

Analisis Rumah Tidak Layak Huni Menggunakan Algoritma X-Means

Analyzing Uninhabitable Houses Using X-Means Algorithm

Ghina Fitria Rohendi^{*1}, Nana Suarna², Gifthera Dwi Lestari³

¹²Teknik Informatika, ³Rekayasa Perangkat Lunak, ¹²³STMIK IKMI Cirebon

*e-mail: ghinafitrirohendi@gmail.com¹, st_nana@yahoo.com², gifthera.ikmi@gmail.com³

Received:	Revised:	Accepted:	Available online:
14.12.2022	03.04.2023	15.04.2023	30.04.2023

Abstrak: Tempat tinggal atau rumah merupakan kebutuhan masyarakat yang sangat mendasar bagi setiap orang selain sandang dan pangan. Rumah yang layak huni dapat diketahui dari struktur bangunan rumah dan fasilitas rumah yang ada. Struktur bangunan rumah meliputi luas lantai, jenis lantai, jenis atap, dan jenis dinding. Sedangkan fasilitas rumah meliputi sumber air minum, sumber penerangan, dan tempat pembuangan akhir kotoran. Dalam penelitian ini penulis akan melakukan pengelompokan rumah tidak layak huni dengan melihat struktur bangunan rumah dan fasilitas rumah yang dimiliki setiap rumah tangga di 4 wilayah yaitu: Batu layang, siantan hilir, siantan tengah dan siantan hulu. Salah satu metode yang dapat digunakan untuk pengelompokan rumah tidak layak huni adalah metode X-means clustering. karena merupakan salah satu penyempurnaan dari metode K-Means clustering. Adapun tahapan metode x-means terdiri dari dua langkah yang diulang sampai selesai. yaitu, 1) Tingkatkan-Params, pada langkah ini menerapkan algoritma k-means pada awalnya untuk k cluster hingga konvergensi. Dimana k sama dengan batas bawah yang disediakan oleh pengguna. 2) Perbaiki Struktur, langkah perbaikan struktur ini dimulai dengan memecah setiap pusat cluster menjadi dua anak dalam arah yang berlawanan di sepanjang vektor yang dipilih secara acak. Setelah itu menjalankan k-means secara lokal di dalam setiap cluster untuk dua cluster. Keputusan masing-masing pusat cluster sendiri dengan membandingkan nilai-nilai Bayesian Information Criterion (BIC). 3) Jika $K \geq k_{max}$ (batas atas) berhenti dan laporkan ke model penilaian terbaik yang ditemukan selama penarian, jika tidak pergi ke langkah 1. Penelitian ini bertujuan untuk menggali informasi dalam memetakan atau mengelompokkan rumah tidak layak huni menggunakan algoritma x-means menjadi beberapa cluster serta untuk mengetahui nilai akurasi terbaik dari hasil uji Davies Bouldin Index (DBI). Hasil yang pengelompokan menggunakan x-means diperoleh terbaik sebanyak 2 cluster, yaitu cluster 0 sebanyak 931 item dan cluster 1 sebanyak 444 item serta Nilai Davies Bouldin Index yang dihasilkan dari algoritma x-means ini sebesar 2,079. Dengan menghitung jarak antara rata-rata cluster diperoleh bahwa cluster 0 merupakan kelompok terbaik dengan jarak terdekat sebesar 9,728.

Kata kunci: Pengelompokan, rumah tidak layak huni, algoritma X-Means

Abstract: A place to live or a house is a very basic community need for everyone besides clothing and food. A livable house can be identified from the structure of the house and the existing home facilities. The structure of the house includes floor area, floor type, roof type, and wall type. Meanwhile, housing facilities include drinking water sources, lighting sources, and sewage disposal sites. In this study, the author will classify uninhabitable houses by looking at the structure of the house buildings and home facilities owned by each household in 4 areas, namely: Batulayang, Siantan downstream, Siantan Tengah and Siantan upstream. One method that can be used for grouping uninhabitable houses is the X-means clustering method. because it is one of the refinements of the K-Means clustering method. The stages of the x-means method consist of two steps that are repeated until they are completed. that is, 1) Increase-Params, in this step apply the k-means algorithm initially for k clusters until convergence. Where k is equal to the lower limit provided by the user. 2) Fix Structure, this structural refinement step begins by breaking each cluster center into two children in opposite directions along a randomly selected vector. After that run k-means locally within each cluster for two clusters. The decision of each cluster center itself by comparing the Bayesian Information Criterion (BIC) values. 3) If $K \geq k_{max}$ (upper limit) stop and report to the best scoring model found during the dance, otherwise go to step 1. This study aims to explore information in mapping or classifying uninhabitable houses using the x-means algorithm into several clusters and to find out the best accuracy value from the Davies Bouldin Index (DBI) test results. The results of grouping using x-means obtained the best as many as 2 clusters, namely cluster 0 as many as 931 items and cluster 1 as many as 444 items and the Davies Bouldin Index value generated from this x-means algorithm was 2.079. By calculating the distance between the average clusters, it is obtained that cluster 0 is the best group with the closest distance of 9.728

Keywords: Clastering, Uninhabitable Houses, X-Means algorithm

1. PENDAHULUAN

Rumah tidak layak huni yang selanjutnya di singkat RTLH adalah rumah yang tidak memenuhi persyaratan keselamatan bangunan, kecukupan minimum luas bangunan dan kesehatan penghuni. Kebutuhan ruang per orang dihitung berdasarkan aktivitas dasar manusia (aktivitas tidur, makan

kerja, duduk, mandi, kakus, cuci dan masak, serta ruang gerak lainnya) didalam rumah. Dari hasil kajian kebutuhan ruang per orang adalah 9 m² dengan perhitungan ketinggian rata-rata langit-langit adalah 2.80 m. Rumah sederhana sehat memungkinkan penghuni untuk dapat hidup sehat, dan menjalankan kegiatan hidup sehari-hari secara layak. (Sumber: Peraturan menteri pekerjaan umum dan perumahan rakyat Republik Indonesia no. 07/PRT/m/2018 tentang bantuan stimulant perumahan swadaya).

