



## ANALISIS KLASTERING K MEANS PADA DATA PENERIMAAN DAERAH TAHUN 1991-1995

Aulia Eka Ayu Putri<sup>1</sup>, Dwi Anisa<sup>2</sup>, Dica Cahyani Putri<sup>3</sup>, Trismayani<sup>4</sup>, Zurnan Alfian<sup>5</sup>

Program Studi Teknik Informatika Universitas Pamulang  
email: [auliaekaayu02@gmail.com](mailto:auliaekaayu02@gmail.com)<sup>1</sup>

Informasi Artikel	ABSTRACT
<p><b>Riwayat artikel :</b> Disubmit : 28 Juni 2025 Direvisi : 25 November 2025 Diterima : 10 Desember 2025 Dipublikasi : 22 Desember 2025</p>	<p><i>Analysis of regional financial data is an important step in understanding the characteristics and fiscal disparities between regions, particularly in the period prior to the implementation of regional autonomy. This study aims to group the types of regional revenue of districts/cities in Indonesia in the period 1991–1995 using the K-Means algorithm and to evaluate the quality of the clustering results in order to obtain representative clusters. The research method used is a data mining approach with the K-Means clustering technique, in which data in the form of the percentage of realisation and average regional revenue from 16 types of revenue are analysed using the Python programming language. The research stages included data pre-processing, normalisation using StandardScaler, determining the optimal number of clusters, and cluster evaluation using Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index. The results showed that the optimal number of clusters was three, with a Silhouette Score of 0.7234, a Calinski-Harabasz Index of 593.93, and a Davies-Bouldin Index of 0.8689, indicating good clustering quality. The cluster distribution shows the imbalance in regional revenue structures, where one cluster is dominant while the other clusters have unique characteristics. In conclusion, the K-Means algorithm is effective in identifying pre-autonomous regional revenue patterns and can be used as a basis for analysing regional structures and fiscal imbalances.</i></p>
<p><b>Keywords:</b> K-Means, clustering, regional revenue, data mining, regional autonomy, Python.</p>	
	<b>ABSTRAK</b>
<p><b>Kata Kunci:</b> K-Means, clustering, penerimaan daerah, data mining, otonomi daerah, Python.</p>	<p>Analisis data keuangan daerah merupakan langkah penting untuk memahami karakteristik dan ketimpangan fiskal antar wilayah, khususnya pada periode sebelum penerapan otonomi daerah. Penelitian ini bertujuan untuk mengelompokkan jenis penerimaan daerah kabupaten/kota di Indonesia pada periode 1991–1995 menggunakan algoritma K-Means serta mengevaluasi kualitas hasil clustering guna memperoleh kluster yang representatif. Metode penelitian yang digunakan adalah pendekatan data mining dengan teknik clustering K-Means, di mana data berupa persentase realisasi dan rata-rata penerimaan daerah dari 16 jenis penerimaan dianalisis menggunakan bahasa pemrograman Python. Tahapan penelitian meliputi pra-pemrosesan data, normalisasi menggunakan StandardScaler, penentuan jumlah kluster optimal, serta evaluasi kluster menggunakan Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index. Hasil penelitian menunjukkan bahwa jumlah kluster optimal adalah tiga kluster dengan nilai Silhouette Score sebesar 0,7234, Calinski-Harabasz Index sebesar 593,93, dan Davies-Bouldin Index sebesar 0,8689, yang menandakan kualitas pengelompokan yang baik. Distribusi kluster memperlihatkan ketimpangan struktur penerimaan daerah, di mana satu kluster bersifat dominan sementara kluster lainnya memiliki karakteristik unik. Kesimpulannya, algoritma K-Means efektif dalam mengidentifikasi pola penerimaan daerah pra-otonomi dan dapat digunakan sebagai dasar analisis struktur serta ketimpangan fiskal daerah.</p>





## PENDAHULUAN

Analisis data keuangan daerah merupakan aspek penting dalam memahami karakteristik dan kinerja fiskal suatu wilayah. Salah satu pendekatan yang banyak digunakan untuk menggambarkan karakteristik sekelompok wilayah berdasarkan kemiripan variabel tertentu adalah teknik *clustering*. Clustering bertujuan untuk mengelompokkan objek atau observasi sehingga anggota dalam satu kelompok memiliki tingkat kemiripan yang tinggi dibandingkan dengan kelompok lainnya.

Dalam ranah *data mining*, clustering merupakan bagian dari proses penemuan pengetahuan dari data yang meliputi tahapan pengumpulan, transformasi, dan analisis data. Data mining didefinisikan sebagai proses mengekstraksi pola atau informasi bernilai dari kumpulan data berukuran besar melalui tahapan pembersihan, integrasi, seleksi, transformasi, dan evaluasi pola (Han & Kamber, 2006; Rosyadi et al., 2022). Pendekatan ini dinilai efektif dalam menganalisis data numerik berskala besar, termasuk data keuangan daerah.

