CHAPTER 4

FINDING AND DISCUSSION

4.1 Introduction

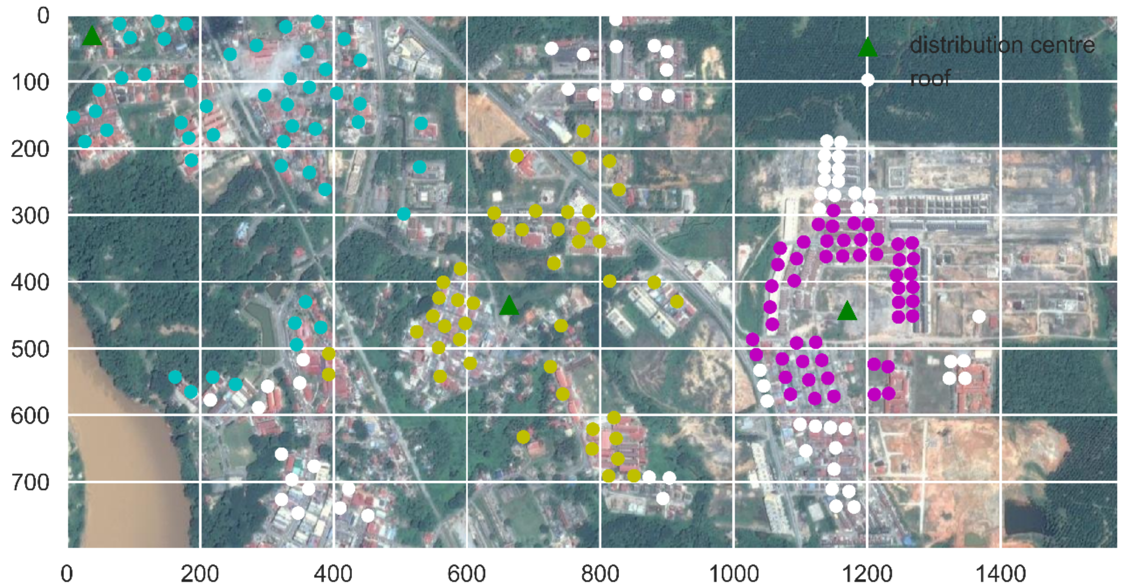
This chapter presents the finding and discussion of experiments conducted for this thesis. Finding towards k nearest neighbor (KNN), Simulated annealing based KNN (SA-KNN), and Genetic Algorithm based KNN (GA-KNN) have been performed. Then, the performance of the experiment is analyzed in term of demand point selected, distribution centre selected, and cost. Each experiment involves the explanation of its procedure, results, and discussion.

4.2 Experimental Result and Analysis

a) **Experiment I :** k-Nearest Neighbor (KNN)

The experiment starting with population size of 200. The conventional KNN is run in the code and number of k was set as 60. As the result, the 60 demand point DP is classified into 3 color represent each DC. When using the KNN show the white color is not cover in distribution centre.

