

# Penerapan Algoritma K-Means Clustering Dalam Pengelompokkan Kepadatan Penduduk

## Application of K-Means Clustering Algorithm in Population Density Grouping

Fenita Delia<sup>1\*</sup> , Hajra Rasmita Ngemba<sup>2</sup> , Syaiful Hendra<sup>3</sup> , Syahrullah<sup>4</sup> , Nouval Trezandy

Lapatta<sup>5</sup> 

<sup>1,2,4</sup>Program Studi Sistem Informasi, Universitas Tadulako, Palu, Indonesia

<sup>3,5</sup>Program Studi Teknik Informatika, Universitas Tadulako, Palu, Indonesia

<sup>1</sup>fenitadelia@gmail.com, <sup>2</sup>hajra.rasmita@gmail.com, <sup>3</sup>syaiful.hendra.garuda@gmail.com, <sup>4</sup>syahroellahms@gmail.com,

<sup>5</sup>nouval@untad.ac.id

\*Penulis Korespondensi

### Artikel Info

#### Riwayat Artikel:

Penyerahan April 27, 2024

Revisi Juli 1, 2024

Diterima Januari 14, 2025

Diterbitkan Februari 21, 2025

#### Kata Kunci:

Codeigniter 4

Nahdlatul Ulama Scholarship

(NUS)

SQL

SDLC

#### Keywords:

Codeigniter 4

Nahdlatul Ulama Scholarship

(NUS)

SQL

SDLC



### ABSTRAK

Kepadatan penduduk yang tidak merata akan memberikan dampak buruk jika tidak diperhatikan. Salah satu cara untuk menanggulangi **masalah** ini dengan kebijakan perencanaan pengelolaan pemerataan kependudukan. Penelitian ini **berfokus** dalam mengelompokkan wilayah kepadatan penduduk yang merupakan rasio antara jumlah penduduk dan luas wilayah di Provinsi Sulawesi Tengah. Pengelompokkan penelitian ini diterapkan dengan teknik **data mining**, yakni K-Means Clustering. Tahapan penelitian yakni pengumpulan data, data understanding, data processing, clustering, cluster review, analisa dashboard dan uji akurasi dengan **aplikasi tableau** dalam memberikan visualisasi kepadatan penduduk di wilayah tersebut. Berdasarkan **hasil** perhitungan algoritma tersebut, menghasilkan tiga cluster dengan cluster 0 merupakan kepadatan penduduk rendah, cluster 1 merupakan kepadatan penduduk tinggi, dan cluster 2 merupakan kepadatan penduduk sedang. Pembentukan cluster berdasarkan visualisasi yang dihasilkan dataset penelitian melalui analisa *Sum Of Square Error*, silhouette coefficient, serta elbow method. Clustering yang dibentuk, dilanjutkan visualisasi dashboard dengan aplikasi tableau. **Hasil** clustering tersebut, berdasarkan perhitungan SSE menghasilkan nilai 4324505738.747303 berarti penetapan jumlah cluster dengan perbedaan signifikan dengan perhitungan jumlah pengelompokkan sebelumnya. Lalu hasil analisa silhouette memberikan nilai rata-rata silhouette paling tinggi pada jumlah cluster yakni 3 dengan nilai 0.6144435666457168, dan metode **elbow** memberikan hasil bahwa elbow point berada di titik 3, berarti jumlah cluster yang optimum dengan 3 cluster.

Ini adalah artikel akses terbuka di bawah [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.



### ABSTRACT

Uneven population density will have a negative impact if not considered. One way to tackle this **problem** is with population equity management planning policies. This research **focuses** on clustering population density areas, which is the ratio between population and area in Central Sulawesi Province. This research clustering is applied with **data mining techniques**, namely K-Means Clustering. The research stages are data collection, data understanding, data processing, clustering, clustering review, dashboard analysis, and accuracy testing with the **tableau application** in providing visualization of population density in the region. Based on the **results** of the algorithm calculation, it produces three clusters,

cluster 0 being low population density, cluster 1 being high population density, and cluster 2 being medium population density. Cluster formation is based on the visualization produced by the **research** dataset through *Sum Of Square Error* analysis, silhouette coefficient, and elbow method. Clustering is formed, followed by dashboard visualization with the tableau application. The clustering results, based on the SSE calculation, produce a value of 4324505738.747303, meaning the determination of the number of clusters with a significant difference with the calculation of the number of previous groupings. Then the **results** of the silhouette analysis provide the highest average silhouette value at the number of clusters, namely 3 with a value of 0.6144435666457168, and the **elbow** method gives the result that the elbow point is at point 3, meaning the optimum number of clusters with 3 clusters.

*Ini adalah artikel akses terbuka di bawah [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.*



DOI: <https://doi.org/10.33050/tmj.v9i3.2270>

Ini adalah artikel akses terbuka di bawah [CC-BY license \(https://creativecommons.org/licenses/by/4.0/\)](https://creativecommons.org/licenses/by/4.0/)

©Penulis memegang semua hak cipta

## 1. PENDAHULUAN

Jumlah penduduk dan luas wilayah memberikan hasil data kepadatan penduduk di suatu daerah, kepadatan penduduk yang tidak sama dan memiliki perbedaan selisih yang signifikan memberikan dampak pada masyarakat wilayah itu sendiri. Kasus kepadatan penduduk menjadi perhatian bagi pemerintahan dalam melihat dampaknya dari berbagai aspek, salah satunya tingginya persaingan antar individu untuk mendapat pekerjaan [1]. Dari kasus kepadatan penduduk yang tidak terkontrol dan manajemen dengan baik, suatu daerah akan menanggung akibatnya. Provinsi Sulawesi Tengah menjadi satu diantaranya provinsi dengan tingginya angka kepadatan penduduk, hal ini berdasarkan sensus penduduk tahun 2020 dengan rata-rata penambahan jumlah penduduk sebanyak 35.000 jiwa pertahun. Ini sangat mempengaruhi terhadap luas wilayah di provinsi Sulawesi Tengah, dan tercatat pada sensus tersebut bahwa terjadi peningkatan kepadatan penduduk hingga 6 jiwa/kilometer persegi [2].

Hal ini menjadi perhatian pemerintah dalam menanggulangi dampak akibat dari kepadatan penduduk, karena semakin naik tingkat kepadatan penduduk, permasalahan yang ditimbulkan juga semakin besar [3]. Dalam halnya pengelolaan kepemilikan tanah terkhususnya di Kota Palu, mayoritas penduduk sudah terdaftar sehingga dalam kebijakan pengelolaan perlu peran dari pemerintahan dalam perencanaan pembangunan, sehingga kependudukan tidak terfokus pada satu titik [4]. Dengan demikian, dilakukannya pengelompokan kepadatan wilayah yakni kecamatan-kecamatan yang ada di Provinsi Sulawesi Tengah untuk mengetahui titik-titik kepadatan penduduk. Untuk mengetahui penyebaran titik kepadatan, diperlukan data jumlah penduduk perkecamatan Provinsi Sulawesi Tengah sesuai data dari Dinas Kependudukan dan Pencatatan Sipil Provinsi Sulawesi Tengah, data luas wilayah sesuai data dari Badan Pusat Statistik Provinsi Sulawesi Tengah, dan data batas desa wilayah Provinsi Sulawesi Tengah dari website Indonesia Geospasial [5].

