



Model to Predict Stock Price with Respect to Day of the Week

¹Dr. Rahul G. Thakkar, ²Mr. Vimal Patel, ³Dr. Manish Kayasth

Assistant Professor, ASPEE Agribusiness Management Institute, Navsari Agricultural University, Navsari, Gujarat, India

Assistant Professor, College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

³ Asst Prof. / Vice-Principal, UCCC & SPBCBA & Udhna Academy College of Computer Application and IT .

Abstract -In this paper, we want to portray the preprocessing steps to enriched numerical data to make it useful for Apriori algorithm in finding out the association rules. Processed data will be applied in association rule mining to generate rules which will give direction about how to do intraday trading on a company's stock. Attributes like company, date, time, open, high, low and close were used for preprocessing of data and data of those attribute will be on tick by tick (one minute) bases.

Above attribute can be used for mining and generating associations with the day of week, which is good for intra-day trading. Database will be build up by past five years tick data of major 45 companies of NSE, which will be use as a raw data for analysis and to generate the association rules using apriori algorithm. Concept of data cleaning, data integration, data transformation (binning) and Data Reduction (Discretization & Concept Hierarchy Generation) will be used to enrich the data to make it appropriate for analysis for apriori algorithm to generate association rules.

Analysis may be helpful in identifying companies for intra-day trading. It can helpful to uncover companies' trend with respect to other factors which will guide to have position in particular stock on particular day.

Keywords: *Intraday data, Binning, Discretization & Concept Hierarchy Generation.*

I. INTRODUCTION

Trend in equity markets (over specific days, weeks, months, and even years) have attracted a widespread attention and considerable interests among the researchers. Over the last few years, a vast number of the literature from the practitioner and academic fields studies day-of-the-week effects or day seasonality on returns of various assets, such as stock market index, debt securities, futures, foreign currencies and even commodities. There are more than fifty research papers published on this topic from different perspectives by 2012. It is on the assumption that a certain pattern of stock markets is formed on the basis of the past stock price, which can be used to predict the future stock price. If the pattern is fixed, informed investors can utilize the pattern to earn a risk-free profit by trading the stocks.

II. LITERATURE REVIEW

The study by Brian Lent, Arun Swami and Jennifer Widom entitled "Clustering Association Rules" focuses on Association rule clustering is useful when the user desires to segment the data and Binning plays an important role in it.

The study of María N. Moreno, Saddys Segrera, Vivian F. López and M. José Polo "Improving the Quality of Association Rules by Preprocessing Numerical Data" focuses on Numerical attribute management is a usual preprocessing task in data mining. Most of the algorithms require the discretization of numerical attributes by splitting the continuous range of values into intervals.

The study of R. Krishnan and Vinod Mishra "Intraday Liquidity Patterns in Indian Stock Market" focuses on NSE stocks volume and spread related liquidity measures exhibit an intra-day U-shaped pattern.

The study by Tirthankar Patnaik and Ajay Shah "Intraday behavior of stock markets: A study of the Indian Equity Markets" focuses on Intraday volatility is usually high at the start and the close of the trading interval and divided the day into three parts, based on the high-volatility period.

The study of Sobhesh Kumar Agarwalla and Ajay Pandey "Whether Cross-Listing, Stock-specific and Market-wide Calendar Events impact Intraday Volatility Dynamics? Evidence from the Indian Stock Market using High-frequency Data" focuses on the intraday variation is more in case of the large cap stocks. Higher volatility is observed in the first and the last one hour of the trade.

The study by Hakan Berument and Halil Kiyamaz entitled "The Day of the Week Effect on Stock Market Volatility" focuses on day of the week effect on stock market volatility by using the S&P 500 market index during the period of January 1973 and October 1997. The findings shows that the day of the week effect is present in both volatility and return equations. While the highest and lowest returns are observed on Wednesday and Monday, the highest and the lowest volatility are observed on Friday and Wednesday, respectively.

The study by Dimitris Kenourgios University of Athens, Department of Economics, Greece entitled "The Day of the Week Effect Patterns on Stock Market Return and Volatility: Evidence for the Athens Stock Exchange" focuses on the day of the week effect on return and volatility for major Athens Stock Exchange (ASE) indexes. Using a conditional variance framework, which extends previous work on the Greek stock market, it was found that the day of the week effect in both the return and volatility equations is present for the emerging ASE over the period 1995-2000.