Figure 15: The map after run KNN algorithm with K set as 60



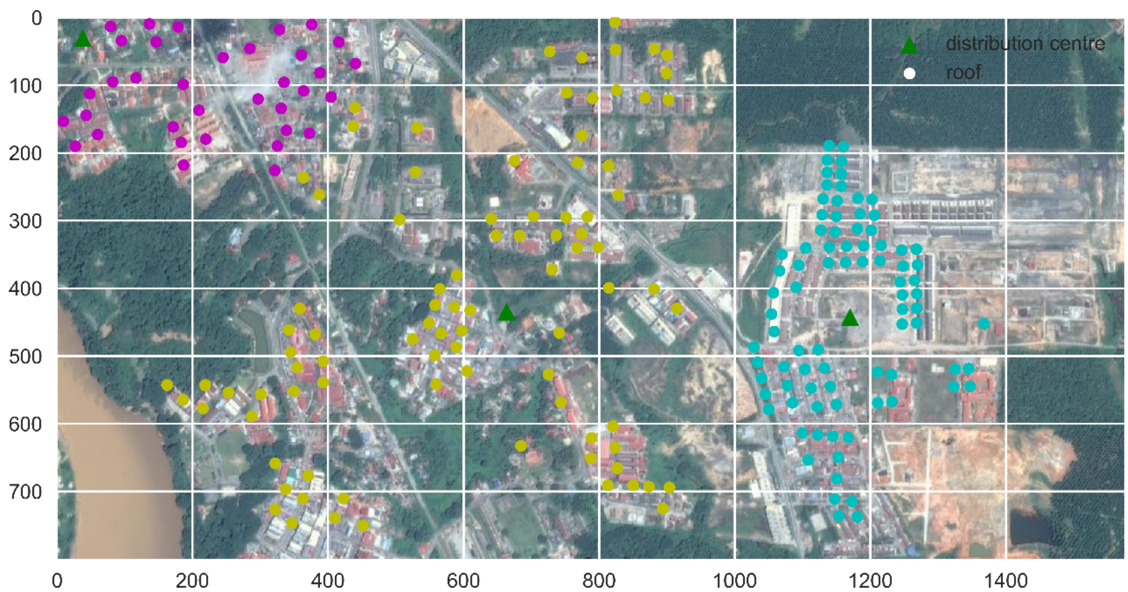
As the population size of DP is 200, there is left-out DP in almost each DC. This means the DC unable to cater all DP in the population. Even if k is set to the higher number, there will be overlap DP covered by each DC. From the result finding in this experiment, there must be improvement of KNN and to meet the constraint.

b) **Experiment II:** Simulated Annealing and Genetic Algorithm based KNN

The experiment is done to GA-KNN and SA-KNN to compare the best solution of DC selection. In order to improve the accuracy of KNN classifier algorithm, the best of k-nearest neighbor for demand point is selected through SA and GA. Firstly the data size of population is initialized. Then the maximum iteration has been determined for validation. Next, the iteration process continues to follow each algorithm procedure and from that the fitness function is evaluated. The DC location is selected, and the value is next use in KNN algorithm for classification.

For SA-KNN experiment set-up as the K value is decided by the algorithm and it seems all DP covered and no left-out DP compared to the KNN previously.

Figure 16: SA-KNN in population of 200 DP



c) **Experiment III:** Genetic Algorithm based KNN

GA-KNN implemented similar process with SA-KNN but differ in the algorithm code. The objective also same which to find the most optimal DC so that the number of DP covered by each DC are nearly equal and balance. 8 potential location has been decided for DC in 200 population and the algorithm will find the best location for DC. Further discussion in the next part of thesis.

