PERBANDINGAN *K-MEANS CLUSTER* DENGAN *DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE* PADA DATA INDIKATOR KEMISKINAN PROVINSI SULAWESI SELATAN

SKRIPSI

SAPRIADI RASYID

H051191003

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2023

PERBANDINGAN *K-MEANS CLUSTER* DENGAN *DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE* PADA DATA INDIKATOR KEMISKINAN PROVINSI SULAWESI SELATAN

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

SAPRIADI RASYID

H051191003

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

SEPTEMBER 2023

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Perbandingan *K-Means Cluster* dengan *Density-Based Spatial Clustering of Application with Noise* pada Data Indikator Kemiskinan Provinsi Sulawesi Selatan

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 27 September 2023

Sapriadi Rasyid

NIM H051191003

PERBANDINGAN *K-MEANS CLUSTER* DENGAN *DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE* PADA DATA INDIKATOR KEMISKINAN PROVINSI SULAWESI SELATAN

Disetujui Oleh:

Pembimbing Utama	Pembimbing Pertama
Siswanto, S.Si., M.Si.	**Sitti Sahriman, S.Si., M.Si.**
NIP. 19920107 201903 1 012	**NIP. 19881018 201504 2 002**

Ketua Program Studi

Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19770808 200501 2 002

Pada 27 September 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama	: Sapriadi Rasyid
NIM	: H051191003
Program Studi	: Statistika
Judul Skripsi	: Perbandingan *K-Means Cluster* dengan *Density-Based Spatial Clustering of Application with Noise* pada Data Indikator Kemiskinan Provinsi Sulawesi Selatan

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Siswanto, S.Si., M.Si. (.....)
2. Sekretaris : Sitti Sahriman, S.Si., M.Si. (.....)
3. Anggota : Anisa, S.Si., M.Si. (.....)
4. Anggota : Andi Kresna Jaya, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 27 September 2023

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur senantiasa dipanjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya. *Alhamdulillahirobbil'alamin*, berkat nikmat kemudahan dan pertolongan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul **"Perbandingan *K-Means Cluster* dengan *Density-Based Spatial Clustering of Application with Noise* pada Data Indikator Kemiskinan Provinsi Sulawesi Selatan"** yang disusun sebagai salah satu syarat akademik untuk memperoleh gelar sarjana pada Program Studi Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyelesaian skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak yang senantiasa turut membantu dalam bentuk moril maupun materil sehingga dengan segala keterbatasan kemampuan dan pengetahuan, penulis dapat menyelesaikan skripsi ini. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih dengan penuh keikhlasan dan setulus-tulusnya kepada Ayahanda Abdul Rasyid dg Muntu dan Ibunda Lebuati dg Te'ne (Almarhuma) yang memberikan dukungan penuh, pengorbanan luar biasa, limpahan cinta dan kasih sayang, kesabaran hati, serta dengan ikhlas telah menemani setiap langkah penulis dengan doa dan restu mulianya. Ucapan terima kasih penulis haturkan kepada Kakak tersayang penulis, Sitti Nadira, Nurlaeli, terutama Yunita Rasyid yang selalu menghibur dan memberikan bantuan dan semangat kepada penulis dalam menyelesaikan skripsi. Serta keluarga besar penulis yang tercinta, terima kasih atas hiburan, doa mulia, serta dukungannya kepada penulis selama ini.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan dan ketulusan juga penulis ucapkan kepada:

1. Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc., selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.

2. Bapak Dr. Eng. Amiruddin, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. Ibu Dr. Anna Islamiyati, S.Si., M.Si., selaku Ketua Departemen Statistika serta segenap dosen pengajar dan staf Departemen Statistika yang telah memberikan arahan, dorongan semangat dan motivasi kepada penulis selama menjadi mahasiswa di Departemen Statistika.
4. Bapak Siswanto, S.Si., M.Si. selaku Dosen Pembimbing Utama sekaligus sebagai Penasehat Akademik penulis yang dengan penuh kesabaran telah memberikan arahan, dorongan semangat dan motivasi selama menjadi mahasiswa di Departemen Statistika. Serta telah membimbing, meluangkan waktu, tenaga, serta pemikirannya sehingga skripsi ini dapat diselesaikan.
5. Ibu Sitti Sahriman, S.Si., M.Si. selaku Dosen Pembimbing Pertama penulis yang dengan penuh kesabaran telah membimbing, meluangkan waktu, tenaga, serta pemikirannya sehingga skripsi ini dapat diselesaikan.
6. Bapak Andi Kresna Jaya, S.Si., M.Si., selaku Dosen Penguji Pertama penulis dan Ibu Anisa, S.Si., M.Si., selaku Dosen Penguji Kedua penulis yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya untuk senantiasa memberikan masukan serta arahan sehingga menjadikan skripsi ini lebih baik, serta memberikan dorongan semangat, dan motivasi kepada penulis dari awal hingga selesainya penulisan.
7. Saudara dan Saudari penulis, Taufiq Akbar, Muhammad Rais, Nurkhaliza, dan Hukma yang senantiasa meluangkan waktu untuk membantu serta memberikan masukan serta pendapatnya dalam penyusunan skripsi.
8. Sahabat SMA penulis, Andi Chaerul Eka Putra, Muhammad Zulkifli, Andi Muh. Reza Pahlevi, dan Abdul Hakim Azis yang selalu mendengarkan segala keluh kesah penulis dalam hal apapun, senantiasa memberikan semangat, dan menemani perjalanan suka dan duka penulis selama ini.
9. Sahabat sepergabutan penulis selama penulisan Skripsi, Agus Hermawan, Muhammad Faturrahman, Fadilah Amirul Adhel, Muhammad Syamsul Bahri, Arief Rahman Nur, Taufik Hidayat, Nurazizah, Yasmin Pratiwi, Muliana, Nursyahfika, Mayashari, dan Nur Aisyah. Terima kasih atas semua hal yang kita lalui bersama selama proses perkuliahan.

