Quant Challenge 2023 – Machine Learning

# Introduction

Equity investors across the world are constantly in search of models that can help them predict future equity returns for a company’s stock. To make their predictions, they use the historical data that is available to them.

One of the things which they rely heavily on is the data released in a company’s earnings. Generally released quarterly, a firm’s earnings release has three major elements - the **income statement**, **balance sheet** and **cash flow statement**, all of which can provide investors with valuable information. In addition, they look at are the historical return series for a stock and its peers over the recent past.

**Passive investing** is a strategy wherein investors attempt to maximize their risk adjusted returns via a buy-and-hold strategy for their portfolio, aiming to get these returns in the long term with minimum buying or selling in the interim.

Let us transport you into a world where you are an investor in the stock market. Your client wants your help to maximize returns via a passive investment strategy in the JPSE (JPMorgan Stock Exchange).

To create a portfolio that helps achieve this, you decide to look at the historical data available to you. For the past several years, you have monthly earnings data (Yes, in our strict universe, companies have to publish earnings monthly. Oh what a pain!! I know) available for various stocks in the JPSE, which gives you stock-specific data that can help you make your decision.

So now, how will you decide which stock to buy or sell?

Well, you can’t really look into the future and see how stock prices will fluctuate - but a good starting point can be trying to make some predictions for the same! For this problem, you simply need to start your journey as a smart investor by creating a model that helps you predict future prices for different stocks that are available to you in the JPSE.

# Question Statement

Given the monthly fundamental indicator values for 100 stocks in the JPSE for the past 20 years, your task is to predict the prices for these stocks as accurately as possible, i.e., aiming to minimize the error in your predictions.

You may assume the following:

1. You have stock-specific data available for 8 financial metrics namely - Current Liabilities, Total Liabilities, Equity, Total Assets, Current Assets, Total revenue, Net Income, Dividend and Shares Outstanding. All these values are available monthly as of the last day of that month (After the market closes). Investors can act on this information only after it is available to them.
2. Most investors convert raw financial metrics into **indicators** that can be humanly understandable and comparable across stocks. Here are a few examples :
3. Current Ratio = Current Assets / Current Liabilities.
4. Dividend Payout = Dividend / Net Income

You might want to look up the conventionally used financial ratios and what they can tell you about a company.

1. Stocks/metrics may also exhibit seasonality or autoregressive behavior. (Aha !! New jargon. Look it up. It isn’t rocket science :) )
2. Well not everything is explainable. There might be one crazy investor with deep pockets who invests basis his whims and fancies and that might move prices erratically

The details of the problem statement are as follow:

1. **Overview**

For this problem, your task is not to make any actual investment in the stock market, but simply to predict the [price] that each of these stocks will have on the last day of the next month. (After the market closes on the last day of month “t”, you will have the data for the financial metrics and the [price]t. You have to predict the price on the last day of the next month [price]t+1 )

1. **Data:**

You are given 2 files.  
Train.csv contains all of the training data(Includes “price”).  
Test\_x.csv contains all of the test data against which your model/predictions will be evaluated. (Does not include “price”)

1. **Performance Evaluation:**
2. Your submissions will be evaluated by checking the Root Mean Square Error (RMSE) for your predictions – the lower the error, the higher your prediction accuracy.
3. Make a presentation of 15 mins explaining your approach to the problem.
4. Focus on what your idea is. In finance, the more explainable your model is, the better. So try to keep your ML idea explainable.
5. Feel free to use any resource you like to train your model. Submit your training code as a Jupyter notebook. You can add illustrations like plots etc in your notebook and use it in your presentation as well.
6. **Evaluation Criterion-**

**You’ll get 15 minutes to present your approach and solution to a panel of senior leaders in JPMorgan QR head office (please adhere to the time limit, you’ll be strictly asked to stop once your time ends). At 10 am on the day of the event, you need to submit 2 things: your code in a jupyter notebook and the ppt.**

**Your solution will be judged on a combination of performance, conceptual clarity, and the final presentation.**