



**ANALISIS PERBANDINGAN METODE *K-MEANS* DAN  
*K-MEDOIDS* DALAM KLASTERISASI TINGKAT PENGANGGURAN  
KABUPATEN KOTA DI PROVINSI KALIMANTAN SELATAN**

**SKRIPSI**

**Untuk memenuhi persyaratan  
dalam menyelesaikan program sarjana Strata-1 Matematika**

**Oleh:**

**ANNISA KAMILAH  
NIM. 2211011220014**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU  
2026**

## HALAMAN PENGESAHAN

### SKRIPSI

#### ANALISIS PERBANDINGAN METODE *K-MEANS* DAN *K-MEDOIDS* DALAM KLASTERISASI TINGKAT PENGANGGURAN KABUPATEN KOTA DI PROVINSI KALIMANTAN SELATAN

Oleh:  
Annisa Kamilah  
NIM. 2211011220014

telah dipertahankan di depan Dosen Penguji pada tanggal 14 Juli 2026  
Susunan Dosen Penguji:

##### Pembimbing I

Thresye, S.Si., M.Si.  
NIP. 197205042000122002

##### Dosen Penguji:

1. Saman Abdurrahman, S.Si., M.Sc. (✓)
2. Sila Rizqina, S.Mat., M.Si. (✓)

##### Pembimbing II

Oni Soesanto, S.Si., M.Si.  
NIP. 1973061262005011003



Banjarbaru, 17 Juli 2026  
Jurusan Matematika FMIPA ULM  
Ketua,

Dr. Na'imah Hijriati, S.Si, M.Si  
NIP. 197911222008012013

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 25 Juli 2026



Annisa Kamilah  
2211011220014

## ABSTRAK

**ANALISIS PERBANDINGAN METODE *K-MEANS* DAN *K-MEDOIDS* DALAM KLASTERISASI TINGKAT PENGANGGURAN KABUPATEN KOTA DI PROVINSI KALIMANTAN SELATAN** (Oleh : Annisa Kamilah; Pembimbing: Thresye, Oni Soesanto, 60 halaman)

Pengangguran menjadi salah satu isu penting dalam bidang ketenagakerjaan karena berpengaruh terhadap pembangunan ekonomi suatu daerah. Variasi tingkat pengangguran yang terjadi antarkabupaten/kota di Provinsi Kalimantan Selatan menunjukkan adanya perbedaan karakteristik dalam bidang ketenagakerjaan. Oleh sebab itu, diperlukan pendekatan yang mampu mengidentifikasi pola kesamaan maupun perbedaan antardaerah sebagai bahan pertimbangan dalam perumusan kebijakan pembangunan di bidang ketenagakerjaan. Analisis *cluster* dilakukan dalam penelitian ini dengan mengimplementasikan algoritma *K-Means* dan *K-Medoids* untuk mengelompokkan kabupaten/kota berdasarkan faktor-faktor tingkat pengangguran. Identifikasi jumlah *cluster* optimal dilakukan menggunakan metode *Silhouette Coefficient*, sedangkan evaluasi kinerja hasil klasterisasi dilakukan melalui perbandingan nilai *Silhouette Coefficient* dan *Davies-Bouldin Index*. Berdasarkan hasil penelitian, jumlah *cluster* optimal untuk kedua algoritma yaitu sebanyak 2 *cluster*. Hasil evaluasi menunjukkan bahwa kualitas klasterisasi yang dihasilkan algoritma *K-Means* lebih baik dibandingkan *K-Medoids*, dengan nilai *Silhouette Coefficient K-Means* sebesar 0,3823 yang lebih tinggi dibandingkan *K-Medoids* sebesar 0,3581, serta nilai DBI *K-Means* sebesar 1,04065 yang lebih rendah dibandingkan *K-Medoids* sebesar 1,2343. Meskipun demikian, nilai *Silhouette Coefficient* sebesar 0,3823 masih tergolong dalam kategori struktur *cluster* lemah, yang menunjukkan bahwa pemisahan antar*cluster* belum terbentuk secara optimal.

