function, it hints for the seasonality and you can find its value by finding the period in the plot.

DATASET DESCRIPTION :

Link to the dataset : https://finance.yahoo.com/quote/%5EGSPC/history?period1=1607519689&period2=1639055689&interval=1wk&filter=history&frequency=1wk&includeAdjustedClose=true

In stock trading, the high and low refer to the maximum and minimum prices in a given time period. Open and close are the prices at which a stock began and ended trading in the same period. Volume is the total amount of trading activity. Adjusted values factor in corporate actions such as dividends, stock splits, and new share issuance.

DATA PREPROCESSING :

Data preprocessing for data mining in python:

In data preprocessing, we have the following steps:

- Importing the libraries
- Importing the Dataset
- Handling of Missing Data
- Handling of Categorical Data

Splitting the dataset into training and testing datasets

1. Importing the libraries:

import numpy as np

import pandas as pd

from pandas import read_csv

from sklearn.impute import SimpleImputer #used for handling missing data

from sklearn.model_selection import train_test_split # used for splitting training and testing data

2. Importing the dataset:

dataset = read_csv('/Users/aishwaryamelige/Downloads/archive (2)/stocks/GNC.csv') print(dataset.describe())

3. Handling the missing data :

The larger the dataset, the more chances are that dataset has missing values. We have to remove the missing values

There are many options we could consider when replacing a missing value, for example:

- A constant value that has meaning within the domain, such as 0, distinct from all other values.

• A value from another randomly selected record. • A mean, median or mode value for the column.
• A value estimated by another predictive model.

Here we define an imputer and calculate the mean accuracy: values = dataset.values

```
X = values[:,0:8]
Y = values[:,8]
imputer = SimpleImputer(missing_values=nan, strategy='mean')
lda = LinearDiscriminantAnalysis()
pipeline = Pipeline(steps=[('imputer', imputer),('model', lda)])
kfold = KFold(n_splits=3, shuffle=True, random_state=1)
result = cross_val_score(pipeline, X, y, cv=kfold, scoring='accuracy')
print('Accuracy: %.3f' % result.mean())
```

4. Splitting the dataset into training and testing datasets:
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=0)¯

PREDICTION ALGORITHM :

Prophet Prediction Model is an additive based model, which is a procedure for forecasting time series data in which the non-linear trends are fit with yearly, weekly, and daily seasonality and as well as the holiday effects. The best results of this model are to be seen when the time series have strong seasonal effects and several season of historical data.The prophet model is vigorous to missing data and the trend shifts and handles the outliers as well. Prophet is an open source software released by Facebook's Core Data Science team. The complex statistical modelling is handled by the stan library and is a prerequisite for prophet. Prophet follows the sklearn model API. The instance of the Prophet class is been created and then call its fit and predict methods.

The data frame in the prophet model have always two columns: ds and y. The ds (date stamp) column should be a format expected by pandas, it can be of any format like YYYY-MM-DD HH:MM: SS for a timestamp and YYYY-MM-DD for a date. The y column must be numeric, and it should represent the measurement or attribute which u need to forecast.

How Prophet Works:

As previously said prophet is an additive model which consists of the following components: $y(t) = g(t) + s(t) + h(t) + \epsilon$□

MODEL SELECTION AND IMPLEMENTATION :

TIME STATIONARY IDENTIFICATION :

Time series analysis relies on stationary identification since it can only be used with that kind of data. One of the most often used statistical tests is the Augmented Dickey-Fuller (ADF) test. If a unit root is present in the series, it may be utilised to establish whether or not the series is stationary. The null hypothesis is rejected if the test statistic is lower than the crucial value (aka the series is stationary). Rejecting the null hypothesis occurs when the test statistic exceeds the crucial value (which means the series is not stationary). Table 2 summarises the results of the ADF test for daily time series prediction. It's clear that the series isn't stationary when the daily, weekly, and monthly test statistics are all above critical.

