Available online at https://journal.rescollacomm.com/index.php/ijqrm/index

International Journal of Quantitative Research and Modeling

e-ISSN 2721-477X
p-ISSN 2722-5046

Vol. 2, No. 3, pp. 147-162, 2021

Potential classification of Smart Village – Smart Economy with Deep Learning methods

Runanto[1*], Muhamad Fahmi Mislahudin[1], Fauzan Azmi Alfiansyah[1], Maudy Khairunnisa Maisun Taqiyyah[2], Eneng Tita Tosida[1*]

[1]*Department of Computer Science, Faculty of Mathemathic and Natural Science, Universitas Pakuan, Bogor City, Indonesia*
[4]*Department of Accounting, Faculty of Economics and Business, Universitas Pakuan, Bogor City, Indonesia*

**Corresponding author email: runanto.065117132@unpak.ac.id, enengtitatosida@unpak.ac.id*

Abstract

Development gap in the city and village is still happening on Indonesia. It happened because of the massive urbanization factors. Poverty in the Indonesian villages are relatively higher than on the urbans. In order to reach the maximal city development, Ministry of Village, Development of Disadvantaged Regions and Transmigration of Indonesia created a sustainable village development program namely Village's Sustainable Development Goals (SDGs) and optimized the village potential data. This study aimed to design the smart village – smart economy classification system by using deep learning methods on village potential data on Indonesia at 2020. The method used in this study is data mining processes namely KDD (Knowledge Discovery and Data mining). The result in this study showed the best models were obtained which consisting of 2 hidden layers and each layer is 128, 128 layers which using target class from the process of calculating the score is able to reach 94.93% of the accuracy from the training process and 96% on the testing process and succeeded to classify the potentials of smart village – smart economy.

Keywords: classification, village potential, smart village, smart economy, deep learning

1. Introduction

Indonesia is an archipelagic country and is strengthened by rural areas which the percentages of the agricultural areas used are more than on the urban areas. On 2018, Statistics Indonesia (Badan Pusat Statistik: BPS) reported that Indonesia had 74.517 villages. Development gaps on the urbans and villages are still happening on Indonesia. It happened because of the massive urbanization factors. Poverty in the Indonesian villages are relatively higher than on the urbans (13.3% on the villages and 7.26% on the urbans). One of the causative factors is the lack of the resources, either human resource or infrastructure. Inequality on the Indonesian villages is addressed through legislation by passing Law No. 6 of 2014 on villages. This law prioritized the development starting from the suburban villages.

In order to reach the maximal city development, Ministry of Village, Development of Disadvantaged Regions and Transmigration of Indonesia created a sustainable village development program namely Village's Sustainable Development Goals, hence forth will be abbreviated into SDGs. SDGs is a role of sustainable development which will be included in the 2021 village funding programs. The SDGs is in accordance with the Presidential Regulations no. 59 of 2017 of national sustainable development goals. It is stated on the Presidential Regulations there is 17 national sustainable development goals. While the SDGs added one more goa. It means SDGs has 18 rural sustainable development goals, which are; concerted effort for creating the village without poverty and famine, equal economic growing village, health-caring village, environmental care village, educational care village, women-friendly village, network village, and culturally responsive village for reaching the sustainable development (Ministry of Village, Development of Disadvantaged Regions, and Transmigration, 2020). One of the strengths needed to reach the SDGs is the beneficial use of information and communications technology (ICT).

The strong globalization flow was greatly driving the ICT which later bringing the changes of the governance on various sectors, starting from business into the governments. This communication technology is seen to be able to play role as an effective and efficient problem-solving solution. for instance, on the governmental sector, it is expected to develop its area by adopting the technology. One of the development efforts by adopting technology is the development of smart village. It is an adopted or derivative concept of smart city. The difference lies only in the location where it applies. If the smart city is implemented to city level, then smart village is implemented to the village level (Subekti and Damayanti, 2019). Smart village has some dimensions which are; smart people, smart environment, smart mobility, smart governance, smart living, and smart economy. Those smart dimensions are adopted from smart city which adapted to the conditional problematic village areas (Santoso et al., 2019).

One of the affecting dimensions in the development of smart village is smart economy. It is focused on the spirit of innovation, entrepreneur, productivity, transforming ability, image and trademark, and flexible labor market, where smart city focused on implementing economic strategy based on digital (Santoso et al., 2019). The indicator of the creation of smart economy of SDGs is that there is the creation of sustainable economy development and proper job (Maja et al., 2020).

According to some studies about smart village in the past (Tosida et al., 2020a), their studies aimed to analyze the potential application of smart village based on big data analytic on Indonesia. The result showed contributed to the analytical processes starting with SLR based on text mining using NVivo 12 plus and showed that the gaps in the analysis of smart village based on big data analytics is potentially to be conducted. The SLR result through cluster analysis also showed the dendrogram of important factors that can be worked on in forming the smart village concept based on big data analytics. The five primary factors are people, project, information, farmer and role model. On Indonesia, it is still needed to improve the strong capacity through the model of city program so that ICT management program is greatly affecting toward the poor people (Agg. $R^2 = 0.5946$) and poor farmer/sailor (adj. $R^2 = 0.606$). The result of doubled regression showed that the ICT management program supported by APBD, PAD, self-help and other funding has the strong impact to the entrepreneur groups and community members.

The next research explained that one of the reasons is the resources gaps, either human or infrastructure. The effort to decrease the farmer poverty level on the village also became the priority program on some countries. One of the concepts that worked to decrease the poverty level is the implementation of smart village concept. The perspective of Citizen Science program is the base development of smart village. Based on big data on Indonesia, it can be creatively examined through clustering methods. The main goal from this study aimed to design the scientific perspective cluster program for citizens to formulating the proper strategies in developing smart village based on Indonesian database. The six clustering models showed the similar models, so that it can map three of citizen science perspective clusters. The results of the clusters show that there are 3 clusters of citizen science potential to develop smart Indonesian villages, namely the very potential (11%), potential (60%) and quite potential (29%) clusters. This citizen science prospects cluster map is visualized on spatial data based on the provinces in Indonesia. Province Bangka Belitung Island, West Java, Central Java, East Java, Yogyakarta, Banten and Bali are potential provinces for developing citizen science in order to construct big data-based smart villages. The results of this cluster were validated with the dendogram structure of the systematic literature review (SLR). The dendogram structure shows the keywords that correspond to the main attributes used in the clustering process. This validation process is an innovative finding in the smart village research ecosystem (Tosida et al., 2020b).

