Vector quantization is a technique for grouping many collections of vector (data) into groups which each group has a number of data (vector) that have the same or close value.

Sample case:

In this project I decided to make 3 groups of customer. They are :

1. Regular Customer

2. Potential Customer

3. VIP Customer

For example, we have 9 customers who has different total purchase for 1 month. So it means that at the end of every month, we have total purchase for each customer.

Customer 1 total purchase is Rp. 100.000

Customer 2 total purchase is Rp. 200.000

Customer 3 total purchase is Rp. 600.000

Customer 4 total purchase is Rp. 700.000

Customer 5 total purchase is Rp. 800.000

Customer 6 total purchase is Rp. 1.000.000

Customer 7 total purchase is Rp. 1.500.000

Customer 8 total purchase is Rp. 1.700.000

Customer 9 total purchase is Rp. 2.000.000

Lets say for each total purchase we don’t take the last 5 digit (but I don’t do this in the application. This is done so it is easy for me to calculate the purchase). So it becomes like this :

Customer 1 total purchase is Rp. 1

Customer 2 total purchase is Rp. 2

Customer 3 total purchase is Rp. 6

Customer 4 total purchase is Rp. 7

Customer 5 total purchase is Rp. 8

Customer 6 total purchase is Rp. 10

Customer 7 total purchase is Rp. 15

Customer 8 total purchase is Rp. 17

Customer 9 total purchase is Rp. 20

we change the data into like this:

K1 = 1

K2 = 2

K3 = 6

K4 = 7

K5 = 8

K6 = 10

K7 = 15

K8 = 17

K9 = 20

The group of this data is called vector.

There are several steps in this algorithm:

1. Starting by decided the amount of codeword. Codeword is middle point for a group of vector in a cluster(group). In this case, the amount of codeword is the amount of customer group which is 3.

Codeword can be called means.

2. We have 3 customer group so we need to calculate the average value of each customer group.

Average value = (1+2+6+7+8+10+15+17+20)/9=9.5. This is Means 2 (M2) for Potential customer.

To find out M1 (Regular Customer), we need to check which vectors below value of M2.

Average value = (1+2+6+7+8)/5 = 4.8 . this is Means 1(M1) for Regular customer.

To Find out M3 (VIP Customer), we need to check which vectors above value of M2.

Average value = (10+15+17+20)/4 = 15.5. This is Means 3 (M3) for VIP customer.

So we have

M1 = 4.8

M2 = 9.5

M3 = 15.5

So vectors which approach M1 is included into Group 1.

vectors which approach M2 is included into Group 2.

vectors which approach M3 is included into Group 3.

3. calculate Euclidean distance which is used to categorized vector.

The Euclidean formula is 

X is vector

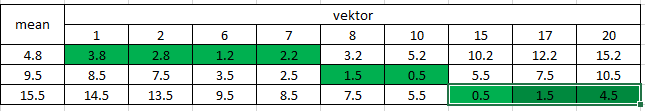
a is means.

So the euclidean of vector K1 with M1 is

The euclidean of vector K1 with M2 is

The euclidean of vector K1 with M3 is

To simplify this you can see the table below:



From table above we know that

Vector 1 has euclidean distance with mean 4.8 is 3.8

Vector 1 has euclidean distance with mean 9.5 is 8.5

And Vector 1 has euclidean distance with mean 15.5 is 14.5

So we take the closest distance value and it is 3.8 (green block). Likewise for vectors 2,6,7,8,10,15,17 and 20 have the closest distance as in the green block.

So it create 3 group:

Group 1 contains vector 1, 2, 6, 7 which are the Regular Customer

Group 2 contains vector 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

4. find out the new means by calculating the average value.

Group 1 = 1, 2, 6, 7 so new M1 is M1=(1+2+6+7) / 4 = 4

Group 2 = 8, 10 so new M2 is M2 = (8+10) / 2 = 9

Group 3 = 15, 17, 20 so new M3 is M3 = (15+17+20) / 3 = 17,33

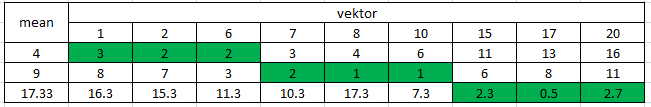
So the new means are

M1 = 4

M2= 9

M3= 17.33

By these new means , we calculate the Euclidean distance again. To simplify we can look the table below



From table above we know that

Vector 1 has euclidean distance with mean 4 is 3

Vector 1 has euclidean distance with mean 9 is 8

Vector 1 has euclidean distance with mean 17.33 is 16.33

So we take the closest distance value and it is 3 (green block). Likewise for vectors 2,6,7,8,10,15,17 and 20 have the closest distance as in the green block.

