

CHAPTER 4

ANALYSIS AND DESIGN

4.1. Analysis

Genetic algorithm is widely known as an optimization algorithm with natural selections. Genetic algorithm in this research will be used to optimize the waste transportation route in Tangerang city. The route will be described as a traveling salesman problem or TSP; here genetic algorithm will find a solution to the traveling salesman problem, by finding the shortest path between each TPS. The goal of this problem is to find the most optimal route and the final destination being the first destination.

4.2. Pre-processing Dataset

Data that has been taken from the website will be processed so it is more suitable for the optimizing phase. Since the original data only shows the table of coordinates, there is a process of making nodes from those coordinates alone.

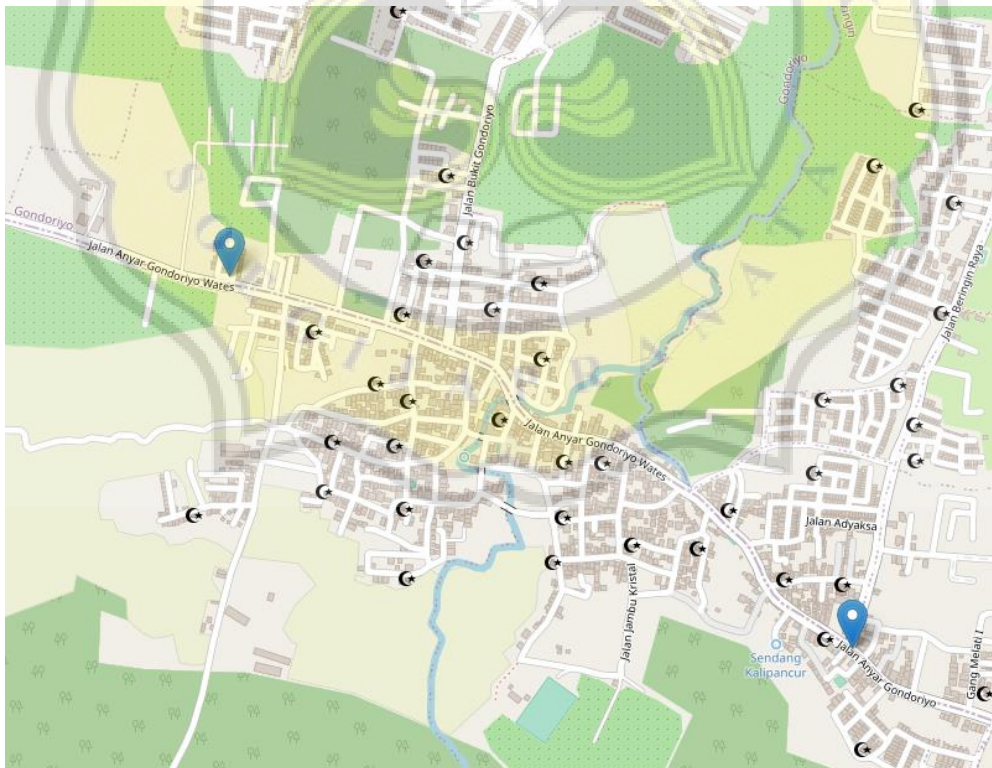


Figure 4.1 Nodes from Data

From here onward a dataset has been created, the dataset consist of the number of node, the longitude and also the latitude. The total dataset is 212. Table 5.1 is the preview of the dataset.

Table 4.1. Table 1 Dataset Examples

index	Y	X
0	-6.189553	106.638313
1	-6.191069	106.633267
2	-6.184849	106.648847
3	-6.174996	106.650157
4	-6.189553	106.638313
...
212	-6.225502	106.637648

4.3. Processing Data

After creating the dataset, path between nodes can be created. This path is used to only for visualization on how messy the route is without genetic algorithm. The path created is based on the nodes number, not by the nearest node available; this is why the path is messy and unorganized.

