

BAB IV HASIL DAN PEMBAHASAN

4.1 Pengolahan Data

Dalam bab ini, penulis akan membahas mengenai proses pengolahan data kanker serviks. Dataset yang digunakan berjumlah 858 data dengan 36 atribut dan akan diproses pemodelan dengan menerapkan data kedalam perhitungan algoritma yang akan dibandingkan yaitu algoritma *K-Means* dan *K-Medoids* pada *tool Rapidminer Studio* dan algoritma terbaik akan diketahui melalui perbandingan nilai *Davies Bouldin Index (DBI)*.

Age	Number of sexual partners	First sexual intercourse	Num of pregnancies	Smokes	Smokes (years)	Hormonal Contraceptives	Hormonal Contraceptives (years)	IUD	IUD (years)	STDs	STDs:condylomatosi	STDs:cervical condylomatosi	STDs:vaginal condylomatosi	STDs:vulvo-perineal condylomatosi	STDs:syphilis	STDs:pelvic inflammatory disease	STDs:genital herpes	STDs:molluscum contagiosum	STDs:AIDS	STDs:HIV	STDs:Hepatitis B	STDs:HPV	
18,0	4.0	15.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
15,0	1.0	14.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
52,0	5.0	16.0	4.0	1.0	37.0	1.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
46,0	3.0	21.0	4.0	0.0	0.0	1.0	15.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
42,0	3.0	23.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
51,0	3.0	17.0	6.0	1.0	34.0	0.0	0.0	1.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
26,0	1.0	26.0	3.0	0.0	0.0	1.0	2.0	1.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
45,0	1.0	20.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
44,0	3.0	26.0	4.0	0.0	0.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
27,0	1.0	17.0	3.0	0.0	0.0	1.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
45,0	4.0	14.0	6.0	0.0	0.0	1.0	10.0	1.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
44,0	2.0	25.0	2.0	0.0	0.0	1.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
43,0	2.0	18.0	5.0	0.0	0.0	0.0	0.0	1.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
40,0	3.0	18.0	2.0	0.0	0.0	1.0	15.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
41,0	4.0	21.0	3.0	0.0	0.0	1.0	0.25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
43,0	3.0	15.0	8.0	0.0	0.0	1.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
43,0	2.0	18.0	4.0	0.0	0.0	1.0	15.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
...
29,0	2.0	20.0	1.0	0.0	0.0	1.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Gambar 4. 1 Dataset Kanker Serviks

Atribut data kanker serviks meliputi *Age*, *Number of sexual partners*, *First sexual intercourse*, *Num of pregnancies*, *Smokes*, *Smokes (years)*, *Hormonal Contraceptives*, *Hormonal Contraceptives (years)*, *IUD*, *IUD (years)*, *STDs*, *STDs multiplay(number)*, *STDs:condylomatosi*, *STDs:cervical condylomatosi*, *STDs:vaginal condylomatosi*, *STDs:vulvo-perineal condylomatosi*, *STDs:syphilis*, *STDs:pelvic inflammatory disease*, *STDs:genital herpes*, *STDs:molluscum contagiosum*, *STDs:AIDS*, *STDs:HIV*, *STDs:Hepatitis B*, *STDs:HPV*. Data tersebut tidak memiliki kelas atau label. Deskripsi atribut atau variabel tersebut dapat dilihat pada tabel di bawah ini:

<i>Age</i>	Usia pasien. Memiliki rentang 13 – 84 tahun
<i>Number of sexual partners</i>	Jumlah pasangan yang pernah berhubungan seksual dengan pasien. Rentang nilai 1 – 28 kali.
<i>First sexual intercourse (age)</i>	usia pertama kali pasien ketika melakukan hubungan seksual. rentang antara 10 – 32 tahun.
<i>Num of pregnancies</i>	jumlah kehamilan yang dialami pasien. rentang nilai 0 – 11 kali.
<i>Smokes</i>	pasien merokok/tidak merokok. Diisi dengan nilai 1 / 0.
<i>Smokes (years)</i>	berapa lama (dalam tahun) pasien yang memiliki kebiasaan merokok. rentang nilai 0 – 37 tahun.
<i>Smokes (packs/year)</i>	berapa bungkus rokok yang dikonsumsi per tahunnya oleh pasien yang memiliki kebiasaan merokok. rentang nilai antara 0 – 37 bungkus/tahun.
<i>Hormonal Contraceptives</i>	pasien menggunakan kontrasepsi hormonal atau tidak. Diisi dengan nilai 0 / 1.
<i>Hormonal Contraceptives (years)</i>	berapa lama (dalam tahun) pasien yang menggunakan kontrasepsi hormonal. Rentang nilai 0 – 22 tahun.
<i>IUD</i>	Apakah pasien menggunakan alat kontrasepsi Intrauterine Device (IUD) atau tidak. Diisi dengan nilai 0 / 1.
<i>IUD (years)</i>	Berapa lama (dalam tahun) pasien yang menggunakan alat kontrasepsi IUD. Rentang nilai 0 – 19 tahun.
<i>STDs</i>	Apakah pasien memiliki riwayat penyakit kelamin, atau sexually transmitted diseases (STDs). Diisi dengan nilai 0 / 1.
<i>STDs (number)</i>	Jumlah penyakit kelamin yang pernah diderita oleh pasien. Rentang nilai 0 – 4 penyakit.
<i>STDs:condylomatosis</i>	Apakah pasien memiliki riwayat penyakit kelamin menular condylomatosis. Diisi dengan nilai 0 / 1.
<i>STDs:cervical condylomatosis</i>	Apakah pasien memiliki riwayat penyakit kelamin menular cervical condylomatosis. Diisi dengan nilai 0 / 1.
<i>STDs:vaginal condylomatosis</i>	Apakah pasien memiliki riwayat penyakit kelamin menular vaginal condylomatosis. Diisi dengan nilai 0 / 1.
<i>STDs:vulvo-perineal condylomatosis</i>	Apakah pasien memiliki riwayat penyakit kelamin menular vulvo-perineal condylomatosis. Diisi dengan nilai 0 / 1.
<i>STDs:syphilis</i>	Apakah pasien memiliki riwayat penyakit kelamin menular syphilis. Diisi dengan nilai 0 / 1.
<i>STDs:pelvic inflammatory disease</i>	Apakah pasien memiliki riwayat penyakit kelamin menular pelvic inflammatory disease. Diisi dengan nilai 0 / 1.
<i>STDs:genital herpes</i>	Apakah pasien memiliki riwayat penyakit kelamin menular genital herpes. Diisi dengan nilai 0 / 1.
<i>STDs:molluscum contagiosum</i>	Apakah pasien memiliki riwayat penyakit kelamin menular molluscum contagiosum. Diisi dengan nilai 0 / 1.
<i>STDs:AIDS</i>	Apakah pasien memiliki riwayat penyakit kelamin menular AIDS. Diisi dengan nilai 0 / 1.
<i>STDs:HIV</i>	Apakah pasien memiliki riwayat penyakit kelamin menular HIV. Diisi dengan nilai 0 / 1.
<i>STDs:Hepatitis B</i>	Apakah pasien memiliki riwayat penyakit kelamin menular hepatitis B. Diisi dengan nilai 0 / 1.
<i>STDs:HPV</i>	Apakah pasien memiliki riwayat penyakit kelamin menular HPV. Diisi dengan nilai 0 / 1.
<i>STDs: Number of diagnosis</i>	Jumlah penyakit kelamin yang pernah didiagnosis pasien. Memiliki rentang nilai 0 – 3.

<i>STDs: Time since first diagnosis</i>	Waktu sejak diagnosis pertama
<i>STDs: Time since last diagnosis</i>	Waktu sejak diagnosis terakhir
<i>Dx: Cancer</i>	Diagnosis sel Cancer
<i>Dx: CIN</i>	Diagnosis Cervical Intraepithelial Neoplasia
<i>Dx: HPV</i>	Diagnosis infeksi Human Papillomavirus
<i>Dx</i>	Diagnosis
<i>Hinselmann</i>	Hasil tes hinselmann. Diisi dengan nilai 0 / 1.
<i>Schiller</i>	Hasil tes schiller. Diisi dengan nilai 0 / 1.
<i>Cytology</i>	Hasil tes cytology. Diisi dengan nilai 0 / 1.
<i>Biopsy</i>	Hasil tes biopsy. Diisi dengan nilai 0 / 1.

