CERTIFICATION OF APPROVAL

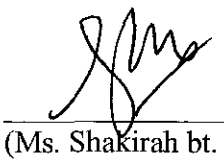
Sales Indicator Support System

By

Nurwaheeda binti Mohamad Ariffin

A project dissertation submitted to the
Business Information System Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons)
(BUSINESS INFORMATION SYSTEM)

Approved by,



(Ms. Shakirah bt. Mohd Taib)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
May 2011

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

NURWAHEEDA BINTI MOHAMAD ARIFFIN

ABSTRACT

Data analysis and knowing customer purchasing behaviour are vital in determining sales of products in retailing industry. However, most of the business owners do not realize the significance of actual analysis and the importance of customer buying patterns but only focusing on data administration. Realising this problem, the author suggests a system that can evaluate customer purchasing behaviour by analysing each of their transactions using Market Basket Analysis approach. The scope of this project is focusing on retailing industry, narrowed down to the grocery stores and mini market. The methodology of the research includes online research, interview with the owner of a grocery store and also the implementation of prototyping model. Sales Indicator Support System is one of the tools that give advantage to the retailers in identifying customer purchasing behavior. The system will create an analysis report for the sales of the products and relationship among the items; giving ideas to the business owners on how they can improve their marketing strategy focusing in store layout and promotion.

ACKNOWLEDGEMENT

This Final Year Project has involved many parties in order to achieve a successful accomplishment for the author thus reaching its goal.

A lot of gratitude to the Almighty God for all the blessing and strength given to me that finally I was able to complete my research and come out with the system that is Sales Indicator Support System.

I would like to express my deepest and sincere gratitude to my supervisor Ms Shakirah bt Mohd Taib for her supervision and support that she truly gave throughout these two semesters. Without her assistance, advice and personal guidance, I could never have acquired as much knowledge and input as I already gained in completing the project.

My special thanks to my beloved parents and family whose supportive spirit never faded in supporting me to accomplish this project. Their understanding and encouragement was in the end made this project possible.

Not to forget many thanks to my friends, who always willing to help and their generosity in sharing ideas and suggestions for the successful completion of the project. Lastly, I would like to express appreciation to everyone who has contributed directly or indirectly in completing this project successfully.

TABLE OF CONTENTS

CERTIFICATION OF APPROVAL i

CERTIFICATION OF ORIGINALITY ii

ABSTRACT iii

ACKNOWLEDGEMENT iv

CHAPTER 1: INTRODUCTION

1.1 Background 1

1.2 Problem Statement 2

1.3 Objectives 2

1.4 Scope of Study 3

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction 4

2.2 Consumer Purchasing Behaviour 4

2.3 Analysis of Customer Behaviour and Marketing 7

2.4 Statistical Method in Analysing Customer Purchasing Behaviours 8

2.5 Data Mining 9

2.6 Data Mining is Sales Analysis 13

CHAPTER 3: METHODOLOGY

3.1 Research..... 14

3.2 Development of Prototyping..... 16

3.3 Project Activities..... 18

3.4 Key Milestones & Gantt Chart..... 19

3.5 Tools..... 20

CHAPTER 4: RESULT & DISCUSSION

4.1 Findings..... 21

4.2 Data Collection..... 22

4.3 Data Pre-Processing..... 22

4.4 Framework for Mining Association Rules..... 24

4.5 System Description..... 25

4.6 Final System of Research..... 26

CHAPTER 5: CONCLUSION & RECOMMENDATION

5.1 Conclusion..... 39

5.2 Recommendation..... 40

REFERENCES..... 41

LIST OF FIGURES

Figure 1	Part of Association Rules	11
Figure 2	Prototyping Model	16
Figure 3	Framework for Mining Association Rules	24
Figure 4	System Description	25
Figure 5	Homepage	26
Figure 6	Analysis of Items	27
Figure 6.1	List of Items and List of Transactions	28
Figure 6.2	Total Count of 1-Itemsets	29
Figure 6.3	Min, Median, Mode	29
Figure 6.4	Pie Chart	30
Figure 6.5	Support of 1-itemset	34
Figure 6.6	Total Count of 2-Itemsets	35
Figure 6.7	Support of 2-Itemsets	36
Figure 6.8	Confidence of 2-Itemsets	37

LIST OF TABLES

Table 1 Co-Occurrence of Products Adapted.....12

Table 2 Prototyping Model.....17

Table 3 Project Activities.....18

Table 4 Key Milestones and Gantt Chart.....19

Table 5 Transactions Data.....23

CHAPTER 1

INTRODUCTION

1.1 Background of Study

In Malaysia nowadays, the competition in retailing industry is increasing as there are a lot of new entrants enter the market place. In order for them to compete within the market, gaining competitive advantage is one of the elements that they need to emphasize on. Competitive advantage is basically gained by the company when the customers probably will choose their products over the competitors.

Analysing the customers' purchasing behaviour is vital to make sure a company can attract more potential customers to make a purchase in future. It is a good opportunity for a company if they can identify their customer purchasing habits as they can improve or modify their products and services in future. Besides, by tracking customers purchasing behaviour, the business owners can create a business strategy on how to increase their profits by fulfilling the customers' needs, demands and preferences. It is strongly important for a company to determine shopping patterns of their customers, so that they can know their actual potential to market the products in a way increase customers' retention.

A lot of organizations do not recognize the importance of data analysis in their organizations as they do not realize the significance of the data that can be gained from their customers. Hence it is becomes difficult for the company to see their potential customers and overall performance especially in the market segment. As a

result, a company should conduct an analysis in order to identify what are customers' current need and preferences and also to identify customer purchasing behaviour.

1.2 Problem Statement

A company who failed to track buying patterns of their customers often does not know their real strength and potential in the market. And for retailing industry, one of the pitfalls in doing business is they do not pay much attention about customer purchasing behaviour. Hence, they probably cannot identify:

- What kind of products customers like to buy
- What products customers like to purchase together
- What is the most common items found among customers

Therefore, besides offering good and high quality products to the customers, retailers should take another step to ensure what actions needs to be taken to increase the level of their performance in the sales of the products.

1.3 Objectives

- To analyse customers purchasing behaviour based on the transactional data
- To develop a Sales Indicator Support System that can integrate data administration and analysis report
- To create an analysis report for the sales of the products and relationship among the items

1.4 Scope of Study

Indicator

An indicator system can be used in monitoring and evaluating the sales patterns in the grocery stores. Indicator can be a measurement, a number, a fact, an opinion or a perception that points at specific condition or situation and measures changes over the time [1]. Indicators are useful for pointing the direction of change, whether the results are positive or negative or the situation is improving or vice versa. It is also an effective tool for monitoring and evaluating progress since it offers additional benefits for summarising information in a single number of figures. It can be used to take corrective measures when the progress in not being achieved. Other than that, it can compare trends in the same location across the time and one of the power tools in project management systems for measuring project outcomes and impacts.

