



Implementation of k-means clustering and weighted products in determining crime-prone locations

Yuni Rahmatika^{*1}, Eko Sedyono², Catur Edi Widodo³

Universitas Diponegoro, Indonesia^{1,3}

Universitas Kristen Satya Wacana, Indonesia²

Article Info

Keywords:

Decision Support System, Data Mining, K-Means Clustering, Weighted Product Method, Crime

Article history:

Received 21 April 2020

Revised 26 May 2020

Accepted 07 July 2020

Published 31 August 2020

Cite:

Rahmatika, Y., Sedyono, E., & Widodo, C. (2020). Implementation of K-Means Clustering and Weighted Products in Determining Crime-Prone Locations. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 5(3). doi:<https://doi.org/10.22219/kinetik.v5i3.1067>

*Corresponding author.

Yuni Rahmatika

E-mail address:

yuni.rahmatika25@gmail.com

Abstract

Clustering algorithms can be used to build geographic mapping systems to determine crime-prone locations. This study aims to establish a geographical mapping system to determine crime-prone locations that can help police control certain locations that often occur crime and provide information to people in crime-prone locations. Criminal groups are calculated based on crime data from November 2018 to October 2019 which occurred in 9 districts in Kudus Regency. The crime grouping process uses the k-means method used to classify based on regional vulnerability and uses a weighted product method that functions as a vulnerability ranking that is vulnerable to crime selection. The grouping results obtained from this study are that there are 1 very vulnerable area, 5 areas in the vulnerable category, and 3 safe areas. While the weighted product method produces Melatilor area as a vulnerable area to be defeated by a score of 0.182093. This research provides benefits for the public to see crime-prone areas so that they can be more vigilant, while for the police to analyze crime so as to speed up the process of resolving crime and increase and improve crime prevention measures.

1. Introduction

Crimes are unlawful acts that refer to human actions, as well as people's views of what is permitted and permitted in law. Crime is a problem for people in the world from time to time and the steps needed to eradicate crime from the world. Current police strategies work to protect crimes after crimes occur [1]. Administration and maintenance at the police station can be made more efficient and effective with the analysis of assistance data that can help law enforcement officials improve the crime resolution process [2]. All government departments around the world give top priority to security and focus on limiting the occurrence of crime. In committing crimes, criminals have the maximum utilization of new practices and innovations [3]. The ongoing pandemic is increasing the number of crime cases. Disasters that occur cause damage to humans, damage to economic assets, and often occur due to widespread looting and difficult to resolve in the community. Natural disasters that occurred in Chile increased the incidence of theft by 15% -30% relative to areas not directly related to the constant impact for 4 years after the disaster [4]. The number of crimes that occur with time and place and various types of crime, it is difficult for the community to determine the location of crime-prone areas and locations that are safe from crime. One solution to improve security in cities is to provide more support in public spaces to stay safe [5]. Increasing the number of crimes Becomes a problem in social aspects such as work, health, and education [6]. The main challenge that is corrected by all police organizations where data is recorded and corrected is not necessary to increase crime data. Complicated and much data is difficult to solve because it requires difficulty and requires a long time [7]. Analysis of crime data can help law enforcement to speed up the process of crime resolution [1].

Grouping crime is a good strategy in improving planning and improving crime prevention [6]. The emergence of the field of prediction of crime, find facts that show geographical patterns. These patterns can be useful for predicting crimes before they occur and allow the police to take proactive actions [8]. Geo-based crime analysis tools used by communities for areas that are vulnerable to crime problems [9]. Geographical mapping can provide insights about handling crime in cities, allocating resources to law enforcement agencies, and improving the environment that is built to prevent crime [10].

Mapping the area involving most of the areas mentioned, different areas also have different potential [11]. In this case, the K-algorithm means a grouping that is used to group crime areas based on the standard of crime that occurred in the region. Clustering is a data mining technique used for grouping data objects into clusters so that the data needed by each cluster is different [12]. Research for risk analysis of killings using the K-mean concludes that homicide declined

from 2009 to 2018, and could be used to improve fighting methods [13]. According to [14] the K-means algorithm can be visualized using Google Maps for a more interactive and easier use. From existing research using only K-means grouping techniques based on defiance that occurs, crime also has several specific alternatives in ranking crime-prone actions based on criteria needed by users, this research combines with weighted product methods.

