Jurnal Teknologi Industri Pertanian 34 (2): 117-126, Agustus 2024
DOI: https://doi.org/10.24961/j.tek.ind.pert.2024.34.2.117
ISSN: 0216-3160 EISSN: 2252-3901

Terakreditasi Peringkat 2
SK Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 158/E/KPT/2021
Tersedia online http://journal.ipb.ac.id/index.php/jurnaltin

AN OPTIMIZE DISTRIBUTION USING HYBRID MODEL: CUSTOMER SEGMENTATION AND TRAVELING SALESMAN PROBLEM

DISTRIBUSI OPTIMAL DENGAN MODEL HYBRID: SEGMENTASI PELANGGAN DAN TRAVELING SALESMAN PROBLEM

Agum Maulana Suyoko*) and Muhammad Asrol

Industrial Engineering Department, BINUS Graduate Program – Master of Industrial Engineering,
Bina Nusantara University, Jakarta, 11480, Indonesia
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah Jakarta 11480 Indonesia
Email: agum.suyoko@binus.ac.id

Paper: Accepted February 20, 2024; Corrected April 22, 2024; Approved May 10, 2024

ABSTRAK

Penelitian ini dilatarbelakangi oleh permasalahan yang ditemukan dalam proses distribusi produk makanan seperti terjadinya pergeseran pola konsumsi pasca pandemic Covid-19 dan potensi kerugian yang timbul akibat keterlambatan proses distribusi makanan ke pelanggan. Tujuan dari penelitian ini adalah untuk mengidentifikasi segmentasi pelanggan untuk meningkatkan efisiensi proses distribusi dan untuk mengidentifikasi rute yang paling efisien pada proses distribusi. Selama proses segmentasi, penting untuk dilakukan identifikasi terhadap atribut yang sesuai. Setelah atribut teridentifikasi, proses segmentasi akan dilakukan dengan menggunakan metode K-Means. Selanjutnya akan dilakukan proses optimasi dengan menggunakan metode Traveling Salesman Problem berdasarkan hasil segmentasi pelanggan. Hasil penelitian menunjukkan proses segmentasi menghasilkan jumlah klaster yang optimal, hal ini ditunjukkan dengan berkurangnya jumlah klaster dari 14 klaster menjadi 9 klaster. Selain itu, hasil proses optimasi menunjukkan bahwa model optimasi sudah optimal. Hasil pada proses optimasi menunjukkan adanya penurunan jarak tempuh keseluruhan sebesar 15,16%, dari total jarak yang sebelumnya 3.386,5 kilometer menjadi 2.940,8 kilometer.

Kata kunci: k-means, proses distribusi, segmentasi pelanggan, traveling salesman problem

ABSTRACT

The research is motivated by issues encountered in the process of distributing products such as shifts in consumption patterns post the Covid-19 pandemic, and potential losses arising from the process of delivering food to customers. The objective of this study is to identify customer segmentation in order to enhance the efficiency of the distribution process and to identify the most efficient route for the distribution process. During the segmentation process, it is critical to identify the suitable attributes. Moreover, once the attributes have been identified, the segmentation process will be implemented using the K-Means method. Subsequently, the optimization process will be executed using the Traveling Salesman Problem. The result showed the segmentation process enable to find the optimal number of cluster, this is shown by reducing the number of clusters from 14 clusters to 9 clusters. Furthermore, the research outcomes regarding the optimization process demonstrate that the model is indeed optimal. The results indicate a 15.16% reduction in the overall travel distance, resulting in a decrease from 3,386.5 kilometers to 2,940.8 kilometers.

Keywords : distribution process, customer segmentation, k-means, traveling salesman problem

INTRODUCTION

In this evolving era, industrial competition has significantly intensified, particularly in the Food and Beverage (F&B) sector, which experienced heightened competition due to shifts in customer behavior brought about by the Covid-19 pandemic. Consequently, numerous companies have adapted their product distribution methods by transitioning to online platforms. According to Putri *et al.* (2021), there is a decrease in consumption patterns in shopping at restaurants which usually reaches 55% of customers, but when the Covid-19 pandemic hit it decreased to 22% of customers. The data presented demonstrates a shift in customer behavior, as individuals have transitioned from their previous preference for in-person shopping to a preference for online shopping.

As the pandemic subsides and enters the era of the "new normal" in Indonesia, people's consumption patterns are undergoing a transformation, reverting to how they were prior to the Covid-19 pandemic. As an illustration, according to Wibawa *et al.* (2022), there is a pattern of increased expenditure for students in the city of Tasikmalaya, which during the pandemic consumption expenditure was 52%, while during the post-pandemic period consumption expenditure was 73%. Based on this data, it is evident that consumption patterns have started to rebound following the Covid-19 pandemic. As a result of this shift, companies will redirect their focus towards competing in direct sales.

