Int. J. Advance Soft Compu. Appl, Vol. 10, No. 2, July 2018
ISSN 2074-8523

Establishing Product Distribution Route of Coca-Cola, Semarang Indonesia

Arum Handini Primandari, Triadi Arie Santoso, and Ayundyah Kesumawati

Universitas Islam Indonesia, primandari.arum@uii.ac.id
e-mail:
Universitas Islam Indonesia, 14611196@students.uii.ac.id
e-mail:
Universitas Islam Indonesia, 136111102@uii.ac.id

Abstract

A company needs to establish the optimal transportation route toward its retailers in order to efficiently distribute its product. The established route includes the length and the order of retailers' visitation which determines the traveling cost. This study proposes methods solving those issues. The available data is the locations including the coordinates of all PT Coca-Cola, Semarang Indonesia retailers. Since the retailers have a tendency to concentrate in the downtown of Semarang, k-medoids algorithm, which is less sensitive to outliers, is employed to cluster the retailers. Clustering is used to define the distribution zone. Meanwhile, Travelling Salesman Problem (TSP) which is utilizing farthest insertion algorithm is used to solve the nearest route and the order of visitation. Instead of being measured by using Euclidean distance matrix, the distance matrix among retailers is measured by employing google maps API to obtain the actual distances. As the result, there are 20 distribution zones and every zone consists of 4–42 retailers. The distribution zone that has 4 retailers has the shortest round trip which distance is 100.773 kilometers. On the other hand, the distribution zone with 42 retailers has the shortest round trip as long as 37.743 kilometers.

Keywords: *kmedoid clustering, TSP, farthest insertion, google maps API, distribution route.*

1 Introduction

A company needs to establish optimal transportation route towards its retailers in order to distribute its product. Once the route is set up, facilities and distribution cost could easily be estimated. The issues related to establishing route is to obtain the most effective way to visit retailers. It is solved by calculating the shortest route and deciding the order of visitation among retailers.

The available data is the retailers' coordinates, that are latitude and longitude, of PT Coca-Cola, Semarang Indonesia. Retailers have a tendency to concentrate in the downtown of Semarang and randomly spread out Semarang's surroundings. Furthermore, there are some retailers located far enough from Semarang city. Thus the data distributes unevenly.

Intuitively, distribution zone is arranged by choosing nearby retailers. Therefore, clustering analysis grouping or clustering objects according to its similarity or intrinsic characteristic is used to establish the distribution zone. K-medoids algorithm is employed because of its robustness to outliers. It is intended for situations in which all variables are the quantitative type. The difference between points is measured by Manhattan distance. Meanwhile, the center of cluster is medoids that is a data point minimize the average dissimilarity between it and all the other the members of the cluster [6]. PAM (Partition Around Medoids) algorithm is employed to find a sequence of medoids that are centrally located in clusters.

K-medoids algorithm requires specifying the number of cluster *k*. One of useful approaches to determine the optimal number of cluster is silhouette algorithm.

In every distribution zone, the shortest round trip and the sequence of visiting retailers are calculated. Travelling Salesman Problem (TSP) utilizing farthest insertion algorithm is used to solve those problems.

2 Related Work

In the transportation problem, clustering is applied to optimize distribution and to decide the shortest distance among area. Clustering analysis is used to optimize the distribution of medical supplies using helicopter and truck in the disaster area [10]. Their research presented two balanced clustering methods for selecting emergency distribution centers (EDCs) and assigning medical aid points (MAPs).

The Coca-Cola retailers' location unevenly spreads out that they are denser in the downtown of Semarang city than its surrounding. Moreover, there is a retailer located far from others that considered as outlier. K-means clustering, the most popular of clustering method, is sensitive to outlier. K-means method takes mean value of all objects within a cluster and sets it as center. While *k*-medoids finds representative object (the medoids) as cluster's center so that it is more robust to outlier than *k*-means [6][7][12]. Besides, the computational time taken by *k*-medoids is less than *k*-means in the case of both normal and uniform distributed data [12].

K-medoids is applied to approximate shortest distance that is improving distance estimation accuracy over local landmark embedding techniques. Their research's result is able to reduce the approximation error of shortest distance up to 29% with respect to the other graph embedding technique.

