



Warehouse Location Optimization with Clustering Analysis to Minimize Shipping Costs in Indonesia's E-Commerce Case

Alvin Trianto Atmojo¹, Manahan Siallagan², Riezan Syauqi Fanhas³

^{1,2,3} School of Business and Management, Institut Teknologi Bandung, Jl. Ganesha No. 10, Bandung, 40132, West Java, Indonesia

ABSTRACT: Due to the growth of the Internet economy, the popularity of online shopping has escalated in recent years. One of the largest e-commerce enterprises in Indonesia, PT. S, is the subject of the research in this article. Instead of typical e-commerce, where anybody may start a store, PT. S is concentrating on social commerce, which makes use of several resellers to offer hand-picked SME brand partners. PT. S must expand the market for inter-island or non-java-to-non-java transactions to fulfill its vision. However, PT. S will have logistical difficulty completing this job. The business used performance indicators to keep track of the logistics process' vision and mission. Gross merchandise value, pickup time service level, and shipping time service level are a few of the performance indicators that pertain to logistics. The process of managing the supply chain will become more complex as a result of the opening of the new warehouse, and the business will need to maximize its use of various selling channels, logistical services, and supply chain management. With the aid of clustering analysis, which assesses demand similarity and proximity, the enterprise can locate a new warehouse. Durairaj and Kasinathan developed the framework template for this study in 2015. Based on the case study, literature review, and clustering method framework, the framework will be modified in several ways, particularly clustering analysis. The alteration concerns framework-integrated theories as an input and as a data source. According to the simulation's findings, shipping costs per kilogram decreased by about 35% for five clusters. But if the corporation does not have a problem with the number of warehouses, according to the simulation's findings, because the cost of transportation will go down as the number of clusters increases, the number of warehouses can be expanded to more than five.

KEYWORDS: Clustering analysis, e-commerce, Optimization, Logistics, Warehouse location.

INTRODUCTION

Over the past few years, internet shopping has exploded in popularity thanks to the expansion of the Internet economy. Many consumers utilize the Internet for research purposes before making a purchase, either online or in a physical store, by reading reviews, comparing pricing, and checking out the newest offerings (Khan, A. G. 2016). The most common type of online business in the e-commerce industry is the online platform, which acts as a reseller for some products and a marketplace for others. Logistics is one of the most expensive processes in the e-commerce business and is widely seen as the primary driver of competitive advantage. In the marketplace model, suppliers typically handle logistics; however, some e-commerce sites have started a program to let other sites use their warehouses and delivery drivers. (Abdul Hafaz Ngah, 2021).

In this journal, the research focuses on PT. S, which is one of the biggest E-commerce companies in Indonesia. PT. S, is focusing on social commerce, which utilizes many resellers to sell our curated SME brand partners, rather than common e-commerce, where everyone can open shop. PT. S has a mission to grow the economy in tier 2 and tier 3 cities. Such as Makassar, Denpasar, and Semarang for tier 2, and Magelang, Prabumulih, and Bangli for tier 3. The majority of Tier 2 and Tier 3 cities are not on Java Island. Based on its vision and mission, PT. S needs to grow the market for inter-island or non-java to non-java transactions. But PT. S will face a logistical challenge to achieve this mission.

The company used performance indicators to track this vision and mission. Performance indicators will help the company identify areas where and how much they can improve, such as customer service or employee satisfaction, which can't be reflected in traditional financial reports (Kaplan & Norton, 2001). Some of the performance indicators that relate to logistics are gross merchandise value, pickup time service level, and shipping time service level (Atmojo et Al., 2023). The company also measures its logistics performance with shipping costs per order in kilograms. This measurement will indicate how much the company and users will pay for its logistics in the delivery of the package from PT. S Vendor to its customers using third party logistics (3PL)



providers. In 2022, PT. S was aiming to reduce shipping costs per order by up to 30%, but it only achieved 75% of the target. This condition will lead to an increase in company budgets for free shipping marketing costs and a reduction in another expanding budget. Management has recommended a number of potential solutions, including the establishment of a second warehouse capable of housing a wide variety of products and nearing customer potential based on the expectation of reducing cost and increasing shipping time service levels, especially in the three big islands of Sumatra, Borneo, and Celebes. The company can't decrease the shipping cost due to the third party related cost.

Opening the new warehouse will lead to complexity in the supply chain process; the company will need to optimize the combination of selling modes, logistic services, and a well-managed supply chain (Xuelian Qina, 2021). According to the CIO, the standards of supply chain management are planning, sourcing, manufacturing, transportation, distribution, and return. The company plans a new warehouse location that meets demand and minimizes shipping costs. The technique that will be used is clustering analysis. Clustering analysis is a data mining technique that aims to group similar objects into clusters based on some similarity or distance measure (Rodriguez, 2019). This research will help the company find a new warehouse location using clustering analysis, which measures demand similarities and distance.

