

Perbandingan Kinerja Algoritma K-Medoids Dan K-Means Untuk Klasifikasi Penyakit Kanker Serviks

Dedi Arbain^{1a*}, Sriyanto^{2b}, Joko Triloka^{3c}

^{a b c}Fakultas Ilmu Komputer, Informatics & Business Institute DarmajayaJl. Z.A. Pagar Alam No. 93, Bandar Lampung - Indonesia 35142 Telp. (0721) 787214 Fax. (0721) 700261

^adediarbain.2221210048@mail.darmajaya.ac.id

^bpaksriyanto@gmail.com

^cjoko.triloka@darmajaya.ac.id

Abstract

K-Medoids and K-Means are unsupervised algorithms that use Cluster Distance Performance to group data. Cluster Distance Performance is a distance measurement method that can help an algorithm group objects based on the similarity of the variables. Several studies have shown that using the right Cluster Distance Performance can improve the performance of the algorithm in clustering. This study aims to compare the results of clustering cervical cancer datasets using the K-Medoids and K-Means Clustering methods. The cervical cancer dataset is 858 records and 36 attributes. The clusters produced by the two methods are 2 classes. K-Means is more effective in dealing with small data sizes. The K-Medoids algorithm model formed 361 data in the positive cluster and 473 data in the negative cluster, while the K-Means algorithm formed 308 data in the positive cluster and 526 data in the negative cluster. In the cervical cancer dataset using K-Medoids it showed a DBI result of -3.517, whereas using K-Means the evaluation results showed a result of -1.108. Thus, clustering using the K-Medoids Clustering method has better results compared to the K-Means Clustering method, because it produces a smaller DBI value of -1.108.

Keywords: Clustering; K-Means; K-Medoids; Cervical cancer.

Abstrak

K-Medoids dan K-Means adalah algoritme unsupervised yang menggunakan Cluster Distance Performance untuk mengelompokkan data. Cluster Distance Performance merupakan metode pengukuran jarak yang dapat membantu sebuah algoritma mengelompokkan objek berdasarkan kemiripan variabel-variabelnya. Beberapa penelitian telah menunjukkan bahwa penggunaan Cluster Distance Performance yang tepat dapat meningkatkan performa algoritma dalam melakukan klastering. Penelitian ini bertujuan untuk mengetahui perbandingan hasil pengklasteran dataset kanker serviks menggunakan metode K-Medoids dan K-Means Clustering. Dataset kanker serviks yang berjumlah 858 Record dan 36 Atribut. Cluster yang dihasilkan kedua metode adalah 2 kelas. K-Means lebih efektif dalam menangani data dengan ukuran kecil. Pemodelan algoritma K-Medoids terbentuk sebanyak 361 data pada klaster Positif dan 473 data pada klaster negatif, sedangkan pada algoritma K-Means terbentuk sebanyak 308 data pada klaster positif dan 526 data pada klaster negatif. Pada dataset kanker serviks menggunakan K-Medoids menunjukkan hasil DBI sebesar -3,517, sedangkan dengan menggunakan K-Means hasil evaluasi menunjukkan hasil sebesar -1,108. Dengan demikian, pengklasteran menggunakan metode K-Medoids Clustering memiliki hasil yang lebih baik dibandingkan dengan metode K-Means Clustering, karena menghasilkan nilai DBI yang lebih kecil yaitu -1,108.

Kata Kunci: Clustering; K-Means; K-Medoids; Kanker Serviks.

1. PENDAHULUAN

Kanker mulut rahim masih merupakan masalah kesehatan perempuan di Indonesia sehubungan dengan angka kejadian dan angka kematiannya yang tinggi[1]. Berdasarkan pernyataan Departemen Kesehatan Republik Indonesia pada tahun 2009 yang dikutip oleh Yustina Ananti, Kanker serviks merupakan penyakit yang disebabkan munculnya sel kanker ganas pada sel leher rahim. Keganasan sel tersebut disebabkan oleh virus yang bernama Human Papiloma Virus (HPV) tipe 16 dan 18 yang umum dijumpai di Indonesia[2]. Adapun faktor risiko terjadinya kanker serviks antara lain adalah aktivitas seksual pada usia muda, berhubungan seksual dengan multipartner, merokok, mempunyai anak banyak, sosial ekonomi rendah, pemakaian pil KB (dengan HPV negatif atau positif), penyakit menular seksual, dan gangguan imunitas. Menurut para ahli kanker, kanker serviks merupakan salah satu jenis kanker yang paling dapat dicegah dan paling dapat disembuhkan dari semua kasus kanker. Salah satu kegiatan deteksi dini kanker serviks yang paling umum di Indonesia adalah menggunakan metode Pap Smear. Sebagai contoh pada Rumah Sakit Umum Daerah (RSUD) Kediri dan Yayasan Kanker Indonesia (YKI) kabupaten Kediri merupakan penyedia layanan untuk tes Pap Smear. Tersedianya data histori rekam medis pasien tidak disertai dengan proses ekstraksi data menjadi sebuah

pengetahuan atau informasi yang dapat berguna untuk keputusan klinis. Penggunaan teknik data mining sangat potensial untuk diterapkan dalam pelayanan kesehatan, khususnya untuk prediksi penyakit kanker serviks[3].

Beberapa penelitian terkait diantaranya adalah yang tertulis pada jurnal dengan judul “Clustering Kanker Serviks Berdasarkan Perbandingan Euclidean dan Manhattan Menggunakan Metode K-Means”[5], Tool yang digunakan adalah Weka. Hasil akurasi ditampilkan dengan bantuan Confusion Matrix dan kurva ROC, Nilai akurasi didapat sebesar 79,30% dengan ROC 79,17% pada K-Means Euclidean Metric sedangkan K-Means Manhattan Metric sebesar 67,83% dengan ROC 65,94%. Selanjutnya jurnal yang berjudul “Perbandingan Tingkat Kepuasan Siswa Terhadap Pelayanan Sekolah Menggunakan Algoritma K-Means Dan K-Medoids”. Didapatkan hasil pada klaster K-Means yaitu klaster puas sebanyak 276 siswa, klaster cukup puas sebanyak 216 siswa dan klaster kurang puas sebanyak 17 siswa. Kemudian pada algoritma K-Medoids yaitu klaster puas sebanyak 324 siswa, klaster cukup puas sebanyak 11 siswa dan klaster kurang puas sebanyak 174 siswa. Kajian-kajian tersebut hanya sebatas menghitung jumlah klaster yang terbentuk dari masing-masing algoritma, namun hasil perhitungan tersebut belum dilakukan validasi klaster dimana tujuan dari validasi klaster adalah untuk mengetahui seberapa baik proses kinerja pengklasteran. Berdasarkan dari analisa kasus dan hasil review jurnal, pada penelitian ini penulis akan menggunakan teknik data mining dengan judul “Perbandingan Algoritma K-Medoids dan K-Means untuk Klasterisasi Penyakit Kanker Serviks”.