Penelitian terdahulu telah menerapkan clustering K-Means, seperti dilakukan oleh pada jumlah penduduk sekelurahan Pematangsiantar yang menunjukkan 29 kelurahan pada cluster jumlah penduduk rendah, 18 kelurahan pada cluster sedang, serta 6 kelurahan pada cluster dengan jumlah penduduk tinggi [6]. Penelitian selanjutnya yang dilakukan menghasilkan clustering menjadi tiga cluster, terdiri dari cluster 1 definisi sangat padat, cluster 2 definisi padat, dan cluster 3 definisi kepadatan penduduk sedang. Perhitungan dilakukan manual dengan standar perhitungan apabila data cluster tidak berubah, maka hasil clustering ditetapkan [7]. Menerapkan algoritma K-Means dalam analisis cluster sebagai perencanaan terhadap pemerataan penduduk dan memberikan hasil berupa pengelompokan atas 3 cluster, yaitu cluster 0, cluster 1, dan cluster 2. Clustering ini berarti kepadatan penduduk dari tidak padat, kepadatan penduduk sedang, dan kategori padat penduduk, sehingga menjadi analisa perhatian pemerintah dalam mengambil kebijakan di sisi pengelolaan kependudukan [8].

Penelitian pendukung selanjutnya dilakukan [9], yang berfokus pada penyebab dari jumlah penduduk padat yakni melimpahnya partikulat lapisan atas tanah yang diakibatkan oleh kendaraan. Yang menghasilkan tiga cluster pada 38 sampel, sehingga cluster 1 terdiri 12 sampel, cluster 2 terdiri 23 sampel, dan cluster 3 terdiri 3 sampel, menerapkan algoritma K-Means pada indikator Indeks Pembangunan Manusia yang menghasilkan

clustering K-Means yang membagi menjadi 3 cluster, yakni cluster pertama atau wilayah tengah terdiri dari 41 kota [10]. Cluster kedua atau wilayah tinggi terdiri dari 21 kabupaten/kota, dan cluster ketiga atau wilayah rendah terdiri dari 92 kabupaten/kota berdasarkan pengukuran dalam upaya membangun kualitas dan pemerataan kehidupan manusia atau ketimpangan nilai Indeks Pembangunan Manusia pulau Sumatera.

Penelitian pendukung selanjutnya dilakukan [9], yang berfokus pada penyebab dari jumlah penduduk padat yakni melimpahnya partikulat lapisan atas tanah yang diakibatkan oleh kendaraan. Yang menghasilkan tiga cluster pada 38 sampel, sehingga cluster 1 terdiri 12 sampel, cluster 2 terdiri 23 sampel, dan cluster 3 terdiri 3 sampel [10], [11] juga menerapkan algoritma K-Means pada indikator Indeks Pembangunan Manusia yang menghasilkan clustering K-Means yang membagi menjadi 3 cluster, yakni cluster pertama atau wilayah tengah terdiri dari 41 kota. cluster kedua atau wilayah tinggi terdiri dari 21 kabupaten/kota, dan cluster ketiga atau wilayah rendah terdiri dari 92 kabupaten/kota berdasarkan pengukuran dalam upaya membangun kualitas dan pemerataan kehidupan manusia atau ketimpangan nilai Indeks Pembangunan Manusia pulau Sumatera [12].

Dalam sistem penerimaan Beasiswa yang ada pada Nahdlatul Ulama masih dilakukan secara manual, hal ini berdampak pada sistem pencatatan yang sering terjadi kesalahan, dan informasi yang diberikan juga belum bisa tersebar secara menyeluruh. Oleh karena itu salah satu lembaga yakni LAKPESDAM PBNU membuat sebuah aplikasi yang beri nama Nahdlatul Ulama Scholarship (NUS). Sehingga pada aplikasi ini diharapkan dapat menyelesaikan permasalahan di atas dan dapat membantu dalam pengelolaan penerimaan beasiswa serta memudahkan akses mahasiswa untuk pendaftaran beasiswa. Aplikasi ini dirancang untuk meningkatkan transparansi, meminimalkan kesalahan pencatatan, serta mempercepat proses penyampaian informasi kepada calon penerima beasiswa, sehingga sistem seleksi menjadi lebih efisien dan terorganisir.

Dengan adanya aplikasi Nahdlatul Ulama Scholarship (NUS), diharapkan proses seleksi beasiswa dapat dilakukan secara lebih objektif dan terukur. Sistem ini memungkinkan pengelolaan data pendaftar secara digital, mulai dari proses pendaftaran, seleksi, hingga pengumuman hasil beasiswa. Selain itu, aplikasi ini juga memberikan kemudahan bagi mahasiswa dalam mengakses informasi terkait persyaratan, jadwal seleksi, serta hasil evaluasi secara real-time. Dengan demikian, NUS berperan sebagai solusi inovatif dalam mendukung pemerataan akses pendidikan dan menciptakan sistem penerimaan beasiswa yang lebih transparan dan akuntabel.

## 2. PERMASALAHAN

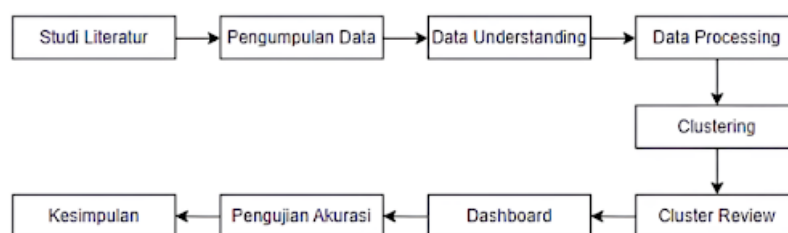
Sulawesi Tengah merupakan provinsi yang terletak ditengah pulau Sulawesi dengan ibukota berada di Kota Palu dengan luas wilayah Provinsi sebesar 61.841,29 km<sup>2</sup>. Penyebaran penduduk di Provinsi ini, termasuk yang tidak merata. Berdasarkan informasi dari Badan Pusat Statistik, beberapa titik wilayah tidak padat penduduk, dan terfokus pada satu-dua titik wilayah. Diketahui juga, berdasarkan sensus penduduk tahun 2020 dengan rata-rata penambahan jumlah penduduk sebanyak 35.000 jiwa pertahun. Ini sangat mempengaruhi terhadap luas wilayah di provinsi Sulawesi Tengah, dan tercatat pada sensus tersebut bahwa terjadi peningkatan kepadatan penduduk hingga 6 jiwa/kilometer persegi. Penyebaran yang tidak merata ini menyebabkan perlunya analisa penyebaran kependudukan lebih lanjut.

Analisa dilakukan dengan penerapan metode data mining yakni clustering terhadap data luas wilayah dan jumlah penduduk tiap kecamatan di Provinsi Sulawesi Tengah yang diperoleh dari kantor Dinas Kependudukan dan Pencatatan Sipil Provinsi Sulawesi Tengah, Badan Pusat Statistik Provinsi Sulawesi Tengah, serta data batas desa wilayah Provinsi Sulawesi Tengah dari website Indonesia Geospasial guna mendapat hasil analisa penyebaran kepadatan penduduk Provinsi Sulawesi Tengah, sehingga menjadi landasan/acuan bagi Pemerintahan Provinsi Sulawesi Tengah dalam pengambilan kebijakan, program, serta ketentuan dalam menanggulangi masalah sosial dari kepadatan penduduk yang tidak diperhatikan optimalnya.