Our prior research is conducted by applying K-Means to cluster the retailers of PT Coca-Cola and the result is 8 clusters. The retailers are unevenly distributed in which one cluster has hundreds of objects, the other only has one object. Another problem is distance measurement that is using Euclidean. It does not represent the actual distance. Hence the result cannot be applied by the company [11].

After the distribution zone is established, the next problem is arranging the shortest route to visit the retailers. The truck that is used for distribute product visits each retailer within distribution zone exactly once then return to the origin retailer. This problem called Traveling Salesman Problem (TSP). The shortest round trip and order of visiting are obtained by using some algorithm to solve TSP. Nearest insertion algorithm is employed to optimize distribution route planning of clean coal [13]. The result showed that the transportation distance saved several kilometers.

An instance of the TSP is given by distance matrix $D = (d_{ij})$ of dimension $n \times n$, either it is symmetric or asymmetric (called ATSP). The notation d_{ij} is the distance between retailer *i* and retailer *j*. If $d_{ij} = d_{ji}$ for every pair retailer in the matrix, then the TSP is symmetric, otherwise it is asymmetric TSP. Algorithm such as arbitrary, nearest, farthest, cheapest insertion or/and two optimal used to solve TSP/ATSP. The comparison of cheapest, farthest, nearest algorithm find that farthest insertion tour length is the shortest among those three [5]. The develop of arbitrary insertion to enhance its performance also success to outperform the prior one [4].

3 Problem Formulation

The problems of this study are establishing distribution zone, calculating shortest route within every distribution zone, deciding the sequence of retailers' visitation, and estimating distribution cost. While the data is 425 retailers in Semarang and its surrounding, the steps of analysis define as follow:

1. Clustering analysis, that is *k*-medoids with PAM algorithm, is employed to set up the distribution zone. The preprocessing process for clustering is scaling the raw data to make it noise free and consistent. This process is intended to enhance clustering's performance [8]. Moreover, silhouette algorithm is used to get the optimal number of cluster *k*.
2. TSP method with farthest insertion algorithm is used to obtain the shortest route and the sequence of retailers' visitation. According to experimental results, the farthest insertion gives the best result than nearest and cheapest

insertion in Euclidean instance [4][9]. However, this research use google maps API which is applied to build the distance matrix and results asymmetric once.

3. Distribution cost defines as follow

$$C = \frac{R}{5.5} \times IDR\ 5{,}150 \tag{1}$$

4 The Proposed Method

This research used *k*-medoids clustering method which is combined with farthest insertion to solve TSP. The google map API is utilized to build distance matrix when solving TSP.

4.1 K-medoids Clustering

PAM (Partition Around Medoids) is the representative of *k*-medoids clustering method which requires *K* the number of cluster and *D* dataset of *N* points. The algorithm defines as follow [1][6][7]

1. Specify the number of cluster *K*
2. Arbitrarily choose *K* number of data points in *D* as initial center, called medoids, of cluster
3. For each remaining data point *p* in *D*, find the nearest medoid (apply Manhattan distance) and assign *p* to the corresponding cluster.
4. Randomly select a non-medoid point p_{rand} and compute the overall cost C of swapping p_i with p_{rand}. If $C < 0$ swap p_j with p_{rand} to form new sequence of medoids.
5. Iterate steps 3 and 4 until the assignments do not change.

The average of silhouette approach measures the quality of clustering. A high average of silhouette indicates a good clustering. The coefficient is expressed as follow [2]

$$Sil(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \tag{2}$$

where $a(i)$ is average dissimilarity between *i* and the other object within a cluster, $d(i,C)$ is average dissimilarity between *i* to all observations of C (other cluster except cluster where *i* belongs to), $b(i)$ is minimum $d(i,C)$. Silhouette coefficient is ranged from -1 to 1. The objects are placed in the right cluster when it has *Sil(i)* valued near to 1, and vice versa. If Sil(i) = 0, then the object is between the clusters.

The distance between two points is defined by Manhattan distance, instead of Euclidean distance. A point $x = (x_1, x_2, ..., x_N)$ and $y = (y_1, y_2, ..., y_N)$ in *N*-dimensional space will have Manhattan distance as follow

$$d(x,y)=\sum_{i=1}^{N}|x_i-y_i| \tag{3}$$

4.2 Travelling Salesman Problem

A travelling salesman problem is visiting all the retailers exactly at once, then return home by taking the shortest route. This problem was solved by graph theory. A cycle in a given graph G that contains every vertex of G is called a Hamiltonian Cycle of G [14].

