Quantitative Analysis of Dairy Product Packaging with the Application of Data Mining Techniques

Ankita Chopra[*],.Yukti Ahuja[#], Mahima Gupta[#]

[*] Assistant Professor, JIMS-IT Department, IP University
3, Institutional Area, Rohini sec-5, New Delhi-110085, India

[#]Assistant Professor, JIMS-Management Department, IP University
3, Institutional Area, Rohini sec-5, New-Delhi-110085, India

Abstract— **Consumers make choices about where to shop based on their preferences for a shopping environment and experience as well as the selection of products at a particular store. This study illustrates how dairy firms and marketing analysts can utilize data mining techniques to better understand customer profiles and behaviour.**
A survey is done on blind dairy product packaging and its competitors. Primary data is collected based on the package design rating of dairy products. The customers are evaluated on the attributes like gender, education, occupation and a quantitative analysis is done with the application of data mining technique like classification on the data set.
The paper presents how data mining discovers and extracts useful patterns from this large data to find observable patterns. The paper demonstrates activity of data mining in improving the quality of decision making process in dairy industry.

Keywords— **Data Mining, Classification, Data mining Application, Knowledge discovery database.**

I. INTRODUCTION

We are in an age often referred to as the information age. In this information age, because we believe that information leads to power and success, and thanks to sophisticated technologies such as computers, satellites, etc. We have been collecting tremendous amount of information.

With the enormous amount of data stored in files, databases, and other repositories, it is increasingly important, if not necessary, to develop powerful means for analysis and perhaps interpretation of such data for the extraction of interesting knowledge that could help in decision making.

Data Mining, also popularly known as knowledge discovery in databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. While data mining and knowledge discovery (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. The following (figure 1.1) shows data mining as a step in an iterative knowledge discovery process.

Figure 1.1: Data Mining is the core of Knowledge Discovery process
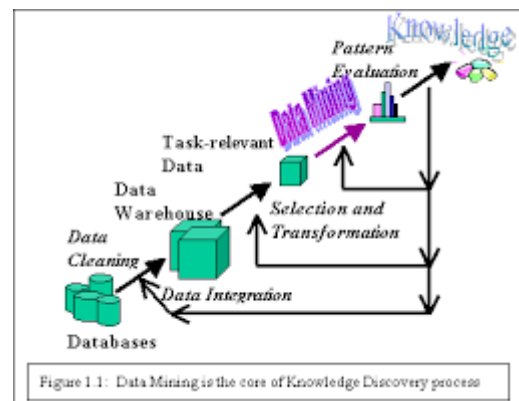
The knowledge discovery in databases process comprises of a few steps leading from raw data collections to some form of new knowledge. The iterative process consists of following steps.

- **Data Cleaning:** Also known as data cleaning, it is a phase in which noise data and irrelevant data are removed from the collection.
- **Data Integration:** At this stage, multiple data sources, often heterogeneous, may be combined in a common source.
- **Data Selection:** At this step, the data relevant to the analysis is decided on and retrieved from the data collection.
- **Data Transformation:** Also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
- **Data Mining:** It is the crucial step in which clever techniques are applied to extract patterns potentially useful.
- **Pattern Evaluation:** In this step, strictly interesting patterns representing knowledge are identified based on given measures.
- **Knowledge Representation:** Is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand & interpret the data mining results.

II RESEARCH METHODOLOGY

For the purpose of the study we designed a questionnaire after rigorous brain storming method. Post this, a team of researchers was deployed to collect primary data from 375 respondents. The data was collected by face to face interviews. Showcards for image packages of the test brand as well as competing brands were created (Show cards are the set of colored print outs) and provided to each interviewer. The team of interviewers were briefed and trained to conduct interviews.

Purposive random sampling technique was used. The areas of data collection were shortlisted to be Delhi and NCR.

After data collection, the data was fed into MS Excel and appropriate Data Mining tools were applied to find meaning results.

The following sections of the paper discuss the Data Mining techniques, application of data mining in dairy industry,

III DATA MINING TECHNIQUES

Various algorithms and techniques like classification, clustering, regression, artificial intelligence, neural networks, association rules, decision trees, genetic algorithm, nearest neighbour method etc. are used for knowledge discovery from databases.

I Association

Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction.

For example, the association technique is used in market basket analysis to identify what products their customers frequently purchase together. Based on this data businesses can have corresponding market campaign to sell more products to make more profit.

