Apple Stock Prediction

Yuba Sigdel & Irene Eshun
UR Data Analytics Bootcamp
Final Project

Problem Statement

How to predict the Next day Stock Price using Previous day Adjusted Closing Price?

- When should you buy an Apple Stock?

Goal

Predict the Next Day Stock Price for Apple Inc. Using Linear Regression.

Given prices for the last N days, we train a model, and predict for day N+1

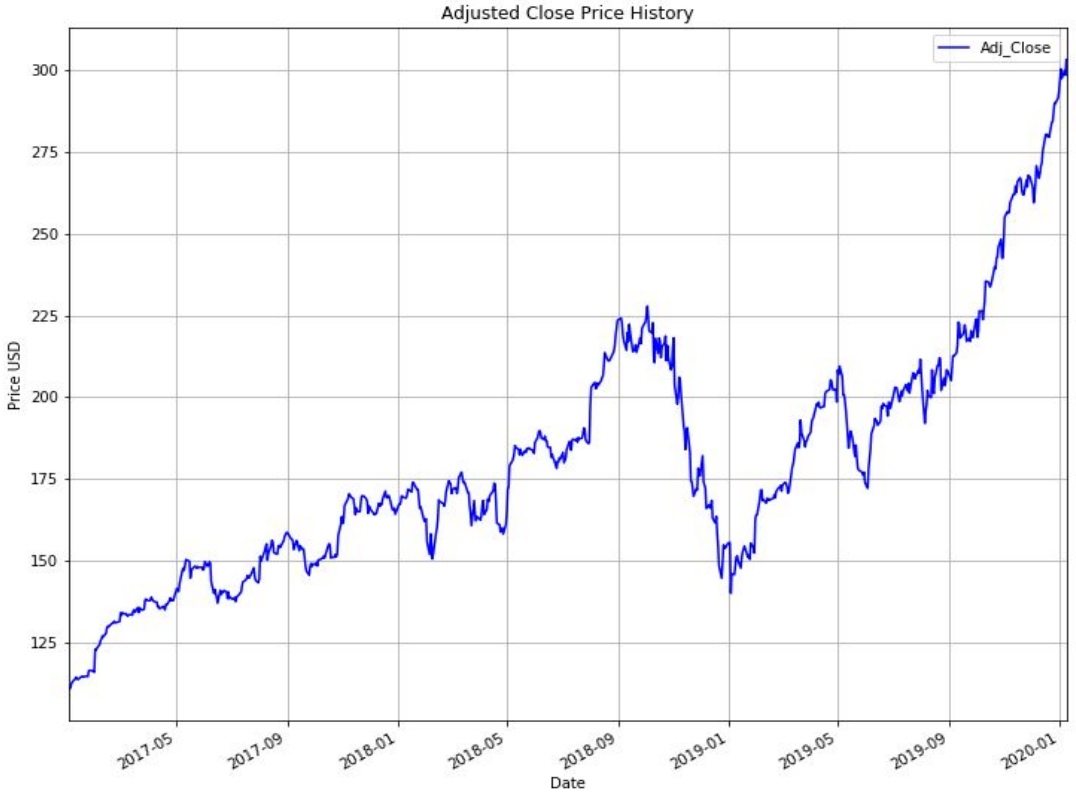
Data Source and ETL

1. Yahoo Finance
2. SEC Site
3. Web Scraping
4. Datetime

Our approach predict daily adjusted closing price of Apple stocks using data from previous N day. We used data from 1/1/2107 to 1/8/2020 which was downloaded from Yahoo Finance.