The idea of insertion algorithm is iteratively adding a city to a partial tour (subroute) until all cities are completely inserted. The farthest insertion algorithm [9]

1. Choose the arbitrary node i as starting node
2. Find a longest node, j. Form a subroute: $R=i-j-i$
3. Find a node, k (outside the subroute R) farthest to any nodes: $d(k,R)=\max_{i\notin R} d(i,R)$
4. Repeat step 3, until Hamilton cycle is formed
5. Find the edge $[i,j]$ of the subroute to insert k, increasing the length of $\Delta f=c_{ik}+c_{kj}-c_{ij}$ is minimized

Notation: R define subroute, $\Delta f=c_{ik}+c_{kj}-c_{ij}$ define increase in route length when k insert between i and j. Define the distance from node k to subroute R as $d(k,R)=\max_{j\in R} c_{kj}$.

5 Numerical Results

PT Coca-Cola Amatil has many factories spread across Indonesia. One of them is located in Bawen, the suburb of Semarang of Jawa Tengah province, Indonesia. It has retailers in order to distribute its product to the consumer. The retailers' location of PT Coca-Cola Amatil Semarang is mapped in Fig. 1. The retailers are denser in the downtown of Semarang than its surrounding. They spread out to the south until Salatiga, to the east until Demak, and to the west until Kendal.

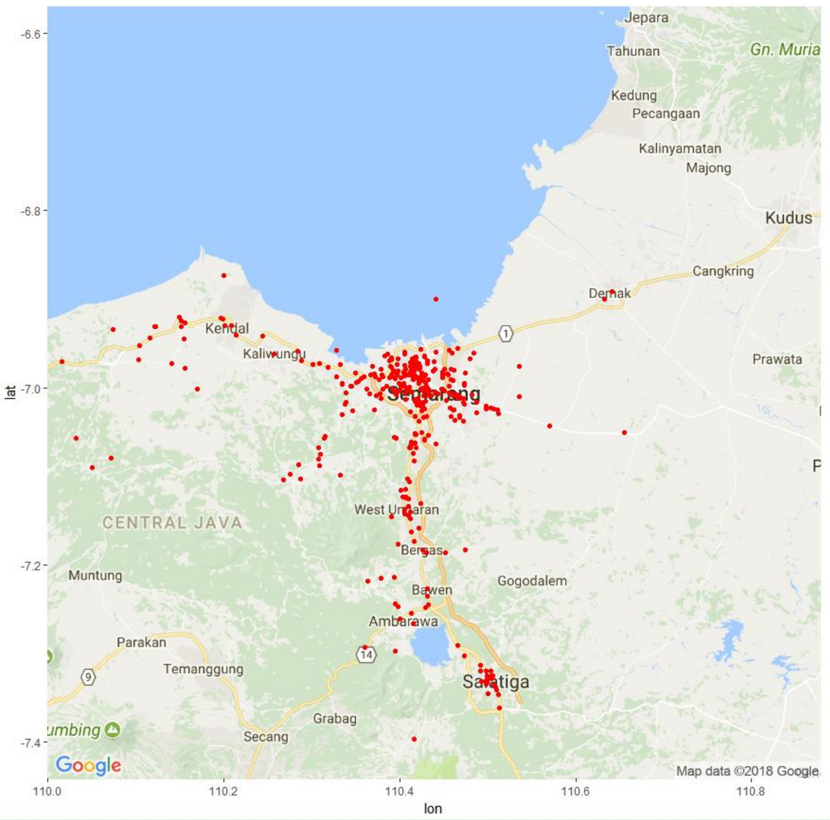


Fig. 1: The Location of Retailers of PT Coca-Cola Amatil Semarang

Taking the first ten data and calculating their Manhattan distance, the visualization distance matrix shows in Fig. 2. The more intense the magenta colour, the farther the distance, e.g. distance between retail number 4 and 7. The more intense the green colour, the nearer the distance, e.g. distance between retail number 5 and 9.

