

PENERAPAN ALGORITMA *K-MEANS* DALAM KLASIFIKASI JUMLAH PENDUDUK JAKARTA SELATAN

Reynold Julian Tamba¹, Aries Suharso², Purwanto³

Universitas Singaperbangsa Karawang

Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

E-mail : 1910631170226@student.unsika.ac.id¹, aries.suharso@staff.unsika.ac.id²,

purwanto.masbro@staff.unsika.ac.id³

ABSTRAK

Penelitian ini merupakan sebuah studi kasus yang dilakukan di Ditjen Kependudukan dan Pencatatan Sipil dengan tujuan untuk mengelompokkan jumlah penduduk berdasarkan kecamatan dan jenis kelamin, dan metode yang digunakan adalah algoritma K-Means. Hasil evaluasi dengan menggunakan metode *silhouette coefficient* menunjukkan bahwa K-Means merupakan algoritma terbaik karena memperoleh nilai 0.7774399276167493. Dalam penelitian ini, terdapat 39 jenis kepadatan penduduk yang dianalisis, di mana 6 tergolong dalam kelompok rendah (cluster 0), 20 berada dalam kelompok menengah (cluster 1), 8 kelompok menengah (cluster 2), dan 6 yang paling tinggi (cluster 3). Penelitian ini memberikan kontribusi penting dalam penentuan jumlah naik atau turunnya kepadatan penduduk setiap tahunnya.

Kata kunci : Algoritma K-Means, *Data Mining*, *Silhouette coefficient*

ABSTRACTS

This research is a case study conducted at the Directorate General of Population and Civil Registration with the aim of grouping the population based on sub-district and gender, and the method used is the K-Means algorithm. The evaluation results using the silhouette coefficient method show that K-Means is the best algorithm because it obtained a value of 0.7774399276167493. In this study, there were 39 types of population density analyzed, of which 6 were classified as low (cluster 0), 20 were in the middle group (cluster 1), and 8 were middle group (cluster 2), and 6 were the highest (cluster 3). This research makes an important contribution in determining the amount of increase or decrease in population density each year.

Keywords: Algorithm K-Means, Data Mining, Silhouette coefficient

1. PENDAHULUAN

Pertumbuhan penduduk memiliki tujuh konsekuensi negatif (Kumalasari, 2011), yaitu dampak terhadap pertumbuhan ekonomi, kemiskinan dan ketimpangan pendapatan, lingkungan, kesehatan, pendidikan, pasokan pangan dan migrasi internasional. Perencanaan yang terstruktur dengan baik diperlukan oleh pemerintah untuk menghindari dampak buruk yang mungkin terjadi akibat pertumbuhan penduduk yang tidak diantisipasi dan diperhitungkan sebelumnya, serta agar pertumbuhan tersebut dapat meningkatkan kualitas suatu wilayah. Penerapan Teknik Data Mining Menggunakan Algoritma K-Means pada data Kependudukan merupakan masalah sosial yang dihadapi oleh hampir setiap negara. Kepadatan Penduduk tidak hanya menjadi masalah bagi pemerintah pusat, tetapi juga menjadi masalah serius bagi pemerintah daerah. Masalah padatnya penduduk yang dihadapi

pemerintah sangat erat kaitannya dengan pendapatan yang rendah dan ketidakmampuan untuk memenuhi kebutuhan dasarnya. Pendapatan yang rendah menyebabkan kurangnya akses ke pendidikan dan fasilitas pemerintah lainnya. Demikian juga pendapatan daerah yang rendah menyebabkan distribusi pendapatan masyarakat berkurang.

Tujuan dari penelitian ini untuk menerapkan data mining dengan menggunakan algoritma K-Means dalam Jumlah Penduduk Jakarta Selatan. Penelitian ini akan dilakukan menggunakan analisis cluster dengan metode K-Means.

2. METODE PENELITIAN

Tahapan penelitian yang menerapkan metodologi Knowledge Discovery in Database (KDD), yang terdiri dari enam tahapan yaitu Selection Data (seleksi data), Preprocessing Data (pemrosesan data), Transformation Data (transformasi data), Data Mining, Evaluation Data (evaluasi), dan Knowledge

Presentation (pengetahuan).

Alur penelitian dilakukan dengan beberapa tahapan-tahapan didalamnya sebagai berikut :

1. Data Selection

Tahapan pertama yaitu pengumpulan dan seleksi data. Dataset yang digunakan adalah informasi data kepadatan penduduk di Kota Jakarta.

2. Data Preprocessing

Tahapan kedua yaitu menentukan banyaknya cluster dari dataset yang digunakan, data ini diperoleh langsung dari Dukcapil Kota Jakarta. Dilakukan pembersihan data atau yang disebut dengan data cleaning yang digunakan untuk menghilangkan duplikasi data dan missing value.