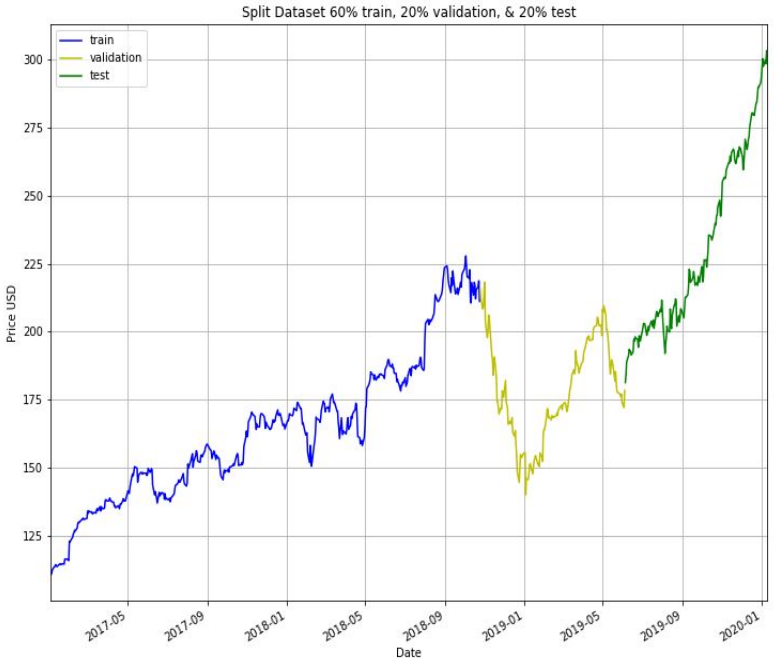
	Date	Open	High	Low	Close	Adj_Close	Volume	Month
0	2017-01-03	115.800003	116.330002	114.760002	116.150002	110.953873	28781900	1
1	2017-01-04	115.849998	116.510002	115.750000	116.019997	110.829704	21118100	1
2	2017-01-05	115.919998	116.860001	115.809998	116.610001	111.393303	22193600	1
3	2017-01-06	116.779999	118.160004	116.470001	117.910004	112.635139	31751900	1
4	2017-01-09	117.949997	119.430000	117.940002	118.989998	113.666824	33561900	1

We plotted the target column “Adjusted Close” to understand how it’s shaping up in our data.

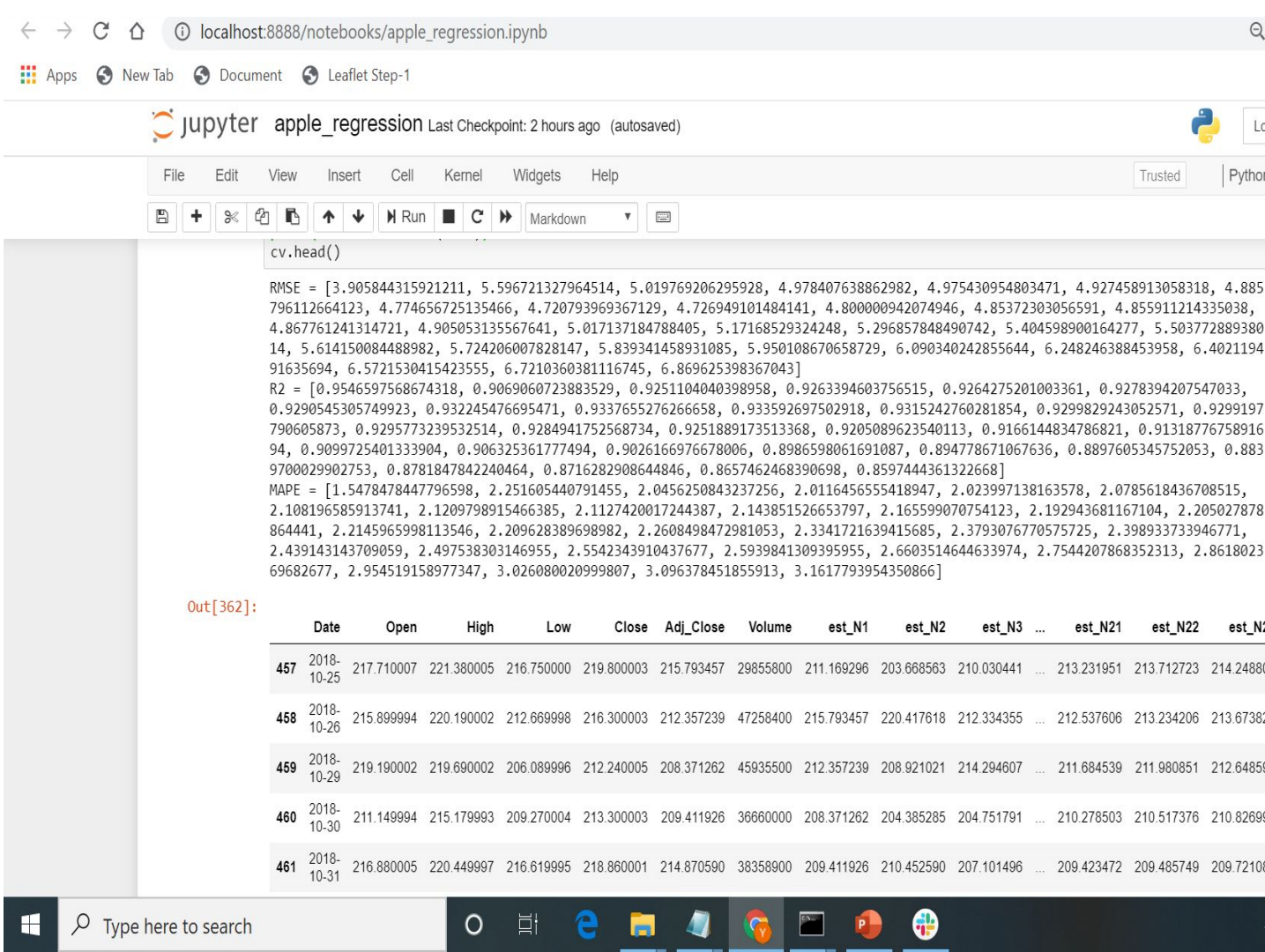


Preprocessing

- **Train 60%**
  Where model will be trained
- **Validate 20%**
  Where hyperparameters will be tuned.
- **Test 20%**
  Where performance of the model will be reports



