International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 3, No. 3, October 2014, pp. 171~177
ISSN: 2252-8776

Decision Support System using Data Mining Method for a Cross Selling Strategy in Retail Stores

Imam Tahyudin, Mohammad Imron, Siti Alvi Solikhatin
Departmen of Information System, STMIK AMIKOM Purwokerto

Article Info

Article history:

Received Jul 18, 2014
Revised Sep 21, 2014
Accepted Sepg 22, 2014

Keyword:

Apriori algorithm
Association
Cross selling
Data mining
Retail stores

ABSTRACT

A sales transaction dataof a retail company which is collect edevery day is enormous. Very large data will bemore meaning fultoin crease the company's profitsif itcanbe extracted properly. Based on the research resultsof Andhika, et al[1], ZhangandRuan[6], Herera et al [7], Witten [11], explained that one of the methods that can gather information from the transaction data is the method of association. With this method it can be determined the patterns of transactions performed simultaneously and repeatedly. Thus, it can be obtained amodel that can be used as a reference for cross selling sales strategy. The purpose of this research is to apply data mining association methods of data mining by using apriori algorithm to create a new sales strategy for cross selling. Based on calculations, Association Rule is implemented by applying Confidence value=0.8while the value of Support=0.1 of the defined minimum value, the total result are 77 rules.

Corresponding Author:

Imam Tahyudin
Departmen of Information System
STMIK AMIKOM Purwokerto
Purwokerto, Indonesia.
Email: imam.tahyudin@amikompurwokerto.ac.id

1. INTRODUCTION

Decision Support System (DSS) began to attract attention among programmers and systems analysts. These systems assist decision-making by managing data and using certain models to solve the problem. Decisionsupport systems to be special because it is able tosolv ethe problem of un structured or semi-structured. In many areas, a decision support system has many perceived benefits and dependency to use the system in creasing with the increasing complexity of the data management proces seach information system.

Application of the DSS to a business interest plays a very important role, such as to provide advice in the preparation of cross-selling strategy atretail stores. One method that can be used to solve the seproblemsis by using association method of apriori algorithms. The study on the application of the method of Apriori Association has been done by previous studies with a variety of objects.

Research conducted by Andhika et al [1] entitled Excavation Association Patterns in Data Warehouse Agent Manufacturing Using Microsoft SQL Server (Case Study: PTXYZ) aimstocreatean integrated system, using the data warehouse and association method of rule mining, that found a pattern of sales transaction data from previous periods regarding the relationship between-a variable that is knowntendency of the product to be purchased by the customer in conjunction with the specific product. The method used is the design ofthe data, starting from the formulation of the problems encountered, and then do a search the required data. Once collected, the data is filtered and transformed that into a form consistent database. Further more, applying association rule mining using Microsoft Visual Studio. The end result is an

application that can provide cross product recommendations-selling of the process of extracting the association of the sales data in XYZ, so it can help XYZ perform data management and decision-making.

Subsequent research conducted by Dongwoon, etal. [4] with the title of High-utility rule mining for cross selling. The objective of this study aims to determine the results of the excavation cross selling with HURM method (high-utility rule mining) to develop three elements ofthe target, effectiveness and opportunities. Average number of transactions carried out buy six items and products. Further research conducted by Lutfi[14], entitled Application of Data Mining Association Algorithm for Improving Sales. This studyaims to determine the groups of items of goods which hare consistently purchased simultaneously. The information obtainedis used for promotions, store layou tsettings to put the goods items in an optimal relation to each other, designing catalogs, and identify customer segments based on the pattern. The method used in this research is to find all frequentitem set, then bring up the strong association rule from frequentitem set. The final result sare expected from the built system that has the ability to see patterns of sales of goods that can then be used to develop new sales strategies.

The next study was conducted by Zhang and Ruan[3]titled modification of association algorithm with its application in the cross-selling strategy in the retail industry. The purpose of this study is to modify the apriori association algorithm by reducing the scale of the can didateitem set Ckand the input output. Based on the result sobtained show that the modified algorithm can improve the performance of the apriori association algorithm efficiently.

