

PAPER • OPEN ACCESS

A hybrid data mining model for Indonesian telematics SMEs empowerment

To cite this article: E T Tosida *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **567** 012001

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

A hybrid data mining model for Indonesian telematics SMEs empowerment

E T Tosida¹, F Andria², I Wahyudin¹, R Widiyanto¹, M Ganda¹, R R Lathif¹

¹Computer Science Department, Faculty of Mathematic & Natural Science, Universitas Pakuan, Bogor, Indonesia

²Management Department, Faculty of Economy, Universitas Pakuan, Bogor, Indonesia

Corresponding author: enengtitatosida@unpak.ac.id

Abstract. The power of information technology and communication (telematics) is one of the vital forces for every country. In the Industrial Revolution 4.0 era, the development of telematics was one of the priorities of the Indonesian government. The development of the field of telematics in Indonesia for a decade is inseparable from the role of SMEs. The role of telematics SMEs in the strength of national development can be mapped through the optimization of National Economic Census data (Susenas). The detailed 2016 Susenas data has not been released by BPS. Therefore, this research still uses 2006 Susenas data. The 2016 Susenas recapitalization shows that Indonesian telematics has a very large power, consisting of 2.6 million players. This great strength needs to be optimized to have high competitiveness so as to be able to support Indonesia's development. The purpose of this study was to conduct hybrid data mining modeling to be used as a decision model in mapping the classification of Indonesian telematics SMEs. The classification map includes the feasibility of assistance for the empowerment of Indonesian telematics SMEs, business prospects and development plans for Indonesian telematics SMEs. The hybrid data mining model with K-Medoids & C4.5 technique shows better performance compared to other models, with an average accuracy rate of 71.87%. This model validation test also involves K-fold cross validations.

1. Introduction

Indonesia has economic power which is supported by SMEs. One of the rapidly developing SMEs is the SMEs in the field of telematics. In the joy of the Indonesian government, the field of telematics is one of the development priorities. This is in line with global economic conditions which are increasingly competing in the era of the 4.0 Industrial Revolution (RI). Moreover, this competition was strengthened by the birth of the Asian Economic Community. Since 2015 competition in the Asian Economic Community has become a benchmark for the economic success of countries in Asia.

Data from Indonesian telematics SMEs show considerable strength. The results of the 2006 National Economic Census alone show 12 thousand SMEs strengths in the field of telematics services [1] and have been mapped through web-based applications [2]. The scope of SMEs in Indonesian telematics services is stated in the Indonesian Minister of Industry Regulation No. 16 / M-Ind / PER / 7/2011. The Efforts to empower SMEs Indonesian telematics services have been modeled through data mining clustering and classification approaches [3][4][5]. This effort was carried out to identify the



strengths and weaknesses of Indonesian SME telematics services. The results of the identification can be used as a reference to focus the target of empowering SMEs, so as to increase competitiveness [6]. The most influential aspect of empowering SMEs in Indonesian telematics services is remuneration [5]. This will be closely related to the level of education of managers and human resources owned by SMEs as a whole [7]. This can be explored through the LVQ approach, and shows the performance of a very good model, with a high accuracy. This aspect is very relevant to the rapidly changing conditions of telematics technology. Empowerment of human resources is needed to answer the challenges of the competitiveness of Indonesian telematics SMEs. Moreover, it is associated with the challenges of MEA and RI 4.0 which are currently being faced by Indonesia and globally by all countries [8] [9] [10] [11]. Therefore research on the condition of SME telematics is very important, because SMEs telematics is one of the sources of development power in Indonesia. The results of the 2016 National Economic Census (Susenas) for telematics SMEs in Indonesia showed a very significant development. However, the results of the Susenas have not been fully accessible, due to constraints on the prevailing policies. The results of the 2016 Susenas in West Java, the number of telematics SMEs reached 62 thousand units. This condition needs to be studied more deeply so that this enormous strength can be optimized to support the nation's development in Indonesia.

In a previous research on the empowerment of Indonesian telematics SMEs, the data mining approach used was limited to single classifier techniques. This technique is able to be used to describe the important aspects and recommendations for empowering Indonesian telematics SMEs through the scheme of eligibility for assistance [3] [4] [5]. Efforts to build a model of feasibility of assistance for the empowerment of more accurate telematics SMEs are no longer significant [12]. Even though there are still many aspects that can be studied more deeply related to efforts to empower SME telematics, including aspects of development plans and business prospects [13]. Therefore in this study the development of a hybrid mining model is proposed for SMEs in Indonesian telematics services. The proposed hybrid mining model is carried out through two stages, namely clustering and classification. The first stage is clustering, which is done to classify telematics SMEs that have homogeneous characteristics related to the level of eligibility for assistance. The second stage, namely classification, is intended to identify the plans and prospects of Indonesian telematics SMEs. Thus the purpose of this study is to build a hybrid mining model for Indonesian telematics SMEs in order to empower them to face the competition of AEC and 4.0 RI.