Salah satu algoritma clustering yang paling banyak digunakan adalah *K-Means*. Algoritma ini termasuk metode non-hierarki yang efisien dan sederhana, serta bekerja dengan membagi data ke dalam sejumlah klaster berdasarkan kedekatan jarak terhadap pusat klaster (*centroid*), yang umumnya dihitung menggunakan jarak Euclidean (Suliman, 2021). Berbagai penelitian menunjukkan bahwa K-Means mampu menghasilkan pengelompokan yang representatif pada data keuangan daerah, seperti pengelompokan wilayah berdasarkan pendapatan asli daerah (PAD), dana perimbangan, dan belanja modal (Siregar et al., 2019; Wuland & Hidayat, 2020).

Dalam konteks pemerintahan Indonesia, analisis keuangan daerah tidak dapat dipisahkan dari kebijakan desentralisasi dan otonomi daerah. Pemberian otonomi yang lebih luas kepada pemerintah daerah bertujuan untuk meningkatkan efektivitas pembangunan dan kemandirian fiskal. Namun, sebelum otonomi daerah diberlakukan secara resmi pada tahun 2001, struktur penerimaan daerah masih sangat dipengaruhi oleh kebijakan pemerintah pusat sehingga menunjukkan pola fiskal yang beragam antar wilayah. Analisis terhadap data historis penerimaan daerah pada periode ini menjadi penting sebagai dasar evaluasi kebijakan fiskal di masa awal desentralisasi. Beberapa penelitian sebelumnya telah menerapkan algoritma K-Means dalam analisis keuangan daerah. Suliman (2021) menunjukkan efektivitas K-Means dalam mengelompokkan wilayah berdasarkan karakteristik pendapatan, sedangkan Sari (2015) memanfaatkan algoritma ini dalam analisis dan prediksi anggaran pendapatan daerah. Penelitian lain oleh Fatkhudin et al. (2023) juga menegaskan bahwa K-Means mampu menangani data numerik berskala besar dengan hasil klaster yang stabil dan informatif. Meskipun demikian, sebagian





besar penelitian tersebut berfokus pada data pasca-otonomi daerah dan belum banyak mengkaji data historis penerimaan daerah sebelum tahun 2000.

Berdasarkan kajian tersebut, masih terdapat celah penelitian (research gap) berupa keterbatasan studi yang menganalisis data penerimaan daerah kabupaten/kota di Indonesia pada periode 1991–1995 menggunakan algoritma K-Means dengan evaluasi kualitas kluster secara komprehensif. Oleh karena itu, kebaruan (novelty) penelitian ini terletak pada pemanfaatan data historis pra-otonomi daerah serta penggunaan beberapa metrik evaluasi kluster, yaitu Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index, untuk memperoleh hasil pengelompokan yang lebih akurat dan bermakna.

Berdasarkan latar belakang dan celah penelitian tersebut, tujuan penelitian ini adalah mengelompokkan jenis penerimaan daerah kabupaten/kota di Indonesia pada periode 1991–1995 menggunakan algoritma K-Means, serta mengevaluasi kualitas hasil clustering guna memperoleh kluster yang representatif sebagai bahan analisis struktur dan ketimpangan fiskal daerah.

## METODE PENELITIAN

### 1. Tools dan Teknologi

Pengolahan data dilakukan menggunakan Bahasa pemrograman python dengan memanfaatkan beberapa pustaka utama:

- a) Pandas (McKinney, 2018) - Merupakan pustaka yang ditulis dalam bahasa Python dengan peran komputasi sebagai pengolah data. (Albanna et al., 2022)
- b) Numpy (Oliphant, 2015) - Adalah library yang digunakan untuk komputasi numerik dalam Python. (Kelly Hermanto et al., 2022).
- c) Matplotlib (Hunter, 2007) - Adalah teknik grafik untuk paket visualisasi data dengan Python yang cakupannya sebagai aspek integral dalam tumpukan data sains dan mudah didukung oleh teknik NumPy, Pandas, dan teknik lainnya yang relevan. (Suharsono & Suparwanto, 2023).  
Tools ini digunakan untuk membuat visualisasi grafis dari data dan hasil analisis.
- d) Sklearn (David Cournapeau, 2007), Scikit-Learn yakni perpustakaan utama untuk bahasa pemrograman Python yang biasanya digunakan dalam proyek pembelajaran mesin. (Rifqi Mulyawan, 17 Juni 2025)
- e) Vscode (Microsoft, 2015) - Merupakan sebuah teks editor ringan dan handal yang dibuat oleh Microsoft untuk sistem operasi multiplatform, artinya tersedia juga untuk versi Linux, Mac, dan Windows. (Kherina Surya Ningsih et al., 2022). Tools ini digunakan untuk menulis, mengedit dan mendebug code. (Populasi, Menggunakan, and Untuk 2025).





