



## Daily news impact on stock using RSS data feed by neural network on big data

Anjali Babubhai Patel

Student, Kalol Institute of Technology and Research Centre, Kalol, Gujarat

### ABSTRACT

*As Prediction in Stock market is a difficult task. As Price Change in Stock Market depends on so many points. We are gating to research over Impact of Daily News on Stock Market Companies. We will read daily news related to Financial Sector (Business/Government etc) using RSS Feed of online news portals. Then we are going to mine this news for its effect by category of news, who will be affected by this news, impacts negative or positive etc. We will generate a model that can be used as a Training Data for Prediction Phase.*

**Keywords:** Stock Prediction, Financial Result, Manipulation, SVM.

### 1. INTRODUCTION

To Find Daily News Impact on Stock using RSS Data Feed By Neural Network and Big Data mean which type of effect on the stock to the daily news and the news is related with the daily news and the stock market prediction.

We will read daily news related to Financial Sector (Business/Government etc) using RSS Feed of online news portal and According to my system As Prediction in Stock market is a difficult task. For today Stock Market as Price Change in Stock Market depends on so many points. Then we are going to mine this news for its effect by category of news, who will be affected by this news, We will generate a model that can be used as a Training Data for Prediction Phase.

So doing this system I have so many difficult tasks on the system of the daily news impact on the stock. But I use one flow for this system my old system have some missing criteria in this system. So I am working on my propose system just like I have to Read Documents from the rich site summary (RSS) for my Pre Process the documents and then I Remove stop word from the news and punctuation, numbers etc.,

After completing this process then I decide my another work is Read Comments from the news of the system according to the Pre Process the Comments after completing this task I

have to do with the Find the Category of News in this system I divide all the news in the different category also like government private etc.

Then Find the Companies/Sectors affected which type also and then Assign the Label of the news which is categorized (Positive/Negative/Neutral) and after this condition Calculate Delta Price of the Stock on that Day. Which are depends on daily news and (Label the News Document using Delta Price and Use NLTP (Natural Language Tool Kit) (for Terms Detections and Terms Weighting) after the system I used in Tagger tool for Tagging then Calculate Positive and Negative. calculation of Each Word using Formula (2) and (3) Used Delta TF-IDF for terms Weighting Terms Reduction Process (Remove terms with a low score) Calculate Public Emotion Score using Comment Data, we Apply SVM (support vector machine) Here data set of my work is Data Set of Base Paper vietstock.vn, hsx.vn, hsn.vn between May 1st, 2014 and April 30th, 2015 by using a web crawler tool. And For Companies like BVH, CII, CSM, DPM, , KDC, MBB, MSN, OGC, PPC, PVD, PVT, REE, SSI, STB, VCB, VIC, VNM Support Vector Mechanism is used to predict Higher accurate prediction. Use SVM (support vector machine) algorithm by this method by this support vector machine algorithm I have to improve my system And the SVM is generally used to classify two classes using a hyper plane. And also support vector machine is technical called kernel trick and this system is used to transform the data based on this transformation also SVM is generally used for the two-class classification.

Support Vector Mechanism (SVM) is used to get more and more accurate data. We have to gain our system with the information according to system consideration also my base system. And the accuracy is higher. Also, support vector machine is technical called kernel trick and this system is used to transform the data based on this transformation also. Today the SVM is a Most used technique in the prediction with the higher accuracy.

In this paper, we compare many other techniques with SVM for the purpose of the prediction accuracy. The other techniques like Neural Network (NN), Artificial NN (ANN), Back Propagation Neural Network (BPNN). In our research, the linear data is used to get more accurate data. For the linear

data, the Radial Bias Function (RBF) Kernel is used to get more and more accuracy

In our research, the linear data is used to get more accurate data. For the linear data, the Radial Bias Function (RBF) Kernel is used to get more and more accuracy.

The SVM uses the kernel function for the classification of the data. SVM is used to make an algorithm for any method we have to use this algorithm for our system to the stock market prediction of daily news they are related with the approximately with the system consideration also and The right choice of the kernel gives the higher accuracy of the prediction. There are two types of the data are available linear and Non-Linear data.

The one more technique is used which is an M-Score equation. The M-Score equation. Shows the manipulation of any company. Which is directly affected by the Stock rate.

The M-Score equation considers various factors of the company to calculate the score. The score decides the company is manipulated or not. If the company is likely manipulated then the effect on the stock is negative, else the company is unlikely manipulated the n the effect on the stock is positive.

