Apriori Application To Pattern Profile Creditor Relationships With Credit Ceiling In Rural Bank

Theresia Widji A
Terpikat Polytechnic
Sambas West Kalimantan
Indonesia

Beta Noranita
Departement of Information System
Postgraduate Programs, Diponegoro University
Semarang Indonesia
Email: beta@undip.ac.id

Abstract-Maintaining the customer to remain loyal and acquiring new clients are the successful capital for financial institutions such as the BPR. This is supported by appropriate marketing strategies. One marketing strategy is obtained from a Ddatabase which is a combination of hidden patterns of relationships between items on the profile of creditors with credit ceiling. Pattern rules creditor profile relationship with the ceiling is presented in the model Apriori application that requires two input values are minimum values and the value of frequent item set minimum confidence. Item profiles creditors and the credit ceilings that appear simultaneously on each transaction are important in finding new marketing strategies such as target new creditors. One model is the application of Apriori Datamining that can be used to process data into a meaningful pattern rules. The results of the application may present Apriori pattern profile creditor relationship with the credit ceiling with graphs the value of support and confidence as well as provide a benchmark in providing the credit limit on case that has ever happened.

Keywords : Customer Profile, Credit Plafond, Marketing Strategy, The Apriori Algorithm

1. INTRODUCTION

Retain customers in order to remain loyal and acquire new customers is critical to the succees of the company supported by marketing strategies. The discovery of interesting pattern in the data company can offer a better marketing strategy (Nan-Chen Hsieh.,2004). The discovery of interesting pattern and build amodel of the database is a techique which is carried by data mining (Mu-Chen Chen et al.,2005). Data mining can be used to identify overall customer behavior patterns derived from customer data and transaction data (Giudici & Passerone.,2002). Data mining techniques search within the database on an ongoing basis without any special aprovision for the hypothesis that the goal of finding a complete, previously unknown, and very useful information including the rule of knowledge, connectedness, and regularity (Chen, Han, & Yu.,1996).

Data mining is astep in the Knowledge Discovery in Databases (KDD), wich involves the application of algorithms for pattern extraction (Mitra, Pal & Mitra.,2002). Extraction is used extensively to analyze the relation between the purchased product and to support market segment and sales promotion (Changchien & Lu.,2001; Changchien, Lee & Hsu,2004).

One of the association rule algorithm is a apriori that the essential aim of finding potential relationship between items or special events are aligned in the database (Agrawal et al.,1993). Relationship items or special events examplified in the database of a bank has been found to predict the pattern of bankcruptcy person (Dasgupta, et al 1994; Desai Crook, & Overstreet,1996) and credit ratings (Kim & Sohn,2004; Lancher et.,al 1995; Sharda & wilson,1996).

Availability profiles in classifying customers is the foundation of a company to provide better service and can maintain a quality customer (Setiono,1998). Customer profile may explain the occurance of each representative the group and provides tools to build the bank's marketing strategy for the better (Nan-Chen Hsieh.,2004).

The problem faced is how the transaction history stored in database rural bank (BPR) Gunung Kawi Semarang can be used and presented in the application that generates a pattern profile relationship with the creditors and the creditceiling becomes the basis of assessment for the provision of ceiling.

2. THEORETICAL

2.1. Data Mining, Association Rules, Apriori Algorithm

Data mining is the process of finding meaningful relationships, patterns, and trends by examining the large collections of data stored in storage by using techniques such as statistical and mathematical patterns. The specificity of the databases used in data mining has millions of records and thousands of variables. Records its not the same as all the variables and stand alone without any relationship with other variable (Larose,p1,2005).