This is the result of 30 sample cross-validation dataset including RMSE,R2 & MAPE score.



❖ Evaluation Metrics

RMSE (root mean square error) the lower the value, the better the performance.

MAPE (mean absolute percentage error) lower the value, the better the performance.

R2 the higher the value, the better the performance.

Linear regression

Linear regression is a linear approach to modeling the relationship between a dependent variable and one or more independent variables

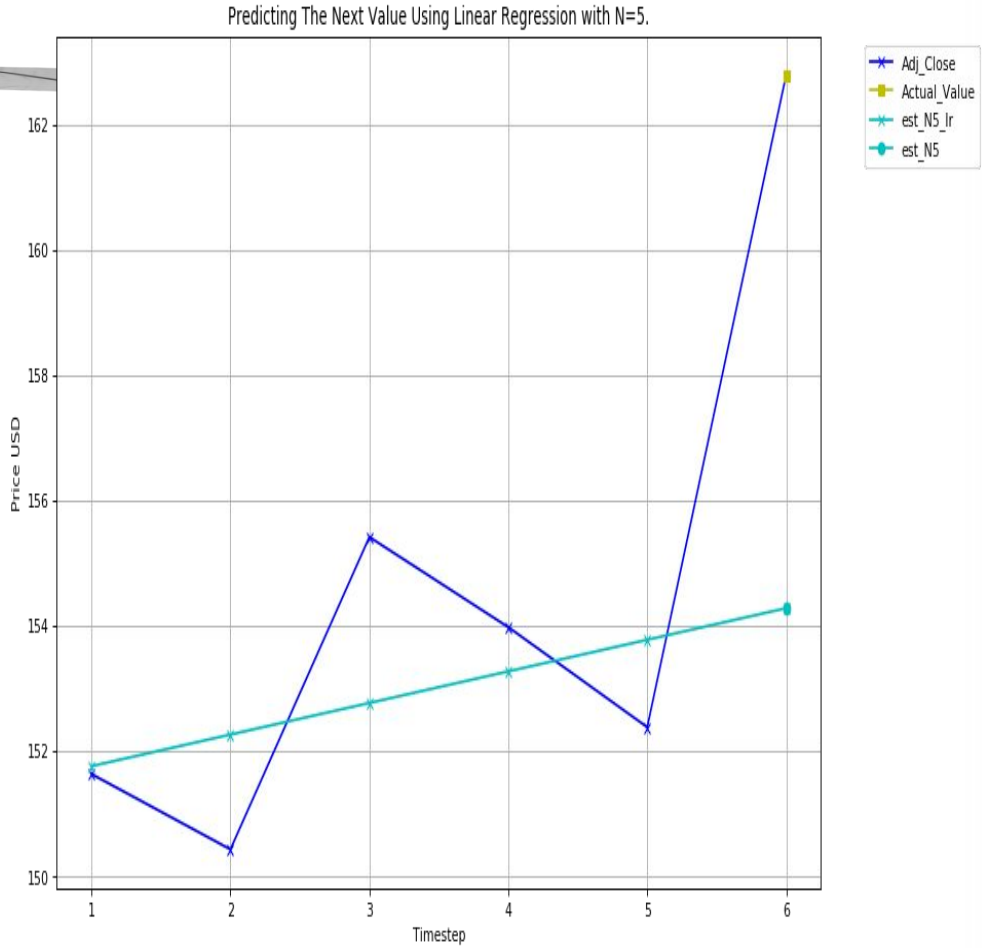
we will fit a linear regression model to the previous N values, and use this model to predict the value on the current day.