Tabel 4. 1 Atribut Dataset Penyakit Kanker Serviks

Untuk menjawab rumusan masalah yaitu mengetahui bagaimana hasil pengklasteran terhadap penyakit kanker serviks menggunakan algoritma *K-Means* dan *K-Medoids*, tahap pemodelan pada penelitian ini adalah yang pertama buka aplikasi *RapidMiner Studio* dan *import* dataset yang telah melalui tahap pembersihan data. Masukkan operator proses *retrieve* yang berfungsi untuk memasukkan dataset yang sudah melalui tahap pembersihan data, dataset yang dimasukkan berupa data file excel. Setelah itu, tambahkan operator *Multiplay*, dikarenakan akan menguji dengan menggunakan dua algoritma secara bersamaan. Selanjutnya ditambahkan operator algoritma yaitu *K-Means* dan *K-Medoids* dengan menentukan nilai *k* pada masing-masing operator adalah 2, Setting parameter *measure types* pada kedua operator algoritma dengan menggunakan *BregmanDivergences* dengan tipe *divergence* adalah *SquaredEuclideanDistance*. Parameter *measure types* ini digunakan untuk memilih jenis ukuran yang akan digunakan untuk mencari tetangga terdekat. Selanjutnya menggunakan operator *Apply Model* digunakan untuk mempelajari informasi *ExampleSet* dan yang terakhir *Performance* untuk evaluasi statistik dari kinerja klastering dan memberikan daftar nilai kriteria dari kinerja klastering tersebut. Langkah berikutnya adalah hubungkan setiap operator kemudian Run untuk melihat hasilnya.

4.2 Pemodelan dengan algoritma *K-Means*:

Proses pemodelan pertama yaitu menggunakan algoritma *K-Means*, pemodelan dilakukan dengan bantuan *framework* data mining yaitu *Rapidminer* untuk mengetahui nilai DBI. Berikut ini adalah hasil pemodelan dengan menggunakan algoritma *K-Means*.

Row No.	id	cluster	Age	Number of s...	First sexual ...	Num of preg...	Smokes	Smokes (ye...	Smokes (pa...	Hormonal c
1	1	cluster_0	18	4	15	1	0	0	0	0
2	2	cluster_0	15	1	14	1	0	0	0	0
3	3	cluster_1	34	1	17	1	0	0	0	0
4	4	cluster_1	52	5	16	4	1	37	37	1
5	5	cluster_1	46	3	21	4	0	0	0	1
6	6	cluster_1	42	3	23	2	0	0	0	0
7	7	cluster_1	51	3	17	6	1	34	3	0
8	8	cluster_0	26	1	26	3	0	0	0	1
9	9	cluster_1	45	1	20	5	0	0	0	0
10	10	cluster_1	44	3	15	2	1	1	3	0
11	11	cluster_1	44	3	26	4	0	0	0	1
12	12	cluster_0	27	1	17	3	0	0	0	1
13	13	cluster_1	45	4	14	6	0	0	0	1
14	14	cluster_1	44	2	25	2	0	0	0	1
15	15	cluster_1	43	2	18	5	0	0	0	0
16	16	cluster_1	40	3	18	2	0	0	0	1

Gambar 4. 2 Hasil klasterisasi *K-Means*

PerformanceVector (*K-Means*)

PerformanceVector:

