E-ISSN : 2986-2124
JURNAL ILMIAH MULTIDISIPLIN AMSIR
Published By : Lembaga Penelitian dan Pengabdian Masyarakat (LP2M)
Institut Ilmu Sosial dan Bisnis Andi Sapada
Parepare-Indonesia

Implementation of K-means and Weight Product (WP) Methods in Determining Work From Home (WFH) Priorities in the New Normal Period in Indonesia

Hadryan Eddy
Fakultas Ilmu Komputer, Tomakaka University, Indonesia
Email: hadryaneddy0@gmail.com

Abstrak
Covid-19 (SARS-CoV-2) adalah jenis virus baru yang dimulai di kota Wuhan, China pada akhir 2019. Seiring berjalannya waktu, perkembangan Virus Covid -19 meningkat pesat sehingga berdampak secara keseluruhan pada suatu organisasi. Saat ini, ada berbagai metode dalam organisasi yang telah dilakukan untuk mengurangi perkembangan COVID-19, salah satunya bekerja dari rumah selama masa New Normal. Penelitian ini bertujuan untuk menerapkan metode K-Means and Weight Product (WP) dalam menentukan prioritas pemberian jadwal WFH kepada seluruh karyawan tergantung kondisi masing-masing karyawan. Metode K-Means digunakan untuk mengelompokkan sejumlah data kasus pasien covid-19 berdasarkan usia dan riwayat penyakit. Hasil keluaran dari proses akan digunakan sebagai kriteria input dan kriteria pentingnya dalam metode WP. Penggunaan metode Elbow memudahkan untuk menentukan nilai K dalam proses clustering pada berbagai data yang digunakan. Dalam penelitian ini, nilai K terbaik adalah 3 berdasarkan hasil evaluasi menggunakan metode siku. Penggunaan 2 kriteria dari hasil pengelompokan data pasien covid-19 dalam pengambilan model keputusan dengan metode WP memberikan hasil keputusan yang lebih objektif dan tepat berdasarkan data/fakta yang telah terjadi. Aspek fungsionalitas sistem sangat baik setelah melalui proses pengujian hasil perhitungan secara manual dan menggunakan sistem, keduanya memiliki hasil yang sama.

Kata kunci- K-Means, Produk Berat, Penambangan Data, COVID-19, Bekerja Dari Rumah

Abstract
Covid-19 (SARS-CoV-2) is a new type of virus that began in the city of Wuhan, China in late 2019. Over time, the development of the Covid -19 Virus has greatly increased so that it has an overall impact on an organization. Currently, there are various methods in organizations that have been carried out to reduce the development of COVID-19, one of which is working from home during the New Normal period. This research aims to apply the K-Means and Weight Product (WP) methods in determining the priority of giving WFH schedules to all employees depending on the conditions of each employee. The K-Means method is used to group a number of covid-19 patient case data based on age and disease history. The output results of the process will be used as input criteria and criteria importance in the WP method. The use of the Elbow method makes it easy to determine the value of K in the clustering process on various data used. In this study, the best K value is 3 based on the evaluation results using the elbow method. The use of 2 criteria from the results of clustering covid-19 patient data in making decision models with the WP method provides more objective and precise decision results based on data / facts that have occurred. The functionality aspect of the system is very good after going through the process of testing the calculation results manually and using the system, both have the same results.

Keywords- K-Means, Weight Product, Data Mining, COVID-19, Work From Home

E-ISSN : 2986-2124
JURNAL ILMIAH MULTIDISIPLIN AMSIR
Published By : Lembaga Penelitian dan Pengabdian Masyarakat (LP2M)
Institut Ilmu Sosial dan Bisnis Andi Sapada
Parepare-Indonesia

Introduction

Covid-19 is a new type of virus that first appeared in the city of Wuhan, China in late 2019. Over time, the development of the spread of the Covid-19 virus has increased very rapidly to various countries in this part of the world, one of which is Indonesia, until it was declared a pandemic by the world health organization WHO. With this pandemic, it indirectly has a comprehensive impact in various sectors. One of the impacts occurred in the industrial sector which had laid off and even made mass dismissals due to declining sales.

