

PENERAPAN ALGORITMA *K-MEANS CLUSTERING ANALYSIS* PADA KASUS PENDERITA HIV/AIDS (STUDI KASUS KABUPATEN BANJAR)

Hayati Noor, Adani Dharmawati, Tri Wahyu Qur’ana

¹Fakultas Teknologi Informasi , Universitas Islam Kalimantan MAB Banjarmasin
adhienk.athie@gmail.com

²Fakultas Teknologi Informasi , Universitas Islam Kalimantan MAB Banjarmasin
adani.dharmawati@gmail.com

³Fakultas Teknologi Informasi , Universitas Islam Kalimantan MAB Banjarmasin
twqurana@gmail.com

Abstrak

Penggabungan data mining dengan kemampuan dalam mengelola dan mengolah database, statistika dan kecerdasan buatan telah banyak diterapkan dalam berbagai bidang. Penerapannya beragam, tergantung pada bagaimana data itu didistribusikan dan dimanfaatkan. Ada yang diterapkan di bidang kemiliteran, pendidikan, kesehatan, keuangan dan masih banyak lagi lainnya. Tujuan utama dari penelitian ini ialah untuk menganalisis jumlah kasus HIV/AIDS yang ada di Kabupaten Banjar dengan penyebaran di 20 Kecamatan didalamnya. Data yang dijadikan sumber berasal dari RSUD Ratu Zalecha Martapura. Analisis didukung dengan teknik clustering dengan pemilihan algoritma k-means dalam mengidentifikasi similaritas antar data. Jumlah kluster yang ditentukan dalam implementasi algoritma k-means adalah 3 kluster. Masing-masing kluster memiliki nilai rata-rata yang berbeda. Masing-masing kluster menunjukkan label tingkat kerawanan terjadinya HIV di tiap kecamatan yang berada di wilayah Kabupaten Banjar.

Kata Kunci : *Clustering, Data Mining, HIV/AIDS, Kesehatan, k-Means*

1. PENDAHULUAN

HIV/AIDS atau *human immunodeficiency virus / aquired immune deficiency syndrome* adalah suatu spektrum kondisi yang disebabkan oleh infeksi virus HIV. Diawali dengan adanya infeksi tersebut, seseorang yang terkena virus ini akan mengalami sakit seperti influenza. Hal tersebut biasanya diikuti dengan tanpa gejala yang cukup berkepanjangan. Indonesia merupakan salah satu negara yang memiliki potensi kenaikan jumlah kasus HIV/AIDS tiap tahunnya. Hal itu dapat dilihat dari data yang diperoleh dari tahun 1987-2014 untuk AIDS dan dari tahun 2005-2014 untuk HIV. Kenaikan signifikan terjadi di hampir seluruh provinsi yang ada di Indonesia. Bukan hanya itu, kasus penderita yang meninggal akibat AIDS juga menjadi faktor penentu kenaikan potensi tersebut. Apabila pemerintah tidak menanggulangi kejadian ini, dikhawatirkan ke depannya Indonesia akan menjadi salah satu negara menyumbang terbesar penderita HIV/AIDS.

Data mining merupakan suatu disiplin ilmu dalam ilmu komputer yang digunakan

untuk tujuan menggali informasi tersembunyi dari sekumpulan data set untuk kemudian digunakan dalam pengambilan keputusan maupun kebijakan yang berguna dalam bidang-bidang tertentu. Merujuk kepada kasus HIV/AIDS yang ada di Indonesia, dan dengan tersedianya data mentah yang dapat diolah, penelitian ini dapat dikembangkan untuk membantu pemerintah dalam melihat daerah atau provinsi dengan potensi HIV/AIDS terburuk. Oleh karena itu, dengan menggunakan metode clustering K-means, diperoleh gambaran dari setiap wilayah di Kabupaten Banjar untuk pengambilan kebijakan dan strategi berikutnya.

Dari latar belakang masalah diatas, dapat disimpulkan bahwa perlu adanya penelitian untuk menggali informasi tersembunyi dari sekumpulan data set untuk kemudian dapat digunakan dalam menganalisis jumlah kasus HIV/AIDS yang ada di Kabupaten Banjar serta penyebarannya di 20 kecamatan didalamnya.