2. KERANGKA TEORI

2.1. Data mining

Data mining adalah proses analitik yang dirancang untuk memeriksa sejumlah data yang besar dalam mencari suatu pengetahuan tersembunyi yang berharga dan konsisten. Tujuan dari data mining yaitu mencari trend atau pola yang diinginkan dalam database besar untuk membantu dalam pengambilan keputusan pada waktu yang akan datang [4]. Secara umum kegunaan data mining dapat dibagi menjadi dua yaitu deskriptif dan prediktif. Deskriptif berarti data mining digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti data mining digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi [6]. Salah satu metode yang terdapat pada data mining yaitu Clustering yang berperan penting dalam mengeksplorasi data, menciptakan prediksi dan mengatasi anomali data. Pengelompokan dalam suatu cluster, dikelompokkan menggunakan teknik reiteratif/ perulangan dengan karakteristik identik suatu dataset [8].

2.2. Kanker serviks

Kanker serviks adalah tumor ganas primer yang berasal dari sel epitel skuamosa. Sebelum terjadinya kanker, akan didahului oleh keadaan yang disebut lesi prakanker atau neoplasia intraepitel serviks (NIS) [6]. Kanker ini adalah kanker yang terjadi pada area leher rahim yaitu bagian rahim yang menghubungkan rahim bagian atas dengan vagina. Kanker serviks disebabkan infeksi virus HPV (Human Papilloma Virus) atau virus papiloma manusia. HPV menimbulkan kutil pada pria maupun wanita, termasuk kutil pada kelamin yang disebut kondiloma akuminatum. Hanya beberapa saja dari ratusan varian HPV yang dapat menyebabkan kanker. Kanker serviks atau kanker leher rahim bisa terjadi jika terjadi infeksi yang tidak sembuh-sembuh untuk waktu lama. Sebaliknya, kebanyakan infeksi HPV akan hilang sendiri, teratasi oleh sistem kekebalan tubuh. Pada tahap awal, penyakit ini tidak menimbulkan gejala yang mudah diamati. Gejala fisik serangan penyakit ini pada umumnya hanya dirasakan oleh penderita kanker stadium lanjut. Apabila kanker sudah menyebar ke panggul, maka pasien akan menderita keluhan nyeri punggung, hambatan dalam berkemih, serta pembesaran ginjal. Akan tetapi apabila ditangani lebih cepat maka kemungkinan penyembuhan terhadap kanker bisa diatasi tergantung dari stadium dari kanker tersebut [7].

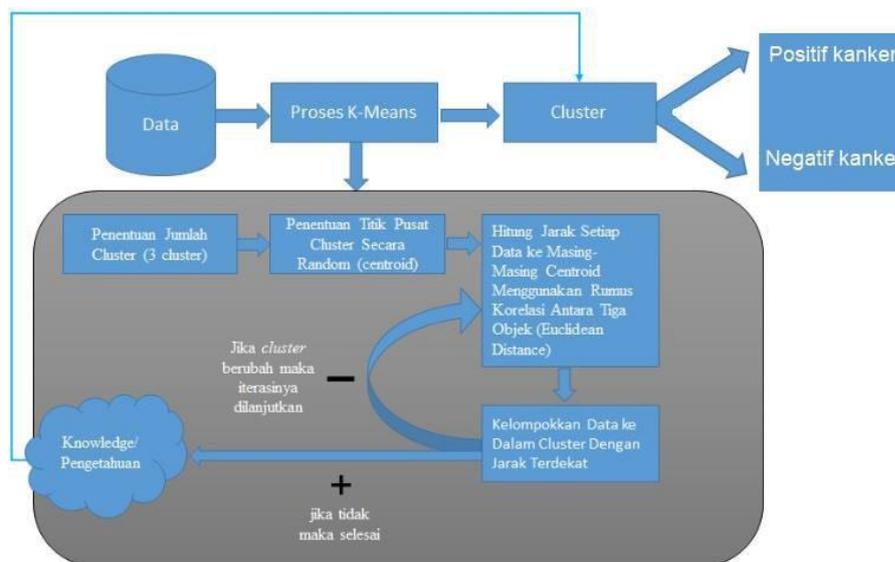
2.3. Clustering

Clustering merupakan suatu proses pengelompokan data, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek [9]. Berbeda dengan proses klasifikasi, clustering tidak mempunyai target variable dalam melakukan. Clustering sering dilakukan sebagai langkah awal dalam proses data mining. Terdapat banyak algoritma klustering yang telah digunakan oleh peneliti sebelumnya seperti K-Means, Improved K-Means, K-Medoids (PAM), Fuzzy C-Means, DBSCAN, CLARANS dan Fuzzy Subtractive [10]. Clustering telah digunakan secara luas dan pentingnya pengelompokan tumbuh dengan cepat dikarenakan jumlah data yang berhubungan dengan eksponen aljabar dalam kecepatan pengolahan komputer sangat banyak. Algoritma clustering berfungsi untuk mengelompokkan data sesuaidengan karakteristik dan mengukur jarak kemiripan antar data dalam satu kelompok walaupun setiap algoritma clustering memiliki kelebihan dan kekurangan masing-masing [11].

2.4. K-Means

K-Means clustering ialah metode analisis cluster yang bertujuan untuk memecah objek menjadi k cluster kemudian diamati di mana setiap objek cluster diperoleh melalui rata-rata terdekat. Algoritma ini merupakan salah satu pembelajaran yang terkenal sederhana dan mudah dipelajari sebagai pemecahan masalah pengelompokan dari sebuah dataset. Algoritma K-Means adalah algoritma evolusi yang mana metode operasinya memiliki artian yang serupa dengan nama algoritmanya[12]. Algoritma ini mengelompokkan pengamatan ke dalam kelompok k, di mana k merupakan parameter input. Tiap data kemudian ditetapkan pada setiap pengamatan cluster berdasarkan kedekatan pengamatan nilai rata-rata cluster[13].

K-Means merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Dengan demikian data yang memiliki kemiripan berada pada satu cluster yang sama dan data yang memiliki ketidaksamaan berada pada cluster yang lain. Algoritma *K-Means*, mungkin yang pertama dari algoritma pengelompokan yang diusulkan, didasarkan pada ide yang sangat sederhana. Diberi serangkaian kumpulan awal, menetapkan setiap titik ke salah satunya, lalu setiap pusat gugus diganti dengan titik rata-rata pada klaster masing-masing. Dua langkah sederhana ini diulang hingga konvergensi. Suatu titik ditugaskan ke cluster yang dekat dalam jarak Euclidean ke titik.



