Indonesian Journal of Electrical Engineering and Computer Science
Vol. 27, No. 3, September 2022, pp. 1557~1566
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v27.i3.pp1557-1566

Machine learning classification of infectious disease distribution status

Irzal Arief Wisky[1], Musli Yanto[2], Yogi Wiyandra[3], Hadi Syahputra[3], Febri Hadi[2]
[1]Department of Information Systems, Faculty of Computer Science, Universitas Putra Indonesia "YPTK", Padang, Indonesia
[2]Department of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia "YPTK", Padang, Indonesia
[3]Department of Computer Systems, Faculty of Computer Science, Universitas Putra Indonesia "YPTK", Padang, Indonesia

Article Info

Article history:

Received Feb 2, 2022
Revised Jun 14, 2022
Accepted Jul 3, 2022

Keywords:

Classification
Infectious diseases
Information dan knowladge
Machine learning
Public health

ABSTRACT

Infectious diseases are common diseases and are caused by microorganisms such as viruses, bacteria, and parasites. Indicators of the spread of this disease can be seen based on the population level and the number of confirmed cases. This study aims to develop a machine learning (ML) analysis model using the K-means cluster, artificial neural network (ANN), and decision tree (DT) methods. The dataset used in this study was obtained based on the number of confirmed patients and the distribution of the population. The analysis process is divided into two stages, namely preprocessing and the classification process. The pre-processing stage aims to produce a classification pattern that can describe the level of distribution status. The classification pattern will be continued at the classification analysis stage using ANN and DT. Classification analysis gave significant results with an accuracy rate of 99.77%. The results of the classification analysis can also describe the level of knowledge distribution based on the decision tree. Overall, the contribution of this research is to develop a classification analysis model that presents the latest information and knowledge. The results of the research presented also have an impact on the control process in environmental management and public health.

This is an open access article under the CC BY-SA license.

Corresponding Author:

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

1. INTRODUCTION

Analysis of the status of the spread of infectious diseases is used as a tool for the public health management process [1]. In general, these infectious diseases consist of influenza, hypertension, diarrhea, tuberculosis and others [2]. Infectious diseases cause pain, paralysis, and even death with a fairly high percentage rate of 69.91% [3]. The transmission rate is spread nationally so that it is one of the main health problems today [4]. To help solve these problems, the classification analysis process can play an active role in developing a model to provide the best alternative solution.

Classification analysis has been developed in various problems to provide the desired results [5]. These various models use several methods in conducting classification analysis [6]. The analysis model can be seen in the concept of machine learning (ML). The model has been able to contribute quite effectively to the classification process [7]. ML works optimally to present output with a fairly good level of accuracy [8]. The development of ML in several studies shows a significant graph in solving problems in the world of health [9]. These problems can be seen from the process of identification, classification, and prediction [10], [11]. In this case, ML will also be used to carry out a classification analysis process on the problem of the status of the spread of infectious diseases.

The method that will be used in the classification analysis process involves K-means clusters, artificial neural networks (ANN), and decision trees (DT). This method can work more effectively in presenting the desired results. K-means cluster is a method that can categorize data based on mathematical calculations [12]. K-means is a very popular method used in the big data concept [13], [14]. This method works on pre-processing to produce a classification pattern [15]. The pattern obtained can be proven to be effective in carrying out the classification analysis process [16]. The results given from pre-processing using K-means clusters will be forwarded to the analysis using the ANN concept.

ANN is a concept that is widely used in ML [17]. This method is a supervised learning concept that can produce fairly good accuracy results [18]. ANN performs an analysis based on the weighted value obtained to produce the output [19]. This concept continues to develop as many problems have been well resolved [20]. In the process of analyzing the classification of the spread of infectious diseases, ANN is expected to provide optimal performance. To get these results, the stages of the training and testing process in learning will be maximized to produce outputs [21]. ANN performance can be seen based on the level of accuracy and error in the output presentation [22]. The outputs obtained in the process will later be re-analyzed in order to present the information and knowledge needed. The ANN output results obtained will be re-analyzed using the DT concept.

In general, DT works by conducting analysis based on previously obtained patterns to present knowledge-based [23]. The DT analysis process will refine the analysis of the spread of infectious diseases in the form of a decision tree. The results represented in the DT represent a previously hidden information and knowledge [24]. Several analytical models have been developed with DT such as the classification process for a disease that aims to support a decision [25].

Overall, this study presents the novelty of the classification analysis model. The model was developed through pre-processing and classification processes on the ML concept. The up-to-date model also provides a structured and systematic analysis process to provide precise and accurate output. With this, this research can to present new knowledge and information that describes the spread of infectious diseases. Furthermore, that this research will also be useful for related parties in environmental and community health management.

2. RESEARCH METHOD

The classification analysis process using the machine learning concept has 2 stages, namely the pre-processing stage and the classification analysis stage. The methods and algorithms used consist of the K-means cluster, ANN, and DT. The description of the research, stages can be seen in Figure 1.