*Coressponding Author

Direct selling requires a distribution process, where the finished product needs to be sent from the factory to the store where the company conducts sales. During this process, the factory dispatches products to each store based on predetermined area coverage. One of the obstacles encountered in the distribution process is the occurrence of traffic congestion on the roadways. This can happen because the number of vehicles is increasing. According to the Indonesian Central Statistics Agency (2022), the number of vehicles in West Java in 2021 is 16,843,145 units. There were 3,698,521 passenger cars, 21,529 buses, 414,995 trucks and 12,708,100 motorcycles. Compared to the previous year, the number of all types of vehicles increased, where passenger cars increased by 1.27%, buses increased by 3.31%, trucks increased by 6.06%, and motorcycles increased by 5.52%. Based on this data, it is evident that the surge in motorized vehicles can lead to congestion in the product distribution process.

Product distribution optimization is very important for the company. A decrease in distribution productivity will result in losses for example in economic value. In an article written by Saragih (2020) entitled Congestion in Urban Areas and Impacts on the Logistics System, losses caused by traffic jams in Indonesia amount to a loss of at least US$ 4 billion or the equivalent of 0.5% of Indonesia's Gross Domestic Product (GDP) (Victoria, 2019). The data described earlier forced the company to carry out an optimization process in the product distribution process. According to Chen *et al.* (2021), logistics optimization significantly contributes to mitigating losses caused by vehicle congestion in supply chain management.

Many previous researchers have conducted studies regarding the clustering process to optimizing the distribution process. In essence, a research process entails conducting a comprehensive literature review to underpin the formulation of the problem and provide insights for the study. Rao *et al.* (2019) conducted a study aimed at optimizing vehicle routing problems, which are significant issues within logistics management systems. This research also focuses on mitigating the loss of vehicle fuel consumption during the distribution process. In a study conducted by Zhang *et al.* (2019), it is highlighted that cluster analysis represents one of the initial optimization methods employed in water network sectorization. The proposed approach utilizes a bottom-up clustering algorithm to identify similarities among nodes intended for partitioning in water network sectorization. Therefore it can be seen that the implementation of customer segmentation can enhance efficiency in product delivery.

Additionally, this research will encompass the optimization of distribution processes, as it holds significant importance for companies seeking to enhance cost savings. Numerous methods are available for employment within the optimization process. Among these approaches, the study conducted by Zhang *et al.* (2018), describes the formulation of a comprehensive optimization framework that encompasses both the vehicle routing problem and the task of determining the most efficient route with minimal energy consumption. The results of this study reveal that the vehicle routing problem method can achieve optimization, specifically in generating the shortest route with reduced fuel energy consumption. Furthermore, in the study conducted by Maiyar and Thakkar (2020), the study addresses the issue of optimizing transportation within food supply chains under conditions of uncertain procurement.

The goals of the optimization process may vary depending on the specific objectives. In the study conducted by Niu *et al.* (2018), optimizing vehicle route problems can lead to minimizing total costs, encompassing fuel consumption, CO2 emissions, and driver wages. Additionally, as indicated by Qi and Hu (2020), the challenge of product distribution is a significant concern for numerous experts due to the many drawbacks associated with this process. Febriana *et al.* (2020) applied genetic algorithm for distribution optimization. Lastly, in accordance with Wang *et al.* (2018), customer segmentation and employing the same vehicle for delivery will result in substantial cost and resource savings. The role of customer segmentation becomes particularly crucial when a company must transport items from a single factory source to multiple customers. This aligns closely with the research's objective of grouping customers and organizing vehicle routes.

Among the multiple optimization approaches, this study will employ the Traveling Salesman Problem (TSP) method. Traveling Salesman Problem (TSP) is chosen due to its versatility, as it can be adapted to address a range of problems. As per (McDougall and Otero, 2017; Kim and Moon, 2018; Roberti and Wen, 2016), this technique is extensively utilized in addressing issues within the realm of emergency management. In the study conducted by Liao *et al.* (2018), the Traveling Salesman Problem (TSP) is segmented into several subproblems containing nodes of smaller scale using a clustering algorithm. Each of these subproblems is addressed by employing an ant colony optimization algorithm, forming the foundational layer. The results demonstrate that the proposed algorithm can achieve solutions with enhanced precision and substantially decrease runtime. The objective of the Traveling Salesman Problem (TSP) is to determine the shortest route that connects all cities, ensuring that each city is visited only once before returning to the starting point, while the Traveling Salesman Problem (TSP) may seem a simple problem, but it stands as one of the most crucial classical optimization challenges and has demonstrated considerable difficulty in resolution (Jaradat and Diabat, 2019). Numerous studies have employed the Traveling Salesman Problem (TSP) method, including a paper authored by Ha *et al*. (2018), the Traveling Salesman Problem (TSP) was applied to reduce the overall operational expenses, which included transportation costs attributed to vehicle wait times. It's evident that the Traveling Salesman Problem (TSP) approach stands out as one of the effective methods concerning the distribution process.

The integration of clustering methods alongside the traveling salesman problem is crucial in executing an optimization process for a distribution problem. In the journal written by Baniasadi et al. (2020), this study discusses the clustered generalized traveling salesman problem (CGTSP), where the set of nodes is divided into clusters of nodes, and these clusters are further divided into subclusters of nodes. The outcomes obtained from this study aim to generate an optimized route that visits exactly one node from each subcluster so that all subclusters from each cluster are visited sequentially. Furthermore, in the paper authored by Rani *et al.* (2018), the challenge of grouping daily visited destinations can be resolved through clustering, which groups nearby destinations for visitation on the same day. Meanwhile, the task of optimizing visitation sequences can be addressed using the Traveling Salesman Problem (TSP) approach. Lastly, in accordance with Bernardino and Paias (2018), considered a problem known as the family traveling salesman problem (FTSP), where nodes are grouped into clusters referred to as families. In FTSP, salespersons are tasked with visiting a predetermined number of nodes within each cluster in a sequential manner.