Gambar 1. Alur Kerja K-Means Clustering

Pada gambar 1 merupakan proses algoritma *K-Means*, dimana tahap pertama adalah mengumpulkan data dan *pre-processing data*. Tahapan selanjutnya adalah proses pengolahan *K-Means* dengan menentukan jumlah *cluster* secara random, titik pusat *cluster*, menghitung jarak setiap data dan mengelompokkan data pada *cluster* sehingga menghasilkan *knowledge* yaitu *clustering*.

2.5. K-Medoids

Algoritma PAM (Partitioning Around Medoids) atau biasa juga disebut dengan algoritma K- Medoids, merupakan algoritma yang diwakili oleh cluster yaitu medoid. Perbedaan antara algoritma K- Medoids dengan algoritma K-Means yaitu algoritma K-Medoids menggunakan objek sebagai perwakilan (medoid) pusat cluster untuk tiap cluster, sementara algoritma K-Means membutuhkan nilai rata-rata (mean) sebagai pusat cluster.

2.6. Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) adalah suatu ukuran yang digunakan untuk menentukan jumlah klaster yang terbaik setelah proses pengklasteran selesai [10]. Pendekatan DBI ini bertujuan untuk memaksimalkan jarak antara klaster yang satu dengan klaster yang lain dan mencoba meminimalkan jarak antar objek dalam suatu klaster. Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0) maka semakin baik klaster yang diperoleh dari pengelompokan k- modes clustering yang digunakan [14]. Dengan menggunakan DBI suatu cluster akan dianggap memiliki skema clustering yang optimal adalah yang memiliki DBI minimal [14].

3. METODOLOGI

3.1. Pengumpulan Data

Data diperoleh dari dataset kanker serviks *UCI Machine Learning Repository* yang berjumlah 858 record dan 36 attribut. Data yang diambil berisikan data attribut dataset kanker serviks.

Tabel 1. Atribut Dataset Penyakit Kanker Serviks

| | |
|---|---|
| <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. |

STDs: Number of diagnosis

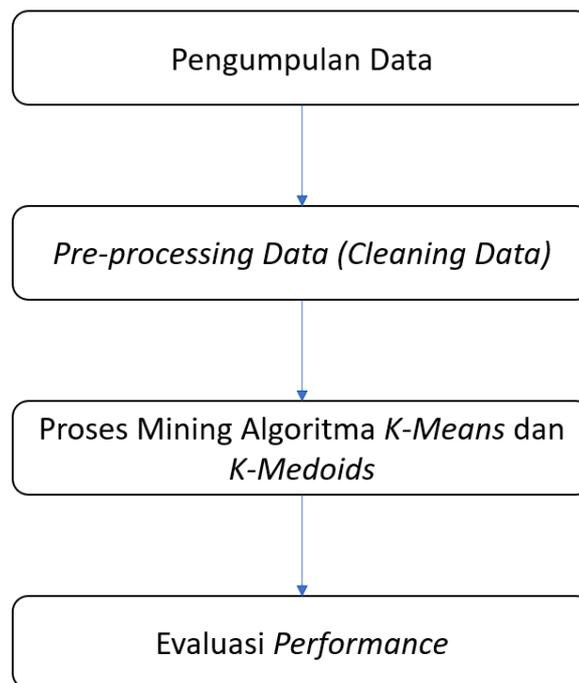
Jumlah penyakit kelamin yang pernah didiagnosis pasien. Memiliki rentang nilai 0 – 3.

3.2. Penerapan Algoritma

Algoritma yang digunakan untuk perhitungan pada penelitian ini adalah algoritma K-Medoids dan K-Means dengan membagi menjadi 2 kluster yaitu kluster negatif (0) dan kluster positif (1). Perhitungan dilakukan menggunakan aplikasi rapid miner.

3.3. Kerangka Kerja Penelitian

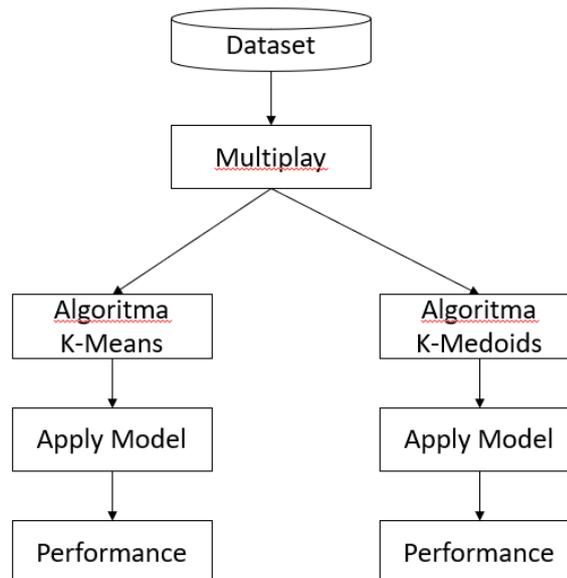
Kerangka kerja penelitian yang dibuat secara sistematis yang mengarah pada tahapan proses yang akan ditempuh pada penelitian ini dapat dilihat pada gambar 2.



Gambar 2. Alur Dalam Tahapan Penelitian

Berdasarkan gambar 2 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. Bagan Permodelan

4. Evaluasi Performance.

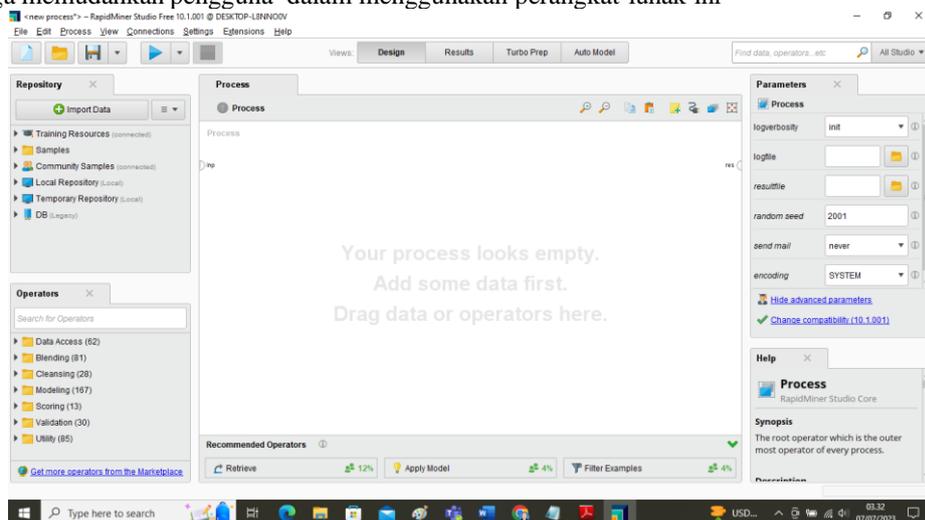
Proses evaluasi *performance* hasil klustering dengan algoritma *K-Means* dan *K-Medoids* yaitu menggunakan metode *Davies-Bouldin Index* (DBI).

