TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 21, No. 1, February 2023, pp. 104~112
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v21i1.24238

SOM-SIS approach to auto summary of clustering results on university academic performance

Rahayu Widayanti[1], Sarifuddin Madenda[2], Eri Prasetyo Wibowo[2], Khoerul Anwar[1]
[1]Departement of Information System, Pradnya Paramita College of Informatics Management and Computer, Malang, Indonesia
[2]Department of Information Technology, Faculty of Computer Science, Gunadarma University, Jakarta, Indonesia

Article Info

Article history:

Received Jun 14, 2021
Revised Nov 02, 2022
Accepted Nov 12, 2022

Keywords:

Academic performance
Quality assurance system
Self organizing map
TOPSIS

ABSTRACT

The analysis of the performance of higher education academic quality in terms of student achievement, study period, and drop out rates is still an intensive study among researchers. Several clustering methods are often used to understand student and graduate groups, in influencing college performance. However, the conventional method has only arrived at the results of clustering, so it is difficult to interpret it as a support for academic decisions, especially in mapping the position of universities against other universities nationally. This article introduces a combination of techniques from self organizing map and technique for oorder preference by similarity to an ideal solution (SOM-SIS), an auto-summarizing technique from clustering results as well as mapping university academic performance. First, the academic performance indicators are grouped using the self organizing map (SOM) method and the results are concluded using the technique for order preference by similarity to an ideal solution (TOPSIS) approach. The SOM-SIS technique was tested using data from one of the universities in Indonesia. As a result, the SOM-SIS technique has a 100% compatibility rate with the higher education quality assurance system, through recommendations from three university experts.

Corresponding Author:

Rahayu Widayanti
Departement of Information System, Pradnya Paramita College of Informatics Management and Computer
LA. Sucipto St. No. 249, Malang, East Java, Indonesia
Email: rahayu@stimata.ac.id

1. INTRODUCTION

Academic quality is one of the most important parts of the existence and sustainability of a university. To improve the quality of academic quality assurance, universities must be able to manage data effectively and find hidden knowledge from the data to support management decision making. Several universities have used data mining as part of education quality assurance with the aim of discovering knowledge and providing timely data for academic decision making [1], [2]. Data mining, commonly known as knowledge discovery in database (KDD) is an activity related to data collection and the use of historical data to find knowledge and information in big data [3]. In data mining, visualization is one of the easiest ways to understand multidimensional data structures and data analysis [4]. One of the clustering and visualization methods that is often used is self-organizing map (SOM) which maps high-dimensional data to low-dimensional space while maintaining the topological structure of the data [5].

The problem that arises is that conventional clustering has not been able to automatically conclude the position of higher education's academic performance compared to others. Likewise, the process of analyzing clustering results in conventional methods usually focuses on the quality of clustering results such as the entropy or F-measure methods [6]–[8]. However, clustering results do not automatically inform a university's academic performance. Meanwhile, a quality assurance system is needed that is able to provide information on academic performance quickly and accurately. Therefore, advanced techniques are needed that are able to summarize the results of academic clustering automatically, quickly and accurately.

This article proposes a new technique, namely self organizing map and similarity to an ideal solution (SOM-SIS) which can summarize the results of clustering through the SOM technique and the technique for order preference by similarity to an ideal solution (TOPSIS) automatically. TOPSIS was developed by Yoon and Hwang [9]. This technique uses the basic concept that the ideal solution chosen must have the shortest distance from the positive ideal solution, and the farthest from the negative ideal solution.

SOM-SIS is applied to three academic parameters that represent university output, namely student achievement [10]–[12], study period [13], [14], and drop out rate [15], [16]. Student learning achievement is measured by the grade point average (GPA) [17], [18], the study period is the length of student learning [19]–[21], while the drop out rate is the student's failure rate [22]. The basis of the 3 academic-parameters refers to the higher education accreditation instrument (HEAI) in Indonesia [23].

The auto-summarizing SOM-SIS method was tested using a dataset of 300 taken from universities in Indonesia. As a result, the position of universities compared to others can be known automatically and accurately from the SOM results. These results were validated by several quality assurance experts in universities with 100% accuracy.