From previous research references, it is evident that utilizing a combination of the clustering method and the traveling salesman problem is feasible and can produce optimal outcomes. Essentially, this research can offer valuable contributions to a company's distribution process. This research has the potential to aid companies with numerous customers spread across various locations in their distribution processes. It could lead to savings in transportation costs as it concentrates on identifying the shortest distance during distribution. The research aims to identify the clusters of the outlet and to identify the most efficient route for the distribution process. The research was conducted in one of a Food and Beverage company in Bandung, Indonesia with a focus on 41 customer outlets to verifying and validating the proposed model.

RESEARCH AND METHODS

Research Stage

The research process initiates with the initial identification stage, followed by the subsequent stages of data collection and processing. Once the data processing is completed, data analysis is conducted, leading to the formulation of conclusions and recommendations. These stages are depicted in Figure 1.

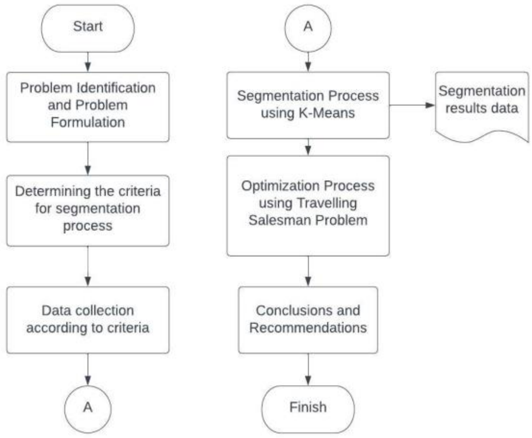

Figure 1. Research Stage

K-means for Customer Segmentation

According to Chen *et al.* (2013) by having data from customers, valuable knowledge can be obtained using data mining techniques. According to Alasadi and Bhaya (2017), data mining is the extraction process of useful patterns and models from huge dataset. According to Mughal (2018), data mining is the analysis of data involves handling large volumes of data, exploring various patterns, employing different methods, algorithms, and tools. Furthermore, in the research conducted by Alalousi *et al.* (2016), there are several explanations related to the use of the clustering method. This study used a K-means clustering methods. In Figure 2 it can be seen related to the flowchart of the algorithm for K-means methods.

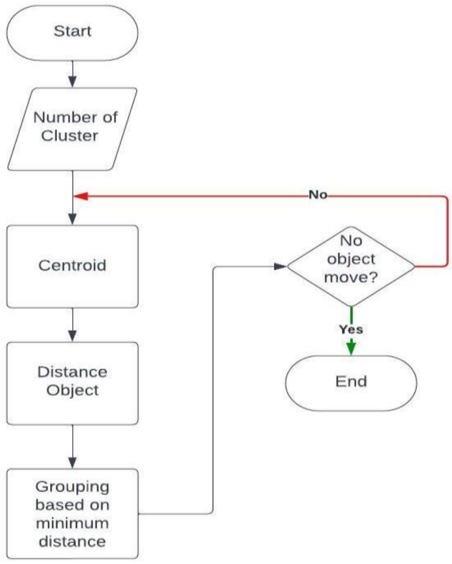

Figure 2. K-means Algorithm Concept
(Source: Alalousi et al, 2016)

Figure 2 is the process of grouping data using a K-means algorithms. The K-means clustering will process the grouping by calculating the centroid and the distance object. Furthermore, according to Agusta

(2007), the algorithm related to the K-means method has the following steps:

a. Determine the number of clusters k
b. Determine the centroid center value with the following formulation:

$$V_{ij} = \frac{1}{N_i}\sum_{k=0}^{N_i} X_{kj} \quad \text{(Equation 1)}$$

Where:

V_{ij}	:	the Ith cluster mean for the jth variable
N_i	:	the number of data members of the i-cluster
i,k	:	index of cluster
X_{kj}	:	the kth data value in the cluster for the jth variable

c. Calculating the distance between the centroid points and points on each object, use the following Euclidean distance formulation:

$$d_{ij} = \sqrt{\sum_{k=1}^{p}(x_{ik} - y_{jk})2} \quad \text{(Equation 2)}$$

d. Calculating the minimum distance of objects, the value obtained in the distance matrix is 0 or 1, where the value is 1 for data allocated to clusters and a value of 0 for data allocated to other clusters
e. Return to the stage in point b, repeat until the centroid value is fixed and cluster members do not move to other clusters.

