DEVELOPMENT OF RECOMMENDATION ENGINE BY AUTHENTICATING ASSOCIATION RULES ON SALES DATA USING LOGISTIC REGRESSION

Sonu Verma, Indian Institute of Foreign Trade, New Delhi, India
Kavita Pabreja, Maharaja Surajmal Institute, GGSIP University, New Delhi, India
Arpan Kumar Patra, Indian Institute of Foreign Trade, New Delhi, India

dx.doi.org/10.18374/JABE-21-2.7

ABSTRACT

In modern era of stiff competition, organizations are facing a lot of challenges to offer best deal to a customer so as to retain him and get good profit too at the same time. Numerous data mining and data analytics techniques have been used in the past to uncover hidden patterns from the buying behavior of customers in order to develop recommendation engine. Though association rule mining algorithms like Apriori, ECLAT (Equivalence Class Transformation), FP-growth (Frequent Pattern growth) find immense applications in the discovery of association rules, but one needs to assess the reliability of rules generated by these techniques. In this paper, an attempt has been made to validate the rules generated by Apriori algorithm by evaluating the Goodness of fit test in logistic regression. We have also analyzed pseudo R squared parameter to determine how much variation in dependent variable is explained by independent variables. The study is based on real dataset collected from the retail outlets in South Delhi area. With the application of Logistic regression algorithm, we are able to uncover certain novel facts about the association rules generated by Apriori algorithm. which will help retailers to optimize store layout, target precise campaigns, plan inventory to customer's need and to offer more profitable advertising and promotions.

Keywords*: Association rule mining, Apriori algorithm, Logistic regression, Pseudo R squared, goodness of fit.*

1. INTRODUCTION

In today's competitive retail scenario, companies around the world employ digital means to record sales transactions in order to gather, study and act on consumer purchase data to strengthen consumer relationships and yield greater consumer lifetime value. It remains a challenge for the retailers to understand the market basket and predict buying behavior of customers. While a strong predominance of unorganized retail players exists in Indian market, there is a steady growth and consumer adoption of organized formats like multi-brand retail, large format stores and exclusive retail outlets in metro and Tier-I cities that employ computerized billing and database storage of sales transaction at point of purchase. In addition, large and popular local stores across Indian cities are also moving from paper invoices to electronic point-in sales.

This huge collection of data, which is structured, semi-structured or unstructured can be used to extract knowledge and explore interesting, non-trivial rules and patterns using data mining techniques so as to provide managers with solutions for complicated problems.

Market Basket Analysis is an important data mining and modelling technique to discover attractive relationships among the high volume of sales transactions. It is based on the theory that if a customer buys a set of certain products, there are probable chances that he or she might buy other set of products. Market basket analysis, an example of mining association rules, helps a retailer in retaining and growing customers through creating, delivering, and communicating a superior customer experience. The common way to analyze market basket data is to search for meaningful association rules based on support and confidence for an itemset.

Customers purchase many items from various different categories of products in a single transaction. To stay competitive, retailers must understand consumer behavior and purchase patterns. Understanding customer behavior can help retailers to retain customers, improve sales, and extend their relationship with