According to the problem and previous conducted research, so far there is no one conducted the research about the classification of potential smart village – smart economy on Indonesia which can equip research (Tosida et al. 2020c). Even though the potential village data are completely provided, so that the it can be used to support the forming of village development strategies. Therefore, this study aimed to design the classification models of smart village – smart economy with the deep learning methods. These models are expected to be able to predict the potentials of smart village – smart economy on Indonesia.

2. Literature Review

2.1. Smart Village

Smart village is defined as the regions and rural communities which are developed above the power and their owned assets, at the same time there is an effort to develop the new opportunity, where the network either traditional or new and the service improved through digital technology, telecommunication, innovation and the usage of better knowledge (European Network for Rural Development, 2018). According to research in the past (Arwildayanto and Utoyo, 2019), conducting the smart village stub (Rintisan Desa Cerdas: RDC) through the empowerment of the Bakti village community, Pulubala sub-district, Gorontalo district Program. The goals of RDC is transforming the ways of thinking, the way of societal working in increasing life quality, economy, skill, expertise, management and village governance supported with innovation and technology transfer. The programs which is held are: financial governance village training, governance village training, mother and children health socialisation, creative economy improvement through environmental usage to family medicinal plants, children’s basic right socialisation, community protection from HIV/AIDS and cervical cancer, reforestation and seed planting program and internet network’s infrastructure improvement program.

2.2. Smart Economy

Smart economy is one of the dimensions that focuses on innovation spirit, entrepreneurship, transformable productivities, image and trademark, and flexible labor market. Smart economy focused on implementing the economic strategy based on digital technology (Santoso et al., 2019). The indicators of the smart economy creation on SDGs are the creation of sustainable economic growth and proper jobs (Maja et al., 2020). According to the research in the (Santoso et al., 2019) which written on book titled "Desa Cerdas: Transformasi kebijakan dan pembangunan Desa Merespon Era Revolusi Industri 4.0" shared the experiences how district Kulonprogo creating the innovation spirit and entrepreneurship on developing the smart economy. Kulonprogo district which is in the special province Yogyakarta (DIY) is still experiencing the poverty issue which they still tried to solve until now. Because of it, the government of Kulonprogo was making a breakthrough by making the bela beli in Kulonprogo (Belabeliku) program in 2013. The program tried to trigger the creation and development of local businesses in the villages. Then, in order to create a strong and big economy, at the beginning of 2016 the Belabeliku program's policies were making digital processes by making the marketplace namely Belabeliku.com. along with the ongoing Belabeliku program cannot be separated from government efforts in creating the Smart Economy in entrepreneurship and innovation aspects. On the innovation aspect, the development of the Bela Beli Kulonprogo program has the innovative spirit which is proven by the continuation of this program when it's firstly initiated. On the entrepreneurship aspect, the Bela Beli Kulonprogo program had become one of the Indonesian first programs where the district government made the marketplace for the local products of its village.

2.3. Deep Learning

Deep learning is the sub-field of artificial intelligence which is focused on designing the big nerve network model that can make decisions based on accurate data. Deep learning greatly fits for the contexts where the data are complex, and the big collections of datasets are provided (Kelleher, 2019). According to the study (Shrestha and Mahmood, 2019) deep learning has important roles in daily life and gives a big impact to various fields such as cancer diagnostics, precession cure, no-driver car, forecasting and speech recognition.

2.3.1. Neural Network Model

The Neural Network (NN) model is a network structure consisting of a few small data processing units arranged into one or several layers. On the architecture of neural network, layer is grouped into 3 kinds, they are; input layer which is the layer that triggers the inputs for the neural network, output layer which is the layer for making the outputs and hidden layer which is one of layer in-between input and output (Fauzan et al., 2018). The computational functions which is performed on every node are:

1. Accumulator functions is to make accumulate the inputs which is multiplied with the weighting of each with the following equation:

$$net = b + w_1x_1 + \cdots + w_nx_n \quad (1)$$

2. The activation function functionalized to process the outputs. The function used in this study is:
   a. Rectified Linear Unit (ReLU) is the efficient functions which use the backward and forward techniques for training processes.

$$f(x) = (0, x) \quad (2)$$

   b. Softmax is the function used for the prediction model which is multiclass. This function normalized the output for every class that valued 0 and 1 and divided with the number of both.

$$S(O_t) = \frac{e^{(O_i)}}{\sum_{j=0}^{k} e^{(O_j)}} \quad k = 0, 1, 2, \dots t \quad (3)$$

The computational process for every node can be modelled as the Figure 1.

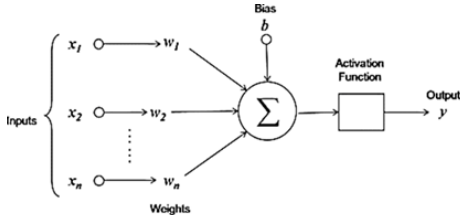


Figure 1. The computational model of a node

Where xi is the input signal to the Wi-I which is the signal weight to– I, b is the input bias, Y is the output node, and $n \geq 1$ is the number of inputs. The architecture of the multi-layer neural network generally can be modelled as the Figure 2.

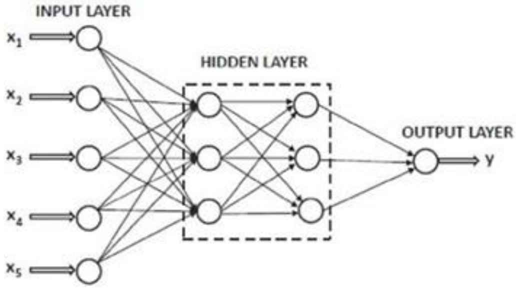


Figure 2. Multi-layer neural network architecture model

As it can be seen from the Figure 2, the input which is originated from the input layers will be received to every node inside the hidden layer. The Output of a node from the hidden layer, asides the final hidden layer, will be continued to all node inside the next hidden layer, whereas the output of the node from the final hidden layer will be received to the output layer.

2.4. Backpropagation algorithm

Backpropagation algorithm is the algorithm learning which is supervised and normally used by perceptron with some layer screens to change the existed weights on its hidden layer. Backpropagation is the easy and simple iterative algorithm which normally works properly. This algorithm is very useful in every application such as introduction, selection, location and final evaluation. The characteristic of Backpropagation includes three layers; input layer where the data are introduced to the network, hidden layer where the data are processed and output layer, where the result of the input layer is inputted (Wanto, 2017).

