ISSN No. 0976-5697

Volume 3, No. 5, Sept-Oct 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Emperical Study Of Investors Preference Towards Investing In Commodity Futures Using Data Mining

Amit Kumar Marab
Department of CSE
GNDEC Bidar, Karnataka, India
amithbhalki@gmail.com

Channaveer Patil
Department of CSE
GNDEC Bidar, Karnataka, India
Channveer.patil@gmail.com

Srinivas Murthy*
Department of ISE
Basavakalyan Engineering College
Basavakalyan, Bidar, India
s.sri108@gmail.com

Abstract: Financial Investment is the allocation of money to assets that are expected to yield some gain over period of time .It is an Exchange of Financial claims like stocks, derivatives and bonds of money .Investors have many options to invest their money ,commodity is one of the market where investors can invest and get returns .In this paper we use Commodities Market characteristics and background , gathered data and applied Apriori algorithm to extract Rules and Patterns for the enhancement of the organization. This will help to analyze investor's preference investing in the commodity futures and measure the service provided by the organization.

Keywords: Commodity Market, Data mining, Weka, Association Rule Mining, Apriori Algorithm

I. INTRODUCTION

Commodity is any item that can be bought and sold. Commodity is a physical substance, such as food grains, Gold sliver and so on, which is interchangeable with another product of same type, which investors buy or sell through future contracts. The price of commodity is subject to supply and demand. Commodity derivatives records high volumes in the world compared to equidity derivatives [1]. With the other asset classes offering attractive returns," Why Commodities" is an inevitable question that pops in one's mind. Commodity Markets have two markets, Future and Spot .Therefore trading is a process where commodity is bought and sold in separate markets to take advantage of price, time and location difference between two different markets. India Infloline is probity research and services private company at Mumbai, under companies' act 1956.Our study proposes architecture based on data mining which aims at extracting Rules and patterns on (IndiaInfo line) data to analyze Investors preference investing in commodity futures and measure the service provided by the organization.

For studying investor's preference, towards investing on commodity future, a field survey was undertaken at India Infoline ltd, Gulbarga city. A questionnaire was prepared considering various parameters of respondents towards commodity futures .The information was also collected using various company brochures, publications and text books. The sample consists of initial 50 respondents with 18 attributes. (Figure 3 demonstrates all attributes loaded in WEKA Explorer). The financial analysts apply various statistical tools and techniques like likert's scale, one way table bars and graphs to represent and study the various factors considered by investors, while investing and measuring the service offered by (India Infoline ltd ,Gulbarga city) have the following problems (from the detailed survey).

a. Infeasible to test all potential hypothesis for large number of attributes
b. Testing Hypothesis with small sample has limited statistical power

Data mining on the other hand requires no hypothesis, can generate association rules independent of sample size, and thus mine association in large dataset with multiple temporal attributes. Thus this paper uses Data Mining to build a Model for reaching the enhancement of the organization which is discussed in Section III. The remaining part of the paper is organized as follows- Section IV discusses the Apriori Algorithm .Section V reveals the Implementation details of the Apriori algorithm on the Respondent data set using WEKA (Waikato Environment for Knowledge Analysis) .Section VI focuses on the analysis of the result using various Rules Generated. Finally Section VII Concludes

II. PROPOSED METHODOLOGY

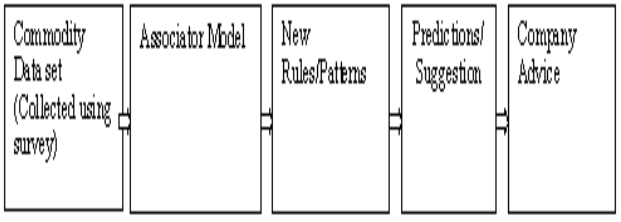

Figure. 1 Proposed Methodology Based On Data Mining

The Architecture and modelling of the current Paper is Depicted in Fig 1. It begins with the collection of respondent's records, which is followed by building an Associator model, of data set in which Class implementing an Apriori-type algorithm is implemented. The algorithm has an option to mine class association rules. Based on the Rules/Patterns Generated, one can come out with suitable suggestions and recommendations to enhance the organization.

