

The Implementation of Data Mining Techniques for Sales Analysis using Daily Sales Data



Norulhidayah Isa¹, Nur Syuhada Mohd Yusof², Muhammad Atif Ramlan³

Universiti Teknologi MARA,

Faculty of Computer and Mathematical Sciences

Terengganu Branch, Kuala Terengganu Campus, Malaysia,

¹norul955@uitm.edu.my, ²syuhada.yus123@gmail.com, ³atif@uitm.edu.my

ABSTRACT

It is important to every business to know their level of market competition, for example customers demand, customers' pattern of buying and their sales performance. This project proposes a sales analysis using prediction and association technique in sales data. Chan Furniture and Electrical Appliances Company have been selected as a case study. From the preliminary study, the researcher found that the Company is facing fluctuating amount of sales throughout the year. The sales obtained are lower than expected. The prediction model is used to predict the product that will be sold for every month. The association model is used to identify the associated products purchased by the customer. Purchasing transactions data is taken from the sales invoices issued by the company. The result obtained from both is used in developing a business strategy for Chan furniture.

Key words: Sales analysis, Prediction, Association, CRISP-DM.

1. INTRODUCTION

The term Business Intelligence (BI) refers to technologies, applications and practices for the collection, integration, analysis, and presentation of business information. The purpose of Business Intelligence is to support better business decision making. Essentially, Business Intelligence systems are data-driven Decision Support Systems (DSS). One of the analysis in BI is sales analysis. Sales analysis is an activity that being done by organization and businesses to study and understand their company business performance.

According to [4], sales analysis report presents the trends that occur in a company's sales volume over time. In its most basic form, a sales analysis report shows whether sales are increasing or declining. While [5] defined sales analysis as examines sales reports to see what goods and services have and have not sold well. The analysis is used to determine how to stock inventory, how to measure the effectiveness of a sales force, how to set manufacturing capacity and to see how the company is performing against its goals.

A sales analysis is important because it helps the senior management of the company in making decisions for the

company in term of inventory management, marketing activities, schemes offers to be rolled and changes in the manufacturing processes if applicable. Besides that, by doing the sales analysis it also displays the current market trends to the company which help the company in preparing their future business plans.

This paper implements data mining techniques which are association and prediction model to perform sales analysis in daily transactional dataset. Chan Furniture and Electrical Appliances Company has been chosen as a case study. Chan furniture and electrical appliances operates by selling furniture and electrical appliances. This organization has 80 branches and their main store (HQ) is located at Karamuning, Sabah. This project will focus on only one branch which is in Jerantut, Pahang.

2. LITERATURE REVIEW

2.1 Prediction and Association Techniques

There are four main techniques that can be used in data mining which are classification, prediction, association and clustering [3]. In this research, the researcher has focused on two techniques which are prediction and association. Table 1 shows the definition of prediction and association techniques proposed by [3].

Table 1: Definition of prediction and association techniques

Technique	Definition
Prediction	"It is normally referring as the act of telling about the future which is different from a simple guessing activity by taking into account the experiences, other opinions, and also relevant information."
Association	"It is popular technique that being used to discover the interesting relationships between variables of the data sets in the large databases. The most common association rule mining used id is the Market Basket analysis."

2.2 Related Works

In this section, the previous studies on similar topic are reviewed.

(a) Research of Clothing Sales Prediction and Analysis Based on ID3 Decision Tree Algorithm

This research paper is written by [6]. This research paper is regarding the sales prediction based on current clothing fashion. In modern society being fashionable is a trend. There are multiple factors influence apparel and fashion sales, including price, type of cloth, and size, together with the traditional factors like season and material. Then, in order to improve the sales and maintain as little inventory as possible, the sellers need to identify appropriate factors which consistent with the specific condition of their stores. This research paper used decision tree and ID3 algorithm to conduct the prediction analysis regarding the clothing sales. It can be concluded that the main aim of this previous work is to create an effective model which can help the fashion industry sellers to reduce unnecessary loss from the business. This research paper explores on how factors of weather, color and size can influence apparel sales and then makes predictions on future sales.

