Optimizing the Classification Assistance through Supply Chain Management for Telematics SMEs in Indonesia using Deep Learning Approach

Eneng Tita Tosida[#1], Irfan Wahyudin[#2], Fredi Andria[*3], Fajar Delli Wihartiko[#4], Andi Hoerudin[#5]

[#] *Computer Science Department, Universitas Pakuan, Jl. Pakuan PO Box 452, Bogor, Indonesia*

[*] *Management Department, Universitas Pakuan, Jl. Pakuan PO Box 452, Bogor, Indonesia*

[1]enengtitatosida@unpak.ac.id

[2]irfan.wahyudin@gmail.com

[3]fredi.andria@unpak.ac.id

[4]fajardelli@unpak.ac.id

[5]andihoerudin@ansoja.com

Abstract— This study aims to optimize the classification process of providing assistance to Indonesian Telematics Small and Medium Enterprises (SMEs) using a deep learning approach. The data used is the 2016 Economic Census data. The research was conducted comprehensively through the process of comparing performance through several approaches. Deep learning performance shows an optimal accuracy rate of 99.03%, higher than other approaches of the Adaboost and Adaboos-Bagging Ensemble (92.0%), LVQ (93.11%) and Backpropagation (89.1%). The deep learning approach still has shortcomings in terms of tracing attributes that affect the provision of assistance. Unlike the case with an ensemble approach that is able to display priority attributes, and these results are also validated with relevant research results. Research development opportunities can be done through the integration of EXPLAIN and IME models in the deep learning model, making it easier for stakeholders to prioritize attributes that affect the delivery of telematics SMEs. This is expected to encourage the improvement of SMEs competitiveness in facing the challenges of the Industrial Revolution 4.0.

Keywords— *Deep-learning, Ensemble, Classification, Industrial Revolution 4.0, Telematics SMEs.*

1. Introduction

Currently, telematics technology continues to develop and continue to provide benefits for sustainable human life. Telematics is also a development priority in Indonesia which also played a role in the Industrial Revolution 4.0 era [1] [2] [3] [4]. In this era, organizations are required to be able to adapt [5] [6] [7] so as to create a climate with high competitiveness [3] [4]. Indonesia has great potential in Micro, Small and Medium Enterprises (MSMEs / SMEs) [8]. This potential is in line with the field of telematics where the number of SME telematics continues to increase along with the needs of the Indonesian people. This has prompted the Indonesian Government to improve the quality of Telematics SMEs in the face of the free trade era in Asia 2015.

This topic is important to study because the strength of the Indonesian economy lies with SMEs. Also one of the progressive SMEs is the telematics sector. This study is a continuation of previous research which has produced a classification model of providing assistance to SMEs in telematics [9]. The assistance of this will rise human capital as the most important thing to develop the Indonesian competitiveness [10].

Research on the description of Indonesian telematics SMEs factors [11] and improvement in business competitiveness has been carried out with various approaches such as clustering [12] and classification (hybrid mining) [13]. The study used data from the 2006 National Economic Census (Susenas) and didnot consider the imbalance of available data. The Susenas process is carried out every 10 years by the Central Statistics Agency (BPS). Susenas data publication is done 1 year later after being validated by the authorized agency. Thus the main drawback of research [11], [12], and [13] lies in the representation of data.

The difference from previous research is in the more updated data set. In this study, the data update problem will be overcome by using the 2016 Economic Census (SE), which was officially released at the end of 2018. The scope of SE 2016 data that has been released by BPS is still limited to 14 provinces. This data set is very important for the representation of conditions and policies taken related to the outcomes generated from this research. But this data has also other limitation, that is related to imbalance data.

The problem of reported imbalance of data severely impedes classification performance, having been overcome by using several algorithms. Classification algorithms in question include sampling methods, const-sensitive learning, and bagging and boosting ([14] [15] [16] [17]). Based on those researchs, the ensemble approach will be used in this study to deal with the imbalance data problem. The use of deep learning for data classification has been carried out with high accuracy results ([18] [19] [20] [21] [22]). This research aims to optimize the classification model of assistance for Indonesian telematics SMEs through a deep learning approach and compared with the results of the implementation of the Adaboost and Bagging algorithm. The results of this study are expected to improve the accuracy of the Indonesian telematics SME classification model. Another benefit of this research can be used as a reference in the decision- making process for providing assistance to Indonesian telematics SMEs.