2. RESEARCH METHODS

The development stage of the SOM-SIS auto-summarizing academic quality assurance system is described in Figure 1. The research stages are divided into two important parts, namely data filtering and mining, as well as clustering and auto-summarizing. The filtering and mining stage begins with cleaning, integrating, selecting, transforming into the desired form [24], and mining data using the SOM algorithm. The next stage consists of processing the SOM results using the TOPSIS algorithm to generate auto-summarizing SOM-SIS.

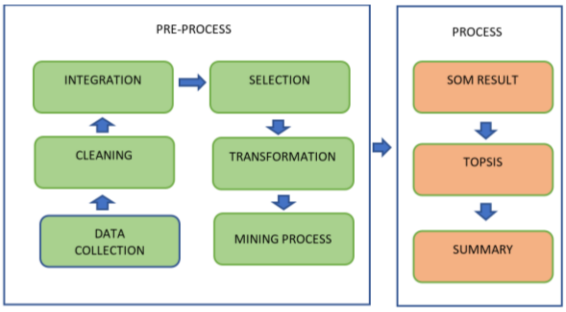


Figure 1. Research block diagram

Data cleaning is the process of removing noise and inconsistent or irrelevant data. Data cleaning is carried out if the data obtained from the university database contains imperfect entries such as missing, invalid, or typographical data. The irrelevant data is then discarded or replaced with the appropriate value. Data preparation is followed by data integration, which is combining data from various databases into one new database. Data from various attributes such as high school, college entrance system, national exam scores, achievement index for semester 1 to semester 4, index cumulative grade for semester 4, final cumulative index, residence, parental salary, study period, priority study program, which consists of several files and then put together in a single file. The data was selected according to the needs of the analyzed parameters, namely student achievement, study period, and drop out rates. Furthermore, the data is transformed through a conversion process, namely changing one data format to another data format so that it can be read by certain systems for the mining process. The conversion process is necessary because the academic data required has different units and types of data, so it needs to be converted into an equivalent numerical form. After conversion, data normalization is then performed using the min-max method, where for each input data the minimum ($Xmin$) and maximum ($Xmax$) values are sought, then the normalization process is carried out so that normal data is obtained [25] using (1).

$$Newdata = (data - Min) \times \frac{(NewMax - NewMin)}{(Max - Min)} + NewMin \quad (1)$$

Newdata is the normalized data, min is the minimum value of the data, max is the maximum value of the data, newmin and newmax are the minimum and maximum limits.

The clustering process using SOM is carried out after all the required data has been normalized. Clustering is done using miniSOM which is a python library that focuses on scientific computing. The clustering process begins with the formation of a SOM network map based on the input data on the system created, then a learning process is carried out with several iterations to produce an ideal weight matrix. Furthermore, the ideal weight matrix is used to map the input data into groups of output data. The learning process is based on the distance between the input data and the weight matrix. After the initialization process, then proceed with the training process. The unsupervised learning algorithm on the kohonen SOM network [26] is (2).

$$D_j = \sum_{i=1}^{n}(W_{ij} - X_i)^2 \tag{2}$$

Where D_j is the euclidean distance, W_{ij} is the weight of the i-th neuron, X_i is the i-th input vector. After getting the winning neurons, then updating the weight values of the winning neurons and neighboring neurons is (3).

$$W_{ij}\,(t + 1) = W_{ij}\,(t) + \alpha\,(t)\,[\,X_i - W_{ij}(t)] \tag{3}$$

Where W_{ij} is the weight for the j-th output neuron and the i-th input neuron, $\alpha\,(t)$ is the learning rate, and the neighbor function. The stages of the Kohonen SOM algorithm are in [26]:

- Initialize weight W_{ij} with random value, learning rate and neighbor function.
- Select input X_i randomly from the input set.
- Calculate the degree of similarity using the eucledian distance D_j (2) for all neurons (j).
- Select the winning neuron, that is, the neuron with the minimum euclidean distance.
- Improved the weight of the winning neuron in the W_{ij} (3) score and the weight of the neighboring neurons.
- Update the learning rate and reduce the neighbor function linearly or exponentially.
- Perform steps 2 to 5 until the epoch value (maximum iteration value) is reached.