III. DATA MINING AND APRIORI ALGORITHM

Data Mining also referred to as Knowledge Discovery in Databases or KDD is the search for Relationships and Global Patterns that exist in the large databases but are "hidden" among the vast amount of data [2][4]. Data mining also refers to extracting information from very large databases [3]. *Classification and Association* are two mechanisms to represent extracted information [3][5].Association Rules are of type A->B where A and B are the sets of attributes (items).

In Computer Science Data Mining, *Apriori* [6] is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation). Other algorithms are designed for finding association rules in data having no transactions (Winepi and Minepi), or having no timestamps (DNA sequencing)[7].As is common in association rule mining, given a set of *item sets* (for instance, sets of retail transactions, each listing individual items purchased), the algorithm attempts to find subsets which are common to at least a minimum number C of the item sets. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as *candidate generation*), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

Apriori uses breadth-first search search and a tree structure to count candidate item sets efficiently[6]. It generates candidate item sets of length k from item sets of length $k-1$. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates. Apriori, while historically significant, suffers from a number of inefficiencies or trade-offs, which have spawned other algorithms. Candidate generation generates large numbers of subsets (the algorithm attempts to load up the candidate set with as many as possible before each scan). Bottom-up subset exploration (essentially a breadth-first traversal of the subset lattice) finds any maximal subset S only after all $2^{|S|}-1$ of its proper subsets. The Pseudo Code for Apriori Algorithm is depicted in fig.2

Apriori (T, ϵ)
$L_1 \leftarrow \{$ **large 1-itemsets** $\}$
$k \leftarrow 2$
while $L_{k-1} \neq \emptyset$
$C_k \leftarrow \{c | c = a \cup \{b\} \wedge a \in L_{k-1} \wedge b \in \bigcup L_{k-1} \wedge b \notin a\}$
for transactions $t \in T$
$C_t \leftarrow \{c | c \in C_k \wedge c \subseteq t\}$
for candidates $c \in C_t$
$count[c] \leftarrow count[c] + 1$
$L_k \leftarrow \{c | c \in C_k \wedge count[c] \geq \epsilon\}$
$k \leftarrow k + 1$
$\bigcup L_k$
return k

Figure. 2 Pseudo Code for Apriori Algorithm

IV. EXPERIMENTAL RESULTS

For the implementation of Apriori algorithm to get new rules on India Infoline data, we used WEKA (Waikato environment for knowledge Analysis), the gathered data was entered in an excel sheet. This file was given as input to Weka Explorer in CSV (Comma Separated Value) format, as shown in the figure3 below. Rules were Generated By changing the Class index value, and minimum support value and we got distinguished Associator Models and Different Rules/Patterns, Which Are depicted in Fig4, 5, 6 respectively. Weka was created by researchers at the University of Waikato in New Zealand. It's a collection of open source of many data mining and machine learning algorithms, including [8]

a. pre-processing on data
b. Classification
c. Clustering
d. Association rule extraction

The Steps of Data Mining using WEKA for finding out the Association Rules on India Infoline data are as follows

Step 1- Click the Explorer on WEKA GUI

Step 2- On the Explorer Window, Click button open file to open data file from where India Infoline data file is stored in ARFF (Attribute Relation File Format) format.

Step3- After loading a data file click, associate and under Associate Choose Apriori algorithm and click start .The fig 4, 5,6 shows Results when Apriori Algorithm is applied over India Infoline Data Set, all Generated Rules are shown in the output pane. The Results vary according to Index Changed