Kata Kunci : Tingkat Pengangguran, *K-Means*, *K-Medoids*, *Silhouette Coefficient*, *Davies- Bouldin Index*

## **ABSTRACT**

**COMPARATIVE ANALYSIS OF THE K-MEANS AND K-MEDOID'S METHODS IN CLUSTERING UNEMPLOYMENT RATES IN DISTRICTS AND CITIES IN SOUTH KALIMANTAN PROVINCE** (By: Annisa Kamilah; Advisors: Thresye, Oni Soesanto, 60 pages)

*Unemployment is a key issue in the labor market because it affects a region's economic development. Variations in unemployment rates across regencies and cities in South Kalimantan Province indicate differences in labor market characteristics. Therefore, an approach is needed that can identify patterns of similarities and differences among regions to serve as a basis for formulating labor market development policies. Cluster analysis was conducted in this study by implementing the K-Means and K-Medoids algorithms to group regencies and cities based on unemployment rate factors. The optimal number of clusters was identified using the Silhouette Coefficient method, while the performance of the clustering results was evaluated by comparing the Silhouette Coefficient and the Davies-Bouldin Index. Based on the research results, the optimal number of clusters for both algorithms was 2. The evaluation results show that the clustering quality of the K-Means algorithm is better than that of K-Medoids, where the K-Means Silhouette Coefficient value of 0.3823 exceeds the K-Medoids value of 0.3581, while the K-Means DBI value of 1.04065 is lower than the K-Medoids DBI value of 1.2343. Nevertheless, a Silhouette Coefficient value of 0.3823 still falls into the category of weak cluster structure, indicating that the separation between clusters has not yet been optimally established.*

*Keywords: Unemployment Rate, K-Means, K-Medoids, Silhouette, Davies- Bouldin Index*

## PRAKATA

Puji syukur penulis panjatkan kehadiran Allah Subhanahu Wa Ta'ala yang telah melimpahkan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Perbandingan Metode *K-Means* dan *K-Medoids* dalam Klasterisasi Tingkat Pengangguran Kabupaten Kota di Provinsi Kalimantan Selatan”. Shalawat serta salam tidak lupa dihaturkan kepada Nabi Muhammad Shalallahu ‘Alaihi Wa Sallam., keluarga, sahabat, serta pengikut beliau hingga akhir zaman.

Penyusunan skripsi ini bertujuan untuk memenuhi persyatan dalam menyelesaikan program sarjana Strata-1 Matematika di Jurusan Matematika Fakultas Matematika dan Ilmi Pengetahuan Alam Universitas Lambung Mangkurat.

Penulis menyadari bahwa penyelesaian skripsi ini tidak terlepas dari bantuan, arahan, serta motivasi dari berbagai pihak. Oleh karena itu, penulis menyampaikan rasa hormat dan terima kasih kepada :

1. Bapak Prof. Sunardi, S.Si., M.Sc., Ph.D. selaku Dekan Fakultas Matematika dan Ilmi Pengetahuan Alam Universitas Lambung Mangkurat.
2. Ibu Dr. Naimah Hijriati, S.Si., M.Si. selaku koordinator Jurusan Matematika beserta seluruh dosen dan staf Jurusan Matematika.
3. Bapak Akhmad Yusuf S.Si., M.Kom. dan Ibu Sila Rizqina, S.Mat., M.Si. selaku dosen Penasehat Akademik yang telah memberikan arahan, motivasi, dan bimbingan selama menjalani perkuliahan.
4. Ibu Thresye, S.Si., M.Si. dan Bapak Oni Soesanto, S.Si., M.Si. selaku dosen pembimbing yang bersedia memberikan bimbingan, nasehat, petunjuk serta motivasi dalam pelaksanaan penyusunan skripsi ini.
5. Mama tercinta penulis yang selalu mendoakan dengan tulus dan memberikan dukungan moral maupun materil tiada henti dan lelah. Serta kepada ayah penulis yang semasa hidupnya selalu mendoakan dan memberikan dukungan serta motivasi untuk semangat melanjutkan sekolah sampai ke perguruan tinggi.
6. Keluarga yang selalu mendoakan dan memberikan dukungan dalam penyusunan skripsi ini.

