



## Spatial Clustering-Based Gas Station Location Determination

Ayasda Rahardian<sup>1</sup>, Eneng Tita Tosida<sup>2</sup>, Ema Kurnia<sup>3</sup>, Hairulnizam Mahdin<sup>4</sup>

<sup>1,2</sup>*Computer Science Study Program Pakuan University, Bogor, Indonesia*

<sup>3</sup>*Informatics Management Study Program Pakuan University, Bogor, Indonesia*

<sup>4</sup>*Faculty Sains Komputer dan Teknologi Maklumat, Universiti Tun Hussein Onn Malaysia*

\*Corresponding author email: [rahardianayasda@gmail.com](mailto:rahardianayasda@gmail.com)

---

### Abstract

The absence of gas stations built in Cibeber Subdistrict is not balanced with the high level of transportation use for ease of mobility among residents. The purpose of this research is to cluster data using K-Means clustering and spatial modeling to provide a potential location for the construction of gas stations in Cibeber District. Based on the research process that has been carried out using RStudio, the potential villages for the construction of gas stations consist of four villages, namely Cikotok, Cibeber, Neglasari, and Wanasari. As for the results of spatial modeling, Cibeber District has a total of 862 potential location points, and within the scope of potential villages, namely four villages, there are 233 potential location points. Then, after being processed with weighted products for optimization and getting the best location results, 3 potential locations were obtained, namely Tegalumbu Village located in Wanasari Village, Nagrak Village located in Cikotok Village, and Cinangga Village located in Cibeber Village.

*Keywords:* K-Means Clustering, Model Builder, Weighted Product, Optimization.

---

### 1. Introduction

Accelerated growth in the transportation sector has a huge impact on human life. The development and progress in the transportation sector can be seen with the increasing number of motorized vehicles and the needs for modernization. The availability of an adequate road network system in an area as a structural form of spatial utilization patterns has a relationship to the development of the potential of the city or local area.

Cibeber District is one part of Lebak Regency, Banten Province, with an area of 400.96 km<sup>2</sup>, which consists of 22 villages or sub-districts. Cibeber Sub-district has a population of 57,815 people with a density of 165 (lebakkab.bps.go.id). The absence of gas stations built in Cibeber Subdistrict is not balanced with the high level of transportation use for the ease of mobility of residents. Taking data based on the questionnaire made from as many as 200 data samples from 22 villages in Cibeber District, 99% have private two-wheeled and four-wheeled vehicles, and 96% are the main means of transportation. The frequency of refueling vehicles in 1-3 days amounted to 56.2%, in 4-7 days to 31.8%, and above one week to 11.9%. A total of 98.5% answered the importance of gas stations in Cibeber Sub-district, with 76.1% citing reasons for cost savings, 19.9% needing certain fuels, and 4% citing other reasons. By considering the results of the questionnaire above, it can be concluded that the people in Cibeber Sub-District really need the existence of gas stations.

However, because Cibeber Sub-district has a very large area, it is necessary to select potential villages where gas stations will be built. To select potential villages by looking at several criteria, K-means clustering is used. K-Means clustering is one of the non-hierarchical data clustering methods that seeks to partition the available data into one or more groups (Aldino et al., 2020).

After the potential villages are obtained, the selection of potential locations is done using spatial modeling with Model Builder. A model is a workflow that assembles a sequence of geoprocessing tools, feeding the output of one tool to another as input. Geoprocessing models automate and document your spatial analysis and data management processes (Stefanidis et al., 2021).

Therefore, using the Geographic Information System (GIS) with clustering is expected to provide information to the public or government about potential land that is strategic for building gas stations and can help provide input to

investors to be able to establish a branch of gas stations in Cibeber District by considering the requirements and criteria based on the data obtained, it can also be a comparison for research using different methods and models.