The study by Rahul G. Thakkar and Dr. Rustom D. Morena entitled “Predicting Stock Movement Using Classification Rules: A Data Mining Approach” focuses on few attributes of quarterly data like income, expenditure, PE, EPS, etc and database was built up of last 5 years of all the companies listed in NSE and Decision Tree is used to mine the database and rules were generated which is useful in selection of stock on the basis of quarterly results generated by the respective company.

The study by Rahul G. Thakkar and Dr. Rustom D. Morena entitled “Data Mining Approach for Predicting Stock Movement by Considering Fundamental and Technical Aspects” focuses on daily bhav copy of last few year of all the companies listed in NSE. Rules are generated by considering both the aspects using Decision Tree with higher accuracy.

The study by Md. Lutfur Rahman Department of Business Administration, Bangladesh entitled “Stock Market Anomaly: Day of the Week Effect in Dhaka Stock Exchange” focuses on the presence of day of the week effect anomaly in Dhaka Stock Exchange (DSE). Several hypotheses have been formulated; dummy variable regression and the GARCH (1, 1) model were used in the study. The result indicates that Sunday and Monday returns are negative and only positive returns on Thursdays are statistically significant. Result also reveals that the mean daily returns between two consecutive days differ significantly for the pairs Monday-Tuesday, Wednesday-Thursday and Thursday-Sunday. Result also shows that the average daily return of every working day of the week is not statistically equal. Dummy variable regression result shows that only Thursdays have positive and statistically significant coefficients.

The study by Thomas Hellström¹ and Kenneth Holmström² Center of Mathematical Modeling (CMM) Department of Mathematics and Physics Mälardalen University Sweden entitled “Predicting the Stock Market” focuses on various approaches of technical and fundamental analysis and the prediction problem is formulated as a special case of inductive learning using time series and data mining.

The study by Aboudou Maman Tachiwou (Corresponding author) S/C MAMAN - WATARA Gaouzou entitled “Day-of-the-Week-Effects in West African Regional Stock Market” focuses on the presence of the day of the week effects in West African regional stock market in the sample for the period September 1998 to December 2007. A daily pattern in stock markets is observed for the two indices of West African Regional stock market return.

The study by Hameed Al-Qaheri, Aboul Ella Hassanien and Ajith Abraham Entitled “Discovering Stock Price Prediction Rules Using Rough Sets” focuses on a generic stock price prediction model using rough set theory. The model was able to extract knowledge in the form of rules from daily stock movements. These rules then could be used to guide investors whether to buy, sell or hold a stock. For comparison purposes, the results obtained using rough sets were compared to those generated by neural networks algorithms. It was shown, using the same constrained data set, that rough set approach have a higher overall accuracy rates and generate more compact and fewer rules than neural networks

The study by Benjamin Liu and Bin Li Griffith Business School, Griffith University, Brisbane, QLD 4111, Australia entitled “Day-of-the-Week Effects: Evidence from Top 50 Australian Stocks” focuses on day-of-the-week effects in the top 50 Australian companies across different industry sectors was made. Unlike other Australian studies, the study was of weekday seasonality using stock return data of individual companies. Utilizing the daily data for the period of January 2001 through June 2010, it was found that weekday anomalies are mixed across companies and industries. It was also found that largest mean weekday returns occur on Monday for 15 companies, most of which are the materials and energy companies. Further tests indicate that returns on Monday are significant larger than the other four days for six companies.

The study by Ehsan Hajizadeh*, Hamed Davari Ardakani and Jamal Shahrabi Industrial Engineering Department, Amirkabir University of Technology, Tehran, Iran entitled “Application of data mining techniques in stock markets: A survey” focuses on application of data mining techniques such as decision tree, neural network, association rules, factor analysis and etc in stock markets. Also, it reveals progressive applications in addition to existing gap and less considered area and determines the future works for researchers.

The study by Rahul G. Thakkar and Dr. Rustom D. Morena entitled “Data Mining Approach to Identify Stock Trend in Relation with Day of the Week by Considering Technical Aspects” focuses on suggesting methodology for generating rules having association of stock with Day of the week using Association Rules by considering tick by tick data.

