

LITERATUR REVIEW: PENERAPAN KLASTERING DENGAN MENGGUNAKAN ALGORITMA K-MEANS DALAM BIDANG KESEHATAN

Giri Adi Nuryanto
UNIVERSITAS STIKUBANK SEMARANG

ABSTRAK

Pengelompokan objek (object clustering) adalah salah satu proses dari objek mining yang bertujuan untuk mempartisi objek yang ada kedalam satu atau lebih cluster objek berdasarkan karakteristiknya. Objek dengan karakteristik yang sama dikelompokkan dalam satu cluster dan objek dengan karakteristik berbeda dikelompokkan kedalam cluster yang lain. Algoritma K-Means Cluster Analysis termasuk dalam kelompok metode cluster analysis non hirarki, dimana jumlah kelompok yang akan dibentuk sudah terlebih dahulu diketahui atau ditetapkan jumlahnya. Algoritma K-Means Cluster Analysis menggunakan metode perhitungan jarak (distance) untuk mengukur tingkat kedekatan antara objek dengan titik tengah (centroid). Algoritma K-Means tidak terpengaruh terhadap urutan objek yang digunakan, hal ini dibuktikan ketika penulis mencoba menentukan secara acak titik awal pusat cluster dari salah satu objek pada permulaan perhitungan. Jumlah keanggotaan cluster yang dihasilkan berjumlah sama ketika menggunakan objek yang lain sebagai titik awal pusat cluster tersebut. Namun, hal ini hanya berpengaruh pada jumlah iterasi yang dilakukan. Algoritma K-Means Cluster Analysis pada dasarnya dapat diterapkan pada permasalahan dalam memahami perilaku konsumen, mengidentifikasi peluang produk baru dipasaran dan algoritma K-Means ini juga dapat digunakan untuk meringkas objek dari jumlah besar sehingga lebih memudahkan untuk mendiskripsikan sifat-sifat atau karakteristik dari masing-masing kelompok. Pada kesempatan kali ini penulis melakukan studi literatur yang berkaitan dengan kesehatan yang dianalisis berdasarkan teknik klustering menggunakan algoritma k-means.

Kata Kunci : Klastering, K-means, Algoritma Klastering, Klaster Analisis, Euclidian distance, kesehatan

ABSTRACT

Object grouping (object clustering) is one of the processes of object mining that aims to partition an existing object into one or more cluster objects based on its characteristics. Objects with similar characteristics are grouped in one cluster and objects with different characteristics are grouped into other clusters. The K-Means Cluster Analysis algorithm belongs to a group of non-hierarchical cluster analysis methods, in which the number of groups to be established is already known or determined in number. The K-Means Cluster Analysis algorithm uses the distance calculation method to measure the degree of closeness between the object and the midpoint (centroid). The K-Means algorithm is not affected by the order of objects used, this is proven when the author tries to randomly determine the starting point of the cluster center of one of the objects at the beginning of the calculation.

The number of cluster members generated amounts to the same when using other objects as the starting point of the cluster center. However, this only affects the number of iterations performed. K-Means Cluster Analysis algorithm basically can be applied to problems in understanding consumer behavior, identifying new product opportunities in the market and K-Means algorithm can also be used to summarize objects of large quantities making it easier to describe the properties or characteristics of each- each group. On this occasion the author conducted a study of health-related literature analyzed based on clustering techniques using the k-means algorithm.

Kata Kunci : Clustering, K-means, Clustering algorithms, Cluster Analysis, Euclidian distance, healthy