4. HASIL DAN PEMBAHASAN

Dalam penelitian ini pengujian metode algoritma *K-Medoids* dan *K-Means*, tools yang digunakan untuk mempermudah perhitungan yaitu dengan menggunakan software RapidMiner Studio. Berikut adalah penerapan tool RapidMiner pada pengujian Algoritma *K-Medoids* dan *K-Means* :

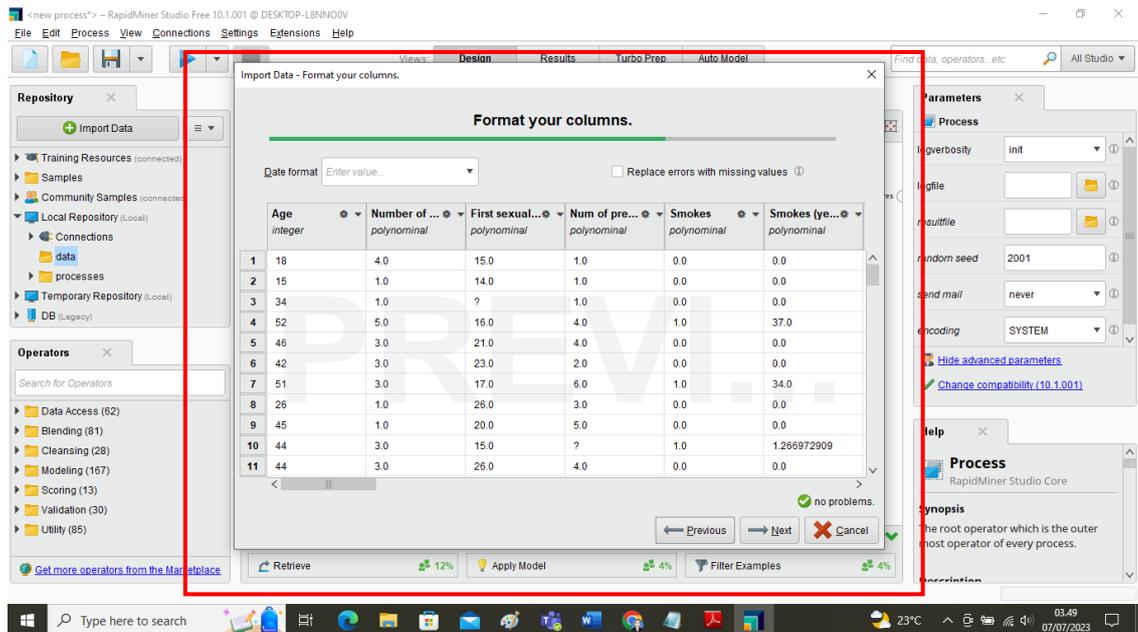
1. Membuka terlebih dahulu aplikasi RapidMiner.

RapidMiner adalah perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari *Institute of Technology Blanchardstown* dan Raif Klinkenberg dari *rapid-i.com* dengan tampilan *GUI* (*Graphical User Interface*) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini



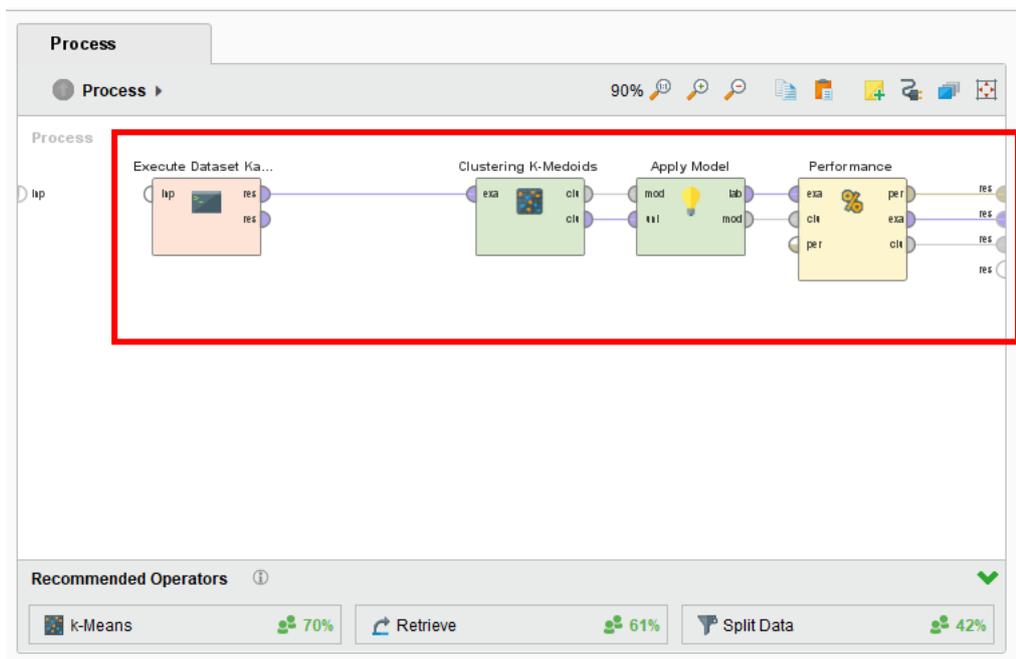
Gambar 4. Tampilan Rapidminer 10

2. Dataset yang akan dimasukkan ke dalam Rapidminer adalah Dataset yang digunakan berasal dari kaggle.com dengan dataset `kag_risk_factors_cervical_cancer.csv`. Pada gambar 5 dibawah ini dapat dilihat cara pengaturan dataset dalam melakukan masukkan ke dalam tool RapidMiner.

