

Cluster analysis for determining distribution center location

Dyah Lestari Widaningrum^{*}, Aditya Andika, Richard Dimas Julian Murphiyanto

Industrial Engineering Department, Faculty of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480

dwidaningrum@binus.edu

Abstract. Determination of distribution facilities is highly important to survive in the high level of competition in today's business world. Companies can operate multiple distribution centers to mitigate supply chain risk. Thus, new problems arise, namely how many and where the facilities should be provided. This study examines a fast-food restaurant brand, which located in the Greater Jakarta. This brand is included in the category of top 5 fast food restaurant chain based on retail sales. There were three stages in this study, compiling spatial data, cluster analysis, and network analysis. Cluster analysis results are used to consider the location of the additional distribution center. Network analysis results show a more efficient process referring to a shorter distance to the distribution process.

1. Background

High level of competition in today's business world has forced companies to rethink their methods in maintaining their competitiveness. Only companies that can deliver products or services on time with the right quality and the right price can survive in this era. One of the ways to maintain their competitiveness is by managing their supply chain risks. Determination of the location of production and distribution facilities is very important in supply chain risk management. Companies can reduce risk by operating multiple facilities in spatially dispersed locations [1]. This paper will elaborate research findings on distribution center location determination.

The research is conducted in the foodservice industry. Food service industry is a growing industry in Indonesia. Data from Indonesia Central Bureau of Statistics show that Gross Domestic Product growth from Food Service Industry is always positive for the last five years (2012 – 2016) [2]. In addition, this industry has a high distribution of outlets. This made a suitable condition as an object of research on spatial data. We use a brand as a case study for this research. This brand is included in the category of top 5 fast food restaurant chain based on retail sales [3]. Nonetheless, the methodology offered in this research can be applied to various other industries, especially those with high distribution activities. The company at the moment only has one distribution center in the Special Capital Region of Jakarta. All shipment of intermediate products (uncooked processed food) to all stores originates from this distribution center. By operating only one distribution center, the company is prone to supply chain risks, especially environmental risks of Jakarta, such as traffic congestion and flooding. A research by INRIX ranks Jakarta as the 19th most congested city in the world and the most congested city in Indonesia [4]. Those risks can create losses for the company, such as late delivery of stores and damaged food. Therefore, it is necessary for the company to create multiple distribution centers as a way to mitigate those risks.



2. Facility Location

The facility location problem (FLP) is the main problem when designing supply and service network of a company [5]. The pioneer of FLP was Weber in 1909 who studied the facility location to minimize the distance between the factory and the customer [6]. There are many factors when deciding new locations, such as available budgets, corporate strategies, current facility locations, competitor facility locations, markets, demographics (ex. age, gender, population), consumer targets, logistics and traffic conditions at candidate locations [7]. Now, various kinds of FLP have been studied, such as warehouses in a supply chain [8, 9].

There are methods to find the best location for second and third warehouse using spatial analysis and network analysis, which place the right warehouse location can be a long-term solution for the company [10]. In the context of warehouses, avoiding back order and delivery lead time is the main goal of increasing the number of warehouses apart from decreasing transportation costs and environmental requirements in terms of emission gas (CO₂) reduction [11, 12].

There are previous studies in determining warehouse's location, Maharjan & Hanaoka used a simplex algorithm with branch and bound [13]. Wang, Fu, Chen, & Zhou used a genetic algorithm for their linear model [14]. Monthatipkul used nonlinear program based on the load distance technique [15]. Fuente and Lozano used cluster analysis in Spain [16]. Aghezzaf used Lagrangian relaxation decomposition [17]. Abo-Elnaga, El-Sobky, & Al-Naser introduced a trust region algorithm for their nonlinear optimization [18]. Hua, Hu, & Yuan used artificial intelligence technique called particle swarm optimization [19].

3. Research Methods

Jakarta as one of the 30 largest urban agglomerations [20], which calls as megacity [21], plays roles as economic wealth and employment driver [22]. It has become hubs of commerce [23] and needs to find a strategy to be more efficient on resource allocation in the complex linkages [22]. Based on this phenomenon, Greater Jakarta was chosen related to the problems studied in this research.