Weighted product method is a ranking method that uses multiplication to link attribute ratings, on the rating of each attribute must be raised first with the weight of the attribute in question [15]. Based on the relative standards stated in the percent weighted product method is known to have an average of 35.75% while the SAW method has an average of 12.64% [16]. In this study, the weighted product method is used to rank locations of crime-prone areas with criteria sorted by priority of interest. Weighted product method is used to calculate the ranking of several crime-prone locations based on predetermined criteria, namely the number of crime incidents, population density and street lighting levels. The system of determining crime-prone locations using the K-means algorithm can classify crimes based on the types of crimes that occur in certain areas that can be visualized using GoogleMaps. However, with the limitations of the K-mean algorithm which can only group data, it can be combined with the weighted product method that can be used by users to filter crime data based on certain criteria so that they can determine the level of regional vulnerability based on the type of crime.

2. Research Method

2.1 Research subject

The subject in this research is how to implement the K-Means clustering algorithm and weighted product to determine crime-prone locations so that it can make it easier for the community or the police to control certain locations that often occur criminal acts and provide information to the public at crime-prone locations based on criteria that have been determined.

2.2 K-Means Clustering

Clustering is a data mining technique used to group data objects into clusters so that the data held by each cluster is different [12]. Grouping is an analytical method used for grouping data objects into grouping data needed by each different grouping and this grouping can be categorized as a hierarchical grouping [17]. Clusters are collections of data objects that are similar to each other in the same cluster and are different from objects in other clusters. The grouping method commonly used is K-Means [18].

The K-Mean algorithm is simple to implement and runs relatively quickly, is easily resolved, and is of general use in practice [19]. K-means is used to develop decision support systems to improve safety using a multi-step process. [20]. So cluster analysis is generally given data that is in one cluster that has a high level of expectation, and data that is in various clusters is low [21]. The k-means algorithm is very easy to implement, mostly in the implementation of very large data that can also overcome the data difficulties that are needed [22].

The basic process of clustering with k-means consists of six steps, the first step is an initialization, which is to determine the value of k as the number of clusters desired and the desired metric (distance) metric. For example, if you want to group into 5 clusters, then $k = 5$.

The second step is to determine the cluster center randomly. Then the third step counts each data to the center of the cluster Mathematically illustrated in Equation 1.

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

In the fourth step that is grouping data into clusters with the shortest distance, then calculate the new cluster center using Equation 2.

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (2)$$

In the last step, step six repeats step two to step four so that no more data is moved to another cluster.

2.3 Weighted Product

Weighted Product Method is a ranking that uses multiplication to link attribute ratings, where the rating of each attribute must be raised first with the weight of the attribute in question [15]. The Weighted Product method requires a normalization process because this method multiplies the results of the assessment of each attribute (taufik 2019). Weighted product methods can provide alternatives with multiplication to link the attribute ratings that aim to provide

the best alternative results [23]. Calculations using the Weighted Product (WP) method must go through the following stages:

First give weights (w) for each attribute needed, then the second step to make weights improvements. Weights can be fixed by the following Equation 3.

$$P_k = \frac{w_j}{\sum w_j} \tag{3}$$

Group each criterion into cost criteria or profit criteria. The higher the value of the profit criteria, the better the results, while the higher the cost criteria, the results will be lower. The value of w is positive for the profit criteria and negative for the cost criteria.

The third stage is to multiply the value of each criterion x_i which has been raised with the weight that has been fixed to produce vector s, the notation in Equation 4.

$$s_i = x_i^{P_k} \tag{4}$$

The fourth or final step is to calculate vector v with Equation 5 which will be used to obtain the ranking results.

$$v_i = \frac{s_i}{\sum s} \tag{5}$$

3. Results and Discussion

3.1 K-Means calculation

The first step is to group data on crime-prone locations, which are sub-districts in Kudus regency based on the type of crime category using the k-means clustering method. In this research, a case study of crime data is used in Kudus regency. Data on crime incidents can be seen in Table 1. The data used are data of criminal acts from November 2018 to October 2019 which occurred in 9 districts in Kudus regency.

