

University of Science and Technology
College of Graduate Studies and Academic Advancement

A Research project in:

DATA MINING IN CRM
Extracting Sales Patterns Using Association Rules

Case Study:

Sales transactions data from bakery store

A thesis submitted in Partial Fulfillment of the Requirements for the
Degree of Master in Information Systems

By:

Esra Saad Said Ahmed Awadallah

Supervisor:

Dr. Atif Ali Mohamed

July 2014

Abstract

Customer satisfaction is the main measure of business successful, so managers work hard to meet customer's needs.

Retailing is an industry with high level of competition. It is a customer-based industry which depends on how it could be aware of what the customers' needs and requirements are.

The purpose of this research is to find frequent itemset then generate association rules, so customer shopping behavior was analyzed to identify frequent pattern on sales transactions datasets. This is done by using data mining technique. Apriori and FP-Growth algorithms were selected as methods in the data mining modeling phase. Using WEKA (Waikato Environment for Knowledge Analysis) software.

The algorithms performance is analyzed based on the execution time for different number of instances in bakery data set. The results show that FP-growth faster than apriori algorithm in evaluating phase.

The generated models discovered correlations between products in a form of association rules with several measures support, confidence, lift, leverage and conviction. By using these rules we can predict related products that might be purchase together.

Based on the rules for combining popular products, marketing personnel or corporate decision makers can formulate more appealing marketing plan or operational rules and actively offer products that might interest the customers.

المخلص

البيع بالتجزئة هو احد الأعمال ذات ميزه تنافسيه عاليه، والتي تعتمد في نجاحها على الإلمام التام بمتطلبات واحتياجات العملاء لذلك يعتبر العميل أصل من أصول المؤسسة.

الغرض الرئيسي من هذا البحث تحليل السلوك الشرائي للعملاء لاستخلاص الأنماط المتكررة في سجلات المبيعات وذلك بتطبيق خوارزميه "Apriori" و خوارزميه "FP-Growth" وهما إحدى الطرق المستخدمة للتقيب عن الارتباط بين البيانات. تم استخدام برنامج ويكا "WEKA" لتطبيق هذه الخوارزميه. تم تقييم الأداء لكل من الخوارزميتين بناء على فترة زمن التنفيذ لأحجام مختلفة من البيانات.

وكنتيجه لتطبيق الخوارزميتين تم إنشاء نموذجين، لكل: المدخل هو سجلات المبيعات والمخرج عبارة عن قواعد تسمى "Association Rules" تبين الارتباط بين المنتجات أو الأصناف، والتي بموجبها يمكن إجراء توقعات للسلوك الشرائي للعملاء.

1.1 Introduction:

The analysis of customer behavior is used to maintain good relationship with customers. It maximizes the customer satisfaction. We can also improve customer loyalty and retention. Based on the historical data of customers' shopping behaviors, prediction is done and data is estimated with the help of data mining techniques.

The aim of this research is to extracting frequent pattern on sales transactions using apriori algorithm for mining association rule, A typical usage scenario for searching frequent patterns is the so called "market basket analysis".

1.2 Problem statement

The research focuses on the problem of frequent pattern mining that can be defined as: given a large dataset of transactions, each consists of set of items. Aim of this research is to find all the frequent itemsets then generates association rules.

1.3 Research objectives

The goal with this research can be split up as following:

- Concept: understanding the concept of CRM, Data Mining, Market Basket Analysis, Association Rules and which used as a tool to meet the objectives of others.
- Method: explain in detailed how these concepts used to extract useful knowledge from transactional datasets.
- Implementation: examine this tool with real datasets.

1.4 Methodology

This research used market basket analysis concept that deal with presence and absence of items in the shopping basket. And discover the association between these items in retail transactions, by generating association rules. To do this, two association rule mining algorithms were tested in WEKA software of version 3.7.4. WEKA software is a collection of open source of many data mining and machine learning algorithms, including pre-processing on data. The results were interpreted and discussed; a comparison was done between these two algorithms. The data is a bakery basket data; it's about 20000 instances for 50 attributes/items.

1.5 Thesis structure

Chapter two takes a look on CRM (customer relationship management) and Data Mining definitions, objectives, benefits and so on. Chapter three explains market basket analysis concept and describes the technique of association rule mining using apriori algorithm. Chapter four details Weka data mining tool, its features and the facilities that provides.

Finally chapter five includes our use of weka to experiments with Apriori algorithm to generate association rules from the sample dataset.

1.6 Related Work

A various journals and articles were studied regarding to the research topic, some of them are described here:

- Isti surjandari: implemented apriori algorithm using weka software to identify associated products which then grouped in mix merchandise, these associations will be applied in the design layout of products in supermarket. Results found new categories association rules; based on it a new design of product placement layout was proposed. This design expected to give more shopping convenience to the customers [1].
- Pramod Prasad, dr.Latesh Malik: elaborated the role of association rule mining in extracting frequent patterns within a dataset and showcases the implementation of the Apriori algorithm in mining association rules from a sales transactions dataset, using weka software. Results show the best association rules for frequent patterns [2].
- Shweta, dr.kanwal garg: implemented apriori to find the best association rules in bank database using data mining tool Weka. Result is extracting useful information that answers complex queries in this large database [3].
- Jiawei Han, Jian Pei, and Yiwen Yin: proposed a novel frequent-pattern tree (FP-tree) structure and developed an efficient FP-tree based mining method, FP-growth, for mining the complete set of frequent patterns by pattern fragment growth. Our performance study shows that the FP-growth method is efficient and scalable for mining both long and short frequent patterns, and is about an order of magnitude faster than the Apriori algorithm and also faster than some recently reported new frequent-pattern mining methods [4].
- Mirzaei.Afshin, Sheikh.Reza: This study elaborates upon the use of association rule mining in extracting patterns that occur frequently within a dataset and showcases the implementation of the FP-Growth algorithm in mining association rules from a real dataset obtained from a supermarket chain containing sales transactions of a retail store. Results show the association rules that give the associations between purchased items[5].

- Ankur Mehay , Dr. Kawaljeet Singh, Dr. Neeraj Sharma: First they found frequent itemsets using Weka tool and Rapid-miner tool. Then generate association rules from the frequent itemsets using apriori and FP-growth algorithms. By analyzing the data, and giving different support and confidence values, we can obtain different number of rules. During analysis it found that FP-growth is much faster for large number of transactions as compare to apriori[6].
- K.Vanitha, R.Santhi: presented the performance comparison of Apriori and FP-growth algorithms. The performance is analyzed based on the execution time for different number of instances and confidence in Super market data set. These algorithms are presented together with some experimental data. The performance study shows that the FP-growth method is efficient and scalable and is about an order of magnitude faster than the Apriori algorithm[7].