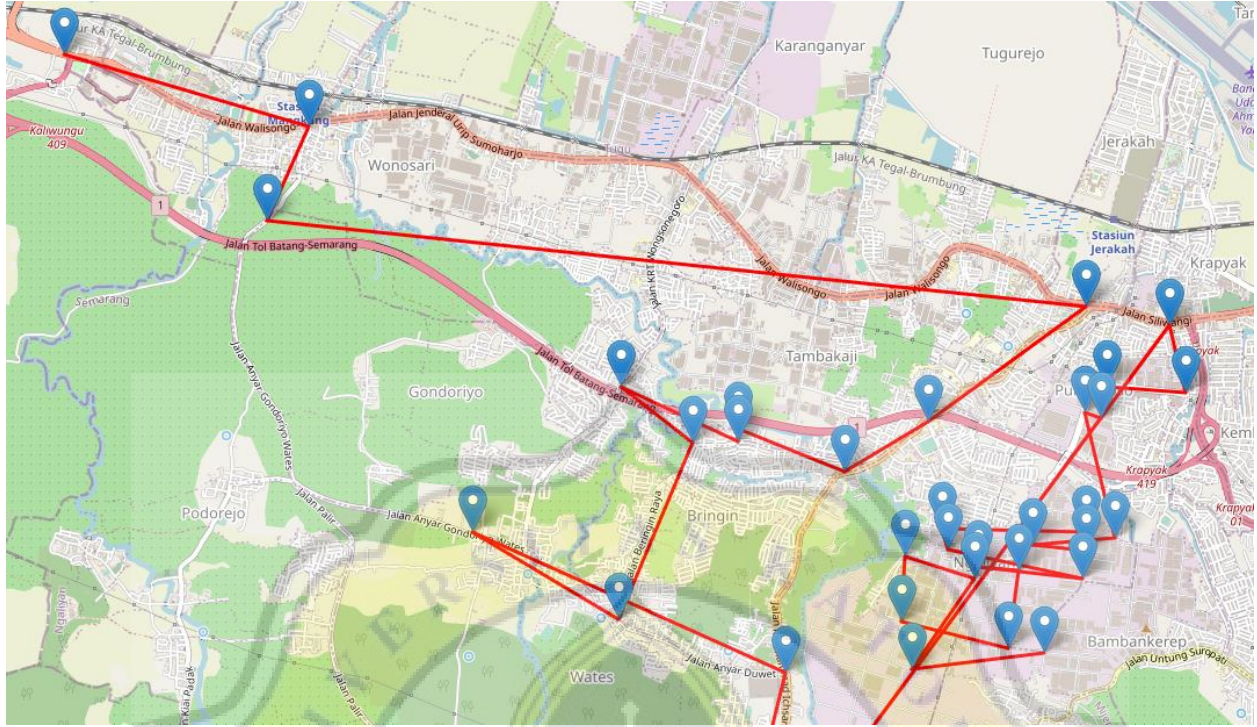


Figure 4.1 Original Path

The length of this path will also be calculated, this has the purpose to estimate the initial length and will it shortened after genetic algorithm has been conducted. To calculate the length from one node to another a formula called *Haversine Formula* will be used.

$$d = 2r \arcsin \sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos\varphi_1 * \cos\varphi_2 * \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}$$

Where d is the distance between two coordinates, r is the radius of the sphere, φ_2, φ_1 are the latitude from one point to another point, and the λ_2, λ_1 are the longitude from one point to another. An example from the data, calculating the distance between index [0] and index [1];

Table 4.2. Data Examples

index	X	Y
0	-6.96829	110.2848411
1	-6.973521	110.302996

From this data, the longitude and also latitude has been found, with the value;

$$\varphi_1, \lambda_1 = -6.96829, 110.2848411$$

$$\varphi_2, \lambda_2 = -6.973521, 110.302996$$

After that the latitude and also longitude can be inserted into the formula to be calculated, since r is the radius of the sphere, and in this case the sphere is the earth, the radius of the earth must be inserted, since the radius of the earth varies from 6356.752 km at the poles to 6378.137 km at the equator the one that will be used as r is 6371, next is to input the values into the formula and solve the equation.

$$d = 2r \arcsin \sqrt{\sin^2\left(\frac{-0.005231}{2}\right) + \cos(-6.96829) * \cos(-6.973521) * \sin^2\left(\frac{0.0181549}{2}\right)}$$

$$d = 2.0865244$$

With this formula the distance from one node to another node can be calculated, but to know the length of the overall pipeline, all of the distance must be added.

