IOP Conf. Series: Materials Science and Engineering **332** (2018) 012018 doi:10.1088/1757-899X/332/1/012018

Application of decision rules for empowering of Indonesian telematics services SMEs

E T Tosida*, O Hairlangga, F Amirudin, M Ridwanah

Computer Science Department, Pakuan University, Jl. Pakuan PO Box 452, Bogor, Indonesia

*Email: enengtitatosida@unpak.ac.id, oktahairlangga.062@gmail.com, fadlanamirudin043@gmail.com, Mridwanah@gmail.com

Abstract. The independence of the field of telematics became one of Indonesia's vision in 2024. One effort to achieve it can be done by empowering SMEs in the field of telematics. Empowerment carried out need a practical mechanism by utilizing data centered, including through the National Economic Census database (Susenas). Based on the Susenas can be formulated the decision rules of determining the provision of assistance for SMEs in the field of telematics. The way it did by generating the rule base through the classification technique. The CART algorithm-based decision rule model performs better than C45 and ID3 models. The high level of performance model is also in line with the regulations applied by the government. This becomes one of the strengths of research, because the resulting model is consistent with the existing conditions in Indonesia. The rules base generated from the three classification techniques show different rules. The CART technique has pattern matching with the realization of activities in The Ministry of Cooperatives and SMEs. So far, the government has difficulty in referring data related to the empowerment of SMEs telematics services. Therefore, the findings resulting from this research can be used as an alternative decision support system related to the program of empowerment of SMEs in telematics.

1. Introduction

The Indonesian government is supporting business independence for *technopreneur* or entrepreneurs who are into the core businesses involving technology-based. Moreover, the start-up of Information and Communication Technology (ICT) or also known as telematics, is currently experiencing a significant increase. Self-reliance of start-ups based on telematics is needed to be one of the supporters of development in Indonesia. Self-reliance in the field of telematics became one of Indonesia's vision in 2024. One of Indonesia's vision in telematics is to realize an independent, competitive, dignified and ethical personality with function as an efficient meta-infrastructure, and ICT that provides opportunities and resources justly in achieving prosperity [1]. The increasing number of SME telematics is a potential that must be supported to have a competitive value, especially in the face of the ASEAN Economic Community (AEC).

Achieving that vision requires synergism from many parties. Particularly central or local governments responsible for implementing relevant national policies encourage independence for ICT-based technopreneurs. But there are many difficulties in deciding to develop technopreuneur through

the provision of assistance. A practical mechanism is needed in determining the provision of assistance to SMEs in desperate need of support. One of the causes of such difficulties is the amount of data that is out of sync. Another cause is the varying qualification standards in determining the feasibility of SMEs for assistance [2]. In fact, the Indonesian government has a centralized and consistent source of data related to SMEs, including SMEs in telematics. The data in question comes from the National Economic Census (Susenas) owned by the Central Bureau of Statistics (BPS). But so far data related to the condition of SMEs in Susenas precisely has not been utilized optimally for the activities of empowerment of SMEs.

Susenas data that holds data about SMEs of telematics services includes many attributes appropriate to the empowerment process [3]. Therefore, this research is proposed to give one way of empowering Indonesian SMEs telematics services through study of feasibility of rule-based aid based on Susenas data. Although the data used in this study is still using Susenas 2006 data, but the principles, ideas and the concept of empowerment through the utilization of Susenas data is relatively new. The Susenas data can be generated to form the decision rule of eligibility through the Data Mining approach, using the classification method. The classification technique used is a decision tree technique that is formed through three techniques namely J48 or C45 algorithm, Classification and Regression Tree (CART) and Iterative Dichotomies Tree (ID3).

Based on the trials conducted with the three classification techniques are expected to find a rule of thumb that is practical. The decision rule is expected to be utilized by the government and other stakeholders in an objective but comprehensive assistance process.

2. Method

The method applied in this research uses the stages of data mining or also known as Knowledge Discovery and Data Mining (KDD) [4]. As a series of processes, data mining can be divided into the several phases presented in Figure 1. These stages are interactive, the user is directly involved or through the knowledgebase mediation.

