

Feasibility study for banking loan using association rule mining classifier

Agus Sasmito Aribowo^{a,1,*}, Nur Heri Cahyana^{a,2}

^a Department of Informatics FTI UPN Veteran Yogyakarta, Indonesia

¹ sasmito_skom@yahoo.com; ² ohmyon_nr@yahoo.co.id

ARTICLE INFO

Article history:

Received March 18, 2015

Revised March 25, 2015

Accepted March 31, 2015

Keywords:

Feasibility

Association Rule Mining Classifier

WIT-Tree

Banking Loan

ABSTRACT

The problem of bad loans in the koperasi can be reduced if the koperasi can detect whether member can complete the mortgage debt or decline. The method used for identify characteristic patterns of prospective lenders in this study, called Association Rule Mining Classifier. Pattern of credit member will be converted into knowledge and used to classify other creditors. Classification process would separate creditors into two groups: good credit and bad credit groups. Research using prototyping for implementing the design into an application using programming language and development tool. The process of association rule mining using Weighted Itemset Tidset (WIT)-tree methods. The results shown that the method can predict the prospective customer credit. Training data set using 120 customers who already know their credit history. Data test used 61 customers who apply for credit. The results concluded that 42 customers will be paying off their loans and 19 clients are decline.

Copyright © 2015 International Journal of Advances in Intelligent Informatics.
All rights reserved.

I. Introduction

Koperasi syariah growing rapidly in Indonesia. The activity of this koperasi is the provision of credit to small businesses and the poor. The Koperasi get greatest revenue from loans, buying and selling. Procedures to apply for a loan is to fill the application form and give some documents, such as personal data last, the data that is being run business and financial reports, family conditions / environment prospective partner. The data is used to predict whether the partners can complete pay the credit (repay financing) to the end or not.

Problems happened when the number of members and the number of credit applicants increasing. Based on interviews with the Koperasi BMT Subulussalam Bantul, Yogyakarta Indonesia, the most koperasi in Bantul area, there are about 60 similar koperasi and a quarter of them have bad credit problems.

Koperasi must identify the characteristics of prospective creditors troubled by studying the pattern of the troubled lender characteristics of existing customer data and the information from previous lending experience. Then, these patterns can be used as learning, if there is a borrower or lender has characteristics in common with the previous customer problems is the management of Koperasi will gain a valuable consideration for a more detailed and more careful in deciding the appropriateness customers receive credit funds.

One of informatics method to recognize patterns in the Koperasi data members, called data mining methods. Appropriate data mining methods used to identify characteristic patterns of prospective lenders are the methods of classification. This research uses a classification method, called Association Rule Mining Classifier. The results of this method is a characteristic pattern of troubled lenders. Classification must recognize new members who apply for credit by classifying new members into two groups, include 1) good credit, and 2) bad credit groups. If a potential creditor can be classified based on their profile earlier then it will be considered to pass a creditor receives money loan and this will certainly reduce the risk of bad debts.

This research aims to develop a decision support system model for classification the credit application to reduce the risk of bad debts using Association Rule Mining classifier method.

II. Related Work

Applied research related to data mining methods for bank credit has been investigated by several researchers. Data mining classification and bi-level programming were used for optimal credit allocation, and findings show that although the objective functions of the leader are worse in the bilateral scenario but agent banks collaboration is attracted and guaranteed [1]. Data mining for credit worthiness using decision tree. C4.5 algorithm used to classify the customer profile of bank credit. The classification is determined by the rate of return risk of credit with 6 criteria, namely: gender, occupation, income, savings, mortgage credit, and education. Data input is derived from credit transactions database. The application is also able to perform analysis or prediction to determine the success of the credit [2]. Application of the method of logistic regression modeling to predict credit risk on an individual basis [3]. Another study is the comparing three credit application scoring model is a model logistic regression (LR), models of classification and regression tree (CART) and a model of neural network (NN) to classify the credit application will be accepted and rejected. The results showed that the Neural Network models more valid and accurate [4], association rule mining with clustering used for loss profit estimation [5]. Research the use of WIT Tree for classification was also performed by Cahyana and Aribowo [6]. However, these studies used three dummy data as data set. An effective approach for mining weighted association rules from the weighted transaction database have been proposed, the results is finding the shorter average execution times, and also could remove some redundant rules [7].