Davies-bouldin index (DBI) metric introduced by Davies and Bouldin in 1979 [27]. DBI is used to evaluate clusters through the process of calculating sum of square within clusters sum of squares within (SSW) as a cohesion metric with i-clusters. The clustering evaluation process using SSW is (4).

$$\text{SSW} = \frac{1}{mi}\sum_{n=1}^{mi} d\left(x_j, c_i\right) \tag{4}$$

Where m_i is the number of input data that is in the i-th cluster, while c_i is the i-th centroid cluster. The sum of square between clusters (SSB) formula is used by measuring the distance between the centroids (weight metrics) for example clusters i (c_i), and clusters j (c_j) as in (5).

$$SSB_{i,j=d\,(\,C_i,C_j)} \tag{5}$$

Furthermore, R_{ij} is the comparison value between cluster i and cluster j. The value is obtained from the components of cohesion and separation. A good cluster must have the smallest cohesion value and the largest separation value as in the (6).

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \tag{6}$$

The DBI value is obtained from the (7).

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j} \; R_{ij} \tag{7}$$

The stages in the TOPSIS method are contained in [9]. The normalized decision matrix is determined as in the (8).

$$R_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^{2}}} \tag{8}$$

Determine the weighted normalization decision matrix, with the criteria weights in Table 1.

Table 1. Criteria weights

No	Criteria	Weight
1	Student performance	5
2	Study period	4
3	Drop out	3

Calculates a weighted normalization matrix as in the (9).

$$Y_{ij} = W_i R_{ij} \tag{9}$$

Determine the positive ideal solution matrix and the negative ideal solution matrix as in the following (10), (11) The positive ideal solution (A^{+}) is determined by:

$$A^{+} = (Y^{1+}, Y^{2\,+}, Y^{3\,+}, \ldots. Y^{n+}) \tag{10}$$

The negative ideal solution (A^{-}) is determined by:

$$A^{-} = (Y^{1^{-}}, Y^{2\,-}, Y^{3\,-}, \ldots. Y^{n-})$$
$$Y^{i+} \begin{cases} \max Yij : if\ j\ \ is\ an\ attribute\ of\ profit \\ \min Yij : if\ j\ \ \ is\ a\ cost\ attribute \end{cases}$$
$$Y^{i^{-}} \begin{cases} \max Yij : if\ j\ \ is\ a\ cost\ attribute \\ \min Yij : if\ j\ \ is\ an\ attribute\ of\ profit \end{cases} \tag{11}$$

The distance between alternative A, and the positive ideal solution is defined as (12).

$$D_i^{+} = \sqrt{\sum_{j=1}^{n} \left(Y_{ij} - Y_i^{+}\right)^2}, i = 1,2,3 \ldots \ldots m \tag{12}$$

The distance between alternative A, and the negative ideal solution is defined as (13).

$$D_i{}^{+} = \sqrt{\sum_{j=1}^{n} \left(Y_{ij} - Y_i{}^{-}\right)^2}, i = 1,2,3 \ldots \ldots m \tag{13}$$

Decision matrix D is used to find the preference value for each given alternative, refers to m alternatives that are evaluated based on the specified criteria, shows the computational performance for the i-th alternative and the j attribute. The closeness of each alternative to the ideal solution is calculated according to the (14).

$$V = \frac{D_i^{-}}{D_i^{-} + D_i^{+}}, i = 1,2,3 \ldots \ldots m \tag{14}$$

The SOM-SIS method starts after the results of clustering using SOM are known, then auto-summarizing about the academic-performance of universities using TOPSIS is made. SOM-SIS is useful for determining the level of university academic performance based on SOM results using the TOPSIS decision support system. Table 2 is the cluster value of each parameter, and Table 3 is the dominant cluster combination from the SOM results.