The policy taken by the government to inhibit the development of the Covid-19 virus and keep the economy running is to implement a new lifestyle. The implementation of this *new normal* means that activities will run as usual by obeying health protocols. In the worker sector, there is a policy to *work* from home to reduce interactions carried out in an organization because the impact of the Covid-19 virus is very large on employee productivity [1].

For this reason, one of the treatments taken to reduce the impact of decreased employee performance is the WFH (*Work From Home*) policy. For some people, WFH has not been fully implemented either from the company, or from the employees themselves who do not necessarily have readiness for the WFH system [2]. Therefore, in determining the priority of providing WFH schedules to each employee, it is necessary to consider and take into account several factors/criteria of each employee so that productivity is still achieved and prevention of transmission of the Covid-19 virus can be minimized.

In previous research conducted [3] using 2 *Deccission Support System* (DSS) methods in ranking 10 prospective employees who registered at PT Warta Media Nusantara. In this study, more accurate ranking results were obtained using the *Weight Product* (WP) method compared to the *Simple Additive Weighting* (SAW) method based on manual calculation of assessment conditions.

This WP method has a working principle that connects between attributes through multiplication operations that are raised by the positive value of the weight of each criterion for *benefit-type* criteria and negative rank for criteria with *cost* types [4]. The weight used as a rank variable is the normalized weight. The level of importance of criteria can be determined based on the opinion of the user or determined based on facts/data that occur related to the problem. In determining the level of importance of criteria can be done through *text* processing to group data that has similarities with other data into the same group.

Grouping a number of data into small groups can be done using the K-Means algorithm. In his research [5], the K-Means algorithm is able to properly group books in the library consisting of several input variables to group books that are often borrowed.

Based on the problem and understanding of several related previous studies, this research aims to combine the K-Means algorithm with the WP method in determining the priority of providing WFH for employees. The K-Means algorithm will be used to classify data related to covid-19 patient cases, the output of this process will be used as input in the form of criteria and the level of importance of criteria in the WP method. The use of K-Means in grouping data is intended to make the determination of the level of importance more accurate so that it is expected that the output of the decision to prioritize the provision of WFH is more precise.

Materials and Methods

Research on K-Means algorithm and WP method has been done before. However, previous studies focused on one of these methods. Therefore, this research wants to combine the two methods in solving the problem of providing WFH for employees in a company.

E-ISSN : 2986-2124
JURNAL ILMIAH MULTIDISIPLIN AMSIR
Published By : Lembaga Penelitian dan Pengabdian Masyarakat (LP2M)
Institut Ilmu Sosial dan Bisnis Andi Sapada
Parepare-Indonesia

Previous research by [6] on the speed and effectiveness of the data clustering process using the K-Means algorithm on complex categorical data. In his research, an approach using partial scale Manhattan Frequency k-Means (MFk-M) was proposed. The dataset used to test MFk-M uses 3 different datasets namely, Global Terrorism Dataset (GTD), Baltimore Crimes dataset, and Mushroom dataset. The treatment of data before clustering process includes data transformation, random centroid determination, observation of cluster results, matrix identification, and centroid updating. From several experimental scenarios, it is found that the use of data sharing techniques during the clustering process can increase the speed of the data clustering process with good cluster results on a large amount of categorical and quantitative data.

Determination of the number of K in the data clustering process using the K-Means algorithm must be absolutely correct and represent the ideal / best number of groups with the right number of K. The application of the elbow method in research [7] is used in evaluating the best number of K in the process of grouping data on customer purchases. The experimental scenario carried out is to use the same 2 datasets with different amounts. In the study, 100 and 300 customer goods purchase datasets were used.