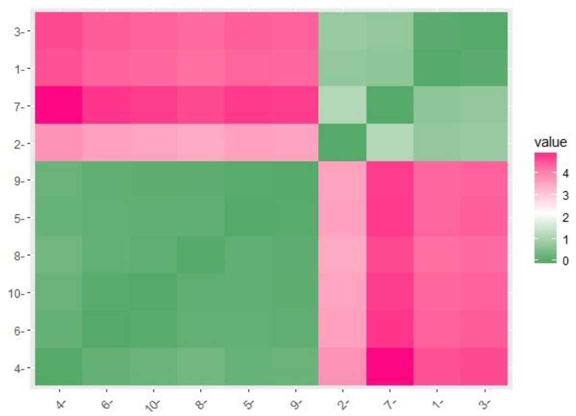


Fig. 2: The Visualization of Manhattan Distance Matrix

The distribution zones are obtained by doing k-medoids clustering with PAM algorithm repeatedly to the scaled raw data. The average of silhouette width is showed in Fig. 3. give the highest value at the number of cluster k = 3. However, considering the number of data, the first process of clustering takes the initial number of cluster k = 17 which is the third highest average silhouette width.

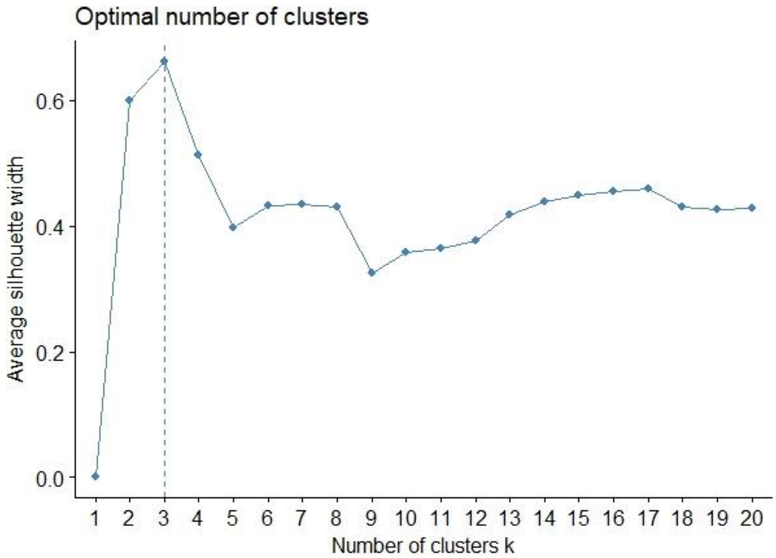


Fig. 3: The Average Silhouette Width

The result of clustering is presented in the table below.

Table 1: The First Clustering Result

Cluster	Freq	Annotation	Cluster	Freq	Annotation
1	59	Dist. zone 1: Need to be clustered	10	14	Dist. zone 10
2	35	Dist. zone 2	11	28	Dist. zone 11
3	23	Dist. zone 3	12	12	Dist. zone 12
4	29	Dist. zone 4	13	12	Dist. zone 13
5	26	Dist. zone 5	14	6	Dist. zone 14
6	56	Dist. zone 6: Need to be clustered	15	21	Dist. zone 15
7	31	Dist. zone 7	16	1	Dist. zone 16
8	42	Dist. zone 8	17	4	Dist. zone 17
9	26	Dist. zone 9			

The first and sixth cluster need to be clustered again because it contains lots of retailers.

