

Implementation of K-Medoid and Hierarchical Clustering for Strengthening Job Satisfaction on The Teacher Profession

1st Eneng Tita Tosida
Computer Science Dept.
Pakuan Univeristy
Bogor, Indonesia
enengtitatosida@unpak.ac.id

2nd Irma Anggraeni
Computer Science Dept.
Pakuan University
Bogor, Indonesia
irmairhamna@unpak.ac.id

3rd Suci Putri Utami
Computer Science Dept.
Pakuan University
Bogor, Indonesia
usuciputri@gmail.com

4th Indra Permana Solihin
Informatic Dept.
UPN Veteran Jakarta
Jakarta, Indonesia
indrapermana@upnvj.ac.id

Abstract—Education is a very important element in social life to support the formation of individual character. The implementation of data mining in the education sector shows an increasing trend. The purpose of this study is to implement strengthening job satisfaction for the teaching profession using the K-medoid and Hierarchical Clustering method. Cluster analysis was carried out on five main factors including salary, promotion, supervisor's supervision, relationships with colleagues and the teaching profession. Tests were carried out on 136 respondents and 39 questions divided into 5 factors. The trial of determining the number of clusters was carried out on 3 trials (K=2, K=3 and K=4). The optimal cluster value is obtained at K=3 with a Davies Bouldin index (DBI) of 2,823 for the K-Medoid method and K=4 with a DBI value of 1,415 for the Hierarchical method. The cluster results show that the priority of strengthening teacher professional satisfaction needs to be increased on the salary indicator by 50% for the K-Medoid method and 65% for the Hierarchical method. While the results of the percentage of priorities that need to be maintained are the indicators of relationships with colleagues by 10% in the K-Medoid method and 5% in the Hierarchical method.

Keywords—Clustering, Data mining, Teacher profession, K-Medoid, Hierarchical.

I. INTRODUCTION

Education in Indonesia is very important for the future progress of the nation. Education is all efforts to foster personality and develop Indonesian human abilities, physically and spiritually that lasts a lifetime, both inside and outside school in the context of developing Indonesian unity [1] and a just and prosperous society based on Pancasila. Formal basic education in Indonesia consists of three levels, namely Elementary School (SD), Junior High School (SMP) and Senior High School (SMA). Junior High School (SMP) is a formal education at the basic education level that has human resources including students and teachers. Teachers are professional educators with the main task of educating, teaching, guiding, directing, training, assessing, and evaluating students in early childhood education through formal education, basic education, and secondary education [2]. Job satisfaction is one of significant factor that can influencing teacher performance. Job satisfaction is basically an individual thing, each individual will have a different level of satisfaction with the value system that applies to them [3]. Teacher job satisfaction is one of the important things to pay attention to and improve. One of the efforts that can be done

to increase teacher job satisfaction in schools is to improve the quality of work life. Teachers' job satisfaction is of utmost important that can impact on students' academic achievement and their future career [4]. Other efforts can be made through strengthening job satisfaction, transformational leadership and self-efficacy to increase commitment to the profession [5].

Based on data from the West Java Publisher article, in 2019 the wages, especially for the State Junior High School in Cirebon Regency, are honorary teachers, which is 300 thousand per month. The salary is not sufficient to meet daily needs. So many honorary teachers choose to add work elsewhere. This causes honorary teachers not to focus on carrying out their duties. In several previous studies, many have analyzed and grouped teacher performance using data mining and clustering [6][7][8][9]. However, there has not been an analysis of the criteria that can affect teacher performance satisfaction.

The object of data was from honorary teachers who teach at Cirebon Junior High Schools. In this study, the data used came from secondary data [5] through the distribution of questionnaires which were then processed using the Knowledge Discovery and Data Mining (KDD) method or better known as data mining. Data mining is the process of discovering new meaningful correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern reasoning technology as well as statistical and mathematical techniques [10]. Data mining is used for processing grouping data in very large quantities [11].

