CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter explains the methodology used for the research stating from data collection and process of proposed algorithm. Briefly, for the experiment conducted with KNN and followed by modified KNN which involve Simulated annealing and Genetic Algorithm based KNN. This chapter also explains more about distribution centre selection by using of all approach mentioned above.

3.2 Methodology Applied

The research tackled the objectives of this research by implementing few approaches. The first approach was by making review towards literatures on optimization applied in disaster relief operation and comparative studies has been made. This was important to find out what made the optimization worked to select optimal distribution centre. The second approach was to develop an optimization algorithm to meet nearly equal demand point and distribution centre. The third approach is to evaluate the fitness function value and compare with the existing algorithm. Table 1 summarises the methodology applied throughout all studies within the thesis; starting from the literature search up to the development and trials conducted.

Table 3: Summary of the methodology applied

Research Question	**Research Objectives**	**Brief Methodology**
What is the current state-of-art of disaster relief operation to decide the distribution centre?	To study the existing algorithm for determine the distribution centre in post disaster.	Reviewing literatures related to optimization applied in disaster relief operation plus making comparative studies.
Is there any chances for further enhancement in order to meet nearly equal demand point between distribution centre?	To propose a new algorithm for minimize the difference between demand point covered by each distribution Centre	Developing the optimization algorithm to meet nearly equal demand point and distribution centre
How can the proposed algorithm show as an effective approach in determining the distribution centre?	To compare the number of demand point covered by distribution centre of the proposed algorithm and existing algorithm	To use SA-KNN and GA-KNN fitness function value and compare with existing algorithm

3.3 Data Collection

3.2.1 Interview with volunteer Representative and Victim

From Interview on December 2015, the flood victims at Kota Bharu, Kelantan, Sakinah Che Harun acknowledged that her family loss almost RM 120000 to recover from the disaster. All the furniture and house equipment are destroyed. According to her, the volunteer has involved in recovery but not all family has been entertained. Most of the victim that far from capital is not receiving the fund and support.

As for Maziana, a committee of humanitarian mission (Islamic relief ARP), she involves many humanitarian missions and also helping disaster victim inside and outside the country. According to her experience during flood disaster mission at end of 2014, as the volunteer she is following the instruction from the lead in the distribution centre and do her task to assist the disaster victim cleaning the house and distribute the fund. Most of the houses need cost to repair and even worse, they need to rebuild the house. But for that, the case extended to the respective agency such as NADMA.

Norazman is the chairman of Tanjung Karang resident also joined in the disaster relief operation. He and his team having difficulty to decide which location is strategic to unload the needs and distribute to the victim. The team only can reach people near to the relief centre and some of them receive the fund and facility redundantly and some of them do not receive any. This scenario show the unbalance distribution make the team not meet the satisfaction.

Information also gathered through telephone conversation with NADMA representative Pn. Hafizah to know about their involvement in disaster relief activity. According to her, there is always not enough fund and aid during disaster relief operation and they are welcoming any organization to take part. In NADMA portal has information about current alert and activity. The organization that would like to contribute in disaster relief operation can register through their system. When discussing about distribution centre determination will be decided by district representative that in charge in that specific area and most of the cases is for situation during disaster happen. Those information collected from interview is to get the clear picture and to feel the real scene of the disaster history and not for experiment purpose.

3.3 Research Framework

The datasets were coordinates from google map in Kuala Krai. The location selected referring to previous disaster history The experiment conducted following below framework. The first section is to retrieve the map to get the coordinate and then implement in K nearest neighbor classification in software of scientific Phyton development tool, Spyder . The second section will be the explanation on optimized KNN involving simulated annealing algorithm SA and genetic algorithm GA. The result of experiment is in the next chapter.