10. Teman sekaligus tempat curhat penulis dalam proses penyusunan Skripsi, Fadhila Febriyanti Najamuddin, Diah Lestari, dan Wahyu Dwi Rahmawati serta akun dan pengikut setia mydestination16.
11. Semua teman seperjuangan di Statistika 2019. Terima kasih atas ilmu, kebersamaan, suka dan duka dalam menjalani perkuliahan di Departemen Statistika. Terima kasih sudah menerima kehadiran penulis.
12. Keluarga posko 22 KKNT Gel. 108, Muhammad Arfani Asra, Senap, dan Selva Ananda Arung. Terima kasih untuk 60 hari terbaik di sepanjang 2022.
13. Pusat Penelitian Kelapa Sawit, Badan Pusat Statistik Kabupaten Pangkep, Pihak MKU, dan UKM LDK-MPM UNHAS, terima kasih sudah hadir di kehidupan saya yang telah mengembangkan bakat dan nilai kompetensi saya selama proses perkuliahan berlangsung.
14. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan, partisipasi, dan apresiasi yang diberikan kepada penulis.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga karya ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 27 September 2023

Sapriadi Rasyid

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama	: Sapriadi Rasyid
NIM	: H051191003
Program Studi	: Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis Karya	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif** (***Non-exclusive Royalty- Free Right***) atas tugas akhir saya yang berjudul:

"Perbandingan *K-Means Cluster* dengan *Density-Based Spatial Clustering of Application with Noise* pada Data Indikator Kemiskinan Provinsi Sulawesi Selatan"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar tanggal 27 September 2023.

Yang menyatakan,

(Sapriadi Rasyid)

ABSTRAK

Program penanggulangan kemiskinan telah dilakukan pemerintah seperti adanya program pemberian bantuan masyarakat miskin. Namun, jumlah penduduk miskin di Sulawesi Selatan masih meningkat. Oleh karena itu, diperlukan metode statistik untuk mengelompokkan masyarakat miskin berdasarkan indikator kemiskinan setiap wilayah sebagai acuan dalam memberikan bantuan. Metode statistik yang tepat yaitu analisis *cluster* dengan meminimalkan perbedaan objek dalam satu *cluster* serta memaksimalkan perbedaan objek antar *cluster*. Penelitian ini menggunakan dua metode yaitu *K-Means* dan *Density-Based Spatial Clustering of Application with Noise* (DBSCAN). Hal ini bertujuan untuk membandingkan metode *K-Means* dengan DBSCAN berdasarkan *Silhouette Coefficient*. Data yang digunakan dalam penelitian ini terdiri dari delapan indikator kemiskinan Provinsi Sulawesi Selatan tahun 2022. Hasil analisis menggunakan *K-Means* memperoleh dua *cluster* optimal, *cluster* 1 terdiri dari 23 kabupaten dan kota, sedangkan *cluster* 2 hanya Kota Makassar. Sedangkan, hasil analisis lanjut pada *cluster* 1 menghasilkan delapan *cluster* baru dengan *Silhouette Coefficient* sebesar 0,507. Berbeda dengan menggunakan metode DBSCAN menghasilkan satu *cluster* yang terdiri dari 23 kabupaten dan kota, sedangkan Kota Makassar terdeteksi sebagai *noise*. Sedangkan, hasil analisis lanjut pada *cluster* tersebut terdiri dari satu *cluster* dengan tiga *noise* dan menghasilkan *Silhouette Coefficient* sebesar 0,318. Dengan demikan, berdasarkan *Silhouette Coefficient* dapat disimpulkan bahwa *K-Means* memiliki hasil pengelompokkan yang lebih baik pada data indikator kemiskinan Provinsi Sulawesi Selatan.

Kata Kunci: *Cluster*, DBSCAN, Kemiskinan, *K-Means*, *Silhouette Coefficient*.

ABSTRACT

The government has implemented poverty alleviation programs, including assistance programs for the poor. Despite these efforts, the number of impoverished individuals in South Sulawesi continues to rise. To address this issue, a statistical method is necessary to cluster the poor based on error indicators for each region, serving as a reference for providing assistance. The appropriate statistical method is cluster analysis by minimizing object differences within one cluster and maximizing object differences between clusters. This study employs two methods, namely K-Means and Density-Based Spatial Clustering of Application with Noise (DBSCAN), to compare their effectiveness based on the Silhouette Coefficient. The data used for analysis includes of eight poverty indicators for the in South Sulawesi province in 2022. The findings reveal that the K-Means method results in two optimal clusters, with cluster 1 comprising 23 regencies and cities, and cluster 2 comprising only Makassar City. While the results of the further analysis on cluster 1 consisted of eight new clusters and produced a Silhouette Coefficient of 0,507. In contrast, the DBSCAN method yields one cluster, encompassing 23 regencies and cities, with Makassar City identified as noise. While the results of further analysis on the clusters consisted of one cluster with three noises and produced a Silhouette Coefficient of 0,318. Consequently, the study concludes that K-Means has a greater Silhouette Coefficient value in the case of clustering poverty indicators for South Sulawesi Province.

