

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

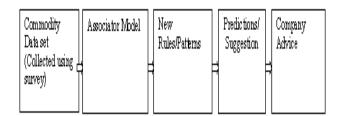


Figure. 1 Proposed Methodology Based On Data Mining

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

$$\begin{array}{c} \underset{k \leftarrow 2 \\ \text{while } L_{k-1} \leftarrow \{ \\ \text{large 1-itemsets} \} \} \\ k \leftarrow 2 \\ \text{while } L_{k-1} \neq \emptyset \\ C_k \leftarrow \{ c | c = a \cup \{ b \} \land a \in L_{k-1} \land b \in \bigcup L_{k-1} \land b \notin a \} \\ \text{for transactions } t \in T \\ C_t \leftarrow \{ c | c \in C_k \land c \subseteq t \} \\ \text{for candidates } c \in C_t \\ count[c] \leftarrow count[c] + 1 \\ L_k \leftarrow \{ c | c \in C_k \land count[c] \} \\ k \leftarrow k + 1 \\ \bigcup L_k \\ \text{return } k \end{array}$$

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

Al Nore Pattern Al Nore 1/5:55 21 2/5:45 13 3 Nore 3/6:53 2 propulsion 4/40xet5 3 Nature, 1/7, nore 4 4 Nore 4/40xet5 4 Nore 4/40xet5 3 Nature, 1/7, nore 4/40xet5 4 Nore 4/40xet5 5 Nore 4/40xet5 4 Nore 4/40xet5 5 Nore 4/40xet5 4 Nore 4/40xet5 5 Nore 4/40xet5 6 Norese, first 4/40xet5 6 Norese, first 4/40xet5 6 Norese, first 4/40xet5 9 Type_Com Cass: Satistation (Non) 10 Statistation 1/1 11 Statistation 1/1 12 Statistation 1/1 13 Statistation 1/1 14 Pate 1/1 15 Statistation 1/1 13 Statistation 1/1 14 Statistation 1/1 15 Statistation 1/1	process Classify	Custer Associate Sele	ct attributes Visualize							
Conse Mark Prediction Prediction Prediction Prediction Prediction Prediction All None All None Prediction Prediction All None Prediction Prediction All None Prediction Prediction All None Prediction Prediction Prediction P	Open file	Open URL.	Open 18	Gen	erate	Un	to 🗌	Edit	Save	
Conse Mark Prediction Prediction Prediction Prediction Prediction Prediction All None All None Prediction Prediction All None Prediction Prediction All None Prediction Prediction All None Prediction Prediction Prediction P	er.									
Junc Solidad athula Peldoric on Paldoric on Statuse: 49 Athulas: 18 Name: Age Water U(N) Definit: 4 User: Age Water U(N) None Mark Mark Pattern A None Mark Mark Pattern A None Mark Mark Pattern A None B Connon B Conn B Statistion B Conn B Statistion B Statistion B Statistion B Statistion B Statistion B Statistion										And
Rédoir on stance: 49 Athute: 18 Nue: Nue: Athute: 18	Unose None									ACDY
Indexes: 49 Athbuts: 13 Mosing D(M) Dotint: 4 Unique: 0(M) totals Intel: None Intel: 1/25-55 21 Al None Intel: 1/25-55 21 Al None Pattern 3/65-55 6 1 5/2 3/65 18 2 Conception 3/65-55 6 1 5/2 1/2 1/2 2 Conception 3/65-55 6 3 Mority Jonne - - 4 Abores: Concolty - - 7 Sorres: (Jripio 8 - 8 Poses: Connon - - 10 Ann, Inc. Comm - - 11 Satis: Com - - 12 Foro: Com - - 13 Satisfaction - - 14 Pattern - - 15 Satisfaction - - 16 Tork: Dest - - 17 Soria: Com - - 18 Satisfaction - - 19 Satisfaction - - <td>rrent relation</td> <td></td> <td></td> <td></td> <td>Selected at</td> <td>etudre</td> <td></td> <td></td> <td></td> <td></td>	rrent relation				Selected at	etudre				
Image: Second										
Image: Second	instances: 49		Attributes: 18		Missing: (1 (0%)	Distinct: 4	ι	hique: 0(0%)	
Al None Inset Pattern a. None Inset Pattern b. None Inset Inset a. Inset Inset Inset a. Inset Inset Inset a. Inset Inset Inset b. Norely jone Ins	tributes				NO.	Label		Count		
a. Nare 1 451 2 Status (i m) 3 Introduction 4 Introduction 5 Mater (i m) 6 Justes (i m) 6 Justes (i m) 7 Sorres (i min) 9 Type (cmm) 10 Anno (i m) 11 Parts (com) 12 Prog (com) 13 Sorres (i min) 14 Prog (com) 15 Sorres (com) 16 Sorres (com) 17 Sorres (com) 18 Sorres (com) 19 Sorres (com) 10 None 10 Sorres (com) 10 Sorres (com)					1	25-35		21		