Traveling Salesman Problem for Distribution Optimization

The Traveling Salesman Problem (TSP) technique is employed to minimize the distribution distance by identifying the nearest distance and route for optimizing the distribution procedure. The formulation of the optimization model is conducted according to the optimization model's objectives. The aim of this study is to determine the most efficient route for the distribution process, which will be achieved by formulating the research using the shortest path problem methodology. The mathematical model to be employed in this investigation is presented below:

$$minimize \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}X_{ij} \quad \text{(Equation 3)}$$

$$constraint\ 1: \sum_{j=1}^{n} X_{ij} \geq d_j \quad \text{(Equation 4)}$$

$$constraint\ 2: \sum_{\substack{1<i<n \\ i\neq j}} X_{ij} = 1 \quad \text{(Equation 5)}$$

In the mathematical model above, it can be seen in Equation 3 that the objective function of the optimization model is to minimize the distance of transportation from i to j knowing that S_{ij} is the distance of transportation from i to j and X_{ij} is a variable, namely the unit of transportation from i to j. While in Equation 4 is a constraint in this study, namely the transportation units sent must have products that are more than or equal to the demand at store j. Meanwhile, in Equation 5 there is a constraint where the transportation unit is only allowed to send to one customer/outlet.

Data Collection

To assess the efficacy of the clustering algorithm, a case study was conducted at a food company situated in Bandung. The company's factory in Bandung serves as the distribution hub for 41 outlets across various areas, including Bandung, Sukabumi, Cirebon, Bogor, Bekasi, Tangerang, Depok, Tasikmalaya, and other regions. The locations of the 41 customers can be seen in Figure 3.

Figure 3. Customer location

The data collection process involved two stages, aligned with each method employed. The initial stage focused on gathering data of each attribute that align with the distribution process. The data to be collected includes the average demand per store, driving distance, and driving time. The data employed for this study is presented in Table 1. Subsequently, data retrieval was performed during the clustering process, prioritizing the predetermined attributes. Company's internal data and data obtained through observations were utilized for this purpose. Information concerning the distance between shops was acquired from the Geographic Information System (GIS) application, whereas the travel time for shops was gathered through observational analysis using the Google Maps application on weekdays during time slots from 08.00 – 10.00, 12.00 – 14.00, and 15.00 – 17.00. Once the data is collected, a validation procedure is conducted in collaboration with the company to assess the accuracy and suitability of the information. Furthermore, as part of the optimization process, there will be a data collection phase to determine the distance from each store.

RESULTS AND DISCUSSION

Actual Customer Condition

In the previous section, it can be seen that there is some data used in this research. Presently, the company conducts the delivery process from one factory to 41 stores.

Table 1. Average demand, driving distance, and driving time

No	Store Coordinates	Location	Average Demand (during pandemic Covid-19)	Average Demand (after pandemic Covid-19)	Driving Distance (km)	Driving Time (min)
1	-6.92075, 106.92271	West Java – Greater Bandung	250	300	105	168
2	-6.81817, 107.13923	West Java – Greater Bandung	200	300	73	109
3	-6.94188, 107.66398	West Java – Greater Bandung	300	350	2	5
…	…	…	…	…	…	…
38	-6.17875, 106.62371	Banten – Greater Jakarta	250	350	183	190
39	-6.26397, 106.65431	Banten – Greater Jakarta	250	300	182	182
40	-6.38128, 106.92484	Banten – Greater Jakarta	250	300	159	159
41	-6.37496, 106.83161	Banten – Greater Jakarta	300	350	167	165

The company groups stores based on 14 different clusters. There are 5 clusters covering the regions of Bandung, Sukabumi, and Cianjur. Additionally, there are 4 clusters covering the areas of Tasikmalaya, Cirebon, Kuningan, Indramayu, Majalengka, and Subang. Moreover, there are 5 clusters covering the Karawang, Bekasi, Purwakarta, Bogor, Depok, Tangerang, Banten, Jakarta, and Cikampek areas. There are variations in the total distance traveled for product distribution across each type of region. In the first type of region, the total distance traveled amounts to 408.2 km, in the second type, it sums up to 1964.4 km, and in the third region, it reaches 1013.9 km. Presently, the company's total distribution distance stands at 3386.5 km.

Customer Segmentation Using K-Means

K-Means method has been applied in many cases including transportation and performance measurement. Anand *et al.* (2014) applied k-means and PCA algorithm to structuring transportation data. Kargari *et al.* (2012) applied k-means for store clustering process to optimize suppl chain and transportation. While in our previous research, Wahyuningtyas and Asrol (2022) elaborated k-means for industry performance and clustered into group for monitoring.

Despite the benefit of k-means in many sector and its conformity for the case study, then k-means is proposed for customer segmentation process. The customer segmentaion based on related attributie is proposed to reducing risk in food distribution and transportation, which the it is mostly found in food distribution (Slamet *et al.*, 2014). Once the necessary attribute of the customer dataset is obtained, the subsequent stage involves performing the clustering process using the K-Means technique. The essential attributes for this study comprise demand outlet per day, driving distances, and driving time. Furthermore, once the required data for the clustering process is acquired, a grouping procedure is executed, categorizing each outlet based on its location type. The grouping is divided into three main types of locations: West Java – Greater Bandung, West Java – North Coast, and Banten – Greater Jakarta. The purpose of this grouping is to ensure that the clustering process results are free from bias and closely aligned with real-world conditions.

