FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

BITI 3143 – EVOLUTIONAL COMPUTING

SEM 2 – 2020/2021

PROJECT TITLE:

OPTIMIZATION OF INTERNET SIGNALS IN RURAL AREAS

STUDENT'S NAME:

1. NURUL 'IZZAH BINTI MUHAMMAD ZAWAVE B031910128
2. NURAFIQAH BINTI KHOSNI B031910083
3. WAN ADDINI AISYAH BINTI WAN ISMAIL @ WAN ROSLAN B031910404

CLASS: 2 BITI (S1G2)

PREPARE FOR: TS. DR. ZERATUL IZZAH MOHD YUSOH

Individual Task Description

Task	Contributors	Responsibilities
Coding	Nurul ‘Izzah	• Chromosome Evaluation • Parent Selection • Crossover • Mutation • Applicant Initialization • Population Initialization • Survival Selection
	Nurafiqah	• Print Results in TXT File • Add Comments
Report	Wan Addini Aisyah	• Problem Description • Problem Instances Description • Algorithm Design Description
	Nurafiqah	• Generate Graphs • Result Analysis • Create EXCEL File for Generated Chromosome

Problem Description

Substations play a vital role in the electricity network. It connects wires and adjusts the voltage, ensuring that power is safe to use. Without them, the power created in the larger power stations would not be able to make its way into our homes. The purpose of the communication substation is to provide rural citizens with high-speed internet connection so that they may access global information and cultural resources, as well as to offer access to local marketplaces so that they may expand their market reach across the state. However, the placement of communication in rural areas is also simultaneously important. Figure 1 depicts the best and worst communication substation location in rural areas. The best location has a shorter distance, while the worst location has a longer distance. Only the best communication substation location can optimize internet coverage. Otherwise, certain locations will not have adequate coverage and would be unable to access the internet. By placing communication substation strategically in rural area, not only the cost can be lowered, but the coverage of each residence may also be improved.

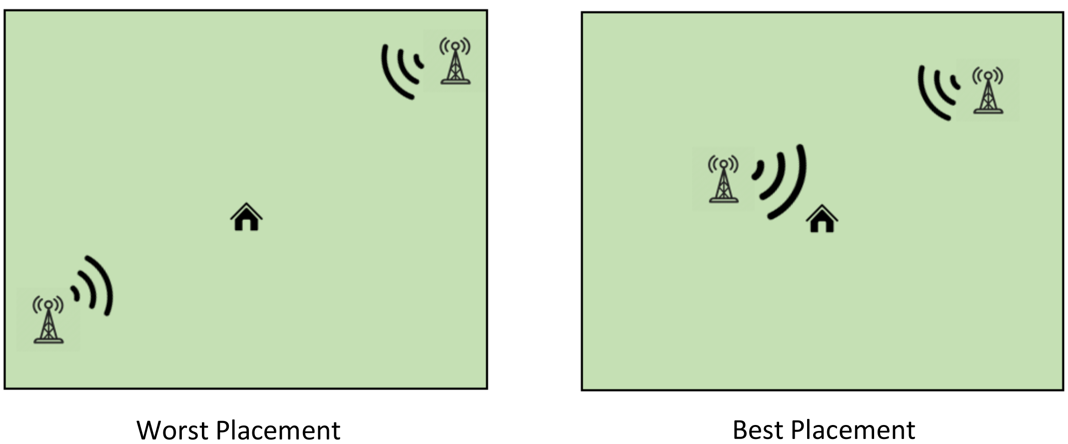


Figure 1 Placement of communication substation in rural area

Description of the Problem's Instances

This project has been developed with the genetic algorithm (GA) to optimize the internet coverage in rural area and place them into a best location. The data of this project has been generated randomly. Given 50 possible location we want to select the best sites, with the lowest costs, while giving us the coverage needed.

The objectives of this project are to provide maximum internet access in rural areas while reducing the number of communication substations in order to save money. The distances from rural area to each substation is inversely proportional with the cost of particular substation construction. Therefore, the lower the distance of rural area to the substation, the higher the cost of that substation construction. Figure 2 depicts the distance between the communication substation and the homes

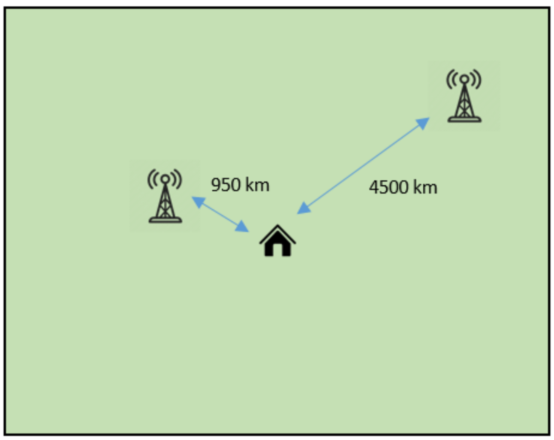


Figure 2 Distance of communication substation

As can be seen, internet signals cover the rural areas over a variety of distances: long and short, with communication substation distances of 950 km and 4500 km, respectively.

The Algorithm Design

This project's context can be considered of as a combinatorial optimization problem. Combinatorial optimization problems can be viewed as searching for the best element of some set of discrete items. However, solving this problem using normal programming approaches is extremely challenging. As a result, we must apply the Genetic Algorithm developed by John

Holland and his associates in the 1960s to solve this problem. This algorithm is a biological evolution model based on Charles Darwin's natural selection hypothesis.

A. Chromosome representation

This project was created using an integer chromosome representation with values in the range from 1 to 50, which is shown in Figure 3. Because there are 8 potential locations in these rural areas, each chromosome has 8 genes. As a result, 8 distance of communication will be chosen randomly by this algorithm from the distance array.