Figure 8: The framework of experiment process SA-KNN

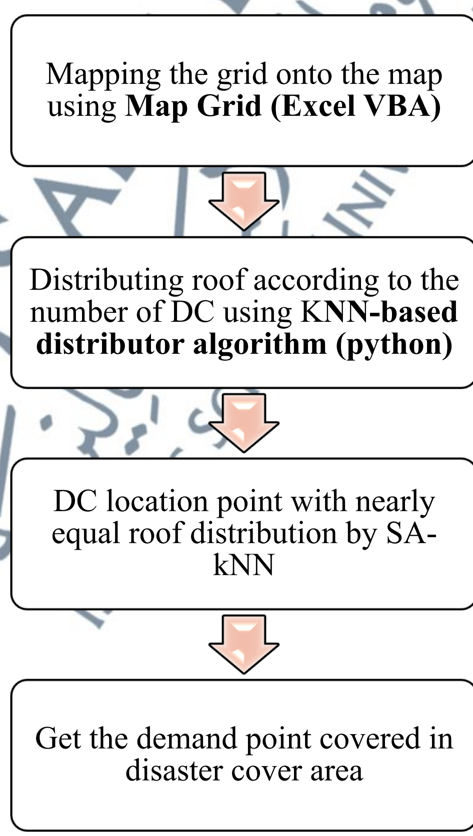
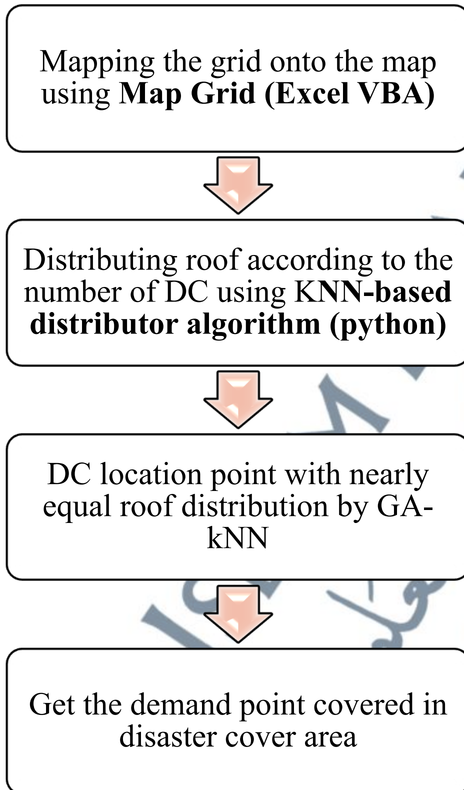


Figure 9: The framework of experiment process GA-KNN



The research process started with insertion of the map. In this thesis, the map taken from google map specifically in Kuala Krai, Kelantan. The map will be extracted into excel VBA to get the coordinate. The coordinate later will be used in the algorithm experiment. At the first place when the map is inserted, the user need to mark the roof and also the potential distribution centre. The coordinate will be generated in the excel table. From the Figure 8, the proposed algorithms where K-Means algorithms. The using K-means algorithm as fitness function for both optimisation algorithms (Simulated Annealing (SA) and followed by Genetic Algorithms (GA) in Figure 9. KNN to modified KNN using phyton programming to run this algorithm. Use the generated coordinate as the input in the algorithm. The distribution centre number need to be decided. The number of roof covered for each distribution centre will be the output of the algorithm.

3.4 Experiment Setting

3.4.1 Problem Formulation and Assumption

For conducting the research, assume in one post-disaster site with a set of disaster victim house as a population demand point and the potential distribution centre as the target point. Each demand point has its own coordinate and do so for distribution centre. The demand point is marked in a given area N = L x L, where L is the side length of the map. From the marked demand point in the map will retrieve the coordinate and will be used to calculate in the k nearest neighbor determination. The consideration of the following assumption about the distribution centre model DC:

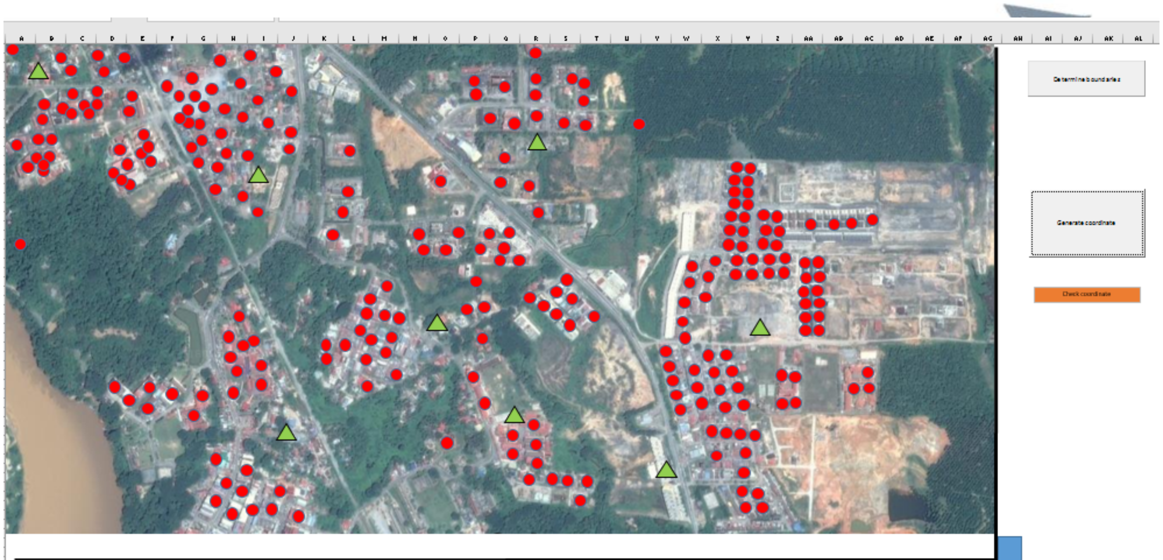
1. Let DP to be the demand point that distributed randomly in a given area N. All demand point is a disaster victim which is demand them to receive fund or basic needs or even support by the volunteer.
2. Let DC to be the potential distribution centre and some of it has tendency to be optimal distribution centre. Distribution centre is a place for volunteer gathered to distribute the needs to the needy in each demand point.
3. Let M to be the total DP covered by the DC.
4. Let Cost to be the value of minimization of the difference between DP equally covered the DC.

For the constraint, we consider each DP can be covered by nearest neighbor of DC and each DC can entertain at least 50 DP. We assume each DC is having own organization to manage volunteer to do the task. The capacity of the DC also decided by the organization referring to the standard and procedure.

3.4.2 Mapping of disaster site

The map will be extract into excel VBA to get the coordinate. The map is retrieve from Google Map. By using map grid algorithm, the location of demand point DP and distribution centre DCs are located by researcher on the map, ensuring high accuracy and avoiding left-out of demand point in the area.

Figure 10: The Structure of Map Gridding



After deciding the DP and DC in the map, click on generate coordinate and the list will be the output. Set the value of triangle and oval to (x,y) as coordinate and save as .csv. file, which will be later fed into distributor algorithm. Figure 10 show the coordinate for experiments which will be using in K-Mean, K-Means with GA and K-Means with SA.

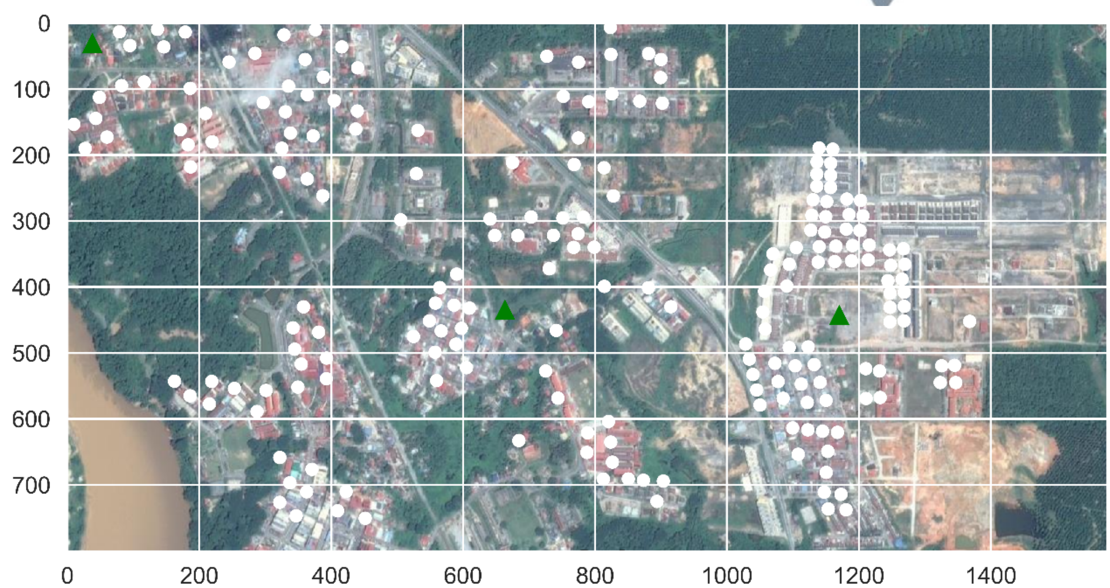
Table 4: List of Coordinate Involved in Experiment