Research on the acceptance of new employee candidates using the WP method [8] uses 5 criteria in determining the ranking of prospective employees. Some of the criteria used are education criteria, GPA, work experience, interview test results and basic ability test results. All five criteria are of the benefit type. The stages of forming a decision model with the WP method include weighting each criterion, normalizing the weight of each criterion, calculating the S vector and calculating the V vector which determines the value of ranking. This model is tested on 4 alternative data / prospective employees, and the largest V vector value will be selected as a prospective employee who is prioritized to pass and get hired.

Further research [9] on determining the location to be used as a home building business using the WP method. Data collection carried out in this study is through 3 ways, namely through interviews, observations and literature studies on related problems. Three alternative data used in testing the decision model include Gading Rejo, Sukoharjo, and Pring sewu. From the results of the experiments that have been carried out, it is found that this system makes it easier for investors to determine the location of housing construction.

This study uses a dataset of 975 covid-19 patient data in Indonesia. The data is secondary data obtained from internet sources collected by someone from primary data sources from government sites, namely http://covid-monitoring.kemkes.go.id/ and https://corona.jatengprov.go.id/.

A. *K-Means Cluster*

The process of clustering large amounts of unlabeled data can be done by *unsupervised* machine learning methods. One of the algorithms that can do this is the K-Means algorithm [10]. This algorithm requires input in the form of a number of data that is still scattered and the value of the number K. This K value serves as a reference for the system in dividing the number of data into K groups. In the division process, a *centroid* value is needed which is used as a reference for data grouping [11]. Data with a value close to the center value will be grouped into that group, otherwise it will be grouped into another group.

B. *Elbow Method*

The grouping of data into K small groups is determined manually. So a method is needed that evaluates the number of K. The elbow method is one of the methods used in evaluating the determination of the number of K in the clustering process. The way this method works is by clustering a number of predetermined K values, for example in the range of 10, the elbow method will cluster from the number k equal to 1 to k equal to 10

E-ISSN : 2986-2124
JURNAL ILMIAH MULTIDISIPLIN AMSIR
Published By : Lembaga Penelitian dan Pengabdian Masyarakat (LP2M)
Institut Ilmu Sosial dan Bisnis Andi Sapada
Parepare-Indonesia

which is illustrated in the graph of the relationship between clusters and error reduction. A significant decrease in error or characterized by the formation of the most angled angle is the best K in the clustering process on the data [7].

C. *Weight Product (WP) Method*

Weight Product is one of the many decision-making methods. The basic concept of this method is a multiplication operation that connects between attributes that rank with normalized weights [12] as in equation (1)

$$w_j = \frac{w_o}{\sum w_o} \quad (1)$$

where :

w_o = criteria weight

w_j = normalized criteria weights

Criteria with benefit types will be positively multiplied and vice versa, criteria with negative types will be multiplied by the negative value of the criteria weight until the S vector results are obtained as in equation (2).

$$\text{Si} = \prod_{k=1}^{n} x_{ij}^{w_j} \quad (2)$$

Where i = 1,2,3

The final process of this WP method is the calculation of the V vector value as in equation (3).

$$vi = \frac{\prod_{j=1}^{n} x_{ij}^{w_j}}{\prod_{j=1}^{n} (x_j^*)^{w_j}} \quad (3)$$

The largest V vector value will be used/selected as an alternative with the first rank/order.

Results

The following is an explanation of several theories and studies related to the concepts and stages of the analysis process in this study.

A. *Data Acqusition*

The following shows the dataset obtained from the government website in the form of 975 data on covid-19 patient cases and also data on the case history of covid-19 patients in Table 1.

Table 1. Raw data

http://covid-monitoring.kemkes.go.id/			https://corona.jatengprov.go.id/		
ID	**Gender**	**Age**	**ID**	**Disease**	**Cases**
1	Female	31	1	Hypertension	304
2	Female	64	2	Stroke	28
.....	.....	.....	....	.....	.....
795	male	50	10	asthma	23

B. *Data Normalization and Filtering*

After the raw data is obtained, the data normalization process is carried out because there are 6 data with *missing values*. The data normalization process uses the mean formula, namely from a number of existing data that are not missing values, the average age value will be sought. From a total of 969 data accumulated to 36894 and divided by 969, a value of 46 is obtained. This number is used to fill in the empty column.

