**STOCK PRICE PREDICTION**

**INTERIM REPORT**

**Problem Statement:**

* **Exploring companies In different sector (IT ,Banking and Health Industry)**
  + **Predict the closing Price of stocks of each company and finding out the Best Model.**
  + **which the**

|  |  |  |  |
| --- | --- | --- | --- |
| Given the Stock Market Dataset ,predict the **Closing Price of the Stock on s given day.**   Showing changes in outcomes (error) on inclusion of features that are widely accepted as market determinants.   |  |  | | --- | --- | | Exploring companies in different sector (**IT Industry , Banking Industry, Health Industry)**   |  | | --- | | Finding out the best technique in each of the above sectors | |  |  | | --- | | Exploring companies in different sector (**IT Industry , Banking Industry, Health Industry)**    Does the technique of predicting the closing price of one company’s stock work with another company’s stock ?   | |

**Features:**

Original Dataset had the following features :

|  |
| --- |
| Opening Price |
| Highest Price |
| Lowest Price |
| Last Price |
| Closing Price |
| Trade Quantity |
| Turnover(Lacs) |

Based on the Study and Research in Stock Market We have come out with the following features :

|  |
| --- |
| Previous Day Closing Price |
| Same Day Opening Price |
| Previous Day Nifty Closing Price |
| 10 Days Moving averages |
| 15 Days Moving averages |
| 20 Days Moving averages |
| 40 Days Moving averages |
| 10 Days Momentum |
| 40 Days Momentum |
| Average Difference between Opening Price and Closing Price (5 Days ) |
| Average Difference between Opening Price and Closing Price (10 Days ) |
| Average Difference between Opening Price and Closing Price (40 Days ) |
| Average Difference between Highest Price and Lowest Price (5 Days ) |
| Average Difference between Highest Price and Lowest Price (10 Days ) |
| Average Difference between Highest Price and Lowest Price (40 Days ) |
| Volatility |
| Average Turnover for 10 Days |

**PREPROCESSING**

|  |  |  |
| --- | --- | --- |
| **Technique** | **Applied** | **Remarks** |
| Normalization | YES | The Dataset was scaled between 0 and 1. |
| PCA | NO | The number of features were not large. |
| Missing Values Correcton | YES | Taking average of surrounding values |
| Equalise no of Datapoints in Datasets | YES | Equal no of Datapoints in each dataset were processed. |

**Techniques Applied**

* **Simple Linear Regression**
* **LASSO Regression**
* **Ridge Regression**

**We performed Regularized Linear Regression. We used grid-search based approach to find the best value of alpha and delta (one which gives minimum MSE) for fitting the model.**

dvalues=[0,0.1,0.05,0.01,0.005,0.001,0.0005,0.0001,0.00005,0.00001];

avalues=[0.1,0.05,0.01,0.005,0.001,0.0005,0.0001,0.00005,0.00001];

**Finding alpha and delta for Regularized Regression**

**LASSO REGULARIZED REGRESSION**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | alpha | 0.1 | 0.05 | 0.01 | 0.005 | 0.001 | 0.0005 | 0.0001 | 0.00005 | 0.00001 |
| Delta |  |  |  |  |  |  |  |  |  |  |
| 0 |  | 0.0371 | 0.0368 | 0.0371 | 0.0372 | 0.0373 | 0.0384 | 0.0560 | 0.0641 | 0.0735 |
| 0.1 |  | 0.0408 | 0.0402 | 0.0442 | 0.0478 | 0.0779 | 0.1249 | 0.3850 | 0.4878 | 0.6012 |
| 0.05 |  | 0.0388 | 0.0391 | 0.0404 | 0.0423 | 0.0556 | 0.0750 | 0.1815 | 0.2240 | 0.2710 |
| 0.01 |  | 0.0374 | 0.0372 | 0.0378 | 0.0382 | 0.0407 | 0.0448 | 0.0747 | 0.0877 | 0.1024 |
| 0.005 |  | **0.0373** | 0.0371 | 0.0374 | 0.0377 | 0.0390 | 0.0416 | 0.0649 | 0.0754 | 0.0871 |
| 0.001 |  | 0.0371 | 0.0369 | 0.0371 | 0.0373 | 0.0376 | 0.0390 | 0.0578 | 0.0663 | 0.0761 |
| 0.0005 |  | 0.0371 | 0.0369 | 0.0371 | 0.0372 | 0.0375 | 0.0387 | 0.0569 | 0.0652 | 0.0748 |
| 0.0001 |  | 0.0371 | 0.036 | 0.0371 | 0.0372 | 0.0373 | 0.0385 | 0.0562 | 0.0644 | 0.0738 |
| 0.00005 |  | 0.0371 | 0.0368 | 0.0371 | 0.0372 | 0.0373 | 0.0385 | 0.0561 | 0.0642 | 0.0736 |
| 0.00001 |  | 0.0371 | 0.0368 | 0.0371 | 0.0372 | 0.0373 | 0.0384 | 0.0560 | 0.0642 | 0.0735 |

