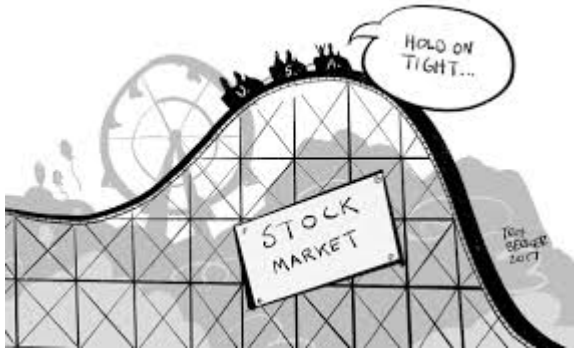
Stock Market Prediction Using Different Machine Learning Algorithms



Vishesh Jain (2K18/CO/391)
Vishu Dhama (2K18/CO/394)

Motivation

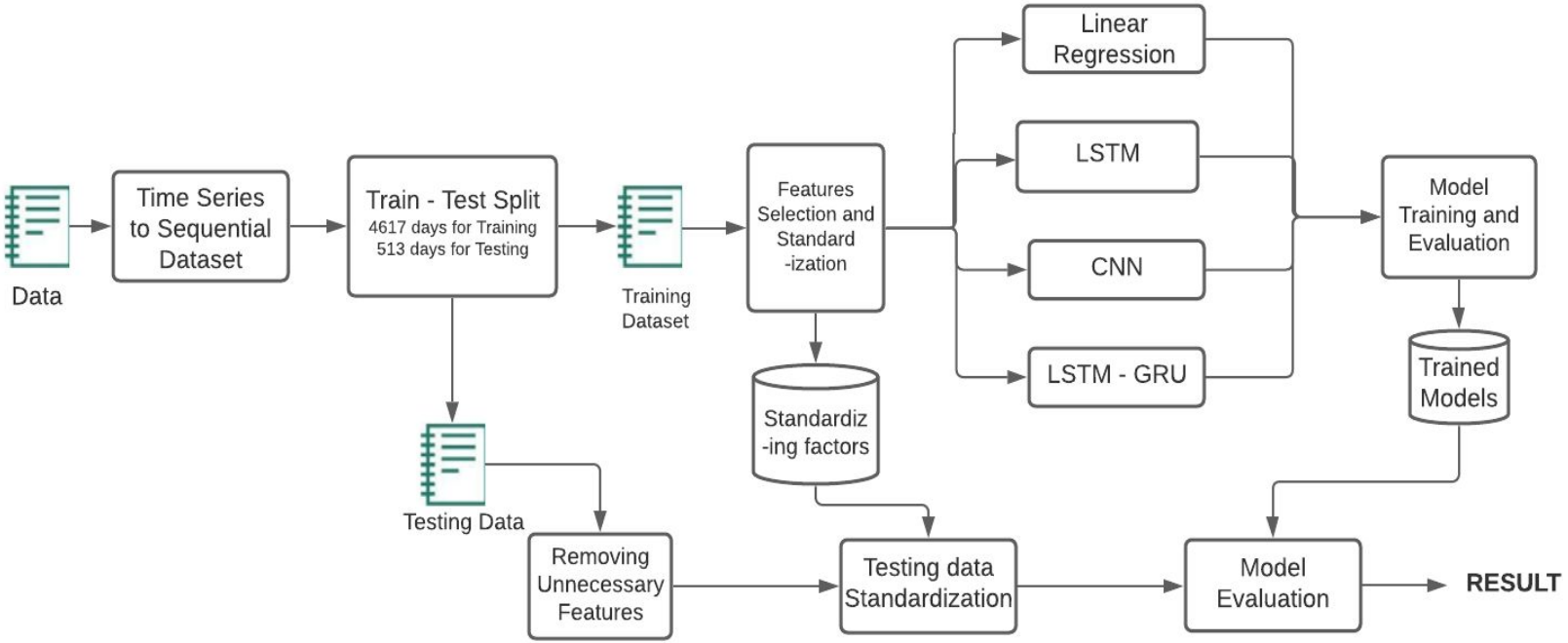
Economy Perspective :

Movements in the stock market can have a profound economic impact on the economy and individual consumers. A collapse in share prices has the potential to cause widespread economic disruption. Most famously, the stock market crash of 1929 was a key factor in precipitating the great depression of the 1930s.