1. Forward Phase (Forward Propagation)
   In this phase, the input pattern of the propagation is computed forward. The processes starting from input to output layers.
1. Back Phase (Back Propagation)
   In this phase, the propagation of every output unit receives the target pattern related to the input pattern for counting the error value, then the error will be propagated back.
2. Weight modification phase
   This phase is conducted for decreasing the error that is occurred.

2.5. Matrix Confusion

According to the research in the past (Han et al., 2011) on the data mining, in order to measure the result of the model performance, it can be done by using the matrix confusion. It has some terms as the representative of the classification process result as followed:

1. True Positive (TP): the data which are predicted to be appropriate (positive) class and actually is appropriate (true) class;
2. True Negative (TN): the data which are predicted to be inappropriate (negative) class and actually is inappropriate (true) class;
3. False Positive (FP): the data which are predicted to be appropriate (positive) class and actually is inappropriate (false) class;
4. False Negative (FN): the data which are predicted to be inappropriate (negative) class and actually is appropriate (true) class.

Table 1. Matrix Confusion

Actual Class	Predictive class	
	Yes	True
Yes	TN	FN
True	FP	TF

As for some parameters which is usually used in measuring performance of a method are Precision, Recall, and Accuracy

a. Precision is the prediction ratio which is correct from the whole results of positive prediction. Precision explained some of the status percentages of the developing village index and the prediction is appropriately to its class. For the Precision, the formula as followed:

$$precision = \frac{TP}{\mathrm{FP+TP}} \tag{4}$$

b. Recall is the prediction ratio which is correct from the whole appropriate data. Recall explained some of the status percentages of the developing village index and the prediction is appropriately to its class according to the whole data. For the Recall, the formula as followed:

$$recall = \frac{TP}{FN+TP} \tag{5}$$

c. Accuracy is the prediction ratio which is correct from the whole data. Accuracy explained some of the status percentages of the developing village index and the prediction is appropriately to its class. For the Accuracy, the formula as followed:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

d. F1-score is the ratio that compared between Precision and Recall, the formula as followed:

$$f1 - score = \frac{2 \times (Recall \times Precission)}{(Recall \times Precission)} \tag{7}$$

3. Materials and Methods

3.1. Materials

In this study, the potential village data of 2020 on Indonesia which is obtained from the Indonesia Statistics is used. The potential village data consisted of 786 attributes and 5 target classes as the parameter of the classification which is targeted to the developing village index of 2020. According to the indexes, there is 5 status of the village, they are: self-sufficient village, disadvantaged village, and very disadvantaged village (Ministry of Village, Development of Disadvantaged Regions, and Transmigration, 2016). This study used the Google Collaboratory and Microsoft office 2020 for helping the data processing also for using the programming language python

3.2. Methods

The method used in this study is data mining processes namely Knowledge Discovery and Data Mining (KDD). It is the process of taking the pattern of the data which will be processed where later the result will be the output of a very important information. KDD can be divided into some steps as can be seen from the Figure 3

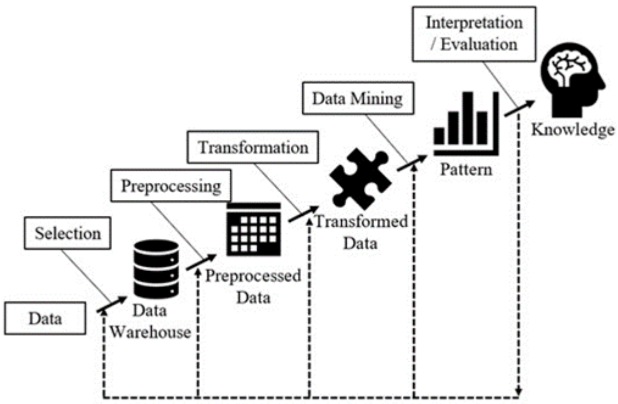


Figure 3. KDD Process (Gullo, 2015)

3.2.1. Selection

Selecting the data from an operational dataset must be done before starting the step of extracting the information on the Knowledge Discovery in Database (KDD). The chosen data used in the data mining process is saved separately from the operational database (Mardi, 2017). The used data resource in this study is the Indonesian potential village data on of 2020 which is taken from Indonesian statistics which consisted of 786 attributes and 5 target classes as the parameter of the classification which is targeted to the developing village index of 2020 of the 5 village status respectively: self-sufficient village, disadvantaged village, and very disadvantaged village (Ministry of Village, Development of Disadvantaged Regions, and Transmigration, 2016).

3.2.2. Preprocessing / Cleaning

In this study, data cleaning is conducted by filling the empty data with mean (for the numerical data) and mode (for the categorical data). This aimed to remove inconsistent data and the outlier based on the determination. The cleaning is also covered the removal of duplicate data, checking the inconsistent data, and fixing the error of the data (Tosida et al., 2020c).

3.2.3. Transformation

Transformation is the process of changing the data into the other shape or reducing into projecting the data in resulting of the presentation which is proper to the task which will be conducted on the data mining process. The transformation generally involves many ways starting from smoothing, aggregation, discretization, attribute construction, generalization, and normalization (Gullo, 2015). The regulation/generalization process is conducted in the data transformation step. The process is used to prevent the model from extracting very specific pattern from the data collections. This aimed to design the model that can predict based on the pattern of the data but not very specifically stuck to the training data, so that the model is able to predict the data which isn't existed on the training data. In this study, the regulation/generalization process is using the dropout method. Dropout is the prevention effort of overfitting and it improves the learning processes (Santoso and Ariyanto, 2018).

3.2.4. Data Mining

Data Mining Process is the process for finding interesting pattern or information on the chosen data by using the specific techniques or methods. This mining process used the deep learning with the neural network model with the backpropagation algorithm as the classification methods.

3.2.5. Interpretation / Evaluation

After the selection, preprocessing, transformation, and data mining processes has been conducted, next, the interpretation or evaluation is conducted. In this process, the result from data mining technique such as either special patterns or predicted model is evaluated to be valued that the existed one is reached. Evaluation from this process used the deep learning methods. The generated information pattern from the data mining should be visualized into the easily understandable one, in order to make some audiences understood, then the presentation of the result from this data mining process is the chart version of the result from the classification phase.