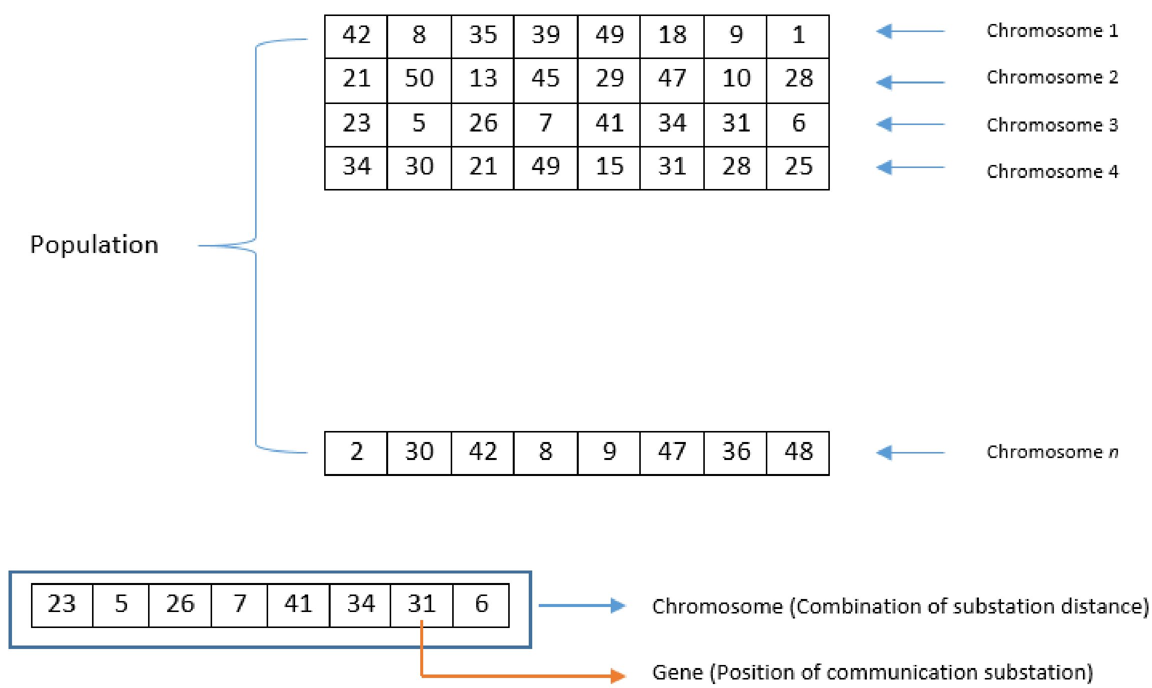


Figure 3 Chromosome Representation

B. Fitness function

The fitness that will be evaluated is the sum of:

1. Total distances from rural area to each substation in each generation.
2. Total cost needed in substation construction.

For the first fitness, it is measured by the summation of all randomly selected distance in the array. There are 50 distance value in the array representing the 50 available sites for the substation construction. The distance value is varying from 800m to 5000m. However, the required total distance was 38600m for one generation. Therefore, the differences of total required distance and the total obtained distance is divided by total required distance before multiply by 0.5. As there are two fitness function evaluated, each one of it will be multiply by 0.5 to correspond to a full probability of fitness function.

In the second fitness function, the total cost of selected substations is calculated. There are also 50 costs in the array representing each substation cost. The cost is varying from RM 80 000 to RM 120 000. However, the required total cost was RM 96 000 for one generation. This factor also will be multiplied by 0.5 then will be added with the first calculation. The complete fitness function is shown as below.

$$calculation\ 1 = \frac{(Total\ required\ distance - Total\ obtained\ distance)}{Total\ required\ distance} \text{ x } 0.5$$

$$calculation\ 2 = \frac{(Total\ required\ cost - Total\ obtained\ cost)}{Total\ required\ cost} \text{ x } 0.5$$

$$fitness\ function = calculation\ 1 + calculation\ 2$$

Figure 4 Complete Fitness Function Calculation

C. Strategy for Parent Selection, Crossover, Mutation and Survivor Selection

The tournament selection is used to choose the parents. Based on their fitness function, two players are chosen at random to compete. One of the parents is chosen as the winner. Repeat until both parents have been chosen and parent 1 is not the same as parent 2. The chromosome with the lowest fitness value will become the parents in this communication substation scenario. Tournament selection was chosen because it is easier to compute, code, and offers a more equal chance of being picked as a parent to a chromosome with a lower fitness value.

The crossover method is 1-point crossover. A crossover point on the parent chromosome is selected. All data beyond that point in the chromosome is swapped between the two parent chromosomes. As seen in the figure below, each child inherits one segment from each parent.

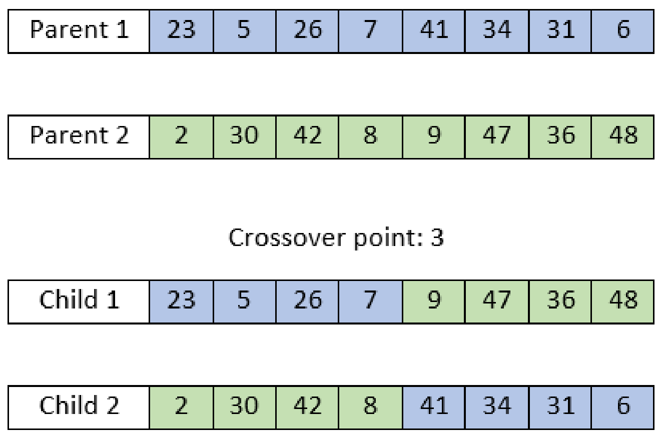


Figure 4 1-Point Crossover process

The mutation strategy used in this genetic algorithm for communication substation is bit-flip mutation. One random gene in a chromosome is selected randomly. The value is exchanged to produce a mutated chromosome as depicted in the figure below.