2. Hybrid Mining Model for Management and Empowerment Strategies, and Data Mining Techniques Used in This Part

Empowerment of SMEs can be done through providing assistance. Assistance is provided in a variety of ways, and is tailored to the main needs of the SMEs. The process of providing assistance to telematics SMEs has a mechanism similar to the process of providing credit assistance at banks. Therefore the basic literature used in this study refers to [12] [14]. Research on the empowerment of SMEs that implement data mining models is still relatively small (only 1%) [14] [15]. The study of the application of hybrid mining models for empowering SMEs is still a potential and novelty for this paper. Special study on the empowerment of Indonesian telematics SMEs on an ongoing basis [1] [2] [3] [4] [5] [13].

The proposed hybrid mining model [12] uses two stages of the mining process. The first stage of the clustering model is used to classify accepted credit proposers and new proposers. The second stage of the classification model (SVM) is used to build a credit scoring model. With this hybrid mining model, rules for determining proposers can be identified that have the potential to be in arrears in credit, so that they can assist management in the process of risk management.

Hybrid mining models also applied into credit scoring [16]. In the research carried out through three mining stages, the first stage performs feature selection and is tested using SVM classification technique, the best feature selection performance is shown by PCA (compared to Genetic Algorithm, information gain ratio and relief. The second stage processes the credit scoring model using artificial neural network (ANN) Adaptive Boosting (There is a boost) with the best FS technique (PCA). The

performance of this hybrid mining model shows a model that is performing strongly especially for the credit scoring model. The application of hybrid mining models to the survey data on job career dynamics has produced a potential model to be implemented [17]. The proposed method is a model-based combination and heuristic clustering. The model is capable of forming cluster sequences without losing information about their dynamics. There are two the research phase, first transformed the categorical data sequences using hidden markov models, so that it formed into probabilistic data. The second stage carried out the clustering process using hierarchical clustering. The identification strategy of factors that influence the management of innovation was carried out through the application of hybrid mining models Search Algorithm (HGA) combined with K-NN can eliminate factors that are irrelevant in determining innovation performance decisions of manufacturing companies [18].

The application of hybrid mining models has been carried out [19] by combining M5 decision tree and evolutionary tree for categorical data and regression algorithms for numeric data, showing better model performance. The model has been validated using 12 data set decision problems. Ada also [20] proposed a hybrid mining model using artificial neural network (ANN) and self-organizing map (SOM) to predict turnover rate technology professionals. This can affect the performance of human resources to improve organizational performance. In this study proposed a hybrid mining model for empowerment of Indonesian telematics SMEs services that have mixed data in the form of numeric and categorical. The study was conducted in two stages, and used a combination of clustering techniques (first stage) and classification (second stage). There are 3 hybrid mining scenarios, namely k-means - interactive dichotomizer three (ID3), k-medoids - C45 and k-medoids - C5. Based on the three scenarios selected which have the highest level of accuracy and tested its relevance to the real conditions and regulations that occur in Indonesia.

Clustering is one of the unsupervised learning techniques and aims to group data or objects into clusters, so that each cluster will be filled with data as closely as possible. A good clustering method will produce high cluster quality with a high degree of similarity in one class and a low level of similarity between classes. The quality of clustering results is very dependent on the size of the similarity used by the method and its implementation. The quality of the clustering method is also measured by its ability to find some or all hidden patterns [21].

k-means is one of the non-hierarchical data clustering methods that partition data into one or more clusters / groups, so that data that has the same characteristics are grouped in the same cluster and data that has different characteristics are grouped into other groups [21]. K-means algorithm is :

1. Determine the number of clusters
2. Calculating the centroid
3. Calculating distance (distance between data and centroid). Use the euclidean distance formula as follows.

$$D_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \dots\dots\dots(1)$$

Information :

D_{ij} = Distance between object i and j

p = Data Dimension

X_{ik} = coordinates of object i in dimension k

X_{jk} = coordinates of object j in dimension k

k = amount of cluster

4. Group by minimum distance

5. If there is a moving object, then do the next iteration again by repeating steps 2 to 4, if it is not there then the calculation is complete.

