Classification of Family based on Financial Conditions and Stock Market Investing Recommendation

Dr Priya S
Assistant Professor,Department of Computer Application
Government First Grade College, Domlur, Bangalore ,Karnataka, India

Abstract

The price variation for the various stocks in the companies happens regularly. This variation must be tracked based on average value of the stock variables. First the set of organizations from yahoo finance are taken. For each of the organization the latest tweets are taken and analysis is performed on the tweets in order to derive the emotional number. The performance of the organization on the stock market will depend upon the social media tweets along with stock data. First for the given company code the stock for the whole year are obtained. Time series process is done on each of the companies on the stock based data.Suggesting the stocks based on performance in the stock market will not be a good attire for all users. Based on the monetary conditions of the consumer the group of organizations must be recommended. By making use of insurance spend, cost to company, rental expenses and other expenses the users are classified into GOLD, Silver or Bronze Category. Once the user group is determined the stock market companies under that specific group are ranked based on descending order stock probability

Keywords: **Stock Data, Machine Learning, Web Crawler, Data Cleaning, Time Series**

1.INTRODUCTION

The communication channels used by the organizations are analyzed. The overall performance of the organization will depend upon previous timeline based data for the stocks as well as emotions on the companies. The company information which are available on the public domain can be used for stock data analysis along with financial data. The company informational tweets are taken out for each company. For each of tweet, a tweet is divided into sentences, for each sentence emotions are analyzed and then a matrix is formed for the emotions. Find the distinct companies from the matrix of emotions and then all emotions score are added to get overall emotion score for the companies. Based on this emotions and historical data the companies on which trading has to be done are determined and then trading is performed.

The social media data is taken are then analyzed to obtain interesting facts related to company performance [1]. The stock price is determined by doing the analysis of sales [2]. Additionally the communication within the channels like jabber along with teams will help in having a better score for the companies [3]. Two clusters of have to be formed known as inter org cluster or an intra org cluster form which data can be captured, analyzed and the information can be communicated to the organizations. Stability factor along with robustness factor can be used for better communication analysis [4].

The communication theory is a subject of great interest. The functional equation can be used to determine the failure or success of the organization based on communications within the organization among the employees [5]. A stock model can be created based on mergers of the companies, acquiring of start up, performing the process changes and determine the bankruptcy of data as well. The stock summary is done and then trade is performed on a platform [6].

The historical data of all stocks are taken out using rules. Different fields like pattern flow, performing the feedback of the employees, public domain information along with events of the company [7-[8].

The analysis of jobs available on the company site, the employee performance is studied, group and individual productivity are used as variables and then a model is created for the stock [9].

The models are created based on communication between subordinates of a CEO, team heads, measuring the quality of information. The auto-correlation is done on the communication processes and emotional score will be generated [10].

2.MOTIVATION

The price variation is a complex entity which effects the stock of the product. The market emotions will also effect the stock. The better price is obtained for each of the companies based on price variation and emotions. Trading parameters help in rise or a fall of the company and are used in stock based applications. The time series data of the stock will be analyzed and then used for ranking the stocks of the companies.

The texts available in the news are analyzed along with comments which have specific commercial nature. On those comments the emotional value is found for all the companies as summary based on crawling on twitter. For each of the individual tweet emotion is found and then total emotional score is found as a summary score for the companies. The problem of the current approach can be summarized as below

1) The data of the stocks are taken from the stock market related API and then ranking of the companies is performed based on stock value. The methods does not consider emotions present in inter and intra-cluster.

2) The classification of users is not done in the existing system based on financial conditions of the users. Hence the process can provide better suggestions and unique suggestions for each user.

3.BACKROUND

The attributes of the stock are growth, intensity, capital investment, advertising factor. The class label is assigned based as either low or medium or high attractive company based on the input parameters. The analysis of the companies based on formal entity parameters, informal entity parameters, nature volume computation are used to compute rewarding stars. The organizations rank will depend upon decisions taken by the managers and the computed rewarding stars [11].