Penelitian yang dilakukan oleh Tiar Pandapotan Purba, dan Topan Himawan, menyimpulkan bahwa, perumahan dan permukiman merupakan salah satu kebutuhan dasar bagi manusia yang mempunyai peran penting dalam peningkatan kualitas sumber daya manusia (SDM). Upaya yang dilakukan dalam menyelesaikan permasalahan penduduk yang kurang mampu dalam materi adalah pemenuhan hak dasar penduduk seperti pemenuhan pangan, layanan kesehatan, layanan pendidikan, pekerjaan dan berusaha, air bersih, dan sanitasi serta hak pemenuhan tempat tinggal. Perumahan adalah salah satu kebutuhan dasar yang sulit dimiliki oleh masyarakat miskin di Indonesia. (Tiar Pandapotan Purba, dan Topan Himawan, 2021).

Penelitian terdahulu yang dilakukan oleh Aditya Wijayanto dan Rifki Adhitama dalam jurnalnya menyimpulkan bahwa Algoritma X-Means merupakan algoritma yang menyempurnakan kekurangan dari algoritma KMeans, dimana algoritma K-means memiliki kelemahan dalam komputasi yang relative lama dan nilai kluster K yang harus ditentukan oleh user. Algoritma pembelajaran ini termasuk dalam kategori unsupervised learning dimana komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Algoritma X-Means mengoptimalkan nilai Bayesian Information Criterion (BIC). Komputasi pada pengoptimalan nilai BIC jauh lebih efisien daripada satu kali iterasi pada algoritma K-Means (Aditya Wijayanto #1, Rifki Adhitama #2, 2019).

Permasalahan dari penelitian ini adalah: 1) Belum adanya teknik penilaian yang objektif dan tepat sasaran untuk menentukan calon penerima bantuan peralatan menjahit. 2) Banyaknya data yang belum diolah untuk mengelompokkan mana yang masuk dalam kategori penerima bantuan dan bukan penerima bantuan. 3) bagaimana untuk melakukan analisa rumah tidak layak huni dengan menggunakan algoritma x-means klustering. Tujuan dari penelitian ini adalah untuk menyediakan layanan informasi dalam pengenalan menggunakan sistem pakar, antara lain: 1) Untuk mengetahui teknik penilaian yang objektif dalam menentukan rumah tidak layak huni. 2) Untuk mengetahui informasi pengelompokan dataset rumah tidak layak huni terbaik berdasarkan wilayah. 3) Untuk mengetahui bagaimana cara menganalisa rumah tidak layak huni menggunakan algoritma X-means.

Metode penelitian yang digunakan yaitu menggunakan algoritma x-means yaitu sebuah algoritma yang dapat mendefinisikan objek dalam suatu pusat kelompok data yang biasanya menjadi titik-tengah dari kelompok data tersebut. Algoritma ini akan melakukan perulangan untuk mendapatkan titik-tengah yang optimal. Metode X-Means ini akan membagi kluster sesuai dengan jumlah kluster yang telah ditentukan atau diinisiasi diawal saat menjalankan algoritma ini. Hasil penelitian pengelompokan Dataset rumah tidak layak huni menggunakan algoritma X-Means, maka hasilnya dapat disimpulkan antara lain: 1) Hasil pengelompokan menggunakan x-means diperoleh terbaik sebanyak 2 cluster, yaitu cluster 0 sebanyak 931 item dan cluster 1 sebanyak 444 item, 2) Nilai Davies Bouldin Index yang dihasilkan dari algoritma x-means ini sebesar 2,079. Dengan menghitung jarak antara rata-rata cluster diperoleh bahwa cluster 0 merupakan kelompok terbaik dengan jarak terdekat sebesar 9,728.

2. TINJAUAN PUSTAKA

a. Pengertian Data mining

Penerapan data mining untuk clustering data penduduk miskin pada kota Jambi menggunakan algoritma k-means (Sunia, Kurniabudi, & Jusia, 2019). Hasil dari penelitian tersebut terdapat lima cluster yang dihasilkan. Selanjutnya, Metode k-means untuk pengelompokan masyarakat miskin menghasilkan tiga cluster dalam penelitiannya (Ramadani et al., 2019). Penerapan data mining untuk clustering data penduduk miskin menggunakan algoritma hard c-means menghasilkan tiga cluster dalam penelitian tersebut. Selanjutnya, implementasi data mining dalam penentuan tingkat

kemiskinan Kabupaten Bone Bolango, Provinsi Gorontalo menggunakan fuzzy c-means juga menghasilkan tiga cluster (Juniar Hutagalung, Usti Fatimah Sar, 2021) .

b. Pengertian Rumah

Pentingnya tempat tinggal atau rumah layak huni untuk setiap orang dapat berlindung, istirahat, dan berkumpul dengan keluarga dengan aman dan nyaman perlu diperhatikan. Di Indonesia masih banyak sekali rumah yang masuk dalam kategori tak layak huni. Pemerintah memberikan bantuan Rumah Tidak Layak Huni (RTLH) berupa uang untuk pembelian bahan bangunan guna pemugaran Rumah Tidak Layak Huni dimana rumah tersebut tidak memenuhi persyaratan keselamatan bangunan dan kecukupan minimal bangunan. (Asep Muhidin, Indarwista Baragigiratri, 2017).