## 3. METODOLOGI PENELITIAN

Clustering K-Means terhadap kepadatan penduduk dimulai dengan studi literatur, mengumpulkan data jumlah penduduk, luas wilayah, serta batas desa di Provinsi Sulawesi Tengah. Selanjutnya, dilakukan pemahaman data melalui analisis atribut yang tersedia dan evaluasi kualitas dataset, sehingga proses pengolahan data meliputi pembersihan, penghapusan data tidak relevan, serta perlakuan khusus untuk mempertahankan dan meningkatkan kualitas data, termasuk pemilihan atribut yang relevan untuk penelitian. Tahapan berikutnya adalah memberikan perlakuan penilaian kecenderungan cluster guna membentuk kelompok atas data penelitian ini. Penentuan jumlah cluster dilakukan dengan metode elbow dan silhouette untuk menentukan jumlah cluster yang optimal [13]. Dengan demikian, pembentukan parameter ukur cluster didasarkan pada jumlah penduduk

dan luas wilayah yang terdapat dalam atribut dataset penelitian, sehingga analisis cluster dapat divisualisasikan secara jelas untuk memahami pola kepadatan penduduk di wilayah tersebut.



Gambar 1. Alur Penelitian

Gambar 1 merupakan tahapan penelitian guna mencapai tujuan penelitian, yang diawali dengan studi literatur hingga visualisasi dashboard kepadatan penduduk Provinsi Sulawesi Tengah, dan pengujian akurasi hasil penelitian.

### 3.1. Clustering

K-means clustering sering digunakan dalam proses pembentukan cluster untuk mencari pola atau struktur yang tersembunyi dalam data dengan cara meminimalkan jarak antara titik data ke pusat cluster yang sesuai [13]. Metode ini diperkenalkan secara luas oleh James B MacQueen [14]. Huruf K bermakna jumlah cluster yang akan dibuat, sementara means memiliki makna nilai sementara atas pusat cluster (centroid) [15]. Clustering dilakukan dengan menerapkan analisa *Sum Of Square Error*, elbow method, serta dengan penilaian silhouette pada dataset penelitian.

Dalam penerapan clustering, digunakan perhitungan euclidean distance guna menghasilkan pengukuran rentang jauh antara objek data dengan pusat cluster, dalam penelitian ini merupakan data jumlah penduduk di setiap kecamatan akan dihitung ke pusat cluster dengan teori euclidean distance. Perhitungan *Sum Of Square Error* didapatkan dari hasil pertambahan atas seluruh rentang setiap data dengan jarak titik pusat cluster data tersebut, memberikan nilai K yang berbeda dalam perhitungan nilai SSE guna mendukung penentuan jumlah cluster paling optimum [16].

Namun, sebelum cluster paling optimal ditetapkan, dilakukan analisa nilai silhouette guna melihat kualitas hasil pengelompokkan dari perhitungan jarak dengan penafsiran nilai silhouette coefficient sebagai berikut.

Tabel 1. Penafsiran Nilai Silhouette Coefficient

<i>Silhouette Coefficient</i>	Penafsiran
0.71 – 1.00	Bentuk yang dihasilkan kuat
0.51 – 0.70	Bentuk yang dihasilkan baik
0.26 – 0.50	Bentuk yang dihasilkan lemah
0.25	Tidak menghasilkan bentuk

Dalam table 1, analisa nilai SSE dapat ditentukan nilai K berdasarkan nilai yang signifikan, sehingga analisa dengan metode elbow dalam menentukan cluster berdasarkan informasi grafik berbentuk siku pada suatu titik [17].

### 3.2. Visualisasi Dashboard

Dalam visualisasi lanjut dilakukan dengan software tableau, software ini menyediakan tools penggambaran data secara grafis yang lebih interaktif dan mudah dipahami, dimana memberikan kebutuhan visualisasi data dalam analisis data, hal ini berdasarkan kemampuan mengkomunikasikan temuan dengan jelas [18]. Visualisasi yang dihasilkan berdasarkan data tabel menjadi visualisasi grafik, diagram, geo mapping, serta menampilkan perbedaan dan perubahan data menjadi lebih jelas dan dipahami [19]. Pada penelitian ini, juga dilakukan pengujian akurasi hasil penelitian dengan tableau. Pengujian akurasi dilakukan guna melihat apakah analisis yang dilakukan mampu memperoleh nilai kembali atau recovery yang baik atau sebaliknya [20].

## 4. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan clustering pada masalah yang diangkat dan dashboard dilakukan dengan software tableau untuk pengamatan analisa sesuai kebutuhan penelitian, serta pengujian akurasi penelitian ini.

Berikut penjabaran hasil dan pembahasan penelitian ini.

#### 4.1. Pengumpulan Data/Data Collection

Tahapan data collection penelitian ini berasal dari data Dinas Kependudukan dan Pencatatan Sipil Provinsi Sulawesi Tengah, Badan Pusat Statistik Provinsi Sulawesi Tengah, serta data pada website Indonesia Geospasial yang mana dibawah naungan Kementerian Perhubungan Republik Indonesia berupa data jumlah penduduk, luas wilayah, dan batas desa wilayah Provinsi Sulawesi Tengah. Berikut 5 contoh sample dataset penelitian yang digunakan penelitian ini.

Tabel 2. Sample Dataset Penelitian

Kabupaten	Kecamatan	Jumlah Penduduk	Luas Wilayah (km <sup>2</sup> )
Banggai	Batui	19.043	1062,36
Poso	Poso Kota	23.830	11,29
Donggala	Dampelas	31.555	603,32
Toli-Toli	Dampal Selatan	22.955	392,67
Kota Palu	Palu Timur	44.428	7,71

Pengelompokan data yang dianalisa pada tabel 2 berupa data jumlah penduduk dan luas wilayah setiap kecamatan Provinsi Sulawesi Tengah. Data yang diperoleh berdasarkan perbaharuan data terbaru yakni tahun 2023.

#### 4.2. Data Understanding

Data yang digunakan memiliki 176 baris dengan atribut-atribut seperti kabupaten, kecamatan, jumlah penduduk, luas wilayah, dan kepadatan penduduk. Pada analisis clustering penelitian ini, menggunakan dua atribut yakni jumlah penduduk dan luas wilayah. Gambar 3 merupakan deskripsi dataset yang digunakan.

```

<bound method NDFrame.head of
0  BANGGAI      BATUI      19043      1862.36
1  BANGGAI      BATUI      19817      579.88
2  BANGGAI      KECITOH    11869      428.78
3  BANGGAI      LUMUK     33866      72.82
4  BANGGAI      LAVALA    7381      228.66
..  ...
171 KOTA PALU    PALU UTARA  25041      29.94
172 KOTA PALU    ULUJADI   36156      48.25
173 KOTA PALU    TATANGGA  93483      14.95
174 KOTA PALU    TAMELET  23182      59.25
175 KOTA PALU    MANTIKULORE  88343      286.88

Kepadatan Penduduk(:liva/km2)
0      17.925185
1      34.226252
2      27.688862
3      454.078558
4      33.887182
..  ...
171      836.372745
172      898.285714
173      3576.128481
174      398.661888
175      388.585883

[176 rows x 5 columns]
Shape : (176, 5)
Kabupaten      object
Kecamatan      object
Jumlah Penduduk  int64
Luas Wilayah (km2)  float64
Kepadatan Penduduk(:liva/km2)  float64
dtype: object

Jumlah Penduduk      Luas Wilayah (km2)      Kepadatan Penduduk(:liva/km2)
count      176.000000      176.000000      176.000000
mean      17848.545455      371.645857      215.143696
std      12099.254919      396.851136      717.682887
min      3332.000000      7.718000      6.775778
25%      9771.000000      145.520000      33.031940
50%      15127.000000      236.485000      55.123333
75%      22751.000000      512.252500      83.778571
max      89343.000000      2486.790000      5762.386512

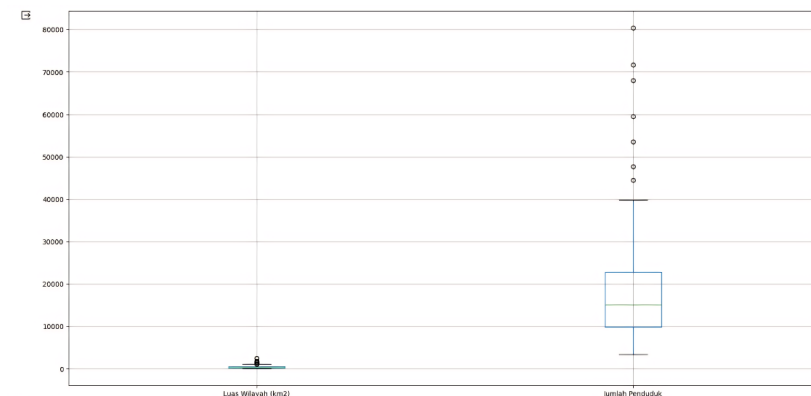
Data Null ?
Kabupaten      0
Kecamatan      0
Jumlah Penduduk  0
Luas Wilayah (km2)  0
Kepadatan Penduduk(:liva/km2)  0
dtype: int64

