



Analisis Komparatif Jarak Euclidean, Manhattan, Canberra, Chebyshev, Cosine pada K-Means untuk Evaluasi Kepuasan Masyarakat

Fakhri¹, Iis Afrianty^{1*}, Elvia Budianita¹, Fadhilah Syafria¹, Siska Kurnia Gusti¹, Salmiyati²

¹ Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

² Fakultas Psikologi, Program Studi Psikologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹12250111381@students.uin-suska.ac.id, ^{2*}iis.afrianty@uin-suska.ac.id, ³elvia.budianita@uin-suska.ac.id,

⁴fadhilah.syafria@uin-suska.ac.id, ⁵siskakurniagusti@uin-suska.ac.id, ⁶salmiyati@uin-suska.ac.id

Email Penulis Korespondensi: iis.afrianty@uin-suska.ac.id

Abstrak—Pemilihan metrik jarak pada algoritma K-Means Clustering dapat memengaruhi kualitas hasil pengelompokan, terutama pada data kepuasan masyarakat yang menggunakan skala Likert. Penelitian ini bertujuan membandingkan performa lima metrik jarak, yaitu Euclidean Distance, Manhattan Distance, Canberra Distance, Chebyshev Distance, dan Cosine Similarity, dalam mengelompokkan tingkat kepuasan masyarakat terhadap layanan publik. Data penelitian diperoleh dari 533 responden pengguna layanan Mal Pelayanan Publik (MPP) Pekanbaru melalui kuesioner yang terdiri atas 23 pertanyaan berdasarkan dimensi SERVQUAL dan indikator Survei Kepuasan Masyarakat sesuai PERMEN PAN-RB Nomor 14 Tahun 2017. Setelah proses data *cleaning*, satu data duplikat dihapus sehingga sebanyak 532 data responden digunakan pada tahap analisis. Jumlah cluster ditentukan menggunakan Elbow Method, sedangkan kualitas cluster dievaluasi menggunakan Davies-Bouldin Index (DBI) dan Silhouette Score. Hasil penelitian menunjukkan bahwa Manhattan Distance dengan $k=2$ menghasilkan nilai DBI terendah sebesar 0,8144, sedangkan Euclidean Distance dengan $k=3$ menghasilkan nilai Silhouette Score tertinggi sebesar 0,5088. Hasil clustering membentuk kelompok responden dengan tingkat kepuasan yang berbeda, yaitu Tidak Puas, Puas, dan Sangat Puas. Penelitian ini memberikan kontribusi berupa evaluasi komparatif lima metrik jarak pada algoritma K-Means menggunakan dua pendekatan evaluasi sekaligus, yaitu Davies-Bouldin Index dan Silhouette Score, pada data kepuasan masyarakat berbasis skala Likert. Hasil penelitian menunjukkan bahwa performa metrik jarak dapat berbeda bergantung pada metode evaluasi yang digunakan, sehingga pemilihan metrik jarak perlu mempertimbangkan karakteristik data dan tujuan analisis. Perbedaan hasil evaluasi menunjukkan bahwa DBI dan Silhouette Score menilai kualitas clustering dari aspek yang berbeda. Berdasarkan hasil penelitian, Manhattan Distance dan Euclidean Distance menunjukkan performa yang lebih baik dibandingkan metrik jarak lainnya pada data yang digunakan, sehingga dapat menjadi pertimbangan dalam analisis kepuasan masyarakat pada layanan publik.

Kata Kunci: K-Means Clustering; Perbandingan Jarak; Davies-Bouldin Index; Silhouette Score; Kepuasan Masyarakat

Abstract—The selection of distance metrics in the K-Means Clustering algorithm can affect the quality of clustering results, particularly on public satisfaction data measured using a Likert scale. This study aims to compare the performance of five distance metrics, namely Euclidean Distance, Manhattan Distance, Canberra Distance, Chebyshev Distance, and Cosine Similarity, in clustering the level of public satisfaction toward public services. The research data were obtained from 533 respondents who used the services of the Mal Pelayanan Publik (MPP) Pekanbaru through a questionnaire consisting of 23 questions based on the SERVQUAL dimensions and the Community Satisfaction Survey indicators in accordance with PERMEN PAN-RB Number 14 of 2017. After the data cleaning process, one duplicate record was removed, resulting in 532 respondent records used in the analysis stage. The number of clusters was determined using the Elbow Method, while cluster quality was evaluated using the Davies-Bouldin Index (DBI) and Silhouette Score. The results show that Manhattan Distance with $k=2$ produced the lowest DBI value of 0.8144, whereas Euclidean Distance with $k=3$ produced the highest Silhouette Score of 0.5088. The clustering results formed groups of respondents with different satisfaction levels, namely Dissatisfied, Satisfied, and Very Satisfied. This study contributes an evaluative comparison of five distance metrics in the K-Means algorithm using two evaluation approaches simultaneously, namely the Davies-Bouldin Index and Silhouette Score, on public satisfaction data based on a Likert scale. The results indicate that the performance of distance metrics may differ depending on the evaluation method used, therefore the selection of distance metrics should consider the characteristics of the data and the objectives of the analysis. The difference in evaluation results indicates that DBI and Silhouette Score assess clustering quality from different aspects. Based on the findings, Manhattan Distance and Euclidean Distance demonstrated better performance compared to other distance metrics on the dataset used, and can therefore be considered in the analysis of public satisfaction toward public services.

Keywords: K-Means Clustering; Distance Comparison; Davies-Bouldin Index; Silhouette Score; Community Satisfaction

1. PENDAHULUAN

Pelayanan publik dapat dimaknai sebagai bentuk layanan yang diselenggarakan oleh pemerintah kepada masyarakat secara efektif, dengan tujuan untuk menghasilkan kepuasan bagi masyarakat yang menerima layanan tersebut (Supriyanto et al., 2021). Pelayanan publik merupakan segala bentuk kegiatan pemberian layanan yang ditujukan kepada masyarakat yang memiliki kebutuhan terhadap suatu instansi tertentu, yang pelaksanaannya berpedoman pada peraturan dan ketentuan yang telah berlaku (Hayuningtyas & Monoarfa, 2023).

Kepuasan masyarakat dapat dipahami sebagai hasil perbandingan antara harapan yang dimiliki masyarakat terhadap suatu layanan dengan kenyataan layanan yang sesungguhnya mereka terima, baik dari segi kualitas maupun kinerja pelayanan tersebut. Dalam proses pelayanan yang melibatkan masyarakat dan pemerintah, diharapkan kedua belah pihak dapat merasakan kepuasan, baik dari sisi pemberi layanan maupun dari sisi penerima layanan itu sendiri (Harahap et al., 2024).



Berdasarkan ketentuan yang tercantum dalam Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 23 Tahun 2017, Mal Pelayanan Publik (MPP) didefinisikan sebagai sebuah wadah yang difungsikan sebagai sentra penyelenggaraan berbagai jenis layanan publik, yang meliputi layanan di bidang barang, jasa, dan/atau urusan administrasi. Kehadiran MPP merupakan bentuk pengembangan dan perluasan dari konsep pelayanan terpadu yang beroperasi di berbagai tingkatan, baik pusat maupun daerah, sekaligus mengakomodasi dan memadukan layanan yang berasal dari BUMN, BUMD, maupun pihak swasta. Keseluruhannya diarahkan untuk mewujudkan penyelenggaraan layanan yang lebih cepat, praktis, terjangkau, aman, serta memberikan kenyamanan bagi seluruh lapisan masyarakat (Prasetyo et al., 2022).

Pelayanan yang berkualitas terbukti berperan penting dalam membentuk kepuasan masyarakat sebagai pengguna layanan. Hal ini dapat dilihat dari lima dimensi utama, yakni keandalan, daya tanggap, jaminan, empati, dan bukti fisik, yang masing-masing memberikan kontribusi nyata terhadap bagaimana pengguna menilai dan merasakan kualitas layanan yang diterima (Diastuti & Ayyustia, 2025).

Algoritma K-Means membangun kelompok data melalui serangkaian proses yang dilakukan secara berulang atau iteratif. Dalam metode K-Means, sebanyak K pola dipilih secara acak untuk dijadikan titik awal atau centroid. Nilai K yang digunakan sebagai pusat awal cluster dihitung dengan memanfaatkan rumus jarak geometris, yang berfungsi untuk menentukan jarak terdekat antara titik centroid dengan data atau objek yang ada. (Kinanti et al., 2024). Di antara berbagai algoritma clustering yang ada, K-Means dikenal sebagai algoritma dengan tingkat kesederhanaan yang paling tinggi, menawarkan kemudahan implementasi, kecepatan komputasi, dan fleksibilitas terhadap berbagai kebutuhan analisis (Oop Sofiyah et al., 2023).

Sejumlah penelitian telah menerapkan K-Means Clustering untuk analisis pengelompokan data. (Kaligis & Yulianto, 2022) membuktikan bahwa K-Means merupakan algoritma clustering paling optimal dibandingkan K-Medoids dan X-Means dalam pengelompokan data kinerja pegawai berdasarkan perbandingan nilai DBI. (Wisanta & Marlim, 2021) menerapkan K-Means menggunakan Euclidean Distance dengan $k=3$ untuk mengelompokkan tingkat kepuasan layanan di MPP Pekanbaru dan mencapai konvergensi pada iterasi ketiga. (Gunadi & Wirawan, 2025) membandingkan Euclidean, Manhattan, dan Chebyshev Distance menggunakan DBI dan Average Within Centroid Distance, dengan Euclidean Distance menghasilkan nilai terbaik berupa DBI sebesar 1,202 dan Average Within Centroid Distance sebesar 16,082 serta konvergensi dalam 4 iterasi. Sebaliknya, (Farida & Lubis, 2025) yang membandingkan Manhattan, Euclidean, dan Canberra Distance menggunakan Silhouette Score justru menemukan Manhattan Distance sebagai metode paling optimal dengan nilai tertinggi sebesar 0,321, sementara nilai DBI ketiga metode menghasilkan angka yang sama yaitu 1,014. (Permata et al., 2025) menunjukkan hasil yang berbeda, dimana Canberra Distance unggul dengan nilai Average Within Cluster Distance terkecil sebesar 0,0880 dibandingkan Euclidean (0,2482), Manhattan (0,3158), dan Mahalanobis Distance (0,3446). (Hidayati et al., 2021) yang membandingkan Euclidean, Manhattan, Chebyshev, dan Minkowski Distance menemukan bahwa Chebyshev Distance menunjukkan performa yang paling stabil pada berbagai variasi jumlah atribut dan cluster. Adapun (Khoirala et al., 2025) membuktikan keunggulan Cosine Similarity yang menghasilkan Silhouette Score sebesar 0,6918, Calinski-Harabasz Index sebesar 3016,65, dan DBI sebesar 0,9508, jauh melampaui Euclidean Distance dengan nilai masing-masing sebesar 0,2364, 1257,17, dan 1,3682. Dalam penentuan jumlah cluster, Elbow Method merupakan metode heuristik yang banyak digunakan, dimana titik siku yang paling mendekati sudut 90° menunjukkan jumlah cluster paling optimal (Herdiana et al., 2025).