Data mining is also called Knowledge Discovery in Databases (KDD) is a process of searching data ini a very large memory of the data for patterns using tools such as classification, association, clustering. The overall process to discover useful knowledge from data with reference to the steps particular, figure 1(Fayyad,1996).
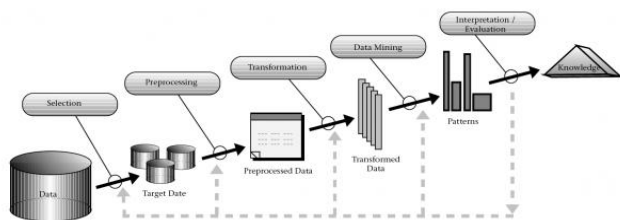
Figure 1. KDD (Fayyad, 1996)

Association rule starts from the transaction that contains one or more products or services and some information that has not been perfect transaction for purpose of analysis the product (Berry Linoff, 1997) and alaso a form of expression X → Y, where X and Y is the collection item. To find a convincing association rule, user specializes in minimum support and minimum confidence. Association rule can be made into two sub-problems (Ramakrishnan.,et al,1997). *Association rule* in the form " *if … then* …" is the knowledge that results from an association rule (Sani Susanto.,dkk, h97,2010).

Ramakrishnan.,et al,1997 (Agrawal et al,1996), when is now apriori algoritm is used to find all frequent itemsets as the basis for the presentation. Xindong Wu.,et al, (p63,2007), Apriori is an algorithm to find all items that have a value set support (support) and the value of certainty (kepastian). The technical term is set to follow in the apriori association rule (Susanto & Suryadi, h97,2010), the first of item goods that become the object of spending activity. The set is an item of goods that become the object of spending activity. The set of items is denoted by I. for instance I = $\{i_1,i_2,…,i_m\}$. the set of items purchased to -i is called the i-transaction denoted by **Ti**, the set of all transaction denoted **D**, such D = {T1,T2,T3…Ti}. Each transaction has a uniqe identifier called TID. Association rule to be gained will form the following implications:

"If A, Then B" Or "A → B".

A is called *anteseden* of implications, B is called C*onsequent*. Association rule must satisfy two properties, first of both A and B is the set of pure I, so A,B $\subset$ I . second, the set A and B are two subsets of each other off, so **:**

A ∩ B = ø

2.2. Reference Literature

"An integrated data mining and behavioral scoring model for analyzing bank customer" proposed assessment to manage the behavior of credit card customer's bank in Taiwan to remain loyal. (Nan-Chen Hsieh, 2004).

"Mining Changes in Customer Behavior in Retail Marketing" proposed marketing managers can establish a relationship with cutomers for the long term if they know and predict changes in customer behavior (Mu-Chen Chen et al., 2005). Datamining can be used to discover the identity of the behavior patterns of customer desire and customer number is large (Giudici & Passerone, 2002). Relationship products purchased by customers are analyzed using the algorithmApriori association rule (Agrawal,

Imielinski,& Swami,1993; Srikant Vu, & Agrawal,1997). Apriori association rule can be used to get to know one of the realtionship between customer profiles represented by Demographic variables and purchase of products from the database of customers and products are examined (Song et al.,2001).

"Application of data mining association rules in determing the inter-item type". Organizations can be flooded with various kinds of data invaolved in any sales transaction. Sales transaction data is collected and stored can provide useful knowledge for management in the conduct of bussiness. Knowledge of association rule between types of items that appear simulataneously on each transaction an important input in efforts to increase turnover and provide the knowledge combination of items that can be included in the promotion (Yogi Yusuf et al.,2006).

"Application Algorithhms of Datamining to incease sales to association". The right sales strategy is very important in bussiness to be able to increase the value of sales. The use of data mining algorithm apriori association rule to construct a system that has the ability to see patterns of sales of goods which can then be used to develop new sales strategies (Emha Taufiq Luthfi, 2009).

2.3. Equation

Association rule can be made into two-subproblems (Ramakrishnan.,et al,1997):
1. Find all combinationss of items that have support greater than the minimum support. The combination of items that most often occur from an itemset is called frequent itemsets.
2. Using frequent itemset to generate the desired rules.

Support of item with equation (Kusrini& Emha.,p150,2009)

$$Support\ A = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ transaksi} \quad (1)$$

$$Support\ (A,B) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Total\ transaksi}$$

$$(2)$$

$$Confidence = \frac{Support\ ABCD\ (\ Jumlah\ Transaksi\ ABCD)}{Support\ AB\ (Jumlah\ Transaksi\ Mengandung\ AB)} \quad (3)$$

The process is easy to do for at least frequent itemset is found. Support dan Confidence of more frequent between the values of 0 % - 100% instead 0 - 1.0 (Emha,2009).