Problem Formulations:

1. Will warehouse location optimization have an influence on minimizing shipping costs per kilogram for the customer in e-commerce?
2. How much is the reduction in shipping costs per kilogram when optimizing the warehouse location?
3. Where are the most ideal warehouse locations between Sumatera, Borneo, and Celebes that the business might take into account?

LITERATURE REVIEW

Warehouse Location Optimization

Numerous industries, including manufacturing and shipping, use warehouses as crucial business facilities in the supply chain for storing finished goods and basic materials. The location of a warehouse refers to the site that is optimal for the company and will benefit them. This location was chosen with consideration, taking into consideration all relevant criteria. This is a complex process, and a single mistake in selecting a location could result in the bankruptcy of the business (iThink Logistic, 2020). There are three main factors determining the location of the warehouse: Infrastructure, government, and Market (Singh, 2018).

Warehouse location optimization for omnichannel, or online retail. Based on the interaction of product category, online demand percentage, and the elasticity of online demand based on distance, warehouses can be placed wherever demand is highest, with the best option being "Ship from Warehouse" rather than the other three distinct types of e-commerce product distribution centers (Millestein, 2021). The cluster method is able to determine the warehouse's location, capacity, and size. The technique was used in Spain by determining the total demand, coordinates of towns, and transport costs (Fuente, 1997). The method is also used to determine the location of a waste recycling center by analyzing the main attributes to be considered, such as benefit at the economic, social, environmental, and management levels (Jiang, 2015).

There are several studies that use Indonesia's geographical area and clustering analysis to determine the center of distribution. Determining the export coffee warehouse distribution center in West Java using Model Optimization (P-Median Method) classifies West Java into 4 clusters that consist of sub-districts and cities (Reha et al., 2020). And determining distribution centers by employing cluster analysis to pinpoint where to put a food distribution facility in the broader Jakarta metro region (Lestari, 2017).

Numerous studies have been conducted to find the warehouse location through clustering analysis. However, there's a limited study that specifies Indonesia's wide geographical area, especially Sumatera, Borneo, and Celebes. Also, the utilization of third-party logistics as a logistic actor for the company makes a difference in this research. We can conclude that the main factors determining warehouse location are distance, market, and shipping price. Those factors will be used as features in clustering analysis.

Clustering Analysis

Clustering analysis is a data mining technique that aims to group similar objects into clusters based on some similarity or distance measure. It is widely used in various fields, such as biology, marketing, image processing, and social network analysis, to



discover patterns and structures in data (Rodriguez, 2019). Clustering analysis can be divided into different types, such as partitioning, hierarchical, density-based, grid-based, and model-based methods. Each type has its own advantages and disadvantages and may be suitable for different scenarios and applications. One of the most popular and simple clustering methods is the k-means algorithm, which partitions the data into k clusters by minimizing the sum of squared distances from each object to its cluster center (Qina, 2019).

K-Means Clustering

K-means clustering is a method of vector quantization that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster center or cluster centroid), serving as a prototype of the cluster. One popular method for initialization is 'k-means++' which selects initial cluster centroids using sampling based on an empirical probability distribution of the points' contribution to the overall inertia. This technique speeds up convergence (Arthur et al., 2007)

The objective of K-means clustering is to minimize the within-cluster sum of squares (WCSS) (i.e., variance). Formally, the objective is to find:

$$\arg_S \min ||x - \mu_i||,$$

where S is the set of observations, k is the number of sets of predictors or clusters, μ_i is the mean of points in S_i , and $||x - \mu_i||$ is the Euclidean distance between x and μ_i .

Weighted K-Means Clustering

Weighted k-means clustering is a variation of the k-means algorithm that takes into account the relative importance of each data point. This is done by assigning a weight value to each data point, which is used when calculating the cluster centroids. By assigning small weight values to outlier data points, k-means will form clusters that are robust to these extreme values (Dubey et Al., 2017). To perform weighted k-means clustering, we need to make a minor tweak to the way the cluster centroids are calculated after each iteration. Instead of using the mean to calculate the centroids, we use the weighted mean (Dubey et Al., 2017).

Web Scraping

Web scraping, sometimes referred to as web extraction or harvesting, is a method for extracting information from the World Wide Web (WWW) and storing it in a filesystem or database for subsequent retrieval or analysis. Web data is frequently scraped using HTTP (Hypertext Transfer Protocol) or a web browser. A human can do this manually, or a bot or web crawler can do it automatically (Mooney et al. 2015; Bar-Ilan 2001).

Obtaining online resources and then extracting the needed information from them are the two consecutive processes that make up the data scraping process. A web scraping application specifically begins by creating an HTTP request to get resources from a selected website. This request may be structured as a URL with a GET query or as a portion of an HTTP message with a POST query. The requested resource will be obtained from the website and then delivered back to the given web scraping application when the targeted website has successfully received and processed the request (Zhao, 2017).

The use of the JavaScript programming language was necessary to make connections between the computer and the PT.X website to scrape the data. The connection process between Javascript and the PT.X website was made through an Application Programming Interface (API). Any software having a specific function is referred to as an application when discussing APIs. Interfaces can be compared to a service agreement between two programs. This agreement specifies the requests and replies that the two parties will use to communicate (Frye, 2023).