4. Results and Discussion

Based on the stages of data mining, namely Knowledge Discovery and Data Mining, the systematic discussion of this study is arranged through the following stages:

4.1 Selection

The selected data is the attributes / variables of the indicators of smart village – smart economy. The indicators of the creation of smart economy on the SDGs is the creation of sustainable economy development and proper job (Maja et al., 2020). The indicators of smart village – smart economy (Ministry of Village, Development of Disadvantaged Regions, and Transmigration, 2021) respectively are;

1. Citizens has the mobile phone with powerful signal;
2. Internet access on village office;
3. There is an internet access for citizens;
4. There are more than one resident economic activities;
5. Economic institution;
6. Trading dimensions;
7. Credit access dimensions;
8. Distribution access.

In the data selection process, the 2020 potential village data is conducted with the number of 62847 data and attributes consisted of 38 attributes and one of them is the target class attribute as the parameter of the classification which is targeted to the 2020 developing village index. The attributes / variables can be seen from the Table 2 and for the target class can be seen from the Table 3.

Table 2. Variables / Attributes of Potential Village

Variable Name	Stuffing Categories	Type	Description
R1003B	1= most of the citizens 2= small part of citizens 3= none	C	There are citizens who use mobile phones
R1004	1= exist 2= non-exist	C	There is internet café (including online gaming) on the village
R1005C	1= very powerful signal 2= powerful signal 3= weak signal 4= no signal	C	There is mobile phone signal on most of the village areas
R1006B	1=functioning 2=rarely functioning 3=not functioning 4=no functions	C	There are internet facilities on the head village's office
R1007A	1=operating 2=rarely operating 3=not operating 4=no operations	C	There is post office on the village
R1007C	1=operating 2=rarely operating 3=not operating 4=no operations	C	There is private expedition (documents or stuff expedition) office
R1201A		N	Small and micro industry (has approximately 20 employers) based on the raw material: textile industry (bag, shoes, slippers)
R1201B		N	Small and micro industry (has approximately 20 employers) based on the raw material: wooden craft industry (tables, chairs, boards)
R1201C		N	Small and micro industry (has approximately 20 employers) based on the raw material: wooden craft industry (utensils and jewelries)
R1201D		N	Small and micro industry (has approximately 20 employers) based on the raw material: fabric industry (weaving craft and convections)
R1201E		N	Small and micro industry (has approximately 20 employers) based on the raw material: pottery / ceramic industry (roof tile, bricks, ceramics, tiles, and porcelain)
R1201F		N	Small and micro industry (has approximately 20 employers) based on the raw material: bamboo and rattan webbing industry (webbing mat, webbing bag, and webbing wallpaper)
R1201G		N	Small and micro industry (has approximately 20 employers) based on the raw material: F&B industry (procession and preservation of meat, fish, fruits, vegetables, oils and fats, milk and food made from milk)
R1201H		N	Small and micro industry (has approximately 20 employers) based on the raw material: other kind of industry
R1204A		N	The number of operating KUD on the village or district
R1204B		N	The number of KUD which is buying or selling the

			farming products or results on the village or district
R1204C		N	The number of KUD which is providing the business credit
R1204D		N	The number of KUD which is doing other kind of stuff
R1205A1		N	The number of small industry cooperatives and folk crafts (Kopinkra)/actively operating micro business
R1205A2		N	The number of actively operating saving and loan cooperatives
R1205A3		N	The number of actively operating other kind of cooperatives
R1206AK2		N	The number of economic facilities and infrastructures on the village or district: store groups (has minimal 10 stores and grouped into one location)
R1206BK2		N	The number of economic facilities and infrastructures on the village or district: traditional market with permanent building (has roof, floors, and walls)
R1206CK2		N	The number of economic facilities and infrastructures on the village or district: traditional market with semi-permanent building (has roof and floors, without walls)
R1206DK2		N	The number of economic facilities and infrastructures on the village or district: traditional market with no buildings (such as: shock market, dawn market, and floating market.)
R1206EK2		N	The number of economic facilities and infrastructures on the village or district: minimarket or supermarket (retailer market with labeled prices on permanent building, self-services, with floor area < 400 m2)
R1206F1K 2		N	The number of economic facilities and infrastructures on the village or district: Grocery shop or store (retailer store selling various kinds of daily needs on permanent building, no self-services)
R1206F2K 2		N	The number of economic facilities and infrastructures on the village or district: general store (groceries only)
R1206GK2		N	The number of economic facilities and infrastructures on the village or district: restaurant or food stalls (taxed fast-food business)
R1206HK2		N	The number of economic facilities and infrastructures on the village or district: FnB stalls (taxed fast food and beverages business)
R1206IK2		N	The number of economic facilities and infrastructures on the village or district: Hotel (providing accommodation and restaurant, lodging with the hotel licenses)
R1207A	1=exist 2=non-exist	C	Facilities of people's business credit (Kredit Usaha Rakyat: KUR) which is accepted by the residents in the last year
R1207B	3=exist 4= non-exist	C	Facilities of food and energy security business credit (Kredit Ketahanan Pangan dan Energi: KPP - E) which is accepted by the residents in the last year
R1207C	5=exist 6= non-exist	C	Facilities of micro business credit (Kredit Usaha Kecil: KUK) which is accepted by the residents in the last year

R1207D	7=exist 8= non-exist	C	Facilities of joint business credit (Kredit Usaha Bersama: KUBE) which is accepted by the residents in the last year
R1208AK2		N	The number of general governmental bank (such as BRI, BNI, MANDIRI, BPD, BTN.) which is operated in the village areas
R1208BK2		N	The number of general private Bank (such as BCA, Permata, Sinarmas, CIMB, etc.) which is operated in the village areas
R1208CK2		N	The number of rural banks (Bank Perkreditan Rakyat: BPR) which is operated in the village areas

Note: C = Categorical, N = Numeric

Table 3. Targeted Class

Scores	Class
0.815 > X	Self-sufficient village
0.707≤ X ≥ 0.815	Developed village
0.599 ≤ X ≥ 0.707	Developing village
0.419 ≤ X ≥ 0.599	Disadvantaged village
X < 0.491	Very disadvantaged village

On the target class, the used data is consisted of two usages as followed:

1. Target class used the data from the 2020 SDGs survey result which is conducted by ministry of village and region: Disadvantages and transmigration
2. Target class used the result from the attribute calculation processes of the potential villages which is selected according to the smart village – smart economy indicators. The indicators have the scores 0 – 5. The scores based on the FGD analytical Hierarchy process (AHP). The index calculations of every dimension are conducted with the scoring methods which later is transformed into an index (Ministry of Village, Development of Disadvantaged Regions, and Transmigration, 2021). The equation used in this study is as followed:

$$I_x = \frac{\sum_i^n Skor\ X}{n_x\ x\ 5} \tag{8}$$

Ix is index and n is the number of indicators

The smart village-smart economy indicators along with village potential attributes can be seen in Table 4.