## 2. Algoritma

- Impor Pustaka: pandas untuk manipulasi data, StandardScaler untuk normalisasi fitur, KMeans untuk algoritma clustering.
- Pengambilan Data: Membaca file CSV (dataset\_penerimaan\_daerah.csv) ke dalam DataFrame df.
- Pra-pemrosesan Data: Memisahkan fitur (X) dari label (Jenis Penerimaan), Menormalkan fitur menggunakan StandardScaler sehingga data memiliki rata-rata 0 dan standar deviasi 1.
- Clustering K-Means: Mengatur jumlah cluster  $k = 3$ , Membuat objek KMeans dan menjalankan metode fit\_predict() untuk melakukan pengelompokan hasil pengelompokan disimpan dalam kolom baru "Cluster" pada DataFrame.
- Simpan Hasil: Menyimpan DataFrame yang sudah dilengkapi label cluster ke file hasil\_clustering\_penerimaan.csv.

```

import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans

# Baca data dari CSV
filename = "dataset_penerimaan_daerah.csv"
df = pd.read_csv(filename)

# Pisahkan kolom fitur dan label
jenis = df["Jenis Penerimaan"]
X = df.drop(columns=["Jenis Penerimaan"])

# Normalisasi (StandardScaler)
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Clustering
k = 3
kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
clusters = kmeans.fit_predict(X_scaled)

# Gabungkan hasil ke DataFrame
df["Cluster"] = clusters

# Simpan hasil ke CSV
output_file = "hasil_clustering_penerimaan.csv"
df.to_csv(output_file, index=False)

```

Gambar 1. Code Python Algoritma Clustering K-Means

## 3. Evaluasi Model

Program ini mengevaluasi kualitas hasil clustering dari file CSV menggunakan tiga metrik:

- Silhouette Score (0.7234): Menunjukkan pemisahan cluster yang baik. Nilai mendekati 1 menandakan cluster yang jelas dan terpisah.
- Calinski-Harabasz Index (593.9333): Menunjukkan rasio variansi antar-cluster terhadap intra-cluster. Nilai tinggi menunjukkan kualitas cluster yang baik.
- Davies-Bouldin Index (0.8689): Mengukur kesamaan antar cluster. Nilai rendah menunjukkan cluster yang terpisah dengan baik dan kompak.





- d) Distribusi Cluster: Cluster 0: 10 anggota (dominant), Cluster 1: 1 anggota, Cluster 2: 1 anggota distribusi ini menunjukkan ketidakseimbangan, yang mungkin perlu dianalisis lebih lanjut.
- e) Koordinat Centroid: Menunjukkan posisi rata-rata setiap cluster dalam ruang fitur. Perbedaan antar centroid menunjukkan pemisahan cluster yang baik dan membantu memahami karakteristik masing-masing cluster.

```

=== Evaluasi Clustering dari File CSV ===
Silhouette Score      : 0.7234
Calinski-Harabasz Index : 593.9333
Davies-Bouldin Index  : 0.0689

=== Jumlah Anggota per Cluster ===
Cluster
0      10
1       1
2       1
2       1
Name: count, dtype: int64

=== Koordinat Centroid Tiap Cluster ===
1991/1992 1992/1993 1993/1994 1994/1995
0  0.461863  0.450671  0.530793  0.554119
1  3.214303  3.218787  3.200553  3.194181
2  -0.367617 -0.366946 -0.373135 -0.374830
PS C:\Users\LENOVO\OneDrive\Documents\DATAMININGJURNAL>