No mutation

Child 1	23	5	26	7	9	47	36	48

Mutation happen

Child 2	2	30	42	8	41	34	31	6

After mutation

Child 2	2	30	42	8	41	29	31	6

Figure 5 Bit-flip Mutation

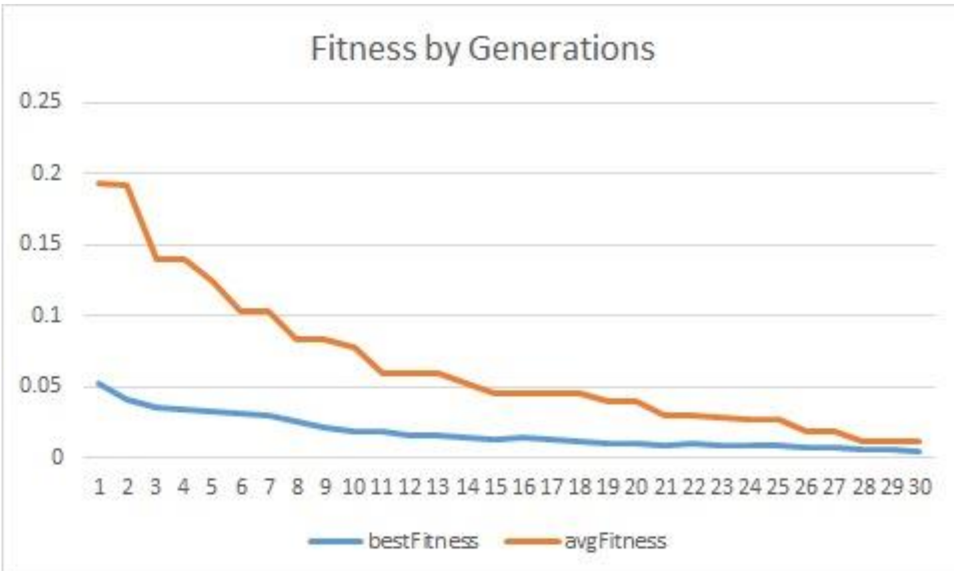
The all children replace parent strategy was used in the survival selection. This is a method of reducing the number of chromosomes in the current generation's population. After the crossover process, the population increases, and therefore we must maintain the chromosome population size for the upcoming generation. The children always take the position of the parents in this method.

D. Termination Strategy

When the genetic algorithm achieves the termination criterion, it will ground to a stop. In this problem, the maximum number of generations allowable is used as the termination condition. In this method, the maximum allowable generation is set at 100. As a result, the process will ground to a stop once 100 chromosome generations have been generated.

E. Measurement Indices

In order to evaluate the algorithm, we used fitness function as stated above. The best chromosome of each population is determined by its fitness. While the average fitness is the mean fitness of all fitness in a population. In the figure below, is the visualization of best fitness and average fitness. Blue line indicates average fitness while orange line indicates best fitness. In this project, as out solution is based on the lower the better, the blue supposed to be above the orange line.