Table 2. Criterion value

Criteria	Description	Value
Student performance	Poor	1
	Fair	2
	Good	3
Study period	On time	1
	Not on time	2
Drop out	No potential	1
	Potential	2

The results of the cluster of three academic parameters using the SOM produce a combination, the TOPSIS method is known as an alternative. The combination of the values of the three academic parameters (alternatives) produces 12 cluster channels. The SOM-SIS base rules are Figure 2.

Table 3. Combination of dominant cluster values

12 Chanel cluster	Cluster Value		
	Student performance	Study period	Drop out
A	1	1	1
B	1	2	2
C	1	1	2
D	1	2	1
E	2	1	1
F	2	2	2
G	2	1	2
H	2	2	1
I	3	1	1
J	3	2	2
K	3	1	2
L	3	2	1

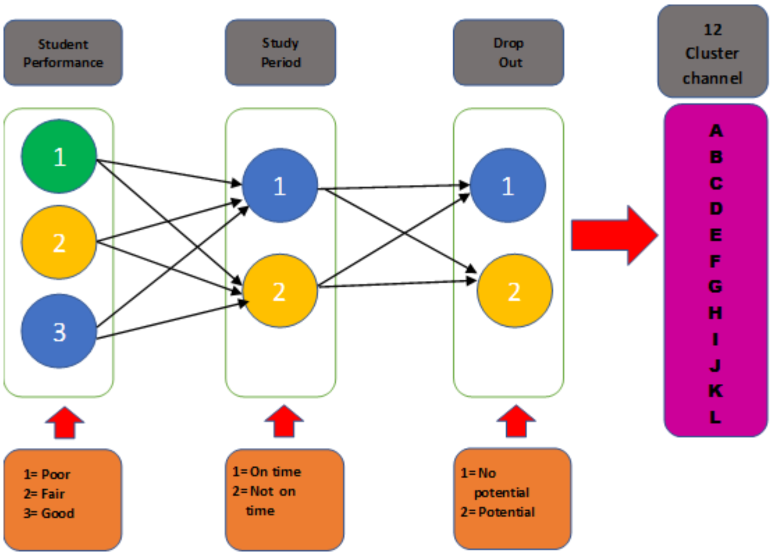


Figure 2. Rules base SOM-SIS

Table 4 is a ranking of preference values which is the final result of the SOM-SIS method. Based on the combination of the dominant cluster values in Table 3, universities can find out the cluster ranking and academic performance by referring to Table 5. Table 5 is the result of automatic conclution, where universities can find out their level on the cluster channel, cluster rankings and college academic performance.

Table 4. Preference value range

Range	Cluster rank	College category
1-4	1	Good
5-8	2	Fair
9-12	3	Poor

Table 5. Autosummarizing SOM-SIS

Preference	Condition of college	Value	Ranking	Cluster rank	College academik performance
V1	A	0.40	8	2	Fair
V2	B	0.00	12	3	Poor
V3	C	0.33	10	3	Poor
V4	D	0.26	11	3	Poor
V5	E	0.62	4	1	Good
V6	F	0.37	9	3	Poor
V7	G	0.53	6	2	Fair
V8	H	0.47	7	2	Fair
V9	I	1.00	1	1	Good
V10	J	0.59	5	2	Fair
V11	K	0.74	2	1	Good
V12	L	0.67	3	1	Good

The validation of the SOM-SIS system was carried out to determine whether the system developed was in accordance with the academic quality assurance system of higher education institutions in Indonesia. Validation was carried out by three quality assurance experts from universities. This validation is done by comparing the SOM-SIS system with HEAI on the assessment matrix [26] using manual calculations and TOPSIS.

3. RESULTS AND ANALYSIS

A total of 300 datasets were collected from a survey of universities in Indonesia for the 2011-2013 academic year. The parameters used for academic quality assurance consist of student performance, study period, and drop out rates. The measurement of student performance uses attributes: high school, university entrance system, school exam scores, parents' salaries, index of cumulative grades for semester 4 and index of final cumulative grades. Measurement of study period using attributes: college entrance system, residence, parental salary, final cumulative achievement index. The measurement of the drop out rate uses the criteria for students who have not graduated up to 8 semesters for undergraduate studies, with attributes: high school, priority study programs, college admissions system, parents' salaries, achievement index for semesters 1 to 4, number of semester credit units.