This study examines a fast-food restaurant brand in Indonesia, which is included in the top 5 global brands in chained fast food based on retail sales [3]. This brand has main distribution center to serve all outlets in Greater Jakarta and some distribution centers (warehouses) in other provinces. The determination of the distribution center branches in the Greater Jakarta area will be the main objective of this research.

There were three stages in this study. The first stage of this study was compiling spatial data. The spatial data collection is consist of borderline and road network. These were obtained from Geospasial untuk Negeri, Indonesia's geospasial portal (<http://tanahair.indonesia.go.id/portal/landingpage>), managed by BIG – Badan Informasi Geospasial. The fast-food restaurant locations are geocoded based on information from this brand website. The second stage was cluster analysis. This analysis is conducted to determine natural groupings [24]. It was grouping restaurants location data. Results from Cluster analysis are used to consider the location of additional distribution centers. The third stage was network analysis, to observe the difference between current condition (with one distribution center) and propose condition (propose some additional distribution centers). These three stages were analyzed using ArcGIS® Desktop 10.0 for Windows®.

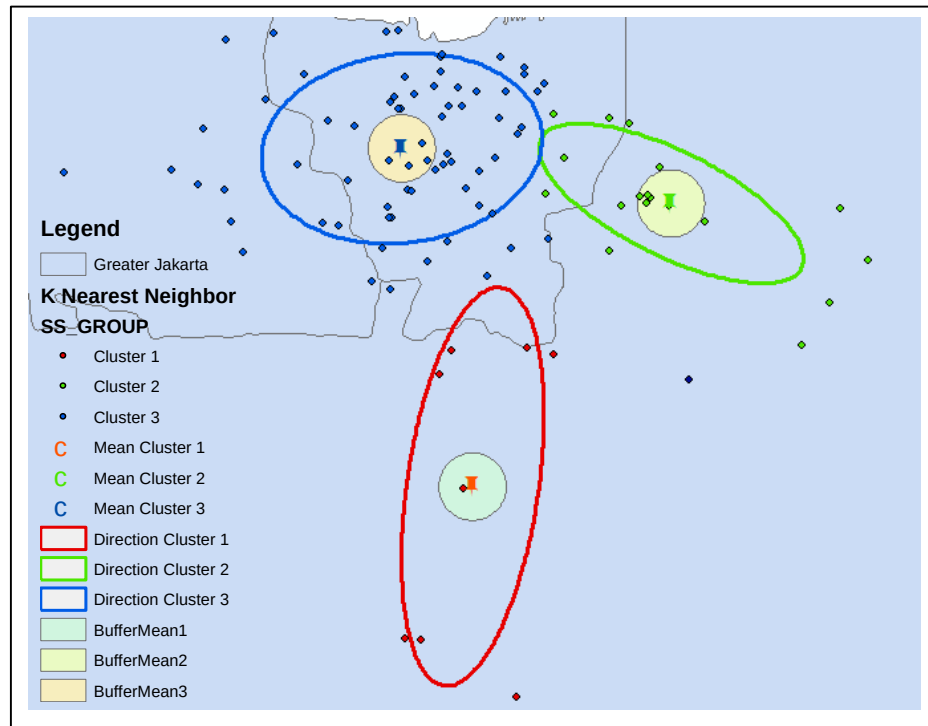
4. Findings and Discussion

The Greater Jakarta area extends along with the urban agglomeration. According to the spread of this object study, these fast food restaurants were spread in Central Jakarta, North Jakarta, South Jakarta, East Jakarta, West Jakarta, Bogor, Depok, Tangerang, and Bekasi, as can be seen in Figure 1.

The number of fast food restaurants that became the main object in the study were 96 outlets. 65% of restaurant locations are located in Jakarta. These locations are used as input features for grouping analysis. Output feature class which was a result from grouping analysis was a new integer field called SS_Group, which define a specific group for each record. For this preliminary study and according to the spread of the location, it was decided to divide it into 3 clusters.