Keywords**:** *Cluster, DBSCAN, Poverty, K-Means, Silhouette Coefficient.*

DAFTAR ISI

HALAMAN SAMPUL ..... **i**
HALAMAN JUDUL ..... **ii**
LEMBAR PERNYATAAN KEOTENTIKAN ..... **iii**
LEMBAR PERSETUJUAN ..... **iv**
HALAMAN PENGESAHAN ..... **v**
KATA PENGANTAR ..... **vi**
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK ..... **ix**
ABSTRAK ..... **x**
ABSTRACT ..... **xi**
DAFTAR ISI ..... **xii**
DAFTAR GAMBAR ..... **xiv**
DAFTAR TABEL ..... **xv**
DAFTAR LAMPIRAN ..... **xvi**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 4
1.5 Manfaat Penelitian 4
BAB II TINJAUAN PUSTAKA **5**
2.1 Analisis *Cluster* 5
2.1.1 Jarak *Euclidean* 5
2.1.2 Multikolinearitas 6
2.2 *Principal Component Analysis* 6
2.2.1 Matriks Korelasi dan Matriks Varian-Kovarian 7
2.2.2 Koefisien Pembobot dan Penentuan Komponen Utama 8
2.3 *K-Means Cluster* 9
2.4 *Density-Based Spatial Clustering of Application with Noise* 10
2.5 *Silhouette Coefficient* 11
2.6 Kemiskinan 12

BAB III METODOLOGI PENELITIAN 14
3.1 Data 14
3.2 Metode Analisis..... 14
BAB IV HASIL DAN PEMBAHASAN..... 17
4.1 Deskripsi Data 17
4.2 Uji Multikolinearitas 19
4.3 Pembentukan Komponen Utama dengan *Principal Component Analysis* 20
4.4 Mengukur Kedekatan Antar Objek 21
4.5 Analisis *Cluster* dengan Metode *K-Means*..... 22
4.5.1 Penentuan Nilai *k* Terbaik..... 22
4.5.2 Penentuan *Cluster* 23
4.5.3 Interpretasi Hasil *Cluster* 26
4.6 Analisis *Cluster* dengan Metode DBSCAN 30
4.6.1 Penentuan *Minimum Points* dan Tingkatan Hirarki Terbaik..... 30
4.6.2 Langkah-langkah Penentuan *Cluster*..... 31
4.6.3 Interpretasi Hasil *Cluster* 33
4.7 Evaluasi dan Perbandingan Hasil *Cluster* 36
4.7.1 Evaluasi Hasil *Cluster* dari *K-Means*..... 36
4.7.2 Evaluasi Hasil *Cluster* dari DBSCAN..... 38
BAB V KESIMPULAN..... 40
5.1 Kesimpulan..... 40
5.2 Saran..... 40
DAFTAR PUSTAKA..... 41
LAMPIRAN..... 45

DAFTAR GAMBAR

Gambar 4.1 Plot korelasi antar variabel 19
Gambar 4.2 Grafik penentuan k menggunakan *Silhouette Coefficient* 23
Gambar 4.3 Hasil *cluster* dengan Metode *K-Means*..... 26
Gambar 4.4 Hasil *cluster* lanjut dengan Metode *K-Means* 27
Gambar 4.5 Grafik penentuan hirarki menggunakan Silhouette *Coefficient*..... 31
Gambar 4.6 Objek p^* awal yang terpilih..... 31
Gambar 4.7 Objek p^* kedua yang termasuk *core point* 32
Gambar 4.8 Hasil *cluster* dengan Metode DBSCAN 34
Gambar 4.9 Hasil *cluster* lanjut dengan Metode DBSCAN 35

DAFTAR TABEL

Tabel 2.1 Kriteria Evaluasi SC 12
Tabel 3.1 Variabel Indikator Kemiskinan 14
Tabel 4.1 Statistika Deskriptif Data Indikator Kemiskinan di Provinsi Sulawesi Selatan Tahun 2022 17
Tabel 4.2 Nilai $Tolerance$ dan VIF Setiap Variabel 19
Tabel 4.3 Proporsi Keragaman Kumulatif dari KU 20
Tabel 4.4 Nilai SC untuk Setiap k 22
Tabel 4.5 *Centroid* Awal 23
Tabel 4.6 Perhitungan Jarak Antara *Centroid* dan Objek Lainnya 24
Tabel 4.7 *Centroid* Iterasi 2 25
Tabel 4.8 Perhitungan Jarak Antara *Centroid* Iterasi 2 dan Objek Lainnya 25
Tabel 4.9 Karakterisitk Deskriptif Metode *K-Means* 27
Tabel 4.10 Karakterisitk Deskriptif Metode *K-Means* pada Analisis Lanjut 28
Tabel 4.11 Nilai SC untuk Setiap Tingkatan Hirarki 30
Tabel 4.12 Karakteristik Deskriptif Metode DBSCAN 34
Tabel 4.13 Karakteristik Deskriptif Metode DBSCAN pada Analisis Lanjut 36

DAFTAR LAMPIRAN

Lampiran 1 Data Indikator Kemiskinan Provinsi Sulawesi Selatan Tahun 2022. 46
Lampiran 2 Hasil Standarisasi Data Indikator Kemiskinan dengan *Z-Score*..... 47
Lampiran 3 Nilai dari Komponen Utama yang Terbentuk..... 48
Lampiran 4 Jarak Karakteristik Antar Objek dengan Jarak *Euclidean*..... 49
Lampiran 5 Dendogram Metode DBSCAN Berbasis Hirarki..... 52
Lampiran 6 Nilai *Silhouette* Setiap Kabupaten dan Kota..... 53

