

CLUSTERING DATA PENDUDUK BERDASARKAN KELOMPOK UMUR PER DESA DI KABUPATEN PURWOREJO MENGGUNAKAN METODE K- MEANS

Ricky Fauzan Afrizal, Martanto, Agus Bahtiar

Teknik Informatika, STMIK IKMI Cirebon

fauzan46.aja@gmail.com

ABSTRAK

Pengolahan data penduduk adalah tanggung jawab pemerintah kabupaten, Pelaksanaannya dari desa. Pada pelayanan tersebut harus dilakukan dengan benar juga cepat supaya masyarakat merasa mendapatkan pelayanan yang memuaskan. pada pembuatan surat untuk keterangan sampai dengan laporan kependudukan di Kelurahan merdeka pada pembuatannya telah menggunakan komputer, sehingga data yang diolah belum terintegrasi kedalam satu kawasan penyimpanan. dari kegiatannya ini menimbulkan permasalahan seperti tidak efisiennya waktu yang digunakan untuk memproses pelayanan kependudukan dalam melayani kebutuhan-kebutuhan masyarakat, Variabel penelitian merupakan salah satu atribut atau nilai dari seseorang, berdasarkan data penduduk kabupaten purworejo berdasarkan umur per desa pada cluster_0 dari pria umur 45-49 berjumlah 48.849, wanita umur 45-49 berjumlah 50.619, jumlahnya 99.468, persentase 0.066, pria umur 50-54 berjumlah 49.705, wanita umur 50-54 berjumlah 53.295, jumlahnya 103.000, persentase 0.069, pria umur 55-59 berjumlah 49.239, wanita umur 55-59 berjumlah 52.782, jumlahnya 102.022, persentase 0.068. dengan jumlah 464 items, dan Cluster 1 dengan jumlah 3 items, jumlah total cluster adalah 467 items. Metode K-means mempunyai kelebihan mudah untuk dipahami dan diimplementasikan.

Kata kunci: Data Penduduk, Data mining, K - means, Rapid Minner

1. PENDAHULUAN

Penelitian ini menggunakan data mining. Di era digital berkembangnya data menjadi big data akan dibutuhkannya suatu metode guna mengolah big data supaya lebih praktis bagi masyarakat, maka akan dibutuhkan suatu metode berikut ini merupakan data mining. Data tersebut merupakan metode yang digunakan untuk mengotomatisasi prosesing data untuk mencari pola dari sekumpulan data dengan ukuran besar.

Data mining dipakai guna mengolah big data pada basis data sehingga menuai informasi yang berguna untuk strategi bisnis. Diantaranya data mining terpisah menjadi 2 bagian, bagian pertama yaitu Deskriptive Mining, yang kedua yaitu Predictive Mining, pada descriptive mining gunanya untuk menentukan karakteristik data, dan pada predictive mining gunanya untuk mencari pola data.

Penelitian terdahulu akan berfokus pada metode k-means bisa digunakan untuk mengelompokkan daerah potensi. Penelitian ini akan berfokus pada teknik clustering yang akan diterapkan pada data penduduk per-desa berdasarkan umur yang nantinya akan mencari clustering dengan nilai terbaik. Metode K-Means akan menjadi pilihan yang tepat karena metode ini terbilang cukup mudah untuk dipahami dan diimplementasikan selain itu proses pembelajaran membutuhkan waktu yang relatif cepat sehingga mempermudah proses penyelesaian, dan juga metode K-Means sudah sering digunakan sebagai teknik clustering. *State of the art* yang diperoleh dari penelitian terdahulu juga harus dicantumkan untuk menunjukkan kontribusi serta bedanya penelitian yang dilakukan sekarang dengan penelitian sebelumnya.

Pada penelitian ini penulis ingin Clustering pada data penduduk menggunakan algoritma K-Means. Metode Clustering ini mengacu pada optimizing threshold, dan parallel threshold. Sequential threshold metode ini melakukan pengelompokan terlebih dahulu untuk dipilih satu objek dasar yang nantinya akan dijadikan nilai utama cluster, Kemudian semua cluster yang ada dalam jarak terdekat dengan cluster ini kemudian akan dilakukan penggabungan, dan seterusnya sampai cluster ini terbentuk beberapa cluster dengan keseluruhan objek yang terdapat didalamnya. Algoritma K-means antara lain karena Algoritma tersebut mempunyai ketelitian terhadap ukuran setiap objek, Sehingga algoritma ini akan relatif efisien untuk pengelompokan objek yang berukuran besar.

