CHAPTER 6

MINIMISATION OF DISTRIBUTION COST USING K-MEANS AND CHEAPEST LINK ALGORITHM

The work of this chapter is done to explore if the new routes can reduce the distribution cost further. To identify the new routes the k-means clustering, and Cheapest Link Algorithm are used. Here, k-means clustering is clustering method and Cheapest Link Algorithm is applied for Vehicle Routing Problem method.

6.1 Methodology and Algorithm

In this part the Capacitated K-means Clustering was used to split delivery locations into similar size groups (i.e. clusters) based on proximity without exceeding a specified total cluster capacity. Each cluster was to be served by a local stockist. The CLA was then used to find delivery routes from dairy (i.e. depot) to stockist in each cluster and from stockist to all other delivery locations within the cluster.

The work is done in two phases. These phases are described as follows:

a) Clustering: To divide the problem into smaller parts, the service locations were grouped into clusters (Shieh and May, 2001). Delivery locations were grouped by their closeness to each other. Instead of solving the CVRP for all the delivery locations together, we solved the CVRP for each cluster separately, thus reducing complexity and increasing the optimality of our heuristic solution.

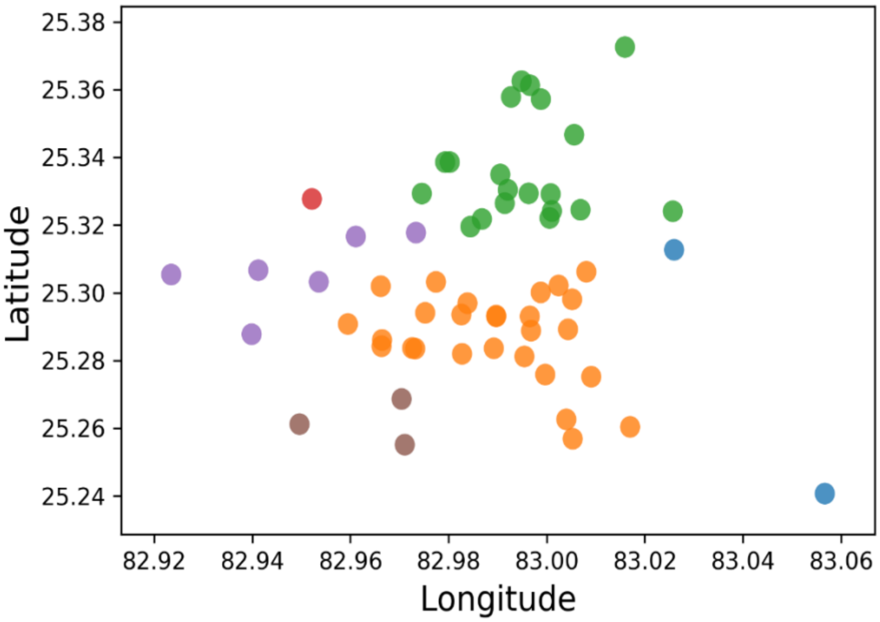
b) The CVRP was solved for
    a. the depot to cluster centres (stockists) and
    b. each cluster centre (stockist) to each delivery location within the cluster.

6.1.1 Clustering

The first phase in clustering was to partition service locations and the depot into clusters based on their proximity to each other. Numerous clustering methods exist, but not all of them proved suitable for this research. Our criteria for choosing a suitable clustering algorithm was, therefore, influenced by the following factors.