3.1. Clustering using the self organizing map

The results of clustering using SOM are presented in Table 6. The cluster results of the three academic parameters show that the "fair" cluster members are the most dominant in the student performance parameters, the "not on time" cluster is more dominant in the study period parameter, while the "not potential" cluster is more dominant in dropout rates. Based on the dominant cluster value, the criterion value of the three academic parameters is (2, 2, 1).

Figure 3(a) is a distribution map of clustering results on student performance parameters which are visualized in colored dots. Blue color indicates student performance is in the "good" cluster, orange indicates "fair", and green "poor". In this parameter, the dominant cluster is "orange". Figure 3(b) is a distribution map of clustering results for the study period parameter, the blue color indicates "on time", while orange "is not on time". In this parameter the dominant cluster is "orange". Figure 3(c) is a distribution map of clustering results for the dropout parameter, the blue color indicates "not potential", while the orange color "potential", in this parameter the dominant cluster is "blue".

Table 6. SOM result

Parameter	Cluster	Number of members	Dominant cluster	Criterion value
Student performance	Good	22	Fair	2
	Fair	204		
	Poor	74		
Study period	On time	94	Not on time	2
	Not on time	206		
Dropout	No potential	167	No potential	1
	Potential	133		

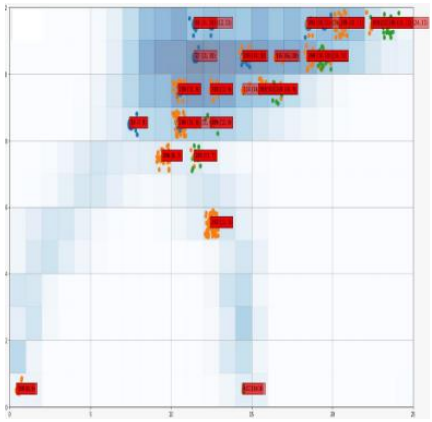
3.2. Clustering evaluation

Clustering evaluation is used to find out how precisely a data is grouped. Clustering evaluation in this study uses the validity test of the DBI. Table 7 shows the average DBI results are quite good with a value of 1.11. The DBI value for study period parameter and the drop out rate is 1.00, which is better than the student performance parameter of 1.34.

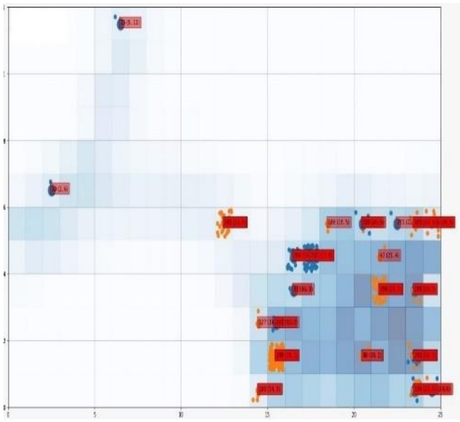
After clustering is done using SOM, the next step is to integrate the clustering results into TOPSIS to determine the preference value of the university. The preference value obtained from the integration of the two methods produces the Autosummarizing SOM-SIS algorithm to determine the academic quality of the university automatically. Based on the results of clustering using three academic parameters, the dominant cluster is worth (2, 2, 1) contained in the channel cluster "H", so the results of auto-summarizing SOM-SIS indicate the level of university academic performance is in "rank 2" with "fair" criteria as Table 8.

Tabel 7. Clustering evaluation

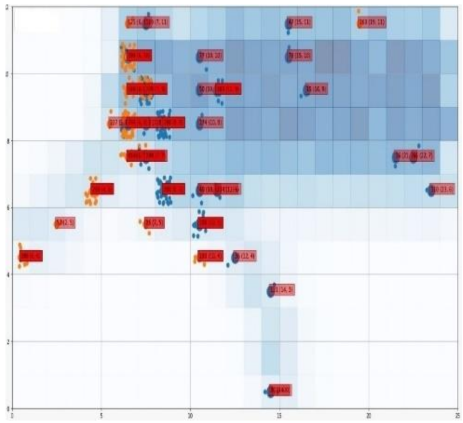
No	Parameters	QE	DBI
1	Student performance	< 2	1.34
2	Study period	< 1	1.00
3	Drop out	< 2	1.00
	Mean		1.11