III. Basic Theory

A. Decision support system

Decision Support System (DSS) proposed in the early 1970s by Scott Morton, who articulate the important concept of decision support systems. The concept of a decision support system is characterized by a computer-based interactive system that helps decision-making by utilizing the data module to select issues that are not structured [8]. Keen and Scott Morton argues that the decision support system is a computer-based support system for management decision makers who deal with the problems of unstructured, decision support systems combine the intellectual resources of individuals with computer capabilities to improve the quality of decisions [8]. Decision Support System is built to support a solution to a problem or to evaluate an opportunity. DSS as it is called DSS applications.

B. Association rule mining using WIT tree

1) Algoritma WIT- FWIs (Weight Itemsets Tidset- Frequent Weight Itemsets)

Apriori algorithm does not consider the divergence of interests every items in the transaction [9]. All item in the itemset have same interest's value. Example bread that give profit \$3 considered equal with milk that give profit only \$1.5. Le *et al.* applied new strategy in mining frequent item set. The strategy is called frequent weight item sets, which every item have different interest's value.

Algoritma WIT-FWIs consist of two basic concepts, weight itemset tid-set tree (WIT- Tree) data structure and frequent weight item sets (FWIs). WIT- Tree is tree data structure for exploring frequent weight item sets and FWIs for processing frequent weight item sets [9].

2) WIT- Tree (Weight Itemsets Tidset Tree)

WIT-Tree algorithm only scan the database once, because the next itemset can built from the intersection *Tidsets* and it can used to compute weight support for the next step [9]. It can reduce time reading large database. Before built WIT-Tree, the process must find *weight* item value in every transaction. If D is database and in its database contain transaction $T = \{t_1, t_2, t_3, \dots, t\}$, every transaction t consist itemset $I = \{i_1, i_2, i_3, \dots, i_n\}$. Every item have *weight* $W = \{w_1, w_2, w_3, \dots, w_n\}$ corresponds with every item in I, then the transaction weight can be calculated. Transaction weight (*tW*) every transaction t_k , defined as [9]:

$$tw(t_k) = \frac{\sum_{i_j \in t_k} w_j}{|t_k|} \tag{1}$$

Transaction weight (tw) in every transaction t_k in T, where T is transaction list in database defined as [9]:

$$ws(x) = \frac{\sum_{t_k \in t(x)} tw(t_k)}{\sum_{t_k \in T} tw(t_k)} \tag{2}$$

Example processing $ws(x)$ using transaction itemset shown on Table 1 and weight on Table 2 [9].

Table 1. A transaction itemset

Transaction ID	Itemset
1	A,B,C,E
2	B,C,E
3	A,B,D,E
4	A,B,C,E
5	B,C,D

Table 2. Weighted Items

Item	Weight (w)
A	0.6
B	0.1
C	0.3
D	0.2
E	0.9

Each transaction ID is calculated for determine the transaction weight, the example for transaction ID number 1 or $tw(t_1) = \frac{0.6+0.1+0.3+0.9}{4} = 0.475$. The results of transaction weight that calculated is stored in table 3.

Table 3. Transaction weight of all transaction in Table 1

Transaction ID	Transaction Weight (tw)
1	0.475
2	0.4
3	0.45
4	0.475
5	0.2
Sum	1.98

Item A appears in transaction {1, 3, 4}, the weight support item A $ws(A) = \frac{0.475+0.45+0.475}{1.98} = 0.7$. The step for build the WIT- Tree [9]:

- a. Build the vertex

Vertex is node–node in WIT Tree. Vertex x is build by user, and $t(x)$ is transaction contain x and ws is weight support x , vertex formulated by :

$$x \times t(x).ws(x) \tag{3}$$

- b. Determining the line from one vertex to another vertex at lower level. Example node A will joined with node {B}, {C},{E} and then $[A] = \{\{AB\},\{AC\},\{AD\},\{AE\}\}$. $[AB]$ will be class equivalence class if join with node $\{AC\},\{AD\},\{AE\}$.

Illustration the WIT Tree with minimum weight support=0,4 shown on Figure 1.

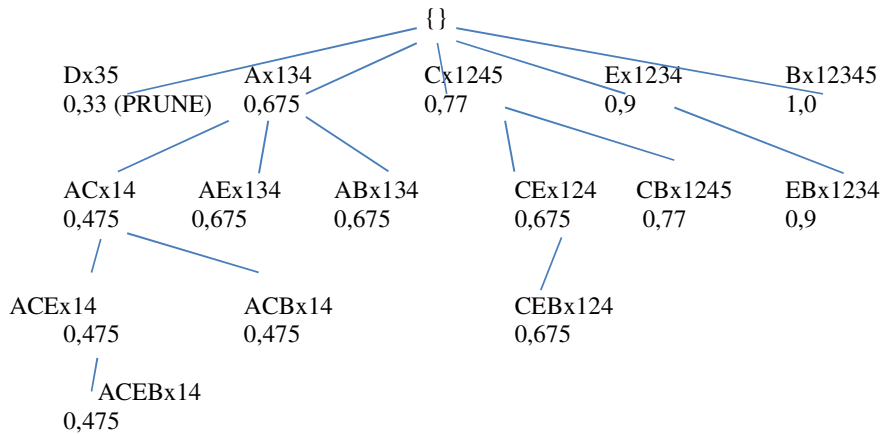


Fig. 1. WIT Tree search for transaction based on Table 1

3) FWIs (Frequents Weigth itemsets) Process

The following steps is FWIs process [9].

