IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 459~468
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp459-468

Deep learning approach analysis model prediction and classification poverty status

Musli Yanto, Yogi Wiyandra, Sarjon Defit
Department of Information Technology, Faculty of Computer Science, Universitas Putra Indonesia YPTK, Padang, Indonesia

Article Info

Article history:

Received Apr 25, 2022
Revised Aug 15, 2022
Accepted Sep 13, 2022

Keywords:

Classification
Data mining
Poverty
Prediction
West Sumatra

ABSTRACT

The problem of poverty is a scourge for every developing country coupled with the economic crisis that occurred during the coronavirus disease (COVID-19) pandemic. The impact of these problems is felt directly by the people in Indonesia, especially in the Province of West Sumatra. This study aims to predict and classify the level of poverty status by developing an analytical model based on the deep learning (DL) approach. The methods used in this study include the K-means method, artificial neural network (ANN), and support vector machine (SVM). The analytical model will be optimized using the pearson correlation (PC) method to measure the accuracy of the analysis. The variable indicator uses the parameters of population (X_1), poverty rate (X_2), income (X_3), and poverty percentage (X_4). The results of the study present prediction and classification output with a validity level of accuracy of 99.8%. Based on these results, it can be concluded that the proposed DL analysis model can present an updated analytical model that is quite effective in carrying out the prediction and classification process. The research findings also contribute to the initial handling of the problem of poverty.

Corresponding Author:

Musli Yanto
Department of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK
Padang, Sumatera Barat, 25145, Indonesia
Email: musli_yanto@upiyptk.ac.id

1. INTRODUCTION

The level of poverty status during the coronavirus disease (COVID-19) pandemic has increased in line with the decline in the global economy [1]. This also has a major impact on several developing countries [2]. Indonesia is also one of the countries that have had the impact of the pandemic, which has caused a high movement in the poverty rate in recent times [3]. The impact is explained that the movement of the poverty rate can be seen by the population density and the level of community economic income [4]. To overcome these problems, several studies have been carried out in an effort to overcome the problem of poverty. One form of application for handling and controlling poverty cases can be seen in the classification process. This process is carried out to be used as recommendations as well as control and monitoring for the government in managing the community's economy.

The classification process has developed by producing a model to be used as a solution to solving a problem. Artificial intelligence (AI) is a concept that is widely used in the classification process. AI is a multidisciplinary prospective study of a branch of science that has many opportunities in research and provides results as an impetus for the development of knowledge [5]. The development of AI also has a major impact on the economy so it has the potential to increase productivity growth [6]. AI can also be applied in financial analysis in the form of an algorithm developed [7].

The application of AI in classified problems covers a variety of problems. Previous research explained that the classification process using machine learning (ML) provides a model and causes the emergence of poverty status based on household income and expenditure [8]. Furthermore, the same study explains that the classification of poverty status using the support vector machine (SVM) provides a classification model with an accuracy of 71.93% [9]. The classification process is also able to overcome the level of poverty by measuring indicator variables that have a significant influence [10]. The ML approach produces a fairly good performance and level of accuracy in classifying poverty status [11]. A data mining approach is used to analyze poverty data based on the community based monitoring system (CBMS) database [12]. The development of the classification process as well as the existing model is also seen in the comparison of the performance of various algorithms for classification [13]. The system is able to process 14 levels of poverty in developing countries [14].

Based on the previous explanation, this study proposes a predictive analysis model and classification of poverty status levels. The model was developed using the deep learning (DL) approach in unsupervised and supervised learning. Unsupervised adopt K-means learning clusters to generate classification patterns. Supervised learning uses the artificial neural network (ANN) method and the SVM which is optimized by the pearson correlation (PC) method in measuring the level of prediction and classification results. The K-means method is able to provide analysis process performance based on patterns from data that have the same group [15]. The same explanation also explains that K-mean cluster is an unsupervised method that works efficiently in grouping big data [16]. To optimize K-means, the sum of squared error (SSE) method can to validate the cluster results [17], [18]. The K-means algorithm has been developed and used as a form of data grouping process on various problems [19]. Clustering is also used with the aim of running data and grouping it to get clustering optimization [20].

Supervised learning is also one of the processes in conducting prediction analysis and classification of poverty status. Basically, supervised learning is used in solving classification problems [21]. This learning concept has a label on the data set used [22]. The learning methods used include ANN and SVM. The ANN method is developing so fast along with previous research in the classification process to produce an optimal solution [23]. The application of ANN also helps a lot and solves a big problem [24]. The same explanation also explains that ANN is a method capable of developing a classification model with a fairly good performance and level of accuracy [25]. The performance of ANN is able to present a very good performance in the prediction process [26], [27]. In addition to ANN, the SVM method is also an alternative method for conducting classification analysis. SVM is a method that is able to solve problems with non-linear data using several kernels [28]. SVM can be recommended to carry out the classification process with a fairly good output [29]. The SVM method is able to show optimal performance based on generalization compared to conventional methods [30]. The application of SVM is also able to classify the poverty level based on the attribute data used with an accuracy of 71.93% [31].