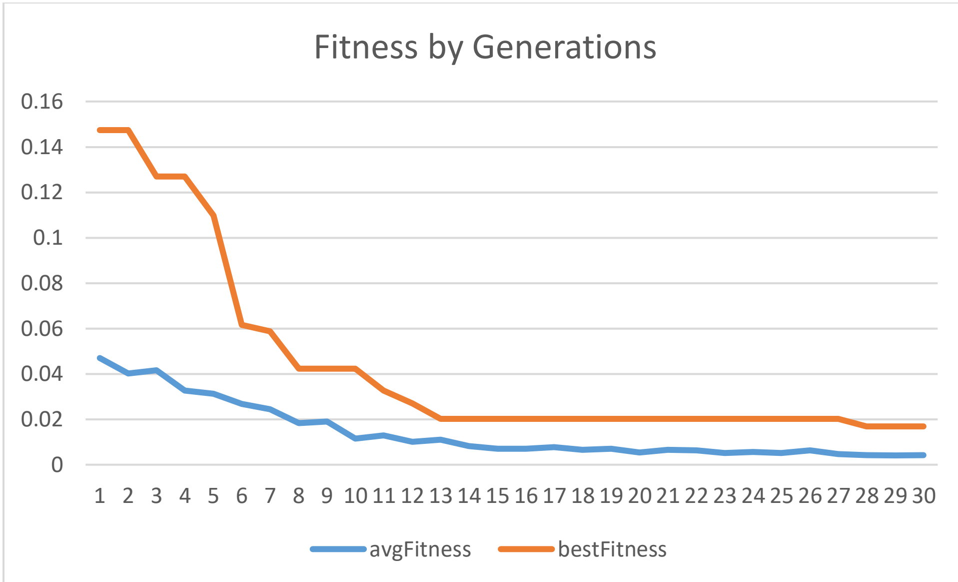


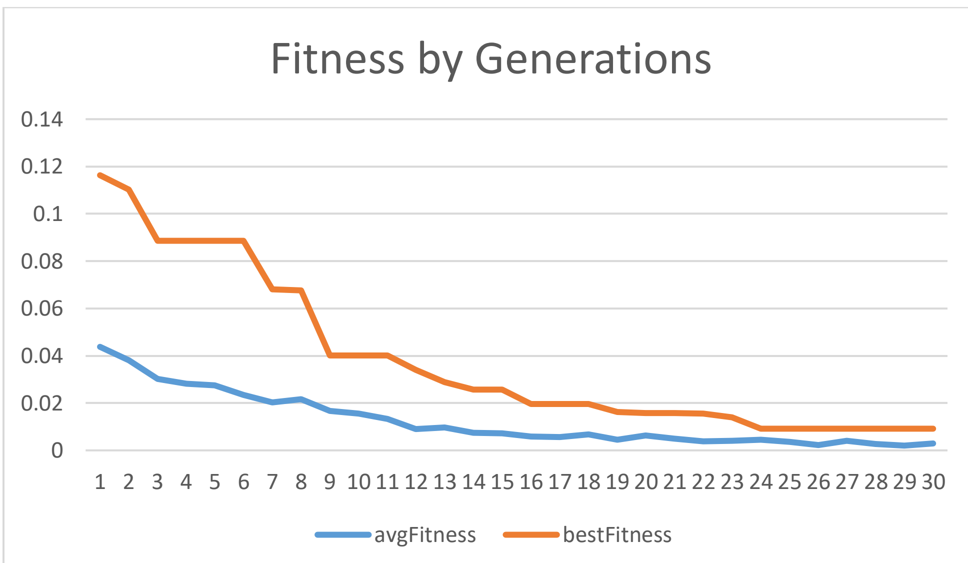
Result Analysis

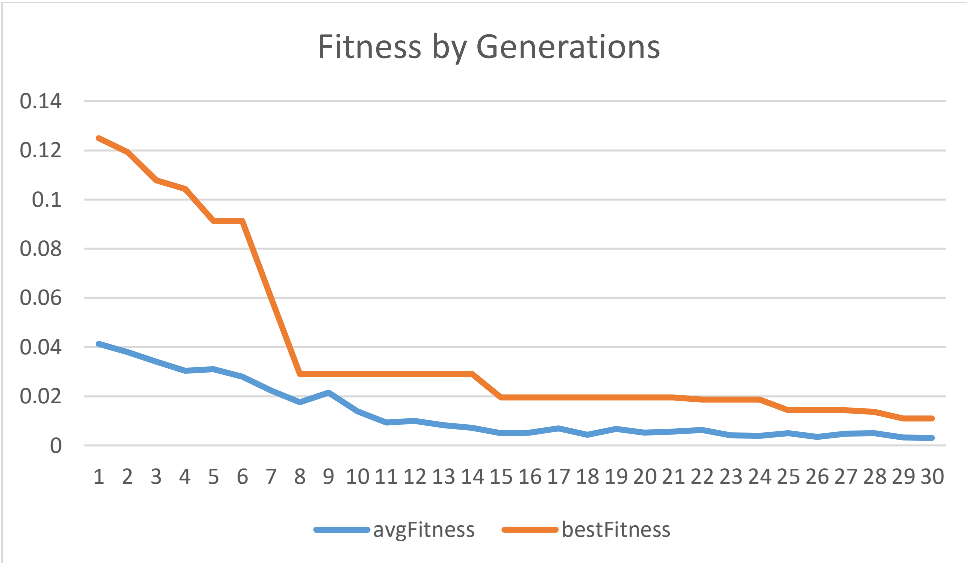
A. 30 Generation With 50 Population Size

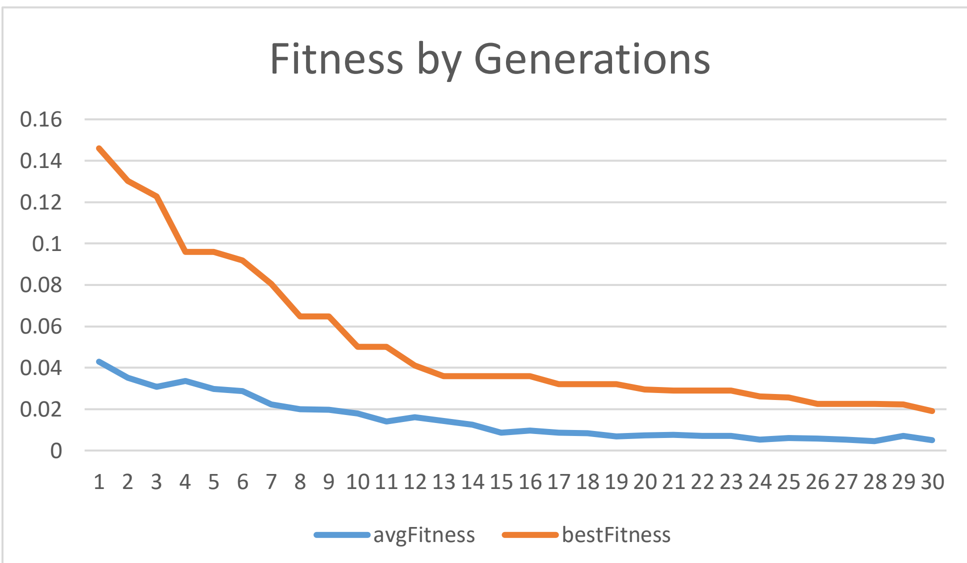
Crossover probability: 0.9

Mutation probability: 0.2


Fitness by Generations
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
avgFitness
bestFitness


Fitness by Generations
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
avgFitness
bestFitness


Fitness by Generations
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
avgFitness
bestFitness


Fitness by Generations
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
avgFitness
bestFitness