```

Gambar 2. Data Understanding  
(Sumber : Peneliti)

Berdasarkan data pada gambar 2 yang diperoleh tersebut, penggunaan fungsi dengan bahasa pemrograman python yang dilakukan pada pandas, seperti fungsi head untuk menampilkan data teratas dan terbawah sehingga diketahui juga jumlah data sebanyak 176 baris, fungsi df.shape menampilkan dimensi atau ukuran baris dan kolom dari dataframe, fungsi df.dtypes menampilkan tipe data tiap kolom dataframe, dengan fungsi ini bisa diketahui tipe data dari jumlah penduduk dan luas wilayah merupakan numerical data untuk dilakukan analisa clustering. df.describe guna menampilkan deskriptif statistic data dan hanya menampilkan kolom dengan tipe numerik. df.isnull. Guna menghitung nilai null atau kosong tak bernilai pada dataframe. Dengan demikian, diketahui dataframe penelitian ini tidak memiliki nilai null atau kosong pada atribut yang dibutuhkan, sehingga dapat dilakukan kelanjutan analisa cluster penelitian ini [21].

Setelah analisa dataframe agar tidak terdapat berbagai noisy data dan sebagainya, selanjutnya dilakukan pengamatan dengan box plot guna melihat pemusatan dan penyebaran data dari nilai tengah, nilai ekstrim ataupun outlier dan pengukuran sebagainya [22]. Box plot merupakan penyajian data informasi secara grafis dengan menampilkan distribusi nilai-nilai data pada yang ditujukan, dalam halnya penyajian distribusi data (skewness) maupun ukuran penyebaran (keragaman) data pengamatan [23].



Gambar 3. Visualisasi Box Plot  
(Sumber : Peneliti)

Selain itu, box plot juga sangat efektif dalam mendeteksi keberadaan pencilan atau outlier dalam sebuah dataset. Pencilan dapat terlihat dengan jelas sebagai titik-titik di luar garis whisker yang menggambarkan rentang data. Ini memberikan informasi tambahan yang sangat berharga bagi analis data untuk mengevaluasi kualitas dan akurasi data yang digunakan dalam penelitian.

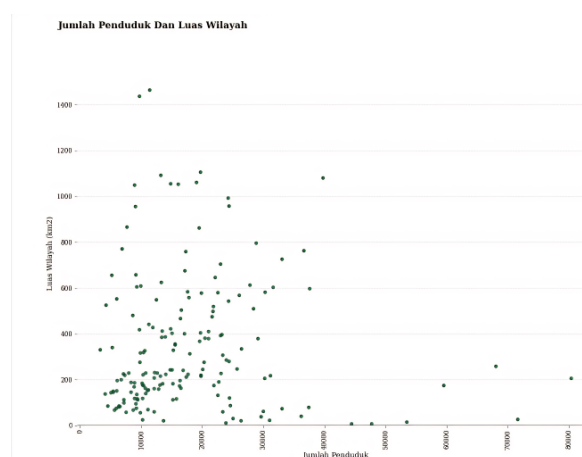
Penggunaan box plot dalam analisis data sangat bermanfaat ketika ada data dengan distribusi yang tidak normal atau skewed. Dengan memvisualisasikan distribusi data secara jelas, box plot memungkinkan kita untuk melihat apakah data lebih cenderung ke kiri (left-skewed) atau ke kanan (right-skewed), yang dapat mempengaruhi keputusan analisis lebih lanjut seperti pemilihan model statistik yang tepat.

Akhirnya, box plot juga memberikan informasi yang sangat berguna untuk perbandingan antara beberapa grup data. Misalnya, dalam penelitian tentang kepadatan penduduk, box plot dapat digunakan untuk membandingkan distribusi kepadatan antara beberapa kecamatan atau wilayah, sehingga memudahkan pengambilan keputusan berdasarkan perbedaan distribusi tersebut.

Berdasarkan visualisasi gambar 3 diatas, diketahui nilai tengah/median, nilai terendah, nilai tertinggi, dan outlier sekalipun, perbandingan nilai yang sangat berbeda antara luas wilayah dan jumlah penduduk ini menunjukkan adanya ketidaksetaraan wilayah padatnya kependudukan di Provinsi Sulawesi Tengah.

### 4.3. Data Processing

Selanjutnya diberlakukan pendeskripsian dataset yang terdiri atas informasi dataset dan analisis dengan scatter plot data, sehingga menghasilkan visualisasi scatter plot sebagai berikut :



Gambar 4. Data Processing  
(Sumber : Peneliti)

Berdasarkan grafik gambar 4 scatter plot tersebut, dapat dilihat outlier yang dapat mempengaruhi hasil

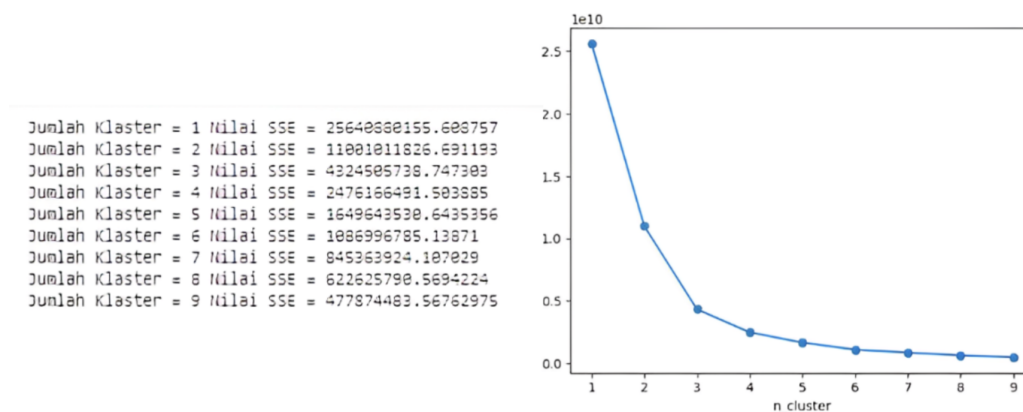
clustering pada data jumlah penduduk yakni pada nilai  $>40000$ . Penanganan pada data outlier penelitian ini menggunakan metode penghapusan demi mengurangi adanya gangguan terhadap analisa pembentukan cluster dan melihat jarak dari penyebaran plot yang outlier pada grafik, sehingga dapat meningkatkan keakuratan hasil clustering.