RESEARCH METHODOLOGY

The process of this study's research technique begins with problem identification. Quantitative research techniques were employed. In the first phase, the author will identify the business issues facing the company, create a problem statement, and define the limitations of the research question. After surveying the relevant literature, a solution and strategy for implementing it in a corporate context are suggested. An informal questionnaire will be used to collect information for the purposes of gaining understanding and validating the business issue from the perspective of stakeholders. An unstructured interview with those in charge



of the decision-making processes currently in place over which warehouses to use was conducted to understand the current situation, management plan, and expectations.

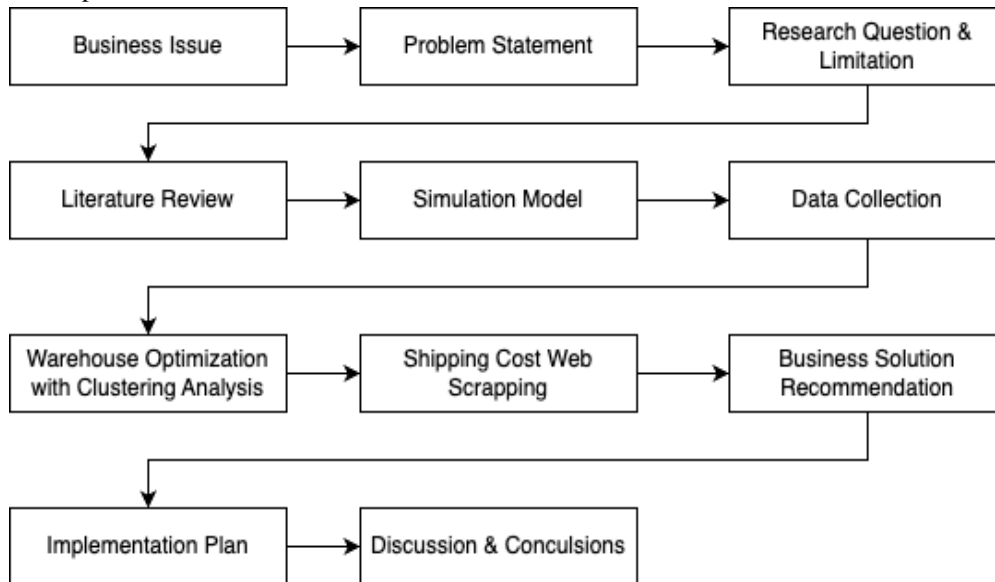


Figure 1. Research Methodology Flow

Secondary data is applied to do a cluster analysis of the market's potential customers. Some valid sources, such as the government's civil registration (Kementerian dalam negeri, Kemendagri) and Indonesia's central Agency on statistics (Badan Pusat Statistik, BPS), can be used to gain information about the location attributes. Company data for performance results, historical orders, and the current warehouse and courier list will be implemented, but with some manipulation in order to protect confidentiality. Clustering analysis with k-means clustering based on distance (Lestari, 2017) and weighted k-means clustering based on distance and customer capacity for point weight (Fuente, 2017) will help with warehouse location optimization. The data collection process will help calculate and find the distance and customer capacity features. The Elbow method will be deployed to find the best k number.

Web scraping technique to collect the shipping prices from a variety of couriers, beginning from the clustering-optimal position and ending at the destination in transaction simulation. After determining the best possible site in terms of shipping costs, we report our findings and recommendations to company management along with the business solution and implementation plan. Analysis of other factors in the implementation plan will be part of future research recommendations.

SIMULATION MODEL

In order to conduct this study, Durairaj and Kasinathan created a framework template. There will be several modifications to the framework based on the case study, literature review, and clustering method framework. The modification is about framework-integrated theories as an input and as a data source due to this research being about business simulation, especially clustering analysis. So, based on Bart De Moor, who defines the clustering analysis framework, we need to understand the data source. Then the implementation will be changed to an attribute or variable that will be a feature or attribute of the clustering analysis. Based on the research, the data source can be equivalent to implementation rather than Integrated theory.