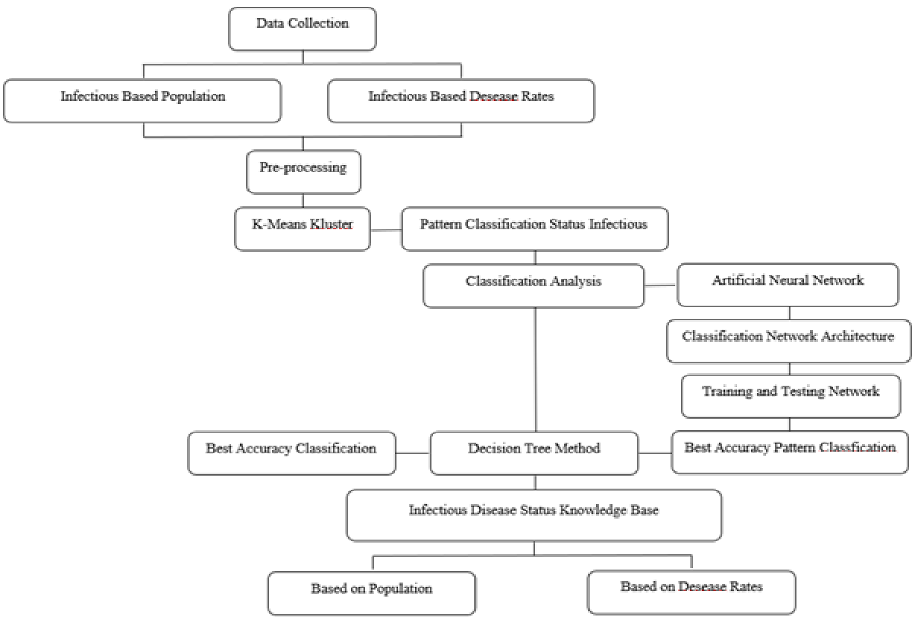

Figure 1. Research stages

Figure 1 explain the analysis process starting with data analysis based on population size and infectious disease. the classification stage starts from pre-processing using the K-means cluster algorithm aimed at obtaining patterns in classification analysis. With the analysis pattern obtained, the classification process will be carried out using an ANN. ANN learning using a feedforward algorithm aims to get the maximum classification results. The classification analysis stage will be continued by using the DT method to obtain information and knowledge. The results of the analysis based on numbers present the output in the form of a DT in the status of the spread of the disease seen from the population and the number of infection cases.

2.1. Data collection

The discussion of this study uses population data and infectious disease figures for 3 periods, 2018, 2019, and 2020. The source of the data used comes from the Pesisir Selatan District Health Office. The data will be analyzed previously to be used as variables in conducting classification analysis. The variables used will be seen based on 2 indicators based on population and the number of cases of infectious diseases. The variables used in the analysis process can be seen in Table 1.

Table 1. Variable of classification analysis of infectious disease distribution status

Population	Variable	Infected number	Variable
Population number	X1	Ispa	X8
Male	X2	Influenza	X9
Famele	X3	Gastritis	X10
(1-12 Year)	X4	Hipertensi	X11
(1-30 Year)	X5	Diarrhea	X12
(31-45 Year)	X6	Rheumatism	X13
> 45 Year	X7	Fever	X14
		Commond cold	X15
		Asthma	X16
		Dengue fever	X17
		Tuberculosis	X18
		Dispepsia	X19
		Skin Allergies	X20

K-means cluster is an initialization algorithm for grouping data [26]. The implementation of K-means clusters can provide results in the form of analysis patterns for recommendations for classification, determination, and prediction processes [27]. K-means works by looking for and finding similar patterns in the data with the output of information and knowledge [28]. This algorithm is an exploratory analysis concept that can be applied in supervised machine learning [29]. The concept of the K-means cluster algorithm can be seen in (1) [30].

$$\sum_{j=1}^{k} \sum_{xi \in} ||Xi - \mu||_2^2 \quad (1)$$

2.2. Artificial neural network (ANN)

ANN is a method that is widely used in machine learning [31]. The ANN method in machine learning gives promising results to produce a comprehensive review [32]. The implementation of this concept can carry out learning in the classification analysis process with better output [33]. ANN is a non-linear concept with mathematical calculations on a modeled problem to produce the output [34]. ANN performance results provide a fairly high level of sensitivity based on network output [35].

2.3. Decision tree (DT)

DT is a classification analysis concept developed to provide decisions based on data filters [36]. The development of this method is used in the classification process by validating the tests carried out [37]. DT is used in solving problems on complex data to produce information and knowledge that is presented in the form of a DT [38]. The performance process still uses mathematical calculations in the development of decision-making systems [39]. The equations in the DT method can be seen in (2) [40].

$$Entropy(S) = -\sum_{i=1}^{c} \mathrm{PS(ci)logPS(ci)} \quad (2)$$

3. RESULTS AND DISCUSSION

3.1. Pre-processing analysis

The pre-processing analysis stage aims to maximize the classification process that will be carried out [41]. This process can provide a better and structured analysis presentation to get better output results [42]. In this pre-processing analysis, the algorithm used is the K-means cluster. This algorithm can group data based on the level of closeness of the relationship in the data [43]. The results of the pre-processing analysis using K-means clusters can be seen in Table 2. Table 2 shows that the results of the cluster provide a classification pattern based on the group of data on the status of the spread of infectious disease numbers. The cluster results show the level of distribution with high status (C1) as many as 8 items, moderate as many as 2 items, and low as 5 items. From Table 1, it can be seen that there are 3 categories of infectious disease distribution status, namely high, medium and low status. With the results of the pre-processing, a classification process will be carried out for the spread of infectious disease numbers.