Table 1. Crime Data

Districts	Em	Ct	Cr	Fr	Be	Ga
Kudus	13	12	9	15	9	12
Jati	0	3	1	1	1	2
Undaan	1	1	1	1	0	0
Kaliwungu	1	1	2	1	0	2
Gebog	0	1	0	1	1	4
Dawe	0	1	2	0	0	2
Bae	1	0	0	1	0	1
Jekulo	0	1	1	0	1	2
Mejobo	0	1	0	0	0	4

- The first iteration
Determine the number of clusters be 3.
- Determination of the initial cluster center
Determine the cluster center at random which can be seen in Table 2.

Table 2. New Cluster Center

Districts	Em	Ct	Ct	Fr	Be	Ga
Jati	0	3	1	1	1	2
Kudus	13	12	9	15	9	12
Bae	1	0	0	1	0	1

- Calculation of the distance of each data to the cluster center
Calculate the distance of each crime data with the initial center of the cluster table using Equation 1.

$$C(1,1) = \sqrt{(13 - 0)^2 + (12 - 3)^2 + (9 - 1)^2 + (15 - 1)^2 + (9 - 1)^2 + (12 - 2)^2} = 25,96151$$

$$C(1,2) = \sqrt{(13 - 13)^2 + (12 - 12)^2 + (9 - 9)^2 + (15 - 15)^2 + (9 - 9)^2 + (12 - 12)^2} = 0$$

$$C(1,3) = \sqrt{(13 - 1)^2 + (12 - 0)^2 + (9 - 0)^2 + (15 - 1)^2 + (9 - 0)^2 + (12 - 1)^2} = 27,69476$$

The calculation continues until the 9th data. The results of the calculation of the distance of each data to the initial center of the cluster are in the [Table 3](#).

Table 3. Calculation of Data Distance to the First Iteration Cluster Center

	C0	C1	C2
25,96151	0	27,69476	
0	25,96151	3,605551	
3,162278	27,38613	1,732051	
2,645751	26,28688	2,44949	
3	26,36285	3,464102	
2,645751	27,29469	2,828427	
3,605551	27,69476	0	
2,236068	27,25803	2,44949	
3,316625	27,22132	3,464102	

- Data grouping

The results of distance calculation will be compared with the closest data distance chosen by the cluster center, this distance cluster shows the data that has the shortest distance in one group. [Table 4](#) shows the grouping of data obtained.

Table 4. Cluster Data Grouping

Districts	C0	C1	C2	Cluster
Kudus	25,96151	0	27,69476	C1
Jati	0	25,96151	3,605551	C0
Undaan	3,162278	27,38613	1,732051	C2
Kaliwungu	2,645751	26,28688	2,44949	C2
Gebog	3	26,36285	3,464102	C0
Dawe	2,645751	27,29469	2,828427	C2
Bae	3,605551	27,69476	0	C2
Jekulo	2,236068	27,25803	2,44949	C0
Mejobo	3,316625	27,22132	3,464102	C0

After grouping members from each cluster, then creating a new cluster center to be calculated based on data from each cluster member. Iteration continues in the same way until there are no changes to the data in a cluster.

K-means grouping using crime data using rapid miner with an initial cluster number of 3 clusters, resulting in 5 items for cluster 0, 1 item for cluster 1 and 3 items for cluster 2. The results of grouping using fast miners can be seen in [Figure 1](#) Meanwhile, [Figure 2](#) shows a graph of crime in Kudus that occurred in each district.

Label	Em	Ct	Cr	Fr	Be	Ga
cluster_1	13	12	9	15	9	12
cluster_0	0	3	1	1	1	2
cluster_2	1	1	1	1	0	0
cluster_2	1	1	2	1	0	2
cluster_0	0	1	0	1	1	4
cluster_0	0	1	2	0	0	2
cluster_2	1	0	0	1	0	1
cluster_0	0	1	1	0	1	2
cluster_0	0	1	0	0	0	4