The study by Faryad Hussain Department of Management Sciences, Kashif Hamid Institute of Business Management Sciences, Rana Shahid Imdad Akash Department of Business Administration, Majid Imdad Khan Department of Management Sciences, Pakistan entitled “Day of the Week Effect and Stock Returns: (Evidence from Karachi Stock Exchange-Pakistan)” focuses on day of the week effect with a stock market anomaly on the equity market practices in Pakistan. The regression analysis is performed to meet the thrust of this study. The modulus-operandi applicable in this research consists of daily stock prices concerned to KSE-100 Index, for the period January 2006 to December 2010. The working week for trade matters consist of five days. Study concludes that Tuesday returns are quite significant and positive. Hence it is inferred that there exists day effect in Pakistani stock market. The returns of Tuesday on an average are greater in comparison to rest of the days.

The study by Jo Ting, Tak-chung Fu, and Fu-lai Chung† Department of Computing, Hong Kong Polytechnic University Hung Hom, Kowloon, Hong Kong entitled “Mining of Stock Data: Intra- and Inter-Stock Pattern Associative Classification” focuses on a pattern-based stock data mining approach which transforms the numeric stock data to symbolic sequences, carries out sequential and non-sequential association analysis and uses the mined rules in classifying/predicting the further price movements is proposed. Two formulations of the problem are considered. They are intra-stock mining which focuses on finding frequently appearing patterns for the stock time series itself and inter-

stock mining which discovers the strong inter-relationship among several stocks. Three different methods are proposed for carrying out associative classification/prediction, namely, Best Confidence, Maximum Window Size and Majority Voting.

The study by Y.L. Hsieh¹, Don-Lin Yang¹, and Jungpin Wu Department of Information Engineering and Computer Science Feng Chia University, Taichung, Taiwan entitled “Using Data Mining to Study Upstream and Downstream Causal Relationship in Stock Market” focuses on the causal relationship of stock market. Most investors use some fundamental knowledge and basic analysis techniques to analyze or predict the trends. However, there are always some other factors beyond our control or unexpected events that might affect the stock market one way or the other.

Individual stock intraday movement and how they are related with the days of the week is a new study area in Data Mining field. Yet only single research has done on top 50 companies of Australian stock market. Other all studies were done in respect to this is to find the relationship between the days of the week effect with overall stock market indices, not only in the USA and other developed markets but also in the emerging markets like Malaysia, Hong Kong, Turkey. For most of the western economies, (U.S.A., U.K., Canada) empirical results have shown that on Mondays the market has statistically significant negative returns while on Fridays statistically significant positive returns. In other markets such as Japan, Australia, Singapore, Turkey and France the highest negative returns appear on Tuesdays. In most developed markets such as the USA's, the United Kingdom's and Canada's, most studies, Cross (1973), Gibbons & Hess (1981), Keim & Stambaugh (1984), Theobald and Price (1984), Jaffe & Westerfield (1985), Harris (1986), Simrlock & Starts (1986), Board and Sutcliffe (1988), and Kohers and Kohers (1995), Tang and Kwok (1997) for six indices [Dow Jones Industrial Average Index(US), Financial Times Index (UK), Nikkei Average Index (Japan), Hang Seng Index (Hong Kong), FAZ General Index (Germany) and All Ordinary Index (Australia)] and many others, have come to the conclusion that Mondays' average returns are negative and Fridays' are positive.

The earliest research on return seasonality can be traced back to as early as the late 1920s (Pettengill, 2003). Up to 2010, there are many published papers on this topic from different perspectives. All research were done in respect to find the relationship between the days of the week effect with overall stock market indices, single research has done on top 50 companies of Australian stock market relationship with day of week. Certain research clearly shows that Data Mining is useful technique to do research in stock market. Not a single research found which use the tick data of one minute for predicting stock intraday movement, research done on 50 Australian companies use daily base data, so using tick data of one minute will be more meaningful as compare to that. There is not a single research found which will give the relationship between intraday movements of a particular stock on fixed interval on a particular day. Many researchers has use time series analysis, statistical techniques and data mining techniques in doing research on stock market data, inspiring to that use of association rule mining (Apriori Algorithm) will lead to the best result in finding out relationship between intraday movements of a particular stock on fixed interval on a particular day.

Many researchers' focuses on the high volatility in the first and the last hour of the trade, generally the shaped found for volume and volatility is 'U' shaped. Some of the researcher has also divided the trading period of whole day into three parts based on the volatility for their research work.

It is also found that to improve quality of the association rules preprocessing numerical data is necessary and done through Discretization & Concept Hierarchy Generation technique and association rule clustering is useful when the user desires to segment the data and in which Binning plays an important role.

III. RESEARCH METHODOLOGY

BUILDING DATABASE

The basic requirement for this research was to get the tick by tick data of past few years. We obtained the tick by tick (one minute) data of 48 companies consisting of date, time, open, high, low, close of each working day from January 2008 to January 2013. We want to use it as a main source of data.