The purpose of this study is to implement the clustering technique for the analysis of strengthening job satisfaction for the teaching profession. Clustering as a data mining technique is a process of grouping an object into a class of objects that have similarities. Several clustering methods that can be used are K-Medoid and Hierarchical. K_Medoids algorithm is an algorithm similar to K Means [12], but in text processing K Medoids are superior [13] as well as in statistical grouping [14]. The K-Medoid algorithm does not use the centroid point but uses the medoid position [15]. In the K-Medoid technique, this is if the new total distance (TD) < initial TD, swap the position of the medoid with the new medoid and become the new medoid and repeat the previous steps until the medoid does not change [12]. Hierarchical clustering is one of the clustering algorithms with the characteristics of each data must be included in a particular cluster [16]. Thus the

contribution of this research is to produce a model of strengthening job satisfaction for the teaching profession based on data mining clustering techniques.

II. RESEARCH METHODS

This study uses the Data Mining method or also known as Knowledge Discovery and Data Mining (KDD). KDD as a series of processes, data mining can be divided into several stages as shown in Figure 1 and the research stage are shown in figure 2. The initial step of the research is to conduct secondary data analysis [10].

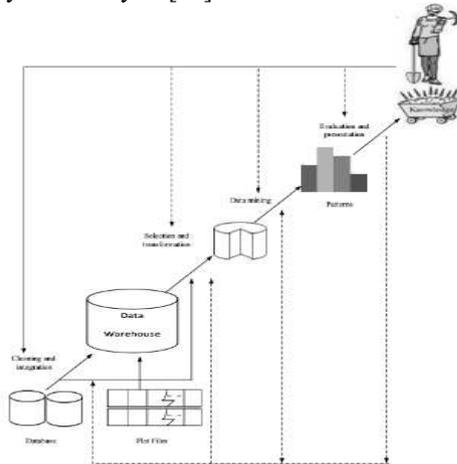


Fig. 1. Data mining stages [10]

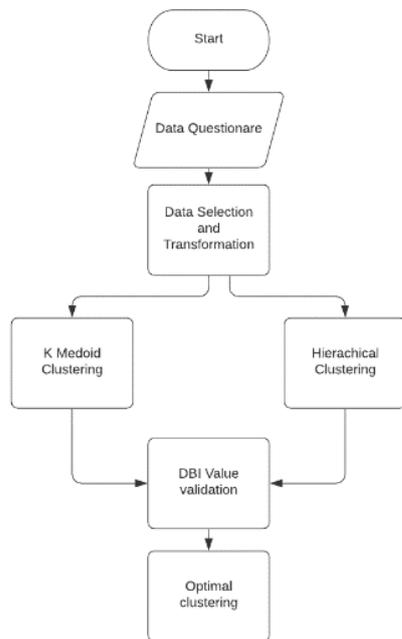


Fig. 2. Research Stages

A. Data Cleaning

There are 136 respondents who are honorary teachers at state junior high schools in Cirebon Regency who filled out a questionnaire about teacher satisfaction. There are 39 question regarding the assessment of factors that affect the strengthening of teacher job satisfaction. In each question, respondents can answer in 4 conditions, there are strongly agree (SA), agree (A), neither agree (NA) or disagree (DA),

strongly disagree (SDA that shown in TABLE I. In this study, the cleaning process was not carried out because the data obtained from the questionnaire had no missing values and the data was also consistent.

TABLE I. FIVE SAMPLES OF 39 QUESTIONS

No	Questions	SA	A	NA	DA	SDA
1	I think the amount of salary I receive is adequate					
2	I feel the honorarium received is too small and inadequate					
3	I feel that the amount of salary I receive can fulfill the needs of my family					
4	I feel that the amount of overtime paid is appropriate					
5	I assess the salary increase received according to the applicable rules					

Respondents fill out 39 questions in the questionnaire and the sample of data as shown in Fig 3.

Out[12]:

responden	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	...	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Jumlah	
0	1	3	3	2	3	3	4	4	4	...	4	4	4	3	4	2	2	2	2	4	139
1	2	4	3	4	4	4	4	4	4	...	4	4	4	5	4	2	2	2	2	2	141
2	3	4	3	4	4	4	4	4	4	...	4	4	4	5	4	2	2	2	2	2	141
3	4	2	3	2	3	3	4	3	3	...	4	3	4	4	4	3	2	2	3	3	130
4	5	2	2	2	3	4	4	4	4	...	4	4	4	5	4	3	1	3	3	3	138
...
131	132	2	2	2	2	2	3	3	3	...	4	3	4	4	4	3	3	3	3	3	124
132	133	4	4	3	4	2	4	4	4	...	4	4	4	4	4	4	4	4	4	4	147
133	134	1	5	1	1	3	4	4	4	...	4	4	5	4	4	4	1	3	2	3	132
134	135	2	2	2	3	3	4	4	4	...	4	4	4	4	4	3	3	3	3	3	140
135	136	4	3	2	3	3	4	2	4	...	4	3	3	3	3	3	3	3	4	3	131

136 rows x 41 columns

Fig. 3. Result Questionnaire Data

B. Data Transformation

The transformation process by changing the data that was previously still in the form of a questionnaire then converted into a .csv format.

