

An Analysis of Public Satisfaction with Government Services: A Multi-Method Approach Using PCA, K-Means Clustering, and Linear Regression

Abuzar Gafari¹⁾, Sarjon Defit²⁾, dan Rini Sovia³⁾

^{1,2,3}Magister Teknik Informatika, Universitas Putra Indonesia YPTK Padang
^{1,2,3} Jl. Raya Lubuk Begalung, Lubuk Begalung Nan XX, Kecamatan Lubuk Begalung Kota Padang Sumbar
E-mail: gafariabuzar.2@gmail.com¹⁾, sarjon_defit@upiypk.ac.id²⁾, rini_sovia@upiypk.ac.id³⁾

ABSTRACT

Flawless performance evaluation results across all service dimensions may potentially obscure the identification of areas for improvement and diminish objectivity in decision-making. This study aims to identify the specific service attributes influencing public satisfaction and to segment respondents based on their satisfaction levels at the Office of the Ministry of Religious Affairs in Payakumbuh City. The research integrates Principal Component Analysis (PCA), K-means clustering, and linear regression. PCA was employed to reduce data dimensionality and establish principal components; K-means clustering was utilized to group respondents based on perceptual similarities regarding service quality; and linear regression was applied to identify the most significant factors influencing public satisfaction within each segment. The data were sourced from the Public Service Survey Information System (SISULAP) application of the Payakumbuh Ministry of Religious Affairs, spanning June 2024 to October 2025, with a total of 1,950 respondents. The findings reveal that service process and efficiency are the primary factors influencing all respondent segments, with the low-satisfaction segment identified as the top priority for service improvement. The regression models demonstrate robust performance across all segments. These findings provide an empirical foundation for data-driven policymaking to enhance public service quality.

Keywords: *Public Satisfaction, PCA, K-means Clustering, Linear Regression, CSI, Service Quality*

Analisis Kepuasan Masyarakat Terhadap Layanan Publik Melalui Pendekatan PCA, K-means clustering, dan Regresi Linear

ABSTRAK

Hasil evaluasi kinerja layanan publik yang sempurna di semua aspek layanan berpotensi mengaburkan identifikasi area perbaikan dan mengurangi objektivitas dalam pengambilan keputusan. Penelitian ini bertujuan mengidentifikasi aspek layanan yang berpengaruh terhadap kepuasan masyarakat serta melakukan segmentasi responden berdasarkan tingkat kepuasan pada Kantor Kementerian Agama Kota Payakumbuh. Penelitian ini mengintegrasikan metode *Principal Component Analysis* (PCA), *K-means clustering*, dan regresi linear. PCA digunakan untuk mereduksi dimensi data dan membentuk komponen utama, *K-means clustering* untuk mengelompokkan responden berdasarkan kesamaan persepsi terhadap kualitas layanan serta regresi linear untuk mengidentifikasi faktor yang berpengaruh terhadap kepuasan masyarakat pada setiap segmen. Data penelitian bersumber dari aplikasi Sistem Informasi Survei Layanan Publik (SISULAP) Kantor Kementerian Agama Kota Payakumbuh periode Juni 2024 hingga Oktober 2025 dengan total data 1.950 responden. Hasil penelitian menunjukkan bahwa faktor proses dan efisiensi pelayanan merupakan faktor utama yang berpengaruh di seluruh segmen responden, dengan segmen kepuasan rendah menjadi prioritas utama untuk perbaikan layanan. Model regresi menunjukkan tingkat kinerja yang baik pada seluruh segmen. Temuan ini memberikan dasar empiris bagi pengambilan kebijakan dalam peningkatan kualitas layanan publik berbasis data.

Kata Kunci: *Kepuasan Masyarakat, PCA, K-means Clustering, Regresi Linear, IKM, Kualitas Layanan*

1. PENDAHULUAN

Kepuasan masyarakat merupakan indikator penting dalam evaluasi kinerja layanan publik dan menjadi dasar utama dalam evaluasi kinerja instansi pemerintah. Di Indonesia, pengukuran kepuasan masyarakat dilakukan secara berkelanjutan melalui Indeks Kepuasan Masyarakat (IKM) merupakan instrumen resmi evaluasi

kinerja layanan publik (Peraturan Menteri PANRB, 2023). Namun, nilai IKM yang tinggi tidak selalu mencerminkan kepuasan masyarakat secara merata. Perbedaan persepsi antar kelompok pengguna layanan sering kali tersembunyi di balik nilai rata-rata, sehingga berpotensi mengaburkan area perbaikan dan menyebabkan pengambilan kebijakan yang kurang tepat

sasaran (Channamallu dkk., 2025; Sansan Yasinta Ar Rohmah dkk., 2025).

Fenomena tersebut juga terjadi pada Kantor Kementerian Agama Kota Payakumbuh. Hasil pengukuran IKM tahun 2024 menunjukkan nilai yang sangat tinggi yaitu 96,60 dengan capaian penilaian yang sempurna pada semua aspek layanan. Kondisi ini berpotensi mengaburkan identifikasi area perbaikan dan mempengaruhi objektivitas dalam pengambilan kebijakan hasil evaluasi kinerja layanan publik. Kondisi tersebut juga mengindikasikan bahwa pendekatan evaluasi berbasis nilai agregat belum sepenuhnya mampu menangkap keragaman pengalaman masyarakat terhadap layanan publik.

