CASE STUDY OF Predicting Stock Prices For Large-cap Technology Companies

TIRTHANKAR HALDER

INTRODUCTION

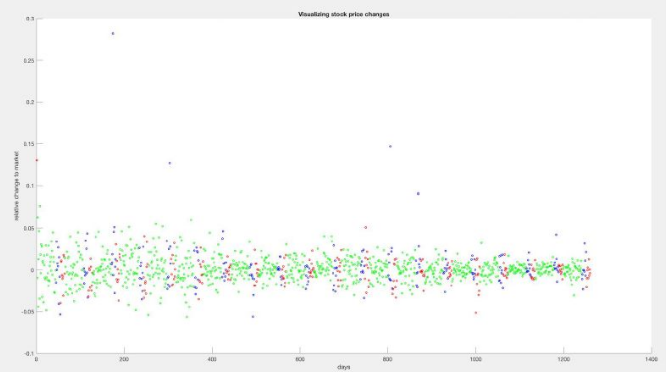
Why stock price prediction is important to us?

:- Companies are looking forward to gain the maximum return with modern technologies.In this way we can say that how the price is impacted ? and How to control the future stock behavior?

- “RISK HAI TO ISHQ HAI” by Harsh Mehta but in modern era the RISK FACTOR is not only influenced by the market value, also influenced by News(specifically news headline with punctuation),the tweets,social media rumour,diverse political and economic factors, change of leadership, investor sentiment and many other factors.
- Here we will analyse by considering all of these things and try to predict the exact price fluctuation.
- The objective is to find the strong correlation between the movement of stock prices and the publication of news articles.

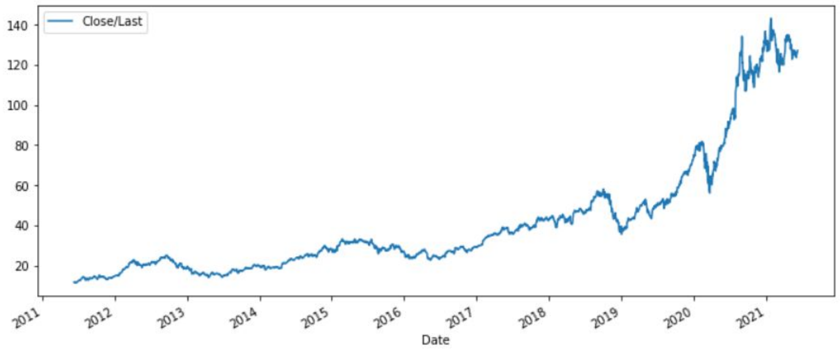
DATASET

9 years worth of stock price data were gathered for roughly 8 large-cap technology companies(e.g., FB, BIDU, NVDA, AAPL) and the NASDAQ-100 Technology Sector Index (NDXT) dating from 2011-06-10 to 2021-06-09.



The plot below gives a simple visualization of the processed data for an example stock (FB). Blue and red points are days immediately before and after an Earnings Report respectively (they tend to have larger percentage changes). Green points are the data we are interested in and serve as
actual inputs for the model.

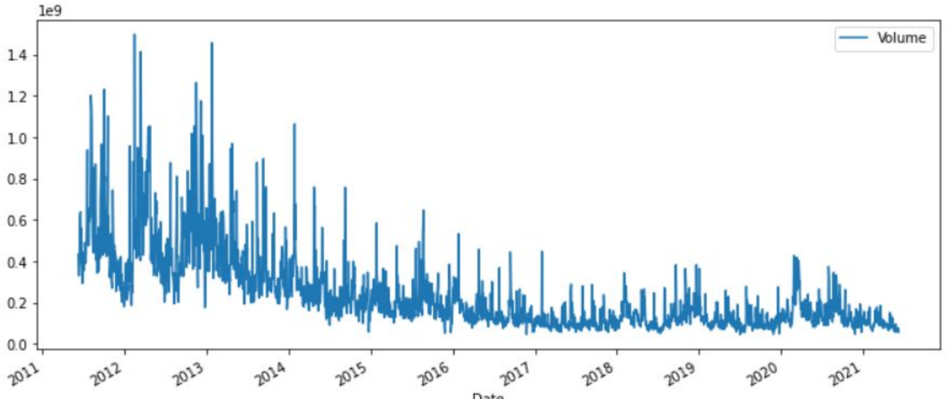
DATASET



- Here we can see that there is changes in stock price according to the time changes,here the close value is more in recent years so that we can say the companies might get the best value in last few years.