The k-medoids algorithm, also known as Partitioning Around Medoids (PAM), is a variant of the k-means algorithm. This is based on the use of medoids rather than the use of means possessed from each cluster, with the aim of reducing the sensitivity of the partition produced in relation to the extreme values present in the dataset [21] [22]. The formula for calculating the distance k-medoids is referred to in equation (1) with Euclidean Distance. PAM algorithms include the method of partitioning clustering to group a set of n objects into a cluster of k clusters. Cluster representation in PAM is an object of a set of objects representing clusters called medoids. Clusters are built by calculating the proximity between the medoid and non-medoid objects using a distance measure. So the method of partitioning can still be done based on the principle of minimizing the number of dissimilarities between each object and medoid that are appropriate and this which forms the basis of the PAM method [21] [22]. Whereas to calculate the value of difference (S) is stated by equation 2 as follows:

$$S = \text{new total cost} - \text{long total cost} \dots\dots\dots(2)$$

Where :

new total cost : Amount of cost / distance of non-medoids

long total cost : total cost / distance of medoids

The process of validating the clustering model can be done through an internal approach, and the Davies Bouldin Index (IDB) is one of them [23]. This validation technique is based on maximizing the inter-cluster distance and minimizing intra-cluster distance. It is denoted by the cluster average (μ_i), the dispersion of points around the cluster average (σ_μ), and the total variant of $\text{var}(C_i)$, then the following equations (3) and (4) are obtained:

$$\mu_i = \frac{1}{n_i} \sum_{x_j \in C_i} x_j \dots\dots\dots(3)$$

$$\sigma_{\mu_i} = \sqrt{\frac{\sum_{x_j \in C_i} \delta(x_j, \mu_i)^2}{n_i}} = \sqrt{\text{var}(C_i)} \dots\dots\dots(4)$$

DB in equation (5) measures the ratio of clusters C_i and C_j with the equation:

$$DB_{ij} = \frac{\sigma_{\mu_i} + \sigma_{\mu_j}}{\delta(\sigma_{\mu_i}, \sigma_{\mu_j})} \dots\dots\dots(5)$$

DB_{ij} measure the compactness of clusters compared to the distance between cluster averages, so IDB is defined by the equation (6):

$$DB = \frac{1}{k} \sum_{i=1}^k \max\{DB_{ij}\} \dots\dots\dots(6)$$

The smallest DB value is the best clustering result. Between clusters are well separated (high cluster average distance), and each cluster is well represented based on its average value (low distribution value).

Classification is one method that is used as a method of supporting decision making by looking at the behavior and attributes of defined groups. This technique can provide classification of new data by manipulating existing data that has been classified and by using the results to provide a number of rules [21].

ID3 algorithm uses the concept of information entropy. This algorithm conducts greedy searches on all possible decision trees. Algorithms in this method use the concept of entropy. The Entropy concept is used to measure "how informative" a node (which is usually called how good). Entropy (S) = 0, if all instances in S are in the same class. Entropy (S) = 1, if the number of samples is positive and the number of negative samples in S is equal. $0 < \text{Entropy}(S) < 1$, if the number of examples positive and negative in S are not the same.

C4.5 algorithm is a group of Decision Tree algorithms. This algorithm has input in the form of Samples and Samples Training. Samples Training is sample data that will be used to build a tree that has been tested for its truth. Whereas Samples is data fields that will later be used as parameters in classifying data [22]. The calculation formula of the C4.5 algorithm is referenced in equations 7 and 8.

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i \log_2 p_i \dots \dots \dots (7)$$

Information :

S = set of case

n = amount of S partition

p_i = proportion from the S_i to S

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \text{Entropy}(S_i) \dots \dots \dots (8)$$

Information :

A = attribute

n = number of partition attributes of A

$|S_i|$ = number of cases on the i -partition

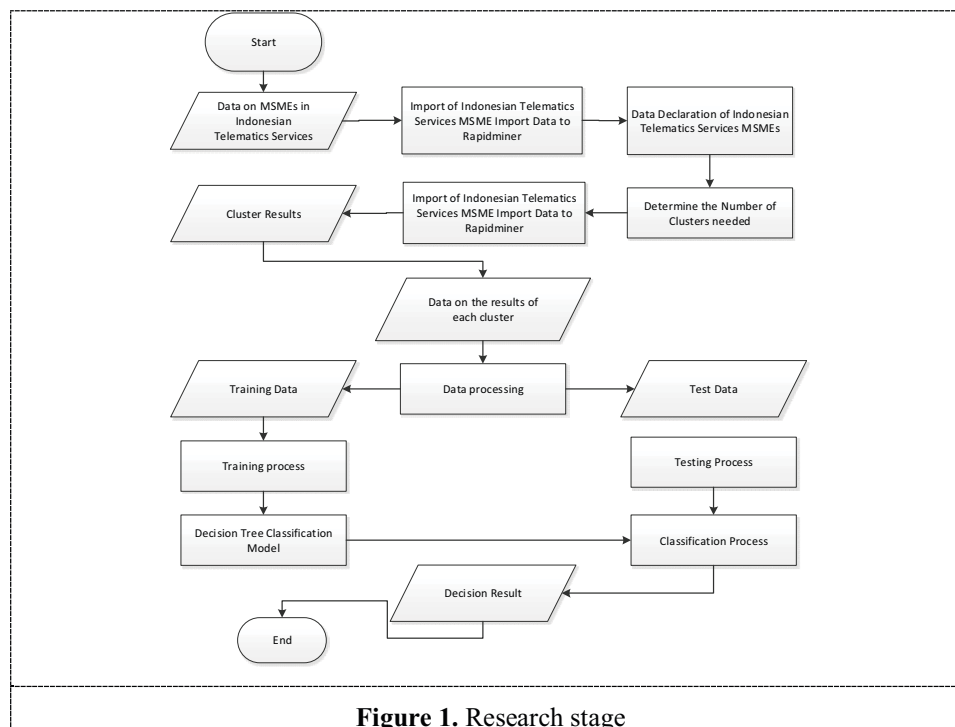
$|S|$ = number of cases in S .