BAB I
PENDAHULUAN

1.1 Latar Belakang

Kemiskinan adalah suatu kondisi kekurangan sehingga tidak mampu memenuhi kebutuhan hidup minimal yang dialami seseorang atau rumah tangga. Hal itu berupa kebutuhan pangan, sandang, perumahan dan kebutuhan sosial yang dibutuhkan untuk memenuhi hidup secara layak. Bahkan, kemiskinan menjadi tantangan terbesar pada pembangunan nasional (Febianto & Palasara, 2019). Program penanggulangan kemiskinan telah dilakukan pemerintah seperti adanya program keluarga harapan, jaminan kesehatan masyarakat hingga membuat program bantuan pendidikan (Kominfo, 2011). Namun, pada September 2022, jumlah penduduk miskin di Sulawesi Selatan masih meningkat sebanyak 16,86 ribu jiwa dalam setahun terakhir (BPS, 2023). Faktor utama yang menyebabkan penduduk miskin meningkat adalah adanya ketidakefektifan dalam memberikan bantuan pada masyarakat miskin. Oleh karena itu, diperlukan metode yang dapat mengelompokkan masyarakat miskin berdasarkan indikator kemiskinan setiap wilayah sebagai acuan dalam memberikan bantuan.

Indikator kemiskinan berdasarkan Rahmawati (2016), Abidin (2017), Rais dkk (2021), dan BPS (2023) terdiri dari beberapa indikator, diantaranya yaitu indeks pembangunan manusia, jumlah penduduk, jumlah angkatan kerja, persentase penduduk miskin, partisipasi angkatan kerja, jumlah pengangguran, kepadatan penduduk serta pengeluaran per kapita. Indikator kemiskinan tersebut dapat menunjukkan karakteristik kemiskinan setiap wilayah berdasarkan kemiripan. Metode statistika yang dapat digunakan untuk mengelompokkan wilayah yaitu analisis *cluster*. Pada umumnya, analisis *cluster* digunakan untuk mengelompokkan objek dengan karakteristik yang mirip ke dalam satu *cluster* dan objek dengan karakteristik yang berbeda antar *cluster* dengan meminimalkan perbedaan antar objek dalam satu *cluster* serta memaksimalkan perbedaan objek antar *cluster* lainnya (Pramana dkk., 2018).

Analisis *cluster* terdiri dari berbagai algoritma, dua diantaranya yaitu algoritma partisi dan algoritma kepadatan. Algoritma partisi yang umum digunakan yaitu algoritma *K-Means* yang mengelompokkan data dalam *k cluster*

pada *dataset* berdasarkan jarak terdekat dari *centroid* yang dipilih berdasarkan rata-rata anggota *cluster*. Metode ini dapat digunakan dengan menentukan banyak kelompok yang diinginkan, tetapi rentan terhadap *noise* (Priati & Fauzi, 2017). Sedangkan, algoritma kepadatan yang populer yaitu *density-based spatial clustering of application with noise* yang mengelompokkan data berdasarkan kepadatan jarak, sehingga dapat mengidentifikasi *noise,* tetapi tidak dapat menentukan banyak kelompok yang diinginkan (Jing dkk., 2019). Menurut Wahyuni dkk (2021) metode kepadatan dapat dilakukan dengan proses hirarki dengan membentuk dendogram yang menggelompokan objek berdasarkan jarak.

Analisis *cluster* pada umumnya memiliki asumsi yaitu data yang akan dianalisis tidak memiliki korelasi yang kuat antar dua atau lebih variabel dengan variabel lainnya. Hal tersebut dikarenakan korelasi yang kuat akan menyebabkan multikolinearitas pada data (Hair dkk., 2010). Multikolinearitas menyebabkan informasi dari data saling tumpah tindih sehingga dapat menyebabkan informasi menjadi kurang akurat (Pramana dkk., 2018). Untuk mengatasi multikolinearitas, perlu dilakukan *principal component analysis* (PCA) untuk mereduksi dimensi data menjadi variabel komponen utama yang saling bebas (Johnson & Wichern, 2007). Dengan demikian, untuk mengeksplor kesamaan dan pola tersembunyi dalam analisis *cluster* maka diperlukan kombinasi PCA agar hasil pengelompokkan lebih optimal (Granato dkk., 2018).

Hasil analisis *cluster* dapat diukur menggunakan *Silhouette Coefficient*. Menurut Dewi & Pramita (2019), hasil analisis *cluster* yang baik yaitu memiliki ketidaksamaan yang tinggi antara *cluster* serta memiliki kesamaan yang tinggi sesama anggota *cluster*. Metode *Silhouette Coefficient* pada dasarnya adalah perbandingan rata-rata jarak karakteristik antar *cluster* dengan sesama *cluster*. *Silhouette Coefficient* yang mendekati nilai 1 menunjukkan hasil *cluster* yang semakin baik. Pada penelitian (Hidayat., 2023) dan (Dewi & Pramita., 2019), *Silhouette Coefficient* dapat menentukan jumlah *cluster* terbaik sehingga dapat membuat hasil *cluster* semakin optimal.

Penelitian sebelumnya mengenai proses *cluster* telah dilakukan oleh Widodo dkk. (2021) dengan melakukan perbandingan *euclidean metric* dan *manhattan metric* menghasilkan nilai akurasi secara berurut sebesar 79,30% dan

67,83%. Selain itu, Rais dkk (2021) telah mengoptimalisasi *K-Means cluster* dengan PCA. Penelitian menghasilkan dua komponen utama dan menghasilkan dua *cluster*. Sedangkan, penelitian mengenai *density-based spatial clustering of application with noise* dilakukan untuk mengelompokkan data Indikator Pembangunan Manusia dengan penentuan *epsilon* menggunakan *K-Nearest Neighbour*. Penelitian menggunakan data Sulawesi Selatan dengan 4 variabel. Penelitian tersebut menghasilkan *noise* sebanyak 5 kabupaten dan kota serta menghasilkan 1 *cluster* yang terdiri dari 19 kabupaten dan kota (Akbar, 2021).