3. Data Mining

Pada tahap ketiga proses ini menggunakan algoritma K-means untuk menghitung jarak euclidean antar objek dengan pusat.

4. Data Transformation

Proses transformasi dilakukan apabila pada tabel dataset terdapat data yang kategorial (non numeric) maka data tersebut akan diganti menjadi data yang numeric sesuai kebutuhan

5. Data Evaluation

Pada tahap kelima dilakukan pengujian hasil dari tahap data mining yang digunakan, pada penelitian ini untuk menguji hasil dengan menggunakan metode elbow. Untuk melihat seberapa baik untuk digunakan dalam penelitian ini.

6. Knowledge Presentation

Pada tahap keenam terbentuknya cluster sebagai informasi atau pengetahuan yang dilakukan pada penelitian ini, yang akan divisualisasikan dengan menggunakan google colaboratory sehingga informasi dan pengetahuan akan lebih mudah dipahami.

3. HASIL DAN PEMBAHASAN

Hasil dari penelitian yang telah dilakukan adalah melakukan pengelompokan tingkat kepadatan penduduk berdasarkan cluster rendah dan tinggi di Kota Jakarta Selatan menggunakan teknik clustering dan algoritma K-Means. sebagai metode standar dalam data mining untuk mengolah data mentah yang didapatkan dari hasil observasi, dengan tahapan-tahapan yakni : Seleksi data, Cleaning data, Transformation data, Proses data mining (Clustering), dan Pengujian data.

3.1 Sub Bab

Tahapan pertama yaitu pengumpulan dan seleksi data (data selection). Dataset yang digunakan adalah informasi data kepadatan penduduk yang bersumber dari Dukcapil Kota Jakarta Selatan. Data yang digunakan untuk dilakukan penelitian ini pada tahun 2019,2021,2022 dan 2023.

Kecamatan	Laki-Laki	Perempuan	Jumlah	Tahun
Jagakarsa	173.233	170.593	343.826	2019
Pasar Minggu	161.827	159.406	321.233	2019
Cilandak	108.884	110.064	218.948	2019
Pesanggrahan	127.224	126.596	253.820	2019
Kebayoran Lama	161.147	159.582	320.729	2019
Kebayoran Baru	77.182	77.499	154.681	2019
Mampang Prapatan	79.384	77.190	156.574	2019
Pancoran	85.691	85.009	170.700	2019
Tebet	118.939	119.384	238.323	2019
Setiabudi	58.182	57.763	115.945	2019
Jagakarsa	185.068	184.144	369.212	2021
Pasar Minggu	167.427	166.577	334.004	2021
Cilandak	112.575	114.981	227.556	2021
Pesanggrahan	133.403	133.906	267.309	2021
Kebayoran Lama	77.402	78.898	156.300	2021
Kebayoran Baru	165.655	165.260	330.915	2021
Mampang Prapatan	80.082	78.730	158.812	2021
Pancoran	88.672	88.830	177.502	2021
Tebet	119.694	121.320	241.014	2021
Setiabudi	58.492	58.567	117.059	2021
Tebet	2.725	2.776	5.501	2022
Setiabudi	1.450	1.449	2.899	2022
Mampang Prapatan	1.714	1.655	3.369	2022
Pasar Minggu	3.668	3.445	7.113	2022
Kebayoran Lama	3.428	3.321	6.749	2022
Cilandak	2.118	2.207	4.325	2022
Kebayoran Baru	1.599	1.654	3.253	2022
Pancoran	1.890	1.803	3.693	2022
Jagakarsa	3.179	3.371	6.550	2022
Pesanggrahan	2.923	2.876	5.799	2022
Tebet	1.468	1.482	2.950	2023
Setiabudi	765	740	1.505	2023
Mampang Prapatan	837	809	1.646	2023
Pasar Minggu	1.844	1.776	3.620	2023
Kebayoran Lama	1.933	1.887	3.820	2023
Cilandak	1.223	1.264	2.487	2023
Kebayoran Baru	882	935	1.817	2023
Pancoran	966	995	1.961	2023
Jagakarsa	1.835	1.854	3.689	2023
Pesanggrahan	1.517	1.485	3.002	2023

Gambar 3.1 Data Selection

3.2 Data Preprocessing

Setelah tahap seleksi data (data selection) dilakukan, tahap selanjutnya masuk kedalam tahap Pre-Processing Data yang dilakukan yaitu pembersihan data atau Data Cleaning. Dalam tahap ini dilakukan penghapusan data yang tidak sesuai dengan kebutuhan penelitian, dimana proses tersebut menghilangkan Noise Data atau Missing Values. Data yang menjadi Missing Value karena terdapat atribut yang tidak memiliki nilai, sehingga perlu dilakukan Pre-Processing Data agar mendapatkan dataset yang bersih dan siap untuk diproses. Pada data yang bersumber dari Dukcapil Kota Jakarta Selatan tidak perlu karena data tersebut sudah siap pakai dan sudah berbentuk angka jadi tidak perlu di preprocessing lagi.