So it creates new group of customer:

Group 1 contains vector 1, 2, 6 which are the Regular Customer

Group 2 contains vector 7, 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

5. repeate the calculation of new means if there is different beetween the previous group customer and the latest group customer. If there is no difference then algorithm finish.

In this case we can see in point 3 the previous group customer is

Group 1 contains vector 1, 2, 6, 7 which are the Regular Customer

Group 2 contains vector 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

But in point 4 the latest group customer is

Group 1 contains vector 1, 2, 6 which are the Regular Customer

Group 2 contains vector 7, 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

At this point we know that there is a difference which vector 7 from group 1 become group 2. So Because of differences, new means and new Euclidean distances must be recalculated.

New means is

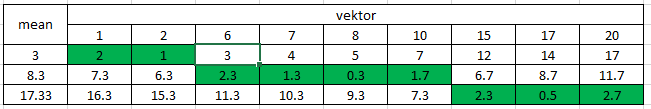
Group 1 = 1,2,6 so the new M1 is M1=(1+2+6) / 3 = 3

Group 2 = 7,8,10 so the new M2 is M2 = (7+8+10) / 3 = 8.3

Group 3 = 15,17,20 so the new M3 is M3 = (15+17+20) / 3 = 17,33

Then the new means are M1 = 3 , M2 = 8.3 and M3 = 17.33

With these new means, we recalculate the Euclidean distance as the table below



From table above we know that

Vector 1 has euclidean distance with mean 3 is 2

Vector 1 has euclidean distance with mean 8.3 is 7.3

Vector 1 has euclidean distance with mean 17.33 is 16.33

So we take the closest distance value and it is 2 (green block). Likewise for vectors 2,6,7,8,10,15,17 and 20 have the closest distance as in the green block.

So we have a new group customer

Group 1 contains vector 1, 2 which are the Regular Customer

Group 2 contains vector 6, 7, 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

At this point there is still a difference between the previous group customer and the latest group customer.

The previous group customer is

Group 1 contains vector 1, 2, 6 which are the Regular Customer

Group 2 contains vector 7, 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

The latest group customer is

Group 1 contains vector 1, 2 which are the Regular Customer

Group 2 contains vector 6, 7, 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

Vector 6 change group from regular customer to potential customer.

Because of that the new means and new Euclidean distances must be recalculated.

New means is

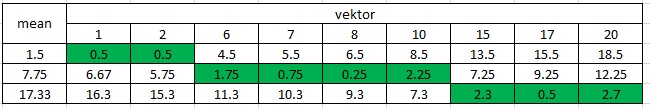
Group 1 = 1,2 so the new M1 is M1=(1+2) / 2 = 1,5

Group 2 = 6,7,8,10 so the new M2 is M2 = (6+7+8+10) / 4 = 7,75

Group 3 = 15,17,20 so the new M3 is M3 = (15+17+20) / 3 = 17,33

Then the new means are M1 = 1.5 , M2 = 7.75 and M3 = 17.33

With these new means, we recalculate the Euclidean distance as the table below



From table above we know that

Vector 1 has euclidean distance with mean 1.5 is 0.5

Vector 1 has euclidean distance with mean 7.75 is 6.67

Vector 1 has euclidean distance with mean 17.33 is 16.33

So we take the closest distance value and it is 0.5 (green block). Likewise for vectors 2,6,7,8,10,15,17 and 20 have the closest distance as in the green block.

So we have a new group customer

Group 1 contains vector 1, 2 which are the Regular Customer

Group 2 contains vector 6, 7, 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

At this point there is no difference between the previous group customer and the latest group customer.

The previous group customer is

Group 1 contains vector 1, 2 which are the Regular Customer

Group 2 contains vector 6, 7, 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

And the latest group customer is

Group 1 contains vector 1, 2 which are the Regular Customer

Group 2 contains vector 6, 7, 8, 10 which are the Potential Customer

Group 3 contains vector 15, 17, 20 which are the VIP Customer

So the algorithm finish and we get the customer segmentation

Regular Customer is Customer 1, Customer 2

Potential Customer is Customer 6, Customer 7, Customer 8, Customer 10

VIP Customer is Customer 15, Customer 17, Customer 20