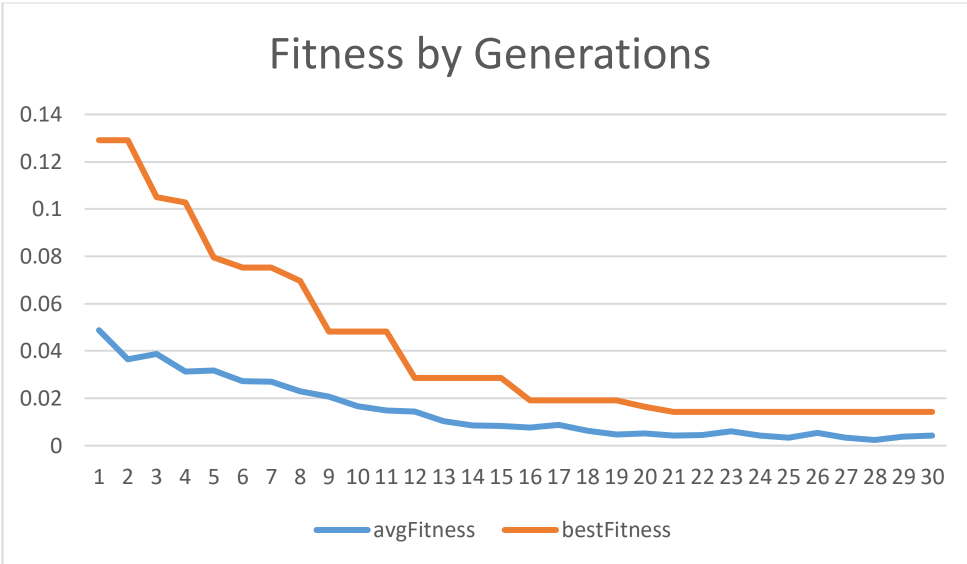
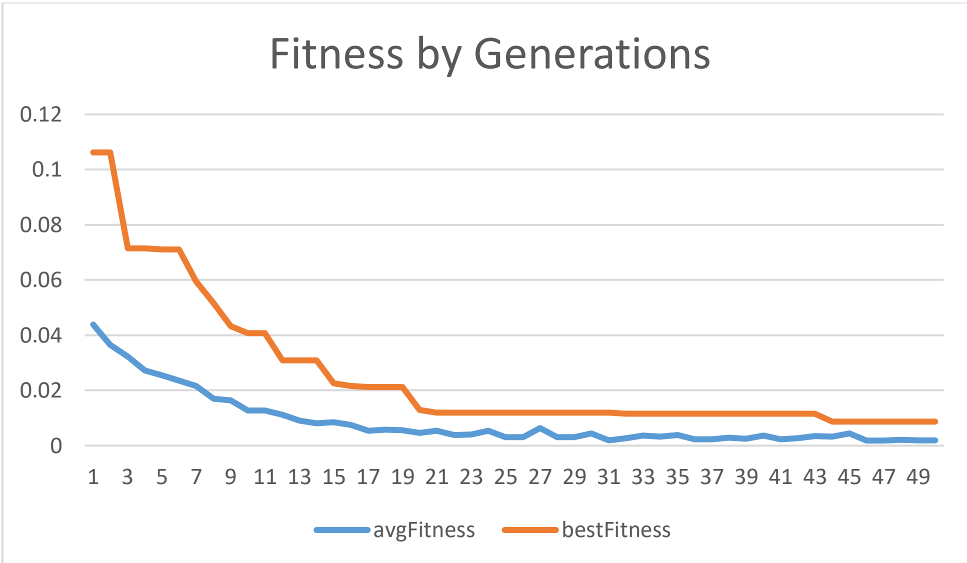


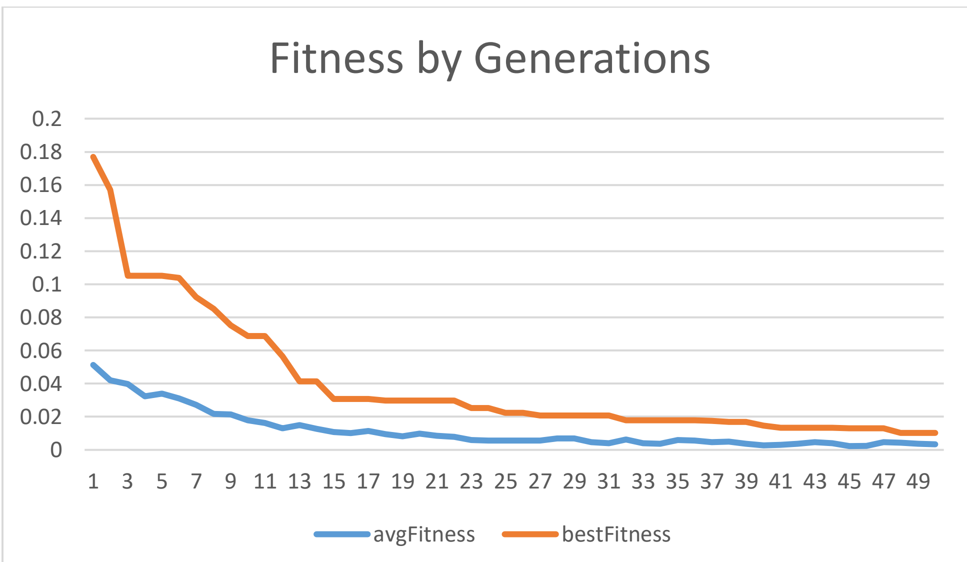
Figure above show the results of 30 generation with 50 population size. As we can see, reaching the end of generation, most of the graph if still have gap between them. Therefore, we considered this result from 30 generation is not good and still need improvement. Thus, we continue to generate result using 40 and 50 generation. The best fitness recorded is 0.01907 with average fitness recorded is 0.004955.