Table 4. Smart Village – Smart Economy Indicators and The Attributes of Potential Village Data

Indicators	Attributes of Potential Village Data
Citizens has the mobile phone with powerful signal	R1003B, R1005C
Internet access on the village office	R1006B
There is internet access for the residents	R1004
There are more than one community economic activities	R1201A, R1201B, R1201C, R1201D, R1201E, R1201F, R1201G, R1201H
Economic institution	R1204A, R1204B, R1204C, R1204D, R1205A1, R1205A2, R1205A3, R1206GK2, R1206HK2, R1206IK2, R1208AK2, R1208BK2
Trading dimensions	R1206AK2, R1206BK2, R1206CK2, R1206EK2, R1206F1K2, R1206F2K2, R1206DK2
Credit access dimensions	R1207A, R1207B, R1207C, R1207D, R1208CK2
Distribution access	R1007A, R1007C

4.2 Preprocessing / Cleaning

In this study, data cleaning is conducted by filling the empty data with mean (for the numerical data) and mode (for the categorical data). This aimed to remove inconsistent data and the outlier based on the determination. The cleaning also covered the removal of duplicate data, checking the inconsistent data, and fixing the error of the data (Tosida et al., 2020c). The preprocessing or cleaning generates in a number of data in total 58182 and a number of attributes in total 38 attributes with one of them is the attribute class as parameter of the classification according to the 2020 potential developing village.

4.3 Transformation

The regulation/generalization process is conducted in the data transformation phase. The process is used to prevent the model from extracting very specific patterns from the data collections. This aimed to design the model that can predict based on the pattern of the data but not very specifically stuck to the training data, so that the model is able to predict the data which doesn't exist on the training data. In this study, the regulation/generalization process is using the dropout method. Dropout is the prevention effort of overfitting and it improves the learning processes (Santoso and Ariyanto, 2018).

4.4 Data Mining

Data Mining Process is the process for finding interesting patterns or information on the chosen data by using the specific techniques or methods. This mining process used deep learning with the neural network model with the backpropagation algorithm as the classification methods. In order to start the data mining process, the data should first be divided into training data and testing data. The training data is 40726 (70%) and the testing data is 17455 (30%). In the target class, the one-hot encoding process is conducted in order to visualize data with categorical type and the values of integers are 0 and 1. the categorical values. The next phase is conducting the training process by using the processed dataset. In this phase, the training will be conducted using the parameter and it can be seen from the Table 5.

Table 5. Parameter of Training

Parameter	**Value**
Epoch	100
Optimizer	Adam
Learning Rate	0.001
Loss	Categorical Cross entropy
Validation Split	0.20
Evaluation Spilt	0.30
Dropout	0.25
Batch size	20

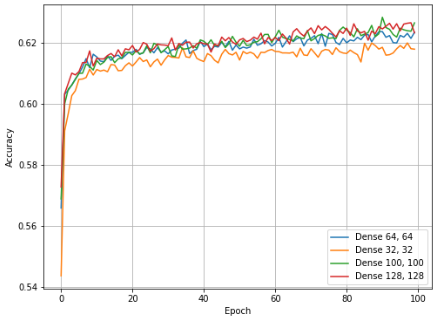
According to Table 5. On the model which is made 100 iteration processes are conducted. Optimization used in this phase is Adam with the learning rate is 0.001 and batch size is 20. Adam is the adaptive optimized version with better performance than the general gradient descent (Zhang, 2018). In order to prevent the overfitting and improve the pace of learning processes on the models, dropout with 0.25 is used. The work system of dropout is removing the neuron which is either hidden or visible layer inside the network (Santoso and Ariyanto, 2018). Whereas in order to measure how good the model performs in classifying, categorical cross entropy as loss function is used. Categorical cross entropy is the loss function used to classify the multiclass.

The implementation of neural network is conducted by using the library Tensorflow to ease the processes of modelling and evaluating the model which is already made. On every training model, the matrix performance of the model is recorded into epoch-accuracy and epoch-loss plot to compare how the model extracted the pattern of the 2020 potential village data. After the next parameter of the training is determined, next the model training process is conducted. The result of the model training process can be seen from the Table 6

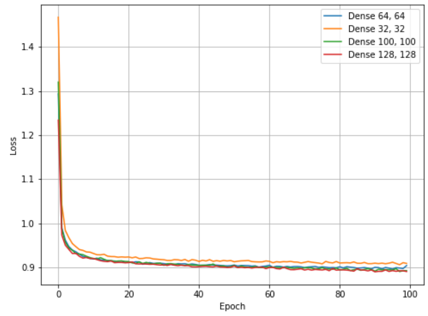
Table 6. Training Model Results

Target Class (Survey of 2020 SGDs)			Target Class (Score Calculation)		
Model	Accuracy	*Loss*	Model	Accuracy	*Loss*
Dense-32,32	61.79%	90.93%	Dense-32,32	89.86%	26.41%
Dense-64,64	62.34%	90.46%	Dense-64,64	93.14%	17.90%
Dense-100, 100	62.65%	89.30%	Dense-100, 100	94.21%	15.58%
Dense-128, 128	62.32%	89.11%	Dense-128, 128	94.93%	14.49%

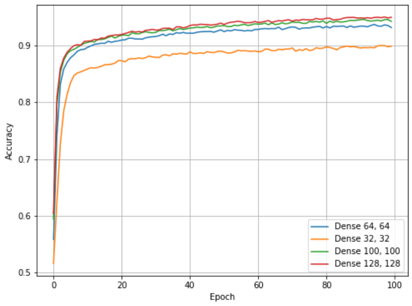
According to Table 6, training result is known that dense model with each hidden layer 128 and 128 units which used the target class from the score calculation processes reached the 94.93% accuracy which is more superior than other models. For next researcher, it can be conducted with the architecture of Convolution Neural Network (CNN) to improve for better accuracy. CNN worked by convolution processing, pooling, and normalization for calculating the weight and bias on to a tensor which later will make the new tensor which consisted of simpler structure than the input tensor (Tosida, Suprehatin, et al., 2020). If the training process on table 6 is visualized with the parameter, it can be plotted like on the Figure 4