Target Market

Target markets for this project are business owners in retailing industry, narrowed down to the mini market and grocery stores. The scope will emphasize on how they can increase their sales of products by analysing and predicting customer purchasing behaviour using an indicator system and also by identifying what are the marketing strategies that can contribute to their success in the business.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter provides information obtained through studies regarding the conceptualization of the key terms for Customer Purchasing Behaviour, Analysis of Customer Behaviour, Statistical Method in Analysing Customer Purchasing Behaviours, Data Mining and Data Mining in Sales Analysis. The topics will be explained in the five following sections.

2.2 Customer Purchasing Behaviour

Customers are smart enough when it comes to the shopping experience. A memorable shopping experience is determined by retail factors, transport mode and customers itself [2].

There are many factors directly will affect the customers to purchase again in future starting from the physical appearance, atmospheric elements and presentation of service quality to the customer, [3]. Besides, a store environment that includes store layout, decoration and display is major factor that can influence the customers to make a purchase [4].

2.2.1 Physical Appearance of the Store

The impression of a store is determined by layout, decorations and display of merchandise and ease of movement in the store [5]. All the factors are probably can give significant impact on store performance environments, atmospheric cues and customer effect on the retail experience. Store layout design can play a key role not only in satisfying buyer's requirement but also in influencing their wants and preferences. A suitable and proper store layout can drive success to their business. An attractive and impressive decoration with good combination of products display can attract customer's attention towards the store.

2.2.2 Atmospheric Elements

Atmospheric elements are referred to the physical properties of the retail that are designed to create an effect on consumer purchases [6]. Atmospheric variables such as music, colour, scent, lightning and crowding can affect consumer behaviour in retail store environments [5]. Lightning, air condition, floor covering and architecture will also affect a shop's image and consumer's choice. Appropriate use of lightning and sound is one of the examples that are effective in establishing a good mentality for the consumers and for enhancing enjoyment. By understanding atmospheric effects towards the customers, the retailers can create new strategies to modify their store environment in order to create positive shopping experience in the future.

2.2.3 Presentation of Service Quality

The retailers should be aware the effect of service offered by their business. Knowledgeable and helpful salesperson is important and can give impact to a customer while deciding to choose a shop or shopping centre [6]. Customers will option for those who give the best customer service in their daily business. The salesperson is expected to learn and know about every single thing in the shops so that they can quickly assist the customers when they seek for help. A company with good customer service can keep and maintain their customers as they successfully meet the customers' expectation.

2.3 Analysis of Customer Behaviour and Marketing

Customer's behaviour is dynamic because the thinking, feelings and actions of individual customers and targeted customers group are constantly changing. Marketing strategies play important roles for a company to learn, attract, grow and retain their customers. Marketing strategies not only adapt to customer needs and demands but also change what customers think and feel about themselves when there are various marketing offerings and reasons for purchase.

Most retailers do not want customers to come to their stores only once and never return in future. Store royalty can be strongly influenced by the arrangement of the environment, particularly reinforcing properties in retails store [7]. Analysis of the store environment and customer's store related affect, cognition and behaviour is critical for successful marketing. To meet customers' expectation, some factors need to be identified by the retailers.

The two popular approaches to study and analyse customer purchasing behaviour is by using Statistical Methods and Data Mining such as Association Rules Analysis.

2.4 Statistical Method in Analysing Customer Purchasing Behaviours

Statistics is a branch of mathematics dealing with analysis and interpretation of masses of data [8]. The use of statistical method has long played an important role in quality control and quality improvement in business. Statistical Method can be used to analyse customer purchasing behaviours by using descriptive statistics and inferential statistics.

2.4.1 Descriptive Statistics

Descriptive statistics is defined as collecting, summarizing and processing data to transform the data into information [9]. It is used to describe the shape of a sample [10]. There are graphical or visual descriptive tools which generally include bar charts, pie charts and histograms. Graphical tools help the users to see how the data behave and summarize the data visually. On top of that, numerical descriptive tools will help the users to summarize data numerically. It will provide the users the statistics consists of the average, median, mode and also largest and smallest data values.

2.4.2 Inferential Statistics

An inference is a deduction or a conclusion reached on basis evidence and reasoning [11]. Inferential statistics is used to draw conclusions and make predictions about the behaviour of the population based on the information in the sample. It involves a decision-making process that allows users to objectively measure either the project are significant or not [12]. The technique of inferential statistics allows the user to make inference and conclusions about the population from the sample [13]. A sample is a piece of population and properly selected from a population. The information gained

from the sample can describe and give reliable information about the population.

2.5 Data Mining

There are a lot of definitions of Data Mining. Data Mining is the discovery of knowledge with software that automatically identified and utilizes information that is hidden in the data [14] and according to Ho [15] data mining is the process of exploring and analyzing large volumes of data in order to discover interesting and hidden pattern, rules and relationships with data. The other definition of data mining that it is a term which is used for collecting, analyzing and making decisions based on the information [16].

There are two types of data mining which are directed and undirected [17]. For directed data, the target field is selected and available data is used to build a model that describes one particular variables of interest in terms of the rest of the available data meanwhile, there is no target field for undirected data and it is used to establish relationship among the variables.

Most of the companies nowadays utilized Data Mining techniques that enable them to determine relationship among internal and external factors such as pricing, product positioning and customer demographics [18].

The data mining technique most closely allied with Market Basket Analysis is the automatic generation of Association Rules [19]. Market Basket Data describes what customer purchases and Association Rules is a technique used to gain better understanding about relationship among items in the single basket. They play an important role in shopping basket data analysis and indirectly can determine the best arrangement for the store layout.

2.5.1 Market Basket Analysis

Market Basket Analysis is a popular technique used by retailers to understand what products are commonly purchased together [15].

Based on Ho [9], Market Basket Analysis uses the information what customers purchase to provide insight into who they are and why they make certain purchase. It also provides insight into the merchandise by telling us which products tend to be purchased together and which are most amenable to the promotion [19].

The importance of data gathered by using Market Basket Analysis is to customize store layouts, cross-selling and identify customer segments based on their buying patterns. Besides, by leveraging Market Basket Analysis the retailers can also drive more profitable marketing strategies such as advertising and promotions, attract more customers and also increase the value of the market basket.

The business owner can also analyse item details and relationship between the items by using Association Rules.

2.5.2 Association Rules (AR)

According to Lili Aunimo [20], Association Rules express regularities that exist in a dataset. It has been used in Market Basket Analysis to capture relationship present among items in large data sets.