c. Algoritma X-Means

Menurut Ahmad Jurnaidi Wahidin, Dana Indra Sensuse mengatakan bahwa: Algoritma k-means merupakan algoritma yang paling banyak digunakan dan sering digunakan sebagai algoritma perbandingan, k-means memiliki kelebihan ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar, namun k-means juga memiliki kelemahan antara lain mengharap pengguna untuk menentukan banyak cluster terlebih dahulu dan juga sifat kaku (hard) yang dimiliki. (Ahmad Jurnaidi Wahidin, Dana Indra Sensuse, 2021)

3. METODE

a. Metode penelitian

Metode penelitian yang digunakan yaitu menggunakan algoritma x-means yaitu sebuah algoritma yang dapat mendefinisikan objek dalam suatu pusat kelompok data yang biasanya menjadi titik-tengah dari kelompok data tersebut. Algoritma ini akan melakukan perulangan untuk mendapatkan titik-tengah yang optimal. Metode X-Means ini akan membagi klaster sesuai dengan jumlah klaster yang telah ditentukan atau diinisiasi diawal saat menjalankan algoritma ini.

Tahapan Algoritma X-Means adalah :

1. Tentukan nilai k sebagai jumlah klaster yang ingin dibentuk
 2. Inisiasi k sebagai centroid yang dapat dibangkitkan secara random
 3. Hitung jarak setiap data ke masing-masing centroid menggunakan persamaan Euclidean Distance
 4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya
 5. Tentukan posisi centroid baru (k)
 6. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama
- Penghitungan jarak antara data dengan centroid menggunakan euclidean distance dengan persamaan sebagai berikut:

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2} \quad (1)$$

Kemudian untuk menentukan centroid berikutnya pada langkah 5 yaitu dengan persamaan:

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x} \quad (2)$$

Dengan C adalah centroid baru yang dihasilkan berdasarkan data pada klaster yang terbentuk

b. Operasionalisasi Variabel

Operasional variabel yang akan dilakukan dalam penelitian ini dibagi menjadi 2 (dua) variabel yaitu terdiri dari variable (X) : Klasterisasi dataset rumah tidak layak, sedangkan untuk variable (Y) adalah algoritma x-means clustering.

4. HASIL DAN PEMBAHASAN**4.1 Hasil****a. Lokasi penelitian**

Lokasi penelitian yang dilakukan oleh peneliti yaitu penelitian secara deskripsi karena sumber data dilakukan melalui referensi atau hasil pengkajian data dari repository BPS. Penulis melakukan Penelitian dalam pengambilan datanya yaitu melalui Data yang diperoleh dari <https://www.bps.go.id/>, seperti pada tampilan gambar 4.1.



Gambar 1.1: Tampilan Lokasi penelitian
(Sumber: <https://www.bps.go.id/>)

b. Populasi dan Sampel

Populasi dan Sampel yang digunakan pada penelitian ini yaitu Dataset rumah yang tidak layak huni yang berjumlah 1374 record dengan rincian terdiri Jumlah: 1) Wilayah Batu Layang = 320. 2) Wilayah Siantan Hilir= 480, 3) Wilayah Siantan Tengah= 190, 4) Wilayah Siantan Hulu = 384

c. Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset sebanyak 1374 record dan terdiri 23 atribut pada tahun 2019 sampai dengan 2020 dataset tersebut bersumber dari Repository Badan pusat statistik.

d. Selection

Data yang dipilih pada penelitian ini adalah data rumah tidak layak huni pada tahun 2019. Data ini diperoleh dari <https://www.bps.go.id/> yang merupakan data publik, dan data ini memiliki 1374 record dan Dari 33 atribut yang ada pada masing-masing wilayah, hanya digunakan 23 atribut. Karena 11 attribute yang tidak digunakan memiliki nilai yang sama pada masing-masing atributnya. Berikut atribut yang dipakai setelah dilakukan seleksi seperti pada table 4.2.

Selection atribut adalah atribut-atribut yang terpilih dari 33 atribut setelah dilakun seleksi atribut yang ganda ternyata hanya ada 23 atribut, seperti pada table 4.2 mulai dari atribut Kode wilayah sampai dengan atribut pada baris ke 23 yaitu atribut Kondisi lantai.

Tabel 4.2: Selection atribut

No	Nama Atribut	Type
1	Kode Wilayah	polynomial
2	No Urut Rtlh	Integer
3	Nama	polynomial
4	Jenis Kelamin	binomial
5	Nomor Ktp	Nominal
6	Alamat	polynomial
7	Pekerjaan	polynomial
8	Besar Penghasilan Atau Pengeluaran Perbulan	polynomial
9	Pernah Mendapatkan Bantuan Perumahan	polynomial
10	Kolom Dan Balok	polynomial
11	Konstruksi Atap	polynomial
12	Kepemilikan Kamar Mandi Dan Jamban	polynomial
13	Sumber Air Minum	Nominal
14	Jarak Sumber Air Minum Ke Tempat Pembuangan Kotoran/Tinja	polynomial
15	Sumber Utama Penerangan Rumah Tangga	polynomial
16	Luas Rumah (M ²)	polynomial
17	Jumlah Penghuni	polynomial
18	Material Atap Terluas	polynomial
19	Kondisi Material Atap	polynomial
20	Material Dinding Terluas	polynomial
21	Kondisi Dinding	polynomial
22	Kondisi Dinding	polynomial
23	Kondisi Lantai	polynomial

e. Preprocessing

Pada langkah preprocessing dilakukan beberapa kegiatan yaitu sebagai berikut:

1. Pembersihan data (Data Cleaning)

Pembersihan data dilakukan untuk membersihkan data dengan cara dihapus record yang berisi data yang tidak konsisten data pada: Dataset rumah terdiri 4 buah wilayah yang bernama batu layang, siantan hilir, siantan tengah dan siantan hulu. Pembersihan data disajikan seperti tampilan pada table 4.3:

Tabel 4.3: Pembersihan data

No.	Nama File	Jumlah Record	Missing
1.	Batu Layang	320	0
2.	Siantan Hilir	480	0
3.	Siantan Tengah	190	1
4.	Siantan Hulu	384	0
Jumlah		1374	1

Record ke-190 yang missing pada wilayah Siantan Tengah di delete untuk menghilangkan missing, sehingga jumlah total record menjadi 189. Masing-masing wilayah dibaca dengan menggunakan operator Read Excel. Hasil untuk Wilayah Batu Layang adalah seperti tampilan gambar 4.4.

Row No.	KODE WILA	NO URUT RT	NAMA	JENIS KELA	NOMOR KTP	ALAMAT	JUMLAH KK	PEKEI
1	6171041004	1	KRISNO SUH...	LAKI-LAKI	6171040405...	JL. KHATULI...	1	KARY...
2	6171041004	2	SAHREL	LAKI-LAKI	6171042005...	JL. KHATULI...	1	KARY...
3	6171041004	3	SURYATI	PEREMPUAN	6171045410...	GG. BERINGL...	1	KARY...
4	6171041004	4	MARSELUS ...	LAKI-LAKI	6108020708...	GG. BERINGL...	1	KARY...
5	6171041004	5	SUNG POH ...	LAKI-LAKI	6171040101...	JL. KHATULI...	1	KARY...
6	6171041004	6	BATIR	LAKI-LAKI	6171041008...	JL. KHATULI...	1	KARY...
7	6171041004	7	ALIYAKBER	LAKI-LAKI	6171040102...	JL. FLORA G...	1	BURU...
8	6171041004	8	RUSDI HASAN	LAKI-LAKI	6171041208...	JL. FLORA G...	1	KARY...
9	6171041004	9	ANI	LAKI-LAKI	6171044306...	JL. FLORA G...	1	BURU...
10	6171041004	10	SAFIE	LAKI-LAKI	6171041203...	JLN. FLORA ...	1	KARY...
11	6171041004	11	EDI SUPRAP...	LAKI-LAKI	6171023108...	JL. KEBANG...	1	KARY...
12	6171041004	12	DWI WAHYU ...	LAKI-LAKI	6171041210...	JL. KEBANG...	1	KARY...
13	6171041004	13	SY. RAMH	PEREMPUAN	6112030405...	R. KERAMC...	1	KARY...

Gambar 4.4: Read Excel.Wilayah Batu Layang

Hasil untuk Wilayah Siantan Hilir adalah seperti tampilan pada gambar 4.5.

Row No.	KODE WILA.	NO URUT RT...	NAMA	JENIS KELA.	NOMOR KTP	ALAMAT	JUMLAH KK...	PEKEI
1	6171041003	1	SALYAH	PEREMPUAN	6171044904	Jln Khabulab.	1	LAIN
2	6171041003	2	ZAINAL	LAQ-LAKI	6171041212	Jln Khabulab.	1	KARY
3	6171041003	3	HAMNAH	PEREMPUAN	6171047112	Gg. Selamat 6	1	LAIN
4	6171041003	4	JASMA	PEREMPUAN	6171044202	Gg. Selamat 6	1	LAIN
5	6171041003	5	NHRILAN	LAQ-LAKI	6171042505	Gg. Teluk Sel.	1	KARY
6	6171041003	6	RASIB	LAQ-LAKI	6171041304	Gg. Teluk Sel.	1	BURL
7	6171041003	7	MOCLIS	LAQ-LAKI	6171021202	Gg. Teluk Sel.	1	BURL
8	6171041003	8	YUSRIAN	LAQ-LAKI	6171042403	Gg. Teluk Sel.	1	BURL
9	6171041003	9	HARYADI	LAQ-LAKI	6171043112	Gg. Teluk Sel.	1	BURL
10	6171041003	10	HIRUDIN	LAQ-LAKI	6171041208	Gg. Teluk Sel.	1	WIRAI
11	6171041003	11	HASAN	LAQ-LAKI	6171041608	Gg. Teluk Sel.	1	BURL
12	6171041003	12	AHMAD	LAQ-LAKI	6171040102	Gg. Teluk Sel.	1	BURL

Gambar 4.5: Wilayah Siantan Hilir

Hasil untuk Wilayah Siantan Tengah adalah seperti tampilan pada gambar 4.6.

Row No.	KODE WILA.	NO URUT RT...	NAMA	JENIS KELA.	NOMOR KTP	ALAMAT	JUMLAH KK...	PEKEI
1	6171041002	1	SITI RAHMIMAH	PEREMPUAN	6171045201	Jl. Panwasal ...	1	LAIN
2	6171041002	2	NASYEH	PEREMPUAN	6171045108	Jl. Panwasal ...	1	LAIN
3	6171041002	3	KAMARUDIN	LAQ-LAKI	6171041409	Jl. Panwasal ...	1	BURL
4	6171041002	4	PURADAH (B.)	PEREMPUAN	6171045002	Jl. Panwasal ...	1	LAIN
5	6171041002	5	SUNIRI	LAQ-LAKI	6171041606	Jl. Panwasal ...	1	KARY
6	6171041002	6	IYAS	LAQ-LAKI	6171042101	Jl. Panwasal ...	1	WIRAI
7	6171041002	7	HASAN NUG.	LAQ-LAKI	6171040101	Jl. Panwasal ...	1	KARY
8	6171041002	8	SAIRI	LAQ-LAKI	6171041706	Jl. Panwasal ...	1	WIRAI
9	6171041002	9	NURANI	PEREMPUAN	6171044405	Gg. Damai	1	LAIN
10	6171041002	10	ERWANSHYAH	LAQ-LAKI	6171040909	Gg. Damai	1	BURL
11	6171041002	11	JANNI	LAQ-LAKI	6171040505	Gg. Damai	1	BURL
12	6171041002	12	NORTIA	PEREMPUAN	6171044303	Gg. Perjuang...	1	LAIN

Gambar 4.6: Statistic siantan tengah

Hasil untuk Wilayah Siantan Tengah adalah seperti tampilan pada gambar 4.7.