```

Gambar 2. Hasil Python Evaluasi

#### 4. Dataset

Dataset ini menggambarkan persentase realisasi dan rata-rata penerimaan daerah kabupaten/kota di Indonesia berdasarkan 16 jenis penerimaan selama tahun anggaran 1991/1992 hingga 1994/1995. Data mencakup kategori seperti sisa anggaran, pendapatan asli daerah (termasuk pajak daerah, retribusi, laba BUMD), bagi hasil pajak/non-pajak, sumbangan, dan penerimaan pembangunan. Data bersumber dari laporan resmi *Statistik Penerimaan Daerah Tahun 1995* oleh Kementerian Keuangan RI, dan memberikan gambaran struktur keuangan daerah pada masa tersebut. (Sari 2015),

```

penerimaan_daerah.csv
1  Jenis Penerimaan,1991/1992,1992/1993,1993/1994,1994/1995,Rata-rata
2  Sisa lebih perhitungan tahun lalu,2.97,2.49,2.33,2.49,2.57
3  Pendapatan asli daerah,13.45,12.66,11.59,11.50,12.63
4  Pajak daerah,3.64,3.43,4.26,4.36,3.92
5  Retribusi daerah,7.48,6.81,6.38,6.52,6.80
6  Bagian laba BUMD,0.41,0.37,0.55,0.62,0.53
7  Penerimaan Dinas-dinas,0.54,0.45,0.5,0.62,0.53
8  Lain-lain,1.38,1.60,1.3,1.14,1.21
9  Bagi hasil pajak/bukan pajak,12.41,12.74,13.59,15.18,13.48
10 Bagi hasil pajak,10.82,11.22,11.94,13.42,11.85
11 Bagi hasil bukan pajak,1.59,1.52,1.65,1.75,1.63
12 Sumbangan dan bantuan,70.28,71.58,71.87,68.35,70.52
13 Sumbangan,29.19,29.83,33.67,33.29,31.5
14 Bantuan,41.09,41.75,38.2,35.06,39.03
15 Penerimaan Pembangunan,0.89,0.83,0.62,0.86,0.80
16 Pinjaman Penda,0.87,0.81,0.61,0.76,0.76
17 Pinjaman BUMD,0.02,0.02,0.02,0.1,0.04

```

Gambar 3. Dataset Penerimaan Anggaran Tahunan





Selain untuk analisis data penerimaan daerah, metode clustering juga dapat diterapkan pada berbagai dataset lain. Misalnya, dalam bidang pendidikan, clustering calon peserta mata kuliah berdasarkan prasyarat yang telah diambil (seperti pada data Prasyarat Mata Kuliah ITI 372) dapat membantu universitas merancang kurikulum yang lebih adaptif dan memberikan rekomendasi mata kuliah yang sesuai.

```

hasil_clustering_peserta.csv
1  AMA214,AMA192,AMA101,ITI171,AMA213,Cluster,Interpretasi
2  1,1,2,2,2,0,Nilai Rendah - Perlu Bimbingan
3  4,2,4,4,2,2,Nilai Sedang - Cukup Siap
4  2,2,2,2,2,0,Nilai Rendah - Perlu Bimbingan
5  3,3,3,2,2,0,Nilai Rendah - Perlu Bimbingan
6  3,2,3,1,3,0,Nilai Rendah - Perlu Bimbingan
7  4,4,4,4,4,1,Nilai Tinggi - Sangat Siap
8  3,4,3,3,4,1,Nilai Tinggi - Sangat Siap
9  2,1,3,2,2,0,Nilai Rendah - Perlu Bimbingan
10 2,2,2,2,1,0,Nilai Rendah - Perlu Bimbingan
11 2,1,2,2,2,0,Nilai Rendah - Perlu Bimbingan
12 4,4,4,4,3,1,Nilai Tinggi - Sangat Siap
13 3,2,3,3,3,2,Nilai Sedang - Cukup Siap
14 4,4,3,3,4,1,Nilai Tinggi - Sangat Siap
15 3,2,3,1,3,0,Nilai Rendah - Perlu Bimbingan
16 4,4,4,4,2,2,Nilai Sedang - Cukup Siap
17 3,3,2,2,2,0,Nilai Rendah - Perlu Bimbingan
18 1,2,3,3,4,0,Nilai Rendah - Perlu Bimbingan
19 3,3,2,3,3,1,Nilai Tinggi - Sangat Siap
20 1,2,2,1,1,0,Nilai Rendah - Perlu Bimbingan
21 3,4,3,3,4,1,Nilai Tinggi - Sangat Siap
    
```

Gambar 4. Dataset Nilai Mata Kuliah Prasyarat ITI 372

Di bidang kesehatan, data biometrik individu seperti tinggi, berat, dan lingkar lengan (Data Mahasiswa Tinggi, Berat, LLB) dapat dikelompokkan untuk mengidentifikasi segmen populasi dengan karakteristik fisik serupa, yang berguna untuk studi nutrisi atau kesehatan Masyarakat.

```

documen.csv
1  Mahasiswa,BMI,Ukuran Kerangka
2  Mahasiswa 1,22.21,11.04
3  Mahasiswa 2,23.87,9.95
4  Mahasiswa 3,19.71,10.93
5  Mahasiswa 4,21.05,10.38
6  Mahasiswa 5,17.93,12.85
7  Mahasiswa 6,17.72,12.00
8  Mahasiswa 7,18.71,11.53
9  Mahasiswa 8,25.86,9.33
10 Mahasiswa 9,19.15,11.80
11 Mahasiswa 10,18.42,11.20
12 Mahasiswa 11,22.94,10.60
13 Mahasiswa 12,26.89,10.44
14 Mahasiswa 13,24.91,10.68
15 Mahasiswa 14,22.99,11.47
16 Mahasiswa 15,26.81,9.17
17 Mahasiswa 16,19.14,12.11
18 Mahasiswa 17,21.09,10.67
19 Mahasiswa 18,18.71,12.36
20
    
```