btnGenCoor	***btnGenCoor***	***1612***	***235***
Oval 23	Oval 23	42.95448685	144.3863068
Oval 24	Oval 24	59.95448685	172.6363068
Oval 25	Oval 25	81.95449066	94.63629913
Oval 26	Oval 26	116.4544907	89.13629913
Oval 27	Oval 27	185.954483	98.63629913
Oval 28	Oval 28	182.954483	184.3863068

Oval 29	Oval 29	171.204483	161.3863068
Oval 30	Oval 30	219.454483	179.6363068
Oval 31	Oval 31	590.2044678	380.3862915
Oval 32	Oval 32	585.9544678	427.3862915
Oval 33	Oval 33	557.9544678	424.3862915
Oval 34	Oval 34	548.7044678	451.3862915
Oval 35	Oval 35	564.4544678	400.8862915
Oval 36	Oval 36	566.4544678	466.6362915
Oval 37	Oval 37	597.2044678	462.3862915
Oval 38	Oval 38	344.5616455	494.2077026
Oval 39	Oval 39	557.4544678	498.8862915
Oval 40	Oval 40	609.4544678	432.1362915
Oval 41	Oval 41	588.9544678	486.6362915
Oval 42	Oval 42	524.7044678	474.8862915
Oval 43	Oval 43	209.204483	136.8863068

In the figure 10, show the DP of disaster area and green show the 3 DC. In Table 4 shown the list of coordinate generated from the map gridding process. The coordinate will be used for further step in kNN and modified kNN experiment. 200 population size has been chose because it is the most obvious value to detect leftout of demand point as discussed in previous paper from experiment conducted by Xiaoliang.

Figure 11: The map with 200 population size of DP and 3 DC



3.5 Simulated Annealing and Genetic Algorithm based KNN

3.5.1 Simulated Annealing Algorithm Procedure

This simulated Annealing SA plan of action is to probabilistically re-sample the problem space where the acceptance of new samples into the currently held sample is managed by a probabilistic function. The information processing objective of the technique is to locate the minimum cost configuration in the search space. This probabilistic decision is based on the Metropolis-Hastings algorithm for simulating samples from a thermodynamic system. Simulated Annealing is inspired by the process of annealing in metallurgy. In this natural process a material is heated and slowly cooled under controlled conditions to increase the size of the crystals in the material and reduce their defects. This has the effect of improving the strength and durability of the material. The heat increases the energy of the atoms allowing them to move freely, and the slow cooling schedule allows a new low-energy configuration to be discovered and exploited.

Simulated annealing is widely used for combinatorial optimization problem. The convergence proof suggests that with a long enough cooling period, the system will always converge to the global optimum. The downside of this theoretical finding is that the number of samples taken for optimum convergence to occur on some problems may be more than a complete enumeration of the search space. Performance improvements can be given with the selection of a candidate move generation scheme neighborhood that is less likely to generate candidates of significantly higher cost. Restarting the cooling schedule using the best found solution so far can lead to an improved outcome on some problems. A common acceptance method is to always accept improving solutions and accept worse solutions with a probability of $P(accept) \leftarrow \exp\frac{e-e\prime}{T}$, where T is the current temperature, e is the energy (or cost) of the current solution and e' is the energy of a candidate solution being considered. The size of the neighborhood considered in generating candidate solutions may also change over time or be influenced by the temperature, starting initially broad and narrowing with the execution of the algorithm. A problem specific heuristic method can be used to provide the starting point for the search. Thus, below is the pseudocode listing of the main Simulated Annealing algorithm for minimizing a cost function.

