ISSN 2395-1621

Market Basket Analysis using Data Mining

#1Basod Sagar, #2Raut Sagar, #3Nitesh Singh, #4Mungase Ganesh

1sagarb4093@gmail.com

#1234Designation, Department of Computer Engineering
H.S.B.P.V.T. COE, Kashti.

ABSTRACT

Market basket analysis (MBA) is a powerful and common practice in modern retailing that has some limitations stemming from the fact that it infers purchase sequence from joint purchasing data. However, internet retailers automatically collect purchase sequence data from their shoppers, and new technology is available for traditional (bricks and mortar) retailers to do the same, making it possible to analyze purchase sequences, rather than inferring them from joint purchases. This study first compares and contrasts traditional market basket analysis with a sequential extension, and then proposes a framework for purchase sequence analysis, which is illustrated utilizing shopping trip data from one grocery store. Market Basket Analysis is a process to analyze the habits of buyers to find the relationship between different items in their market basket. The discovery of these relationships can help the merchant to develop a sales strategy by considering the items frequently purchased together by customers. In this research,the data mining with market basket analysis method is implemented, where it can analyze the buying habit of the customers.

Keywords: **Market basket analysis, Data Mining, Business Intelligence**

ARTICLE INFO

Article History

Received:30th September 2015

Received in revised form :

2nd October 2015

Accepted:6th October, 2015

Published online :

9th October 2015

I. INTRODUCTION

A Smart Analytical Tool for Predicting Customer Behaviour Business Intelligence.By employing Basket Analysis, strategic decisions and policies will be formulated by utilizing large amounts of authentic data.In the world of automation advanced technology and ever expanding need for standards in multidimensional economies has led to a large amount of data being made available in the form of raw datasets. However efficient data analysis and potential decisions that can be evaluated from the data are not meeting the corresponding needs. Traditional database from various international surveys is in raw format that is neither useful for general public nor to the governing bodies unless analyzed properly. The paper employs Business Intelligence techniques and tools to analyze transactions between Customer and Retailer. The proposed system shall provide a comprehensive analysis and reflect user behaviour .The aim is to analyse the available data into a smart analytical format that would prove helpful for the retailer to increase sales.

II. LITERATURE REVIEW

We have referred a number of IEEE papers and Journals that have helped us with our initial researches. The papers have made it easier for us to move forward with the development of the system with all the theoretical details we needed for the initial stages. They have helped us understand:

Business Intelligence

Business intelligence is a comprehensive term that encompasses multiple dimensions crucial to the evolution of strategic decisions in various sectors and domains. BI is an umbrella term that includes the applications, infrastructure, tools and efficient practices that enable access and analysis of information and optimize decisions and performance. As stated by Howard Dresner, the father of Business Intelligence, it involves concepts and methods to improve business decision making by using factbased support system.

III. DATA MINING

Overview:
Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Continuous Innovation:
Although data mining is a relatively new term, the technology is not. Companies have used powerful computers to sift through volumes of supermarket scanner data and analyze market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost. Example
For example, one Midwest grocery chain used the data mining capacity of Oracle software to analyze local buying patterns. They discovered that when men bought diapers on Thursdays and Saturdays, they also tended to buy beer. Further analysis showed that these shoppers typically did their weekly grocery shopping on Saturdays. On Thursdays, however, they only bought a few items. The retailer concluded that they purchased the beer to have it available for the upcoming weekend. The grocery chain could use this newly discovered information in various ways to increase revenue. For example, they could move the beer display closer to the diaper display. And, they could make sure beer and diapers were sold at full price on Thursdays. Data, Information, and Knowledge Data mining is primarily used today by companies with a strong consumer focus retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data. The input for the market basket analysis is a dataset of purchases. A market basket is composed of items bought together in a single trip to a store. The most significant attributes are the transaction identification and item identification. While ignoring the quantity bought and the price. Each transaction represents a purchase, which occurred in a specific time and place. This purchase can be linked to an identified customer(usually carrying a card) or to a nonidentified customer The dataset with multiple transactions can be shown in a relational table (transaction,item). Corresponding to each attribute there is a set called domain. The table(transaction, item) is a set of all transactions T={T1, T2, T3, …, Tn} where each transaction contains a subset of items Tk
= {Ia, Ib, Ic …}.