Istilah data mining memiliki beberapa pandangan, seperti knowledge discover ataupun pattern recognition. Kedua istilah tersebut sebenarnya memiliki ketepatannya masing-

masing, istilah *knowledge discovery* atau penemuan pengetahuan tepat karena digunakan tujuan utama dari data mining memang untuk mendapat pengetahuan yang masih tersembunyi di dalam bongkahan data [1,2].

Istilah *pattern recognition* atau pengenalan pola pun tetap untuk digunakan karena pengetahuan yang hendak digali memang berbentuk pola-pola yang juga masih perlu digali dari dalam bongkahan data yang tengah dihadapi [2].

Banyak definisi bagi istilah ini dan belum ada yang dibakukan atau disepakati semua pihak. Namun demikian, istilah ini memiliki hakikat (*notion*) sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki, kegiatan inilah yang menjadi garapan atau perhatian utama dari disiplin ilmu data mining [3].

Clustering atau juga dikenal dengan istilah *cluster analysis* adalah salah satu metode atau teknik dalam data mining yang tepat digunakan untuk mengolah data yang tidak diketahui label atau kelasnya. Cara kerja teknik ini ialah mengelompokkan sekumpulan data ke dalam kelas-kelas atau kluster-kluster, yang mana objek-objek yang ada pada kelas tersebut memiliki similaritas yang tinggi jika dibandingkan dengan objek lain yang ada dalam kelas tersebut, namun memiliki similaritas yang rendah jika dibandingkan dengan objek yang ada di kelas/kluster lain (Han dan Kamber, 2006) [4].

Dalam sekelompok data, akan ada beberapa objek data yang ditempatkan ke dalam kluster yang sama berdasarkan sifat dan karakteristik data tersebut. Namun, dalam kasus tertentu, akan ada suatu objek data yang memiliki karakteristik dan sifat yang sangat berlainan dengan data lain, dan biasanya, data ini ditempatkan ke dalam kluster tersendiri. Jumlah anggota kluster yang menempati kluster itu biasanya sangat sedikit jika dibandingkan dengan rasio jumlah data yang ada. Data tersebut dikenal dengan istilah data anomali atau *outlier*. Salah satu keunggulan teknik clustering ialah dapat mendeteksi data *outlier* (*outlier detection*) karena tidak adanya label/kelas ketika data ini pertama kali diolah (Han dan Kamber, 2006) [4].

Algoritma *k-means* menggunakan masukan berupa parameter, jumlah *k*, dan sekumpulan data set dari sekian objek untuk

dimasukkan ke dalam kelas/kluster sehingga similaritas intrakluster semakin tinggi sedangkan similaritas antarkluster semakin rendah. Similaritas kluster diukur berdasarkan nilai rata-rata dari keseluruhan objek yang ada di kluster tersebut, yang bisa dipandang sebagai pusat kluster.

Cara kerja *k-means* adalah sebagai berikut. Pertama, algoritma ini akan memilih sejumlah *k* objek dari beberapa objek yang ada pada sekumpulan data. Masing-masing objek yang terpilih, merepresentasikan nilai rata-rata dari sejumlah *k* kluster. Kemudian objek yang tersisa, masing-masing akan di-assign ke dalam kluster yang sudah ditentukan berdasarkan similaritasnya dengan nilai rata-rata masing-masing kluster. Setelah itu, dilakukan perubahan nilai rata-rata kluster yang sudah menjadi beberapa anggota kluster. Ketiga langkah di atas dilakukan secara iteratif (berulang) hingga tidak ada lagi perubahan nilai rata-rata dan seluruh data sudah didistribusikan ke masing-masing kluster yang ada [4].

2. METODE PENELITIAN

Penelitian ini merupakan penelitian eksperimen yang mana penelitian ini akan melakukan pengujian dengan melakukan percobaan terhadap berbagai model algoritma *k-means clustering analysis* untuk penderita penyakit HIV/AIDS di daerah Kabupaten Banjar, Kalimantan Selatan. Pengukuran kinerja dilakukan dengan menghitung nilai error yang terjadi dari berbagai percobaan tersebut dan membandingkannya untuk mencari model yang memiliki nilai error paling kecil.

Model yang diusulkan yaitu algoritma *k-means clustering analysis* untuk menganalisis jumlah kasus HIV/AIDS yang ada di Kabupaten Banjar serta penyebarannya di 20 kecamatan di dalamnya.

Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam data set, termasuk variabel yang diperlukan dalam proses penelitian ini. Dengan metode eksperimen maka terdapat tahapan pada penelitian ini, yaitu:

a) Pengumpulan Data

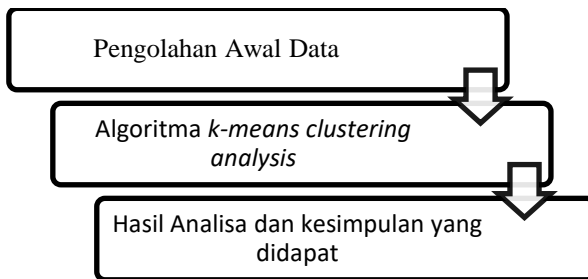
Data yang digunakan adalah data penderita penyakit HIV/AIDS di daerah Kabupaten Banjar yang diperoleh secara tidak langsung

dari Rumah Sakir Ratu Zalecha di Kota Martapura, Kabupaten Banjar.

b) Pengolahan Data

Data yang telah diperoleh tersebut akan diterapkan ke dalam beberapa model algoritma *k-means clustering analysis* untuk menganalisis jumlah kasus HIV/AIDS yang ada di Kabupaten Banjar serta penyebarannya di 20 kecamatan didalamnya. Model-model tersebut akan diimplementasikan dengan menggunakan RapidMiner 5.3.015.

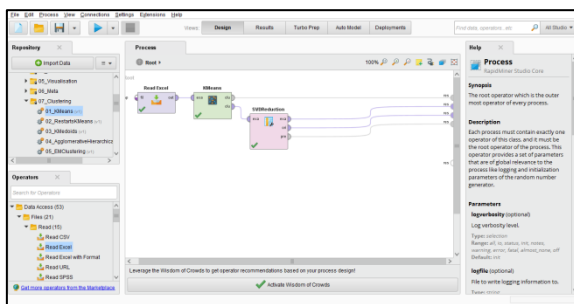
Model yang diusulkan yaitu algoritma *k-means clustering analysis* untuk menganalisis jumlah kasus HIV/AIDS yang ada di Kabupaten Banjar serta penyebarannya di 20 kecamatan didalamnya.



Gambar 2. 1 Model Yang Diusulkan

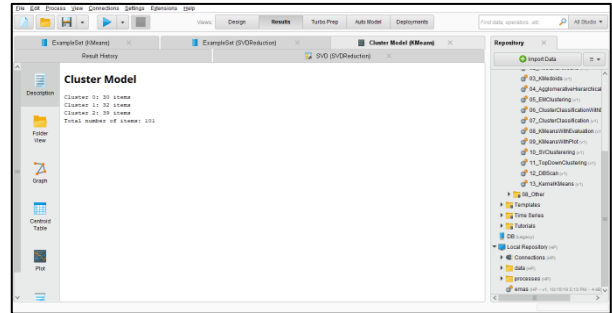
3. HASIL DAN PEMBAHASAN

Berikut merupakan model yang digunakan dalam penelitian ini yang menggunakan aplikasi Rapid Miner.

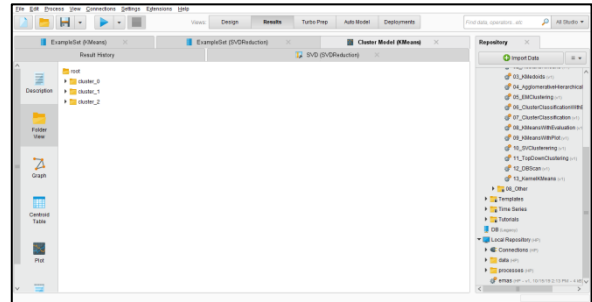


Gambar 3. 1 Model Penelitian

Berdasarkan dari model yang digunakan dan data yang ada, maka terbentuk terbentuk 3 cluster. Cluster pertama berisi 30 data, cluster kedua berisi 32 data, sedangkan cluster ketiga berisi 39 data.