This plot is an example for N=5

#N = number of samples to use to predict the next value (N-fold)

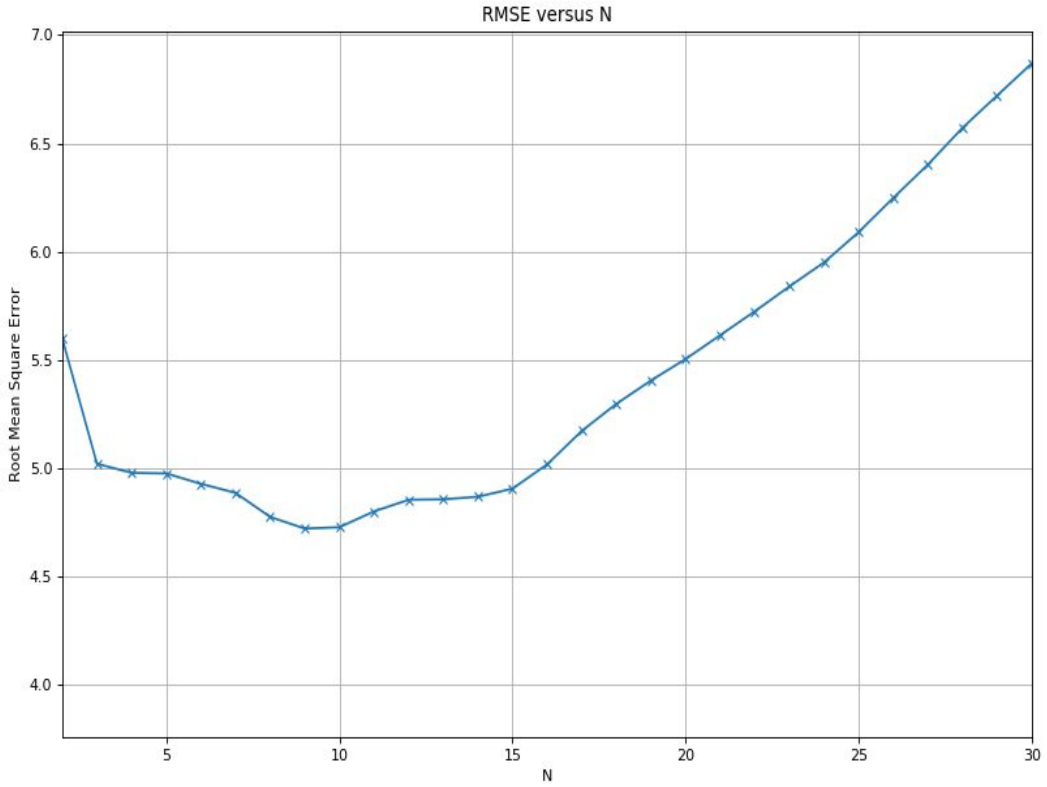
The actual adjusted closing prices are shown as dark blue cross, and we want to predict the value on day 6 (yellow square).

We fit a linear regression line (light blue line) through the first 5 actual values, and use it to do the prediction on day 6 (light blue circle)



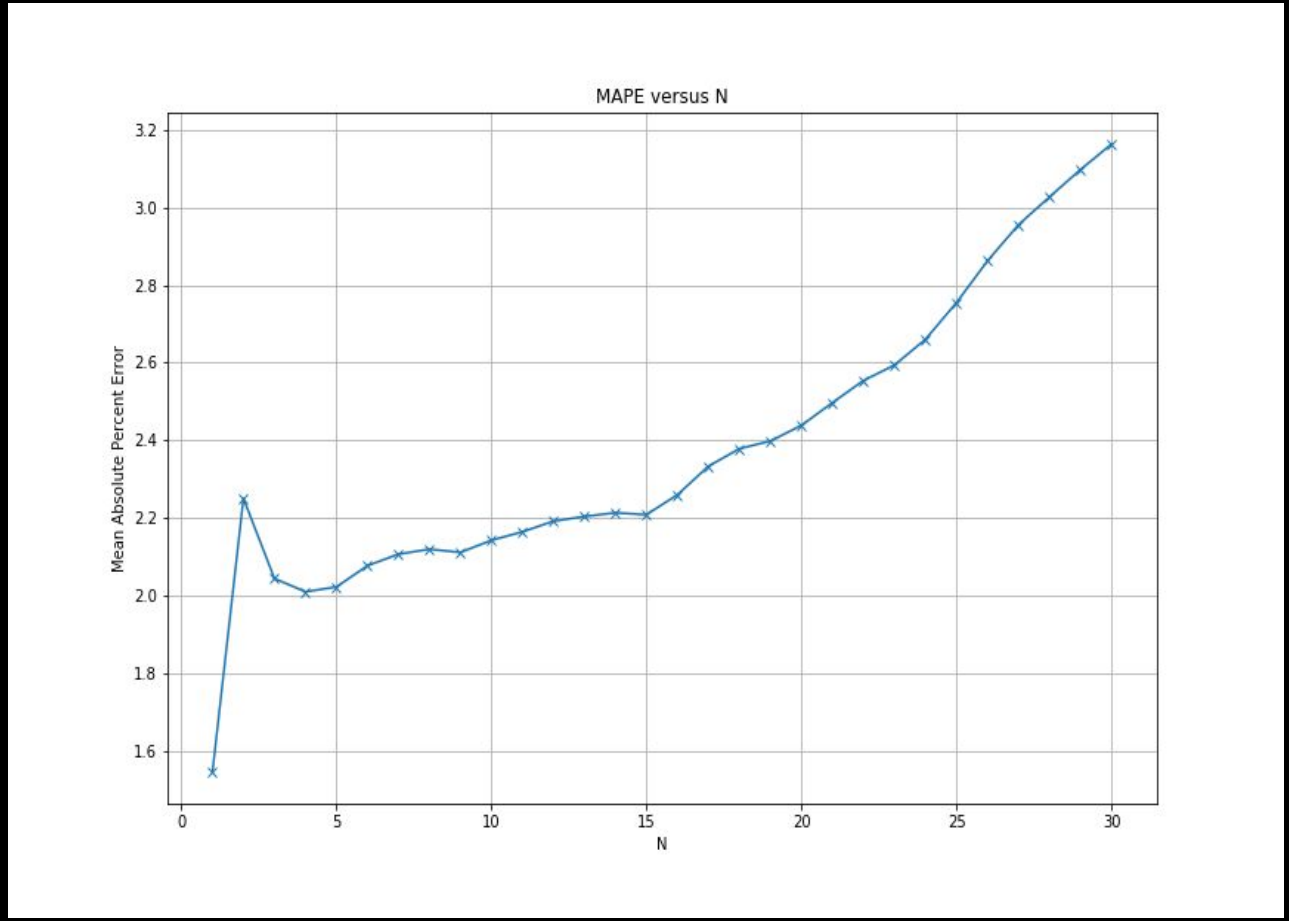
This plot shows the RMSE between the actual and predicted values on the validation set, for various values of N.

