

# Portfolio management and Stock prediction and recommendation using data mining

VirtiSavla, KhushbooMadalani, NithyaRaghuram  
Department of Computers,  
K.J. Somaiya College Of Engineering, Mumbai, India.

**Abstract**—There are several factors which affect the price of a stock. Some of them are daily news articles, volume of that stock traded, sentiment in the market, profit of the company etc. Due to the advancement in technology a large amount of data about the stocks is generated every day in the form of news articles, analyst reviews, twitter data etc. The increasing amount of data is making it increasingly difficult to manually analyze the data to make strategic decisions. In this paper we present a way in which news articles and analyst reviews from multiple sites are mined and analyzed to predict whether the price of a stock will increase or decrease in the near future with an accuracy of 70%. Also, volume of a stock traded contributes in determining the future of that stock. We also present a method to recommend stocks based on their volume traded and news about them.

**Index Terms**—classify, predict, recommend, volume

## I. INTRODUCTION

The news articles which affect the price of a stock are of different types. Some are outcomes of board meetings, information about major acquisitions made by the company, financial analyst recommendations etc. To draw a potent inference from these news articles there is a critical need for a system that can congregate news from various financial sites and classify them. In this paper we present a way in which real time news articles are fetched dynamically from the web and classified as positive or negative to estimate the future of a stock.

Also, traders invest in many different stocks belonging to disparate sectors. This makes it difficult to keep a track of the holdings. We present a clean portfolio management system which enables the user to manage their stocks efficiently. We also present the concept of threshold value in which user will receive timely alert when the price of the stock breaches the threshold value entered by the user. This concept greatly eases the arduous job of keeping a watch on the stock market for the user. As soon as the user receives the alert he can decide his call on the stock i.e hold/buy/sell.

Volume traded plays an important role in determining the movement of a stock price. We present a method of recommending stocks belonging to a particular sector by

considering the volume of each stock traded and predicting their future by using data mining.

## II. PROPOSED APPROACH

### A. Building a Classifier for Stock news using data mining

#### 1. Gathering Stock news articles

The first step is to train the system. In-order to train the system a training dataset consisting of 1000 positive and 1000 negative news articles for around 30 stocks is created manually. Articles were taken from website [www.moneycontrol.com](http://www.moneycontrol.com), which covered a time period of 18 months. The stocks chosen cover sectors like IT, Banking, Pharmaceuticals, Airline sector, Oil and gas sector, Steel sector and Automobile sector. The collected data is stored in a repository for future analysis.

For prediction of increase/decrease of stock price, we require a test dataset. Test dataset is created by dynamically fetching the news article from the web. These articles are fetched by crawling the websites like [www.moneycontrol.com](http://www.moneycontrol.com), [www.economicstimes.com](http://www.economicstimes.com). The collected articles are stored in .arff format in a repository for future analysis.

#### 2. Processing of the Stock news articles.

Our goal is to classify each news article as positive or negative. The first step in processing the news is to remove the stop words like a, the, in etc. These words do not contribute in classifying a news article as positive or negative.

The next step is to classify these articles as positive or negative using the open source software WEKA[7] by accessing its library. Random forest algorithm[8] has been used for text classification. This tool matches the test data with the training data set to provide classification. Larger the training data set more accurate would be the classification. If the classification of the test data is positive then the price of the stock will increase in future however if it is negative then the price will decrease.

The validation of this approach can be done by comparing the price of the stock at the time of prediction with that day's closing price.

**B. User's Portfolio management**

In this approach the name of the stock, quantity , purchase date, purchase price and investment amount is stored in the database. Here a threshold value is accepted from the user for each stock in the portfolio and stored in the database. The current price of each stock in the user's portfolio is dynamically fetched from the web and matched with the threshold value entered by the user. When the price fetched from the web becomes equal to the threshold value an alert is sent to the user.