Based on this explanation, this study presents an update in the development of predictive and classification analysis models using the DL approach. The novelty is presented in the form of a measurement process for the output obtained. The analysis model presents a systematic pre-processing process to provide an appropriate analysis pattern based on the clustering of the data set used. The pattern of analysis is continued in the DL learning process using the ANN and SVM methods. Tests on the analysis pattern are carried out for validation between outputs and indicators in the prediction and classification process. The overall results of this research will make a major contribution to the development of knowledge and can be used as a form of control and handling for the government on the problem of poverty.

2. RESEARCH METHOD

Predictive analysis and classification of poverty status levels are described in several stages of the research framework. A quantitative approach is used to perform mathematical calculations on a dataset to obtain the desired results. This research framework presents an overview in the form of a predictive and classification analysis model which was developed with several stages starting from analysis, pre-processing, as well as the prediction and classification process. The analysis stage is carried out to determine indicators that can be used as variables in the analysis process. Indicators based on field survey results are reproduced using the regression method to measure weaknesses and strengths. After the variables are obtained, it is continued at the pre-processing stage using K-means to produce a classification pattern. The results of the K-means cluster process can divide the categories of poverty status to form a prediction and classification pattern. The pattern will then be reused to perform prediction analysis and classification using the ANN and SVM methods optimized by the PC method. This optimization is used as a measurement and validation of the analysis process so that the output has the maximum level of accuracy. The overall results of this analysis are optimal models for predicting and classifying poverty levels. The stages of the research can be seen in the framework of the prediction and classification analysis model presented in Figure 1.

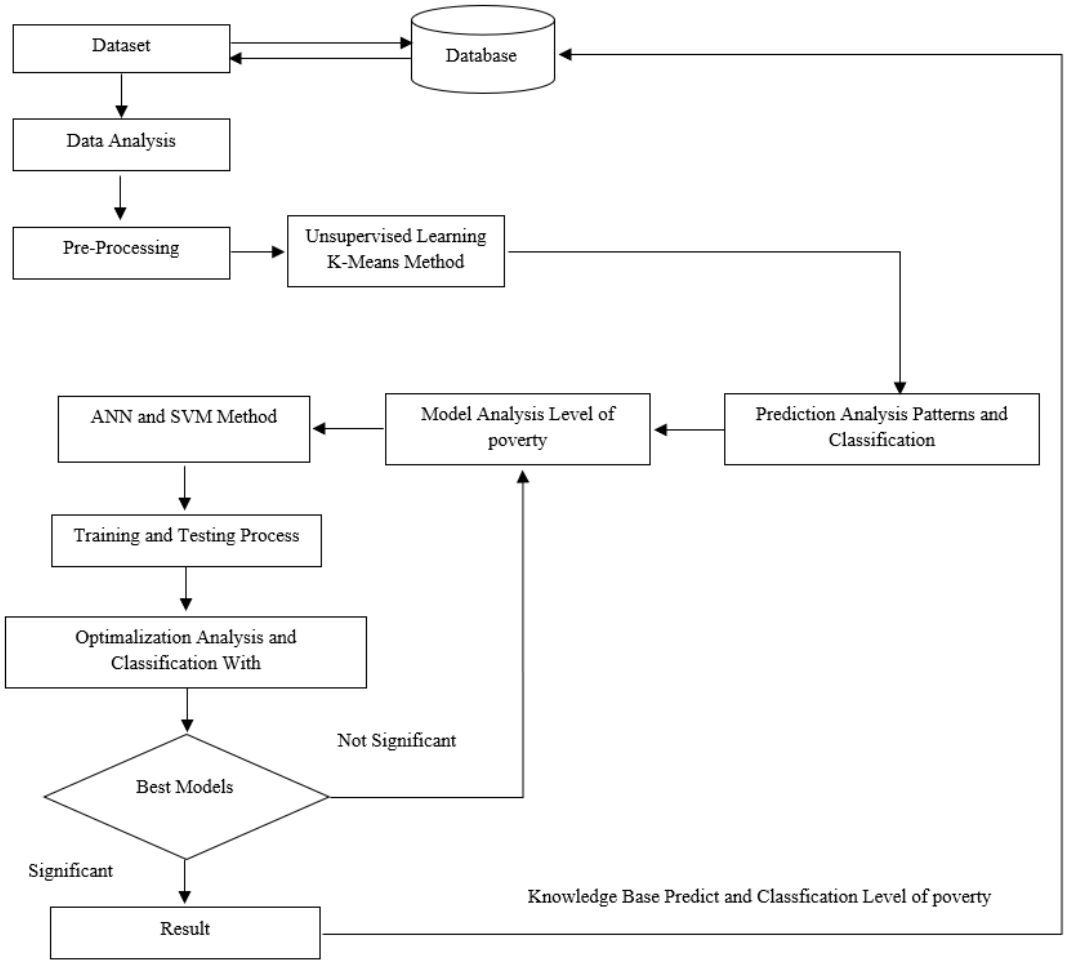

Figure 1. Predictive analysis and classification model framework

2.1. K-means cluster

K-means cluster is a method used in grouping data [32], [33]. Basically, the cluster concept is one of the techniques in which there is unsupervised learning in data mining by grouping data features that have similar features [34]. Among them are i) determining the number of clusters, ii) selecting the initial centroid randomly according to the number of existing clusters, iii) calculating the distance from the data to the centroid using the Euclidean distance formula, iv) updating the centroid by calculating the average value of each cluster, and v) return to step 3 if there is still data that has moved clusters or the centroid value has changed [35]. The calculation of the centroid distance can be expressed as (1) [36].

$$D_e = \sqrt{(M_{ix,} - C_{ix})^2 + (M_{iy} - C_{iy})^2} \quad (1)$$

D_e is the Euclidean distance value to measure the distance from each centroid. $M_{ix,}$ M_{iy} is a value for the coordinates of the object in the data. $C_{ix,}$ C_{iy} is the value of the center coordinate of the centroid.