We going to use N=5 since gives the lowest RMSE.

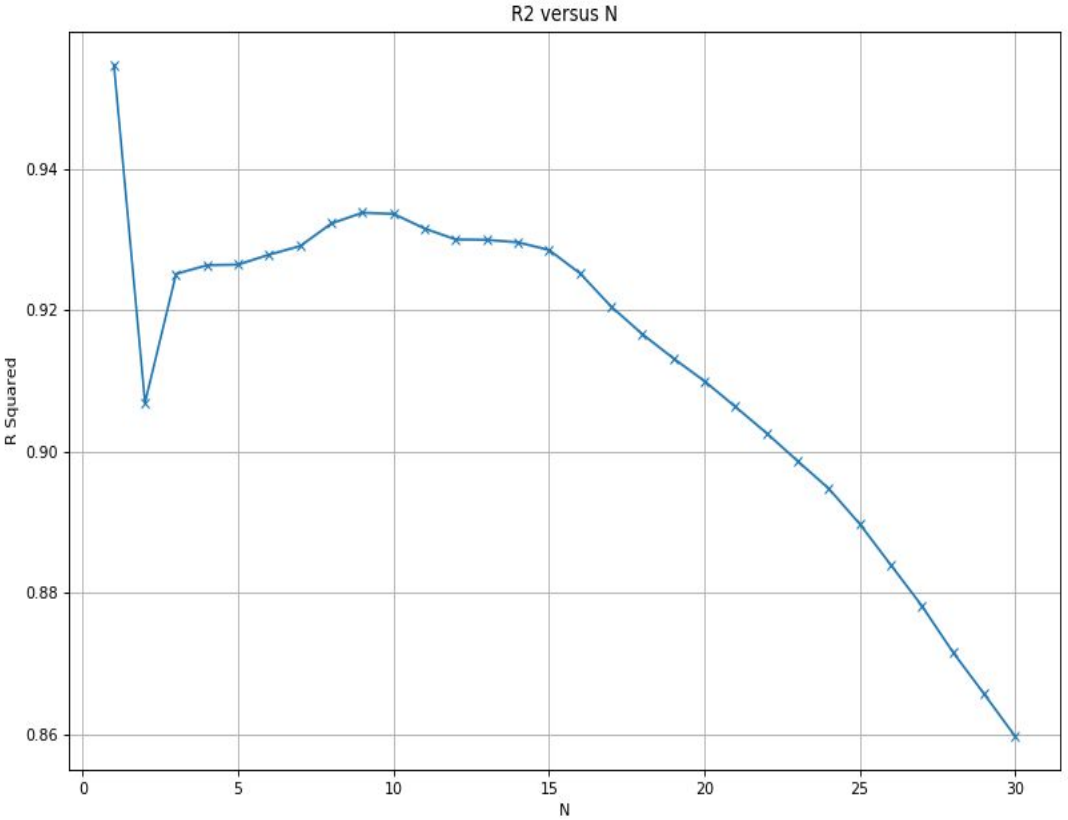


➔ This plot shows the MAPE between the actual and predicted values on the validation set, for various values of N.