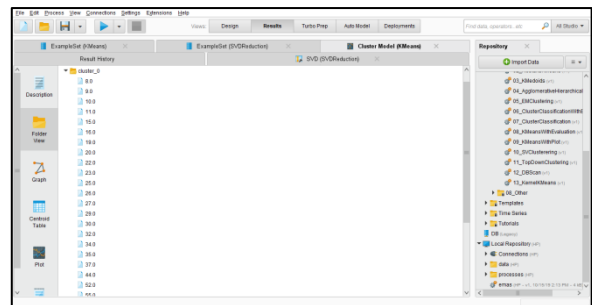


Gambar 3. 2 Hasil Cluster Yang Terbentuk

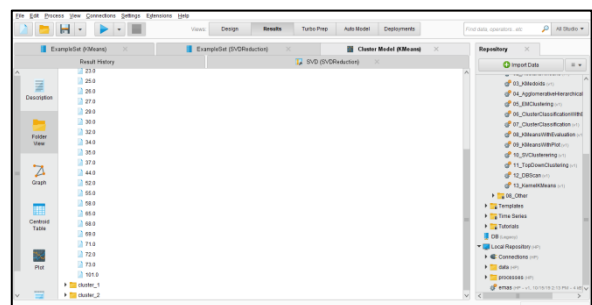


Gambar 3. 1 Data Hasil Cluster

Berikut merupakan rincian data dari cluster pertama, yaitu cluster_0.

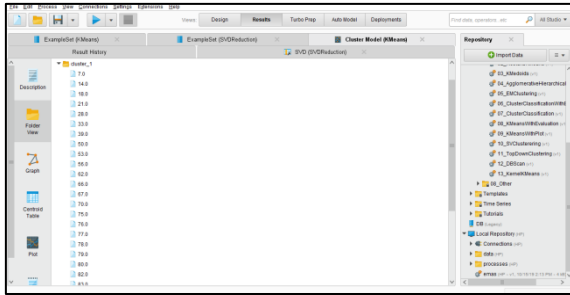


Gambar 3. 2 Rincian Data Cluster Pertama

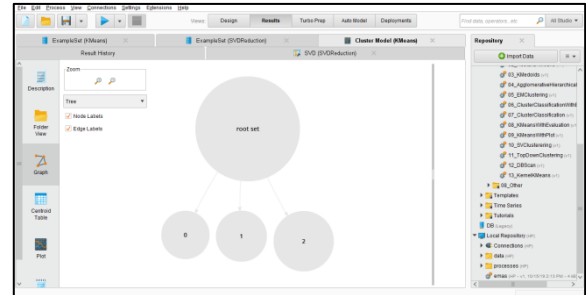


Gambar 3. 3 Rincian Data Cluster Pertama (Lanjutan)

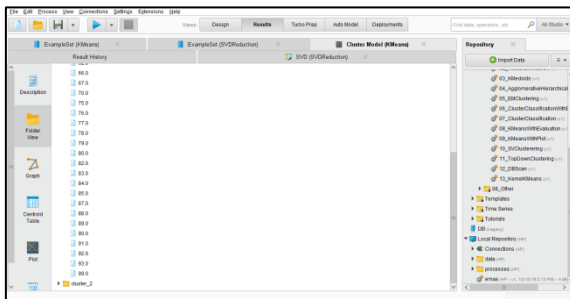
Berikut merupakan rincian data dari cluster kedua, yaitu cluster_1.



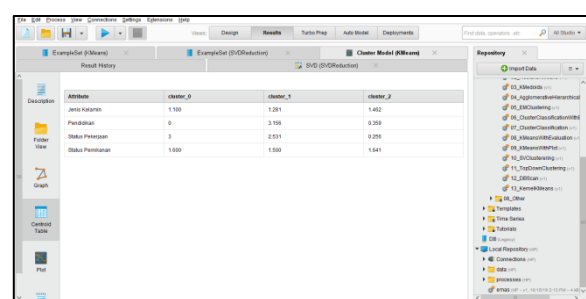
Gambar 3. 4 Rincian Data Cluster Kedua



Gambar 3. 8 Graph

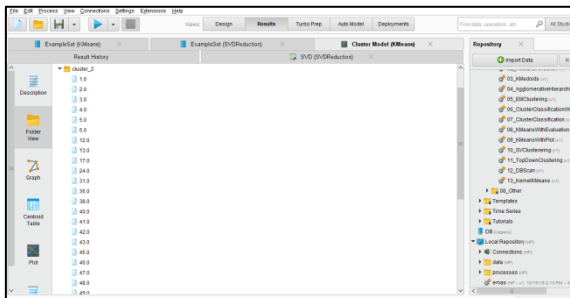


Gambar 3. 5 Rincian Data Cluster Kedua (Lanjutan)