2.2. Artificial neural network

ANN is one method that can be used in the classification process by presenting a fairly good level of performance [37]. ANN performance gives maximum results in dealing with problems such as classification and prediction [38], [39]. Based on the concept that has been explained that ANN can do learning by adopting a mathematical calculation process [40]. Learning is capable of a model that is applied in the form of an algorithm to produce decisions [41]. The model is presented in an architectural pattern based on the input layer, hidden layer, and output layer [42]. Overall, the ANN concept aims to provide optimal output from the learning process carried out [43].

2.3. Support vector machine

SVMs are a technique found in ML that can carry out the classification process and other learning activities [44]. A SVM is a concept that performs classification based on hyperplane to describe points in space that have been categorized [45]. SVM is a discriminatory classifier and is formally characterized by an optimal hyperplane [46]. This hyperplane is a dividing line between data segments, where each data segment will be placed on both sides. For example, multiple row data classification has been carried out with two different data sets [47].

2.4. Pearson correlation

PC is a statistical concept capable of performing calculations in a measurement process [48]. PC can be combined with several methods to give better results [49]. PC-based techniques can also be used to select optimized features in reviewing output from a model [50]. The PC calculation can be expressed as (2) [51].

$$\mathrm{X,Y} = \frac{con\,(X,Y)}{\sigma X \sigma Y} = \frac{E\,((X-\mu X)(Y-\mu Y))}{\sigma X \sigma Y}$$

$$= \frac{E(X,Y)-E(X)E(Y)}{\sqrt{E(X^2}-E^2(X)\sqrt{E(Y^2}-E^2(Y)} \quad (2)$$

The value of cov(X, Y) is the covariance between X and Y. The value of X, Y is the value of the standard deviation of the variables X and Y. The value of E(X) is the expected value of X.

3. RESULTS AND DISCUSSION

The process of prediction analysis and classification of poverty status using the DL approach begins with data analysis. Dataset analysis is sourced from the Central Statistics Agency of West Sumatra Province in 2020 and 2021. The data analysis process aims to determine the parameters and indicators used in the prediction and classification process. The dataset analysis can be seen in Table 1.

Table 1. Dataset analysis

West Sumatra Poverty Status Indicator 2020				West Sumatra Poverty Status Indicator 2021			
Total population	Poverty rate	Income	Poverty percentage	Total population	Poverty rate	Income	Poverty percentage
90.373	12.990	2.126.15	14.37	92.021	13.220	2.326.01	14.37
460,716	34.920	2.422.10	7.58	463,923	36.510	2.232.85	7.87
371,105	32.890	2.554.31	8.86	373,414	29.740	2.422.38	7.96
233,810	16.550	2.107.61	7.08	237,376	16.650	2.411.37	7.01
347,407	18.480	2.607.63	5.32	348,219	16.200	2.763.59	4.65
413,272	33.200	2.234.52	8.03	415,613	29.480	2.268.98	7.09
487,914	32.920	2.785.18	6.75	491,282	33.100	2.339.39	6.74
379,514	26.470	2.254.91	6.97	382,817	26.640	2.125.74	6.96
278,480	20.310	1.894.51	7.29	281,211	20.220	2.464.50	7.19
168,411	11.850	2.587.11	7.04	171,075	12.490	2.537.75	7.30
241,571	15.420	2.896.95	6.38	247,579	15.490	2.375.03	6.26
435,612	31.830	2.220.30	7.31	443,722	31.530	2.443.75	7.11
939,112	44.040	2.836.98	4.69	950,871	42.440	3.278.68	4.46
69,776	2.290	2.971.92	3.28	71,010	2.290	3.019.50	3.22
61,898	1.480	2.766.77	2.39	62,524	1.350	2.510.62	2.16
52,994	3.110	2.723.80	5.87	53,693	3.000	3.188.46	5.59
128,783	6.320	2.744.42	4.91	130,773	6.000	2.838.49	4.59
133,703	7.690	2.969.86	5.75	135,573	7.680	2.913.36	5.66
87,626	4.400	2.661.84	5.02	88,501	4.200	2.519.09	4.75

Table 1 explains that the indicators of the analysis process use the variables of population (X_1), poverty rate (X_2), (X_3), and poverty percentage (X_4). After the data analysis process is carried out, the analysis process is continued at the analysis pre-processing stage to find prediction patterns and classifications using the unsupervised learning concept. The results of the K-means cluster process using Weka Software can be seen in Figure 2.

Figure 2 explains that the pre-processing analysis using K-means gives a fairly good result in describing the prediction and classification patterns. This result is quite significant based on the visualization of the data clusters formed based on the data groups in Table 1. The cluster results provide results with 3 levels of poverty status, namely cluster 2 (high status), cluster 1 (medium status), and cluster 0 (low status). With

these results, the process of predictive analysis and classification can be carried out to achieve maximum results. The predictive analysis process is carried out by adopting supervised learning. The method used is an ANN. The ANN method works to carry out the analysis process by the concept of human thinking [52]. The network architecture model can be seen in Figure 3.