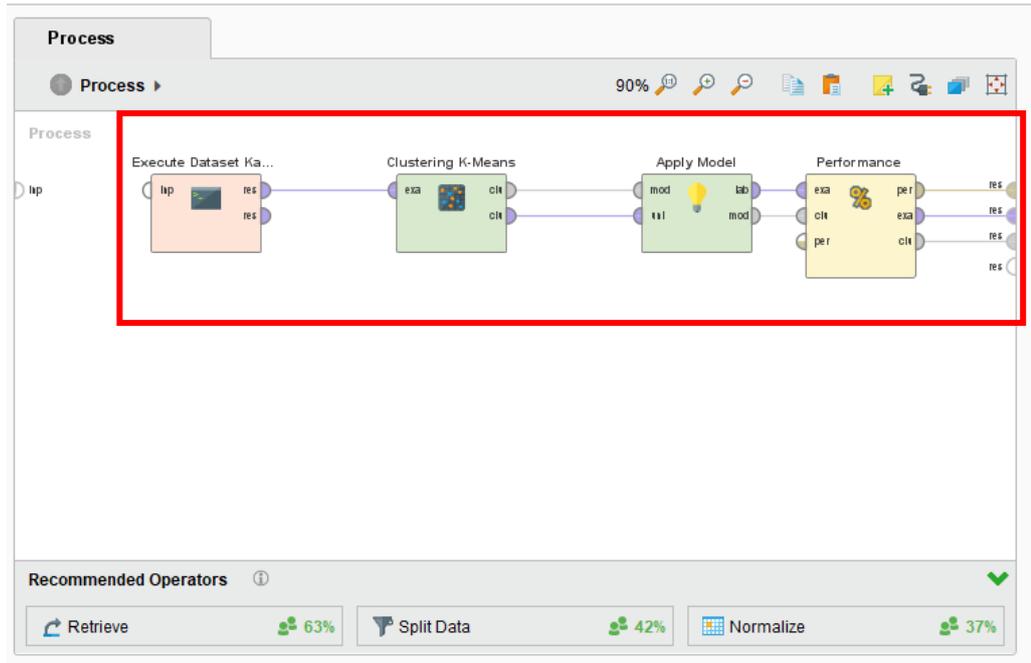


Gambar 5. Pengaturan dan Proses Impor Dataset ke dalam RapidMiner

3. Peneliti juga membuat desain clustering menggunakan algoritma K-Medoids dan K-Means seperti yang ditunjukkan pada gambar 6 dan 7. Dataset yang telah diimport ke Rapidminer, disambungkan ke Algoritma Clustering K-Medoids dan K-Means maka selanjutnya disambungkan lagi ke Apply Model, lalu ke Performance dan terakhir disambungkan ke Result.



Gambar 6. Desain Clustering Menggunakan Algoritma K-Medoids



Gambar 7. Desain Clustering Menggunakan Algoritma K-Means

- Proses pemodelan kedua yaitu menggunakan algoritma K-Medoids dan K-Means, sama dengan pemodelan pertama, pemodelan kedua dilakukan dengan bantuan framework data mining yaitu Rapidminer untuk mengetahui nilai DBI. seperti yang ditunjukkan pada gambar 8 dan 9.

| Row No. | id | label | Age | Number of s... | First sexual ... | Num of preg... | Smokes | Smokes (ye... | Sm... |
|---------|----|-----------|-----|----------------|------------------|----------------|--------|---------------|-------|
| 1 | 1 | cluster_0 | 18 | 4 | 15 | 1 | 0 | 0 | 0 |
| 2 | 2 | cluster_0 | 15 | 1 | 14 | 1 | 0 | 0 | 0 |
| 3 | 3 | cluster_1 | 34 | 1 | 17 | 1 | 0 | 0 | 0 |
| 4 | 4 | cluster_1 | 52 | 5 | 16 | 4 | 1 | 37 | 37 |
| 5 | 5 | cluster_1 | 46 | 3 | 21 | 4 | 0 | 0 | 0 |
| 6 | 6 | cluster_1 | 42 | 3 | 23 | 2 | 0 | 0 | 0 |
| 7 | 7 | cluster_1 | 51 | 3 | 17 | 6 | 1 | 34 | 3.40 |
| 8 | 8 | cluster_1 | 26 | 1 | 26 | 3 | 0 | 0 | 0 |
| 9 | 9 | cluster_1 | 45 | 1 | 20 | 5 | 0 | 0 | 0 |
| 10 | 10 | cluster_1 | 44 | 3 | 15 | 2 | 1 | 1.267 | 2.80 |
| 11 | 11 | cluster_1 | 44 | 3 | 26 | 4 | 0 | 0 | 0 |
| 12 | 12 | cluster_1 | 27 | 1 | 17 | 3 | 0 | 0 | 0 |
| 13 | 13 | cluster_1 | 45 | 4 | 14 | 6 | 0 | 0 | 0 |
| 14 | 14 | cluster_1 | 44 | 2 | 25 | 2 | 0 | 0 | 0 |

Gambar 8. Hasil klusterisasi K-Medoids

| Row No. | id | label | Age | Number of s... | First sexual ... | Num of preg... | Smokes | Smokes (ye... | Smo |
|---------|----|-----------|-----|----------------|------------------|----------------|--------|---------------|------|
| 1 | 1 | cluster_0 | 18 | 4 | 15 | 1 | 0 | 0 | 0 |
| 2 | 2 | cluster_0 | 15 | 1 | 14 | 1 | 0 | 0 | 0 |
| 3 | 3 | cluster_1 | 34 | 1 | 17 | 1 | 0 | 0 | 0 |
| 4 | 4 | cluster_1 | 52 | 5 | 16 | 4 | 1 | 37 | 37 |
| 5 | 5 | cluster_1 | 46 | 3 | 21 | 4 | 0 | 0 | 0 |
| 6 | 6 | cluster_1 | 42 | 3 | 23 | 2 | 0 | 0 | 0 |
| 7 | 7 | cluster_1 | 51 | 3 | 17 | 6 | 1 | 34 | 3.40 |
| 8 | 8 | cluster_0 | 26 | 1 | 26 | 3 | 0 | 0 | 0 |
| 9 | 9 | cluster_1 | 45 | 1 | 20 | 5 | 0 | 0 | 0 |
| 10 | 10 | cluster_1 | 44 | 3 | 15 | 2 | 1 | 1.267 | 2.80 |
| 11 | 11 | cluster_1 | 44 | 3 | 26 | 4 | 0 | 0 | 0 |
| 12 | 12 | cluster_0 | 27 | 1 | 17 | 3 | 0 | 0 | 0 |
| 13 | 13 | cluster_1 | 45 | 4 | 14 | 6 | 0 | 0 | 0 |
| 14 | 14 | cluster_1 | 44 | 2 | 25 | 2 | 0 | 0 | 0 |

Gambar 9. Hasil Klasterisasi K-Means

- Setelah pengujian berjalan maka didapatkan informasi bahwa lama waktu pemrosesan clustering dengan Algoritma K-Medoids sebesar 30 s, sedangkan pemrosesan clustering dengan Algoritma K-Means sebesar 0 detik, seperti yang tercatat pada gambar 10 dan 11.