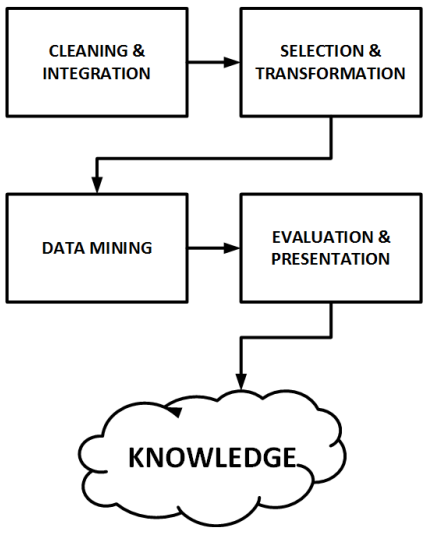

Figure 1. Steps of Data Mining [4]

Explanation of the steps undertaken in this study are as follows:

2.1. Data Cleaning and Integration

Data cleansing is the process of eliminating noise and inconsistent data or data irrelevant. As well as data integration is the combination of data from different databases into the new database. Not infrequently the data needed for data mining is not only derived from the database but also from multiple databases or text files, but this stage was done in previous studies by [2].

2.2. Data Selection and Transformation

The existing data in the database is often not all used, therefore only the appropriate data to be analyzed to be retrieved from the database. As well as the transformation of the data is modified or merged data into a format suitable for processing in data mining. For this study using entropy-based method for selection of the data. Selection of these attributes also according [5]. function to sort the attributes of the most influential in the process of classification so that it becomes more compact and optimized.

2.3. Mining Process

This stage is a major process when the method is applied to find valuable and hidden knowledge from data. To process the data mining of this study using Decision Tree algorithm, J48, CART and ID3. The performance of the three algorithms are analyzed through a high degree of accuracy and analytical relevance to SMEs competence condition Indonesian telematics services. The end result of this mining process is used to examine the decision-making rules of eligibility of assistance.

2.4. Pattern Evaluation

The process of identifying unique patterns and engaging into knowledgebased found done through the stages of evaluation pattern. In this stage the results of data mining techniques in the form of a distinctive pattern and prediction models were evaluated to assess whether the hypothesis that there is indeed reached. If the result is not in accordance with the hypothesis there are several alternatives that can be taken such as making feedback to improve the process of data mining, or accept this result as an out-of-expectation that may be useful. Evaluation of this study using the confusion matrix and proceed with the analysis of the relevance of SMEs competence condition Indonesian telematics services until 2006.

2.5. Knowledge Presentation

This stage is the visualization and presentation of knowledge about the methods used to obtain the knowledge acquired. The last stage of the data mining process is how to formulate a decision or action of the analytical results obtained. There are times when it must involve people who do not understand the data mining. Therefore presenting the results of data mining in the form of knowledge that can be understood by everyone is a necessary stage in the process of data mining. In this presentation, the results in the form of meaningful charts decision tree that contains multiple rules or rule is implemented in the form of a web-based system.

3. Result and Discussion

The result of this research is the classification system of SMEs business assistance of telematics service using J48, CART and ID3 algorithm then the algorithm is executed using R studio application using 80% training data and 20% test data from 8798 total data, thus generating decision tree to determine the feasibility decision a business entity to receive assistance. After obtaining the output of decision tree, rule or rule of the decision tree is implemented using adobe dream weaver then applied in the framework and database designed and created in the MySQL application.

This study is constituted by [2] and is a continuation of [6]. Data cleaning and integration were performed in accordance with the results of the study [6]. The selection of attributes is done gradually through a specific strategy, resulting from the resulting decision tree indicating a rule that is too large.

The first stage attribute selection is done by summarizing some attribute classes. Among them are the provincial attribute class which originally amounted to 31 provinces to 5 islands (Sumatra, Java, Kalimantan, Papua and Bali-Nusa Tenggara). The attributes of the owner's education which originally amounted to seven classes were summarized into four classes, as well as business group attributes (initially seven classes into four classes). This attribute class summary is intended as one of the rule base summary strategies that will be generated by the inference engine. The second attribute selection is done by trimming many attributes into 10 attributes with entropy-based techniques, which are implemented using R studio software. The result of attribute selection is presented in Table 1.