Table 1. Variable-Wise Summary

Variable	Mean	St. Dev.	Min	Max	R ²
X	-6.241610	0.095669	-6.651329	-6.116170	0.608186
Y	106.840304	0.114535	106.520950	107.167997	0.555199

**Figure 1.** Clustered Location and Standard Deviation Ellipse of fast food restaurant

Grouping Analysis tool from ArcGIS® Desktop 10.0 for Windows® computed the mean, standard deviation, minimum and maximum value for each variable. Coordinates X and Y become the variable in this study, considering the clustering factor is based on the proximity of location between one point to another. R² value was also computed to analyze the appropriateness of variables on separating the points. R² value for each variable was greater than 0.5, which indicates that these variables have divided the location of the restaurant into 3 clusters effectively [24]. Calinski-Harabasz Pseudo F-statistic was also computed by this program to exercise the grouping effectiveness, and it has shown that 15 clusters have the highest Pseudo F-statistic. This number found as a non-manageable number of segments. The researchers have taken the stopping rule to determine the number of clusters based on strategic and tactical perspective [25]. 3 clusters are considered the most representative, given the background of the problem. The addition of two new locations of the distribution center is the most reasonable decision, considering the related resources needed for realization.

The three clusters tend to have directional trends, based on standard deviational ellipse as can be seen in Figure 1. These orientations of location distribution were calculated based on each coordinate's standard deviations and its mean center [26]. The first cluster has a straight directional to the south, as the second cluster to the east. It found that the third cluster has no specific direction, seemingly distributed in all directions. The third cluster has 69 members, while the second cluster has 19 members and the first cluster has 8 members. Although the three clusters have a disproportionate amount, the coverage of the three clusters based on Thiessen polygon is relatively similar, as can be seen in Figure 2.

