Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Use of Data Mining in Make in India: In the Field of Tourism and Hospitality

Peeyush Vyas[1]
[1]Asst. Professor – CE/IT department of Vadodara Institute of Engineering, Kotambi, Vadodara.

Abstract: ***To encourage national as well as multi-national companies, Make in India is an imagination thrown by the Government of India in the year 2014. The basic objective of the Government of India is to call business units from all over the world to invest in Indian Manufacturing industry. The field of Tourism and Hospitality is the third largest foreign exchange earner for the country and has accounted for 6.88% of the GDP during 2012-13. Also, with a share of 1.58% of the world's tourism receipt, India is the 15th in the world as far as International Tourism Receipt is concerned. In this paper there is a small effort to predict benefits of reasons to invest in the field of tourism and hospitality with the help of different data mining techniques. Useful functionalities of data mining like Classification, Association Rule Mining, Apriori algorithm etc. have been used to understand the different aspects of statistics, employment and economic growth, issues related to sector policies, FDI policies and financial support. Data mining tool Weka has been used to predict the results.***
Keywords: Make-In-India, Tourism, Hospitality, Association Rule Mining, Classification, Apriori, Weka

I. INTRODUCTION.

In today's world, India is the fastest growing large economy and has held a great position as it overtook China in the year 2015. With the 25 bio-geographic zones, 32 world Heritage sites and various attractive beaches, India certainly offers a geographical diversity. So, it offers a diverse portfolio of niche products like eco-tourism, rural tourism, religious tourism, wellness sports, cruises and many more. The main objective of The Government of India is to invite business entities from all over the world to invest in Indian Tourism and Hospitality sector and for this Government of India is making efforts to generate some rules and regulations to invite investment from foreign investors. Directly or indirectly, domestic companies with some innovative ideas in the field of tourism and hospitality can be given necessary guidance, help and facilities to grow their business at domestic as well as international level.

II. WHY TO INVEST

A. The following points justify investing in this field-

1) As far International Tourism Receipts are concerned, India's position is the 15th in the world with a share of about 1.58% of world's tourism receipts.

2) With the 32 world Heritage Sites, 25 bio-geographic zones and attractive beaches, India offers a very good geographical diversity.

3) India has a diverse portfolio of niche tourism products like cruises, adventure, medical, wellness, sports, MICE, eco-tourism, film, rural and religious tourism.

4) Tourism and hospitality is the third largest foreign exchange earner of the country with 6.88% of the GDP.

5) Tourism in India accounted for 6.88% of the GDP during 2012-13, and tourism the third largest foreign exchange earner for the country.

6) In 2014, the total Fees from tourism were USD 20.236 Billion.

7) In the harmonized list for grant of infrastructure status with the carrying investment of Indian Rs. 200 Crore.

III. FINANCIAL SUPPORT

A. Key Provisions of Budget

1) Tours conducted by Indian tour operators outside India for foreigners are exempted from service tax.

2) For developing SwadeshDarshan and INR 100 crore for PRASAD for Beautification of Pilgrimage Centers, a provision of INR 600 crore is made.

3) Visa on arrivals it to be increased from current 43 countries to 150 countries etc.

B. Tax Incentives

1) For establishing new hotels in the 2-star category and above across India, an investment-linked deduction is proposed under Section 35 AD of the Income Tax Act.

C. State Incentives

1) It includes land at subsidized cost, reduction in stamp duty, exemption on sale/lease of land, incentives on power tariff, loans on the concessional rates, special incentive packages for mega projects.

2) For setting up projects in special areas like the North-East, Jammu & Kashmir, Himachal Pradesh and Utterakhand, special incentives are provided.

D. FDI Policy

1) Under the automatic route in tourism and hospitality, 100% Foreign Direct Investment (FDI) is allowed subject to applicable rules and regulations.

2) 100% Foreign Direct Investment allowed in tourism construction projects, including development of hotels, resorts and recreational facilities.

E. Sector Policy

1) The vision of National Tourism Policy, 2002 is to enhance employment potential within the tourism sector and to foster economic integration through developing linkage with other sectors.