Meskipun berbagai penelitian telah membandingkan performa metrik jarak pada algoritma K-Means, hasil yang diperoleh masih menunjukkan temuan yang berbeda-beda. Beberapa penelitian melaporkan bahwa Euclidean Distance menghasilkan performa terbaik, sementara penelitian lain menunjukkan keunggulan Manhattan Distance, Canberra Distance, maupun Cosine Similarity. Perbedaan hasil tersebut mengindikasikan bahwa performa suatu metrik jarak sangat dipengaruhi oleh karakteristik data yang digunakan. Namun, kajian mengenai pengaruh pemilihan metrik jarak pada data kepuasan masyarakat berbasis skala Likert masih relatif terbatas, khususnya pada konteks layanan publik. Selain itu, penelitian terdahulu umumnya hanya berfokus pada satu metrik evaluasi sehingga belum memberikan gambaran yang menyeluruh mengenai kualitas cluster yang dihasilkan. Oleh karena itu, diperlukan penelitian yang membandingkan beberapa metrik jarak menggunakan lebih dari satu pendekatan evaluasi untuk memperoleh pemahaman yang lebih komprehensif mengenai kualitas hasil clustering.

Dalam konteks Mal Pelayanan Publik (MPP) Pekanbaru, penelitian sebelumnya masih berfokus pada penggunaan Euclidean Distance, sehingga belum diketahui bagaimana performa metrik jarak lainnya ketika diterapkan pada data kepuasan masyarakat dengan karakteristik yang sama. Padahal, perbedaan metrik jarak dapat menghasilkan struktur cluster yang berbeda dan berpotensi memengaruhi interpretasi hasil analisis.

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk menganalisis dan membandingkan performa lima metrik jarak, yaitu Euclidean Distance, Manhattan Distance, Canberra Distance, Chebyshev Distance, dan Cosine Similarity, pada algoritma K-Means Clustering dalam mengelompokkan tingkat kepuasan masyarakat terhadap layanan di Mal Pelayanan Publik (MPP) Pekanbaru. Evaluasi kualitas cluster dilakukan secara komprehensif dengan menggunakan Davies-Bouldin Index dan Silhouette Score secara bersamaan, sehingga dapat diperoleh gambaran yang lebih menyeluruh mengenai kekompakan dan pemisahan antar cluster dari masing-masing metrik jarak yang diuji. Penelitian ini memberikan gambaran mengenai karakteristik dan performa masing-masing metrik jarak pada data kepuasan masyarakat berbasis skala Likert serta menunjukkan pengaruh penggunaan metrik evaluasi yang berbeda

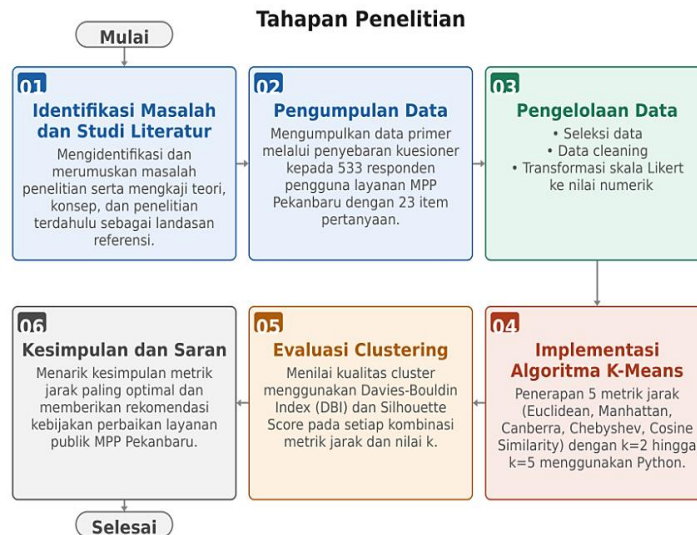
terhadap hasil clustering. Temuan yang diperoleh dapat menjadi bahan pertimbangan dalam pemilihan metrik jarak pada analisis data kepuasan masyarakat dengan karakteristik yang serupa.

2. METODOLOGI PENELITIAN

Berikut ini merupakan uraian dari metode yang digunakan dalam penelitian ini.

2.1 Tahapan Penelitian

Dalam sebuah penelitian, terdapat alur kerja yang tersusun secara sistematis sebagai panduan bagi peneliti dalam menjalankan proses penyelidikan ilmiah secara terorganisir. Tahapan ini diilustrasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian Analisis Perbandingan Jarak

Gambar 1 mengilustrasikan alur tahapan yang dilalui dalam pelaksanaan penelitian ini, yang secara berurutan meliputi tahap identifikasi masalah, pengumpulan data, seleksi data, pra-pemrosesan data, penerapan algoritma K-Means, hingga tahap evaluasi hasil pengelompokan.

2.2 Pengumpulan Data

Pengumpulan data bertujuan untuk memperoleh informasi yang relevan dengan topik penelitian. Teknik pengambilan sampel yang digunakan adalah convenience sampling, yaitu teknik non-probability sampling yang memilih responden berdasarkan kemudahan akses dan ketersediaan responden untuk berpartisipasi dalam penelitian (Golzar & Noor, 2022). Sebelum dilakukan proses clustering, instrumen penelitian diuji validitas dan reliabilitasnya guna memastikan kelayakan data yang digunakan dalam analisis. Uji validitas dilakukan dengan membandingkan nilai r-hitung setiap item pertanyaan terhadap nilai r-tabel seperti pada Tabel 1 berikut ini.

Tabel. 1 Hasil Uji Validitas dan Reliabilitas Instrumen

Pengujian	Hasil
Jumlah Responden	532
Jumlah Item	23
Rentang r-hitung	0,405 – 0,769
r-tabel ($\alpha = 0,05$)	0,361
Validitas Item	Seluruh item valid
Cronbach's Alpha	0,944
Interpretasi Reliabilitas	Sangat reliabel

Berdasarkan Tabel 1, seluruh 23 item kuesioner dinyatakan valid karena nilai r-hitung masing-masing item lebih besar dari nilai r-tabel sebesar 0,361, dengan rentang nilai r-hitung antara 0,405 hingga 0,769. Hasil uji reliabilitas menunjukkan nilai Cronbach's Alpha sebesar 0,944, yang melebihi batas minimum 0,70, sehingga instrumen penelitian dinyatakan reliabel dan layak digunakan sebagai alat pengumpulan data dalam penelitian ini. Dalam penelitian ini, responden merupakan masyarakat yang telah menggunakan layanan di Mal Pelayanan Publik (MPP) Pekanbaru sehingga dapat memberikan informasi yang sesuai dengan tujuan penelitian. Instrumen pengumpulan data menggunakan kuesioner yang disusun dengan menggabungkan metode SERVQUAL yang mencakup 5 dimensi layanan (Supawanhar et al., 2024) dan 9 indikator berdasarkan PERMEN PAN-RB Nomor 14 Tahun 2017.

Berikut 5 dimensi SERVQUAL sebagai berikut:



1. Tangibles

Bukti fisik ialah segala aspek yang dapat diamati secara kasat mata oleh pelanggan, mulai dari kondisi gedung, kelengkapan fasilitas, peralatan yang dioperasikan, hingga penampilan dan kerapian para staf, semuanya termasuk ke dalam dimensi bukti fisik. Keseluruhan tampilan nyata yang ditampilkan oleh suatu perusahaan tersebut memiliki dampak yang cukup signifikan terhadap cara pelanggan menilai dan mengevaluasi mutu layanan yang mereka peroleh.

2. Reliability

Keandalan berkaitan dengan kemampuan perusahaan dalam memenuhi janji layanannya secara konsisten, akurat, dan tepat sasaran. Aspek ini tidak hanya berlaku pada persoalan besar, karena hal-hal kecil pun turut memengaruhi penilaian pelanggan terhadap perusahaan.

3. Responsiveness

Daya tanggap mencerminkan kesediaan dan komitmen perusahaan untuk memberikan pelayanan secara cepat dan tepat waktu. Dimensi ini tidak sekadar berbicara tentang kecepatan layanan, melainkan juga kesiapan perusahaan dan karyawannya dalam merespons setiap kebutuhan pelanggan.

4. Assurance

Keyakinan merupakan kemampuan perusahaan dalam membangun kepercayaan pelanggan melalui pengetahuan yang memadai, sikap sopan santun, serta kompetensi karyawan dalam meyakinkan pelanggan bahwa mereka berada di tangan yang tepat.

5. Empathy

Empati tercermin dari kemampuan karyawan dalam berkomunikasi secara efektif dan jelas kepada pelanggan terkait layanan yang tersedia. Komunikasi yang baik dan penuh perhatian akan memberikan kesan positif yang berpengaruh pada penilaian pelanggan terhadap perusahaan.

Berikut ini 9 indikator berdasarkan PERMEN PAN dan RB Nomor 14 tahun 2017 antara lain:

1. Persyaratan

Persyaratan adalah sejumlah ketentuan yang harus dipenuhi oleh setiap pemohon layanan, mencakup aspek yang bersifat teknis maupun yang berkaitan dengan kelengkapan administrasi dalam proses pengurusan suatu jenis layanan tertentu.

2. Prosedur

Prosedur merupakan rangkaian alur dan mekanisme yang telah distandarisasi dan ditetapkan secara resmi dalam penyelenggaraan layanan, yang mengikat dan berlaku bagi seluruh pihak yang terlibat, baik penyelenggara maupun penerima layanan, termasuk pula dalam proses penyampaian keluhan atau pengaduan.

3. Waktu Pelayanan

Waktu pelayanan adalah durasi yang diperlukan untuk menyelesaikan suatu proses layanan dari awal sampai selesai pada setiap jenis layanan yang tersedia.

4. Biaya/Tarif

Biaya atau tarif adalah pungutan yang dibebankan kepada pengguna layanan sebagai syarat untuk memperoleh layanan dari penyelenggara, yang nominalnya ditetapkan atas dasar kesepakatan bersama.

