J.P. Morgan Quant Mentorship Program 2021

Introduction:

Forecasting the stock market has always been an area of interest for researchers in the field of mathematics, finance, and computer science. Due to the non-linearity of the price movement it is very difficult for traders to predict the stock market price movement. Specifically, machine learning and deep learning models are employed for prediction of financial time series data. Stock market analysis involves two kinds of decision making strategies: fundamental analysis, and technical analysis. Fundamental Analysis makes of use of the company data like financial records, economic reports, and other publicly release statements. Technical analysis focuses on using historic price and volume data to predict the future price movements.

Machine learning methods can be trained over technical analysis tools like momentum indicators used by traders to measure the strength or weakness of a security's price. They identify market trends and complement traditional pricing data like Opening Price, High Price, Low Price, Closing Price, and Trading Volume.

Following are the most common technical indicators:

1. **Simple Moving Average (SMA)**
   Moving averages is a calculation to analyze data points by calculating and creating a series of averages of different subsets from entire data set. The simplest form of moving average is the simple moving average. It is the arithmetic mean of the given set of data points. The formula for calculating the simple moving average over a period of n days of a security is as follows:

$$SMA = \frac{A_1 + A_2 + \cdots + A_n}{n}$$

   - A_i is the price of the security on the i^{th} day
   - n is the number of days

2. **Exponential Moving Average (EMA)**
   The exponential moving average is a type of weighted moving average that gives more weight to recent prices, giving lower importance to older market trends. The formula for calculating exponential moving average is as follows:

$$EMA = \left[v_t \times \left(\frac{s}{1+n}\right)\right] + EMA_y \times \left(1 - \left(\frac{s}{1+n}\right)\right)$$

   - EMA_y was the EMA yesterday
   - v_t is the price of the security today
   - s is the smoothing factor
   - n is the number of days

3. **Moving Average Convergence Divergence (MACD)**

The moving average convergence divergence (MACD) is a trend-following momentum indicator that shows the relationship between two exponential moving averages. The MACD is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA.

$$MACD = 12\ day\ EMA - 26\ day\ EMA$$

4. Range

Range is the difference between the low and high prices for an index or security over a time period.

$$range_i = \max(h_i, h_{i-1}, \dots, h_{i-n-1}) - \min(l_i, l_{i-1}, \dots, l_{i-n-1})$$

- h_i is the high price on the i^{th} day
- l_i is the low price on the i^{th} day
- n is the number of days

5. Stochastic Oscillators (%K and %D)

Stochastic Oscillator aims to detect trends in prices by comparing the closing price of the security with the highest price and lowest price achieved by the security during the last 14 days.

Fast Stochastic Oscillator (%K)

The fast stochastic oscillator can be calculated using the following formula:

$$\%K = \left(\frac{C - L_{14}}{H_{14} - L_{14}}\right) \times 100$$

- C is the most recent closing price
- L_{14} is the lowest price in the 14 day period
- H_{14} is the highest price in the 14 day period

Slow Stochastic Oscillator (%D)

It is a 3-day moving average of the fast stochastic oscillator

$$\%D_i = \frac{\%K_i + \%K_{i-1} + \%K_{i-2}}{3}$$

6. Money Flow Index (MFI)

The money flow index is technical oscillator that uses price and volume data for measuring the price momentum of an asset by comparing its total inflow and outflow levels over a specified period of time. The formulation of Money Flow Index is as follows:

$$MFI = 100 - \frac{100}{1 + MFR}$$

$$MFR = \frac{14\ Period\ Positive\ Money\ Flow}{14\ Period\ Negative\ Money\ Flow}$$

$$Raw\ Money\ Flow = Typical\ Price \times Volume$$

$$Typical\ Price = \frac{High + Low + Close}{3}$$

The steps involved for calculating MFI are as follows:

a) Calculate the typical price for each of the last 14 days
b) Calculate raw money flow by multiplying typical price by the volume traded for that day
c) Money flow corresponding to the increase in typical prices is added to positive flow and decrease in typical price is added to negative flow
d) Money flow ratio is calculated as the ratio of the positive and negative money flow
e) Money flow index can be calculate by substituting the value of money flow ratio.

Problem Statement:

For this case study you have to design and implement a system which predicts the movement (increase or decrease) of price of stock. You can solve it any programming language of your choice (Python, MATLAB, C++, etc.) by writing classes/functions.

Data Description:

You have been provided with the **NSE Index NIFTY 50** data for the time period 1st October 2016 – 30th September 2021. The input contains the following variables: Date, Open Price, Close Price, Adjusted Close Price, High Price, Low Price, and Volume.

The training set should be used to build you principal component regression mode. The test set should be used to see how well your model performs on unseen data.

Task 1:

Compute the technical indicators mentioned above using the following configurations:

- SMA (n = 10) for Closing Price
- EMA (n = 10, s = 2) for Closing Price
- MACD for Closing Price
- Range (n = 10)
- Stochastic Oscillators (%K and %D)
- Money Flow Index

Subsequently split the dataset into training and test dataset. Use the data from 1st October 2016 to 30th September 2020 as the training dataset and the subsequent data as the test data set.

Solution format:

- If you have written the code, please provide a standalone function (no-arguments) that generates the technical indicators
- In the answers report, show us the plots of all the technical indicator variables across time. Explain the patterns you observe with a valid reason.

Task 2:

Perform Principal Component Analysis on the set of 11 variables - traditional data inputs (Close Price, High Price, Low Price, and Volume) and the technical indicators (SMA, EMA, MACD, Range, %K, %D, and MFI)

for reducing the dimensionality of the dataset, and to obtain a lower-dimension matrix of input data, to regress on the daily change of close prices.

Solution format:

- If you have written the code, please provide a standalone function (no-arguments) that performs PCA
- Provide the heat map of the Covariance Matrix, if there is no way to plot in the language of your choice, output the entire covariance matrix
- Output the list of Eigen values and Eigen vectors of the covariance matrix in order of importance
- A labelled plot depicting the percentage of variance explained by each component

Task 3:

Based on the principal components obtained in the previous tasks, perform Multi Linear Regression using principal components that explain 95% variance and 99% variance. The dependent prediction variable would be the change in closing price i.e. $dependent\ variable_i = closing\ price_{i+1} - closing\ price_i$. You can use the sign (positive/negative) of the dependent variable to determine the direction of price movement.

Solution format:

- If you have written the code, please provide a standalone function (no-arguments) that performs Linear Regression for multiple variables
- Output the regression coefficient associated with each principal component
- Output the accuracy of your model i.e. the percentage of times the models predicts the correct direction of price movement.

Instructions for Final Submission:

Any programming language you are comfortable with is permitted. You are permitted to use standard libraries in your chosen language. The program submitted must compulsorily have a main function that must call all the tasks described in the question when run. The program should also be well documented. The program file name must be of the format {CandidateName}_MAIN. If there are additional modules created by you that are imported in the script in the main function, name those files as {CandidateName}_Module_{ScriptName}. Report your findings and attach the results mentioned in the tasks above in a word document and submit it in the format {CandidateName}_JPMQuantMentroship LAAssignment.docx .

You may also include a short description of your program in the solution document if you wish.

Finally, the zip the solution code and documents together and name it

{FirstName}_{LastName}_{CollegeName(Short)}_JPMQuantMentroshipLAAssignment

Mail the compressed file to the indicated email address: jpmqrmentorship.mumbai@jpmorgan.com