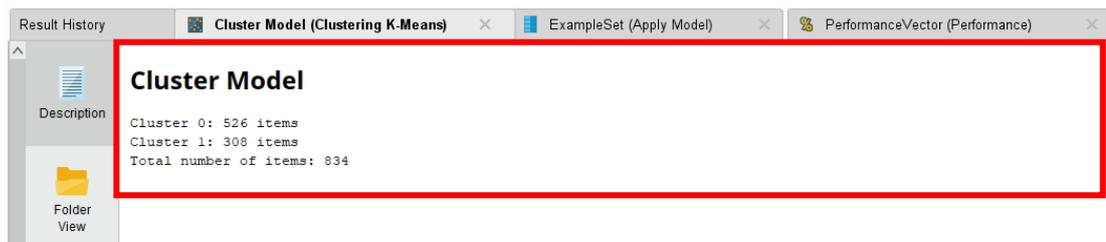
Gambar 10. Result History Clustering menggunakan Algoritma K-Medoids

Gambar 11. Result History Clustering menggunakan Algoritma K-Means

6. Pada Gambar 12 dan 13 menunjukkan bahwa pada penelitian ini, kluster yang digunakan berjumlah 2 kelas. Clustering data berdasarkan algoritma K- Medoids, Kluster positif memiliki data sebanyak 361, kluster negatif sebanyak 473 dengan jumlah dataset sebanyak 834 data. Sedangkan berdasarkan algoritma K- Means, kluster positif memiliki data sebanyak 308, kluster negatif sebanyak 526 data, dengan jumlah dataset sebanyak 834 data.

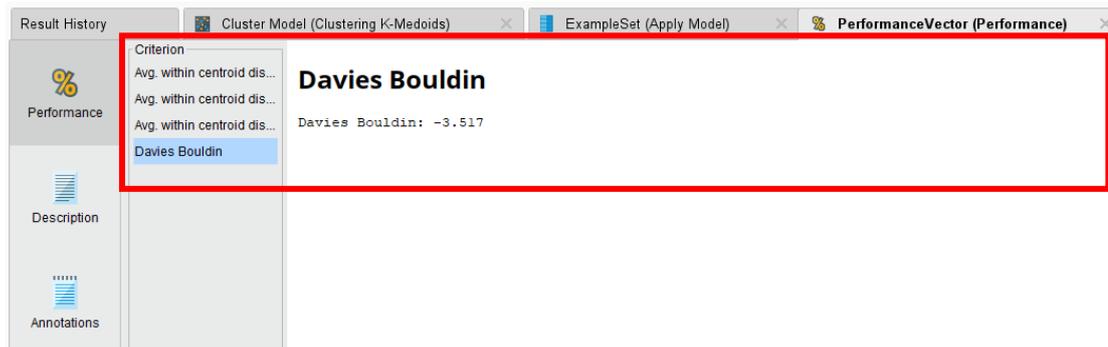


Gambar 12. Hasil Cluster Model K-Medoids

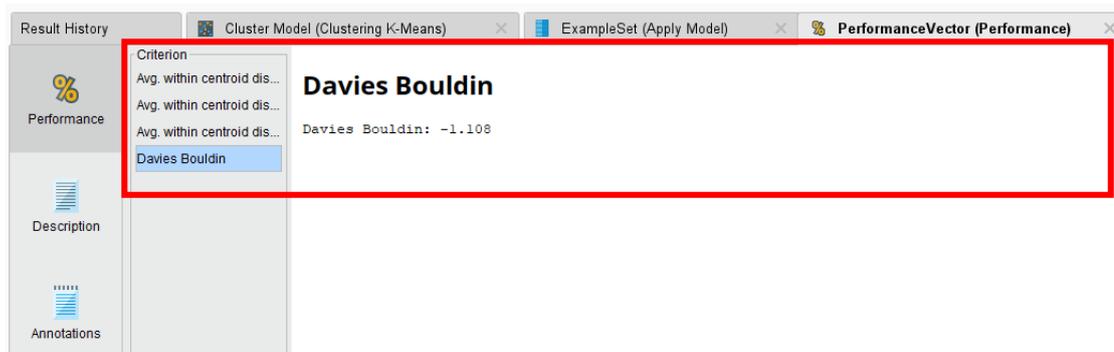


Gambar 13. Hasil Cluster Model K-Means

7. Dapat dilihat juga Davies Bouldin Index (DBI) yang ada didalam Performance Vector K Medoids dan K-Means pada gambar 14 dan gambar 15. DBI merupakan metode validasi cluster dari hasil clustering. nilai DBI untuk validasi dengan metode K- Medoids adalah sebesar -3,517. Sedangkan nilai DBI dengan metode K-Means adalah sebesar -1,108. Semakin rendah nilai DBI, maka cluster tersebut semakin baik.