Figure 17: Potential DC is marked to be the selection of optimal DC

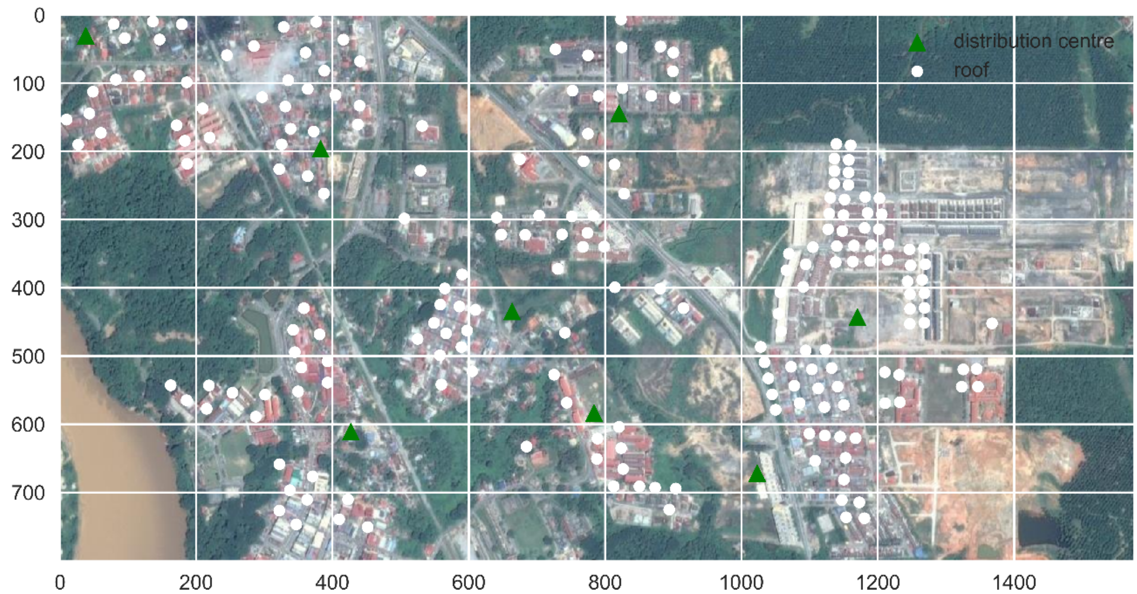
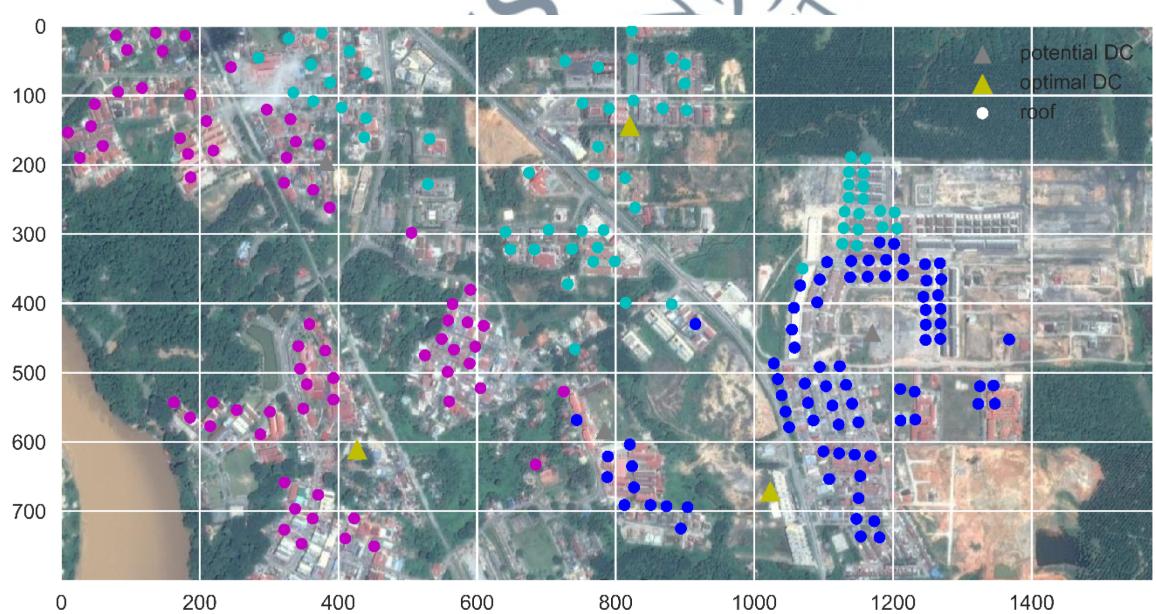


Figure 18: GA-KNN in population of 200 DC



The KNN, SA-KNN, and GA-KNN have been evaluated using 5 datasets created of DP population and was implemented using phyton-based simulation which is Spyder. The result obtained with simulation parameter that were adjusted using Microsoft Excel as the simple database which is to store simple data and Anaconda software. The performance analysis calculated the mean of average different runs for both SA-KNN and GA-KNN. In table 5, show the summarise parameter use in the population and distribution centre. Number of population, distribution centre and deployment area value as stated.

Table 5: Summarize the parameter and value

Parameter	***Value***
Number of Population	100, 200, 300, 400, 500
Distribution Centre	DC1, DC2, DC3,… DC8
Deployment area	5000m *3000m