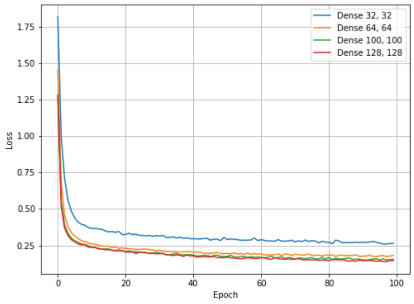
(a) Epoch-Accuracy model with the targe class (2020 survey SDGs)



(b) Epoch-loss model with the target class (2020 survey SDGs)



(c) Epoch-Accuracy model with the target class (score calculation)



(d) Epoch-loss model with the target class (score calculation)

Figure 4. Graphic of Accuracy and Loss Model

According to the Figure 4, it can be concluded that the model had extracted the pattern data properly. however, on the dense model with the target class of 2020 SDGs survey was running into underfitting. it is because an error occurred while extracting the pattern of the data which is marked with" loss" is getting higher and the accuracy is getting low. The dense model with the score calculation class had extracted the pattern data properly, where on the dense-128,128 model reached 94.93% accuracy and is more superior than other models. In order to validate whether the dense-128,128 model can classify properly and no happened or underfitting happened, the analysis process is needed to be conducted by comparing the matrix of accuracy and loss using validation data. The comparison result of the matrix of accuracy and loss using validation data can be seen from the Figure 5

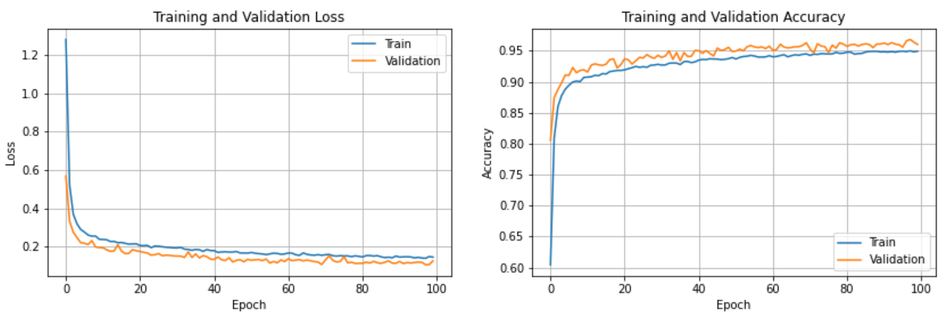


Figure 5. Graphics of Accuracy and Loss Model Training and Validation

According to the Figure 5, every epoch on training process can extract the pattern of the potential village data properly and the validation process which is conducted generated the classifications properly. The Dense-128.128 can be modelled like on the Figure 6

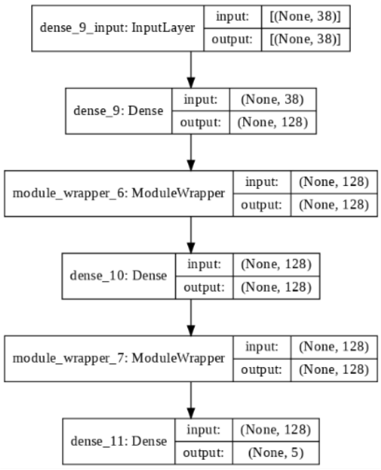


Figure 6. Dense-128, 128 Model with Target Class (Score Calculation)

4.5 Interpretation / Evaluation

In this study, not only the comparison of the Matrix of Loss and accuracy using the validation data had been analyzed, but also the more superior model than other models which are Dense-128.128 with the score calculation class had been analyzed. In order to see the model whether they can classify properly and no overfitting or underfitting happened, evaluation process is conducted to the Dense-128.128 with the score calculation target class. The evaluation process is the process to test the trained model. the model trials are conducted on the testing data as many as 17455 (30%) data using the confusion matrix. The result of the trials can be seen from the Figure 7

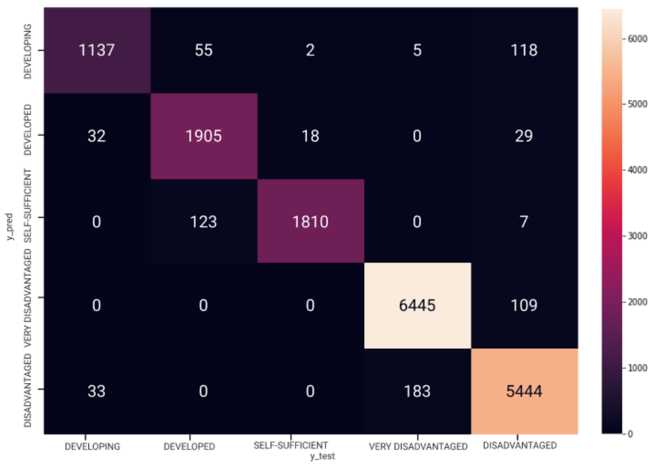


Figure 7. The Result of The Model Trials Using Confusion Matrix

According to Figure 7, the result of the trials of smart village – smart economy classification using the trial data, there are 1) 1810 self-sufficient villages, 2) 1905 developed villages 3) 1377 developing villages, 4) 5444 disadvantaged village and 5) 6445 very disadvantaged villages. In order to see the accuracy, it can be seen from the Figure 8.

	precision	recall	f1-score	support
BERKEMBANG	0.95	0.86	0.90	1317
MAJU	0.91	0.96	0.94	1984
MANDIRI	0.99	0.93	0.96	1940
SANGAT TERTINGGAL	0.97	0.98	0.98	6554
TERTINGGAL	0.95	0.96	0.96	5660
accuracy			0.96	17455
macro avg	0.96	0.94	0.95	17455
weighted avg	0.96	0.96	0.96	17455

Figure 8. Accuracy of Model Trial Results

According to Figure 8, it can be concluded that the classification result of the trials reached 96%. Also smart village is the village which is able to manage their resources self-sufficiently, efficiently, effectively and sustainably to serve the community (Andari and Ella, 2021). The indicators of the smart village creation on SDGs is the creation of sustainable economy development and proper job (Maja et al., 2020). Based on the reference, then the more self-sufficient the village is, the more potential the village will be smart village – smart economy. Then the target classes with some of groups are obtained, they are:

1) The very disadvantaged village as non-potential village;
2) The disadvantaged village as less-potential village;
3) The developing village as the quite potential village;
4) The developed village as the potential village;
5) The self-sufficient village as the very potential village.