2.4. Algorithm

Step algorithm (Xindong Wu.,et al,p63,2007) :
1. Support each item was calculate and items that appear is determined. Each step then generates candidate frequent itemsets are then used to generate new potential frequent itemset called candidate itemset.
2. An attractive minimum support collected, the frequent itemsets that been determined, and became a candidate

for the next. This process is repeated as long as no new frequent itemset.

Psaudocode algoritma apriori (leo willyanto, 2003) :

```
F1 = {frequent1-item sets};
 K = 2;
While (F_{k-1} tidak kosong)
{
        C_k = apriori_generate(F_{k-1});
        Untuk semua transaksi dalam T
        {
                Subset (C_k,t);
        }
        F_k = {C in C_k s.t c.count >= min_suprt};
}
Answer = Union dari semua set F_k;
Apriori_generate (F(k-1))
{
        Join F_{k-1} dengan sehingga F_{k-1},
        C1 = (i1,i2,…,i_{k-1}) dan C2 = (j1,j2,…,j_{k-1})
        Join bersama-sama jika ip = jp untuk 1<=p<=k-1,
        Dan kemudian kandidat baru, C, punya bentuk
C=(i1,i2,…,i_{k-1},j_{k-1}).
        C kemudian ditambahkan ke struktur hash tree.
}
```

3. METHODOLOGY

Object using the data as many as 738 kreditors *record* period Januari 2008 to Januari 2011 was taken random. Association technique used by the program using apriori algorithm in Delphi XE and Database MySql.

Presentation of the pattern pattern profile creditor relationship with the credit ceiling that be supported Toshiba notebook, Intel (R) Core (TM) i3 CPU M 380 @ 2,556 GHz, RAM 2GB, System type 32-Bit Operating System Windows 7. For operating the apriori algorithm used Delphi XE, Database MySQL, Power Designer 6.32 bit.

Making application in accordance with the waterfall model for software development. Application created a tool are not absolute, because it can can not be separated from the interference of users that determine frequent items.

Application through several stages in the KDD according to cases that will be processed as follows:

a. Selection of Data

TABLE 1. CHOOSEN ITEM

Attribute	Information
Id	not null
Id-creditor	
Account	
Sex	Indeks
Date of birth	
Age	Indeks
Company	
Job	
Revenue	Indek
Credit- ceiling	Indeks

Installement payment	
Period of time	
Usefulness	
Economy-sector	Indeks
value of Collateral	Indeks
collectibilitation	

b. Data Cleaning

Data cleaning stages cleaning is done by not to include attributes thar are not used and deleting incomplete data filling.

c. Transformation

The data has been selected, processed into a simpler data for simplicity in works by doing the classification of items selected profile, see Table 2. Job classification, Table 3. revenue Classification, Table 4. Sex classification, Tabel 5. Age classification, Tabel 6. Collateral Classification and Tabel 7. Credit-Ceiling Classification .

Table 2
Job Classification

No	Type of Job	Code
1	Agricalture, farm, forestry	1001
2	Fishery	1002
3	Mining and digging	1003
4	processing Industry	1004
5	Electricity, gas, water	1005
6	Construction	1006
7	Retail	1007
8	Food and beverages	1008
9	Transportation, Ware housing dan Comunication	1009
10	Financial intermediarist	1010
11	Real Estate	1011
12	Goverment administration	1012
13	Education services	1013
14	Health services and social activity	1014
15	Social services,cultural, entertaint, and personal	1015
16	Personal services house hold	1016
17	Bussines are not clear boundaries	1018
18	Not the field of bussiness – home	1019
19	Not the field of bussiness – other	1020

TABEL 3 REVENUE CLASSIFICATION

No	Revenue (A) (Million IDR)	Initial
1.	≤ 2	A1
2.	$> 2 - \leq 5$	A2
3.	>5	A3

TABLE 4. SEX CLASSIFICATION

No	Sex	Initial
1.	Male	L
2.	Female	P

TABLE 5. AGE CLASSIFICATION

No	Age	Initial
1.	Produktif (15 years old – 64 years old)	Pro
2.	Non Produktif (< 15 years old , > 64 years old)	NonPro