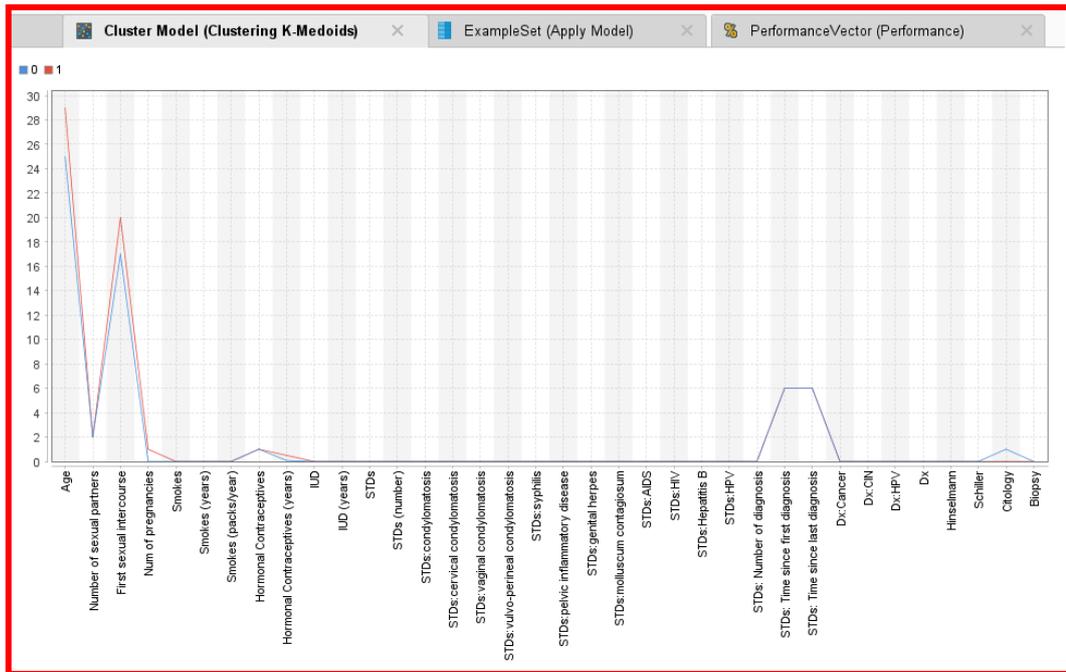


Gambar 14. Davies Bouldin Index untuk Performance Vector K-Medoids

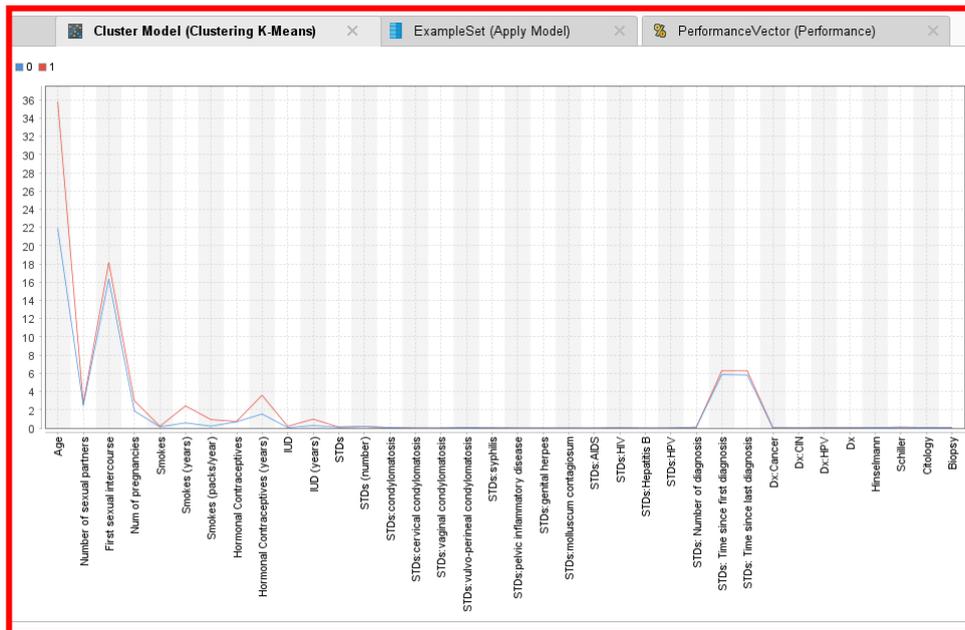


Gambar 15. Davies Bouldin Index untuk Performance Vector K-Means

8. Dapat dilihat juga grafik Visualisasi Algoritma K-Medoids dan K-Means pada gambar 16 dan gambar 17.

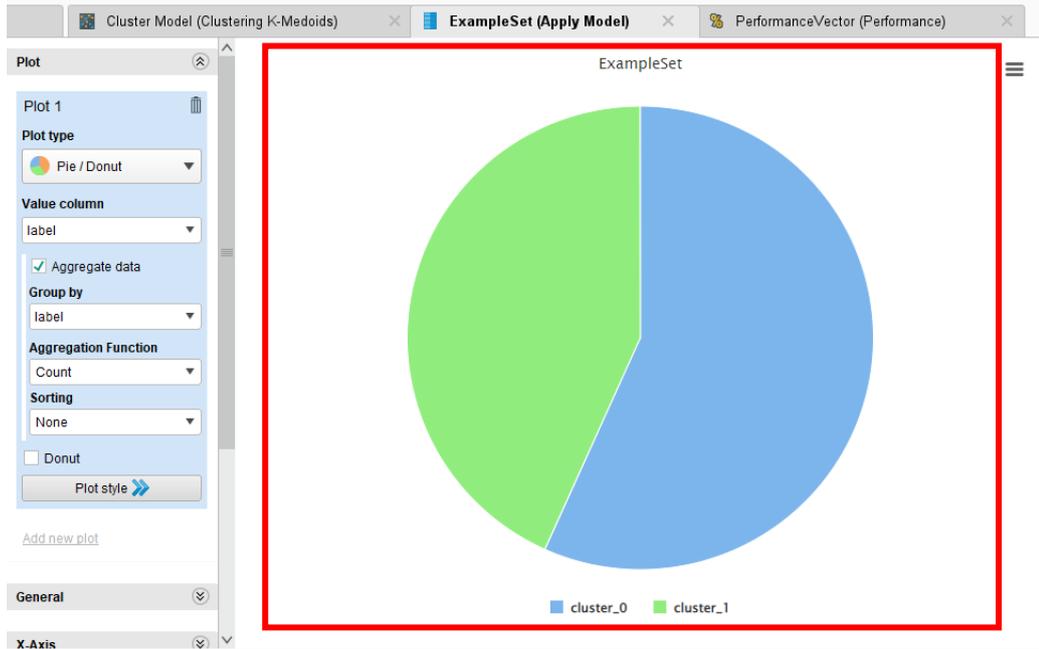


Gambar 16. Visualisasi Algoritma K-Medoids

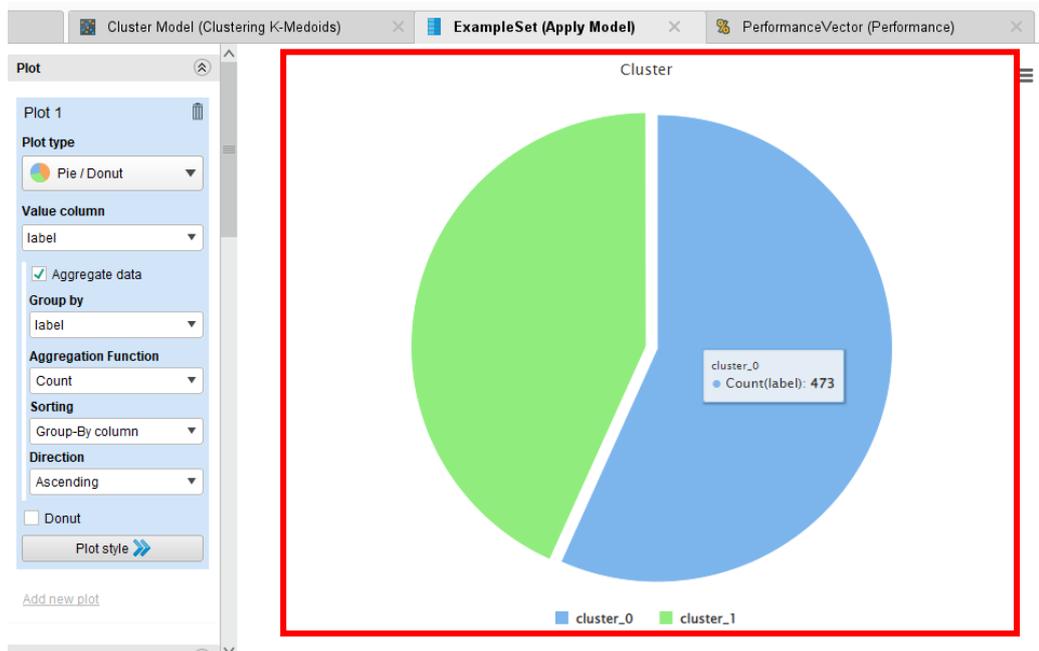


Gambar 17. Visualisasi Algoritma K-Means

9. Dapat dilihat juga Grafik Penyebaran Anggota Menggunakan algoritma K-Medoids dan K-Means pada gambar 18 dan gambar 19.