C. Data Mining

Data from the questionnaire was processed using the K-Medoid and Hierarchical methods. K-Medoids is a partition clustering method to cluster a set of n objects into a number of k clusters. This algorithm uses objects in a collection of objects that represent a cluster. Objects that represent a cluster are called medoids. Clusters are built by calculating the closeness between medoids and non-medoids objects [10].

Agglomerative hierarchical clustering method is a cluster analysis method that aims to group objects based on their characteristics, starting with individual objects until these objects are combined into a single cluster [10]. Hierarchical clustering is a grouping method that can be represented in the form of a tree or hierarchy called a dendrogram. K-Medoid and Hierarchical algorithms are used to find medoids in a cluster.

In the K Medoid process, determining the desired number of clusters randomly selects the data to be used as the initial medoid as much as the number of clusters specified in this study, the sample k values used are 1, 3 and 5 as iteration 1. Next, calculate the distance of the object to each medoid that has been selected. We use Euclidean Distance. The next step

is to calculate the total distance by determining the iteration on the medoid which will stop if the cluster members formed from the medoid do not change again, we choose 1, 2 and 4 as the new medoid. The next step is to calculate the total deviation which serves to determine the cluster members formed in each medoid.

The first hierarchical algorithm is carried out calculating the distance matrix using the Euclidean distance formula, find the minimum distance or single link method. summarizing the cluster data, then grouping it gradually (hierarchically).

Clustering is the process of grouping data points into two or more groups so that data points belonging to the same group are more similar to each other than in different groups, only based on the information available with the data points. The clustering process uses the Python programming language and Jupyter Notebook as a tool for data processing and visualization of cluster results in the form of a website. Test validation for determining the optimal number of clusters can be done by using the DBI method.

The main purpose of this study is to implement clustering to strengthen job satisfaction for the teaching profession. The results of cluster analysis are visualized in the form of diagrams and graphs. The data used came from 136 respondents and 39 variables using 5 categories. These categories include salaries, promotions, supervisory supervisors, relations with colleagues and the work of the teaching profession.

III. RESULTS AND DISCUSSION

A. Clustering Results with K-Medoid and Hierarchical Method

The results of clustering using the K-Medoid method on secondary data which are divided into 5 categories, where the most dominant cluster is in cluster 4 in the form of data visualization using RapidMiner, can be seen in Figure 4.

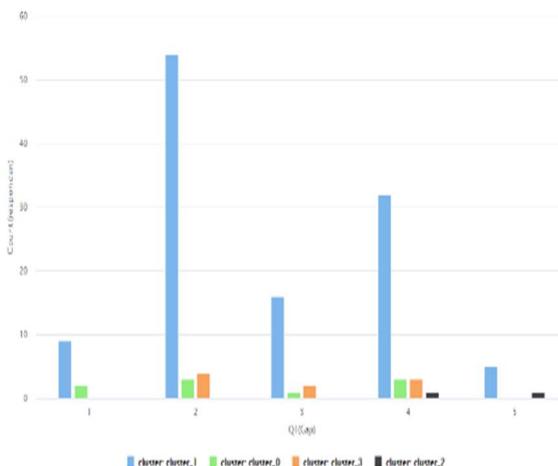


Fig. 4. Cluster 1 of K-Medoid clustering (Salary)

The results of the K-Medoid clustering in Figure 4 show that Cluster 1 is dominated by respondents who are dissatisfied with the salaries of the teaching profession, especially honorary teachers. Cluster 0 is dominated by respondents who are hesitant about promotion and supervisor supervision related to job satisfaction of the teaching profession. Cluster 2 is dominated by respondents who agree on the strength of co-worker relationships that can strengthen the job satisfaction of the teaching profession. Cluster 3 is

dominated by respondents who agree that the element of work will strengthen job satisfaction for the teaching profession.

Hierarchical clustering results are divided into 5 categories. The most dominant cluster result is Cluster 4, as can be seen in Figure 5.