Table 2. Results of pre-processing K-means cluster

Population (X1-X7)							Infected Number (X8-X20)													Y
150	788	720	332	252	209	564	21	14	15	16	46	56	20	96	26	9	3	0	71	High
215	109	105	473	351	306	805	26	75	48	87	57	12	16	34	67	8	4	0	56	High
137	684	694	303	219	201	516	32	15	86	71	10	29	89	35	74	6	9	0	28	High
151	758	757	333	242	219	567	32	65	16	92	55	15	0	10	67	0	0	0	0	High
264	134	130	582	430	378	991	52	22	37	29	0	31	54	24	46	4	4	0	18	High
160	807	794	352	258	230	599	71	73	14	16	0	68	17	0	0	0	0	0	0	High
451	225	225	994	722	655	169	30	13	11	11	52	12	0	0	0	0	0	0	0	Low
303	147	155	666	471	451	113	73	0	26	28	73	25	45	49	88	3	8	3	29	High
525	257	268	115	823	779	196	80	34	69	54	86	63	38	0	13	5	5	0	25	Midd
505	252	253	111	806	734	188	86	0	41	23	10	36	22	0	98	4	6	0	17	Midd
314	154	160	692	495	464	117	57	31	56	41	0	30	0	91	0	0	0	0	11	Low
465	230	234	102	736	680	173	36	23	28	19	38	11	18	0	24	0	0	0	0	Low
367	176	190	809	565	553	137	40	69	26	14	87	31	21	75	81	6	5	1	72	Low
727	347	380	160	111	110	272	27	0	12	72	56	56	34	87	87	6	1	0	12	High
485	240	244	106	770	709	181	29	56	18	16	79	16	24	47	68	0	0	0	13	Low

3.2. Classification analysis

The classification process in the discussion aims to see the status of the spread of infectious diseases based on infection numbers and population. In this case, the analysis process begins by using the ANN method with a feedforward algorithm. The ANN method is a concept that can carry out learning with better outputs [44]. ANN can also be implemented in the case of the classification of a disease by using the concept of feedforward learning. The results given have a fairly high level of accuracy [45]. Basically, this method learns the pattern of network architecture formed by the training and testing process [46]. The study aims to obtain the best network architecture pattern that will be used in the classification analysis process [47]. The results of the best classification of the ANN network architecture pattern can be seen in Figure 2.

Figure 2 is the result of the best classification ANN network architecture through the learning process by training and testing the previous classification pattern. The ANN architectural pattern has 3 layers, namely the input layer, the hidden layer, and the output layer [48]. The architecture is shown in Figure. 2 consists of a layer of 20 units of the input layer, 5 layers of hidden layers of five units namely (50, 35, 25, 15, and 10), and 1 layer of the output layer of one unit. This architectural pattern will be used to carry out the classification process on the status of the spread of infectious diseases. The results of the classification process using ANN can be seen in the learning output graph in Figure 3.

Figure 3 describes the results of the classification analysis using ANN which has a fairly good output. These results can be seen from the performance value of 0.0731% so that the ANN learning process approaches the maximum results in the classification process. ANN output can also be proven the level of relationship based on the linearity value of the input used [49]. In this case, the level of relationship between input and output units is 96.98%. These results are sufficient to illustrate that ANN is able to perform classification analysis on the status of the spread of infectious diseases.

The analysis process will still be continued with the aim of exploring knowledge based on the classification pattern that has been analyzed with ANN. The DT method can present output in the form of knowledge-based [50]. In concept, DT performs analysis to find information and knowledge hidden in a pile of data [51]. The classification analysis process using the DT concept will focus on two directions, namely based on the population level and the number of distribution figures. The purpose of this two-way classification analysis is to find information and knowledge from a different perspective. The results of the analysis provided by DT can be used as a reference to follow up the handling process for related parties. The results of the DT classification analysis based on the population level can be seen in Figure 4.

Figure 4 explains that DT is capable of generating information and knowledge in the form of a DT image. The classification results presented can be seen that the population with the age category>45 years has the highest risk for transmission. then for the population aged 31-45 and under 30 years, it also has a relatively moderate level of probability. To ensure the results obtained in Figure 4, the analysis process will also be seen based on the rate of spread of infectious diseases. The results of the analysis can be described in Figures 5 and 6.

Figure 5 is the result of a DT that describes information and knowledge about the status of the spread of infectious diseases. These results are based on indicators that have been analyzed previously. Figure 6 is a form of classification rule that presents knowledge in the spread of infectious diseases. Overall the results of the classification analysis developed in the new model are able to provide significant results.