Figure 1. K-Means Calculation Results

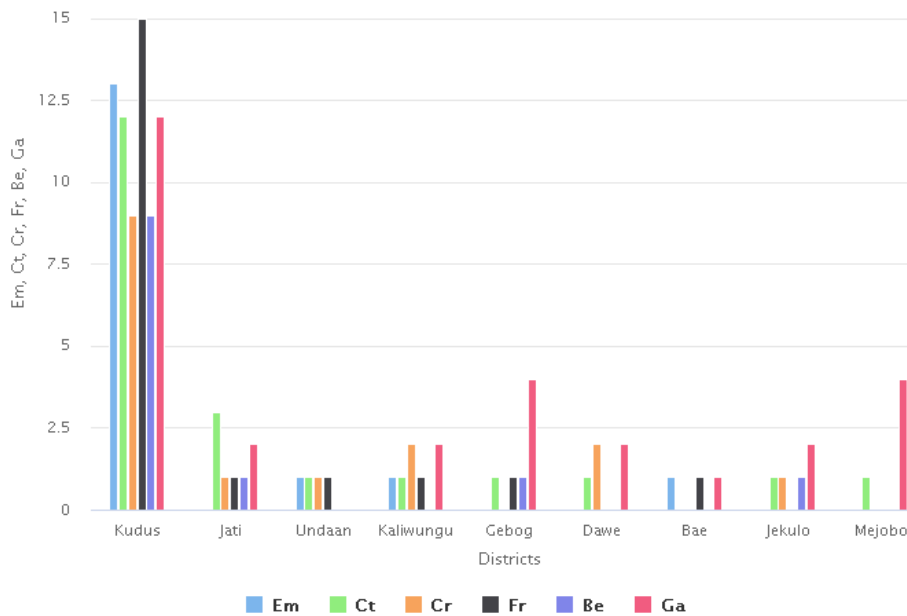


Figure 2. Graph of Crime

3.2 Weighted product calculation

Hazard level criterion data is used to calculate rankings on the weighted product method. In the process of ranking the weighted product method, the criteria used are the criteria for the number of criminal cases that occur in Table 5, population density in Table 6, and in Table 7 the criteria for distance from the nearest police station.

- Criteria for the number of crime incidents

The criteria for crime incidents are the number of criminal incidents that occurred in the area in Kudus regency.

Table 5. Criteria Incidents

Category	Code
<5 incidents	1
5 -10 incidents	2
> 10 incidents	3

- Population density criteria

The criterion for population density is the number of houses in the area where crime is prone.

Table 6. Criteria Population Density

Category	Code
< 5 houses	1
5 -15 houses	2
> 15 houses	3

- Distance criteria

The distance criteria used is the distance between the crime scene and the nearest police station.

Table 7. Criteria Distance

Category	Code
< 1,5 km	3
1,5 km – 3 km	2
> 3 km	1

Decision making gives preference weights to the following criteria for many events, population density and distance: $W = (50, 20, 30)$

Data on beating crime can be seen in Table 8 as a sample to calculate which sub-district ranking is more prone to beating crime.

Table 8. Beating Sample Data

Crime scene	Amount	Population	Distance
Kauman kudus	2	3	1
Keramat kudus	1	2	2
Demaan kudus	2	2	1
Melati lor kudus	3	2	1
Glantengan kudus	1	2	1
Terban jekulo	1	1	3
Loram kulon jati	1	2	3

The weighting improvement is calculated using Equation 3 from the previous preference $W = (50, 20, 30)$. w_j is the weight w index to j . So for w_1 is 50, w_2 is 20 and w_3 is 30. $\sum w_j$ is the sum of w . The calculation of weight is:

$$w_1 = \frac{50}{50 + 20 + 30} = 0,5$$

$$w_2 = \frac{20}{50 + 20 + 30} = 0,2$$

$$w_3 = \frac{30}{50 + 20 + 30} = 0,3$$

The next stage is the calculation of vector s using Equation 4 by multiplying the value of each criterion raised with the weight that has been fixed. The vector s calculation is:

$$s_1 = 2^{0,5} \times 3^{0,2} \times 1^{0,3} = 1,76173$$

$$s_2 = 1^{0,5} \times 2^{0,2} \times 2^{0,3} = 1,414214$$

$$s_3 = 2^{0,5} \times 2^{0,2} \times 1^{0,3} = 1,624505$$

$$s_4 = 3^{0,5} \times 2^{0,2} \times 1^{0,3} = 1,989604$$

$$s_5 = 1^{0,5} \times 2^{0,2} \times 1^{0,3} = 1,148698$$

$$s_6 = 1^{0,5} \times 1^{0,2} \times 3^{0,3} = 1,390389$$

$$s_7 = 1^{0,5} \times 2^{0,2} \times 2^{0,3} = 1,597138$$

Then after determining the vector value using Equation 5 that will be used is to calculate Preferences (V_i) for ranking. The results of Calculating V_i are Preference can be seen in Table 9.