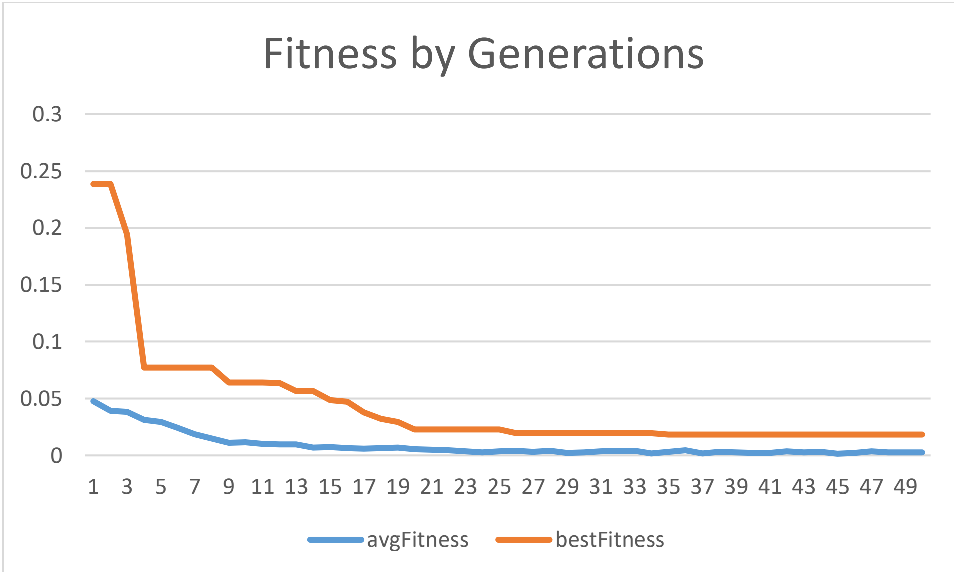
B. **50 Generations With 50 Population Size**

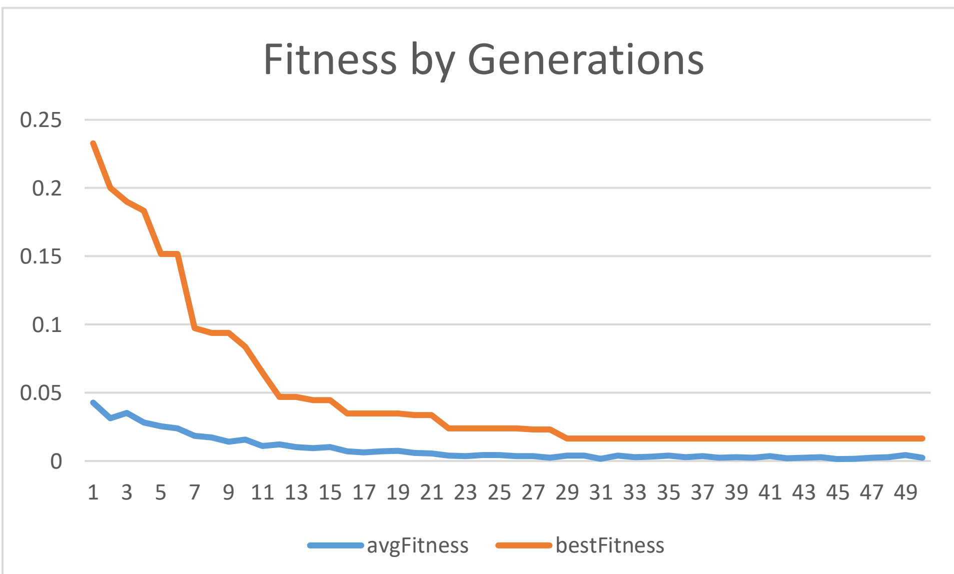
Crossover probability: 0.9

Mutation probability: 0.2


Fitness by Generations
0.12
0.1
0.08
0.06
0.04
0.02
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
avgFitness
bestFitness


Fitness by Generations
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
avgFitness
bestFitness


Fitness by Generations
0.3
0.25
0.2
0.15
0.1
0.05
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
avgFitness
bestFitness


Fitness by Generations
0.25
0.2
0.15
0.1
0.05
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
avgFitness
bestFitness