- a. First compute $ws(x)$ from every 1-itemset. The result are $ws(A)=0.675$, $ws(B)=1.0$, $ws(C)=0.77$, $ws(D)=0.33$, and $ws(E)=0.9$. Set of $ws(x)$ value from 1-itemset larger than *minimum ws* are $\{\phi\}=\{A,C,E,B\}$.
- b. A added to FWIs, and then FWIs contain A, or $FWIs=\{A\}$.
- c. A merged with C become new itemset ACx14 (it means itemset AC supported by transaction number 1 and 4) with $ws(AC)=0,475$. Because $ws(AC)$ larger than *minws* (0,4), then AC added to [A]. Then $[A]=\{AC\}$.
- d. A merged with E, and become new itemset AEx134 with $ws(AE)=0,675$. Then $[A]=\{AC, AE\}$.
- e. A merged with B, and become new itemset ABx134 with $ws(AB)=0,675$. Then $[A]=\{AC, AE, AB\}$.
- f. After finished grouping member of class [A] then gruping every member of class [A] recursifely.
- g. The first of member class [A] is grouping member of class [AC]. Added AC to FWIs, and then $FWIs=\{A,AC\}$.
- h. AC merged sith AE, and become new itemset ACEx14 with $ws(ACE)=0,475$. And the new member of class [AC]= $\{ACE\}$. AC join with AB, and become new itemset ACB with $ws(ACB)=0,475$. And the member of [AC]= $\{ACE,ACB\}$.
- i. FWIs member with $ws(x)$ larger *minimum Ws*=0,4 are $FWIs=\{A, AC, AE, AB, ACE, ACB, ACEB, C, CE, B, E, EB, B\}$.

IV. Results and Discussion

Data mining models to identify the characteristics of members who successfully complete the debt repayments and then generate the rule for predicted other members described in Fig. 2. The model requires the input of training data that example as table 4. Training data will be formed into WIT-tree. Then it will be processed using Frequent Weight Itemset to select the itemset that exceeded the minimum support. Itemset that $ws(x)$ larger than minimum support called frequent itemset, and it is become the knowledge that will be used to classify the data testing. Knowledge rules as in table 5 and the data testing in Table 6.

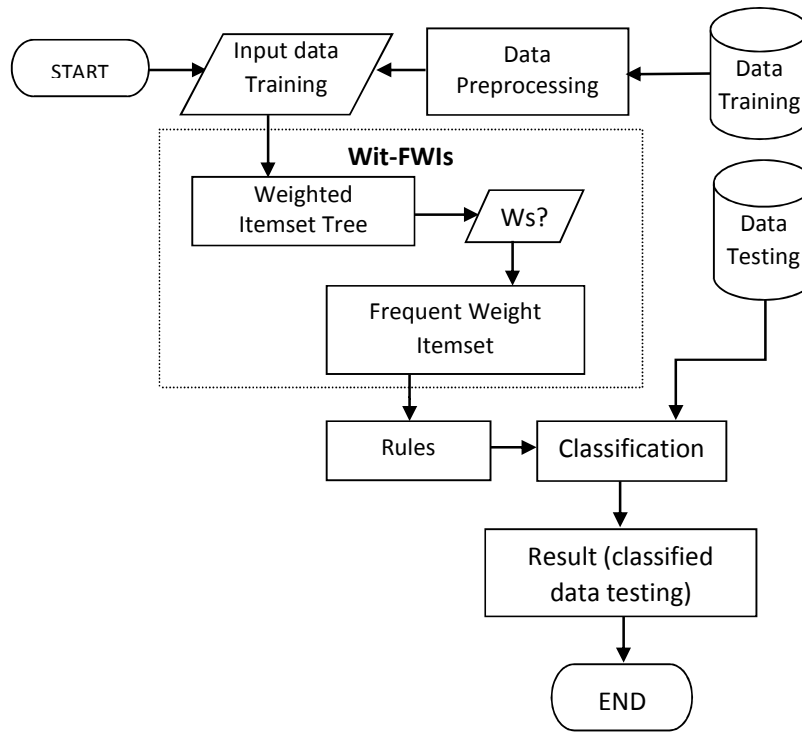


Fig. 2. Data Mining Model

Table 4. Dataset customer profile (as data training)