Dalam kajian kepuasan masyarakat, pendekatan analisis yang digunakan umumnya berasal dari dua ranah utama, yaitu statistika dan data mining. Pendekatan statistika memberikan landasan teoritis yang kuat dalam menganalisis hubungan antar variabel, seperti melalui *Principal Component Analysis* (PCA) dan regresi linear (Kotronoulas dkk., 2023; Zhu dkk., 2024). Sementara itu, data mining memiliki keunggulan dalam menemukan pola tersembunyi dan melakukan segmentasi data secara efektif, khususnya melalui algoritma *K-means clustering* (Chitra & Heikal, 2024; Tan, 2024; G. Wang, 2025). Meskipun demikian, sebagian besar penelitian sebelumnya masih menerapkan kedua pendekatan tersebut secara terpisah (Nguyen dkk., 2024; Sarkar dkk., 2024).

Berbeda dengan penelitian sebelumnya, penelitian ini mengintegrasikan ketiga metode tersebut dalam satu kerangka analisis terpadu untuk mengkaji kepuasan masyarakat terhadap layanan publik di Kantor Kementerian Agama Kota Payakumbuh. Pendekatan ini memungkinkan identifikasi faktor kepuasan masyarakat lebih spesifik pada setiap segmen responden, sehingga mampu mengungkap keterbatasan evaluasi kepuasan berbasis nilai agregat yang tidak dapat tertangkap oleh analisis dengan metode konvensional. Dalam kerangka tersebut, PCA digunakan untuk mereduksi sembilan indikator IKM menjadi beberapa komponen utama (Celestin, 2025), *K-means clustering* diterapkan untuk mengelompokkan masyarakat berdasarkan tingkat kepuasan (Q. Wang, 2024), dan regresi linear digunakan untuk mengidentifikasi faktor-faktor yang paling berpengaruh pada setiap segmen kepuasan (Bach & Thiel, 2024; Zhao dkk., 2025).

Pendekatan terintegrasi ini mampu memberikan pemahaman yang komprehensif tentang kepuasan masyarakat, tidak hanya dalam mengidentifikasi faktor yang paling berpengaruh, tetapi juga dalam mengungkap kelompok masyarakat yang belum terlayani secara optimal. Hasil penelitian ini dapat menjadi dasar empiris bagi pengambilan kebijakan peningkatan kualitas layanan publik yang lebih terfokus, berbasis data, dan berkelanjutan.

2. RUANG LINGKUP

Evaluasi kepuasan layanan publik umumnya dilakukan dengan menggunakan nilai rata-rata dari hasil survei Indeks Kepuasan Masyarakat (IKM). Pendekatan tersebut memiliki keterbatasan karena berpotensi mengaburkan perbedaan persepsi kepuasan antar kelompok responden, sehingga informasi mengenai variasi tingkat kepuasan masyarakat tidak dapat tergambarkan secara komprehensif. Untuk mengatasi keterbatasan tersebut, penelitian ini menganalisis kepuasan masyarakat terhadap layanan publik pada Kantor Kementerian Agama Kota Payakumbuh menggunakan pendekatan analisis multidimensi dan segmentasi.

Analisis dilakukan terhadap sembilan indikator IKM yang merepresentasikan kualitas pelayanan publik, yaitu persyaratan, prosedur, waktu pelayanan, kesesuaian produk layanan, biaya, kompetensi pelaksana, perilaku pelaksana, penanganan pengaduan, serta sarana dan prasarana. Melalui integrasi metode *Principal Component Analysis* (PCA), *K-means clustering*, dan regresi linear, penelitian ini bertujuan untuk mengidentifikasi komponen utama kualitas pelayanan, segmentasi tingkat kepuasan masyarakat, serta faktor-faktor yang mempengaruhi kepuasan pada setiap segmen yang terbentuk. Berdasarkan kerangka tersebut, ruang lingkup penelitian ini meliputi cakupan permasalahan, batasan penelitian, serta hasil yang diharapkan dari penelitian

1. Cakupan Permasalahan

Penelitian ini membahas analisis kepuasan masyarakat terhadap pelayanan publik pada Kantor Kementerian Agama Kota Payakumbuh dengan pendekatan statistik *multivariat* dan teknik *data mining*. Permasalahan yang dikaji mencakup identifikasi struktur faktor utama yang merepresentasikan kualitas pelayanan publik, segmentasi responden berdasarkan tingkat kepuasan, serta analisis faktor-faktor yang mempengaruhi kepuasan masyarakat terhadap layanan yang diberikan.

2. Batasan-batasan penelitian

Penelitian ini menggunakan data Survei Indeks Kepuasan Masyarakat (IKM) yang bersumber dari Sistem Informasi Survei Layanan Publik (SISULAP) pada Kantor Kementerian Agama Kota Payakumbuh dengan periode pengumpulan data Juni 2024 hingga Oktober 2025 dan jumlah responden sebanyak 1.950. Variabel yang dianalisis terbatas pada sembilan indikator IKM, yaitu persyaratan, prosedur, waktu pelayanan, kesesuaian produk layanan, biaya, kompetensi pelaksana, perilaku pelaksana, penanganan pengaduan, serta sarana dan prasarana. Analisis dilakukan menggunakan metode *Principal Component Analysis* (PCA), *K-means clustering*, dan regresi linear dengan bantuan bahasa pemrograman *Python*