We going to use N=5 since it gives the lowest MAPE.

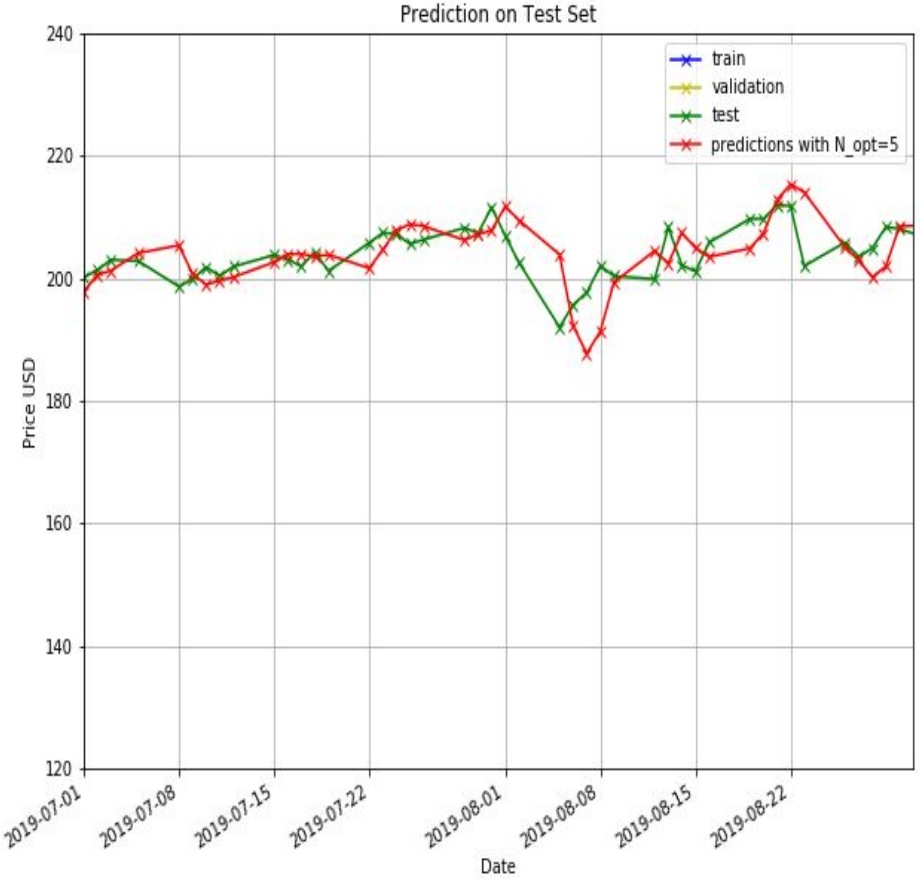


This plot shows the R2 between the actual and predicted values on the validation set, for various values of N.



For test set we pick test data of range of date(2019, 2, 1 to 2019, 3, 31).