C5.0 algorithm is one of the data mining algorithms that are specifically applied to the decision tree technique. C5.0 is an improvement of the previous algorithm formed by Ross Quinlan in 1987, namely ID3 and C4.5. In this algorithm the selection of attributes will be processed using information gain. In selecting attributes for object breakers in several classes attributes must be chosen which produce the greatest information gain. Attributes with the highest information gain value will be selected as parent for the next node [21].

3. Methods

This research is based on the initial research in the form of a map-based Information System of SME telematics services [2] and factor analysis that influence the provision of assistance through the Multi Attribute Decision Making (MADM) [1][13]. Pre-process data also refers to previous research using single classifier [4] [5] and single clustering [3]. The data used is Susenas 2006 based, because until now the 2016 Susenas data has not been accessed in detail.

The hybrid mining model is proposed to delve deeper into the conditions of Indonesian telematics SMEs, through the determination of the factors presented in the Susenas results. The Susenas results provide data on the condition of SMEs, including HR conditions, basic technology, remuneration, training, business difficulties, partnerships, prospects, and business development plans [24][25][26] [27]. The focus of this study is aimed at predicting prospects and business development plans that are currently not widely discussed by other studies. This factor can be used as a reference for the second stage of the empowerment process after SMEs telematics services are grouped through the clustering process. The establishment of hybrid mining models is carried out using Rapidminer. The overall stages of research are presented in Figure 1.



The stages of the research were carried out for all scenarios. The scenario that produces the best hybrid mining model performance is then implemented in HTML, PHP, and My SQL programming as the database.

4. Result and Discussion

The characteristics of SMEs in Indonesian telematics services in 2016 did show significant changes but were still dominated by internet cafe businesses. Because the detailed data of the 2016 Susenas results are not yet accessible, this paper is still based on the 2006 Susenas results. The basic characteristics of the 2006 and 2016 Susenas data patterns still have similarities, only the dominance of SMEs in Indonesian telematics services shifts from kiosks to internet cafes. This condition shows that the growth of SMEs in Indonesian telematics services does not yet have a strong foundation for the basic framework of RI 4.0 [7] [8]. However, from the results of the mining hybrid model this time it can be shown the condition of telematics service SMEs that have business prospects and development plans. That description became the novelty of this research.

The results of the clustering process as the first stage of the formation of hybrid mining models indicate that the number of clusters that are more optimal for testing this model are 5 clusters. This is also in line with the results of the study [3]. The use of a number of 5 clusters is also proven by the results of the validation test using IDB [23] for all three scenarios. The IDB level trial was conducted on three alternative number of clusters, shown in Table 1.

Table 1. Test the IDB validity of the clustering process

Amount of cluster	IDB value		
	k-means – ID3	k-medoids – C45	k-medoids – C5
3	3.397	2.084	2.074
5	3.262	2.060	2.050
7	3.448	2.075	2.083

This clustering process generates groups of SME telematics services that are clustered based on the similarity of their characters. The intended character shows the level of feasibility to be given assistance [25][26]. The clustering results are input to be processed in the second stage, namely the classification process. The classification process is used to see the prospects and plans for the development of Indonesian telematics services SMEs that have not been widely studied so far in previous studies.

The process of forming the classification learning machine this time was tested using the k-fold cross validation scenario and without k-fold cross validation. The scenario of the distribution of training data and test data is very important to see the opportunities for even distribution of data in each scenario performed [19]. Moreover, the data set used is data imbalance, which is indicated to affect the performance of the hybrid mining model. The k-fold cross validation scheme used is 5-fold cross validation, this is adjusted also by the number of clusters used. The performance of hybrid mining models is shown in Table 2.