Clusterer output

```
Initial starting points (random):

Cluster 0: 168411,11850,2587119,7.04
Cluster 1: 53693,3000,3188468,5.59

Missing values globally replaced with mean/mode

Final cluster centroids:
                                        Cluster#
Attribute                 Full Data           0           1
                             (38.0)      (22.0)      (16.0)
=============================================================
Jumlah Penduduk             284823   400779.2727   125383.125
Data Kemiskinan          18562.8947    27255.4545    6610.625
Upah Buruh            2561752.4737 2396953.6364 2788350.875
Persentase Kemiskinan        6.4692        7.6032        4.91



Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      22 ( 58%)
1      16 ( 42%)
```

Figure 2. K-Means cluster process results

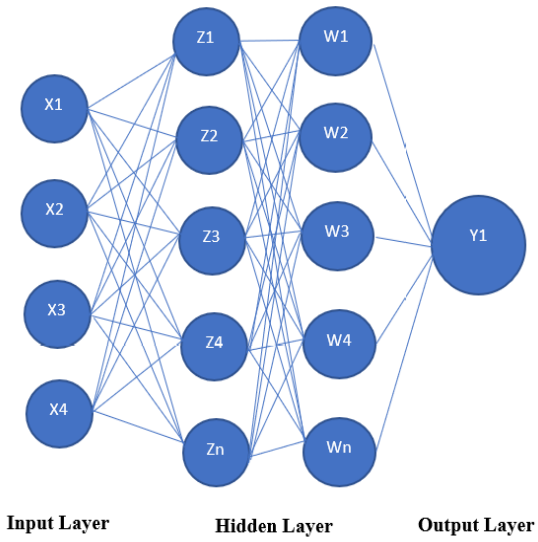

Figure 3. ANN architectural models

Figure 3 is a form of ANN network architectural model design in the prediction process. The input layer consists of 4 layers based on predetermined variable indicators. The hidden layer has a multilayer shape and consists of several layers. The output layer consists of one layer to describe the prediction results obtained. The algorithm used in training and testing the ANN network is using the feedforward algorithm. This algorithm is able to provide an analytical model with a fairly good level of accuracy and mean square error (MSE) value in the prediction process [53]. The results of the training and testing of ANN can be seen in the Table 2.

Table 2. Results of ANN training and testing

Network architecture	Single hidden layer					
	Training			Testing		
	Accuracy	MSE	Performance	Accuracy	MSE	Performance
(4-4-1)	99.981000	0.019000	0.000190	99.925800	0.074200	0.003900
(4-6-1)	99.981000	0.019000	0.001000	99.901200	0.098800	0.005200
(4-8-1)	97.350300	2.649700	0.026497	99.937300	0.062700	0.003300
(4-10-1)	99.981000	0.019000	0.001000	99.862300	0.137700	0.007200
(4-11-1)	99.981000	0.019000	0.001000	99.930500	0.069500	0.003700
Network architecture	Multi hidden layer					
	Training			Testing		
	Accuracy	MSE	Performance	Accuracy	MSE	Performance
(4-4-4-1)	99.969100	0.030900	0.0016	99.969100	0.030900	0.001600
(4-8-4-1)	99.983900	0.016100	0.000845	99.983900	0.016100	0.000845
(4-4-4-4-1)	99.880100	0.119900	0.006300	99.926900	0.073100	0.003800
(4-4-8-4-1)	99.973500	0.026500	0.001400	99.973500	0.026500	0.001400
(4-10-8-4-1)	99.835800	0.164200	0.008600	99.923000	0.077000	0.004100

Table 2 describes the results of training and testing of several network architecture patterns in the prediction process. The network architecture used adopts single layer forms such as 4-4-1 (4 input layers, 4 hidden layers, and 1 output layer) and multilayers such as 4-4-4-1 (4 input layers, 4 hidden layers_1, 4 hidden layers_2, and 1 output layer). The network architecture with the best performance is obtained in the 4-8-4-1 pattern with an accuracy rate of 99.98%. The results of the ANN analysis process in making predictions can be seen in Figure 4.