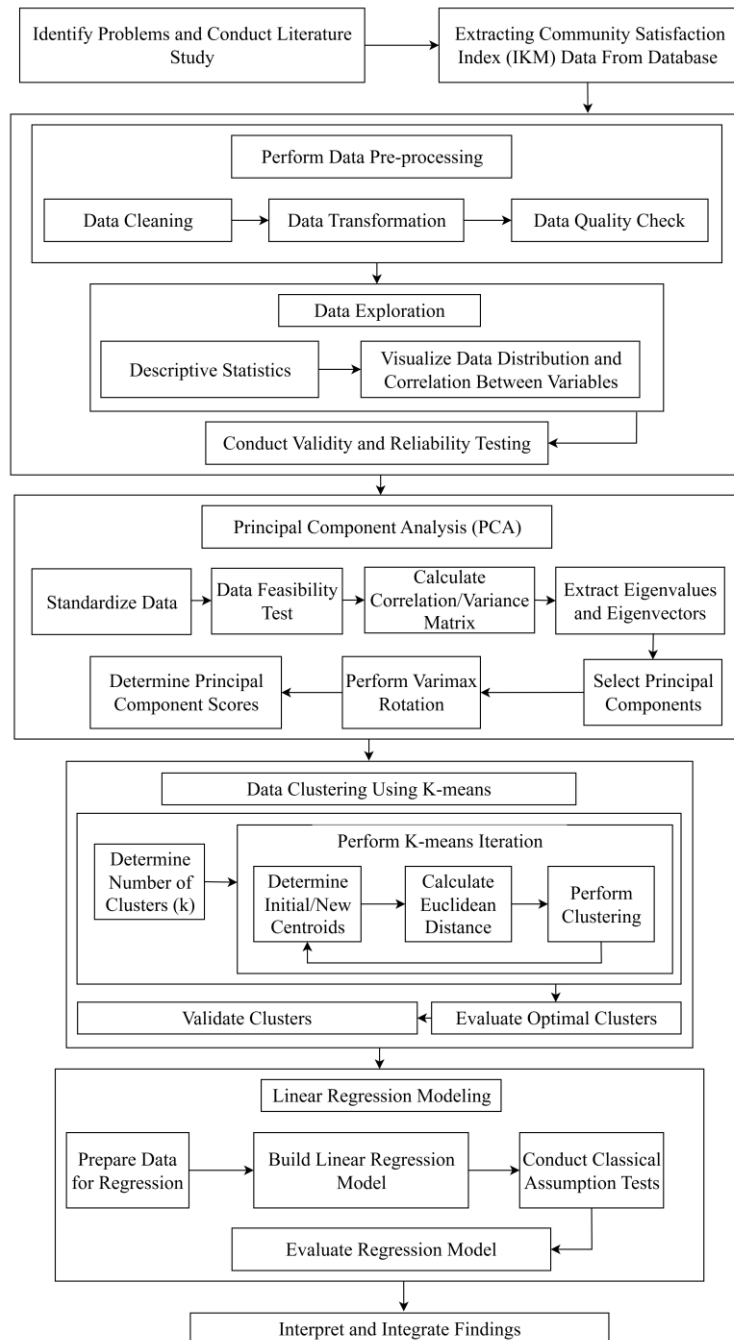
3. Rencana hasil yang diharapkan

Penelitian ini diharapkan menghasilkan identifikasi komponen utama yang merepresentasikan kualitas

pelayanan publik, klasifikasi segmentasi kepuasan masyarakat terhadap layanan publik, serta model analisis yang menunjukkan faktor-faktor yang mempengaruhi kepuasan masyarakat pada setiap segmen yang terbentuk. Hasil penelitian diharapkan dapat menjadi dasar empiris bagi instansi dalam merumuskan strategi peningkatan kualitas pelayanan publik yang lebih terarah, berbasis data, dan sesuai dengan karakteristik tingkat kepuasan masyarakat

3. BAHAN DAN METODE

Penelitian ini menerapkan pendekatan kuantitatif dengan mengombinasikan tiga teknik analisis, yaitu *Principal Component Analysis (PCA)*, *K-means clustering*, dan regresi linear. Keterkaitan antar metode serta urutan proses analisis untuk menjawab tujuan penelitian dirangkum dalam kerangka kerja penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Figure 1. Research Framework

Gambar 1 merupakan kerangka kerja penelitian yang menggambarkan alur analisis data dimulai dari tahap pra-pemrosesan data untuk membersihkan serta menguji validitas dan reliabilitas data, dilanjutkan dengan reduksi dimensi menggunakan PCA dari banyak variabel menjadi beberapa komponen utama, kemudian dilakukan segmentasi responden menggunakan *K-means clustering*, dan diakhiri dengan analisis faktor yang mempengaruhi kepuasan masyarakat menggunakan regresi linear

3.1 Pra-Pemrosesan dan Eksplorasi Data

Tahapan pra-pemrosesan data merupakan tahap penting untuk membersihkan (Bylemans dkk., 2025), mentransformasi, dan menstandarisasi data (Paap dkk., 2024) yang diperoleh dari basis data aplikasi Kantor Kementerian Agama Kota Payakumbuh. Tahap ini bertujuan untuk memastikan data memenuhi asumsi model statistik dan siap dianalisis menggunakan PCA, *K-means clustering*, dan *Regresi Linear*. Selanjutnya tahap eksplorasi data dilakukan untuk memahami struktur dasar data sebelum analisis formal dilakukan (Lokanan, 2024). Tahap ini bertujuan untuk mendapatkan gambaran awal data yang diperoleh dari basis data, bukan untuk menarik sebuah kesimpulan. Proses ini juga berfungsi sebagai pengendali kualitas data untuk meminimalkan kesalahan interpretasi di masa akan datang. Kegiatan eksplorasi data meliputi analisis statistik deskriptif dan visualisasi distribusi data (Malakar & Nepal, 2024)

