# **UCS1603 Introduction to Machine Learning**

**Stock Price Prediction**

**Abstract**

Prediction of stock prices has been an important area of research for a long time. Prediction of stock prices helps the user analyse when they can buy a certain company’s stocks or sell them, in some cases doing nothing with the stock is also a valid option. In this project, we are predicting the closing price of a certain company’s stock on a particular day using the closing prices for the past 2 months.

**Dataset**

Current Stock Prices of a particular company for which we need to predict Dataset in the form of

[ open, high, low, close, volume, %change }

Open - stock price at the start of the day

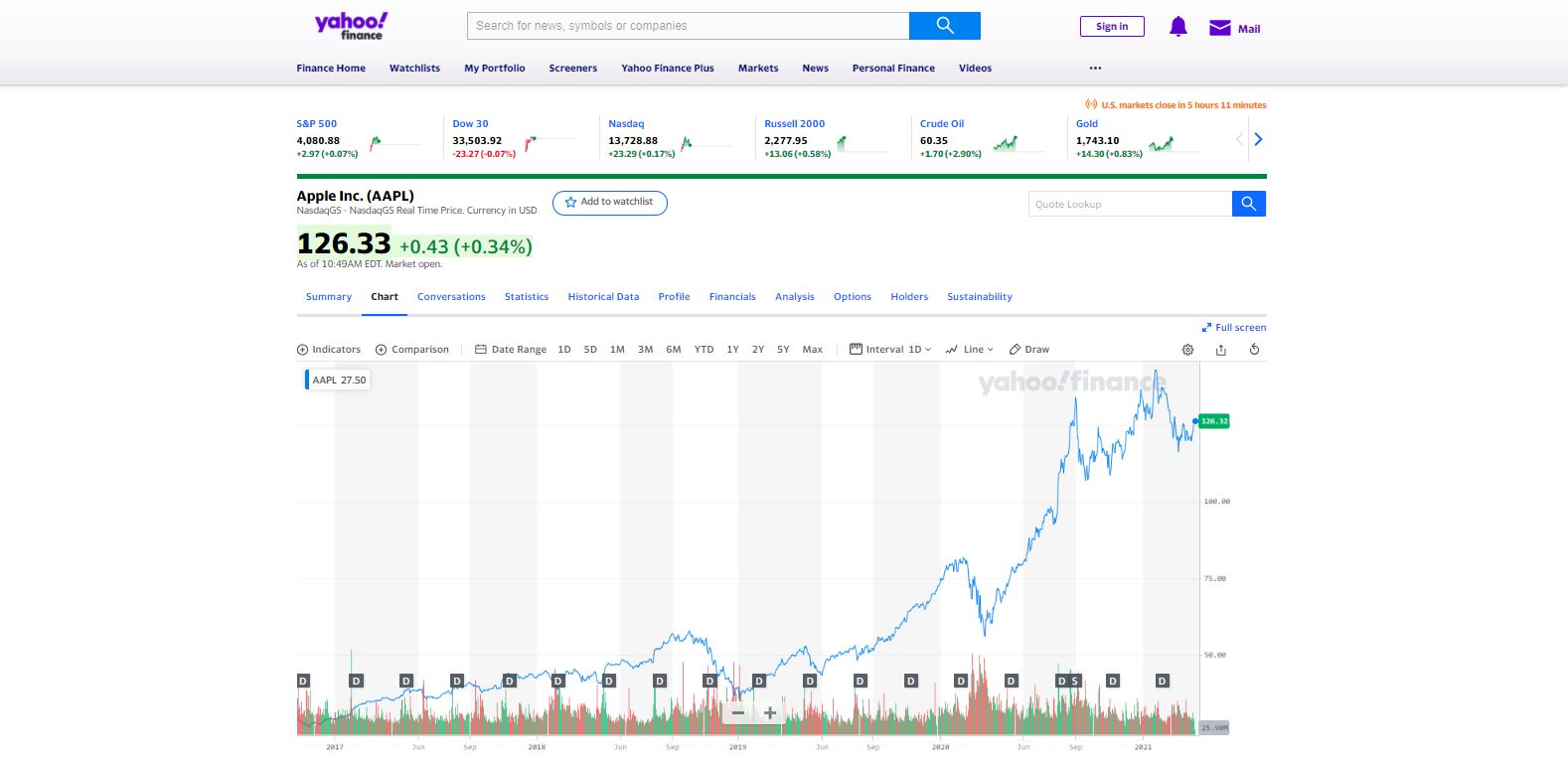
High - highest recorded price on that day