$$V_1 = \frac{1,76173}{1,76173 + 1,414214 + 1,624505 + 1,989604 + 1,148698 + 1,390389 + 1,597138} = 0,161238$$

Table 9. Preference	
Preference	Value
V_1	0,161238
V_2	0,129432
V_3	0,148679
V_4	0,182093
V_5	0,105132
V_6	0,127252
V_7	0,146174

From the above calculation, the value of V_4 determines the greatest value so that in other words V_4 is the highest alternative with the weight given by the decision-maker for the selection of the most prone to crime.

4. Conclusion

The results of this research discussion show the information system for determining crime-prone locations that were built with K-means Clustering is able to classify the level of rioting locations of crime that have occurred in Kudus Regency. Geographic based crime grouping is a good strategy in improving planning and improving crime prevention. This system can make it easier for the police to regulate certain locations that often occur criminal acts and provide information to people prone to crime in accordance with the requests that have been determined. Clusters form three groups namely vulnerable, vulnerable, and safe. While the product weighting method can be used to rank crime-prone locations that are wanted by users based on weighting and selection of crime assistance policy, population subsidence, and distance from the nearest police office. The test results show the system can work well according to design. The results of manual calculations with calculations using the system have the same results as the system can be accepted valid. The results of the grouping obtained from this study there are 1 very vulnerable area, 5 areas in the vulnerable category, and 3 safe areas. While the weighted product method produces Melatilor area as a vulnerable area to be defeated by a score of 0.182093. This research provides benefits for the public to see crime-prone areas so that they can be more vigilant, while for the police to analyze crime so as to speed up the process of resolving crime and increase and improve crime prevention measures.

Suggestions for further research development are broader uses in other areas of serious concern, such as the grouping and mapping of pandemics that are currently taking place, providing information to communities where pandemic events are more prevalent.

Notation

The example of notation can be described with the following description:

d_{ik}	: object distance i and k
C_{ij}	: coordinates of object i on dimension j
C_{kj}	: coordinates of object k on dimension j
i	: 1, 2, 3, ..., m
j	: 1, 2, 3, ..., n
k	: 1, 2, 3, ..., k
x_{ij}	: a record i of each document selected as a member of the cluster
p	: number of cluster members to k
P_k	: the value of weighting improvement to k
w_j	: weight value to j
$\sum w_j$: the total value of weight j
s_i	: vector value of s to i
x_i	: i value criteria
v_i	: vector value v
$\sum s$: the number of vector values s
Em	: embezzlement of valuables
Ct	: theft with another incriminating case
Cr	: motor vehicle theft
Fr	: fraud
Be	: beatings
Ga	: gambling

References

- [1] V. Jain, Y. Sharma, A. Bhatia, and V. Arora, "Crime Prediction using K-means Algorithm," *GRD Journals - Glob. Res. Dev. J. Eng.*, 2017.
- [2] T. Win and E. E. Phyo, "Predicting of Crime Detection Using K-Means Clustering Algorithm," vol. 6, no. 3, pp. 8–12, 2019.
- [3] S. B. M. Vijayalakshmi and H. Goyal, "Crime Pattern Recognition And Prediction Using Optimised K-Means And Svm," vol. 118, no. 22, pp. 581–586, 2018.
- [4] J. García, "Journal of Economic Behavior and Organization The lasting effects of natural disasters on property crime : Evidence from the 2010 Chilean earthquake R," Vol. 175, Pp. 114–154, 2020. <https://doi.org/10.1016/j.jebo.2020.04.008>
- [5] P. Piroozfar, E. R. P. Farr, E. Aboagye-Nimo, and J. Osei-Berchie, "Crime prevention in urban spaces through environmental design: A critical UK perspective," *Cities*, 2019. <https://doi.org/10.1016/j.cities.2019.102411>
- [6] Q. Zhu, F. Zhang, S. Liu, and Y. Li, "An anticrime information support system design: Application of K-means-VMD-BiGRU in the city of Chicago," *Inf. Manag.*, 2019. <https://doi.org/10.1016/j.im.2019.103247>
- [7] S. Aarthi, M. Samyuktha, and M. Sahana, "Crime hotspot detection with clustering algorithm using data mining," in *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019*, 2019. <https://doi.org/10.1109/ICOEI.2019.8862587>