F. Other Important Policies

1) Guidelines for approval of convention centers, motel projects, guesthouses etc.
2) Guidelines for assistant to central agencies in tourism infrastructure development.
3) Scheme for large revenue generating projects.
4) Scheme for PPP(Public-Private-Partnership) in infrastructure development
5) Guidelines for approval of convention centers etc.

G. Foreign Investors

1) Accor (France)
2) The Four Seasons Group (Canada)
3) Starwood Hotels (USA)
4) Thomas Cook (UK)
5) Marriott Hotels (USA
6) Expedia (USA)
7) Premier Travel Inn (UK)
8) Cox & Kings (UK)
9) Mandarin Oriental (Hong Kong)
10) Jumeirah (UAE)

H. Agencies

1) Hotel Association of India
2) Association of Tourism Trade Organizations, India
3) Federation of Hotel & Restaurants Associations of India
4) Indian Association of Tour Operators
5) Travel Agents Association of India

IV. DATA MINING

Data mining, or knowledge discovery, is the computer-based process of analyzing huge sets of data and then pull out the meaningful information of the data. To make business proactive and to get knowledge-driven decisions, data mining tools are used. Data

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

mining refers to the application of algorithms for extracting patterns from data without the additional steps of the KDD process. KDD (Knowledge Discovery in Databases) refers to the overall process of discovering useful knowledge from data. It includes the choice of encoding schemes, preprocessing, sampling, and projections of the data prior to the data mining step. It involves the following steps

A. To learn the application domain including prior knowledge and goals of the end-users.
B. To select data set and create appropriate target data set on which discovery is to be performed.
C. To process and clean data i.e. to remove noise or outliers and handling missing data fields.
D. To reduce and transform data using various methods like dimensionality reduction or transformation methods.
E. To seek different functions of data mining like classification, clustering, regression, summarization etc.
F. To search and evaluate patterns while removing redundant patterns and to represent knowledge in the form of patterns and charts.
G. To make the optimum utilization of the discovered knowledge.

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems.

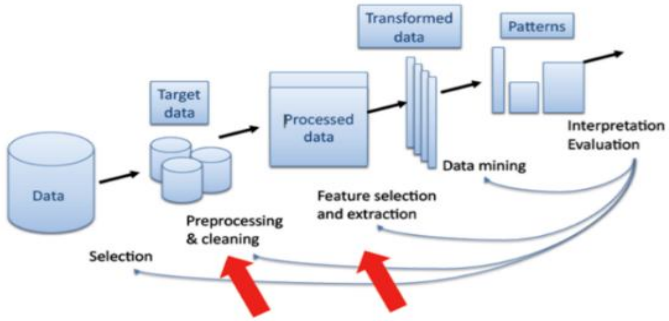

Fig.1

V. DATA MINING IN TOURISM AND HOSPITALITY

Here, we will discuss two types of machine learning activities which are common in field of tourism and hospitality and these are association rule mining and classification. In association rule mining the learning methods look for facilities provided by the Tourist Company and associations or relationship between the features of tourist behavior. For example, the algorithm may try to find out if companies who are giving services in India are also arranging the international tours or foreign tourists who are interested in getting the experience of village life of India can prefer to stay in such a place where in the hotel or in the resort itself there is a artificial villages created and give the feel of actual village life. Hence, there is no target attribute or variable in this type of data mining and it is called as unsupervised learning. Similarly, another machine learning is classification which is a type of supervised learning which contains a specific target variable. For example, with the help of classification we may divide the investors companies in two groups – high investment or low investment. In this case target variable will be investment on setting the related company in India. Similarly, we can have companies of foreign investors or Indian investors, companies having setup in India or setup in foreign etc. Classification algorithm will establish specific attributes of a company that qualify it as a high investor or low investor company, company with international setup or Indian setup etc.