II Classification

Classification is a data mining technique based on machine learning basically classification is used to classify each item in a set of data into one predefined set of classes or groups.
Classification Techniques
- Regression
- Decision Trees
- Neural Networks

III Clustering

Clustering is "the process of organizing objects into groups whose members are similar in some way"
- Hierarchical Methods
- Partitioning Methods
- Model based clustering methods

IV Prediction

The predication as it name implies is one of a data mining technique that discovers relationship between independent variables and relationship between dependent and independent variables.

IV APPLICATION OF DATA MINING TECHNIQUES IN DAIRY INDUSTRY:

Survey on dairy products packaging has been done and the transactional data that has been collected is listed in the below table.

Customer is evaluated on the basis of following attributes like age, education, occupation and package design rating for the products is done based on the class-labelled attribute Rating.

Table I: Sample dataset used for data mining

S.NO	Gender	Occupation	Education	Package Design Brand-Ranking Dairy product 1(class labelled attribute)
1	2	Govt. employee	Graduate	2
2	2	Business	MBA	2
3	2	Army	Graduate	2
4	2	IGI Airport	Retired	2
5	2	Business	MBA	2
6	1	Senior Mgr.	Post Grad	3
7	2	Business	Graduate	2
8	1	Govt. Job	Graduate	1
9	2	Business	9	3
10	2	CISF	Graduate	2
11	2	Govt. Job	Graduate	1
12	2	Bank Mgr.	MA	3
13	2	Teacher	MSC	3
14	1	Business	12	3
15	1	Business	BA	1
16	1	Job	10	3
17	1	Job	Graduate	2
18	1		M Com	2
19	1	Business	8	3
20	1	Shop Owner	Graduate	2
21	1	Business	Graduate	1
22	1	Business	Graduate	2
23	1	Business	10	3
24	1	Welding	8	1
25	1	Job	12	2
26	1	Auto Driver	12	3
27	1	Job	Graduate	1
28	1	transport employee	10	2
29	1	Auto Driver	Illiterate	3
30	1	Metro Employee	Illiterate	3
31	1	Chef	12	2
32	1	Business	12	2
33	1	Railway Eng.	12	2
34	1	Job	10	2
35	1	Business	BA	2
36	1	Police	BA	1
37	1	Govt. Job	12	2
38	1	Business	Illiterate	2
39	1	Job	M.Com	2

Table II: Sample dataset used for data mining

S.NO	Gender	Occupation	Education	Package Design Ranking-Dairy Product II(class labelled attribute)
1	2	Govt. employee	Graduate	2
2	2	Business	MBA	3
3	2	Army	Graduate	2
4	2	IGI Airport	Retired	2
5	2	Business	MBA	3
6	1	Senior Mgr.	Post Grad	3
7	2	Business	Graduate	3
8	1	Govt. Job	Graduate	2
9	2	Business	9	2
10	2	CISF	Graduate	2
11	2	Govt. Job	Graduate	3
12	2	Bank Mgr.	MA	2
13	2	Teacher	MSC	2
14	1	Business	12	3
15	1	Business	BA	3
16	1	Job	10	2
17	1	Job	Graduate	1
18	1		M Com	2
19	1	Business	8	2
20	1	shop Owner	Graduate	3
21	1	Business	Graduate	4
22	1	Business	Graduate	2
23	1	Business	10	2
24	1	Welding	8	2
25	1	Job	12	3
26	1	Auto Driver	12	2
27	1	Job	Graduate	2
28	1	transport employee	10	1
29	1	Auto Driver	Illiterate	2
30	1	Metro Employee	Illiterate	2
31	1	Chef	12	1
32	1	Business	12	2
33	1	Railway Engg.	12	3
34	1	Job	10	2
35	1	Business	BA	2
36	1	Police	BA	3
37	1	Govt. Job	12	2
38	1	Business	Illiterate	2
39	1	Job	M.Com	3

Table III: Sample dataset for data mining

S.NO	Gender	Occupation	Education	Package Design Ranking-Dairy Product III(class labelled attribute)
1	2	Govt. employee	Graduate	1
2	2	Business	MBA	3
3	2	Army	Graduate	2
4	2	IGI Airport	Retired	3
5	2	Business	MBA	3
6	1	Senior Mgr.	Post Grad	3
7	2	Business	Graduate	2
8	1	Govt. Job	Graduate	1
9	2	Business	9	2
10	2	CISF	Graduate	2
11	2	Govt. Job	Graduate	2
12	2	Bank Mgr.	MA	1
13	2	Teacher	MSC	1
14	1	Business	12	2
15	1	Business	BA	2
16	1	Job	10	3
17	1	Job	Graduate	2
18	1		M Com	3
19	1	Business	8	3
20	1	Shop Owner	Graduate	3
21	1	Business	Graduate	5
22	1	Business	Graduate	1
23	1	Business	10	2
24	1	Welding	8	3
25	1	Job	12	3
26	1	Auto Driver	12	3
27	1	Job	Graduate	3
28	1	transport employee	10	3
29	1	Auto Driver	Illiterate	3
30	1	Railway Engg.	12	1

