PERBANDINGAN KINERJA *K-MEDOIDS* DAN *DENSITY BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE* MENGGUNAKAN UJI *SILHOUETTE COEFFICIENT*

SKRIPSI

TAUFIQ AKBAR

H051181307

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2023

PERBANDINGAN KINERJA *K-MEDOIDS* DAN *DENSITY BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE* MENGGUNAKAN UJI *SILHOUETTE COEFFICIENT*

SKRIPSI

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

TAUFIQ AKBAR

H051181307

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2023

LEMBAR PERNYATAAN KEOTENTIKAN

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

Perbandingan Kinerja *K-Medoids* dan *Density Based Spatial Clustering of Application with Noise* Menggunakan Uji *Silhouette Coefficient*

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

Makassar, 08 Maret 2023

Taufiq Akbar
NIM H051181307

PERBANDINGAN KINERJA *K-MEDOIDS* DAN *DENSITY BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE* MENGGUNAKAN UJI *SILHOUETTE COEFFICIENT*

Disetujui Oleh:

Pembimbing Utama	Pembimbing Pertama
Dr. Dr. Georgina Maria Tinungki, M.Si.	**Siswanto, S.Si., M.si.**
NIP. 196209261987022001	**NIP. 199201072019031012**

Ketua Program Studi

Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 197201171997032002

Pada 08 Maret 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Taufiq Akbar

NIM : H051181307

Program Studi : Statistika

Judul Skripsi : Perbandingan Kinerja *K-Medoids* dan *Density Based Spatial Clustering of Application with Noise* Menggunakan Uji *Silhouette Coefficient*

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 : Dr. Dr. Georgina Maria Tinungki, M.Si.. (.....)
2. Sekretaris : Siswanto, S.Si., M.Si. (.....)
3. Anggota : Dr. Nurtiti Sunusi, S.Si., M.Si. (.....)
4. Anggota : Sitti Sahriman, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 08 Maret 2023

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji hanya milik Allah Subhanallahu Wa Ta'ala atas segala limpahan rahmat dan hidayah-Nya yang telah diberikan kepada penulis sampai saat ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah Shallallahu 'Alaihi Wa sallam. Alhamdulillaahirobbil'aalamiin, berkat rahmat dan kemudahan yang diberikan oleh Allah Subhanallahu Wa Ta'ala, penulis dapat menyelesaikan skripsi yang berjudul "**Perbandingan Kinerja *K-Medoids* dan *Density Based Spatial Clustering of Application with Noise* Menggunakan Uji *Silhouette Coefficient***" sebagai salah satu syarat memperoleh gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis tidak akan sampai pada titik ini, jikalau tanpa dukungan dan bantuan dari pihak yang selalu ada, peduli dan menyayangi penulis. Oleh karena itu, penulis haturkan rasa terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya untuk orang tua penulis yang telah memberikan dukungan penuh, pengorbanan, kesabaran hati, cinta dan kasih sayang, serta dengan ikhlas telah mengiri setiap langkah penulis dengan doa dan restunya.

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

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.,** selaku Rektor Universitas Hasanuddin berserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin,** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi S.Si., M.Si.,** selaku Ketua Departemen Statistika Universitas Hasanuddin, segenap Dosen Pengajar dan Staf yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Ibu Dr. Dr. Georgina Maria Tinungki, M.Si.,** selaku Pembimbing Utama dan **Bapak Siswanto S.Si., M.Si.,** selaku Pembimbing Pertama yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya di

tengah berbagai kesibukan dan prioritasnya untuk senantiasa memberikan arahan, dorongan, dan motivasi kepada penulis mulai dari awal hingga selesainya penulisan tugas akhir ini.