The overall performance of the corporation with respect to stock will depend upon multiple methods of communication. the primary method is communication with the friends or customers at once. the second one method is communique with the intermediate man or woman to the clients or other peers. The verbal exchange can be associated with corporation collaboration in and outside the enterprise [12]

The world settings are found based on financial market. The web application logs of a social media and any company survey application are taken as an input and then stock market value is predicted. The emotions score are done on anxiety level, worry level of the stock[13]. The amount of time an employee or a customer has used the companies assets are a factor to determine the emotional guidance score. The communication scale and performance variation of stocks of the companies must be taken into consideration and combined in order to have better stock market decisions [14]

The profits for an organization are gained based on better quality of service. The changes in the price of stock market and good communication process on the public domain results in better stock positions on the stock market[15]. The risk factor of a company depends upon two parameters namely accounting based parameter along with market based parameter. The stake holders will communicate the information on news websites, internal and external communication channels. The information can be analyzed to check the impact on the price of the stocks [16]. The stock price changes, properties along with structure of stock values over time are used for performing the training of the machine learning model. For the model to have a positive feedback the internal communication factor also plays an important role [17].

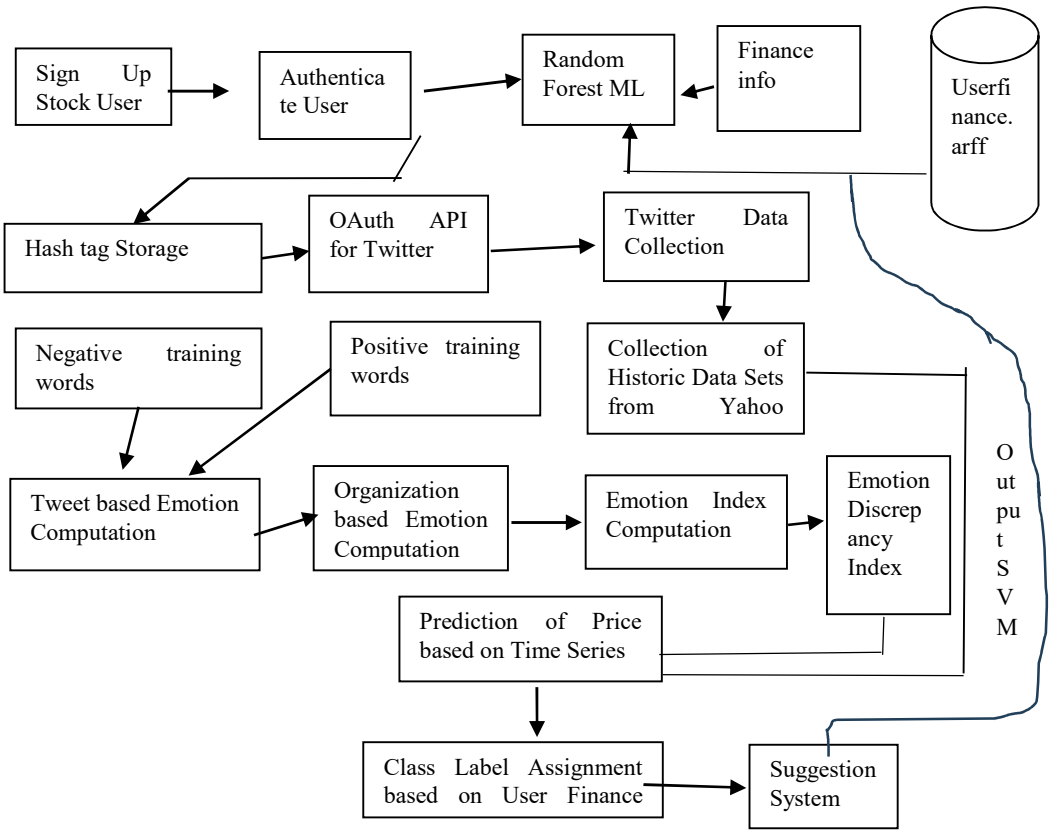

Figure1: User Classification and Stock Recommendations

4.METHODOLOGY

The process involved in classification of user is done with the help of machine learning process namely Random Forest based system and Support Vector Machine. The input training vector will have the financial information of the user and output is the class label which can be one of GOLD, SLIVER, PLATINUM or BRONZE.From the social media data such as twitter the companies information is fetched out and then emotional score is computed which can fall under either positive and negative. Based on the class of the user a group of organizations are combined and then sorted based on descending order of probable price.All the different modules to achieve the complete methodology are described as below

A. **Sign Up**

This module is responsible for providing the demographic information along with user id. If the user id is not unique then user is allowed to sign up otherwise sign-up otherwise sign up fails.