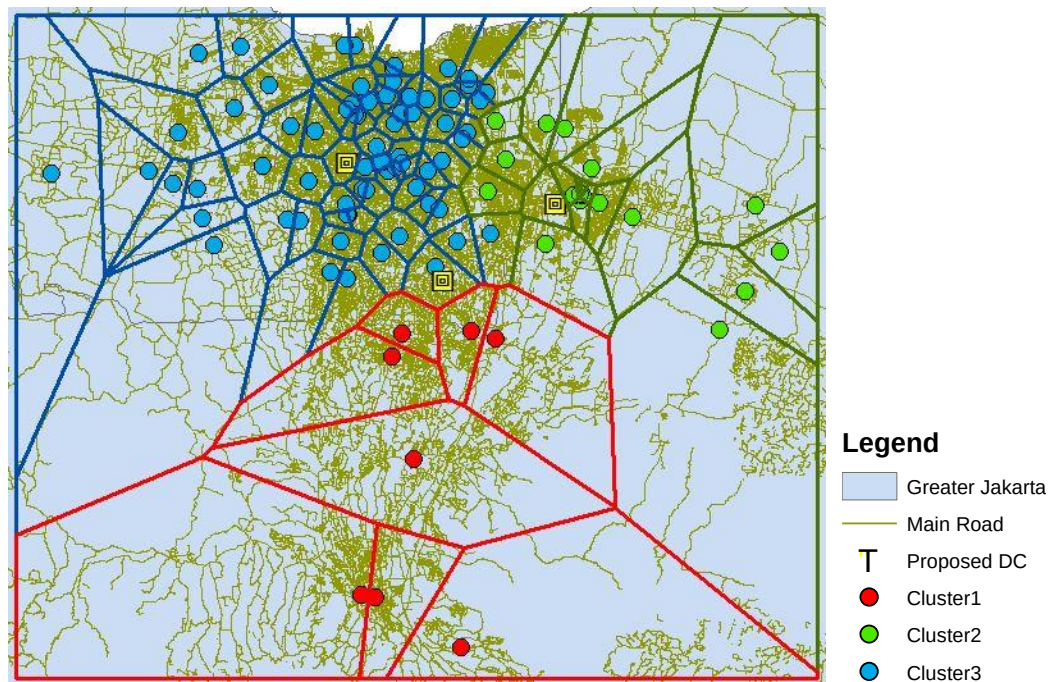


Figure 2. Thiessen Polygon Coverage for Each Restaurant

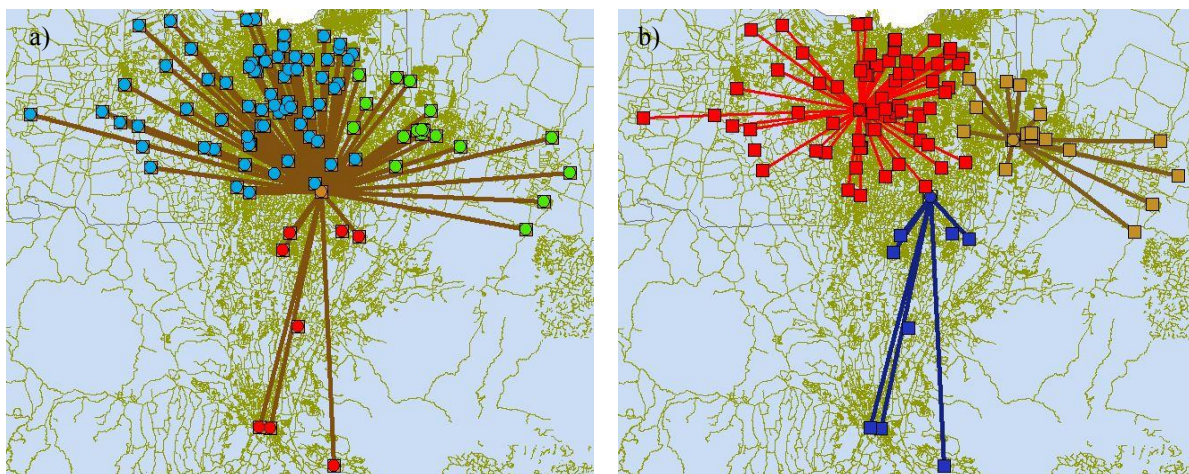


Figure 3. Network Analysis; a) Distribution from 1 Distribution Center;
b) Distribution from 3 Distribution Centers

Figure 3 illustrates the straight lines between origin and destinations, with only one distribution center and with three distribution centers. The total distance that must be taken in the distribution process becomes shorter with the addition of distribution center. Reduced distribution distance will save energy and have a positive impact on the environment. To optimize an eco-friendly system in the distribution process, we also have to incorporate diverse variable in dynamic system. Further research can be done to compare costs as well as various advantages and disadvantages.

5. Conclusion

The results show a mechanism that practitioners can apply in dividing the region, by cluster analysis of spatial data. The development of Geographic Information System has an impact on the ease of conducting cluster analysis of spatial data. The availability of spatial data allows various realm to perform this analysis.

The determination of the number of clusters to be one of the important issues in this mechanism, as explained by Hair, et al. [25]. As in the case of location determination to become an additional hub of distribution, the determination of the number of clusters can not only be based on recommendations of computational results but must be based on consideration of operational ease and availability of resources.

Network analysis performed shows a more efficient process referring to a shorter distance to the distribution process. Further research can compare the origin of the cost-matrix with a various number of distribution centers, as well as relate it to various other influencing factors, such as traffic congestion, volume, and frequency distribution.

Acknowledgment

This work was supported by Bina Nusantara University Research Grant 2017.

