

Data Mining Applications for Sales Information System by Using Market Basket Analysis on Stationery Company

Alexander Setiawan¹, Gregorius Satia Budhi², Djoni Haryadi Setiabudi³ Ricky Djunaidy⁴

Informatics Department, Faculty of Industrial Technology Petra Christian University Surabaya

Email : alexander@petra.ac.id¹, greg@petra.ac.id², djonihs@petra.ac.id³, m26413109@john.petra.ac.id⁴

Abstract—Stationery company located in Banjarmasin, is a company that sells items such as stationery, office equipment, and others. Business processes performed by these company are ordering and purchasing goods from suppliers and selling goods to the customer. Based on the existing situation required the completion of the method by using Data Mining association rules and administrative information systems based website for the company to be computerized and improve the control of business processes in the company. Based on the existing situation, namely the completion of the necessary sales information systems based data mining for the company to be computerized and improve the control of business processes in the company.

The end result of making this application is that it can make the process of purchasing, sales, and process of data mining market basket analysis methods. Additional features to support existing data as a reminder of stock items, inventory recommendations are also included.

Keywords—Point of Sales; Data Mining; Market Basket Analysis; Association Rule

I. INTRODUCTION

The company is still using conventional methods in the process of selling goods, using recording transactions manually. Perceived company sales year - last year it has increased very significantly. Sellers also have difficulty in selling if there are customers who buy goods very much and must record into a memorandum for each customer transaction while others had to wait to be served. Sellers also sometimes forget in cutting inventory on card stock when the transaction ends, so that the data card stock to perform inventory taking are often incompatible with its contents. With this required computerized sales information system that can help facilitate the sales process and the recording of sales transactions (memorandum) to the customer.

This company also provides a bundle purchase goods for sale. Application of bundle sales processes have been implemented long ago. Sales of goods bundle is based on the moon - in particular, such as the one for the month of June-July is a time where customers prepare their children to go to school. At that time also company presents bundle (package) schools with contents 3 pack of books, 3 roll plastic cover, 3 roll cover of chocolate and then there are also two dozen pens, one dozen pencils, 3 erasers and 3 tip-x at a cheaper price than sales ordinary. With sales like this, bundling sales proved to be more sought after and more selling than buying items one by one in this company.

Based on these cases, the application will be made along with the sale of goods Market Basket Analysis, which is to look for linkages between the products in the company using Association Rule Stationery Asia. With sales of computerized processes, obtain useful data for data processing to the data mining process. The results offered by the combination of regular evidence-based sales transactions that are running. With the known linkage of purchase by the customer, the company has a marketing strategy, that strategy is the product placement in the company, provide inventory for products that are related to each other, and develop a package (bundle) for the customer.

II. LITERATURE REVIEW

A. Data Mining

Data Mining is often defined as "write many of the reports and queries". But in fact the data mining activities do not report creation and query altogether. Data mining is done with a special tool, which executes data operations that have been defined based on the model of analysis. Data mining is the extraction of information or patterns that are important or interesting from the data residing on large data base which had been unknown but potentially useful information [1]

Data Mining analyze the data to find information hidden in large amounts of stored data. Data mining is a process that is different from the usual statistical analysis. Function-common functions that are applied to data mining are: [2]

- Association, is the process of finding associative rule between a combination of items.
- Sequence, the process of finding associative rule between a combination of items at a time and apply more than one period.
- Clustering, is the process of grouping a number of data / objects, into groups so that each group of data containing similar data.
- Classification, the discovery process model or function or distinguish concepts or classes of data, with the aim to be able to estimate the class of an object that the label is not known.
- Regression, is the process of estimating the value predicted by the patterns in the data set.
- Solution, is a process of discovery of the root problem and solving of the problems faced by business or at least as information in decision making.

B. Market Basket Analysis

Market basket analysis is an association in data mining to find attributes that appear in one time [4]. This process can determine customer buying patterns by finding relationships between different items in a sales transaction. The results that have been obtained can be used by retail companies such as sales company / supermarket develop a marketing strategy to see the items that may be purchased simultaneously by consumers.

Analysis of the association or the association rule is a data mining techniques to discover the rules of associative between a combination of items [3]. Examples of associations that possibility, namely how likely buyers buy bread with jam simultaneously. Association rules can be determined by two parameters, support and confidence. Support that is a percentage of the value of supporting combinations of items in the database, while confidence is certainty value the strong relationship between items in the rules of association.