E-ISSN : 2986-2124
JURNAL ILMIAH MULTIDISIPLIN AMSIR
Published By : Lembaga Penelitian dan Pengabdian Masyarakat (LP2M)
Institut Ilmu Sosial dan Bisnis Andi Sapada
Parepare-Indonesia

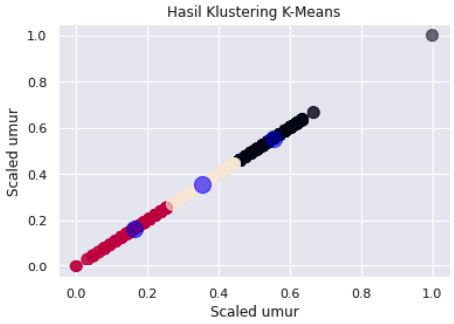
The next stage is the filtering process, which will select covid-19 patient data whose age is in the range of 17-57 years referring to government policies related to working age in Indonesia. The results of the filtering process obtained 604 data as shown in Table 2.

Table 2. Age data normalization and filter results

ID	Gender	Age
1	Female	31
2	Male	56
.....	.....	.....
604	Male	50

C. *Age Clustering*

After the age data has been normalized and filtered according to the data requirements for the problem, the next step is the clustering process. The clustering process is carried out using the help of the sklearn library from python which is run using google colab. The following are the results of the age clustering process shown in Figure 1.



Age clustering results

From a total of 604 data on the age of covid-19 patients, grouping is carried out into 3 age range groups which will be used as input variables for criteria and criteria importance in the decision support system.

D. *Disease History Clustering*

From a total of 10 disease histories suffered by covid-19 patients, clustering is carried out into 3 groups of disease histories according to the number of cases. The following are the results of the disease history clustering shown in Figure 2.

	riwayat_penyakit	kasus	kluster
0	hipertensi	304	1
1	diabetes melitus	307	1
2	stroke	28	0
3	Rheumatoid Arthritis	2	2
4	gagal jantung	37	0
5	ginjal kronis	41	0
6	jantung koroner	33	0
7	lain-lain	12	2
8	PPOK	6	2
9	asma	23	0
10	kanker paru	3	2

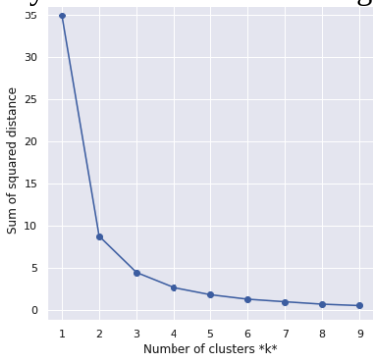
Figure 2. Results of disease history clustering

E. *Testing the Clustering Process*

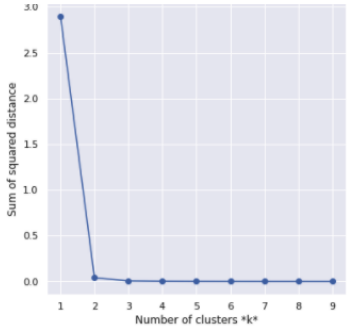
From the two processes of clustering age and disease history data, the researcher first determined the value of K is 3. Then to find out whether the value of K is equal to 3 will produce good cluster results or not. Testing is done using the elbow method by evaluating

E-ISSN : 2986-2124
JURNAL ILMIAH MULTIDISIPLIN AMSIR
Published By : Lembaga Penelitian dan Pengabdian Masyarakat (LP2M)
Institut Ilmu Sosial dan Bisnis Andi Sapada
Parepare-Indonesia

cluster results from 1-10. The test results for the age cluster can be seen in Figure 3 and the test results for the disease history cluster can be seen in Figure 4.



Evaluation of age data clustering



Evaluation of clustering of disease history data

Based on the results of testing the quality of clusters for age and medical history data, the value of 3 for K is said to be good. The principle of the elbow method assessment is that the more angles are formed, the optimal/best value is at that point.