b. Name 1 \$42 2 Scopetion 3 Mothly Done 4 Aboretis 5 Nature of prive 6 Averages Consolity 7 Sores of prive 8 Poses conn 9 Type Conn 10 Ren Dire Conn 11 Satistic Conn 12 Prog Com 13 Gravity Done 14 Prog Com 15 Satistic Conn 16 Prog Com 17 Sorie Con 18 Satistic Con 19 Satistic Con 10 Satistic Con 15 Satistic Con 16 Prog Con 17 Sorie Con 18 Satistic Con 18 Satistic Con 19 Satistic Con 10 Satistic Con 10 Satistic Con 10 Satistic Con	Al	None	Invert	Pattern						
1 66 2 toropetion 3 Mothly Jonne 4 IV (Jonos 5 Nature (J Inv 6 Averences Connodity 7 Sorver, J Info 8 Projection 9 Type Com 9 Type Com 11 Patas Com 12 Proje Com 13 Conk Jee 14 Pred Com 15 Store 16 Tore 17 Sories 18 Toroledge Tore 19 Satification										
2 copetin 3 torfwily Joure 4 nr (Jours 5 start (f inv 6 warens (Jonnoly 9 type Com 9 type Com 9 type Com 10 ent inv (Comn 11 stars (Com 13 Strok (be 14 Pet Com 15 star Une 18 Statistion 18 Statistion 19 type Com 10 ent inv (Com 10 ent inv (Com 11 stars (Com 13 Statistion 14 Pet Com 15 star Une 16 broke 17 Stroke 18 Statistion 18 Statistion		ie –			4	Above55		4		
3 britly / trane 4 m / tytors 5 bitru of jn // 6 weresets (annody) 7 Saves of jrite 8 brits (ann 10 en /n // Comm 11 states (ann 12 eq. Cmm 13 Grad, Die 14 Pet Cam 15 Safe Tore 15 Safe Tore 16 wereset 17 Saves 18 Safedan 19 protection 19 protection 10 Safedan 10 Safedan 10 Safedan 11 Safedan 12 Pet Cam 13 Grad, Die 14 Pet Cam 15 Safe Tore 15 Safedan 16 Pet Cam 17 Saves 18 Safedan 19 Pet Cam 10 Pet Cam 10 Pet Cam 10 Pet Cam 10 Pet Cam 11 Safedan 10 Pet Cam 10 Pet Cam 11 Safedan 11 Safedan 12 Pet Cam 13 Conscient (Safedan 14 Pet Cam 15 Safedan 15 Safedan 16 Pet Cam 17 Pet Cam 18 Safedan 19 Pet Cam 10 Pe										
S Nature (of jov 6 Heaveres (Connodity) 7 Source (of prin 8 Posse (connodity) 9 Type (conn 90 Type (conn 101 Anto (connom) 112 Anto (connom) 121 From (condity) 132 Group (connom) 133 Group (connom) 134 Pred (connom) 135 From (condity) 136 Tronke (connom) 137 Protec 138 Stafastion										
€ Ivertes (comody 7 Sover (F) fré 8 Pose (comody) 9 (ref. com 10 Int (comon 11 Post (comon 12 Post (comon 13 Grank (bet 14 Pote (comon 15 Sér Ure 16 Porte 18 Post (comon 18 Saldadion										
T Sove of Jride 8 Descession 9 Type Com 10 In Jry, Comm 11 Satis Com 12 Sove Joint 13 Cons Joint 14 Bet or Com 15 Safe Time 16 Desc Safe Safe 17 Sovie 18 Safe Safe 19 Safe Safe 10 Renoe										
B Poses runn 9 Poses runn 9 Dess Satistation (lun) v Toute 10 Han Jrv Comm 11 Posts Com 12 Eror, Com 13 Const Lee 16 Post Com 15 Satistation 17 Sorie 18 Satistation Renore Bat										
9 Type Cmm Cless Seldsation (kin) v Vextex / Ve										
10 Writ Jyn Comm 11 Status Com 12 Froz. Com 13 Grad. Jtes 14 Pré Com 15 Safe Une 16 Pré Com 17 Srice 18 Batisfation					Class: Salist	ection (Nom)			Vis.	ualize Al
11 Saturs_Com 12 Grow Loss 13 Grow Loss 14 Pet Com 15 Safe Zare 16 Donoledge Jaw 17 Savice 18 Safe Saton					1					_
12 Cook Units 13 Cook Units 14 Pré Com 15 Sée Units 16 Pré Com 17 Srice 18 Pré doin 18 Série Restore Estrore										
14 Pre_Com 15 Sdr Dre 16 Trowidz Jin 17 Sorico 18 Biskistation					21	_				
15 Ser 2 ne 15 Ser 2 ne 15 Ser 2 ne 17 Ser 2 ne 17 Ser 2 ne 18 Set 5 set 5 ne 18 Set 5 set 5 ne 19 Set 5 set 5 ne	13 Core	ult_Dec								
IS provide prov	14 Pret	Corn					\$			
17 Bonie 18 Batiston Resole Batistication										
Renove										
	18 Sats	action								
									4	
		9	anove							
	tus									
									Log A	ø