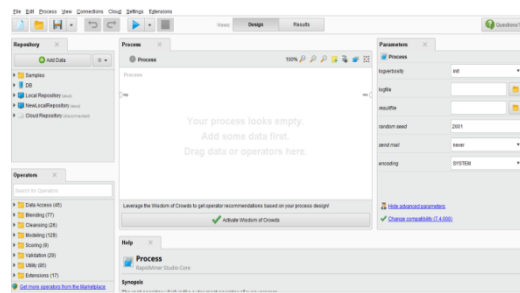
Row No.	KODE WILA.	NO URUT RT...	NAMA	JENIS KELA.	NOMOR KTP	ALAMAT	JUMLAH KK...	PEKEI
1	6171041001	1	KARJAN	LAQ-LAKI	6171040810	Jl. Gusti Sibat ...	1	BURL
2	6171041001	2	SUHRI	LAQ-LAKI	6171040209	Jl. Pant Pang...	1	BURL
3	6171041001	3	MUTIRAH	PEREMPUAN	6171045004	Jl. Gusti Sibat ...	1	LAIN
4	6171041001	4	FAISAL	LAQ-LAKI	6171043011	Jl. Gusti Sibat ...	1	KARY
5	6171041001	5	KASIRAN	LAQ-LAKI	6171040705	Jl. Gusti Sibat ...	1	LAIN
6	6171041001	6	MUNIWI	PEREMPUAN	6171044512	Jl. Gusti Sibat ...	1	LAIN
7	6171041001	7	RIWANTEN	PEREMPUAN	6171045003	Jl. Gusti Sibat ...	1	LAIN
8	6171041001	8	MURNI UMAR	LAQ-LAKI	6171040302	Jl. Gusti Sibat ...	1	LAIN
9	6171041001	9	MARYANI	PEREMPUAN	6171044305	Jl. Gusti Sibat ...	1	LAIN
10	6171041001	10	BUDI SABRAN	LAQ-LAKI	6171041205	Jl. Gusti Sibat ...	1	KARY
11	6171041001	11	SARILAN	LAQ-LAKI	6171040103	Jl. Gusti Sibat ...	1	WIRAI
12	6171041001	12	MARLIHA	PEREMPUAN	6171044802	Jl. Gusti Sibat ...	1	LAIN

Gambar 4.7: Statistic siantan hulu

2. Data Mining

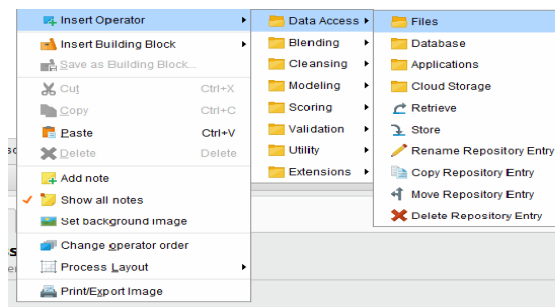
Proses mencari atau menggali informasi menarik dalam sebuah data dengan jumlah besar yang terpilih dengan menggunakan teknik atau metode tertentu. Penelitian ini menggunakan algoritma *x-means* untuk mengelompokkan penduduk miskin di indonesia berdasarkan provinsi. Dalam penelitian ini pengolahan data penduduk miskin menggunakan bantuan *software* RapidMiner dengan menggunakan algoritma *x-means clustering*.

Dengan menggunakan algoritma *k-medoids clustering* dapat dilihat pada gambar 4.8.

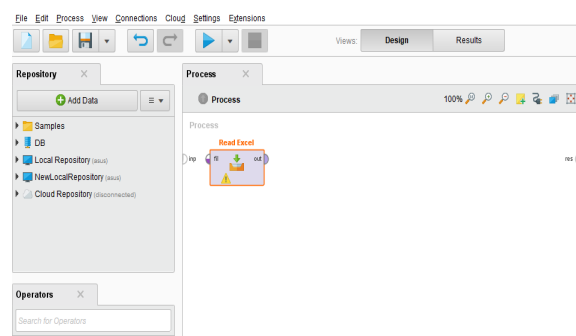


Gambar 4.8: Halaman Utama RapidMiner

Pada gambar 4.18 menampilkan halaman utama RapidMiner untuk membuat halaman kerja pada RapidMiner. Import data dilakukan untuk memasukan data yang akan diuji dalam bentuk format .xls atau .xlsx. Adapun langkah untuk import data yaitu Klik kanan pada layar → Insert Operator → Data Access → Files → Read → Read Excel.



