

# STOCK PRICE PREDICTION

RISHABH GUPTA(2014086)  
KUSHAGRA MAHAJAN(2014055)

## INTRODUCTION

In this project, machine learning algorithms were used to predict the closing price of stocks of companies across various sectors (IT, Pharmaceutical ,Banking Industry).

Stock price prediction has always been one of the most challenging problems affecting the lives of millions of people around the world. Recently the problem has gained great popularity due to its unpredictable behaviour. This work is our attempt to explore the performance of some of the common techniques on stock datasets.

## DATASETS

We explore three datasets taken from Quandl.com :- HDFC Bank, TCS, CIPLA Pharmaceuticals Pvt. Ltd.

The datasets provided stock values for the past 8 years(3000 samples).

- 80%(2400 samples) of the data was chosen for training.
- 10% of the training data(240 samples) were used for validation.
- 20% of the total data was used for testing.

## FEATURE EXTRACTION

The original data had the following features: Opening, Highest, Lowest, Last, Closing prices, Total trade Quantity and Turnover.

The features used for Learning Model are given in Table-3 .

## EVALUATION METRICS

Mean Squared Error was used as an Evaluation Metrics for our Models.

## PRE-PROCESSING

<u>Technique</u>	<u>Applied</u>	<u>Remarks</u>
Normalization	YES	The Dataset was scaled between 0 and 1.
PCA	NO	The number of features were not large.
Missing Values Correction	YES	Taking average of surrounding values

## TECHNIQUES APPLIED

We have explored

- Simple Linear Regression
- Regularized Linear Regression(LASSO and RIDGE).

We have used Cross Validation Technique with Grid Search(over alpha and delta values) to hyper-tune the parameters(alpha and delta) and come up with the parameters which gives minimum MSE.

We have compared both Stochastic and Batch Gradient Descent to train our model .

## 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].

## PROGRESS AND OBSERVATIONS

GRID SEARCH was applied with CROSS VALIDATION to find out the best value of parameters

**Table-1** : Best Value of Alpha and Delta by applying cross validation and grid search

Stock		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

Using them Best Value from table-1 across every stock, Stochastic and Batch Gradient Descent algorithm was applied.

**Table-2** : MSE corresponding to best parameters with Stochastic and Batch Gradient Descent

Stock	Parameters	Gradient Descent	MSE
CIPLA	Alpha=0.05	Stochastic	0.002474
	Delta=0	Batch	<u>0.002279</u>
TCS	Alpha=0.1	Stochastic	<u>0.001067</u>
	Delta=0.1	Batch	0.001116
HDFC	Alpha=0.00005	Stochastic	<u>0.001425</u>
	Delta=0.01	Batch	0.001754

To explore and prove the importance of good feature selection we eliminated an important feature(NIFTY closing price) and recalculated the MSE.

**Table-2** : MSE comparison when NIFTY feature was removed.

Stock	Gradient Descent	MSE (With Nifty)	MSE (Without Nifty)
CIPLA	Stochastic	0.002474	0.0027
	Batch	0.002279	0.00248
TCS	Stochastic	0.001067	0.001198
	Batch	0.001116	0.001284
HDFC	Stochastic	0.001425	0.001465
	Batch	0.001754	0.001774

## RESULTS

The Best Model corresponding to each stock is as follows:

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

- We observe that the same model does not work across different stocks. Hence different stocks move differently with features and historical data.
- The stock Price of TCS is most predictable among the 3 stocks as it gives the least MSE.
- 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.
- The MSE values were found to be consistently higher when NIFTY feature was removed .This shows that the choice of features is an important determinant in good performance from the machine learning models.

## REFERENCES

- <https://www.quandl.com/data/NSE?keyword=>
- [http://web.itu.edu.tr/~cataltepe/pdf/2011\\_ICMFECataltepe.pdf](http://web.itu.edu.tr/~cataltepe/pdf/2011_ICMFECataltepe.pdf)
- [http://www.ntu.edu.sg/home/elpwang/PDF\\_web/13\\_ICONIP.pdf](http://www.ntu.edu.sg/home/elpwang/PDF_web/13_ICONIP.pdf)

**Table-3** : Features used in our Learning Model.

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