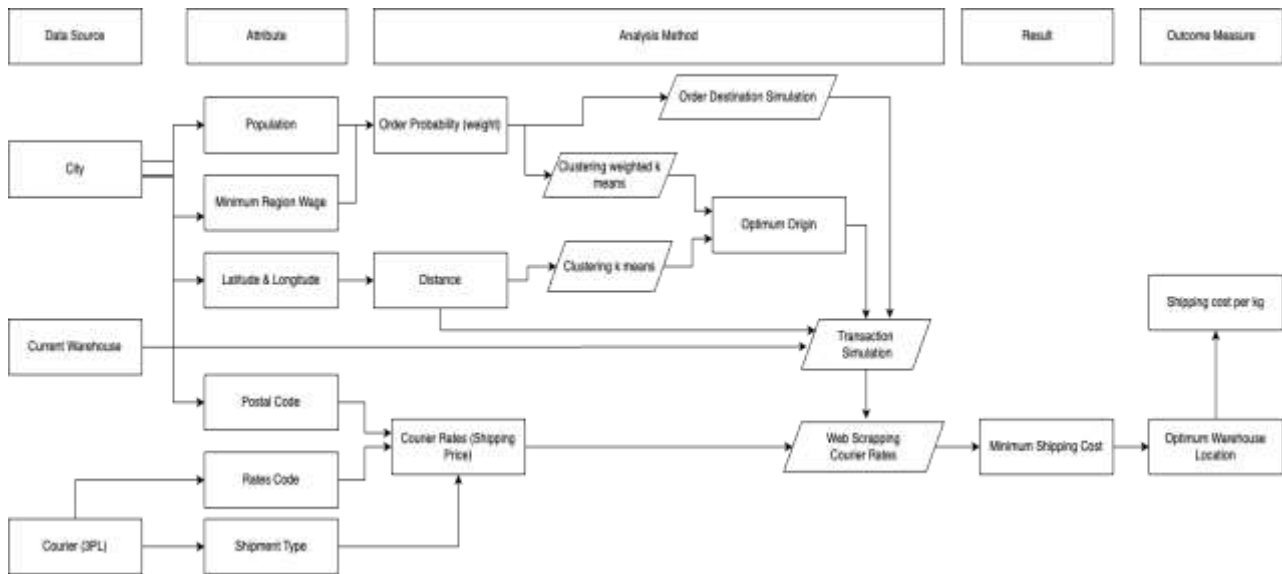


Figure 2. Research Simulation model

The figure shows the new conceptual framework or simulation model that will be implemented in this research. The Data source is a step toward understanding the data and variables that will be considered in this research. There are several explanations that need to be searched for, such as City, Current warehouse, and Courier (3PL). The city contains attributes such as name, province, island, population, minimum regional wage, coordinates (latitude and longitude), and postal code. The current warehouse contains the city name and postal code of the current location of the company warehouses. Every 3PL has an attribute rate code or distinctive code for the shipment type that they own. Shipment Type refers to how fast and big the package is, such as regular, express, same day, or cargo. Both combinations will generate the rate price and determine the shipping cost from origin to destination based on the postal code, rate code, shipment type, and package weight.

Table 1. Example of the city attribute list

No	Island	Province Name	City Name	Latitude	Longitude	Postal Code	Minimum regional wage (Rp)	Population	Customer Capacity
1	Sumatera	Bangka Belitung	Kabupaten Bangka	-2.74	106.44	33214	3.510.076	334.344	1.173.540.895.744
2	Sumatera	Bangka Belitung	Kabupaten Bangka Barat	-1.99	105.22	33312	3.510.076	209.413	735.420.656.188
3	Sumatera	Bangka Belitung	Kabupaten Bangka Selatan	-3.08	106.11	33783	3.510.076	202.263	710.497.926.188
4	Sumatera	Bangka Belitung	Kabupaten Bangka Tengah	-2.74	106.44	33684	3.510.076	205.510	721.555.912.560



Figure 3. City distribution

There are 293 cities spread across three major non-Java islands. The average regional minimum wage is Rp. 3,109,802.55. The average population is 333,619 people, and the total population is 97,750,491 people. Kalimantan Island has the highest average regional minimum wage compared to other islands, but Sumatra Island has a larger population, so Sumatra has more customer capacity. Thus, orders from Sumatra will be higher than those from other islands.

Feature Selection

After collecting the data from the second source, We need to sanitize the data to check for null or missing data. We also check the format of the data to minimize mistakes in data calculation. The collected data only contains attributes. To collect the feature, some calculations will be deployed.

Order Probability

The total multiplication of the wage region and the population of all currently available city data, where the probability of all cities in the data being equal to 1, then divides the calculation of the regional wage by the population in a city.

$$P(A_i) = \frac{wage(A_i) \cdot pop(A_i)}{\sum_{i=0}^{totCity} wage(i) \cdot pop(i)}$$

Where P is the probability of ordering in city A, wage (A_i) is the minimum region wage in city A, pop (A_i) is the total population in city A, wage (i) is the minimum region wage in city I from all city data, and population (i) is the population of city I from all city data.

Distance

The distance calculation process uses the Euclidean method with the input latitude (lat) and longitude (long) from cities A and B, which will be entered into the analysis method of determining the cluster midpoint and also determining the closest warehouse origin to a destination.

$$d(A, B) = \sqrt{(A_{lat} - B_{lat})^2 + (A_{long} - B_{long})^2}$$

where d is distance, A is city A, B is city B, lat is latitude and long is longitude

Courier rates

Courier rates are the shipping costs that 3PL users will pay based on the destination city, the city of origin, the package weight, the type of shipment, the type of package, and the package volume, as well as insurance and other administrative costs. In this research, only the city of origin and destination attributes were used because the indicator that was taken into account was the shipping cost per kg. Since each courier has their own data, there is no need for additional calculations, and there will be no collection of all courier rate data. Instead, a web scraping application will help with this process.