Gambar 4.9: Memanggil Operator Read Excel



Gambar 4.10: Operator Read Excel yang sudah dipanggil

Setelah operator read excel telah dipanggil, langkah selanjutnya yaitu import data untuk diproses dengan cara klik Import Configurations Wizard → klik file yang akan digunakan. Jika data telah berhasil di import dan tidak ada yang error maka klik finish untuk melanjutkan ke langkah selanjutnya. Berikut tampilan import data :

Open in Turbo Prep Auto Model Filter (320 / 320 examples): all

Row No.	KODE WILA...	NO URUT RT...	NAMA	JENIS KELA...	NOMOR KTP	ALAMAT	JUMLAH KK...	PEKE...
1	6171041004	1	KRISNO SUH...	LAKI-LAKI	6171040405...	JL. KHATULL...	1	KARY...
2	6171041004	2	SAHREL	LAKI-LAKI	6171042006...	JL. KHATULL...	1	KARY...
3	6171041004	3	SURYATI	PEREMPUAN	6171045410...	GG. BERINGL...	1	KARY...
4	6171041004	4	MARSELUS...	LAKI-LAKI	6108020708...	GG. BERINGL...	1	KARY...
5	6171041004	5	SUNG POH...	LAKI-LAKI	6171040101...	JL. KHATULL...	1	KARY...
6	6171041004	6	BATJAR	LAKI-LAKI	6171041008...	JL. KHATULL...	1	KARY...
7	6171041004	7	ALYAKSER	LAKI-LAKI	6171040102...	JL. FLORA G...	1	BURL...
8	6171041004	8	RUSDI HASAN	LAKI-LAKI	6171041208...	JL. FLORA G...	1	KARY...
9	6171041004	9	ANI	LAKI-LAKI	6171044306...	JL. FLORA G...	1	BURL...
10	6171041004	10	SAPTE	LAKI-LAKI	6171041203...	JLN. FLORA...	1	KARY...
11	6171041004	11	EDI SUPRAP...	LAKI-LAKI	6171023108...	JL. KEBANG...	1	KARY...
12	6171041004	12	DWI WAHYU...	LAKI-LAKI	6171041210...	JL. KEBANG...	1	KARY...
13	6171041004	13	EV. DAHM...	BEDEHDIAM...	6110030406...	B. KEDAMIC...	4	KADY...

ExampleSet (320 examples, 0 special attributes, 33 regular attributes)

Gambar 4.11: Tampilan import data

4.2 Pembahasan

Pada proses *clustering k-means* terdapat beberapa operator yang digunakan untuk menganalisis data. Operator tersebut adalah :

1. *Read Excel*, digunakan untuk membaca data yang akan dianalisis.
2. *Clustering*, digunakan untuk mengelompokkan kedalam beberapa *cluster*.
3. *Cluster Performance Distance*

Karena file yang dibaca atau diinputkan pada rapidminer lebih dari satu file, maka diperlukan sebuah operator Append untuk menggabungkan 4 file excel menjadi 1 file excel seperti tampak pada gambar 4.12 sebagai berikut.



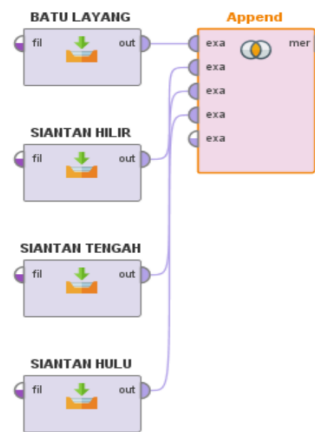
Gambar 4. 12: Operator Append

Parameter yang digunakan pada operator *Append*, tampak pada tabel 4.11.

Tabel 4.11: Parameter Operator Append

Parameter	Pilihan
data management	auto

Model proses pada pembacaan data pada Rapidminer tampak pada gambar 4.3 sebagai berikut.



Gambar 4.13: Model Proses Pembacaan Data

Open in Turbo Prep Auto Model Filter (1,375 / 1,375 examples): all

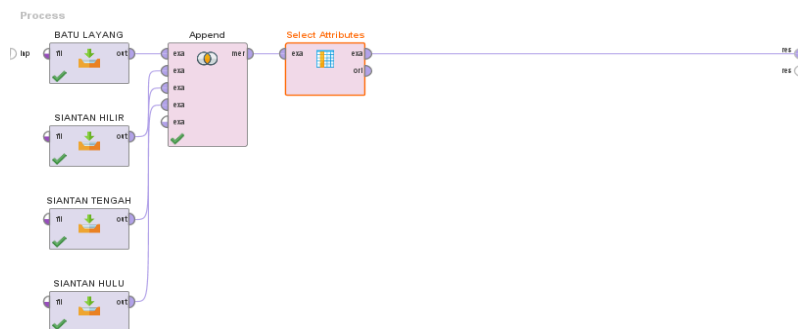
Row No.	KODE WILA...	NO URUT RT...	NAMA	JENIS KELA...	NOMOR KTP	ALAMAT	JUMLAH KK...	PEKEJ...
1	6171041004	1	KRISNO SUH...	LAKI-LAKI	6171040405...	JL. KHATULI...	1	KARY...
2	6171041004	2	SAHREL	LAKI-LAKI	6171042006...	JL. KHATULI...	1	KARY...
3	6171041004	3	SURYATI	PEREMPUAN	6171045410...	GG. BERINGI...	1	KARY...
4	6171041004	4	MARSELUS ...	LAKI-LAKI	6108020708...	GG. BERINGI...	1	KARY...
5	6171041004	5	SUNG POH ...	LAKI-LAKI	6171040101...	JL. KHATULI...	1	KARY...
6	6171041004	6	BATIAR	LAKI-LAKI	6171041008...	JL. KHATULI...	1	KARY...
7	6171041004	7	ALIYAKSER	LAKI-LAKI	6171040102...	JL. FLORA G...	1	BURL...
8	6171041004	8	RUSDI HASAN	LAKI-LAKI	6171041208...	JL. FLORA G...	1	KARY...
9	6171041004	9	ANI	LAKI-LAKI	6171044306...	JL. FLORA G...	1	BURL...
10	6171041004	10	SAFTIE	LAKI-LAKI	6171041203...	JLN. FLORA ...	1	KARY...
11	6171041004	11	EDI SUPRAP...	LAKI-LAKI	6171023108...	JL. KEBANG...	1	KARY...
12	6171041004	12	DWI WAHYU ...	LAKI-LAKI	6171041210...	JL. KEBANG...	1	KARY...
13	6171041004	13	SY. D'ALMI	PEREMPUAN	6112020105...	JL. KEBANG...	1	KARY...