Table 2. Comparison of the accuracy of hybrid mining models

Cluster	Accuraction level (%)					
	k-means – ID3		k-medoids – C45		k-medoids – C5	
	a	b	A	B	a	b
1	36.45	91.90	58.36	68.42	44.87	49.21
2	35.03	90.99	59.41	68.55	48.18	51.87
3	33.72	97.58	59.78	80.59	0.08	94.24
4	35.12	93.42	59.20	73.68	46.42	48.99
5	35.53	94.96	59.00	68.11	59.36	70.48

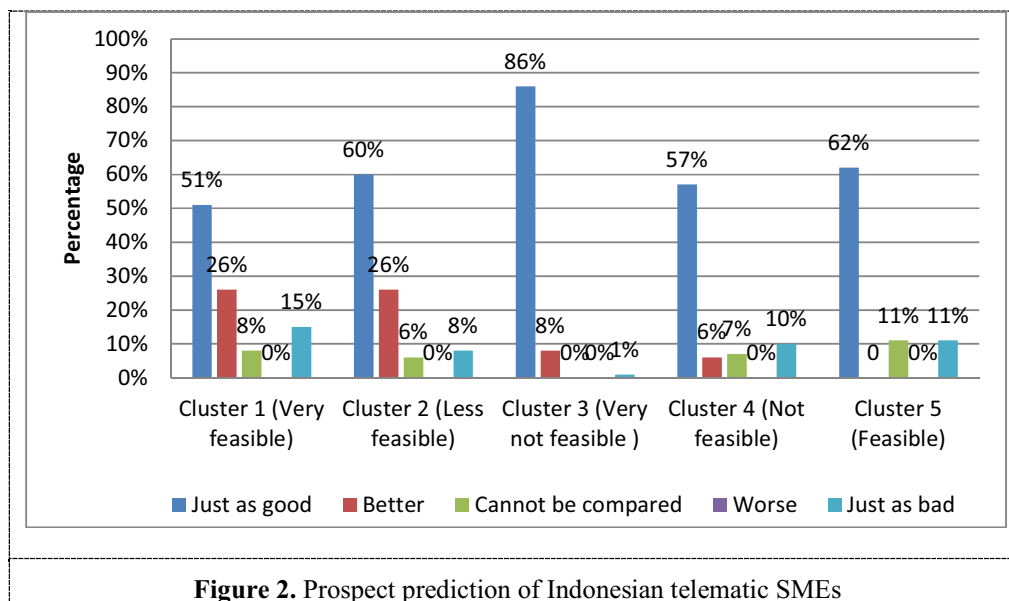
^ak-fold cross validation

^bno k-fold cross validation

Based on Table 2. It shows that the 5-fold cross validation scheme has a significant effect on the performance of the hybrid mining model. All scenarios show that the accuracy of the hybrid mining model has a lower level of accuracy when the 5-fold cross validation is applied. K-means - ID3 hybrid mining model shows a very significant difference in performance after being tested 5-fold cross validation. The decline in model performance in the first scenario averaged 58.6%. This condition shows the model's instability, and has a high sensitivity to changes in the distribution of training data and test data. In the second scenario (k-medoids - C45) showed almost the same condition, but the decline in performance was relatively lower (12.72%). The second scenario provides better model performance, and this is also in line with research [12] [19]. The third scenario (k-medoids - C5) gives an unexpected picture, because the 5-fold cross validation test results provide better performance. However, there are indications that the distribution of clustering results results in an imbalance

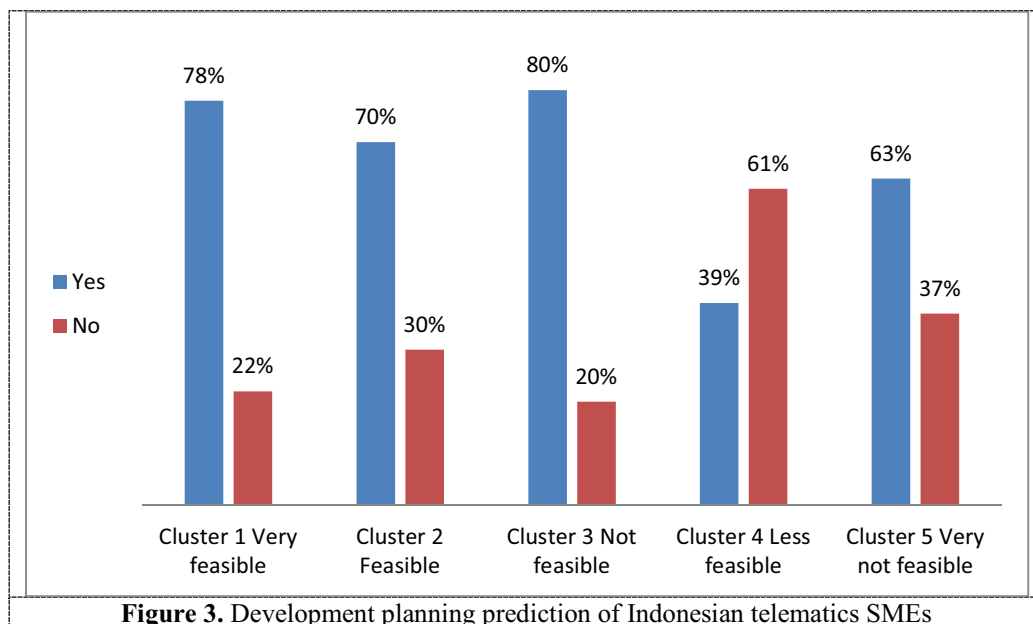
resulting in the performance of the third scenario being more unstable. This is also indicated by the performance of the model which only achieved an average accuracy rate of 39.78%. The decline in the performance of the hybrid mining model for the third scenario also showed a significant difference to reach 23.17%.