(b) Sales Prediction Model Using Classification Decision Tree Approach for Small Medium Enterprise Based on Indonesian E-Commerce Data [7].

In this research, data Mining has been employed in this paper to describe the patterns of the market and to predict the potential region of the national market commodities of the e-commerce site. The algorithm that being used in this research paper are decision trees and CHAID. The attributes used in this research paper were price, type of shoes, insurance, product viewer, city of seller, rating of speed, service and accurate has impact on product sales. The researcher found that West Java, Jakarta and East Java were the first three largest location of footwear sales while as for the sales pattern on tokopedia.com it shows that the higher the price the higher the amount of sale generated by the e-commerce website. The decision tree generated by the researcher has eleven nodes with seven terminal nodes and the depth of the classification tree is three.

(c) Market Basket Analysis of Beauty Products by [8]

The main purpose of this research was to see how different products in a beauty shop assortment interrelate and how to exploit these relations in marketing activities. Association rule mining has been conducted by researcher using apriori algorithm. As a result, the researcher found that some of the shoppers may purchase a single product during a shopping trip by randomly pick while others buy more than one product on the same time with their own target of use for efficiency reasons. The support and confidence value used in this research were 0.1 and 0.4 respectively. Support and confidence is a matrix used to evaluate generated associated rule as discussed in [12]

(d) Analyze Market Basket Data Using FP-Growth and

Apriori Algorithm by [9].

For this case study the researchers had used the data from an electronic store. From the data they came out with the variety of association rules and the algorithm that being used in the study were FP-Growth and also Apriori algorithm. It was also stated that different algorithm used needs different amount of times to generate all the rules. In this study the researchers had used few sets of support and confidence values.

(e) Analyzing Customer Baskets [10].

In this research, the researcher mainly focusses on the activity of association rule mining for customer basket and in discovering the frequent item sets whose items have the highest correlations. This study only uses one set of support and confidence values which were 0.01 for its support and 0.7 for the confidence.

(f) Sales Analysis Using Product Rating in Data Mining Techniques [11].

In this paper, sales analysis can provide a frequent pattern of same type of product sales which is like market basket analysis within same type of products or goods and by representing the data related with the project graphically. So, that they can easily and quickly view type of customer. This step is being done by the researchers to aid them to target the customers which will increases sales and profit of organization. This journal stated that the researchers had used the apriori algorithm to conduct the research project. The aim of the project was to identify the products rating by pre-processing the sales data with minimum threshold using Association rule by using the apriori algorithm and then rank the product.

3. METHODOLOGY

Business Intelligence framework (figure 1) as proposed by [3] has been adapted in this research.

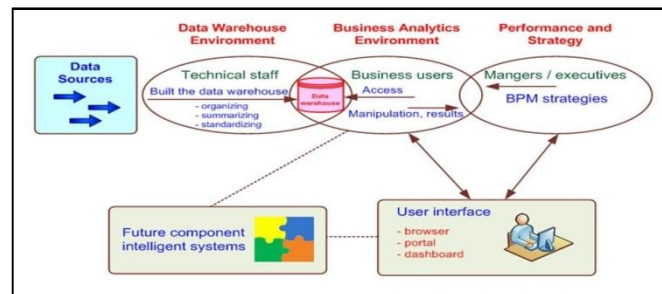


Figure 1: Business Intelligence Framework [3]

There are five main components in the business intelligence framework, which are data sources, data warehouse environment, business analytics environment, performance and strategy and user interface. For the data sources and data warehouse, attributes of the data set were collected and stored.

The main component is business analytics environment where the manipulation of the data happens in this environment. Data mining process was implemented in this environment. Cross Industry Standard Process for Data Mining (CRISP-DM) has been used to carry out the data mining.

Next is performance and strategy component. Result from business analytics environment will be analyzed and useful information will be extracted to propose the business and sales strategy. Figure 2 shows the adapted Business Intelligence Framework.