2. Material and Method

2.1. Material

The making of this classification model is based on material that related to SME's telematics data [23] with a summary of the attributes as shown in Table 1. This summary carried out from SE 2016 data, especially for Indonesian Telematics SMEs. This data covered 14 Provinces.

Table 1. Description of the Indonesian Telematics SMEs data after pre-processing

Num.	Attributes	Range
A.	**General Information**	
1.	Owner's Education	1. not an elementary school/equivalent graduate 2. Elementary School/equivalent graduate 3. Secondary High School/equivalent graduate 4. High School/equivalent graduate/Paket C 5. Vocational High School 6. Diploma 1/Diploma 2/Diploma 3 7. Diploma 4/Bachelor 8. Post-graduate/ Doctoral-degree
2	Legal Entity Status	1. Public Company 2. CV 3. Firm 4. Cooperatives / Pension Funds 5. Foundation 6. Special permission from the authorized agency 7. Representatives of Foreign Institutions 8. Incorporate
3	Year Operation	1. < 2007 2. ≥ 2007
4	Computer user	1.Yes, 2. No
5	Internet user	1.Yes, 2. No
B.	**Financial**	
6	Sale	1.Micro & Small Enterprises 2.Medium Enterprises 3.Big Enterprises
7	Remuneration	1. Salary and Benefit 2. Wages and Overtime 3. Transportation Fee 4. Bonus 5. Tuition
C	**Business Obstacles & Prospects**	
8	Trouble	1. Capital 2. Raw material 3. Marketing
9	Cooperative	1.Yes, 2. No
10	Partnership	1.Yes, 2. No
11	Workshop Acceptance	1.Managerial 2. Skill/Production Techniques 3. Marketing 4. Others
12	Marketing	1.Export 2. District 3. Outside of the Province 4. Province
13	Company condition 1 year ago	1. Grow up 2. Constant 3. Decrease 4. Cannot be compared
14	Business prospects for the next 3	1. Better 2. Equally good 3. The same is bad

	years	4. Worse
15	The main way to manage a business	1. Franchise 2. Conventional 3. M-L-M
16	Average working day	1. ≤ 15 days 2. ≥ 15 days
D.	**Target Class**	**1. Receive Help** **2. Not Receiving Help**

2.2. Method

The method applied in this classification uses data mining stages or also called Knowledge Discovery and Data Mining (KDD [24]), as a series of processes, data mining can be divided into several stages which are presented in Figure 1.

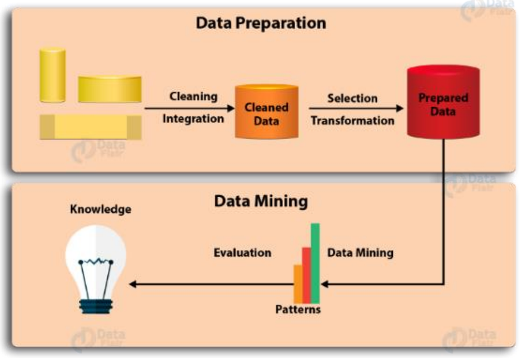

Figure 1. Research Stages

2.2.1. Data Cleaning

Data cleaning is the process of removing noise and inconsistent or irrelevant data. The results of the cleaning of the telematics service business data set obtained from Susenas in 2016 through filling in blank data with mean values (for numerical data), mode (for categorical data), erasing inconsistent data and outliers in accordance with the provisions [25]. The amount of initial data is 42619 data after going through the data cleaning stage to 27550.

2.2.2. Selection and Transforming Data

The data in the database are often not all used, therefore only the data that is suitable for analysis will be retrieved from the database. Meanwhile, data transformation is the process by which data is changed or merged into a format suitable for processing in data mining. The attribute selection process is carried out using AHP [11] by selecting the 16 most influential attributes. Data transformation is done by changing the data size of 16 x 1 to 4 x 4 size.

2.2.3. Mining Process

The mining process is the main stage in this research, which is carried out through a deep learning approach using the Convolutional Neural Network (CNN) algorithm [26], using the python programming language and several supporting libraries for machine learning namely sci-kit learn, hard, numpy, matplotlib and TensorFlow . The formation of a predictive model/classification architecture of assistance for Indonesian telematics SMEs is carried out through the architecture shown in Figure 2.