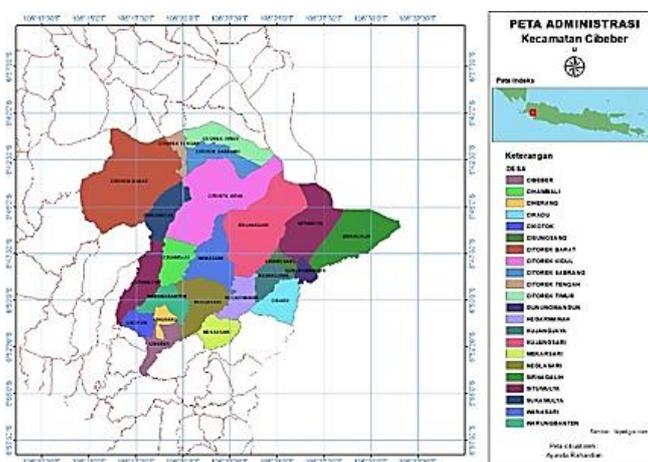
## 2. Materials And Methods

### 2.1. Material

This research conducts the process of determining the location of gas stations in Cibeber District using K-Means clustering and spatial modeling. The tools used are RStudio for K-Means clustering and ArcGIS for spatial modeling. The data used are population, number of vehicles, number of health facilities, number of schools, Cibeber District map, land use map, land slope map, river map, and road map. Cibeber District data can be seen in Table 1 and the Cibeber District Administration Map in Figure 1.

**Table 1: Cibeber Sub-district Data**

Village	Total Population	Number of Vehicles	Number of Health Facilities	Number of Schools
Kujangjaya	2756	975	0	5
Neglasari	3755	1031	1	5
Mekarsari	2437	712	0	7
Sukamulya	2753	684	0	4
Warungbanten	2638	853	0	5
Cikadu	2376	793	0	6
Sirnagalih	1961	587	0	3
Cihambali	1861	610	1	3
West Citorek	2417	822	0	3
Gunungwangun	1395	374	0	2
Wanasari	2209	695	1	8
Citorek Sabrang	1821	304	0	1
Ciherang	3269	986	0	4
Cibeber	4046	1159	1	6
Central Citorek	3525	925	0	5
Cisungsang	2049	643	1	3
Kujangsari	1775	574	0	3
Cikotok	2464	779	2	8
East Citorek	2391	649	0	0
Hegarmanah	1847	565	1	3
Citorek Kidul	1740	497	0	5
Situmulya	2388	886	0	2



**Figure 1: Cibeber Sub-district Map**

## 2.2. Methods

The methods used in this research are K-Means clustering, spatial modeling, and weighted product. The stages of using the method can be seen in Figure 2.

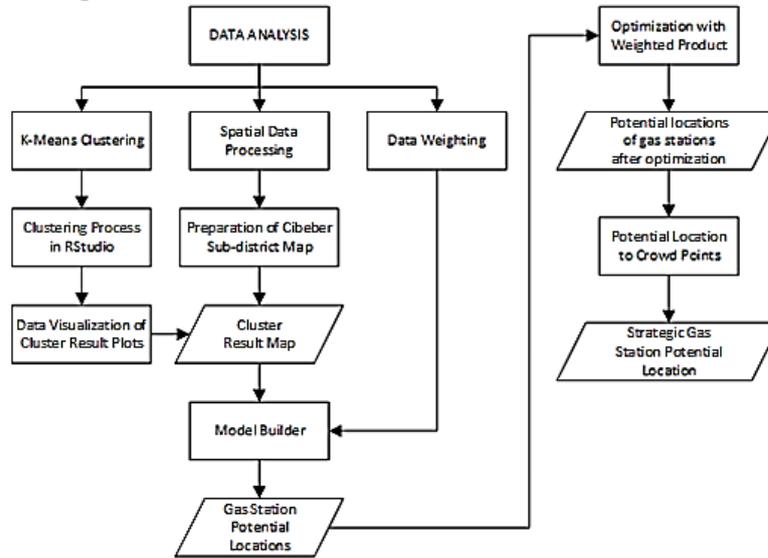


Figure 2: Flowchart of the method

### 2.2.1. K-Means Clustering with Rstudio

This K-Means clustering process uses the R Studio application to generate clusters. The K-means clustering algorithm is an iterative clustering analysis algorithm that has been considered an effective way to solve urban road planning problems by scholars over the past few decades; however, it is very difficult to determine the number of clusters and initialize the central cluster sensitively (Ran et al., 2021). The data used in this K-Means clustering with RStudio is the data in Table 1.

### 2.2.2. Spatial Modeling using ArcGIS

ArcGIS spatial analysis Model building is completed in Model Builder, which can connect data and spatial processing tools to handle complex GIS tasks. In the model generator, input data, output data, and corresponding spatial processing tools are represented by a visual graphical language (Jin et al., 2023). Therefore, researchers created this spatial modeling using Model Builder in order to obtain the potential locations of gas stations. The model builder can be seen in Figure 3.

