

BLUEPRINT FOR STOCK PRICE MOVEMENT USING CROWD DATA

~ Karamjit (ikrmjit@gmail.com)

1. Rough approach to overall problem

First step is to understand the problem we need to solve and prepare a rough model of our project and make a step by step approach to reach our goal.

2. Data Collection

Identify the sources to collect the relevant data. For the project following two sources were used :

- Data collection via Yahoo finance API
- Data collection of StockTwits message

3. Exploratory data analysis (EDA)

To get a rough idea of our datasets EDA is performed.

4. Data Preprocessing

Usually raw data comes with a lot of inconsistencies and it needs to be treated before further analysing and model applications etc.

Following are some steps to be applied :

- Data cleaning
 - Handling Missing Values
 - Scaling and Normalization
 - Parsing Dates
 - Character Encodings
 - Inconsistent Data Entry
- Divide data in training and testing data

5. Model fitting

- Different models are used to fit our data (Pure Bag of Words Model, Mixed Features Model, k-NN Regression etc.) and compared and best model is used to fit our training data to train our model
- Then this model is tested with testing data to know its ability on new unseen data and its accuracy is checked to know the amount of errors.

6. Final report of the project

Final reports using this model are made and various data visualisation tools used for better understanding of results.