Analysis of the hybrid mining model is then carried out for the selected model, the model formed from the second scenario. Based on the results of the implementation of the hybrid mining model, the conditions for the prospects of SMEs in Indonesian telematics services were obtained as shown in Figure 2.



Prediction of the prospect of SMEs in Indonesian telematics services based on the results of the hybrid mining model capable of describing conditions relevant to government policies related to the provision of assistance and empowerment of SMEs [24][25][27]. In cluster 1, it is a group of SMEs that are very feasible to be given assistance, there are 26% of SMEs that have better business prospects. This can be used as a reference for the government or even those who will provide assistance so that this group is the priority of the beneficiaries. In this study, the types of assistance that are really needed by SMEs have not been identified. But the latest findings related to the focus and priority groups that will be given assistance are seen from the business prospect factors, which have not been able to be discussed in previous studies [3][4][5].

The description of other conditions that can be extracted from the results of implementing this hybrid mining model is being able to predict business development plans for SMEs in Indonesian telematics services. This condition is detailed in Figure 3. The business development plan is an important factor for evaluating the feasibility of providing assistance to SMEs, so that the assistance provided is truly a generator of empowerment to increase the competitiveness of SMEs. The assistance provided is not only transactional and grants, but must be able to provoke and generate SMEs to improve the quality of their business, so they can be highly competitive. Moreover, the field of telematics as one of the priority areas for MEA competition and facing IR 4.0 is very relevant if the strength of its SMEs needs to be improved. The ability to develop a business development plan is certainly very influenced by the creativity of human resources owned by SMEs. This cycle plays an important role in empowering SMEs, so it is expected that Indonesian telematics SMEs are able to make a real contribution in the development of human resources in particular, even the economic development of Indonesia in general [6][7][13][24][27].



In Figure 3 shows that the focus of providing assistance can be assigned to cluster 1 and cluster 2 which are predicted to have business development plans. The types of business development plans of each SMES are certainly different. This finding is also one of the novelty, because in previous studies [3] [4] [5] have not been able to dig up information to this information. The shortcomings of this system have not been able to show in detail the types of business development plans of the SMES in question.

This hybrid mining model is implemented on a user friendly interface system. This is intended to facilitate stakeholders who will use this system. The process of providing assistance and empowerment of Indonesian telematics services SMEs can be done easily, accurately and is able to provide more detailed recommendations regarding the priorities of beneficiaries (especially those related to prospects and business development plans of each assessed SMES). This system interface is shown in Figure 4.



The user interface for loading data is shown in Figure 5. The process of loading data can be input independently directly using this feature, (specifically for small amounts of data, suppose data only amounts to under 10 data). If the data to be loaded is grouped data, the user can use the "import data"

facility so that the loading process is more effective and efficient. This feature has also not been developed in the previous system [5].

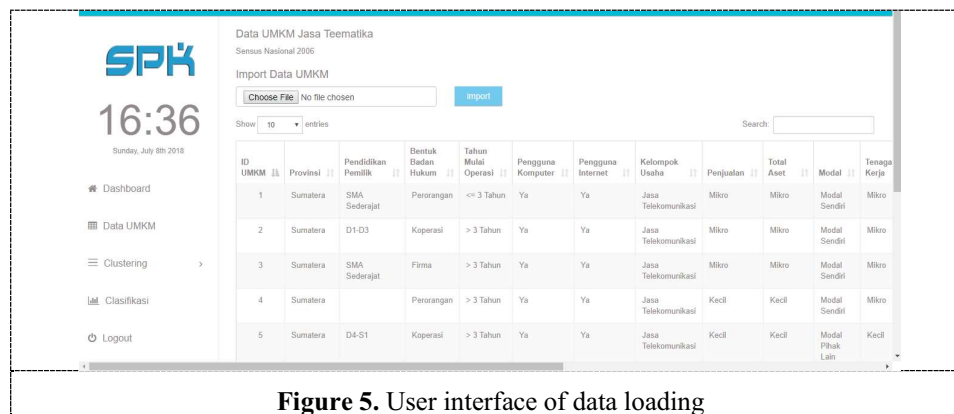


Figure 5. User interface of data loading

In this system the process of supporting decisions using hybrid mining models is carried out simultaneously. The clustering process (using the k-medoids technique) in this system produces SME clusters grouped based on the level of eligibility for assistance. The next process of the system will immediately display the prioritized clusters to predict prospects and business development plans using the C45 classification technique. The use of simple clustering and classification techniques is able to provide an accuracy value of 71.87%.