Berdasarkan uraian, maka penelitian ini akan menggunakan *Silhouette Coefficient* dalam menentukan jumlah *cluster* optimal (banyaknya k) pada metode *K-Means cluster* serta proses perbandingan *cluster* menggunakan *Silhouette Coefficient* terbesar. Dengan demikian, peneliti ini mengangkat judul "**Perbandingan *K-Means Cluster* dengan *Density-Based Spatial Clustering of Application with Noise* pada Data Indikator Kemiskinan di Provinsi Sulawesi Selatan**". Diharapkan, pengelompokkan wilayah berdasarkan indikator kemiskinan dapat membantu pemerintah dalam menanggulangi kasus kemiskinan setiap wilayah di Provinsi Sulawesi Selatan.

1.2 Rumusan Masalah

Rumusan masalah yang dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana hasil pengelompokkan menggunakan metode *K-Means cluster* dengan *density-based spatial clustering of application with noise* berdasarkan data indikator kemiskinan di Provinsi Sulawesi Selatan?
2. Bagaimana perbandingan metode *K-Means cluster* dengan *density-based spatial clustering of application with noise* berdasarkan *Silhouette Coefficient* pada data indikator kemiskinan di Provinsi Sulawesi Selatan?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan terdiri dari delapan variabel mengenai indikator kemiskinan setiap wilayah di Provinsi Sulawesi Selatan pada tahun 2022.
2. Metode penentuan jarak (*distance*) yang digunakan yaitu jarak *euclidean* serta metode evaluasi yang digunakan menggunakan *Silhouette Coefficient*.

3. Metode penanganan multikolinearitas menggunakan PCA dengan proporsi kumulatif minimal 80%.
4. Metode *density-based spatial clustering of application with noise* yang digunakan hanya berdasarkan nilai *minimum points* sedangkan penentuan *epsilon* dilakukan secara hirarki.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh hasil pengelompokkan menggunakan *K-Means cluster* dengan *density-based spatial clustering of application with noise* berdasarkan data indikator kemiskinan di Provinsi Sulawesi Selatan.
2. Memperoleh hasil perbandingan metode *K-Means cluster* dengan *density-based spatial clustering of aplication with noise* berdasarkan *Silhouette Coefficient* pada data indikator kemiskinan di Provinsi Sulawesi Selatan.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Sebagai sumber informasi bagi peneliti selanjutnya mengenai hasil pengelompokkan menggunakan metode *K-Means cluster* dengan metode *density-based spatial clustering of application with noise* berbasis hirarki dalam mengelompokkan wilayah berdasarkan data indikator kemiskinan di Provinsi Sulawesi Selatan serta sebagai sumber informasi peneliti selanjutnya untuk mengatasi multikolinearitas pada data dengan menggunakan *principal component analysis*.
2. Sebagai sumber informasi bagi pemerintah terhadap pengelompokan wilayah berdasarkan data indikator kemiskinan sehingga dapat menjadi acuan dalam pelaksanaan kebijakan dalam menanggulangi kemiskinan di Provinsi Sulawesi Selatan.

BAB II
TINJAUAN PUSTAKA

2.1 Analisis *Cluster*

Analisis *cluster* atau *clustering analysis* merupakan usaha untuk mengidentifikasi kelompok objek yang mirip dan membantu menemukan pola penyebaran serta pola hubungan dalam sekumpulan data yang besar. Hal penting dalam proses pengelompokan adalah menyatakan sekumpulan pola ke kelompok yang sesuai untuk menemukan kesamaan dan perbedaan sehingga dapat menghasilkan kesimpulan yang berharga (Merliana dkk., 2015). *Clustering* ini mengelompokkan sejumlah besar data menjadi beberapa kelas berdasarkan kemiripan pada objek data ke dalam satu *cluster* dan sebaliknya meminimalkan kemiripan terhadap *cluster* yang lain. Pada umumnya, teknik *cluster* mempunyai dua cara dalam proses pengelompokkan yaitu *hierarchical clustering* dan *non-hierarchical clustering* (Tan dkk., 2006). Selain itu, metode analisis *cluster* membutuhkan suatu ukuran ketidakmiripan atau jarak yang didefinisikan untuk setiap pasang objek yang akan dikelompokkan.

2.1.1 Jarak *Euclidean*

Metode analisis *cluster* membutuhkan suatu ukuran ketidakmiripan atau jarak yang didefinisikan untuk setiap pasang objek yang akan dikelompokkan. Objek dikatakan memiliki karakteristik yang sama jika jarak antara objek tersebut kecil. Pengukuran jarak yang digunakan dalam penelitian ini adalah jarak *euclidean* sesuai Persamaan (2.1) berikut (Pangkey dkk., 2018):

$$d_{p,q} = \sqrt{\sum_{i=1}^{m} \left(x_{ip} - x_{iq}\right)^2} \tag{2.1}$$

dengan:

$d_{p,q}$: jarak antara objek p dengan objek q

x_{ip} : nilai variabel ke-i pada objek p

x_{iq} : nilai variabel ke-i pada objek q

m : banyaknya variabel yang diamati

2.1.2 Multikolinearitas

Analisis *cluster* menurut Ilyas (2018) memiliki asumsi bahwa tidak terjadi multikolinieritas. Suatu peristiwa yang terjadi apabila terdapat korelasi yang kuat antara dua atau lebih variabel disebut masalah multikolinearitas. Gujarati (2009) menyatakan bahwa salah satu cara untuk mengidentifikasi masalah multikolinieritas adalah dengan menghitung nilai *Variance Inflation Factor* (VIF) yang dirumuskan berdasarkan Persamaan (2.2) berikut:

$$VIF_i = \frac{1}{Tolerance} = \frac{1}{1 - R_i^2} \tag{2.2}$$

dengan:

$$R_i^2 = \frac{JK_R}{JK_T} = \frac{\sum_{i=1}^{n}\left(\widehat{X}_i - \bar{X}\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{2.3}$$