B. **Authenticate User**

The stock market user is allowed to authenticate and if successful authentication is completed then based on type of user the functionality is provided.

C. Machine Learning User Classification

The financial factors of the user are collected and then machine learning algorithm is trained. The random forest based machine learning method is executed by running multiple decision tree methods in parallel. Each of this decision tree will be based on information gain with entrophy computation to have better accuracy C4.5 decision tree will be used. The class label at the output will be one of GOLD, SLIVER, PLATINUM or BRONZE.

The step summary for the random forest method can be described in the following way
1. All the data collection from history is taken and various financial attributes are seggregated
2. The count of number of unique rows are found from the dataset.
3. The entire data set is divided into multiple independent datasets in a random manner.
4. For each of the datasets C4.5 method is executed.
5. The class labels are found out for each of the decision trees
6. The map of class label with count of trees is found
7. The class label which corresponds to maximum count is the final predicted user classification class.

D. **Hashtag Save**

The twitter system model consist of a hash tag. Each hashtag is having a set of stream of tweets. The module will collect the hashtags related to financial performance of the organization. Each hashtag-tuple will have name of company along with hashtag.

E.OAUTH API

The communication between the twitter based system and the finance recommendation system is done based combination of multi factors namely secret key variable along with oauth value variable. Once the Oauth system gets executed only for a specific session the communication is valid.

F. Twitter Data Collection

The stream of hashtags are read. For each of hashtag the latest tweets are obtained. The entire collected information is saved into a matrix format.TweetID, Tweet, Hashtag, Language, Company.

Table1: Twitter Data Information

Name	Description
Tweet ID	Unique ID for the tweet
Tweet	Tweet performed by the company
Hashtag	Hashtag to which tweet belongs
Language	Language used in the tweet
Company	Name of the company

G. Collection of Historic Data Sets from Yahoo API

The yahoo finance API secret key and token are collected after registering. From the yahoo finance API real time data for the stocks for the entire year are collected and various parameters like Volume, Daily Turn Over, Marketing Index along with Price are taken. These steps are executed for all the companies on which stock market analysis have to be performed.

H. Training Keywords - Positive Emotion

The data set related to positive emotions are taken into consideration as a training data

I. **Training Keywords - Negative Emotion**

The data set related to negative emotions are taken into consideration as a training data
The keywords are used to train the naive bayes method

J. Tweet Based Emotion Computation

The Naive Bayes method will make use of positive and negative keywords as training for the model. Once the model is created for the new tweets the positive and negative score is created. The class label for the tweet is generated by using the steps as shown in Figure2

Process: Tweet Emotion Analysis

Input: Tweet data, Positive and Negative training vectors

Steps

1. Find the real time tweets which are related to organization List
2. Find the size of tweet list
3.
4. For each of the element in the list
    a. Find the tweet information
    b. The entire tweet information is sub divided into set of sentences based on a uniquely configured separator.
    c. Find the emotion score for each sentence based on training data
    d. Emotion Score of postive and negative are compared and based on which ever score is higher the emotion is assigned a factor of 1.
    e. Summation of score of emotions for all sentences of a tweet is performed
5. The total emotion score for the tweet is saved in the format containing attribute {TweetId, Positive Emotion, Negative Emotion, Organization Name}

K.Organization Emotion Computation

The process will compute the emotional score across entire organization. The organization emotional values computation can be summarized in Figure 3

Algorithm: Organization Based Sentiments

Input: Tweet based emotional Analysis

Output: organization based Sentiment Score

Process

1. Obtain the unique organization names from social media
2. For each of the organization compute the following

$$GE(c) = \sum_{k=0}^{n} \mathrm{GE(k)}$$

$$BE(c) = \sum_{k=0}^{n} \mathrm{NE(k)}$$

Where, GE(k) good emotion score of kth tweet for organization c, NS(k) is the negative score of kth tweet for organization c, GE(c) is the total good emotions for organization and BE(c) is the total bad emotions

