# Machine Learning approach in Stock monitoring

By Soppa Prasunna - 1901199


# Problem definition

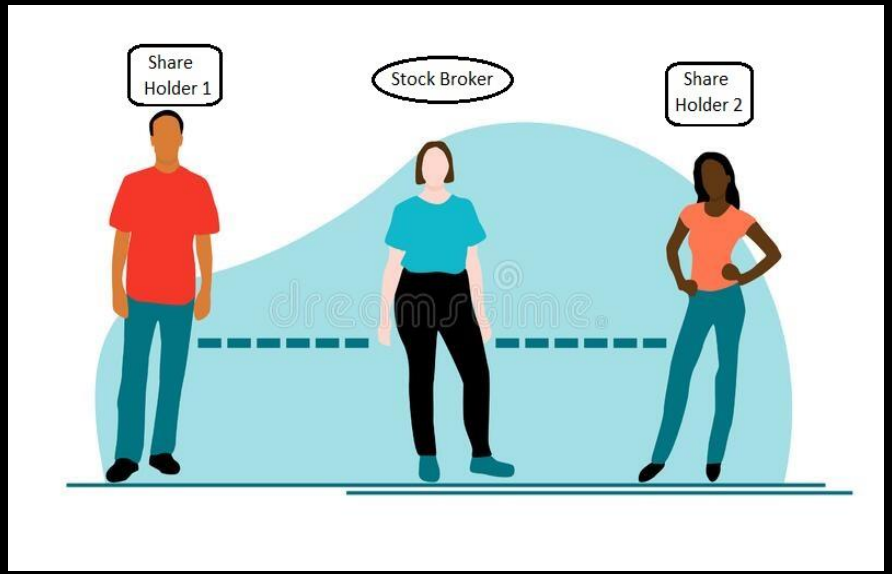
To find a predictive tool to predict stock's annual return, excess return, etc. for a weighted stock selection model which can affect portfolio performance that a manager expect to achieve.

# What is STOCK MARKET?

# Introduction

- Stock markets are trading institutions where stocks (equity) and other financial instruments such as bonds are offered for trade.
- In most stock exchanges, the common and easily accessible market is the equity market (stocks), where the entry investment can be as low as USD1.
- The performance of stock markets is measured on a daily basis by some key indicators such as 'share index'. Such an index is important in not only gauging the performance of trades in the stock exchange but also the economic performance of the particular country as a whole.

- Shareholders however do not directly execute the trade, nor is there any meeting between buyers and sellers for negotiations.
- Shareholders trade by giving instructions to their Stockbrokers, who in turn execute the orders.
- In their advisory role, some Stockbrokers base their advice on the fundamentals of the various stocks or undertake technical analysis.
- Stockbrokers need to be empowered, through better predictive tools, to enable them have some capability to provide the best advice to their clients.



- A predictive tool that Stockbrokers can use to guide on exact price movements, as a basis of investment, is therefore desirable. This can be a Machine Learning approach. Due to the importance of stock markets, investment is usually guided by some form of prediction.
- We are going to use Multivariate Linear Regression to predict the stocks as there are multiple features and the output is real-valued.


# Multivariate Linear Regression

Multivariate Regression is a supervised machine learning algorithm involving multiple data variables for analysis. A Multivariate regression is an **extension of multiple regression with one dependent variable and multiple independent variables**. Based on the number of independent variables, we try to predict the output.

The formula for linear regression:-

$$h = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \cdots$$

Here, ‘h’ is called the hypothesis. This is the predicted output variable. **Theta0 is the bias term** and all the other theta values are coefficients.