#### 4.4. Clustering

Selanjutnya untuk menentukan jumlah cluster yang terbentuk, dilakukan analisa *Sum Of Square Error* (SSE), metode elbow, serta pengamatan berdasarkan standar silhouette coefficient [24]. Analisa *Sum Of Square Error* (SSE) diterapkan guna menganalisa informasi mengukur error jarak data ke centroid (titik tengah) [25]. Dalam penelitian ini, hasil clustering penelitian ini menggunakan metode pengukuran Eulidean Distance.

Dalam melakukan analisis clustering, langkah pertama yang dilakukan adalah pengolahan data yang sesuai dengan metodologi yang telah ditentukan. Salah satu komponen penting dalam clustering adalah pemilihan teknik yang tepat untuk mengevaluasi hasil dari setiap cluster yang terbentuk. Dengan menggunakan metode Sum of Squared Error (SSE), kita dapat menentukan seberapa baik hasil clustering yang dilakukan dengan melihat seberapa dekat data dengan pusat cluster yang telah ditentukan. Selain itu, evaluasi dengan menggunakan silhouette coefficient dan metode elbow juga memberikan gambaran tentang jumlah cluster yang optimal untuk pengelompokan data lebih lanjut.

Berikut nilai SSE berdasarkan pembagian clusternya masing-masing menggunakan tools pandas google colab.



Gambar 5. Perhitungan *Sum Of Square Error*  
(Sumber : Peneliti)

Perhitungan pada gambar 5 *Sum Of Square Error* dilakukan dengan tools pandas GoogleColab. Berdasarkan perhitungan jarak euclidian distance nilai SSE yang dihasilkan pada penelitian ini, dan nilai K yang digunakan adalah 3. Perbandingan antara nilai SSE satu dengan yang lain, pembeda yang paling signifikan pada jumlah cluster 3 dengan nilai SSE senilai 4324505738.747303. Selisih nilai SSE bisa dilihat dengan jelas pada visualisasi perbandingan nilai SSE dan jumlah cluster penelitian ini. Hal ini bisa menentukan jumlah pembagian cluster pada penelitian ini yakni dengan dibagi menjadi 3 cluster.

Selanjutnya analisa kualitas pengelompokan perhitungan jarak yang merupakan analisa nilai rata-rata silhouette pada cluster berdasarkan selisih nilai SSE paling memungkinkan terhadap cluster yang dibentuk. Analisa nilai rata-rata silhouette, dilakukan perkiraan terhadap klaster yang memungkinkan dibentuknya cluster, serta jika dilihat dari perbandingan nilai SSE pada jumlah cluster yang dibentuk, tertuju pada jumlah cluster 3, 4, dan 5. Berikut pembentukan jumlah cluster berdasarkan nilai silhouette penelitian ini.

```

Jumlah klaster = 3 Nilai rata-rata Silhouette = 0.6144435666457168
Jumlah Klaster = 4 Nilai rata-rata Silhouette = 0.5838193594137112
Jumlah Klaster = 5 Nilai rata-rata Silhouette = 0.5643120467003521

```

Gambar 6. Perhitungan Silhouette  
(Sumber : Peneliti)

Dengan demikian, pada gambar 6 nilai rata-rata silhouette paling tinggi pada jumlah cluster yakni 3 dengan nilai 0.6144435666457168 jika dibandingkan dengan nilai rata-rata silhouette pada jumlah kluster 4 dan 5 dengan nilai 0.5 sekian dikeduanya. Dengan nilai 0.6144435666457168 tersebut memiliki arti bahwa clustering yang dibentuk merupakan interpretasi dengan struktur yang dihasilkan baik dalam silhouette coefficient.

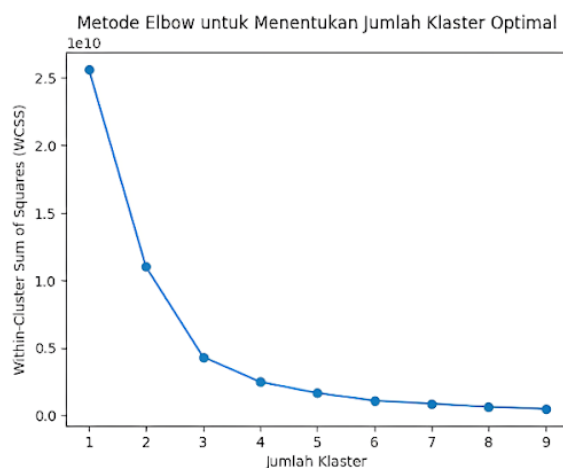
Hasil analisis menggunakan metode elbow menunjukkan bahwa titik elbow berada pada jumlah cluster 3. Titik ini menandakan bahwa menambah jumlah cluster lebih dari tiga tidak memberikan peningkatan signifikan dalam pengurangan SSE. Oleh karena itu, jumlah cluster tiga merupakan jumlah yang optimal untuk analisis clustering ini. Dengan memanfaatkan elbow point ini, kita dapat yakin bahwa pembagian cluster yang menggunakan tiga kelompok ini menghasilkan pemisahan data yang paling representatif berdasarkan karakteristik kepadatan penduduk di wilayah tersebut.

Selain itu, analisa tambahan menggunakan silhouette coefficient semakin memperkuat keputusan jumlah cluster tiga. Silhouette coefficient ini memberikan gambaran yang jelas mengenai kualitas pengelompokan yang terbentuk. Cluster dengan nilai silhouette tinggi menunjukkan bahwa data dalam cluster tersebut lebih homogen dan terpisah dengan jelas dari cluster lainnya. Dengan nilai silhouette rata-rata 0.614, clustering yang dihasilkan menunjukkan bahwa pengelompokan yang dilakukan sangat baik, mencerminkan kualitas pembagian yang optimal antara kepadatan rendah, sedang, dan tinggi.

Untuk memperkaya hasil analisis, dilakukan pula validasi terhadap model clustering yang telah dibentuk. Proses validasi ini mencakup evaluasi melalui berbagai metode dan perbandingan dengan data yang ada. Selain SSE dan silhouette, pengujian lainnya juga dilakukan untuk memastikan bahwa model clustering yang digunakan memberikan hasil yang dapat diandalkan dan akurat. Penggunaan software tableau dalam visualisasi hasil clustering memungkinkan pemangku kebijakan untuk melihat distribusi kepadatan penduduk secara lebih jelas dan interaktif, memberikan insight yang berguna untuk perencanaan dan pengambilan kebijakan lebih lanjut [26].

Secara keseluruhan, penerapan metode K-Means Clustering dalam penelitian ini berhasil memberikan gambaran yang lebih terperinci tentang kepadatan penduduk di Provinsi Sulawesi Tengah. Dengan adanya pembagian yang jelas menjadi tiga cluster, para pembuat kebijakan dapat lebih mudah merumuskan strategi yang lebih tepat dalam menangani masalah kepadatan penduduk dan merancang kebijakan pembangunan yang lebih merata dan efisien. Rekomendasi lebih lanjut akan terus dibutuhkan untuk memantau perkembangan kependudukan di masa mendatang, agar pengelolaan sumber daya dan infrastruktur dapat dilakukan dengan lebih baik.

Namun analisa pembagian cluster tidak berhenti dengan nilai rata-rata silhouette tersebut, menganalisa pembagian cluster penelitian ini juga didukung dengan metode elbow sebagai berikut :



Gambar 7. Elbow Method  
(Sumber : Peneliti)

Metode elbow diterapkan pada penelitian ini untuk menentukan jumlah cluster pada metode algoritma K-Means, hal ini diketahui optimum dengan metode elbow. Berdasarkan gambar 7, dapat dilihat elbow point

berada di titik 3. Dengan demikian, selanjutnya dilakukan pembentukan cluster dengan algoritma K-Means.