TABLE 6. COLLATERAL CLASSIFICATION

No	Collateral (B) (Million IDR)	Initial
1.	<5	B1
2.	> 5 – < 15	B2
3.	15 - < 30	B3
4.	30 – 50	B4
5.	> 50	B5

TABLE 7. CREDIT-CEILING CLASSIFICATION

No	Credit-Ceiling (C) (Million IDR)	Inisial
1.	≤ 5	C1
2.	> 5 – 15	C2
3.	≥ 15 – < 30	C3
4.	> 30 - < 50	C4
5.	50 – 100	C5
6.	≥ 100	C6

d. Using Support and Confidence

The use of equation of support and confidence to do after getting the items to becombined and have been simpler form. It would obviously be simulated case studies of 738 records of items that have been assigned the economic sector, revenue, sex, age,collateral with a credit ceiling, see Table 8.

TABLE 8. CREDITOR TRANSACTION

Economic-sector	Revenue (IDR)	Sex	Age	Collateral (IDR)	Ceiling (IDR)
Zigent car	20.000.000	L	32	90.000.000	55.000.000
Yayasan Nurul Islami	1.500.000	L	45	71.470.000	3.000.000
Yayasan Maarif NU	2.500.000	P	54	7.000.000	5.000.000
Yayasan Budi Luhur	1.5000.000	P	51	112.067.000	50.000.000
Yan's Sell	4.500.000	L	40	7.000.000	4.000.000
WR. Mergo Roso	3.000.000	L	58	7.000.000	4.000.000
Wirausaha counter Hp	484.000	L	29	8.000.000	5.000.000
Wiraswasta Jual	2.400.000	P	47	371.000	3.753.750
Sembako	2.700.000	P	44	1.335.134	6.435.000
Wiraswasta Jual blanjan					

Each item *scanned* and creditor profile support is computed. Frequent the set is used as a base value for the ruke of asscociation "if – then". Item profiles creditors who used the items that have a frequent occurence is 2, if less then the set value is no used for the next process is the combination, see figure 1.

tid		7	15	A1	A2	L	P	Pro	B1	B2	B5	C1	C2	C5
1						1		1		1				1
2			1	1		1		1				1	1	
3			1		1		1	1	1				1	
4				1			1	1		1				1
5		1			1		1	1		1		1		
6				1		1		1		1				
7		1		1			1	1				1		
8		1			1		1	1	1			1		
9		1			1		1	1	1				1	
Σ		4	2	3	5	5	4	9	2	4	3	4	3	2
S		0.44	0.2	0.3	0.6	0.556	0.4	1	0.222	0.4	0.3	0.4	0.33	0.2

Figure 1. Qualify for Frequent-itemset

Candidate profile creditor items delivered as A transaction and the addition of new items is defined as transaction B. Rules of this support continue until there are no items creditor profile so there is no longer a creditor profile items can be combined and this combination indicates the transaction has been completed. Combination

of items performed from 2 items to 6 items, Figure 2. Images that separates column items that have not passed minimal frequent and minimal confidence in the column item that has qualified.