2. THEORY

SVM

Support Vector Machine (SVM) is used for the classification of the data. It helps to find a proper higher plane to minimize the error rate. It is one kind of learning machine technique for the statistical data. Support Vector Machine supports the high dimensional data for classification.

Support Vector Machine considers the mathematical representation for the classification of the data.

Here a training set of instance-label pairs  $(x_i, y_i)$ ,  $i = 1$  where  $x_i \in R^n$  and  $y \in \{1, -1\}$

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} w^T w + c \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & y_i (w^T x_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, \dots, l. \end{aligned}$$

The training vector  $x_i$  is mapped by a higher dimensional space by the function  $\phi$  as is  $\phi(x_i)$ .  $C > 0$  is the penalty parameter of the error term. So the problem is solved as follows:

$$\begin{aligned} \min_{\alpha} \quad & F(\alpha) = \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\ \text{subject to} \quad & y^T \alpha = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l \end{aligned}$$

Where the  $e$  is the vector of all,  $C > 0$  is upper bound,  $Q$  is a one by one +ve semidefinite matrix,  $Q_{ij} = y_i y_j K(x_i, x_j)$ , and  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$  is a kernel function. Thus a decision function is:

$$\begin{aligned} \sin(w^T \phi(x) + b) &= \sin\left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b\right) \\ w &= \sum_{i=1}^l \alpha_i y_i \phi(x_i) \end{aligned}$$

Here, the Gaussian kernel is,

$$K(\tilde{x}, \bar{x}) = \exp\left(-\frac{\|\tilde{x} - \bar{x}\|^2}{2\sigma^2}\right)$$

And the Radial Bias Function Kernel is:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0$$

3. NEURAL NETWORK

Application of *Neural Networks* to the task of stock index prediction. We have to describe the theory behind an ANNs and the Neural Network model which types of and its salient features according to the system. The accuracy rate of the E-audit is better than the manual audit. This consumes less time than the manual processor. It is very fast compared to the manual audit.

Neural network absorbs voice image and many other forms of data and turning it into valuable into variable information.

The number of previous data sets used for testing the data was varied for the double hidden *layered of the* neural network. The manipulation considers two types such as likely manipulate and unlikely manipulated. Likely manipulated company gives a negative effect on the stock market and the unlikely manipulated company gives the positive effect on the stock market.

$$\begin{aligned} \phi_{AE} \leq \phi_n \leq 0.00057^\circ \\ \phi_n \leq 0.0057^\circ \\ ARW \times \sqrt{\tau} \leq 0.00057^\circ \\ ARW \times \leq \frac{\sqrt{2} \times 0.00057^\circ}{\sqrt{\tau}} \\ RND \leq \frac{\sqrt{2} \times 0.00057^\circ}{60 \times \sqrt{\tau}} \end{aligned}$$

When  $\tau$  can reach 0.035 seconds

$$\begin{aligned} ARW \times \leq \frac{0.00057^\circ}{\sqrt{\frac{0.035 \text{ seconds}}{3600 \text{ seconds/hour}}}} \\ ARW \times \leq 0.18 \frac{^\circ}{\sqrt{\text{hour}}} \\ RND \leq 0.18 \frac{^\circ}{\sqrt{\text{hour}}} \times \frac{\sqrt{2}}{60} \\ RND \leq 0.0043 \frac{^\circ/\text{sec}}{\sqrt{\text{Hz}}} \end{aligned} \tag{4}$$

The reason I asked is simply that I want to know how the ANN correlates the independent variables in the input layer and hidden layers. and the real application of the system is the task of the stocks and behind this system consideration the actual flows of the system

- DEPI (depreciation index)  
DEPI= Depreciation rate<sub>t-1</sub>/Depreciation rate<sub>t</sub> ,

Where, Depreciation rate = Depreciation / (Depreciation + PPE).

Which shows the Declining depreciation rates could indicate understated depreciation as a means of manipulating earnings.

- SGAI (sales, general, and administrative expenses index)
- $SGAI = (SGA_t / Sales_t) / (SGA_{t-1} / Sales_{t-1})$

Which shows an increase in fixed SGA expenses suggests decreasing administrative and marketing efficiency, which could predispose a company to manipulate earnings.

- $Accruals = (Income\ before\ extraordinary\ items - Cash\ from\ operations) / Total\ assets.$   
Accruals shows the Higher accruals can indicate earnings manipulation.

- LEVI (leverage index)  
 $LEVI = Leverage_t / Leverage_{t-1},$

Where the Leverage is calculated as debt to assets.

Which shows Increasing leverage could predispose a company to manipulate earnings.

