

STOCK PRICE PREDICTION USING MACHINE LEARNING

Sakshi Modase¹, Komal Mhaske², Akansha Narake³, Pratiksha Moholkar⁴

¹⁻⁴Computer Engineer, Satara-415509, Maharashtra, India

⁵Guided By – Prof. Dr. S.B. Chaudhari

Abstract – The main objective of this project is to find the best model to predict the value of the stock market. During the process of research and considering various techniques and variables that must be taken into account, we found out that techniques like random forest, support vector machine were not explored fully. In, this project we are going to present and review a more feasible method to predict the stocks with higher accuracy. The first thing we have taken into consideration is the dataset of the stock market prices which is the most recent one.

The dataset will be pre-processed and tuned up for real analysis. Hence, our project paper will also focus on data pre-processing of the raw dataset. Secondly, after pre-processing the data, we will review the use of random forest, support vector machine on the dataset and the outcomes it generates. In addition, the proposed project examines the use of the prediction system in real-world settings and issues associated with the accuracy of the overall values given. The project also presents a machine-learning model to predict the longevity of stock in a competitive market.

The efficient prediction of the stock will be a great asset for the stock market companies and will provide practical solutions to the problems that stock investors face.

Key Words: Machine learning, Dataset, Linear regression, knn, random forest, stock price prediction.

1. INTRODUCTION

Stock market plays a very important role in fast economic growth of the developing country like India. So our country and other developing nation's growth depend on performance of stock market. If stock market rises, then countries economic growth would be high. If stock market falls, then countries economic growth would be down. In other words, we can say that stock market and country growth is tightly bounded with the performance of stock market.

In any country, only 10% of the people engaging themselves with the stock market investment because of the dynamic nature of the stock market. There is a misconception about the stock market i.e., buying and selling of shares is an act of gambling. So, this

misconception can be changed and bringing the awareness across the people for this. The prediction techniques in stock market can play a crucial role in bringing more people and existing investors at one place. The more promising results of the prediction methods can change the mindset of the people. Data mining tools also helps to predict future trends and behaviours; helping organizations in active business solutions to knowledge driven decisions.

Intelligent data analysis tools produce a data base to search for hidden information that may be missed due to beyond expert's predictions. Extraction which was previously unknown, implicit and potentially useful information from data in databases, is an effective way of data mining. It is commonly known as knowledge discovery in databases (KDD). Although data mining and knowledge discovery in databases (or KDD) both are used as similar often, Data mining is actually part of knowledge discovery. Data mining techniques also play important role in stock market which can search uncover and hidden patterns and increasing the certain level of accuracy, where traditional and statistical methods are lacking. There is huge amount of data are generated by stock markets forced the researchers to apply data mining to make investment decisions. The following challenges are addressed by data mining techniques in stock market analysis

2. LITERATURE SURVEY

Many algorithms of data mining have been proposed to predict stock price. Neural Network, Genetic Algorithm, Association, Decision Tree and Fuzzy systems are widely used. In addition, pattern discovery is beneficial for stock market prediction and public sentiment is also related to predicting stock price.

There is a certain correlation among them. A review of previous studies on stock price forecasting shows the prevalent use of technical indicators with artificial neural networks (ANNs) for stock market prediction over the last two decades. Kunhuang and Yu [3] used backpropagation neural network with technical indicators to forecast fuzzy time series, the study findings showed that ANN has better forecast ability than time series model. One of the well-researched and most important algorithms in the field of Data mining is Association Rule Mining (ARM), Clustering & Decision Tree for financial forecasting. Strong rule generation is an important area of data mining. Association rule mining aims to obtain associations among item sets in data repositories.

Now, Apriori algorithm play a vital role in identifying frequent item sets and generating rule sets from it. Decision trees are excellent for making financial decisions where a lot of complex data needs to be taken into account. They provide an effective framework in which alternative decisions and the implications of taking those decisions can be laid down and evaluated.

They also form an accurate, balanced image of the risks and rewards that can result from a particular choice. In this research, the problem of discovering association rules was first introduced in 1993 and an algorithm called AISN was proposed for mining association rules [5]. For last fifteen years many algorithms for rule mining have been proposed. Wanzhong Yang [1] also proposed one innovative technique to process the stock data named Granule mining technique, which reduces the width of the transaction data and generates the association rules. R.V.Argiddi [2] has proposed fragment based mining which deals mainly with reducing the time and space complexity involved in processing the data in association rule mining technique.