(a)



(b)



(c)

Figure 3. SOM distribution map parameters (a) student performance, (b) study period, and (c) drop out rate

Table 8. Result of dominant cluster value combination

No	Criteria	Discription	Matrix value	12 chanel cluster	Cluster rank	College academik performance
1	Student performance	Fair	2	H	2	Fair
2	Study period	Not on time	2			
3	Drop out	No potential	1			

3.3. Similarity to an ideal solution validation

SOM-SIS was validated by 3 higher education quality assurance experts using the HEAI score matrix through manual calculations and TOPSIS. The results of the validation carried out by higher education quality assurance experts showed a conformity level of 100%. The conclusion is that the SOM-SIS system is able to provide accurate conclusions in terms of cluster rankings and college academic performance, as Table 9.

Based on the results of clustering academic data of students at a university in Indonesia using three academic parameters, the SOM-SIS system can summarize them well, so it can be seen that the academic performance of the university is in “rank 2” with the criteria of “fair”. The SOM-SIS system can help quality assurance in universities to summarize the results of academic clustering and conclude the academic performance of universities compared to others. Knowledge of SOM-SIS results will assist university management in making academic decisions to improve its performance.

Table 9. SOM-SIS validation with HEAI

Comparison				
Method	SOM-SIS result	Matrix HEAI (manual score)	Matrix HEAI (TOPSIS)	Conformity level
Academic performance	Fair	Fair	Fair	100%

4. CONCLUSION

In this study, SOM-SIS was able to automatically summarize the results of the clustering of 300 datasets of universities in Indonesia. The results of clustering using three academic parameters can be well summarized by the SOM-SIS system, so that the academic performance of the college can be known. The SOM-SIS system can help higher education quality assurance to summarize the results of academic clustering and conclude the academic performance of universities compared to others. Knowledge of SOM-SIS results will assist university management in making academic decisions to improve its performance. The results of the validation carried out by three higher education quality assurance experts showed a 100% conformity level. The conclusion is that the SOM-SIS system is able to summarize the results of clustering and determine cluster rankings and college academic performance accurately.

In future research, it is necessary to use larger data to support a better analysis of academic quality assurance. The use of data transformation methods such as one hot encoding or integer encoding can be used to obtain precise cluster results. Several other academic parameters such as graduate competence, acceptance in work, need to be added to determine the academic performance of higher education institutions for the better.

AKNOWLEDGMENTS

Rahayu Widayanti's work is fully supported by scholarships from the Gunadarma Education Foundation and the Pradnya Paramita Foundation. This work was supported by Prof. Dr. E. S. Margianti, SE, MM, as the Chancellor of Gunadarma University, and Dr. TB Moh. Akhriza, S. Si, MMSI, as the Chairman of Pradnya Paramita College of Informatics Management and Computer.