PENDAHULUAN

Analisis Cluster merupakan teknik mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis Cluster mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam cluster yang sama. Cluster-cluster yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi. Homogenitas internal yaitu kesamaan antar anggota dalam satu cluster. Heterogenitas eksternal yaitu perbedaan antara cluster yang satu dengan cluster yang lain. Analisis ini menggunakan set variabel yang ditentukan oleh peneliti itu sendiri. Fokus dari Analisis Cluster adalah membandingkan objek berdasarkan set variabel. Set variabel cluster adalah suatu set variabel yang merpresentasikan karakteristik yang dipakai objek-objek. Solusi Analisis Cluster bersifat tidak unik, anggota cluster untuk tiap penyelesaian/solusi tergantung pada beberapa elemen prosedur dan beberapa solusi yang berbeda dapat diperoleh dengan mengubah satu elemen atau lebih. Solusi cluster secara keseluruhan bergantung pada variabel-variabel yang digunakan sebagai dasar untuk menilai kesamaan. Penambahan atau pengurangan variabel-variabel yang relevan dapat mempengaruhi substansi hasil analisis cluster. Pada tulisan ini penulis menggunakan metode K-Means Cluster Analysis sebagai solusi untuk pengklasifikasian karakteristik dari objek. Alasan penggunaan algoritma K-Means diantaranya ialah karena algoritma ini memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Selain itu algoritma K-Means ini tidak terpengaruh terhadap urutan objek.

DASAR TEORI

Definisi Data

Dalam Webster's New World's Dictionary tertulis bahwa datum (bentuk tunggal data) merupakan suatu yang diketahui atau dianggap. Dengan demikian, data dapat memberi gambaran tentang suatu keadaan atau persoalan. Jadi, dapat disimpulkan bahwa data adalah sesuatu yang nyata diketahui atau dianggap yang dipakai untuk keperluan suatu analisa, diskusi, presentasi ilmiah atau tes statistic.

Data Kualitatif

Data kualitatif secara sederhana dapat disebut data yang bukan berupa angka. Ciri utama data kualitatif didapat dengan cara menghitung, sehingga tidak memiliki nilai desimal. Selain itu data kualitatif memiliki ciri-ciri tidak bisa dilakukan operasi matematika, seperti penambahan, pengurangan, perkalian dan pembagian. Contoh data kualitatif adalah data gender, data golongan darah, data tempat tinggal atau data jenis pekerjaan. Agar dapat dilakukan proses pada data kualitatif atau non metric, data tersebut harus diubah ke dalam bentuk angka, proses ini dinamakan kategorisasi.

Data Kuantitatif

Data kuantitatif dapat disebut sebagai data berupa angka dalam arti sebenarnya. Jadi, berbagai jenis operasi matematika dapat dilakukan pada data kuantitatif. Data kuantitatif merupakan data yang didapat dengan jalan mengukur sehingga bisa mempunyai nilai desimal. Contoh data kuantitatif adalah tinggi badan, usia, penjualan barang, dan sebagainya.

Jenis Data Menurut Sumbernya

Pembagian jenis data menurut sumbernya didasarkan pada sumber perolehan data tersebut, yaitu data internal dan data eksternal. Data internal adalah data yang dikumpulkan oleh suatu organisasi untuk menggambarkan keadaan atau kegiatan organisasi yang bersangkutan serta berguna untuk keperluan kegiatan harian. Misalnya, data penjualan, data produksi suatu perusahaan, data keuangan, data kepegawaian, data pendidikan dan lain sebagainya. Data eksternal adalah data yang dikumpulkan untuk menggambarkan suatu keadaan atau kegiatan di luar organisasi tersebut. Contoh dari data eksternal seperti data jumlah penduduk dan data pendapatan nasional yang didapat dari kantor pusat statistik setempat.

Jenis Data Menurut Cara Memperolehnya

Berdasarkan cara memperolehnya, data dapat dibedakan menjadi dua jenis, yaitu data primer dan data sekunder. Data primer adalah data yang dikumpulkan dan diolah sendiri oleh organisasi atau perorangan langsung dari objeknya. Misalnya suatu perusahaan ingin mengetahui konsumsi rata-rata suatu produk terhadap penduduk di suatu daerah dengan cara melakukan wawancara langsung kepada penduduk setempat. Data sekunder adalah data yang diperoleh dalam bentuk jadi dan telah diolah oleh pihak lainnya. Biasanya data sekunder ini dalam bentuk publikasi.

Data Mining

Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan. Data mining merupakan metode pengolahan data berskala besar.

Clustering

Pada dasarnya clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu hierarchical clustering dan non-hierarchical clustering.

Hierarchical clustering adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip.