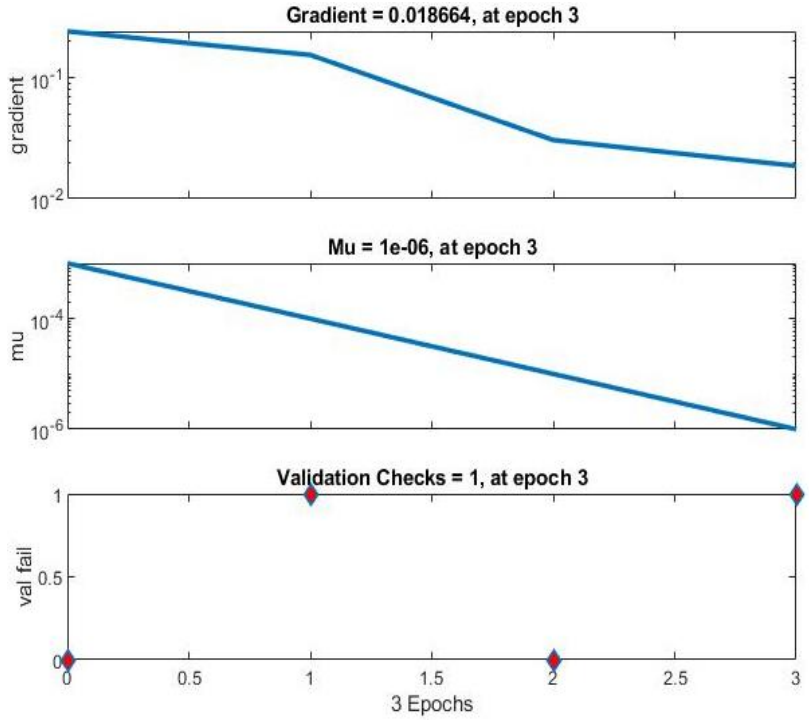

Figure 4. Results of ANN analysis process

Figure 4 explains that the results of the analysis process look quite significant based on the output graph presented. After the results of the prediction process using ANN, the analysis process is continued to carry out the classification process using the SVM method. This method is a method used on linear and nonlinear data. The results of the SVM analysis can be seen in Figure 5.

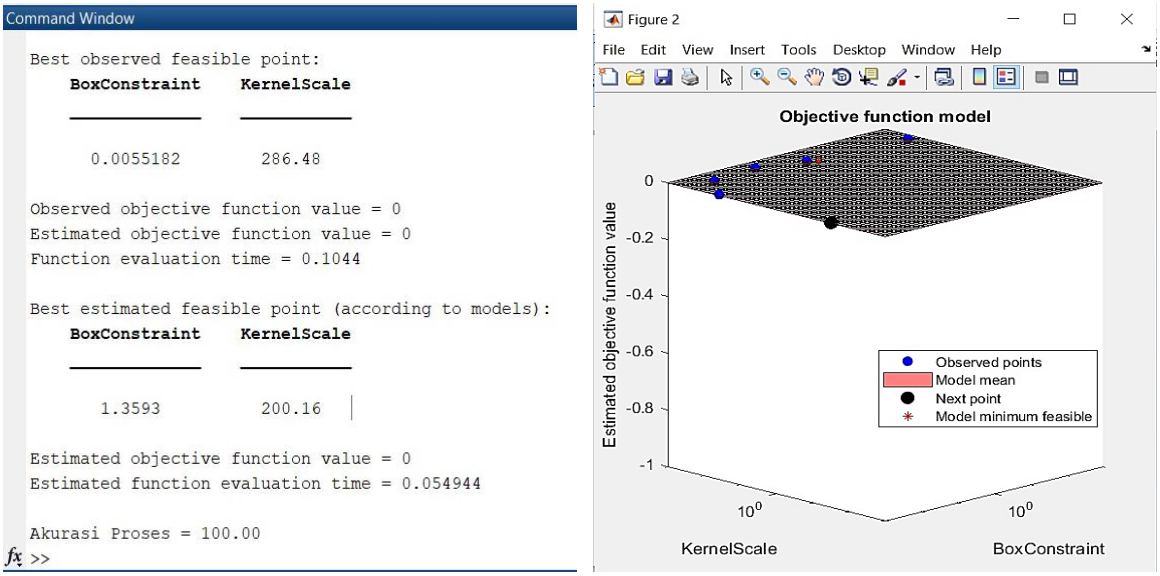

Figure 5. SVM analysis results

Figure 5 explains that the SVM analysis process presents a fairly good classification result in classifying poverty status. After the prediction analysis and classification process is carried out, then the process is continued for the validation process using the PC method. This method is a statistical method used to see the relationship between regression and correlation of each indicator variable to the resulting output. The results of the validation process using a PC can be seen in the Table 3.

Table 3. Results of regression and correlation test process

Model summary									
Model	R	R Square	Adjusted R Square	Std. error of the estimate	Change statistics				
					R Square	F Change	df1	df2	Sig. F
1	.994[a]	.988	.984	165.464.736	.988	283.078	4	14	.000

a. Predictors: (constant), X_4, X_1, X_3, X_2

Correlations						
Control variables			X_1	X_2	X_3	X_4	Y
-none-[a]	X_1	Correlation	1.000	.917	.013	.002	.904
	X_2	Correlation	.917	1.000	-.305	.002	.993
	X_3	Correlation	.013	-.305	1.000	-.501	-.312
	X_4	Correlation	.002	.002	-.501	1.000	.277
	Y	Correlation	.904	.993	-.312	.300	1.000

a. Cells contain zero-order (pearson) correlations.

Table 3 explains that the regression test process on the ANN pattern in making predictions produces a result of 98.8%. These results prove that the indicator variables used have a relationship with each other. The correlation results presented also provide significant results based on the indicator correlation test process with the resulting prediction output. The results of the correlation test of the population indicator (X_1) have a correlation of 90.4%, poverty data (X_2) is 99.3%, labor wages (X_3) are 31.2%, and poverty percentage (X_4) 27.7% affects the predicted output (Y). Based on the results of the regression and correlation test validation, the prediction analysis model and classification of poverty status levels provides output with a fairly good level of accuracy.