Gambar 5. Dataset Tinggi, Berat, LLB





Bahkan dalam konteks makanan, analisis pola rating makanan dalam 2 aplikasi online dapat dianalisis untuk mengidentifikasi tiap aplikasi dengan rating yang berbeda, yang dapat membantu dalam perbandingan diantara 2 aplikasi tersebut. (Farhan Nugraha, Martano, and Hayati 2024)

1	Menu	Rating_App#1	Rating_App#2	Cluster
2	Ayam Bakar	2	5	1
3	Soto Betawi	2	5.5	1
4	Seblak	5	3.5	0
5	Dim Sum	6.5	2.2	0
6	Goyong	7	3.3	0
7	Gado-gado	3.5	4.8	1
8	Gudeg	4	4.5	1
9	Bakso	5	4.0	0
10	Nasi Goreng	6	5.5	0
11	Rendang	4.5	4.5	1
12	Sate Ayam	5	5.0	0
13	Pecel Lele	3.5	4.0	1
14	Mie Ayam	3	4	1

Gambar 6. Dataset Rating Makanan Online

## 5. Pengolahan Data

- Pengumpulan Data:** Data yang digunakan berasal dari file csv. Dataset ini berisi informasi berbagai jenis penerimaan daerah beserta nilai-nilai kuantitatif lainnya yang relevan untuk dianalisis.
- Pemisahan fitur & label:** Untuk analisis kluster, data dipisahkan menjadi, Fitur (X): kolom-kolom numerik (selain Jenis Penerimaan) dan Label (y): kolom Jenis Penerimaan hanya digunakan untuk identifikasi, bukan untuk pemodelan.
- Normalisasi Data:** Data dinormalisasi menggunakan metode StandardScaler, yaitu menstandarkan fitur agar memiliki rata-rata 0 dan standar deviasi 1. Tujuannya adalah untuk menghindari dominasi fitur yang memiliki skala besar dalam proses clustering.
- Clustering dengan K-means:** Algoritma K-Means digunakan untuk melakukan proses pengelompokan data ke dalam 3 kluster ( $k = 3$ ), yang diasumsikan mewakili kelompok jenis penerimaan yang memiliki karakteristik serupa.
- Penyimpanan Hasil:** Hasil pengelompokan disimpan dalam kolom baru bernama Cluster pada DataFrame dan kemudian diekspor ke file CSV untuk dokumentasi lebih lanjut. (Setiadi et al. 2023)



## HASIL DAN PEMBAHASAN

### 1. Data Selection

Data selection merupakan tahapan dimana dilakukan proses pengumpulan dataset awal dengan melakukan pemilihan variabel. Data yang akan digunakan yaitu data distribusi persentase realisasi penerimaan daerah kabupaten/kota seluruh Indonesia menurut jenis penerimaan tahun 1991-1995 yang didapat dari sumber Mardiasmo (1999). Pemilihan periode waktu ini penting mengingat data tersebut mencerminkan struktur fiskal sebelum implementasi desentralisasi fiskal yang dimulai pada tahun 2001, sebagaimana dijelaskan dalam laporan Fiscal Policy Agency (2022) bahwa periode pra-otonomi menunjukkan pola ketergantungan daerah yang tinggi terhadap pemerintah pusat.

### 2. Data Mining

Penelitian ini menggunakan teknik data mining berupa clustering dengan algoritma K-Means untuk menganalisis data persentase realisasi dan rata-rata penerimaan daerah kabupaten/kota di Indonesia tahun 1991/1992–1994/1995. Tujuannya adalah mengelompokkan jenis penerimaan ke dalam beberapa kluster dengan karakteristik keuangan yang mirip. Penentuan jumlah kluster optimal dilakukan menggunakan metode Elbow, yang menunjukkan nilai optimal sebanyak 3 kluster. Model K-Means kemudian dibangun menggunakan bahasa pemrograman Python melalui Visual Studio Code.

Hasil klasterisasi menunjukkan pembagian yang mencerminkan tingkat variasi penerimaan daerah. Pemilihan  $k=3$  ini didukung oleh evaluasi Davies Bouldin Index sebesar 0.8689, di mana nilai DBI yang lebih rendah mengindikasikan kualitas kluster yang lebih baik (Muningsih et al., 2021). Penelitian Khalif et al. (2024) dalam analisis kemiskinan di Indonesia juga menggunakan metode serupa dan menemukan bahwa tiga kluster optimal mampu mengidentifikasi pola disparitas regional dengan baik.