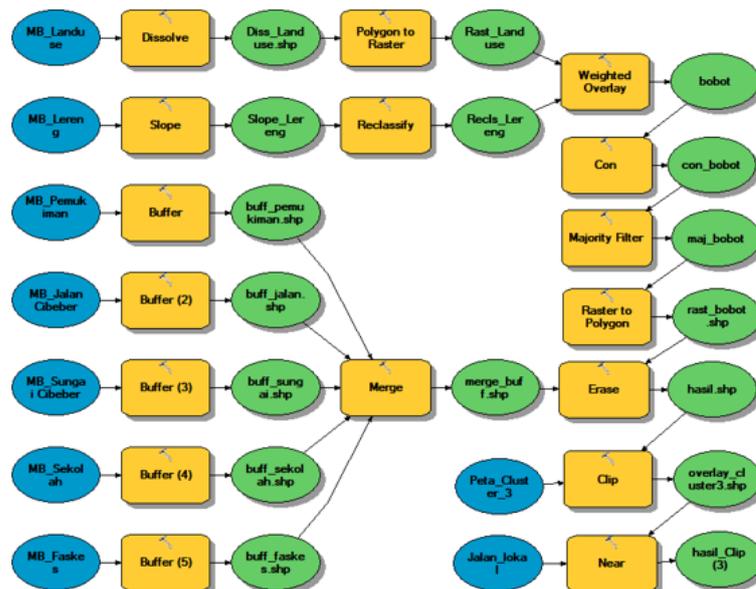


Figure 3: Model Builder Process

### 2.2.3. Weighted Product

The weighted product method is known to be more thorough because it obtains twice the valuation value, and the weighted value results in ranking order. This makes the ranking information more precise with specific results and has a faster average execution time (Khrisna et al., 2020). Weighted Product was used as an optimization of the potential location results obtained in the Model Builder. The weighted product stage can be seen in Figure 4.

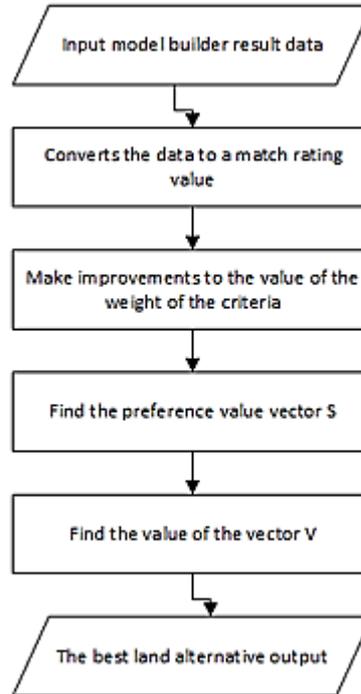


Figure 4: Weighted Product Process

## 3. Results And Discussion

### 3.1. K-Means Clustering with RStudio

Data processing with K-Means Clustering uses Cibeber District Data, namely data on population, number of vehicles, number of schools, and number of health facilities. Stages in RStudio for K-Means Clustering processing, starting with the installation of packages in RStudio, import files containing Cibeber District data in csv format. Furthermore, the selection of numeric data, data standardization to equalize data units, and finally the clustering process.

The results of K-Means clustering processing with RStudio obtained Cluster 1 members; the villages included in this cluster are Kujangjaya, Mekarsari, Sukamulya, Warungbanten, Cikadu, West Citorek, Central Citorek, Ciharang, and Situmulya. Then in Cluster 2, the villages included in this cluster are Sirnagalih, Cihambali, Gunungwangun, Citorek Sabrang, East Citorek, Cisungsang, Kujangsari, Hegarmanah, and Citorek Kidul. While only four villages in Cibeber Sub-district have high potential for the construction of gas stations included in Cluster 3, namely Neglasari, Wanasari, Cibeber, and Cikotok Villages. The data visualization of the cluster result plot can be seen in Figure 5.