CLEANING

We want to check for the missing values. If missing values are found in the database then past 10 transactions will be taken for that particular field and average will be calculated for filling that missing price. Care will be taken to check that the new price falls within the high and low prices of that minute.

We want to convert the database into five minutes data from one minute data to make it more meaningful.

DERIVING ATTRIBUTE

On the basis of date we will derive DAY attribute which possesses value which depicts day of the week. This attribute will be derived to find the correlation between the stock trend & day of the week.

CALCULATING PERCENTAGE CHANGE

To generate a trend, we need to calculate percentage change for open, high, low, close price of each stock on each day as compare to that of previous record in which we have to eliminate the first record of each company on daily basis.

DATA DIVISION

We want to partitioned records on the basis of the time for each day. Each partition will be termed as a session and it is divided as given below:

Session I: First hour of the trading.

Session II: Duration between Session I and Session II.

Session III: Last hour of the trading.

These sessions will be created on each company on daily basis to know the relation between stock trend and the day of the week.

DATA TRANSFORMATION

We want to use Binning technique for removing noise from the data on the open, high, low and close attributes on the basis of the sessions that we will create. In smoothing by bin means, each value in a bin is replaced by the mean value of the bin. Smoothing is a form of data cleaning and data transformation, which works to remove the noise from data.

DATA REDUCTION

We want to use Data Reduction (Discretization & Concept Hierarchy Generation) Techniques to obtain a reduced representation of the data set that is much smaller in volume; get closely maintains the integrity of the original data. Technique will be use to reduce the number of values for a given continuous attribute. Reducing the number of values for an attribute is especially beneficial if association rule mining is to be applied to the preprocessed data. A concept Hierarchy for a given numeric attribute defines a discretization of the attribute. Concept Hierarchy can be used to reduce the data by collecting & replacing low-level concepts. We want to reduce the data by collecting & replacing low-level concepts such as numeric values into three classes (low, medium and high) for the binning values that will be obtained for the open, high, low and close attributes.

DECIDING VALUATION

On the basis of steps of preprocessing of data like cleaning of data, data Integration, Binning technique and Discretization and Concept Hierarchy Generation techniques on the tick by tick data we will derive attributes like open_val, high_val, low_val and close_val which will be used in knowing the day trend of the stock and also on session basis.

IDENTIFYING TREND IN INDIVIDUAL SESSION.

Trend will be identified on the basis of derived attributes open_val, high_val, low_val and close_val. If all the values of the derived attributes in all records are high then we can say that there is a positive trend; if all the values of the derived attributes in all records are low then we can say that there is a negative trend; and if there is a mix of high and low records then we will obtain the number of low and high valuation in the record and decide the trend for that session.

On the basis of this we can know the trend of the stock on session basis. This will be useful in knowing the trend of the stock in different sessions of the day of the week except the first session.

IDENTIFYING TREND FOR WHOLE DAY

Trend of whole day depends on the trend which is identified in different sessions on that day. If all sessions have a high trend then we can say that a strong positive trend exists on that day; if all sessions have a low trend then we can say a strong negative trend exists on that day; and if it is a mixed trend then we have to get summation of result that we obtained to know the trend of each session. If it is positive with more number of sessions in high trend then we can say that a positive trend exists for that day of week; if it is negative with more number of sessions in low trend then we can say a negative trend exists for that day of week.

APRIORI ALGORITHM

Apriori is an influential algorithm for mining frequent Item set. We may use it for attribute open, high, low and close. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent item set; in our case it can be open, high, low, close and day.

L1= {frequent items};

for(k= 1; Lk!=∅; k++) do begin

Ck+1= candidates generated from Lk;

for each transaction t in database do

increment the count of all candidates in Ck+1that are contained in t

Lk+1= candidates in Ck+1with min_support

end

return

ACCURACY (HOLD-OUT METHOD)

We need strong association rules which satisfy both minimum **support** and minimum **confidence**. We want to use Hold-out method for determining accuracy in which two thirds of the data are allocated to the training set and the remaining one third is allocated to the test set. The training set is used to derive the classifier, the accuracy of which is estimated with the test set.

IV. CONCLUSION

Complete work will be based on the stock market data. The main goal of the research methodology that we have depicted is to derive a set of rules which can be used to select stocks on intra-day basis. For daily traders it will be beneficial to know the trend of the stock beforehand so that they may take appropriate position in the stock market building up confidence to trade on a particular day.