7. Arsella Putri Nabila dan Syintia Amanda selaku sahabat penulis yang telah mengingatkan dalam kebaikan dan membantu penulis dalam menjalani masa perkuliahan.
8. Teman-teman serta seluruh pihak lain yang tidak dapat penulis sebutkan satu persatu, yang telah memberikan dukungan dan bantuan dalam proses penyusunan skripsi ini.

Semoga segala kebaikan kalian memperoleh balasan dan pahala dari Allah Subhanahu Wa Ta'ala. Penulis menyadari adanya ketidaktelitian, kekurangan dan kesalahan dalam penulisan skripsi ini. Oleh karena itu, penulis menerima kritik dan saran yang bersifat membangun. Semoga penulisan skripsi ini dapat bermanfaat bagi pembaca dan semua pihak.

Banjarbaru, 25 Juli 2026



Annisa Kamilah  
2211011220014

## DAFTAR ISI

|                                                               | Halaman     |
|---------------------------------------------------------------|-------------|
| <b>HALAMAN PENGESAHAN</b>                                     | <b>ii</b>   |
| <b>PERNYATAAN</b>                                             | <b>iii</b>  |
| <b>ABSTRAK</b>                                                | <b>iv</b>   |
| <b><i>ABSTRACT</i></b>                                        | <b>v</b>    |
| <b>PRAKATA</b>                                                | <b>vi</b>   |
| <b>DAFTAR ISI</b>                                             | <b>viii</b> |
| <b>DAFTAR TABEL</b>                                           | <b>x</b>    |
| <b>DAFTAR GAMBAR</b>                                          | <b>xii</b>  |
| <b>DAFTAR LAMPIRAN</b>                                        | <b>xiii</b> |
| <b>ARTI LAMBANG DAN SINGKATAN</b>                             |             |
| Error! Bookmark not defined.                                  |             |
| <b>BAB I PENDAHULUAN</b>                                      | <b>1</b>    |
| 1.1 Latar Belakang.....                                       | 1           |
| 1.2 Tujuan Penelitian.....                                    | 3           |
| 1.3 Sistematika Penulisan.....                                | 4           |
| <b>BAB II TINJAUAN PUSTAKA</b>                                | <b>5</b>    |
| 2.1 <i>Data Mining</i> .....                                  | 5           |
| 2.1.1 Pengertian <i>Data Mining</i> .....                     | 5           |
| 2.1.2 Teknik <i>Data Mining</i> .....                         | 6           |
| 2.1.3 <i>Cross-Industry Standard Process (CRISP-DM)</i> ..... | 7           |
| 2.1.4 Uji Multikolinearitas.....                              | 9           |
| 2.1.5 <i>Outlier</i> dan Metode Deteksi <i>Outlier</i> .....  | 9           |
| 2.1.6 <i>Winsorisasi Outlier</i> .....                        | 10          |
| 2.1.7 Normalisasi Data.....                                   | 11          |
| 2.2 <i>Clustering</i> .....                                   | 11          |
| 2.2.1 Jenis-Jenis <i>Clustering</i> .....                     | 12          |
| 2.2.2 <i>K-Means Clustering</i> .....                         | 13          |
| 2.2.3 <i>K-Medoids Clustering</i> .....                       | 15          |
| 2.3 <i>Silhouette Coefficient</i> .....                       | 16          |