#### 4.5. Cluster Review

Selanjutnya menganalisa pembagian cluster pada masing-masing kecamatan. Mulai dari inisiasi objek standard scaler bertujuan agar dataset memuat pengelompokkan dengan penskalaan dalam mendistribusi deskripsi dataset pada data x (jumlah penduduk), melakukan penentuan cluster optimal, inisiasi model K-Means, dan penerapannya pada standarisasi data jumlah penduduk, serta menambah kolom cluster pada dataframe [27].

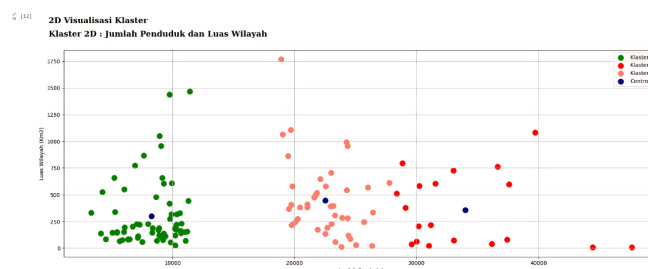
Kabupaten	Kecamatan	Jumlah Penduduk	Luas Wilayah (km <sup>2</sup> )	Kepadatan Penduduk (Jiwa/km <sup>2</sup> )	Cluster	
0	BANGGAI	BATUI	19043	1062.36	17.925185	0
1	BANGGAI	BUNTA	19817	579.00	34.226252	0
2	BANGGAI	KINTOM	11869	428.78	27.600662	1
3	BANGGAI	LUWUK	33066	72.82	454.078550	1
4	BANGGAI	LAMALA	7301	220.66	33.087102	1
...	...	...	...	...	...	...
171	KOTA PALU	PALU UTARA	25041	29.94	836.372745	1
172	KOTA PALU	ULUJADI	36156	40.25	898.285714	2
173	KOTA PALU	TATANGA	53463	14.95	3576.120401	2
174	KOTA PALU	TAWAELI	23342	59.75	390.661088	1
175	KOTA PALU	MANTIKULORE	80343	206.80	388.505803	2

176 rows × 6 columns

Gambar 8. Penambahan Kolom Cluster  
(Sumber : Peneliti)

Dengan demikian, pada gambar 8 dapat dianalisa cluster masing-masing kecamatan berdasarkan data jumlah penduduk dan luas wilayah dan cluster masing-masing kecamatan, seperti kecamatan Palu Utara masuk dalam cluster 1 dan Kecamatan Ulujadi masuk dalam cluster 2. Sehingga, tahapan selanjutnya dengan melakukan visualisasi scatter plot chart dua dimensi pada data jumlah penduduk, luas wilayah, dan cluster. Selain itu, pemilihan metode K-Means Clustering dalam penelitian ini terbukti efektif dalam memberikan gambaran yang jelas mengenai sebaran kepadatan penduduk. Meskipun terdapat beberapa titik data yang termasuk outlier, metode ini tetap dapat mengidentifikasi cluster dengan akurat, terutama setelah dilakukan penghapusan terhadap data yang mencurigakan [28].

Pada langkah berikutnya, hasil dari visualisasi scatter plot ini memberikan gambaran yang lebih jelas mengenai distribusi kepadatan penduduk di setiap kecamatan [29]. Dapat dilihat bahwa masing-masing cluster menggambarkan kepadatan yang berbeda, yang memberikan informasi penting dalam merumuskan kebijakan perencanaan wilayah. Untuk memperdalam analisis, setiap kecamatan yang teridentifikasi dalam masing-masing cluster dapat dianalisis lebih lanjut untuk melihat faktor-faktor yang mempengaruhi kepadatan, seperti infrastruktur, tingkat kemiskinan, dan akses ke fasilitas umum [30]. Visualisasi ini memudahkan pihak terkait dalam mengambil keputusan yang lebih tepat untuk meningkatkan kesejahteraan masyarakat.



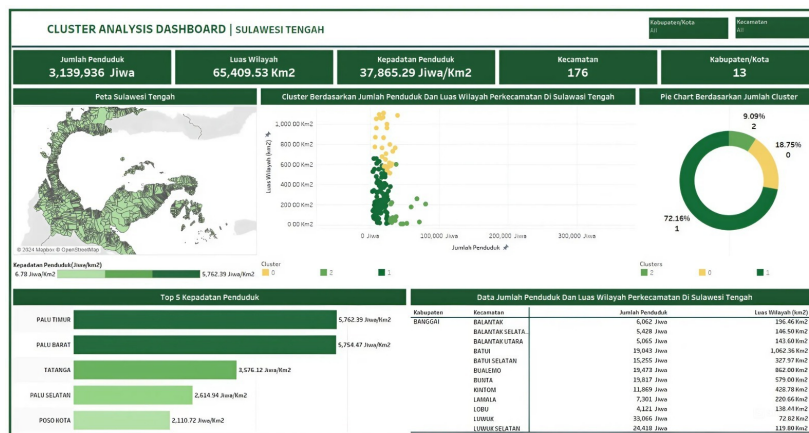
Gambar 9. Visualisasi Clustering  
(Sumber : Peneliti)

Berdasarkan grafik gambar 9, terdapat tiga cluster dan nilai centroid pada masing-masing cluster tersebut. Cluster yang terbentuk dapat dianalisa bahwa cluster 0 merupakan cluster dengan kepadatan jiwa yang rendah dimana dengan kepadatan penduduk dibawah dari 10.000 jiwa dengan luas wilayahnya dibawah 1.500 km<sup>2</sup>, lalu cluster 1 dengan kepadatan penduduk tinggi dengan kepadatan mencapai angka dibawah 50.000 jiwa

dengan luas wilayah dibawah 1.250 km<sup>2</sup>, dan cluster 2 dengan kepadatan sedang yang mencapai 30.000 jiwa penduduk dengan luas wilayah dibawah 1.800. Dengan demikian, clustering K-Means sudah diterapkan dan dilakukan eksport data dengan tujuan diterapkannya visualisasi dengan mapping dan tools lainnya menggunakan aplikasi tableau.

**4.6. Dashboard**

Penelitian ini menggunakan perangkat lunak tableau, untuk menampilkan tampilan dashboard dengan informasi seputar penyebaran kepadatan penduduk Provinsi Sulawesi Tengah.

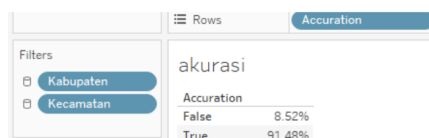


Gambar 10. Dashboard Kepadatan Penduduk Provinsi Sulawesi Tengah (Sumber : Peneliti)

Dashboard analisis gambar 10 cluster kepadatan penduduk di Provinsi Sulawesi Tengah memberikan informasi yang interaktif mempercepat penyampaian informasi, dari informasi mapping penyebaran kepadatan penduduk sesuai dengan cluster, kabupaten, kecamatan, serta jumlah penduduk dan luas wilayah terbagi dalam visualisasi mapping dengan peta Sulawesi Tengah. Informasi clustering bisa dilihat dengan pie chart, mapping, maupun grafik beserta tabel. Berdasarkan penyediaan visualisasi tersebut, dapat dianalisa bahwasanya dari 176 kecamatan di Provinsi Sulawesi Tengah, kecamatan Palu Timur merupakan kecamatan dengan kepadatan penduduk tertinggi dengan kepadatan 5,762.39 Jiwa/Km<sup>2</sup>, disusul kecamatan Palu Barat, Tatanga, Palu Selatan, dan Poso Kota. Diketahui juga, kecamatan dengan kepadatan penduduk tinggi alias dalam cluster 1 sebanyak 72.16%, kemudian kecamatan dengan kepadatan penduduk sedang sebanyak 9.09%, dan kecamatan dengan kepadatan penduduk rendah sebanyak 18.75%. Dengan hasil tersebut, demi pemerataan populasi penduduk Provinsi, hal ini menjadi acuan dan rekomendasi pemangku kepentingan pemerintah Provinsi Sulawesi Tengah untuk mengambil kebijakan dalam mendukung pemerataan penduduk Provinsi Sulawesi Tengah.