Figure 3 is the result of the final combination. The combination of the final rules's pattern profile obtained is a creditor who frequently appeared together as many as 44, 18, 12 (> minimal frequent) and is a form of knowledge in marketing strategy to target new creidtors.
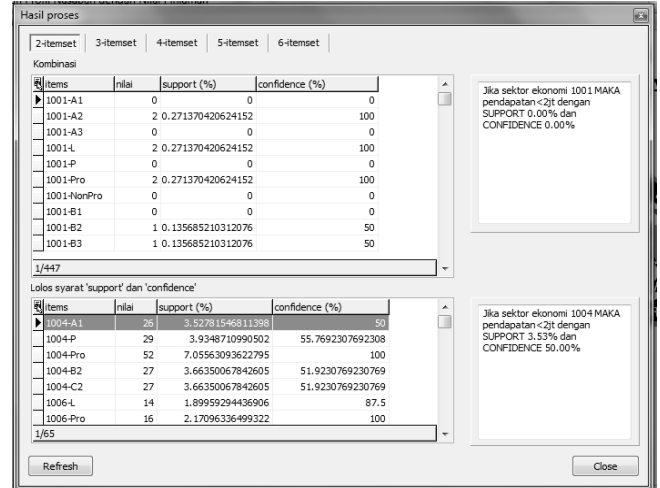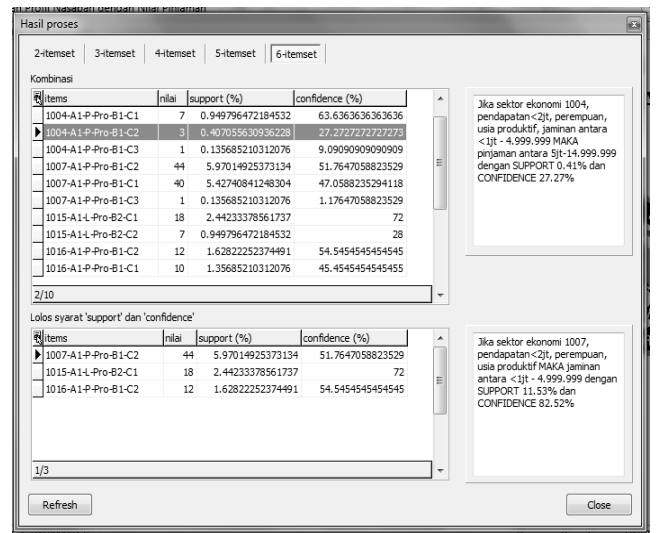

Figure 2. Combination of 2-item

Figure 3. Combination of 6-items

Figure 4. is a graph of the support and confidence resulting from a combination of late (done test 50 x). the end of six combination only for pattern rule to 4, 7 and 9 who pass the prequisites set out in the beginnng of the process.
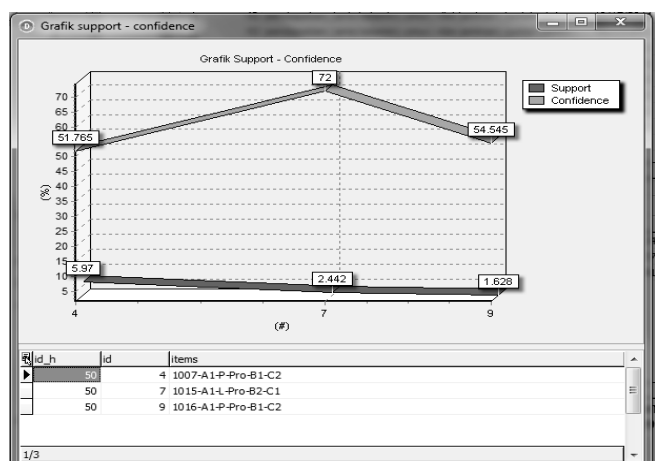
Figure 4. Graph of the Support dan Confidence

Application for the calculation by using apriori would be maximized if used for new case that will happen. If there as a creditor with a profile that has ever happened, it is known that as a basis for consideration in providing a number of ceiling, Figure 5. By filling in the items available, it will know what the ceiling will be given to the new creditor and for the pattern formed can also be seen how often appears simulataneously occur.
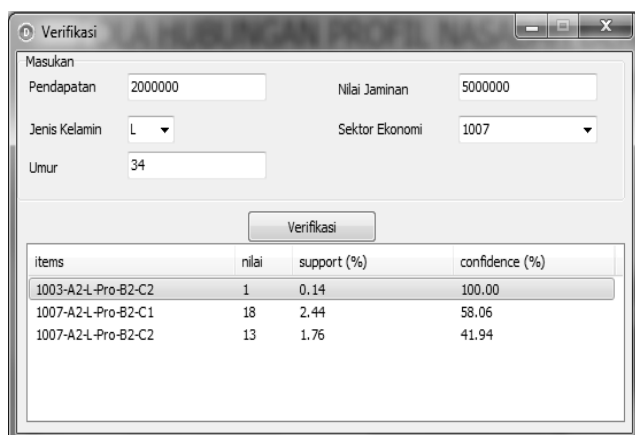