Clustering Analysis

K-means clustering, an unsupervised learning technique, is used to refer to clustering (Saroj, 2016). Unsupervised learning methods are used for data sets that have never been labeled or classified. Each data point is processed in line with the identical points that are stored in the data set. The model, which is either a distance-based method or a centroid-based strategy, is used to calculate the distance to assign points to each cluster. After obtaining a K value, split the data into K categories to enhance the similarity of the data within each category and make the data simpler to distinguish.

The Main Goal and Steps of K-Means

The primary objective of the iterative K-Means method is to cluster the points as closely as possible to their respective cluster centroids (Jain, 2009).

First, choose K items to serve as the first cluster's center in accordance with our study goal. Divide the data with a close distance into one category by calculating the Euclidean distance between the data and the cluster center (Jain, 1999).

$$S_i^{(t)} = \{ x_p : ||x_p - m_i^{(t)}||^2 \leq ||x_p - m_j^{(t)}||^2 \forall j, 1 \leq j \leq k \}$$

Second, recalculate the cluster centers of the newly divided clusters, and then divide the clusters again using the original divisions.

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

Third, continue this repeating process until the cluster center is no longer present. The process can be stopped under three different conditions: first, the newly created cluster's centroid must remain intact; second, all of the points must remain in the same cluster; and third, it must complete the maximum number of iterations. We can halt the procedure if the newly created cluster's centroid has not altered. The algorithm has not learned any new patterns if, even after many iterations, all clusters have the same centroid. Training can be halted at this point. Another indication that training should be explicitly ended is if, even after performing several iterations, these points are still in the same cluster. In this case, training should be terminated.

When the designated number of iterations has been reached, we may finally halt training. Let's say we decide to use 200 iterations. 200 rounds of the method will be repeated before stopping. Here is a diagram that will explain the process.

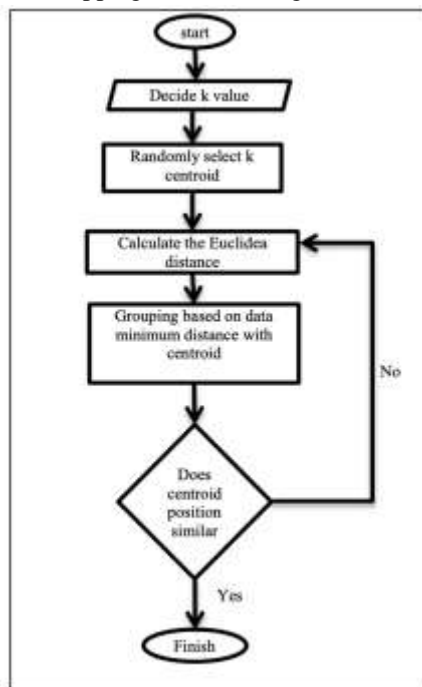


Figure 4. Diagram of the clustering algorithm



Elbow Method

There is a flaw in this technique that is connected to the selection of the first K locations that must be addressed in order for the centroid to stop shifting. This issue is resolved by calculating the algorithm's performance for various centroids. The distance between the centroid of each cluster and the data point can be determined during evaluation as long as convergence takes place. The total computed distances will then be used as a performance indicator. The size of the goal function will shrink as the number of cluster centroids rises. Typically, the elbow approach is utilized to choose the optimal K.

The elbow approach works well for modest values of k. The elbow technique figures out the squared variation between several k values. The average distortion degree decreases as the k value rises. Each category has fewer samples, and the samples are located nearer the center of gravity. The k value corresponding to the elbow is the place where the improving impact of the distortion degree drops the fastest as the k value rises.

The elbow we introduce WCSS (inside-Cluster Sum-of-Squares), a variable that quantifies the variation present inside each cluster. The better the clustering, the lower the overall WCSS. But at some point, there's no significant difference in WCSS between k and k-1. Hence, we can decide the number of clusters that we will use.

□ Elbow Method Formula for WCSS:

$$WCSS = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2 + \sum_{P_i \text{ in Cluster 3}} \text{distance}(P_i, C_3)^2$$

RESULTS

This simulation was performed in mid-2023 with a personal computer with a processor called the Apple M1 and 8 GB of memory. Clustering Simulation was implemented for 400 iterations to get similar centroid conditions for each cluster. The following is an example of the clustering results that have been carried out