Table 1. Attributes Selection Result

No	Attribute	Value	
1	The form of legal entity	1. CV 2. Firm 3. Special permission from the relevant agencies 4. Cooperative	5. Individual 6. Limited Firm 7. Foundation
2	Business group	1. Telecomunication Services 2. Hardware consultation	3. Software consultation 4. others
3	Sales	1. Mikro 2. Small	3. Medium
4	Difficulty	1. Raw material / merchandise 2. Fuel oil /energy 3. Workmanship 4. Others 5. Marketing	6. Capitals 7. No difficulty 8. Transportation 9. Wage laborers
5	Cooperative	1. Yes	2. No
6	Partnership	1. Yes	2. No
7	Receive training	1. Yes	2. No
8	Marketing	1. Eksport 2. District	3. Outside of the district 4. Province
9	Business prospects	1. Better 2. Worse 3. Sama baik	4. Just as bad 5. Not comparable
10	Plan	1. No 2. Yes, expand business premises 3. Yes, opening branch	4. Yes, improve skills 5. Yes, others

The selected classification techniques J48, CART and ID3 have equality in the inference process, and each has deficiencies and advantages [5]. The purpose class for the inference process of SMEs assistance in telematics services is to convert from the attribute "never received help or never received help". Based on the results of this conversion is expected to obtain the base rules for determining eligible SMEs to be given assistance. Recapitulation accuracy of all three classification techniques using a variety of model scenario are shown in Table 2. Base the rules resulting from the three classification techniques showed different rules. The resulting rule base for J48 algorithm, CART and ID3 shown successively in Figures 2, 3 and 4.