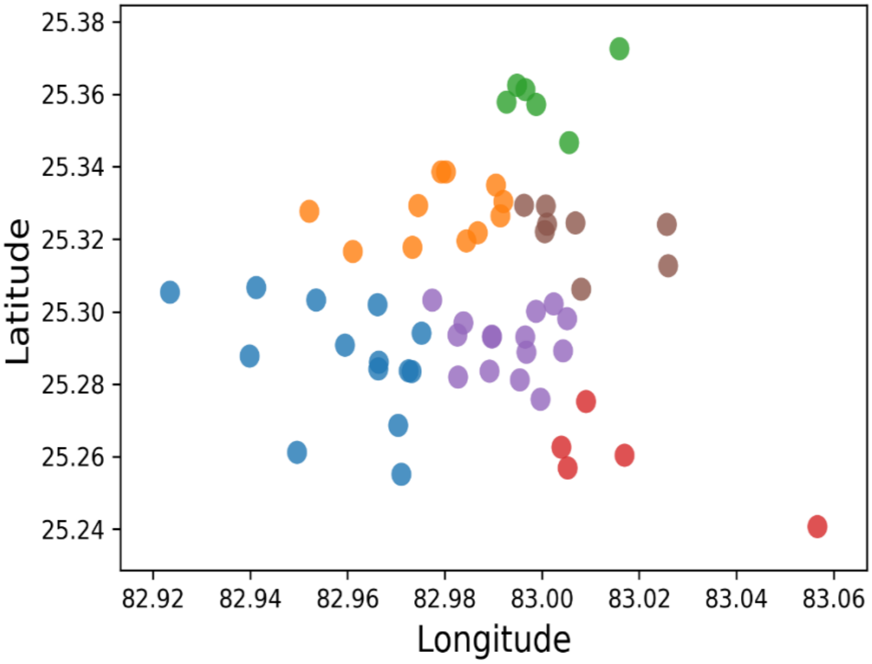
a) Even cluster sizes: If some clusters were small, then others would be big, and the larger the cluster, the more complicated it would be to solve the CVRP for it. Evenly distributed cluster sizes were, therefore, preferred.
b) Not too many clusters: Having too many clusters would make the CVRP for the depot to cluster centres complex. With increasing complexity, the optimality of the solution would diminish.

For comparison, both agglomerative and k-means clustering were used on the same dataset. The k-means clustering method produced better results for our case. Figures 6.1 and 6.2, in which each colour represents a cluster, depict this comparison visually.



$\sigma_{\text{cluster size}} = 9.99$

Figure 6.1: Agglomerative Clustering



$\sigma_{\text{cluster size}} = 3.8$

Figure 6.2: The K-means Clustering Method

Here, lower standard deviation values are better. Keeping all the factors in mind, we chose the K-means Clustering algorithm. In Figures 6.1 and 6.2, the longitude and latitude appear on the x-axis y-axis, respectively. The coloured points in the Figures represent the delivery locations. Locations grouped into the same cluster were assigned the same colour.

6.1.2 Capacitated Vehicle Routing Problem (CVRP)

After successfully grouping a total of 58 demand locations through K-means Clustering based on the proximity of all the demand points to each other, a total of six clusters were identified. Each cluster had a centre point designated as a stockist.

Furthermore, to find the best route to deliver milk from the depot to stockists and then to all other service locations, the work was done in two parts:

1. First, we had to find optimal routes for deliveries from the depot (dairy) to each cluster centre/stockist.
2. Then, we had to determine the optimal routes for deliveries from each stockist to service locations in that cluster.

The constraints used for the k-means clustering and the CLA are

1. capacity constraint,
2. distance constraint, and
3. demand constraint.

The pseudocode used for the problem is presented below:

$I = Total\ number\ of\ simulation\ instances$

$G = Number\ of\ clusters$

$C_j\ = \boldsymbol{jth}\ cluster$

$C\ = \ Set\ of\ all\ clusters$

$D_{ji}\ = \ Demand\ of\ location\ \boldsymbol{i}\ in\ cluster\ \mathbf{j}$

$Pji\ = \ Cost\ of\ delivery\ to\ all\ demand\ locations\ in\ cluster\ Cj\ in\ simulation\ instance\ i.$