Figure 5. Verification of application apriori

4. RESULT AND DISCUSSION

The pattern rule with credit-ceiling profile kreditor generated " if-then " from combination of items eligible creditors who minimum frequent itemset 10 and minimum confidence of 50%, has three patterns of association rules. From the pattern of relationship creditor profile relationship with the credit ceiling is formed as follows:

1. **If** creditor have sector economic in retail trade (1007), revenue of less than 2 million rupiahs, female are still productive age has a value of collateral amounted to lessthan five million rupiahs, **Then** received credit ceiling of 5 million to 15 million rupiahs and have support 5.97 %, confidence 51.74%.

2. **If** creditor have sector economic in social services, social, cultural, entertainment (1015), revenue of less then 2 million rupiahs, male still productive age has collateral value of more than 5 million dollars to less than 15 million rupiahs, **Then** the credit ceiling received the highest is 5 million rupiahs with support 2,44% and confidence 72 %.

3. **If** creditor have sector economic in personel services serving household (1016), revenue of less than 2 million rupiahs, female is still productive age has value of collateral mounted to less than 5 million rupiahs, **Then** received the credit limit of 5 million dollars to 15 million rupiahs with support is 1,62 % and confidence 54.54%.

The three patterns of the resulting rules can be viewed target product promotion strategy, with the goal of success to maintain creditor and rigth on target markets. Can be seen that the three patterns of relationship with the creditor profile of credit ceiling has a different economic sector, but has the same income that is equal to less than 2 million rupiahs, more dominant creditor is female with collateral to 5 million rupiahs, may receive a credit limit of 15 million rupiahs.

The use of aplication can help to allocation ceiling, due to a basic assessment of the new creditor profile is to have a previous history of the creditors and how often the pattern of the old creditors profile appear as it did on the new creditor profile.

5. CONCLUSION

Apriori application model has been produced to present a pattern of relationship with the creditor profle of the credit ceiling combination of items in the database so that creditors remain loyal and precise in targeting target new creditors.

This application profile to find the pattern rule creditor relationship with the credit ceiling among the items in database are accompanied by support and confidence values for each pattern of relationship with graph of support and confidence for the rule's pattern profile creditor relationship with the credit ceiling.

The application provides an apriori basis for assessment againts the ceiling of the new creditor with similar items.

If frequent itemset entered the greater the number of items that occur together will be less, because the number of items that appear simultaneously are few in number.

BIBLIOGRAPHY

[1] Adomavicius,Gediminas; & Tuzhilin,Alexander., 1999. User profiling in Personalization Applications Through Rule discovery and validation. KDD-99 San Diego CA USA, 1-58113-143-7/99/08.

[2] Anonim., 2011. Perkembangan Beberapa Indikator Utama Sosial-Ekonomi Indonesia (BPS).

[3] Berry, Michael J.A; & Linoff,Gordon S.,1997. *Data mining Techniques for Marketing, sales, Customer Support*. United States of America:Wiley Publishing,Inc.

[4] Berry, Michael J.A; & Linoff,Gordon S.,2004, *Data mining Techniques for Marketing, sales, customer Relationshhip Management*. Second Edition United States of America: Wiley Publishing,Inc.

[5] Budi, Gregorius, et al., 2006. *Penggunaan Metode Market Basket Analysis Untuk Meningkatkan Kemampuan Search Engine Online Buku Tugas Akhir Pada Perpustakaan Uk Petra*. Paper No. 74 UTY.

[6] Chen, Mu-Chen, et al, 2005. *Mining Changes in customer behavior in retail marketing*. Published by Elseiver Ltd. Doi: 10.1016/j.eswa.2004.12.033

[7] Giudici, Paolo; & Passerone, Gianluca., 2002. *Data Mining of Association Structures to Model Consumer Behaviour.* Published by Elseiver Science. PII S 0167-9473(01)00077-9

[8] Indrajani., 2011. *Pengantar dan Sistem Basis Data.* Jakarta: PT. Gramedia

[9] Jogiyanto.,1999. *Analisis dan Desain Sistem Informasi: Pendekatan terstruktur teori dan praktek aplikasi bisnis.* Yogyakarta: Andi

[10] Kadir, Abdul., 1999. *Konsep & Tuntunan praktis Basis Data.*Yogyakarta: Andi Offset.