METODE

Metode yang digunakan dalam penulisan artikel ini adalah studi literatur. Studi literatur merupakan langkah awal dalam penelitian ini, studi literatur dilakukan untuk melengkapi pengetahuan dasar dan teori-teori yang berasal dari buku-buku, jurnal, maupun media internet.

Literature Review

Ada beberapa referensi yang diambil penulis sebagai bahan pertimbangan untuk penelitian yang dilakukan, referensi tersebut diambil dari beberapa jurnal yang membahas permasalahan dengan topik kesehatan, antara lain :

1. Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means
2. Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai
3. Implementasi Algoritma K-means untuk Clustering penyakit HIV/AIDS di Kabupaten Semarang
4. Clustering menggunakan metode k-means untuk menentukan status gizi balita
5. Penerapan algoritma k-means untuk menentukan tingkat kesehatan bayi dan balita pada kabupaten dan kota di Jawa Tengah
6. Penerapan algoritma k-means clustering analysis pada penyakit menular manusia (studi kasus kabupaten majalengka)

Review Jurnal Ke-1

Judul	Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means
Jurnal	Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta
Tahun	2014
Penulis	Nielza Atthina dan Lizda Iswari
Tujuan penelitian	Menentukan Rentang Derajat Kesehatan Penduduk pada suatu Daerah
Metode penelitian	Klustering menggunakan algoritma K-means
Objek penelitian	angka kelahiran kasar, angka kematian kasar, angka kematian bayi, angka kematian balita, dan angka kematian ibu.
Hasil penelitian	Setelah data selesai diklaster, Pakar dapat memasukkan analisis terkait hasil klasterisasi data. Untuk memasukkan hasil analisis, Pakar harus memilih tahun dan klaster tertentu sehingga akan tampil data indikator kesehatan sesuai kriteria yang dimasukkan.
Kesimpulan	klasterisasi data kesehatan penduduk untuk menentukan rentang derajat kesehatan daerah dengan metode k-Means ini dapat disimpulkan k-Means dalam mengelompokkan data berdasarkan metode minimum maksimum untuk normalisasi data dan metode Euclidean Distance untuk menghitung jarak antar data

--	--

Review Jurnal Ke-2

Judul	Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai
Jurnal	Jurnal Teknologi Informasi & Komunikasi Digital Zone
Tahun	2016
Penulis	Taslim Fajrizal
Tujuan penelitian	Distribusi obat pada instansi layanan kesehatan dapat dikelompokkan sesuai kebutuhan berdasarkan data distribusi obat pada tahun tahun sebelumnya dan dapat digunakan sebagai acuan perencanaan obat untuk tahun berikutnya.
Metode penelitian	Klustering menggunakan algoritma K-means
Objek penelitian	Puskesmas rumbai
Hasil penelitian	Terdapat 133 jenis obat pada apotik puskesmas rumbai bukit, setelah dilakukan analisa kluster dapat disimpulkan bahwa banyak diantara obat yang ada masuk dalam kelompok cluster kurang.
Kesimpulan	Dari hasil analisa cluster perlu dilakukan lagi penelitian lanjutan agar clusterisasi data obat dapat dilakukan secara lebih valid dengan menetapkan nilai centroid terbaik.

Review Jurnal Ke-3

Judul	Implementasi Algoritma K-means untuk Clustering penyakit HIV/AIDS di Kabupaten Semarang
Jurnal	Universitas Dian Nuswantoro Semarang
Tahun	2017
Penulis	Ardila Rifani dan Desi Purwanti
Tujuan penelitian	Mengetahui penyebaran HIV/AIDS di kabupaten Semarang
Metode penelitian	Klustering menggunakan algoritma K-means
Objek penelitian	Penderita HIV/AIDS di kabupaten Semarang
Hasil penelitian	hasil dari analisis , kecamatan-kecamatan pada Kabupaten Semarang di bagi menjadi 3 cluster.
Kesimpulan	Dari penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma k-means dapat diterima untuk pengklasteran data tersebut.