$$P_i = \sum_{j=1}^{G} P_{ji}$$

$$P_m = \min(S)\ , S = \{P_1, P_2, P_3, \ldots., P_I\}$$

$AD_{max} = \ Maximum\ Capacity\ of\ large\ vehicle$

```
$P_m = \infty,\ i = 0$
while $i \leq I$ do
    createClusters()
    solveVRP_factoryToStockist()
    solveVRP_stockistToDeliveryPoints()
    $P_i = \sum_{j=1}^{G} P_{ji}$
end

return $P_m$ =min$(S)$,  $S = P_1, P_2, P_3, \ldots,\ P_I$
```

```
Procedure solveVRP_factoryToStockist():
    $depot\ = Dairy\ Factory$
    $locations\ = All\ Stockists$
    $D_j$ = solveVRP($C_j$)
    return
```

```
Procedure solveVRP_stockistToDeliveryPoints():
    for $C_j \leftarrow C$ do
        $depot\ = Cluster\ Center$
        $locations\ = All\ Locations\ in\ Cluster$
        $D_j$ = solveVRP($C_j$)
    end

    return
```

```
Procedure createClusters():
    $D_{max} = \infty;$
    while $D_{max} > AD_{max}$ do
        $Create\ G\ clusters\ using\ K\text{-}means;$
        $D_{max}\ = max(D);$
    end
    return
```

The algorithm is explained below in simple language for easier understanding.

Step 1. Minimum_Cost = Infinity; Iterations = 0

Step 2. Create clusters.

1. Create clusters using simple k-means clustering.
2. Calculate the total demand for each cluster.
3. If the total demand for any cluster is more than the capacity of the largest vehicle, then go to step 2.1.
4. If the cluster contains a factory, make the factory a cluster centre.

Step 3. Solve the VRP for factory to stockists (cluster centres).

1. Set capacity constraints for each vehicle type.
2. Set travel distance constraints for each vehicle type.
3. Set the demand of each stockist/cluster centre equal to the total demand of the cluster.
4. Solve the VRP with the milk factory as the depot and the stockists as delivery locations.

Step 4. Solve the VRP for each cluster separately (i.e. cluster centre to delivery points).

1. Set capacity constraints for each vehicle type.
2. Set travel distance constraints for each vehicle type.
3. Solve the VRP with the cluster centre as the depot and other points in the cluster as delivery locations.

Step 5. Calculate the total cost based on travel distance, vehicle labour, and stockist storage cost.

Step 6. If Total_Cost < Minimum_Cost, then assign Minimum_Cost = Total_Cost.

Step 7. Iterations = Iterations + 1.

Step 8. If Iterations < Max_Iterations, repeat step 2.

Some additional considerations were considered:

1. Large delivery trucks would only be used for delivery from the depot to stockists. Given the on-the-ground reality that large delivery trucks were hard to find for a reasonable rate, we had to limit their use in the overall delivery plan.
2. Small delivery trucks were cheaper and easier to find but had much less capacity than large trucks. The maximum distance they could travel was also less compared to large trucks.
3. We faced a time constraint of two hours from the depot to service locations.

The CLA was used to solve the CVRP. The flowchart for CLA (see Figure 6.3) illustrates how to create a Hamilton circuit (Berge, 1973), a set of directed edges (paths in the VRP) such that each vertex (delivery point) is visited only once. One can think of vertices as delivery locations and edges as the delivery route. The weights associated with each edge represent the cost of that route. In the context of delivery, the pseudocode in next page can find an optimal solution to the VRP.

```
Procedure solveVRP($Cluster$):
    $uncoloredPaths$ = Set of all possible paths between locations in
     $Cluster$
    $D = 0$
    while isNotEmpty($uncoloredPaths$) do
        $selectedPath$ = removeSmallest($uncoloredPaths$)
        $circuitComplete$ = isCircuitComplete($selectedPath$)
        $allLocationsVisited$ = areAllLocationsVisited($selectedPath$)
        if $circuitComplete$ and $allLocationsVisited$ then
            break
        end

        if $circuitComplete$ or
         $threeColoredPathsFromSingleLocation(selectedPath)$ then
            continue
        end

        color($selectedPath$)
        $D = D +$ cost($selectedPath$)
    end
    return $D$
```