- [8] G. Hajela, M. Chawla, and A. Rasool, "A Clustering Based Hotspot Identification Approach for Crime Prediction," in *Procedia Computer Science*, 2020. <https://doi.org/10.1016/j.procs.2020.03.357>
- [9] S. Curtis-Ham and D. Walton, "Mapping crime harm and priority locations in New Zealand: A comparison of spatial analysis methods," *Appl. Geogr.*, 2017. <https://doi.org/10.1016/j.apgeog.2017.06.008>
- [10] S. N. Sadeek, A. J. M. Minhuz Uddin Ahmed, M. Hossain, and S. Hanaoka, "Effect of land use on crime considering exposure and accessibility," *Habitat Int.*, 2019. <https://doi.org/10.1016/j.habitatint.2019.102003>
- [11] D. M. Fenimore, "Mapping harmspots: An exploration of the spatial distribution of crime harm," *Appl. Geogr.*, 2019. <https://doi.org/10.1016/j.apgeog.2019.06.002>
- [12] U. R. Raval and C. Jani, "Implementing & Improvisation of K-means Clustering Algorithm," *Int. J. Comput. Sci. Mob. Comput.*, 2016.
- [13] A. Joshi, A. S. Sabitha, and T. Choudhury, "Crime Analysis Using K-Means Clustering," *Proc. - 2017 Int. Conf. Comput. Intell. Networks, CINE 2017*, Vol. 7, No. 04, Pp. 33–39, 2018. <https://doi.org/10.1109/CINE.2017.23>
- [14] S. Sivaranjani, S. Sivakumari, and M. Aasha, "Crime prediction and forecasting in Tamilnadu using clustering approaches," in *Proceedings of IEEE International Conference on Emerging Technological Trends in Computing, Communications and Electrical Engineering, ICETT 2016*, 2017. <https://doi.org/10.1109/ICETT.2016.7873764>
- [15] S. H. Kusumadewi, "Fuzzy Multi-Attribute Decision Making (Fuzzy MADM)," *Graha Ilmu Yogyakarta*, 2006.
- [16] K. Savitha and D. C. Chandrasekar, "Vertical Handover decision schemes using SAW and WPM for Network selection in Heterogeneous Wireless Networks," *Glob. J. Comput. Sci. Technol.*, 2011.
- [17] X. Wu, B. Wu, J. Sun, S. Qiu, and X. Li, "A hybrid fuzzy K-harmonic means clustering algorithm," *Appl. Math. Model.*, 2015. <https://doi.org/10.1016/j.apm.2014.11.041>
- [18] K. Kouser, "A comparative study of K Means Algorithm by Different Distance Measures," *Int. J. Innov. Res. Comput. Commun. Eng.*, 2013.
- [19] E. Prasetyo, *DATA MINING Mengolah Data Menjadi Informasi Menggunakan Matlab*. 2014.
- [20] K. Dhalmahapatra, R. Shingade, H. Mahajan, A. Verma, and J. Maiti, "Decision support system for safety improvement: An approach using multiple correspondence analysis, t-SNE algorithm and K-means clustering," *Comput. Ind. Eng.*, 2019. <https://doi.org/10.1016/j.cie.2018.12.044>
- [21] X. Lan, Q. Li, and Y. Zheng, "Density K-means: A new algorithm for centers initialization for K-means," in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 2015. <https://doi.org/10.1109/ICSESS.2015.7339213>
- [22] B. Bangoria, N. Mankad, and V. Pambhar, "A survey on Efficient Enhanced K-Means Clustering Algorithm," *IJSRD -International J. Sci. Res. Dev.*, vol. 1, no. 9online, pp. 2321–613, 2013.
- [23] D. M. Khairina, M. R. Asrian, and H. R. Hatta, "Decision support system for new employee recruitment using weighted product method," in *Proceedings - 2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2016*, 2017. <https://doi.org/10.1109/ICITACEE.2016.7892459>