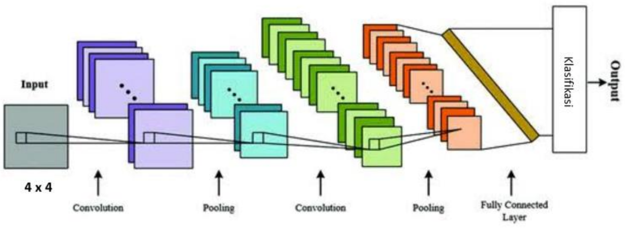

Figure 2. CNN Architecture

The strategy for this problem can explain based on Figure 2, by next source code :

```
import keras
from keras.models import Sequential
from keras.layers.core import Dense, Dropout,
Activation, Flatten
from keras.layers.convolutional import
Convolution2D, MaxPooling2D
from sklearn.model_selection import
train_test_split
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
data = pd.read_excel('dataset/FINALOK.xlsx')
data.head()
plt.plot(data)
plt.show()
X_values = data.iloc[:,0:16].values
X_values[:5]
X_values.shape
y_values = data['Decision'].values
y_values
X_values = X_values.reshape(-1,4,4,1)
x_train, x_test, y_train, y_test =
train_test_split(X_values,y_values,test_size=0.2,
random_state=4)
#CNN model
model = Sequential()
model.add(Convolution2D(filters = 32,
```

```
                kernel_size = 3,
                data_format='channels_last',
                activation='relu',
                input_shape=(4,4,1),
                padding = 'same'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Flatten())
model.add(Dense(10, activation='relu'))
model.add(Dense(1))
model.compile(loss='mse', optimizer = 'adam',
metrics = ['mae'])
model.summary()
model.fit(x_train,y_train, epochs=100)
model.evaluate(x_test,y_test)
```

2.2.4. Pattern Evaluation

Because the resulting model is a black-box model, this study only shows the results of the model accuracy. The pattern evaluation is proposed through the implementation of the EXPLAIN and IME models [27]. This model is able to display attributes with positive or negative influences that have not been seen, especially in the black-box machine learning model, as well as the CNN model of classification of assistance for these telematics SMEs.

3. Result and Discussion

This research was conducted using the Knowledge Data Discovery (KDD) approach [24]. The stages of the research include pre-processing data (cleaning and data integration, data selection and transformation), data mining, pattern evaluation and knowledge presentation (especially for comparative research)). The results of pre-processing data have reduced the amount of data that was originally 42619 to 27550. The process of data mining is done using the CNN deep learning approach and compared with two other classification scenarios, namely through the adaptive-boosting (AdaBoost) and adaptive-boosting (Adaboost) -bagging algorithms [15] [16]. Distribution of training data and test data is done by the k-fold cross-validation strategy by applying a 5-fold scenario (with the comparison of training data and test data is 70%; 30%. Implementation of the comparison model (using adaptive-boosting (AdaBoost) and adaptive-boosting (AdaBoost) -bagging) is done using Sublime Text, Visual Studio Code, Python, R Studio, and Anaconda Navigator.

The 2016 SE data released by BPS consists of telematics SMEs spread across parts of Indonesia (Sumatra, Banten, Jakarta, West Java, Central Java, Yogyakarta, South Sulawesi, Central Kalimantan, and Gorontalo). The attributes of the process of providing assistance to SME telematics consist of 18 attributes, with each class having very diverse and unbalanced attributes. This data imbalance condition can be overcome by AdaBoost and AdaBoost-bagging algorithms so that it produces a good level of accuracy (92%) [15] [16] and this system can also be used to identify attributes that have a high influence on the classification process.

The level of accuracy achieved in this study increased significantly when compared to the previous algorithm [25]. The results of the classification of assistance for Indonesian telematics SMEs using three single-class scenarios show a lower level of accuracy. The results of the study [26] carried out the process of optimizing the classification model using Learning Vector Quantization (LVQ), showing higher accuracy results (93.11%) when compared to the ensemble approach (Adaboost and Adaboost Bagging). However, this study still uses the 2006 Susenas data and still has shortcomings in identifying the attributes that most influence the process of providing assistance to Indonesian telematics SMEs. Two comparative scenarios applied in this study (AdaBoost and AdaBoost-bagging) do not show differences in terms of accuracy. Adaboost classification uses the weak learner principle, which is to collect a low accuracy classifier which is then aggregated into a strong classifier. Bagging technique is done to improve the accuracy of the dataset by creating a bootstrap and combining several models with the final result in the form of max-pooling [16].