In the above table attributes are given the following values:
Gender-{male (1), female (2)}
Occupation-{unskilled worker(1),skilled worker(2),petty trader(3),shop owner(4),businessman(5),no of employess(6), 10+(7),self-empployed(8),clerk(9),supervisor(10),salaried-juniorlevel(11),salaried-seniorlevel(12)}
Education-{Illiterate(1),literate but no school(2),upto 4yrs(3),5 to 9(4),HSC(5),HSC+(6),Graduate/postgrad-gen(7),grad/postgrad-prof(8)}
Package design ranking as class labelled attribute takes three values -1(poor),2(good),3(excellent){1,2,3}

Classification:

The following hypothesis are done on the above data

H1: Package design rating is dependent on age of customer

H2: Package design rating is dependent on occupation of customer

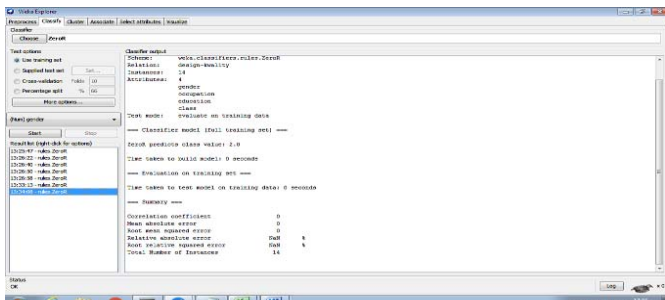H3: Package design rating is dependent on education of customer.