Penelitian ini akan mencoba untuk mengclustering data penduduk per desa dikabupaten purworejo menggunakan algoritma K-means. Algoritma K-Means memakai clustering yang basisnya titik centroid yang dimana dapat menentukan tiga parameternya yaitu jumlah total cluster, inisialisasi cluster, dan jarak sistem. Yang diurutkan dengan objek penelitian dibantu dengan Rapidminner ataupun bisa dengan cara manual, Data yang mempunyai representasi persamaan nilai dalam data yang mempunyai perbandingan kelompok lainnya dan data satu kelompok sehingga memungkinkan pengelompokan data yang berbeda pasti memiliki tahapan variasi kecil.

2. TINJAUAN PUSTAKA

Penelitian ini Menggunakan Metode K-Means. Penelitian tersebut dilakukan oleh Indriyani, F.,

Irfiani, E. Penelitian ini akan membahas tentang Objek penelitian yaitu Genta Corp yang berjualan jenis- jenis peralatan outdoor yang berketempatan di kota Bogor. Pengambilan datanya diambil dari data persediaan barang selama tiga bulan beserta items data barang sebanyak tiga puluhan, Jenis datanya kuantitatif yang cara penghitungannya dilakukan secara langsung sebagai variabel angka atau bilangan. [1]

Penelitian selanjutnya yang berjudul “Kajian data minning profil siswa baru dalam penentuan strategi promosi dengan metode two step clustering”. Penelitian tersebut dilakukan oleh. Iis ismawati , Jurnal Ilmiah Teknologi Informasi Terapan 15 Agustus 2019. [2]

Penelitian selanjutnya dilakukan oleh. Ari Sulistiyawati, Eko Supriyanto. Penelitian ini bertujuan untuk tentang sistem informasi berbasis web digunakan untuk mengelola data yang akurat dalam laporan hasil pembelajaran dan pengelompokan siswa dengan mengimplementasikan Clustering. Penerapan dalam algoritma K-Means dengan berpartisipasi data ke dalam bentuk satu kelompok atau lebih, dimana data ini memiliki representatif kemiripan dengan satu kelompok, dan data yang memiliki perbedaan kelompok yang lain. [3]

3. METODE PENELITIAN

Dengan menggunakan metode eksperimen yang dimana penelitian ini melakukan uji coba terhadap variabel dataset yang akan dilakukan penelitian guna menerima informasi yang diinginkan. Metode kuantitatif ini diklaim karena menyajikan sebuah informasi penelitian secara numerik serta dengan memakai grafik pada analisisnya. menggunakan melakukan penelitian kuantitatif rata-rata di sampel yg diambil random, konklusi penelitian dapat digeneralisasikan ke populasi sampel, secara kuantitatif dengan metode algoritma K-Means.

3.1. Metode pengumpulan data

Penelitian ini akan mencoba untuk mengclustering data penduduk per desa di kabupaten purworejo menggunakan algoritma K-means. Memakai metode clustering titik centroid dengan menentukan 3 parameternya ialah jumlah cluster, jarak sistem dan cluster.

Salah satunya adalah untuk meminimalisasi guna dari objek pada dasarnya minimalisasi jenis dalam suatu kelompok dan maksimalkannya.

3.2. Tahapan penelitian

Dalam tahapan penelitian ini menggunakan pendekatan data science dengan tahapan Knowledge Discovery in Database (KDD) dan menggunakan algoritma Random Forest dalam memprediksi data stunting. Data mining atau (KDD) suatu proses penemuan pengetahuan ini diterapkan pada database dalam jumlah yang besar.

Berikut tahapan yang akan dilakukan menggunakan metode KDD (knowledge discovery in database), yaitu sebagai berikut:

1. Data Selection

Selection adalah salah satu proses dimana pengambilan informasi/ pengumpulan informasi, pemilihan informasi serta pengecekan informasi. Pada tahap ini, data digunakan sebagai bahan penelitian adalah 494 data dari masing-masing kecamatan yang terdiri dari data sebanyak 25,820 Pria, dan 27,836 Wanita. Jumlah keseluruhan mencapai 53, 565 Jiwa berdasarkan umur 45 sampai umur 59. Lalu dipilih beberapa atribut yang akan digunakan pada saat pengclustering di RapidMiner.