Avg. within centroid distance: 2.226

Avg. within centroid distance_cluster_0: 1.012

Avg. within centroid distance_cluster_1: 4.248

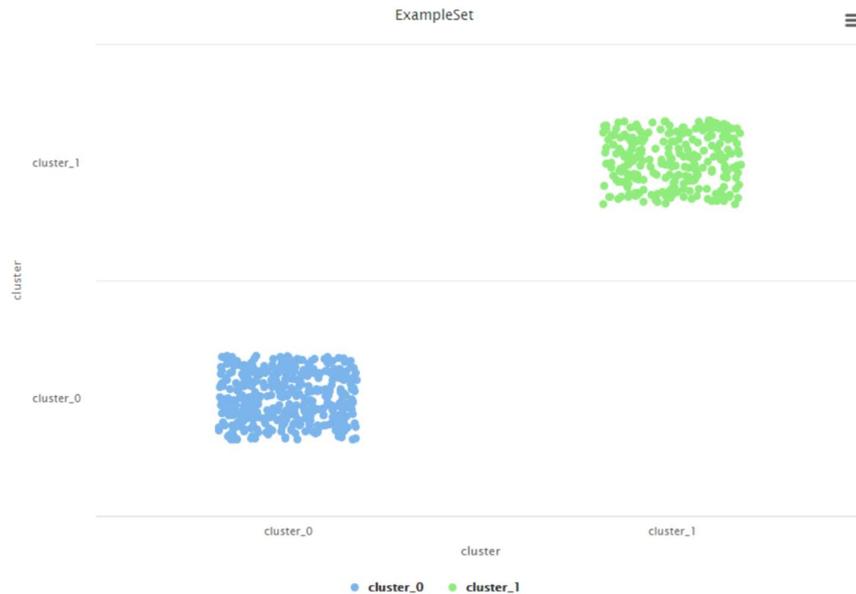
Davies Bouldin: 0.030

Cluster Model

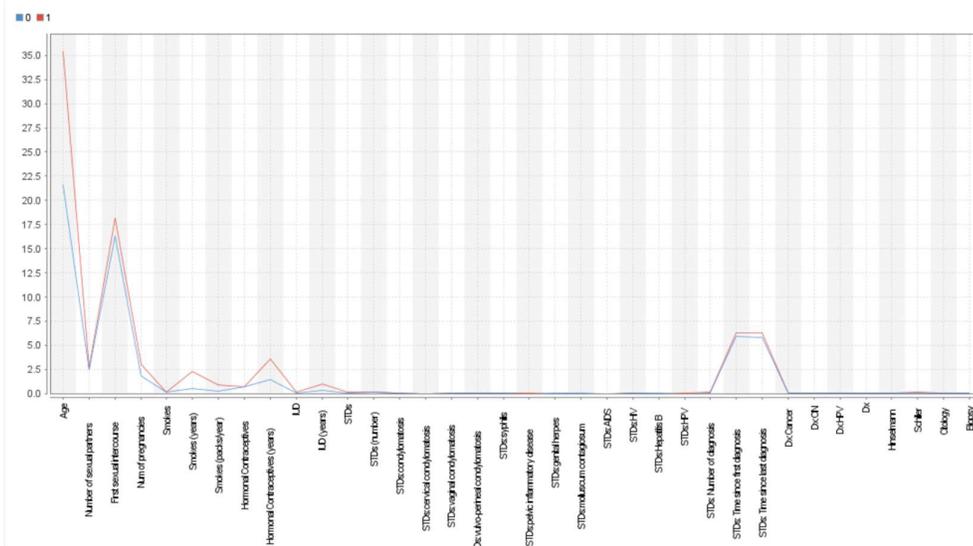
Cluster 0: 536 items

Cluster 1: 322 items

Total number of items: 858



Gambar 4. 3 Grafik Peyebaran Anggota Menggunakan Algoritma K-Means



Gambar 4. 4 Visualisasi Algoritma K-Means

Pada pemodelan dengan algoritma *K-Means*, dengan nilai k yang sudah ditentukan di awal yaitu 2, kluster positif dihasilkan sebanyak 322 pasien dan kluster negatif dihasilkan 536 pasien, Nilai *Davies Bouldin Index* yang dihasilkan sebesar 0.030.

4.3 Pemodelan dengan algoritma *K-Medoids*

Proses pemodelan kedua yaitu menggunakan algoritma *K-Medoids*, sama dengan pemodelan pertama, pemodelan kedua dilakukan dengan bantuan *framework* data mining yaitu *Rapidminer* untuk mengetahui nilai DBI. Berikut ini adalah hasil pemodelan dengan menggunakan algoritma *K-Means*.

Row No.	id	cluster	Age	Number of s...	First sexual ...	Num of preg...	Smokes	Smokes (ye...	Smokes (pa...	Hormonal t
1	1	cluster_0	18	4	15	1	0	0	0	0
2	2	cluster_0	15	1	14	1	0	0	0	0
3	3	cluster_1	34	1	17	1	0	0	0	0
4	4	cluster_1	52	5	16	4	1	37	37	1
5	5	cluster_1	46	3	21	4	0	0	0	1
6	6	cluster_1	42	3	23	2	0	0	0	0
7	7	cluster_1	51	3	17	6	1	34	3	0
8	8	cluster_1	26	1	26	3	0	0	0	1
9	9	cluster_1	45	1	20	5	0	0	0	0
10	10	cluster_1	44	3	15	2	1	1	3	0
11	11	cluster_1	44	3	26	4	0	0	0	1
12	12	cluster_1	27	1	17	3	0	0	0	1
13	13	cluster_1	45	4	14	6	0	0	0	1
14	14	cluster_1	44	2	25	2	0	0	0	1
15	15	cluster_1	43	2	18	5	0	0	0	0
16	16	cluster_1	40	3	18	2	0	0	0	1

Gambar 4. 5 Hasil klasterisasi *K-Medoids*

PerformanceVector (*K-Medoids*)

PerformanceVector:

Avg. within centroid distance: 3.202

Avg. within centroid distance_cluster_0: 1.632

Avg. within centroid distance_cluster_1: 5.301

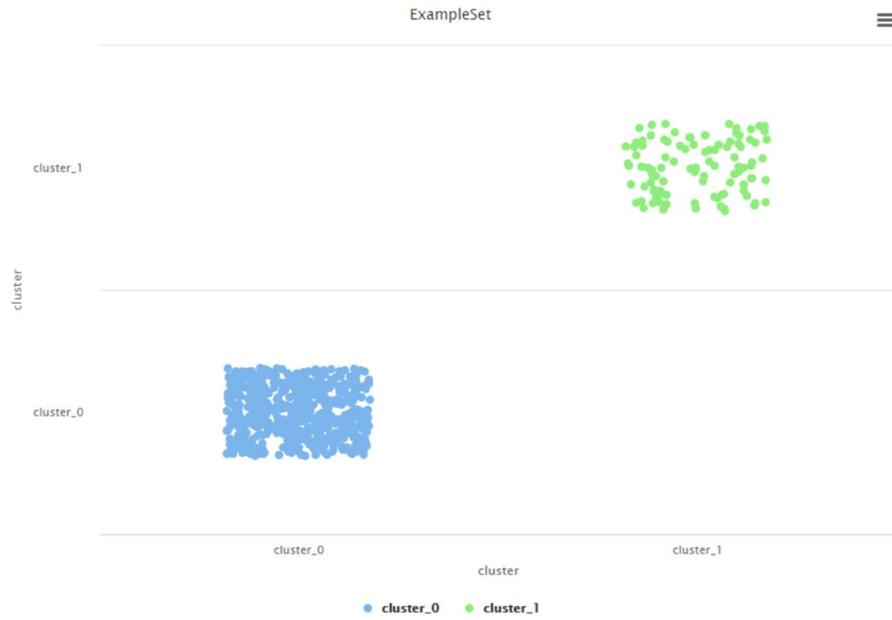
Davies Bouldin: 0.096

Cluster Model

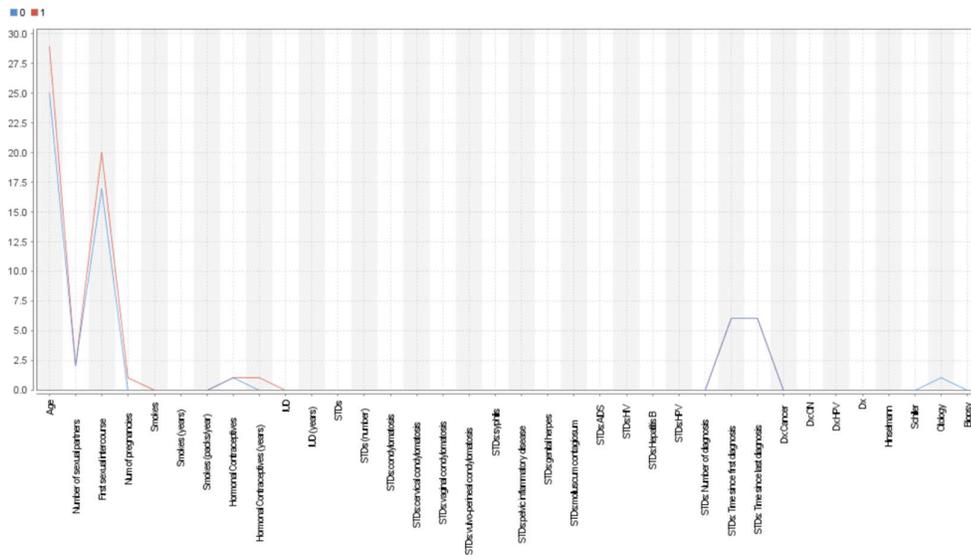
Cluster 0: 491 items

Cluster 1: 367 items

Total number of items: 858



Gambar 4. 6 Grafik Peyebaran Anggota Menggunakan Algoritma K-Medoids



Gambar 4. 7 Visualisasi Algoritma K-Medoids

Pada pemodelan dengan algoritma *K-Medoids*, kluster positif dihasilkan sebanyak 367 pasien dan kluster negatif dihasilkan sebanyak 491 pasien, Nilai *Davies Bouldin Index* yang dihasilkan sebesar 0.096.

4.4 Perbandingan perhitungan algoritma *K-Means* dan *K-Medoids*

Setelah dilakukan perhitungan dengan menggunakan algoritma *K-Means* dan *K-Medoids*, dapat dilihat perbedaan dalam pengklasteran data. Perbedaan hasil kedua algoritma tersebut dapat dilihat pada tabel berikut:

	Jumlah Kluster Positif	Jumlah Kluster Negatif
<i>K-Means</i>	322	536
<i>K-Medoids</i>	367	491

Tabel 4. 2 Hasil Klasterisasi

Pada tabel diatas, pemodelan algoritma *K-Means* terbentuk sebanyak 322 data pada kluster Positif dan 536 data pada kluster negatif, sedangkan pada algoritma *K-Medoids* terbentuk sebanyak 367 data pada kluster positif dan 491 data pada kluster negatif.