One disadvantage of this system is the process of predicting prospects and business development plans for SMEs that are assessed, still carried out in parallel (each one data inputted). The input data will be read by the learning machine, then stored by the system and will improve the next hybrid mining model that is formed. This developed system has not been able to generate predictive processes in groups, because the hybrid mining model implemented is still carried out separately using rapidminer. The lack of this system can be used as input for the construction of subsequent research.

5. Conclusion

The empowerment of Indonesian telematics SMEs will have a positive impact on increasing the nation's competitiveness, especially in the face of MEA and IR 4.0 competition. Indonesia has enormous strength because it is supported by a large number of telematics SMEs. The strength of this SMEs has proven to be a pillar of the Indonesian economy in the face of the monetary crisis. Therefore the government and other stakeholders need to have a good scheme in the process of empowering SMEs, especially SME telematics. The results of this study provide an alternative process for assessing the feasibility of assistance for SMEs in Indonesian telematics services using hybrid mining models.

The hybrid mining model that is implemented into decision support systems is the best model of the three scenarios analyzed. The use of k-medoids and C45 techniques which are simple data mining techniques are still able to produce model performance with an accuracy rate of 71.87%. The system is built with a user friendly interface, making it easier for stakeholders to use it. The lack of this system has not been able to show more specific business development (which refers to the development of certain fields). Another disadvantage of this study is that the model formation process is still separate using rapidminer, so the prospect prediction process and business development plan are still carried out in stages. The system has not been able to generate predictive processes for large groups of data, because the process of synchronizing new data into the renewal of hybrid mining models is still simultaneous. The results of this study have been able to provide novelty which is able to provide priority information for SME groups of telematics services that deserve to be given assistance by showing predictions of prospects and business development plans. Thus the focus of empowerment

will be more targeted and the nature of the provision of assistance will be mutually reinforcing in line with the prediction of prospects and business development plans owned by the SMEs.

Acknowledgments

1. DRPM RistekDikti, 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. Agro Industrial Engineering IPB and Computer Science IPB who have become the companion research team in this research grant scheme
4. Vocational Schools (200 schools) in Jabodetabek, BPS, ICT Associations and any stakeholders for active participation in the activities of interviews and user requirement

References

- [1] Tosida, ET, Seminar KB, Herdiyeni Y. 2016. Atribut selection of Indonesian Telematic Services MSMEs Assistance Feasibility, using AHP. *Kursor* 8 (2), 2016. DOI: 10.21107/kursor.v8i2.1299.
- [2] Tosida ET, Maryana S, Thaheer H, Damin FA. 2015. Visualization model of Small Medium Enterprises (SMEs) telematics services potentiality map in Indonesia. *International Conference on Information & Communication Technology and Systems (ICTS)*, 2015. DOI: 10.1109/ICTS.2015.7379890.
- [3] Tosida ET, Maryana S, Thaheer H. 2017a. Implementation of Self Organizing Map (SOM) as decision support: Indonesian telematics services MSMEs empowerment. *IOP Conference Series: Materials Science and Engineering* 166 (1), 012017.
- [4] Tosida ET, Hairlangga O, Amirudin F, Ridwanah M. 2018a. Application of Decision Rules for Empowering of Indonesian Telematics Services SMEs. *IOP Conference Series: Materials Science and Engineering* 332 (1), 012018.
- [5] Tosida ET, FD Wihartiko, I Lumessa. 2018. Learning Vector Quantization Implementation to Predict The Provision of Assistance for Indonesian Telematics Services SMEs. *International Journal of Engineering & Technology*, Vol. 7 No 3. 20
- [6] Jaswadi, Iqbal M, Sumiadji. 2015. SME Governance in Indonesia – A Survey and Insight from Private. *Procedia Economics and Finance* 31 (2015) 387 – 398.
- [7] Andra Benesova, Jiri Tupa. 2017. Requirements for Education and Qualification of People in Industry 4.0. *Procedia Manufacturing* 11 (2017) 2; 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy. doi: 10.1016/j.promfg.2017.07.366
- [8] Jian Qin, Ying Liu, Roger Grosvenor. 2016. A Categorical Framework of Manufacturing for Industry 4.0 and Beyond. *Procedia CIRP* 52 (2016) 173-178. Elsevier. doi: 10.1016/j.procir.2016.08.005
- [9] Sommer L. 2015. Industrial Revolution - Industry 4.0: Are German Manufacturing SMEs the First Victims of this Revolution?. *Journal of Industrial Engineering and Management JIEM*, 2015 – 8(5): 1512-1532. DOI: 10.3926/jiem.1470.
- [10] Jones C, Pimdee P. 2017. Innovative Ideas: Thailand 4.0 and the fourth industrial revolution. *Asian International Journal of Social Sciences*, 17(1), 4–35. DOI: 10.29139/aijss.20170101.
- [11] Pereira T, L Barreto, A Amaral. 2017. Network and Information Security Challenges within Industry 4.0. *Paradigm*. 2017. *Procedia Manufacturing* 13 (2017) 1253-1260.
- [12] Chen W, Xiang G, Liu Y, Wang K. 2012. Credit risk Evaluation by hybrid data mining technique. *System Engineering Proceedings* 3 (2012), 194-200. DOI: 10.1016/j.sepro.2011.10.029.
- [13] Tosida ET, H Thaheer, S Maryana. 2017. Strategi Peningkatan Daya Saing melalui Framework Rantai Nilai untuk Kompetensi Usaha Jasa Telematika Indonesia. *Jurnal Penelitian Pos dan Informatika*, 5 (1), 1-18.