The next research was conductedby Tang, etal[6]with the title the use of data mining to accelerate cross-selling. The purposeof this study was to determine the pattern of cross-selling sales to be taken into considerationto make sales acceleration strategy. The method used apriori association algorithm method with XL Miner software. The results showed that by using the parameters of minimum support and minimum confidence. Both of these parameter suseful and in fluential seriously.

The next research was conducted by Yang, etal. [15] with the title the use of decision tree and association algorithm for predictive cross selling opportunities. The purpose of this study is to predict the cross selling opportunities with in novative approaches effectively. The method used is the method of decision tree and association algorithms. The result sobtained show that the approach can improveprediction accuracy and helps telecomvend ors in making policy for cross selling.

Research conducted byYusuf, et al[13] with the title ofthe application of data mining in the determination of association rules be tweentypes of items. The study aimed to determine the association between the type of product, the types of products that appear the same on every transaction so that the transaction data is an important input in making efforts to increase the sales. The method used is the association method with the apriori algorithm. The result sobtained show that the sales transaction data gives three rules that meet the 80% confidence limit.

2. ASSOCIATION METHOD AND APRIORI ALGORITHM

2.1. Association Rule Mining

Association rule mining is a method used to determine the general patterns and repetitions in a set of transactions in large amounts. Association rule studied the frequency of a number of items tha to ccur together in a transaction database based on two me asures called support and confidence. Both ofthese measuresto identifythe occurrence and association rules from the item set. The formation of association rules on item set if the support and confidence values greater than the minimum support and confidence specified by the analyst [5].

Association rule canbe use donone or more than one data dimension. If it is in one-dimensional, association rules tha to ccur only in volves one-dimensional logical data from multiple dimensions of data in data warehouses and datam arts. Inmulti dimensional association rules that occurinvolving more than one dimension of the logical data from multiple dimensions of data in data warehouses and data marts.

A procedure is to look for relationships between items in aspecified data set [9]. Association Rule Mining includes two stages:

a. Looking for the most common combination of an item set (frequent itemset).
b. Generate the Association Rule of frequent item set that has been made before.

Generally there are two measures of confidence (interestingness measure) used in deter mining anassociation rule, namely the Support and Confidence [9].

2.2. Apriori Algorithm for Finding Frequent Item Sets

Apriori algorithm is an efficient method for selecting strong rules contained in the transaction group[10]. The first phase of the algorithm generates frequent item set appears in a systematic and robust second phase generates rules from the item set.

An Association rule can be explained as follows: Ois a setof items where O={o1, o2, ..., on}. Tiis thei-the transaction that contains a set of items. Dis the set of all transactions so that D={T1, T2, ..., Tm}. Association rule to generate will bes haped following implications:
"If A, then B" or"A ⇒B"

A is the antecedent (predecessor) of the implications, while Bis the consequent (follower) of the implications. A and Barepuresub sets of I so that A, B⊂I.A and Baretwodisjointsetsso A∩B=∅.

There are twosizesin deter mining whe the rapair of item scan be expressed as anassociation rules. This size is expressed as support and confidence.

a. Supportis a requirement on how of tenan/a set of items mustappear to be expressed as a rule. Support denoted bys. supp{A ⇒B)} =f(A ∪B) numberofTminD
b. Confidence shows the level of confidence predecessor items (antecedents) and a follower items (consequent) will appearin the same transaction. Confidenceis denoted byp. conf{A ⇒ B} = $\boldsymbol{ff}$($\boldsymbol{AA}$∪$\boldsymbol{BB}$) $\boldsymbol{ff}$($\boldsymbol{AA}$)

Item set is a set comprising some or all ofthe items that are members of I.Anitem set consisting of kitemsis calleda k-item set. A frequent item set (frequent item set) is an item set which has a frequency of numbers φ. Frequent item set which has kelements is called ak-item set (frequent k-itemset).