2. Data Preprocessing

Preprocessing adalah sebagian proses persiapan informasi pada saat sebelum dikerjakannya proses data mining. Data yang dilakukan preprocessing umumnya dicoba dengan sebagian sesi serupa cleaning, reduction, integration. Saat sebelum proses clustering, diperlukan proses preprocessing terlebih dahulu dengan hapus data yang tidak perlu, data yang hilang atau data missing.

3. Data minning

Data Minning adalah proses pengolahan data algoritma yang konsisten dengan alur data mining. Algoritma yang digunakan dalam penelitian ini ialah K-means yang termasuk dalam clustering.

4. Evaluasi

Pada tahap evaluasi digunakan untuk mengevaluasi hasil dari menggunakan algoritma K-means untuk mengetahui kesimpulan dari proses data mining. Nilai performance yang digunakan adalah Davies Bouldin Index.

5. Knowledge

Tahapan Knowledge ini menampilkan data yang dihasilkan dari proses clustering, proses ini memaparkan hasil sehingga menjadi data yang mudah dipahami.

4. HASIL DAN PEMBAHASAN

Penelitian ini yaitu diperolehnya cluster-cluster dengan sifat-sifat yang memiliki kesamaan sangat tinggi berdasarkan parameter-parameter status sosial di masyarakat. Data yang diperoleh dari penelitian ini yaitu jumlah penduduk di kabupaten purworejo berdasarkan umur perdesa sebanyak 469 data. Kemudian data tersebut dikelompokkan menjadi dua cluster, pengelompokan ini menggunakan metode K-Means dan diimplementasikan dengan menggunakan aplikasi RapidMiner.

4.1. Data selection

Pada bagian ini yaitu proses KDD yaitu data selection. Pada tahapan ini akan dilakukan seleksi data yang akan digunakan dalam proses pengelompokan.

Tahapan ini bertujuan untuk diprediksi melalui tools RapidMiner, tahapan seleksi data dilakukan di Microsoft Excel. Pada tahap ini, data digunakan sebagai bahan penelitian adalah 494 data dari masing-

masing kecamatan yang terdiri dari data sebanyak 25,820 Pria, dan 27,836 Wanita. Jumlah keseluruhan mencapai 53,565 Jiwa berdasarkan umur 45 sampai umur 59.

Tabel 1. data selection

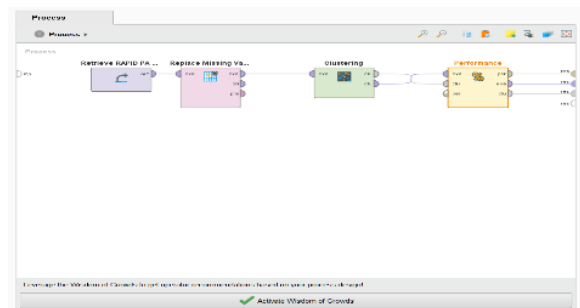
pria umur 45-49	27.19495	393.5	185.3333	113.52	65.10606
wanita umur 45-49	28.15162	409.5	196.3333	117.96	67.31818
Jumlah	55.34657	803	381.6667	231.48	132.4242
Persentase	0	0	0	0	0
pria umur 50-54	27.93141	371.5	180.6667	113.28	66.82576
wanita umur 50-54	29.79061	424	216.6667	122.12	70.96212
jumlah2	57.72202	795.5	397.3333	235.4	137.7879
persentase2	0	0	0	0	0
pria umur 55-59	27.57762	363	190.1667	113	65.70455
wanita umur 55-59	29.83032	403.5	205.5	120.7	70.15152
jumlah3	57.40794	766.5	395.6667	233.7	135.8561
persentase3	0	0	0	0	0

4.2. Data mining

Dalam tahap proses data mining dilakukan menggunakan tools Rapidminer menggunakan algoritma K-means. Tahapan ini digunakan untuk mengetahui result dari nilai terbaik dipformance yang dapat dilihat pada gambar 1.