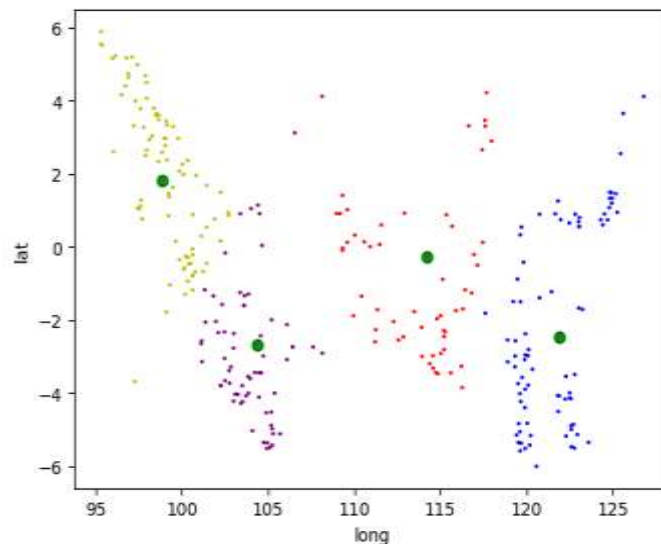


Figure 5. Example of Simulation Results

The figure above is the result of clustering for a weighted scenario with 4 clusters in the 400th iteration. The x axis is longitude, and the y axis is latitude. The dark green point is the centroid, or warehouse, obtained in the simulation. The light green, purple, red, and blue colors used to represent each centroid's area. If you look closely, the island of Sumatra has two centroids, which are divided into the north and south of the island, while Kalimantan and Sulawesi have one warehouse. Each centroid has a latitude and longitude attribute so that the name of the city and the name of the province can be obtained for further processing.



Table 2. Warehouse Location Weighted Cluster

Province Name	City Name	Postal Code
Kalimantan Tengah	Kabupaten Murung Raya	73911
Sulawesi Tengah	Kabupaten Morowali	94974
Sumatera Selatan	Kabupaten Musi Rawas Utara	31654
Sumatera Utara	Kabupaten Karo	22111

Clustering is a simulation process that will produce different results at each simulation iteration. However, there will be a point where the difference in results will no longer be significant. One method to determine clustering is to use the elbow method. The method will look for differences in inertia.

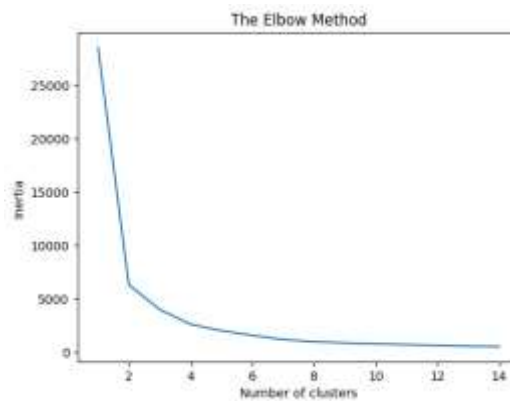


Figure 6. Elbow Method k-means clustering

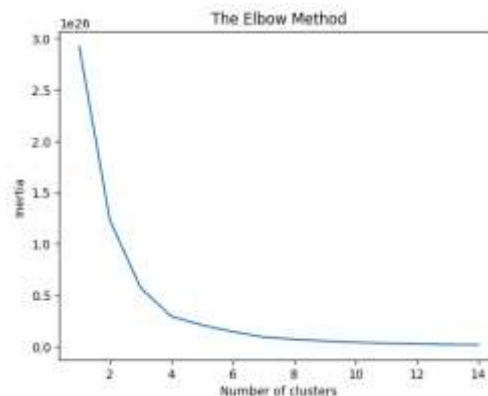


Figure 7. Elbow Method weighted k-means clustering

Based on the two figures above, it is found that clusters 1-3 have a significant difference in inertia and will be sloping in cluster 4, and so on. It can be seen that the difference in inertia between the 5th and 6th clusters is not significant. Therefore, this research will limit the number of clusters used to simulate transactions and get shipping costs to five clusters, valid for regular k-means and weighted k-means. On the other hand, because there were other costs, it would be better if the company built fewer warehouses that were efficient.

After getting warehouse locations represented by centroids based on the simulation in the previous stage, At this stage, an order probability simulation is carried out, which is the result of the opportunity for an order to appear from a city based on purchasing power or customer capacity.



To determine the number of orders, it will be based on the average monthly orders from Java to non-Java, namely 1,000 orders per month (manipulated data). In addition, the company targets sales of up to 10,000 orders per month, or ten times that. Therefore, there are four scenarios for demand estimation. When ordering only 100, then 500, then 1,000, and 10,000. The calculation of the shortest distance between the optimal warehouse for each scenario and the estimated demand scenario was performed. This is done because the company has an algorithm that will look for the nearest destination when placing an order. So there is no need to calculate the shipping price for each optimal origin. The following is an example of a transaction simulation result from a weighted k-means scenario with four clusters or origin warehouses.

Table 3. Example of transaction simulation results

No	Warehouse City Name	Warehouse Postal Code	Warehouse Coordinate	Destination Postal Code	Destination Coordinate
1	Karo	22111	3.02, 98.27	22911	0.84, 99.25
2	Morowali	94974	-2.48, 121.91	96266	0.73, 121.99
3	Morowali	94974	-2.48, 121.91	90511	-5.05, 119.63
4	Musi Rawas Utara	31654	-2.37, 103.03	36514	-1.25, 103.38
5	Morowali	94974	-2.48, 121.91	91911	-2.97, 120.09

To determine the shipping cost, the company's application was crawled in mid-June 2023 by using the postal code of origin from the warehouse and the postal code of destination. There are 12 warehouse location scenarios and 4 scenarios for the number of orders that occur. The table below shows the shipping cost per kg for each warehouse and the number of orders against the current cost per kg in percentage. The percentage is used to disguise the current company shipping cost per order. Higher percentages mean higher shipping costs, the opposite happens. Thus, we will find the lowest percentage.