2.5.2.1 Association Rules Concept

A general form of an Association Rule is [21]:

Body → Head [Support, Confidence]

Each is shown in Figure 1:

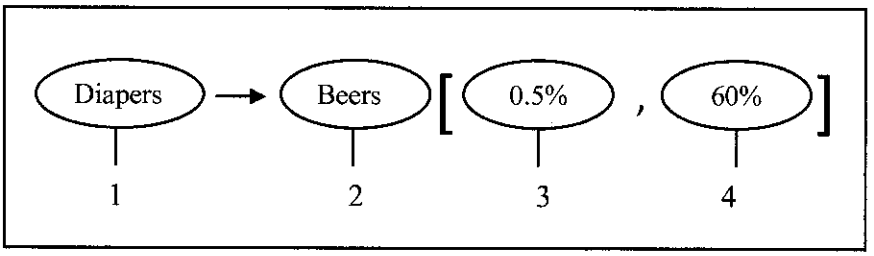


Figure 1: Part of Association Rules [21]

Pirjo Moen [21] defined that:

1) Antecedent is left-hand side, body
2) Consequent is right-hand side, head
3) Support is frequency
4) Confidence is strength

Based on the relationship shown in Figure 1, an antecedent is an item found in the data and a consequent is an item that is found in combination with the

antecedent. Support is the fraction transactions that contain both 1 and 2 meanwhile confidence are how often 2 appear in transactions that contain 1.

An example on how the Association Rules works is as below:

Table 1: Co-Occurrence of Products Adapted [19]

	OJ	WINDOW CLEANER	MILK	SODA	DETERGENT
OJ	4	1	1	1	2
WINDOW CLEANER	1	2	1	1	0
MILK	1	1	1	0	0
SODA	2	1	0	3	3
DETERGENT	1	0	0	1	2

From the Table 1 above, it illustrates five transactions in a grocery store that carries five products. Each of these transactions gives the information about which products are purchased with other products.

The co-occurrence table above tells the numbers of times that any pair of products was purchased together. For instance, the box where the 'Soda' row intersects the 'OJ' columns has a value of "2", meaning that there are two transactions contain both soda and orange juice. And some simple patterns that been also highlighted from the co-occurrence table are:

- Orange juice and soda are more likely to be purchased together than any other two items
- Milk is never purchased with soda or detergent
- Detergent is never purchased with window cleaner or milk

These observations are examples of associations and may suggest a formal rule like: "If a customer purchases soda, then the customer also purchases orange juice" [19].

The transaction that includes both soda and orange juice is two of five and it supports the rules (two out of five or forty percent). This rule also has a strong confidence since both the transactions that contain soda also contain orange juice [20]. In other meaning, out of three, there are two transactions that contain both soda and orange juice. The rule "if soda, then orange juice" has a confidence of sixty seven percent [19].

2.6 Data Mining in Sales Analysis

Besides using camera, machine recognition, in-store observation or in-store interviewing [22], Data Mining would be more useful in tracking consumer buying habits. The most important part for the retailers in using Data Mining in determining consumer purchasing behaviour is they must be able to understand the connections between the system and rules used in line with their business plan and marketing strategy [23]. After all, the 'pattern' that is produced by Data Mining represent novel, hidden and previously implicit knowledge that may be used for increasing sales revenue in a profitable manner [24].

CHAPTER 3

METHODOLOGY

3.1 Research

3.1.1 Online Database

Throughout the project, research method being used in from the discovering idea through the internet, related journals and books. This project requires intensive research and information searching on these following topics:

- The Importance of Understanding Customer Purchasing Behaviour
- Why Use Data Mining, Market Basket Analysis and Association Rules in Analysing Consumer Purchasing Behaviour
- Layout and Promotions in Retails Industry

3.1.2 Interview

The author will make an appointment with the business owner to organize an interview session in order to get more information about their business. The types of question that will be asked during the interview include but not limited to:

- In which industry is their business
- How they conduct their business
- What types of products they offer to the customers
- How they organize the inventory, store layout and promotions

3.1.3 Gathering the Data

Besides having interview, the author will also collect some data from their business after getting permission from the business owner. This would be done by collecting the details of transactions from the customers in the specific duration of time.

3.2 Development of Prototyping

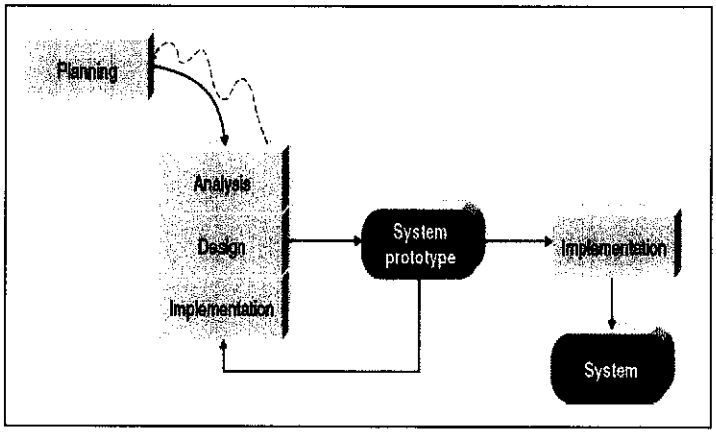


Figure 2: Prototyping Model

In developing Sales Indicator Support System, the methodology that will be used is Prototyping, as shown in Figure 2. Prototyping is important in building fast, more reliable and better quality system. It quickly provides a system for the user to interact with. It also save time and promptly can refine the real requirement.

There are four phases involves in prototyping which are planning, analysis, design and implementation. The objectives, how to do and deliverables for each of the phases are illustrated in the Table 2.

Table 2: Prototyping Model

Phase	Objectives	How to do	Deliverables
Planning	Why build the system, how to structure the system	Identify opportunity, analyze feasibility, develop work plan	Project Plan
Analysis	Identify the target audience, when and where it will be used	Develop analysis strategy, determine business requirement	System Proposal
Design	How will the system work?	Design physical system, interface, program and databases	System Specification
Implementation	Delivery and support of completed system	Construct system	Final system

3.3 Project Activities

Table 3: Project Activities

No	Task	Status
1	Research in the topic	Done
2	Preparing the Extended Proposal	Done
3	Proposal Defense and Progress Evaluation	Done
4	Submission of the Interim Report	Done
5	Develop the Work Plan	Done
6	Development Stage	Done
7	Design the Interface	Done
8	Build the System	Done
9	Testing Stage	Done
10	Submission of the Technical Report	Done

Based on Table 3, project activities will be divided into four phases. The first phase will include the research in the topic, preparing the extended proposal, participating in the proposal defense and progress evaluation and also the submission of the Interim Report. Once the tasks in Phase 1 are completed, Phase 2 is going to be embarked with developing the work plan. An analysis will be done in order to collect information that is required for the project. Phase 3 is the development stage that includes interface and system design. The system will consist of three modules; Module A, Module B and Module C, specifically for Statistics, Graphs and Final Result. The Statistics part is designed to calculate and display the data meanwhile Final Result part will shows the analysis of customer purchasing behaviour based on the calculation of the customers' transactions. The last phase is testing the performance of the system. The purpose of the testing is to ensure the system is ease of use and meet the expectations and needs of the project.