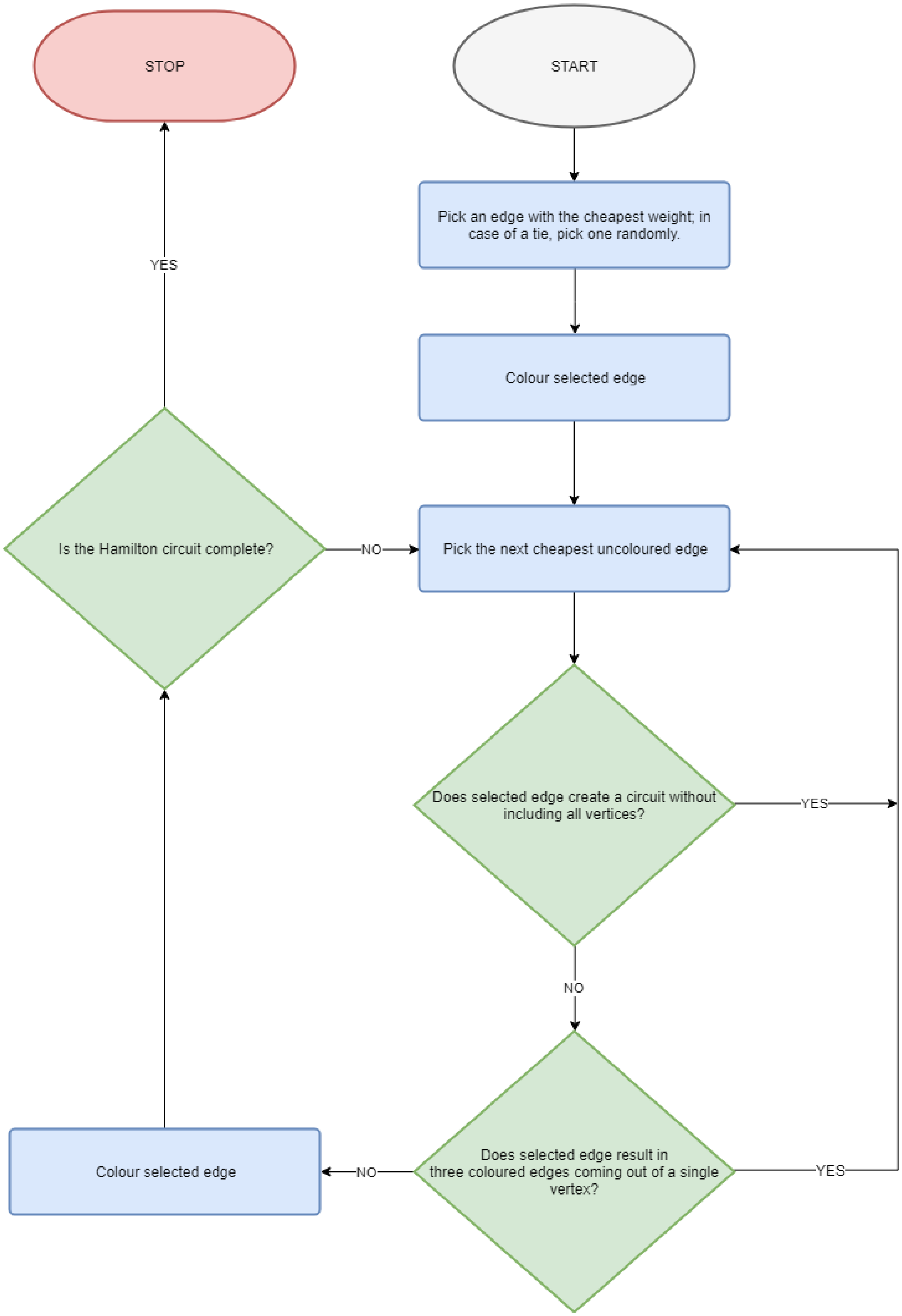


Figure 6.3: Flowchart for the Cheapest Link Algorithm (CLA)

6.2 Results

Figure 6.4 represents the clusters and routes made by k-means and the CLA. The algorithm identified a total of six clusters. Different colours represent the different clusters. Our problem consists of 58 total location points and point 0 represents the depot (dairy).