No	Graduate	Skill	Wellfare	Personality	Credit History
0001	High School	Ok	Poor	Good	Succed
0002	Primary School	Not Ok	Very Poor	Good	Not Succed
0004	Primary School	Ok	Poor	Moderate	Not Succed
0005	High School	Ok	Poor	Good	Succed
0006	Primary School	Not Ok	Poor	Good	Not Succed
0007	Primary School	Ok	Very Poor	Moderate	Not Succed
0008	Primary School	Ok	Poor	Bad	Not Succed
0009	Primary School	Not Ok	Poor	Good	Not Succed
0010	Primary School	Not Ok	Poor	Good	Not Succed
0011	Primary School	Ok	Poor	Bad	Not Succed
0012	High School	Ok	Poor	Good	Succed

Attribute class indicates success on the relevant credit. The second step a data mining process from Table 5 with the association rule mining algorithm, WIT Tree. The result of WIT Tree process is the rules (Table 6).

Table 5. The rule result from WIT-Tree process

No. Rule	Rule	Weight Support
R1.	(Graduate=HIGH SCHOOL)^(Wellfare=POOR)^(Personality=Good)→Succed	50%
R2.	(Graduate=JUNIOR HIGH SCHOOL)^(Wellfare=POOR)^(Personality=Bad)→not Succed	45%
R3.	(Graduate=HIGH SCHOOL)^(Wellfare=Moderate)^(Personality=Good)→Succed	40%
R4.	(Graduate=JUNIOR HIGH SCHOOL)^(Wellfare=POOR)^(Personality=Good)→Succed	35%
R5	(Graduate=PRIMARY SCHOOL)^(Wellfare=Very POOR)^(Personality=Moderate)→Not Succed	47%

Table 6. Data testing and prediction

No	Graduate	Skill	Wellfare	Personality	Rule	Prediction
1.	JUNIOR HIGH SCHOOL	Ok	POOR	Good	R2.	Succed
2.	HIGH SCHOOL	Not Ok	Moderate	Good	R3.	Succed
3.	JUNIOR HIGH SCHOOL	Ok	POOR	Good	R4.	Succed
4.	PRIMARY SCHOOL	Ok	Very POOR	Moderate	R5	Not Succed

The system implemented using Java J2SE programming languages and MySQL database. The results of WIT Tree formation exist in Fig. 3. The results of classification shown in Fig. 4.

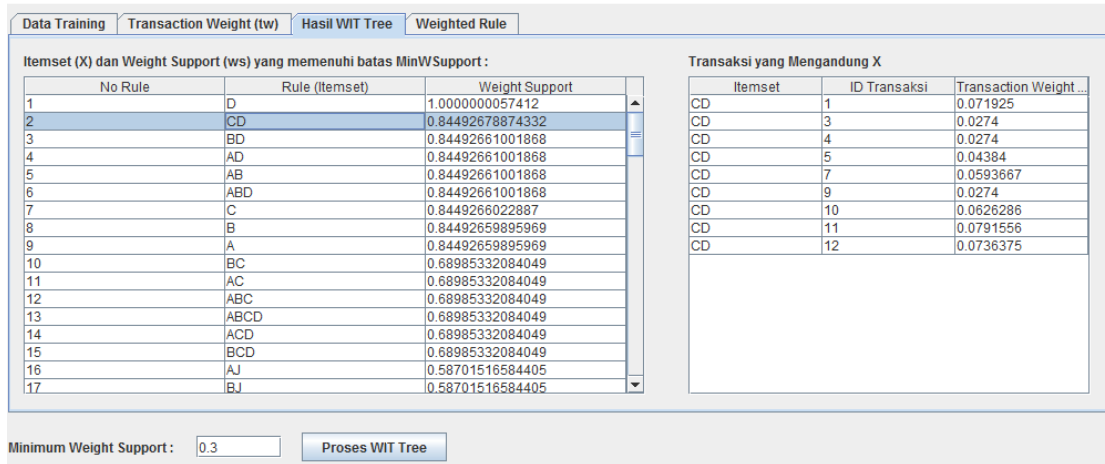


Fig. 3. Association process result, the rule and the TidSet

Training data used to build the rule is 120 customers who already know their credit history. Data testing used 61 customers who apply for credit. From the results of these calculations can be concluded that 42 customers will be paying off their loans and 19 clients do not succeed.