5. Produk Spesifikasi Jenis Pelayanan

Produk spesifikasi jenis pelayanan adalah hasil akhir yang diperoleh penerima layanan dari suatu proses pelayanan yang mengacu pada aturan dan standar yang berlaku, di mana setiap jenis layanan menghasilkan output yang berbeda sesuai karakteristiknya.

6. Kompetensi Pelaksana

Kompetensi pelaksana adalah kemampuan yang wajib dimiliki petugas layanan, mulai dari pengetahuan dan keahlian hingga keterampilan serta pengalaman di bidang tugasnya.

7. Perilaku Pelaksana

Perilaku pelaksana merupakan gambaran nyata dari cara bersikap dan nilai-nilai etika yang ditunjukkan oleh petugas dalam mengemban tanggung jawabnya ketika berinteraksi dan memberikan layanan kepada masyarakat yang membutuhkan.

8. Sarana dan Prasarana

Sarana dan prasarana mencakup ketersediaan fasilitas yang cukup untuk mendukung pelaksanaan pelayanan agar berjalan sesuai standar yang berlaku.

9. Penanganan Pengaduan, Saran dan Masukan

Penanganan pengaduan, saran, dan masukan adalah mekanisme yang ditetapkan dalam mengelola dan menindaklanjuti setiap laporan, keluhan, maupun usulan yang disampaikan oleh masyarakat.

Kuesioner yang digunakan memiliki 23 pertanyaan pada dimensi Reliability terdapat sebanyak 6 pertanyaan, dimensi Responsiveness 4 pertanyaan, dimensi Assurance 5 pertanyaan, dimensi Empathy 4 pertanyaan dan dimensi Tangibles sebanyak 4 pertanyaan terlihat pada Tabel 2 berikut.

Tabel 2. Instrumen Pertanyaan Pengukuran Kepuasan Masyarakat

No	Dimensi Kepuasan Layanan	Variabel	Pertanyaan Kuisisioner
1	<i>Reliability</i>	Q1	Persyaratan pelayanan disampaikan petugas secara jelas dan mudah dipahami.
		Q2	Prosedur pelayanan dilaksanakan sesuai ketentuan yang berlaku.
		Q3	Pelayanan dilakukan petugas dengan cepat dan akurat.
		Q4	Petugas teliti dalam memeriksa hasil produk dari jenis pelayanan yang diajukan.
		Q5	Petugas memberikan layanan secara tepat dan benar.
		Q6	Pengaduan ditangani oleh petugas dengan bijak.
		Q7	Petugas berinisiatif ketika saya kebingungan terhadap alur prosedur layanan.
2	<i>Responsiveness</i>	Q8	Pelayanan tepat waktu.
		Q9	Petugas melayani dengan sigap dan cepat.
		Q10	Pengaduan ditindaklanjuti dengan cepat.
		Q11	Petugas menyampaikan kejelasan mekanisme dan prosedur dengan percaya diri.
3	<i>Assurance</i>	Q12	Saya merasa yakin terhadap hasil pelayanan karena petugas melayani dengan baik dan profesional.
		Q13	Petugas memiliki kemampuan yang memadai dalam memberikan pelayanan.
		Q14	Petugas bersikap sopan dan ramah.
		Q15	Petugas memberi penjelasan yang lengkap tentang tindak lanjut pengaduan saya.
		Q16	Petugas mendengarkan kebutuhan saya dengan sungguh-sungguh saat memberikan pelayanan.
4	<i>Empathy</i>	Q17	Petugas menunjukkan kepedulian dengan berusaha membantu mencari solusi atas permasalahan yang saya hadapi.
		Q18	Petugas mendengarkan, mencatat, dan menanggapi pengaduan saya.
		Q19	Petugas menanyakan kembali kebutuhan saya untuk memastikan pengaduan atau saran saya dipahami dengan benar.
5	<i>Tangibles</i>	Q20	Formulir yang disediakan untuk mengisi pengajuan mudah dipahami.
		Q21	Hasil produk jenis layanan sesuai kebutuhan yang sudah diajukan.
		Q22	Fasilitas yang disediakan seperti ruang tunggu dan loket antrian berfungsi dengan baik.
		Q23	Fasilitas pelayanan membuat saya merasa nyaman saat berada di area pelayanan

Sebelum disebarkan, instrumen kuesioner diuji validitasnya oleh pakar yang memiliki kompetensi di bidang pelayanan publik untuk memastikan kesesuaian setiap item pertanyaan dengan tujuan penelitian. Penyebaran kuesioner dilakukan melalui dua metode, yaitu secara online menggunakan Google Form dan secara offline melalui pembagian kuesioner secara langsung kepada responden di lokasi MPP Pekanbaru.

2.3 Seleksi Data

Seleksi data merupakan tahapan preprocessing yang dilakukan untuk memilih data yang relevan dengan tujuan penelitian serta menghilangkan data yang tidak diperlukan sehingga kualitas analisis dapat ditingkatkan (Lopez-miguel, 2021). Pada penelitian ini, data yang digunakan adalah jawaban responden terhadap 23 item kuesioner yang telah memenuhi kriteria kelengkapan dan kevalidan pengisian.

2.4 Pre-prosesing

Tahap data preprocessing mencakup serangkaian proses persiapan data yang meliputi pembersihan data dari nilai yang tidak valid, transformasi data ke dalam format yang sesuai, serta pengaturan dan penyusunan data sehingga dapat siap digunakan untuk keperluan analisis pada tahap berikutnya. Setiap tahap tersebut bertujuan untuk memberikan pemahaman yang lebih mendalam terhadap tingkat kepuasan masyarakat terhadap layanan di MPP Pekanbaru.

a. Data Cleaning

Data *cleaning* merupakan proses membersihkan data dari duplikasi, *missing value*, serta data yang tidak relevan yang dapat disebabkan oleh kesalahan teknis maupun pengisian kuesioner yang tidak valid. Tahap ini dilakukan untuk meningkatkan kualitas data sehingga data yang digunakan dalam proses analisis lebih akurat, konsisten, dan sesuai dengan tujuan penelitian.

b. Data Transformation

Data *transformation* adalah suatu proses pengolahan data yang bertujuan untuk mengubah data yang ada agar sesuai dengan kebutuhan analisis yang akan dilakukan. Proses ini mengkonversikan data ke dalam format yang lebih mudah untuk diolah pada tahap selanjutnya, khususnya dalam bentuk numerik atau angka. Dalam pelaksanaannya, transformasi data dilakukan dengan mengubah data yang semula berupa huruf atau teks menjadi representasi angka sebagaimana yang ditampilkan pada Tabel 3 di bawah ini.



Tabel 3. Pembobotan Jawaban Kuesioner

Jawaban	Inisialisasi Jawaban
Sangat Tidak Setuju	1
Tidak Setuju	2
Netral	3
Setuju	4
Sangat Setuju	5

2.5 Data Mining

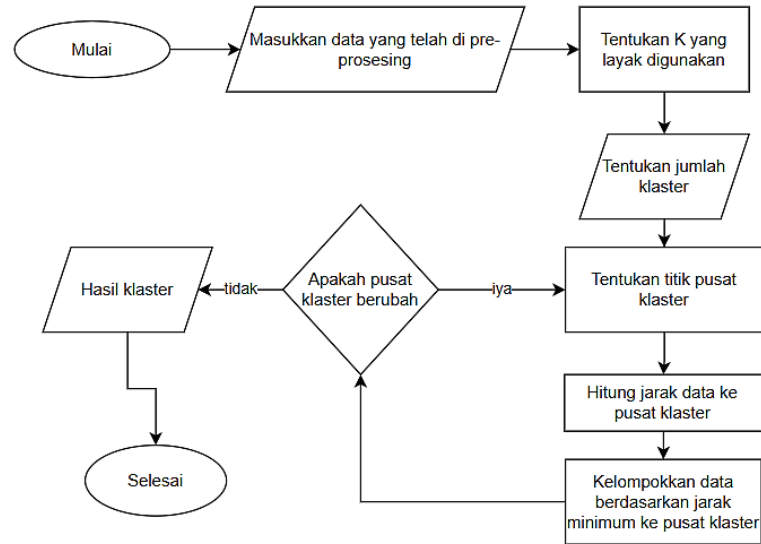
Data mining merupakan teknik yang memungkinkan pengguna untuk memproses dan menganalisis kumpulan data berukuran besar secara efisien dalam waktu yang cukup singkat (Baha'uddin & Fatah, 2024).

2.6 Clustering

Clustering merupakan suatu proses pengelompokan record, observasi, maupun objek tertentu ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik yang dimiliki. Pada proses ini, objek yang memiliki karakteristik serupa akan ditempatkan dalam kelompok yang sama, sedangkan objek yang memiliki karakteristik berbeda akan ditempatkan pada kelompok yang berbeda. Tujuan utama clustering adalah untuk menemukan pola atau struktur yang tersembunyi di dalam data sehingga data yang memiliki jumlah besar dapat lebih mudah dipahami dan dianalisis. Melalui proses pengelompokan tersebut, setiap kelompok yang terbentuk dapat merepresentasikan karakteristik tertentu yang membedakannya dari kelompok lain, sehingga informasi yang dihasilkan dapat digunakan sebagai dasar dalam pengambilan keputusan maupun penyusunan strategi berdasarkan pola yang ditemukan dalam data (Agung Nugraha et al., 2022). Keunggulan metode ini terletak pada kemampuannya dalam menggali dan mengidentifikasi pola-pola yang tidak tampak secara kasat mata di dalam kumpulan data berukuran besar maupun yang bersifat kompleks. Dalam konteks pemasaran misalnya, clustering dimanfaatkan sebagai sarana untuk mengklasifikasikan konsumen berdasarkan kecenderungan dan kebiasaan mereka dalam berbelanja. Konsumen yang kerap melakukan pembelian terhadap produk-produk dengan kategori serupa akan dikelompokkan ke dalam satu segmen yang sama, sehingga memungkinkan perusahaan untuk merancang dan mengimplementasikan strategi promosi yang lebih terarah dan berdampak optimal. Lebih dari itu, penerapan clustering juga merambah ke sejumlah bidang lainnya, seperti dunia kesehatan, sektor pendidikan, bahkan keamanan siber, di antaranya untuk mengidentifikasi anomali maupun pola perilaku yang tidak lazim dalam sebuah sistem jaringan komputer. Kemampuan metode ini dalam menyusun dan mengorganisasikan data secara mandiri tanpa bergantung pada pelabelan atau kategorisasi awal menjadikannya sebagai perangkat analisis yang sangat diandalkan, baik dalam tahap eksplorasi data maupun dalam proses pengambilan keputusan yang dilandasi oleh data. Pada akhirnya, clustering memberikan kontribusi nyata dalam memudahkan interpretasi terhadap data dalam jumlah besar dengan cara menyederhanakan dan menyusunnya ke dalam kelompok-kelompok yang lebih mudah dipahami, sehingga kegiatan analisis data dapat dilaksanakan secara lebih efisien dan memiliki arah yang jelas (Ikhsan et al., 2025).

