INTI JOURNAL | eISSN:2600-7320
Vol.2020:15

Market Basket Analysis for E-Commerce using Association Rule Mining

Kayalvily Tabianan[1*], Sarasvathi Nagalingham[1], Leong Kai Cheng[2]

[1]Centre for Emerging Technologies in Computing (CETC), INTI International University, Nilai, Negeri Sembilan, Malaysia
[2]Faculty of Information Technology, INTI International University, Nilai, Negeri Sembilan, Malaysia.

***Email:** kayalvily.tabianan@newinti.edu.my

Abstract

Nowadays, shopping with ecommerce has become the most common lifestyle for everyone in modern era. In order to make the research to be successful, it requires to discover the best research effort to improve the algorithm. In order to make this research successful, author will need to identify the best algorithm for finding the item sets frequently bough together and top sales product on each country to predict the sales. The author has developed an ecommerce system which has back-end system to display the performance of the product and using Association Rule Mining on the datasets. By using this system, they can know the hidden product relationships which product has the potential to be purchase together. For develop the system author has uses KDD research methodology which can help to extract the minimal support, confidence and lift from the datasets.

Keywords

Market Basket Analysis, E-Commerce, KDD methodology

Introduction

During the Movement Control Order, customers have had to rely on e-commerce platforms to make purchases. So naturally, the e-commerce landscape as a whole is growing. When there are thousand of user complete a transaction, market basket analysis using Association Rule Mining on the datasets, it will finds the association between different objects in a set, find frequent patterns in a transaction database, relational databases or any other information repository (DataCamp Community, 2019).

International Conference on Innovation and Technopreneurship 2020
Submission: 8 July 2020; Acceptance: 3 August 2020

By using the Association Rule Mining, they can obtain three code effectiveness measurements known as help, trust, raise and affinity. Help also implies how much historical data the policy follows, so confidence indicates how sure they are that the law applies. Help can be measured as a percentage of rows representing either A and B or A and B combined likelihood. Next, it also helps the seller to understand there products well, it might be able to perform a lot of market strategy sells by understanding the product user selling. This utilizes the purchasing data and maximize sales and marketing effectiveness. Market Basket Analysis looks for mixtures of items that often happen in transactions and are used prolifically since before the intro of electronic level-of-sale technologies that have enabled huge amounts of data to be collected. Market basket analysis uses only purchases with more than one product, as no single purchase comparisons can be created. The interaction between things does not necessarily suggest a cause and effect, but rather a co-occurrence test.

Methods

Research Methodology

One of the research methodologies be use on the proposed project is Knowledge Discovery Database Process which is known as the process to extract useful data from a huge data source. These methodologies is having a huge range of used by data mining technique. The KDD process's unifying aim is to derive information awareness in the sense of large databases. The cycle of information acquisition is repeated, collaborative and is made up of nine phases. The first phase is to understand the production and awareness of the application client and sets the scene to consider what to do with conversion, algorithms, and representation. The proposed system requirements are defined in as much details as possible in this stage. The second phase is chosen and construct a data set, In this stage, the author has validate the data by determining whether it is needed to be clean or it is contain NULL value and choose among from all the data which will helping through the proposed system. The third phase is pre-processing and maintenance, in this phase, the author starts to cleaning dirty data such as cleaning the null value and eliminating outliers value. The fourth phase is data transformation. In this phase, it involve dimension reduction, such as choice and elimination of features and record testing, and conversion of attributes such as discretization of numerical attributes and functional transformation. The fifth phase is choose a suitable data mining task which is mostly relies on the objectives. In this phase, whereby generalizing from a sufficient number of training instances, a prototype is built explicitly or implicitly. The inductive approach's underlying assumption is that the learned template extends to future cases. The sixth phase is data mining algorithm with the strategy which is discover on previous step and decides with the tactics. This stage involves selecting the specific pattern search method, including multiple inducers. Lastly, this phase is to determine to presenting the research what author done and determining the enhancement with maintenance of the purposes systems.