**4.7. Pengujian Analisa**

Pengujian terhadap analisa clustering penelitian ini, dilakukan analisa pada software tableau dengan mengkombinasikan dataset kabupaten dan kecamatan untuk melihat nilai prediksi akurasi analisa penelitian ini. Berikut pengujian yang dilakukan pada software tableau.



Gambar 11. Akurasi Test (Sumber : Peneliti)

Berdasarkan gambar 11 hasil akurasi tersebut, diketahui ketepatan analisis clustering K-Means penelitian ini memberikan hasil 91.48% dan ini merupakan hasil yang cukup optimal dengan mendekati nilai maksimal. Dalam penelitiannya terhadap Analisa Amonia terhadap air tanah, juga melakukan pengukuran akurasi. Pengukuran akurasi dikatakan baik, yaitu jika terdapat pada rentang nilai 90% [31].

Dengan uji akurasi terhadap dashboard kependudukan wilayah Provinsi Sulawesi Tengah, menggabungkan data penelitian dan penerapan metode clustering k-means yang digunakan untuk analisa clustering kepadatan penduduk di Provinsi Sulawesi Tengah dapat diterima sebagai bahan acuan analisa kelanjutan program kebijakan pemerintah dalam kependudukan Provinsi Sulawesi Tengah.

## 5. KESIMPULAN

Berdasarkan pembahasan dan hasil penerapan algoritma K-Means clustering dengan pengelompokkan pada kepadatan kependudukan perkecamatan di Provinsi Sulawesi Tengah, diterapkannya melalui beberapa tahapan penelitian, baik dari tinjauan pustaka terhadap penelitian ini, mengumpulkan data, pemahaman data, memproses data sebelum dilakukannya clustering, penggunaan penilaian *Sum Of Square Error*, metode elbow, serta silhouette coefficient sehingga dilakukan review cluster yang dibentuk, serta visualisasi dashboard kepadatan penduduk Provinsi Sulawesi Tengah dengan aplikasi tableau.

Informasi yang termuat dashboard tersebut, seperti mapping penyebaran kependudukan, scatter plot analysis, hingga dekripsi data padat penduduk. Dengan clustering tersebut, cluster dibentuk menjadi tiga cluster, yakni cluster 0 sebanyak 18.75% dengan kepadatan rendah, lalu cluster 1 sebanyak 72.16% dengan kepadatan penduduk yang tinggi, cluster 2 sebanyak 9.09% dengan kepadatan penduduk menengah/sedang.

Dengan hasil tersebut, kemudian dilakukan uji akurasi dengan aplikasi tableau berdasarkan dataset yang digunakan penelitian ini, sehingga didapatkan nilai 91.48% yang berarti nilai yang cukup optimal untuk hasil visualisasi yang diberikan pada penelitian ini. Dengan demikian, penelitian ini bisa menjadi acuan, landasan dan rekomendasi bagi para pemangku kepentingan pemerintah Provinsi Sulawesi Tengah dalam mengambil kebijakan dan program mendukung pemerataan penduduk serta pengelolaan kependudukan lainnya, sehingga dapat menanggulangi atas permasalahan kependudukan Provinsi Sulawesi Tengah. Dan tak hanya itu, diharapkan terus dilanjutkan analisa terhadap permasalahan kependudukan lainnya, terkhususnya Provinsi Sulawesi Tengah.


## 6. SARAN


Penelitian terhadap kependudukan bisa dapat terus dilakukan dengan analisa melanjut demi melihat perkembangan terhadap kependudukan baik dari kepadatan penduduk ataupun permasalahan kependudukan lainnya. Diharapkan analisa dapat terus dilakukan dengan perkembangan data di Provinsi Sulawesi Tengah demi tata pengelolaan perkotaan yang baik demi menanggulangi masalah yang ditimbulkan dari ketidakperhatian terhadap tata kelola wilayah yang baik.


## 7. DEKLARASI


### 7.1. Tentang Penulis

Fenita Delia (FD)  <https://orcid.org/0009-0007-8161-8820>

Hajra Rasmita Ngemba (HR)  <https://orcid.org/0000-0001-8727-9267>

Syaiful Hendra (SH)  <https://orcid.org/0000-0002-0037-5895>

Syahrullah (SS)  <https://orcid.org/0009-0005-4810-9181>

Nouval Trezandy Lapatta (NT)  <https://orcid.org/0000-0002-0959-9596>

### 7.2. Kontribusi Penulis

Konseptualisasi: FD; Metodologi: HR ; Perangkat Lunak: SH; Validasi: SS dan NT; Analisis Formal: FD dan HR; Investigasi: SH ; Sumber Daya: SS; Kurasi Data: NT; Penulisan Draf Asli Persiapan: FD dan HR ; Penulisan Tinjauan dan Penyuntingan: SH dan SS ; Visualisasi: NT; Semua penulis, FD, HR,SH dan SS, telah membaca dan menyetujui versi naskah yang diterbitkan.

### 7.3. Pernyataan Ketersediaan Data

Data yang disajikan dalam studi ini tersedia atas permintaan dari penulis terkait.

#### 7.4. Pendanaan

Penulis tidak menerima dukungan finansial untuk penelitian, kepenulisan, dan/atau penerbitan artikel ini.

#### 7.5. Deklarasi Konflik Kepentingan

Penulis menyatakan bahwa mereka tidak memiliki konflik kepentingan, konflik kepentingan finansial yang diketahui, atau hubungan pribadi yang dapat memengaruhi pekerjaan yang dilaporkan dalam makalah ini.