3.4 Key Milestones and Gantt Chart

Table 4: Key Milestone & Gantt Chart

NO	TASKS	MONTH								
		1	2	3	4	5	6	7	8	9
1	Research in the topic									
2	Preparing the Extended Proposal									
3	Proposal Defense and Progress Evaluation									
4	Submission of the Interim Report									
5	Develop the Work Plan									
6	Development Stage									
7	Design the Interface									
8	Build the System									
9	Testing Stage									
19	Submission of the Technical Report									

As shown in Table 4, within duration eight months, there are several tasks that need to be completed in developing the system. It is started by doing the research in the topic and it will be done in duration four months starting from February until May. The preparation for extended proposal, proposal defense and interim report will begin from February until April.

After doing research and complete all the documentation, the author will develop a work plan starting from March until May 2011 and the milestones reach the last stage which is development stage that consists of designing the interface, build the system, testing the functionality of the system and also preparation for technical report.

3.5 Tools

The system is planned to develop using HTML and PHP language for the main part of the system and MySQL would be used for the system's database. The rationale behind the selection of web based approach is because PHP provides support for MySQL, can easily integrate with the latest features, have great flexibility during and after initial project and it also provides affective and economical web solutions.

CHAPTER 4

RESULT & DISCUSSION

4.1 Findings

An interview had been conducted in order to gain some information from a grocery store, which is 'Kedai Runcit Rahmath', located in Jalan IB1, 32600 Sri Iskandar, Tronoh, Perak. The owner of the store is Mr Basir b. Mohamad, 40 years old. 'Kedai Runcit Rahmath' offers a variety of items to the customers like groceries, stationeries, cosmetics, toiletries and also packaging food.

The interview is being done by interviewing one of the workers, Mr Rafeeq Ahmad. Most of the questions asked are respectively focusing on how they conduct their operations and manage the store layout and promotion. Regarding to Mr Rafeeq, they never study what types of products that had been purchased by most of the customers and there are also no specific procedures that need to be followed in arranging items in the store. The arrangement is basically based on their preferences and normal practices in the market. Other than that, the promotion only will be held if they get quantity offer from the suppliers and all the specific items offered during the period of the time are not depends on any particular condition.

4.2 Data Collection

Among the steps involved in performing Data Mining are data collection and data pre-processing.

For the data collection, this process required a collection of data from a specific location. The author had collected some data from a grocery store which is 'Kedai Runcit Rahmath', located in Bandar Seri Iskandar, Tronoh, Perak. The method that had been used is by recording the actual customers' transactions data and after two days of observation, the author managed to collect a list of hundred transactions that consists of sixty seven types of items that had been purchased by the customers. The number of transactions will represent the transactions ID.

4.3 Data Pre-Processing

Data pre-processing is a process to transform the data that had been collected into a format that will be more easily and effectively processed. One of data pre-processing techniques is data reduction.

4.3.1 Data Reduction

The purpose of data reduction is to avoid impractical and infeasible analysis because the huge dataset and also to reduce the volume of dataset but still manage to maintain the integrity. Attribute Subset Selection is one of the

strategies in Data Reduction. It reduces the dataset size by removing irrelevant and redundant attributes.

4.3.2 Data Selection

For the purpose of this research, only ten common items in ten transactions are selected for the mining process. Both items and transactions are shown as Table 5 below:

Table 5: Transactions Data

TID	Purchased Items
1	Paper, Tape, Cereal, ShowerCream
2	Tape, SoftDrink, ShowerCream
3	File, Pencil
4	Marker, SoftDrink, Chocolate, Biscuits
5	SoftDrink, Chocolate
6	Paper, Marker, SoftDrink, Chocolate
7	Pencil, Cereal
8	File, SoftDrink, Chocolate
9	SoftDrink, Chocolate, Biscuits
10	Cereal, Biscuits

The ten items are Paper, Tape, Cereal, ShowerCream, SoftDrink, Chocolate, Biscuits, Pencil, Cereal and File.

4.4 Framework for Mining Association Rules

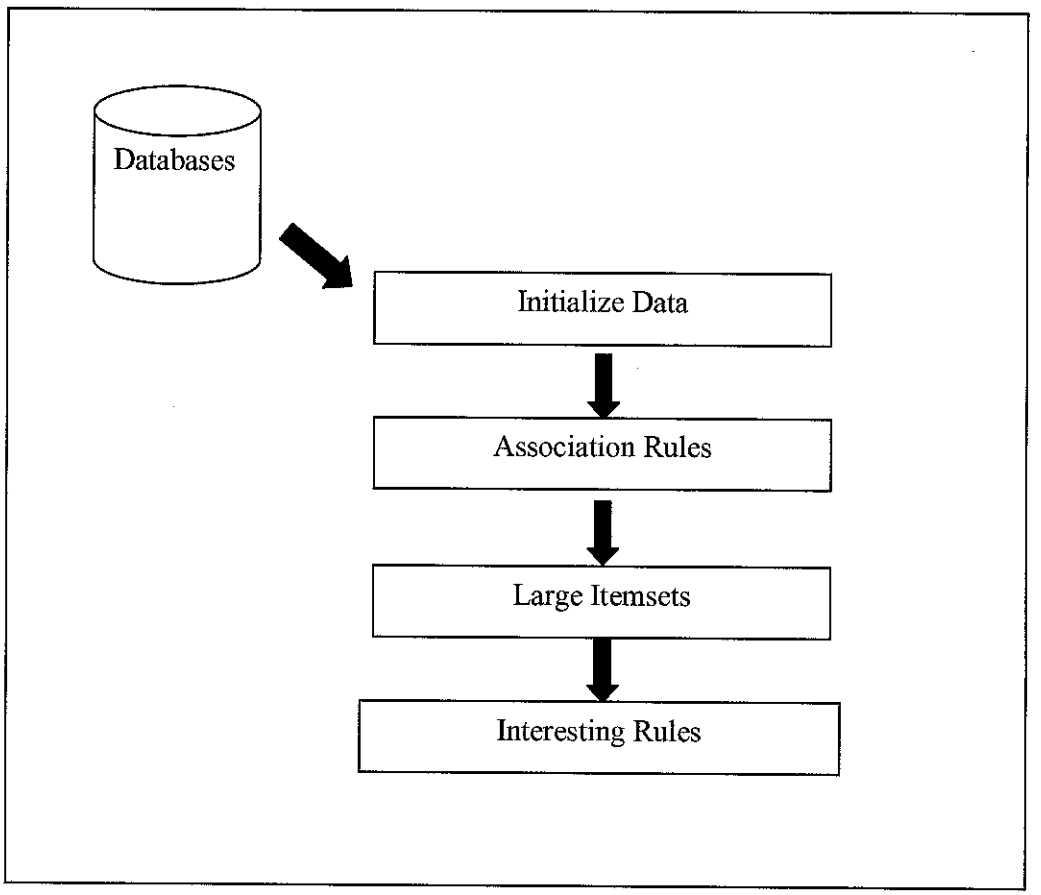


Figure 3: Framework for Mining Association Rules

After selecting the items, all the data will be stored in a database. The system then will initialize the data by using Association Rules algorithm. The function of Association Rules algorithm is to calculate Support and Confidence for each of the transactions. Support and Confidence are the two measures used in order to find interesting rules from the large itemsets.