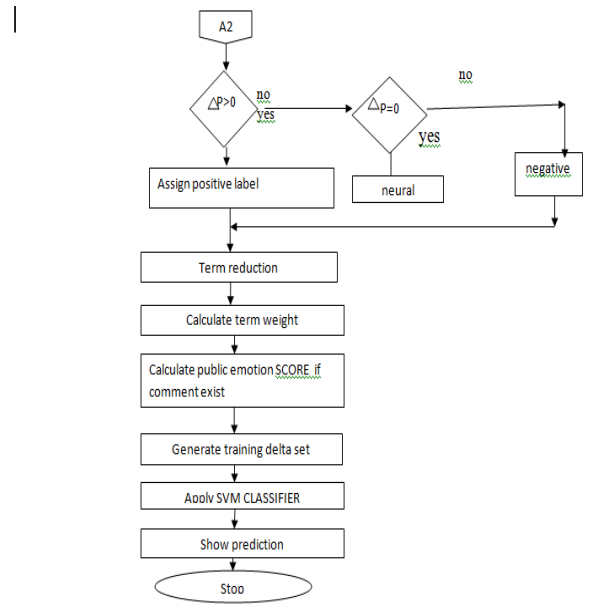
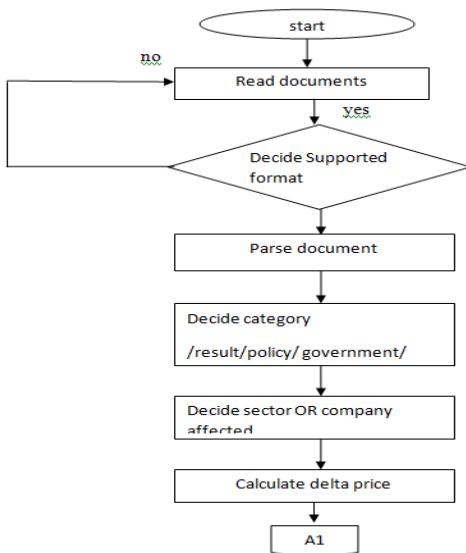
This all sub equations affected on the M-Score equation.

If the resulting value of the M-Score equation is less than -2.22 then the company unlikely to be a manipulator or the value is greater than -2.22 then the company likely to be a manipulator.

#### 4. METHODOLOGY

The Methodology shows how the whole research work with the help of the SVM and Neural network equation. By using the proper methodology we can get the impact of the financial result on the stock prediction with the higher accuracy.

##### Flow Of the research:



##### a. Read Documents

From the various companies, we have to get the quarterly result of the company and store it. And then use this all the results to get the accurate prediction. And then read all the documents from the company result also. This all stored results are used as the data set of the prediction process and decide which type of the documents.

##### b. Decide Supported Format

in to the company decision of the news, we have to decide some type of the format from the format of the decided the format by the programmer in this criteria we have to see the system consideration and find out the documents which depend on the system and which properly follows the actual format also decide supported format.

##### c. Parse Category

The classification process is done using various methods. But in this study, the classification process is done using the Support Vector Machine (SVM). Here we have to decide the category of documents. And the system consideration. The SVM classifies the linear and nonlinear data using the kernel function. In this study, we are using linear data so the classification is done based on the linear kernel function. The kernel function for the linear data is Radial Basis Function (RBF) Kernel. For the whole study, we are using the RBF kernel as a kernel function.

RBF Kernel is:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0$$

It classifies the company by the sector or the group and the financial information of any company by the profit or loss. Which depends on the company and then decide the format for the system. If the company is going to make a profit then it is considered as positive or the company is going to make a loss then it is considered as the negative.

#### d. Decide Category

In the next step of the methodology, the calculation of the score is done. And the system we have to decide the supported format the actual category of the system and we have some specific category of the news and the system have some decided some category and the news are depends on the decided system and then decide the category.

So the earning manipulation is also affected by the stock prediction process. Because the manipulated company gives the negative effect on the company's stock rate and manipulator company gives the positive effect on the company's stock value.

#### e. Decide Sector

Read the stock price using API. The system and which decide the sector and Inc. is a leading provider of an APIs for the real-time and the historical delta on the stock. Which are depends on the format and which type of the company news is related to the system? This API helps to read both real-time and historical data to make the prediction process more powerful and the decide the actual the sector of the system.