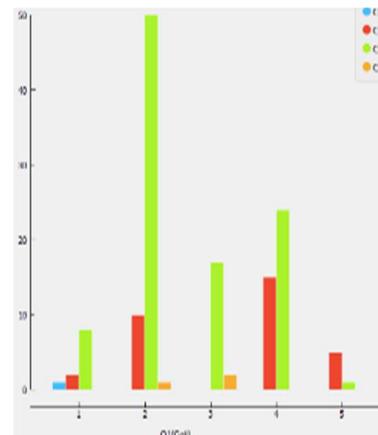


Fig. 5. Cluster 1 of Hierarchical clustering (Salary)

Based on Figure 5. cluster 1 is dominated by respondents who strongly disagree on the salary and promotion can strengthen the satisfaction of the teaching profession. Cluster 2 is dominated by respondents who strongly agree that relationships with co-workers will strengthen the job satisfaction of the teaching profession. Cluster 3 is dominated by respondents who strongly agree that supervisory supervision can strengthen the satisfaction of the teaching profession. Cluster 4 is dominated by respondents who doubt that the element of work can strengthen job satisfaction.

The distribution of data from the generalization of K-Medoid clustering shows that the job satisfaction of the teaching profession is influenced by a salary factor of 50%, promotion by 15%, and the other side is the supervisory factor of superiors, relationships with colleagues and elements of the work itself by 35%. The results of the generalization of the K-Medoid cluster seen as Figure 6.

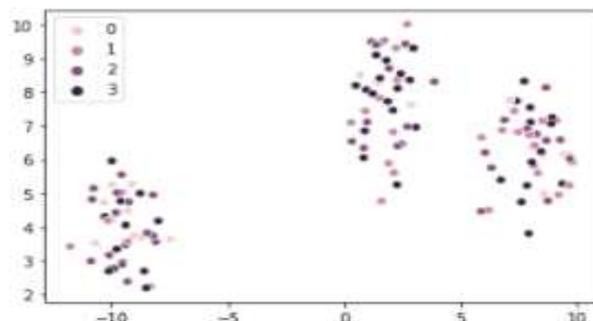


Fig. 6. Scatter plot of the generalization results of the K-Medoid

Based on the results of the generalization of the K-Medoid clustering, it can be recommended to stakeholders to review the salaries and promotions of the teaching profession, especially the salaries of honorary teachers.

The distribution of data from the generalization of Hierarchical clustering shows that the job satisfaction of the teaching profession is influenced by salary of 65%, promotion of 15%, and the rest is on supervisory supervision, relationships with colleagues and work elements by 20%. The

results of the generalization of Hierarchical clustering can be seen in Figure 7.

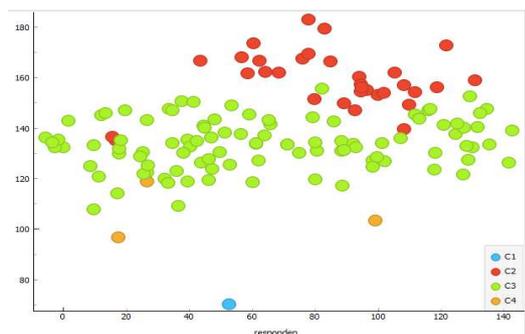


Fig. 7. Scatter Plot of Cluster Results Hierarchical method

The cluster results in general from the hierarchical method show an interesting pattern because outliers are detected. This outlier data can be studied more deeply by optimizing the number of clusters with DBI technique. The final results of Hierarchical clustering can be used as recommendations for stakeholders to review the salaries of the teaching profession, especially the salaries of honorary teachers.

B. Discussion

In this study, the results of K-Medoid and Hierarchical clustering were compared with the results of the study [5]. Research [5] uses the method of Scientific Identification Theory to Conduct Operation Research in Education (SITOREM). The results of this study indicate that the order of factors that need to be improved in strengthening the job satisfaction of the teaching profession is salary (SITOREM weight 3.13), promotion (SITOREM weight 3.51), supervisor supervision (SITOREM weight 3.67), job elements (SITOREM weight 3. SITOREM 3.52) and the last order is the relationship with colleagues (SITOREM weight 4.14). Details of the comparison of SITOREM results with K-Medoid and Hierarchical clustering can be seen in TABLE II.

