

Kriti Stock Market Prediction



Problem Statement

Building a Stock Price Prediction Model with Machine Learning and Deep Learning Techniques.

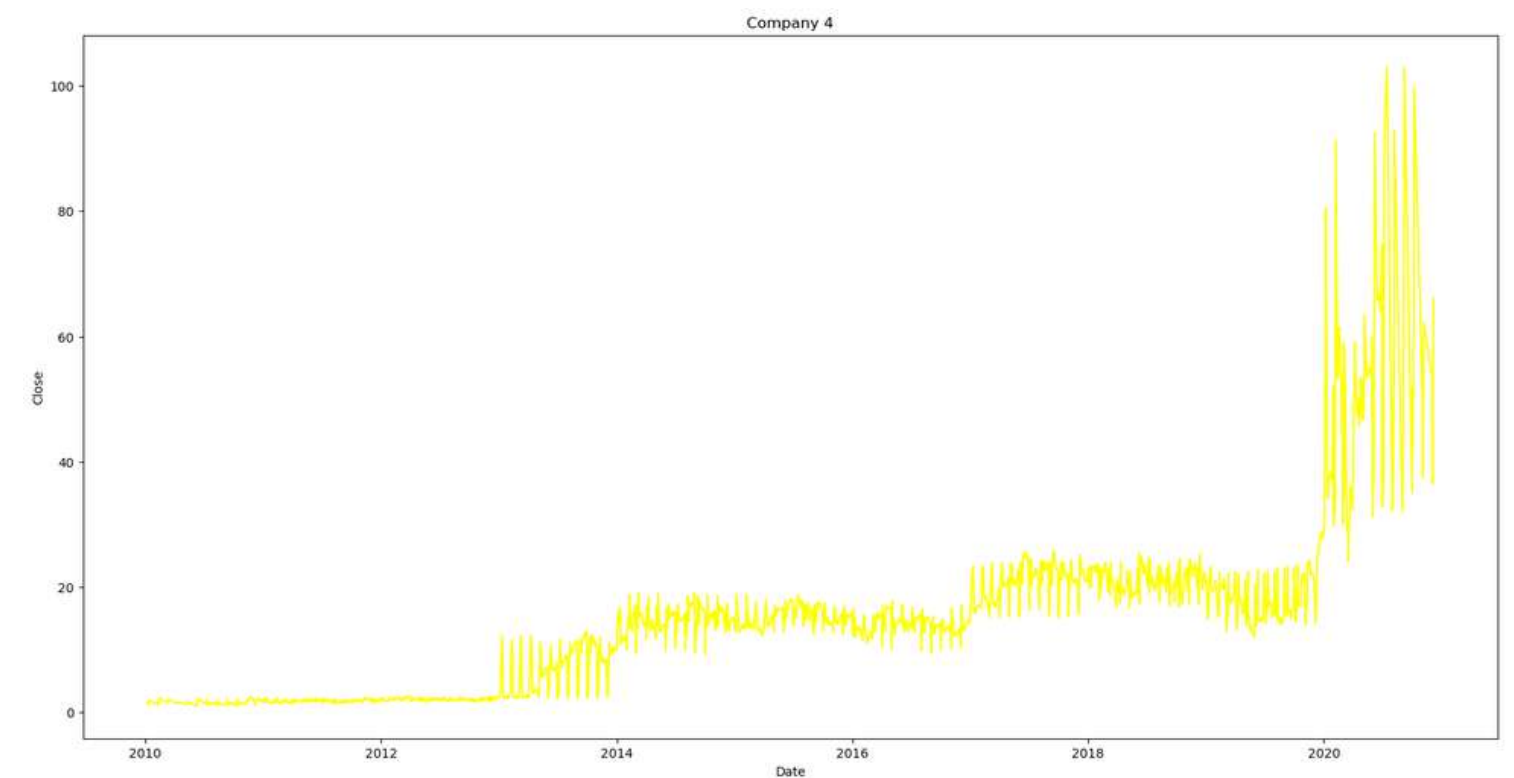