Once the location type is determined, the clustering process is subsequently performed according to these three specific location types. The clustering process was conducted utilizing the R-Studio software, employing the K-means method, and the determination of the number of clusters was performed through the elbow method. Figure 4 illustrates the programming language processing conducted with R-Studio.

```r
library(cluster)
library(factoextra)
library(tidyverse)
tableoutlet=read.delim("clipboard")
tableoutlet
summary(tableoutlet)
view(tableoutlet)

# Identify the columns to be chosen for clustering
tablefix=tableoutlet[,4:6]
view(tablefix)

# Identify the optimal number of clusters using the elbow method
fviz_nbclust(tablefix,kmeans, method = "wss")
# Result: West Java - Greater Bandung Location K = 3
# Result: West Java - North Coast Location K = 4
# Result: Banten - Greater Jakarta Location K = 3


# Identify the clustes using K-Means method
# K is input based on the elbow results
final=kmeans(tablefix,3)
print(final)


# Visualize the clustering outcomes in a graphical format
fviz_cluster(final,data=tablefix)

# Visualize the clustering outcomes in a table format
finalakhir=data.frame(tablefix,final$cluster)
view(finalakhir)
finalakhir
```

Figure 4. Programming language using R-Studio

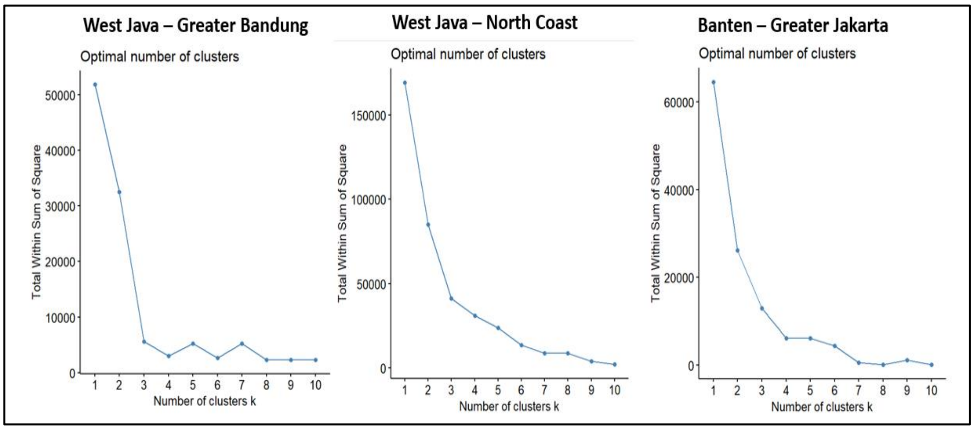

Figure 5. Optimal Number of Cluster

Furthermore, the number of clusters is determined through the utilization of the elbow method. The outcomes of identifying the number of clusters are visible in Figure 5. Subsequently, the clustering process is performed based on the determined number of clusters using the elbow method. This approach involves mapping attributes by selecting the curve's elbow point as the designated number of clusters. The results are illustrated in Figure 5. In that figure, it can be seen in the West Java – Greater Bandung location the optimal number of clusters is 3 clusters, in the West Java – North Coast location the optimal number of cluster is 3 clusters, and in the Banten – Greater Jakarta location the optimal number of cluster is 4 clusters.

Once the ideal number of clusters is identified, you can proceed with the clustering process using the K-Means method. Clusterization was conducted separately for each location type, which includes West Java – Greater Bandung, West Java – North Coast, and Banten – Greater Jakarta. The outcomes of this clustering process is location West Java – Greater Bandung has 3 clusters where the first cluster has 2 outlets, second cluster has 6 outlets, dan third cluster has 8 outlets. Moreover, the West Java – North Coast location has 3 clusters where the first cluster has 3 outlets, the second cluster has 6 outlets, and the third cluster has 4 outlets. Furthermore, the Banten - Jabodetabek location has 4 clusters where the first cluster has 1 outlet, the second cluster has 3 outlets, the third cluster has 3 outlets, and the fourth cluster has 5 outlets.

After the clustering process is carried out, it can be seen that there is a new mapping for each type of location. The total number of clusters generated was 10 clusters, Figure 6 displays the mapping results with 3 clusters for the West Java – Greater Bandung location types. Meanwhile, Figure 7 shows the 3 cluster types found in the West Java – North Coast location types. Lastly, Figure 8 exhibits the 4 clusters identified for the Banten – Greater Jakarta location types.