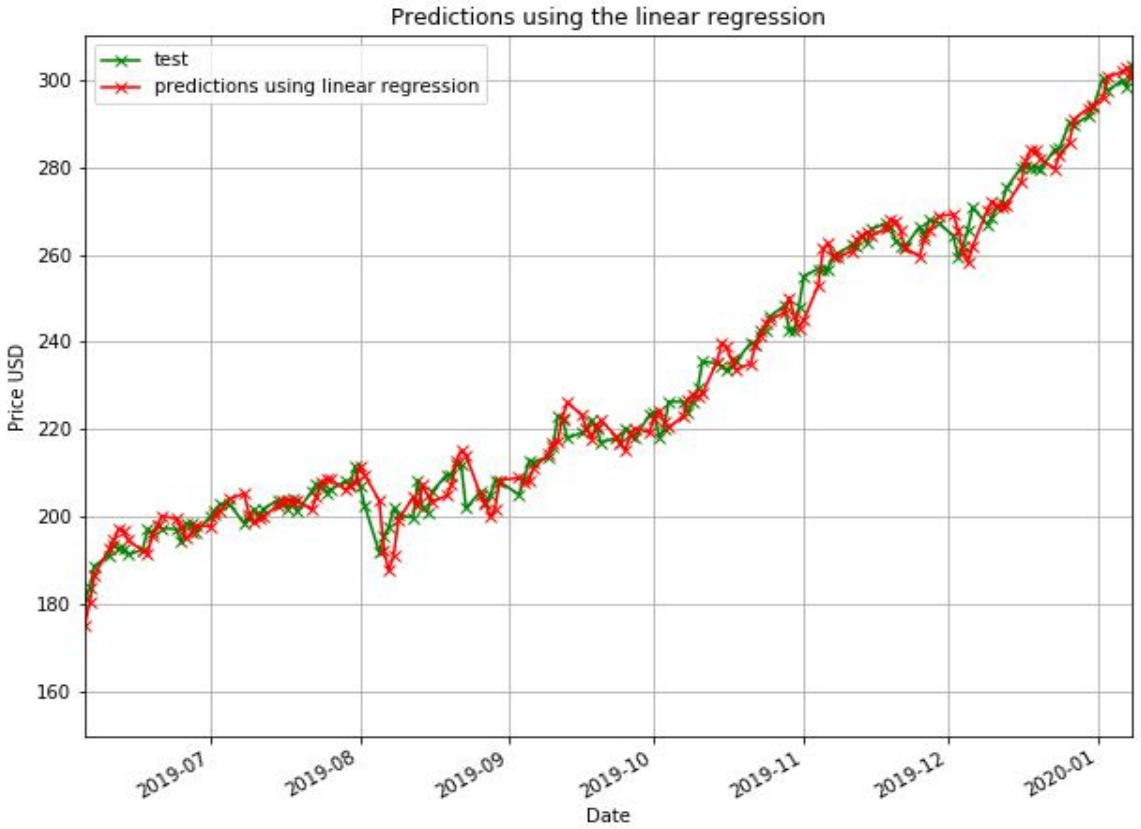
We Set optimum N= 5 for this plot.



Story for illustration purposes only

This plot shows the predictions using linear regression method.

It can be observed that this method does not capture changes in direction (ie. downtrend to uptrend and vice versa) very well.



Findings

The lowest RMSE is 3.905 On the validation set. It came when using N=1, ie. using value on day t-1 to predict value on day t.

The hightest R2 is 0.954 on the validation set. It came when using N=1, ie. using value on day t-1 to predict value on day t.

We use N_opt=5 because our goal is to use linear regression

The RMSE is 3.953 on the test set using N_opt=5.

The MAPE is 1.370% on the test set using N_opt=5.

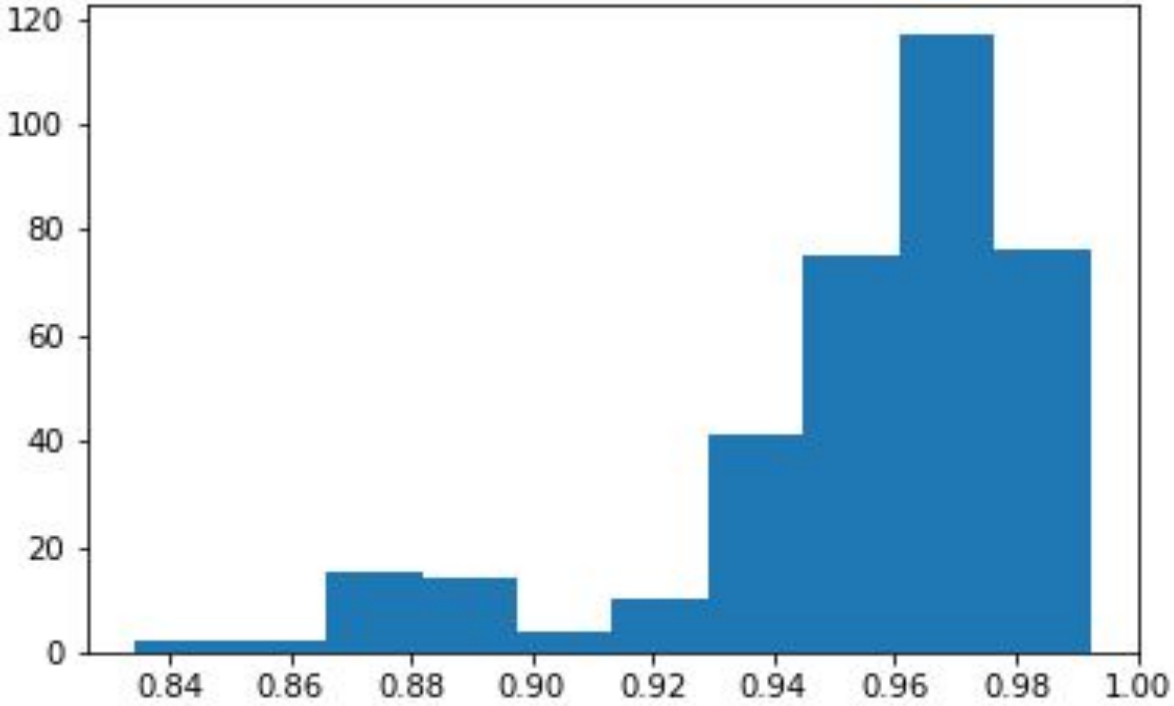
The R2 is 0.985 on the test set using N_opt=5.