F. *Clustering Result*

After testing the clustering results on age and disease history data, the test results show that the highest K value that can be used as clustering is 3, so the two data will be grouped into 3 data which will become input data for the level of importance of criteria in the decision support system. The following are the results of the clustering process shown in Table 3.

Table 3. Clustering results

Age Cluster		**Disease History Cluster**	
Data	**Cluster**	**Data**	**Cluster**
17-33 years old	1	Rheumatoid arthritis, lung cancer, COPD, etc.	1
34-45 years old	2	Stroke, heart failure, chronic kidney disease, coronary heart disease, asthma	2
46-57 years old	3	Hypertension, diabetes mellitus	3

Table 3 represents the results of grouping age and disease history data. These results will be used as additional criteria in the decision support system using the WP method.

G. *Application of WP method*

After obtaining the results of grouping data on the age of covid-19 patients and the history of diseases suffered by most covid-19 patients, the next stage is to implement the data as input in a decision support system (SPK) using the WP method.

E-ISSN : 2986-2124
JURNAL ILMIAH MULTIDISIPLIN AMSIR
Published By : Lembaga Penelitian dan Pengabdian Masyarakat (LP2M)
Institut Ilmu Sosial dan Bisnis Andi Sapada
Parepare-Indonesia

This study uses 5 criteria in making the WP model. The following 5 criteria and their types can be seen in Table 4.

Table 4. Criteria

Code	Criteria	Type	Weight
C1	Work interests WFO (Work From Office)	Cost	2
C2	WFH Facilities	Benefit	2
C3	Age	Benefit	3
C4	Access to Office	Cost	1
C5	Disease History	Benefit	4

The two criteria of age and medical history obtained from the clustering results will be used to determine the level of importance of the criteria. The level of importance other than the criteria of age and medical history can be seen in Table 5.

Table 5. Level of Importance

Description	Value/Weight
Low	1
Simply	2
High	3

The following is the level of importance for criteria derived from the data mining process using the K-Means algorithm shown in Table 6. .

Table 6. Importance of age and medical history

Age		Disease History	
Description	Weight	Description	Cluster
17-33 years old	1	Rheumatoid arthritis, lung cancer, COPD, etc.	1
34-45 years old	2	Stroke, heart failure, chronic kidney disease, coronary heart disease, asthma	2
46-57 years old	3	Hypertension, diabetes mellitus	3

In this study, 5 alternative data are used for the decision model testing process made using the WP method. The 5 alternative data will be assessed against the 5 criteria based on the conditions of each alternative. The following shows the suitability rating of the five alternatives with each criterion in Table 7.

Table 7. Suitability rating table

Alternative	C1	C2	C3	C4	C5
A1	3	3	2	3	1
A2	3	1	3	1	2
A3	2	3	1	1	1
A4	1	2	2	3	3
A5	2	1	1	2	2

Criteria weight data from each alternative is normalized before being used as a rank in connecting between criteria attributes. Below are the results of the weight normalization process shown in Table 8.

E-ISSN : 2986-2124
JURNAL ILMIAH MULTIDISIPLIN AMSIR
Published By : Lembaga Penelitian dan Pengabdian Masyarakat (LP2M)
Institut Ilmu Sosial dan Bisnis Andi Sapada
Parepare-Indonesia

Table 8. Weight normalization table

Criteria	Weight	Normalization	Normalized Weight
C1	2	2/12	0.1667
C2	2	2/12	0.1667
C3	3	3/12	0.2500
C4	1	1/12	0.0833
C5	4	4/12	0.3333
Total	**12**		

The next stage is the calculation of the S vector value. At this stage, the multiplication operation is carried out between the criteria attributes which are raised by normalized weights. The following are the results of the S vector calculation can be seen in Table 9.