In addition tothe sizeof the support and confidence of an item set is frequent, the third measure that can be considered is the value of the lift. Lift size is deter mined as follows:

$$\boldsymbol{ll} = \text{lift}\{A \Rightarrow B\} = \boldsymbol{ff}(\boldsymbol{AA} \cup \boldsymbol{BB})\ \boldsymbol{ff}(\boldsymbol{AA})\ \boldsymbol{ff}(\boldsymbol{BB})$$

Lift Value illustrates the following points:

a. If the value of the lift<1, then A and Bhave the same low frequency of occurrence in the data as expected based on the independent as sumption. In other words, A and Bhave a negative dependence and the influence of substitution betweenAandB.
b. If the value of the lift=1, then A and B at the same frequency of occurrence is frequent in the dataas expectedbased on the independent as sumption. A and B can be said tobe independent from one another.
c. If the value of the lift>1, then A and B at the same frequency of occurrence of more frequent data as expectedbased on the independent as sumption. In other words, A and B are positive inter dependence, and there is a complementary effect between A and B.

Lift is calculated only for the 2-itemset because the lift value tends to be higher for large item set compared to slight item set To that end, the lift is notsui table to determine the influence of different sizes item set. Apriori algorithm to perform frequentitemset to obtain as sociation rules. As the name implies, this algorithm us esprior knowledge of frequent item set properties onwhichwe had known before, to processfurther information. Aprioriusesan iter ativeapproac his referred to as level-wise searchwherek-itemset is usedto find the(k +1) –itemset [8].

3. RESULTS AND DISCUSSION

Sample data used comes from retail stores Alfamart Jl MT Haryono Cilacap, based on the collection of secondary data, it is obtained that every day there are about 700 transactions occur, in a yeart here are at least 255.600 transactions. This is just one branch of Alfamart alone. Number of Branches in Cilacap district, there are about 230 branches Alfamart. Then, after the completion of data used each year, what for is the data? Will it bediscarded? Is that just kept until piling numbers? Of course, although it is only kept on file an nual, there arecosts to beincurred by the company for its maintenance. If the data that has beenaccumulatedis not used, but there are maintenance costs to be incurred, the company wouldget loss.

For the process of extracting data with as sociation analysis, we take asample of transaction data Alfamart Jl MT Haryono, the transaction occurred on August 2, 2013

Table1. Shopping cart transaction data

ID Transaksi	ITEM
101-1355	KINO LRTN PET 200ML
101-1356	SAMPOERNA A MILD MRH 16[PB]
101-1356	DJARUM 76
	GG SURYA 16 [PB]
	SAMPOERNA A MILD MRH 12[PB]
101-1361	DJARUM L.A LIGHT 16
	GG FILTER MERAH 12
101-1363	WALL'S CRNT BLACK&WHT NEW 110ML
	WALL'S FEAST COKLAT 65ML
	WALL'S POPULAIRE STRAWBERRY 90ML
101-1364	BINTANG ZERO KLG 330ML
101-1365	MARLBORO BLACK MENTHOL 20
101-1367	A GULA TEBU LOKAL 1KG
	ABC SPC GRADE COCOPANDAN 585ML
	KAPAL API KOPI SPECIAL 65G
	KG AST MERAH MINI 700G
	SARIWANGI TEH ASLI 25'S
	WALL'S FEAST COKLAT 65ML
101-1368	GG FILTER MERAH 12
101-1369	DUNHILL FINE CUT MILD 20
101-1370	BIG STROBERI PET 3.1L

Association analysisis use fulfor finding important relationships hidden amongvery large datasets. Open relationshipal ready represented in the form of association rulesor a rule set of items that frequently appear. The ruleindicatesa strong relationship between the sale of one item to another because many customers buy two items of the product. Thus, the mini market management can use this as away to identify potential opportunities for cross-selling strategy (cross-selling) the goods ares old.