4.5 System Description

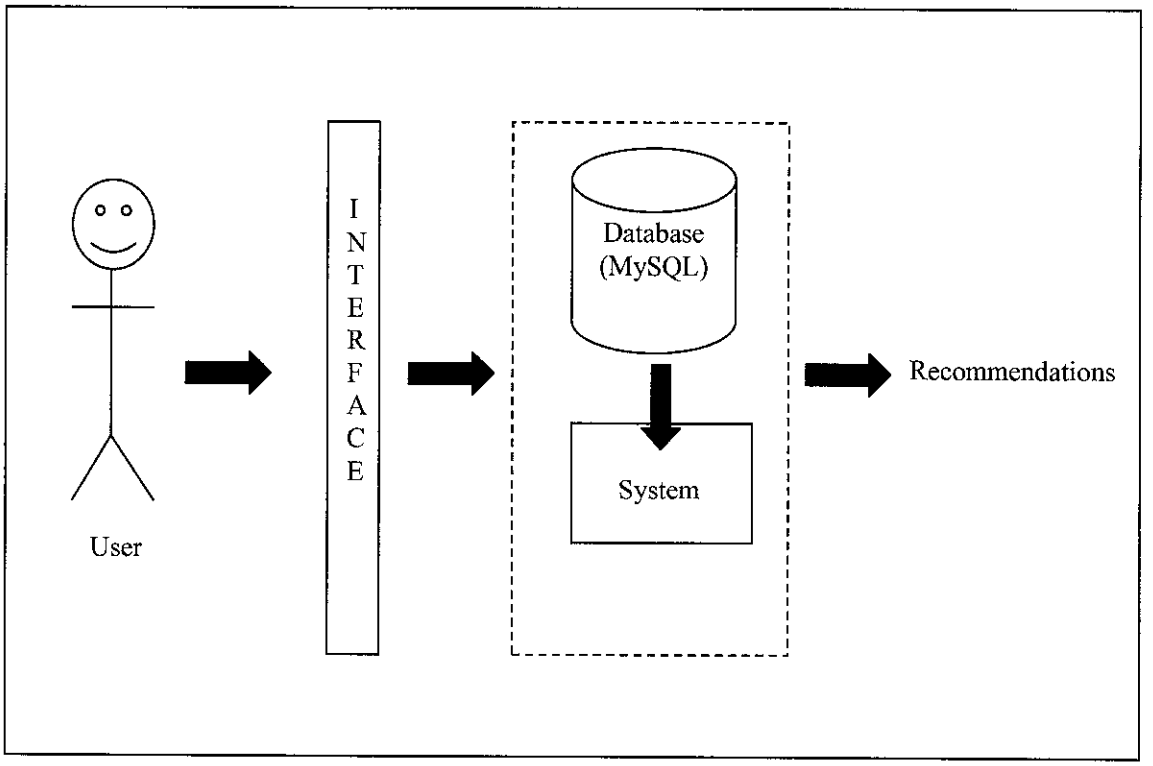


Figure 4: System Description

The system of research will be consisting of the system itself in which all computational processes will be done, a database and also user interface for the user to interact with the system.

4.6 Final System of Research

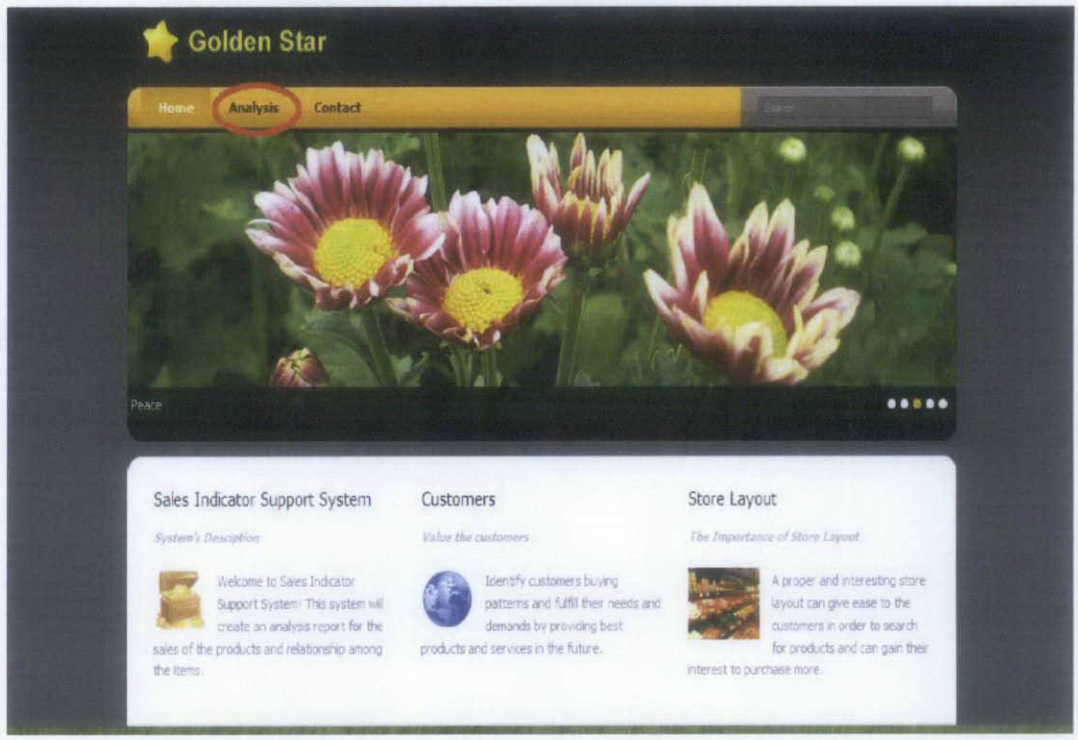


Figure 5: Homepage

The final system of research will show the all analysis results when the users click on the analysis tab from the homepage.

The page then will be directed to the Statistics, Graph and Final Result page.