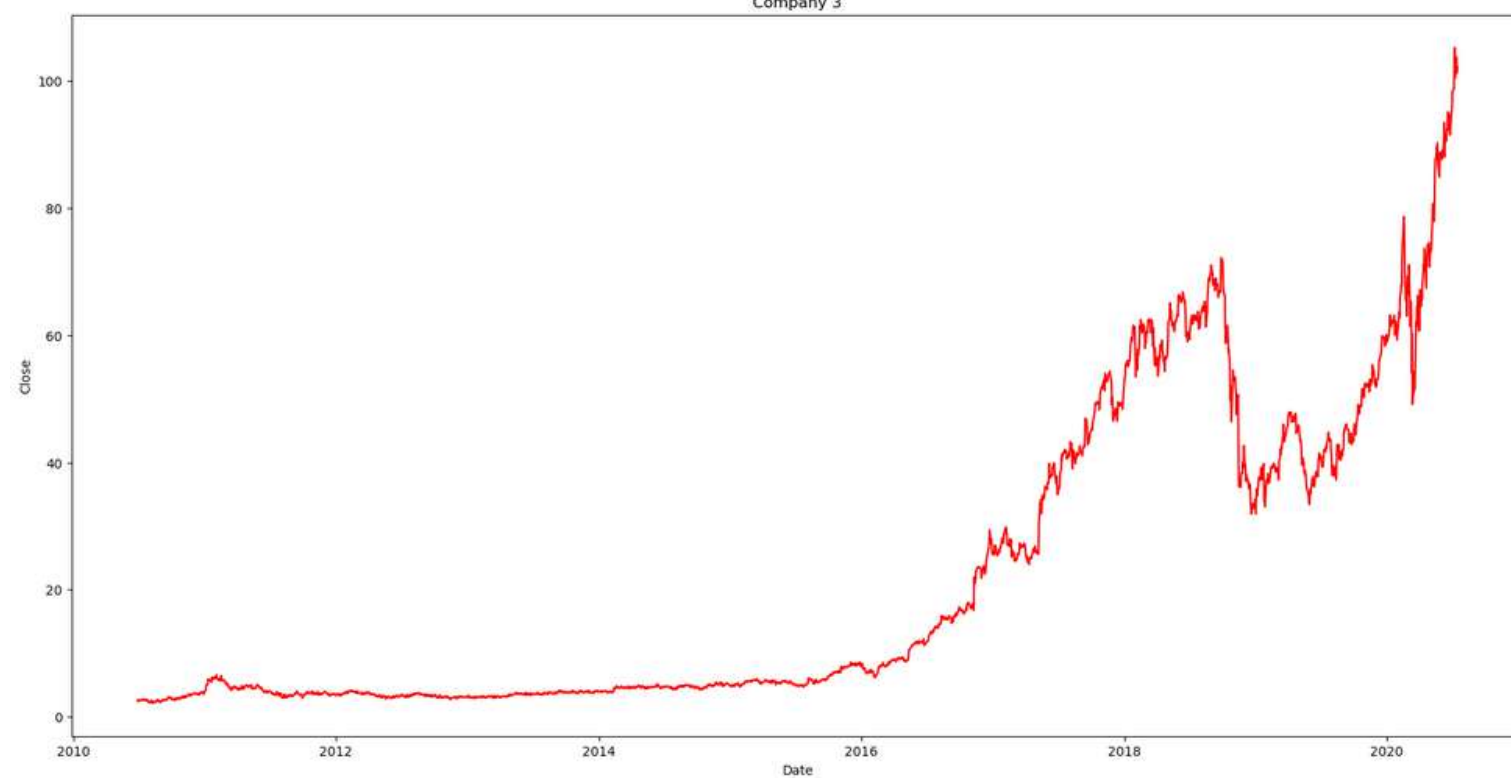
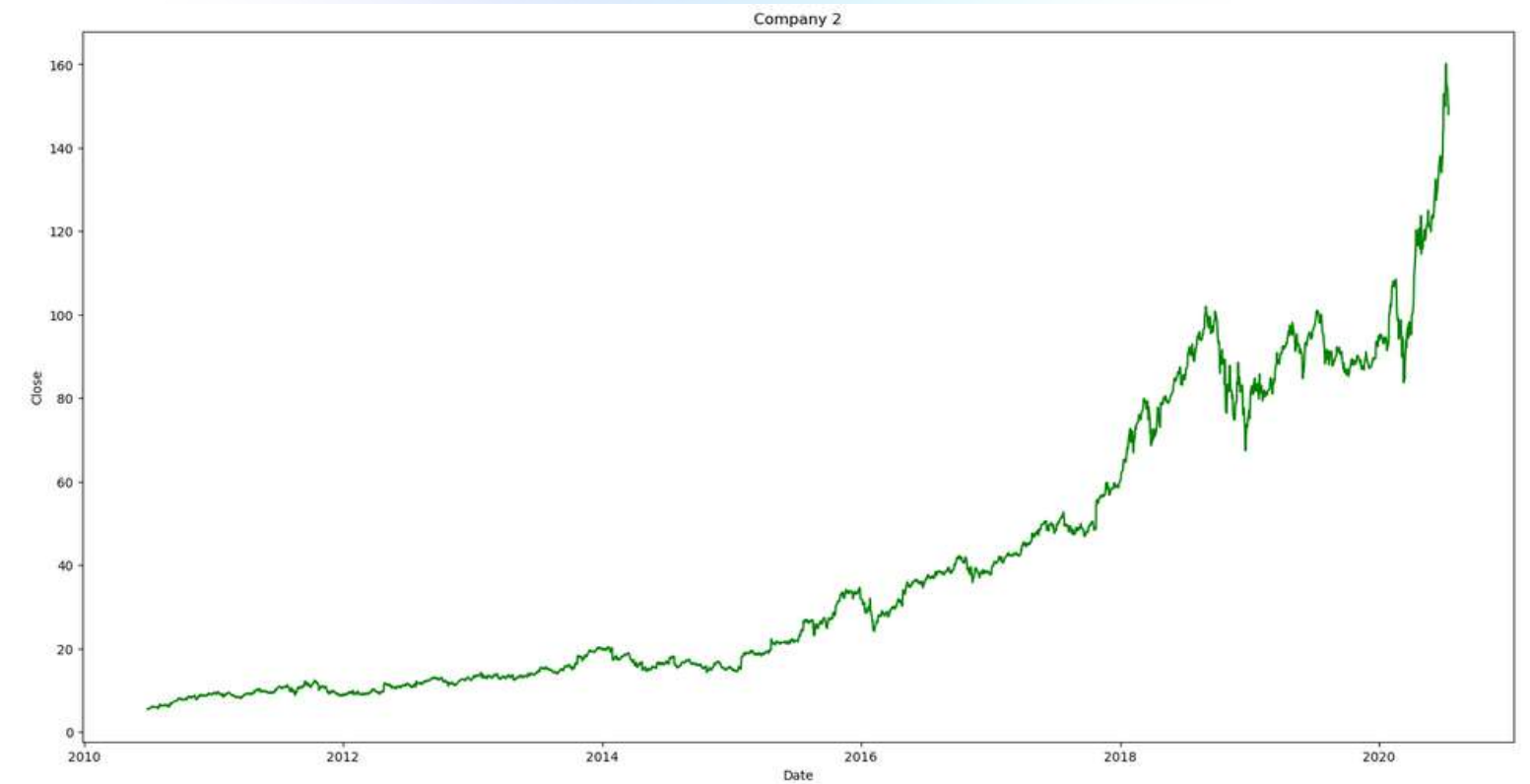