|                                                                                                                          |           |
|--------------------------------------------------------------------------------------------------------------------------|-----------|
| 2.4 <i>Davies-Bouldin Index (DBI)</i> .....                                                                              | 18        |
| 2.5 Pengangguran .....                                                                                                   | 19        |
| <b>BAB III PROSEDUR PENELITIAN</b>                                                                                       | <b>24</b> |
| 3.1 Metode Analisis .....                                                                                                | 24        |
| 3.2 Data Penelitian.....                                                                                                 | 25        |
| 3.3 Prosedur Penelitian .....                                                                                            | 25        |
| 3.4 Diagram Alir Penelitian.....                                                                                         | 27        |
| <b>BAB IV HASIL DAN PEMBAHASAN</b>                                                                                       | <b>28</b> |
| 4.1 Implementasi Algoritma <i>K-Means</i> dan <i>K-Medoids</i> .....                                                     | 28        |
| 4.1.1 <i>Preprocessing</i> Data .....                                                                                    | 28        |
| 4.1.2 Klasterisasi dengan Algoritma <i>K-Means</i> .....                                                                 | 36        |
| 4.1.3 Klasterisasi dengan Algoritma <i>K-Medoids</i> .....                                                               | 40        |
| 4.2 Identifikasi Jumlah <i>Cluster</i> Optimal.....                                                                      | 45        |
| 4.2.1 Jumlah <i>Cluster</i> Optimal untuk Algoritma <i>K-Means</i> .....                                                 | 45        |
| 4.2.2 Jumlah <i>Cluster</i> Optimal untuk Algoritma <i>K-Medoids</i> .....                                               | 47        |
| 4.3 Perbandingan Algoritma <i>K-Means</i> dan Algoritma <i>K-Medoids</i> .....                                           | 50        |
| 4.3.1 Perhitungan Nilai <i>Davies Bouldin Index</i> Algoritma <i>K-Means</i> .....                                       | 50        |
| 4.3.2 Perhitungan Nilai <i>Davies Bouldin Index</i> Algoritma <i>K-Medoids</i> .....                                     | 53        |
| 4.3.3 Perbandingan Metode Berdasarkan Nilai <i>Silhouette Coefficient</i> dan<br><i>Davies Bouldin Index (DBI)</i> ..... | 56        |
| <b>BAB V PENUTUP</b>                                                                                                     | <b>60</b> |
| <b>5.1 Kesimpulan</b> .....                                                                                              | 60        |
| <b>5.2 Saran</b> .....                                                                                                   | 60        |
| <b>DAFTAR PUSTAKA</b>                                                                                                    | <b>62</b> |
| <b>LAMPIRAN</b>                                                                                                          |           |

## DAFTAR TABEL

|                                                                                                       | Halaman |
|-------------------------------------------------------------------------------------------------------|---------|
| <b>Tabel 3.1</b> Variabel Penelitian .....                                                            | 25      |
| <b>Tabel 4.1</b> Data Penelitian.....                                                                 | 28      |
| <b>Tabel 4.2</b> Nilai $R^2$ Setiap Variabel .....                                                    | 29      |
| <b>Tabel 4.3</b> Nilai VIF Setiap Variabel.....                                                       | 29      |
| <b>Tabel 4.4</b> Nilai $R^2$ dan VIF Setelah Variabel $X_2$ Dihapus.....                              | 31      |
| <b>Tabel 4.5</b> Data Variabel $X_6$ .....                                                            | 31      |
| <b>Tabel 4.6</b> Deteksi <i>Outlier</i> dengan IQR.....                                               | 32      |
| <b>Tabel 4.7</b> Data Penelitian Setelah Winsorisasi .....                                            | 33      |
| <b>Tabel 4.8</b> Data Normalisasi dengan <i>Min-Max</i> .....                                         | 35      |
| <b>Tabel 4.9</b> <i>Centroid</i> Awal .....                                                           | 36      |
| <b>Tabel 4.10</b> Perhitungan Jarak Objek Data ke <i>Centroid</i> Awal .....                          | 37      |
| <b>Tabel 4.11</b> Pengelompokan <i>Cluster</i> Data .....                                             | 37      |
| <b>Tabel 4.12</b> <i>Centroid</i> Baru .....                                                          | 38      |
| <b>Tabel 4.13</b> Pengelompokan <i>Cluster</i> dengan <i>Centroid</i> Baru.....                       | 39      |
| <b>Tabel 4.14</b> Hasil Klasterisasi .....                                                            | 39      |
| <b>Tabel 4.15</b> Hasil Klasterisasi dari $k = 2$ sampai $k = 7$ Algoritma <i>K-Means</i> .....       | 40      |
| <b>Tabel 4.16</b> <i>Medoid</i> Awal .....                                                            | 41      |
| <b>Tabel 4.17</b> Hasil Perhitungan Jarak Objek Data Ke <i>Medoid</i> Awal .....                      | 41      |
| <b>Tabel 4.18</b> Pengelompokan <i>Cluster</i> Data .....                                             | 42      |
| <b>Tabel 4.19</b> <i>Medoid</i> Baru .....                                                            | 43      |
| <b>Tabel 4.20</b> Pengelompokan <i>Cluster</i> dengan <i>Medoid</i> Baru .....                        | 43      |
| <b>Tabel 4.21</b> Hasil Klasterisasi dengan <i>K-Medoids</i> .....                                    | 44      |
| <b>Tabel 4.22</b> Hasil Klasterisasi dari $k = 2$ sampai $k = 7$ Algoritma <i>K-Medoids</i> ...       | 45      |
| <b>Tabel 4.23</b> Hasil Perhitungan <i>Silhouette Coefficient</i> untuk 2 Cluster <i>K-Means</i> ...  | 47      |
| <b>Tabel 4.24</b> Hasil Perhitungan <i>Silhouette Coefficient K-Means</i> .....                       | 47      |
| <b>Tabel 4.25</b> Hasil Perhitungan <i>Silhouette Coefficient</i> untuk 2 <i>Cluster K-Medoids</i> .. | 49      |
| <b>Tabel 4.26</b> Hasil Perhitungan <i>Silhouette Coefficient K-Medoids</i> .....                     | 50      |
| <b>Tabel 4.27</b> Hasil Perhitungan Jarak antarcentroid.....                                          | 52      |