3.2 Principal Component Analysis (PCA)

PCA digunakan untuk mereduksi dimensi data dan mengatasi multikolinearitas antar variabel. Tahapan awal PCA yaitu melakukan standarisasi data bertujuan agar seluruh variabel memiliki rentang skala yang sama (Guo dkk., 2025; Musfiroh dkk., 2023) menggunakan persamaan (1):

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Diketahui Z pada Persamaan (1) adalah data standar, (x) menyatakan nilai observasi, sedangkan (μ) merupakan nilai rata-rata dan (σ) nilai standar deviasi. Selanjutnya dibentuk matrik kovarians dari data yang telah terstandarisasi untuk melihat hubungan antar variabel (Chan & Gibberd, 2025) yang kemudian dilanjutkan dengan penentuan *eigenvalue* dan *eigenvector* (Delaosa dkk., 2025). Komponen utama dipilih berdasarkan kriteria kumulatif varians (Kalantan dkk., 2025). Untuk memudahkan proses interpretasi dilakukan rotasi varimax terhadap komponen utama yang terbentuk dan hitung Skor komponen utama menggunakan persamaan (2):

$$PC=ZV \quad (2)$$

Diketahui PC pada Persamaan (2) adalah komponen utama, Z merupakan matrik data terstandarisasi dan V

matriks *eigenvector* terpilih. Skor komponen utama yang dihasilkan menjadi input data untuk *K-means clustering*.

3.3 K-means Clustering

K-means clustering merupakan algoritma yang digunakan untuk mengelompokkan data berdasarkan kesamaan persepsi responden dengan input data berupa skor komponen utama hasil PCA (Deldadehas dkk., 2025; Sánchez Vines dkk., 2025). Proses diawali dengan penentuan jumlah kluster dan inialisasi *centroid*, kemudian hitung jarak data ke titik terdekat (Anilshi, 2024) menggunakan persamaan (3):

$$d_e = \sqrt{\sum_{j=1}^m (x_j - c_j)^2} \quad (3)$$

Diketahui d_e (3) adalah *Euclidean Distance*, x_j merupakan nilai data pada variabel fitur ke j , c_j menjelaskan nilai *centroid* pada variabel ke j , m merupakan jumlah total fitur dalam data, j merupakan indek variabel. Setelah nilai jarak ditemukan, lakukan pengelompokan data, kemudian proses iteratif dilakukan dengan memperbaharui nilai *centroid* hingga mencapai konvergensi. Kualitas kluster yang terbentuk dievaluasi dan divalidasi untuk memastikan kluster optimal dan stabil. Hasil kluster akan jadi input model regresi linear.

3.4 Pemodelan Regresi Linear

Regresi linear digunakan untuk menganalisa korelasi antara variabel dependen dengan variabel independen berdasarkan hasil kluster yang telah terbentuk (Setiawan dkk., 2024). Model regresi terbagi dua, pertama *regresi linear* sederhana digunakan untuk menjelaskan satu variabel bebas terhadap variabel terikat (Grünwald dkk., 2024). Kedua, *regresi linear* berganda digunakan untuk menjelaskan lebih dari dua variabel bebas terhadap satu variabel terikat (Jullia & Finatariyani, 2024). Dengan demikian regresi linear berganda sangat tepat digunakan untuk analisa data yang memiliki banyak variabel (Dulger Altner dkk., 2024). Model regresi linear berganda dirumuskan sebagaimana yang terlihat pada persamaan (4):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (4)$$

Diketahui Y pada Persamaan (4) adalah kepuasan masyarakat, X_1, X_2, \dots, X_n merupakan komponen utama hasil PCA, β_0 sebagai konstanta, berikutnya $\beta_1, \beta_2, \dots, \beta_n$ adalah koefisien regresi yang menyatakan pengaruh setiap komponen utama terhadap kepuasan masyarakat dan ε adalah *error term*. Koefisien regresi β_i merupakan besarnya perubahan rata-rata pada variabel dependen saat adanya satu peningkatan pada variabel X_i atau variabel independen, dengan anggapan variabel bahwa nilai X lain bernilai tetap (Fathi dkk., 2025). Sebelum interpretasi hasil dilakukan uji asumsi klasik meliputi uji

normalitas, uji heterokedastisitas dan uji autokorelasi (M. Wang dkk., 2024; Yildirim, 2024).. Evaluasi model dilakukan untuk menilai kinerja model yang dihasilkan (Allgaier & Pryss, 2024).

3.5 Interpretasi dan Integrasi Hasil

Pada tahapan ini dilakukan analisis data untuk menjelaskan atau menafsirkan hasil dari temuan penelitian yang telah dipaparkan pada bagian pembahasan. Hasil tersebut akan diverifikasi kembali terhadap rumusan masalah, sejauh mana pertanyaan pada rumusan masalah dapat terjawab dengan baik oleh hasil penelitian. Serta sejauh mana tujuan yang ditetapkan di awal tercapai dengan baik berdasarkan data yang dihasilkan. Setelah interpretasi temuan selesai dilanjutkan dengan mengintegrasikan temuan untuk membangun sebuah narasi yang akan dijadikan sebuah kesimpulan akhir yang utuh. Narasi kesimpulan tersebut diperoleh dari hasil penggabungan analisis data menggunakan PCA, *K-means clustering*, dan regresi linear.