The result of the trials of potential smart village – smart economy village can be seen from the Figure 9.

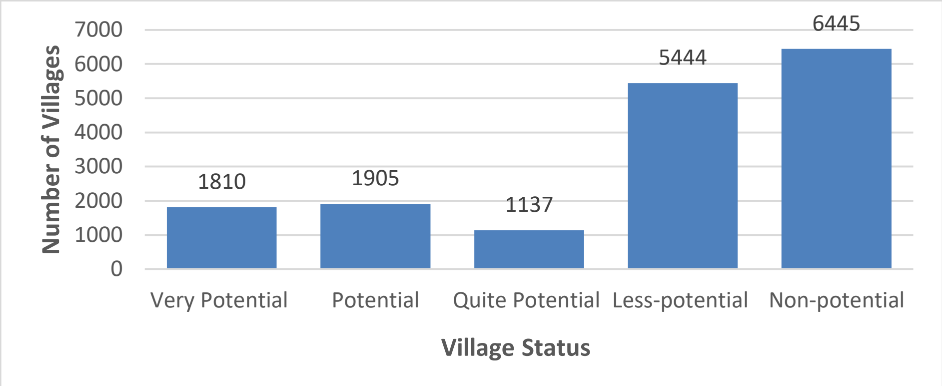


Figure 9. The result of the predicted classification of the potential smart village – smart economy

According to the Figure 9, it can be seen the result of the predicted classification of the potential Smart Village – Smart Economy was using the dense model with each hidden layer 128 and 128 units using the score calculation target class. From 38 variables of potential village and using the trial data from the potential village data can be classified into 5 groups of data according to the predetermined targets.

The result of the trials of the smart village – smart economy potential village classification is also visualized into the comparison table of classification and the table can be seen below as the Table 7.

Table 7. The comparison of classification result

No	Actual target	Target prediction
1	Disadvantaged	Disadvantaged
2	Disadvantaged	Disadvantaged
3	Disadvantaged	Disadvantaged
4	Disadvantaged	Disadvantaged
5	Very disadvantaged	Very disadvantaged
6	Very disadvantaged	Very disadvantaged
7	Very disadvantaged	Very disadvantaged
8	Disadvantaged	Disadvantaged
9	Self-sufficient	Self-sufficient
10	Very disadvantaged	Very disadvantaged

4.5.1 Affecting variables / attributes of the potential of smart village – smart economy

In order to see the affecting variables / attributes of the potential of smart village – smart economy, the mutual information method as known by acronym MI is used. The method has the ability to calculate the consisted information in the terms and its contribution to properly make decision for classification on a class (Irham et al., 2019). In order to see the affecting variables / attributes of the potential of smart village – smart economy, it can be seen from the Figure 10.

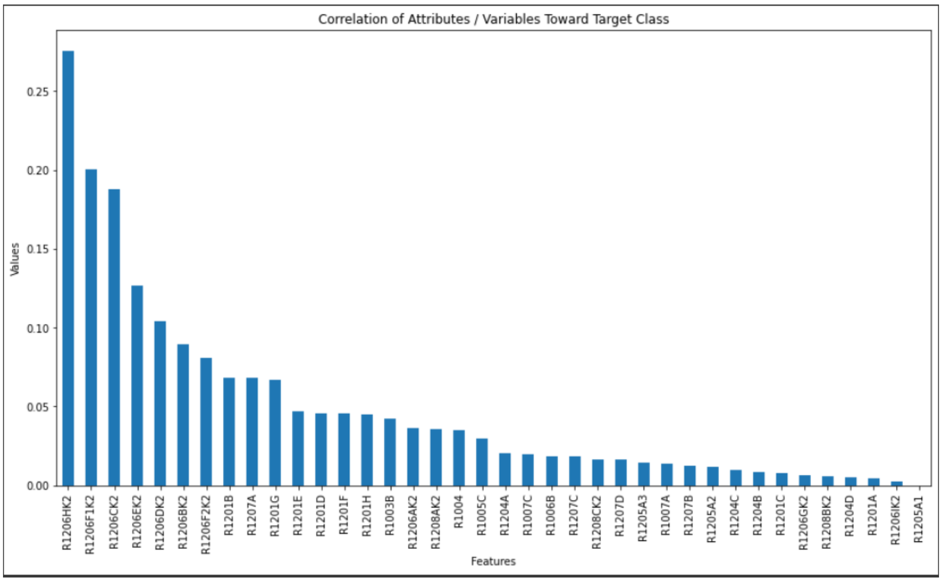


Figure 10: The correlation of attributes / variables toward target class

According to the Figure 9, 10 affecting variables / attributes of the potential of smart village – smart economy which also can be seen from the Table 8.

Table 8. The most affecting variables / attributes of the potential of smart village – smart economy

Variable Name	Description
R1201B	Small and micro industry (has approximately 20 employers) based on the raw material: wooden craft industry (tables, chairs, boards)
R1201G	Small and micro industry (has approximately 20 employers) based on the raw material: F&B industry (procession and preservation of meat, fish, fruits, vegetables, oils and fats, milk and food made from milk)
R1206BK2	The number of economic facilities and infrastructures on the village or district: traditional market with permanent building (has roof, floors, and walls)
R1206CK2	The number of economic facilities and infrastructures on the village or district: traditional market with semi-permanent building (has roof and floors, without walls)
R1206DK2	The number of economic facilities and infrastructures on the village or district: traditional market with no buildings (such as: shock market, dawn market, and floating market.)
R1206EK2	The number of economic facilities and infrastructures on the village or district: minimarket or supermarket (retailer market with labeled prices on permanent building, self-services, with floor area < 400 m2)
R1206F1K2	The number of economic facilities and infrastructures on the village or district: Grocery shop or store (retailer store selling various kinds of daily needs on permanent building, no self-services)
R1206F2K2	The number of economic facilities and infrastructures on the village or district: general store (groceries only)
R1206HK2	The number of economic facilities and infrastructures on the village or district: FnB stalls (taxed fast food and beverages business)
R1207A	Facilities of people's business credit (Kredit Usaha Rakyat: KUR) which is accepted by the residents in the last year

According to the Table 8, it can be concluded that the most affecting smart village – smart economy indicators are: there are more than one economic community activities, trading dimensions, credit access dimension and economic institution. There are lots of less-potential village to be smart village – smart economy, it is caused the score of the indicators for smart village – smart economy is low, those indicators respectively are: there are more than one economic activity, trading dimension, credit access dimension, economic institution on the village.