Figure 12: Pseudocode of Simulated Annealing Algorithm

Input: ProblemSize, $iterations_{max}$, $temp_{max}$
Output: S_{best}
$S_{current} \leftarrow$ CreateInitialSolution(ProblemSize)
$S_{best} \leftarrow S_{current}$
For ($i = 1$ To $iterations_{max}$)
  $S_i \leftarrow$ CreateNeighborSolution($S_{current}$)
  $temp_{curr} \leftarrow$ CalculateTemperature(i, $temp_{max}$)
  If (Cost(S_i) $\leq$ Cost($S_{current}$))
    $S_{current} \leftarrow S_i$
    If (Cost(S_i) $\leq$ Cost(S_{best}))
      $S_{best} \leftarrow S_i$
    End
  ElseIf (Exp($\frac{CostS_{current} - CostS_i}{temp_{curr}}$) > Rand())
    $S_{current} \leftarrow S_i$
  End
End
Return (S_{best})

Source: Kirkpatrick, S., & Vecchi, M. P.,1983

3.5.2 Genetic Algorithm Procedure

Individuals of a population contribute their genetic material called the genotype proportional to their suitability of their expressed genome (called their phenotype to their environment, in the form of offspring. This algorithm inspired by the population genetics which may include the mutation and recombination of chromosomes. The next generation is created through a process of mating that involves recombination of two individuals genomes in the population with the introduction of random copying errors called mutation. This iterative process may result in an improved adaptive-fit between the phenotypes of individuals in a population and the environment.

Binary strings referred to as 'bitstrings' are the classical representation as they can be decoded to almost any desired representation. Real-valued and integer variables can be decoded using the binary coded decimal method, one's or two's complement methods, or the gray code method, the latter of which is generally preferred. Problem specific representations and customized genetic operators should be adopted, incorporating as much prior information about the problem domain as possible. The size of the population must be large enough to provide sufficient coverage of the domain and mixing of the useful sub-components of the solution and the Genetic Algorithm is classically configured with a high probability of recombination (such as 95%-99% of the selected population) and a low probability of mutation. Algorithm below shows a pseudocode listing of the Genetic Algorithm for minimizing a cost function.

Figure 13: Pseudocode of Genetic Algorithm

```
Input: Population_size, Problem_size, P_crossover, P_mutation
Output: S_best
Population ← InitializePopulation(Population_size, Problem_size)
EvaluatePopulation(Population)
S_best ← GetBestSolution(Population)
While (¬StopCondition())
  Parents ← SelectParents(Population, Population_size)
  Children ← ∅
  For (Parent_1, Parent_2 ∈ Parents)
    Child_1, Child_2 ← Crossover(Parent_1, Parent_2, P_crossover)
    Children ← Mutate(Child_1, P_mutation)
    Children ← Mutate(Child_2, P_mutation)
  End
  EvaluatePopulation(Children)
  S_best ← GetBestSolution(Children)
  Population ← Replace(Population, Children)
End
Return (S_best)
```

Source: Holland J.H., 1984

3.6 Fitness function

This section describe the design and implementation of SA and GA based KNN for disaster victim demand point population. In the implementation of SA and GA according to fitness function that clarify in Equation 1, after choose the potential best k parameter that clarify number of nearest neighbors of demand point position in order to minimize the difference between each DP covered by DC.