Review Jurnal Ke-4

Judul	Clustering menggunakan metode k-means untuk menentukan status gizi balita
Jurnal	Jurnal Informatika, Vol. 15, No.2, Bulan Desember 2015
Tahun	2015

Penulis	Windha Mega Pradnya Duhita
Tujuan penelitian	Menentukan kondisi gizi balita dengan metode k-means
Metode penelitian	Klustering menggunakan algoritma K-means
Objek penelitian	50 balita di desa Karang Songo kedalam 5 cluster status gizi
Hasil penelitian	Kebenaran algoritma K-Means dalam mengelompokkan status gizi balita memiliki nilai akurasi sebesar 34%
Kesimpulan	Algoritma K-Means hanya memiliki nilai akurasi 34% benar. Namun nilai ini bisa berubah seiring dengan penambahan data uji.

Review Jurnal Ke-5

Judul	Penerapan algoritma k-means untuk menentukan tingkat kesehatan bayi dan balita pada kabupaten dan kota di Jawa Tengah
Jurnal	Universitas Dian Nuswantoro Semarang
Tahun	2012
Penulis	Erga Aprina Sari
Tujuan penelitian	Menentukan tingkat kesehatan bayi dan balita pada kabupaten dan kota di Jawa Tengah dengan klustering menggunakan algoritma k-means
Metode penelitian	Klustering menggunakan algoritma K-means
Objek penelitian	Bayi dan balita pada kabupaten dan kota di Jawa Tengah
Hasil penelitian	Kabupaten/kota memiliki tingkat kesehatan pada indikator buruk karena jumlah penderitanya banyak.
Kesimpulan	Analisa dilakukan pada tiap indikator agar penanganan dari masalah kesehatan tersebut menjadi lebih terfokus dan tepat sasaran.

Review Jurnal Ke-6

Judul	Penerapan algoritma k-means clustering analysis pada penyakit menular manusia (studi kasus kabupaten majalengka)
Jurnal	Jurnal Sistem Informasi (Journal of Information System), Volume 14, Issue 1, April 2018
Tahun	2018
Penulis	Ade Bastian, Harun Sujadi, dan Gigin Febrianto
Tujuan penelitian	Melakukan uji clustering analysis pada penyakit menular manusia
Metode penelitian	Klustering menggunakan algoritma K-means
Objek penelitian	Pasien dari Puskesmas Cigasong dan Talaga, Majalengka
Hasil penelitian	Hasil dari tiap cluster, penderita penyakit menular yang berasal dari Puskesmas Cigasong dan Talaga, terlihat bahwa penyakit menular didominasi oleh penyakit ISPA dan Diare.
Kesimpulan	Algoritma K-means mampu mengelompokkan objek besar dengan sangat cepat sehingga mempercepat proses pengelompokan

KESIMPULAN

K-Means Clustering merupakan metode klasterisasi berdasarkan persamaan karakteristik, dan merupakan metode yang sangat berguna karena mampu mentranslasi ukuran persamaan yang intuitif menjadi ukuran yang kuantitatif.

SARAN

Untuk menampilkan hasil yang maksimal diperlukan lebih banyak data. Semakin banyak data yang dijadikan sampel akan semakin baik hasil yang diperoleh.

DAFTAR PUSTAKA

Atthina, Nielza & Iswari, Lizda. 2014. Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means. Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta

Fajrizal, Taslim. 2016. Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai. Jurnal Teknologi Informasi & Komunikasi Digital Zone, Volume 7, Nomor 2, November 2016

Rifani, Ardila & Purwanti, Desi. 2017. Implementasi Algoritma K-means untuk Clustering penyakit HIV/AIDS di Kabupaten Semarang. Universitas Dian Nuswantoro Semarang

Mega, Windha P.D. 2015. Clustering menggunakan metode k-means untuk menentukan status gizi balita. Jurnal Informatika, Vol. 15, No.2, Bulan Desember 2015

Aprina Sari, Erga. 2012. Penerapan algoritma k-means untuk menentukan tingkat kesehatan bayi dan balita pada kabupaten dan kota di Jawa Tengah. Universitas Dian Nuswantoro Semarang

Bastian, Ade dkk. 2018. Penerapan algoritma k-means clustering analysis pada penyakit menular manusia (studi kasus kabupaten majalengka). Jurnal Sistem Informasi (Journal of Information System), Volume 14, Issue 1, April 2018