Individual Perspective ：

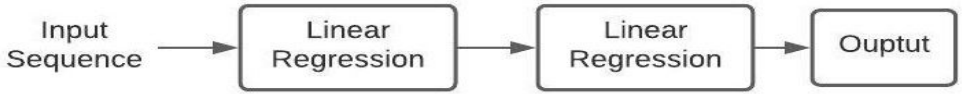
Stock markets promote investment. The raising of capital allows companies to grow their businesses, expand operations and create jobs in the economy. This investment is a key driver for economic trade, growth and prosperity.

Experimental Design

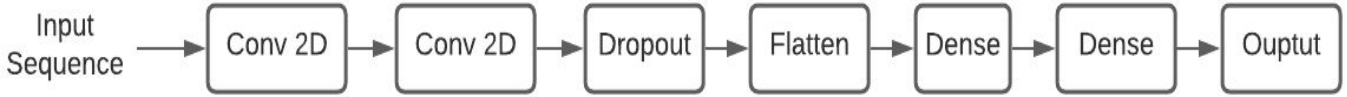


Model Architectures

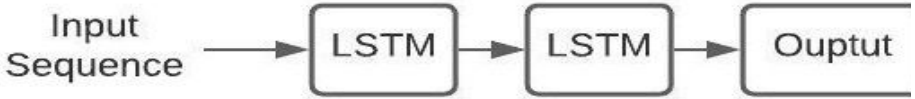
1. Linear Regression



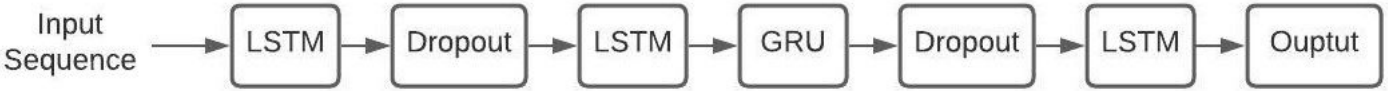
2. Convolutional Neural Network



3. Long- Short Term Memory Model