References

- [1] J. H. Dorfman, *Economics and Management of the Food Industry*. New York: Routledge 2014.
- [2] Badan Pusat Statistik, "Tabel Dinamis," 2017.
- [3] Minister of Agriculture and Agri-Food, "Foodservice Profile - Indonesia," Her Majesty the Queen in Right of Canada, Ottawa, Canada 2016.
- [4] G. Cookson and B. Pishue, "INRIX Global Traffic Scorecard 2016," INRIX Research 2017.
- [5] L. Tang, C. Zhu, Z. Lin, J. Shi, and W. Zhang, "Reliable Facility Location Problem with Facility Protection," *PLoS ONE*, vol. 11, pp. 1-24, 2016.
- [6] R. G. Askin, I. Baffo, and M. Xia, "Multi-commodity warehouse location and distribution planning with inventory consideration," *International Journal of Production Research*, vol. 52, pp. 1897-1910, 2013.
- [7] B. Bozkaya, S. Yanik, and S. Balcisoy, "A GIS-Based Optimization Framework for Competitive Multi-Facility Location-Routing Problem," *Networks and Spatial Economics*, vol. 10, pp. 297-320, 2010.
- [8] Z.-J. M. Shen, R. L. Zhan, and J. Zhang, "The Reliable Facility Location Problem: Formulations, Heuristics, and Approximation Algorithms," *INFORMS Journal on Computing*, vol. 23, pp. 470-482, 2011.
- [9] S. C. K. Chu and L. Chu, "A modeling framework for hospital location and service allocation," *International Transactions in Operational Research*, vol. 7, pp. 539-568, 2000.
- [10] M. Musavi and A. Bozorgi-Amiri, "A multi-objective sustainable hub location-scheduling problem for perishable food supply chain," *Computers & Industrial Engineering*, 2017.
- [11] S. Pan, E. Ballot, and F. Fontane, "The reduction of greenhouse gas emissions from freight transport by pooling supply chains," *International Journal of Production Economics*, vol. 143, pp. 26-94, 2010.
- [12] H.-Y. Mak and Z. J. M. Shen, "Risk diversification and risk pooling in supply chain design," *IIE Transactions*, vol. 44, pp. 603-621, 2012.
- [13] R. Maharjan and S. Hanaoka, "Warehouse location determination for humanitarian relief distribution in Nepal," in *World Conference on Transport Research*, Shanghai, 2017, pp. 1151-1163.
- [14] B. Wang, X. Fu, T. Chen, and G. Zhou, "Modeling Supply Chain Facility Location Problem and Its Solution Using a Genetic Algorithm," *Journal of Software*, vol. 9, pp. 2335-2341, 2014.
- [15] C. Monthatipkul, "A non-linear program to find an approximate location of a second warehouse: A case study," *Kasetsart Journal of Social Sciences*, vol. 37, pp. 190-201, 2016.
- [16] D. de la Fuente and J. Lozano, "Determining warehouse number and location in Spain by cluster analysis," *International Journal of Physical Distribution & Logistics Management*, vol. 28, pp. 68-79, 1998.
- [17] E. Aghezzaf, "Capacity Planning and Warehouse Location in Supply Chains with Uncertain Demands," *Journal of the Operational Research Society*, vol. 56, pp. 453-462, 2005.

- [18] Y. Abo-Elnaga, B. El-Sobky, and L. Al-Naser, "An active-set trust-region algorithm for solving warehouse location problem," *Journal of Taibah University for Science*, vol. 11, pp. 353-358, 2017.
- [19] X. Hua, X. Hu, and W. Yuan, "Research optimization on logistics distribution center location based on adaptive particle swarm algorithm," *International Journal for Light and Electron Optics*, 2016.
- [20] Department of Economic and Social Affairs, Population Division, United Nations. World Urbanization Prospects: The 2014 Revision [Online].
- [21] Department of Economic and Social Affairs, Population Division, United Nations. (2016, The Worlds Cities in 2016 - Data Booklet.
- [22] United Nations Human Settlements Programme, "World Cities Report 2016," United Nations, Nairobi, Kenya 2016.
- [23] Department of Economic and Social Affairs, Population Division, United Nations. (2014, Population Facts. *Our urbanizing world*.
- [24] ESRI (Geographic Information System Company). (2017, How Grouping Analysis Works.
- [25] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, Seventh ed.: Pearson Prentice Hall, 2010.
- [26] D. W. Allen, *GIS Tutorial 2: Spatial Analysis Workbook*. California: ESRI Press, 2016.