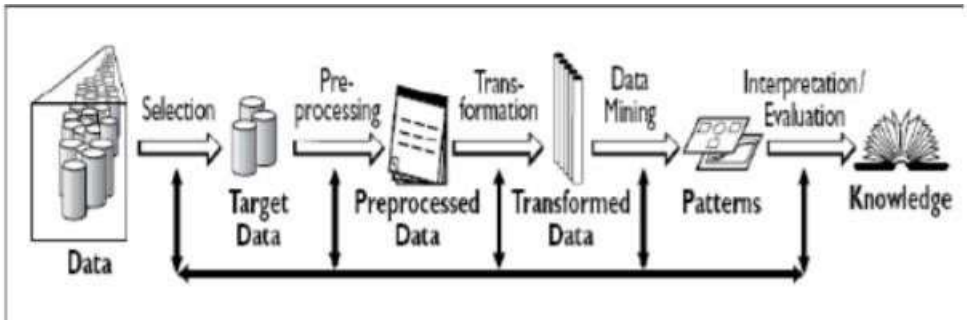

Figure 1. KDD Process

Questionnaire

The questionnaire is a well-established tool within social science research for acquiring information on participant social characteristics, present and past behavior, standards of behavior or attitudes and their beliefs and reasons for action with respect to the topic under investigation (Bulmer, 2004). The proposed system is using google forms to collect the questionnaire data from thirty users which is chosen by author.

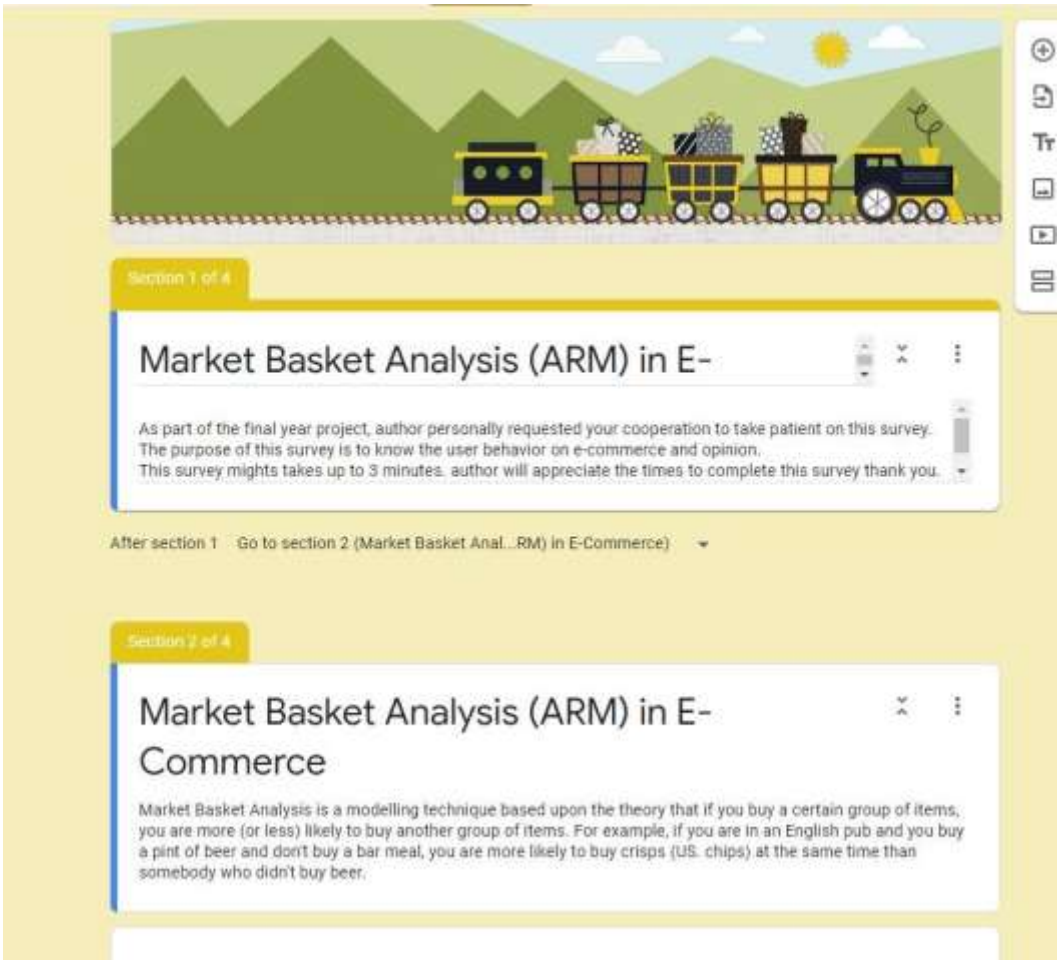

Figure 2. Google forms Questionnaires

Observation

Observation is a way of gathering data by observing, as the name implies. The observation data gathering approach is known as a participatory analysis since, when taking notes and/or documenting, the author must immerse himself in the environment where his respondents are. Author will observe on the result of the questionnaire and understand usually how people do would like to purchase more product usually on eCommerce, so Author will understand what is the best method to perform Association Rule Mining on eCommerce and understand what categories putting together will reach the maximum impact of selling product