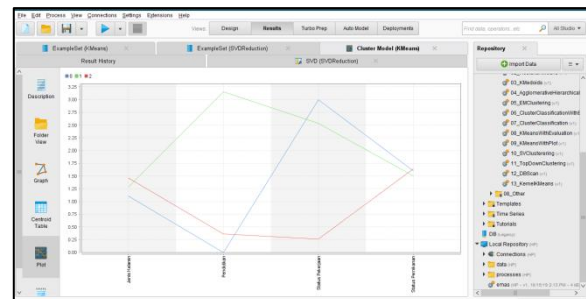


Gambar 3. 9 Centroid Table

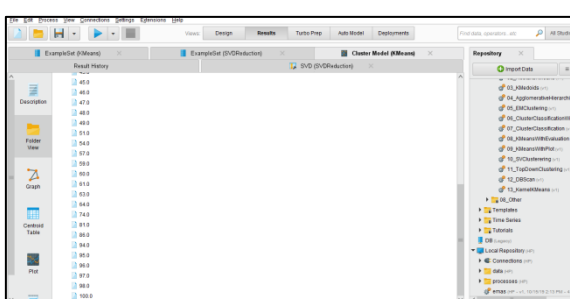
Berikut merupakan rincian data dari cluter ketiga, yaitu cluster_2.



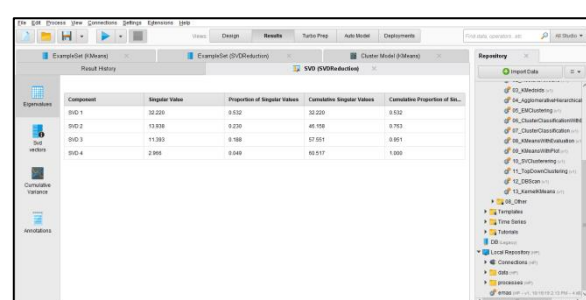
Gambar 3. 6 Rincian Data Cluster Ketiga



Gambar 3. 10 Plot



Gambar 3. 7 Rincian Data Cluster Ketiga (Lanjutan)



Gambar 3. 11 SVD Reduction

4. KESIMPULAN

Hasil dari penelitian ini yaitu dari seluruh data ODH yang ada, terbentuk 3 cluster dengan rincian sebagai berikut:

- a. Cluster pertama terdiri dari 30 data. 90% diantaranya berjenis kelamin laki-laki, dan 10% sisanya perempuan. Serta 53,33% dari cluster ini berstatus Kawin, 43,33% berstatus Belum Kawin, dan sisanya 3,33% berstatus Cerai. Dari segi pekerjaan, 80% merupakan karyawan swasta, 6,67% sebagai ibu rumah tangga.
- b. Cluster kedua terdiri dari 32 orang, yang didominasi oleh pria sebesar 71,88%. Dari segi pendidikan, terdapat 68,75% lulusan SMA sederajat. 15,63% lulusan SMP sederajat. Dari cluster ini terdapat 59,38 merupakan karyawan swasta dan 9,38 yaitu ibu rumah tangga.
- c. Cluster ketiga terdiri dari 39 orang, yang juga didominasi oleh pria sebanyak 52,85%. 46,15 diantaranya berstatus kawin, 43,59% bertatus belum kawin, sisanya berstatus cerai.

Berdasarkan hasil penelitian diatas, ditemukan bahwa ketiga cluster yang terbentuk memiliki kriteria masing-masing yang nantinya diharapkan akan menjadi fokus pihak terkait dalam upaya menekan angka ODHA di Kabupaten Banjar. Disamping itu diharapkan juga penelitian ini dapat dikembangkan lebih jauh misalnya untuk meneliti faktor-faktor apa saja yang mempengaruhi dan berkaitan erat dengan angka ODHA di Kabupaten Banjar.

5. REFERENSI

- [1] Berry, M.W. and Browne, M., 2006. Lecture notes in data mining. World Scientific.
- [2] Susanto, S. and Suryadi, D., 2010. Pengantar data mining: mengagali pengetahuan dari bongkahan data.
- [3] Luthfi, K. and Taufiq, E., 2009. Algoritma Data Mining. Yogyakarta: Andi.
- [4] Han, Jiawei; Kamber, Micheline, Data Mining: Concepts and Techniques 2nd Edition, (2006) Morgan Kaufmann Publishers, San Fransisco. Hal 38