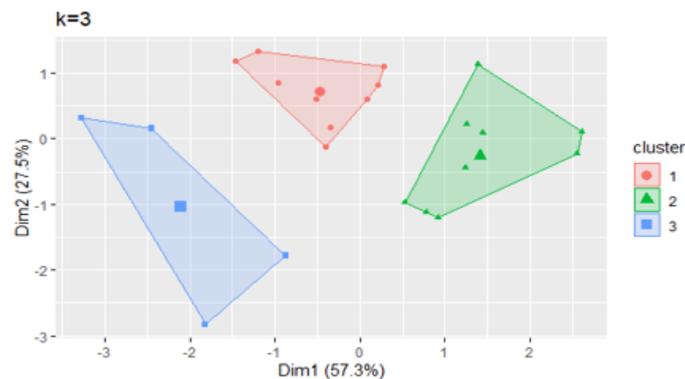
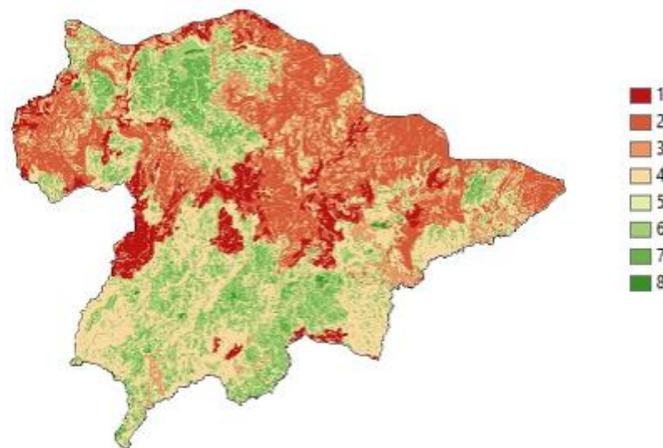


Figure 5: Data Visualization of Cluster Result Plot



### 3.2.3. Weighted Overlay Results

Weighted overlay analysis is a technique usually applied in geographic information systems to select a suitable system for a particular use (Zakaria et al., 2022). According to Papadopoulou and Hatzichristos (2019), it requires the analysis of many different factors in raster layers with varying scales of value and relative importance. Weighted Overlay is the result of two shapefiles, namely land use and slope, which have their respective weights. In this process, the result is an area that has the highest value and an area with a value of 1 to 9 from the weights that have been entered, where 9 is the highest value. However, there are no areas with a value of 9. This is because the weight of land use does not have a value of 9, and there are also limited areas that can fulfill it. The weighted overlay results can be seen in Figure 7.



**Figure 7:** Weighted Overlay Results

### 3.2.4. Conditional and Majority Filter Results

Conditional features are used to help control the desired output value; the higher the value selected, the better the resulting region. For the proposed region, a river threshold value of 100 m is used. A reference stream has been used for the stream network. All steps were executed using the conditional tool of Arc GIS (Arefin et al., 2020). Since the results of the weighted overlay are still on a scale of 1 to 8, and the highest value is chosen, areas with a value equal to or above 7 were selected using this conditional feature.

Furthermore, due to the large number of noise areas in the conditional results, the Majority Filter feature was used for better depiction. Class-specific sediment distribution maps are automatically filtered in ArcGIS using the "Majority Filter" to remove noise or artifacts from the ensemble model (Galvez et al., 2022). The majority filter uses a 4x4 grid cell window to determine the most common (majority) value to replace smaller cells or pixels in the raster image (ESRI 2021). In the Number of neighbors to use section, researchers used eight, and the replacement threshold used the majority.

### 3.2.5. Merge Buffer Result

Merge Buffer is the result of merging the shapefiles that have been buffered, namely the shapefiles of settlements, roads, schools, health facilities, and roads, with the classification in the table. To facilitate analysis, after cleaning, all individual visitor traces for each sampling location were merged using the Merge tool in ArcGIS, resulting in one merged shapefile for each sampling location (Sisneros-Kidd et al., 2021). This merge buffer serves to summarize the overlay processed shapefile with the results from the majority filter using the erase feature.

### 3.2.6. Erase Result

The Erase feature is good for deleting parts of the shapefile and taking only the parts needed (Apriyanti et al., 2021). The Erase feature is a feature to delete the shapefile resulting from the majority filter by the merge buffer shapefile. The results of this erase feature produce shapefiles that have potential locations for gas stations to be built. In this Erase result, 862 location points are obtained, which are spread across 22 villages in Cibeber Subdistrict.

### 3.2.7. Results of Overlay with Cluster Map

In the Erase result, this potential area is still within the scope of the Cibeber sub-district level and not yet in a cluster of villages suitable for the construction of gas stations. Therefore, an overlay process was carried out between

the potential map and the map of cluster three results, consisting of four villages. Clipping is done by overlaying polygons on one or more target features (layers) and extracting from the target feature (or features) only the target feature data located within the area outlined by the clip polygon (Okwudili, 2021). The results of potential location points obtained after overlaying the Cluster 3 map totaled 233 points.