ExampleSet (1,375 examples, 0 special attributes, 33 regular attributes)

Gambar 4.14 :Data Integrasion

3. Data Selection

Seleksi data dilakukan untuk memilih atribut yang akan dianalisa menggunakan operator Select Attribute.



Gambar 4.15: Data Selection

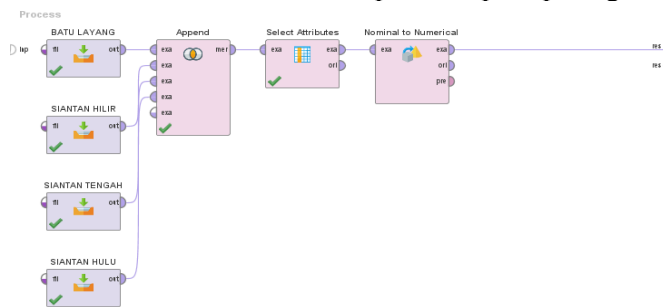
Parameter:

Tabel 4.16: Parameter-1

Parameter	Keterangan
Attribute filter type	Subset
Attribute	Select Attribute: (alamat, besar penghasilan, jarak sumber air minum, jenis kelamin, jumlah penghuni, kepemilikan kamar mandi, kode wilayah, kolom dan balok, kondisi dinding, kondisi lantai, kondisi material atap, konstruksi atap, luas rumah, material atap terluas, material dinding terluas, material lantai terluas, nama, no urut rtlh, pekerjaan, pernah mendapat bantuan, sumber air minum, sumber utama penerangan)

4. Data Transformation

Transformasi data dilakukan untuk mengubah data bertipe nominal menjadi numerik dengan menggunakan operator **Nominal to Numeric** model proses seperti pada gambar 4.16



Gambar 4.16 :Data Transformations

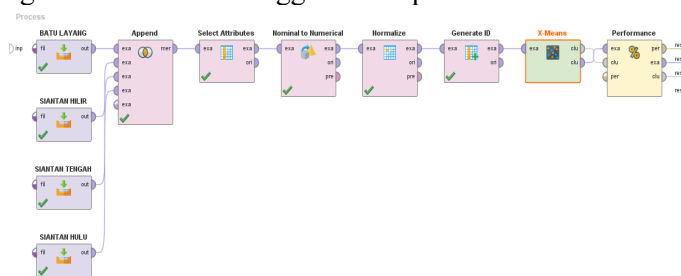
Untuk Parameter hasil data transformasi seperti pada table 4.17

Tabel 4.17 :Parameter-2

Parameter	Keterangan
Attribute filter type	Subset
Attribute	Select Attribute: (seluruh atribut)
Coding type	Unique integer

5. Data Mining

Data mining menggunakan algoritma X-means dan untuk mengukur besarnya Davies Bouldin yang diperoleh dari algoritma x-means menggunakan operator Cluster Distance Performance.



Gambar 4.18: Data Mining

Parameter X-Means tampilannya seperti pada table 4.19

Tabel 4.19: Parameter x-means

Parameter	Keterangan
k min	2
k max	20
measure type	NumericalMeasurement
numerical measure	EuclideanDistance
clustering algorithm	KMeans

Cluster distance performance tampilannya seperti pada table 4.20

Tabel 4.20: Cluster distance performance

Parameter	Keterangan
Main criterion	Davies Bouldin

6. Interpretasi Data

Hasil dari data mining yang diperoleh dengan menggunakan algoritma x-means sebagai berikut.

- Hasil pengelompokan menggunakan x-means diperoleh terbaik sebanyak 2 cluster, yaitu cluster 0 sebanyak 931 item dan cluster 1 sebanyak 444 item.

Cluster Model

```
Cluster 0: 931 items
Cluster 1: 444 items
Total number of items: 1375
```

- Nilai Davies Bouldin Index yang dihasilkan dari algoritma x-means ini sebesar 2,079. Dengan menghitung jarak antara rata-rata cluster diperoleh bahwa cluster 0 merupakan kelompok terbaik dengan jarak terdekat sebesar 9,728