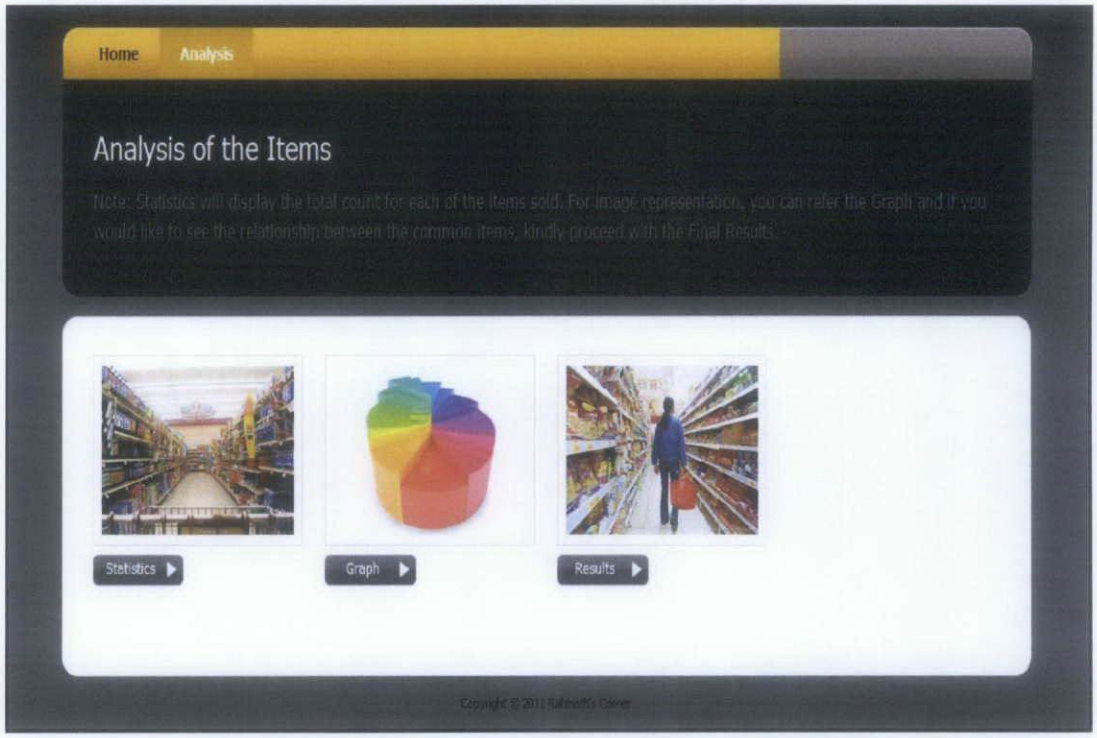


Figure 6: Analysis of the Items

4.6.1 Statistics

When the users click Statistics part, the system will automatically show the list of the items and also the list of the transactions occurred.

```
List of Items
Array
(
[0] => File
[1] => Paper
[2] => Tape
[3] => Marker
[4] => Pencil
[5] => SoftDrink
[6] => Cereal
[7] => Chocolate
[8] => Biscuits
[9] => ShowerCream
)

List of Transactions
Array
(
    [0] => Paper,Tape,Cereal,ShowerCream
    [1] => Tape,SoftDrink,ShowerCream
    [2] => File,Pencil
    [3] => Marker,SoftDrink,Chocolate,Biscuits
    [4] => SoftDrink,Chocolate
    [5] => Paper,Marker,SoftDrink,Chocolate
    [6] => Pencil,Cereal
    [7] => File,SoftDrink,Chocolate
    [8] => SoftDrink,Chocolate,Biscuits
    [9] => Cereal,Biscuits
)
```

Figure 6.1: List of Items & List of Transactions

There are ten items and ten transactions that had been selected for the project as shown in Figure 6.1. After that, system will calculate the total amount for each of the items (1-itemset) that had been purchased by the customers, as shown in Figure 6.2.

```
Step 1: Total Count of 1-itemset
Array
(
    [File] => 2
    [Paper] => 2
    [Tape] => 2
    [Marker] => 2
    [Pencil] => 2
    [SoftDrink] => 6
    [Cereal] => 3
    [Chocolate] => 5
    [Biscuits] => 3
    [ShowerCream] => 2
)
```

Figure 6.2: Total Count of 1-itemset

Based on Figure 6.2, the item that had been purchased by most of the customers is SoftDrink that shows total of 6, meanwhile File, Paper, Tape, Marker, Pencil and ShowerCream shows the same amount which is two. After the calculation of total count for 1-itemset, the system will also calculate the min, median and mode for the dataset.

4.6.1.1 Min, Median, Mode

Min, median and mode of the data are used to measures the central tendency. Mean is useful for predicting future result when there are no extreme values in the dataset and median is more valuable than the mean when there are extreme values in the data set as it is not affected by extreme values. And mode can be very useful when the most items, characteristics or value of a data set is required.

The results for all these three measurements are shown in Figure 6.3.

```
Min    : 2.9

Median : 2

Mode   : 2
```

Figure 6.3: Min, Median, Mode

4.6.2 Graph

Other than that, if the users click graph, the result of the calculations will be displayed in a pie chart, give better visual representation to the users. This simply indicates which products had been purchased by most of the customers and what types of products that need promotion in future in order to increase the sales of the products.

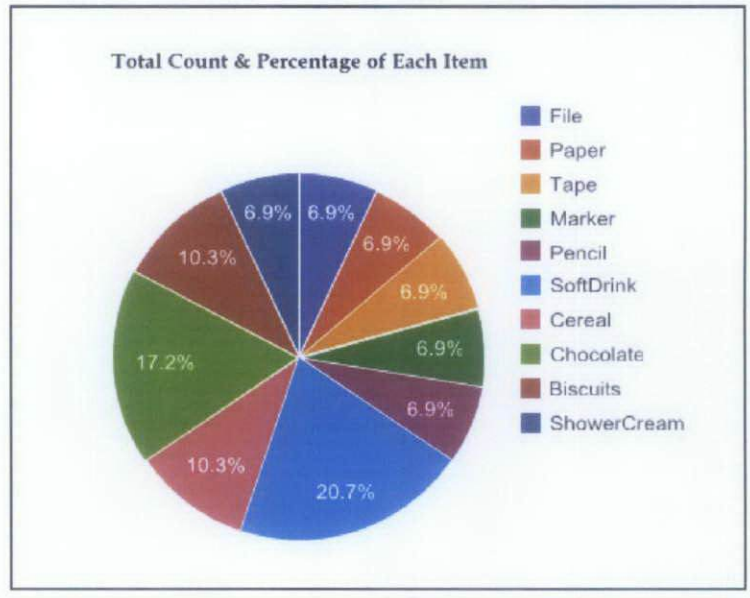


Figure 6.4: Pie Chart

4.6.3 Final Result

The last part is the final result. This part will include all the calculation of the Support and Confidence which is the measures used in order to describe Association Rules. It will also describe the relationship between the items. This system will used minimum support of 20% and minimum Confidence of 50%.