4.4. Genetic Algorithm

After getting the rough result of the route length, genetic algorithm will be used to solve the traveling salesman problem. The solution of genetic algorithm is represented by chromosome, this chromosome is made from a set of genes that represents solutions to unsolved problems. There are a few main process of genetic algorithm after initializing the parameter,

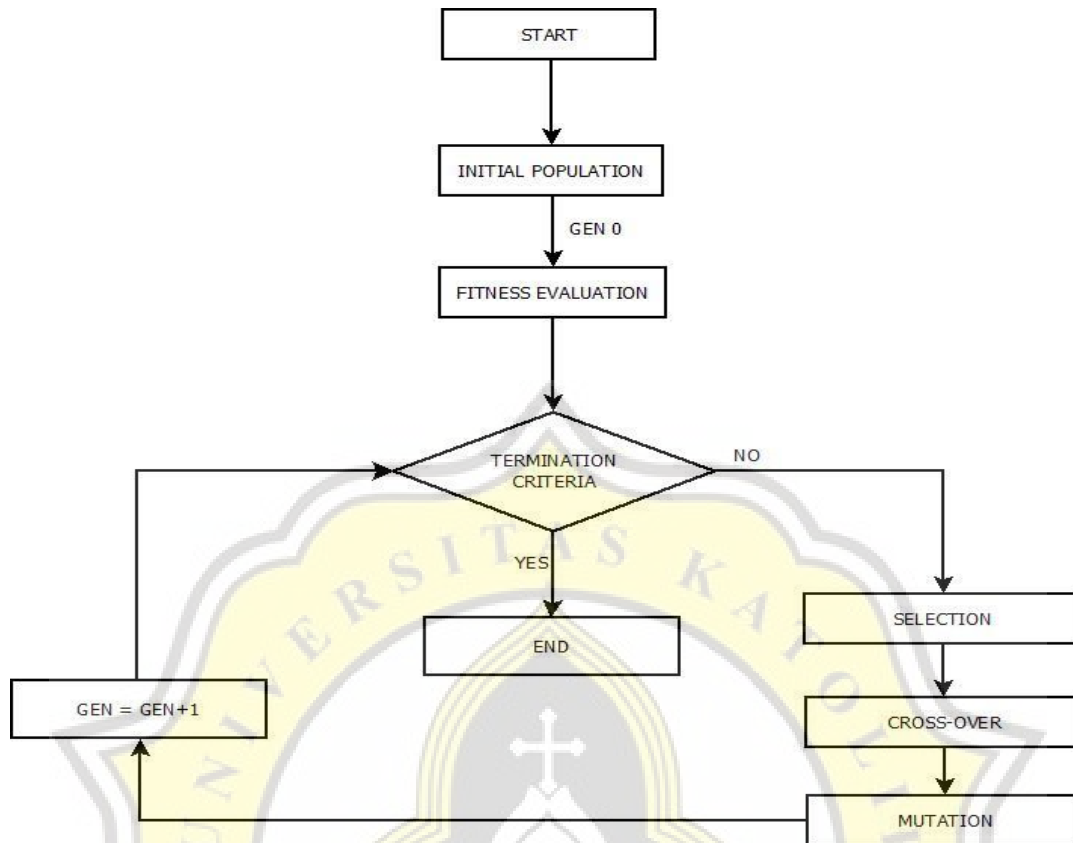


Figure 4.1 Genetic Algorithm Process

4.4.1. Initial Population

Initial population is the first step in the process of genetic algorithm, without the initial population the process cannot go forward; this is why this step is important. Population in genetic algorithm can be defined as a set of chromosomes. Here the set of genes is represented by strings.

The initial population in this research will be based from the index number of nodes that has been created in the beginning. In this research the index number consist from the 0 – 481, but for example index 0 – 40 will be displayed;

[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, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40].