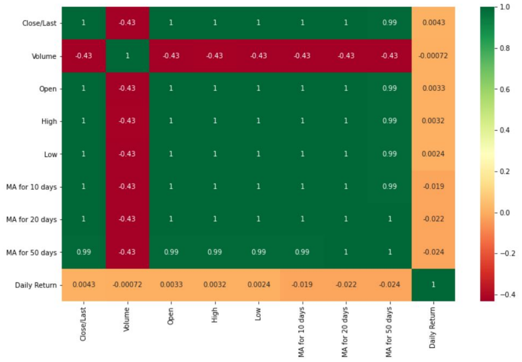
- But in this scenario we can check that companies stock sale is decreasing.In this time series that is clearly represented.

  Lastly,we saw that there is two scenario-stock value increasing and stock sale quantity decreasing ,later we shall consider this thing to make our decisions.



FEATURES

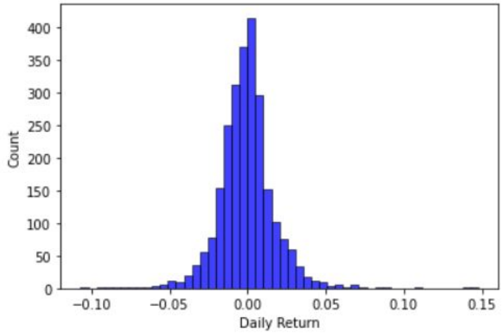
- The data were preprocessed to exclude 10 trading days both before and after a company’s quarterly Earnings Report from the data.
- We can see that the remaining days’ prices have lower volatility.
- If we make differentiate daily NDXT percentage changes from every stock’s daily percentage price changes to obtain an estimate of its relative change within the technology industry.



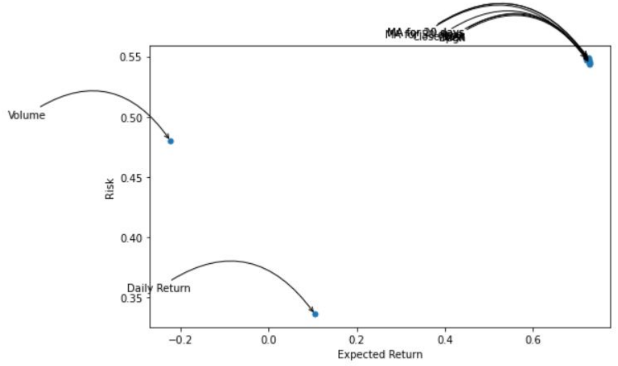
- There is correlation between several features,have to consider that thing also

If we perform this operation we can obtain the less affect the entire industry at once.

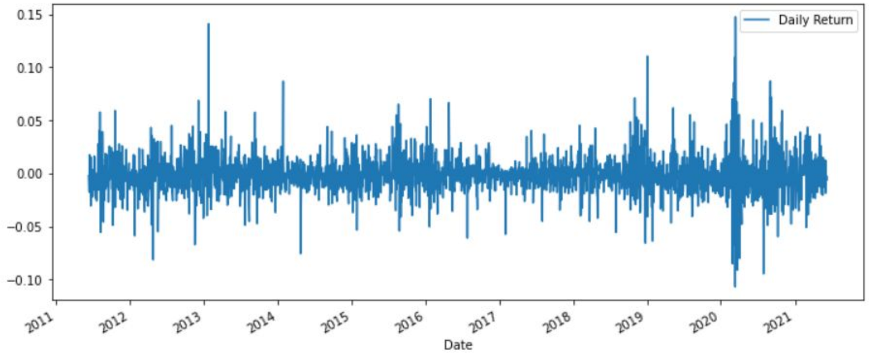
FEATURES



- In the right side we see that how risk factor will changes according and how it affects the return.