Figure.3. WEKA Explorer with attributes (Dataset)

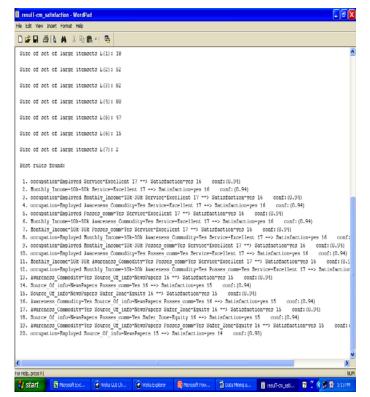


Figure. 5.Associator Model with Minsup=0.20 and Class Index=17

0 # 8 & 8 A 3 & 8 A 4		
Knowledge Inv		
Service		
Satisfaction		
Associator model (full training :	set)	
Apriori		
Minimum support: 0.35 (17 instances)		
Minimum metric <confidence>: 0.9</confidence>		
Number of cycles performed: 113		
Generated sets of large itemsets:		
Size of set of large itemsets $L\left(1\right)$:	7	
Size of set of large itemsets $L\left(2\right)$: (20	
Size of set of large itemsets L(3): 3	30	
Size of set of large itemsets L(4): 2	25	
Size of set of large itemsets L(5): 1	11	
Size of set of large itemsets L(6): 2	2	
Best rules found:		
	Stocks 22 ==> Pref_Comm=High_Return 22 conf:(1)	
	ns-Stocks 22> Pref_Comm-High_Return 22 conf:(1)	
	mc=10k-30k Inv Options=Stocks 22 ==> Pref Comm=High Return 22 conf:(1) Stocks Awareness Commodity=Yes 22 ==> Pref Comm=High Return 22 conf:(1)	
	Stocks Posses commerces 22> Pref Commerligh Return 22 conf: (1)	
	ns-Stocks Awareness_Commodity-Yes 22> Pref_Comm-High_Return 22 conf:(1)	
	ns=Stocks Posses_comm=Yes 22 ==> Pref_Comm=High_Return 22 conf:(1)	
	me=10k-30k Inv_Options=Stocks Awareness_Commodity=Yes 22 ==> Pref_Comm=High_Return	
	me=10k-30k Inv_Options=Stocks Posses_comm=Yes 22 ==> Pref_Comm=High_Return 22 Stocks Awareness Commodity=Yes Posses comm=Yes 22 ==> Pref_Comm=High_Return 22	CO C
<pre>constant contraction - copyroyed inv_operons</pre>	several availables computer to reace commence as> rich committee as	
ior Help, press F1		
🛃 Start 🛛 🛱 Murusuli Exu 🛃 Z. Jorod Th	1) + 👩 Marasalt Pow 🗿 2 Marasalt + 🗐 result -an jedi 🗐 on result 🛛 🖗 🕻 🕼	ñ

He Edit View Insert Format Help	
122日 春日 美術園の 勢	
Size of set of large itemsets L(1): 18	^
Size of set of large itemsets $L\left(2\right):$ 52	
Size of set of large itemsets L(3): 82	
Size of set of large itemsets L(4): 80	
Size of set of large itemsets L(S): 47	
Size of set of large itemsets L(6): 15	
Size of set of large itemsets L(7): 2	
Best rules found:	
 occupation=Employed Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) Monthly_Income=10b-30k Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) occupation=Employed Monthly Income=10b-30k Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) occupation=Employed MonthlyIncome=10b-30k Noses_comm=Pes Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) occupation=Employed MonthlyIncome=10b-30k Noses_comm=Pes Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) occupation=Employed MonthlyIncome=10b-30k Noses_comm=Pes Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) occupation=Employed MonthlyIncome=10b-30k Noses_comm=Pes Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) occupation=Employed MonthlyIncome=10b-30k Noses_comm=Pes Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) occupation=Employed Maratenes_Commodit=Pes Poses_comm=Pes Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) Awareness_Comodit=Noses_comm=Fes Service=Excellent 17 =>> Satisfaction=Pes 16 conf:(0.94) Marateness_Comodit=Nose Poses_comm=Fes Satisfaction=Pes 15 conf:(0.94) Source_of_ind=MiessPapers Poses_comm=Fes Satisfaction=Pes 15 conf:(0.94) Marateness_Comodit=Per Source_of_ind=MiessPapers Poses_comm=Fes Satisfaction=Pes 15 conf:(0.94) Marateness_Comodit=Per Source_of_ind=MiessPapers Poses_comm=Fes Satisfaction=Pes 15 conf:(0.94) Marateness_Comodit=Per Source_of_ind=MiessPapers Poses_comm=Fes Satisfaction=Pes 15 conf:(0.94) Marateness_Comodit=Per Source_of_ind=M	0.94) (0.94) nf:(0.5 faction
	•
	>
iarkho, persifi 17 start Bhorowhite Griede Gullon Griede Equiver Bhorowhite	NUM

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, Itd 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