### 3.2.8. Near Feature Result

Based on the partnership.pertamina.com website, the criteria for the location of gas stations on local land is at least 1000 m<sup>2</sup>, with a minimum face width of 20 m<sup>2</sup> and a minimum side width of 50 m<sup>2</sup> from the local road. To overcome this problem, researchers used the Near feature on local roads with a radius of 50m. The Near Tool in ArcGIS is used to determine the distance from each district to the nearest public health facility. The Near Tool determines the Euclidean distance between input features (district centers) and near features (public health facilities) (Shawky, 2016).

The potential location points obtained from this near feature totaled 21 points. The most potential locations were obtained in Cikotok Village with 7 potential location points, followed by Neglasari Village with 6 points, Cibeber Village with 4 points, and Wanasari Village with 4 points. Distribution of potential location points in Cluster 3 villages. Results of the Near feature can be seen in Figure 8.



Figure 8: Near Feature Results

### 3.2.9. Ground Truthing

Ground truthing is an activity carried out to assist in the classification process and improve the quality and accuracy of the Model Builder results. Candidate areas are ground truth boundaries, and grounding is correct if the most suitable area is the same as ground truthing (Liu et al., 2019). Ground truthing results from 21 potential location points; as many as 17 points are suitable for the construction of gas stations, and 4 points are not suitable for the construction of gas stations. The four points are declared unsuitable because they have a land slope above 15%.

### 3.2.10. Optimization with Weighted Product

After designing a weighted product, the weighted product process is carried out. Hypothesis testing is carried out using the weighted product method so that it can be known whether the results of the test are acceptable or not and the best alternative can be known (Fitria et al., 2019). In the first stage of determining the suitability rating table, 18 data points have a land value of 6 and a slope of 9, and 3 data points have a land value of 7 and a slope of 9, namely locations 3, 14, and 21. Then the next weight improvement is to calculate the S vector value later. Weight improvement is done by dividing the values of the criteria from land use and DEM by 10, and the weight of land use is 0.53 and DEM is 0.47.

Furthermore, calculating the S vector value is done by calculating the value of each data point multiplied by the weight of each criterion. For 18 data points that have land value 6 and slope 9, the S vector value is 7.259, and for 3 data points that have land value 7 and slope 9, the S vector value is 7.877. The total number of S vectors obtained from 21 data points is 154.29. The total of this S vector is to be used in the calculation of vector V. The calculation of vector V is done by multiplying the vector S value of each data point by the total vector S. Then the result of vector V in 18 data points is 0.047046 and in 3 data points is 0.051052.

From the results of the above calculations, the highest V vector value is 0.051052, and it can be concluded that the best alternative potential land candidates for gas stations are locations 3, 14, and 21. The results of optimization with weighted products can be seen in Table 3.

**Table 3: Optimization Results with Weighted Product**

No.	Coordinate Point	Location	Address	Compatibility
3	-06.8197124°, 106.3810571°		Tegalumbu Village, Wanasari Village	Suitable
14	-06.8539865°, 106.3089117°		Kp. Nagrak, Cikutok Village	Suitable
21	-06.8939220°, 106.2968399°		Kp. Cinangga, Cibeber Village	Suitable

### 3.2.11. Potential Locations to Crowded Points

Based on research conducted by Elsa Juliany with the title "Analysis of the Effect of Product Availability, Facilities, Service Quality, and Location on Consumer Purchasing Decisions in Purchasing Peralite Type Fuel Oil in Rantauprapat", location selection must also look at crowd factors and be easily accessible to consumers. Therefore, the location obtained from the weighted product is reviewed based on the distance from the crowd centers, such as markets, schools, and settlements. The selection of markets, schools, and settlements is based on the closest distance to the location. The crowd points can be seen in Table 4.

**Table 4: Point to Crowd**

No.	Coordinate Point	Distance to Crowd point		
		Settlements	School	Market
3	-06.8197124°, 106.3810571°	450m	240m	13.9km
14	-06.8539865°, 106.3089117°	1km	160m	950m
21	-06.8939220°, 106.2968399°	600m	260m	6.7km

Based on Table 4, it can be seen that the distance from the nearest settlement, school, and market to the location. The distances from the three crowd points are averaged to get the most strategic location close to the crowd point. The average calculation can be seen below.