C. Association Rules

Association rules are usually expressed in the form: {bread, butter} → {milk} (support = 40% confidence = 75%). This rule means 75% of transactions includes bread, butter also contains milk. While 40% of all transactions containing the third item. Association analysis is defined as a process to find all association rules that meet the minimum requirements to support (minimum support) and the minimum requirements for confidence (minimum confidence) [3]. If support ≥ minimum support and confidence ≥ minimum confidence, then the rule can be expressed as interesting rule [2].

D. Apriori Algorithm

Apriori algorithms including the type of association rules in data mining. In addition to Apriori, which is included in this group are the method and the Generalized Rule Induction Based Hash Algorithm. Association analysis is data mining techniques to discover an association rule between a combination of items. Examples of association rules from the analysis of the purchase in a supermarket is able to know how big the possibility someone to buy bread together with milk. Based on this knowledge, the owner can self-regulate the placement of the goods or designing a marketing campaign by using a combination of discount coupons for certain items. [3]

Based on the parameters of support (supporting values) and confidence (value certainty), the association basic methodology can be divided into two phases, namely: Analysis of the highest frequency pattern and establishment of association rules. Algorithm pseudocode can be seen on Fig 1.

Apriori Pseudocode

```

Apriori (T, ε)
L1 ← { large 1-itemsets that appear
in more than ε transactions }
k ← 2
while Lk-1 ≠ ∅
    Ck ← Generate(Lk-1)
    for transactions t ∈ T
        Ct ← Subset(Ck, t)
        for candidates c ∈ Ct
            count[c] ← count[c] + 1
    Lk ← { c ∈ Ck | count[c] ≥ ε }
    k ← k + 1
return ∪ Lk
    
```

← Join step and prune step

Fig. 1 Apriori Algorithm in Pseudocode

III. ANALYSIS AND DESAIN DATA MINING

A. Data Flow Diagram (DFD)

Data Flow Diagram (DFD) describes the data sources, data flow, data processing, data storage, and a purpose that data in a data flow diagram system serves to document the system used now and planning the design of the new system. Manufacture begins with making the DFD context diagram is a picture of the overall information system. After making a context diagram, can be made level 0 and level diagram to describe sub processes further and more detailed data flow. In this DFD there are 4 (four) external entities that provide input and output on the data mining such as, supplier, customer, owner and employee.

B. Entity Relationship Diagram (ERD)

Entity Relationship Diagram (ERD) is a technical overview to describe the schema of a database. Named as the ERD as it shows the various entities are modeled and the important relationship between the entity to entity. Entity is all physical resources, events, and agents where necessary in an organization or company in getting the data [5].

IV. RESULT AND DISCUSSION

A. Market Basket Analysis Sales Transaction

The process is based on the analysis of sales transactions in the system. The transaction data that used to be a comparative study on the purchased goods in Table I. From the results of the existing deal then create a table that contains a total of goods produced. Table number of items can be seen in Table II.

TABLE I. TRANSACTION TABLE

Transaction	Item Purchased
1	D, E, C
2	C, A
3	C, B, D
4	A, C, E, D
5	E, D
6	E, C, B
7	B, D, A

TABLE II. TABLE OF ITEM THAT HAVE BEEN CLASSIFIED

Transaction	A	B	C	D	E
1	0	0	1	1	1
2	1	0	1	0	0
3	0	1	1	1	0
4	1	0	1	1	1
5	0	0	0	1	1
6	0	1	1	0	1
7	1	1	0	1	0
Summary	3	3	5	5	4

From Table II and then specify the value in this example the value Φ rated 3. What does it mean for the transaction $k = 1$, all data is greater than the value of Φ then $F1 = \{\{A\}, \{B\}, \{C\}, \{D\}, \{E\}\}$. for $k = 2$ (2 elements) then set that may be formed is $\{A, B\}, \{A, C\}, \{A, D\}, \{A, E\}, \{B, C\}, \{B, D\}, \{B, E\}, \{C, D\}, \{C, E\}, \{D, E\}$. For the calculation of the amount concerned can be seen in Fig 2.