REFERENCES

- [1] Brian Lent, Arun Swami & Jennifer Widom. Clustering Association Rules. IEEE journal 1997.
- [2] María N. Moreno, Saddys Segrera, Vivian F. López and M. José Polo. Improving the Quality of Association Rules by Preprocessing Numerical Data. Journal of CEDI 2007, ISBN 978-84-9732-602-5.
- [3] Some Evidence on Intraday Volatility Patterns of Indian Market. Published in World Finance Conference in Italy.
- [4] Martin Becker* Ralph Friedmann. Intraday overreaction of stock prices. Published in journal of www.researchgate.net of germany.
- [5] Tirthankar Patnaik and Ajay Shah. Intraday behaviour of stock market:A study of Indian equity markets. Published in journal of Social Science Research Network, may 2004.
- [6] Sobhesh Kumar Agarwalla, Ajay Pandey. Whether Cross-Listing, Stock-specific and Market-wide Calendar Events impact Intraday Volatility Dynamics? Evidence from the Indian Stock Market using High-frequency Data. Published in journal of Social Science Research Network, W.P. No. 2012-11-03, Nov 2012.
- [7] Benjamin Liu, Bin Li. Day-of-the-Week Effects: Another Evidence from Top 50 Australian Stocks. Published in European Journal of Economics, Finance and Administrative Sciences ISSN 1450-2275 Issue 24 (2010).
- [8] Kenneth R. French. Stock Returns and the Weekend Effect. Published in Journal of Financial Economics 8 (1980) 55-69.
- [9] Y.L. Hsieh¹, Don-Lin Yang¹, and Jungpin Wu². Using Data Mining to Study Upstream and Downstream Causal Relationship in Stock Market. Published in journal of <http://www.atlantis-press.com/journals/>.
- [10] Jo Ting, Tak-chung Fu, and Fu-lai Chung. Mining of Stock Data: Intra- and Inter-Stock Pattern Associative Classification. Published in journal of citeseerx.ist.psu.edu.
- [11] Aboudou Maman Tachiwou. Day-of-the-Week-Effects in West African Regional Stock Market. Published in International Journal of Economics and Finance Vol. 2, No. 4; November 2010
- [12] Faryad hussain, Kashif Hamid. Day of the Week Effect and Stock Returns: (Evidence from Karachi Stock Exchange-Pakistan). Published in Far East Journal of Psychology and Business Vol 3 No 1 April 2011.
- [13] Hakan Berument and Halil Kiymaz. The Day of the Week Effect on Stock Market Volatilit. Published in Journal of economics and finance, Volume 25, Number 2, Summer 2001.
- [14] Dimitris Kenourgios, Aristeidis Samitas. The Day of the Week Effect Patterns on Stock Market Return and Volatility: Evidence for the Athens Stock Exchange. Published in International Research Journal of Finance and Economics ISSN 1450-2887 Issue 15 (2008).
- [15] Md. Lutfur Rahman. Stock Market Anomaly: Day of the Week Effect in Dhaka Stock Exchange. Published in International Journal of Business and Management Vol 4, No. 5, May 2009.
- [16] Rahul G. Thakkar and Dr. Rustom D. Morena. Decision Tree for Data Mining of Stock Market Database;
- [17] Rahul G. Thakkar and Dr. Rustom D. Morena. Predicting Stock Movement Using Classification Rules: A Data Mining Approach
- [18] Rahul G. Thakkar and Dr. Rustom D. Morena. Data Mining Approach to Identify Stock Trend in Relation with Day of the Week by Considering Technical Aspects.
- [19] Rahul G. Thakkar and Dr. Rustom D. Morena. Data Mining Approach for Predicting Stock Movement by Considering Fundamental and Technical Aspects. International Conference on Data Management.
- [20] V. A. Avdhani. Security Analysis & Portfolio Management. 8th Edition.
- [21] Robert Davis Edwards, John Magee, W. H. C. Bassetti. Technical analysis of stock trends.
- [22] David Louis Olson, Dursun Delen. Advanced Data Mining Techniques. PHI publishing.
- [23] Jiawei Han, Micheline Kamber. Data mining: concepts and techniques. Edition: 2, illustrated.
- [24] Jack D. Schwager. Getting started in technical analysis
- [25] S. Sumathi, S. N. Sivanandam. Introduction to data mining and its applications.