4. PEMBAHASAN

Bagian ini menyajikan hasil implementasi metode PCA, *K-means clustering* dan regresi linear. Analisis diawali dengan menyajikan hasil pra-pemrosesan data, dilanjutkan dengan pembentukan skor komponen utama, klasterisasi menggunakan *K-means clustering* dan hasil pemodelan regresi linear. Selanjutnya hasil klasterisasi data dan model regresi dianalisis secara komprehensif untuk menginterpretasikan hubungan antar variabel. Pembahasan difokuskan pada keterkaitan hasil analisis dengan tujuan penelitian. Dengan demikian temuan tidak hanya disajikan secara statistik, namun juga dapat diinterpretasikan dan memberi pemahaman yang lebih bermakna.

4.1 Pra-Pemrosesan Data

Bagian ini menyajikan hasil analisis pra-pemrosesan data yang dilakukan untuk memastikan kualitas data yang digunakan dalam analisis. Data yang diperoleh dari basis data aplikasi Sistem Informasi Survei Layanan Publik (SISULAP) Kantor Kementerian Agama Kota Payakumbuh terlebih dahulu melalui proses pembersihan data untuk menangani data yang hilang. Proses analisis statistik deskriptif dan eksplorasi data dilakukan untuk mendapatkan gambaran awal distribusi data. Kemudian proses standarisasi diterapkan untuk menyamakan skala antar variabel. Sampel hasil akhir tahap pra-pemrosesan dapat dilihat pada Tabel 1.

Tabel 1. Data Standarisasi Z-score
Table 1. Z-score Standardization Data

Requirements	Procedures	Service Delivery Time	Conformity of Service Output	Cost	Implementer Competence	Implementer Behavior	Complaint Handling	Summary
0,276	0,273	0,304	0,315	0,350	0,297	0,266	0,334	0,332
0,276	0,273	0,304	0,315	0,350	0,297	0,266	0,334	0,332
0,276	0,273	0,304	0,315	0,350	0,297	0,266	0,334	0,332
0,276	0,273	0,304	0,315	0,350	0,297	-3,638	-2,894	0,332
0,276	0,273	0,304	-3,026	0,350	-3,278	0,266	0,334	0,332

Hasil pra-pemrosesan data yang disajikan pada Tabel 1 terlihat bahwa sebagian besar memiliki nilai *z-score* relatif sama dan positif, namun juga terdapat nilai negatif yang menandakan nilai yang diberikan masih jauh dari nilai rata-rata persepsi kepuasan layanan. Sementara nilai negatif cukup besar pada beberapa aspek layanan, menunjukkan adanya keberagaman.

4.2 Analisis Principal Component Analysis (PCA)

Bagian ini menyajikan hasil analisis data terstandarisasi menggunakan PCA untuk mereduksi dimensi data serta membentuk skor komponen utama. Penentuan jumlah komponen utama dilakukan berdasarkan kriteria kumulatif varian minimal $\geq 70\%$, dengan tetap mempertimbangkan efektivitas reduksi dimensi dan. Sampel hasil akhir analisis PCA dapat dilihat pada Tabel 2 sebagai dasar untuk melakukan

analisis segmentasi menggunakan algoritma *K-means clustering*.

Tabel 2. Skor Komponen Utama
Table 2. Z-score Standardization Data

Score_PC1	Score_PC2	Summary
0,911	0,086	0,377
0,911	0,086	0,377
0,911	0,086	0,377
-1,441	-1,708	-0,526
-1,494	0,436	-0,526

Hasil analisis PCA pada tabel 2 merupakan transformasi data hasil reduksi sembilan variabel menjadi dua komponen utama yaitu skor PC1 dan skor PC2. Skor PC1 dan PC2 bernilai positif menunjukkan

beberapa responden memberikan penilaian baik dan tinggi dibandingkan dengan nilai rata-rata persepsi responden. Sementara skor bernilai negatif menunjukkan adanya responden yang memberikan penilaian lebih rendah daripada nilai rata-rata kepuasan layanan.

4.3 Analisis K-means Clustering

Bagian ini menyajikan hasil klasterisasi data menggunakan algoritma K-means clustering. Proses yang dilakukan pertamakali yaitu menetapkan jumlah klaster $k=4$ serta menginisialisasi nilai *centroid* awal. Proses iterasi dilakukan mulai dengan menghitung jarak data terhadap titik *centroid* terkecil dan dikelompokkan berdasarkan nilai jarak terdekat data dengan *centroid*. Selanjutnya, nilai *centroid* baru dihitung berdasarkan nilai rata-rata masing-masing anggota klaster yang terbentuk, iterasi berlanjut hingga konvergensi tercapai. Sampel hasil akhir klasterisasi menggunakan *K-means clustering* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Klasterisasi
Table 3. Cluster Analysis Results

Score PC1	Score PC2	Sum	Cluster_Id	Caption
0,911	0,086	0,377	Cluster 1	Sangat Puas
0,911	0,086	0,377	Cluster 1	Sangat Puas
0,911	0,086	0,377	Cluster 1	Sangat Puas
-1,441	-1,708	-0,526	Cluster 3	Puas
-1,494	0,436	-0,526	Cluster 1	Sangat Puas

Berdasarkan hasil klasterisasi data pada Tabel 3 terlihat bahwa data telah dikelompokkan dalam beberapa klaster. Pada umumnya data berada pada klaster 3 dengan skor komponen utama (PC1) relatif tinggi dan bernilai positif dengan kategori Sangat Puas. Distribusi anggota klaster yang dihasilkan secara lebih rinci dapat dilihat pada Tabel 4.