T	A	B	&&	T	A	C	&&	T	A	D	&&	T	A	E	&&
1	0	0	F	1	0	1	F	1	0	1	F	1	0	1	F
2	1	0	F	2	1	1	F	2	1	0	F	2	1	1	F
3	0	1	F	3	0	1	F	3	0	1	F	3	0	0	F
4	1	0	F	4	1	1	T	4	1	1	T	4	1	1	T
5	0	0	F	5	0	0	F	5	0	1	F	5	0	1	F
6	0	1	F	6	0	1	F	6	0	0	F	6	0	1	F
7	1	1	T	7	1	1	T	7	1	1	T	7	1	0	F
	Σ		1		Σ		1		Σ		2		Σ		1
T	B	C	&&	T	B	D	&&	T	B	E	&&				
1	0	1	F	1	0	1	F	1	0	1	F	$\Phi = 3$			
2	0	1	F	2	0	0	F	2	0	0	F	Syarat $\Sigma \geq \Phi$			
3	1	1	T	3	1	1	T	3	1	0	F				
4	0	1	F	4	0	1	F	4	0	1	F				
5	0	0	F	5	0	1	F	5	0	1	F				
6	1	1	T	6	1	0	F	6	1	1	T				
7	1	1	T	7	1	1	T	7	1	0	F				
	Σ		2		Σ		2		Σ		1				
T	C	D	&&	T	C	E	&&	T	D	E	&&				
1	1	1	T	1	1	1	T	1	1	1	T				
2	1	0	F	2	1	0	F	2	0	0	F				
3	1	1	T	3	1	0	F	3	1	0	F				
4	1	1	T	4	1	1	T	4	1	1	T				
5	0	1	F	5	0	1	F	5	1	1	T				
6	1	0	F	6	1	1	T	6	0	1	F				
7	0	1	F	7	0	0	F	7	1	0	F				
	Σ		2		Σ		2		Σ		2				

Fig 2. Calculation k=2

From the input purchases based on case studies in company, meaning a score of 'F' in the column indicates && goods are not sold together, while the 'T' which means the goods are sold together. With written terms ($\Sigma \geq \Phi$) then found the next set is $F2 = \{\{C, D\}, \{C, E\}, \{D, E\}\}$. The combination of the set F2 combined into third candidate item set. Item sets that can be merged are itemset-itemsets that have similarities in k-1 first item. For k = 3 elements, which may be formed is set $\{C, D, E\}$, the calculation to find the set F3 in Table III.

TABLE III. CALCULATION OF K=3 IN THE SET {C, D, E}

T	C	D	E	&&
1	1	1	1	T
2	1	0	0	F
3	1	1	0	F
4	1	1	1	T
5	0	1	1	F
6	1	0	1	F
7	0	1	0	F
			Σ	2

The table of the results were obtained $F3 = \{\}$, since no qualified ($\Sigma \geq \Phi$), then set F4, F5, F6, F7 ... is also a null set. Having obtained all of which qualify ($\Sigma \geq \Phi$), then determine (ss-s) is antecedent and (s) that is consequent So that derived table that can be seen in Table IV.

On the set there $F2 = \{\{C, D\}, \{C, E\}, \{D, E\}\}$ It can be arranged :

- If (ss-s) = C, if s = D → If buy C then buy D
- If (ss-s) = D, if s = C → If buy D then buy C
- If (ss-s) = C, if s = E → If buy C then buy E
- If (ss-s) = E, if s = C → If buy E then buy C
- If (ss-s) = D, if s = E → If buy D then buy E
- If (ss-s) = E, if s = D → If buy E then buy D

TABLE IV. PERCENTAGE TABLE SUPPORT AND COFIDENCE

IF antecedent Then consequent	Support	confidence
C → D	(3/7)x100% = 42.86%	(3/5)x100% = 60%
D → C	(3/7)x100% = 42.86%	(3/5)x100% = 60%
C → E	(3/7)x100% = 42.86%	(3/5)x100% = 60%
E → C	(3/7)x100% = 42.86%	(3/4)x100% = 75%
D → E	(3/7)x100% = 42.86%	(3/5)x100% = 60%
E → D	(3/7)x100% = 42.86%	(3/4)x100% = 75%

Once can support and confidence for each candidate, his confidence level of 70% is taken up, so that in the table can be as follows:

Having obtained the multiplication of support and confidence, choose the greatest multiplicative result. Result in the greatest of the multiplication of the multiplication is a rule that is used at the time of selling. Because the result of the multiplication of the 2nd sales over the same value, then everything can be used as a rule.

- If you buy E will buy D with 42.86% support and 75% confidence
- If you buy E will buy C with 42.86% support and 75% confidence

Based on the results of these calculations, the display program issued a corresponding rule in Figure 3.

Frequent Itemsets	
Time:	0 second(s)
{D} = 5	
{E} = 4	
{C} = 5	
{A} = 3	
{B} = 3	
{D, E} = 3	
{D, C} = 3	
{E, C} = 3	
Association Rules	
Time:	0 second(s)
E Juga membeli item D : Support = 42.86%, Conf = 75%	
E Juga membeli item C : Support = 42.86%, Conf = 75%	

Fig 3. The results of frequency program items and association rules

Yield data association_rule Data in the table can be seen in Fig 4.

id rule	antecedent	consequent	support	confidence
1	E	1B D	1B	42.86 75
2	E	1B C	1B	42.86 75

Fig 4. results rule data is stored in the database

Association rule is formed by bonding two rules are antecedent and consequent. In Fig 4 can be read with the antecedent of the consequent with the value of the support and confidence values. Data released in accordance with a minimal amount of support and the minimum confidence that has been entered by the user.