Table 4. Transaction shipping price results using web scraping

Warehouse Location	Number of Order				Average
	100 orders	500 orders	1,000 orders	10,000 orders	
Java (Benchmark)	102.48%	98.64%	97.87%	98.14%	99.28%
Management Proposed	77.34%	74.28%	75.23%	75.81%	75.67%
Cluster 1	123.01%	121.61%	122.50%	123.44%	122.64%
Cluster 2	98.95%	94.83%	99.29%	98.91%	98.00%
Cluster 3	98.00%	91.58%	97.68%	94.66%	95.48%
Cluster 4	80.78%	79.50%	81.80%	81.95%	81.01%
Cluster 5	67.58%	64.82%	66.09%	66.36%	66.21%
Weighted Cluster 1	123.01%	122.95%	122.62%	123.40%	123.00%
Weighted Cluster 2	98.95%	94.83%	99.29%	98.86%	97.99%
Weighted Cluster 3	81.61%	78.53%	80.65%	81.95%	80.69%
Weighted Cluster 4	84.82%	75.73%	81.42%	78.73%	80.17%
Weighted Cluster 5	80.07%	67.56%	69.12%	68.46%	71.30%



It can be concluded that the feature distance greatly influences the determination of the location of the warehouse in the simulation process, but the significance of the customer capacity feature needs further research. This shipping price can reduce the current shipping price by up to 35%. This percentage also achieves the KPI of reducing cost per order (CPO) by more than 116% .

Besides the calculation of shipping cost reduction in number, The simulation also provides the best location for the company's warehouses, which can reduce shipping costs. To minimize the cost of shipping, the company must open up to five warehouses in the location that is defined by the simulation. Here, the location of that company can be considered.

Table 5. Optimized warehouse location

Province Name	City Name	Postal Code
Gorontalo	Gorontalo Utara	96254
Kalimantan Tengah	Katingan	74413
Sulawesi Selatan	Wajo	90911
Sumatera Selatan	Musi Banyuasin	30711
Sumatera Utara	Tapanuli Tengah	22611

DISCUSSIONS AND CONCLUSIONS

The simulation of warehouse locations in 293 cities on the islands of Sumatra, Borneo, and Celebes to reduce shipping costs from five courier providers for PT. S has been performed. There are several features that are used for determining the location, such as distance and customer capacity. The population's buying power and the minimum regional wage represent the customer capacity. But the distance is more significant to find the cluster centroid or warehouse location based on the fact that there is no big difference in shipping reduction between k-means and weighted k-means. Thus, warehouse location optimization has an influence on minimizing shipping costs per kilogram in PT. S's case.

The simulation gave results showing a reduction in shipping costs per kilogram of around 35% for five clusters. But, based on the simulation results, if the company doesn't have any issue with the number of warehouses, The number of warehouses can be increased to more than five because, when the number of clusters is increased, the shipping cost will decrease. The fact that shipping costs will be lower when the origin and destination locations are close can explain this situation. But there will be new costs, such as warehouse rent, inventory handling, and inventory costs.

There will be five locations that are the most ideal warehouse locations between Sumatra, Borneo, and Celebes that management might take into consideration. In Sumatra, there will be two warehouses in the north and south: Tapanuli Tengah, North Sumatra, and Musi Banyuasin, South Sumatra. Then, in Borneo, there was only one warehouse in Katingan, Central Borneo. For Celebes, there are also two warehouses in Gorontalo Utara, Gorontalo, and Wajo in South Celebes.

This research can be considered to find the optimal location for warehouses with the condition that demand is stable throughout the year and other factors such as the remote area and real buying power, the warehouse cost for each location, and the type of product. For further research, historical data and seasonal demand can be considered for simulation, as can the coverage area of the warehouse and demand location, such as East Indonesia.

REFERENCES

1. Abdul Hafaz Ngah, M. M.-M.-C. (2021). Online Sellers' Reuse Behaviour for Third-Party Logistics Services: An Innovative Model Development and E-Commerce. *Sustainability*, 13.
2. Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation research part c: emerging technologies*, 58, 240-250.
3. Singh, R., Chaudhary, N., & Saxena, N. (2018). Selection of warehouse location for a global supply chain: A case study. *IIMB Management Review*, 30. doi:10.1016/j.iimb.2018.08.009



4. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), 651-666.
5. Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027-1035).
6. Barreto, Sérgio & Ferreira, Carlos & Paixão, José & Santos, Beatriz. (2007). Using clustering analysis in a capacitated location-routing problem. *European Journal of Operational Research*. 179. 968-977. 10.1016/j.ejor.2005.06.074.
7. Blanca Hernández, J. J. (2010). Customer behavior in electronic commerce: The moderating effect of e-purchasing experience. *Journal of Business Research* 63.
8. Chuan Zhang, H.-M. M. (2022). Introduction of the marketplace channel under logistics service sharing in an e-commerce platform. *Computers & Industrial Engineering* 163.
9. Crew, J. (2021). Behind the explosive rise of Indonesia's Tier 2 and 3 cities. *Tech in Asia*. Retrieved October 10, 2022, from <https://www.techinasia.com/explosive-rise-indonesias-tier-2-3-cities>
10. Dubey, A., & Choubey, A. P. D. A. (2017). A systematic review on k-means clustering techniques. *Int J Sci Res Eng Technol (IJSRET, ISSN 2278-0882)*, 6(6).
11. Durairaj, K., & Umar, I. N. (2015). A Proposed Conceptual Framework in Measuring Social Interaction and Knowledge Construction Level in Asynchronous Forum among University Students. *Procedia - Social and Behavioral Sciences*, 176, 451-457. doi:10.1016/j.sbspro.2015.01.496
12. Frye, M. K. (2023). What is an API? Retrieved June 24, 2023, from <https://www.mulesoft.com/resources/api/what-is-an-api>
13. Fuente, D., & Lozano, J. (1998). Determining warehouse number and location in Spain by cluster analysis. *International Journal of Physical Distribution & Logistics Management*, 28, 68-79. doi:10.1108/09600039810205962
14. Ikotun, A., Ezugwu, A., Abualigah, L., Abuhaija, B., & Heming, J. (2022). K-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data. *Information Sciences*, 622. doi:10.1016/j.ins.2022.11.139
15. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), 264-323.
16. Khan, A., G. (2016). Electronic Commerce: A Study on Benefits and Challenges in an Emerging Economy. *Global Journal of Management and Business Research: B Economics and Commerce*, (16)1.
17. Lestari, Dyah & Andika, Aditya & Murphiyanto, Richard. (2017). Cluster analysis for determining distribution center location. *IOP Conference Series: Earth and Environmental Science*. 109. 012021. 10.1088/1755-1315/109/1/012021.
18. Meilatinova, N. (2021). Social commerce: Factors affecting customer repurchase and word-of-mouth intentions. *International Journal of Information Management* 57.
19. Mengyao, C. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. *Geoscience and Remote Sensing Vol. 3: 9-16*. DOI: <http://dx.doi.org/10.23977/geors.2020.030102>.
20. Mitchell A. Millstein, Canser Bilir, James F. Campbell. (2022). The effect of optimizing warehouse locations on omnichannel designs, *European Journal of Operational Research*, Volume 301(2),576-590, <https://doi.org/10.1016/j.ejor.2021.10.061> .
21. Mooney, S. J., Westreich, D. J., & El-Sayed, A. M. (2015). Epidemiology in the era of big data. *Epidemiology*, 26(3), 390.
22. Pengky Frusman, D. W. (2020). Design and Implementation of Warehouse Management Improvement Strategy Using Barcode Systems Approach at PT Latinusa Tbk. *Ictom 04 – The 4th International Conference on Technology and Operations Management*.
23. Prashant Premkumar, S. G. (2020). Trends in third party logistics – the past, the present & the future. *INTERNATIONAL JOURNAL OF LOGISTICS: RESEARCH AND APPLICATIONS*.
24. Qin, X., Liu, Z., & Tian, L. (2019). The Strategic Analysis of Logistics Service Sharing in an E-commerce Platform. *Omega*, 92, 102153. doi:10.1016/j.omega.2019.102153
25. Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. D. F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PloS one*, 14(1), e0210236.



26. Saroj, K. (2016). Review: Study on simple k mean and modified K mean clustering technique. *International Journal of Computer Science Engineering and Technology*, 6(7), 279-281.
27. Xuelian Qina, Z. L. (2020). The strategic analysis of logistics service sharing in an e-commerce platform. *Omega* 92 .
28. Yu, S., De Moor, B., & Moreau, Y. (2023). Clustering by heterogeneous data fusion: framework and applications. *Neural Information Processing 2009 Workshop Learning from Multiple Scores (NIPS 2009 Workshop)*.
29. Zhao, B. (2017). Web Scraping. In *Encyclopedia of Big Data* (pp. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-32001-4_483-1
30. Trianto Atmojo, A., & Wibisono, D. (2023). Proposed Design of Performance Management Framework for 3PL (Third-Party Logistics) Aggregator. In *International Journal of Current Science Research and Review* (Vol. 06, Issue 06). Everant Journals. <https://doi.org/10.47191/ijcsrr/v6-i6-05>

Cite this Article: Alvin Trianto Atmojo, Manahan Siallagan, Riezhan Syauqi Fanhas (2023). Warehouse Location Optimization with Clustering Analysis to Minimize Shipping Costs in Indonesia's E-Commerce Case. International Journal of Current Science Research and Review, 6(8), 5927-5939