- [14] Sadatrasoul, SM, GholamianMR, Siami M, HajimaohammadiZ. 2013. Credit scoring in banks and financial institutions via data mining techniques: A literatur review. *Journal of AI and Data Mining*. Vo. 1., No. 2, 2013, 119-129.
- [15] Zhu, Y., C. Xie, Gang-Jin Wang, Xin-Guo Yan. 2016. Comparison of individual, ensemble and integrated machine learning methods to predict China's SME credit risk in supply chain finance. *Neural Comput & Applie*. The Natural Computing Applications Forum 2016. CorssMark, Springer. DOI 10.1007/s00521-016-2304-x.
- [16] Koutanaei FN, H Sajedi, M Khanbabaei. 2015. A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *Journal of Retailing and Consumer Services*, Vol. 27, November 2015, 11-23. DOI: 10.1016/j.jretconser. 2015.07.003.
- [17] LucaDe Angelis, José G.Dias. 2014. Stochastics and Statistics Mining categorical sequences from data using a hybrid clustering method. *European Journal of Operational Research*, Vol. 234, Issue 3, 1 May 2014, 720-730. DOI: 10.1016/j.ejor.2013.11.002.
- [18] Eroglu DY, Kilic K. 2017. A novel Hybrid Genetic Local Search Algorithm for feature selection and weighting with an application in strategic decision making in innovation management. *Information Sciences*, Vol. 405, 18-32. DOI: 10.1016/j.ins.2017.04.009.
- [19] Kim K, Hong J. 2017. A hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis. *Pattern Recognition Letters*, Vol. 98, 39-45. DOI: 10.1016/j.patrec. 2017.08.011.
- [20] Fan CY, Fan PS, Chan TY, SH Chang . 2012. Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals. *Expert Systems with Applications*, Vol. 39, Issue 10, 8844-8851. DOI: 10. 1016/j.eswa.2012.02.005.
- [21] Han, J., Kamber M, Pei J. 2012. *Data mining: Concepts and techniques*. Third Edition. Morgan Kaufmann, Waltham, MA 02451, USA.
- [22] Mishra, D., Hiranwal, S. 2014. Analysis & implementation of item basedcollaboration filtering using K-Medoid. *Paper presented at 2014 International Conference on the Advances in Engineering and Technology Research (ICAETR)*.
- [23] Zaki MJ, Wagner M Jr. 2013. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Troy, Belo Horizonte.
- [24] Palacios-Marqués, D., Soto-Acosta, P., Merigó, J.M. 2015. Analyzing the effects of technological, organizational and competition factors on Web knowledge exchange in SMEs. *J. Telematics and Informatics* 32(1): 23–32. DOI: 10.1016/j.tele.2014.08.003.0736-5853/@2014.
- [25] Hadighi SA, Sahebjamnia N, Mahdavi I, Shirazi MA. 2013. A framework for strategy formulation based on clustering approach: A case study in a corporate organization. *Knowledge-Base Systems*, 49 (2013), 37-49. DOI: 10.1016/j.knosys.2013.04.008.
- [26] Lee H, Lee S, Byungun Y. 2011. Technology clustering based on evolutionary patterns : The case of information and communications technologies. *J. Technological Forecasting & Social Change*, 78 (2011) 953-967. DOI : 10.1016/j.techfore.2011.02.002.
- [27] McGuirk H, Lenihan H, Hart M. 2015. Measuring the impact of innovation human capital on small firms propensity to innovative. *J. Research Policy*, Vol. 44 (4): 965-976. DOI: 10.1016/j.respol.2014.11.008.