3.1. Modeling with Rapid Miner 5.0

Modeling with RapidMiner using data from sales transactionsin August 2013 to December 2013. Processing is done by first binaring existing transaction data. Existingraw transaction data can not be processed because RapidMiner can only read the data in the form of binary data, which is represented by"0" or"1".

Data transactions are processed real data, and nomanipulation. Despite only using transaction datain a day, but the patternsand rules can be obtained is expected to represent the spending behavior of consumers. The generation of association rules canbe continued with daily transaction data processing based on receipt number to find items that couldbe used ascross-selling strategy.

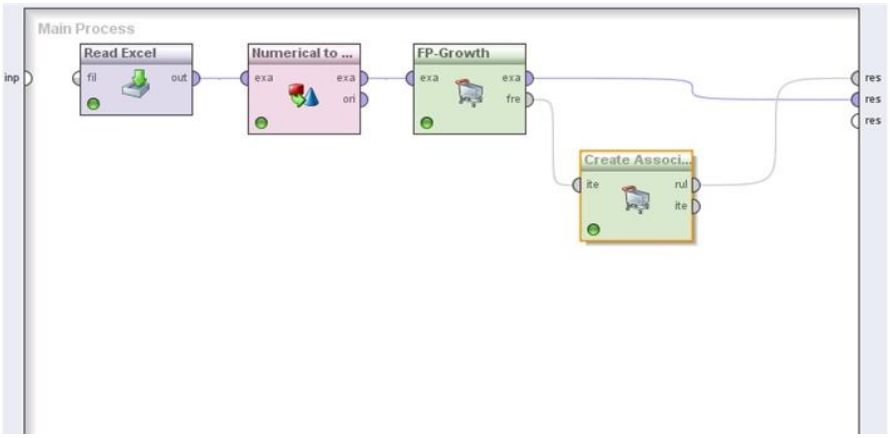

Figure 1.Processof Association Rule Modeling with RapidMiner

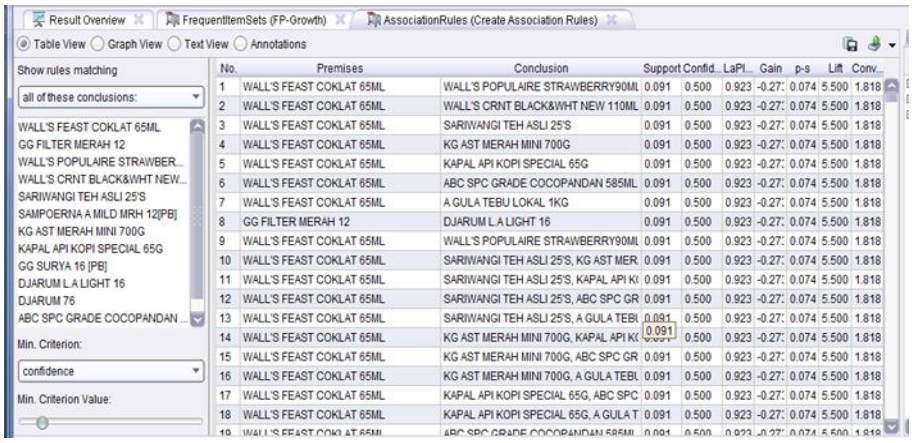

Figure 2.Results of Application of Association Rule Table View

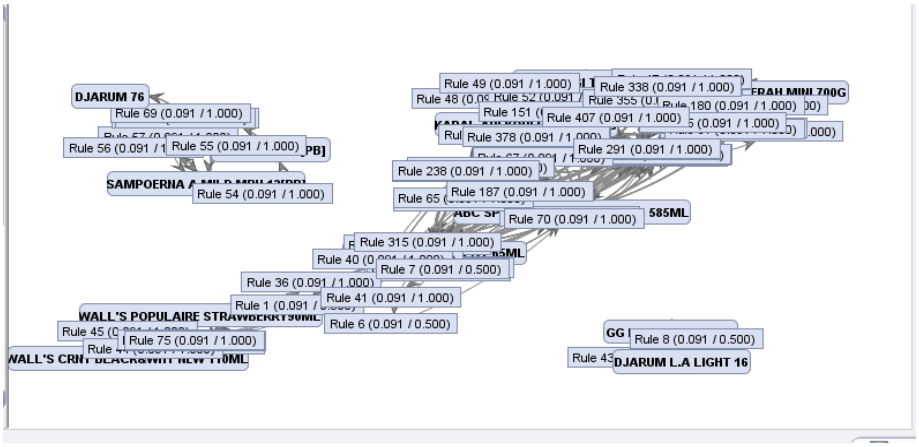

Figure 3.Results of Application of Association Rule Graph View

Calculation of Associaton Rule applied by applying Confidence value=0.8 while the value of Support=0.1. Of the defined minimumvalue, the result a total is 77rules. Rapid Miner can accommo date up to thousands of rules in accord ance with the minimum support and minimum confidence tha twe specify.

3.2. Conclusions of Cross-selling Item

Several conclusions about cross-selling products that can be recommended from the process of extracting the association data include:

a. Wall's Feast Chocolate with Sariwangi The Asli
b. Gudang Garam Filter Merah with Djarum LA Light
c. Sariwangi The Asli in various size with Kapal Api Kopi with different variants and sizes
d. Gula Tebu Lokal with Kapal Api Kopi with different variants and sizes
e. Gula Tebu Lokal with Sariwangi The Asli with different variants and sizes
f. Gula Tebu Lokal with ABC Special Grade Coco pandan
g. Djarum 76 with Gudang Garam 16
h. Sariwangi The Asli, Kapal Api Kopi with Wall's Feast Chocolate

4. CONCLUSION

From the data processing transaction dated August 2, 2013on one branch Alfamart mini market in Cilacap, it is obtained the results of association rules or link age samong products that can be used for the consideration of cross-selling strategies as well as to determine the layout of the items in the shop window. Modeling of Asssociaton Rule only takes a sample of data from the transaction for one day. Data processing