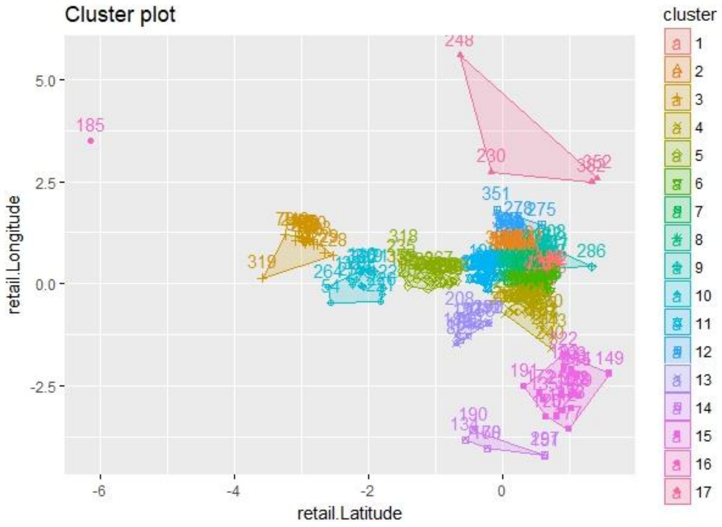
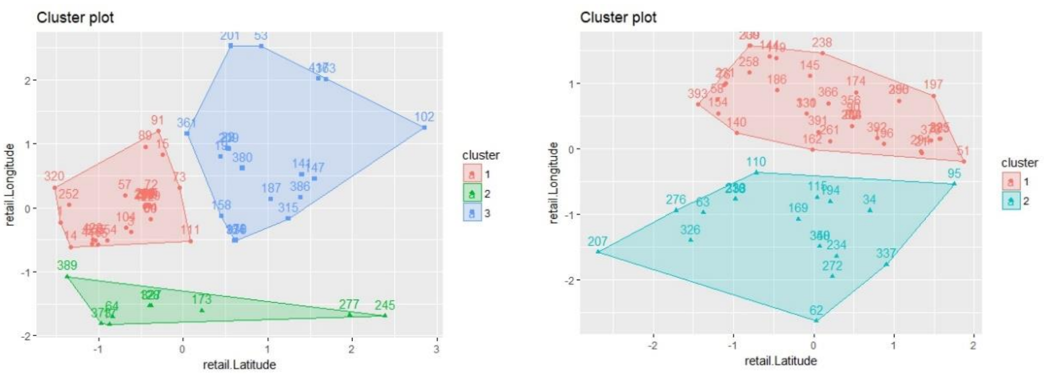


Fig. 4: The Cluster Plot of First Clustering

Based on Fig. 4., the sixteenth cluster only has 1 member because it is located far enough from others retailers. It is SAT H571 Juwiring located in Klaten district, Jawa Tengah.

Taking the similar clustering process, the first distribution zone is divided into three cluster which have frequency 30, 9, and 20. Meanwhile, the sixth distribution zone splits into two group which have frequency 37 and 19. Thus, the total number of cluster is 20 clusters.



(a) (b)

Fig. 5: (a) The Cluster Plot of Second Clustering (Dist. Zone 1); (b) The Cluster Plot of Third Clustering (Dist. Zone 6)

After clustering, the next step is solving TSP to obtain the shortest round trip and the order of visiting retailers within a cluster. Asymmetric matrix distance representing actual distance is built by employing GoogleApis on Google maps. Driving mode is activated while using it.

The members of seventeenth distribution zone are Sat H263 A Yani Gubug, Sat H117 Gubug Grobogan, Agung Swalayan, and Aneka Jaya Demak. The distance matrix of this zone which is built using google maps is represented in table 2.

Table 2: Distance Matrix of 17th Distribution Zone

	Sat H263 A Yani Gubug	Sat H117 Gubug Grobogan	Agung Swalayan	Aneka Jaya Demak
Sat H263 A Yani Gubug	0	34486	23611	25419
Sat H117 Gubug Grobogan	34655	0	40075	41884
Agung Swalayan	23631	39926	0	1458
Aneka Jaya Demak	24531	40827	1337	0

The TSP is visiting the four retailers exactly at once and coming back to the start point. Asymmetric matrix such as table 2, is solved using farthest insertion algorithm.

The shortest route of the eighth cluster, distribution zone that has the most member, is 37.743 kilometers. Distribution zone located in Semarang's surrounding such as third cluster with 23 members has the shortest route, that is around 57.589 kilometers. The longest round trip is 100.773 kilometers which is the shortest route of the seventeenth distribution zone with only 4 members. The order of retailers' visitation in this zone is Agung Swalayan, Aneka Jaya Demak, SAT H117 Gubug Grobogan, and SAT H263 A. Yani Gubug. The path of visitation showed in Fig.5.