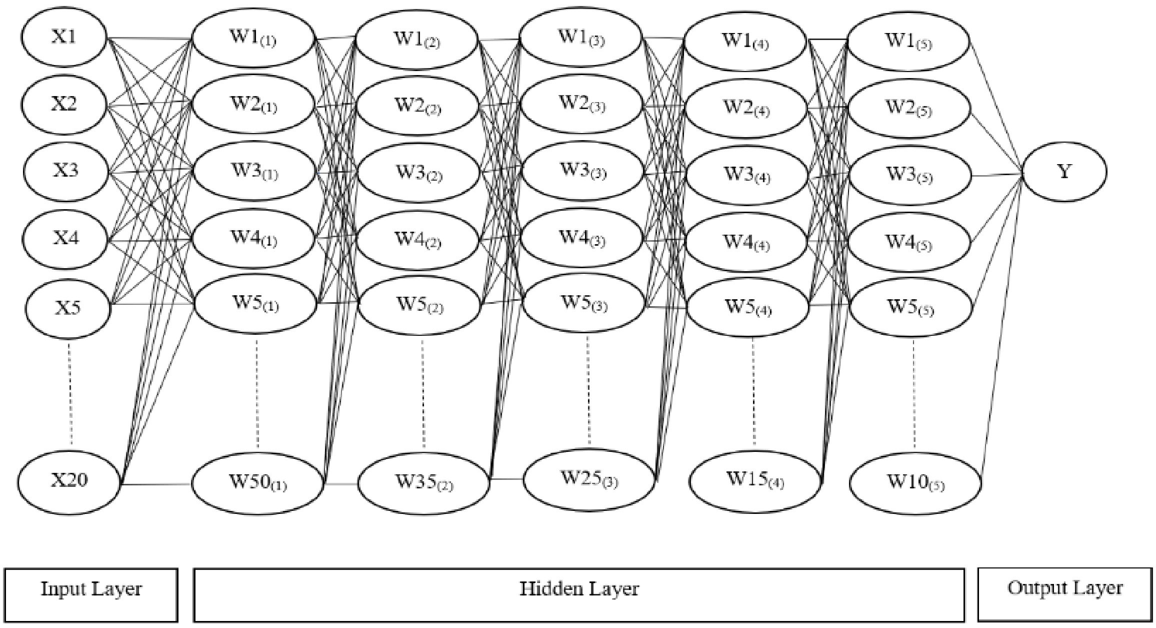

Figure 2. Architecture of ANN classification

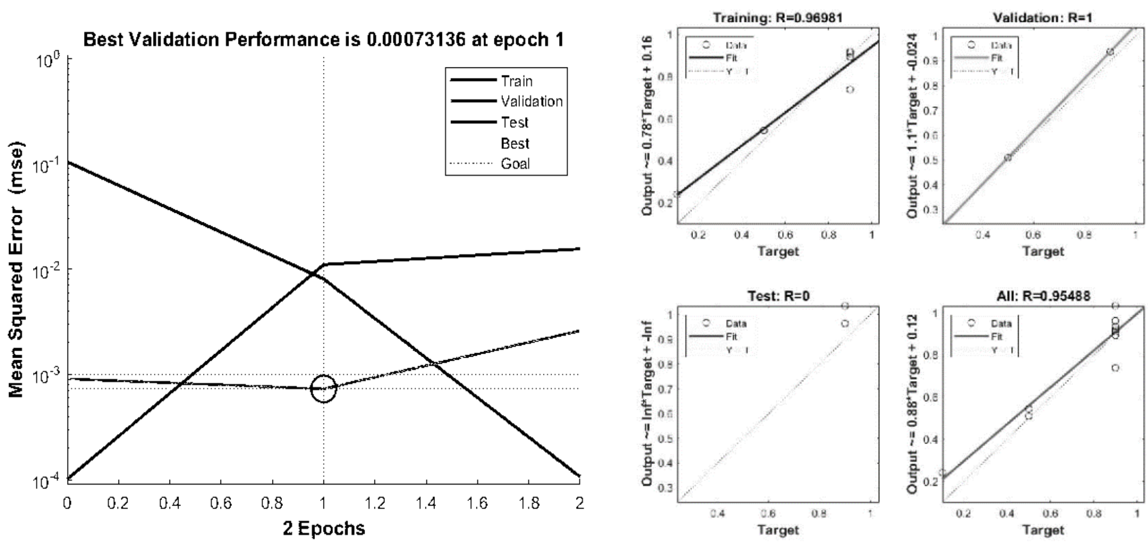

Figure 3. Graph of learning outcomes

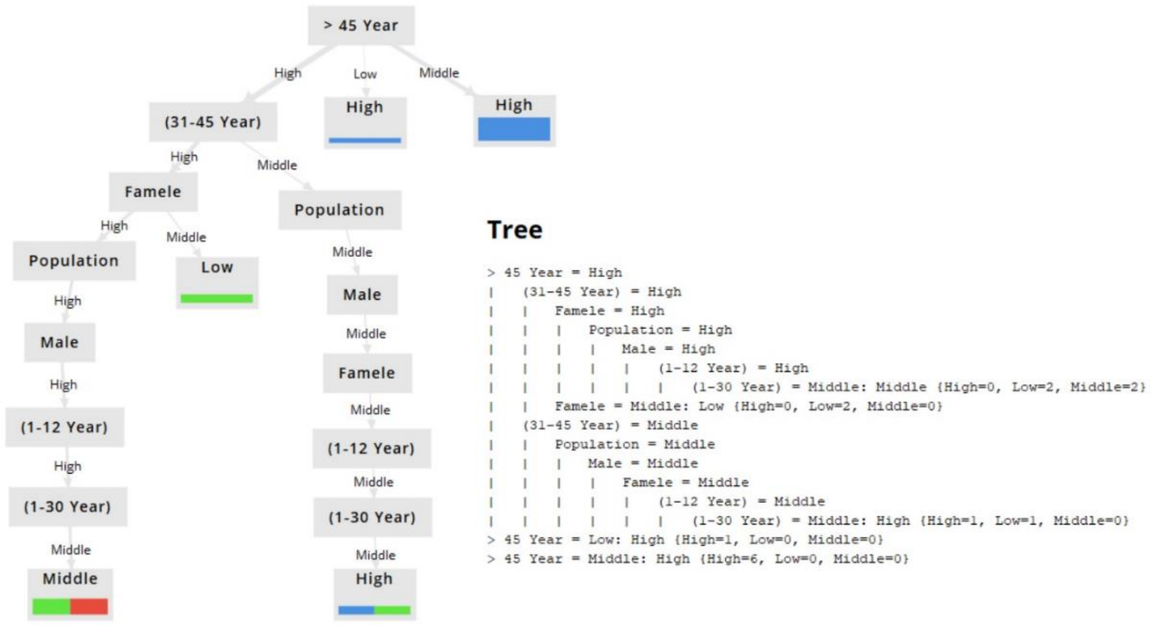

Figure 4. Results of DT classification analysis based on population level

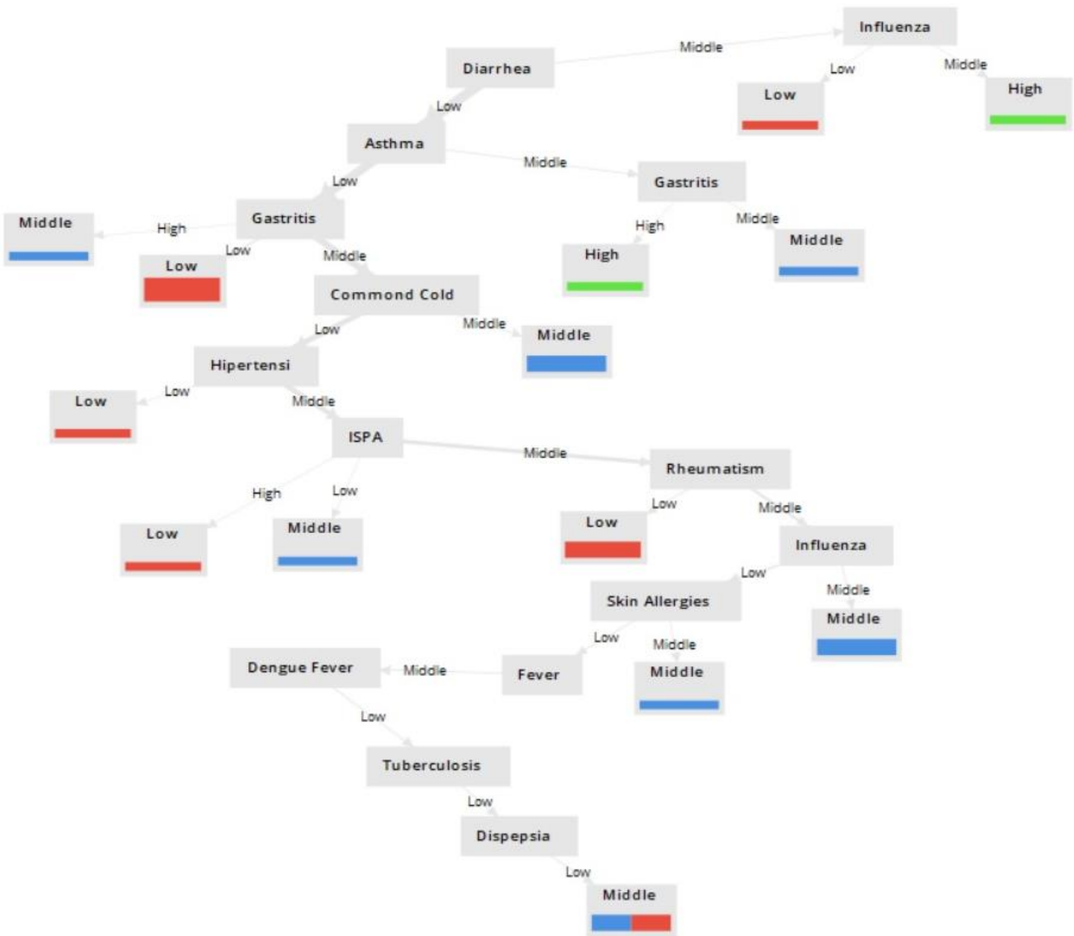

Figure 5. Decision tree of classification analysis results

```
Tree

Diarrhea = Low
|   Asthma = Low
|   |   Gastritis = High: Middle {Middle=1, High=0, Low=0}
|   |   Gastritis = Low: Low {Middle=0, High=0, Low=3}
|   |   Gastritis = Middle
|   |   |   Commond Cold = Low
|   |   |   |   Hipertensi = Low: Low {Middle=0, High=0, Low=1}
|   |   |   |   Hipertensi = Middle
|   |   |   |   |   ISPA = High: Low {Middle=0, High=0, Low=1}
|   |   |   |   |   ISPA = Low: Middle {Middle=1, High=0, Low=0}
|   |   |   |   |   ISPA = Middle
|   |   |   |   |   |   Rheumatism = Low: Low {Middle=0, High=0, Low=2}
|   |   |   |   |   |   Rheumatism = Middle
|   |   |   |   |   |   |   Influenza = Low
|   |   |   |   |   |   |   |   Skin Allergies = Low
|   |   |   |   |   |   |   |   |   Fever = Middle
|   |   |   |   |   |   |   |   |   |   Dengue Fever = Low
|   |   |   |   |   |   |   |   |   |   |   Tuberculosis = Low
|   |   |   |   |   |   |   |   |   |   |   |   Dispepsia = Low: Middle {Middle=1, High=0, Low=1}
|   |   |   |   |   |   |   |   Skin Allergies = Middle: Middle {Middle=1, High=0, Low=0}
|   |   |   |   |   |   |   Influenza = Middle: Middle {Middle=2, High=0, Low=0}
|   |   |   Commond Cold = Middle: Middle {Middle=2, High=0, Low=0}
|   Asthma = Middle
|   |   Gastritis = High: High {Middle=0, High=1, Low=0}
|   |   Gastritis = Middle: Middle {Middle=1, High=0, Low=0}
Diarrhea = Middle
|   Influenza = Low: Low {Middle=0, High=0, Low=1}
|   Influenza = Middle: High {Middle=0, High=1, Low=0}
```