However, the amount of data successfully classified using the Adaboost and Adaboost-Bagging methods show differences, as can be seen in Figure 3. The amount of data from the SME classification data, telematics using the adaptive boosting method, produces positive classes totaling 15184 data, negative classes totaling 11184. While the amount of data from the classification using adaptive boosting and bagging results in the number of positive classes 10284 and negative classes 16289. The results of the composition of the different target classes for the two ensemble approaches have no significant effect because overall performance shows the same level of accuracy.

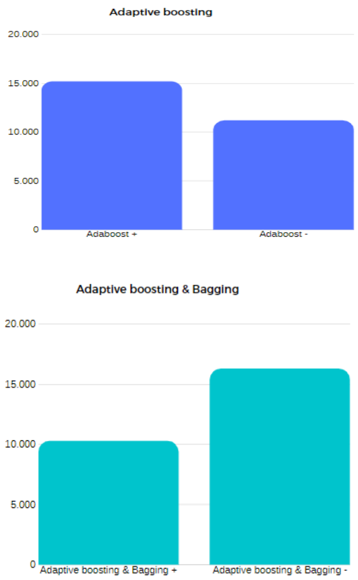

Figure 3. Composition of the results of the classification of the two ensemble techniques as a performance comparison

The visualization of the results of the classification using the ensemble approach can be shown in Figure 4. The system can visualize well. The number of positive class targets from the Susenas data (Receiving assistance) is visualized with +1 sign as many as 11607 records and the number of negative classes Not Receiving Aid -1 is 14761 records.

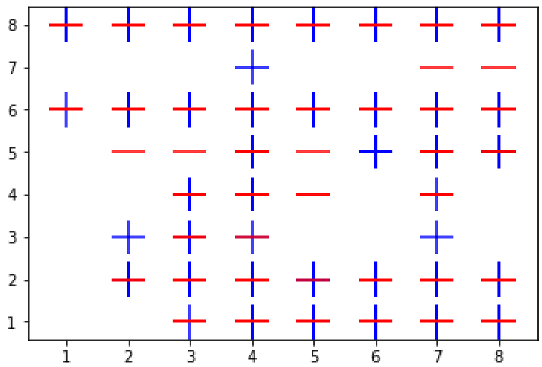

Figure 4. Visualization of the results of the ensemble approach

Unlike the case with the results of the deep learning approach using CNN, it appeared to show a very significant increase inaccuracy. The accuracy comparison is shown in Table 2.

Table 2. Comparison of the accuracy of the classification of Indonesian telematics SME assistance

	Accuracy (%)*				
Techniques	A	B	C	D	E
Data on SME 2006 telematics	93.11	89.1	64.0	87.0	-
2016 Telematics SME Data	-	-	92.0	92.0	99.03

*Information :
A = ANN / LVQ
B = ANN / Backpropagation
C = Ensemble - Adaboost
D = Ensembe - Adaboost Bagging
E = Deep learning – CNN

Accuracy performance parameters for CNN implementation are carried out using Mean Absolute Error (MAE) 0.0097, and epoch is set at epoch 100 so that the achievement level reaches an optimal accuracy. The weakness of the deep learning approach is the difficulty of tracking attributes that have a real influence on the classification process.

In the classification process using ensemble, it can be seen the influential attributes through alpha values. The results of the search for influential attributes show two attributes that correspond to the results of the study [10]. The most influential attribute on assistance is based on the AHP approach through the involvement of three competent experts. The two attributes in question are the year of operation and use of the internet which ranks first and second, while the difference in priority attributes of assistance lies in operating profit. This shows the seriousness of the government in the selection of assistance based on increasing SME profits. This effort is also expected to encourage telematics SMEs to be more serious in managing their businesses so that they have more competitiveness after being given assistance.
The efforts to improve machine learning performance based on artificial neural networks have been proposed [27]. This model is able to display attributes that have a positive or negative

effect, using the EXPLAIN and IME models. The EXPLAIN method has limitations in capturing the results of a dependency prediction. The limitations of the EXPLAIN method can be overcome by the IME method. The shortcomings of this model still need to be improved for the case of large data sets. Additional analysis of the decision-making process needs to be done which will further enhance visualization and presentation and identify relevant attributes. In B2B sales forecasting, further work is needed to identify slippage with regression.