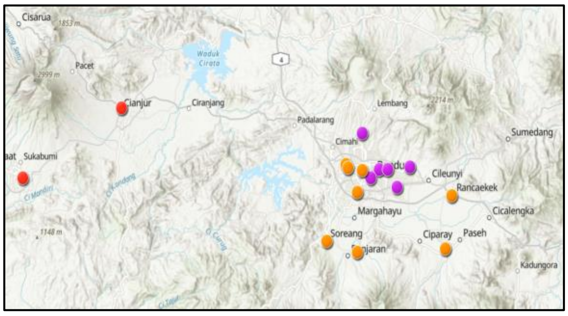

Figure 6. West Java – Greater Bandung cluster mapping

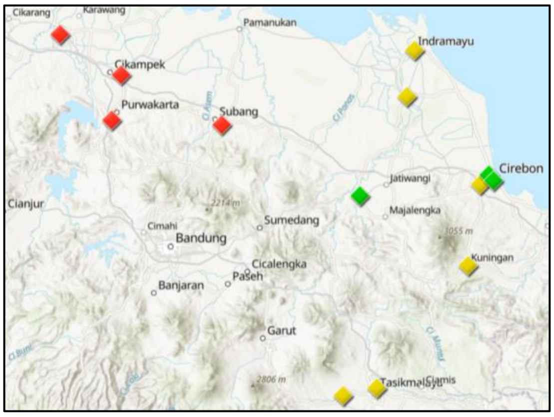

Figure 7. West Java – North Coast Cluster Mapping

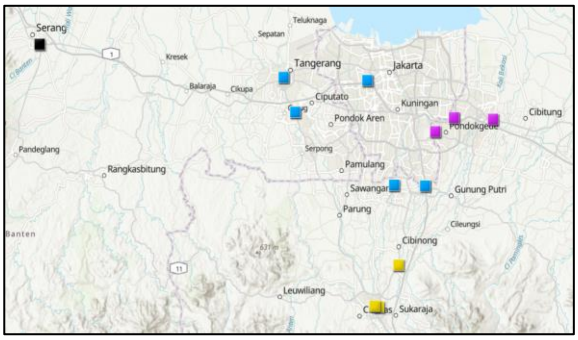

Figure 8. Banten – Greater Jakarta cluster mapping

Optimize Food Distribution Based On Customer Segmentation

Once the clustering process is completed, the subsequent stage involves data processing associated with the optimization procedure. The process of optimization is implemented using the Travelling Salesman Problem (TSP). TSP has been applied for many cases, especially for transportation and distriobution modelling and optimization. Palhares and Araujo (2018) set TSP and vehicle routing problems for dairy distribution process. Sawik *et al*. (2017) extend the TSP for multiobjective transportation optimization with environmental impact considerations. In this case, TSP is applied for food distribution to optimize cost and fulfill customer demand.

For applying TSP, the initial phase involves generating a distance matrix for each cluster based on the designated location type. This step simplifies data processing by providing a prior understanding of the distances between the outlets. A total of nine distance matrices are generated, West Java – Greater Bandung has 3 matrix, West Java – North Coast has 3 matrix, and Banten – Greater Jakarta has 4 matrix. Table 2 provides an illustration of a matrix that has been generated, specifically representing the matrix for cluster 2 in the West Java – North Coast region.

Once all the distance matrices have been established, the subsequent phase involves data processing using Excel Solver. Figure 9 displays the functions employed when utilizing the Excel Solver. The figure illustrates the goal of minimization and the application of a constraint, ensuring that each sorting iteration remains unique due to the restriction that each store can only be visited once. The Excel Solver function is applied to all clusters within each location category.

Table 3 displays the outcomes of processing Traveling Salesman Problem (TSP) data through Excel Solver. The table provides an optimal delivery sequence, which is from store 13 to store 12, then to store 3, followed by store 7, store 1, and lastly to store 11. Additionally, the table indicates a total distance covered, which amounts to 190.7 kilometers

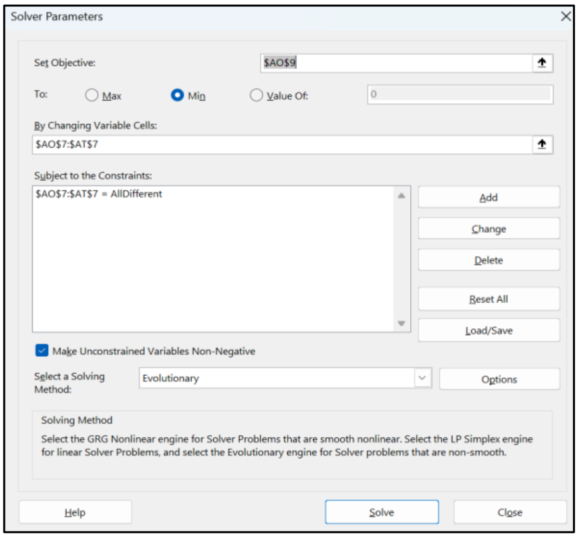

Figure 9. Excel solver parameter

Table 2. Distance Matrix of West Java – North Coast Cluster 2

Store Number	Store Coordinates	Factory	1	3	7	11	12	13
1	-6.47565, 108.30364	126	0	75.6	31.6	20.7	129	140
3	-6.98568, 108.49608	162	75.6	0	49.2	94.9	74.4	84.9
7	-6.70882, 108.43592	125	31.6	49.2	0	50.8	116	126
11	-6.32699, 108.32535	142	20.7	94.9	50.8	0	149	159
12	-7.32534, 108.22027	106	129	74.4	116	149	0	14.8
13	-7.34948, 108.11313	105	140	84.9	126	159	14.8	0

Table 3. Traveling salesman problem result of West Java – North Coast Cluster 2

Row Number	6	5	2	3	1	4
Total Index	14.8	74.4	49.2	31.6	20.7	
Total Distance	190.7					
New Route	13	12	3	7	1	11