### DAFTAR PUSTAKA

- [1] D. Yunianto, "Analisis pertumbuhan dan kepadatan penduduk terhadap pertumbuhan ekonomi," *FORUM EKONOMI*, vol. 23, no. 4, pp. 687–698, 2021. [Online]. Available: <http://journal.feb.unmul.ac.id/index.php/FORUM EKONOMI>
- [2] B. P. S. Tengah, "Hasil sensus penduduk 2020 provinsi sulawesi tengah," 2021, accessed: Mar. 18, 2024. [Online]. Available: <https://sulteng.bps.go.id/pressrelease/2021/01/21/936/hasil-sensus-penduduk-2020-provinsi-sulawesi-tengah.html>
- [3] L. Y. Hutabarat, I. Gunawan, I. Purnamasari, M. Safii, and W. Saputra, "Penerapan algoritma k-means dalam pengelompokan jumlah penduduk berdasarkan kelurahan di kota pematangsiantar," *Jurnal Ilmu Komputer dan Teknologi*, vol. 2, no. 2, pp. 20–26, 2021. [Online]. Available: <http://creativecommons.org/licenses/by/4.0/>
- [4] H. R. Ngemba, S. Hendra, and I. G. N. A. K. D. Arsana, "Implementasi enkripsi data md5 dan sha-256 pada sistem informasi peminjaman buku tanah," *Techno.COM*, vol. 22, no. 3, pp. 654–664, 2023.
- [5] N. Oktaviany, N. Suarna, and W. Prihartono, "Implementasi algoritma k-means clustering dalam mengelompokkan kepadatan penduduk di provinsi dki jakarta," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 119–126, 2024.
- [6] P. Marpaung, I. Pebrian, and W. Putri, "Penerapan data mining untuk pengelompokan kepadatan penduduk kabupaten deli serdang menggunakan algoritma k-means," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 6, no. 2, pp. 64–70, 2023.
- [7] A. A. Arifyanti, F. S. Darusman, and B. W. Trenggono, "Population density cluster analysis in dki jakarta province using k-means algorithm," *Journal of Information Systems and Informatics*, vol. 4, no. 3, 2022. [Online]. Available: <http://journal-isi.org/index.php/isi>
- [8] K. H. Kirana, M. Iryanti, M. G. Nugraha, E. Agustine, and D. Fitriani, "Application of k-means cluster analysis for magnetic susceptibility zoning of urban topsoil in bandung city," in *Journal of Physics: Conference Series*. Institute of Physics, 2022.
- [9] M. Muttaqin, "Cluster analysis using k-means method to classify sumatera regency and city based on human development index indicator," *Seminar Nasional Official Statistics*, vol. 2022, no. 1, pp. 967–976, 2022.
- [10] Y. D. Putra, M. Sudarma, and I. B. A. Swamardika, "Clustering history data penjualan menggunakan algoritma k-means," *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 2, p. 195, 2021.
- [11] F. A. Fernaldy, A. A. Arifyanti, and D. S. Y. Kartika, "Klasterisasi tracer study alumni universitas xyz menggunakan algoritma k-means," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, 2025.
- [12] R. B. Sippan and N. Setiyawati, "Pemetaan dan klasterisasi daerah rawan bencana alam di provinsi sulawesi tengah menggunakan k-means," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1031–1045, 2025.
- [13] L. P. Refialy, H. Maitimu, and M. S. Pesulima, "Perbaikan kinerja clustering k-means pada data ekonomi nelayan dengan perhitungan sum of square error (sse) dan optimasi nilai k cluster," *Techno.COM*, vol. 20, no. 2, pp. 321–329, 2021.
- [14] M. A. Subhartha and M. K. S. Budhi, "Pengaruh infrastruktur terhadap pertumbuhan ekonomi provinsi sulawesi tengah," *Public Service and Governance Journal*, vol. 6, no. 1, pp. 143–163, 2025.
- [15] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, "Analisis silhouette coefficient pada 6 perhitungan jarak k-means clustering," *Techno.COM*, vol. 20, no. 2, pp. 186–197, 2021.
- [16] T. Santhi, A. M. Sari, D. K. A. M. Putra, G. S. Mahendra, and M. P. Ariasih, "Implementasi business intelligence menggunakan tableau untuk visualisasi prediksi kelulusan mahasiswa," *Jurnal Software Engineering and Information System (SEIS)*, vol. 3, no. 2, pp. 51–58, 2023. [Online]. Available: <https://ejournal.umri.ac.id/index.php/SEIS/index>

- [17] M. Rahman, M. K. F. Fili *et al.*, “Analisis hasil rekapitulasi pilkada daerah khusus jakarta (dkj) 2024 menggunakan metode support vector machine,” *JICode: Jurnal Informatika dan Komputer*, vol. 2, no. 1, pp. 100–111, 2025.
- [18] I. Maria, S. V. Sihotang, R. A. Sunarjo, and A. W. Handaru, “Pemberdayaan komunitas melalui pelatihan pengelolaan keuangan sederhana untuk kesejahteraan ekonomi,” *ADI Bisnis Digital Interdisiplin Jurnal*, vol. 5, no. 2, pp. 33–40, 2024.
- [19] L. P. Refialy, H. Maitimu, and M. S. Pesulima, “Perbaikan kinerja clustering k-means pada data ekonomi nelayan dengan perhitungan sum of square error (sse) dan optimasi nilai k cluster,” *Techno.COM*, vol. 20, no. 2, pp. 321–329, 2021.
- [20] V. A. Ekasetya and A. Jananto, “Klusterisasi optimal dengan elbow method untuk pengelompokan data kecelakaan lalu lintas di kota semarang,” *Dinamika Informatika*, vol. 12, no. 1, pp. 20–28, 2020.
- [21] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, “Analisis silhouette coefficient pada 6 perhitungan jarak k-means clustering,” *Techno.COM*, vol. 20, no. 2, pp. 186–197, 2021.
- [22] I. Geospasial, “Download shapefile batas desa data kependudukan tahun 2019 seluruh indonesia,” 2019, accessed: Mar. 18, 2024. [Online]. Available: <https://drive.google.com/open?id=1v2rEPfN7x-XYWhNslvBiLuu4y7ZSwcSQ>
- [23] T. Santhi, A. M. Sari, D. K. A. M. Putra, G. S. Mahendra, and M. P. Ariasih, “Implementasi business intelligence menggunakan tableau untuk visualisasi prediksi kelulusan mahasiswa,” *Jurnal Software Engineering and Information System (SEIS)*, vol. 3, no. 2, pp. 51–58, 2023. [Online]. Available: <https://ejournal.umri.ac.id/index.php/SEIS/index>
- [24] L. Community, “What are the methods for handling outliers in data mining?” 2024, accessed: Mar. 18, 2024. [Online]. Available: <https://linkedin.com/advice/1/what-methods-handling-outliers-data-mining-skills-data-mining#:~:text=One%20way%20to%20handle%20outliers,%2C%20or%20clustering-based%20measures>
- [25] S. A. Ramadhan and I. Musfiroh, “Review artikel: Verifikasi metode analisis obat,” *Farmaka*, pp. 87–92, 2021.
- [26] S. Soesmono, R. Pertiwi, B. Saputri, N. Putri, and E. Widodo, “Pengelompokan provinsi di indonesia berdasarkan tingkat pengangguran tahun 2023 menggunakan k-medoids: Pengelompokan provinsi di indonesia berdasarkan tingkat pengangguran tahun 2023 menggunakan k-medoids,” *Emerging Statistics and Data Science Journal*, vol. 3, no. 1, pp. 498–515, 2025.
- [27] A. Setiawan, “Box-plot: Memahami dan membuat box and whisker plots dalam eksplorasi data,” 2024, accessed: Mar. 18, 2024. [Online]. Available: <https://www.smartstat.info/materi/statistika/eksplorasi-data/mengenal-box-plot-box-and-whisker-plots.html>
- [28] D. C. Nabilla, “Analisis pengaruh indeks pembangunan manusia, produk domestik regional bruto, dan upah minimum kerja terhadap kemiskinan di provinsi daerah istimewa yogyakarta (2013-2022),” Ph.D. dissertation, Universitas Islam Indonesia, 2025.
- [29] N. P. L. Yashati and N. N. Yuliarmi, “Faktor penentu jumlah anak ideal di kecamatan kuta,” *Jurnal Visi Manajemen*, vol. 11, no. 1, pp. 87–106, 2025.
- [30] D. Yuliana and R. Rosyadi, “Pengaruh jumlah penduduk, investasi, dan indeks pembangunan manusia terhadap ketimpangan pendapatan di kalimantan barat,” *Journal of Economics and Business UBS*, vol. 14, no. 1, pp. 27–47, 2025.
- [31] Y. M. Wibowo and A. D. R. Sari, “Analysis of nitrite (no<sub>2</sub><sup>-</sup>) and nitrate (no<sub>3</sub><sup>-</sup>) levels in well water by viewing light spectrophotometer: Analisis kadar nitrit (no<sub>2</sub><sup>-</sup>) dan nitrat (no<sub>3</sub><sup>-</sup>) pada air sumur secara spektrofotometer sinar tampak,” *Jurnal Kimia dan Rekayasa*, vol. 5, no. 2, pp. 65–70, 2025.