5. Conclussion

According to the result of potential smart village – smart economy classification with deep learning methods. With the help of potential village data from Indonesian Statistics about 2020 developing village with 38 variables and 5 target classes as the parameter of the classification referring to the 2020 developing village index, the 5-village status respectively are: self-sufficient village, disadvantaged village and very disadvantaged village. The model which can classify the potential of smart village – smart economy with deep learning method has successfully been made. The model itself is Dense with each hidden layer 128 and 128 units which using the score calculation target class. The parameters used in the training process are epoch with the value of 100, optimizer with Adam Algorithm, learning rate with 0.001 value, loss functions which is used for categorical cross entropy, validation split with the value of 0.20, evaluation split with the value of 0.30 and dropout with the value of 0.25.

In this study, the model generated the accuracy as much as 94.93% on the training process and 96% on the testing process also successfully classifying the potential smart village – smart economy. According to the trial result of potential smart village – smart economy classification from the potential village data, the villages can be classified into 5 groups according to the determined target by using the trial data, they are: 1) 6445 villages as non-potential, 2) 5444 villages as the less-potential, 3) 1137 villages as quite potential, 4) 1905 villages as the potential and 5) 1810 villages as the very potential. There are lots of less-potential village to be smart village – smart economy, it is caused the score of the indicators for smart village – smart economy is low, those indicators respectively are: there are more than one economic activity, trading dimension, credit access dimension, economic institution on the village.

Acknowledgments

We would like to acknowledge the Universitas Pakuan for facilitating the PKM-RE and Directorate General of Instruction and Student Affairs Ministry of Education, Culture, Research and Technology Republic of Indonesia for funding the PKM-RE.

References

Andari, R. N., & Ella, S. (2021). *Model Desa Cerdas Untuk Membangun Indonesia Maju*. Aceh: Syiah Kuala University Press.

Arwildayanto, A., & Utoyo, S. (2019). Rintisan Desa Cerdas (RDC) melalui Program Pemberdayaan Masyarakat Desa Bakti Kecamatan Pulubala Kabupaten Gorontalo Provinsi Gorontalo. *PengabdianMu: Jurnal Ilmiah Pengabdian Kepada Masyarakat*, *5*(1), 24-32.

European Network for Rural Development. (2018). *Smart Villages: Revitalising Rural Services*. European Network for Rural Development.

Fauzan, M., Adiwijaya, A., & Mubarok, M. S. (2018). Klasifikasi Multi-Label pada Topik Berita Berbahasa Indonesia menggunakan Artificial Neural Network. *eProceedings of Engineering*, *5*(3), 7764-7773

Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. *Physics Procedia*, *62*, 18-22.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*, Amsterdam: Elsevier.

Irham, L. G., Adiwijaya, A., & Wisesty, U. N. (2019). Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, *3*(4), 284-292.

Kelleher, J. D. (2019). *Deep learning*. London: MIT press.

Maja, P. W., Meyer, J., & Von Solms, S. (2020). Development of Smart Rural Village Indicators in Line With Industry 4.0. *IEEE Access*, *8*, 152017-152033.

Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4. 5. *Edik Informatika*, *2*(2), 213-219.

Ministry of Village, Development of Disadvantaged Regions, and Transmigration. (2020). *Dari TPB ke SDGs desa*. Retreived from https://sdgsdesa.kemendesa.go.id/dari-tpb-ke-sdgs-desa/

Ministry of Village, Development of Disadvantaged Regions, and Transmigration. (2021). *Pemutahiran Data IDM Berbasis SDGs Desa 2021*. Ministry of Village, Development of Disadvantaged Regions, and Transmigration.

Ministry of Village, Development of Disadvantaged Regions, and Transmigration. (2016). *Peraturan Menteri Desa, Pembangunan Daerah Tertinggal dan Transmigrasi nomor 2 tahun 2016 tentang Indeks Desa Membangun*. Ministry of Village, Development of Disadvantaged Regions, and Transmigration.

Santoso, A. D., Fathin, C. A., Effendi, K. C., Novianto, A., Sumiar, H. R., Angendari, D. A. D., & Putri, B. P. (2019). Desa Cerdas: Transformasi Kebijakan dan Pembangunan Desa Merespon Era Revolusi Industri 4.0. *Monograf.(EA Purwanto & D. Permadi, Eds.). Yogyakarta: cfds Fisipol UGM*.

Santoso, A., & Ariyanto, G. (2018). Implementasi deep learning berbasis keras untuk pengenalan wajah. *Emitor: Jurnal Teknik Elektro, 18*(1), 15-21.

Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access*, *7*, 53040-53065.

Subekti, T., & Damayanti, R. (2019). Penerapan model smart village dalam pengembangan desa wisata: Studi pada desa wisata boon pring sanankerto turen kabupaten malang. *Journal of Public Administration and Local Governance*, *3*(1), 18-28.

Tosida, E. T., Herdiyeni, Y., & Suprehatin, S. (2020a). The Potential for Implementing a Big Data Analytic-based Smart Village in Indonesia. In *2020 International Conference on Computer Science and Its Application in Agriculture (ICOSICA)* (pp. 1-10). IEEE.

Tosida, E. T., Suprehatin, S., Herdiyeni, Y., & Solihin, I. P. (2020b). Clustering of Citizen Science Prospect to Construct Big Data-based Smart Village in Indonesia. In *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (pp. 58-63). IEEE.

Tosida, E. T., Wahyudin, I., Andria, F., Djatna, T., Ningsih, W. K., & Lestari, D. D. (2020c). Business Intelligence of Indonesian Telematics Human Resource: Optimization of Customer and Internal Balanced Scorecards. *Journal of Southwest Jiaotong University*, *55*(2).

Village Potential Data 2020 Central Bereau of Statistics (BPS) Indonesia, Jakarta, 2020.

Wanto, A. (2017). Optimasi Prediksi Dengan Algoritma Backpropagation Dan Conjugate Gradient Beale-Powell Restarts. *Jurnal Nasional Teknologi dan Sistem Informasi*, *3*(3), 370-380.

Zhang, Z. (2018, June). Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)* (pp. 1-2). IEEE.