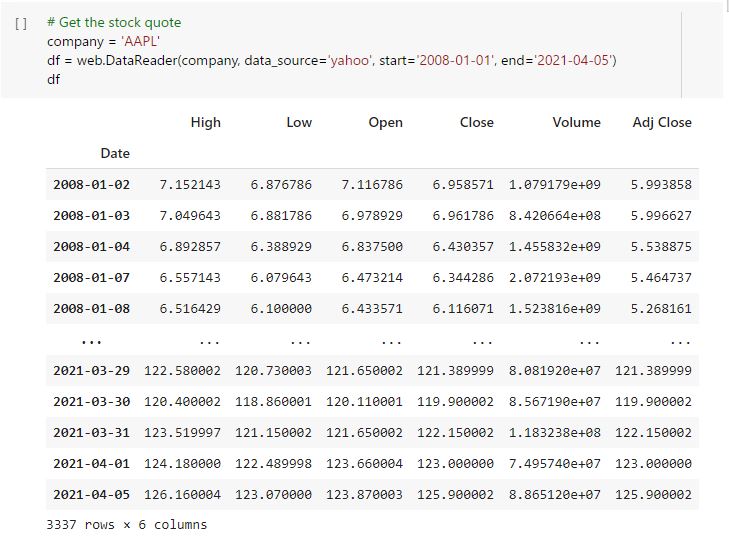
Low - lowest recorded price on that day

Close - stock price at the end of that day

Volume - total investment on that day

%change - percentage change in the stock price with respect to day before

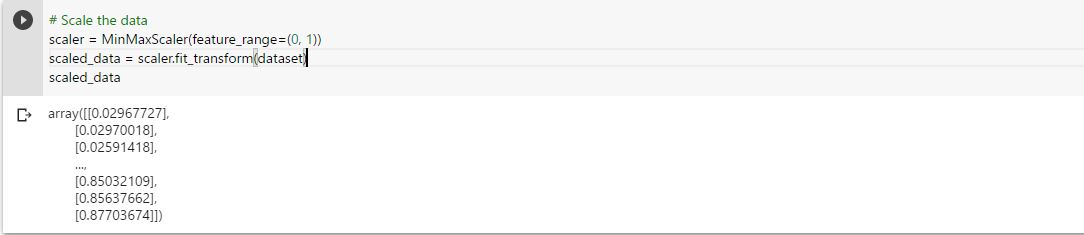




**Preprocessing**

Machine learning algorithms just see numbers. If there is a vast difference in the range, say few ranging in thousands and few ranging in the tens, and it makes the underlying assumption that higher ranging numbers have superiority of some sort. So these more significant numbers start playing a more decisive role while training the model.

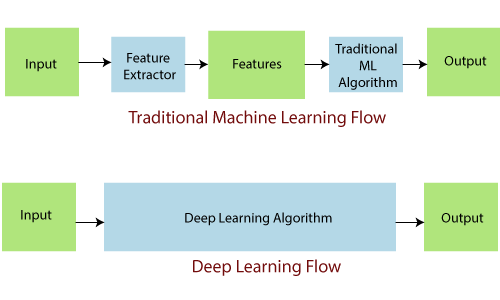
Here, the closing prices might be in a wide range and in order to bring them within a certain range [0, 1], we use MinMaxScaler.

****

**Methodology**

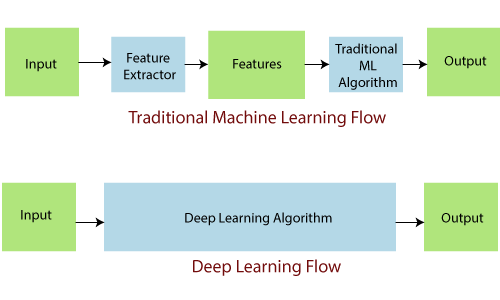
**Linear regression**

Linear regression is a linear model, e.g. a model that assumes a linear relationship between the input variables (x) and the single output variable (y). More specifically, that y can be calculated from a linear combination of the input variables (x).

****

**LSTM**

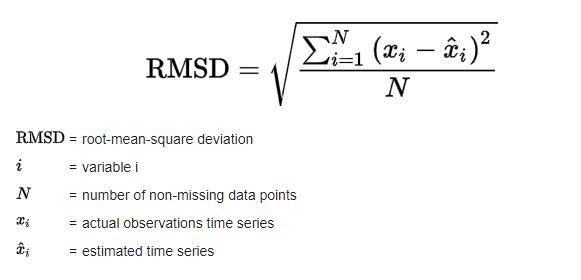
Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning Order dependence in sequence prediction problems. This is a behaviour required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning.

****

**Performance Metrics**

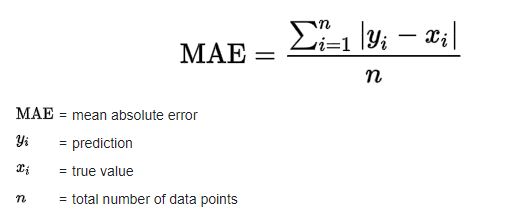
**RMSE**

Root Mean Square Error(RMSE) is the standard deviation of the residuals (predictionerrors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of the best fit.



**MAE**

Mean absolute error (MAE) measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures *accuracy* for continuous variables. The equation is given in the library references.



**R-Squared**

R-squared is a statistical measure of how clase the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.

R-squared = 1 - (First Sum of Errors / Second Sum of Errors)

**Inference**

****

We hereby infer that the linear regression model performs better than the LSTM model for the given dataset. And therefore it is efficient in predicting the stock prices.

This will help the user in understanding how the future trend will be and make appropriate choices in his upcoming investments.