Based on the discussion that has been carried out, this study provides an updated analytical model for the prediction and classification process. The update is presented in the validation measurement process of the

analysis pattern and the resulting output. This measurement can test the entire analysis process carried out so that the output obtained provides a much better level of accuracy. Furthermore, the output of this research also presents the novelty of the concept of systematic analysis. The concepts and methods used have been able to be adopted quite well in DL learning to provide significant results. With this, the proposed analytical model is expected to provide a precise and accurate presentation of the analysis process so that the output of the analysis results can be taken into consideration by related parties in making decisions.

4. CONCLUSION

The development of the analytical model proposed in the prediction and classification process using DL provides fairly good performance with a validation level of 99.8% on the analysis pattern and output. These results prove that the development of the DL analysis model can present a structured model based on the stages of the process that have been carried out. This development is presented at the pre-processing stage which has a major contribution to forming a precise and accurate analysis pattern. Overall, the results of this study have an updated analytical model presented in the validation process to measure the accuracy of the results and provide an improvement in the analysis process. Based on these findings, prediction analysis and classification models can be used as a form of control and handling of poverty problems.

ACKNOWLEDGEMENTS

The author's gratitude to Dr. Hj. Zerni Melmusi, SE, MM, Ak, CA, as the head of Yayasan Perguruan Tinggi Komputer YPTK Padang who has provided support for this research.

REFERENCES

[1] J. Niehues, "Armut in Deutschland — Einfluss der Coronapandemie noch nicht ablesbar," *Wirtschaftsdienst*, vol. 102, no. 3, pp. 181–184, Mar. 2022, doi: 10.1007/s10273-022-3130-7.

[2] P. Singh and F. Pattanaik, "Economic status of women in India: paradox of paid–unpaid work and poverty," *International Journal of Social Economics*, vol. 46, no. 3, pp. 410–428, Mar. 2019, doi: 10.1108/IJSE-05-2018-0277.

[3] E. S. Rahayu, Setyowati, and R. Rahwadwiati, "The impact of climate change on poverty in the pandemic of Covid-19 in Bengawan Solo Watershed, Wonogiri District, Central Java," *IOP Conference Series: Earth and Environmental Science*, vol. 824, no. 1, p. 12109, Jul. 2021, doi: 10.1088/1755-1315/824/1/012109.

[4] J. Jung, J. Manley, and V. Shrestha, "Coronavirus infections and deaths by poverty status: the effects of social distancing," *Journal of Economic Behavior & Organization*, vol. 182, pp. 311–330, Feb. 2021, doi: 10.1016/j.jebo.2020.12.019.

[5] Y. K. Dwivedi *et al.*, "Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *International Journal of Information Management*, vol. 57, p. 101994, Apr. 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.

[6] J. Furman and R. Seamans, "AI and the economy," *Innovation Policy and the Economy*, vol. 19, pp. 161–191, Jan. 2019, doi: 10.1086/699936.

[7] X. Ren, "Application and innovation of traditional financial big data based on AI algorithm," in *2021 4th International Conference on Information Systems and Computer Aided Education*, Sep. 2021, pp. 482–485, doi: 10.1145/3482632.3482734.

[8] A. Alsharkawi, M. Al-Fetyani, M. Dawas, H. Saadeh, and M. Alyaman, "Poverty classification using machine learning: the case of Jordan," *Sustainability*, vol. 13, no. 3, p. 1412, Jan. 2021, doi: 10.3390/su13031412.

[9] M. P. Naviamos and J. D. Niguidula, "A study on determining household poverty status," in *Proceedings of the 3rd International Conference on Software Engineering and Information Management*, Jan. 2020, pp. 79–84, doi: 10.1145/3378936.3378969.

[10] P. Bilton, G. Jones, S. Ganesh, and S. Haslett, "Classification trees for poverty mapping," *Computational Statistics & Data Analysis*, vol. 115, pp. 53–66, Nov. 2017, doi: 10.1016/j.csda.2017.05.009.

[11] A. Abu Bakar, R. Hamdan, and N. S. Sani, "Ensemble learning for multidimensional poverty classification," *Sains Malaysiana*, vol. 49, no. 2, pp. 447–459, Feb. 2020, doi: 10.17576/jsm-2020-4902-24.

[12] J. A. Talingdan, "Performance comparison of different classification algorithms for household poverty classification," in *2019 4th International Conference on Information Systems Engineering (ICISE)*, May 2019, pp. 11–15, doi: 10.1109/ICISE.2019.00010.

[13] D. Ispriyanti, A. Prahutama, and Mustafid, "Analysis of space and classification poverty in Semarang City using spatial-logistic regression," *Journal of Physics: Conference Series*, vol. 1217, no. 1, p. 12106, May 2019, doi: 10.1088/1742-6596/1217/1/012106.

[14] D. Saccone and M. Deaglio, "Poverty, emergence, boom and affluence: a new classification of economies," *Economia Politica*, vol. 37, no. 1, pp. 267–306, 2020, doi: 10.1007/s40888-019-00166-4.

[15] M. Capó, A. Pérez, and J. A. Lozano, "An efficient approximation to the K-means clustering for massive data," *Knowledge-Based Systems*, vol. 117, pp. 56–69, Feb. 2017, doi: 10.1016/j.knosys.2016.06.031.

[16] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11218 LNCS, pp. 139–156, doi: 10.1007/978-3-030-01264-9_9.

[17] A. N. Matheri, F. Ntuli, J. C. Ngila, T. Seodigeng, and C. Zvinowanda, "Performance prediction of trace metals and cod in wastewater treatment using artificial neural network," *Computers & Chemical Engineering*, vol. 149, p. 107308, 2021.