Predicting Next Day Stock Price for Apple Inc. Using Random Forest Regression

Random Forest Regression

Random forest algorithm creates multiple decision trees and merges them together to obtain a more stable and accurate prediction..

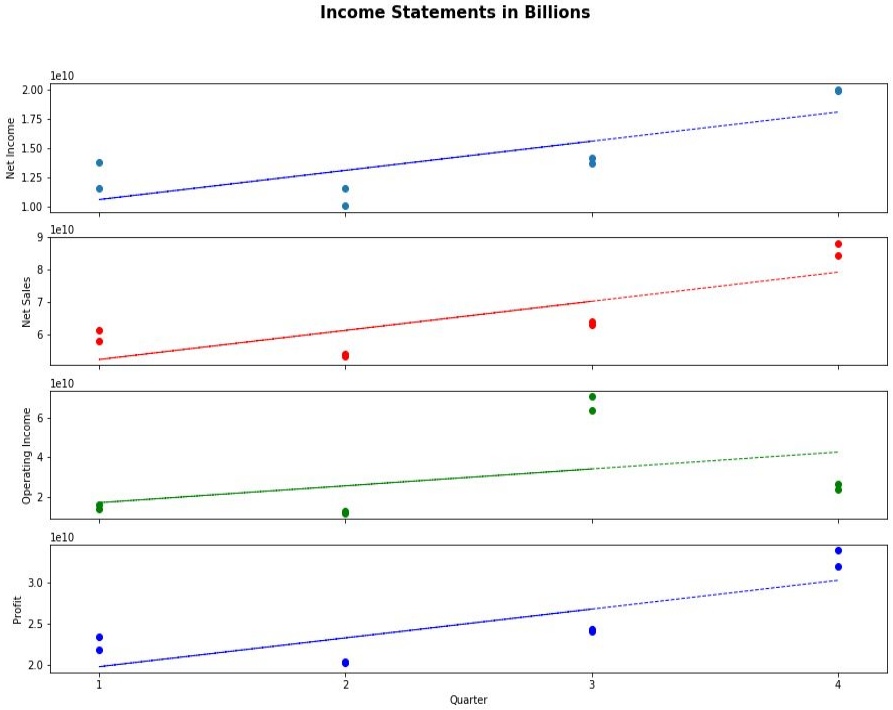
Every decision tree forecasts a response for an occurrence and the endmost response is decided through voting

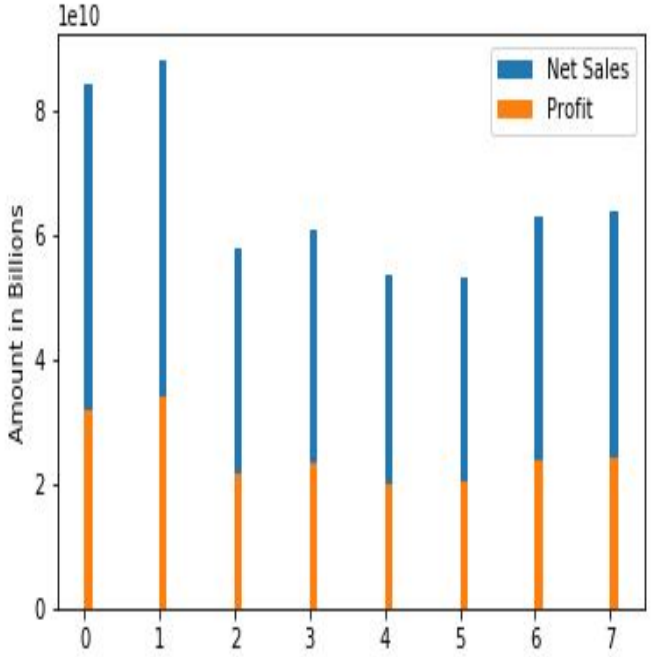


R2 = 0.98

MSE = 4.64

Income Statement Analysis





Moving Forward/Lessons

A. Create a graph for Random forest score over time
B. The less samples, the less accurate the R2 and MSE
C. Long Short Term Memory (LSTM): uses selective remembering patterns for long durations of time. This will help determine the patterns in this stock typical behavior