Table 4. Final Path for All Location

Location Type	Cluster	Method	Final Path	Total Distance (Km)
West Java – Greater Bandung	1	Travelling Salesman Problem (TSP)	F-14-16-F	209.3
West Java – Greater Bandung	2	Travelling Salesman Problem (TSP)	F-11-7-6-5-15-1-F	56.9
West Java – Greater Bandung	3	Travelling Salesman Problem (TSP)	F-3-4-10-2-8-9-12-13-F	111.8
West Java – North Coast	1	Travelling Salesman Problem (TSP)	F-4-9-10-F	275.8
West Java – North Coast	2	Travelling Salesman Problem (TSP)	F-13-12-3-7-1-11-F	437.7
West Java – North Coast	3	Travelling Salesman Problem (TSP)	F-2-8-5-6-F	271.3
Banten – Greater Jakarta	1	Travelling Salesman Problem (TSP)	F-5-F	474
Banten – Greater Jakarta	2	Travelling Salesman Problem (TSP)	F-2-6-7-F	298.5
Banten – Greater Jakarta	3	Travelling Salesman Problem (TSP)	F-8-9-10-F	393.7
Banten – Greater Jakarta	4	Travelling Salesman Problem (TSP)	F-11-12-1-3-4-F	411.8
			Total Distance	2,940.8

A similar procedure is conducted for every cluster in each location, involving the establishment of a distance matrix and the processing of Traveling Salesman Problem (TSP) method using Excel Solver. Table 4 displays the outcomes of all optimization procedures, presenting the overall distance from each cluster in the table. Table 4 provides an overview of the outcomes of all optimization processes, providing the comprehensive distances from each cluster within the table. In the location category of West Java – Greater Bandung, the total distance achieved is 397.3 kilometers. Meanwhile, in the West Java – North Coast location category, the total distance covered amounts to 984.8 kilometers. In the Banten – Greater Jakarta location category, the total distance extends to 1,578 kilometers.

Within Table 4, the distance from the factory to the outlet has not been included. Specifically, the distance from the factory to outlet 13 is 105 kilometers, and from shop 11 to the factory, it's 142 kilometers. Consequently, the total optimal distance calculated using the Traveling Salesman Problem (TSP) method is 437.7 kilometers.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The data processing is conducted through both the clustering technique and the optimization approach. The process of clustering is executed utilizing the K-Means methodology, while the optimization process is performed utilizing the shortest-path model. Based on the research conducted, several conclusions were obtained according to the following.

The clustering process generates 10 clusters, reflecting the grouping based on multiple attributes like customer demand, driving distance, and delivery time. This led to a shift in the company's cluster count from its initial 14 shipping clusters to 10 shipping clusters. The optimization procedure is implemented through the Traveling Salesman Problem (TSP), involving the computation of both a distance matrix and the processing of Traveling Salesman Problem (TSP) method using Excel Solver. The outcomes of the optimization process generated a cumulative distance of 2,940.8 kilometers, representing a reduction in distance by 445.7 kilometers compared to the previous total distance of 3,386.5 kilometers.

Recommendations

The managerial implication process may involve reducing the number of clusters from the previous count of 14 down to 10 clusters. This will lead to a decrease in the number of transportation modes utilized by the company, requires decisions regarding appropriate strategies to manage these reductions. Each mode of transportation has a capacity ranging from 1500 to 2500 per box. Upon comparing the number of requests with the transportation capacity, no cluster surpasses the transportation capacity. However, it's important to acknowledge that in the event of significant shifts in demand, the clustering and optimization procedures must be re-executed to achieve optimal outcomes. This could serve as a recommendation for future research aimed at developing an efficient system that adjusts to varying attribute conditions.

REFERENCES

Agusta Y. 2007. K-means penerapan, permasalahan dan metode terkait. *Jurnal Sistem dan Informatika* 3 : 47-60.

Alalousi A, Razif R, AbuAlhaj M, Anbar M, Nizam S. 2016. A preliminary performance evaluation of K-means, KNN and EM unsupervised machine learning methods for network flow classification. *International Journal of Electrical and Computer Engineering*. 6(2) : 778.

Alasadi SA and Bhaya WS. 2017. Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*. 12(16):, 4102-4107.

Anand S, Padmanabham P, and Govardhan A. 2014. Application of factor analysis to k-means clustering algorithm on transportation data. *International Journal of Computer Applications*, *95*(15).

Baniasadi P, Foumani M, Smith-Miles K, Ejov V. 2020. A transformation technique for the clustered generalized traveling salesman problem with applications to logistics. *European Journal of Operational Research*. 285 (2): 444-457.

Bernardino R and Paias A. 2018. Solving the family traveling salesman problem. *European Journal of Operational Research*. 267(2): 453-466.

Chen CH, Chiang RD, Wu TF, Chu HC. 2013. A combined mining-based framework for predicting telecommunications customer payment behaviors. *Expert Systems with Applications*. 40(16): 6561–6569.

Chen J, Liao W, and Yu C. 2021. Route optimization for cold chain logistics of front warehouses based on traffic congestion and carbon emission. *Computers & Industrial Engineering*, 161, 107663.