Proses klasterisasi pada data keuangan daerah ini konsisten dengan temuan penelitian sebelumnya yang menunjukkan bahwa algoritma K-Means efektif dalam mengidentifikasi pola pada data fiskal. Studi Etrisia et al. (2023) tentang inflasi regional di Indonesia menggunakan pendekatan yang sama dan memperoleh hasil yang stabil dengan Silhouette Coefficient 0.66, yang menunjukkan bahwa metode ini reliabel untuk data ekonomi Indonesia.





```

=== Hasil Clustering ===
                Jenis Penerimaan Cluster
0  Sisa lebih perhitungan tahun lalu      0
1                Pajak daerah             0
2          Retribusi daerah                0
3          Bagian laba BUMD               0
4  Penerimaan Dinas-dinas                 0
5                Lain-lain                 0
6          Bagi hasil pajak                2
7  Bagi hasil bukan pajak                  0
8                Sumbangan                 1
9                Bantuan                   0
10 Pinjaman Pemda                         0
11 Pinjaman BUMD                          0

```

Gambar 7. Hasil Clustering

### 3. Interpretation/Evaluation

Setelah melakukan pemodelan dengan algoritma K-Means clustering yang menghasilkan 3 cluster pada dataset 1991-1995, evaluasi hasil dilakukan menggunakan tiga metrik utama untuk memastikan kualitas klasterisasi.

Evaluasi kuantitatif menunjukkan kualitas klasterisasi yang baik. Silhouette Score sebesar 0.7234 mengindikasikan pemisahan yang jelas antar klaster, nilai ini menunjukkan performa yang lebih baik dibandingkan dengan penelitian Sari et al. (2023) yang memperoleh nilai 0.58 pada analisis inflasi dengan metode K-Means. Calinski-Harabasz Index 593.93 menunjukkan rasio variansi antar-klaster terhadap intra-klaster yang tinggi, mengonfirmasi homogenitas dalam klaster dan heterogenitas antar klaster yang kuat. Davies-Bouldin Index 0.8689 menunjukkan tingkat kesamaan antar klaster yang rendah, sejalan dengan prinsip yang dikemukakan oleh Muningsih et al. (2021) bahwa nilai DBI yang lebih kecil menandakan klaster yang lebih terpisah dengan baik.

Distribusi anggota klaster menunjukkan pola yang tidak merata namun bermakna. Klaster 0 dengan 10 anggota mencerminkan jenis penerimaan yang dominan dalam struktur fiskal daerah seperti Pendapatan Asli Daerah (PAD), dana perimbangan, dan sumbangan daerah. Klaster 1 dan 2 yang masing-masing hanya berisi 1 anggota menunjukkan karakteristik