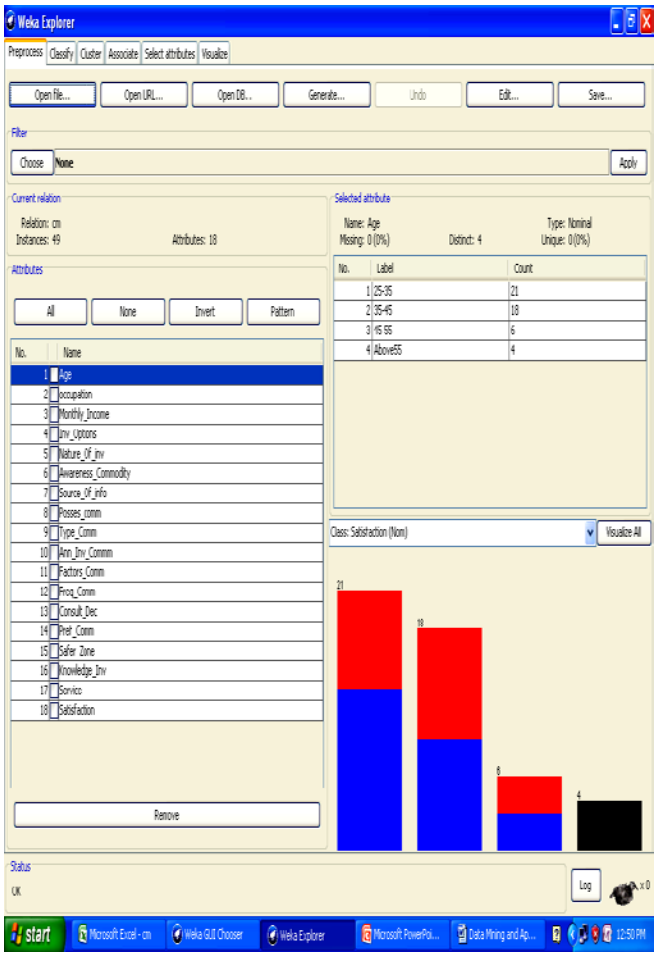

Figure.3. WEKA Explorer with attributes (Dataset)