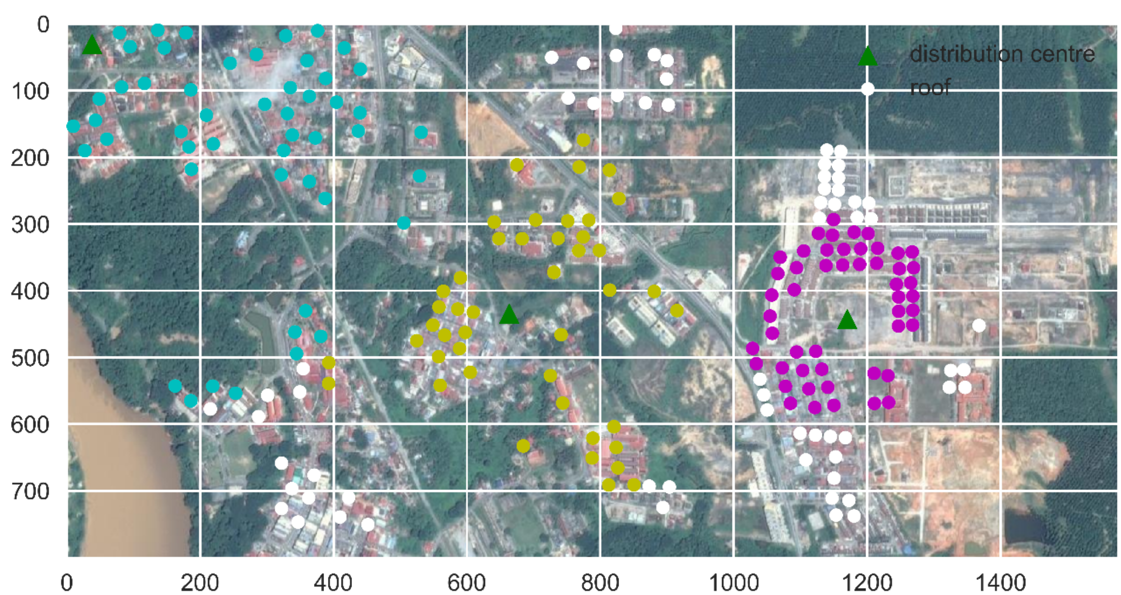
4.3. Result Experiment for population size

The dataset 100, 200, 300, 400 and 500 of population size. The conventional KNN is run in the code and number of k was set. As the result, each number of demand point DP population is classified into 3 color represent each DC. Table 6 summarize the obtained results, where DP as population demand point, DC is total of DC covered, K is the value of DC that we set, M as total DP covered by DC, and LO for left-out DP which not covered. For example, case of population size of DP is 200, there is left-out DP in almost each DC. This means the DC unable to cater all DP in the population. Even if k is set to the higher number, there will be overlap DP covered by each DC. The white circle shows the left-out DP that not covered after run KNN algorithm with K set as 60 in figure 19.

Table 6: Result obtained from KNN algorithm

DP	*DC*	*K*	*M*	*LO*	*Overlap*
100	2	60	100	0	20
200	3	60	180	20	0
300	5	60	300	0	0
400	7	60	400	0	20
500	8	60	480	20	0

Figure 19: The white circle shows the left-out DP that not covered



In the figure 20, the comparison between SA-KNN and GA-KNN impact the cost in covering the area of distribution. For the 3 allocated distribution centre, It clearly shown SA-KNN in higher impact compare to cost reading for GA-KNN.