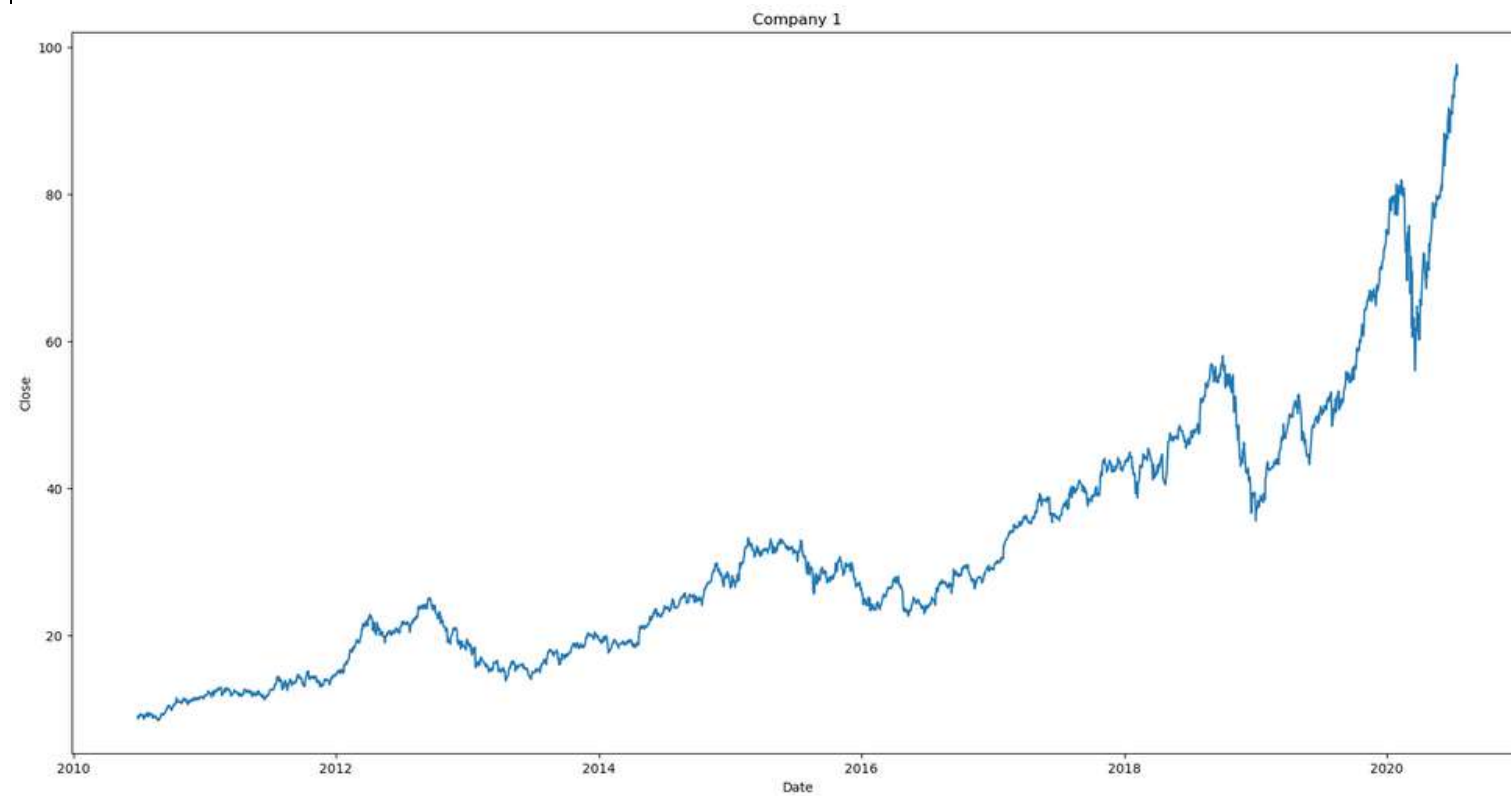
- DATASET DESCRIPTION:

- The dataset consists of train and test data for 4 companies.
- The features are present in the dataset are :
 - Date
 - Open
 - High
 - Low
 - Close
 - Adj Close
 - Volume.
 - Where 'Close' is the target column.



Data Visualization

||





MODELS USED

Linear Regression with just one feature

Gave Public LB Score 6.53,
Private LB score 6.21

Auto ML, SVRegressor

Didnt give good score in
validation dataset, so didn't
submit

Linear Regression with multiple features

Gave public LB score 6.50,
Private LB score 6.44

LSTM and GRU

LB score was around 50 for new
dataset.

XG Boost, Catboost, Random Forest

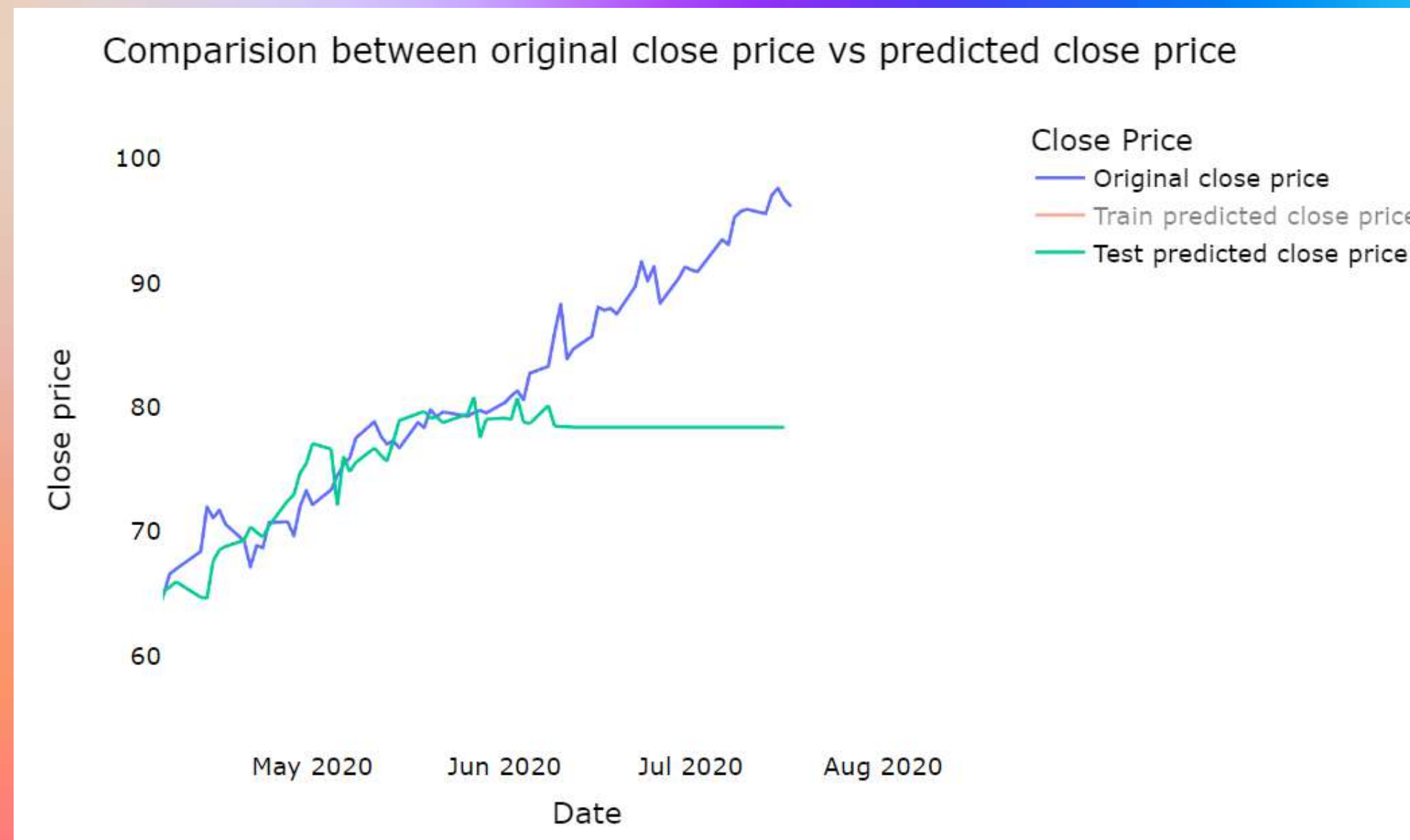
Didn't work well with the given
time series data

Lasso and Ridge regression

Gave similar scores as Linear
Regression, hence didn't use for
the final dataset.



Problem with Boosting Algorithms



FEATURES ADDED



Cyclic Time features

Features extracted from time stamp

Moving Averages

But was rejected by MI score

Other features

Lags on Adj Close values



Lag features

- Last 15 days:
 - Open
 - Adjusted Close
 - High & Low
 - Volume
- Sum of last 2 days Volume

Features extracted from timestamps

- Day of Week
- Date of Month
- Week of Year
- Month
- Year
- Quarter, Is_Month_Start, Is_Month_End, Is_Quarter_Start, Is_Quarter_End, Is_Year_End

Cyclic Features

- Month_sin, Month_cos
- Day_sin, Day_cos
- Date_sin, Date_cos
- Week_sin, Week_cos

