Performance Based Market Basket Analysis for
an Improvement of Customer Relationship Management

Akkapop Prasompon

640631128

Introduction:

In the early time of modern trading of mankind, most of the trading activity will focus on the process of how to maximize the capacity of supply chain in order to produce the highest quantity of product and deliver to the client in the shortest period of time. This period is known as the "product Era" and "Sales Era". After a period of time, the trends of the world tend to shift towards the trends which pay more attention on the customer need and this era is known as the "Marketing Era" [1]. This trend of focusing on the need of customer has created the marketing strategy known as CRM (Customer Relationship Management). CRM is the strategy that not only connect the seller with the customer, but it can also be considered as a tool that capable of analyze the behavior of the customer for further business marketing development.

In the present period of time there are countless number of the software to help with the marketing solution for the business which cover all area of marketing area, one area that is significantly important is the customer relationship management and the example of the system which is available for this task is the software for customer relationship and data analytic. This software has play an important role for the business by helping in the process of collecting the business data which include the sell, shipping and customer preference data and analyze for better business strategy in the real time manner. Moreover, it has also pay important role in the process of minimizing the expenses and mistakes coursing by the company's human resource too [2].

The data analytic system is the system that is capable of study the behavior of the data input by the process of recognize the pattern created by that data for the purpose of predicting the next action that might happen accordingly, in the marketing area the data is the behavior of the human that are collected, and the system will predict the action that is most likely to be performed. In February 2021 the scientist from south Korea name Seunghyun Choi and the fellow scientist has published an article [3] which mention about the application of sequential pattern strategy in order to classify the event that could have taken place from the monitoring process of human being in the limited area and predict the most possible upcoming event by the help of object detection

algorithm. Wherewith the algorithm will simulate the pattern of event that could have happen accordingly which related to the data selected.

There are many studies related to the technology that apply the knowledge about customer behavior prediction, for example in 2013 scientist name Laura Badea from Bucharest University has applied Artificial Neural Networks (ANN) in the study [4] in order to detect the character of individual which is highly possible to have chance to deposit money from the information which includes race, gender, educational background, age and wage. And the result of the algorithm came out were considered to be positively promising with 73 to 78 percentage of accuracy rate of prediction. Even though this algorithm were able to perform the task of prediction in the rate that could considered to be in a good performance, but this technology contains some downturn that has to be considerate too which include the time consuming and or the higher chance that could leads to the overfitting scenario compared to the other method available.

One of the simplest method that is available nowadays that is wildly applied in the CRM technique in order to identify the relationship between each type of product that have high possibility to be purchased analyzed from the products that is already put in the basket is called market basket analytic [5]. This technique is process with the help of the association rule algorithm. This technique is considered to be an efficient tool for the organization for analyzing the purchasing behavior of the customer for further marketing strategy. Market basket analytic can be applied via various tools and countless analyzing method according to the situation which could leads to the different level of accuracy of the result according to the data set.

Literature Review:

Customer Relationship Management (CRM) began from the concept of Relationship Marketing in the late 20th century which considers the customer as a part of marketing enhancement. Companies must change their marketing strategy from "Short-term Transaction-oriented goal" which focus on maximizing sale amount to "Long-term Relationship-building goal" which straight forward to creating long-term relationship with their customers, Professor Philip Kotler from Northwestern University talked about a new type of marketing [6]. CRM is defined as an infrastructure that needs to increase the customer's value to convince and motivate the valuable customer to remain loyalty with the brand and come back to purchase our products again and again [7].

Nowadays, technologies have been involved in CRM in the aspect of collecting and analyzing data to establish the relationship between company and customer. The use of analytic tools to gain the hidden information of their customer and improve their relationship management strategies has been widely talked about in the academic perspective, data mining is one of the various approaches that influenced CRM analysis, according to the study [8], applications of data mining consist of (1) detecting the hidden pattern from existing data, (2) extract the pattern from existing data for predicts a future event, and (3) using extracted data to find unusual data elements. The article also mentioned analytic approaches of using data mining in different situations such as fraud detection for a banking system, customer churn analysis for telecommunication service, or customer management, they also mentioned data mining approaches for retail marketing which is a topic of this study that the data mining technique is capable of building a market basket analysis, sale forecasting, database marketing, and merchandise planning. In this study, we aim to use data mining tools to analyze an association rule by market basket analysis (MBA) from the market product transaction data.

MBA is the methodology to analyze the pattern of buying of a product/products can influence the buying of another product, the relationship between these set of products will be represented in the form of association rule [9], those set of products that associated between each other will be defined as itemset, and the accuracy of the model will be represented by confidence value. The study of Roshan Gangurde and team [10] had implemented an artificial neural network algorithm to performs the market basket analysis, they used Feed Forward Neural Network (FFNN) algorithm to train the transaction data with the java programming language, a number of

nodes in the hidden layer and output layer they used are equal to the summation of combination value from the total number of possible input which ignoring the order of data and repetition is not allowed. The study used four variables which are X1, X2, X3, and X4, they assigned weight for two variables equal to one and the other two variables equal to zero and negative five, respectively. The result of the analysis showed the association between four groups of itemset that might appear in the same market basket, they concluded that their algorithm is able to detect an association rule among sale transaction data, they also claimed that their algorithm has an ability to handle with a data cleanness and continuously challenge. However, this study does not mention the result in a computational time effectiveness aspect when applied to the real-world data. The study by Agus and Triyoso [11] also studied the association rule detection by using a Neural Network, the result showed an association rule of products after executed algorithm. Although the study has identified an epoch, number of data, and efficacy of the model, it still has not mentioned computation time in both result and conclusion.