Tabel 4. Distribusi Anggota Klaster
Table 4. Cluster Membership Distribution

Cluster 1	Cluster 2	Cluster 3	Cluster 4
1.675	95	108	72

Berdasarkan data pada Tabel 4, terlihat bahwa distribusi anggota pada klaster 1 merupakan klaster yang memiliki anggota terbanyak dengan jumlah 1.675 responden yang memiliki label Sangat Puas. Klaster 3 dengan label Puas terdiri dari 108 responden. Selanjutnya, klaster 2 memiliki 95 responden dengan label Cukup Puas. Sementara itu, klaster 4 merupakan klaster dengan jumlah responden paling sedikit, yaitu 72 responden, yang merepresentasikan kelompok dengan tingkat kepuasan rendah. Hasil segmentasi ini menunjukkan bahwa secara keseluruhan responden berada pada kelompok sangat puas, namun keberadaan klaster dengan tingkat kepuasan rendah menunjukkan

perlu perbaikan pada aspek layanan tertentu agar kualitas pelayanan publik dapat lebih merata.

4.4 Analisis Regresi Linear

Bagian ini menyajikan hasil analisis regresi linear berganda untuk mengidentifikasi faktor yang berpengaruh pada setiap klaster yang telah terbentuk pada tahap analisis K-means clustering. Pada setiap klaster dilakukan perhitungan koefisien dan konstanta untuk menemukan faktor yang berpengaruh terhadap kepuasan layanan. Hasil akhir analisis model regresi linear berganda dapat dilihat pada Tabel 5.

Tabel 5. Hasil Regresi Linear Berganda
Table 5. Multiple Linear Regression Results

Model	R-squared	Adj. R-squared	coef	std err
1	0.941	0.940	0.3890	0.013
2	0.976	0.975	0.3901	0.008
3	0.977	0.977	0.3958	0.007
4	0.985	0.985	0.4603	0.001

Berdasarkan data pada Tabel 5 terlihat model 1 memperoleh nilai $R^2 = 0.941$ dan Adjusted $R^2 = 0.940$ yang menunjukkan sekitar 94% variasi kepuasan masyarakat pada klaster 3 berhasil dijelaskan oleh model selebihnya dijelaskan oleh faktor diluar model. Model 2 memperoleh nilai $R^2 = 0.976$ dan Adjusted $R^2 = 0.975$, menunjukkan bahwa pada model 2 variasi kepuasan layanan dapat dijelaskan 97.5%. Model 2 memperoleh nilai $R^2 = 0.977$ dan Adjusted $R^2 = 0.977$, model 3 dapat menjelaskan variasi kepuasan layanan sebesar 97.7%. Model 4 dapat menjelaskan variasi sebesar 98% dengan nilai $R^2 = 0.985$ dan Adjusted $R^2 = 0.985$. Sementara nilai koefisien rata-rata lebih besar dari 0.3890 menunjukkan bahwa faktor tersebut merupakan faktor yang berpengaruh terhadap kepuasan layanan pada setiap model. Dengan kata lain model regresi yang dibangun layak dan andal untuk digunakan dalam analisis kepuasan masyarakat.

Hasil evaluasi model regresi linear berganda secara keseluruhan menggunakan perhitungan MSE memperoleh nilai sebesar 0.18 dan RMSE sebesar 0.134. Nilai ini menunjukkan bahwa kesalahan dalam model prediksi relatif kecil dan memiliki akurasi model yang baik.

5. KESIMPULAN

Nilai Indeks Kepuasan Masyarakat (IKM) agregat sebesar 96,60 ditemukan berpotensi mengaburkan identifikasi area perbaikan yang sebenarnya dibutuhkan oleh pengguna layanan. Melalui integrasi *Principal Component Analysis* (PCA), *K-means clustering*, dan regresi linear, penelitian ini berhasil menjawab keterbatasan evaluasi konvensional dengan mengidentifikasi empat segmen kepuasan yang memiliki karakteristik evaluasi berbeda, termasuk kelompok responden dengan tingkat kepuasan rendah pada dimensi

layanan tertentu. Akurasi model yang divalidasi dengan nilai RMSE rendah (0,134) dan koefisien determinasi yang tinggi membuktikan bahwa kerangka kerja ini konsisten dalam merepresentasikan variasi kepuasan antar segmen. Temuan ini memberikan kontribusi pada pengembangan kerangka evaluasi layanan publik berbasis segmentasi yang lebih kaya dan presisi dibandingkan pendekatan nilai tunggal, serta menawarkan dasar metodologis bagi pengambilan kebijakan yang lebih tepat sasaran. Sebagai pengembangan lebih lanjut, penelitian di masa depan dapat mengeksplorasi penggunaan algoritma clustering lain seperti DBSCAN untuk membandingkan stabilitas kluster, serta menambahkan variabel demografi sebagai pemoderasi guna mendalami pengaruh karakteristik responden terhadap dimensi kualitas layanan yang terbentuk.