[18] Andrizal, Lifwarda, A. Hidayat, R. Susanti, N. Alfitri, and R. Chadry, "The identification of car engine combustion based on data pattern gas emission using sum square error method," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no. 5, pp. 1512–1519, Oct. 2019, doi: 10.18517/ijaseit.9.5.3928.

[19] L. E. Aik, T. W. Hong, and A. K. Junoh, "Distance weighted K-means algorithm for center selection in training radial basis function networks," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 1, pp. 54–62, 2019, doi: 10.11591/ijai.v8.i1.pp54-62.

[20] S. Tangwannawit and P. Tangwannawit, "An optimization clustering and classification based on artificial intelligence approach for internet of things in agriculture," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 1, p. 201, 2022, doi: 10.11591/ijai.v11.i1.pp201-209.
[21] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: a survey and review," *Advances in Intelligent Systems and Computing*, vol. 937, pp. 99–111, 2020, doi: 10.1007/978-981-13-7403-6_11.
[22] H. Rosenbusch, F. Soldner, A. M. Evans, and M. Zeelenberg, "Supervised machine learning methods in psychology: a practical introduction with annotated R code," *Social and Personality Psychology Compass*, vol. 15, no. 2, Feb. 2021, doi: 10.1111/spc3.12579.
[23] F. Luna-Perejón, J. M. Montes-Sánchez, L. Durán-López, A. Vazquez-Baeza, I. Beasley-Bohórquez, and J. L. Sevillano-Ramos, "IoT device for sitting posture classification using artificial neural networks," *Electronics*, vol. 10, no. 15, p. 1825, Jul. 2021, doi: 10.3390/electronics10151825.
[24] Pham Le Quoc Han, K. I. Bushmeleva, A. S. Uvaysova, and S. U. Uvaysov, "Artificial neural networks in the problem of classification of defects of electronic devices," *Journal of Communications Technology and Electronics*, vol. 66, no. 9, pp. 1068–1074, Sep. 2021, doi: 10.1134/S1064226921090047.
[25] R. Al Ahsan and L. Wuttisittikulkij, "Artificial neural network (ANN) based classification of high and low PAPR OFDM signals," in *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Jun. 2021, pp. 1–4, doi: 10.1109/ITC-CSCC52171.2021.9501484.
[26] W. N. H. F. W. Zurey, S. Ismail, and A. Mustapha, "Forecasting accuracy: a comparative study between artificial neural network and autoregressive model for streamflow," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 3, pp. 464–472, 2020, doi: 10.11591/ijai.v9.i3.pp464-472.
[27] A. S. Hamzah and A. Mohamed, "Classification of white rice grain quality using ANN: a review," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 4, p. 600, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp600-608.
[28] J. Nalepa and M. Kawulok, "Selecting training sets for support vector machines: a review," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 857–900, Aug. 2019, doi: 10.1007/s10462-017-9611-1.
[29] A. Naresh Kumar, M. Suresh Kumar, M. Ramesha, B. Gururaj, and A. Srikanth, "Support vector machine based fault section identification and fault classification scheme in six phase transmission line," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 4, pp. 1019–1024, 2021, doi: 10.11591/IJAI.V10.I4.PP1019-1024.
[30] A. Iranmehr, H. Masnadi-Shirazi, and N. Vasconcelos, "Cost-sensitive support vector machines," *Neurocomputing*, vol. 343, pp. 50–64, May 2019, doi: 10.1016/j.neucom.2018.11.099.
[31] M. P. Naviamos and J. D. Niguidula, "Classification and visualization of poverty status: analyzing the need for poverty assistance using SVM," in *Advances in Computer, Communication and Computational Sciences*, S. K. Bhatia, S. Tiwari, S. Ruidan, M. C. Trivedi, and K. K. Mishra, Eds. Singapore: Springer Singapore, 2021, pp. 571–581.
[32] Y. Abo-Elnaga and S. Nasr, "K-means cluster interactive algorithm-based evolutionary approach for solving bilevel multi-objective programming problems," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 811–827, Jan. 2022, doi: 10.1016/j.aej.2021.04.098.
[33] A. Triayudi, W. O. Widyarto, L. Kamelia, Iksal, and Sumiati, "Clg clustering for dropout prediction using log-data clustering method," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, pp. 764–770, 2021, doi: 10.11591/ijai.v10.i3.pp764-770.
[34] K. Chowdhury, D. Chaudhuri, A. K. Pal, and A. Samal, "Seed selection algorithm through K-means on optimal number of clusters," *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 18617–18651, Jul. 2019, doi: 10.1007/s11042-018-7100-4.
[35] D. Zhao *et al.*, "K-means clustering and kNN classification based on negative databases," *Applied Soft Computing*, vol. 110, p. 107732, 2021.
[36] S. Arlis and S. Defit, "Machine learning algorithms for predicting the spread of Covid–19 in Indonesia," *TEM Journal*, vol. 10, no. 2, pp. 970–974, 2021, doi: 10.18421/TEM102-61.
[37] T. Kogut and A. Slowik, "Classification of airborne laser bathymetry data using artificial neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1959–1966, 2021, doi: 10.1109/JSTARS.2021.3050799.
[38] G. Shobana and D. R. N. Priya, "Cancer drug classification using artificial neural network with feature selection," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Feb. 2021, vol. 14, pp. 1250–1255, doi: 10.1109/ICICV50876.2021.9388542.
[39] M. Yanto, S. Sanjaya, Y. Yulasmi, D. Guswandi, and S. Arlis, "Implementation multiple linear regresion in neural network predict gold price," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 3, pp. 1635–1642, Jun. 2021, doi: 10.11591/ijeecs.v22.i3.pp1635-1642.
[40] O. Adigun and B. Kosko, "Bidirectional backpropagation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 5, pp. 1982–1994, 2020, doi: 10.1109/TSMC.2019.2916096.
[41] S. Y. Heng *et al.*, "Artificial neural network model with different backpropagation algorithms and meteorological data for solar radiation prediction," *Scientific reports*, vol. 12, no. 1, pp. 1–18, 2022.
[42] M. O. Moreira, P. P. Balestrassi, A. P. Paiva, P. F. Ribeiro, and B. D. Bonatto, "Design of experiments using artificial neural network ensemble for photovoltaic generation forecasting," *Renewable and Sustainable Energy Reviews*, vol. 135, p. 110450, 2021.
[43] P. Soto *et al.*, "Atari: a graph convolutional neural network approach for performance prediction in next-generation wlans," *Sensors*, vol. 21, no. 13, p. 4321, 2021.
[44] G. Battineni, N. Chintalapudi, and F. Amenta, "Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM)," *Informatics in Medicine Unlocked*, vol. 16, 2019, doi: 10.1016/j.imu.2019.100200.
[45] J. D. J. H. Harikiran, "Hyperspectral image classification using support vector machines," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 4, pp. 684–690, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp684-690.
[46] P. Nanglia, S. Kumar, A. N. Mahajan, P. Singh, and D. Rathee, "A hybrid algorithm for lung cancer classification using SVM and neural networks," *ICT Express*, vol. 7, no. 3, pp. 335–341, 2021.
[47] D. K. Choubey, S. Tripathi, P. Kumar, V. Shukla, and V. K. Dhandhania, "Classification of diabetes by kernel based SVM with PSO," *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, vol. 14, no. 4, pp. 1242–1255, 2021.
[48] H. Pan, X. You, S. Liu, and D. Zhang, "Pearson correlation coefficient-based pheromone refactoring mechanism for multi-colony ant colony optimization," *Applied Intelligence*, vol. 51, no. 2, pp. 752–774, Feb. 2021, doi: 10.1007/s10489-020-01841-x.
[49] P. Waldmann, "On the use of the pearson correlation coefficient for model evaluation in genome-wide prediction," *Frontiers in Genetics*, vol. 10, Sep. 2019, doi: 10.3389/fgene.2019.00899.
[50] I. Jebli, "Prediction of solar energy guided by pearson correlation using machine learning," *Energy*, vol. 224, 2021, doi: 10.1016/j.energy.2021.120109.
[51] H. Zhu, X. You, and S. Liu, "Multiple ant colony optimization based on pearson correlation coefficient," *IEEE Access*, vol. 7, pp. 61628–61638, 2019, doi: 10.1109/ACCESS.2019.2915673.