As in granule mining, fragment-based approach fragments the data sets into fragments for processing thereby reducing the input size of data sets fed to the algorithm. In contrast to granule mining, in fragment-based mining the condition and decision attributes are summed for obtaining generalized association rules. analysis variables has been largely ignored like Price-earnings ratio, Moving average, rumours etc. Kannika Nirai Vaani M, E Ramaraj [4],[5] has now proposed new approach to generate association rules (E-Rules) i.e. Providing faster generation of frequent item sets to offer interesting and useful rules in an effective and optimized way with the help of Genetic Algorithm approach. From the above literature review, technical

indicators with different data mining techniques had been widely used, while there are only few studies of the use of fundamental indicators. The impact of fundamental But Prashant S. Chavan, Prof. Dr. Shrishail. T.Patil [10] has said that hybridized parameters gives better & more accurate results that applying only single type of input variables. A.A. Adebiyi, C.K. Ayo, M.O Adebiyi and S.O. Otokiti [11] has proposed predictive model has the potential to enhance the quality of decision making of investors in the stock market by offering more accurate stock prediction using hybrid parameters.

They used ANN for this but their performance is not always satisfactory. Robert K. Laia, Chin-Yuan Fanb, Wei-Hsiu Huang b, Pei-Chann Chang [6],[9] has proposed forecasting model that integrates a data clustering technique, fuzzy decision tree (FDT) and genetic algorithm (GA) to construct a decision-making system based on historical data and technical indices. Public information such as news, blogs, twitter mood, social networking sites and stock articles can also affect stock market trend. Web has been treated as a great source of financial information; many papers proposed stock price predicting approaches based on analysing web sentiments using text mining. Fung et al. (2002) predicted stock trend changes by analysing news articles with data mining and text mining techniques and a new statistical based piecewise segmentation was proposed. Schumaker and Chen [12] examined different textual representations of news articles to predict future stock price, which was compared to linear regression with SVM. Bollen et al. [13] analysed the daily text content in Twitter by mood tracking tools, and investigated to utilize these public mood time series for predicting the changes of DJIA closing values, which is based on a self-organizing fuzzy neural network