4.3. Data Preprocessing

Dalam tahap ini data dilakukan proses Preprocessing yang dimana akan dilakukan pembersihan atau menghilangkan data yang missing atau kosong sebelum ke tahap berikutnya, karena data yang peneliti gunakan tidak terdapat data yang missing atau kosong maka tidak dilakukan penghapusan data yang kosong, untuk tabel centroid dapat pada tabel 2



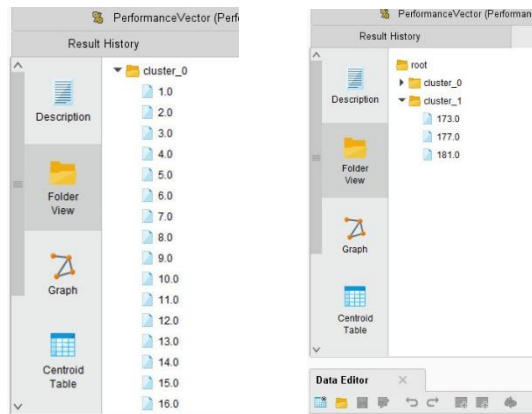
Gambar 1. data mining

Tabel 2. data preprocessing

pria umur 45-49	27.19495	393.5	185.3333	113.52	65.10606
wanita umur 45-49	28.15162	409.5	196.3333	117.96	67.31818
Jumlah	55.34657	803	381.6667	231.48	132.4242
Persentase	0	0	0	0	0
pria umur 50-54	27.93141	371.5	180.6667	113.28	66.82576
wanita umur 50-54	29.79061	424	216.6667	122.12	70.96212
jumlah2	57.72202	795.5	397.3333	235.4	137.7879
persentase2	0	0	0	0	0
pria umur 55-59	27.57762	363	190.1667	113	65.70455
wanita umur 55-59	29.83032	403.5	205.5	120.7	70.15152
jumlah3	57.40794	766.5	395.6667	233.7	135.8561
persentase3	0	0	0	0	0

4.4. Evaluasi

Evaluasi hasil clustering yang dilakukan menggunakan RapidMiner yaitu menjadi 5 kelompok dan masing-masing memiliki nilai, berikut gambar yang bisa ditampilkan pada gambar berikut



Gambar 2. Folder cluster

Dapat dilihat pada Gambar 4.8 terdapat dua cluster yang masing-masing didalamnya mempunyai items sebagai berikut :

Cluster_0 : 464 items

Cluster_1 : 3 items

Jumlah total cluster : 467 items

Pada tahap evaluasi yaitu membahas kelompok cluster, pada gambar 4.8 yaitu 2 kelompok cluster, Cluster 0 dengan jumlah 464 items, dan Cluster 1 dengan jumlah 3 items, jumlah total cluster adalah 467 items.

4.5. Hasil Clustering

Berikut ini hasil percobaan dari 2 cluster sampai 5 cluster:

Tabel 3. Hasil Clustering 2 Sampai 5 Cluster

K-Means	AVG	DBI	Cluster
2	1839.530	0.023	Cluster 0: 464 items Cluster 1: 3 items
3	676.106	0.043	Cluster 0: 343 items Cluster 1: 121 items Cluster 2: 3 items
4	359.259	0.044	Cluster 0: 142 items Cluster 1: 3 items Cluster 2: 44 items Cluster 3: 278 items
5	288.078	0.046	Cluster 0: 277 items Cluster 1: 2 items Cluster 2: 50 items Cluster 3: 132 items Cluster 4: 6 items

dapat dilihat bahwa jumlah cluster yang paling mendekati nilai 0 adalah 2 cluster dengan nilai Davies Bouldin index 0.023, dan nilai Avg. within centroid distance 1839.530. nilai Davies Bouldin Index (DBI) nilai 0 menandakan semakin baik cluster yang diperoleh. Makin kecil nilai DBI artinya hasil cluster yang optimal.