VI. ASSOCIATION RULE MINING AND MARKET BASKET ANALYSIS

Market Basket Analysis can be effectively used in a company of Tourism and Hospitality. The analysis of data depends upon, which type tour packages are purchased by a consumer. Some association rules can be generated showing that which tour packages are purchased together. On these available facts, companies can decide some sort of available association between different tour packages that are sold for various purposes. The sample data(assumed) by various companies offering different tour packages in India as well as abroad can be shown as below –

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

TABLE-I
SAMPLE DATA

Sno.	Company	Tour Packages Offered	Investment capacity to setup branch in India Rs. In Lakhs	Facilities Provided
1	Accor	International, South, Goa, Himachal	40	Budget, 3 Star
2	Thomas Cook	Rajasthan, Goa, Kerla	60	3 star
3	Marriot Hotels	International, South, Himachal, Rajasthan	80	5 star
4	Expedia	Goa, Kerla, Rajasthan	25	Budget, 3star, 5 star
5	Cox & Kings	International, South, Kerla	35	3star, 5 star
6	Jumeirah	International, Rajasthan,Himachal	65	Budget, 3 star, 5 star

```
=== Run information ===

Scheme:       weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     Hotel1-CSV
Instances:    10
Attributes:   2
              Transaction
              Companies
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20

Size of set of large itemsets L(2): 10

Best rules found:

 1. Companies=Accor ,Thomas Cook, Marriott Hotels 1 ==> Transaction=T1 1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
 2. Transaction=T1 1 ==> Companies=Accor ,Thomas Cook, Marriott Hotels 1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
 3. Companies=Expedia, Thomas Cook, Marriott Hotels  1 ==> Transaction=T2 1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
 4. Transaction=T2 1 ==> Companies=Expedia, Thomas Cook, Marriott Hotels  1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
 5. Companies=Accor, MARRIOTT HOTELS  1 ==> Transaction=T3 1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
 6. Transaction=T3 1 ==> Companies=Accor, MARRIOTT HOTELS  1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
 7. Companies=Accor, Thomas Cook, Marriott Hotels, Cox & Kings , Jumeirah 1 ==> Transaction=T4 1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
 8. Transaction=T4 1 ==> Companies=Accor, Thomas Cook, Marriott Hotels, Cox & Kings , Jumeirah 1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
 9. Companies=Accor ,Thomas Cook, Marriott Hotels, Cox & Kings  1 ==> Transaction=T5 1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
10. Transaction=T5 1 ==> Companies=Accor ,Thomas Cook, Marriott Hotels, Cox & Kings  1    <conf:(1)> lift:(10) lev:(0.09) [0] conv:(0.9)
```

Fig.2

Now, to select a company out of the above let us take an example of the different transactions by various customers according to their choices, facilities and packages –

TABLE II
TRANSACTION SAMPLES BY VARIOUS CUSTOMERS

Sno.	Transaction	Companies
1	T1	Accor ,Thomas Cook, Marriott Hotels
2	T2	Expedia, Thomas Cook, Marriott Hotels
3	T3	Accor, MARRIOTT HOTELS
4	T4	Accor, Thomas Cook, Marriott Hotels, Cox & Kings , Jumeirah
5	T5	Accor ,Thomas Cook, Marriott Hotels, Cox & Kings
6	T6	Expedia, Marriott Hotels
7	T7	Accor, Marriott Hotels, Cox & Kings
8	T8	Expedia, Thomas Cook, Marriott Hotels, Cox & Kings
9	T9	Accor, Cox & Kings , Jumeirah
10	T10	Accor, Marriott Hotels, Cox & Kings , Jumeirah