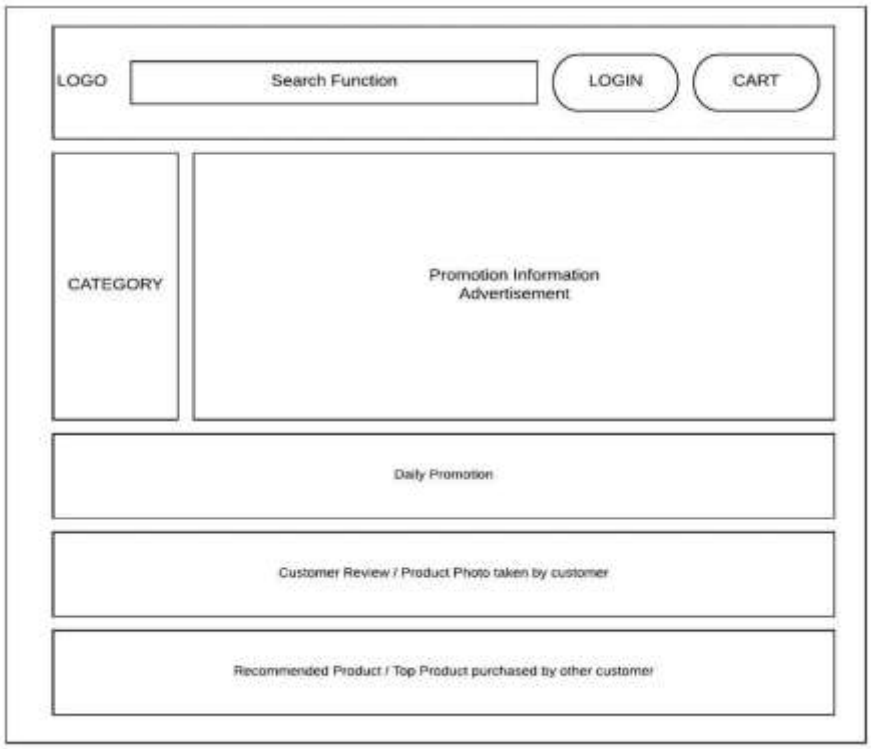

Figure 3. Observation Diagram

Results and Discussions I. Acceptance Testing

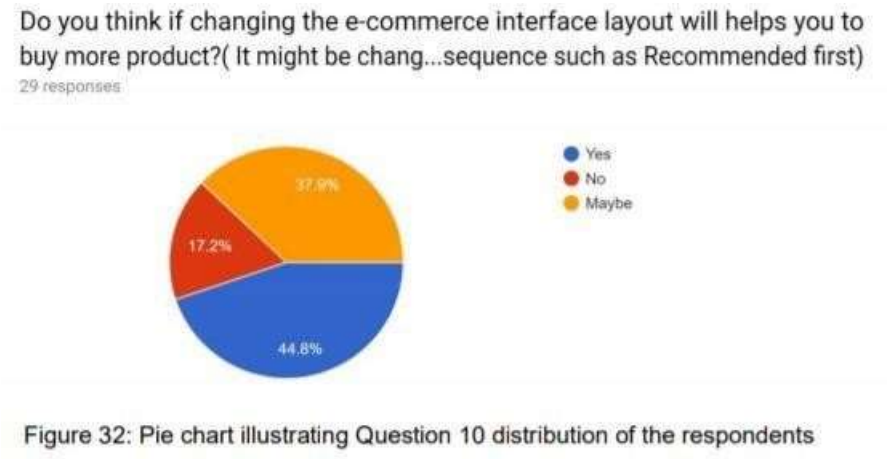

Figure 4. Pie chart to evaluate ecommerce interface design

To identify whether the e-commerce layout will affect how respondent changes their mind while purchasing product.) Based on the survey conducted, 44.8% (13 respondents) of user voted Yes, 37.9% (11 respondents) of user voted No, However, the result of Yes remained significantly higher than for the result No.