4.6. Analisa Hasil Clustering

Pada pembahasan hasil clustering ini, didapatkan hasil dari pengujian dataset penduduk yang dilakukan

menggunakan tools rapidminer yang dikelompokkan menjadi 2 cluster. Berikut merupakan hasil dari data yang sudah dikelompokkan :

Tabel 4. Centroid Cluster 0

Atribut	Cluster_0
Pria umur 45-49	48.849
Wanita umur 45-49	50.619
Jumlah	99.468
Persentase	0.066
Pria umur 50-54	49.705
Wanita umur 50-54	53.295
Jumlah2	103
Persentase2	0.069
Pria umur 55-59	49.239
Wanita umur 55-59	52.782
Jumlah3	102.022
Persentase3	0.068

Pada Tabel Data penduduk kabupaten purworejo berdasarkan umur per desa pada cluster_0 dari pria umur 45-49 berjumlah 48.849, wanita umur 45-49 berjumlah 50.619, jumlahnya 99.468, persentase 0.066, pria umur 50-54 berjumlah 49.705, wanita umur 50-54 berjumlah 53.295, jumlahnya 103.000, persentase 0.069, pria umur 55-59 berjumlah 49.239, wanita umur 55-59 berjumlah 52.782, jumlahnya 102.022, persentase 0.068.

Tabel 5. Centroid Cluster 1

Atribut	Cluster_1
Pria umur 45-49	345.333
Wanita umur 45-49	364.000
Jumlah	709.333
Persentase	0.071
Pria umur 50-54	328.667
Wanita umur 50-54	381.333
Jumlah2	710.000
Persentase2	0.072
Pria umur 55-59	327.333
Wanita umur 55-59	369.000
Jumlah3	696.333
Persentase3	0.071

Pada Tabel Data penduduk kabupaten purworejo berdasarkan umur per desa pada cluster_1 dari pria umur 45-49 berjumlah 345.333, wanita umur 45-49 berjumlah 364, jumlahnya 709.333, persentase 0.071, pria umur 50-54 berjumlah 328.667, wanita umur 50-54 berjumlah 381.333, jumlahnya 710.000, persentase 0.072, pria umur 55-59 berjumlah 327.333, wanita umur 55-59 berjumlah 369.000, jumlahnya 696.333, persentase 0.071.

5. KESIMPULAN

Menurut hasil dan Pembahasan maka dapat disimpulkan se bagai berikut: Hasil cluster terbaik darin pengelompokan Penduduk berdasarkan umur yaitu memiliki dua cluster dengan nilai Davies Bouildin Index -0.278. Jumlah anggota cluster dengan nilai anggota yang sama, Cluster 0 mempunyai 464

items, dan Cluster 1 mempunyai 3 items. kelompok cluster yaitu 2 kelompok cluster, Cluster 0 dengan jumlah 464 items, dan Cluster 1 dengan jumlah 3 items, jumlah total cluster adalah 467 items. Semakin banyak data penduduk yang diinput maka semakin banyak cluster centroid yang terbentuk diproses metode k-means. Pengclusteran dengan menggunakan metode K-Means didapatkan untuk data penduduk.

DAFTAR PUSTAKA

- [1] K. Fatmawati and A. P. Windarto, "DATA MINING: PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDARAH DENGUE (DBD) BERDASARKAN PROVINSI," 2018. [Online]. Available: <https://www.depkes.go.id/>.
- [2] M. Hariyanto and R. Tahara Shita, "17 Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance," 2018.
- [3] S. Budi, "Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means Text Mining For Movie Review Sentiment Analysis Using K-Means Algorithm." [Online]. Available: <http://www.cs.cornell.edu/People/pabo/movie-review-data/>.
- [4] M. R. Muttaqin and M. Defriani, "Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 121–129, Aug. 2020, doi: 10.33096/ilkom.v12i2.542.121-129.
- [5] R. R. Putra and C. Wadisman, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 1, no. 1, pp. 72–77, Mar. 2018, doi: 10.31539/intecom.v1i1.141.
- [6] A. Bastian, H. Sujadi, and G. Febrianto, "PENERAPAN ALGORITMA K-MEANS CLUSTERING ANALYSIS PADA PENYAKIT MENULAR MANUSIA (STUDI KASUS KABUPATEN MAJALENGKA)," 2018.
- [7] C. Selvi, D. Sembiring, L. Hanum, and S. Parsaoran Tamba, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENENTUKAN JUDUL SKRIPSI DAN JURNAL PENELITIAN (STUDI KASUS FTIK UNPRI)," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022.
- [8] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan*, vol. 15, no. 2, 2020.
- [9] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [10] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 2, p. 437, Apr. 2020, doi: 10.30865/mib.v4i2.2080.