**C. Classification of current Stock news articles.**

The test dataset created dynamically by fetching the real time news articles from the web is now used for predicting the future of the stock (increase/decrease). For classification of training and test dataset a popular data mining tool WEKA[7] is used. WEKA[7] provides various classification algorithms like J48 Decision tree, random forest, naïve Bayes etc. First the training dataset is pre-processed. Then the dataset is converted from string to word vector. Same operations are performed on the test dataset. Our dataset is trained to classify each article as either positive or negative. The algorithm used for classification is random forest. Random forest is an ensemble classifier that uses many decision trees for classification. Random forest algorithm classifies each article in our test data as positive or negative. Thus, total positive and total negative articles are calculated. If number of positively classified articles are more than total number of negatively classified articles the result of prediction is that the stock price will INCREASE else the stock price will DECREASE.

**D. Recommendation of Stocks using its volume traded**

**1. Acquiring data about volume of a stock traded from the web**

The first step is to fetch the volume of each stock stored in the database belonging to the sector entered by the user. Volume of a stock is a number which shows how many shares of a stock was bought or sold on that day. Volume of a stock can be obtained from sites like [www.moneycontrol.com](http://www.moneycontrol.com) or [www.finance.yahoo.com](http://www.finance.yahoo.com) . The stocks now need to be arranged in descending order of the volume. Stocks with greater volumes are better.

**2. Predicting the future of each stock**

The next step is to predict the future of each stock. The stocks with prediction as decrease are discarded from the list. Among the ones present in the list the stock with the greatest volume

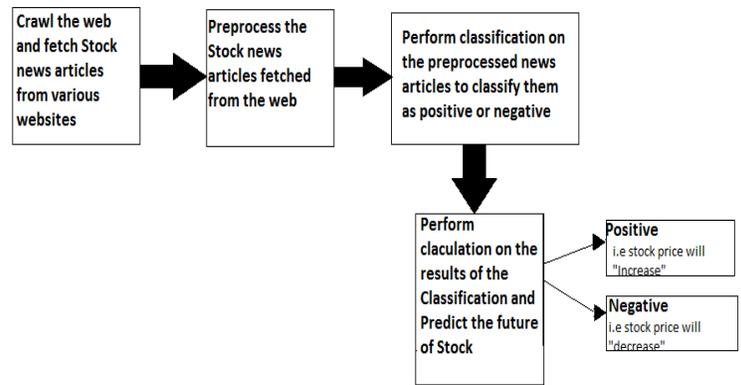
is recommended. The pseudo code for recommendation module is given in Fig.a.

**Pseudo Code:**

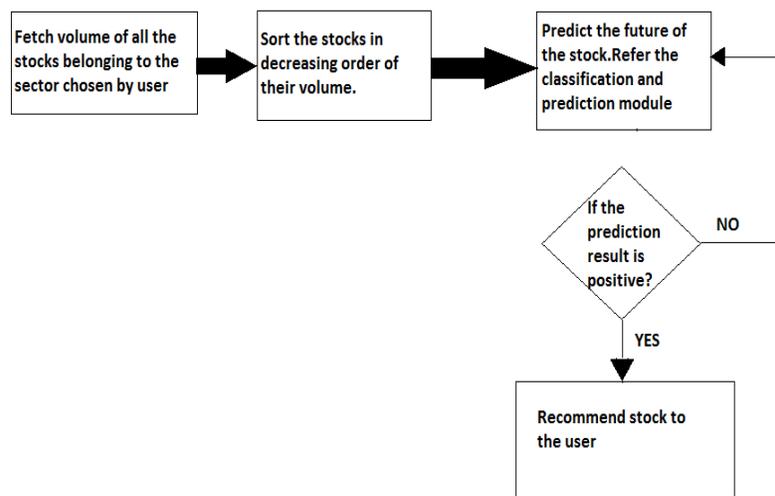
1. Initially fetch stocks belonging to sector chosen by the user.
2. for each stock
  - fetch the volume of the stock.
3. Sort(stocks according to their volume in decreasing order )
4. Predict the future of stock
5. If "increase":
  - Recommend stock to the user;
- Else:
  - Repeat steps 4 and 5.

**Fig. a Pseudo Code**

The entire prediction and recommendation methodology is shown in Fig.1 and Fig.2.