Figure 20: The Result of Fitness value for 200 DP in SA-KNN

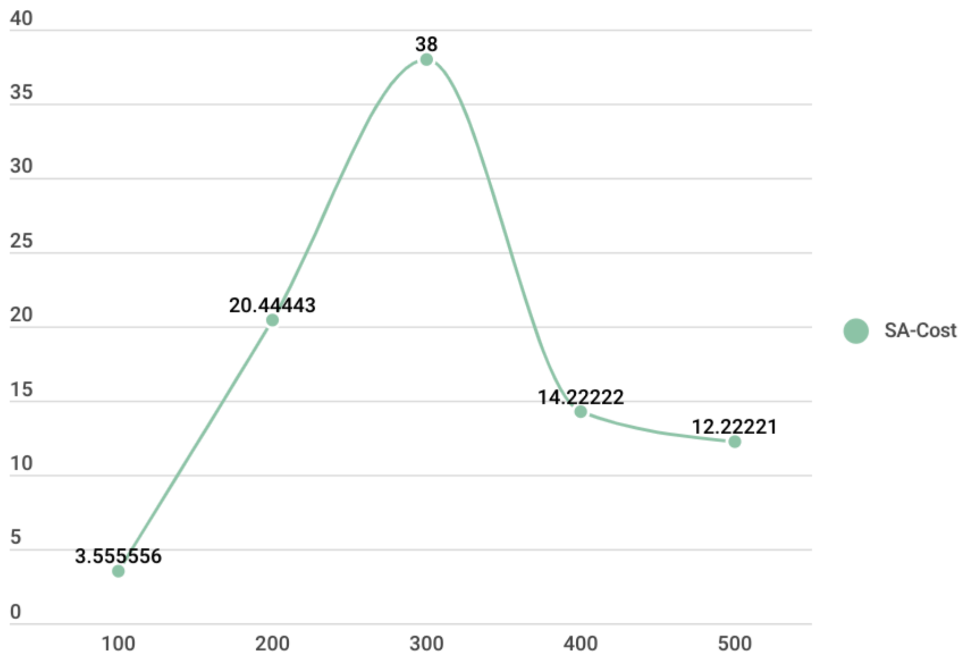


Figure 21: The Result of Fitness value for 200 DP in GA-KNN

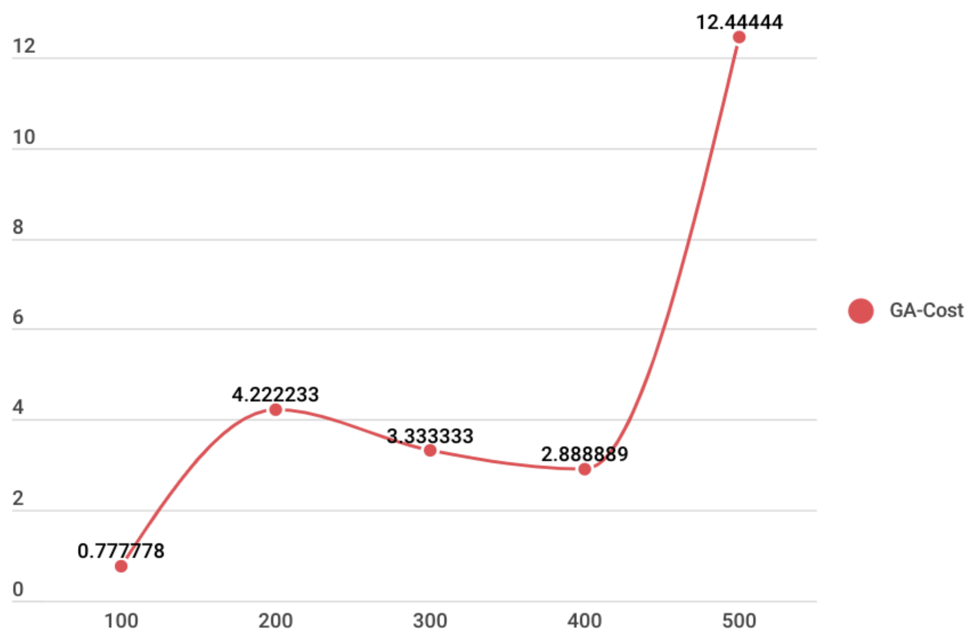


Figure 22: Result of SA Experiment with 200 DP

Figure 23: Result of GA Experiment with 200 DP

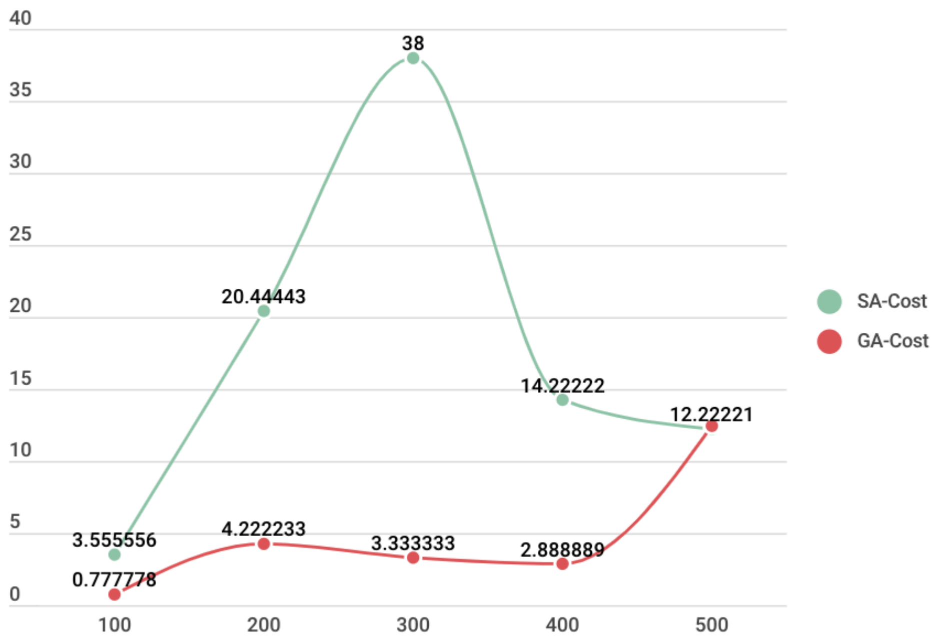
From population 200, there is the 16% difference in fitness value for SA-KNN and GA-KNN algorithm. In Figure 21 shows all the DP are covered and no overlaps between each DC. The purple circles in the population map is the DP covered by dc3. We can see the distribution of yellow circle is dominated by dc2 and the blue circles are represent DP covered by DC1. But the distribution of the DP for each DC is not balance which may affect the logistic cost if in real implementation. From figure 21, clearly illustrate the GA-KNN produced the least cost which means the minimization meet the objective function. As in figure 23, the distribution of purple, light blue, and dark blue are most equal. All DP covered and the optimal position of DCs has been selected.

Table 7: Result obtained from the SA-KNN and GA-KNN algorithm

DP	SA-Cost	GA-Cost	Min
100	3.555556	0.777778	0.777778
200	20.44443	4.222233	4.222233
300	38.00000	3.333333	3.333333
400	14.22222	2.888889	2.888889
500	12.22221	12.44444	12.22221
Total	78.89%	21.11%	

From Table 7, demand point with 300 population result the highest value of cost for SA-KNN compared to GA-KNN with value only 3.3333. This shows the obvious value gap between the number of demand points can be covered by one distribution centre.

Figure 24: Fitness Value of each Algorithm SA and GA-KNN



For overall result obtained in Figure 24, the graph clearly show the green line is having higher value of fitness compared to red line . The mean value of SA-KNN is 79% higher than GA-KNN with 21%. Thus, giving affirmation that *k*-means algorithm is sensitive to the outliers in GA-KNN will produce minimisation and lower the gap between distribution centre with the demand point in one population area,