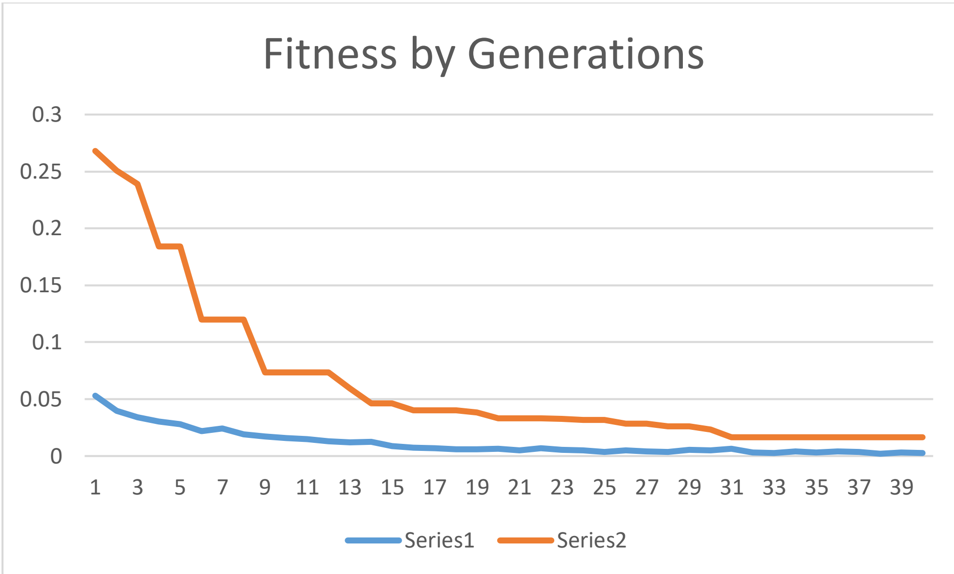
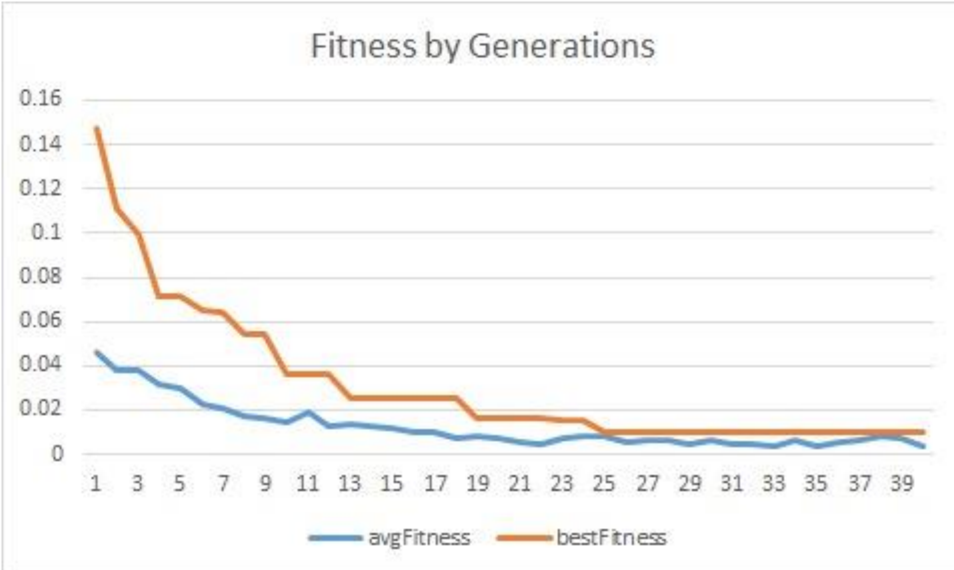


Figure above show the result for running 40 generation with 50 population size. As the second analysis result, we could see an improvement in the ending gap which is it managed to get smaller the previous 30 generation result, but it is still considered not good. Plus, we could see there are many 'lucky' scenarios in the graph. Whereas, there are line that go down instantly where it supposed to go down constantly. The best fitness recorded is 0.010234 with average fitness recorded is 0.003269. Therefore, we considered to conduct another result analysis using 40 generation.

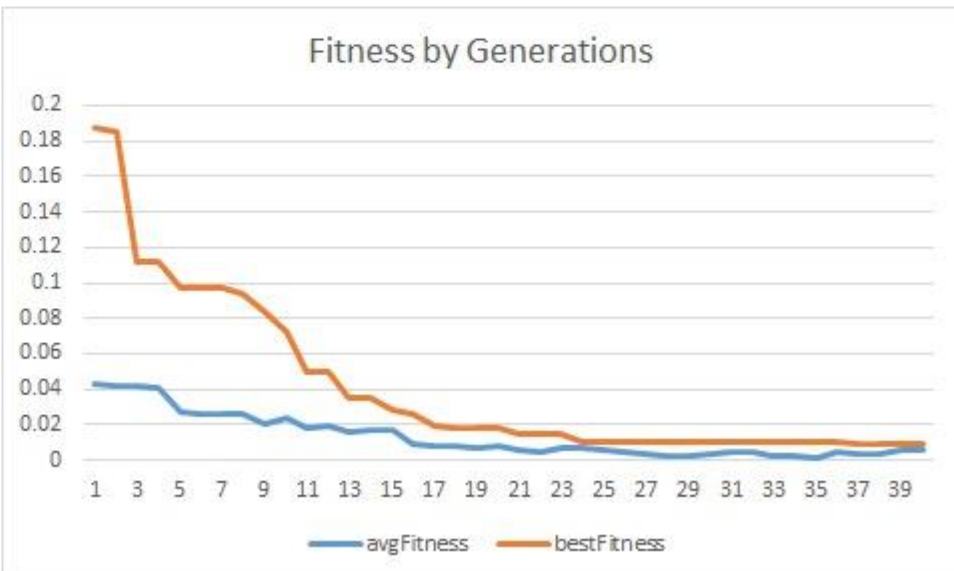
C. **40 Generation With 50 Population Size**