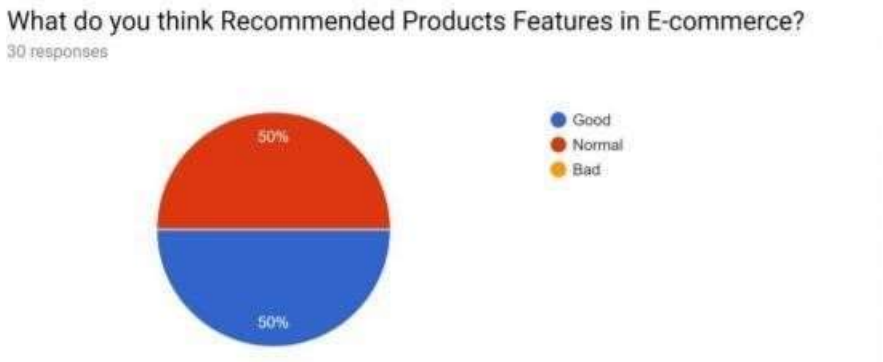

Figure 5. Pie chart to evaluate recommended product features

To clearly get an opinion on the recommended product features. Author is surprised that the recommend products is both 50% on good and normal, which means recommend product still has their own good feedback towards user.

II. Graphical User Interface Design E-commerce User Interface

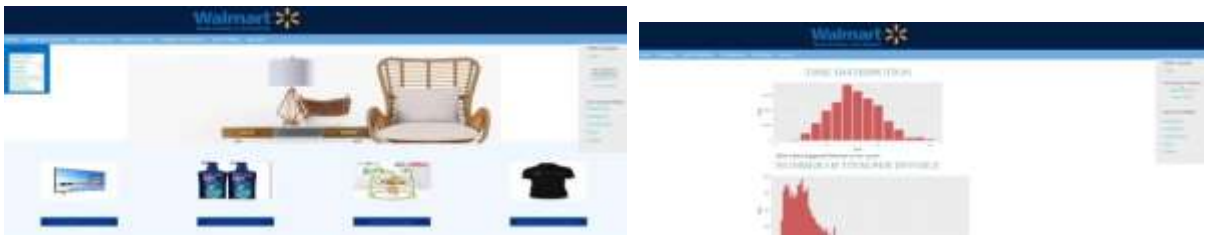

Figure 6. Ecommerce Interface

Firstly, the system will display the basic ecommerce homepage, which will allow user to browse product, category, login, logout, forgot password, check shopping out, checkout and visit the social media of Walmart. Also, for the category of the product it has total of 4 categories which is sorted on the Walmart datasets. Author also allow user to view product without login, but when user wan to purchase product, system will require user to login or register. Author also add up some extra features such as forgot password, it will send the password to user registered email. Furthermore, for the shopping cart, it is saved on MYSQL, so system allow different account to access different shopping cart. Lastly for the checkout features, author only design user can check out via cash or card. Furthermore the E-commerce Admin, , it has total 5 Categories, the first will be the home page of the admin, it shows the shopping time distribution, number of items per invoice distribution and top 10 best sellers, also additionally, it will be providing the CSV files of the customer purchase below.