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -142.557
Avg. within centroid distance_cluster_0: -132.829
Avg. within centroid distance_cluster_1: -162.954
Davies Bouldin: -2.079
```

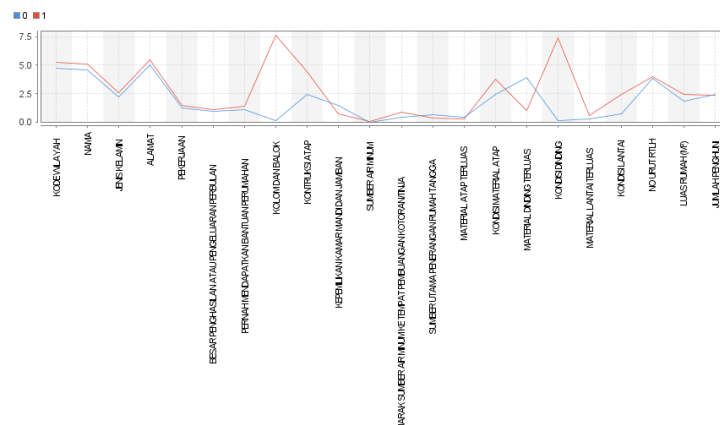
Tabel 4.21: Performance vector

	0	1
Avg. within centroid distance	142.557	142.557
Avg. within centroid distance	132.829	162.954
distance	9.728	20.397

c. Plot

Dari grafik plot diperoleh data bahwa beberapa karakteristik sebagai berikut.

- Nilai dari cluster 0 selalu lebih kecil dari pada cluster 1 kecuali pada atribut “kepemilikan kamar mandi dan jamban”, “sumber utama penerangan rumah tangga”, “material atap terluas”, “material dinding terluas”, dan “jumlah penghuni”.
- Pada atribut “kolom dan balok”, “konstruksi atap”, “sumber utama penerangan rumah tangga”, “material dinding”, dan “kondisi dinding” memiliki nilai yang berlawanan arah.



Gambar 4.42: Grafik Plot

5. KESIMPULAN

Dataset rumah tidak layak huni menggunakan algoritma X-Means, maka hasilnya dapat disimpulkan sebagai berikut :

1. Hasil pengelompokan menggunakan x-means diperoleh terbaik sebanyak 2 cluster, yaitu cluster 0 sebanyak 931 item dan cluster 1 sebanyak 444 item
2. Nilai Davies Bouldin Index yang dihasilkan dari algoritma x-means ini sebesar 2,079. Dengan menghitung jarak antara rata-rata cluster diperoleh bahwa cluster 0 merupakan kelompok terbaik dengan jarak terdekat sebesar 9,728.

DAFTAR PUSTAKA

- Ahmad Jurnaidi Wahidin, Dana Indra Sensuse. (2021). Perbandingan Algoritma K-Means, X-Means Dan K-Medoids Untuk Klasterisasi Awak Kabin Lion Air. *urnal ICT : Information Communication & Technology*, 298-302.
- Ai Rohmah, Falentino Sembiring, Adhitia Erfina. (2021). Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: Smk Yaspim Gegerbitung). *Sismatik (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, 290-298.
- Asep Muhidin, Indarwista Baragigiratri. (2017). Pemetaan Penduduk Calon Penerima Bantuan Renovasi Rumah Desa Pesangkalan Menggunakan Algoritma Clustering K-Means. *Paper Knowledge . Toward a Media History of Documents*.
- Benri Melpa Metisen, Herlina Latipa Sari. (2017). Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila. *Jurnal Media Infotama*, 110-118.
- Bustami Yusuf, Rike Mahara, Hendri Ahmadian , Sri Wahyuni4, Khairan AR. (2022). Analisis Clustering Penduduk Miskin Di Provinsi Aceh Menggunakan Algoritma K-Means Dan X-Means. *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, 26-35.
- Deni Triyansyah dan Devi Fitriana. (2018). Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing. *Jurnal Telekomunikasi dan Komputer*, 163.
- Erene Gernaria Sihombing. (2017). Klasifikasi Data Mining pada Rumah Tangga Menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa Menggunakan K-Means Clustering Method. *Computer Engineering, System and Science Journal*, 74-82.
- Faiza Rini1), Novhirtamely Kahar2), 3)Juliana. (2016). Penerapan Algoritma K-Means Pada Pengelompokan Data Siswa Baru Berdasarkan Jurusan Di Smk Negeri 1 Kota Jambi Berbasis Web. *Seminar Nasional APTIKOM*, 94-99.
- Hasyrif SY, Rismayani, Asrul Syam. (2019). Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare di Kota Makassar. *ISITI : Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, 73-82.
- Istiqomah Sumadikarta, Evan Abeiza. (2018). Penerapan Algoritma K-Means Pada Data Mining Untuk Memilih Produk Dan Pelanggan Potensial (Studi Kasus : PT Mega Arvia Utama). *Jurnal Satya Informatika*, 1-12.
- Lisa Novia Ningsi, Poningsih, Heru Satria Tambunan. (2021). Implementasi Data Mining Kluster Pada Rumah Tangga Yang Memiliki Akses Hunian Layak Berdasarkan Provinsi. *Kajian Ilmiah Informatika dan Komputer*, 228-234.
- Ni Putu Eka Merliana, Ernawati, Alb. Joko Santoso. (2018). Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering. *Unisbank (Sendi_U)*, 978-979.
- Nur Wakhidah. (2015). Clustering Menggunakan K-Means Algorithm. *Jurnal Transformatika*, 33.
- Olivia Immanuela Massiel, Tesa Nur Padilah2i. (2021). Klasterisasi Angka Usia Muda Melek TIK Berdasarkan Algoritma K-Means Menurut jumlah Provinsi Indonesia. *Jurnal komputer,surnal sains*, 759-767.
- R. P. A. Sormin, F. Y. Rumlawang, L. J. Sinay. (2017). Aplikasi Metode Fuzzy C-Means Untuk Pengklasteran Kelayakan Rumah Application of Fuzzy C-Means Algorithm for Clustering House Feasibility. *Jurnal Ilmu Matematika dan Terapan*, 135-146.

- Rifki Adhitama, Auliya Burhanuddin , Ridho Ananda. (2020). Penentuan Jumlah Cluster Ideal Smk Di Jawa Tengah Dengan Metode X-Means Clustering Dan K-Means Clusterin. *JIKO (Jurnal Informatika dan Komputer)*, 1-5.
- S Ramadani, I Ambarita, A M H Pardede. (2019). Metode K-Means Untuk Pengelompokan Masyarakat Miskin Dengan Menggunakan Jarak Kedekatan Manhattan City Dan Euclidean (Studi Kasus Kota Binjai).