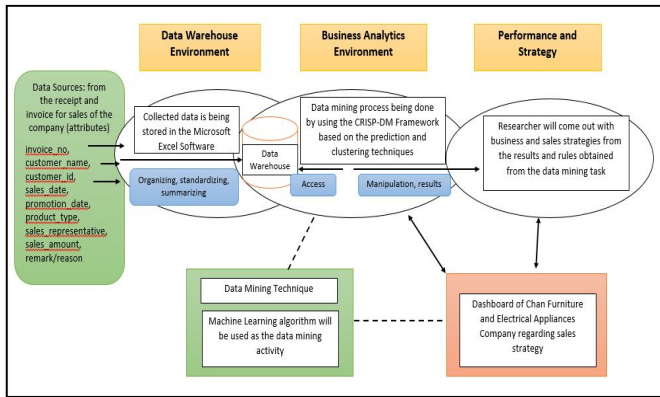


Figure 2: Adapted Business Intelligence Framework

3.1 CRISP – DM

CRISP-DM was used in conducting the data mining process. This framework splits the overall data mining project processes into six phases, and it allows the researchers for needing to go back and forth within the phases between a different stage. Figure 3 shows the CRISP – DM framework.

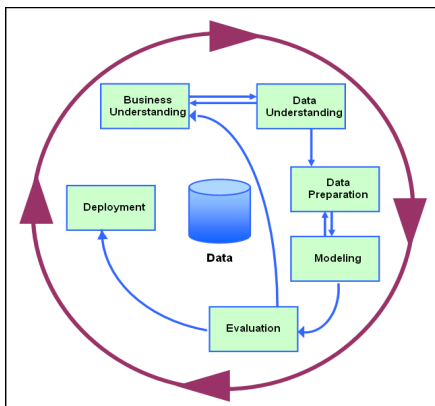


Figure 3: CRISP-DM

3.1.1 Business Understanding

During this phase, the current process, and problem faced by the company will be identified. An interview session is conducted with the staffs of the company especially the manager of the store which is Encik Suhaimi Bin Salleh. The interview session was conducted on August 2018. It is important to truly understand the processes that the company

undergo in handling their ‘sales’ business operation in order to get a clear view about the situation of the company which later on can affect the process in executing the possible solution that can aid in solving the problem of the company.

3.1.2 Data Understanding

The data was collected via the process of interviewing the staff of Chan Furniture and Electrical Company and the data collected is from the year of 2014 until 2017. The data collected is in the form of the hardcopy of sales invoices. Nine attributes have been extracted from the invoices which are invoice no, customer name, customer id, sales date, promotion date, product type, sales representative, sales amount, and remark/reason. Table 2 shows the metadata.

Table 2: Metadata

Attribute Name	Attribute Type	Attribute Format	Description
Invoice no	Integer	E.g.: 57*173260	Represent the invoice series number
Customer Name	Polynomial	E.g.: - 'FAUZI'	Represent the customer name
Customer ID	Integer	E.g.: 860923062441	Represent the customer identification card number
Sales_ date	Date	DD/MM/YY	Represent the date of an invoice
Product type	Polynomial	Eg:- -sofa set -tv	Represent the products or goods types
Sales_ amount	Integer	Eg:- '2198'	Represent the amount of sales

3.1.3 Data Preparation

In this phase the datasets are being prepared for the data mining process. There are four steps in the data pre-processing which are data i) consolidation, ii) data cleaning, iii) data transformation and iv)reduction. For data consolidation, the relevant data that used in this project is collected from the identified sources which means for this project, the data is being collected from the Chan Furniture Company. During data cleaning there were few subtasks that being done by the researcher which are handling missing values, identifying and reducing noise in the dataset and finding and eliminating erroneous data. Data transformation step was mainly to transform the data from the original form into new form that was necessary for the data mining purpose. In this project, the attribute that being transformed was the ‘items’ and ‘date’. For ‘items’ attribute the researcher had converted the text form of the data into polynomial form while for ‘date’ researcher had separate the date information into three which are day, month and year. Figure 4 shows the data

before being transformed while Figure 5 shows the data transformation results.