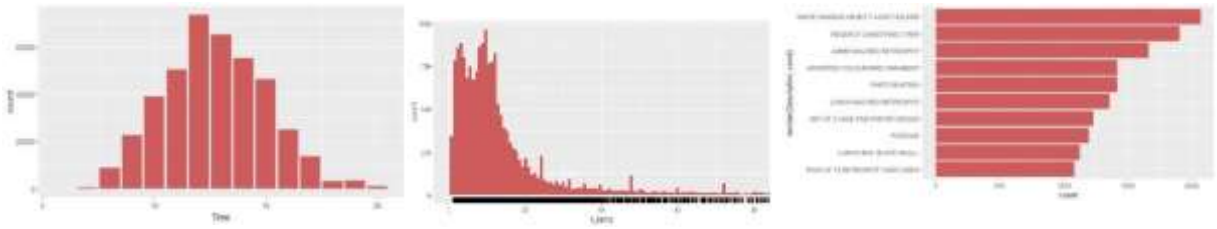

Figure 7. Basic Analysis

For the shopping time distribution, The above figure has clearly showed that most order of the datasets is occurred 10:00-15:00 the time which people often purchase online. This figure is an extra figure that author found it interest, for the number of item per invoice is people often purchase less than 10 times which means less than 10 items in each invoice. Also for the top 10 best sellers will let sellers to understand which product will be the top sales and sellers can prepare for the stocks incase sold out. For the modelling part, it shows the distribution of object based on item matrix which means the transaction or item set absolute item frequency plot we can understand that white hanging heart t-light holder and regency cake stand 3 tier have the top sales so if the seller would like to increase the sales

of set 3 of tins pantry design the seller could put the sells nearby regency cake stand 3 tier, it would help to let the buyer have to view the product and become potential buyer too.

Figure 8. Apriori Algorithm

This categories APRIORI, it is a function which is from a package of R Studio name arules, it is an algorithm to locate repeated itemsets for Boolean association principle in a dataset. The name of the algorithm is APRIORI since it requires advance knowledge of both the characteristics of commonly set products. It will apply an iterative approach which it is known as k frequent itemsets are used to find k+1 itemsets

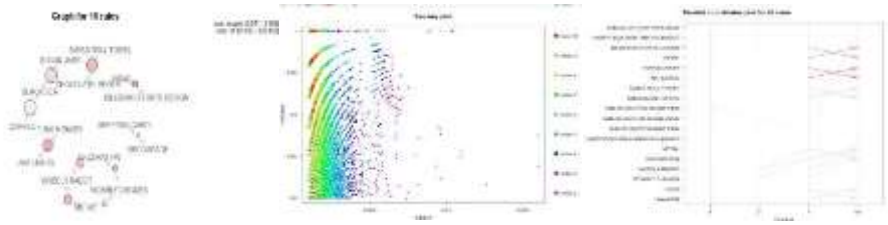

Figure 9. Market Basket Analysis with Association Rule Mining

This analysis is the majority part of the proposed systems which is using support, confidence and lift to find out the market basket analysis. Before this process, author already done filter, cleaning the data and make the product data to data frame and creating some rules which is using the apriori algorithm in Arules library to mine frequent itemsets and association rules. Author setup that support must =0.001 and confident =0.8 to return all the rules that have a support of at least 0.1% and confidence of at least 80%. It diagrams easily to shows that the product relationship with each other.

Conclusion

The proposed system described in this paper has been successfully designed and tested by author, the research part is to investigate what objective author will achieve and what requirement and the concept of the market basket analysis with association rule mining. The author has faced a lot of challenges during the implementation stage, however author solved it one by one by doing more research, asking lectures and friends. The most important objective is that author learned the skills of analyzing and solving the unexpected situation such as facing error. After completing this project author has different vision to see the way to achieve solution with more various vision. In conclusion, author has learned a lot of knowledge on this project and achieved main objective of the project and with some extra features that would make the project way more complicated and friendly

References

Techopedia.com. (2019, October 20). What is Market Basket Analysis? - Definition from Techopedia. Retrieved from technopedia.com: https://www.techopedia.com/definition/32063/market-basketanalysis

DataCamp Community. (2019, October 19). Market Basket Analysis using R – DataCamp.com. Retrieved from Datacamp.com: https://www.datacamp.com/community/tutorials/market-basketanalysis-r

Medium. (2019, October 21). Association Rule Mining. – Medium.

Retrieved from Towards Data Sciences: https://towardsdatascience.com/association-rulemining-be4122fc1793

Nation, E. and Nation, E. (2019, October 22). The Evolution of E-Commerce, where to next? • ECN | E-Commerce Nation. - ECN | E-Commerce Nation. Retrieve from Nation, E: https://www.ecommerce-nation.com/evolution-e-commerce-where-to-next/

Demaitre, E. (2019, October 22). Mobile robots in e-commerce: Experts weigh in on applications, demand. - The Robot Report. Retrieved from therobotreport.com: https://www.therobotreport.com/mobile-robots-ecommerce

Rai, A. (2019, October 23). Association Rule Mining: An Overview and its Applications. upGrad blog. Retrieved from upgrad.com:https://www.upgrad.com/blog/associationrulemining-anoverview-and-its-applications/