In conclusion, the result of the analysis presented in this chapter show that KNN is an acceptable method to target the distribution centre location but there is probability to have overlap or left-out demand point. Thus GA-KNN is the one to achieve optimal distribution centre and cater the DPs surround it. The SA might stuck in local optima and effect the SA-KNN to get minimum value and supposedly meet the objective function.

4.3 Summary

Using map grid algorithm, the DPs and DCs are located by researcher on the map, ensuring high accuracy and avoiding left-out roofs in the area. The experiments outcome suggested the use of GA-KNN to help in finding the most optimal DCs location with nearly-equal DP distribution between DC. From managerial perspective, the proposed algorithm renders the researchers the flexibility to locate and find the most optimal DC location from several possible options instead of having fixed and limited option of DC exemplified in experiment 1. In addition, nearly-equal DPs distribution per DC facilitates the real-world aid distribution in disaster area, time-wise and cost-wise.

This work proposes Genetic Algorithm based KNN to better solve the problem of finding the best location of distribution centre with reducing the difference between the demand point covered by each DC. Moreover, after choosing the location of DC in disaster map, it will determine best of nearest neighbors' nodes. 5 datasets have been evaluate to be input in this experiment. Finally, the performance of the Genetic Algorithm based k-nearest neighbor (GA-KNN) is compared with another algorithm which is Simulated Annealing optimization-based k-nearest neighbor (SA-KNN).