2.7 K-Means

Algoritma K-Means pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976 dan hingga kini tetap menjadi salah satu metode clustering non-hierarki yang paling banyak digunakan dalam analisis data. Algoritma ini bekerja dengan membagi sekumpulan objek data ke dalam k kelompok yang jumlahnya telah ditentukan sebelum proses pengelompokan dimulai. Setiap kelompok dibentuk berdasarkan tingkat kemiripan antar data, sehingga objek yang memiliki karakteristik serupa akan berada dalam cluster yang sama, sedangkan objek yang memiliki karakteristik berbeda akan berada pada cluster yang berbeda (Yusuf et al., 2022). Algoritma K-Means clustering merupakan salah satu teknik pengelompokan data yang bersifat non-hierarki, yang dirancang untuk menghimpun data-data dengan kesamaan tertentu ke dalam satu wadah kelompok yang sama. Derajat kesamaan antara satu data dengan data lainnya diukur dengan mengacu pada nilai jarak yang terdapat di antara setiap objek yang dibandingkan. Objek-objek yang memiliki nilai jarak paling kecil atau paling berdekatan satu sama lain akan disatukan ke dalam kelompok yang identik. Kondisi kedekatan jarak tersebut mengindikasikan bahwa objek-objek yang bersangkutan memiliki keserupaan karakteristik yang cukup tinggi apabila dibandingkan satu dengan yang lainnya (Iqbal et al., 2023). Di antara berbagai algoritma clustering yang ada, K-Means dikenal sebagai algoritma dengan tingkat kesederhanaan yang paling tinggi. Algoritma ini menawarkan sejumlah kelebihan yang membuatnya unggul, di antaranya adalah kemudahannya dalam hal implementasi dan pengoperasian, kecepatan proses komputasi yang tidak membutuhkan waktu lama, serta kemampuannya yang fleksibel untuk disesuaikan dengan beragam kebutuhan analisis. Berbagai keunggulan tersebut menjadikan K-Means sebagai salah satu pilihan utama yang paling dominan digunakan dalam menyelesaikan berbagai permasalahan analisis pada bidang data mining (Oop Sofiyah et al., 2023). Berikut proses K-Means yang dapat dilakukan dengan langkah-langkah seperti pada Gambar 2 di bawah ini.



Gambar 2. Flowchart K-Means

Berdasarkan dari Gambar 2 *Flowchart* K-Means maka dijelaskan:

- Tentukan K yang layak digunakan (Pengujian k yang layak digunakan dari k=2 hingga k=10).
- Menentukan nilai k menjadi kluster yang di uji.
- Menentukan nilai k menjadi centroid yang dapat dihasilkan secara random.
- Menghitung jarak data ke masing-masing centroid.
- Kelompokkan setiap data berdasarkan jarak dengan centroid terdekat.
- Menentukan centroid baru, maka persamaan untuk menghitung centroid baru digunakan Persamaan 1.

$$C_j = \left(\frac{1}{n}\right) \sum_{i=1}^n x = i \quad (1)$$

Keterangan: C_j merupakan centroid baru pada kluster ke-j, n adalah jumlah data dalam kluster, dan x merupakan data ke-i dalam kluster.

- Mengulangi kembali ke langkah 3 apabila pusat kluster berubah, maka selesai.

Metode perhitungan jarak yang digunakan pada penelitian ini antara lain:

- Euclidean Distance*

Dalam analisis data, salah satu cara mengukur jarak yang paling sering dipakai adalah *Euclidean distance*. Metode ini digunakan untuk menentukan jarak antara dua titik data, misalnya x dan y, dalam ruang berdimensi-d, yang perhitungannya dapat dilihat pada Persamaan 2.

$$d = \sqrt{\{(x_2 - x_1)^2 + (y_2 - y_1)^2\}} \quad (2)$$

Keterangan: d merupakan jarak euclidean antara dua titik, x adalah data, dan y merupakan centroid.

- Manhattan Distance*

Manhattan distance dilakukan dengan memanfaatkan dua titik koordinat dari objek yang dibandingkan. Adapun rumus yang digunakan dalam perhitungan *Manhattan distance* untuk membandingkan kedua metode perhitungan jarak yang diterapkan adalah sebagai berikut:

$$d = \sum_{i=1}^n |x_i - y_i| \quad (3)$$

Keterangan: d merupakan jarak manhattan antara dua titik/objek, n adalah jumlah atribut/variable/dimensi yang dibandingkan, dan $|x_i - y_i|$ selisih absolut antara atribut dari dua objek.

- Canberra Distance*

Canberra Distance menekankan nilai-nilai kecil dengan cara melakukan normalisasi pada setiap dimensi, sehingga metode ini menjadi sensitif terhadap variasi kecil dan cocok digunakan untuk data yang bersifat jarang (sparse).

$$d_{C(x,y)} = \frac{\sum_{i=1}^n (|x_i - y_i|)}{|x_i| + |y_i|} \quad (4)$$

Keterangan: $d_{C(x,y)}$ merupakan jarak Canberra antara titik x dan y, $|x_i - y_i|$ selisih absolut pada dimensi ke-i, $|x_i| + |y_i|$ merupakan jumlah nilai absolut dari dimensi ke-i pada x dan y, dan n adalah jumlah dimensi.

- Chebyshev Distance*

Chebyshev Distance dikenal pula dengan sebutan Maximum distance, hal ini dikarenakan dalam penerapannya metode ini bekerja dengan cara mencari nilai maksimum dari selisih jarak antara dua titik data x dan y dalam ruang berdimensi d.



$$d_{\{\infty\}}(x,y) = \max_{\{i=1,2,\dots,n\}} |x_i - y_i| \quad (5)$$

Keterangan: $d_{\{\infty\}}(x,y)$ merupakan jarak chebyshev antara titik x dan y , \max adalah nilai maksimum dari seluruh dimensi, $|x_i - y_i|$ merupakan selisih absolut pada dimensi ke- i , n adalah jumlah dimensi.

e. *Cosine Similarity*

Cosine Similarity mengukur kemiripan dua vektor dengan melihat sudut yang terbentuk di antara keduanya dalam ruang multidimensi, tanpa mempertimbangkan panjang vektornya.

$$\cos(\theta) = \frac{x \cdot y}{\|x\| \cdot \|y\|} \quad (6)$$

Keterangan: $\cos(\theta)$ merupakan nilai tingkat kemiripan antara dua data/vector, $x \cdot y$ adalah hasil perkalian antara dua vector, $\|x\|$ merupakan panjang (norma) vektor x , $\|y\|$ merupakan panjang (norma) vektor y .

2.8 Evaluasi

Evaluasi dalam penelitian ini bertujuan memastikan bahwa pengelompokan data menggunakan algoritma K-Means telah berjalan dengan baik. Performa cluster dinilai menggunakan Davies-Bouldin Index (DBI) dan Silhouette Score. Davies-Bouldin Index (DBI) dan Silhouette Score merupakan metrik yang sering digunakan untuk mengevaluasi kualitas hasil clustering. Nilai DBI yang lebih kecil menunjukkan kualitas cluster yang lebih baik karena setiap cluster memiliki tingkat kekompakan yang tinggi dan pemisahan yang lebih jelas terhadap cluster lainnya. Sementara itu, Silhouette Score digunakan untuk mengukur tingkat kesesuaian suatu data terhadap cluster tempat data tersebut berada dibandingkan dengan cluster lain yang paling dekat. Nilai Silhouette Score berada pada rentang -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa data telah terkelompok dengan baik, sedangkan nilai yang mendekati 0 mengindikasikan adanya tumpang tindih antarcluster. Oleh karena itu, semakin tinggi nilai Silhouette Score yang diperoleh, semakin baik kualitas hasil clustering yang dihasilkan (Sutirta & Noviandi, 2024). Elbow Method merupakan salah satu teknik yang digunakan untuk menentukan jumlah cluster optimal dalam algoritma K-Means. Metode ini bekerja dengan menghitung nilai inerti atau Within-Cluster Sum of Squares (WCSS) untuk setiap nilai k yang diuji. WCSS mengukur total kuadrat jarak antara setiap data dengan centroid clusternya masing-masing, sehingga dapat menggambarkan tingkat kekompakan cluster yang terbentuk. Semakin kecil nilai WCSS yang diperoleh, semakin dekat jarak antar data dengan centroid cluster, yang menunjukkan bahwa cluster yang terbentuk semakin homogen. Penentuan jumlah cluster dilakukan dengan mengamati pola penurunan nilai WCSS pada setiap nilai k dan mencari titik perubahan yang signifikan atau dikenal sebagai titik elbow (Hartono & Lusiana, 2026).

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian serta pembahasan yang diperoleh dari setiap tahapan penelitian yang telah dilakukan. Hasil yang diperoleh akan dianalisis untuk mengetahui keterkaitannya dengan tujuan penelitian serta untuk memberikan pemahaman yang lebih mendalam mengenai temuan yang dihasilkan. Selain itu, pembahasan dilakukan dengan menginterpretasikan hasil analisis dan menghubungkannya dengan penelitian terdahulu yang relevan.