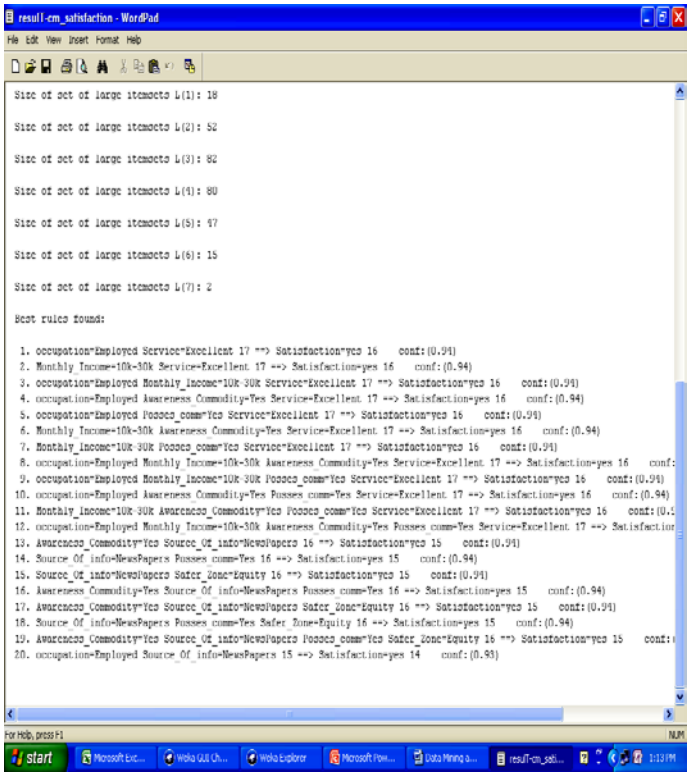

Figure. 5.Associator Model with Minsup=0.20 and Class Index=17

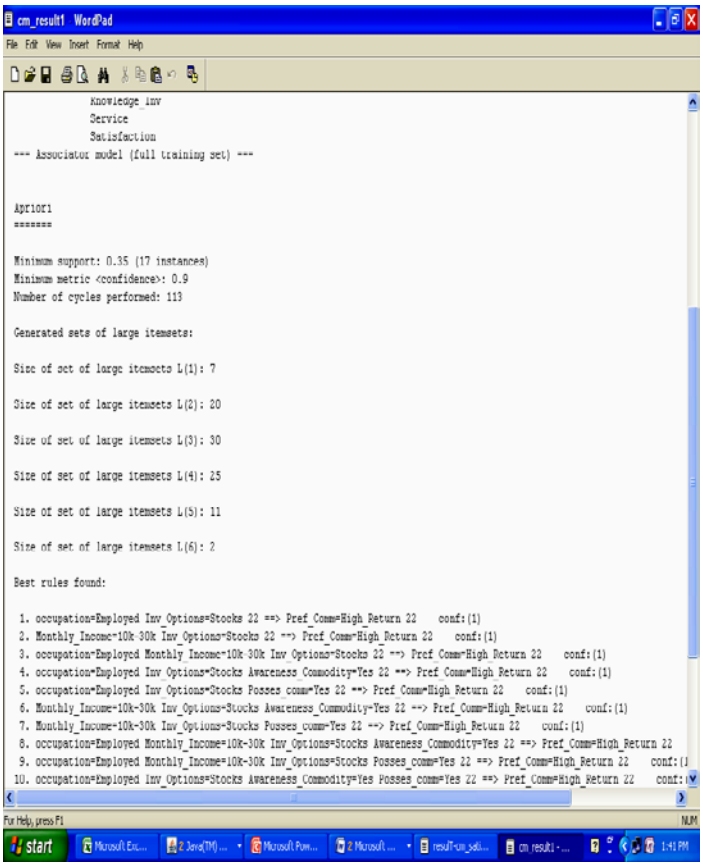

Figure. 4.Associator Model with Minsup=0.35 and Class Index=14

Figure-6 Associator Model with Minsup=0.20 and Class Index=18

V. RESULTS ANALYSIS

The Rules Generated are as follows-

a. Rule1 (Figure 4) can be interpreted as: The 35% of the respondents (Min sup=0.35) who were Employed and whose Investment option was stock preferred 100 %(confidence) to invest in High returns commodity futures
b. Rule 3(Figure4) can be interpreted as 35% of the respondents who had monthly income of 10,000 to 30,000 INR and whose investment option was stock preferred 100% (Confidence)in High returns commodity futures
c. Rule 10(Figure 5) Can be interpreted as 20% percent of the respondents whose monthly income was 10,000 to 30,000 INR and whose annual investment in commodity was 2, 50,000 to 3, 50,000 INR believed that service provided by India Info line was 100% (Confidence) Excellent
d. Rule 13(Figure 6) can be interpreted as, 20 %(Min sup=0.2) of the respondents who have awareness on commodity futures, who gather information on commodity futures through newspapers have satisfaction (94%) towards Brokerage charges of India Infoline ltd, (Gulbarga)

From the overall project ,the researcher has come to the conclusion that Investors preference investing in High Returns Commodity futures is very good(100%) .To conclude many of the respondents (80%) were not aware of commodity futures and the respondents who were aware (remaining 20%) were satisfied with brokerage charges of India Infoline, ltd Gulbarga and their level of satisfaction was 94%

VI. CONCLUSION

In today's situation, the organizations have a lot of difficulties to measure their level of service and customers satisfaction. In this paper discusses the problems of statistical analysis methods like likert's scale and one way table in the context of Financial Analysis .To overcome problems of these financial tools and techniques, a Model based on Data Mining has been proposed .Conclusions are made on the generated rules, for analysing investors preference and measuring the service of the organization by doing Data Mining on India Infoline data. These generated Rules can be used for Enhancement of the organization

VII. REFERENCES

[1] Prasanna Chandra, "Financial Management" 5th Edition, Publication: Tata Mc graw Hill, 2000

[2] Agarwal R, Imielinski T,Swami AN."Mining Association Rules between Sets of Items in Large Databases." SIGMOD. June 1993,22(2):207-16

[3] Jiawei Han, Micheline Kamber, and Jian Pei, Data Mining: Concepts and Techniques 3rd edition, Morgan Kaufmann, 2011.

[4] Chakrabarti, S. (2000): "Data mining for hypertext: A tutorial survey". SIGKDD explorations, **1**(2), pp. 1–11.

[5] Jae-won Park, Nam-Yong Lee, "A Conceptual Model of ERP for Small and Medium-Size Companies Based on UML",IJCSNS International Journal of Computer Science and Network Security, VOL.6 ,No.5A,May 2006.

[6] Apriori Algorithm http://en,wikipedia.Org/wiki/Apriori Algorithm ,Accessed Date: 13 November ,2011

[7] Akerkar, R. A.; Lingras, P. (2008). An Intelligent Web: Theory and Practice, 1st edn. Johns and Bartlett, Boston J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[8] Weka- Open source Machine learning Software in Java http://www.cs.waikato.ac.nz/ml/weka/, Accessed Date: 14 November, 2011