Gambar 18. Grafik Peyebaran Anggota Menggunakan Algoritma K-Medoids



Gambar 19. Grafik Peyebaran Anggota Menggunakan Algoritma K-Means

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

Tabel 2. Hasil Klasterisasi

| | Jumlah Kluster Positif (1) | Jumlah Kluster Negatif (0) | Davies-Bouldin Index (DBI) |
|------------------|----------------------------|----------------------------|----------------------------|
| K-Medoids | 361 Pasien | 473 Pasien | -3,517 |
| K-Means | 308 Pasien | 526 Pasien | -1,108 |

Pada tabel 2, Pemodelan algoritma K-Medoids terbentuk sebanyak 361 pasien pada kluster Positif dan 473 pasien pada kluster negatif, sedangkan pada algoritma K-Means terbentuk sebanyak 308 pasien pada kluster positif dan 526 pasien pada kluster negatif. Pada dataset kanker serviks menggunakan K-Medoids menunjukkan hasil DBI sebesar -3,517, sedangkan dengan menggunakan K-Means hasil evaluasi menunjukkan hasil sebesar -1,108.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat diketahui bahwa kedua algoritma pengelompokan yaitu K-Medoids dan K-Means dengan dilakukan eksperimen dengan dataset yang berukuran kecil, maka K-Means lebih efektif dalam menangani data dengan ukuran kecil. Pemodelan algoritma K-Medoids terbentuk sebanyak 361 pasien pada kluster Positif dan 473 pasien pada kluster negatif, sedangkan pada algoritma K-Means terbentuk sebanyak 308 pasien pada kluster positif dan 526 pasien pada kluster negatif. Pada dataset kanker serviks menggunakan K-Medoids menunjukkan hasil DBI sebesar -3,517, sedangkan dengan menggunakan K-Means hasil evaluasi menunjukkan hasil sebesar -1,108. Hal ini disebabkan karena K-Medoids memiliki performa komputasi yang rendah dibandingkan dengan K-Means. Untuk penelitian selanjutnya dapat dilakukan evaluasi pada kedua algoritma klustering di atas dengan dilakukan eksperimen dengan dataset berukuran besar.

DAFTAR PUSTAKA

- [1] I. Rasjidi, "Epidemiologi Kanker Serviks," *Indones. J. Cancer*, vol. 3, no. 3, pp. 103–108, 2009, doi: 10.33371/ijoc.v3i3.123.
- [2] E. Suwandy Salim, H. Bindan, E. Pranoto, and A. Dharma, "Analisa Metode Random Forest Tree dan K-Nearest Neighbor dalam Mendeteksi Kanker Serviks," 2020.
- [3] T. Praningsih and I. Budi, "Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN Cervical Cancer Disease Prediction System Using CART, Naive Bayes, and k-NN," 2017.
- [4] H. Sariangah, W. Wanayumini, and R. Rosnelly, "Penentuan Kelas Menggunakan Algoritma K Medoids Untuk Clustering Siswa Tunagrahita," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 1, p. 83, Jan. 2021, doi: 10.30865/mib.v5i1.2547.
- [5] S. Widodo, H. Brawijaya, and S. Samudi, "Clustering Kanker Serviks Berdasarkan Perbandingan Euclidean dan Manhattan Menggunakan Metode K-Means," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 687, Apr. 2021, doi: 10.30865/mib.v5i2.2947.
- [6] B. Serasi Ginting and M. Simanjuntak, "Pengelompokan Penyakit Pada Pasien Berdasarkan Usia Dengan Metode K-Means Clustering (Studi Kasus : Puskesmas Bahorok)," *Algoritma. J. Ilmu Komput. dan Inform.*, vol. 6341, no. November, p. 2, 2021.
- [7] A. J. Wahidin and D. I. Senses, "Perbandingan Algoritma K-Means, X-Means Dan K-Medoids Untuk Klasterisasi Awak Kabin Lion Air," *J. ICT Inf. Commun. Technol.*, vol. 20, no. 2, pp. 298–302, 2021, doi: 10.36054/jict-ikmi.v20i2.387.
- [8] D. Transaksi Bongkar Muat di Provinsi Riau, I. Kamila, U. Khairunnisa, P. Studi Sistem Informasi, and F. Sains dan Teknologi UIN Sultan Syarif Kasim Riau, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019.
- [9] R. Rousyati, F. F. Wati, D. Pratomanto, and A. Crisna, "Pengelompokan Siswa Penyandang Disabilitas Berdasarkan Tingkat Tunagrahita Menggunakan Algoritma K-Medoids," *Indones. J. Softw. Eng.*, vol. 5, no. 1, pp. 134–142, 2019, doi: 10.31294/ijse.v5i1.6550.
- [10] V. A. P. Sangga, "Perbandingan Algoritma K-Means dan Algoritma K-Medoids dalam Pengelompokan Komoditas Peternakan di Provinsi Jawa Tengah Tahun 2015," *Tugas Akhir Jur. Stat. Fak. Mat. dan Ilmu Pengetah. Alam Univ. Islam Inndonesia Yogyakarta*, vol. 53, no. 9, pp. 1689–1699, 2018.
- [11] Y. Hilda Susanti, E. Widodo, and R. Artikel, "Perbandingan K-Means dan K-Medoids Clustering terhadap Kelayakan Puskesmas di DIY Tahun 2015," *Seminar Nasional Integrasi Matematika dan Nilai Islami*, vol. 1, no. 1, pp. 116–122, 2017.
- [12] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [13] M. Nuranti, P. Pamulang, M. N. Aini, and U. Enri, "Edumatic: Jurnal Pendidikan Informatika Komparasi Distance Measure Pada K-Medoids Clustering untuk Pengelompokan Penyakit Ispa," vol. 5, no. 1, pp. 99–107, 2021, doi: 10.29408/edumatic.v5i1.3359.
- [14] F. Tempola, M. Muhammad, and A. Mubarak, "PENGUNAAN INTERNET DIKALANGAN SISWA

SD DI KOTA TERNATE: SUATU SURVEY, PENERAPAN ALGORITMA CLUSTERING DAN VALIDASI DBI USE OF THE INTERNET IN THE ELEMENTARY SCHOOL STUDENTS IN TERNATE CITY: A SURVEY, IMPLEMENTED OF CLUSTERING ALGORITHM AND VALIDATION DBI,” vol. 7, no. 6, 2020, doi: 10.25126/jtiik.202072370.