R_i^2 adalah koefisien determinasi yang diperoleh dari meregresikan variabel ke-i dengan variabel lainnya. Nilai VIF yang lebih besar dari 10 menunjukan adanya masalah multikolinieritas (Gujarati, 2009). Multikolinearitas dapat menyebabkan informasi dari sebuah data saling tumpah tindih sehingga dapat menyebabkan informasi kurang akurat (Pramana dkk., 2018). Dengan demikian, diperlukan metode statistika untuk mereduksi variabel untuk mengatasi multikolinearitas. Metode statistika yang umum digunakan untuk mereduksi variabel pada data multivariat yaitu menggunakan *Principal Component Analysis* (PCA) atau disebut analisis komponen utama (Fortuna & Maturo, 2019).

2.2 *Principal Component Analysis*

Principal Component Analysis (PCA) merupakan salah satu teknik dalam analisis multivariat untuk menganalisis data (Abdi & Williams, 2010). Tujuan PCA adalah untuk mengekstrak informasi penting dari data dan mengaturnya ke variabel baru yang disebut *principal component* (komponen utama). Selain itu, PCA juga berguna untuk mengidentifikasi data yang telah tersedia (Kurita, 2019). Menurut Jolliffe (2002), PCA digunakan untuk meringkas data dengan jumlah variabel yang lebih kecil. PCA menjadi metode pilihan tidak hanya untuk eksplorasi data tapi juga untuk *preprocessing* data sebelum menerapkan metode statistika atau pembelajaran mesin yang kompleks (Kherif & Latypova, 2019).

Jika $Z = (Z_1, Z_2, \dots, Z_m)$ adalah nilai $X = (X_1, X_2, \dots, X_m)$ yang telah terstandarisasi menjadi variabel baru pada $KU = (KU_1, KU_2, \dots, KU_g)$ (Abdi & Williams, 2010). Dengan nilai Z dirumuskan pada Persamaan (2.4) berikut:

$$Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_i^2}} \tag{2.4}$$

dengan:

X_i : nilai variabel i pada objek, $i = 1, 2, \dots, m$

μ_i : nilai rata-rata variabel i

σ_i^2 : nilai varian variabel i

Z_i : nilai standar i, $|Z_i| \geq 3.00$ dikategorikan sebagai *outlier* (Ferdinand, 2002).

Maka diperoleh Persamaan (2.5) sebagai berikut:

$$KU_f = \boldsymbol{b_{1g}} Z_1 + \boldsymbol{b_{2g}} Z_2 + \cdots + \boldsymbol{b_{mf}} Z_m \tag{2.5}$$

dengan demikian, persamaan PCA dapat ditulis menjadi Persamaan (2.6) berikut:

$$\begin{aligned} KU_1 &= \boldsymbol{b_{11}} Z_1 + \boldsymbol{b_{21}} Z_2 + \cdots + \boldsymbol{b_{m1}} Z_m \\ KU_2 &= \boldsymbol{b_{12}} Z_1 + \boldsymbol{b_{22}} Z_2 + \cdots + \boldsymbol{b_{m2}} Z_m \\ &\vdots \\ KU_g &= \boldsymbol{b_{1g}} Z_1 + \boldsymbol{b_{2g}} Z_2 + \cdots + \boldsymbol{b_{mg}} Z_m \end{aligned} \tag{2.6}$$

dengan:

p : banyaknya variabel

KU_f : komponen utama ke-f, $f = 1, 2, \dots, g$

$\boldsymbol{b_{mg}}$: vektor eigen variabel m pada komponen utama g (Abdi & Williams, 2010).

2.2.1 Matriks Korelasi dan Matriks Varian-Kovarian

Matriks korelasi yaitu matriks dengan elemen diagonal utama bernilai 1 yang disebut matriks simetris. Korelasi (r) bertujuan untuk mengukur keeratan linier antara dua variabel. Kumpulan korelasi dapat membentuk sebuah matriks yang disebut matriks korelasi. Dengan demikian, matriks korelasi dapat ditulis berdasarkan Persamaan (2.7) berikut (Astutik dkk., 2018):

$$\boldsymbol{R_{m \times m}} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix} \tag{2.7}$$

dengan:

r_{ij} : nilai korelasi antara X_i dan X_j

$\boldsymbol{R_{m\times m}}$: matriks korelasi berukuran $m \times m$.

Matriks korelasi digunakan dalam menentukan komponen utama bila terdapat variabel satuan ukur yang berbeda. Sedangkan, matriks varian-kovarian digunakan jika semua variabel memiliki satuan ukur yang sama atau telah terstandarisasi. Dengan demikian, pengukuran sebaran data dapat digunakan menggunakan rumus varian dan kovarian untuk mengetahui seberapa erat hubungan antara satu variabel dengan variabel lainnya. Nilai varian-kovarian dapat diperoleh dengan mengikuti Persamaan (2.8) berikut (Astutik dkk., 2018):

$$S_{ij} = \frac{\sum_{i,j=1}^{m}\left(Z_{ij} - \overline{Z}_i\right)\left(Z_{ij} - \overline{Z}_j\right)}{m-1} \tag{2.8}$$

Sehingga, kumpulan nilai varian dan kovarian akan membentuk matriks varian-kovarian baru yang dapat ditulis berdasarkan Persamaan (2.9) berikut:

$$\boldsymbol{S_{m\times m}} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1m} \\ S_{21} & S_{22} & \dots & S_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & \dots & S_{mm} \end{bmatrix} \tag{2.9}$$

dengan:

S_{ij} : nilai varian-kovarian

$\boldsymbol{S_{m\times m}}$: matriks varian-kovarian berukuran $m \times m$ (Astutik dkk., 2018).