|                                                             |    |
|-------------------------------------------------------------|----|
| <b>Tabel 4.28</b> Hasil Perhitungan Nilai Rasio .....       | 52 |
| <b>Tabel 4.29</b> Hasil Perhitungan Jarak antarmedoid ..... | 55 |
| <b>Tabel 4.30</b> Hasil Perhitungan Nilai Rasio .....       | 55 |
| <b>Tabel 4.31</b> Deskripsi Cluster .....                   | 57 |

## DAFTAR GAMBAR

|                                                                              | <b>Halaman</b> |
|------------------------------------------------------------------------------|----------------|
| <b>Gambar 3.1</b> Diagram Alir Penelitian .....                              | 27             |
| <b>Gambar 4.1</b> Nilai Korelasi antarvariabel .....                         | 30             |
| <b>Gambar 4.2</b> Deteksi Outlier pada Data Penelitian Menggunakan IQR ..... | 33             |
| <b>Gambar 4.3</b> Penanganan Outlier dengan Winsorisasi .....                | 34             |

## DAFTAR LAMPIRAN

- Lampiran 1.** Script Google Colab Algoritma *K-Means*, *Silhouette Coefficient* dan *DBI*
- Lampiran 2.** Script Google Colab Algoritma *K-Medoids*, *Silhouette Coefficient* dan *DBI*
- Lampiran 3.** Perhitungan Data Normalisasi
- Lampiran 4.** Perhitungan Jarak Objek Data ke *Centroid* Awal
- Lampiran 5.** Perhitungan *Centroid* Baru
- Lampiran 6.** Perhitungan Jarak Objek Data Ke *Medoid* Awal
- Lampiran 7.** Perhitungan nilai *Silhouette Coefficient* untuk  $k=2$  Algoritma *K-Means*
- Lampiran 8.** Perhitungan nilai *Silhouette Coefficient* untuk  $k=2$  Algoritma *K-Medoids*
- Lampiran 9.** Perhitungan *Excel* Algoritma *K-Means* dari  $k=3$  sampai  $k=7$
- Lampiran 10.** Perhitungan *Excel* Algoritma *K-Medoids* dari  $k=3$  sampai  $k=7$
- Lampiran 11.** Perhitungan *Excel* Nilai *Silhouette Coefficient*  $k=2$  sampai  $k=7$  Algoritma *K-Means*
- Lampiran 12.** Perhitungan *Excel* Nilai *Silhouette Coefficient*  $k=2$  sampai  $k=7$  Algoritma *K-Medoids*