6. SARAN

Instansi penyelenggara layanan publik disarankan untuk mulai mengadopsi pendekatan evaluasi berbasis segmentasi secara sistematis dan tidak lagi mengandalkan Indeks Kepuasan Masyarakat (IKM) agregat sebagai satu-satunya parameter evaluasi kinerja. Secara praktis, strategi pelayanan harus disesuaikan dengan karakteristik spesifik setiap kelompok, bagi segmen dengan tingkat kepuasan menengah dan rendah, fokus perbaikan perlu diarahkan pada penyederhanaan prosedur, percepatan waktu layanan, serta penguatan sistem informasi layanan digital untuk menyelaraskan transparansi persyaratan dan biaya antara petugas dengan masyarakat. Implementasi transformasi digital ini krusial untuk memastikan aspek-aspek kritis layanan dapat terpenuhi secara optimal. Bagi peneliti selanjutnya, pengembangan model dapat dilakukan dengan menguji kerangka analisis ini pada berbagai jenis layanan publik dan wilayah yang berbeda guna meningkatkan validitas temuan. Selain itu, penggunaan algoritma clustering lain seperti DBSCAN serta penambahan variabel demografi sebagai variabel moderasi sangat disarankan untuk mendapatkan perbandingan stabilitas kluster dan melihat pengaruh karakteristik sosial terhadap dimensi persepsi yang dibentuk melalui PCA.

7. REFERENSI

Allgaier, J., & Pryss, R. (2024). Cross-Validation Visualized: A Narrative Guide to Advanced Methods. *Machine Learning and Knowledge Extraction*, 6(2), 1378–1388. <https://doi.org/10.3390/make6020065>

Anilshi, A. (2024). Penerapan Algoritma K-Means Clustering Nilai. 720–734.

Bach, A., & Thiel, F. (2024). Collaborative online learning in higher education—quality of digital interaction and associations with individual and group-related factors. *Frontiers in Education*, 9(November), 1–12. <https://doi.org/10.3389/educ.2024.1356271>

Bylemans, J., Everts, T., Brys, R., & Duncan, R. P. (2025). From anarchy to clarity, data pre-processing and statistical choices influence quantitative environmental DNA (eDNA) analyses. *Methods in Ecology and Evolution*, 16(7), 1322–1333. <https://doi.org/10.1111/2041-210X.70064>

Celestin, P. D. M. (2025). Principal Component Analysis For Simplifying Multivariate Financial Data In Portfolio Risk Analysis. *SSRN Electronic Journal*, 9(02), 171–179. <https://doi.org/10.2139/ssrn.5188109>

Chan, T. S. T., & Gibberd, A. (2025). Feasible model-based principal component analysis: Joint estimation of rank and error covariance matrix. *Computational Statistics and Data Analysis*, 201(July 2024), 108042. <https://doi.org/10.1016/j.csda.2024.108042>

Channamallu, S. S., Kermanshachi, S., Rosenberger, J. M., Pamidimukkala, A., & Hladik, G. (2025). Determinants of user satisfaction in smart parking applications. *Transport Economics and Management*, 3(April), 214–221. <https://doi.org/10.1016/j.team.2025.05.001>

Chitra, J., & Heikal, J. (2024). Customer segmentation using the K-Means Clustering algorithm in Foreign Banks in Indonesia. *Indonesia Accounting Research Journal*, 11(4), 230–241.

Delaosa, C., Pestana, J., Proudler, I. K., & Weiss, S. (2025). Impact of space-time covariance matrix estimation on bin-wise eigenvalue and eigenspace perturbations. *Signal Processing*, 233(February). <https://doi.org/10.1016/j.sigpro.2025.109946>

Deldadehasl, M., Karahoodi, H. H., & Haddadian Nekah, P. (2025). Customer Clustering and Marketing Optimization in Hospitality: A Hybrid Data Mining and Decision-Making Approach from an Emerging Economy. *Tourism and Hospitality*, 6(2), 1–19. <https://doi.org/10.3390/tourhosp6020080>

Dulger Altiner, D., Yıkımsı, S., Şimşek, M. A., Türkol, M., Tokatlı Demirok, N., & Celik, G. (2024). Impact of Thermosonication Treatment on Parsley Juice: Particle Swarm Algorithm (PSO), Multiple Linear Regression (MLR), and Response Surface Methodology (RSM). *ACS Omega*, 9(27), 29585–29597. <https://doi.org/10.1021/acsomega.4c02749>

Fathi, H., Cremona, M. A., & Severino, F. (2025). Selection of functional predictors and smooth coefficient estimation for scalar-on-function regression models. <http://arxiv.org/abs/2506.17773>

Grünwald, P., Lardy, T., Hao, Y., Bar-Lev, S. K., & de Jong, M. (2024). Optimal E-Values for Exponential Families: the Simple Case. *Mdl*, 1–28. <http://arxiv.org/abs/2404.19465>

Guo, G., Song, H., & Zhu, L. (2025). The iterated score regression estimation algorithm for PCA-based missing data with high correlation. *Scientific Reports*, 15(1), 1–27.