Figure 6. Classification rules for the spread of infectious diseases

The analytical model presented is also quite effective in presenting the up-to-date process of ML classification analysis. The update of the model can be seen based on the output of the analysis stages that have been carried out. The overall analysis results have been validated by measuring the level of accuracy and error as well as testing the performance and sensitivity of the analytical model. With these results, the proposed analytical model is able to provide an update on the previous model in describing the classification of the status of the spread of infectious diseases.

4. CONCLUSION

The development of classification analysis using ML gives quite good results. Overall, this study presents an updated analysis model for the classification of the status of the spread of infectious diseases. The analysis process provides output in two directions, namely classification based on data on the number of infected cases and population distribution. These results are obtained through pre-processing in order to obtain a precise and accurate analysis pattern. Classification analysis provides an accuracy rate of 99.77% and an error rate of 0.33%. Furthermore, the output of the classification results is also able to describe the DT with an accuracy of 91.67%. The DT will be used as an information and new knowledge for related parties. The knowledge gained can also be useful in carrying out environmental and community health management processes.

REFERENCES

[1] L. Tang *et al.*, "A review of multi-compartment infectious disease models," *International Statistical Review*, vol. 88, no. 2, pp. 462–513, Aug. 2020, doi: 10.1111/insr.12402.

[2] C. N. Fries, E. J. Curvino, J. L. Chen, S. R. Permar, G. G. Fouda, and J. H. Collier, "Advances in nanomaterial vaccine strategies to address infectious diseases impacting global health," *Nature Nanotechnology*, vol. 16, no. 4, 2021, doi: 10.1038/s41565-020-0739-9.

[3] Kemenkes, *Main results of 2018 basic health research (in Indonesian: Hasil Utama Riset Kesehatan Dasar 2018)*, Jakarta: The National Institute of Health Research and Development, 2018. [Online]. Available: https://drive.google.com/open?id=1MRXC4lMDera5949ezbbHj7UCUj5_EQmY

[4] M. I. Hadi, Munawir, M. Y. Alamudi, D. Suprayogi, and M. Widiyanti, "Detection of emerging infectious disease in chiroptera brachjatis and rhinolopus boorneensis as reservoirs of zoonotic diseases in Indonesia," *Indian Journal of Forensic Medicine and Toxicology*, vol. 14, no. 3, pp. 2027–2032, Jul. 2020, doi: 10.37506/ijfmt.v14i3.10725.

[5] A. D. Association, "2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2021," *Diabetes Care*, vol. 43, no. Supplement 1, pp. S14–S31, Jan. 2020, doi: 10.2337/dc20-S002.

[6] G. Vergnaud, "A classification of cognitive tasks and operations of thought involved in addition and subtraction problems," in *Addition and Subtraction: A Cognitive Perspective*, Routledge, 2020, pp. 39–59.

[7] G. S. Randhawa, M. P. M. Soltysiak, H. El Roz, C. P. E. de Souza, K. A. Hill, and L. Kari, "Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study," *PLOS ONE*, vol. 15, no. 4, Apr. 2020, doi: 10.1371/journal.pone.0232391.

[8] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *Journal of Big Data*, vol. 7, Dec. 2020, doi: 10.1186/s40537-020-00327-4.

[9] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019, doi: 10.1056/NEJMra1814259.

[10] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," *Expert Systems with Applications*, vol. 117, pp. 345–357, Mar. 2019, doi: 10.1016/j.eswa.2018.09.029.

[11] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in Genetics*, vol. 9, Nov. 2018, doi: 10.3389/fgene.2018.00515.

[12] M. Kushwaha, H. Yadav, and C. Agrawal, "A review on enhancement to standard K-means clustering," in *Social Networking and Computational Intelligence*, 2020, pp. 313–326.

[13] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[14] T. Xie and T. Zhang, "The fast clustering algorithm for the big data based on K-means," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 18, no. 06, Nov. 2020, doi: 10.1142/S0219691320500538.

[15] R. Vohra, J. Pahareeya, A. Hussain, F. Ghali, and A. Lui, "Using self organizing maps and K means clustering based on RFM model for customer segmentation in the online retail business," in *Intelligent Computing Methodologies. ICIC 2020. Lecture Notes in Computer Science*, vol. 12465, D. S. Huang and P. Premaratne, Eds. 2020, pp. 484–497.

[16] 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, May 2021, doi: 10.18421/TEM102-61.

[17] G. Matar, J.-M. Lina, and G. Kaddoum, "Artificial neural network for in-bed posture classification using bed-sheet pressure sensors," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 101–110, Jan. 2020, doi: 10.1109/JBHI.2019.2899070.

[18] A. Talukder and B. Ahammed, "Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh," *Nutrition*, vol. 78, Oct. 2020, doi: 10.1016/j.nut.2020.110861.

