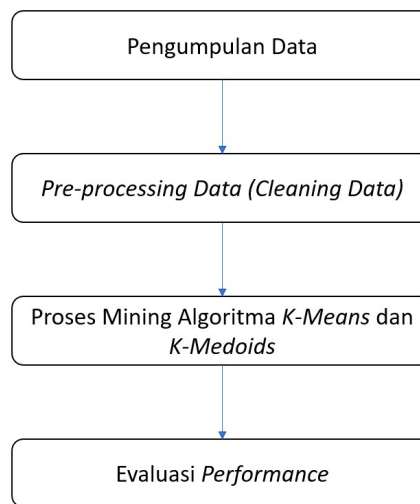


BAB III METODOLOGI PENELITIAN

3.1 Metode Penelitian

Pada bab ini akan membahas langkah-langkah dari proses penelitian yang akan dilaksanakan. Dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan, maka dibuat alur dalam tahapan penelitian yang akan dilakukan sebagai berikut:



Gambar 3.1 Alur Dalam Tahapan Penelitian

Berdasarkan gambar 3.1 diatas akan dijelaskan mengenai alur dalam tahapan penelitian sebagai berikut:

1. Pengumpulan Data.

Data yang digunakan merupakan dataset publik berupa data yang diperoleh dari *UCI Machine Learning Repository* dengan nama file *Cervical cancer (Risk Factors) Data Set*. Dataset dikumpulkan dari *'Hospital Universitario de Caracas' di Caracas, Venezuela* melalui link

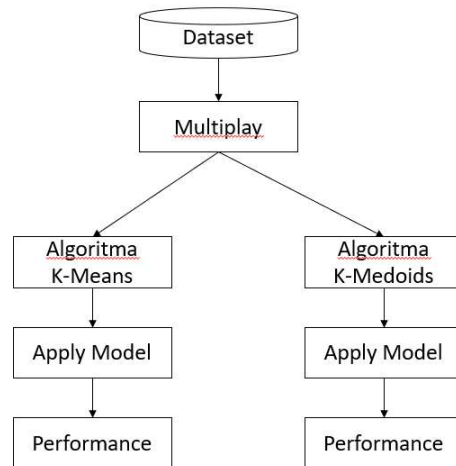
<https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29> yang terdiri dari informasi demografis, kebiasaan, dan catatan medis historis. Beberapa pasien memutuskan untuk tidak menjawab beberapa pertanyaan karena masalah privasi (*missing value*). Jumlah data adalah 858 data dengan 36 atribut.

2. *Pre-processing Data (Cleaning Data)*.

Pre-processing Data adalah proses persiapan data. Data yang memiliki nilai kosong akan diganti dengan nilai rata-rata dari nilai pada atribut yang langsung diimplementasikan pada tool RapidMiner menggunakan operator *Replace Missing Value*. Data berjumlah 858 record data dengan 36 atribut.

3. Proses Mining Algoritma *K-Means* dan *K-Medoids*.

Pada tahap ini, data yang sudah melalui tahap *pre-processing data* akan diuji menggunakan algoritma *K-Means* dan *K-Medoids* dengan bantuan tool *RapidMiner Studio*. Dalam melakukan pemodelan pada tool *RapidMiner*, yang dilakukan oleh peneliti adalah memasukkan dataset yang telah terseleksi kemudian menggunakan operator *Multiplay* karena data akan diuji menggunakan dua algoritma dalam satu waktu, selanjutnya menggunakan algoritma *K-Means* dan *K-Medoids*, dan operator *Apply Model*. Selanjutnya untuk yang terakhir adalah *Performance*.



Gambar 3. 2 Bagan Permodelan

4. Evaluasi Performance.

Proses evaluasi *performance* hasil klustering dengan algoritma *K-Means* dan *K-Medoids* yaitu menggunakan metode *Davies-Bouldin Index (DBI)*.

3.2 Penerapan algoritma *K-Means*

K-Means Clustering merupakan algoritma pengklasteran yang paling sederhana dibanding algoritma pengklasteran yang lain. *K-Means* Clustering mempunyai kelebihan mudah diterapkan dan dijalankan, relatif cepat, mudah untuk diadaptasi, dan paling banyak dipraktikkan dalam tugas data mining. Pengklasteran merupakan suatu metode untuk mengelompokkan dokumen di mana dokumen dikelompokkan dengan konten untuk mengurangi ruang pencarian yang diperlukan dalam merespon suatu query (Grossman dan Frieder, 2004).