FEATURES





FEATURES

Moving Averages

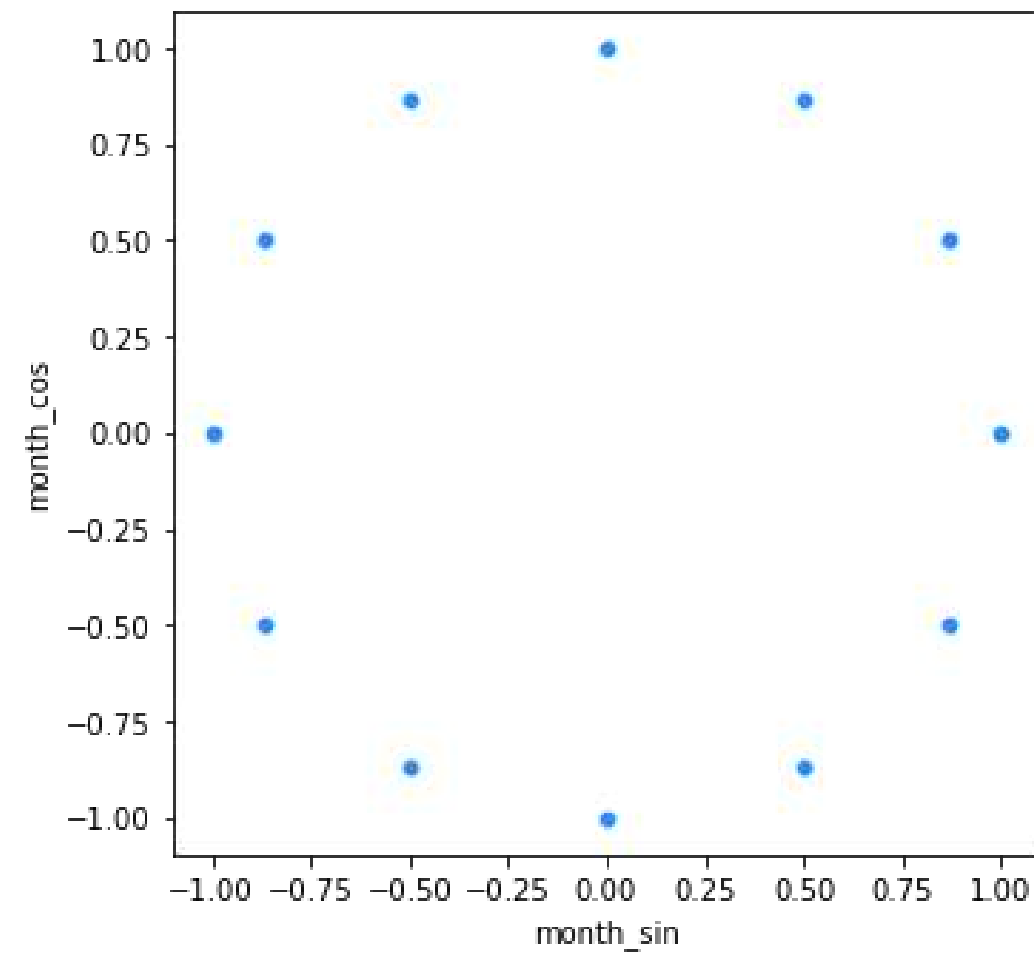
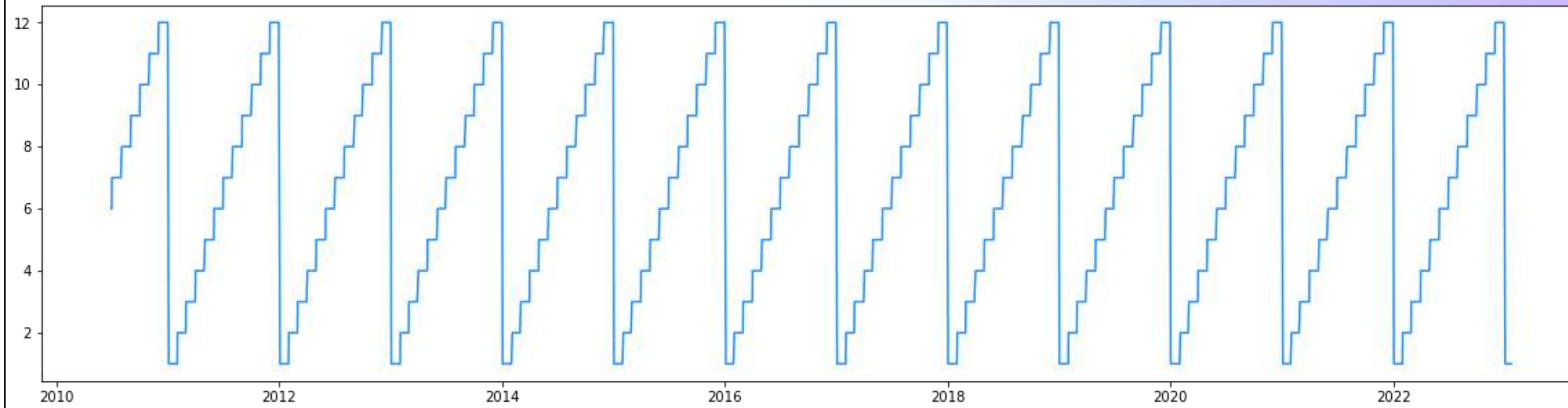
- Last 7 and 21 days MA of Adj Close
- MACD
- Bollinger Bands
- Exponential Moving Average
- Momentum