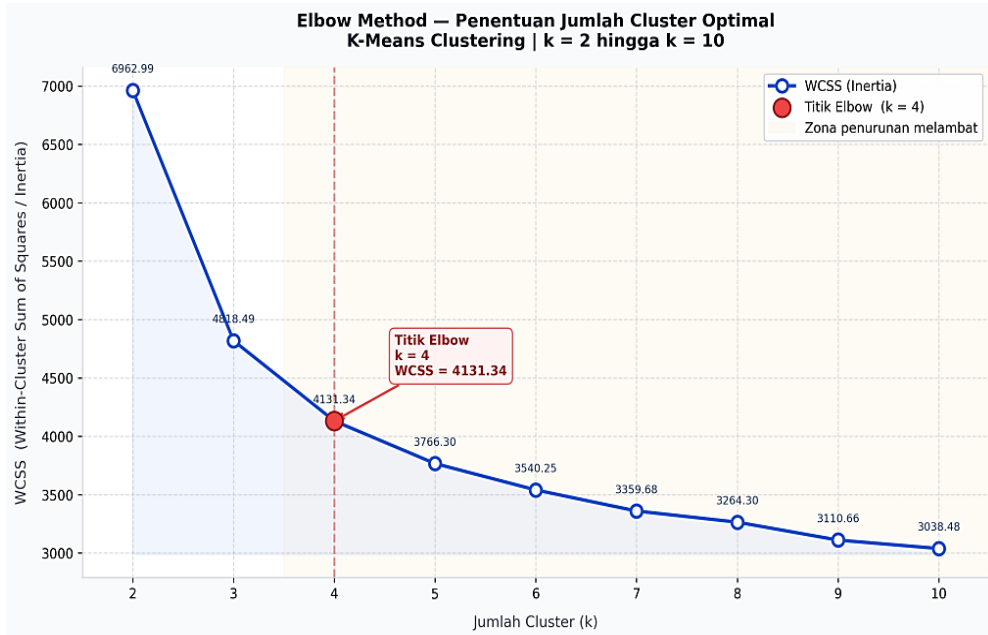
3.1 Hasil Seleksi dan Preprocessing Data

Tahap seleksi data dilakukan dengan memilih data yang benar-benar relevan dan dibutuhkan dalam proses analisis. Data yang digunakan pada penelitian ini merupakan hasil jawaban responden terhadap 23 pertanyaan kuesioner yang mencakup lima dimensi kepuasan layanan, yaitu Reliability, Responsiveness, Assurance, Empathy, dan Tangibles, yang diperoleh dari 533 responden yang pernah menggunakan layanan di Mal Pelayanan Publik (MPP) Pekanbaru. Setelah tahap seleksi, data kemudian melewati proses preprocessing yang terdiri dari dua tahapan, yaitu data cleaning dan data transformation. Pada tahap data cleaning, seluruh data yang terkumpul diperiksa untuk mendeteksi adanya duplikat, missing value, maupun data yang tidak valid akibat kesalahan teknis atau pengisian yang tidak sesuai. Hasil pemeriksaan menunjukkan bahwa ditemukan duplikat data sebanyak 1 responden yang mengisi 2 kali dan tidak ditemukan adanya missing value pada data, sehingga dari seluruh 533 data responden data yang bersih dan layak untuk digunakan sebanyak 532 data responden untuk tahap analisis berikutnya. Selanjutnya pada tahap data transformation, data yang semula berupa teks jawaban responden dikonversikan ke dalam bentuk numerik agar dapat diproses oleh algoritma K-Means Clustering. Proses transformasi dilakukan dengan mengacu pada skala Likert, di mana jawaban "Sangat Tidak Setuju" diinisialisasi dengan nilai 1, "Tidak Setuju" dengan nilai 2, "Ragu-Ragu" dengan nilai 3, "Setuju" dengan nilai 4, dan "Sangat Setuju" dengan nilai 5. Setelah kedua tahapan preprocessing ini selesai dilakukan, data telah berada dalam format numerik yang siap digunakan sebagai masukan pada proses clustering menggunakan algoritma K-Means.

3.2 Penentuan Nilai K Menggunakan Elbow Method

Sebelum melakukan pengujian menggunakan berbagai metrik jarak, terlebih dahulu dilakukan analisis Elbow Method untuk menentukan rentang nilai k yang layak diuji pada proses clustering. Tahap ini bertujuan untuk mengidentifikasi jumlah cluster yang dapat merepresentasikan struktur data secara memadai tanpa menghasilkan jumlah cluster yang berlebihan. Gambar 3 menunjukkan grafik Elbow Method yang memplot nilai Within Cluster Sum of Squares (WCSS)

terhadap jumlah cluster $k=2$ hingga $k=10$. Berdasarkan grafik tersebut, penentuan nilai k dilakukan dengan mengamati titik perubahan penurunan WCSS yang mulai melandai (elbow point). Nilai k yang berada di sekitar titik tersebut kemudian digunakan sebagai acuan dalam proses pengujian dan evaluasi clustering pada tahap selanjutnya.



Gambar 3. Grafik Penentuan K Terbaik

Berdasarkan Gambar 3, grafik Elbow Method menunjukkan pola penurunan WCSS yang tidak sepenuhnya linear, melainkan terdapat perubahan kemiringan kurva yang dapat diidentifikasi secara kuantitatif. Penurunan WCSS paling tajam terjadi dari $k=2$ ke $k=3$ sebesar 30,80% (dari 6.962,99 menjadi 4.818,49), kemudian melambat secara signifikan menjadi 14,26% dari $k=3$ ke $k=4$ (dari 4.818,49 menjadi 4.131,34). Setelah $k=4$, laju penurunan WCSS terus berkurang secara konsisten menjadi 8,84% pada $k=5$, 6,00% pada $k=6$, 5,10% pada $k=7$, 2,84% pada $k=8$, 4,71% pada $k=9$, dan 2,32% pada $k=10$. Perubahan laju penurunan yang paling drastis terjadi antara $k=3$ ke $k=4$ menuju $k=5$, di mana rasio penurunan turun dari 14,26% menjadi 8,84%, menandakan bahwa $k=4$ dengan WCSS sebesar 4.131,34 merupakan titik transisi optimal sebelum kurva memasuki zona penurunan yang stabil dan melambat.

Perlu dipahami bahwa karakteristik kurva yang terlihat lebih gradual dibandingkan data kontinu merupakan fenomena yang wajar dan umum terjadi pada data berbasis skala Likert dengan rentang nilai terbatas (1–5). Data ordinal berskala sempit secara inheren menghasilkan variasi antar kluster yang lebih kecil, sehingga penurunan WCSS cenderung tidak membentuk sudut siku-siku yang tajam secara visual. Oleh karena itu, penentuan titik elbow pada data jenis ini lebih tepat dilakukan melalui analisis persentase perubahan laju penurunan WCSS secara kuantitatif daripada hanya mengandalkan inspeksi visual grafik semata. Berdasarkan analisis tersebut, pengujian pada rentang $k=2$ hingga $k=5$ telah mencakup titik elbow dan zona transisi yang relevan, sehingga dapat dipertanggungjawabkan secara metodologis.

3.3 Penerapan Algoritma K-Means menggunakan Python

Implementasi K-Means dilakukan menggunakan bahasa pemrograman Python. Proses algoritma K-Means diawali dengan tahap inisialisasi, yaitu menentukan nilai k dan memilih k titik secara acak sebagai centroid awal. Selanjutnya dilakukan penghitungan jarak setiap data terhadap seluruh centroid menggunakan metrik jarak yang diuji, yaitu Euclidean Distance, Manhattan Distance, Canberra Distance, Chebyshev Distance, atau Cosine Similarity. Berdasarkan hasil perhitungan jarak tersebut, setiap data kemudian ditetapkan masuk ke dalam kluster dengan centroid terdekat. Setelah seluruh data terpenetapan ke dalam klasternya masing-masing, posisi centroid diperbarui dengan menghitung ulang rata-rata seluruh data dalam setiap kluster. Proses penghitungan jarak, penetapan kluster, dan pembaruan centroid ini diulang secara iteratif hingga tidak ada perubahan keanggotaan kluster atau kondisi konvergensi tercapai. Pengujian dilakukan pada $k=2$ hingga $k=5$, dan hasil evaluasi menggunakan DBI dan Silhouette Score ditampilkan pada Tabel 4 berikut.

Tabel 4. Hasil Evaluasi Setiap K

Metode	K	DBI	Silhouette Score
Euclidean	2	0.8397	0.4837
Manhattan	2	0.8144	0.4798
Canberra	2	0.8896	0.467
Chebyshev	2	1.2267	0.4493



Metode	K	DBI	Silhouette Score
Cosine	2	0.8377	0.4839
Euclidean	3	0.9821	0.5088
Manhattan	3	0.9087	0.5072
Canberra	3	1.0803	0.4991
Chebyshev	3	1.5089	0.1853
Cosine	3	1.7044	0.3589
Euclidean	4	1.034	0.3812
Manhattan	4	1.0012	0.3841
Canberra	4	1.1226	0.3877
Chebyshev	4	1.3459	0.1948
Cosine	4	1.5627	0.2634
Euclidean	5	1.2863	0.3557
Manhattan	5	1.0812	0.3446
Canberra	5	1.2547	0.3872
Chebyshev	5	1.3284	0.2578
Cosine	5	1.5693	0.2795

Berdasarkan Tabel 4, Manhattan Distance dengan $k=2$ menghasilkan nilai DBI terendah sebesar 0,8144, sedangkan Euclidean Distance dengan $k=3$ menghasilkan Silhouette Score tertinggi sebesar 0,5088. Perbedaan hasil yang diperoleh dari evaluasi DBI dan Silhouette Score menunjukkan bahwa kedua metode tersebut menilai kualitas clustering dari sudut pandang yang berbeda. Pada penelitian ini, Manhattan Distance dengan $k=2$ menghasilkan nilai DBI terbaik, yang menunjukkan bahwa responden dalam setiap cluster memiliki karakteristik kepuasan yang relatif serupa. Sementara itu, Euclidean Distance dengan $k=3$ menghasilkan nilai Silhouette Score tertinggi, yang menunjukkan pemisahan antarcluster yang lebih jelas. Hasil ini mengindikasikan bahwa tidak terdapat satu metrik jarak yang selalu memberikan hasil terbaik pada seluruh kriteria evaluasi. Oleh karena itu, pemilihan metrik jarak perlu disesuaikan dengan tujuan dan karakteristik data yang dianalisis.

3.4 Hasil Pengujian

Hasil yang didapat dari dataset ditampilkan pada tahap ini menggunakan algoritma K-Means, seperti dari Tabel 5 berikut.

Tabel 5. Keanggotaan Evaluasi DBI Sesuai Kluster Terbaik

Metode	Kluster	Jumlah
Euclidean	Kluster 0	249
	Kluster 1	283
Manhattan	Kluster 0	266
	Kluster 1	266
Canberra	Kluster 0	257
	Kluster 1	275
Chebyshev	Kluster 0	88
	Kluster 1	444
Cosine	Kluster 0	250
	Kluster 1	282

Berdasarkan hasil pada Tabel 5, metode Manhattan Distance menghasilkan distribusi kluster yang paling seimbang dengan jumlah responden yang sama pada Kluster 0 (Puas) dan Kluster 1 (Sangat Puas), masing-masing sebanyak 266 responden. Distribusi yang relatif seimbang juga ditunjukkan oleh metode Canberra, Cosine Similarity, dan Euclidean Distance dengan selisih jumlah anggota kluster yang tidak terlalu besar. Kondisi ini menunjukkan bahwa keempat metode tersebut mampu membagi data ke dalam kelompok yang relatif proporsional. Sebaliknya, metode Chebyshev Distance menghasilkan distribusi kluster yang kurang seimbang, dengan 444 responden (83,5%) berada pada Kluster 1 dan 88 responden (16,5%) berada pada Kluster 0. Ketidakeimbangan tersebut menunjukkan bahwa sebagian besar data terkonsentrasi pada satu kluster sehingga kemampuan metode dalam membedakan karakteristik responden menjadi lebih terbatas. Hasil ini diduga dipengaruhi oleh karakteristik Chebyshev Distance yang hanya mempertimbangkan selisih terbesar pada satu dimensi, sehingga informasi dari dimensi lainnya kurang berperan dalam proses pengelompokan data kepuasan masyarakat yang diukur menggunakan skala Likert.