#### f. Calculate Delta Price

the news articles are believed to cause the movements based on delta closing price within the day the articles are established. Delta closing price for a specific day is formulated as a change in closing price from

the previous day  $P_{i-1}$  to the closing price in the day  $P_i$ :

$$\Delta P = P_i - P_{i-1}$$

The label of the articles

- Upward  $\Delta P > 0$
- Neutral  $\Delta P = 0$
- Downward  $\Delta P < 0$

#### g. term reduction

Calculating the delta price and then which depends on the term and reduction and depends on the reduction of the term. After this whole process the resulting data equal to the training data then the process becomes stop and if the data is going to be testing data then the process should go to the next stage. The training data is considered as the historical data and the testing data is considered as the real-time data. So the training data is neglected and the testing data is proceeding. So, either there is a testing data then the process going to stop or for testing data going to the next step.

#### h. Term weight

After getting the testing data the whole data classified using SVM. The SVM classified the data based on the profit or loss. And make a prediction about the stock value of the company. The prediction process is based on the score and the company information. The relation of the score and stock effects more on the prediction. This relation makes the prediction more accurate. It leads the predicted value near to the actual.

#### i. Calculate public emotion

After the prediction, we have to calculate the accuracy regarding the previous data or the historical data. And then the news how affect on the system and which is depends on the public emotion about the news's when the company's results declare then we have to check about how our prediction is near to the actual value. So the accuracy will calculated by the comparison of the actual data vs historical data. Higher accuracy gives the lower error rate. The error rate is decreased with the increase of the accuracy.

### 5. CONCLUSION

As Prediction in Stock market is a difficult task. As Price Change in Stock Market depends on so many points. We are gating to research over Impact of Daily News on Stock Market Companies. We will read daily news related to Financial Sector (Business/Government etc) using RSS Feed of online news portals. Then we are going to mine this news for its effect by category of news, who will be affected by this news is impact negative or positive? Etc. We will generate a model that can be used as a Training Data for Prediction Phase After this have to find the relation between the score and price movement using Support Vector Machine (SVM) because SVM gives high accuracy to the high dimension data. SVM is a useful technique for data classification, regression, and prediction. Using this method the prediction will be more accurate

### 6. REFERENCES

- [1] Minh Dang, Duc Duong, Improvement Methods for Stock Market Prediction using Financial News Articles IN 2016.
- [2] Chayanin Wong, and In-Young Ko, Predictive Power of Public Emotions as Extracted from Daily News Articles on the Movements of Stock Market Indices IN 2016
- [3] Felix Ming Fai Wong Zhenming Liu Mung Chiang Stock Market Prediction from WSJ: Text Mining via Sparse Matrix Factorization IN 2014.
- [4] Stefan Lauren Dra. Harlili S. M.Sc Stock Trend Prediction Using Simple Moving Average Supported by News Classification IN 2014.
- [5] Germ'an G. Creamer, Yong Ren, Jeffrey V. Nickerson. Impact of dynamic corporate news networks on asset return and volatility IN 2013
- [6] Hiral R. Patel Satyen M. Parikh Dhara N. Darji Prediction Model for Stock Market using News based different Classification, Regression and Statistical Techniques IN 2016
- [7] B. Dbouk and I. Zaarour "Financial Statements Earnings Manipulation Detection Using a Layer of Machine Learning", International Journal of Innovation, Management and Technology, Vol. 8, No. 3, June 2017
- [8] Bingyu Cao, Deping Zhan, Xianbin Wu, "Application of SVM in Financial Research", 2009 International Joint Conference on Computational Sciences and Optimization
- [9] Yauheniya Shynkevich1, T.M. McGinnity1,2, Sonya Coleman1, Ammar Belatreche1 Predicting Stock Price Movements Based on Different Categories of News Articles IN 2016
- [10] Anushree Priyadarshini and Sonali Agarwal, "A Map Reduce based Support Vector Machine for Big Data Classification", International Journal of Database Theory and Application Vol.8, No.5 (2015)
- [11] M. Cao and J. Wei, "Stock market returns: A note on temperature anomaly," JBank.Finance, vol. 29, no. 6, pp.

1559–1573, Jun. 2005SVM Ensembles”, IJACSA, Vol. 7, No. 5, 2016

[12] Wanjawa, Barack Wamkaya & Muchemi, Lawrence,” ANN Model to Predict Stock Prices at Stock Exchange Markets ”

[13] Shobana.V, Maheshwari.S, Savithri.M, “Study on Big data with Data Mining”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 4, April, 2015

[14] Mr. Sachin Sampat Patil, Prof. Kailash Patidar, Assistant Prof. Megha Jain, “Stock Market Prediction Using Support Vector Machine”, JAN-FAB, 2016