B. Combination of Goods Based on Rule

The improved recommendation by the Joint Assessment Process is the support and confidence of the initial compilation process scan data. In Figure 5 is the result of the combination of goods by the rules.

rule	support	confidence	action
binder clip no 155_jytko_pcs (Related to) isi staples no 10_jytko_pcs	0.33	33.33	Filter Data
binder clip no 200_jytko_pcs (Related to) binder clip no 155_jytko_pcs	0.33	100	Filter Data
binder clip no 200_jytko_pcs (Related to) binder clip no 155_jytko_pcs	0.33	100	Filter Data
binder note a5_jytko_pcs (Related to) isi binder a5 50lbr_jytko_pcs	0.33	100	Filter Data
buku 38_titlar_dama_pak (Related to) tp ev ke 01_heriko_pcs	0.33	29.57	Filter Data
cartridge 802 hitam_hj_pcs (Related to) cartridge 802 warna_hj_pcs	0.33	66.67	Filter Data
cartridge 802 warna_hj_pcs (Related to) cartridge 802 hitam_hj_pcs	0.33	100	Filter Data
double tape 12mm_damanu_pcs (Related to) stjerfiam 4040_alibaku_pcs	0.33	66.67	Filter Data
double tape 24mm_damanu_pcs (Related to) kertas hafan_uinstan_pcs	0.33	40	Filter Data
gelas agar_stairtdo_pak (Related to) sendok beang sb_tumbuh hembang_pak	0.33	33.33	Filter Data

Fig 5. The results of association rule that can be used

C. The Recommendations and The Report on The Composition of The Resulting Rule

Several recommendations are based on the combination of the composition of the cabinet there. Once the user performs the scan data on the composition of the goods can be produced based on the total confidence of the biggest rule to rule the smallest. In Figure 6 is an example of a recommendation from the cabinet.

Item	Quantity	Related Items
stapler hid 10 max pcs	10	isi staples no 10_jytko_pcs, kertas hrs 14 70g mitsage pak, kantong-an 40 isi 50 lampu pak
binder clip no 200_jytko_pcs	200	cartridge 802 warna hp pcs, binder clip no 155_jytko_pcs, map kertas kuno pak, lamban hitam 48mm diamatu pcs
stapler hid 10 henko pcs	10	kertas hrs 14 70g mitsage pak, kertas hrs 14 70g mitsage pak, sampul colial kwarto fancy pcs
lamban hitam 48mm diamatu pcs	48mm	lamban hitam 48mm diamatu pcs, map kertas kuno pak, tinta printer dp 40 dataprint pcs
tinta printer dp 40 dataprint pcs	40	isi staples no 2 max pcs, isi staples no 10 max pcs

Fig 6. Goods entering cabinets in accordance with the order of the rule.

V. CONCLUSIONS

From the creation and design of the program can be concluded as follows:

- Applications can make the sales process is automatically integrated with the database.
- Applications can perform the purchase process is automatically integrated with the database.
- Applications can perform recording in accordance with the stock purchase and sales.
- The application can notify the owner of the stock item is running low.
- Applications can be provided reports as required by the manager.

- Applications can perform data mining process based on existing sales data. Then program can help decide when to make the process of bundling.
- According to the questionnaire, 85% of consumers rate the appearance is very good, 15% of consumers rate the appearance is very good, 100% of users assess the accuracy of the data generated very good, 25% of users rate the application, simply, 75% of users rated ease of application excellent, 100% of users evaluate reports produced good, 10% of users to assess the suitability of the needs of both, 90% of consumers rate the suitability to the needs of very good

REFERENCES

- [1]. Rifa, Noverino (2004). *Bussiness Intelligence*, ITB.
- [2]. Han, Jiawei & Kember, Michelin, 2001, *Data Mining Concepts & Techniques*, Simon Fraser University Academic Press, US.
- [3]. Kusriani, Emha T.L., *Algoritma Data Mining*, Yogyakarta: Penerbit Andi.
- [4]. giudici, P. and Figini, S. (2009) Market Basket Analysis, in *Applied Data Mining for Business and Industry*, Second Edition, John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9780470745830.ch7.
- [5]. Romney, M, & Steinbart, P. (2012). *Accounting Information System*. New Jersey: Prentice Hall.