Tabel 6. Evaluasi Silhouette Score Sesuai Kluster Terbaik

Metode	Kluster	Jumlah
Euclidean	Kluster 0	34
	Kluster 1	269
	Kluster 2	229



Metode	Klaster	Jumlah
Manhattan	Klaster 0	241
	Klaster 1	266
	Klaster 2	25
Canberra	Klaster 0	218
	Klaster 1	269
	Klaster 2	45
Chebyshev	Klaster 0	88
	Klaster 1	444
Cosine	Klaster 0	250
	Klaster 1	282

Berdasarkan hasil pada Tabel 6, metode Euclidean dengan $k=3$ menghasilkan distribusi klaster sebanyak 34 responden pada Klaster 0, 269 responden pada Klaster 1, dan 229 responden pada Klaster 2. Metode Manhattan menghasilkan 241 responden pada Klaster 0, 266 responden pada Klaster 1, dan 25 responden pada Klaster 2. Berdasarkan hasil pada Tabel 5, Manhattan Distance dengan $k=3$ menghasilkan Klaster 2 yang hanya berisi 25 responden (4,7%). Keberadaan klaster kecil ini perlu dianalisis untuk menentukan apakah kelompok tersebut merupakan segmen yang valid atau sekadar penciliran data. Berdasarkan analisis nilai centroid, Klaster 2 memiliki rata-rata skor yang berbeda secara konsisten dari dua klaster lainnya pada seluruh 23 dimensi, bukan hanya pada satu atau dua dimensi tertentu. Pola perbedaan yang konsisten pada seluruh dimensi ini mengindikasikan bahwa 25 responden tersebut memiliki karakteristik penilaian yang kohesif dan berbeda secara sistematis, sehingga bukan merupakan penciliran acak melainkan segmen yang valid. Kelompok ini kemungkinan merepresentasikan masyarakat dengan pengalaman layanan yang sangat spesifik dan berbeda dari mayoritas responden, misalnya pengguna yang mengakses jenis layanan tertentu dengan tingkat ekspektasi yang lebih tinggi. Meskipun proporsinya di bawah 5% sehingga perlu diinterpretasikan dengan kehati-hatian dalam konteks kebijakan, keberadaan klaster kecil ini justru menunjukkan kemampuan Manhattan Distance dalam mendeteksi segmen minoritas yang tersembunyi di dalam data, yang merupakan informasi berharga bagi pengambil kebijakan dalam mengidentifikasi kelompok pengguna layanan dengan karakteristik khusus. Sementara itu, metode Canberra menghasilkan 218 responden pada Klaster 0, 269 responden pada Klaster 1, dan 45 responden pada Klaster 2. Dari ketiga metode tersebut, Manhattan memiliki distribusi paling merata antara dua klaster utama, meskipun Klaster 2 hanya berisi 25 responden. Metode Euclidean dan Canberra juga menunjukkan adanya kelompok kecil pada salah satu klasternya. Adapun metode Chebyshev dan Cosine menggunakan $k=2$ sebagai jumlah klaster terbaik berdasarkan evaluasi Silhouette Score. Chebyshev menghasilkan distribusi yang tidak seimbang dengan 444 responden pada Klaster 1 dan 88 responden pada Klaster 0, sedangkan Cosine menghasilkan distribusi yang lebih proporsional dengan 250 responden pada Klaster 0 dan 282 responden pada Klaster 1.

Penetapan kategori klaster dilakukan berdasarkan nilai rata-rata centroid terhadap skala Likert. Pada hasil clustering $k=2$, Klaster 0 memiliki rata-rata centroid sebesar 3,748 sehingga dikategorikan sebagai Puas, sedangkan Klaster 1 memiliki rata-rata centroid sebesar 4,900 sehingga dikategorikan sebagai Sangat Puas. Pada kelompok Puas, item Q8 dan Q10 memiliki nilai terendah sehingga perlu menjadi perhatian untuk perbaikan layanan, sedangkan Q11 memperoleh nilai tertinggi. Pada kelompok Sangat Puas, hampir seluruh item memperoleh nilai di atas 4,92, namun item Q21 memiliki nilai paling rendah sehingga menunjukkan bahwa aspek tersebut masih belum optimal.

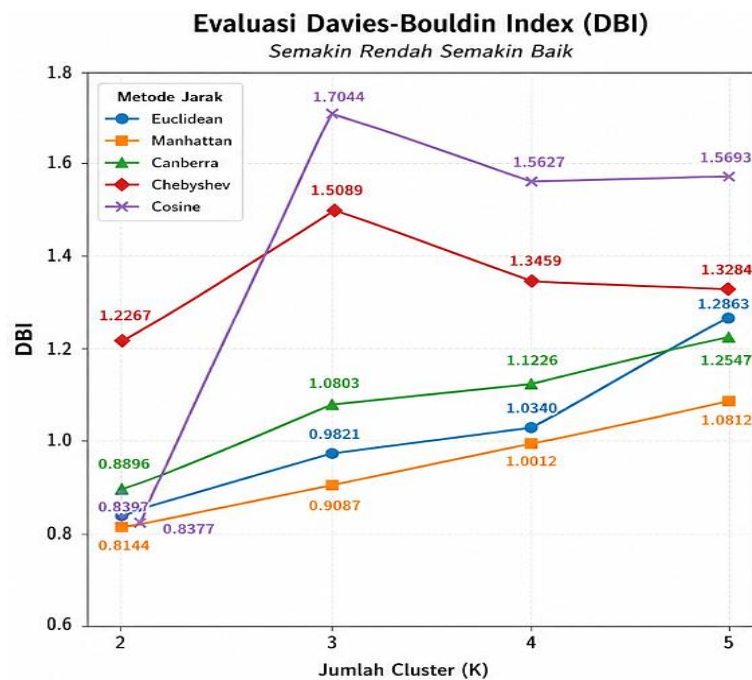
Pada $k=3$ menggunakan metode Euclidean, penetapan kategori didasarkan pada nilai rata-rata centroid masing-masing klaster. Klaster 0 dengan 34 responden (6,4%) memiliki rata-rata centroid sebesar 2,373 yang mendekati skor 2 pada skala Likert, sehingga dikategorikan sebagai kelompok Tidak Puas. Pada klaster ini, item Q6 mencatat nilai terendah sebesar 1,853 dan Q11 mencatat nilai tertinggi sebesar 2,735, namun secara keseluruhan seluruh item masih berada di bawah skor 3, yang mengindikasikan ketidakpuasan menyeluruh terhadap layanan MPP Pekanbaru. Klaster 1 dengan 269 responden (50,6%) memiliki rata-rata centroid sebesar 4,895 sehingga dikategorikan sebagai kelompok Sangat Puas. Hampir seluruh item pada klaster ini memperoleh nilai di atas 4,92, dengan Q10 sebagai nilai tertinggi sebesar 4,970. Satu-satunya pengecualian adalah item Q21 yang hanya memperoleh nilai 3,684, jauh di bawah item lainnya, sehingga menjadi indikator yang perlu mendapat perhatian khusus. Klaster 2 dengan 229 responden (43,0%) memiliki rata-rata centroid sebesar 3,943 yang mendekati skor 4, sehingga dikategorikan sebagai kelompok Puas. Pada klaster ini, item Q8 mencatat nilai terendah sebesar 3,790 dan item Q1 mencatat nilai tertinggi sebesar 4,052, menunjukkan bahwa penilaian responden pada kelompok ini relatif merata dan konsisten di seluruh aspek layanan.

Pada $k=3$ menggunakan metode Manhattan, Klaster 0 dengan 241 responden (45,3%) memiliki rata-rata centroid sebesar 3,915 yang mendekati skor 4 pada skala Likert, sehingga dikategorikan sebagai kelompok Puas. Item Q8 mencatat nilai terendah sebesar 3,763 dan item Q2 mencatat nilai tertinggi sebesar 4,037, menunjukkan distribusi penilaian yang relatif seragam di seluruh aspek layanan. Klaster 1 dengan 266 responden (50,0%) memiliki rata-rata centroid sebesar 4,900 sehingga dikategorikan sebagai kelompok Sangat Puas. Serupa dengan hasil Euclidean, hampir seluruh item memperoleh nilai di atas 4,92 dengan Q10 sebagai nilai tertinggi sebesar 4,974, sementara item Q21 kembali menjadi pengecualian dengan nilai terendah sebesar 3,680. Klaster 2 dengan 25 responden (4,7%) memiliki rata-rata centroid sebesar 2,139 yang mendekati skor 2, sehingga dikategorikan sebagai kelompok Tidak Puas. Item Q6 mencatat nilai terendah sebesar 1,440 dan Q11 mencatat nilai tertinggi sebesar 2,520, namun seluruh item masih berada di bawah skor 3 yang mengindikasikan ketidakpuasan pada semua dimensi layanan.

Secara keseluruhan, mayoritas responden berada pada kategori Puas dan Sangat Puas, sehingga menunjukkan bahwa layanan yang diberikan oleh MPP Pekanbaru secara umum telah memenuhi harapan masyarakat. Temuan ini mengindikasikan bahwa berbagai aspek pelayanan yang meliputi keandalan, daya tanggap, jaminan, empati, dan bukti fisik telah dinilai positif oleh sebagian besar responden. Meskipun demikian, hasil analisis centroid menunjukkan bahwa beberapa indikator masih memperoleh nilai yang relatif lebih rendah dibandingkan indikator lainnya, yaitu Q8, Q10, dan Q21. Kondisi tersebut menunjukkan bahwa masih terdapat aspek layanan yang perlu mendapatkan perhatian untuk meningkatkan kepuasan masyarakat secara lebih merata. Keberadaan kelompok Tidak Puas pada kedua metode, meskipun jumlahnya relatif kecil yaitu 34 responden (6,4%) pada metode Euclidean Distance dan 25 responden (4,7%) pada metode Manhattan Distance, tetap menjadi informasi yang penting untuk diperhatikan. Kelompok ini menunjukkan bahwa masih terdapat sebagian masyarakat yang merasakan kualitas layanan belum sesuai dengan harapan mereka. Oleh karena itu, evaluasi terhadap aspek-aspek layanan yang memperoleh nilai lebih rendah perlu dilakukan secara berkelanjutan agar kualitas pelayanan dapat terus ditingkatkan. Dengan memperhatikan kebutuhan dan pengalaman kelompok responden yang belum puas, pengelola MPP Pekanbaru dapat memperoleh masukan yang lebih spesifik dalam merumuskan strategi peningkatan layanan sehingga kepuasan masyarakat dapat meningkat secara menyeluruh.