2.2.2 Koefisien Pembobot dan Penentuan Komponen Utama

Nilai eigen dan koefisien pembobot komponen utama atau dikenal vektor eigen dapat diselesaikan menggunakan Persamaan (2.10) (Astutik dkk., 2018):

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{b_i} = 0 \tag{2.10}$$

Hasil $\boldsymbol{b_i}$ yang diharapkan pada Persamaan (2.10) adalah non-trivial sehingga diperoleh solusi berdasarkan Persamaan (2.11) berikut:

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0 \tag{2.11}$$

Sehingga akan menghasilkan akar karakteristik λ_l yaitu $\lambda_1 > \lambda_2 > \cdots > \lambda_v > 0$ sehingga setiap karakteristik λ_l bergantung nilai b_i (Astutik dkk., 2018).

dengan:

$\boldsymbol{b_i}$: vektor koefisien pembobot (vektor eigen) pada variabel ke-i

$\boldsymbol{I}$: Matriks identitas

λ : nilai eigen

λ_l : nilai eigen ke-l dengan $l = 1, 2, \dots, v$.

Matriks $\boldsymbol{A}$ adalah matriks korelasi atau matriks varian-kovarian yang bersifat definit positif (Astutik dkk., 2018). Selain itu, untuk menentukan komponen utama dapat mengikuti kriteria dengan mempertimbangkan proporsi kumulatif dari varian total yang dapat ditentukan dengan memilih k komponen utama sebagai kontribusi terbesar dari varian data serta mempunyai proporsi kumulatif lebih dan mendekati nilai 80% dengan nilai perhitungan berdasarkan Persamaan (2.12) berikut (Abdi & Williams, 2010):

$$\frac{\sum_{l=1}^{u} \lambda_l}{\sum_{l=1}^{v} \lambda_l} > 0.80 \; ; u \leq v \tag{2.12}$$

dengan:

$\sum_{l=1}^{u} \lambda_l$: total varian komponen utama hingga ke-u, $l = 1, 2, \dots, u$

$\sum_{l=1}^{v} \lambda_l$: varian total, $l = 1, 2, \dots, v$

2.3 *K-Means Cluster*

Analisis *cluster non-hierarchical* merupakan teknik gerombol yang dilakukan dengan mempartisi data ke dalam sejumlah *cluster* yang telah ditetapkan. Salah satu analisis *cluster non-hierarchical* yaitu *K-Means* merupakan salah satu metode yang paling umum digunakan (Fani dkk., 2021). Konsep dasar *K-Means* yaitu pencarian pusat *cluster* secara *iteratif*. Pusat *cluster* ditetapkan berdasarkan jarak setiap data ke pusat *cluster*. Proses *clustering* dimulai dengan mengidentifikasi data yang akan dikelompokkan sebanyak *k cluster*. Dengan demikian, *K-Means* adalah metode pengelompokan dengan membentuk *k cluster* dengan *centroid* merupakan rata-rata anggota *cluster* (Nugraha dkk., 2018). Secara sederhana, algoritma *K-Means* dapat dilakukan dengan mengikuti langkah-langkah sebagai berikut (Merliana dkk., 2015):

1. Tentukan k sebagai jumlah *cluster* yang ingin dibentuk.
2. Tentukan *k centroid* (titik pusat *cluster*) awal secara acak.
3. Hitung jarak setiap data ke setiap *centroid* menggunakan Persamaan (2.1).
4. Setiap data memilih *centroid* yang terdekat.

5. Tentukan posisi *centroid* baru dengan cara menghitung nilai rata-rata dari data-data yang memilih pada *centroid* yang sama.
6. Kembali ke langkah 3 jika posisi *centroid* baru dan *centroid* lama tak sama.
7. Proses pengelompokkan dinyatakan selesai jika posisi *centroid* baru dan *centroid* lama telah sama.

2.4 *Density-Based Spatial Clustering of Application with Noise*

Algoritma *Density-Based Spatial Clustering of Application with Noise* (DBSCAN) adalah metode pengelompokan data spasial berdasarkan kepadatan hal ini dapat ditemukan berdasarkan satu kondisi kerapatan. Dalam DBSCAN dikenal pula istilah *noise* yang memiliki jarak yang besar (Hahsler dkk., 2019). Adapun istilah dasar lainnya dalam DBSCAN yaitu (Jing dkk., 2019)

a) *Epsilon* yaitu radius DBSCAN dalam menentukan ketetanggan. Digunakan untuk menentukan *core points*, *border points*, dan *noise*.
b) *Minimum points* yaitu kumpulan poin minimum. Sebagai *core points*, jumlah titik di lingkungannya harus melebihi atau sama dengan *minimum points*.
c) *Core points* yaitu poin yang jumlahnya memenuhi syarat *minimum points*.
d) *Direct arrived density* yaitu terdapat titik p dan q. Jika titik q berada di lingkungan *eps* dari titik p dan titik p adalah *core points*, maka akan disebut *direct arrived density* antara titik p dan titik q.
e) *Arrived density* yaitu jika serangkaian titik P_i $(i = 1, 2, 3, 4, \dots, n)$ di suatu daerah dalam satu *cluster*. Dengan, P_1 dan P_2 memiliki *core points* yang berbeda maka P_1 dan P_2 akan disebut *arrived density*.