**Fig.1 Prediction Module Block Diagram**



**Fig.2 Recommendation Module Block Diagram**

### III. RESULTS

We have conducted analysis of various stocks to test our system. The accuracy of prediction is 70%. The experiments conducted are shown in Fig.3.

	A	B	C	D	E	F	G
1	Date	Stock name	Prediction	Actual	Price during pred	price after pred	result
2	26-02-2016	Infosys	Increase	Increase	1125	1130.4	Correct
3	26-02-2016	Sunpharma	Decrease	Decrease	882	871	Correct
4	26-02-2016	Axis bank	Decrease	Increase	384	386	Wrong
5	29-02-2016	Axis bank	Decrease	Decrease	389	375	Correct
6	01-03-2016	Ashok Leyland	Increase	Decrease	85	87	Correct
7	01-03-2016	Axis bank	Increase	Increase	379	391	Correct
8	01-03-2016	TCS	Increase	Increase	2182	2275	correct
9	04-03-2016	TCS	Increase	Decrease	2373	2359	Wrong
10	04-03-2016	Lupin	Decrease	Increase	1790	1795	Wrong
11	04-03-2016	Axis bank	Decrease	Decrease	417	415	Correct
12	09-03-2016	Axis bank	Increase	Increase	409	415	Correct
13	10-03-2016	Infosys	Decrease	Decrease	1164	1145	Coorrect
14	10-03-2016	Wipro	Increase	Decrease	538	536	Wrong
15	10-03-2016	Ashok Leyland	Decrease	Increase	96.45	97.8	Wrong

Fig.3. Results of the stockprediction

### IV. FUTRURE WORK

Our system gives accurate prediction for stocks which generate latest news. However, it does not provide accurate prediction for stocks which do not generate news daily. Low volume stock do not generate new. As a future work we propose the use of input from social media like twitter to predict the future of such stocks. The recommendation module considers only the volume traded and text classification for recommendation. We also propose to consider the percentage returns in that stock in the past five years for recommendation. Current system predicts the future of the stock as increase/decrease, as future work we propose to predict the price of the stock. Thus a hybrid approach which can include traditional data mining methods like neural networks, stock news, Twitter data and expert input can be combined to improve the stock future prediction.

### V. CONCLUSION

In this paper, we introduced a novel method to predict future of the stock using news article. We found interesting relationships between current news article and current market trend . We also introduced a novel methodology in which we consider volume of the stock along with the future prediction to recommend stocks to the user .

### VI. REFERENCES

[1] VivekSehgal and Charles Song, ” SOPS: Stock Prediction using Web Sentiment” , Department of Computer Science University of Maryland College Park, Maryland, USA,2007  
 [2] AnuragNagar1 , Michael Hahsler2, ” Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams”, 1Computer Science and Engineering, Southern Methodist University, Dallas, TX, USA 2 Engineering Management, Information and Systems, Southern Methodist University, Dallas, TX, USA

[3] Gabriel Pui Cheong Fung, Jeffrey Xu Yu, and Wai Lam, ” News Sensitive Stock Trend PredictionDepartment of Systems Engineering & Engineering Management The Chinese University of Hong Kong, 2002.  
 [4] Robert P. Schumaker and Hsinchun Chen, “Textual Analysis of Stock Market Prediction Using Financial News Articles”, Artificial Intelligence Lab, Department of Management Information Systems The University of Arizona, Tucson, Arizona 85721, USA,2009.  
 [5] AzadehNikfarjam, EhsanEmadzadeh, SaravananMuthaiyah,” Text mining approaches for stock market prediction” , Faculty of IT, MMU Cyberjaya, Malaysia, 2010.  
 [6] Leo Breiman,” RANDOM FORESTS”, Statistics Department University of California Berkeley, CA 94720, January 2001.  
 [7] An Introduction to the WEKA Data Mining System  
 [8] Eric Debreuve / Team Morpheme,” An introduction to random forests”, University Nice Sophia Antipolis / CNRS / Inria,2013