- <https://doi.org/10.1038/s41598-025-93333-6>
- Jullia, M., & Finatariyani, E. (2024). Pengaruh Pertumbuhan Perusahaan, Kepemilikan Manajerial dan Kepemilikan Institusional Terhadap Nilai Perusahaan. *AKADEMIK: Jurnal Mahasiswa Humanis*, 4(3), 913–923. <https://doi.org/10.37481/jmh.v4i3.1024>
- Kalantan, Z. I., Alharbi, L. S., Al-Zahrani, M. H., & Saleh Binhimd, S. M. (2025). Robust Dimensionality Reduction: A Bootstrap-Based Evaluation of PCA with Applications in Nutritional and Environmental Sciences. *Contemporary Mathematics (Singapore)*, 6(1), 923–942. <https://doi.org/10.37256/cm.6120256016>
- Kotronoulas, G., Miguel, S., Dowling, M., Fernández-Ortega, P., Colomer-Lahiguera, S., Bağcıvan, G., Pape, E., Drury, A., Semple, C., Dieperink, K. B., & Papadopoulou, C. (2023). An Overview of the Fundamentals of Data Management, Analysis, and Interpretation in Quantitative Research. *Seminars in Oncology Nursing*, 39(2), 1–9. <https://doi.org/10.1016/j.soncn.2023.151398>
- Lokanan, M. (2024). Harnessing Exploratory Data Analysis for Robust Financial Fraud Detection and Model Enhancement. *Preprint on Research Square / ResearchGate (Article Title: "Harnessing Exploratory Data Analysis for Robust Financial Fraud Detection and Model Enhancement")*. <https://www.researchsquare.com/article/rs-5635767/v1>
- Malakar, I., & Nepal, B. (2024). Conceptualizing Exploratory Data Analysis in Applied Statistics. *Patan Gyansagar*, 6(1), 46–63. <https://doi.org/10.3126/pg.v6i1.67406>
- Musfiroh, M., Novitasari, D. C. R., Intan, P. K., & Wisnawa, G. G. (2023). Penerapan Metode Principal Component Analysis (PCA) dan Long Short-Term Memory (LSTM) dalam Memprediksi Prediksi Curah Hujan Harian. *Building of Informatics, Technology and Science (BITS)*, 5(1), 1–11. <https://doi.org/10.47065/bits.v5i1.3114>
- Nguyen, H. M., Ho, T. K. T., & Ngo, T. T. (2024). The impact of service innovation on customer satisfaction and customer loyalty: a case in Vietnamese retail banks. *Future Business Journal*, 10(1), 1–15. <https://doi.org/10.1186/s43093-024-00354-0>
- Paap, K. R., Anders-Jefferson, R. T., Balakrishnan, N., & Majoubi, J. B. (2024). The many foibles of Likert scales challenge claims that self-report measures of self-control are better than performance-based measures. *Behavior Research Methods*, 56(2), 908–933. <https://doi.org/10.3758/s13428-023-02089-2>
- Sánchez Vinces, B. V., Schubert, E., Zimek, A., & Cordeiro, R. L. F. (2025). A comparative evaluation of clustering-based outlier detection. In *Data Mining and Knowledge Discovery* (Vol. 39, Issue 2). Springer US. <https://doi.org/10.1007/s10618-024-01086-z>
- Sansan Yasinta Ar Rohmah, Salsa Sahbani, & Poni Sukaesih. (2025). Public Satisfaction Index of Public Services at Pamarayan Community Health Center, Serang District, Banten Province. *Proceeding of International Conference on Business, Economics, Social Sciences, and Humanities*, 8, 35–46. <https://doi.org/10.34010/icobest.v8i.678>
- Sarkar, M., Puja, A. R., & Chowdhury, F. R. (2024). Optimizing Marketing Strategies with RFM Method and K-Means Clustering-Based AI Customer Segmentation Analysis. *Journal of Business and Management Studies ISSN: 2709–0876*, 8–22. <https://doi.org/10.32996/jbms>
- Setiawan, A., Utami, E., Ariatanto, D., & others. (2024). Cattle Weight Estimation Using Linear Regression and Random Forest Regressor. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 8(1), 72–79.
- Tan, W. (2024). *Big Data and Data Mining : Techniques for Discovering Hidden Insights*. 05(03), 12–15.
- Wang, G. (2025). Customer segmentation in the digital marketing using a Q-learning based differential evolution algorithm integrated with K-means clustering. *PLoS ONE*, 20(2 February), 1–21. <https://doi.org/10.1371/journal.pone.0318519>
- Wang, M., Kang, X., Liang, J., Wang, K., & Wu, Y. (2024). Heteroscedasticity identification and variable selection via multiple quantile regression. *Journal of Statistical Computation and Simulation*, 94(2), 297–314. <https://doi.org/10.1080/00949655.2023.2243533>
- Wang, Q. (2024). College Employment Recommendation Based on Improved K-Means Clustering and SimRank Algorithm in College Employment Management. *IEEE Access*, 12(July), 154230–154243. <https://doi.org/10.1109/ACCESS.2024.3450965>
- Yildirim, H. (2024). The Multicollinearity Effect on the Performance of Machine Learning Algorithms : Case Examples in Healthcare Modelling. *Academic Platform Journal of Engineering and Smart Systems*, 12(3), 68–80.
- Zhao, Y., Lang, P., & Wang, H. (2025). Dynamic study on the influencing factors and spatial and temporal evolution of cross-travel integration in Guangxi based on multiple regression computational analysis. *International Journal for Housing Science and Its Applications*, 46(3), 2371–2381. <https://doi.org/10.70517/ijhsa463198>
- Zhu, Y., Jiang, W., & Alonso, G. (2024). *Efficient Tabular Data Preprocessing of ML Pipelines*. <http://arxiv.org/abs/2409.14912>