**Cost function and Gradient descent:-**When theta values are initiated in the beginning, the formula is not trained to predict the dependent variable. The hypothesis is far away from the original output variable 'Y'. This is the formula to estimate the cumulative distance of all the training data:

$$J(\theta_0, \theta_1, \theta_2, ..) = \frac{1}{2m}\sum(h_i - y_i)^2$$

This is called **the cost function**.The idea of a machine learning algorithm is to minimize the cost function so that the difference between the original output and the predicted output is closer. To achieve just that, we need to optimize the theta values.Here is how we update the theta values. We take the partial differential of the cost function with respect to each theta value and deduct that value from the existing theta value

$$\theta_0 = \theta_0 - \alpha \frac{d}{d\,\theta_0} J(\theta_0)$$

$$\theta_1 = \theta_1 - \alpha \frac{d}{d\,\theta_1} J(\theta_1)$$


Here, alpha is the learning rate and it is a constant. I am not showing the same formula for all the theta values. But It is the same formula for all the theta values. After the differentiation, the formula comes out to be:

$$\theta_0 = \theta_0 - \alpha \frac{1}{m} \Sigma (h_i - y_i)$$

$$\theta_i = \theta_i - \alpha \frac{1}{m} \Sigma (h_i - y_i) X_i$$

This is called gradient descent.

From the Dataset:-

Input features:

- Large B/P
- Large ROE
- Large S/P
- Large Return Rate in the last quarter
- Large Market Value
- Small systematic Risk

Target Variables:

1)Annual Return
2)Excess Return
3)Systematic Risk
4)Total Risk
5)Abs. Win Rate
6)Rel. Win Rate

Data Set Characteristic - Multivariate

Attribute Characteristic - Real

Associated Task - Regression


# Result

## Approach 1:

Scores of the model:

| | Training | Testing | | M.S.E |
|---|---|---|---|---|
| Annual Return:- | 0.727575 | 0.688261 | | 0.005404 |
| Excess Return:- | 0.768669 | 0.779566 | | 0.003857 |
| Systematic Risk:- | 0.450503 | -0.6137 | | 0.017643 |
| Total Risk:- | 0.589259 | 0.160344 | | 0.013049 |
| Abs. Win Rate:- | 0.345117 | -0.05055 | | 0.015907 |
| Rel. Win Rate:- | 0.574806 | 0.472092 | | 0.006679 |


Contd…

Hyper parameter tuning is done.

| | MSE of Train set | MSE of Test set |
|---|---|---|
| Annual Return - | 0.004625 | 0.007903 |
| Excess Return - | 0.006619 | 0.008991 |
| Systematic Risk - | 0.013758 | 0.008229 |
| Total Risk - | 0.007897 | 0.003529 |
| Abs. Win Rate - | 0.006229 | 0.011350 |
| Rel. Win Rate - | 0.006815 | 0.010948 |

Hyper parameters are tuned with different combination of values. These Mean Squared values are deduced when hyper parameters have its best value.


# Analysis:

- Since the values of target variables are continuous, we can use Regression method to predict the output and relation between independent variables and dependent variables is linear, we could use Linear Regression for this purpose.
- Since there are multiple dependent variables, we came to a conclusion that we have to use Multivariate Linear Regression.
- Regression is applied to two different periods. Hyper parameter tuning is done for the second period while on the third period only splitting of dataset is done.
- We came to a conclusion that all the input features contribute to each dependent variable(targets).
- When hyper parameters are at its best value, MSE is low which means our model fits pretty good.

# References


- Using mixture design and neural networks to build stock selection decision support systems; Yi-Cheng Liu1 • I-Cheng Yeh2.
- ANN Model to Predict Stock Prices at Stock Exchange Markets; Article August 2016, Barack Wamkaya Wanjaw University of Nairobi • Lawrence Muchemi University of Nairobi.
- Stock Closing Price Prediction using Machine Learning Techniques; Article 2020, Mehar Vijha • Deeksha Chandolab • Vinay Anand Tikkiwalb • Arun Kumarc.
- Study of Machine learning Algorithms for Stock Market Prediction; Vol. 9 Issue 06, June-2020, Ashwini Pathak MPS in Analytics • Sakshi Pathak B.Tech. Information Technology.


# *THANK YOU*