TABLE II. COMPARISON RESULTS OF PERCENTAGE CLUSTER RESULTS ON RESEARCH FACTORS

Method	Indicator (%)				
	A	B	C	D	E
SITOREM [5]	26,0	23,0	19,0	17,0	15,0
K-Medoid	50,0	15,0	13,0	12,0	10,0
Hierarchical	65,0	15,0	9,0	6,0	5,0

Based on TABLE II, indicator A is salary, B is promotion, C is supervisor's supervision, D is work element and E is relationship with co-workers. In TABLE II it can be seen that the order of the prioritized factors from the two studies gave the same results. The results of the priority factors for strengthening the job satisfaction of the teaching profession based on the clustering method, both the K-medoid method and the Hierarchical method, produced the same pattern as the results of the SITOREM analysis [17]. This pattern can be used as validation technique to give recommendation of priority of factors.

Optimization of K-value in this clustering model used by the Davies-Bouldin Index (DBI) technique. DBI is one of the internal evaluations of the cluster, where whether or not the cluster results are good or not can be seen from the cohesion

and separation. This research produces clusters, so that it can be used to provide recommendations regarding what indicators or factors are prioritized for strengthening the job satisfaction of the teaching profession. To determine the level of validation of the cluster results, DBI analysis was performed. Determination of the most optimal cluster using DBI was carried out through experiments, and the results are as shown in Figures 7 and 8.

Validation through DBI is in accordance which states that determining the optimal number of clusters can be done by testing DBI validation . The number of clusters with the smallest DBI value can be concluded as the optimal number of clusters. Based on Figure 8 and Figure 9 below. To see the results of DBI values in detail, it can be seen in TABLE III. Based on the comparison using the two methods based on the DBI value, it shows the most optimal cluster results, namely using the Hierarchical method because the smaller the DBI value, the better the cluster results.

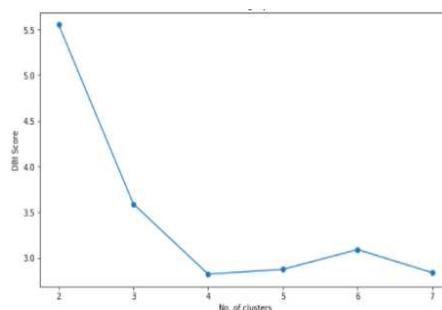


Fig. 8. Graph of DBI Value using the K-Medoid algorithm

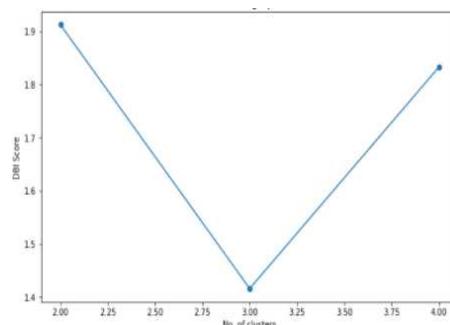


Fig. 9. Graph of DBI Value using Hierarchical Algorithm

TABLE III. OPTIMIZATION OF K-VALUE USING DBI

Number of Cluster	DBI Value	
	K- Medoid	Hierarchiecal
2	5,551	3,969
3	3,480	1,415*
4	2,823*	2,619

Noted : *Optimal K-value

Based on TABLE III shows that the hierarchical method is better than K-Medoid. This is indicated by the smallest DBI value at the position of the optimal number of clusters is three. This condition is also in accordance with the results of the cluster generalization plot in Figure 6, which shows that the hierarchical method in this case is able to detect outlier data. This outlier data is synchronized with the IDB results so that it eliminates the number of clusters from 4 to 3. The hierarchical clustering method shows better performance than K-Medoid because it is also supported by its ability to

visualize clustering results in the form of dendograms so that it is easier to interpret into knowledge [10].

This study also proposes the results of the comparison of patterns in TABLE II, as a proposed factor weight. This is in accordance with [3][5] which states that the dominant problem in job satisfaction in the honorary teacher profession in Indonesia lies in the salary for honorary teachers. This is supported by [1] which states that the shortage of teachers in Indonesia in 2021 will reach 1,090,678 people, the number of non-civil servant teachers will reach 937,228 people, and 728,461 people with honorary status. So the weight of the salary factor in assessing the job satisfaction of the teaching profession becomes very dominant (60%). This condition is also in accordance with the government's strategy of recruiting honorary teachers through the government employee program with a work agreement or known as PKKK. So that honorary teachers can get income rights in the form of salaries and allowances with the same amount as civil servants according to their level and group of positions. This strategy is expected to increase teacher motivation as well as improve the quality of education.