[52] E. M. Golafshani, A. Behnood, and M. Arashpour, "Predicting the compressive strength of normal and high-performance concretes using ANN and ANFIS hybridized with grey wolf optimizer," *Construction and Building Materials*, vol. 232, p. 117266, Jan. 2020, doi: 10.1016/j.conbuildmat.2019.117266.
[53] M. Yanto, S. Arlis, J. Na, Y. Yuhandri, and D. M. Putra, "Hybrid method air quality classification analysis model," *TEM Journal*, vol. 11, no. 2, pp. 829–836, 2022, doi: 10.18421/TEM112.

BIOGRAPHIES OF AUTHORS

Musli Yanto Born in Jakarta on July 7, 1989. A lecturer at the Putra Indonesia University YPTK Padang. The educational history of the Informatics Engineering undergraduate program was completed in 2012 and the Informatics Engineering Masters program in 2014. His areas of expertise include artificial intelligence (AI), expert systems (ES), data mining (DM), decision support systems (DSS). He can be contacted at email: musli_yanto@upiyptk.ac.id.

Yogi Wiyandra Educational background comes from a computer system study program. Advanced education for the postgraduate program in Information Technology, Faculty of Computer Science, was completed in 2014. In my work, I am a lecturer at the Putra Indonesia University, YPTK Padang, majoring in computer systems. Research history includes expert systems, decision support systems, and several branches of artificial intelligence. He can be contacted at email: yogiwiyandra@gmail.com.

Sarjon Defit Place of Birth: Padang Sibusuk/07 August 1970. He is the Chancellor of Putra Indonesia University YPTK Padang. Currently active as a lecturer in Computer Science. The educational history of SI at the College of Informatics and Computer Management (STMIK "YPTK" Padang) with a graduation in 1993. An educational history of S2 at Universiti Teknologi Malaysia, Johor Bahru, graduated in 1998. Then a Doctoral Education History at Universiti Teknologi Malaysia, Johor Bahru, graduated in 2003. The field of science consists of data mining, artificial intelligence, decision support systems, and others. He can be contacted at email: sarjon_defit@upiyptk.ac.id.