4. LSTM extended Model (along with Dropout and GRU layers)

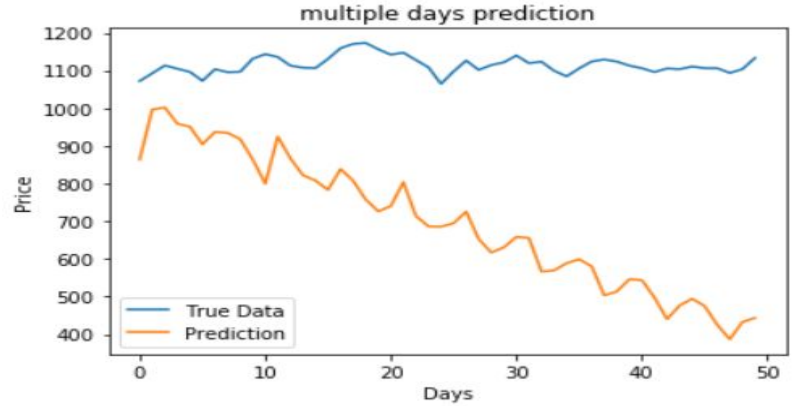


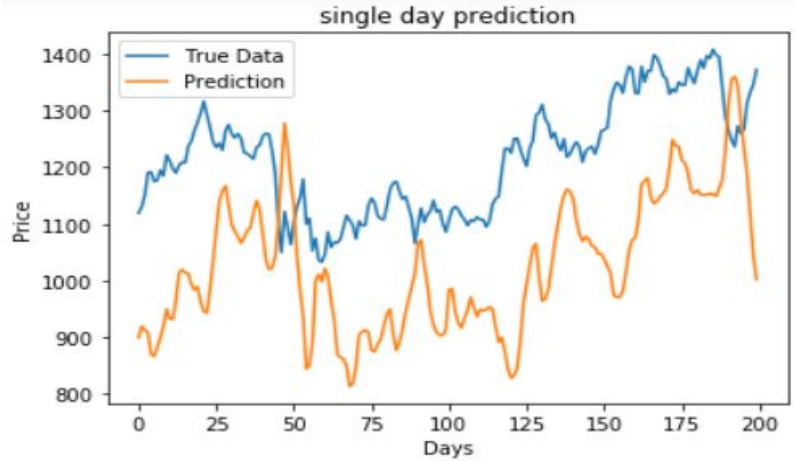
Linear Regression

No of Layers : 2
Layer 1 : crux of a day from OHLC Features
Layer 2 : Closing price from crux of past 10 days

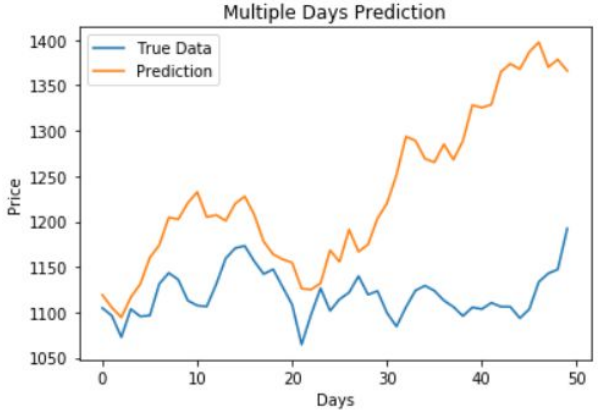
Days

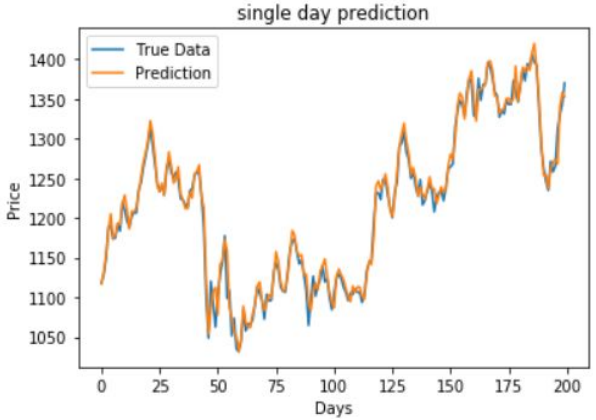
	1	2	3	4	5	10	50
MAE	0.388	0.136	0.141	0.194	0.228	0.407	1.309
RMS	0.442	0.165	0.17	0.224	0.256	0.446	1.389
MSE	0.195	0.027	0.029	0.05	0.067	0.199	1.929
R2	-1.513	0.645	0.633	0.375	0.174	-1.288	-8.101