4.6.3.1 Use a Minimum Support of 20%

Support is the number of times itemsets appear in dataset or can be described as the frequency of the itemsets in dataset.

- An itemset is called frequent if its support is equal or greater than minimal support
- Any itemsets does not meet the minimum support count is eliminated from further analysis
- The frequent 1-itemset is joined with itself to create 2-itemset

4.6.3.2 Effect on the support

- If the support is too high, only a few patterns will appear and could miss interesting rare occurrence
- If the support is too low, too many associations will appear

4.6.3.3 Use a Minimum Confidence of 50%

The confidence is the conditional probability that, when item A present in a transaction, item B will also be present.

An itemset have strong confidence if the confidence is equal or greater than the minimal confidence.

4.6.3.4 Rule Measures: Support and Confidence

The computation is as follows:

Support:

$s\{A \rightarrow B\}$ = total numbers of transactions that contain both A and B divided by the numbers of transactions

Confidence

$c\{A \rightarrow B\}$ = frequency {A,B}/frequency{A}

= the total numbers of transactions that contain both A and B divided by the numbers of times that A appears in the dataset

The final analysis will only consider rules derived from itemsets with high support and high confidence.

4.6.3.5 Implementation Steps

Step 1:

For 1-itemsets, the Apriori Algorithm will analyze all the transactions in a dataset and count the total and support count each of the item. Any item that has support count less than the minimum support count is removed.

Step 2:

The remaining itemsets are joined to create 2-itemsets that each comprise of two items. The support count of each of 2-itemsets is calculated and any 2-itemsets that occur with support count greater or equal to minimum support count are used to create 3-itemsets.

Step 3:

All the itemsets generated with a support count greater than minimum support count, form a set of Frequent Itemsets. These itemsets are then used to generate Association Rules with a confidence greater than or equal to minimum confidence.

4.6.3.6 Support, Confidence and Interesting Rules

When the user clicks Final Result, the systems will show the support and confidence for all itemsets. The results from total count of 1-itemsets will be used to create 2-itemsets. If the support counts of 1-itemsets meet or above the minimum support which is 20%, the items will be used in order to generate 2-itemsets. The support count for 1-itemsets is shown in Figure 6.5.

```
Step 1: Support of 1-itemset
Array
(
    [File] => 20%
    [Paper] => 20%
    [Tape] => 20%
    [Marker] => 20%
    [Pencil] => 20%
    [SoftDrink] => 60%
    [Cereal] => 30%
    [Chocolate] => 50%
    [Biscuits] => 20%
    [ShowerCream] => 20%
)
```

Figure 6.5: Support of 1-Itemset

- Support for 1-itemset = (total count of item/list of transactions) x 100

Support count in Figure 6.5 shows that all the items are meet and above the minimum support. The next step is to create 2-itemsets and calculates the total amount for each of the itemsets. The results are shown in Figure 6.6.

```
Step 2: Total Count of 2-Itemsets
Array
(
[File->Paper] => 0
[File->Tape] => 0
[File->Marker] => 0
[File->Pencil] => 1
[File->SoftDrink] => 1
[File->Cereal] => 0
[File->Chocolate] => 1
[File->Biscuits] => 0
[File->ShowerCream] => 0
[Paper->Tape] => 1
[Paper->Marker] => 1
[Paper->Pencil] => 0
[Paper->SoftDrink] => 1
[Paper->Cereal] => 1
[Paper->Chocolate] => 1
[Paper->Biscuits] => 0
[Paper->ShowerCream] => 1
[Tape->Marker] => 0
[Tape->Pencil] => 0
[Tape->SoftDrink] => 1
[Tape->Cereal] => 1
[Tape->Chocolate] => 0
[Tape->Biscuits] => 0
[Tape->ShowerCream] => 2
[Marker->Pencil] => 0
[Marker->SoftDrink] => 2
[Marker->Cereal] => 0
[Marker->Chocolate] => 2
[Marker->Biscuits] => 1
[Marker->ShowerCream] => 0
[Pencil->SoftDrink] => 0
[Pencil->Cereal] => 1
[Pencil->Chocolate] => 0
[Pencil->Biscuits] => 0
[Pencil->ShowerCream] => 0
[SoftDrink->Cereal] => 0
[SoftDrink->Chocolate] => 5
[SoftDrink->Biscuits] => 2
[SoftDrink->ShowerCream] => 1
[Cereal->Chocolate] => 0
[Cereal->Biscuits] => 1
[Cereal->ShowerCream] => 1
[Chocolate->Biscuits] => 2
[Chocolate->ShowerCream] => 0
[Biscuits->ShowerCream] => 0
```

Figure 6.6: Total Count of 2-Itemsets

From the total count of 2-itemsets in Figure 6.6, the support for 2-itemsets will be calculated.

- Support for 2-itemsets $\{A \rightarrow B\}$

  = (total count of items A and B/list of transactions) x 100

From the calculation, there are only six itemsets that meet and above the minimum support. The results are shown in Figure 6.7.

```
[Tape->ShowerCream] => 20%
[Marker->SoftDrink] => 20%
[Marker->Chocolate] => 20%
[SoftDrink->Chocolate] => 50%
[SoftDrink->Biscuits] => 20%
[Chocolate->Biscuits] => 20%
```

Figure 6.7: Support of 2-Itemsets

All of these six itemsets then are used in order to calculate the confidence.

- Confidence for 2-itemsets {A → B}

  = (total count of items A and B/total count A) x 100

```
[+] Tape-ShowerCream (2) / Tape (2) = 1
    100% = when customer purchase Tape, there is also co-occurrence of buying ShowerCream

[+] Marker-SoftDrink (2) / Marker (2) = 1
    100% = when customer purchase Marker, there is also co-occurrence of buying SoftDrink

[+] Marker-Chocolate (2) / Marker (2) = 1
    100% = when customer purchase Marker, there is also co-occurrence of buying Chocolate

[+] SoftDrink-Chocolate (5) / Chocolate (5) = 1
    100% = when customer purchase Chocolate, there is also co-occurrence of buying SoftDrink
```

Figure 6.8: Confidence of 2-Itemsets

Based on the calculation of the confidence from six 2-itemsets, there are only four itemsets that get 100%. The results finally show the interesting rules that describe relationship between common purchase items and the itemsets that are frequently associated together across the entire population.