The above transactional data set is converted into .arff format and given as input to the WEKA Tool and classification data mining technique is applied to evaluate the above said hypothesis and derive observable patterns:
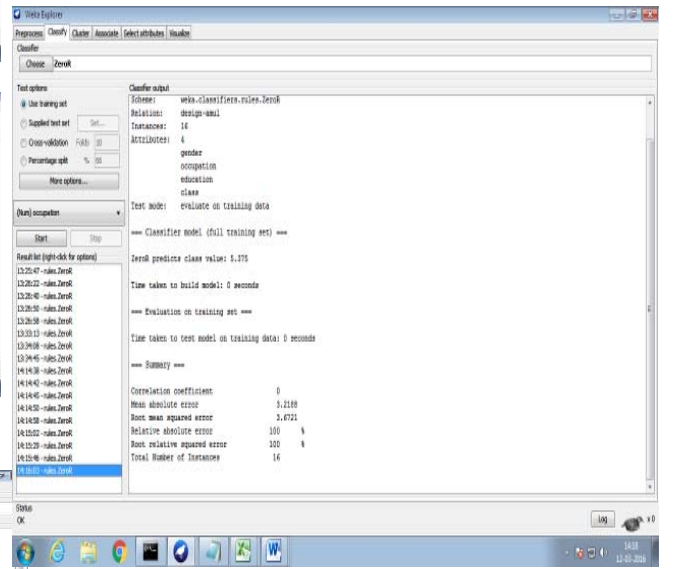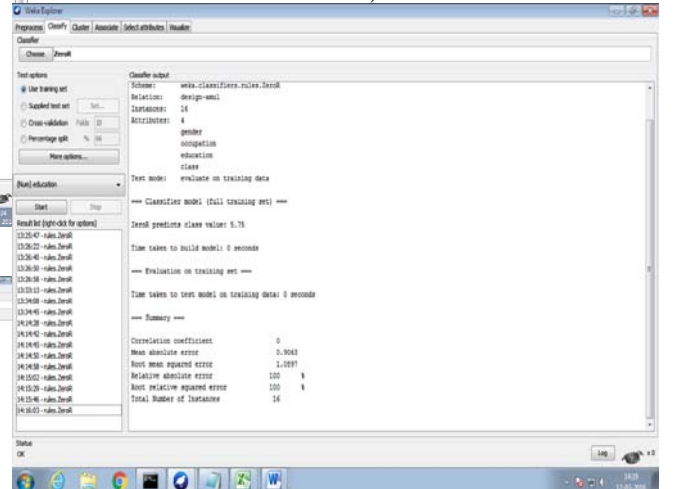
I: **For Dairy product I:**
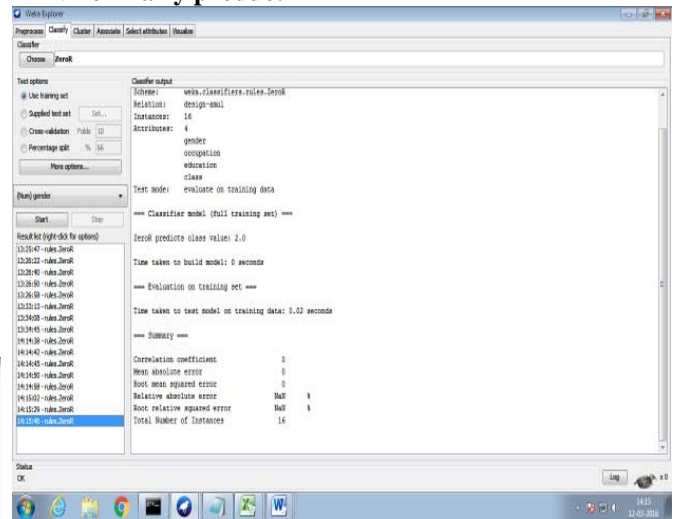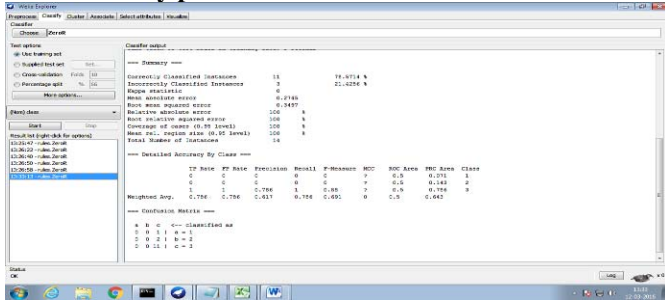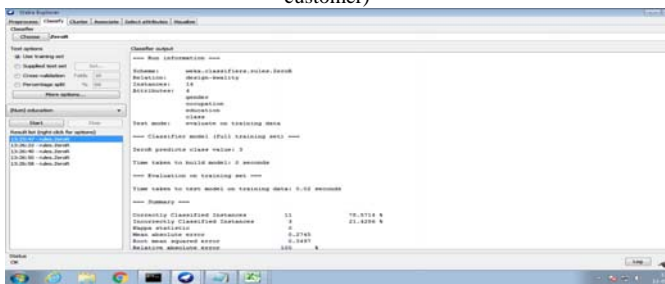

I(i)\H1:fails(package design ranking   is not dependent on age of customer)

I(ii)H2:fails(package designing is not dependent on occupation of customer)

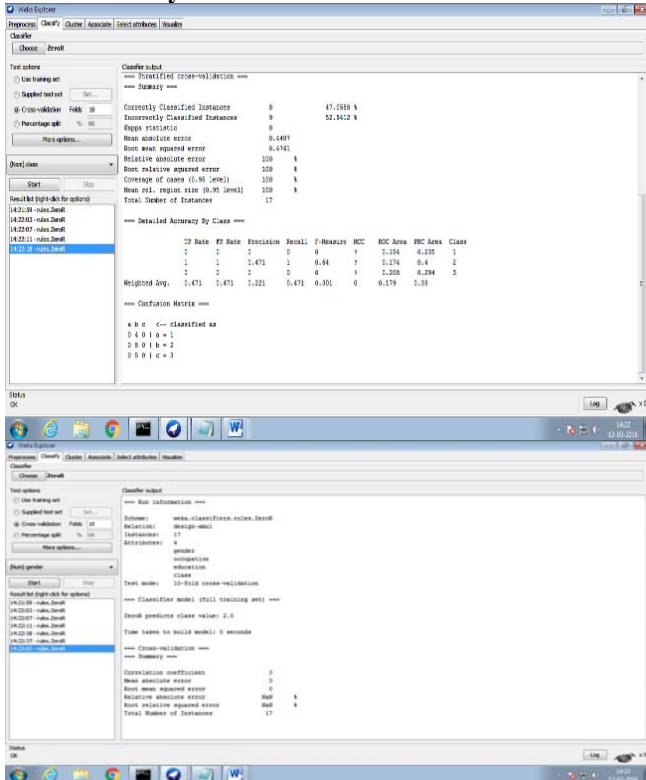I(iii)H3:Success(package design ranking is dependent on education of customer)