MacQueen (1967) mengatakan bahwa *K-Means* merupakan metode pengklasteran secara partitioning yang memisahkan data ke dalam kelompok

yang berbeda. Dengan partitioning secara iteratif, *K-Means* mampu meminimalkan rata-rata jarak setiap data ke klasternya.

K-Means Clustering merupakan algoritma pengklasteran iteratif yang melakukan partisi set data ke dalam sejumlah k kluster yang sudah ditetapkan di awal. *K-Means* Clustering sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, *K-Means* Clustering menjadi salah satu algoritma yang paling penting dalam bidang data minig (Prasetyo, 2012).

K-Means Clustering digunakan sebagai alternatif metode kluster untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan dengan metode hierarki (Sitepu dkk, 2011).

Berikut ini langkah-langkah yang terdapat pada algoritma *K-Means*:

1. Tentukan jumlah kluster (k), tetapkan pusat kluster secara acak.
2. Hitung jarak setiap data ke pusat kluster
3. Kelompokkan data ke dalam kluster dengan jarak yang paling pendek.
4. Hitung pusat kluster baru.
5. Ulangi langkah 2 (dua) sampai 4 (empat) sehingga tidak ada lagi data yang pindah ke cluster lain.

Proses clustering dimulai dengan mengidentifikasi data yang diklasterisasi, digunakan rumus formula *Euclidean Distance* seperti yang terlihat pada rumus berikut ini:

$$d_{ij} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

- D_{ij} : jarak antara i ke pusat data kluster j
 X_{ki} : data ke- i pada atribut data ke- k
 X_{kj} : data ke- j pada atribut data ke- k

No	Age	Number of sexual partners	First sexual intercourse	Num of pregnancies	Smokes	Smokes (years)	...	STDs: HPV
1	18	4	15	1	0	0	...	0
2	15	1	14	1	0	0	...	0
3	34	1	17	1	0	0	...	0
4	52	5	16	4	1	37	...	0
5	46	3	21	4	0	0	...	0
6	42	3	23	2	0	0	...	0
7	51	3	17	6	1	34	...	1
8	26	1	26	3	0	0	...	0
...
858	29	2	20	1	0	0	...	0

Tabel 3. 1 Dataset Penyakit Kanker Serviks

Data pada tabel diatas adalah data pasien penyakit kanker serviks yang akan diuji dengan tahapan sebagai berikut:

1. Tahap pertama untuk *clustering* pada algoritma *K-Means* adalah menentukan pusat centroid setiap kluster. Pada iterasi-1, pusat centroid ditentukan secara random sebagai berikut:

C1	15	1	14	1	0	0	0	0	...	0
C2	45	1	20	5	0	0	0	0	...	0

Tabel 3. 2 Pusat Sample Cluster Random

2. Selanjutnya adalah menghitung jarak terhadap pusat kluster menggunakan rumus :

$$D(ij) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + (X3i - X3j)^2}$$

$$D(1,C1) = \sqrt{\begin{matrix} (18-15)^2 + (4-1)^2 + (15-14)^2 + (1-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (6-6)^2 + (6-6)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{matrix}}$$

$$= 4$$

$$D(1,C2) = \sqrt{\begin{matrix} (18-45)^2 + (4-1)^2 + (15-20)^2 + (1-5)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (6-6)^2 + (6-6)^2 + (0-1)^2 + (0-0)^2 + (0-1)^2 + (0-1)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{matrix}}$$

$$= 28$$

Perhitungan diatas untuk data ke-1, dan perhitungan yang sama dapat dilakukan hingga data ke-858, sehingga didapatkan hasil jarak untuk iterasi-1 sebagai berikut:

No	C1	C2
1	4	28
2	0	31
3	19	12
4	64	53
5	35	15
6	29	6
7	51	36
8	18	21
...
858	15	17

Tabel 3. 3 Hasil Iterasi-1

- Selanjutnya adalah menentukan jarak minimum atau jarak terdekat. Jarak diantara centroid 1 (C1) dan centroid 2 (C2) diambil jarak yang terpendek. Dan untuk menghitung kelompok data kluster pada excel digunakan rumus $=IF(C1 < C2, "Positif", "Negatif")$ sehingga didapatkan hasil sebagai berikut:

No	Jarak Terdekat	Kelompok Data
1	4	Positif
2	0	Positif
3	12	Negatif

4	53	Negatif
5	15	Negatif
6	6	Negatif
7	36	Negatif
8	18	Positif
...
858	15	Positif

Tabel 3. 4 Hasil Jarak Terdekat dan Kelompok Data

4. Selanjutnya menghitung nilai centroid baru berdasarkan data yang tergabung pada setiap klasternya. Persamaan untuk perhitungan nilai centroid adalah sebagai berikut:

$$C = \frac{\sum m}{n}$$

C = centroid data,

m = anggota data yang termasuk ke dalam centroid tertentu,

n = data yang menjadi anggota centroid tertentu.

Contoh perhitungan untuk variabel Age pada kluster positif:

$$C1 = \frac{18 + 15 + 26 + 27 + 30 + 30 + 23 + 30 + \dots + 29}{594} = 22$$

Lakukan perhitungan yang sama seperti diatas kepada seluruh variabel pada kluster positif.

Contoh perhitungan untuk variabel Age kluster negatif:

$$C2 = \frac{34 + 52 + 46 + 42 + 51 + 45 + 44 + 44 + \dots + 33}{264} = 37$$

Lakukan perhitungan yang sama seperti diatas kepada seluruh variabel pada kluster negatif.

Data hasil perhitungan seluruh variabel dapat dilihat pada tabel berikut:

C1	22	2	16	2	0	1	0	1	...	0
C2	37	3	18	3	0	2	1	1	...	0

Tabel 3. 5 Titik centroid baru iterasi-2

Hitung iterasi selanjutnya hingga anggota kluster baru dengan anggota kluster sebelumnya sama atau tidak berubah. Berdasarkan perhitungan yang telah dilakukan pada penelitian ini, data penentuan kluster diperoleh pada iterasi-13 dan hasilnya dapat adalah sebagai berikut:

No	C1	C2	Kelompok Data
1	13	22	Positif
2	14	25	Positif
3	17	13	Negatif
4	61	54	Negatif
5	31	20	Negatif
6	25	15	Negatif
7	47	38	Negatif
8	17	19	Positif
...
858	14	15	Positif

Tabel 3. 6 Tabel Hasil Clustering Menggunakan Algoritma K-Means

Penjelasan tabel diatas adalah data pasien masuk ke dalam kluster positif jika jarak atau nilai C1 lebih kecil dari C2 (nilai jarak C1 lebih pendek), dan data pasien masuk ke dalam kluster negatif jika jarak atau nilai C2 lebih kecil dari C1 (nilai jarak C2 lebih pendek). Sebagai contoh data ke-1. Pasien masuk ke dalam kluster positif karena jarak atau nilai C1 lebih pendek dari pada C2.

Nilai Centroid pada iterasi-13 adalah sebagai berikut:

C1	22	2	16	2	0	1	0	1	...	0
C2	35	3	18	3	0	2	1	1	...	0

Tabel 3. 7 Titik centroid baru iterasi-13

- b. Setelah mengetahui nilai *Sum of Square Within Cluster* (SSW), selanjutnya akan dihitung nilai *Sum of Square Between Cluster* (SSB).

Cara perhitungannya adalah dengan persamaan:

$$SSB_{i,j} = d(c_i, c_j)$$

Centroid 1:

$$= \sqrt{\begin{matrix} (22-22)^2 + (2-2)^2 + (16-16)^2 + (2-2)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (1-1)^2 \\ + (1-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (6-6)^2 + (6-6)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{matrix}}$$

$$= 0$$

Centroid 2:

$$= \sqrt{\begin{matrix} (22-35)^2 + (2-3)^2 + (16-18)^2 + (2-3)^2 + (0-0)^2 + (1-2)^2 + (0-1)^2 + (1-1)^2 \\ + (1-4)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (6-6)^2 + (6-6)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{matrix}}$$

$$= 14.31$$

		Centroid ke-	
		1	2
Centroid ke-	1	0	14.31
	2	14.31	0

Tabel 3. 9 Nilai *Sum of Square Within Cluster* (SSB)