Table 9: S vector calculation table

Alternative	Calculations	S-vector
A1	(3^(-0.1667)) x (3^0.1667) x (2^0.25) x (3^(-0.0833)) x (1^0.3333)	1.085169039
A2	(3^(-0.1667)) x (1^0.1667) x (3^0.25) x (1^(-0.0833)) x (2^0.3333))	1.380713072
A3	(2^(-0.1667)) x (3^0.1667) x (1^0.25) x (1^(-0.0833)) x (1^0.3333))	1.069913194
A4	(1^(-0.1667)) x (2^0.1667) x (2^0.25) x (3^(-0.0833)) x (3^0.3333))	1.756748044
A5	(2^(-0.1667)) x (1^0.1667) x (1^0.25) x (2^(-0.0833)) x (2^0.3333))	1.059463094
	Total	**6.352006442**

From the results of the S vector calculation, the total S vector value of all alternatives is used as the denominator in the calculation of vector V. The following are the results of the V vector calculation shown in Table 10.

Table 10. Vector calculation table V

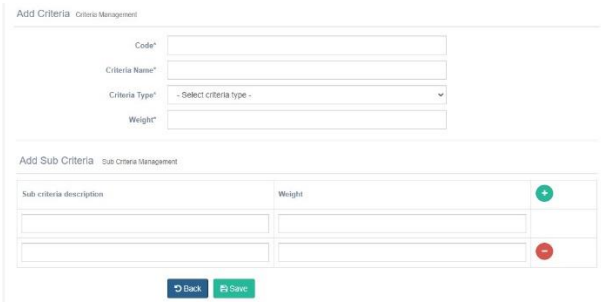
Alternative	Calculations	Preference Value
A1	1.085169039/6.352006442	0.1708387812
A2	1.380713072/6.352006442	0.217366447
A3	1.069913194/6.352006442	0.1684370448
A4	1.756748044/6.352006442	0.2765658473
A5	1.059463094/6.352006442	0.1667918797

The results of the V vector calculation will produce a preference value, this value is used as a reference in determining the priority of determining the WFH schedule. The highest preference value will be selected as prioritizing the WFH schedule.

H. *Decision Support System Testing*

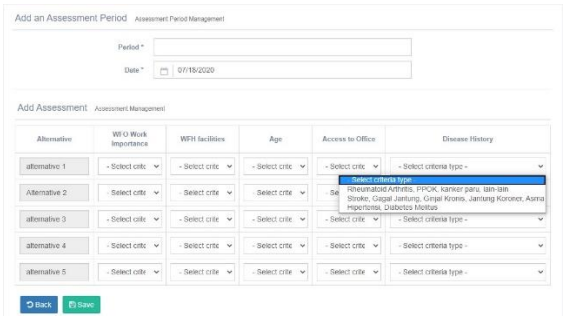
Testing is carried out on several aspects contained in this decision support system with the WP method, including the following.

E-ISSN : 2986-2124
JURNAL ILMIAH MULTIDISIPLIN AMSIR
Published By : Lembaga Penelitian dan Pengabdian Masyarakat (LP2M)
Institut Ilmu Sosial dan Bisnis Andi Sapada
Parepare-Indonesia



Criteria management interface

Figure 5 represents the interface for adding criteria and sub criteria. Based on the experience of several people who have tried this system, 90% agree to say that the criteria management process is very easy and dynamic based on the needs of users. Because this system uses a dynamic addition / subtraction of sub criteria form without reloading the page, so users feel easy and comfortable with the interface.



Assessment interface

The assessment interface as shown in Figure 6 provides convenience for users with clear information and the use of dropdown users will find it easier to provide an assessment by selecting the criteria provided. In addition, from the aspect of flexibility, the interface of this assessment system will adjust to the number of alternatives and the latest criteria from the system.

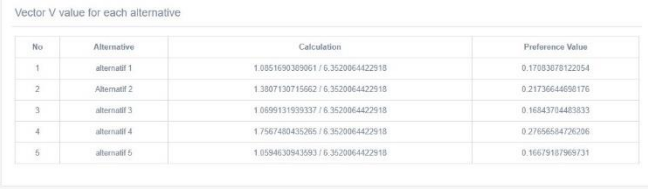


Figure 7. System calculation results

In testing aspects of system functionality, a comparison of system output results with the results of manual calculations as in Figure 7. The results of calculations using the system and manually as in Table 10 show the same results. The use of criteria from the clustering results provides better and more precise decision results because the criteria and the importance of the criteria used are based on data and facts that occur, so that the decisions produced are more objective.