**RIDGE REGULARIZED REGRESSION**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | alpha | 0.1 | 0.05 | 0.01 | 0.005 | 0.001 | 0.0005 | 0.0001 |
| Delta |  |  |  |  |  |  |  |  |
| 0 |  | 0.0371 | 0.0368 | 0.0371 | 0.0372 | 0.0373 | 0.0383 | 0.056 |
| 0.1 |  | 0.0371 | 0.0369 | 0.0372 | 0.0373 | 0.0382 | 0.0399 | 0.0575 |
| 0.05 |  | 0.0371 | 0.0369 | 0.0371 | 0.0377 | 0.0392 | 0.1073 | 0.0568 |
| 0.01 |  | 0.0371 | 0.0368 | 0.0371 | 0.0377 | 0.0386 | 0.0386 | 0.0562 |
| 0.005 |  | 0.0374 | 0.0368 | 0.0371 | 00372 | 0.0380 | 0.0465 | 0.0715 |
| 0.001 |  | 0.0374 | 0.0368 | 0.0371 | 0.0372 | 0.0373 | 0.0385 | 0.0560 |
| 0.0005 |  | 0.0371 | 0.0368 | 0.0371 | 0.0372 | 0.0373 | 0.0384 | 0.0560 |
| 0.0001 |  | 0.0371 | 0.0368 | 0.0371 | 0.0372 | 0.0373 | 0.0384 | 0.0560 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Stocks** | **CIPLA** | | **TCS** | | **HDFC** | |
|  | **LASSO** | **RIDGE** | **LASSO** | **RIDGE** | **LASSO** | **RIDGE** |
| **Alpha** | 0.05 | 0.05 | 0.1 | 0.1 | 0.00005 | 0.00001 |
| **Delta** | 0 | 0 | 0.1 | 0 | 0.01 | 0 |
| **MSE** | 0.03688 | 0.03688 | 0.00139 | 0.00144 | 0.00146 | 0.00188 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Stocks** |  | **Alpha** | **Delta** | **MSE** |
| **CIPLA** | **LASSO** | 0.05 | 0 | 0.033688 |
| **RIDGE** | 0.05 | 0 | 0.3688 |
| **TCS** | **LASSO** | 0.1 | 0.1 | 0.00139 |
| **RIDGE** | 0.10 | 0 | 0.00144 |
| **HDFC** | **LASSO** | 0.00005 | 0.01 | 0.00146 |
| **RIDGE** | 0.00001 | 0 | 0.00188 |

**With Nifty**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Stock** | **CIPLA** | | **TCS** | | **HDFC** | |
|  | **Alpha=0.05** | **Delta=0** | **Alpha=0.1** | **Delta=0.1** | **Alpha=0.00005** | **Delta=0** |
| **Gradient Descent** | **Stochastic** | **Batch** | **Stochastic** | **Batch** | **Stochastic** | **Batch** |
|  |  |  |  |  |  |  |
| **MSE** | **0.002474** | **0.002279** | **0.001067** | **0.001116** | **0.001425** | **0.001754** |
|  |  |  |  |  |  |  |

**Without Nifty**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Stock** | **CIPLA** | | **TCS** | | **HDFC** | |
|  | **Alpha=0.05** | **Delta=0** | **Alpha=0.1** | **Delta=0.1** | **Alpha=0.00005** | **Delta=0** |
| **Gradient Descent** | **Stochastic** | **Batch** | **Stochastic** | **Batch** | **Stochastic** | **Batch** |
|  |  |  |  |  |  |  |
| **MSE** | **0.0027** | **0.00248** | **0.001198** | **0.001284** | **0.001465** | **0.001774** |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**Challenges:**

**Finding out the appropriate features was a challenge.Lot of papers were consulted to come up with relevant features in stock market prediction.Some of the relevant works explored were:**

Feature Investigation by Hui Lin[1].

Feature Selection for Stock Market Analysis by Yuqinq He, Kamaladdin Fataliyev, and Lipo Wang[2].

Reasoning :

LASSO works better in situations where we need to find out the values of the parameters and perform feature selection. LASSO will select only one feature from a group of highly correlated features.

CIPLA : The closing price of cipla stock is linearly related

**Results :**

|  |  |  |  |
| --- | --- | --- | --- |
| **STOCK** | **BEST MODEL** | **GRADIENT DESCENT** | **MSE** |
| **CIPLA** | Simple Linear Regression | Batch | 0.001425 |
| **TCS** | LASSO Linear Regression | Stochastic | 0.001067 |
| **HDFC BANK** | LASSO Linear Regression | Stochastic | 0.002279 |

**Feature Modelling**

**To explore and prove the strong relation between the closing price of a stock (present in NIFTY) and the Nifty Index, we evaluated the MSE values after removing Nifty Index Value as a feature. The MSE values were found to be consistently higher showing that Nifty Index is a crucial feature in determining the closing price of the stock and choice of features is an important determinant in good performance from the machine learning models.**