Location

$$\begin{aligned}
 3 &= (450\text{m} + 240\text{m} + 13.9\text{km})/3 \\
 &= (450 + 240 + 13,900)/3 \\
 &= (14,590)/3 \\
 &= 4,863\text{m}
 \end{aligned}$$

$$\begin{aligned}
 14 &= (1\text{km} + 160\text{m} + 950\text{m})/3 \\
 &= (1,000 + 160 + 950)/3
 \end{aligned}$$

$$\begin{aligned}
 &= (2,110)/3 \\
 &= 703.3\text{m} \\
 21 &= (600\text{m} + 260\text{m} + 6.7\text{km})/3 \\
 &= (600 + 260 + 6,700)/3 \\
 &= (7,560)/3 \\
 &= 2,520\text{m}
 \end{aligned}$$

After the average results of the potential locations to the crowd point are completed, it is found that location no. 14 with coordinates  $-06.8539865^\circ$ ,  $106.3089117^\circ$ , which is located in Cikotok Village, is the closest potential location of the gas station to the crowd point with an average distance of 703.3m.

#### 4. Conclusion

Based on the research process that has been carried out using RStudio, the potential villages for the construction of gas stations consist of four villages, namely Cikotok, Cibeber, Neglasari, and Wanasari. These four villages have potential because they meet the four attributes in the highest clustering process, especially in Cikotok and Cibeber villages, which are mobility centers with markets and terminals. As for the results of spatial modeling by measuring based on the classification of land suitability of gas stations and weighting from experts on the criteria, in Cibeber Sub-district there are 862 potential location points, and within the scope of potential villages, namely four villages, there are 233 potential location points.

However, because gas stations must be located at least within the local road area, locations within a 50-meter radius of the local road were selected, and 21 potential locations were obtained in four potential villages. In this study, an accuracy test was carried out with Ground Truthing by taking coordinate points from the spatial modeling results. After testing the accuracy with ground truthing, the results were optimized using weighted products to get the best land results. The criteria used in the weighted product calculation are slope and land use, according to experts. The results obtained three best location points, namely Kp. Tegalumbu located in Wanasari Village, Kp. Nagrak located in Cikotok Village, and Kp. Cinangga located in Cibeber Village.

#### References

- Aldino, A. A., Darwis, D., Prastowo, A. T., & Sujana, C. (2021). Implementation of K-Means Algorithm for *Clustering Corn Planting Feasibility Area in South Lampung Regency*. *Journal of Physics: Conference Series*, 1751(1). <https://doi.org/10.1088/1742-6596/1751/1/012038>
- Apriyanti, D, Eng, M., Joko Hartadi, I. M., & Waskito Putro, R. S. (2021). Impact and Countermeasures of Abrasion Using Satellite Imagery. In *Scientific Journal of Geomatics Engineering IMAGI* (Vol. 1).
- Arefin, R., Mohir, M. M. I., & Alam, J. (2020). Watershed prioritization for soil and water conservation aspect using GIS and remote sensing: PCA-based approach at northern elevated tract of Bangladesh. *Applied Water Science*, 10(4). <https://doi.org/10.1007/s13201-020-1176-5>
- ESRI (2021) *Majority Filter* (Spatial Analyst)-ArcMap | Documentation. Environmental Systems Research Institute, Inc. website. <https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/Majority-filter.htm>. Accessed May 24, 2023
- Fitria, R., & Rakhmah, S. N. (2019). Application of the *Weighted Product* Method to Analyze Digital Transaction Payment Selection Factors in Online Ojek Applications. *Jurnal Teknika*, 11 (No. 2).
- Galvez, D. S., Papenmeier, S., Sander, L., Bartholomä, A., & Wiltshire, K. H. (2022). Ensemble mapping as an alternative to baseline seafloor sediment mapping and monitoring. *Geo-Marine Letters*, 42(3). <https://doi.org/10.1007/s00367-022-00734-x>
- Jin, F., Wang, C., Zhou, X., & Wang, C. (2023). Application of ArcGIS Spatial Analysis Techniques on Topography Teaching. In *Proceedings of the 2nd International Conference on Internet, Education and Information Technology (IEIT 2022)* (pp. 413-418). Atlantis Press International B.V. [https://doi.org/10.2991/978-94-6463-058-9\\_67](https://doi.org/10.2991/978-94-6463-058-9_67)
- Juliany, E. (2021). Analysis of the Effect of Product Availability, Facilities, Service Quality and Location on Consumer Purchasing Decisions in Purchasing Peralite Type Fuel Oil in Rantau Prapat. *Business Economics Management and Accounting (EBMA)*, Volume 2, Number 1. <https://doi.org/https://doi.org/10.36987/ebma.v2i1.3142>