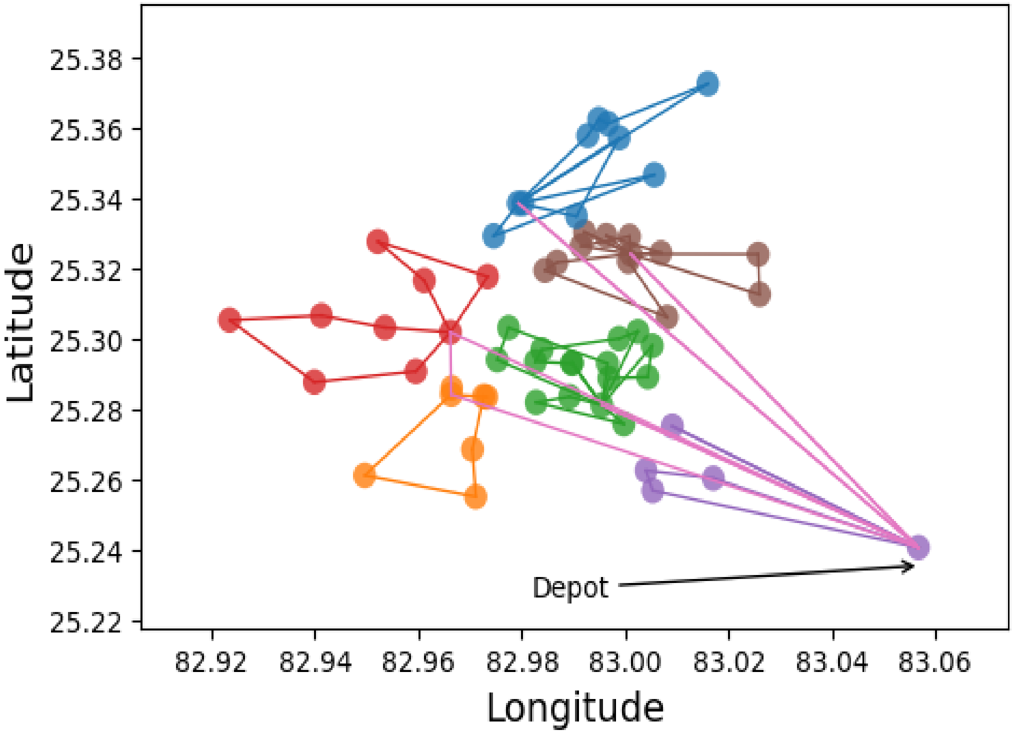


Figure 6.4: Delivery routes—Each Colour Represents A Cluster

The high- and low-capacity vehicles cost INR 25/km and INR 20/km, respectively, and the labour costs for high- and low-capacity vehicles was INR 300 and INR 100, respectively. The main source of high-capacity vehicles was the dairy (the depot), and the destinations were the stockists (the centre of each cluster). The source of low-capacity vehicles were the stockists (the centre of each cluster), and the destinations were other customers (the delivery locations). All the points had their own demands, and the co-operative dairy provided a refrigeration system to the points identified as stockists. The equated monthly instalment (EMI) for refrigerators amounted to INR 25/day for each stockist. Here, the Python programming language was used to solve the problem.

Table 6.1 displays the stockist points (1, 5, 40, 23, and 45) that supply the milk using high-capacity vehicles (2,700 litres). Here, point 0 is depot itself.

Table 6.1: Route for Larger Vehicles Leaving from the Dairy

Starting Point	Points Visited	Demand Fulfilled	Cost (Rupees)
0	1	2,490	1,264.50
0	5	2,691	1,088.52
0	40, 23	1,848	1,256.45
0	45	2,607	765.92
Total		9,636	4,375.39

Table 6.2 contains information on the supply from the stockist point—in this case, the point 0 (dairy/depot). Here, a low-capacity vehicle (500 litres) was used for the location point 44, along with another smaller vehicle for the location points 9, 12, and 43.