Febriana K, Wahjuni S, and Ismayana A. 2019. Optimasi Distribusi Truk Pengangkut Sampah Menggunakan Algoritma Genetika Pada Sistem Pengelolaan Sampah Kota Bogor. *Jurnal Teknologi Industri Pertanian*, *29*(3).

Ha QM, Deville Y, Pham QD, Hà MH. 2018. On the min-cost traveling salesman problem with drone. *Transportation Research Part C: Emerging Technologies*, 86: 597-621.

Jaradat A and Diabat W. 2019. Solving traveling salesman problem using firefly algorithm and k-means clustering. *In 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 586-589). IEEE.

Kargari M and Sepehri MM. 2012. Stores clustering using a data mining approach for distributing automotive spare-parts to reduce transportation costs. *Expert Systems with Applications*, *39*(5), 4740-4748.

Kim S and Moon I. 2018. Traveling salesman problem with a drone station. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 49(1): 42-52.

Liao Ea and Liu C. 2018. A hierarchical algorithm based on density peaks clustering and ant colony optimization for traveling salesman problem. *Ieee Access*, 6, 38921-38933.

Maiyar LM and Thakkar JJ. 2020. Robust optimisation of sustainable food grain transportation with uncertain supply and intentional disruptions. *International Journal of Production Research*. 58(18): 5651-5675.

Marsudijono, Supriyadi D, Muharam IR, Utami S, Rohaeti E. 2022. Statistik Transportasi Provinsi Jawa Barat 2021. *Badan Pusat Statistik*, *Jawa Barat*.

McDougall JA and Otero LD. 2017. Optimal transfer point locations in two-stage distribution systems. *IEEE Access*, 6: 1974-1984.

Mughal MJH. 2018. Data mining: Web data mining techniques, tools and algorithms: An overview. *International Journal of Advanced Computer Science and Applications*. 9(6).

Niu Y, Yang Z, Chen P, Xiao J. 2018. Optimizing the green open vehicle routing problem with time windows by minimizing comprehensive routing cost. *Journal of cleaner production*, 171: 962-971.

Palhares RA and Araújo MCB. 2018. Vehicle routing: application of travelling salesman problem in a dairy. In *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 1421-1425). IEEE.

Putri A, Syarweni N, and Firdaus RZ. 2021. Perubahan Pola Konsumsi Masyarakat Indonesia pada Masa Pandemi Covid-19. *Lembaga Penerbit Politeknik Negeri Indonesia: Jakarta*.

Qi C and Hu L. 2020. Optimization of vehicle routing problem for emergency cold chain logistics based on minimum loss. *Physical Communication*, 40: 101085.

Rani S, Kholidah KN, and Huda SN. 2018. A development of travel itinerary planning application using traveling salesman problem and k-means clustering approach. *In Proceedings of the 2018 7th International Conference on Software and Computer Applications* (pp. 327-331).

Rao Z and Yuan J. 2019. Common distribution path of cold chain logistics of fresh agricultural products. *Revista de la Facultad de Agronomia de la Universidad del Zulia*. 36(5): 1499-

Roberti R and Wen M. 2016. The electric traveling salesman problem with time windows. *Transportation Research Part E: Logistics and Transportation Review*. 89: 32-52.

Saragih NI. 2020. Kemacetan Di Wilayah Perkotaan dan Dampak Terhadap Sistem Logistik.

Sawik B, Faulin J, and Pérez-Bernabeu E. 2017. Multi-objective traveling salesman and transportation problems with environmental aspects. In *Applications of management science* (pp. 21-55). Emerald Publishing Limited.

Slamet AS, Siregar HH, and Kustoyo A. 2014. Vehicle routing problem (VRP) dengan algoritma genetika pada pendistribusian sayuran dataran tinggi. *Jurnal Teknologi Industri Pertanian*. 24(1): 1-10

Wahyuningtyas D and Asrol M. 2022. The implementation of K-Means clustering in coal-fired power plant performance. *International Journal of Emerging Technology and Advanced Engineering*. 12(3): 1-8.

Wang Y, Zhang J, Assogba K, Liu Y, Xu M, Wang Y. 2018. Collaboration and transportation resource sharing in multiple centers vehicle routing optimization with delivery and pickup. *Knowledge-Based Systems*, 160: 296-310.

Wibawa GR and Pauzy DM. 2022. Pengaruh Pandemi Covid-19 Terhadap Pola Konsumsi Mahasiswa di Kota Tasikmalaya. *Sains Manajemen: Jurnal Manajemen UNSERA*, 8(1): 10-18.

Victoria AO. 2019: Bank Dunia: Indonesia Rugi Rp 56 Triliun Per Tahun Akibat Kemacetan, https://katadata.co.id/berita/2019/10/03/bank-dunia-indonesia-rugi-rp-56-triliun-per-tahun-akibat-kemac

Zhang D, Wang X, Li S, Ni N, Zhang Z. 2018. Joint optimization of green vehicle scheduling and routing problem with time-varying speeds. *PloS one*. 13(2): e0192000.

Zhang K, Yan H, Zeng H, Xin K, Tao T. 2019. A practical multi-objective optimization sectorization method for water distribution network. *Science of The Total Environment*. 656: 1401-1412.