- Khrisna Wardhani, A., & Lutfina, E. (2020). Application Culinary decision support system in Kudus city with *Weighted Product* method based on mobile phone. *Journal of Computer Science and Engineering (JCSE)*, 1(1), 10-16. <https://doi.org/10.36596/jcse.v1i1.17>
- Liu, X., Wang, Z., Shao, J., Wang, X., & Li, H. (2019). *Improving Referring Expression Grounding with Cross-modal Attention-guided Erasing*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1950-1959.
- Okwudili John, U. (2021). Mapping and Analysis of the rate of compliance of Egbu-Uratta road right of way in Owerri, Nigeria. *International Journal of Advances in Engineering and Management (IJAEM)*, 3, 856. <https://doi.org/10.35629/5252-0311856861>
- Papadopoulou CA. and Hatzichristos T. 2019. A GIS-based spatial multicriteria decision analysis: Crisp and Fuzzy methods. AGILE, 2019-Limassol, June 17-20.
- Ran, X., Zhou, X., Lei, M., Tepsan, W., & Deng, W. (2021). A novel K-means *clustering* algorithm with a noise algorithm for capturing urban hotspots. *Applied Sciences (Switzerland)*, 11(23). <https://doi.org/10.3390/app112311202>
- Seja, S. M., Mbilinyi, B., & Hieronimo, P. (2022). Integrated Analytical Hierarchical Process and Geographical Information System for Allocation of Compatible Land Uses along Uluguru Mountain Slopes. *Current Urban Studies*, 10(04), 787-820. <https://doi.org/10.4236/cus.2022.104045>
- Shawky MANSOUR (2016) Spatial analysis of public health facilities in Riyadh Governorate, Saudi Arabia: a GIS-based study to assess geographic variations of service provision and accessibility, *Geo-spatial Information Science*, 19:1, 26-38, DOI: 10.1080/10095020.2016.1151205
- Shehab, Z. N., Jamil, N. R., Aris, A. Z., & Shafie, N. S. (2021). Spatial variation in the impact of landscape patterns and land use on water quality across an urbanized watershed in Bentong, Malaysia. *Ecological Indicators*, 122. <https://doi.org/10.1016/j.ecolind.2020.107254>
- Shokr, M.S.; Abdellatif, M.A.; El Baroudy, A.A.; Elnashar, A.; Ali, E.F.; Belal, A.A.; Attia, W.; Ahmed, M.; Aldosari, A.A.; Szantoi, Z.; Jalhoum, M.E.; Kheir, A.M.S. Development of a Spatial Model for Soil Quality Assessment under Arid and Semi- Arid Conditions. *Sustainability* **2021**, 13, 2893. <https://doi.org/10.3390/su13052893>
- Sisneros-Kidd, A. M., D'Antonio, A., Monz, C., & Mitrovich, M. (2021). Improving understanding and management of the complex relationship between visitor motivations and spatial behaviors in parks and protected areas. *Journal of Environmental Management*, 280. <https://doi.org/10.1016/j.jenvman.2020.111841>
- Stefanidis, S., Chatzichristaki, C., & Stefanidis, P. (2021). AN ArcGIS TOOLBOX FOR ESTIMATION AND MAPPING SOIL EROSION. In *Article in Journal of Environmental Protection and Ecology*. [https://www.researchgate.net/publication/350411893\\_ArcEPM](https://www.researchgate.net/publication/350411893_ArcEPM)
- Zakaria, Y. S., Shaibu, A.-G., & Baatuuwie, B. N. (2022). Assessment of Physical Suitability of Soils for Vegetable Production in the Libga Irrigation Scheme, Northern Region, Ghana Using the Analytic Hierarchy Process and Weighted Overlay Analysis. *Turkish Journal of Agriculture - Food Science and Technology*, 10(8), 1395-1403. <https://doi.org/10.24925/turjaf.v10i8.1395-1403.5004>