[11] Karismariyanti, Magdalena, dkk., 2007. *Aturan Asosiasi Spasial Pada Basis data Spasial.* Seminar Nasional Aplikasi Teknologi Informasi (SNATI 2007). ISSN: 1907-5022. Yogyakarta.

[12] Kristanto, harianto., 2004. *Konsep & Perancangan Database.* Yogyakarta: Andi offset.

[13] Kusrini; & Luthfi, Emha Taufiq., 2009. *Algoritma Datamining.* Yogyakarta: Andi Offset.

[14] Kusnawi., 2007. *Pengantar Solusi Data Mining.* Seminar Nasional Teknologi. ISSN: 1978-9777. Yogyakarta.

[15] Lacher, dkk., 1995. Theory and Methodology *A Neural for Classifying the Financial Healt of a Firm.* Elseiver Science. SSDI 0377-2217(93)E0274-2

[16] Larose, Daniel T., 2005. *Datamining Methods and Models.* United States of America : John Willey & Son, Inc.

[17] Luthfi, Emha Taufiq.,2009. *Penerapan datamining algoritma asosiasi untuk meningkatkan penjualan.* Jurnal Dasi volume 10 nomor 1. ISSN: 1411-3201

[18] Margahny; & Mitwaly.,2005. *Fast Algorithm for Mining Association Rules.* AIML 05 Conference, CICC, Cairo, Egypt

[19] Mitra, Sushmita., 2002. *Data Mining in Soft Computing framework : a Survey.* IEEE Transaction On Neural Networks, Vol. 13, No.1.

[20] Nan-Chen Hsieh., 2004. *An Integrated data mining and behavioral scoring model for analyzing bank customers.* Expert System with Application 27 (2004) 623-633, published by Elseiver Ltd, doi: 10.1016/j.eswa.2004.06.007

[21] Pramudiono, I., 2007. *Pengantar Data Mining : Menambang permata Pengetahuan di Gunung Data.*

[22] Sucahyo, Yudho Giri., 2003. *Penerapan Data Mining Permasalahan Apa Saja Yang Bisa Diselesaikannya.* Artikel Poluper IlmuKomputer.com

[23] Sugiyono., 2005. Statistika Untuk Penelitian. Bandung: Alfabeta.

[24] Sulianta, Feri; & Juju, Dominikasi., 2010. *Data Mining: Meramalkan Bisnis Perusahaan .* Jakarta: PT Elex media Komputindo.

[25] Susanto, sani; & Suryadi, Dedy.,2010. *Pengantar Data Mining, menggali pengetahuan dari bongkahan data.* Yogyakarta: Andi.

[26] Srikant, Ramakrisnan, et al.,1997. *Mining association Rules with item constraints.* American Association for Artificial Intelligence.

[27] Tsai, Chieh-yuan, at al., 2006. *A Change detection Model for Credit Card Usage Behavior.* Proceedings of the 5ᵗʰ WSEAS Int. Conf. On Computational Intellingence, Man-Machine Systems and Cybernetic, Venice,Italy, november 20-22, 2006 276

[28] Willyanto, Leo Santoso., 2003. *Pembuatan Perangkat Lunak Data Mining untuk Penggalian kaidah Asosiasi Menggunakan Metode Apriori.* Jurnal Informatika. Volume 4 nomor 2. November 2003 : 49-56.

[29] Wu, Xindong; & Kumar, Vipin., 2008. *The Top ten Algorithms in Datamining.* London: CRC Press

[30] Yusuf, Yogi, dkk.,2006. *Penerapan Datamining Dalam Menentukan Aturan Asosiasi Antar Jenis Item.* Seminar Nasional Aplikasi Teknologi Informasi (SNATI 2006), ISSN:1907-5022. Yogyakarta.