IV. APRIORI ALGORITHM

The first step of the Apriori algorithm generates sets of market baskets. Ik is defined as the set of frequent items with k items bought together. Firstly, the algorithm filters the items with a frequency higher than Threshold, generating I1. In the following stages, for each Ik it generates the Ik+1 candidates, such as IkÍIk+1. For each Ik+1 candidate, the algorithm removes the baskets, which are lower than the Threshold.The cycle ends when it reaches Imax_k.In the second step, the Apriori algorithm generates sets of market baskets and then generates association rules Left_Right. For each rule, the support measure and the confidence measure are calculated.The outputs of the Apriori algorithm are easy to understand and many new patterns can be identified. However, the sheer number of association rules may make the interpretation of the results difficult. A second weakness of the algorithm is the computational times when it searches for large itemsets, due to the exponential complexity of the algorithm.Searching for frequent itemsets performed by Apriori algorithm to get the items that often appear in the database and the pair of items in one transaction. Pair of items that exceed the minimum support will be included into the frequent itemsets are selected. Frequent itemsets that exceed the minimum support will generate association rules after decoding. One frequent itemsets can generate association rules and find the confidence, which is uses a hybriddimension association rules. Results from the mining process show a correlation between the data (association rules) including the support and confidence that can be analyzed. The support is defined as percentage of transactions that contained in the rule and is given by Support = (# of transactions involving A and B) / (total number of transactions). The other factor is confidence it is the percentage of transactions that contain B if they contain A Confidence = Probability (B if A) = P(B/A)Confidence = (# of transactions involving A and B) / (total number of transactions that have A).

V. CONCLUSION

We have shown market basket analysis using association rules to determine the customer behavior pattern.To determine association rules we have used apriori algorithm. In this paper we have also shown an efficient tool analysis for predicting customer behavior in order to increase sales .The tool designed is quite efficient consumes less time and memory space .Future Scope for our Proposed System would be:
Recommending items to individual customers (groceries delivery). Making the tool more smarter and faster. Automatically generating recommendations for the individual shopping online (Ecommerce websites).

REFRENCES

[1] http://www.com.nus.edu.sg/dm2/ “Data Mining II Interser=tingness and Interaction”.Available:Dec 5,2000

[2] Jiawei Han and Micheline Kambler “Data Mining: Concept and Techniques”, Morgan Kaufmann Publishers 2000 ,ISBN 1558604898

[3]http://siva.bpa.arizona.edu/data_mining/Data_Mining.htm"Introduction to Data Mining:. Availabe: Dec 5,2000

[4] R. Agrawal, R. and R. Srikan, "Fast algorithms for mining association rules", Proceedings of the 20th international conference on very large data bases, 478499, 1994.

[5] Trnka., "Market Basket Analysis with Data Mining Methods", International Conference on Networking and Information Technology (ICNIT) ,2010.

[6] Chiu, K.S.Y., Luk, R.W.P, Chan, K.C.C., and Chung, K.F.L, "Marketbasket Analysis with Principal Component Analysis:An Exploration", IEEE International Conference on Systems, Man and Cybernetics, Vol.3, 2002.

[7] Weng, S.S., Liu, J.L.: Featurebased recommendations for onetoone marketing, Expert Systems with Applications, Vol. 26, 2004, pp. 493508. 2

[8] Berry, M.J.A., Linoff, G.S.: Data Mining Techniques: for Marketing, Sales and Customer Relationship Management (second edition), Hungry Minds Inc.,