Crossover probability: 0.9

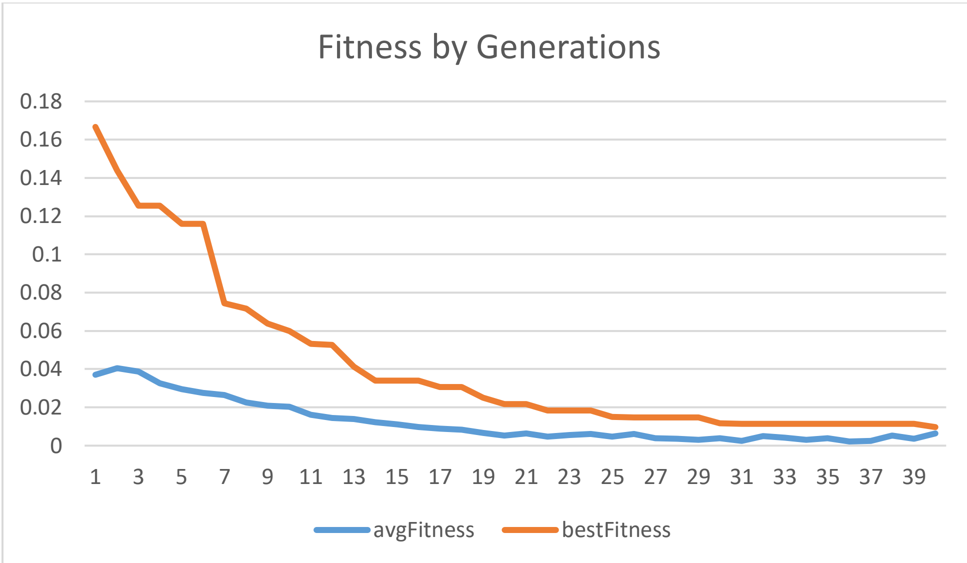
Mutation probability: 0.2

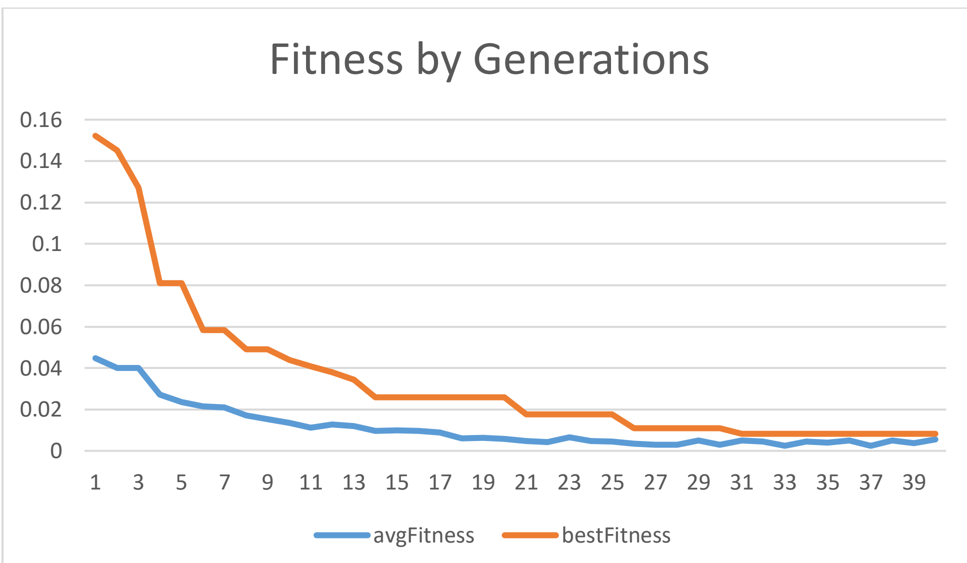


Fitness by Generations
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39
avgFitness
bestFitness



Fitness by Generations
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39
avgFitness
bestFitness


Fitness by Generations
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39
avgFitness
bestFitness


Fitness by Generations
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39
avgFitness
bestFitness

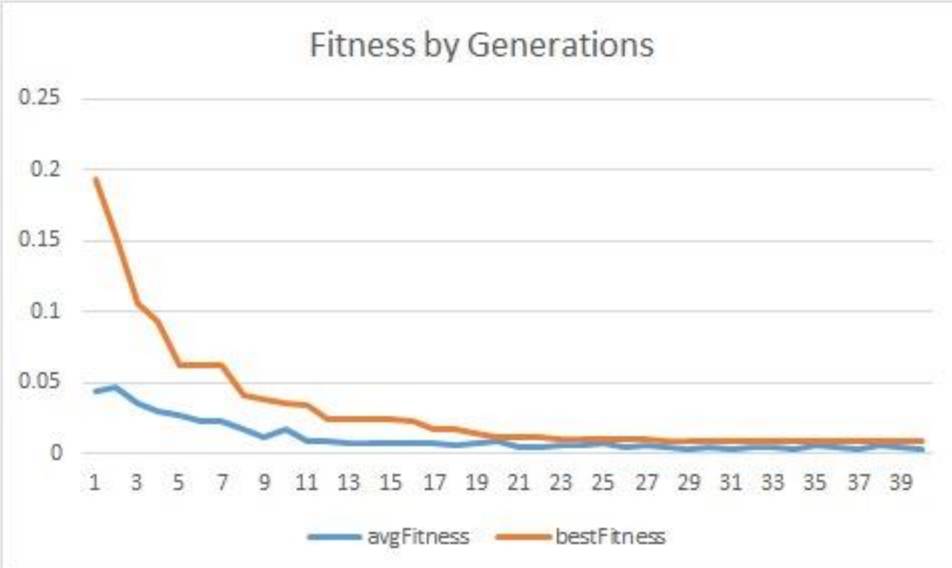


Figure above show the result of 40 generation with 50 population size. Considering the previous result analysis with 30 and 50 generation, this is the best result so far. The best fitness recorded is 0.008703 with average fitness recorded is 0.00200642. We could see that the gap is decreasing and the line go down consistently. Therefore, this is recorded as the best generation evolution.

D. Algorithm Analysis

In this project, we managed to get the best fitness is 0.008703 while the average fitness is 0.00200642. The stated result is obtained from the result of 40 generation with 50 population size. Based on the previous graph in result analysis,

The Best Solution

Run	Best Fitness	Run	Best Fitness
1	0.009692	**11**	0.016851
2	0.008053	**12**	0.009159
3	0.009985	**13**	0.011994
4	0.008586	**14**	0.011408
5	0.01047	**15**	0.012954
6	0.008247	**16**	0.008247
7	0.015202	**17**	0.012498
8	0.019182	**18**	0.009236
9	0.009692	**19**	0.008703
10	0.018009	**20**	0.012929

Sites	1	2	3	4	5	6	7	8
No Substation	24	5	5	29	29	24	37	11

The best solution

POP_SIZE = 50

MAX_GENERATION = 40

FITNESS = (Distance fitness * 0.5) + (Cost fitness * 0.5) = 0.008703

This is the best result that we can conclude from all the 20 runs test that we compile. This means that in sites 1 until 8, only 6 substations which are substations number 24, 5, 29, 24, 37 and 11 will be placed in this rural area.