

# Predicting Equity Prices of Cross-Listed Stocks

Sanketh Kokkodu Balakrishna

Teammate: Lopamudra Pal

## 1 Problem

Stock Prices are considered to be very dynamic and susceptible to quick changes because of the underlying nature of the financial domain and in part because of the mix of known parameters (Previous Days Closing Price, P/E Ratio etc.) and unknown factors (Election Results, Rumors etc.). What makes it highly challenging is the number of uncertainties which model individual stocks and affect their price. In this project, we will apply machine learning techniques to build regression models which can predict stock prices of cross-listed stocks on the National Stock Exchange of India(NSE) based on its performance on the New York Stock Exchange(NYSE) and its historical performance on the NSE. Stocks listed on exchanges in emerging markets like NSE are known to follow prices of their cross-listed counterparts on exchanges in developed markets like NYSE, and accurately predicting prices on the NSE will give leverage and an advantage to investors in emerging markets.

## 2 Methodology

There are three main components in our approach:

1) Data Collection and pre-processing 2) Model selection and Training 3) Hyperparameter tuning, Prediction, and testing.

- Models/Algorithms -

Some of the models we plan to experiment with are Linear Regression, Decision Tree Regression, Support Vector Regression, Neural Networks and Ensemble methods.

- Code Libraries - pandas, scikit-learn, numpy, matplotlib.
- Hardware Platforms - The hardware platforms we are targeting is general purpose computers, although we may incorporate some part of the code in a distributed environment.

## 3 Data Sets

We do not have a fixed dataset, and need to create one by scrapping financial data from Yahoo Finance. All the historical data for the stocks which are trading at NSE and NYSE can be freely obtained at <https://finance.yahoo.com/>.

The dataset has 5 features for each stock, but we may create additional features as necessary. Work has to be done in cleaning and creating a data frame for relevant data.

## 4 Experiments

We'll experiment with different regression techniques on the data. The dataset will be divided into training data, validation data and testing data for each stock, and the trend of the objective function on the testing data will be recorded. The best objective function will be used to create a prediction model for respective stocks. Cross-validation techniques, feature selection, hyperparameter selection, basis and kernel expansion will be used to improve Mean Square Error(MSE). We'll also experiment with Bagging, Boosting and Stacking to achieve accurate predictions. In addition, we may explore the co-relation between index-specific stocks and analyze the effect of factors like news on short-term prices[7].

## 5 Related work and Novelty

A lot of work has been carried out in the field of stock market prediction analysis using machine learning algorithms based on historical data. [1] describes different ML techniques applied to stock market datasets to predict stock prices. It also describes useful parameters and indicators to recognize patterns in stock prices for successful prediction. [2]] analyses the results of applying hybrid ML models based on Genetic Algorithms(GA) and Support Vector Machines(SVM) for stock market prediction. [3] proposes a neuro-genetic stock prediction system based on financial correlation between companies.[4] discusses recent developments in stock market prediction models and also investigates the effects of global events on stock prices. [5] gives us an overview of planning and visualisation of data and results of our experiments. [6] cites the advantages of modeling the stock market problem as a Neural Network.

**Novelty in our approach:** Very less work has been done on applied machine learning techniques to NYSE data to predict prices of the same stock in emerging economies. Stocks specific to NSE have rarely been studied with respect to its performance in a different exchange. These aspects make our project different from previous work.

## 6 Collaboration Plan

### 6.1 Timeline

**Week 1:** Acquire and assemble data. Create data frame and clean data.

**Week 2 & 3:** Build basic algorithms/models and check for regression outputs.

**Week 4:** Apply feature selection, create new features and analyze technical indicators and outputs.

**Week 5:** Implement optimization techniques and experiment with how providing additional features such as news, affects prices. Work on visualizing the results, preparing Report and documentation.

### 6.2 Individual responsibilities

#### Sanketh

- Studying related research papers
- Cleaning the data and creating a dataset using dataframe
- Working in collaboration to pick key stocks to analyze
- Building Neural Network models to predict prices
- Implementing ensembles to improve performance
- Working on visualizing a part of the results
- Working collaboratively in identifying innovative ideas to improve performance of models
- Preparing the project report collaboratively

#### Lopamudra

- Studying previous research papers and summarizing the work done in related field
- Figuring out key problems our team should be working on
- Working in collaboration with Sanketh in picking stocks which may be of interest
- Working to identify the explicit/hidden features and analysing the correlation between the features and the target
- Working on regression methods like Support Vector Regression, Neural Networks
- Working on ensemble methods like Bagging and Random Forests
- Working on feature selection methods like Recursive Feature Elimination and Statistical Dependence Filtering
- Experimenting with hyperparameter selection methods like k-fold Randomised Search cross-validation techniques
- Working on visualizing a part of the results
- Working collaboratively in identifying innovative ideas to improve performance of models
- Preparing the project report collaboratively

## References

- [1] Vatsal H. Shah. *Machine Learning Techniques for Stock Prediction*,  
<http://vatsals.com/Essays/MachineLearningTechniquesforStockPrediction.pdf>
- [2] Rohit Choudhry, and Kumkum Garg. *A Hybrid Machine Learning System for Stock Market Forecasting*,  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.307.6153rep=rep1type=pdf>
- [3] Yung-Keun Kwon, Sung-Soon Choi, Byung-Ro Moon. *Stock Prediction Based on Financial Correlation*,  
<http://dl.acm.org/citation.cfm?id=1068351f>
- [4] Paul D. Yoo, Maria H. Kim, Tony Jan. *Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation*,  
<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1631572>
- [5] Suchismita Naik. *A Narrative Data Visualization Of The Indian Stock Market*,  
<http://dl.acm.org/citation.cfm?id=3014382CFID=740526300CFTOKEN=15991829>
- [6] Birgul Egeli, Meltem Ozturan, Bertan Badur, *Stock Market Prediction Using Artificial Neural Networks*,  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.1114rep=rep1type=pdf>
- [7] Robert P. Schumaker, Hsinchun Chen. *Textual analysis of stock market prediction using breaking financial news: The AZFin text system*,  
<http://dl.acm.org/citation.cfm?id=1462204CFID=740526300CFTOKEN=15991829>