3.5 Analisis Hasil Perhitungan Jarak

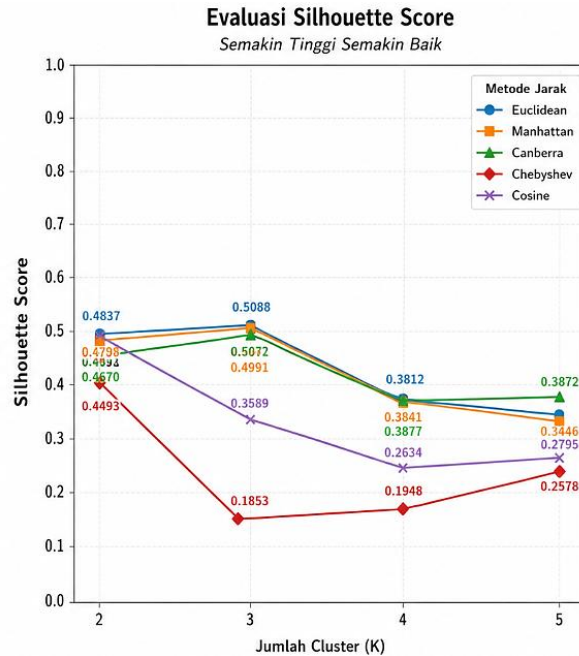
Hasil evaluasi kualitas clustering menggunakan Davies-Bouldin Index (DBI) untuk setiap metrik jarak pada $k=2$ hingga $k=5$. Semakin rendah nilai DBI yang dihasilkan, semakin baik kualitas kluster yang terbentuk karena menunjukkan kluster yang lebih kompak dan terpisah satu sama lain seperti pada Gambar 4 berikut ini.



Gambar 4. Grafik Evaluasi menggunakan DBI

Berdasarkan Gambar 4, hasil evaluasi menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa nilai terbaik diperoleh pada $k=2$. Metode Manhattan Distance menghasilkan nilai DBI terendah sebesar 0,8144, diikuti oleh Cosine Similarity sebesar 0,8377 dan Euclidean Distance sebesar 0,8397. Nilai DBI yang rendah menunjukkan bahwa cluster yang terbentuk memiliki tingkat kekompakan yang tinggi dan pemisahan yang baik antarcluster. Pada $k=3$ hingga $k=5$, seluruh metode cenderung mengalami peningkatan nilai DBI, yang mengindikasikan bahwa kualitas cluster menurun seiring bertambahnya jumlah cluster. Manhattan Distance menunjukkan performa yang relatif stabil dengan nilai DBI yang konsisten lebih rendah dibandingkan sebagian besar metode lainnya pada setiap nilai k yang diuji. Sebaliknya, Chebyshev Distance dan Cosine Similarity menghasilkan nilai DBI yang relatif tinggi, terutama pada $k=3$ hingga $k=5$, yang menunjukkan bahwa cluster yang terbentuk kurang optimal dari sisi kekompakan dan pemisahan antarcluster. Hasil ini menunjukkan bahwa pada data kepuasan masyarakat yang digunakan dalam penelitian ini, penggunaan dua cluster memberikan hasil yang lebih baik dibandingkan jumlah cluster yang lebih banyak. Secara keseluruhan, kombinasi Manhattan Distance dan $k=2$ menghasilkan kualitas clustering terbaik berdasarkan evaluasi Davies-Bouldin Index.

Selain evaluasi menggunakan DBI, kualitas clustering juga diukur menggunakan Silhouette Score untuk setiap metrik jarak pada $k=2$ hingga $k=5$. Semakin tinggi nilai Silhouette Score yang dihasilkan, semakin baik kualitas pemisahan antar kluster yang terbentuk, dengan nilai mendekati 1 mengindikasikan bahwa setiap data telah berada pada kluster yang paling tepat sebagaimana ditampilkan pada Gambar 5 di bawah ini.



Gambar 5. Grafik Evaluasi menggunakan Silhouette Score

Berdasarkan Gambar 5 hasil evaluasi menggunakan Silhouette Score menunjukkan pola yang berbeda dibandingkan DBI. Pada $k=2$, seluruh metode menghasilkan nilai yang cukup tinggi pada rentang 0,44–0,48, yang menunjukkan bahwa cluster yang terbentuk memiliki pemisahan yang baik. Nilai tertinggi pada $k=2$ diperoleh metode Cosine Similarity sebesar 0,4839 dan Euclidean Distance sebesar 0,4837. Pada $k=3$, nilai Silhouette Score meningkat pada beberapa metode, dimana Euclidean Distance menghasilkan nilai tertinggi sebesar 0,5088, diikuti Manhattan Distance sebesar 0,5072 dan Canberra Distance sebesar 0,4991. Namun, metode Chebyshev Distance mengalami penurunan signifikan menjadi 0,1853. Selanjutnya pada $k=4$ dan $k=5$, sebagian besar metode mengalami penurunan nilai Silhouette Score yang menunjukkan bahwa kualitas pemisahan cluster menjadi kurang optimal seiring bertambahnya jumlah cluster.

Berdasarkan hasil evaluasi menggunakan Davies-Bouldin Index (DBI) dan Silhouette Score, jumlah cluster terbaik berada pada $k=2$ dan $k=3$. Pada evaluasi DBI, metode Manhattan Distance dengan $k=2$ menghasilkan nilai terbaik sebesar 0,8144, yang menunjukkan bahwa cluster yang terbentuk lebih kompak dan memiliki pemisahan yang baik. Sementara itu, pada evaluasi Silhouette Score, metode Euclidean Distance dengan $k=3$ menghasilkan nilai tertinggi sebesar 0,5088, yang menunjukkan kualitas pemisahan cluster terbaik dibandingkan metode lainnya. Metode Manhattan Distance terlihat lebih stabil pada evaluasi DBI karena menghasilkan nilai yang relatif rendah pada setiap pengujian. Di sisi lain, metode Euclidean Distance menunjukkan performa terbaik pada Silhouette Score terutama pada $k=3$. Sebaliknya, metode Chebyshev Distance menghasilkan performa paling rendah karena memiliki nilai DBI yang tinggi dan Silhouette Score yang rendah pada sebagian besar pengujian. Secara keseluruhan, metode Manhattan Distance dan Euclidean Distance merupakan metode yang paling optimal dalam penelitian ini karena mampu menghasilkan cluster yang lebih baik dibandingkan metode jarak lainnya.

3.6 Pembahasan

Hasil penelitian ini menunjukkan persamaan dan perbedaan apabila dibandingkan dengan penelitian terdahulu. (Wisanta & Marlim, 2021) yang menerapkan K-Means dengan Euclidean Distance pada data kepuasan layanan MPP Pekanbaru menghasilkan konvergensi pada iterasi ketiga dengan $k=3$, sedangkan penelitian ini menunjukkan bahwa Manhattan Distance menghasilkan nilai DBI terbaik, sementara Euclidean Distance menghasilkan nilai Silhouette Score tertinggi pada dataset yang sama. Hasil tersebut menunjukkan bahwa kualitas cluster yang dihasilkan dipengaruhi oleh metrik evaluasi yang digunakan, sehingga eksplorasi metrik jarak selain Euclidean memberikan perspektif yang lebih luas dalam analisis kepuasan masyarakat di MPP Pekanbaru. Keunggulan Manhattan Distance pada penelitian ini juga sejalan dengan temuan (Farida & Lubis, 2025) yang menemukan Manhattan Distance sebagai metrik terbaik berdasarkan Silhouette Score. Namun, hasil ini berbeda dengan (Gunadi & Wirawan, 2025) yang menemukan Euclidean Distance sebagai metrik terbaik berdasarkan DBI, serta (Khoirala et al., 2025) yang menemukan Cosine Similarity unggul dengan Silhouette Score sebesar 0,6918, dan (Permata et al., 2025) yang menemukan Canberra Distance sebagai metrik terbaik berdasarkan Average Within Cluster Distance. Perbedaan temuan antar penelitian tersebut disebabkan oleh perbedaan karakteristik dataset, metrik evaluasi yang digunakan, serta distribusi jawaban responden pada masing-masing penelitian. Adapun performa Chebyshev Distance yang tidak optimal pada penelitian ini bertentangan dengan temuan (Hidayati et al., 2021) yang menyatakan Chebyshev Distance stabil pada berbagai variasi atribut, yang semakin memperkuat bahwa performa setiap metrik jarak sangat bergantung pada karakteristik data



yang digunakan sehingga tidak terdapat satu metrik jarak yang secara universal optimal untuk seluruh konteks analisis clustering.