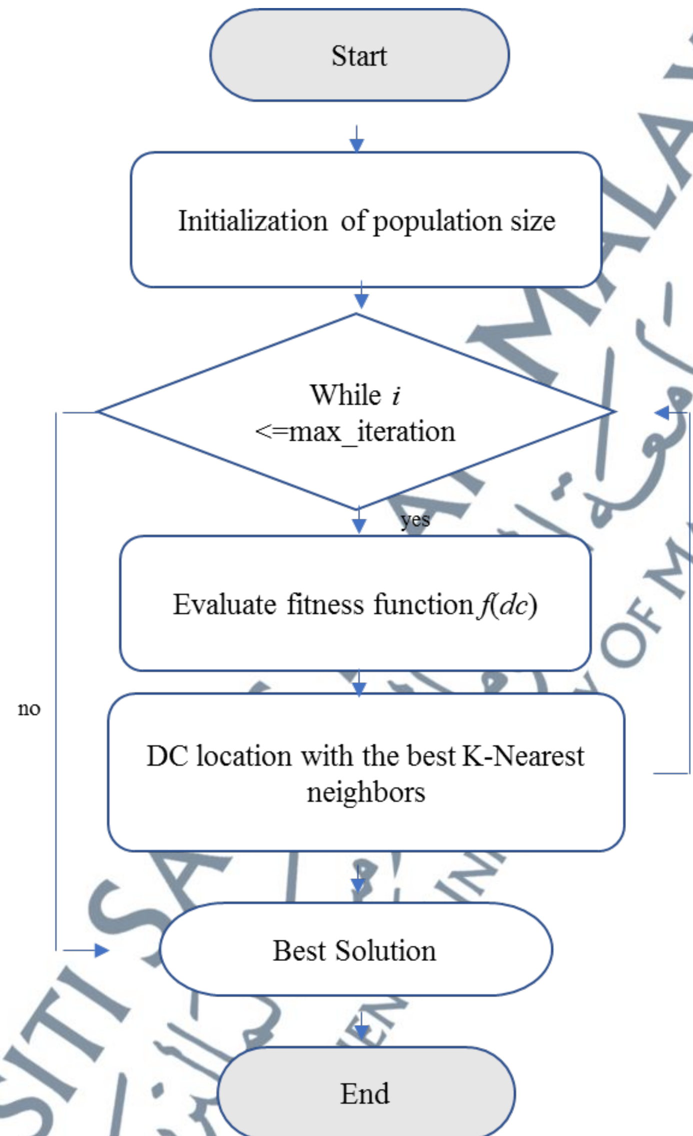
$$f(n_i) = \min|dcn_i - dc_{n+1}| \qquad (1)$$

Where *dcn* is the set of neighbor of distribution centre in one DP population size. Each potential DC supposedly to cover the most equal distribution of DP and having the closest gap with the mean. This is to achieve the objective function which is to minimize the difference between demand point covered by each distribution Centre.

3.6.1 Flow Chart

Figure 14 show the flow chart based on the KNN with Simulated Annealing where KNN as fitness function. It is start from initialization of population size and iteration has been made until maximum point. The Fitness function will be the output and come out with best DC location with kNN algorithm.

Figure 14: Flowchart of SA and GA based KNN



3.6.2 Modified KNN with Simulated Annealing and Genetic Algorithms.

a) Simulated Annealing based KNN

The experiment is done to hybrid SA-KNN (algorithm 2). KNN (algorithm 1) was the fitness function using in the SA. Firstly, the data size of population is initialized.

Then the maximum with 10000 iteration has been determined for validation. The DC location is selected and the value is next use in KNN algorithm for classification.

```
Input: Coordinate DC
       Coordinate DP
       Start and end temperature TZero, TFinal
Define lamda (equation 1)
              Create C = Algorithm 1
              Let t = TZero,
              Let Fit = F(C)
              For i = 1 to ITER
                Let Fnew = F(Cnew)
                Diff = fnew - fit, P = exp(Diff/t)
                If Fnew ≥  fit
                   Fit = Fnew, C = Cnew
                Else
                  If random value (0,1) ≤ P
                    Fit = Fnew, C = Cnew
                  End If
                End If
                t = lamda × t
              End For
              Output: Highest Fit
```

Algorithm 1: Pseudo Code of Hybrid Simulated Annealing Algorithm

The parameter λ is calculated based on swift, (S et.al., 2004) shown in equation 1:

$$\lambda = \frac{\exp((\log(TFinal) - \log(Tzero))}{j} \quad \textbf{(2)}$$

b) **Genetic Algorithm based KNN**

```
Input: Coordinate DC
       Coordinate DP
```

```
Random initialise population DP
Determine fitness of population (Algorithm 1)
For Highest Fit
    Select parents from population
    Crossover and generate new population
    Perform mutation on new population
    Calculate fitness (algorithm 1) for new population
End For
Output: Highest Fit
```

Algorithm 2: Pseudo Code of Hybrid Genetic Algorithm

3.7 Summary

This experiment conducted to see if Genetic Algorithm based kNN have better solution in solving finding the best location of distribution centre with reducing the difference between the demand point covered by each DC. The finding and discussion will be further elaborated in chapter 4.