BUILDING MODEL AND PREDICTION :

The data is divided into train and test sets for the model's construction. The three ARMA models were built using training data (daily, monthly, and weekly data) and projected time series using the proper sequence from step 1 of ARIMA model building (daily, monthly, weekly test data). For forecasting purposes, alpha 0.05 corresponds to a 95 percent confidence level. A price prediction is made, and then compared to the actual price, followed by a thorough study.

EXPLANATION :

ARIMA and PROPHET models for stock price prediction are presented here for construction and forecasting for the first time. Daily, monthly, and weekly forecasts are used to discover the best forecasting period. MAPE is used to compare the error analysis findings of forecasting techniques across three distinct time periods in order to determine the best accurate model. The results of the experiments showed that the PROPHET model has an ARIMA error rate that is quite similar to reality.

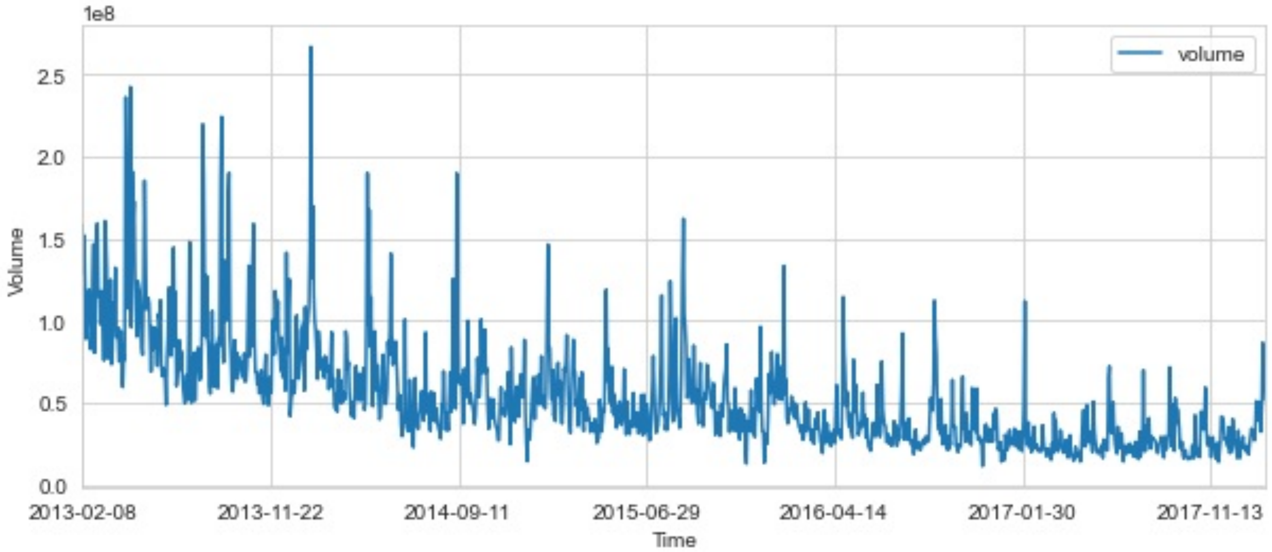
DATA EXPLORATION AND VISUALISATION :

Historical Closing Price of Apple Stock

The above plot shows the increase in price of Apple stock over 5 years from 2008 to 2013.

This shows that there is a good investment opportunity in the stock market for Apple stocks.