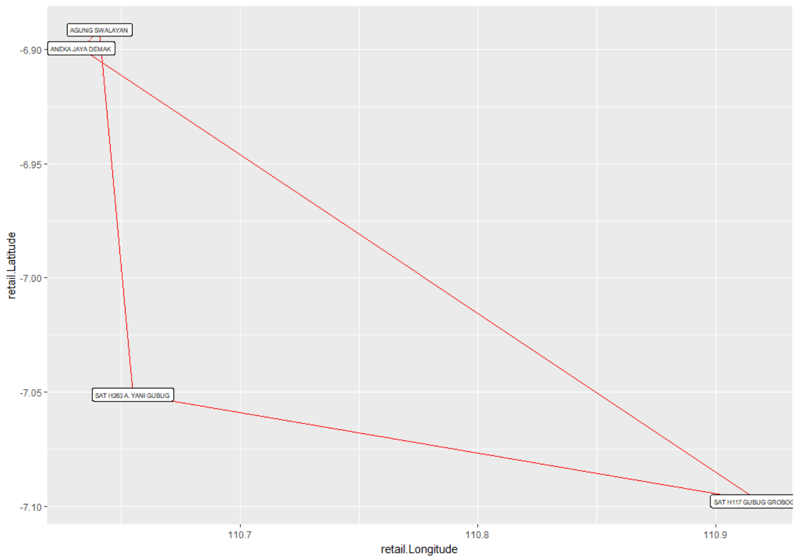


Fig. 6: The Order of Retailers' Visitation (Dist. Zone 17)

Estimation distribution cost for the longest round trip, that is 100.773 kilometers, is IDR 94,400. The most densely populated zone has IDR 35,400 as distribution cost.

6 Conclusion

As the result of *k*-medoids clustering with PAM algorithm, there are total 20 distribution zones. There is a distribution zone which only consists of one retail because it is located far enough from others and considered as outliers. The actual shortest round trip for distribution zone with the most member (42 retailers) is 37.743 km. Its distribution cost is IDR 35,400. The longest round trip is 100.773 km with IDR 94,400 as distribution cost.

Clustering analysis and TSP method are combined to build product distribution route. It is applicable to the company because the distance measured with google map approach, instead of Euclidean distance. Not only the shortest route, this proposed method also provide the sequence of retailers' visitation.

References

[1] Agarwal, S., & Mehta, S. (2017). Approximate Shortest Distance Computing Using k-Medoids Clustering. *Analysis of Data Science, 4*(4), 547–564.

[2] Aranganayagi, S., & Thangavel, K. (2007). Clustering Categorical Data using Silhouette Coefficient as a Relocating Measure. *International Conference on Computational Intelligence and Multimedia Applications* (pp. 13-17). IEEE Computer Science.

[3] Bai, P., Jia, R., & Wang, Z. (2014). Analysis of passenger transportation influencing Factors under clustering analysis ----Shaanxi Province. *Seventh International Joint Conference on Computational Sciences and Optimization.* Beijing, Cina: IEEE. doi:10.1109/CSO.2014.33

[4] Brest, J., & Zerovnik, J. (2005). A Heuristic for The Asymmetric Traveling Salesman Problem. *The 6th Metaheuristics International Conference*, (pp. 145-150). Vienna, Austria.

[5] Hahsler, M., & Hornik, K. (2007). TSP – Infrastructure for the Traveling Salesperson Problem. *Journal of Statistical Software, 23*(2).

[6] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York, USA: Springer.

[7] Jin, X., & Han, J. (2015). K-Medoids Clustering. In C. Sammut, & G. I. Webb (Ed.), *Encyclopedia of Machine Learning and Data Mining.* US: Springer US.

[8] Mohamad, I. B., & Usman, D. (2013). Standardization and Its Effects on K-Means Clustering Algorithm. *Research Journal of Applied Sciences, Engineering and Technology, 6*(17), 3299-3303.

[9] Righini, G. (2000). *The Largest Insertion Algorithm for the Travelling Sallesman Problem.* Retrieved from Dipartimento di Informatica Sede di Crema: http://crema.di.unimi.it/

[10] Ruan, J., Wang, X., Chan, F., & Shi, Y. (2016). Optimizing The Intermodal Transportation of Emergency Medical Supplies Using Balanced Fuzzy Clustering. *International Journal of Production Research, 54*(16), 1-19.

[11] Santoso, T., & Primandari, A. (2017). *Clustering Outlet PT Coca-Cola Region Semarang Menggunakan K-Means.* Yogyakarta: Fakultas MIPA, UII.

[12] Velmurugan, T., & Santhanam, T. (2010). Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points. *Journal of Computer Science, 6*(3), 363-368.

[13] Wang, Y. (2018). Distribution Route Planning of Clean Coal Based on Nearest Insertion Method. *IOP Conf. Series: Earth and Environmental Science*, 1-8.

[14] Wilson, R. J. (1996). *Introduction to Graph Theory.* Malaysia: Longman.