C. Web Visualization Results

The implementation of clustering to measure the strengthening factor of job satisfaction in the teaching profession is visualized in a web-based application, and one of the dashboard displays is shown in Figure 10. This shows that the data mining approach is very relevant to be used in education and other social research fields. The advantage of the data mining approach is that it is able to process large data

and the interpretation of knowledge can be visualized more attractively, so this can be proposed as a recommendation for stakeholders in the education sector, as well as implemented to small and medium enterprises empowering [18][19].

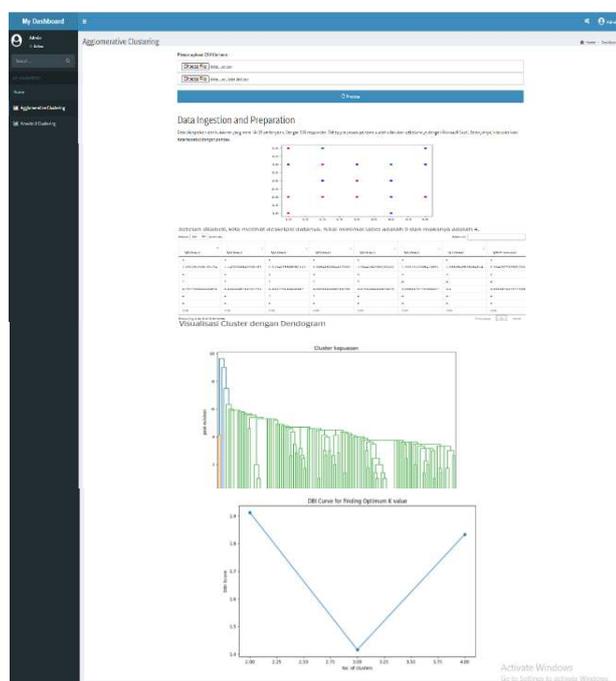


Fig. 10. Clustering application dashboard for strengthening job satisfaction in the teaching profession

IV. CONCLUSION

The application of the K-Medoid and Hierarchical algorithms on the implementation of clustering for strengthening work decisions for the teaching profession was made using the Python programming language and using a MySQL database with the KDD (Knowledge Discovery and Data Mining) research method. The K-Medoid and Hierarchical algorithms produce a priority order of factors that affect the strengthening of job satisfaction in the teaching profession. The results of the priority of these factors show the same order of priority as in the SITOREM analysis.

Factors or variables tested in this study using secondary data. The tested variables consist of 5 categories, namely salary, promotion, supervisory supervisor, relationship with co-workers and the work itself. The test was carried out using 136 respondents and 39 questions.

K-Medoid and Hierarchical clustering implementations were validated using DBI. The validation process was carried out through 3 experiments with modeling $K=2$, $K=3$ and $K=4$. The validation results show that the optimal cluster value is $K=3$ with a DBI value of 2.823 for the K-Medoid method and $K=4$ with a DBI value of 1.415 for the Hierarchical method. The first priority that needs to be improved for strengthening job satisfaction in the teaching profession is the salary variable, with a percentage of 50% through K-Medoid clustering and 65% through Hierarchical clustering. The priority variable that needs to be maintained is the relationship with colleagues by 10% through K-Medoid clustering and 5% through Hierarchical clustering. Based on the comparison using the two clustering algorithms, the most optimal cluster results are using Hierarchical clustering. This is indicated by the smaller DBI value, so the better the results of the cluster.