DATE	PURCHASE ID	ITEMS	CATEGORY
18/1/2014	1	home theatre	Electrical
18/1/2014	2	sofa set	Furniture
19/1/2014	3	bed set	Furniture
20/1/2014	4	tv	Electrical
21/1/2014	5	mattress	Furniture
11/2/2014	6	bed only	Furniture
10/11/2014	7	wardrobe	Furniture
10/11/2014	7	mattress	Furniture
20/2/2014	8	fridge	Electrical

Figure 4: Data before being transformed

Month	Year	PURCHASE ID	ITEMS	CATEGORY
January	2014	1	home_t	Electrical
January	2014	2	sof_set	Furniture
January	2014	3	bd_set	Furniture
January	2014	4	tv	Electrical
January	2014	5	mattress	Furniture
February	2014	6	bd_only	Furniture
November	2014	7	wardrobe	Furniture
November	2014	7	mattress	Furniture
February	2014	8	fridge	Electrical

Figure 5: Data transformation results

3.1.4 Modelling

There are number of data mining tools available in the market as discussed [13]. In this research, Rapid Miner has been used for the mining process since it provides better visualization for the process [13]. The process covers from data preparation process until generating the rules and result for the prediction and association model. Decision Tree and Performance operator used for the highest items sold prediction results while FP-Growth and create association rule operator is used for frequent item set mining process.

Prediction

Prediction is generally referred to as the action of telling about the future consequences of related activities. There are two categories that lies in prediction which are classification and regression. Classification is used in this research with decision tree algorithm with the objective to find the items most likely to be bought by the customers for every month together with the quantity. Rapid Miner software is being used to develop this prediction model. The dataset used in this modelling process had been divided into two where 80% from the rows of data were the training data and the rest 20% is testing data. Decision tree algorithm is popular being used for doing some simple predicting task. A decision tree model is a tree which look like a collection of nodes that intended to create a determination on the values affiliation to a class or in other words an estimate of numerical target value. Figure 6 shows the process flow for the model building of prediction technique using the Rapid Miner.

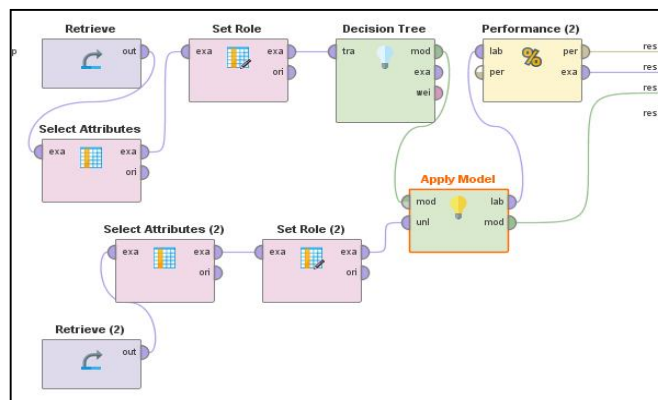


Figure 6: The process flow of the data modelling process for prediction technique using the Rapid Miner software

Association

For association technique, the FP-Growth algorithm was being used to trace out the trends of buying patterns of the customers of the Chan Furniture and Electrical Appliances Company. In order to use the FP-Growth algorithm for this project, the 'FP-Growth' operator of the Rapid Miner software was used. This operator had performed the data associations' process using the algorithm. Several sets of support and confidence had been used in this project in order to get variety of results from the association between the items bought by the customers. Figure 7 shows the process flow of the data modelling process for association technique using the Rapid Miner software.

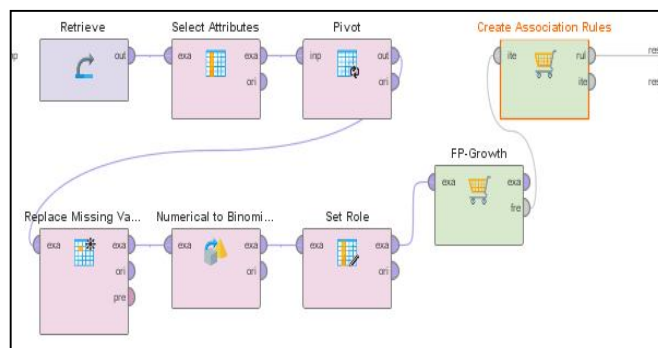


Figure 7: The process flow of the data modelling process for association technique using the Rapid Miner software

3.1.5 Evaluation

To test the developed prediction model, the accuracy of the model is generated by using the Rapid Miner software. The accuracy level also depends on the condition of the data that being used. To test the developed association model, three minimum support and minimum confidence value were used as shown in Table 3.