[19] A. I. Lawal and M. A. Idris, "An artificial neural network-based mathematical model for the prediction of blast-induced ground vibrations," *International Journal of Environmental Studies*, vol. 77, no. 2, pp. 318–334, Mar. 2020, doi: 10.1080/00207233.2019.1662186.

[20] M. Sharma and R. Garg, "An artificial neural network based approach for energy efficient task scheduling in cloud data centers," *Sustainable Computing: Informatics and Systems*, vol. 26, Jun. 2020, doi: 10.1016/j.suscom.2020.100373.

[21] A. Maleki, M. S. Shadloo, and A. Rahmat, "Application of artificial neural networks for producing an estimation of high-density polyethylene," *Polymers*, vol. 12, no. 10, pp. 1–16, Oct. 2020, doi: 10.3390/polym12102319.

[22] E. Checcucci *et al.*, "Artificial intelligence and neural networks in urology: current clinical applications," *Minerva Urologica e Nefrologica*, vol. 72, no. 1, pp. 49–57, Feb. 2020, doi: 10.23736/S0393-2249.19.03613-0.

[23] M. A. Febriantono, S. Hadi Pramono, R. Rahmadwati, and G. Naghdy, "Classification of multiclass imbalanced data using cost-sensitive decision tree C5.0," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 1, pp. 65–72, Mar. 2020, doi: 10.11591/ijai.v9.i1.pp65-72.

[24] D. Phiri, M. Simwanda, V. Nyirenda, Y. Murayama, and M. Ranagalage, "Decision tree algorithms for developing rulesets for object-based land cover classification," *ISPRS International Journal of Geo-Information*, vol. 9, no. 5, May 2020, doi: 10.3390/ijgi9050329.

[25] M. P. Romero *et al.*, "Decision tree machine learning applied to bovine tuberculosis risk factors to aid disease control decision making," *Preventive Veterinary Medicine*, vol. 175, Feb. 2020, doi: 10.1016/j.prevetmed.2019.104860.

[26] V. P. Murugesan and P. Murugesan, "A new initialization and performance measure for the rough K-means clustering," *Soft Computing*, vol. 24, no. 15, pp. 11605–11619, Aug. 2020, doi: 10.1007/s00500-019-04625-9.

[27] S. Souabi, A. Retbi, M. K. Idrissi, and S. Bennani, "A recommendation approach in social learning based on k-means clustering," in *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Jun. 2020, pp. 1–5, doi: 10.1109/ISCV49265.2020.9204203.

[28] N. Mohd Ariff, M. A. Abu Bakar, and Z. H. Zamzuri, "Academic preference based on students' personality analysis through k-means clustering," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 16, no. 3, pp. 328–333, Jun. 2020, doi: 10.11113/mjfas.v16n3.1640.

[29] M. Capó, A. Pérez, and J. A. Lozano, "An efficient K-means clustering algorithm for tall data," *Data Mining and Knowledge Discovery*, vol. 34, no. 3, pp. 776–811, May 2020, doi: 10.1007/s10618-020-00678-9.

[30] W. Chang *et al.*, "Analysis of university students' behavior based on a fusion K-means clustering algorithm," *Applied Sciences*, vol. 10, no. 18, p. 6566, Sep. 2020, doi: 10.3390/app10186566.

[31] W. Gao and C. Su, "Analysis on block chain financial transaction under artificial neural network of deep learning," *Journal of Computational and Applied Mathematics*, vol. 380, Dec. 2020, doi: 10.1016/j.cam.2020.112991.

[32] Sonu and R. P. Bhokal, "Study of artificial neural network," *International Journal of Mathematics Trends and Technology*, vol. 47, no. 4, pp. 253–259, Jul. 2017, doi: 10.14445/22315373/IJMTT-V47P535.

[33] M. Yanto, S. Sanjaya, Yulasmi, D. Guswandi, and S. Arlis, "Implementation multiple linear regresion in neural network predict gold price," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 3, pp. 1635–1642, Jun. 2021, doi: 10.11591/ijeecs.v22.i3.pp1635-1642.

[34] L. N. Araujo, J. T. Belotti, T. A. Alves, Y. de S. Tadano, and H. Siqueira, "Ensemble method based on artificial neural networks to estimate air pollution health risks," *Environmental Modelling & Software*, vol. 123, Jan. 2020, doi: 10.1016/j.envsoft.2019.104567.

[35] A. Tamashiro *et al.*, "Artificial intelligence-based detection of pharyngeal cancer using convolutional neural networks," *Digestive Endoscopy*, vol. 32, no. 7, pp. 1057–1065, Nov. 2020, doi: 10.1111/den.13653.

[36] M. Syamala and N. Nalini, "A filter based improved decision tree sentiment classification model for real-time amazon product review data," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 1, pp. 191–202, Feb. 2020, doi: 10.22266/ijies2020.0229.18.

[37] S. A. Shah, P. Brown, H. Gimeno, J.-P. Lin, and V. M. McClelland, "Application of machine learning using decision trees for prognosis of deep brain stimulation of globus pallidus internus for children with dystonia," *Frontiers in Neurology*, vol. 11, Aug. 2020, doi: 10.3389/fneur.2020.00825.