Convolutional Neural Network Model

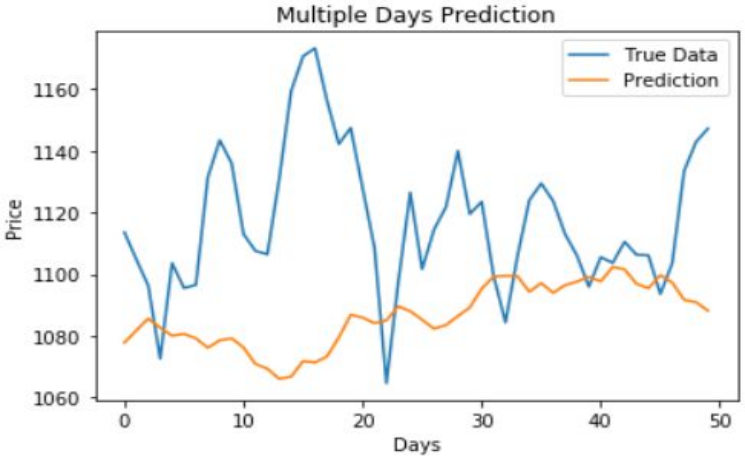


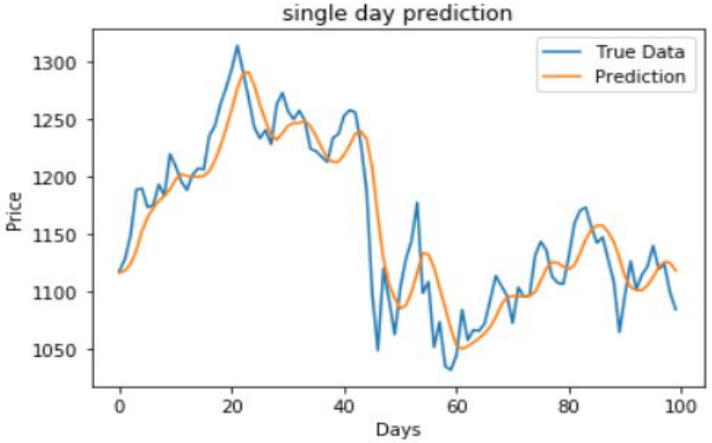


	1	2	3	4	5	10	50
MAE	0.02	0.045	0.06	0.072	0.083	0.128	0.466
RMS	0.031	0.062	0.082	0.1	0.117	0.177	0.571
MSE	0.001	0.004	0.007	0.01	0.014	0.031	0.326
R2	0.988	0.951	0.911	0.866	0.818	0.587	-1.693

Long-Short Term Memory Model

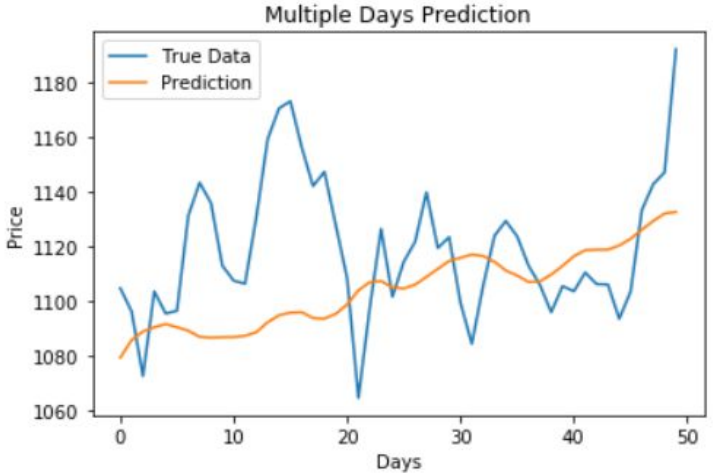
	1	2	3	4	5	10	50
MAE	0.048	0.062	0.077	0.091	0.105	0.166	0.433
RMS	0.065	0.082	0.1	0.119	0.137	0.214	0.591
MSE	0.004	0.007	0.01	0.014	0.019	0.046	0.35
R2	0.938	0.898	0.84	0.761	0.664	-0.162	-60.894

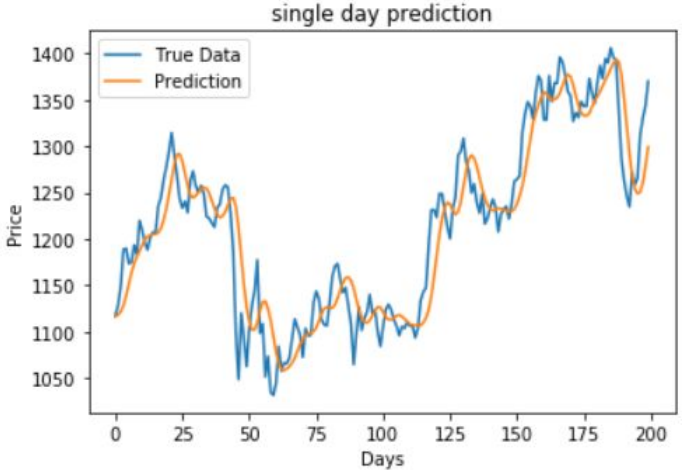




LSTM extended Model (along with Dropout and GRU layers)

	1	2	3	4	5	10	50
MAE	0.057	0.062	0.071	0.082	0.092	0.135	0.333
RMS	0.08	0.088	0.1	0.114	0.127	0.184	0.469
MSE	0.006	0.008	0.01	0.013	0.016	0.034	0.22
R2	0.917	0.902	0.875	0.839	0.803	0.615	-0.025