Table 6.2: Route for Smaller Vehicles Leaving from the Dairy

Starting Point	Points Visited	Demand Fulfilled	Cost (Rupees)
0	44	252	394.40
0	9, 12, 43	403	480.92
Total		655	875.32

Similarly, Tables 6.3, 6.4, 6.5, 6.6 and 6.7 display the supply of milk from the stockist points 1, 5, 23, 40, and 45, respectively, to smaller retailer points with low-capacity (500 litres) vehicles.

For example, Table 6.3 lists Figures regarding the supply of milk from the stockist point 1 to other smaller retailer points (7, 57, 54, 58, 2, 56, 55, 51, and 8).

Table 6.3: Route from Stockist Point 1

Starting Point	Points Visited	Demand Fulfilled	Cost (Rupees)
1	7	148	113.24
1	57, 54	498	305.74
1	58, 2, 56, 55	437	426.36
1	51, 8	454	309.46
Total		1,537	1,154.80

Table 6.4: Route from Stockist Point 5

Starting Point	Points Visited	Demand Fulfilled	Cost (Rupees)
5	52, 53	472	509.40
5	3, 10	406	176.64
5	6	486	166.54
5	18, 19, 11, 20	368	291.64
5	21, 4	463	166.12
Total		2,195	1,310.34

Table 6.5: Route from Stockist Point 23

Starting Point	Points Visited	Demand Fulfilled	Cost (Rupees)
23	13, 14, 32	261	495.46
23	38, 16, 17, 25, 24	383	393.60
Total		644	889.06

Table 6.6: Route from Stockist Point 40

Starting Point	Points visited	Demand Fulfilled	Cost (Rupees)
40	39	203	118.03
40	33, 30, 35, 34, 41	476	396.12
Total		679	514.16

Table 6.7: Route from Stockist Point 45

Starting Point	Points visited	Demand Fulfilled	Cost (Rupees)
45	37, 22, 46	459	332.04
45	49, 36, 50	499	241.48
45	47	353	188.06
45	27, 26, 15, 31, 42	450	289.62
45	48, 29, 28	374	217.20
Total		2,135	1,268.40

Table 6.8: Total Delivery Cost

Points	*0*	*1*	*5*	*23*	*40*	*45*	*Gross Total*
Total Cost	*5,250.71*	*1,154.80*	*1,310.34*	*889.06*	*514.16*	*1,268.40*	*10,387.46*

Table 6.8 contains data on the gross total cost of delivery per programme. The total calculated cost of the delivery from the k-means clustering and CVRP methods was INR 10,387.46—INR 2,990.84 lower than the present delivery cost of INR 13,378.3 (from Table 5.1), which is a reduction of approximately 22.35% per day in delivery costs. Thus, the new route yielded significant savings in distribution costs, which could represent substantial savings for the co-operative dairy.

The k-means clustering and the CLA methods suggested optimal routes with the algorithm runtime in seconds. A short algorithm runtime means that new delivery locations can be added or removed, and new delivery routes can be calculated as frequently as needed. Here, the average runtime of the complete programme was approximately 30 seconds (Code in **Appendix F**).

6.3 Conclusions

In previous chapter the existing routes being used by the co-operative dairy were examined. Therefore, to check if new routes could reduce the distribution cost further, new routes identification in this chapter using the Capacitated Clustering method (k-means) and the Capacitated Vehicle Routing Problem method (Cheapest Link Algorithm). The clustering model (k-means clustering) and CVRP (Cheapest Link Algorithm) were implemented with the constraints of distance, vehicle capacity, and demand. Initially, clustering was completed with the k-means clustering method because it generated evenly sized clusters, and then the CVRP was solved for the clusters and for the depot to cluster centres. This approach led to a reduction of 22.35% in distribution costs per day, resulting in significant savings. The research revealed that vehicle routing and clustering can significantly reduce the distribution costs in an organisation (Blumenfeld et al., 1987). Applying clustering to divide a problem can result in short algorithm runtimes and optimal solutions.