and modeling can be applied to larger data to find a combination and association rules and also a linkage of inter-complex products.

Several conclusions about cross-selling products that can be recommended from process of extracting the association data include:

1. Wall's Feast Chocolate with Sariwangi The Asli
2. Gudang Garam Filter Merah with Djarum LA Light
3. Sariwangi The Asli in various size with Kapal Api Kopi with different variants and sizes
4. Gula Tebu Lokal with Kapal Api Kopi with different variants and sizes
5. Gula Tebu Lokal with Sariwangi The Asli with different variants and sizes
6. Gula Tebu Lokal with ABC Special Grade Cocopandan
7. Djarum 76 withGudangGaram 16
8. Sariwangi The Asli, Kapal Api Kopi with Wall's Feast Chocolate

ACKNOWLEDGMENT

This study has been completed thanks to the helpand support from various parties that cannot be mentioned one by one. Especially thank you to my little family, children and my wife. Research erssaya big thank you to family STMIK AMIKOM Purwokerto who has given support to complete this study. Hope fully this research can be useful.

REFERENCES

[1] Andhika, Rully. Retno., (2013). *Excavation Association Patterns in Data Warehouse Agent Manufacturing Using Microsoft SQL Server (Case Study: XYZ.),*Engineering Journal of POMITS Vol. 2 no. 2 February. ISSN 2223-3539.(2301-9271 Print).

[2] Cox, Roger & Brittain, Paul.(2000). *Retail Mangement 4th Edition. London* : Pearson Education Limited.

[3] Changsheng, Jing Ruan, (2009). *A Modified Apriori Algorithm with Its Application in Instituting Cross-Selling Strategies of the Retail Industry,* 2009 International Conference on Electronic Commerce and Business Intelligence, 978-7695-3661-3/09, ©2009 IEEE, DOI 10.1109/ECBI.2009.121

[4] Dongwon, Sung-Hyuk Park, Songchun Moon, (2011). *High-Utility Rule Mining for Cross-Selling,* Proccedings of the 44th Hawaii International Conference on System Sciences, 1530-1605/11, ©2011 IEEE

[5] Goldie, Dana Indra Sensue, (2012). *Application of Data Mining Method of Market Basket Analysis Of Product Sales Data Book with Apriori Algorithm Using Frequent Pattern and Growth: Case Studies Printing PT. Scholastic*, Journal of Telematika MKOM, Vol.4.1, Maret 2012. ISSN: 2085-725X.