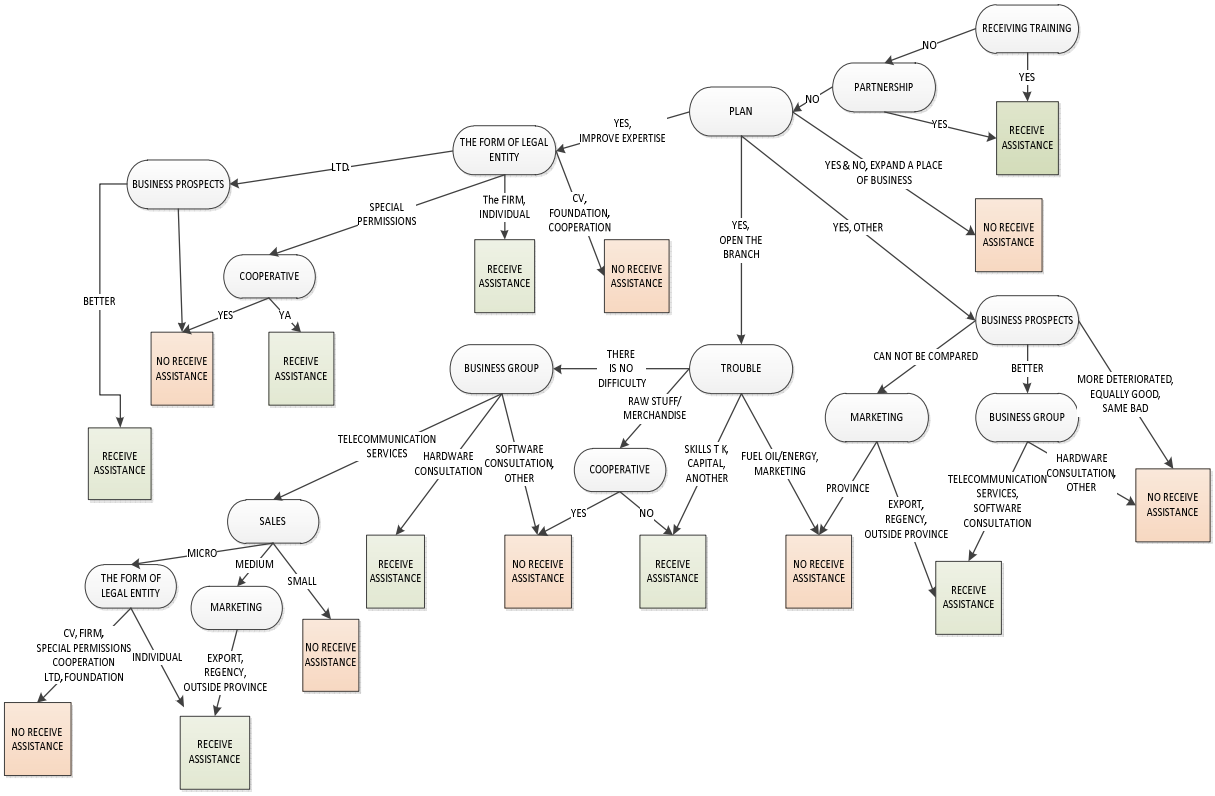

Figure 2. The rule base of SMEs assistance of Indonesian telematics services using J48 Algorithm

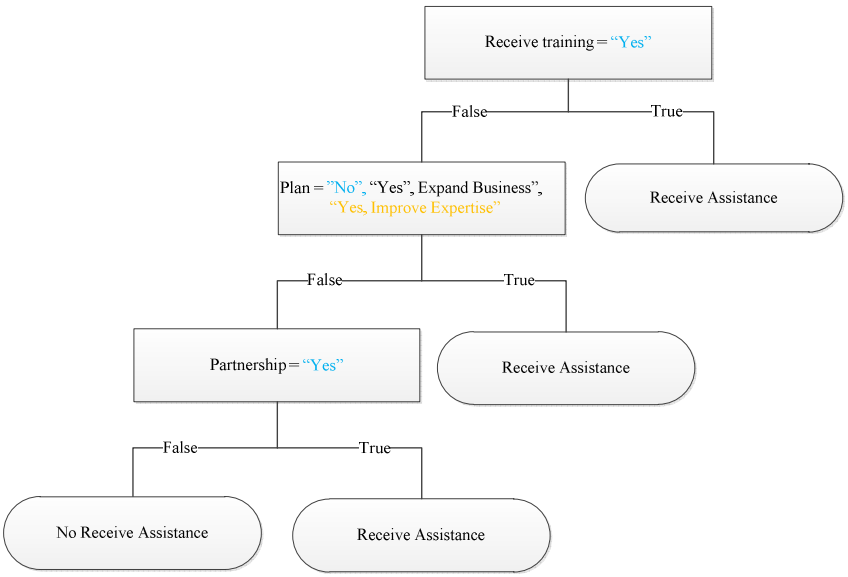

Figure 3. The rule base of SMEs assistance of Indonesian telematics services using CART Algorithm

