math 571 Data preparation and analysis

Stock trend predictioN USING TIME SERIES FORECASTING AND MACHINE LEARNING ALGORITHMS

March 1, 2018

# Overview

## Project Background and Description

Stock market prediction has been an area of interest for every investor from early days. Prediction, as we know, cannot be very accurate as there are multiple factors that affect the stocks positively or negatively. This project aims to predict the trend [positive, negative or neutral] of stock price of a company by using a combination of Time Series Forecasting and Machine learning algorithms. The idea is to compare and analyse the performance various statistical models in predicting the stock trend.

Team members: Ayshwarya, Geethanjali, Sai Ishwarya, Tina.

## Data Sources

Google stock trends : [https://drive.google.com/drive/u/0/my-drive](https://drive.google.com/file/d/0B3JerM8xca5wUmFubWhTXzM5dVNUOEhRSGtnVHBySzFnTTF3/view?usp=sharing)

## Steps to complete analysis

*Data Collection: -*

Source of data will be downloaded from Google finance which will contain Apple’s opening and closing stock prices.

*Data Pre-processing: -*

Stock price data is not looked for weekends and holidays as the price does not change on those days. Further pre-processing includes converting non stationary time series to stationary time series by taking first or second differences of the time series values. Then data transformations like log, square root transformation to be done on the data. The data will be checked for autocorrelation if any and handled appropriately. Finally, the data is split into train and test data.

*Stock price prediction and analysis: -*

These features are then given as input to the random forest algorithm, KNN algorithm and time series forecasting to train and test the data. We will train respective models. After these models are trained, they are tested with test data to check accuracy of the model.

Stock price is predicted(regression), an increase or decrease(ordinal classification) can be predicted and evaluate if there is any large downward swing (binary classification).

*Evaluation metrics: -*

Here we will use F1-score, RMSE, confusion matrix and ROC curve to evaluate the performance of our model.

*Deployment: -*

Planned to deploy this project using a web service.

## Success metrics, Statistical models and KPIs

Statistical models: Random forest algorithm, KNN algorithm and Time series forecasting to train and test the data.

Model Evaluation: F1-Score, RMSE, Confusion matrix and ROC curve

## Deliverables

1. Source code
2. Relevant graphs and metrics
3. Deployed with web service

## Project Planning

Asana is being used for the project planning.