From the itemsets, it can be described as, when the customers:

- buy tape there is always co-occurrence to buy shower cream
- buy marker there is always co-occurrence to buy soft drink
- buy marker there is always co-occurrence to buy chocolate
- buy soft drink there is always co-occurrence to buy chocolate

Since these four 2-itemsets already get 100% of confidence, hence there is no itemsets of three should be created.

4.7 Discussion

The Association Rules represents rules where some set of items is associated to another set of items. The rules express that a certain product is often bought in combination with a certain set of products.

The Association Rules finds the support count and confidence of itemsets and generates all the subsets of each frequent itemset. It eliminates those itemsets that do not meet the minimum support and confidence. In the end, the interesting rules found will be displayed by the system.

The system will give ideas to the business owners regards to store layout and product arrangement in their store. As an example, based on the results from Figure 6.8, the business owners probably can put tape and shower cream together in one place. Same goes to the marker and soft drink, marker and chocolate and also soft drink and chocolate.

4.7.1 The Implications:

When two items are arranged together in one area, it can gain customers interest to purchase the products together. The customers also do not have to waste their time walking around the store just to look for the product that can indirectly leads to customers' satisfaction. When this situation happened, the customers most likely will come again to the store in future. The business owners can maintain customer royalty and at the same time can increase the sales of products in order to gain higher profit.

CHAPTER 5

CONCLUSION & RECOMMENDATION

5.1 Conclusion

One of the strategies for the retailers to increase their daily sales is they must know how to attract more customers to come to their store and make a purchase. And to do that, they must identify what are the factors that can influence the customers to choose their store over the competitors. One of the strategies is it can be done by analyzing customer purchasing behavior in order to identify customers' needs and demands.

To study customer purchasing behaviour, the terms that can be used are Data Mining and Association Rules. Besides, by using Market Basket Analysis, retailers can also increase their competitiveness as they can directly focus on the customer's buying habits and use the knowledge to conduct their operations. Statistical Analysis can also be used as the guidance for decision making process.

In conclusion, in order to overcome the problems outlined in the problem statement, the creation of Sales Indicator Support System is recommended. The creation of the system will offer ease to the retailers to predict customers purchasing behaviour by analysing customers' transactions data using a tool that integrate data administration and analysis report. In the end, retailers will have basic ideas on how they can improve their marketing strategy by organizing store layout and promotion for their future business.

5.2 Recommendation

For the enhancement, other Data Mining techniques can be used for the purpose to capture and identify customer's demographics. Demographics can be described as a statistical view of a population, generally including age, gender, family size, income and education that are used by business to identify target markets for their goods and services.

By having all the information, the business owners can study further on how they can sell and promote their products to the specific target market effectively. They can also directly identify and indicate changing needs in the marketplace and adjust to them.

REFERENCES

[1] CIDA (1996) *Guide to Gender-Sensitive Indicators*, CIDA: Hull

[2] Muhammad F.Ibrahim & Ng C.Wee (2002), The Importance of Entertainment is Shopping Center Experience: Evidence from Singapore, Journal of Real Estate Portfolio Management, 8, 239-254

[3] Omer Torlak, Cevahir Uzkurt & Mujdat Ozmen (2010), Dimensions of service quality in grocery retailing: a case from turkey [Electronic version], Management Research Review, 33, 413-8269

[4] Garvin, Alana Nicole, "Experiential retailing: extraordinary store environments and purchase behavior " (2009). Masters Theses and Doctoral Dissertations, Paper 226

[5] Paul w. Ballantine, Richard Jack, Andrew G.Parsons (2010), The Journal of international Journal of Retail & Distribution Management, 38, 8, 641-653

[6] BLACKWELL, D'SOUZA, TAGHIAN, MINIARD, ENGEL (2006), Purchasing, Consumer Behaviour (1st ed), Australia, Thomson

[7] J. Paul Peter, Jerry C. Olson (2001), Consumer Behaviour's Role in Marketing Strategy, Introduction to Consumer Behaviour and Marketing Strategy (6th ed), United States, Mc Graw Hill

[8] Marilyn K. Pelosi, Theresa M.Sandifer (2000), What Managers Should Know about Statistics, Doing Statistics for Business (1st ed), New York, Wiley

[9] Statistical Method and Interpretation I, Statistics for Business and Economics, 2007 retrieved on 18 April 2011 from http://bss.sfsu.edu/sylee/chap01.pdf

[10] Anup Cheriyan, Solving Insurance Business Problems Using Statistical Methods, Ibexi Soltions retrieved on 4 March 2010 from http://www.ibexi.com/papers/StatisticalMethods.pdf

[11] Definition of Inference retrieved on 18 April 2011 from http://www.wordreference.com/definition/inference

[12] DeCaro, S. A. (2003), A student's guide to the conceptual side of inferential statistics, retrieved on 10 April 2011 from http://psychology.sdecnet.com/stathelp.htm

[13] EMRA (1998) derived on 21 April 2011 from https://www.indiana.edu/~istr561/knuth/pdfs/17_Inferential_Statistics.pdf

[14] Tim France, Dave Yen, Jyun-Cheng Wang & Chia-Ming Chang (2002), Integrating search engine with data mining for customer-oriented information search [Electronic version], Information Management & Computer Security, 10, 242 -254

[15] Sandeep Mitta (2008), Market Basket Analysis, Monday, 04 August 2008 retrieved on 28 February 2011 from http://www.cartesianconsulting.com/insights/index2.php?option=com_content&do_pdf=1&id=122

[16] Data Mining – How to Use Data Mining to Analyze Your Customers (n.d.) retrieved on 3 March 201 from http://www.exforsys.com/tutorials/data-mining/how-to-use-data-mining-to-analyze-your-customers.html

[17] What is Data Mining retrieved on 7 March 2011from http://www.ibm.com/developerworks/opensource/library/os-weka1/index.html

[18] Data Mining: What is Data Mining? retrieved on 28 February 2011 from http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm

[19] Michael J.A. Berry & Gordon S.Linoff (2004), Data Mining Techniques (2nd ed), United States : Willey

[20] Association Rules retrieved on 3 March 201 from http://www.dbtechnet.org/labs/kdd_lab/AssociationRules.pdf

[21] Pirjo Moen, Data Mining Methods (2005) retrieved on 3 March 2011 from http://www.cs.helsinki.fi/u/ronkaine/tilome/luentomateriaali/TiLoMe-140305.pdf

[22] Varpu Uotila & Patrik Skogster (2007), Space Management in a DIY store analysing consumer shopping paths with data-tracking devices, Facilities, 25, 363-374

[23] Data Mining – How to Visualize Data Mining Results retrieved on 3 March 201 from http://www.exforsys.com/tutorials/data-mining/how-to-visualize-data-mining-results.html

[24] Rajesh Natarajan & B Shekar (2005), Issues relevant to e-commerce, Interestingness of association rules in data mining, Research Paper on Sadhana, 30, 291 – 309