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

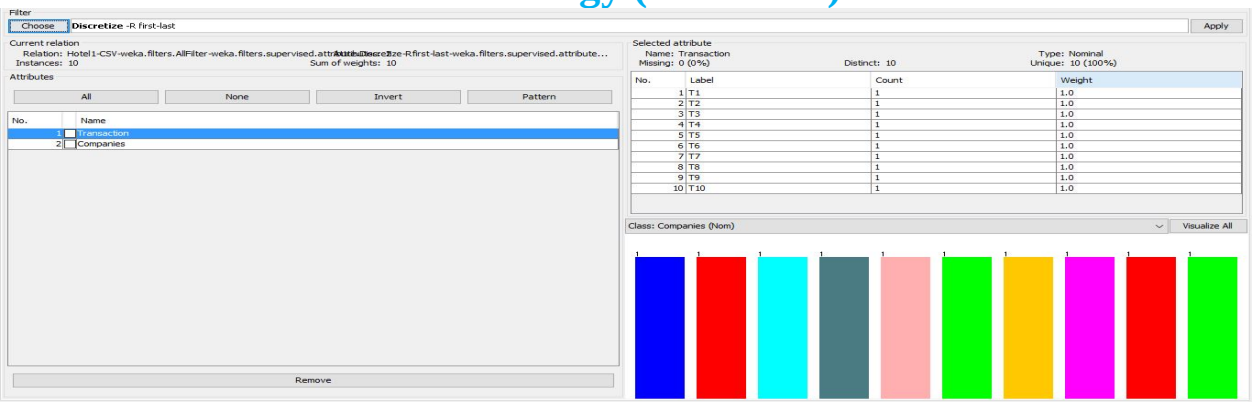

Fig.3

Companies can target their customers while providing a package containing more facilities related to tour. Local and foreign customers can be targeted for various Indian as well as international tour packages according to their budgets and interests. According to the interest of the customers it will become easy for Indian Government to select a company which can be allowed to set up its business in India.

An association rule can be viewed as a set of tuples and each tuple containing data items. In the above example there are ten examples and six data items: (Accor, Thomas Cook, Marriot Hotels, Expedia, Cox & Kings and Jumeirah} as {S1,S2,S3,S4,S5,S6}. Here, the main problem is to find out all the instances where customers who bought a subset of a frequent item set, most of the time also bought remaining data items in the same frequent set. For a given frequent item set, say {S1,S2,S3}, if a customers who buys a subset formed by S1 and S2 also buys S3, 75% of times then there is some sense to apply the rule. This percentage is called confidence of the rule. Confidence of rule "B given A" is a measure of how much more likely it is that B occurs when A has occurred. It can be defined as the ratio of the number of transactions that include all data items in a particular frequent item set to the number of transactions that include all items in the subset.

$$\text{Confidence } I = \frac{P(X \cap Y)}{P(X)}$$

Range [0, 1]; If I=1 then most interesting and if I=0 then least interesting.

On the other hand, a support measures how often the collection of items in an association occurs together as a percentage of all the transactions.

$$\text{Support } I = \frac{(X \cap Y)}{N}$$

Range [0, 1]; If I=1 then most interesting and if I=0 then least interesting.

To clear the above rules, let us consider the following examples where we want to find out the required association rule.

Support = 30% means only the items that ate bought together by at least 3 customers are considered.

Confidence = 90% means in 90% of transactions, the association rule is to be true.

Case 1: (S2, S4) → S3

(S2, S4) were bought by 5 customers but only 3 of them also bought S3, the confidence is 60%.

Case 2: (S5, S6) → S2

(S5, S6) were bought by 3 customers and all 3 of them bought S2 as well.

So, the confidence is 100% and hence this rule has strong confidence and it is to be considered as it is more than 90%.

A. Apriori Algorithm Pseudo code

Join Step: C_k is generated by joining $L_{k-1}$with itself

Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-i

itemset

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

B. Pseudo-code

C_k: Candidate itemset of size k
L_k : frequent itemset of size k
L_1 = {frequent items};
for (k = 1; L_k !=∅; k++) do begin
C_{k+1} = candidates generated from L_k;
for each transaction t in database do
increment the count of all candidates in C_{k+1} that are contained in t
L_{k+1} = candidates in C_{k+1} with min_support
End
Return $U_k L_k$

VII. CLASSIFICATION

The database can be further subdivided into more similar groups just to improve predictive accuracy. The segmentation can be done using different attributes like Budget, Meal Plan, Sight Scene, Offers, Facilities and buying behavior of the customer. It helps in the further classification maps data into predefined groups. On these attributes, classes are defined and used in the classification algorithms. Hence, Indian Government can more accurately predict the company to install its business in India and hence a suitable tour package can be predicted for an appropriate customer depending upon Budget, Meal Plan, Sight Scene, Offers, Facilities and buying behavior of customer. Using these attributes, Indian government can decide a segment and the patterns can be stored in the databases.