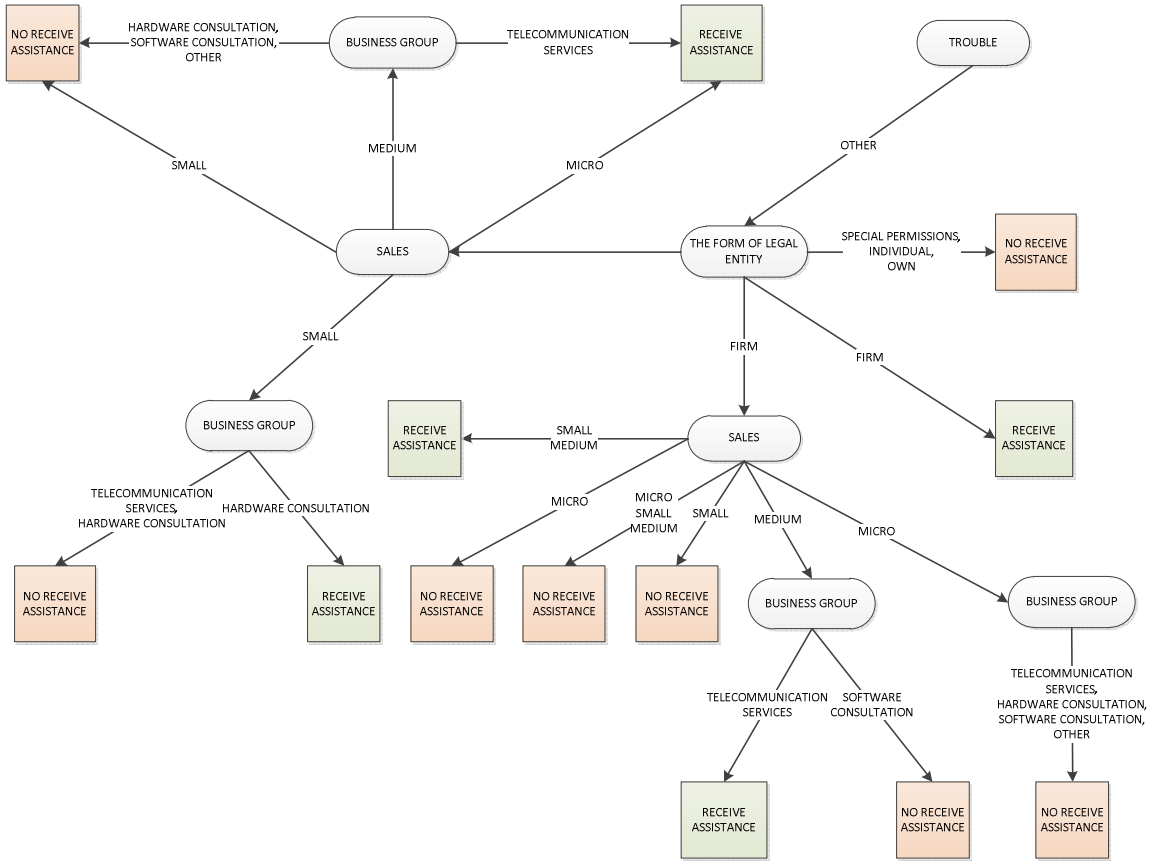

Figure 4. The rule base of SMEs assistance of Indonesian telematics services using ID3 Algorithm

The built-in model scenario is intended to improve model accuracy. Different model scenarios are used for each algorithm. In J48 algorithm the scenarios used are decision pruning process (scenario 1) and without pruning (scenario 2). In the CART and ID3 algorithms use different scenarios ie without data reduction (scenario1) and data reduction (scenario 2). Based on the information in Table 1 shows that each algorithm has a different performance scheme but produces almost the same decision rule. The performance of the J48 and ID3 algorithm-based decision model has better accuracy in the train data than the test data, for both model scenarios. Other conditions are shown in the performance of CART algorithm-based decision model. In the CART model the test data results in a higher accuracy for both model scenarios. This suggests that the CART algorithm has a better ability to classify the process of providing assistance to MSMEs Indonesian telematics services. This is an interesting finding, especially considering the data source used is Susenas data covering the condition of SMEs of Indonesian telematics services that represent all provinces.

Table 2. Model Performance

No	Algorithm	Accuracy (Scenario 1)		Accuracy (Scenario 2)	
		Training data	Testing data	Training data	Testing data
1	J48	64.88%	57.04%	79.71%	55.00%
2	CART	62.21%	73.97%	61.18%	63.92%
3	ID3	99.00%	24.00%	82.00%	26.00%

The performance results of this decision rule model are also in line with the policies adopted by regulators. Regulators in question in this case include Indonesian Communication & Information Ministry, Indonesian Cooperation and SMEs Ministry. Based on the referral of activities undertaken in Indonesian Cooperation and SMEs Ministry related to the selection for assistance for SMEs is based on training activities [7]. Training for start-ups of telematics has been facilitated through the techno park scheme [8]. This shows that CART technique has pattern match with realization of activity in Indonesian Cooperation and SMEs Ministry. But the rule base generated from these three classification techniques needs to be analyzed more deeply because it indicates the existence of unbalanced data on some attributes. Therefore obtaining the resulting accuracy is expected to be improved through the process of balancing data [9]. This research gives a very important result to be used as a reference in the process of providing assistance for SMEs telematics services. Because so far the government has difficulty in referring data related to the empowerment of SMEs telematics services. Therefore, the findings resulting from this research can be used as an alternative supporting decision related to the SMEs empowerment program in general.