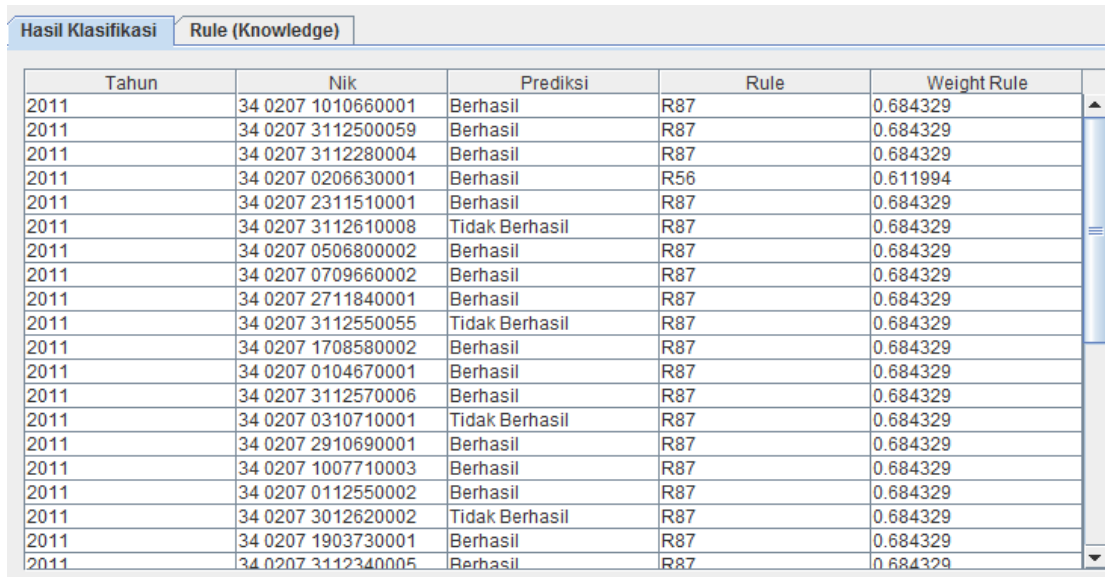


Fig. 4. Classification result

V. Conclusion and Future Work

Association rule mining classifier using WIT-Tree can be used to build rules from data member of Koperasi. Rules can be used for classify of new customers. This model can predict the prospective customer credit. The results of this model can reduce the risk of bad debts.

In this paper, we only improve the phase of mining koperasi data member by WIT-tree. In the future, we will study how to mine with more efficient association rules

References

- [1] S. Sadatrasou, M. Gholamian, and K. Shahanaghi, "An application of data mining classification and bi-level programming for optimal credit allocation," *Decis. Sci. Lett.*, vol. 4, no. 1, pp. 35–50, 2015.
- [2] A. H. Kusuma, "Penerapan Data Mining dalam Klasifikasi Nasabah Bank dan Analisis Resiko Persetujuan Proposal Kredit," Yogyakarta, 2008.
- [3] X. Liu and X. Zhu, "Study on the Evaluation System of Individual Credit Risk in commercial banks based on data mining," in *Communication Systems, Networks and Applications (ICCSNA), 2010 Second International Conference on*, 2010, vol. 2, pp. 308–311.
- [4] Y. Bee Wah and I. R. Ibrahim, "Using data mining predictive models to classify credit card applicants," in *Advanced Information Management and Service (IMS), 2010 6th International Conference on*, 2010, pp. 394–398.
- [5] M. Mittal, S. Pareek, and R. Agarwal, "Loss profit estimation using association rule mining with clustering," *Manag. Sci. Lett.*, vol. 5, no. 2, pp. 167–174, 2015.
- [6] N. H. Cahyana and A. S. Aribowo, "Decision Support System Untuk Kelayakan Kredit Menggunakan Association Rule Mining Classifier Di Koperasi Syariah BMT Subulussalam Bantul," Yogyakarta, 2014.
- [7] Z. Yue, W. Lijuan, and W. Ning, "Efficient weighted association rule mining using lattice," in *Control and Decision Conference (2014 CCDC), The 26th Chinese*, 2014, pp. 4913–4917.
- [8] E. Turban, *Decision support and expert systems 4th Ed.* Prentice-Hall, Inc., New Jersey, 1995.
- [9] B. Le, H. Nguyen, and B. Vo, "Efficient algorithms for mining frequent weighted itemsets from weighted items databases," in *Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2010 IEEE RIVF International Conference on*, 2010, pp. 1–6.