[6] Tang, Hewen; Zengfang Yang, Pingzhen Zhang, Honglin Yan, (2008). *Using Data Mining to Accelerate Cross-Selling,* International Seminar on Business and Information Management. 978-0-7695-3560-9/08, ©2008 IEEE, DOI 10.1109/ISBIM.2008.186

[7] Herera, francisco, 2010. *Data Mining and Soft Computing*, Dept. of Computer Science and A.I.University of Granada, Spain.

[8] Kusrini. (2007). *Concept and Application. Decision Support System*. Yogyakarta : Andi Publisher

[9] Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques.* San Fransisco: Morgan Kauffman. ISBN 13: 978-1-55860-901-3

[10] Vercellis, Carlo.(2009). *Business Intelligence : Data Mining and Optimization for Decision Making*. West Sussex : John Wiley & Sons Ltd.

[11] Witten, Ian, H; Eibe, Frank; Hall, A.M. 2011. *Data Mining : Practical Machine Learning Tools and Technique*, 3rd ed., Asma Sthepan and Burlington, Eds. United States of America, Morgan Kaufman.

[12] Xue-Cheng Yang, Jun.W.; Xiao-Hang.Z., Ting-Jie.L.,(2008).*Using Decision Tree And Association Rules To Predict Cross Selling Opportunities,* Proccedings of the Seventh International Conference on Machine Learning and Cybernetics, 978-14244-2096-4/08, ©2008 IEEE

[13] Yusuf, Y., F. Rian.; Gerry T., (2006). *Application of Data Mining in the Determination of Rules of Association Between Type Item,*National Seminar Application of Information Technology (SNATI 2006), ISSN: 1907-5022

[14] Luthfi, T.E., (2009). *Application of Data Mining Algorithms to Improve Sales Association*. Jurnal DASI. Vol. 10 No. 1 Maret 2009. ISSN : 1411-3201.

[15] Yang X,C; Zhang H X; Lu TJ, (2008). *Using Decision Tree and Association Rules to Predict Cross Selling Oportunities*. Proceedings of seventh international conference on machine Learning and Cybernetics, Kuningan, 12 – 15 July 2008. 978-1-4244-2096-4/08/$25.00@2008 IEEE

BIOGRAPHYIES OF AUTHORS

Imam Tahyudin was born in Indramayu, West Java, Indonesia, on July 12, 1983. He Received S.Si degree from Faculty of Science and Technology in 2006 and M.M. degree from faculty of Economic in 2010 from Jenderal Soedirman University, Purwokerto, Indonesia. He was graduated the M.Eng. degree in Department of Computer Engineering STMIK AMIKOM Yogyakarta, Indonesia, in the field of information system in 2013. He is lecturer in the department of information system STMIK AMIKOM Purwokerto, Indonesia. His research interests are in knowledge management and data mining and artificial intellegent.

Mohammad Imron was born in Brebes, Central Java, Indonesia, on September 29, 1983. He Received S.Kom degree from STMIK AMIKOM Purwokerto in 2010. He currently study M.Kom. degree in Department of Computer Engineering Dian Nuswantoro University, Semarang, Indonesia, in the field of information system Since 2012. He is lecturer in the department of information system STMIK AMIKOM Purwokerto, Indonesia. His research interest is information System.

Siti Alvi Solikhatin was born in Ngawi, East Java, Indonesia, on November 8, 1989. She is student in Departmen of Information System, STMIK AMIKOM Purwokerto, Indonesia. Her research interest is data mining and artificial intellegent.