Volume traded for Apple Stock

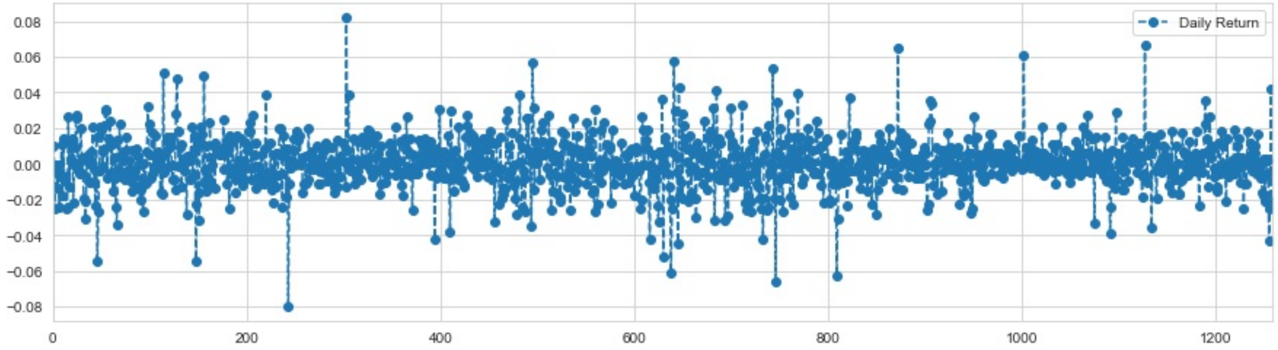


There has been significant dip in the volume traded for the Apple stock, owing to the increase in the price of the stock over the years.