Table 3: Minimum Support and Confidence

Ref	Minimum Support	Minimum Confidence
[8]	0.1	0.4
[10]	0.01	0.7
-	0.01	0.2

3.1.6 Deployment

In this phase the results and rules from the obtained data mining task is interpreted into a proper knowledge and information. The information then is used in creating the dashboard interface. Figure 8 shows the sample of the dashboard interface.



Figure 8: Dashboard interface

4. RESULT AND DISCUSSION

From the prediction and association modelling process rules and results, researcher come out with a set of information that is useful in proposing the new business strategy to the company later. The results from the predicted highest items sold for every month is tabulated based on each category as shown in Table 3,4 and 5.

As for the association modelling result three experiment have been conducted which produce different sets of association rules. Different sets of support and confident values was used to the same dataset for this experiment.

Table 4: Predicted sales items (category electrical)

Month	Electrical	
	Items	Quantity
January	Washer	4
February	Washer	4
March	TV	6
April	Fridge	7
May	TV	5
June	Washer	6
July	TV	5
August	Fridge	6
September	TV	4
October	Washer	4
November	TV	9
December	Washer	4

Table 5: Predicted sales items (category electronic)

Month	Electronic	
	Items	Quantity
January	Laptop	1
February	Smartphone	1
March	Smartphone	2
April	Smartphone	1
May	Smartphone	1
June	Smartphone	3
July	Laptop	2
August	Laptop	2
September	Smartphone	2
October	Smartphone	2
November	Smartphone	2
December	Smartphone	1

Table 6: Predicted sales items (category furniture)

Month	Furniture	
	Items	Quantity
January	Wardrobe	3
February	Sofa Set	3
March	Kitchen Cabinet	3
April	Bedroom Set	4
May	Sofa Set	6
June	Mattress	5
July	Sofa Set	5
August	Bed Set	3
September	Wardrobe	2
October	Sofa Set	3
November	Sofa Set	3
December	Bedroom Set	3

Experiment 1 of Association Modelling

For experiment 1, the value for support and confidence used are 0.1 and 0.4 respectively. Only one rules generated from this experiment. Table 7 and 8 shows the rule generated in experiment 1.

Table 7: Rule generated in experiment 1

Premises	Conclusion	Support	Confidence
Mattress	Bedroom Set	0.026	0.474

Table 8: Rule generated in experiment 2

Premises	Conclusion	Support	Confidence
Speaker	TV	0.003	1
Sofa Set, Gas Cabinet	Dining Set	0.003	1
Dining Set	Kitchen Cabinet	0.003	1
Display Cabinet, Coffee Table	Sofa Set	0.003	1
Sofa Set, Dining Set, Kitchen Cabinet	Gas Cabinet	0.003	1

Experiment 2 of Association Modelling

For experiment 2, the value for support and confidence used are 0.01 and 0.7 respectively. There are several rules generated from this experiment. Table 7 shows the top 5 rules generated in experiment 2.

Experiment 3 of Association Modelling

For experiment 3, the set of values for support and confidence used are 0.01 and 0.2 respectively. There are several rules generated from this experiment. The Table 9 below shows the top 5 rules generated in experiment 3.

Table 9: The top 5 rules generated in experiment 3

Premises	Conclusion	Support	Confidence
Mattress	Bedroom Set	0.026	0.474
Home Theatre	TV	0.017	0.667
Coffee Table	Sofa Set	0.011	0.667
Kitchen Cabinet	Dining Set	0.010	0.350
TV Stand	TV	0.007	0.333

From the three experiment that being done, the researcher gained variety of association rules as the results from the modelling process. For experiment 1 only one association rule that was generated while for experiment 2 there are sum of 12 rows of rules that had been generated. As for experiment 3, there were 42 association rules that had been generated. The researcher have used the rules gained from all the experiment that is suitable and favorable, in order to come out with the business and marketing strategy for the Chan Furniture Company.