4. Conclusion

The classification process of providing assistance to SMEs using 2016 Susenas data has shown optimal results. Optimization is done by using the CNN deep learning approach which shows an accuracy level of 99.03. This performance far exceeds the classification optimization using the Adaboost and Adaboost-Bagging ensemble which only reaches an accuracy level of 92.0%. However, this deep learning approach still has shortcomings because it is still difficult to trace the attributes that influence the classification process. Unlike the case with the ensemble approach which is able to be traced using alpha values and validated with the results of research using AHP.

The existence of two attributes (internet usage and year of operation) of three that are consistent and become priority attributes that influence the classification process, shows that the performance of the chosen ensemble approach is good. Research development opportunities using the EXPLAIN and IME models that are integrated into the classification process through a deep learning approach are very potential to be done. Optimum performance and ease of tracking attributes that affect the classification process will be a very significant invention for the future.

Acknowledgments

1. DRPM RistekDikti, as the main sponsor, which gives us Competitive Grants Scheme
2. Computer Science Department, Mathematics and Natural Science Faculty, PakuanUniversity,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 became the companion research team in this research grant scheme
4. Indonesian Communication & Information Ministry, Indonesian Cooperation and SMEs Ministry and Bandung Technopark for active participation in the activities of interviews anduser requirement.

References

[1] Jian Qin, Ying Liu, Roger Grosvenor. A Categorical Framework of Manufacturing for Industry 4.0 and Beyond. Procedia CIRP 52 (2016) 173-178. Elsevier. 2016. DOI: 10.1016/j.procir.2016.08.005

[2] Sommer L. 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. – Online ISSN: 2013-0953 – Print ISSN: 2013-8423. 2015. http://dx.doi.org/10.3926/jiem. 1470.

[3] Jaswadi, Mohammad Iqbal, Sumiadji. SME Governance in Indonesia – A Survey and Insight from Private. Procedia Economics and Finance 31 (2015) 387 – 398. Elsevier. 2015.

[4] Jones C, Pimdee P. Innovative Ideas: Thailand 4.0 and the fourth industrial revolution. Asian International Journal of Social Sciences, 17(1), 4 – 35. 2017. https://doi.org/10.29139/aijss.20170101

[5] Pereira T, L Barreto, A Amaral. Network and Information Security Challenges within Industry 4.0. Paradigm. Procedia Manufacturing 13 (2017) 1253-1260. Manufacturing Engineering Society International Conference 2017, MESIC 2017, 28-30 June 2017, Vigo (Pontevedra), Spain. Elsevier. 2017.

[6] Daniel Palacios-Marqués, Pedro Soto-Acosta, José M. Merigó. Analyzing the effects of technological, organizational and competition factors on Web knowledge exchange in SMEs. J. Telematics and Informatics 32 (1): 23–32. 2015. Elsevier, http://dx.doi.org/10.1016/j.tele.2014.08.003.0736-5853/@2014.

[7] Andra Benesova, Jiri Tupa. Requirements for Education and Qualification of People in Industry4.0. Procedia Manufacturing 11 (2017) 2; 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy. 2017. DOI: 10.1016/j.promfg.2017.07.366

[8] Tosida ET, Maryana S, Thaheer H, Damin FA. Visualization model of Small Medium Enterprises (SMEs) telematics services potentiality map in Indonesia. *Published* in: International Conference on Information & Communication Technology and Systems (ICTS), 2015. Added to IEEExplorer. January 14th, 2016. DOI: 10.1109/ICTS.2015.7379890.

[9] Tosida ET, O Hairlangga, F Amirudin, M Ridwanah. Application of Decision Rules for Empowering of Indonesian Telematics Services

SMEs. IOP Conference Series: Materials Science and Engineering 332 (1), 012018. IOP Publishing. 2018. http://iopscience.iop.org/article/10.1088/1757-899X/332/1/ 012018/pdf.

[10] 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. 2015. doi:http://dx.doi.org/10.1016/j.respol.2014.11.008.