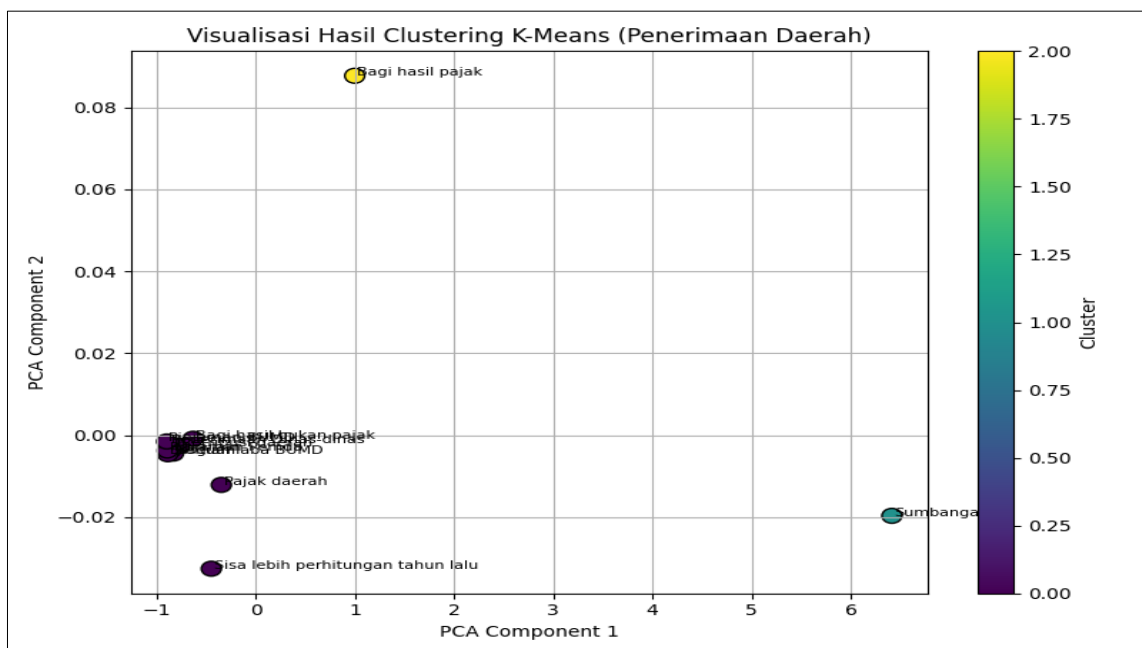




unik yang kemungkinan mewakili sisa anggaran tahun sebelumnya dan penerimaan pembangunan yang memiliki pola fluktuatif berbeda dari jenis penerimaan lainnya.

Temuan ini memiliki implikasi penting dalam konteks struktur fiskal pra-desentralisasi. Seperti yang dijelaskan oleh Wibowo (2010) dalam analisis desentralisasi fiskal di Indonesia, mayoritas pemerintah daerah pada periode tersebut memiliki kapasitas yang relatif rendah untuk menghasilkan pendapatan sendiri dan masih sangat bergantung pada transfer dari pemerintah pusat. Pola klasterisasi yang dihasilkan dalam penelitian ini mengonfirmasi temuan tersebut, di mana penerimaan daerah didominasi oleh beberapa jenis penerimaan utama (terwakili dalam Klaster 0), sementara jenis penerimaan lainnya menunjukkan karakteristik yang sangat berbeda.

Visualisasi hasil clustering melalui diagram PCA mempermudah pemahaman terhadap pola pengelompokan penerimaan daerah. Penelitian Rahmanesta et al. (2024) dalam analisis industri mikro dan kecil di Indonesia menunjukkan bahwa visualisasi PCA efektif untuk mengidentifikasi pola spasial dalam data regional. Dalam konteks penelitian ini, visualisasi membantu mengidentifikasi kesenjangan struktur penerimaan antar wilayah, yang dapat dijadikan acuan dalam evaluasi kebijakan fiskal daerah.



Gambar 8. Hasil Visualisasi Clustering K-Means





Koordinat centroid yang dihasilkan menunjukkan posisi rata-rata setiap kluster dalam ruang fitur. Perbedaan signifikan antar centroid mengindikasikan pemisahan kluster yang baik dan membantu memahami karakteristik keuangan dari masing-masing kelompok. Kluster dengan nilai centroid yang lebih tinggi pada dimensi PAD dan bagi hasil pajak menunjukkan wilayah dengan kemandirian fiskal yang relatif lebih baik, sementara kluster dengan centroid yang lebih tinggi pada dimensi sumbangan dan subsidi mencerminkan wilayah yang masih sangat bergantung pada transfer pusat.

Hasil penelitian ini memberikan baseline empiris untuk memahami evolusi struktur penerimaan daerah pasca implementasi otonomi daerah. Seperti yang diungkapkan dalam laporan "Two Decades of Fiscal Decentralization Implementation in Indonesia" oleh Fiscal Policy Agency (2022), desentralisasi fiskal yang dimulai tahun 2001 membawa perubahan signifikan dalam struktur penerimaan daerah. Perbandingan dengan data historis periode 1991-1995 ini penting untuk mengevaluasi dampak kebijakan desentralisasi terhadap kemandirian fiskal daerah dalam jangka panjang.

## SIMPULAN

Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan data penerimaan daerah kabupaten/kota di Indonesia pada periode 1991–1995 (Nabil and Dwiridotjahjono 2024). Data diambil dari laporan resmi Kementerian Keuangan berjudul Statistik Penerimaan Daerah Tahun 1995, yang mencakup 16 jenis penerimaan, seperti sisa anggaran, PAD, dana perimbangan, dan penerimaan pembangunan. (Saleh 2020)

Proses data mining terdiri dari tiga tahap: pengumpulan, transformasi, dan analisis data. Implementasi dilakukan menggunakan Python dan Visual Studio Code, dengan langkah-langkah seperti impor pustaka, membaca file CSV, normalisasi data, clustering dengan  $k=3$ , dan menyimpan hasil ke file baru. (Ramadhan and Wibowo 2024)

Hasil clustering menunjukkan distribusi tidak merata: cluster 0 berisi 10 anggota, sedangkan cluster 1 dan 2 masing-masing hanya 1 anggota. Kluster yang terbentuk mencerminkan kelompok daerah berpendapatan tinggi, sedang, dan rendah, serta menyoroti kesenjangan fiskal dan perlunya otonomi daerah. Penelitian ini menunjukkan efektivitas K-Means dalam mengidentifikasi pola pertumbuhan fiskal di awal era desentralisasi. (K-medoids 2024)