3. The total emotions for each organization will be computed as {organization Name, Emotion Good , Emotion Bad}

L.Emotion Score Computation

The Emotion Index Computation is done based on Number of Positive emotions along with Number of Negative emotions. The calculation can be done based on equation (1)

$$ESC = \ln((1 + Ng)/(1 + Nb)) \quad \ldots\ldots (1)$$

Where ,

$ESC = emotional\ \ score\ \ computatio\ \ n$

$Ng = number\ \ of\ \ good\ \ emotions$

$Nb = number\ \ of\ \ bad\ \ emotions$

M.Emotion Discrepancy Measure

The Emotion Discrepancy Measure (EDM) will give the emotion score which can effect the stock recommendations and can be computed by making use of equation (2)

$$EDI = |1 - |Np - Nn/Np + Nn|| \quad .....(2)$$

Where,

Np = number of positive emotions

Nn = number of negative emotions

EDI = emotional descriptive index

N. Price Change Computation

The process change for the organization stocks will vary based on volume, amount, price,daily turnover and emotional discrepancy index

$$MCi = 0.0628 * MCi - 1 + 0.0455 * Vli - 1 + 0.0283 * Ai - 1 + 0.0316 * Mi - 1 + 0.0423 * EDI + b \quad .....(3)$$

Where,

$MCi - 1 = monetory\ amount\ in\ previous\ iteration$

$Vli = volume\ during\ ith\ iteration$

$Ai = stock\ amount\ on\ ith\ index$

$Mi = marketing\ index$

$EDI = emotion\ \ description\ \ index$

$b = beta$

$i = ith\ time\ in\ the\ stock\ graph\ or\ time\ series\ data$

O. Class Label Assignment

The price values for all organizations are found out using equation(3). All the price values are sorted in the descending order. The entire company slots are divided into four different slots with first slot being GOLD, second slot being SILVER, third slot is PLATINUM and fourth slot is BRONZE.

P. Suggestion System

The suggestion system will provide the organization a set of companies stock which have a dependency on user classified group. Once the user group is determined then for that specific

group the stock prices are sorted and corresponding organizations are suggested to the end consumers.

5.EXPERIMENTAL RESULTS

This section describes the results of application which has been created and then application results

A. **Set Up For Application**

This part describes the set up with web server, database server and third party integration. The various components in the set up are provided in Figure 4. From the Figure 4 it can be seen that the user will hit the web application through the browser and the request will reach to the tomcat server. The stock market application is deployed on the tomcat server. The request comes to the web container which will the forward the request to the controller which is part of the stock market application. The stock market application has various services like registration, authentication, user financial classification, stock suggestions. Since the emotional score of the organizations has to be computed the stock market application will interact with twitter application using Oauth authentication mechanisms. The tweets data is analyzed and then emotional score is generated. The historical data of the organization stock is also required for a period of a year or more and this data is obtained by having an integration with yahoo finance API.