5. BUSINESS STRATEGY

In this part the researcher uses the information from the modelling process result and rules to visualize the business strategy that is mean to be proposed to the Chan Furniture and Electrical Appliances Company. The business strategies include the promotions, marketing and also inventory planning strategy. Figure 9 shows example of marketing and promotion strategy visualization.

REFERENCES

- Nerkar, A. D. (2016). **Business Analytics (BA): Core of Business Intelligence (BI)**. International Journal of Advanced Engineering, Management and Science, 2(12).
- Ishikiriya, C. S., Miro, D., & Gomes, C. F. S. (2015). **Text Mining Business Intelligence: a small sample of what words can say**. Procedia Computer Science, 55, 261-267.
<https://doi.org/10.1016/j.procs.2015.07.044>



Figure 9: Marketing and Promotion strategy sample

From the Figure 9 shows the implementation of both information and result from the prediction and association modelling in proposing the business strategy. As for prediction part, it shows the prediction on highest items to be sold by month helps in focusing on which items to be promoted on every month. As for the association modelling part, the researcher used the results and rules to come out with offer or items to be packaged sold with the predicted items based on month. The sides' items to be offered with the main predicted items is gained from the association rules where the items are correlated with each other.

6. CONCLUSION

As a conclusion this paper explained regarding all the activities and processes that being done by the researcher in order to achieve all the objectives that have been set at the early stage of the project. The main objective of this research to predict the items that will be sold the highest for every month and to understand the correlations between the purchased items by the customers. Even there was some limitations for this project, it still can successfully be done according to the current needs and requirements. As for the objectives, all the objectives have been achieved by implementing the business intelligence by the researcher.

- Sharda R., Delen D., & Turban E., (2014). **Business Intelligence and Analytics**. Pearson, Global Edition.
- Akers, H. (2018). **What Is a Sales Analysis Report?**. Retrieved October 11, 2018, from <https://smallbusiness.chron.com/sales-analysis-report-58460.html>
- Reddy N., (2015). **Sales Analysis Techniques**. Retrieved December 1, 2018, from <http://bnreddybi.blogspot.com/2015/10/sales-analysis-techniques.html>

6. Sun, F., Liu, Y., Xurigan, S., & Zhang, Q. (2015, January). **Research of Clothing Sales Prediction and Analysis Based on ID3 Decision Tree Algorithm**. In 2015 International Symposium on Computers & Informatics. Atlantis Press.
<https://doi.org/10.2991/isci-15.2015.42>
7. Alamsyah, A. (2015). **Sales Prediction Model Using Classification Decision Tree Approach for Small Medium Enterprise Based on Indonesia E-Commerce Data**. In Seminar & Conference on Business & Technology in ICT Industry.
8. Gancheva, V. (2013). **Market basket analysis of beauty products**. Erasmus University Rotterdam.
9. Mehay, A., Singh, K., & Sharma, N. (2013). **Analyze market basket data using fp-growth and apriori algorithm**. International Journal on Recent and Innovation Trends in Computing and Communication, 1(9), 693-696.
10. Kouzis-loukas, M. (2014). **Analysing Customer Baskets**. Erasmus School of Economics.
11. Bhagwat, S., Jethliya, V., Pandey, A., & Islam, L. (2015). **Sales analysis using product rating in data mining techniques**. International Journal of Research in Engineering and Technology, 4(02), 2319-1163.
<https://doi.org/10.15623/ijret.2015.0402024>
12. N. Isa, N.A.Kamaruzzaman M.A. Ramlan, N.Mohamed, M.Puteh, **Market Basket Analysis of Customer Buying Patterns at Corm Café**, International Journal of Engineering and Technology, Vol 7, No4.42 (2018).
13. Priti S. Patel and Dr. S.G. Desai, **A Comparative Study on Data Mining Tools**, International Journal of Advanced Trends in Computer Science and Engineering, 4(2), March - April 2015, 28 – 30.