REFERENCES

[1] Y. Sanghvi, H. Gupta, H. Doshi, D. Koli, A. Ansh, and U. Gupta, "Comparison of self organizing maps and sammon's mapping on agricultural datasets for precision agriculture," in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2015, pp. 1–5, doi: 10.1109/ICIIECS.2015.7193206.
[2] A. Triayudi and I. Fitri, "Comparison of the feature selection algorithm in educational data mining," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 6, pp. 1865-1871, 2021, doi: 10.12928/telkomnika.v19i6.21594.
[3] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, no. 1, pp. 135–146, 2007, doi: 10.1016/j.eswa.2006.04.005.
[4] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, "Self organizing maps in matlab: the SOM toolbox," in *Proceeding of the matlab DSP Conference*, 1999, pp. 35–40. [Online]. Available: http://lib.tkk.fi/Diss/2002/isbn951226093X/article2.pdf
[5] P. Hartono and Y. Take, "Pairwise elastic self-organizing maps," in *2017 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM)*, 2017, pp. 1–7, doi: 10.1109/WSOM.2017.8020006.
[6] D. Barbará, Y. Li, and J. Couto, "COOLCAT: an entropy-based algorithm for categorical clustering," in *Proceedings of the eleventh international conference on Information and knowledge management - CIKM '02*, 2002, pp. 582–589, doi: 10.1145/584792.584888.
[7] J. Li and O. Zaïane, "Combining usage, content,and structure data to improve web site recommendation," *International Conference on Electronic Commerce and Web Technologies; EC-Web 2004: E-Commerce and Web Technologies*, 2004, vol. 3182, pp. 305–315 doi: 10.1007/978-3-540-30077-9_31.
[8] T. Nizam and S. I. Hassan, "Exemplifying the effects of distance metrics on clustering techniques: f-measure, accuracy and efficiency," *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2020, pp. 39-44, doi: 10.23919/INDIACom49435.2020.9083687.
[9] V. Balioti, C. Tzimopoulos, and C. Evangelides, "Multi-criteria decision making using TOPSIS method under fuzzy environment. application in spillway selection," in *Proceedings EWaS3 2018*, 2018, vol. 2, no. 11, p. 637, doi: 10.3390/proceedings2110637.
[10] K. S. Priya and A. V. S. Kumar, "Improving the student's performance using educational data mining," *International Journal of Advanced Networking and Applications*, vol. 4, no. 4, pp. 1680–1685, 2013.
[11] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students performance," *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 6, 2011, doi: 10.14569/IJACSA.2011.020609.
[12] E. N. Ogor, "Student academic performance monitoring and evaluation using data mining techniques," in *Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007)*, 2007, pp. 354–359, doi: 10.1109/CERMA.2007.4367712.
[13] A. Supriyanto, D. Maryono, and F. Liantoni, "Predicted student study period with C4.5 data mining algorithm," *IJIE (Indonesian Journal of Informatics Education)*, vol. 4, no. 2, 2020, doi: 10.20961/ijie.v4i2.46265.
[14] A. Wibowo, D. Manongga, and H. D. Purnomo, "The utilization of naive bayes and C.45 in predicting the timeliness of students' graduation," *Scientific Journal of Informatics*, vol. 7, no. 1, pp. 99–112, 2020, doi: 10.15294/sji.v7i1.24241.
[15] F. Araque, C. Roldán, and A. Salguero, "Factors influencing university drop out rates," *Computers & Education*, vol. 53, no. 3, pp. 563–574, 2009, doi: 10.1016/j.compedu.2009.03.013.
[16] J. -P. Vandamme, N. Meskens, and J. -F. Superby, "Predicting academic performance by data mining methods," *Education Economics*, vol. 15, no. 4, pp. 405–419, 2007, doi: 10.1080/09645290701409939.
[17] T. T. York, C. Gibson, and S. Rankin "Defining and Measuring Academic Success," *Practical Assessment, Research, and Evaluation*, vol. 20, no. 5, pp 1-20, 2015, doi: 10.7275/hz5x-tx03.
[18] S. B. Kotsiantis and P. E. Pintelas, "Predicting students marks in hellenic open university," in *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, 2005, pp. 664–668, doi: 10.1109/ICALT.2005.223.
[19] D. E. Ukpong and I. N. George, "Length of Study-Time Behaviour and Academic Achievement of Social Studies Education Students in the University of Uyo," *International Education Studies*, vol. 6, no. 3, pp. 172-178, 2013, doi: 10.5539/ies.v6n3p172.
[20] D. Raju, "Predicting student graduation in higher education using data mining models," Ph.D. dissertation, Department of Educational Studies, Psychology, Research Methodology, and Counseling, University of Alabama, Tuscaloosa, Alabama, 2012. [Online]. Available: https://ir.ua.edu/handle/123456789/1395
[21] A. Yaqin, A. D. Laksito, and S. Fatonah, "Evaluation of backpropagation neural network models for early prediction of student's

graduation in XYZ university," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 11, no. 2, pp. 610-617, 2021, doi: 10.18517/ijaseit.11.2.11152.