3.3 Data Transformation

Pada tahapan ini, melakukan proses transformasi

data (Transformasi Data) jika pada tabel dataset terdapat data yang kategorial (non numeric) akan diganti menjadi data yang numeric sesuai kebutuhan dalam penelitian namun setelah dilakukan seleksi data dan pembersihan. Peneliti tidak menemukan data-data yang semestinya di transformasi.

3.4 Data Mining

Setelah selesai melakukan tahap Transformation Data, berikutnya data dapat masuk ke dalam penerapan data mining. Teknik data mining yang digunakan pada penelitian ini yaitu dengan menggunakan algoritma k-means dihitung menggunakan metode elbow pada tools Google Colaboratory.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan maka dapat diambil kesimpulan sebagai berikut.

1. Penerapan algoritma K-Means mampu mengetahui pengelompokan tingkat kepadatan penduduk di Jakarta Selatan pada tahun 2019,2021,2022 dan 2023 berdasarkan jenis kelamin. Pada pengelompokan tingkat kepadatan penduduk di Jakarta Selatan didapatkan sebanyak 39 jenis kepadatan penduduk masuk ke dalam cluster 0 yang merupakan tingkat rendah, terdapat 20 jenis yang masuk ke dalam cluster 1, terdapat 8 jenis kepadatan yang masuk ke dalam cluster 2, 6 jenis yang masuk ke dalam cluster 3 yang merupakan tingkat tinggi terjadinya kasus kepadatan penduduk.

2. Jadi untuk tiap tahun dan kota itu memiliki cluster yang berbeda, fitur yang paling berpengaruh pada hasil cluster adalah jumlah penduduk wanita, pria dan jumlah keseluruhan. Dan peneliti melihat setiap tahunnya pada setiap kota/kecamatan di Jakarta Selatan itu mengalami penurunan.

4.2 Saran

Dari hasil penelitian yang telah dilakukan, maka terdapat saran yang dapat digunakan pada penelitian selanjutnya yaitu:

1. Diharapkan di penelitian selanjutnya dapat menambahkan jumlah dataset yang lebih banyak, dikarenakan dengan semakin banyaknya data yang digunakan, maka akan semakin berpengaruh untuk mendekati

tujuan yang ingin dicapai.

2. Pada penelitian selanjutnya dapat menggunakan algoritma atau metode clustering lainnya dalam melakukan pengelompokan.
3. Penelitian berikutnya dapat menggunakan tools lain dalam melakukan proses data mining dengan kemampuan yang tepat pada bagian ini, nomer bab penutup akan menyesuaikan seberapa jumlah bab pada bagian isi penelitian. Kesimpulan berisikan jawaban dari penelitian Tersebut, serta saran pengembangan untuk penelitian selanjutnya. Judul bab terakhir ini juga dibuat seragam: Kesimpulan.

DAFTAR PUSTAKA

- [1].Kumalasari, M., (2011). Analisis Pertumbuhan Ekonomi, Angka Harapan Hidup, Angka Melek Huruf, Rata-Rata Lama Sekolah, Pengeluaran Perkapita, dan Jumlah Penduduk Terhadap Tingkat Kemiskinan di Jawa Tengah. UNDIP-IR. Jurnal Thesis.
- [2] Abil Fida, D. , Sambiono, F. , Shiddiqi, S. , Ukamah, and P. S. Muhammadiyah, “Transaksi E-Commerce sebagai Pertahanan UMKM di Tengah Pandemi Covid-19 dalam Perspektif Islam,” Jurnal Imtiyaz, vol. 5, no. 02, pp. 53–63, 2021.
- [3] Berliana and D. P. Ompusunggu, “Pemanfaatan E-Commerce Bagi UMKM pada Era Industri 4.0,”Matriks Jurnal Sosial dan Sains, vol. 5, no. 1, pp. 115–120, May 2023, doi: 10.59784/matriks.v5i1.520.
- [4] N. Leksono Putri Handayani, P. Fitrijanti Soeparan, P. Pratama Kendal Jl Raya Utama Timur No, and K. Kendal, “Peran Sistem Pembayaran Digital Dalam Revitalisasi UMKM,” Journal of Economics and Business Management, vol. 1, no. 3, 2022
- [5] D. S. Fuadi, A. S. Akhyadi, and I. Saripah, “Systematic Review: Strategi Pemberdayaan Pelaku UMKM Menuju Ekonomi Digital Melalui Aksi Sosial,” Diklus: Jurnal Pendidikan Luar Sekolah, vol. 5, no. 1, pp. 1–13, Mar. 2021, doi: 10.21831/diklus.v5i1.37122.