[38] F. Tamin and N. M. S. Iswari, "Implementation of C4.5 algorithm to determine hospital readmission rate of diabetes patient," in *2017 4th International Conference on New Media Studies (CONMEDIA)*, Nov. 2017, pp. 15–18, doi: 10.1109/CONMEDIA.2017.8266024.
[39] M. M. Ghiasi, S. Zendehboudi, and A. A. Mohsenipour, "Decision tree-based diagnosis of coronary artery disease: CART model," *Computer Methods and Programs in Biomedicine*, vol. 192, Aug. 2020, doi: 10.1016/j.cmpb.2020.105400.
[40] J. M. Chen, "An introduction to machine learning for panel data," *International Advances in Economic Research*, vol. 27, no. 1, pp. 1–16, Feb. 2021, doi: 10.1007/s11294-021-09815-6.
[41] S. Sengan, P. Vidya Sagar, R. Ramesh, O. I. Khalaf, and R. Dhanapal, "The optimization of reconfigured real-time datasets for improving classification performance of machine learning algorithms," *Mathematics in Engineering, Science and Aerospace*, vol. 12, no. 1, pp. 43–54, 2021.
[42] L. Kurasinski and R.-C. Mihailescu, "Towards machine learning explainability in text classification for fake news detection," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2020, pp. 775–781, doi: 10.1109/ICMLA51294.2020.00127.
[43] J. Xu, D. Han, K.-C. Li, and H. Jiang, "A K-means algorithm based on characteristics of density applied to network intrusion detection," *Computer Science and Information Systems*, vol. 17, no. 2, pp. 665–687, 2020, doi: 10.2298/CSIS200406014X.
[44] S. Benouar, A. Hafid, M. Kedir-Talha, and F. Seoane, "Classification of impedance cardiography dZ/dt complex subtypes using pattern recognition artificial neural networks," *Biomedical Engineering / Biomedizinische Technik*, vol. 66, no. 5, pp. 515–527, Oct. 2021, doi: 10.1515/bmt-2020-0267.
[45] M. A. Aslam, C. Xue, M. Liu, K. Wang, and D. Cui, "Classification and prediction of gastric cancer from saliva diagnosis using artificial neural network," *Engineering Letters*, vol. 29, no. 1, pp. 10–24, 2021.
[46] F. Paulu and J. Hospodka, "Design of fully analogue artificial neural network with learning based on backpropagation," *Radioengineering*, vol. 30, no. 2, pp. 357–363, Jun. 2021, doi: 10.13164/re.2021.0357.
[47] E. A. Frimpong, A. Oluwasanmi, E. Y. Baagyere, and Q. Zhiguang, "A feedforward artificial neural network model for classification and detection of type 2 diabetes," *Journal of Physics: Conference Series*, vol. 1734, no. 1, Jan. 2021, doi: 10.1088/1742-6596/1734/1/012026.
[48] T. Chuntama, C. Suwannajak, P. Techa-Angkoon, B. Panyangam, and N. Tanakul, "Classification of astronomical objects in the galaxy M81 using machine learning techniques ii. an application of clustering in data pre-processing," in *2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Jun. 2021, pp. 1–6, doi: 10.1109/JCSSE53117.2021.9493825.
[49] V. G. Ramachandran, M. Afshar Alam, S. Zafar, and S. S. Biswas, "An implementation of ANN through data mining using regression analysis and clustering for prediction of irreversible dementia," *International Journal of Recent Technology and Engineering*, vol. 8, no. 1 Special Issue 4, pp. 859–861, 2019.
[50] G. Pappalardo, S. Cafiso, A. Di Graziano, and A. Severino, "Decision tree method to analyze the performance of lane support systems," *Sustainability*, vol. 13, no. 2, Jan. 2021, doi: 10.3390/su13020846.
[51] W. Chen *et al.*, "Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods," *Science of The Total Environment*, vol. 701, Jan. 2020, doi: 10.1016/j.scitotenv.2019.134979.

BIOGRAPHIES OF AUTHORS

Irzal Arief Wisky was educational background comes from the information systems study program. Further education for the post-graduate program in Information Technology, Faculty of Computer Science, was completed in 2013. In my job, I am a lecturer at Putra Indonesia University, YPTK Padang, majoring in information systems. The history of research includes decision support systems and expert systems. He can be contacted at email: irzal.arief12@gmail.com

Musli Yanto He is a lecturer at Universitas Putra Indonesia YPTK Padang. Educational background since 2014 has completed postgraduate studies in the field of informatics engineering. His skills include Data Science Analysis, Algorithms and Programming, Big Data and Artificial Intelligence. 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@upiyptk.ac.id

Hadi Syahputra , a graduate of Putra Indonesia University YPTK. In 2013 he obtained a Master's Degree in Computer with a concentration in Information Technology. His career as a lecturer at Putra Indonesia University YPTK began in 2013 until now. 2015 until now he has been trusted as a sports coach at Putra Indonesia University YPTK. Working in the field of networking and Micro Controllers with several published articles. Monitoring DNS Query with Pi-Hole Firewall Using Raspberry B+ Integrated with Mikrotik Router RB 931-2nd is an article that has been published in an international journal in 2021. He can be contacted at email: hadisaputra@upiyptk.ac.id

Febri Hadi educational background comes from a Informatic of Engineering study program. Advanced education for the postgraduate program in Information System, Faculty of Computer Science, was completed in 2013. In my work, I am a lecturer at the Putra Indonesia University, YPTK Padang, majoring in information systems. Research history includes expert systems, decision support systems, and data mining. He can be contacted at email: febri_hadi@upiyptk.ac.id