This is the percent changes in daily return which can be obtained by checking the changes in close value of stock.



- As the graph represents the the daily return of the company and clearly noticeable that there is positive return as well as negative return

MODEL IMPLEMENTATION

- Linear Regression: $$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

  ŷ is the predicted value.
  n is the number of features.
  xi is the ith feature value.
  θj is the jth model parameter (including the bias term θ0 and the feature weights θ1, θ2, , θn).

  We can apply standard linear regression with the standard least mean square errors.

- Logistic Regression: $$\hat{p} = h_\theta(\mathbf{x}) = \sigma(\theta^T \cdot \mathbf{x})$$

  p^= hθ(x) that an instance x belongs to the positive class.

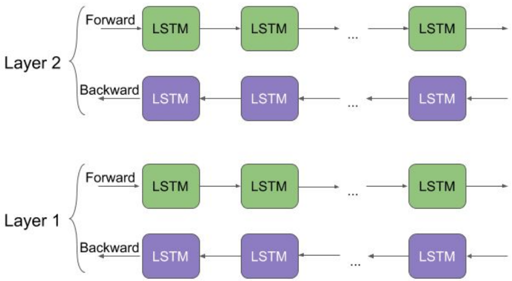
  We can implement this model to the dataset and compare with previous result.

MODEL IMPLEMENTATION

- Naive Bayes:

$$P(C_i \mid x_1, x_2 \dots, x_n) = \left(\prod_{j=1}^{j=n} P(x_j \mid C_i)\right) \cdot \frac{P(C_i)}{P(x_1, x_2 \dots, x_n)}$$

If we use this model with the multinomial event model and Laplace smoothing.



- Deep Learning:The neural network used has 1 input layer, 1 hidden layer, and 1 output layer. The hidden layer has 20 hidden units.
  As we can see data will overfit in most of the cases, by using L2 regularisation with certain lambda value it may lowers.
- There are several pre trained model to do this news analysis like Predicting Stock Prices Using Facebook's Prophet Model,Time-Series Forecasting: Predicting Microsoft (MSFT) Stock Prices Using ARIMA Model,Time-Series Forecasting: Predicting Apple Stock Price Using An LSTM Model

RESULTS(Source from paper)

Model	Training Error (probability of wrong prediction)	Test Error (probability of wrong prediction)	Test Daily % Gain
Naive Bayes (News only)	0.0248 (403 samples)	0.445 (200 samples)	N/A
Linear Regression (Prices only)	0.4601 (589 samples)	0.4365 (126 samples)	0.0311
Logistic Regression (Prices only)	0.4584 (589 samples)	0.4762 (126 samples)	0.0018
Neural Network (Prices + News)	0.0084 (403 samples)	0.4100 (100 samples)	0.0423

The trained Naive Bayes model achieved 97.52% accuracy on training data and 55.4% accuracy on test data.

We can observe that here the more amount of overfitted case occurs.

The linear regression model (and similar the logistic regression model) resulted in 56.35% prediction accuracy and a daily percentage gain of 0.0311%.

Neural network implementation in this case is most efficient way to predict the results.In this case the training error was extremely low at 0.84%, and a test error at 41.00%.

CONCLUSION & FUTURE WORK

- The accuracy of the final neural network correlates highly with the Naive Bayes prediction accuracy.
- We can improve the naive bayes predict model or we can work further with Neural Network Design.
- By implementing some either dropout and max pooling layers or using some efficient activation function,besides adding some layer,optimization function it can improve the result.
- We can add full news instead of headline,also we can build a separate model to classify the news and make some decision over it and after if we implement that,it may be a way to get better result.

THANK YOU