According to the study [12], they mentioned other algorithms that can be used for finding association rules. This study will demonstrate two algorithms containing Apriori Algorithm which is a basic approach that used for finding an association rule between itemset due to simplicity of implementation, and the secondary, Frequent Pattern Growth (FP-Growth) algorithm which is a concept of finding association rule that has been improved from apriori algorithm. There is a study about customer churn predicting [13] that applied both apriori and FP-growth algorithms to find frequent itemset, the study mentioned these algorithms that an apriori algorithm has more efficient in the generating of the candidate but still has a chance of bottleneck problem from performing a complex data and multiple database scanning, while FP-growth has a better ability to handle this problem. In the last section of the study, they showed the result of customer churn predicting using market basket analysis, but they do not mention about execution time in terms of performance analysis as well.

Methodology:

The objective of this study is to compare the difference in the efficiency perspective and performance perspective between an apriori algorithm and frequent pattern algorithm by using sale transaction data of market or grocery. The data that will be used in this study is separated into two datasets, we use 25,900 transactions of online retail data from UCI Machine Learning Repository [14] to mimic the large size of data, and 9,835 transactions of grocery data from an arules library [15] to mimic the small size of data.

The interpreter software that will be used for creating a model consist of four different tools including (1) RStudio – The IDE for executing an R programming language, suitable for statistical computation, (2) Jupyter Lab – The IDE for executing a python language which is a programming language that widely found in data mining project, (3) Matlab – The computer program that used for mathematical problem analysis, developed from C and C++ programming language, and (4) Weka – The software tools that collect machine learning algorithms which is ready for implementing.

In order to standardize the performance of the model evaluation, the device used for running the algorithm will be the same device for every execution. The device in this project will be a computer laptop with the following specification:

- Processor: AMD Ryzen 5 4800H mobile processor 3.0GHz
- Memory: 16Gb DDR4
- Storage: M.2 PCIe SSD
- Graphics: NVDIA GeForce GTX 1650; 4Gb

The result from market basket analysis will be eight sets of pos-processed data consisting of four results from the large dataset and four results from the small data set. The evaluation of the result will consider the performance of data analysis via confidence value and performing time cost for each analysis. All the results will be compared to each other, conclude, and visualized in the form of a table and graph in the last section of the study.

References:

[1] R. J. Keith, The Marketing Revolution, 1960.

[2] M. Sneha and K. Krishna Prasad, "Analysis of Business Strategies of Salesforce.Com Inc.," *IJCSBE,* vol. 2, no. 1, pp. 37-44, 2018.

[3] S. Choi, C. Kim and S. Youm, "Human behavioral pattern analysis-based anomaly detection system in residential space," *The Journal of Supercomputing,* vol. 77, pp. 9248-9265, 2021.

[4] L. M. Badea, "Predicting Consumer Behavior with Artificial Neural Networks," *Procedia Economics and Finance,* vol. 15, pp. 238-246, 2014.

[5] A. Kadlaskar, "A Comprehensive Guide on Market Basket Analysis," 2 October 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-market-basket-analysis/.

[6] Payne, Handbook of CRM: Achieving excellence in customer management, 2006.

[7] J. Dyche, The CRM Handbook: A Business Guide to Customer Relationship Management, 2004.

[8] C. Rygieski and J.-C. Wang, "Data mining techniques for customer relationship management," *Technology in Society,* vol. 24, pp. 483-502, 2002.

[9] B. Lantz, Machine Learning with R, Packt Publishing, 2013.

[10] R. Gangurde and D. B. Kumar, "Optimized Predictive Model using Artificial Neural Network for Market Basket Analysis," *IJCSC,* vol. 9, no. 1, pp. 42-52, 2018.

[11] A. Mansur and T. Kuncoro, "Product Inventory Prediction at Small Medium Enterprise Using Market Basket Analysis Approach - Neural Networks," *Procedia Economics and Finance,* vol. 4, pp. 312-320, 2012.

[12] K. Shaukat, S. Zaheer and I. Nawaz, "Association Rule Mining: An Application Perspective," *International Journal of Computer Science and Innovation,* pp. 29-38, 2015.

[13] M. M. Aung, T. T. Han and S. M. Ko, "Customer Churn Prediction using Association Rule Mining," *IJTSRD,* vol. 3, no. 5, 2019.

[14] D. Chen and S. L. Sain, "Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining," [Online]. Available: https://archive.ics.uci.edu/ml/datasets/online+retail.

[15] M. Hahsler, K. Hornik and T. Reutterer, "Implications of probabilistic data modeling for mining association rules," 2006. [Online]. Available: https://rdrr.io/cran/arules/man/Groceries.html.