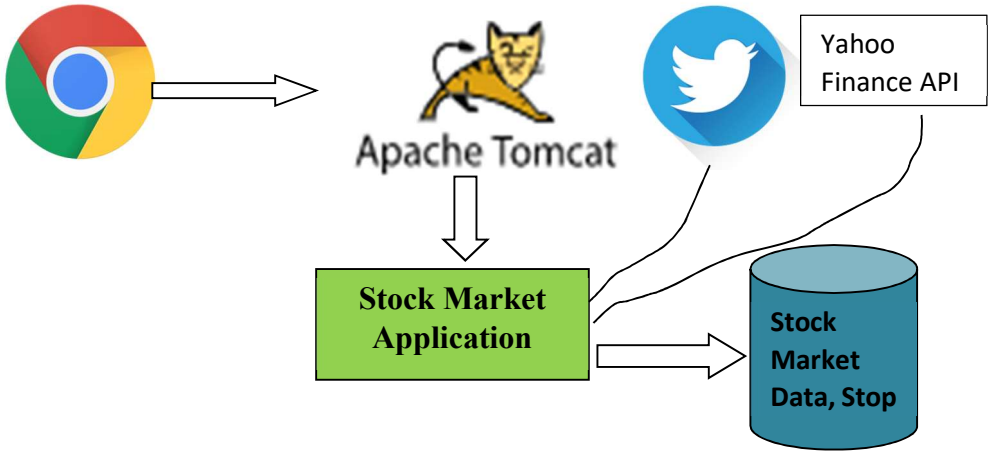

Figure4: Stock Market Suggestions Application Set Up

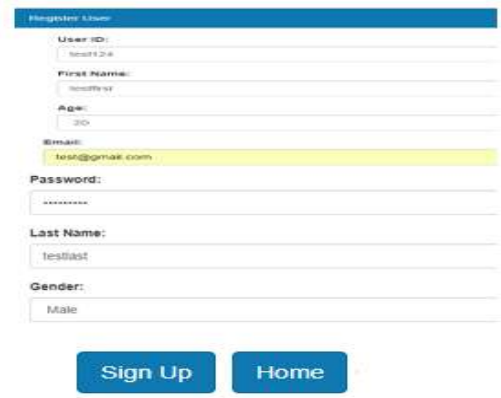

Figure5: Functional Consumer Creation

Figure5 suggests the end purchaser plays the registration in which the input may be consumer id, first name, age, email, password, last name and gender. If no customers have the equal person identity or same email, then registration procedure may be completed in any other case the registration procedure will fail.

Figure6: Sign in Process

Figure 6 shows the module which is responsible to allow both administrator or the customer to accessthe application. If the authentication is successful, then consumer is authorized to access diverse capability of utility otherwise failure happens.

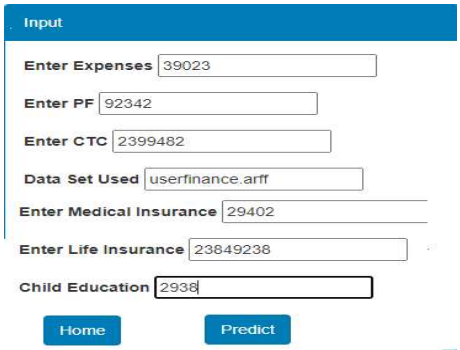

Figure7: Input for User Class Label Determination

Figure7 indicates the input for person financial computations. All the financial information related to job, spent on the life style are taken as an input parameters along with the training file name.

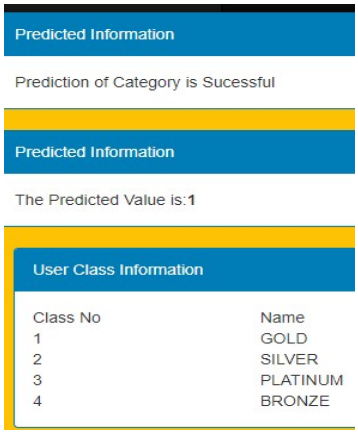

Figure8: Class Label Output Random Forest

Figure8 shows the output for the consumer category. This time the user has been predicted as being belong to GOLD using the machine learning process.

User Finance Class

Company Type	Company Type Detail
1	GOLD

Figure9: Category Class Label for User

Figure9 indicates the grid in which customer or user can see the category associated with him/herself after the ML algorithm has been processed.

Table 2: Average Stock Market Value using Yahoo Finance for a period of one year

Company Name	Volume	Price	Marketing Index	Daily Turn Over
YATRA	109.87	4238	78	98000
AMAZON	167.89	3264	89	67890
GOOGLE	755883.6019	513.37	10.699	5.8923

Table 2 suggests the stock market third party integration output in real time retrieval for diverse organization's

Table3: Emotion Computation and Stock Price Prediction

Company	YATRA	AMAZON	GOOGLE
Volume	0.011468	0.00499	0.00240
Price	0.2432	0.1435	0.0001
Marketing Index	0.01480	0.01233	0.00018
Daily Turnover	0.108	0.0667	0.0015
Emotion Dependency Index	0.222	0.120	0.106
Prediction	0.6644	0.5966	0.5046

Table3 indicates the sentiment score. The common volume, average rate, common marketplace index, average day by day turnover, emotion discrepancy index based on the social media information of organization and stock related attributes are also defined along with predicted performance store among the set of organizations.