3. PROPOSED METHODOLOGY

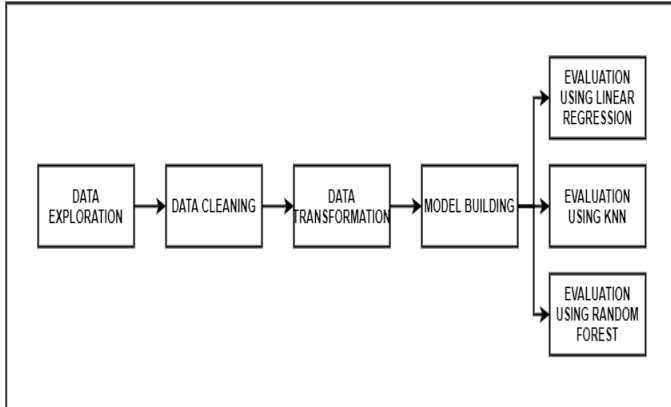


Fig 1. High Level Design

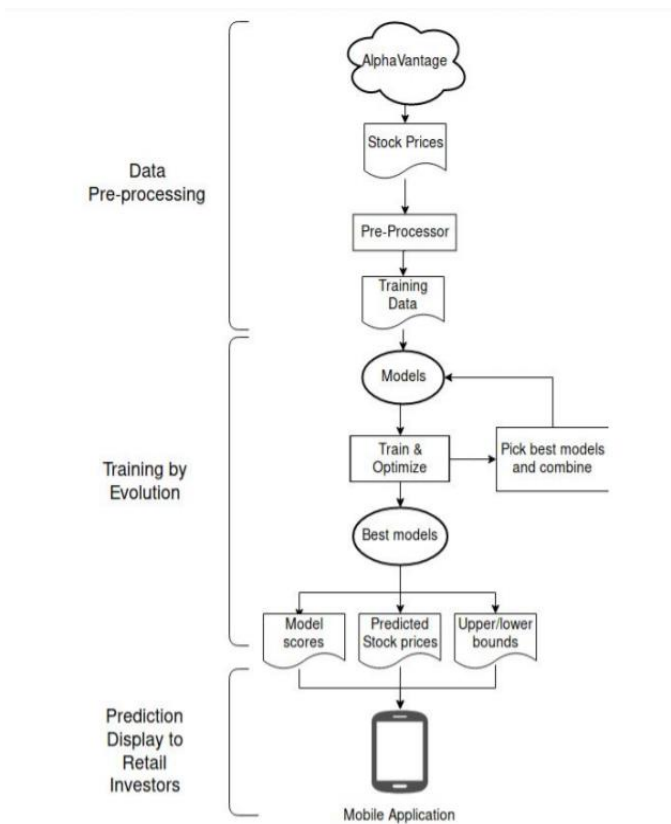


Fig. System Architecture

1. The working is separated into three main stages: Initial, Middle, Last stage.
2. The Initial stage is identified with Data Exploration, Data Cleaning and Data Transformation.
3. The centre stage comprises of data modelling.
4. The final stage comprises of data analysis using multiple models.
5. Data exploration is similar to initial data analysis, visual exploration to understand what is in a dataset and the characteristics of the data, rather than through traditional data management systems.
6. Data Cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.
7. Data transformation is the process of converting data from one format to another, typically from the format of a source system into the required format of a destination system.
8. Once the first stage is cleared then we move to data modelling. Data modelling is the process of producing a descriptive diagram of relationships between various types of information that are to be stored in a database. One of the goals of data modelling is to create the most efficient method of storing information while still providing for complete access and reporting.
9. After this the data is processed using algorithms and results are obtained.
10. This result are the test results generated by training the models on the train dataset.
11. Once the dataset is processed then we can make use of the actual dataset to predict stocks.

4. EXPECTED OUTCOMES

The movement in the stock market is usually determined by the sentiments of thousands of investors. Stock market prediction, calls for an ability to predict the effect of recent events on the investors. These events can be political events like a statement by a political leader, a piece of news on scam etc. It can also be an international event like sharp movements in currencies and commodity etc. All these events affect the corporate earnings, which in turn affects the sentiment of

investors. It is beyond the scope of almost all investors to correctly and consistently predict these hyper parameters. All these factors make stock price prediction very difficult. Once the right data is collected, it then can be used to train a machine and to generate a predictive result

5. CONCLUSION AND FUTURE SCOPE

The aim of our project is to help the stock brokers and investors for investing money in the stock market. The prediction plays a very important role in stock market business which is very complicated and challenging process due to dynamic nature of the stock market. Concluding this project, since a single source of data is used for analysis purpose and multiple models are used for analysis of the same data source it should be noted that it is possible to compare recognition accuracies of this study to find the best prediction model.

In any case, the small feature set implemented in this work seems to be enough for further research. The best prediction model is the model with highest prediction accuracy or highest confidence value. Future work can include assessment of additional models probably with an exhaustive search of optimum classification parameters, increasing dataset size and number of parameters.

6. REFERENCES

[1] Wanzhong Yang. "Granule Based Knowledge Representation for Intra and Inter Transaction Association Mining", Queensland University of Technology, July 2009.

[2] R.V.Argiddi, S.S.Apte "a study of association rule mining in fragmented item-sets for prediction of transactions outcome in stock trading systems" IJCET-2012.

[3] Kunhuang, H. & Yu, T.H.K. (2006). The application of neural networks to forecast fuzzy time series. *Physical A: Statistical Mechanics and Its Applications*, 363(2), 481-491.

[4] KANNIKA NIRAI VAANI M, E RAMARAJ " AN ENHANCED APPROACH TO DERIVE DISJUNCTIVE RULES FROM ASSOCIATION RULE MINING WITHOUT CANDIDATE ITEM GENERATION" IJCA-2013.

[5] R. Agrawal, T. Imielinski, and A.Swami (1993) "Mining association rules between sets of items in large databases". (ACM SIGMOD '93).

[6] Pei-Chann Chang , Chin-Yuan Fan , Jun-Lin Lin (2011) "Trend discovery in financial time series data using a case based fuzzy decision tree" Elsevier Science Direct.

[7] R.V Argiddi, S.SApte " Future Trend Prediction of Indian IT Stock Market using Association Rule Mining of Transaction data" IJCA-2012.

[8] Anandhavalli M, Suraj Kumar Sudhanshu, Ayush Kumar and Ghose M.K. (2009) "Optimized association rule mining using genetic algorithm" *Advances in Information Mining*.

[9] Robert K. Lai , Chin-Yuan Fan , Wei-Hsiu Huang , Pei-Chann Chang. "Evolving and clustering fuzzy decision tree for financial time series data forecasting" Elsevier Science Direct(2009).

[10] Prashant S. Chavan, Prof. Dr. Shrishail. T. Patil" Parameters for Stock Market Prediction" IJCTA | (Mar-Apr 2013).

[11] A.A. Adebisi, 1C.K. Ayo, 1M.O Adebisi, and 2S.O. Otokiti "An Improved Stock Price Prediction using Hybrid Market Indicators" ICT. (2002).

[12] Robert P. Schumaker and Hsinchun Chen. (2009) "Textual analysis of stock market prediction using breaking financial news: The AZF in text system" *ACM Trans. Inf.*

[13] Johan B, Huina Mao and Xiaojun Zeng. "Twitter mood predicts the stock market" *Journal of*

BIOGRAPHIES-



Sakshi Modase.
BE(Computer)
Satara-415509



Komal Mhaske
BE (Computer)
Pune-411028



Akansha Narake
BE (Computer)
Pune-411028



Pratiksha Moholkar
BE (Computer)
Pune-411028