- c. Tahap selanjutnya adalah menghitung nilai rasio dengan persamaan:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

$$= \frac{(5.11 + 6.21)}{14.31}$$

$$= 0.60$$

- d. Menghitung nilai *Davies-Bouldin Index* (DBI) dengan persamaan:

$$\begin{aligned}
\text{DBI} &= \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \\
&= \frac{1}{2} \times 0.60 \\
&= 0.30
\end{aligned}$$

3.3 Penerapan algoritma *K-Medoids*

K-Medoids Clustering, juga dikenal sebagai Partitioning Around Medoids (PAM), adalah varian dari metode *K-Means*. Hal ini didasarkan pada penggunaan medoids bukan dari pengamatan mean yang dimiliki oleh setiap klaster, dengan tujuan mengurangi sensitivitas dari partisi sehubungan dengan nilai ekstrim yang ada dalam dataset (Vercellis, 2009).

K-Medoids Clustering hadir untuk mengatasi kelemahan *K-Means* Clustering yang sensitif terhadap outlier karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data (Han dan Kamber, 2006).

K-Medoids Clustering menggunakan metode pengklasteran partisi untuk mengklasterkan sekumpulan n objek menjadi sejumlah k klaster. Algoritma ini menggunakan objek pada kumpulan objek yang mewakili sebuah klaster. Objek yang mewakili sebuah klaster disebut dengan *medoids*. Klaster dibangun dengan menghitung kedekatan yang dimiliki antara *medoids* dengan objek *non medoids* (Setyawati, 2017).

Metode *K-Medoids* tidak menentukan nilai rata-rata dari objek yang ada dalam cluster sebagai titik acuan, namun menggunakan medoids (median), yang merupakan objek yang paling terletak dipusat sebuah cluster (Keatas et al., 2019).

Langkah-langkah perhitungan dengan metode *K-Medoids* adalah sebagai berikut:

1. Nilai k ditentukan adalah 2
2. Alokasikan setiap data (objek) ke kluster terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)}$$

- d_{ij} : jarak antara objek i dengan j
 x_{ia} : nilai objek i pada variable ke-
 a
 x_{ja} : nilai objek j pada variable ke-
 a
 p : banyaknya variable yang
 diamati

3. Pilih secara acak objek pada masing-masing kluster sebagai kandidat medoid baru.
4. Hitung jarak setiap objek yang berada pada masing-masing kluster dengan kandidat medoid baru.
5. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoid.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan kluster beserta anggota kluster masing-masing.

No	Age	Number of sexual partners	First sexual intercourse	Num of pregnancies	Smokes	Smokes (years)	...	STDs: HPV
1	18	4	15	1	0	0	...	0
2	15	1	14	1	0	0	...	0
3	34	1	17	1	0	0	...	0
4	52	5	16	4	1	37	...	0
5	46	3	21	4	0	0	...	0
6	42	3	23	2	0	0	...	0

7	51	3	17	6	1	34	...	1
8	26	1	26	3	0	0		0
...
858	29	2	20	1	0	0		0

Tabel 3. 10 Dataset penyakit kanker serviks

Dalam menentukan medoid awal, data dipilih secara acak (*random*) sebagai berikut:

C1	45	1	20	5	0	0	0	0	...	0
C2	29	2	20	1	0	0	0	1	...	0

Tabel 3. 11 Medoid awal iterasi-1

Setelah menentukan titik medoid awal, selanjutnya adalah menghitung jarak masing-masing objek ke medoid awal dengan persamaan *eunclidian distance*:

$$D1C1 = \sqrt{\begin{aligned} &(18 - 45)^2 + (4 - 1)^2 + (15 - 20)^2 + (1 - 5)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 \\ &+ (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 \\ &+ (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 \\ &+ (0 - 0)^2 + (6 - 6)^2 + (6 - 6)^2 + (0 - 1)^2 + (0 - 0)^2 + (0 - 1)^2 \\ &+ (0 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 \end{aligned}}$$

$$= 28$$

$$D1C2 = \sqrt{\begin{aligned} &(18 - 29)^2 + (4 - 2)^2 + (15 - 20)^2 + (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 1)^2 \\ &+ (0 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 \\ &+ (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 \\ &+ (0 - 0)^2 + (6 - 6)^2 + (6 - 6)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 \\ &+ (0 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 \end{aligned}}$$

$$= 12$$

Lakukan perhitungan yang sama dengan cara diatas hingga data ke-858, hasil perhitungannya adalah sebagai berikut:

No	Cost 1	Cost 2
1	28	12
2	31	15
3	12	6
4	53	58
5	15	22
6	6	13
7	36	42

...
858	17	0
	18103	9108
	27210	

Tabel 3. 12 Nilai cost pada iterasi-1

Selanjutnya adalah menentukan jarak minimum atau jarak terdekat. Jarak diantara Cost 1 dan Cost 2 diambil jarak yang terpendek. Dan untuk menghitung kelompok data klaster pada excel digunakan rumus =IF(Cost1<Cost2,"Positif","Negatif") yaitu jika nilai Cost1 lebih kecil dari Cost2 maka data akan dimasukkan kedalam klaster positif, dan jika nilai Cost1 lebih besar dari Cost2, maka data akan dimasukkan kedalam klaster negatif sehingga didapatkan hasil sebagai berikut:

No	KEDEKATAN	KLASTER
1	12.33	Klaster 2
2	15.33	Klaster 2
3	6.08	Klaster 2
4	53.22	Klaster 1
5	15.36	Klaster 1
6	5.83	Klaster 1
7	35.67	Klaster 1
8	10.05	Klaster 2
...
858	0.00	Klaster 2

Tabel 3. 13 Jarak Kedekatan Objek Terhadap Medoids, dan Hasil Klaster

Lakukan perhitungan yang sama untuk iterasi selanjutnya. Penentuan medoid baru adalah sebagai berikut:

C1	39	1	24	4	0	0	0	0	...	0
C2	21	2	18	2	0	0	0	1	...	0

Tabel 3. 14 Medoid Baru Iterasi-2

Menghitung jarak masing-masing objek ke medoid (medoid baru) dengan persamaan euclidian distance:

$$D1C1 = \sqrt{\begin{aligned} & (18-39)^2 + (4-1)^2 + (15-24)^2 + (1-4)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ & + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ & + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ & + (0-0)^2 + (6-6)^2 + (6-6)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ & + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{aligned}}$$

= 23

$$D1C2 = \sqrt{\begin{aligned} & (18-21)^2 + (4-2)^2 + (15-18)^2 + (1-2)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 \\ & + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ & + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ & + (0-0)^2 + (6-6)^2 + (6-6)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ & + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{aligned}}$$

= 5

Lakukan perhitungan yang sama dengan cara diatas hingga data ke-858, hasil perhitungannya adalah sebagai berikut:

No	Cost 1	Cost 2
1	23	5
2	26	7
3	9	13
4	55	61
5	17	29
6	4	22
7	38	46
...
858	11	8
	14912	8763
	23675	

Tabel 3. 15 Nilai Cost Pada Iterasi-2

Selanjutnya adalah menentukan jarak minimum atau jarak terdekat. Jarak diantara Cost1 dan Cost2 diambil jarak yang terpendek. Dan untuk menghitung kelompok data klaster pada excel digunakan rumus =IF(Cost1<Cost2,"Positif","Negatif") yaitu jika nilai Cost1 lebih kecil dari Cost2 maka data akan dimasukkan kedalam positif dan jika nilai C1 lebih besar dari C2, maka data akan dimasukkan kedalam negatif sehingga didapatkan hasil sebagai berikut:

No	KEDEKATAN	KLASTER
1	5.00	Negatif
2	7.48	Negatif
3	9.11	Positif

4	54.77	Positif
5	16.97	Positif
6	4.24	Positif
7	37.68	Positif
8	11.92	Negatif
...
858	8.00	Negatif

Tabel 3. 16 Jarak Kedekatan Objek Terhadap Medoids, dan Hasil Kluster

Hitung total simpangan, yaitu menghitung selisih data antara jumlah *cost* baru dengan jumlah *cost* lama). Dengan ketentuan jika $S < 0$, maka tukar nilai objek dengan menentukan medoid baru. Ulangi langkah tersebut hingga mendapatkan nilai $S > 0$.