[11] Tosida, ET, KB Seminar, Y Herdiyeni. Atribut selection of Indonesian Telematic Services MSMEs Assistance Feasibility, using AHP. Kursor 8 (2). 2016. DOI: http://dx.doi.org/10.21107/kursor.v8i2.1299.

[12] Tosida ET, S Maryana, H Thaheer. Implementation of Self Organizing Map (SOM) as decision support: Indonesian telematics services MSMEs empowerment. IOP Conference Series: Materials Science and Engineering 166 (1), 012017. IOP Publishing. 2017.http://iopscience.iop.org/article/10.1088/1757-899X/166/1/012017/meta.

[13] Tosida ET, F Andria, I Wahyudin, R Widianto, M Ganda, RR Lathif. A Hybrid Data Mining Model for Indonesian Telematics SMEs Empowerment. IOP Conf. Ser. : Sci. Eng. 567. 012001. 2019.

[14] Breiman L. Heuristics of instability and stabilization in model selection, Annals of Statistics, 24, Vol. 24, No. 6, 2350-2383. 1996.

[15] Herrera, Francisco. Class Imbalance in Boosting Bagging Problems and Hybrid Based Approaches. 2012. DOI : 10.1007/s00500-018-3629-4

[16] X.F. Lei, Wang, Z, Y.T. Li. The Risk Assessment Model of Small and Medium-Sized Enterprises of Science and Technology-Based On Boosting. 2015. https://doi.org/10.2991/cisia-15.2015.128.

[17] Fatimah Moch Abdul Mukid, Agus Rusgiyono. Credit Scoring Analysis using Bagging and K-Nearest Neigbhor (*Analisis Credit Scoring menggunakan Metode Bagging dan k-nearest Neighbor*). ISSN: 2339-2541 JURNAL GAUSSIAN Volume 6, NOMOR 1 hal 161-170. 2017.

[18] Rakhlin, A. Diabetic retinopathy detection through the integration of deep learning classification framework. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. 2018. doi:http://dx.doi.org/10.1101/225508

[19] Murata, M., Ariji, Y., Ohashi, Y., Kawai, T., Fukuda, M., Funakoshi, T., Ariji, E. Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography. Oral Radiology, 1-7. 2018. DOI: http://dx.doi.org/10.1007/s11282-018-0363-7

[20] Azimi, S. M., Britz, D., Engstler, M., Fritz, M., & Mücklich, F. Advanced steel microstructural classification by deep learning methods. *Scientific Reports (Nature Publisher Group)*, *8*, 1-14. 2018. doi:http://dx.doi.org/10.1038/s41598-018-20037-5

[21] Donati, L., Iotti, E., Mordonini, G., & Prati, A. Fashion product classification through deep learning and computer vision. Applied Sciences, 9 (7). 2019. Doi : http://dx.doi.org/10.3390/app9071385

[22] Zeng, Z., Mao, C., Vo, A., Janna, O. N., Khan, S. A., Clare, S. E., & Luo, Y. Deep learning for cancer type classification. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. 2019. DOI : http://dx.doi.org/10.1101/612762

[23] Tosida ET, Hermawan, F Andria, I Wahyudin, T Djatna. Development Strategy for Telematics Small and Medium Industries in Indonesia. Jurnal Penelitian Pos dan Informatika (JPPI), Vol. 9 No. 1. 2019. DOI http://dx.doi.org/10.17933/ jppi.2019.090104

[24] Han J, Kamber M, & Pei J. Data Mining Concepts and Techniques. San Fransisco: Morgan Kauffman. 2012.

[25] Tosida ET, O Hairlangga, F Amirudin, M Ridwanah. Application of Decision Rules for Empowering of Indonesian Telematics Services SMEs. IOP Conference Series: Materials Science and Engineering 332 (1), 012018. IOP Publishing. 2018. http://iopscience.iop.org/article/10.1088/1757-899X/332/1/ 012018/pdf.

[26] Shresta A, A Mahmood. Review of Deep Learning Algorithms and Architectures. IEEEAccess Multidisciplinary, Open Access Journal, Vol 7, pp 53040-53065. 2019. DOI : 10.1109/ACCESS.2019.2912200.

[27] Bohanec Marko, Mirjana Kljajic Bor´stnar, Marko Robnik-ˇ Sikonja. Explaining machine learning models in sales predictions. Expert Systems With Applications, pp 3-37. 2016. DOI 10.1016/j.eswa.2016.11.010