TABLE III
FACILITIES OFFERED BY COMPANIES

Company Name	Meal Plan Offered	Sight Scene	Offers	Special Facilities	Buying Behavior
Accor	EP	Yes	10%	Online Booking, Foreign exchange	Yes
Thomas Cook	CP	No	20%	Travel Insurance	No
Marriot Hotels	MAP	Yes	0%	Leisure travel and Corporate travel facilities	No
Expedia	AP	Yes	15%	Foreign exchanges	Yes
Cox & KIngs	AP	No	10%	Trade fairs, , insurance	No
Jumeirah	MAP	Yes	25%	Business travel	Yes

In this example, suppose we want to classify the companies on the attributes like Meal Plan, Sight Scene, Offers, and facilities. The different tour packages classification can be done using these attribute as follows -
If Meal Plan = ‘EP’ .AND. Sight Scene =’Yes’ .AND. Offers>0% Then Buying Behavior =’Yes’
If Meal Plan = ‘CP’ .AND. Sight Scene =’No’ .AND. Offers>0% Then Buying Behavior =’No’
If Meal Plan = ‘MAP’.AND. Sight Scene =’Yes’ .AND. Offers=0% Then Buying Behavior =’Yes’
If Meal Plan = ‘AP’ .AND. Sight Scene =’Yes’ .AND. Offers>0% Then Buying Behavior =’Yes’
If Meal Plan = ‘AP’ .AND. Sight Scene =’No’ .AND. Offers>0% Then Buying Behavior =’No’
If Meal Plan = ‘MAP’ .AND. Sight Scene =’Yes’ .AND. Offers>0% Then Buying Behavior =’Yes’

A. Definition

Given a database D = {t1, t2, t3, ……t_n} of n tuples and a set of classes C = {C1, C2, C3,…..C_m} then the classification problem is to define a mapping f : D → C where each ti is assigned to one class. A class Cj contains precisely those tuples mapped to it that is Cj = { ti | f(ti) = Cj, 1≤ i ≤ n, and ti € D}