Other Features

- Scaled Volume
- Average of High and Low
- $\text{Sqrt}(\text{High} * \text{Low})$ (experimental)



CYCLIC FEATURES





**Linear Regression
was all we needed!**



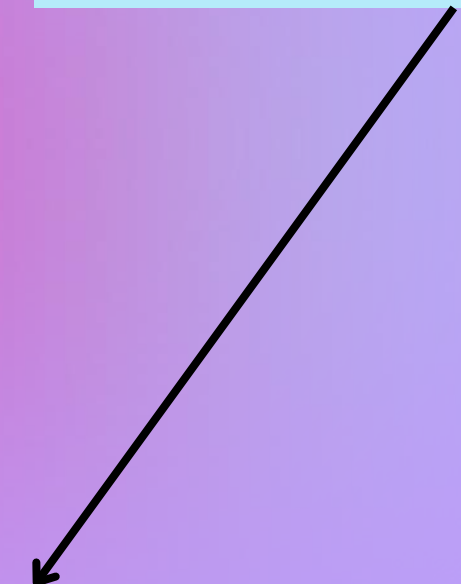
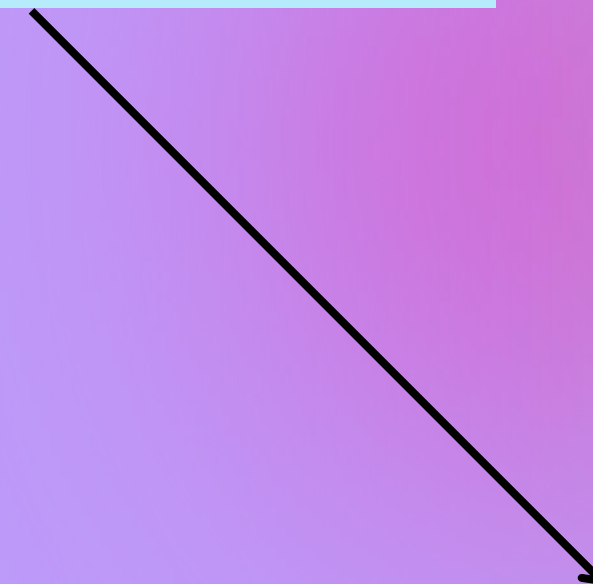
We used Buy Today, Sell Tomorrow (BTST) strategy.

**Next Level:
Ensembling !!**

**Linear Regression
with just one feature**

**Linear Regression
with multiple
features**

Final Submission





Stacking the regressor

Didn't improve the performance of base models

- **Because of overlapping predictions between base models**



A Second level model

Neural Network trained on predicted train values with true close as a feature.

Gave less valid score than the base models, so didn't go with it.



Weighted Average Ensembling

0.45

0.55



Thank You!