II: **For Dairy product II**

II(i)H1:fails(package design ranking is not dependent on gender of customer)

II(ii)H2:fails((package design ranking is not dependent on occupation of customer)

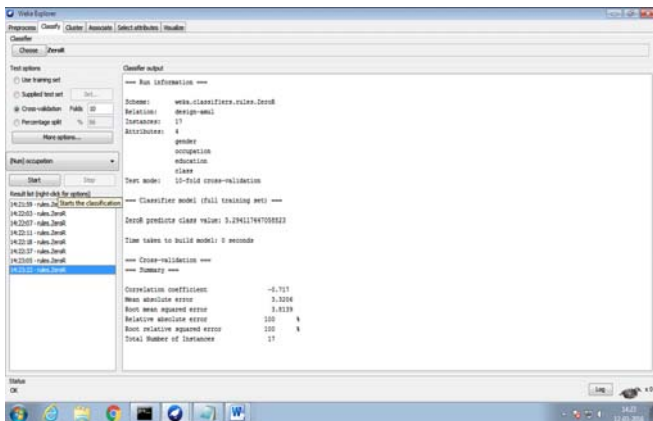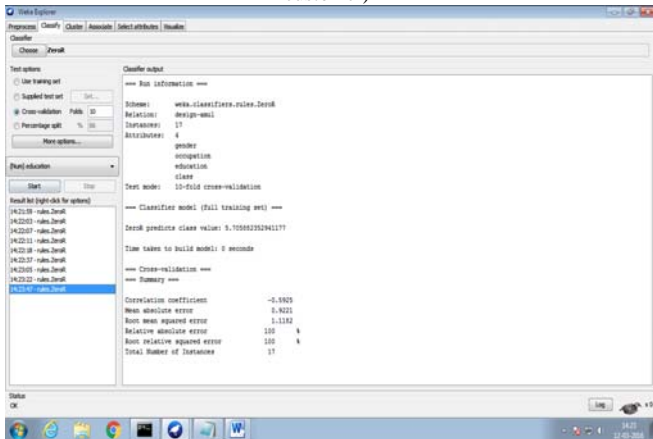II(iii)H3:Success (package design ranking is dependent on education of customer)

III: **For Dairy Product III:**

III(i)H1:fails(package design ranking is not dependent on gender of customer)

III(ii)H2:fails (package design ranking is not dependent on occupation of customer)

III(iii)H3:success(Package design ranking is dependent on education of customer).

CONCLUSIONS & FUTURE WORK

From the above generated patterns we can say that hypothesis-3 is true for all the customers and knowledge derived is if feedback is taken by the customers who are well educated then it helps market analysts to market their product with improved package design and in turn generate more profits for the dairy industry.

Future work would be to study what frequent patterns customers follow to buy a particular dairy product which in turn increases profitability of dairy industry.

REFERENCES

[1] N. Katsaras, P. Wolfson, J. Kinsey and B. Senauer, 'Data Mining A segmentation analysis of U.S. grocery shoppers', *The Retail Food Industry Center University of Minnesota*, March 2001.

[2] K. Kumar, B. K. Chauhan, J.P. Pandey, A. K. Tomer, 'Data mining and knowledge discovery from research problems', *International Journal of Advanced Technology & Engineering Research (IJATER),* 1st International Conference on Research in Science, Engineering & Management (IOCRSEM 2014), Pg. 61-66.

[3] B. M. Ramage, 'Data Mining Techniques and Applications', *Indian Journal of Computer Science and Engineering,* Vol. 1 No. 4 301-305, Pg.301-305.

[4] K. M. Raval, 'Data Mining Techniques', *International Journal of Advanced Research in Computer Science and Software Engineering',* Volume 2, Issue 10, October 2012.

[5] L.Rokach and O. Maimon, 'Clustering Methods', D*ata Mining and Knowledge Discovery Handbook',*Pg. 321-353.

[6] G.S. Linoff, M. J. A. Berry, 'Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 3rd Edition'.

[7] L.Rokach and O. Maimon, 'Data mining for improving the quality of manufacturing: a feature set decomposition approach'.

[8] J.Vasilev, 'Data mining of transactional data for sales of dairy products', *Theoretical and Applied Economics,* Volume XXI (2014), No. 12(601), pp. 3-12.