To create the population the first thing that's needed to do is to determine how much population is needed. In this research the number of population is 200, from here the original index number from 0 – 40 is considered as one chromosome, and the content of that chromosome is called gene, to get the initial population the gene of a chromosome will be randomized,

Table 4.3. Chromosome and Gene(s) Examples

Chromosome	Gene(s)
0	[30, 16, 9, 28, 4, 18, 29, 39, 15, 27, 26, 11, 31, 33, 13, 35, 25, 38, 5, 1, 32, 2, 22, 17, 40, 14, 7, 12, 20, 37, 6, 23, 36, 34, 10, 19, 21, 0, 8, 3, 24]
1	[22, 29, 26, 6, 15, 8, 35, 19, 9, 2, 12, 27, 34, 17, 20, 40, 5, 13, 30, 39, 31, 36, 24, 37, 25, 16, 3, 4, 11, 10, 23, 32, 21, 1, 33, 7, 18, 38, 0, 28, 14]
2	[0, 28, 5, 8, 34, 22, 6, 15, 39, 18, 30, 14, 17, 35, 37, 13, 3, 1, 7, 2, 26, 19, 31, 25, 4, 12, 38, 16, 40, 20, 11, 23, 24, 33, 9, 32, 29, 21, 27, 36, 10]
...	
200	[29, 27, 12, 20, 30, 2, 18, 35, 1, 14, 3, 16, 25, 6, 5, 8, 10, 37, 38, 0, 34, 36, 40, 4, 13, 7, 19, 17, 9, 22, 11, 28, 15, 26, 21, 33, 24, 23, 32, 31, 39]

This is only the example of the initial population; if the number of population is set to 200 then the gene of a chromosome will be randomized until 200 times.

4.4.2. Fitness Evaluation

The fitness evaluation has the function to determine whether an individual is fit for the process or not. If the individual is not able to compete with the other then it will be considered not fit for the job. In this research, first thing to do is to calculate the distance matrix of each node. By finding the Euclidean distance, it will give us the length of the line segment. The formula for Euclidean distance is;

$$d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$$

Where d is the distance between the coordinates, (x_1, y_1) are the coordinates of one point and (x_2, y_2) are the coordinates of the other point. For example;

$$I_1 = (-6.96829, 110.2848411)$$

$$I_2 = (-6.973521, 110.302996)$$

$$d = \sqrt{[(-6.973521 - (-6.96829))^2 + [(110.302996 - 110.2848411)^2]}$$

$$d = 0,01889348$$

This calculation is done until the matrix is complete. From this matrix distance between nodes can be found. For example the distance between index [0] and index [1] is 0.018893484459198227. From here the fitness of the entire chromosome in the population can be calculated. After creating the distance matrix, the initial population will be calculated with the distance of the nodes, so it can create a fitness value. The next step is to determine the minimum fitness value. By setting the fitness value, the chromosome that does not fit can be eliminated.

4.4.3. Termination Criteria

The termination criteria could be used to stop the genetic algorithm process altogether or to continue to process, this is determined by some common criteria such as if a solution from the minimum criteria has been achieved, or the limit number of generation has been reached, or the computation time is too long, or maybe all of the above. If the termination criteria has not been met, then the process will continue into selection. For example, in this research the termination criteria have been limited to 1000 generations. So after generating the fitness 1000 times the process will stop.

4.4.4. Selection

Selection is one of three core processes of genetic algorithm. The purpose of this process is to select individuals from the current generations which will be used for the next one. Based on their fitness score, two pairs of individuals will be selected for the next process, which is crossover. Selection that will be used in this research is the tournament selection, based on the trial and error of the data; this specific selection method gives the best result and also shorter computational time compared to another method.

Based on the figure below, the result of tournament is proven to be producing the best fitness. All of the method uses the same parameter of 40 nodes, 200 population, 1000 generations, 0.6 crossover probability, and also 0.4 mutation probability.