Conclusion

From the results of the research that has been done, it is found that the K-Means method can group data well if the K value is determined correctly. In this study, the evaluation of the

K value using the elbow method can help facilitate in determining the right K value, so that the clustering results obtained are precise and spread evenly into data groups. Additional criteria in making decision models using the WP method, namely age criteria and medical history which are the results of the clustering process, provide output results for WFH prioritization decisions that are more objective and precise. This decision support system has good functionality aspects as evidenced by the test results between manual calculations and manual calculations. Decision from the system gives the same result. In addition to the functionality aspect, this system has ease of use based on the assessment of several people who demo this application dominantly agree with the ease of use of this application.

References

[1] M. A. H. Duta Mustajab, Azies Bauw, Abdul Rasyid, Andri Irawan*, Muhammad Aldrin Akbar, "Working from Home Phenomenon as an Effort to Prevent COVID-19 Attacks and Its Impacts on Work Productivity," *Int. J. Appl. Bus. Tijab*, vol. 4, no. 9, pp. 13-21, 2020, doi: 10.1017/CBO9781107415324.004.

[2] A. Purwanto, U. P. Harapan, and S. Ekploratif, "Exploratory Study of the Impact of Work From Home (WFH) on Teacher Performance During the Covid-19 Pandemic," vol. 2, pp. 92-100, 2020.

[3] A. Setyawan, F. Y. Arini, and I. Akhlis, "Comparative Analysis of Simple Additive Weighting Method and Weighted Product Method to New Employee Recruitment Decision Support System (DSS) at PT Warta Media Nusantara," *Sci. J. Informatics*, vol. 4, no. 1, pp. 34-42, 2017, doi: 10.15294/sji.v4i1.8458.

[4] A. Awasthi, S. S. Chauhan and S. K. Goyal, A multi-criteria decision making approach for location planning for urban distribution centers under uncertainty, Mathematical and Computer Modeling, vol. 53, 2011, pp. 98-109.

[5] S. P. Tamba, M. D. Batubara, W. Purba, M. Sihombing, V. M. Mulia Siregar, and J. Banjarnahor, "Book data grouping in libraries using the k-means clustering method," *J. Phys. Conf. Ser.*, vol. 1230, no. 1, 2019, doi: 10.1088/1742-6596/1230/1/012074.

[6] S. Ben Salem, S. Naouali, and Z. Chtourou, "A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach," *Comput. Electr. Eng.*, vol. 68, no. April, pp. 463-483, 2018, doi: 10.1016/j.compeleceng.2018.04.023.

[7] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration of K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.

[8] D. M. Khairina, M. R. Asrian, H. R. Hatta, C. Science, and I. Technology, "Recruitment Using Weighted Product Method," *Proc. Int. Conf. Inf. Tech., Comput. Electr. Eng.*, pp. 297-301, 2016.

[9] Oktafianto *et al.*, "Determining housing location using weighted product," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 3563-3568, 2018, doi: 10.14419/ijet.v7i4.18838.

[10] Madadipouya K, 2017 A Survey on Data Mining Algorithms and Techniques in Medicine JOIV Int. J. Informatics Vis. 1, 3 p. 61-71.

[11] Dubey A K Gupta U and Jain S, 2018 Comparative Study of K-means and Fuzzy C-means Algorithms on Breast Cancer Data Int. J. Adv. Sci. Eng. Inf. Technol. 8, 1 p. 18-29.

[12] S. Kusumadewi, S. Hartati, A. Harjoko, and R. Wardoyo, Fuzzy MultiAttribute Decision Making (FuzzyMADM).Yogyakarta: Graha, 2006.