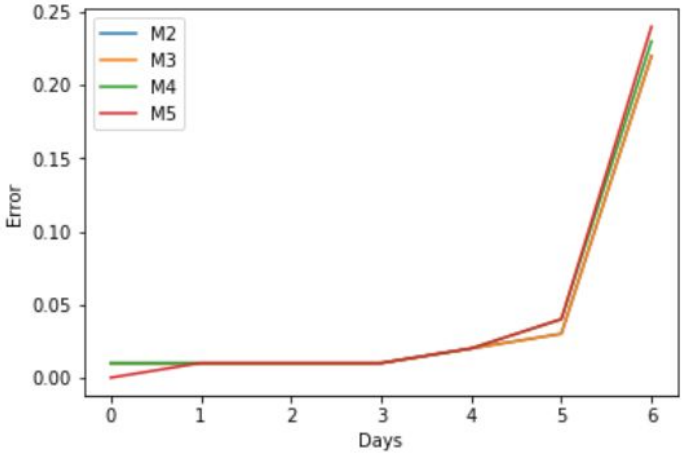
Variation of Sequence Length

Following are the Performance of different sequences :

M1 : Model with Sequence Length 5
M2 : Model with Sequence Length 10
M3 : Model with Sequence Length 15
M4 : Model with Sequence Length 20

Model	1	2	3	4	5	10	50
M1	0.91	0.9	0.87	0.83	0.78	0.55	-0.48
M2	0.92	0.9	0.87	0.84	0.8	0.61	-0.03
M3	0.92	0.91	0.88	0.85	0.81	0.64	-0.03
M4	0.93	0.9	0.86	0.82	0.78	0.58	-0.09
M5	0.95	0.91	0.87	0.83	0.79	0.6	-0.12

R2 Score



Model	1	2	3	4	5	10	50
M1	0.01	0.01	0.01	0.01	0.02	0.04	0.32
M2	0.01	0.01	0.01	0.01	0.02	0.03	0.22
M3	0.01	0.01	0.01	0.01	0.02	0.03	0.22
M4	0.01	0.01	0.01	0.01	0.02	0.04	0.23
M5	0.0	0.01	0.01	0.01	0.02	0.04	0.24

Mean Squared Error

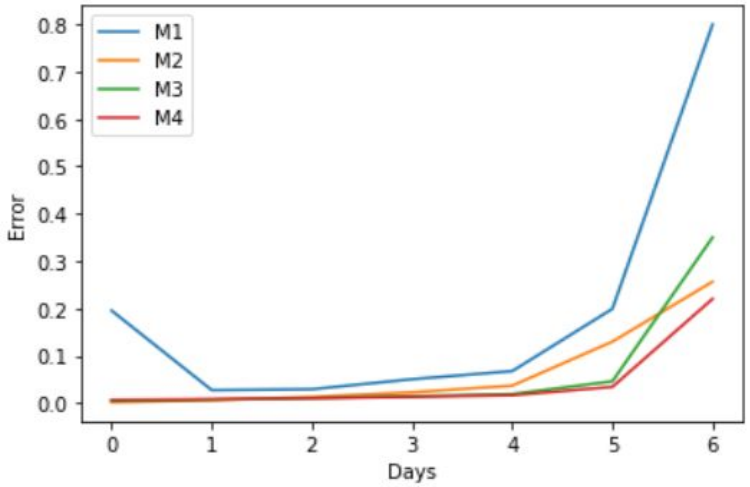
Result and Analysis

Linear regression model has high Mean Squared Error and negative R2 score but it can be seen that this model performs better for long term prediction as compared to next day prediction .

M2 model that is model with Conv2d layer outperformed for within 5 days prediction but on predicting the days further ahead it's loss in prediction increased.

Model with only Lstm and dense layers (M3) performed very well for next day prediction but when multiple days are predicted it performed even worse than Linear Regression Model .

Last model(M4) with additional dropout layers , and giving satisfactory performance for more than one days prediction and the reason may be that dropout layers can reduce the impact of recent days during prediction by dropping some percentage of neurons and that is why having a bit less accurate for first 5 days prediction .



Mean Squared Error