Adapun algoritma DBSCAN yaitu (Nurhaliza & Mustakim, 2021):

1. Menginisiasi parameter *minimum points*.
2. Menentukan p sebagai titik awal secara acak.
3. Menghitung semua jarak titik yang *density reachable* terhadap p menggunakan Persamaan (2.1).
4. Jika titik memenuhi syarat *core point* yaitu jumlah titiknya melebihi atau sama dengan *minimum points* maka akan terbentuk sebuah *cluster*.
5. Mengulangi langkah 2–4 untuk setiap titik yang akan diamati. Jika p merupakan titik border dan tidak ada titik yang *density reachable* terhadap p, maka proses dilanjutkan ke titik yang lain.

Secara sederhana, algoritma ini memerlukan dua parameter masukan yaitu radius *epsilon* dan *minimum points* (Starczewski & Cader, 2019). Dalam menentukan parameter *minimum points*, menurut Hahsler dkk. (2019), setidaknya adalah jumlah variabel dari sebuah *dataset* yang dianalisis ditambah satu. Sedangkan, baru-baru ini proses DBSCAN telah berkembang dengan membuat representasi tanpa memerlukan radius *epsilon* dengan menggabungkan proses hirarki *cluster* yang disebut sebagai metode *Hierarchical Density-Based Spatial Clustering of Application with Noise* (Stewart & Al-Khassaweneh, 2022).

2.5 *Silhouette Coefficient*

Silhouette Coefficient (SC) merupakan salah satu metode yang digunakan untuk menentukan jumlah *cluster* yang optimal (Widyadhana dkk., 2021). Selain itu, SC digunakan untuk mengevaluasi hasil *cluster* berdasarkan Persamaan (2.13) hingga Persamaan (2.16) berikut (Batool & Hening, 2021):

$$SC = \frac{1}{n}\sum_{p=1}^{n} s(p) \tag{2.13}$$

$$s(p) = \frac{b(p) - a(p)}{\max\big(a(p); b(p)\big)} \tag{2.14}$$

$$a(p) = \frac{1}{n_C - 1}\sum_{q \in C}^{n_C} d(p,q) \tag{2.15}$$

$$b(p) = \min_{C \neq D}\left(\frac{1}{n_D}\sum_{q \in D}^{n_D} d(p,q)\right) \tag{2.16}$$

dengan:

$d(p,q)$: jarak karakteristik objek p dengan objek q

$a(p)$: rata-rata dari jarak karakterstik objek p dengan semua objek yang berada dalam satu *cluster* yang sama

n_C : banyaknya data yang berada dalam satu *cluster* yang sama

$b(p)$: rata-rata dari jarak karakterstik objek p dengan semua objek yang berada dalam *cluster* yang berbeda

n_D : banyaknya data yang berada dalam satu *cluster* yang berbeda

$s(p)$: nilai *silhouette* pada objek p.

SC memiliki selang interval yaitu: $-1 \leq SC \leq 1$ yang menunjukkan bahwa nilai yang mendekati nilai 1 menunjukkan hasil yang baik. Kriteria evaluasi metode SC ditunjukkan pada Tabel 2.1 berikut (Batool & Hening, 2021):

Tabel 2.1 Kriteria Evaluasi SC

Nilai SC	Kriteria
$0{,}70 < SC \leq 1{,}00$	Struktur kuat
$0{,}50 < SC \leq 0{,}70$	Struktur baik
$0{,}25 < SC \leq 0{,}50$	Struktur lemah
$SC \leq 0{,}25$	Struktur buruk

2.6 Kemiskinan

Kemiskinan merupakan ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari beberapa aspek. Jadi, penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan (BPS, 2023). Menurut Abidin (2017) menyatakan bahwa indikator kemiskinan dipengaruhi oleh terbatas-nya jumlah lapangan kerja yang dapat ditinjau dengan melihat jumlah penduduk, jumlah angkatan kerja dan kepadatan penduduk. Hal itu membuat banyak penduduk tidak mempunyai pekerjaan (pengangguran). Oleh karena itu, pengangguran menjadi faktor dalam mengukur kemiskinan disuatu daerah. Mengacu pada kajian maka indikator kemiskinan diantaranya (BPS, 2023):

1. Indeks pembangunan manusia (IPM) adalah alat ukur pemerintah yang didasari 3 dimensi dasar yaitu umur panjang dan hidup sehat, pengetahuan dan standar hidup layak.
2. Jumlah penduduk adalah total semua orang yang berdomisili di wilayah geografis Republik Indonesia selama 6 bulan atau lebih dan atau mereka yang berdomisili kurang dari 6 bulan tetapi bertujuan untuk menetap.
3. Jumlah angkatan kerja adalah jumlah penduduk usia kerja (lima belas tahun ke atas) yang bekerja, atau punya pekerjaan namun sementara tidak bekerja dan pengangguran.
4. Persentase penduduk miskin adalah jumlah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan terhadap jumlah

penduduk.

5. Partisipasi angkatan kerja adalah persentase banyaknya angkatan kerja terhadap banyaknya penduduk yang berumur lima belas tahun keatas.
6. Jumlah pengangguran adalah total angkatan kerja yang tidak bekerja atau yang sedang mencaru pekerjaan.
7. Kepadatan penduduk adalah perbandingan total penduduk disuatu daerah terhadap luas daerah.
8. Pengeluaran per Kapita adalah biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan dibagi dengan banyaknya anggota rumah tangga tersebut.