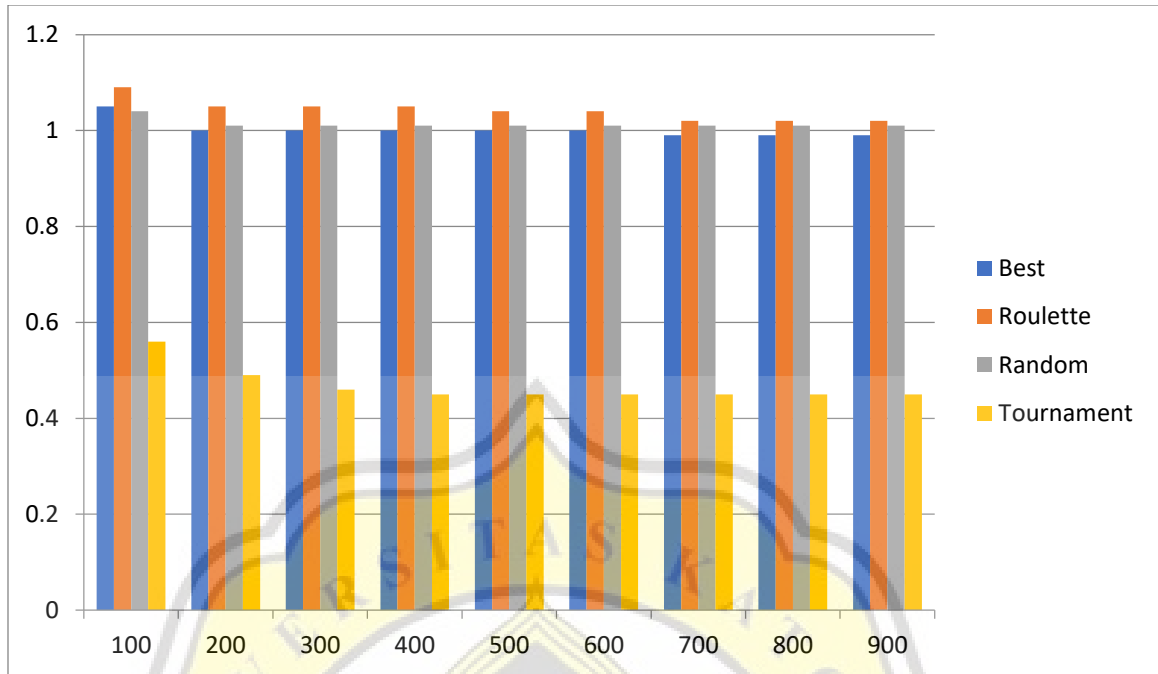


Figure 4.1 Selection Method Comparison

4.4.5. Crossover

Crossover is a process where the gene/s of a chromosome is swapped with another gene/s from a different chromosome; crossover will stop when the swapping process reached its crossover point. There is a crossover probability, that determines will a chromosome endure crossover or not, the common crossover probability is 60%, from here, a random number that range from 0.0 until 1 called i will be generated. i here will represent each chromosome available.

$$Crossover = IF(i < cross_{prob}; true; false)$$

Where i is a random number between 0.0 until 1 and crossprob is the probability that has been determined, which is 80% / 0.8. Crossover probability is chosen based on trial and error from the data, since 0.8 provide the best result, this number is chosen, all of the trial have the same parameter of 40 nodes, 200 population, 1000 generations, and 0.4 mutation probability.