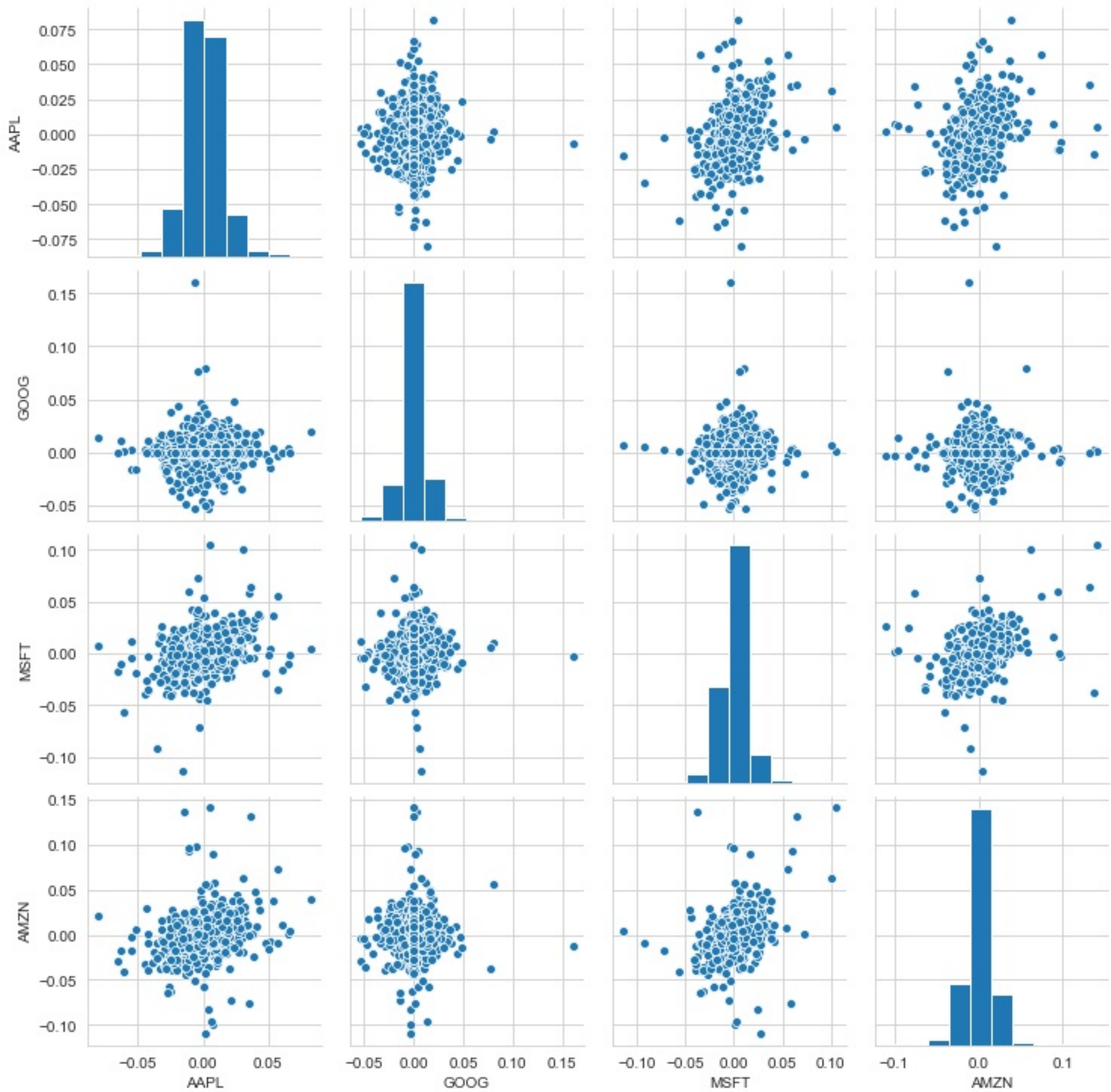
Moving Average for Apple Stock

The above graph indicates the price of the Apple stock at the close of the stock market for the drive years 2008-2013. Dose, dose_10, dose_50 indicate slight differences due to corporate interferences. The prices still seem to increase over time.

Daily Returns



The above plot shows the daily return of the Apple stocks. The plot shows the volume of the returns per day.



This is a “Pair Plot”. It shows us the distribution of Apple, Google, Microsoft and Amazon stocks with themselves and with each other.

IMPLEMENTATION:

```
#!/usr/bin/env python
# coding: utf-8

# ### Import Python Modules

# In[3]:


import pandas as pd
import numpy as np
from pandas import Series,DataFrame
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('whitegrid')
from datetime import datetime
import os


# ### Importing the Stock Prices

# In[4]:


AAPL=pd.read_csv('individual_stocks_5yr/AAPL_data.csv')
GOOG=pd.read_csv('individual_stocks_5yr/GOOG_data.csv')
MSFT=pd.read_csv('individual_stocks_5yr/MSFT_data.csv')
AMZN=pd.read_csv('individual_stocks_5yr/AMZN_data.csv')


# In[5]:


AAPL['date']=pd.to_datetime(AAPL['date'])
GOOG['date']=pd.to_datetime(AAPL['date'])
MSFT['date']=pd.to_datetime(AAPL['date'])
AMZN['date']=pd.to_datetime(AAPL['date'])


# ### Forecasting Apple Stock Price

# In[8]:


AAPL.head()


# In[9]:


df_prophet=AAPL[['date','close']]
df_prophet=df_prophet.sort_values('date')
df_prophet


# **Renaiming the Column names to Suite Prophet Algorithm**

# In[10]:


df_prophet=df_prophet.rename(columns={'date':'ds','close':'y'})
df_prophet


# **Creating the Prophet Model**

# In[1]:
```

RESULTS AND DISCUSSION :

The experiment was done for 1 machine learning models.From the error percentage its shows that the prophet models perform better than that of another model.

However, as Stock Market is very dynamical system, hence the patterns and dynamics presented by the model may not be same always as stock market data is way more dynamic in real world scenario. This causes to the machine learning model to not able to capture the dynamic change in data points.

Results:

The accuracy and result of the model found is depicted as follows using root mean Square error function. rms=np.sqrt(np.mean(np.power((np.array(valid['Close'])-np.array(forecast['Prediction'])),2))) rms=44.954584993246954

Prophet Model

```
#fit the model                    o
model = Prophet () model.fit(train)
#predictions
close prices = model.make_future_dataframe(periods=len(valid))
forecast = model. predict(close prices)
```

Sample Code for the Prophet Model is shown below:

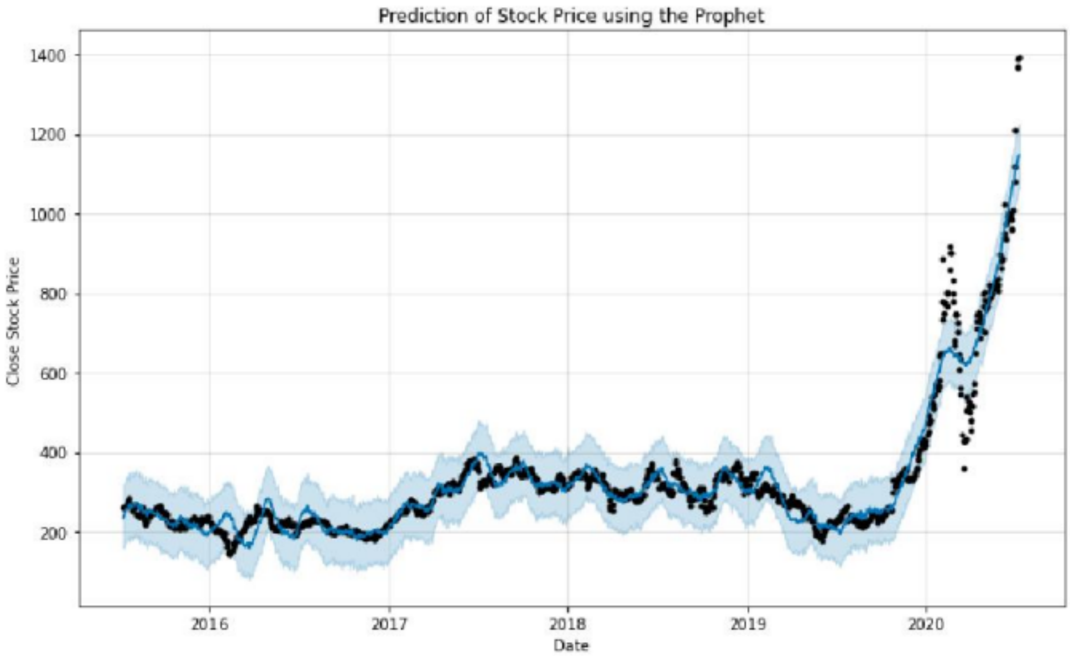
The model.fit () is used to fit the dataset and train model, and predict the closing price of the stock.

The model. make_future_dataframe() is used for forecasting of the closing price and it takes arguments ‘periods’ which is no of days which needs to be forecasted.

Another argument which can be given is ‘freq’ which describes forecasting in days, months or years. E.g.: freq=’D’ D stands for days.

freq=’M’ M stands for Months. Freq=’Y’ Y stands for Years.

The accuracy and result of the model found is depicted as follows using root mean Square error function. #rmse
forecast_valid = forecast['yhat'][987:] rms=np.sqrt(np.mean(np.power((np.array(valid['y'])-np.array(forecast_valid)),2))) rms
57.494461930575149



CONCLUSION:

These past history values are used by the model to capture the increasing trend in the series. Although the predictions done by this model are far off the real-world stock pricing system, hence these predictions are still not close to the real values.

Like most other time series models, Prophet tends to capture the trend and seasonality from the past data and tries forecasting the future. The Prophet model performs better on various time series datasets, only scenarios like this it gets results off to that of its actual calibre. Overall, it turns out that the stock prices do not have a seasonality or a particular trend. It mainly depends on the current upcoming in the markets and owing to which the prices of the stock may go high or comes below. Hence, we can depict that techniques like Prophet are not that good of for forecasting of the stock prices and similar kind of problems.

FUTURE ENHANCEMENTS:

The implementation of this paper can be extended by integrating the technical analysis and fundamental analysis techniques. Through the evaluation of social media analysis particularly on public opinions using fundamental analysis techniques can be incorporated in order to obtain better results. In this way we can provide the improved results for investors in the stock market to choose the better timing for profitable investment decisions.

REFERENCES :

[1]. Qinkun Xiao School of Electronic Information Engineering Xi’an Technological University Xi’an, China - Time Series Prediction using Graph Model.

[2]. Sreelekshmy Selvin, Vinayakumar R, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P Centre for Computational Engineering and Networking (CEN), Amrita School of Engineering,Coimbatore- Stock Price Prediction using RNN , LSTM and CNN window model

[3].https://www.kaggle.com/arindamgot/eda-prophet-mlp-neural-network-forecasting?select=train.csv