[22] G. Dekker, M. Pechenizkiy, and J. Vleeshouwers, "Predicting students drop out: A case study," in *International Conference on Educational Data Mining*, 2009, pp. 41–50. [Online]. Available: https://files.eric.ed.gov/fulltext/ED539082.pdf

[23] BAN PT. Regulation number 3 of 2013. (2019). *Concerning higher education accreditation instruments* (in Bahasa). [Online]. Available: https://www.banpt.or.id/wp-content/uploads/2019/10/Lampiran-5-PerBAN-PT-5-2019-tentang-IAPS-Pedoman-Penilaian.pdf

[24] J. Han, M. Kamber and J. Pei, *Data mining-concepts and techniques*. Elsevier Publishers, Morgan kaufmann, 2012. [Online]. Available: http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf

[25] Henderi, T. Wahyuningsih, and E. Rahwanto "Comparison of min-max normalization and z-score normalization in the k-nearest neighbor (kNN) algorithm to test the accuracy of types of breast cancer," *IJIIS: International Journal of Informatics and Information Systems*, vol. 4, no. 1, pp. 13–20, 2021, doi: 10.47738/ijiis.v4i1.73.

[26] T. Kohonen, "Self-organizing neural projections," *Neural Networks*, vol. 19, no. 6–7, pp. 723–733, 2006, doi: 10.1016/j.neunet.2006.05.001.

[27] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.

BIOGRAPHIES OF AUTHORS

Rahayu Widayanti is a lecturer and researcher in the Department of Information Systems, Pradnya Paramita College of Informatics Management and Computer, Indonesia. She was born in Malang, Indonesia August 26, 1968. Completed: Bachelor of Economics and Development Studies from Brawijaya University, Indonesia in 1992, master's degree in Management from Brawijaya University, Indonesia in 2010, master's degree in Information Systems from Gunadarma University, Indonesia in 2015, and Doctoral degree in Information Technology from Gunadarma University, Indonesia in 2021. His current research interests include Data Mining, Decision Support Systems, Forecasting Systems, Management Information Systems. She can be contacted at email: rahayu@stimata.ac.id.

Sarifuddin Madenda is a researcher and professor at Gunadarma University, Indonesia. He was born in Raha, Indonesia 7th April 1963. He finished his: Bachelor's degree in Physic Instrumentation at University of Indonesia in 1989, master's degree in Electronics from INSA-Lyon, France in 1992, and Doctoral degree in Electronics and Image Processing from Universite de Bourgogne, France in 1995. He is also lecturer at UQO, Quebec - Canada since 2003. He is Director of Doctoral Program in Information Technology. His research interests are focused on Image and Video processing, Multimedia data compression, Content based image and video retrieval, Steganography (Encryption, Decryption, Coding) and decoding of multimedia secret documents, Real time system architecture (FPGA and ASIC design). He can be contacted at email: sarif@staff.gunadarma.ac.id.

Eri Prasetyo Wibowo (b. 1966), is a Professor at Gunadarma University, Indonesia. He was born in Kendal, Indonesia in 1966. He received the B.S degree in Electronics and Instrumentation at Gadjah Mada University, Indonesia, in 1991, the M.S. degree in Information System at Gunadarma University, Indonesia in 1994 and the Ph.D. degree in Electronics Informatics from Universite de Bourgogne, France in 2005. He is member of IEEE and EACOVIROE Project to promote Master VIBOT which founded by Erasmus Mundus. His research interests are focused on sensor CMOS design for face tracking and recognition, designed and developed ADC for video application, VLSI design and real time image processing. He can be contacted at email: eri@staff.gunadarma.ac.id.

Khoerul Anwar graduated with a Bachelor of Engineering from the National Institute of Technology, graduated with a Master's degree in Telecommunication and Informatics from Brawijaya University and a Doctorate from Gadjah Mada University in the field of computer science. The research focus is on Digital Image Processing, Computer vision and Artificial intelligent. Has produced four international journals, one indexed journal Q1 Scopus, one journal at Thomson and two other journals indexed on the database of reputable journals. Currently active as a researcher and lecturer at Pradnya Paramita College of Informatics Management and Computer, Indonesia. He can be contacted at email: alqhoir@stimata.ac.id.