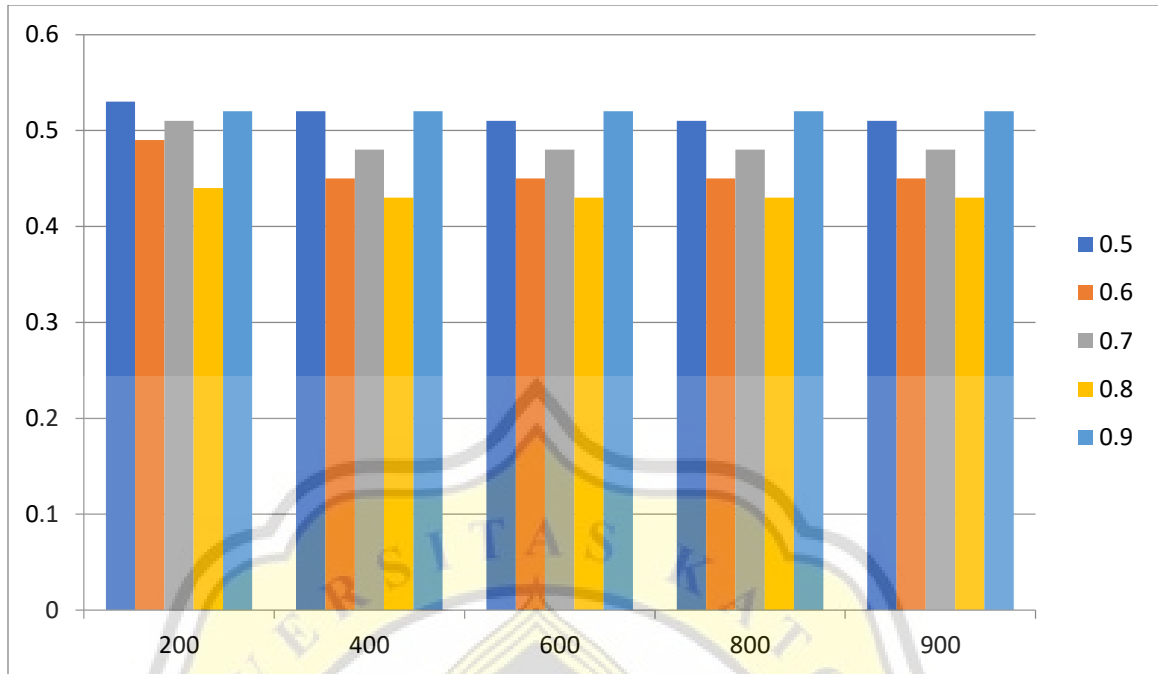


Figure 4.1 Crossover Probability Comparison

If value < crossover probability then crossover will happen, otherwise the crossover will not occur in that specific chromosome, for example;

Table 4.4. Crossover Probability

Chromosome	<i>i</i> (Random)	Crossover
0	0.9	No
1	0.3	Yes
2	0.9	No
...		
200	0.1	Yes

Next step is to actually do the crossover. The list of chromosome will be sorted by the fitness value, then for a chromosome that has higher *i* value then crossover probability, that chromosome will be crossover with another chromosome that fits the qualifications. A crossover point will be generated randomly; the function of this point is to stop the process of crossover, for example;

Table 4.5. Before and After Crossover Comparison

Before Crossover		
Chromosome	<i>i</i> (Random)	Gene(s)
1	0.3	[22, 29, 26, 6, 15, 8, 35, 19, 9, 2, 12, 27, 34, 17, 20, 40, 5, 13,

		30, 39, 31, 36, 24, 37, 25, 16, 3, 4, 11, 10, 23, 32, 21, 1, 33, 7, 18, 38, 0, 28, 14]
200	0.1	[29, 27, 12, 20, 30 , 2, 18, 35, 1, 14, 3, 16, 25, 6, 5, 8, 10, 37, 38, 0, 34, 36, 40, 4, 13, 7, 19, 17, 9, 22, 11, 28, 15, 26, 21, 33, 24, 23, 32, 31, 39]
After Crossover		
1	0.3	[29, 27, 12, 20, 30 , 8, 35, 19, 9, 2, 12, 27, 34, 17, 20, 40, 5, 13, 30, 39, 31, 36, 24, 37, 25, 16, 3, 4, 11, 10, 23, 32, 21, 1, 33, 7, 18, 38, 0, 28, 14]
200	0.1	[22, 29, 26, 6, 15 , 2, 18, 35, 1, 14, 3, 16, 25, 6, 5, 8, 10, 37, 38, 0, 34, 36, 40, 4, 13, 7, 19, 17, 9, 22, 11, 28, 15, 26, 21, 33, 24, 23, 32, 31, 39]

After doing this process, the chromosome will then go through mutation as the final process of the generation.

4.4.6. Mutation

Mutation is the process of tweaking gene in a chromosome; the function of this process is to create a new solution to the problem, in this research, mutation has the function to create a new path that maybe has not been found yet. Since mutation has many variants, the one that will be used in this research is the swap mutation, where two random genes will be swapped.

$$Mutation = IF(e < mutationprob; true; false)$$

Same with crossover, mutation will also have its own mutation probability. The common mutation probability is 0.4 or 40%, also the same process with crossover, mutation will generate a random number in the range of 0.0 until 1 that represent each chromosome, called e . The mutation probability is chosen based on the best result from the trial and error, which will be shown below; the parameter is all the same, with 40 nodes, 1000 generations, and also 0.6 crossover probability.