REFERENCES

- [1] Butcher, N., Khairina, N. N., Kumala, C., Loots, S., "The Struggle Against COVID-19 in Indonesian Education : Responses, Requirements, and Policy Needs for Learning Recovery", Report of World Bank, Washington, DC. <https://openknowledge.worldbank.org/handle/10986/36323> License: CC BY 3.0 IGO, 2021.
- [2] Sulfaemi Wahyu Bagja, "Perundang-Undangan Pendidikan Bogor," Program Studi Administrasi Pendidikan STKIP Muhammadiyah Bogor, 2016.
- [3] Setyaningsih, S. and Sunaryo, W., "Optimizing Transformational Leadership Strengthening, Self Efficacy, and Job Satisfaction to Increase Teacher Commitment", *International Journal of Instruction*, 14 (4), 427-438, <https://doi.org/10.29333/iji.2021.14425a>, 2021.
- [4] Udemba, N., "Relationship between Self Efficacy and Job Performance and Satisfactin of Secondary School Teachers in Anambra State", *International Journal of Multidisciplinary Research and Analysis*, 4 (5), 537-542, doi : 10.47191/ijmra/v4-i5-06, 2021.
- [5] Abdullah, T and Setyaningsih, S., "Penguatan Kepuasan Kerja, Kepemimpinan Transformasional dan Efikasi Diri Untuk Meningkatkan Komitmen Terhadap Profesi," Thesis, Postgraduate School, Universitas Pakuan, Bogor, 2019
- [6] Panjaitan, M., Herryance and Sitompul, D., "Implementasi Algoritma K-Means dan Analytical Hierarchy Process (AHP) untuk Klasterisasi Guru dan Memilih Guru Terbaik (Studi Kasus: SMA Santo Yoseph Medan)," 2017
- [7] Sukrianto, D., "Penerapan Data Mining untuk Kinerja Dosen Menggunakan Metode K-Means Clustering (Studi Kasus di AMIK Mahaputra Riau). *Jurnal PI-Cache*, Volume 5, pp. 54-63, 2016
- [8] Imantika, D., Bachtiar, F.A., and Rokhmawati R.I., "Penerapan metode k-means clustering dan analytical hierarchy process (ahp) untuk pengelompokan kinerja guru dan karyawan pada SMA brawijaya smart school," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. J-PTIIK*, vol. 3, no. 8, pp. 7382-7390, 2019
- [9] Rezki, D.A., "Analisa Kinerja Guru Sekolah di Dinas Pendidikan Dengan Menggunakan Metode Rought Set (Studi Kasus : Dinas

- Pendidikan Provinsi Sumatera Utara)," *J. Comput. Syst. Informatics*, vol. 1, no. 3, pp. 172–177, 2020
- [10] Witten, I.H., Frank, E., Hall, M.A., and Pal, C.J., "Data Mining : Practical Machine Learning Tools and Techniques", Fourth Ed., Morgan Kaufmann, 2017, <https://doi.org/10.1016/C2015-0-02071-8>
- [11] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai yang Layak Mengikuti Asessment Center," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 3, no. 1, pp. 87–93, 2018.
- [12] Pramesti, Dyang F., Furqon, M. T., Dewi, C., "Implementasi Metode K-Medoid Clustering untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(9), 2017
- [13] Vishwakarma, S., Nair, P.S., and Rao, D.S., "A Comparative Study of K Means and K Medoids Clustering for Social Media Text Mining". *International Journal of Advance Scientific Research And Engineering Trends*, Vol 2, Issue 11, 2017.
- [14] Arbin, N., Mokhtar, N.Z., Suhaimi, N.S., and Othman, Z., "Compaarative Analysis Between K Means and K Medoids for Statistical Clustering". *Third International Conference on Artificial Intelligence, Modelling and Simulation*, 2015.
- [15] Rani, U., and Sahu, S., "Review of Clustering Technique for Finding Similarities in Articles", *International Journal Computer Applications*, Vol 155, No 6, 2016.
- [16] G. Abdurrahman, "Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage", *JUSTINDO (Jurnal Sist. Dan Teknol. Inf. Indones.* 4 (13), 2019
- [17] Setyaningsih, S., Ishlah, M.S.N., and Hardhienata, S., "Optimization of enhancement of lecturer professional commitment using path analysis and sitorem method", *International Journal of Recent Technology and Engineering*, 8 (2 Special Issue 7), pp 5-10, 2019.
- [18] Tosida, E.T., Hairlangga, O., Amirudin, F., and Ridwanah, M. "Application of Decision Rules for Empowering of Indonesian Telematics Services SMEs", *IOP Conference Series : Material Science and Engineering* 332 (1), 012018, 2018.
- [19] Tosida, E.T., Andria, F., Wahyudin, I., Widiyanto, R., Ganda, M., and Lathif, R.R, "A Hybrid Data Mining Model for Indonesian telematics SMEs Empowerment, *IOP Conference Series : Materials Science and Engineering* 567 (1), 012001, 2019.