Figure10: Company Stock Performance Prediction

The summary graph of the performance of an organization is done. The more horizontal flow the better is the gain which company can provide.

6. CONCLUSION

The existing emotion analysis system have been studied. This is followed by proposed stock market organizational suggestions with the help of both social media data and real time stock data. The forecasting of the stock data is done with the help of time series methods and specific organizations are suggested to specific category of people.

REFERENCES

[1] Z. Zhang, Z. Yin, J. Wen, L. Sun, S. Su and P. Yu, "DeepBlue: Bi-layered LSTM for tweet popUlarity Estimation," in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2021.3049529.

[2] R. Y. Kim, "Using Online Reviews for Customer Sentiment Analysis," in IEEE Engineering Management Review, vol. 49, no. 4, pp. 162-168, 1 Fourthquarter,Dec. 2021, doi: 10.1109/EMR.2021.3103835.

[3] Y. Hu, "Joint Analysis of Stock Price Predictive Models Using Past Stock Values and Public Sentiments on Twitter," 2021 5th Annual International Conference on Data Science and Business Analytics (ICDSBA), 2021, pp. 479-482, doi: 10.1109/ICDSBA53075.2021.00098.

[4] B. Collingsworth, R. Menezes, and P. Martins, “Assessing organizational stability via network analysis,” in Proc. IEEE Symp. Comput. Intell. Financial Eng., Mar. 2009, pp. 43–50.

[5] D. J. Barrett, “Change communication: Using strategic employee communication to facilitate major change,” Corporate Commun., Int. J.,vol. 7, no. 4, pp. 219–231, 2002.

[6] K. Yates. (2003). Effective Employee Communication Linked to Greater Shareholder Returns, Watson Wyatt Study Finds. Accessed: Nov. 2003.

[7] M.-C. Wu, S.-Y. Lin, and C.-H. Lin, “An effective application of decision tree to stock trading,” Expert Syst. Appl., vol. 131, no. 2, pp. 270–274,2006.

[8] E. Vamsidhar, K. V. S. R. P. Varma, P. S. Rao, and R. Satapati, “Prediction of rainfall using backpropagation neural network model,” Int. J. Comput. Sci. Eng., vol. 2, no. 4, pp. 1119–1121, 2010.

[9] C. W. Down, G. C. Phillip, and A. L. Pfeiffer, “Communication and organizational outcomes,” in Handbook of Organizational Communication, G. Goldhaber and G. Barnett, Eds. Norwood, NJ, USA: Ablex, 1988.

[10] P. G. Clampitt and C. W. Downs, “Employee perceptions of the relationship between communication and productivity: A field study,”J. Bus. Commun., vol. 30, no. 1, pp. 5–28, 1993.

[11] G. S. Hansen and B. Wernerfelt, “Determinants of firm performance: The relative importance of economic and organizational factors,” Strategic Manage. J., vol. 10, no. 5, pp. 399–411, 1989.

[12] W. W. Burke and G. H. Litwin, “A causal model of organizational performance and change,” J. Manage., vol. 18, no. 3, pp. 523–545, 1992.

[13] E. Gilbert and K. Karahalios, “Widespread worry and the stock market,” in Proc. 4th Int. AAAI Conf. Weblogs Social Media (ICWSM), 2010, pp. 59–65

[14] R. Katz, "The effects of group longevity on project communication and performance", Admin. Sci. Quarterly, vol. 27, no. 1, pp. 81-104, 1982.

[15] V. A. Zeithaml, L. L. Berry and A. Parasuraman, "Communication and control processes in the delivery of service quality", J. Marketing, vol. 52, no. 2, pp. 35-48, 1988.

[16] R. B. Higgins and B. D. Bannister, "How corporate communication of strategy affects share price", Long Range Planning, vol. 25, no. 3, pp. 27-35, 1992.

[17] J. Diesner, T. L. Frantz and K. M. Carley, "Communication networks from the Enron email corpus ‘It’s always about the people. Enron is no different", vol. 11, no. 3, 2006.