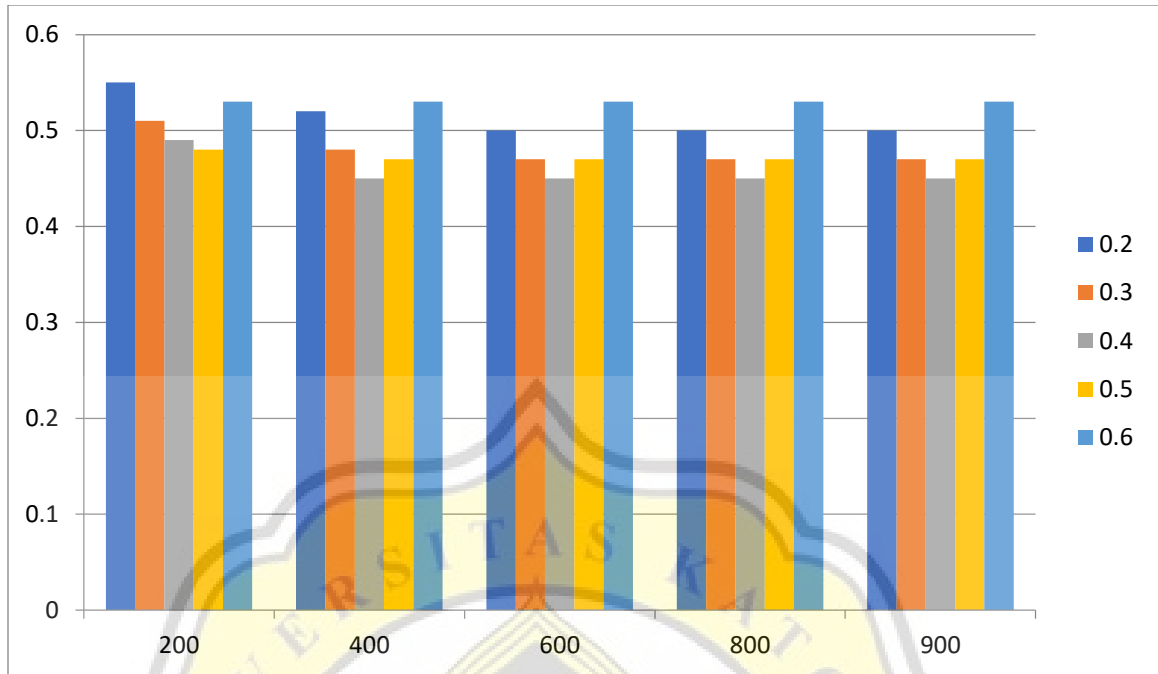


Figure 4.1 Mutation Probability Comparison

Table 4.6. Mutation Probability and Comparison

Chromosome	e(Random)	Mutation
1	0.1	Yes
200	0.5	No
Before Mutation		
Chromosome	e (Random)	Gene(s)
1	0.1	[29, 27, 12, 20, 30, 8, 35, 19 , 9, 2, 12, 27, 34, 17, 20, 40, 5, 13, 30, 39, 31, 36, 24, 37, 25, 16, 3, 4, 11 , 10, 23, 32, 21, 1, 33, 7, 18, 38, 0, 28, 14]
After Mutation		
1	0.1	[29, 27, 12, 20, 30, 8, 35, 11 , 9, 2, 12, 27, 34, 17, 20, 40, 5, 13, 30, 39, 31, 36, 24, 37, 25, 16, 3, 4, 19 , 10, 23, 32, 21, 1, 33, 7, 18, 38, 0, 28, 14]

After doing this process the chromosome will now get a new set of genes, this set of genes will be tested for its fitness value, if the fitness value turns out to be bad, then the result of the crossover and mutation will be deleted, otherwise the new chromosome will replace the previous best solution. This process is then repeated again and again until the maximum number of generations has been reached. After that, the best solution is obtained.

4.5. Evaluating Data

After the process of genetic algorithm has ended, an optimum route based on genetic algorithm has been produced, the route will be compared with brute force. Brute force can be counted as another optimization algorithm that solves problem in a simple way that relies on pure computational power and tries all possibilities rather than sophisticated techniques to improve the efficiency. For brute force to optimize a problem, it needs to find all the possible answers to produce the most optimum solutions.

The result of the path and also the distance of both algorithm will be shown, this has the purpose to compare both algorithm, which algorithm produces the most optimum route.