4. KESIMPULAN

Penelitian ini menganalisis dan membandingkan lima metrik jarak pada algoritma K-Means Clustering untuk mengelompokkan tingkat kepuasan masyarakat terhadap layanan di Mal Pelayanan Publik (MPP) Pekanbaru dengan melibatkan 532 responden setelah proses data cleaning. Jumlah kluster yang diuji ditentukan berdasarkan hasil Elbow Method dengan rentang $k=2$ hingga $k=5$, di mana titik elbow terdeteksi pada $k=4$ sebagai batas transisi optimal sebelum penurunan WCSS melambat secara konsisten. Hasil evaluasi membuktikan bahwa pemilihan metrik jarak berpengaruh signifikan terhadap kualitas clustering, di mana Manhattan Distance dengan $k=2$ menghasilkan nilai DBI terendah sebesar 0,8144 yang menunjukkan kluster paling kompak dan terpisah, sementara Euclidean Distance dengan $k=3$ menghasilkan Silhouette Score tertinggi sebesar 0,5088 yang menunjukkan pemisahan antar kluster paling jelas. Perbedaan metrik terbaik antara kedua evaluasi ini bukan merupakan inkonsistensi, melainkan kontribusi metodologis penelitian ini yang membuktikan bahwa DBI dan Silhouette Score mengukur aspek kualitas clustering yang berbeda, sehingga pemilihan metrik evaluasi harus disesuaikan dengan tujuan analisis. Chebyshev Distance terbukti paling tidak optimal untuk data skala Likert karena hanya mempertimbangkan selisih satu dimensi tertinggi, mengabaikan 22 dimensi lainnya, dan menghasilkan distribusi kluster yang sangat tidak seimbang (444 vs 88 responden). Secara substantif, hasil clustering Manhattan Distance dengan $k=2$ menunjukkan distribusi seimbang antara kelompok Puas (266 responden) dan Sangat Puas (266 responden), yang mengindikasikan bahwa layanan MPP Pekanbaru secara umum telah memenuhi ekspektasi masyarakat. Namun demikian, analisis mendalam terhadap nilai centroid per indikator mengungkap tiga titik lemah yang konsisten lintas kluster, yaitu Q8 (ketepatan waktu penyelesaian layanan), Q10 (tindak lanjut terhadap pengaduan masyarakat), dan Q21 (ketersediaan produk layanan). Temuan ini memberikan rekomendasi kebijakan yang konkret bagi manajemen MPP Pekanbaru yaitu dengan ditetapkan standar waktu penyelesaian layanan yang terukur dan dipantau secara berkala untuk mengatasi kelemahan pada Q8, perlu dibangun mekanisme tindak lanjut pengaduan yang transparan dan dapat dilacak oleh masyarakat untuk mengatasi kelemahan pada Q10, dan perlu dilakukan evaluasi kelengkapan dan ketersediaan produk layanan secara periodik untuk mengatasi kelemahan pada Q21. Penelitian ini memberikan kontribusi dalam analisis penggunaan metrik jarak pada algoritma K-Means untuk data kepuasan masyarakat. Hasil penelitian menunjukkan bahwa perbedaan metrik evaluasi dapat menghasilkan penilaian kualitas cluster yang berbeda, sehingga pemilihan metrik jarak dan metode evaluasi perlu disesuaikan dengan tujuan analisis. Selain itu, penelitian ini memberikan gambaran mengenai performa lima metrik jarak pada data kepuasan masyarakat berbasis skala Likert. Temuan yang diperoleh memberikan gambaran mengenai performa metrik jarak yang digunakan pada data kepuasan masyarakat serta dapat menjadi bahan pertimbangan dalam pemilihan metrik jarak pada penelitian dengan karakteristik data yang serupa. Untuk penelitian selanjutnya, disarankan agar pengelompokan dilakukan secara spesifik berdasarkan jenis layanan di MPP sehingga dapat diidentifikasi layanan mana yang paling memengaruhi kepuasan masyarakat, serta mengeksplorasi metrik jarak lainnya seperti Minkowski atau Mahalanobis Distance pada dataset serupa untuk memperkaya panduan metodologis dalam analisis clustering data layanan publik.

REFERENCES

- Agung Nugraha, Odi Nurdiawan, & Gifthera Dwilestari. (2022). Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 1–7. <https://doi.org/10.36040/jati.v6i2.5755>
- Baha'uddin, M., & Fatah, Z. (2024). Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas. *Gudang Jurnal Multidisiplin Ilmu*, 2(10), 86–94. <https://doi.org/10.59435/gjmi.v2i11.1040>
- Diastuti, A., & Ayyustia, R. (2025). *Strategi Peningkatan Mutu Pelayanan Publik melalui Inovasi Digital di Kelurahan Sunter Agung*. 2(2), 109–126. <https://doi.org/https://doi.org/10.62383/presidensial.v2i2.721>
- Farida, J. I., & Lubis, A. H. (2025). *Grouping Of Tourism Locations In Indonesia Using Distance Variations In The K-Means Algorithm*. 8(3), 42–50. <https://doi.org/10.31289/jite.v8i3Spc.14528>
- Golzar, J., & Noor, S. (2022). Convenience Sampling. *International Journal of Education & Language Studies*, 1(2), 72–77. <https://doi.org/10.22034/IJELS.2022.162981>
- Gunadi, I. G. A., & Wirawan, I. M. A. (2025). Studi Perbandingan Algoritma Euclidean, Manhattan dan Chebyshev Distance untuk Optimasi Metode K-Means Clustering pada Pengelompokan Data Penerimaan Mahasiswa Baru. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 22(2), 116–127. <https://doi.org/10.23887/jptk-undiksha.v22i2.98863>
- Harahap, M. I., Sulawati, S., & Sibuea, S. R. (2024). Pengaruh Kualitas Pelayanan Pembuatan Surat Keterangan Catatan Kepolisian (SKCK) Terhadap Kepuasan Masyarakat dengan Menggunakan Metode Regresi Linier Berganda di Polrestaes Medan. *Journal of Research and Development on Public Policy*, 3(2), 213–233. <https://doi.org/10.58684/jarvic.v3i2.164>
- Hartono, B., & Lusiana, V. (2026). Analisis Metode Elbow SSE, Silhouette Score, dan Jaccard Stability dalam Pemilihan Jumlah Kluster Data yang Optimal TIN: Terapan Informatika Nusantara. 6(8), 1521–1532.



<https://doi.org/10.47065/tin.v6i8.9271>

- Hayuningtyas, H. Y., & Monoarfa, A. (2023). Analisis Pelaksanaan Pelayanan Publik. *JIANA (Jurnal Ilmu Administrasi Negara)*, 21(1), 31–39. <https://doi.org/10.46730/jiana.v21i1.8090>
- Herdiana, I., Kamal, M., Triyani, Estri, M., & Renny. (2025). *A More Precise Elbow Method for Optimum K-Keans Clustering*. 1–22. <https://doi.org/10.48550/arXiv.2502.00851>
- Hidayati, R., Zubair, A., Pratama, A. H., & Indana, L. (2021). Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering. *Tecno.COM*, 20(2), 186–197. <https://doi.org/10.33633/tc.v20i2.4556>
- Ikhsan, T., Haerani, E., Wulandari, F., & Syafria, F. (2025). *Clustering Data Penduduk Menggunakan Algoritma K-Means TIN: Terapan Informatika Nusantara*. 5(12), 679–687. <https://doi.org/10.47065/tin.v5i12.7328>
- Iqbal, M., Syaripuddin, & Nurul, M. (2023). *Implementasi Algoritma K-Means Clustering dengan Jarak Euclidean dalam Mengelompokkan Daerah Penyebaran*. 2(1), 47–56. <https://doi.org/10.30872/basis.v2i1.1019>
- Kaligis, G. B., & Yulianto, S. (2022). Analisa Perbandingan Algoritma K-Means, K-Medoids, dan X-Means Untuk Pengelompokan Kinerja Pegawai. *Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, 01(03), 179–193. <https://doi.org/10.24246/itexplore.v1i3.2022.pp179-193>
- Khoirala, A., Pily, E., Rio, U., & Tashid. (2025). Komparasi K-Means Clustering dengan Euclidean dan Cosine Similarity untuk Segmentasi dan Rekomendasi Produk pada Data E-Commerce Annisa. *The Indonesian Journal of Computer Science*, 14(2), 3032–3055. <https://doi.org/10.33022/ijcs.v14i2.4713>
- Kinanti, R., Jasmir, & Fachruddin. (2024). Penerapan Metode Clustering K-Means Untuk Menentukan Prioritas Penerima Bantuan Program Beras Untuk Rakyat Miskin (RASKIN) Studi Kasus: Kecamatan Siluak. *Jakakom*, 4(2), 2808–5469. <https://doi.org/10.33998/jakakom.v4i2>
- Lopez-miguel, I. D. (2021). Survey on Preprocessing Techniques for Big Data Projects. *Engineering Proceedings*, 7(1). <https://doi.org/10.3390/engproc2021007014>
- Oop Sofiyah, S., R., N., & Danar Dana, R. (2023). Analisis Efektivitas Pelayanan Publik Menggunakan K-Means Clustering Di Kecamatan Sukagumiwang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1291–1296. <https://doi.org/10.36040/jati.v7i2.6536>
- Permata, R. P., Alifah, A. N., & Sanjaya, I. M. W. A. (2025). Optimizing K-Means Clustering through Distance Metric Simulation for Strategic Enrollment Segmentation in Private Universities. *Jurnal Matematika Murni Dan Aplikasi*, 10(2), 616–629. <https://doi.org/10.18860/cauchy.v10i2.33089>
- Prasetyo, D., Muchsin, S., & Suyeno. (2022). Efektifitas Mal Pelayanan Publik Dalam Peningkatan Kualitas Pelayanan Publik. *Jurnal Respon Publik*, 16(10), 1–6. <https://jim.unisma.ac.id/index.php/rpp/article/view/18732>
- Supawanhar, Putri, S., & Febriansah, R. (2024). Pengukuran Indeks Kepuasan Masyarakat Perizinan E - Mal Pelayanan Publik (MPP) Kota Bengkulu. *Journal of Governance and Public Administration*, 1(4), 563–578. <https://doi.org/10.70248/jogapa.v1i4.1237>
- Supriyanto, D. F., Prabowo, S., Widodo, A. S., & Eldo, D. H. A. P. (2021). Analisis Kinerja Pelayanan Publik di Masa Pandemi (Studi Pelayanan Kartu Tanda Penduduk Elektronik (E-KTP) Kabupaten Karawang). *Jurnal Ranah Publik Indonesia Kontemporer (Rapik)*, 1(1), 40–49. <https://doi.org/10.47134/rapik.v1i1.8>
- Sutirta, N. F., & Novianti. (2024). Perbandingan Manhattan dan Euclidean Distance Untuk Pengelompokan Penyakit Jantung Menggunakan Algoritma K-Means. *ICIT Journal*, 10(1), 61–70. <https://doi.org/10.33050/icit.v10i1.2860>
- Wisanta, E. H., & Marlim, Y. N. (2021). Analisis Algoritma K-Means Untuk Clustering Kepuasan Pelayanan: Mall Pelayanan Publik Pekanbaru. *Seminar Nasional Informatika*, 223–228. <https://ejournal.pelitaIndonesia.ac.id/ojs32/index.php/SENATIKA/article/view/1160>
- Yusuf, B., Mahara, R., Ahmadian, H., Wahyuni, S., & AR, K. (2022). Analisis Clustering Penduduk Miskin Di Provinsi Aceh Menggunakan Algoritma K-Means Dan X-Means. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 5(1), 26–35. <https://doi.org/10.32672/jnkti.v5i1.3961>