4. Conclusion

The process of SMEs empowering through the provision of assistance for Indonesian telematics services SMEs can be modeled with Data Mining approach using J48, CART and ID3 classification techniques. This research is a new study, as far as we examine the various studies, especially for the determination of the rules of assistance for SMEs in the field of telematics services, one of them [10]. Although the approach is still using the classic decision rule approach but the strength of this research lies in the optimization of Susenas data which has been very rarely used for related research needs. Moreover, Susenas 2006 data is not equipped with the output of determinant class of assistance. The determination of the output class is done through the optimization of one of the attributes assumed as the output class (the attribute in question is 'ever get help'). Model performance is demonstrated by the accuracy of various schemes and data mining techniques used. The CART algorithm-based decision rule model performs better than C45 and ID3 models. The high level of performance model is also in line with the regulations applied by the government. This becomes one of the strengths of research, because the resulting model is consistent with the existing conditions in Indonesia. The weakness of this research is Susenas data which used is very old data (2006). This happens because the Susenas process in Indonesia is still done every ten years. In 2016 CBS has implemented the process of preparing the list of instruments and respondents and supporting components. Implementation of the census to the SMEs of Indonesian telematics services is only implemented in 2017, so that when the implementation of this study the data cannot be accessed and published to the public.

Acknowlegments

1. DRPM Ristek Dikti, as the main sponsor, which gives us Competitive Grants Scheme
2. Computer Science Department, Mathematics and Natural Science Faculty, Pakuan University,and Research Institute Pakuan University, for supporting, coordinating and facilitating to achieve this grants.
3. Indonesian Communication & Information Ministry, Indonesian Cooperation and SMEs Ministry and Bandung Technopark for active participation in the activities of interviews anduser requirement.

References

[1] Indonesian Communication & Information Ministry. Preparation of a long-term binding roadmap for development of the ICT sector. 2045 Towards 100 years of independent Indonesia. Research and Development Resources, Tools and Implementation of Post and Information Technology, Research and Human Resources. (2016).

[2] Tosida ET, Maryana S, Thaheer H, Damin FA. Visualization model of small and medium enterprises (SMEs) telematics services potentiality map in Indonesia. InInformation &

Communication Technology and Systems (ICTS), 2015 International Conference on 2015 Sep 16 (pp. 151-156). IEEE (2015).

[3] Tosida, E.T., Seminar, K.B. and Herdiyeni, Y. Attribut Selection of Indonesian telematic services MSMEs assisstance feasibility, using AHP. *Kursor*, *8*(2), 97-106 (2016).

[4] Han, J., Kamber M, Pei J. Data mining : Concepts and techniques. Third Edition. Morgan Kaufmann is an imprint of Elsevier, 225Wyman Street,Waltham, MA 02451, USA. (2012).

[5] Sohn SY, Kim JW. Decision tree-based technology credit scoring for strart-up firms : Korean case. *J. Expert Systems with Applications* 39 (4) : 4007-40112. Elsevier. doi:10.1016/j.eswa.2011.09.075.(2012).

[6] Tosida, E. T., S. Maryana, and H. Thaheer. Implementation of Self Organizing Map (SOM) as decision support: Indonesian telematics services MSMEs empowerment. In *IOP Conference Series: Materials Science and Engineering*, vol. 166, no. 1, p. 012017. IOP Publishing (2017).

[7] McGuirk H, Lenihan H, Hart M. Measuring the impact of innovation human capital on small firms propensity to innovative. *J. Research Policy, Vol.44 (4)* : 965-976. Elsevier. doi:http://dx.doi.org/10.1016/j.respol.2014.11.008. (2015).

[8] Soenarso Wisnu S, Nugraha D, Listyaningrum E. Development of science and technology park (stp) in indonesia to support innovation-based regional economy: Concept and early stage development. *J.World Technopolis Review (WTR).* 2 (1) : 32-42.World Technopolis Association. dioi:10.7165/wtr2013.2.1.32. (2013).

[9] Sun Z, Song Q, Zhu X, Sun H, Xu, Zhou Y. A novel ensemble method for classifying imbalanced data. Patter Recognition 48(5) : 1623-1637. http://dx.doi.org/10.1016/j.patcog.2014.11.014. (2015).

[10] Sadatrasoul, SM, Gholamian MR, Siami M, HajimaohammadiZ. Credit scoring in banks and financial institutions via data mining techniques : A literatur review. *Journal of AI and Data Mining*. Vol. 1, No. 2, 2013, 119-129. [Jan 5th, 2015]. Link : http://jad.shahroodut.ac.ir/article_124_0.html. (2013).