## ARTI LAMBANG DAN SINGKATAN

|               |                                                                          |
|---------------|--------------------------------------------------------------------------|
| CRISP-DM      | : <i>Cross-Industry Standard Process for Data Mining</i>                 |
| IQR           | : <i>Interquartile Range</i>                                             |
| Q1            | : kuartil pertama                                                        |
| Q3            | : kuartil ketiga                                                         |
| DBI           | : <i>Davies-Bouldin Index</i>                                            |
| $VIF_i$       | : <i>Variance Inflation Factor</i> untuk variabel ke- $i$                |
| $R_i^2$       | : Koefisien determinasi dari regresi variabel ke- $i$ terhadap variabel  |
| S             | : Simpangan                                                              |
| SSW           | : <i>Sum of Square Within Cluster</i>                                    |
| SSB           | : <i>Sum of Square Between Cluster</i>                                   |
| $x'$          | : Nilai hasil normalisasi                                                |
| $x$           | : Nilai asli yang akan dinormalisasi                                     |
| $\min(v)$     | : Nilai minimal dari sebuah atribut                                      |
| $\max(v)$     | : Nilai maksimal dari sebuah atribut                                     |
| $V_k$         | : <i>Centroid</i> pada <i>cluster</i> ke $k$                             |
| $X_i$         | : Data ke $i$                                                            |
| $N_k$         | : Banyaknya objek/jumlah data yang menjadi anggota <i>cluster</i> ke $k$ |
| $d(x_i, y_j)$ | : Jarak <i>Euclidean</i> antara data ke- $i$ ke <i>centroid</i> $j$      |
| $x_{ik}$      | : Data ke- $i$ pada atribut data ke- $k$                                 |
| $c_{ik}$      | : <i>Centroid</i> ke- $j$ pada atribut data ke- $k$                      |
| $d(x_i, m_j)$ | : Jarak <i>Euclidean</i> antara data ke- $i$ ke <i>medoid</i> $j$        |
| $m_{ik}$      | : <i>Medoid</i> ke- $j$ pada atribut data ke- $k$                        |
| $A$           | : Jumlah data di <i>cluster</i> A                                        |
| $i, j$        | : Indeks dari data                                                       |
| $d(i, j)$     | : Jarak antara data $i$ dengan $j$                                       |
| $d(i, C)$     | : Jarak antara data $i$ dengan semua data di <i>cluster</i> $C$          |

|               |                                                                                                      |
|---------------|------------------------------------------------------------------------------------------------------|
| $C$           | : banyaknya data di <i>cluster</i> $C$                                                               |
| $b(i)$        | : Nilai minimum dari jarak rata-rata antara data ke- $i$ dengan semua data di <i>cluster</i> berbeda |
| $s(i)$        | : <i>Sillhoutte Coefficient</i>                                                                      |
| $a(i)$        | : rata-rata data ke- $i$ di <i>cluster</i> $A$                                                       |
| $m_i$         | : Jumlah data dalam <i>cluster</i> ke- $i$                                                           |
| $c_i$         | : <i>Centroid</i> pada <i>cluster</i> ke- $i$                                                        |
| $d(x_j, c_i)$ | : Jarak antara setiap data $x_j$ dengan <i>centroid</i> $c_i$                                        |
| $d(c_i, c_j)$ | : Jarak antar <i>centroid</i>                                                                        |
| $k$           | : Jumlah <i>cluster</i> yang digunakan                                                               |
| $R_{i,j}$     | : Rasio antara <i>cluster</i> ke- $i$ dan <i>cluster</i> ke- $j$                                     |
| $C_i$         | : <i>Cluster</i> ke - $i$                                                                            |
| $Ci$          | : <i>Centroid</i> (pusat <i>cluster</i> ) ke $i$ , dengan $i = 1, 2, \dots$                          |
| $Mi$          | : <i>Medoid</i> ke $i$ , dengan $i = 1, 2, \dots$                                                    |