## DAFTAR RUJUKAN

- Farhan Nugraha, Muhamad, Martanto Martano, and Umi Hayati. 2024. "Clustering Data Indonesian Food Delivery Menggunakan Metode K-Means Pada Gofood Product List." *JATI (Jurnal Mahasiswa Teknik Informatika)* 8(3): 3484–92. doi:10.36040/jati.v8i3.9727.
- Fatkhudin, Aslam, Ahmad Khambali, Fenilinas A Artanto, and Nabil A Putra Zade. 2023. "Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika)." *Jurnal Minfo Polgan* 12(2): 777–83. <https://jurnal.polgan.ac.id/index.php/jmp/article/view/12494/1682>.
- K-medoids, Menggunakan Algoritma K-means. 2024. "Klasterisasi Tingkat Kemiskinan Kabupaten / Kota Di Indonesia." : 197–208. doi:10.33364/algoritma/v.21-2.1788.
- Kannia, Novranza, and Frieyadie Frieyadie. 2022. "Analisa Pemilihan Aplikasi Pemesanan Makanan Online Menggunakan Metode Analytical Hierarchy Process (AHP)." *Jurnal Sosial Teknologi* 2(11): 972–83. doi:10.36418/jurnalsostech.v2i11.479.
- Nabil, Nibrosun., and Jojok Dwiridotjahjono. 2024. "Al-Kharaj : Jurnal Ekonomi , Keuangan & Bisnis Syariah Al-Kharaj : Jurnal Ekonomi , Keuangan & Bisnis Syariah." *Al-Kharaj : Jurnal Ekonomi , Keuangan & Bisnis Syariah* 6(2): 2547–62. doi:10.47467/alkharaj.v7i1.5084.
- Populasi, Analisis, Jakarta Menggunakan, and Python Untuk. 2025. "Analisis Populasi Jakarta 2010 – 2025 Menggunakan Python Untuk Visualisasi Dan Pemodelan Data." 5: 1–14.
- Ramadhan, Raehan, and Arief Wibowo. 2024. "IMPLEMENTASI ALGORITMA K-MEANS UNTUK IMPLEMENTATION OF K-MEANS ALGORITHM FOR CLUSTERING JAKARTA URBAN VILLAGES BASED ON BULOG FOOD ASSISTANCE." 3(September): 390–97.
- Saleh, Rahmat. 2020. "Pengaruh Pendapatan Asli Daerah (Pad) Dan Dana Perimbangan Terhadap Kemandirian Keuangan Pemerintah Daerah." *Jurnal Informasi, Perpajakan, Akuntansi, Dan Keuangan Publik* 15(2): 111–34. doi:10.25105/jipak.v15i2.6226.
- Sari, Rahayu Mayang. 2015. "Prediksi Data Anggaran Pendapatan Belanja Daerah Menggunakan Algoritma K-Means." *Sains dan Teknologi Informasi* 1(2): 1–6.
- Setiadi, Dede, Bambang Irawan, Agus Bahtiar, Teknik Informatika, Sistem Informasi, Cirebon Indonesia, Knowledge Discovery, Algoritma K-means, and Produksi Perikanan Budidaya. 2023. "Penerapan Algoritma K-Means Clustering Pada Pembesaran." 7(6): 3320–27.





- Suliman, Suliman. 2021. "Implementasi Data Mining Terhadap Prestasi Belajar Mahasiswa Berdasarkan Pergaulan Dan Sosial Ekonomi Dengan Algoritma K-Means Clustering." *Simkom* 6(1): 1–11. doi:10.51717/simkom.v6i1.48.
- Tsabita, H. 2024. "Klasterisasi Provinsi Di Indonesia Berdasarkan Ketimpangan Pembangunan Ekonomi Dengan Pendekatan Algoritma K-Means Dan DBSCAN." 14(2). <https://repository.its.ac.id/110521/>.
- Wiradiansya, Edoar, Lena Elfianty, and Jhoanne Fredricka. 2024. "Klasterisasi Data Kendaraan Bermotor Berdasarkan Tunggakan Pajak Pada Kantor Samsat Kabupaten Bengkulu Selatan Menggunakan Metode K-Means Clustering." 6: 164–73.
- Wu, Hao, Shu Li, Yingpin Chen, and Zhenming Peng. 2019. "Bayesian Acoustic Impedance Inversion with Gamma Distribution." *Journal of Physics: Conference Series* 1237(2). doi:10.1088/1742-6596/1237/2/022034.