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

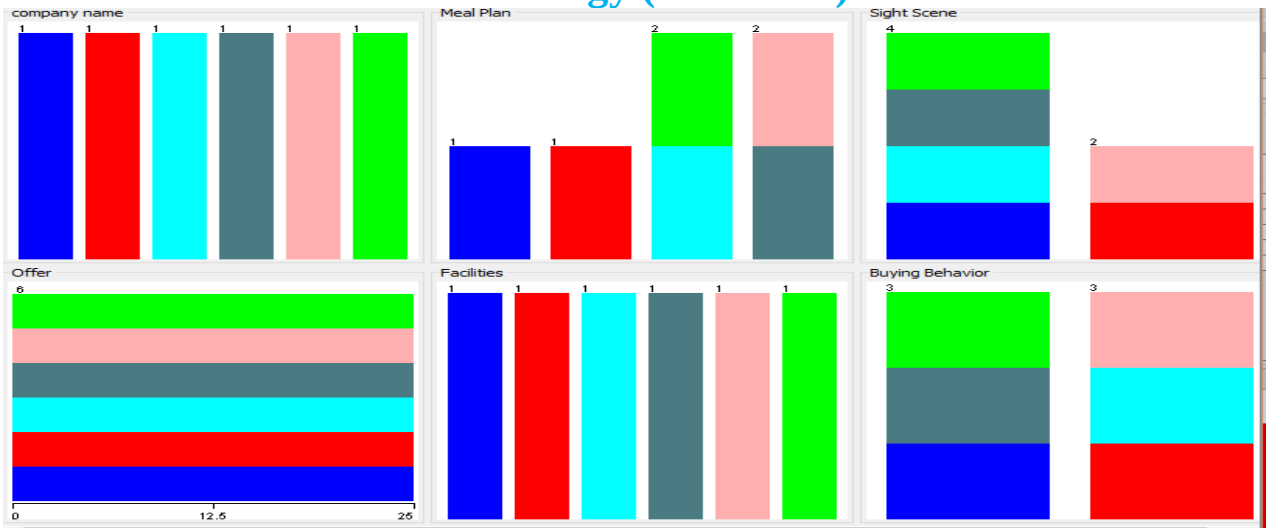

Fig.4

```
Number of Rules : 2
Non matches covered by Majority class.
        Best first.
        Start set: no attributes
        Search direction: forward
        Stale search after 5 node expansions
        Total number of subsets evaluated: 22
        Merit of best subset found:    50
Evaluation (for feature selection): CV (leave one out)
Feature set: 3,6

Time taken to build model: 0.02 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances           5               83.3333 %
Incorrectly Classified Instances         1               16.6667 %
Kappa statistic                          0.6667
Mean absolute error                      0.4111
Root mean squared error                  0.4208
Relative absolute error                 82.2222 %
Root relative squared error             84.1515 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances                6

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                   1         0.333       0.75      1         0.857      0.833    Yes
                   0.667     0           1         0.667     0.8        0.833    No
Weighted Avg.      0.833     0.167       0.875     0.833     0.829      0.833

=== Confusion Matrix ===

 a b   <-- classified as
 3 0 | a = Yes
 1 2 | b = No
```

Fig.5

VIII. CONCLUSION

Data Mining can be used in decision making in a tour & hospitality company just to choose a company to install and invest its business in India and to raise the business of the company. Here, data mining techniques as discussed in this paper are required to understand by the tour and hospitality company. Association rule mining and classification techniques are certainly useful to decide company to be fit in the concept of Make in India and also by the customers to decide a company of their choice. Association rule is very much helpful in this regard as it uses two mathematical measures ‘Confidence’ and ‘Support’ whereas ‘Classification’ can be used to target new companies and customers.

REFERENCES

[1] A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. Decision Support Systems , 42 (2006), 1503-1520.
[2] J.Blanchard,F.Guillet,and H.Bri and.Exploratory visualization for association rulerummaging. In KDD 03WorkshoponMultimediaDatAMining(MDM-03),2003
[3] Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification Tina R. Patil, Mrs. S. S. SherekarSantGadgebaba Amravati

Volume 5 Issue V, May 2017
ISSN: 2321-9653

University, Amravati.
[4] Data Mining in Tourism Indranil Bose The University of Hong Kong, Hong Kong
[5] L. Cavique.A scalable algorithm for the market basket analysis. Journal of Retaiing and Consumer Services,14(6):400407,2007
[6] Mr. A. B. Devale and Dr. R. V. Kulkarni "A REVIEW OF DATA MINING TECHNIQUES IN INSURANCE SECTOR" Golden Research Thoughts Vol- I , ISSUE - VII [January 2012]
[7] Martin Staudt, Anca Vaduva and Thomas c, "Metadata Management and Data Warehouse", Technica Report, Information System Research, Swiss Life, University of Zurich, Department of Computer Science , July 1999, vaduva@ifi.unizh.ch
[8] Y.Cho,J.Kim,andS.Kim.A personalized recommender system based on web usage mining and decision tree induction. Expert Systems with Applications,23(3):329342,2002
[9] S.Brin,R.Motwani,and C.Silverstein.Beyondmar ketbaskets:general-Izing association rules to correlations. Proceedings of the ACMSIGMOD, pages265276,1997
[10] Jiawei Han, Laks V. S. Lakshmanan and Raymond T.NG, "Constraint-Based Multidimensional Data Mining", IEEE, August 1999. Chen, Y.-L., Chen, J.-M., & Tung, C.-W. (2006)
[11] Mr. Peeyush Vyas "Use of Data Mining in Population Survey' VIER Journal of Engineering Research-Vadodara, Vol. I [January 2014] ISSN 2349-9079