TIME SERIES FORECASTING

Predict the Live Stock prices using yahoo finance



Presented to: Presented by

Mohammadreza (Reza) Ebrahimi Sai Vinay Thattukolla,

Abhinav Karre,

Akshay Kumar Reddy nalla,

Praharsha Vadlapatla,

Shreya Reddy Vurelly,

Shivani Reddy Donthireddy

Table of Contents

[Introduction: 3](#_Toc119010454)

[Time series: 4](#_Toc119010455)

[Uses of Time Series 4](#_Toc119010456)

[Components for Time Series Analysis 5](#_Toc119010457)

[Problem statement 5](#_Toc119010458)

[Libraries 6](#_Toc119010459)

[Data Characteristics 6](#_Toc119010460)

[How data is related to the problem being solved? 7](#_Toc119010461)

[Moving Averages: 7](#_Toc119010462)

[Visualizations: 8](#_Toc119010463)

[Correlation plot: 8](#_Toc119010464)

[DM Models: 9](#_Toc119010465)

[Evaluation Metrics: 10](#_Toc119010466)

[Conclusion: 10](#_Toc119010467)

[Results: 10](#_Toc119010468)

[Finding: 11](#_Toc119010469)

[Current Problem: 11](#_Toc119010470)

[References: 12](#_Toc119010471)

# Introduction:

In this project we will predict the live stock prices of the top 100 companies listed in wikipedia using Yahoo finance and scikit learn. Along the way, we will create a new dataset with company names and its corresponding stock symbol appended to the data. There are many models on the google which will predict the future prices of the data (Tomorrow) based on the current (Today) and past (previous day) data. But in this project we will forecast the prices of the stocks from second to second in the day itself. This means our model will predict the prices starting from the 9:30 am to 4:00 pm on any day, which is the time peroid the stock market functions. Also we will include the evaluation metrics to judge the behaviour of the model for th etime series data. Finally we will see how the daily news have an important impact on the stock prices by taking the twitter acquisition example.

The main purpose in doing this is project is, it contains some concepts that I believe has a greater impact and applications in the real world. The majority of people turn to the performance of a country’s stock market as the best indicator of how well that economy is doing. Stock markets cover all industries across all sectors of the economy. This means they serve as a metric of what cycle the economy is in and the hopes and fears of the population who generate growth and wealth. Stock market enable companies to be traded publicly and raise capital. Stock markets promote funds and investments to the company who needs some capital to expand their operations across the globe. For investors, stock markets provide a way to invest money in order to potentially earn a share of the company’s profits and losses too. But what makes the stock market unpredictable is its ability to move the prices based on the daily activities and its has other important factor is the time component.

# Time series:

According to the statisitics, time series is a collection of observations of well-defined data items obtained through repeated measurements over time. For example, measuring the value of retail sales each month of the year would comprise a time series.

# Uses of Time Series

* The most important real world appliaction of time series is that it helps us to predict the future behaviour of the stock or securities market based on the past experience.
* It is helpful in business perspective as it helps in comparing the actual current performance with the expected one.
* We can compare the changes in the values of different variables at different times or places, etc.

# Components for Time Series Analysis

There are four categories of the components of time series. They are

* Trend
* Seasonal Variations
* Cyclic Variations
* Random or irregular Movements

Seasonal and Cyclic Variations are the periodic changes or short-term fluctuations in the price of the stock.

# Problem statement

* In this project, we are predicting whether the future price of a stock goes upward or downward (depicted as “Target”-(1/0)) based on past values of the data.
* There are many websites predicting the stock prices. But they are predicting the stock prices of tomorrow based on today data. In my Knowledge, this is the first project where we are predicting the price of the stock change from second to second.
* Stock prices change from milli second to milli second based on situations happening around the world. That’s why it is important to develop a model that can help to predict the stock price from second to second.
* This is famous for the Intraday Traders.

# Libraries

1. Yahoo finance
2. Pandas
3. Numpy
4. Matplotlib
5. Scikit Learn
6. Datetime
7. Seaborn (Visualization)

# Data Characteristics

* Independent Variables: DateTime, Stock Symbol, Company, Open, High, Close, Low, AdjClose, Volume.
* Dependent Variable: Target
* Model Type: Binary Classification
* Method: Supervised Learning

### How data is related to the problem being solved?

* When we see the data characteristics, we can observe that the attributes are Datetime which is a time series component has instances starting from 9:30 am to 4:00 pm, is the time period in which stock market functions.
* Open, Close, High and Low are stock prices of each company with respect to time.
* AdjClose is the adjusted closing price factors in anything that might affect the stock price after the market closes.

When we observe the price variation of the each stock you can see how the price increase or decrease with respect to time. We can observe the trend in the data. Seasonal and cyclicity can be observed if we increse no of instances in the data.

# Moving Averages:

A Moving average is a tool used by the technical analysts to track the price movements of a stock. It plots average prices over a defined period of time, with the moving average typically overlaid onto a candlestick or bar chart or line plot. Moving averages helps the analysts to predict the behaviour of the stock smoothly.

Chart, line chart

Description automatically generated

# Visualizations:

Chart

Description automatically generatedChart

Description automatically generated with low confidenceChart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generatedA picture containing chart

Description automatically generated

Analysis:

When we observe line plots of each attributes, there is a peak at the start of the open, high and low and graph is following a decresing trend. Since the graph of close and adjclose is similar, this shows that the there is no trend impact news happening during that time.

# Correlation plot:

A screenshot of a computer

Description automatically generated with medium confidence

# DM Models:

After preprocessing the data, we have applied a various machine learning algorithms like logistic regression, support vector machines, Random forest classifier.

|  |  |  |
| --- | --- | --- |
| **Model** | **Accuracy** | **Interpretability** |
| Random forest | 60% | Medium |
| Logistic Regression | 77% | Low |
| Support Vector Machines | 61% | High |

|  |  |  |
| --- | --- | --- |
| **class** | **Actual yes** | **Actual no** |
| Predicted yes | 5656 | 468 |
| Predicted no | 2182 | 3308 |

# Evaluation Metrics:

We can see that the true positives is 5656 and tru negative is 3308.We can find the precision and recall from the confusion matrix.

# Conclusion:

### Results:

* From the accuracy and confusion matrix, Logistic regression is the best classifier when compared to the random forest and Support vectors with 77% accuracy.
* The model is telling that 77% of times the price of the stock goes up in the next second out of 100 times. Though it is not an accurate model, I would consider this as the better model.

# Finding:

* For our analysis, we have taken the recent news that musk had officially took over the Twitter. We have recorded the change is price of the ‘TESLA’ when Elon musk bought the twitter on the 27- oct- 2022.

Chart, line chart

Description automatically generated

# Current Problem:

Since there are random fluctuations in the stock price trend, it is hard to predict the exact value.

Solution:

We have written a custom logic that used moving averages to predict the trend of the data. With the help of moving average, we used that data to create a model. If all the variables are held constant, then my model will be able to predict the future price of the data base don past and current data.

CEO Solution:

Using my model, CEO can predict the future price of the stock. This way investors can decide whether to invest in company or not.

# References:

1. <https://www.byteacademy.co/blog/time-series-python>

2. <https://www.analyticsvidhya.com/blog/2021/07/stock-prices-analysis-with-python/>

3.<https://codeit.us/blog/machine-learning-time-series-forecasting#:~:text=Time%20series%20forecasting%20in%20machine%20learning,-Before%20anything%20else&text=This%20involves%20creating%20assumptions%20and,processing%20current%20and%20previous%20data>

*THANK YOU*