a. Iterasi-2

$$\begin{aligned} S &= b - a \\ &= 23,675 - 27,210 \\ &= -3,535 \end{aligned}$$

e. Iterasi-6

$$\begin{aligned} S &= b - a \\ &= 18,095 - 18,546 \\ &= -452 \end{aligned}$$

b. Iterasi-3

$$\begin{aligned} S &= b - a \\ &= 22,432 - 23,675 \\ &= -1,243 \end{aligned}$$

f. Iterasi-6

$$\begin{aligned} S &= b - a \\ &= 17,341 - 18,095 \\ &= -753 \end{aligned}$$

c. Iterasi-4

$$\begin{aligned} S &= b - a \\ &= 18,613 - 22,432 \\ &= -3,820 \end{aligned}$$

g. Iterasi-6

$$\begin{aligned} S &= b - a \\ &= 17,655 - 17,341 \\ &= 314 \end{aligned}$$

d. Iterasi-5

$$\begin{aligned} S &= b - a \\ &= 18,546 - 18,613 \\ &= -67 \end{aligned}$$

Setelah mendapatkan nilai simpangan $S > 0$, tahap selanjutnya adalah evaluasi klaster dengan menggunakan metode *Davies Bouldin Index (DBI)*. Urutan perhitungannya adalah sebagai berikut:

- e. Hitung nilai *Sum of square within cluster (SSW)*.

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$$

Klaster 1:

$$= \sqrt{\begin{aligned} &(18-24)^2 + (4-0)^2 + (15-15)^2 + (1-3)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ &+ (0-5)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ &+ (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ &+ (0-0)^2 + (6-6)^2 + (6-6)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ &+ (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{aligned}}$$

= 8.10

Klaster 2:

$$= \sqrt{\begin{aligned} &(40-32)^2 + (3-1)^2 + (15-18)^2 + (3-0)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 \\ &+ (3-7)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ &+ (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ &+ (0-0)^2 + (6-6)^2 + (6-6)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ &+ (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{aligned}}$$

= 9.98

Lakukan perhitungan untuk seluruh data berdasarkan klaster masing-masing, dan hasilnya adalah sebagai berikut:

No	Klaster	Jarak data-klaster	SSW	
1	Klaster 2	7.48	Klaster 1	8.10
2	Klaster 2	8.37		
3	Klaster 1	9.70	Klaster 2	9.98
4	Klaster 1	7.48		
5	Klaster 1	8.06		
6	Klaster 1	8.37		
7	Klaster 2	7.55		
8	Klaster 1	7.42		
...		
858	Klaster 2	7.21		

Tabel 3. 17 Nilai Jarak Data-Klaster dan SSW

- f. Setelah mengetahui nilai *Sum of Square Within Cluster* (SSW), selanjutnya akan dihitung nilai *Sum of Square Between Cluster* (SSB).

Cara perhitungannya adalah dengan persamaan:

$$SSB_{i,j} = d(c_i, c_j)$$

Centroid 1:

$$= \sqrt{\frac{(43-43)^2 + (3-3)^2 + (17-17)^2 + (3-3)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + (5-5)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}{(0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}} = 0$$

Centroid 2:

$$= \sqrt{\frac{(43-29)^2 + (3-2)^2 + (17-20)^2 + (3-1)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + (5-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}{(0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}} = 9.43$$

		Centroid ke-	
		SSB	
Centroid ke-	1	0	9.43
	2	9.43	0

Tabel 3. 18 Nilai Sum of Square Within Cluster (SSB)

g. Tahap selanjutnya adalah menghitung nilai rasio dengan persamaan:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} = \frac{(8.10 + 9.98)}{9.43} = 1.92$$

h. Menghitung nilai *Davies-Bouldin Index* (DBI) dengan persamaan:

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) = \frac{1}{2} \times 1.92 = 0.96$$

3.4 Alat dan Bahan

Penelitian ini menggunakan perangkat keras laptop *Lenovo Yoga 6 12ARE05* dengan *processor AMD Ryzen 5 PRO 4650U with Radeon Graphics 2.10 GHz, Installed RAM 16.0 GB (15.4 GB usable)*, dan untuk perangkat lunak menggunakan *Microsoft Excel* dan aplikasi *RapidMiner Studio 9.10.008* yang digunakan dalam pengolahan data.