5. **Ibu Sitti Sahriman, S.Si., M.Si.,** dan **Ibu Dr. Nurtiti Sunusi, S.Si., M.Si.,** selaku Tim Penguji yang telah memberikan banyak saran dan kritik yang membangun dalam penyempurnaan penyusunan tugas akhir ini serta waktu yang telah diberikan kepada penulis.
6. **Ibu Dr. Dr. Georgina Maria Tinungki, M.Si.,** selaku Penasehat Akademik penulis. Terima kasih atas segala bantuan, nasehat serta motivasi yang senantiasa dan selalu diberikan kepada Penulis selama menjalani pendidikan di Departemen Statistika.
7. Teman-teman mahasiswa departemen **Statistika angkatan 2018**, terima kasih sudah menemani, berjuang bersama di dunia perkuliahan.
8. Teman-teman "**LT.5"** yang tidak memberikan kontribusi apapun kepada penulis, kecuali hanya memberikan semangat yang tidak signifikan kepada penulis :v.
9. Printer Epson E1150 yang senantiasa menemani penulis di ruang asisten.
10. Penghuni tetap ruang asisten yakni **Adhyaksa Prananda, Haksar, Keping, Hapis, Noor-man-to, Ikhchwan, Ika dan Real,** semoga loker menyertai saudara-saudara sekalian**.**
11. Ucapan terimakasih yang belum sempat tersampaikan kepada **Nuki, Juni, Yustika, Hajrah, Nuge, Naura, Ica, Miskha dan Pitto damayanti** yang telah menaikkan kadar gula darah penulis dengan memberikan 2 dos brownis brocyl kepada penulis.
12. Teman mabar dan gabut dikala covid-19 menyerang **Acca, Udin, Safitri Poet** dan **Violet.**
13. Kepada semua orang-orang baik yang telah membantu penulis, baik itu dukungan materil maupun emosional, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi Allah Subhanallahu Wa Ta'ala.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Makassar, 08 Maret 2023

Taufiq Akbar

NIM H051181307

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

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

Nama	: Taufiq Akbar
NIM	: H051181307
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-ekslusif (Non-exclusive Royalty- Free Right**) atas tugas akhir saya yang berjudul:

"**Perbandingan Kinerja *K-Medoids* dan *Density Based Spatial Clustering of Application with Noise* Menggunakan Uji *Silhouette Coefficient***"

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 pada tanggal, 08 Maret 2023

Yang menyatakan

(Taufiq Akbar)

ABSTRAK

Analisis Klaster adalah analisis yang bertujuan mengelompokkan objek-objek pada data berdasarkan karakteristiknya. Pengolompokan hasil dari analisis klaster dapat dikatakan baik, jika masing-masing klaster bersifat homogen, dalam hal ini dapat divalidasi menggunakan uji *silhouette coefficient*. Namun, adanya pencilan pada data dapat memengaruhi hasil pengelompokan yang tidak valid. Untuk menghidari hal tersebut, maka digunakan metode yang *robust* terhadap pencilan, seperti *K-Medoids* dan *Density Based Spatial Clustering of Application with Noise*. *K-Medoids* adalah metode yang menggunakan objek representatif (*medoids*) sebagai pusat dari klaster, sedangkan *Density Based Spatial Clustering of Application with Noise* adalah metode yang menggunakan prinsip kepadatan dalam mengelompokkan objek-objek pada data. Tujuan penelitian ini, yaitu membandingkan hasil dan kinerja dari kedua metode menggunakan uji *silhouette coefficient*. Data pada penelitian ini adalah data indikator pembangunan manusia di Provinsi Sulawesi Selatan Tahun 2021. Dari hasil analisis, metode *K-Medoids* menghasilkan 2 kelompok, yaitu kelompok kabupaten/kota yang memiliki indikator pembangunan manusia yang sedang terdiri dari 21 kabupaten/kota dan kelompok yang tinggi terdiri dari 3 kabupaten/kota, sedangkan *Density Based Spatial Clustering of Application with Noise* menghasilkan 1 kelompok yang memiliki karakteristik yang sama, yakni terdiri dari 19 kabupaten/kota dan 5 kabupaten/kota sisanya teridentifikasi sebagai *noise*. Berdasarkan uji *silhouette coefficient*, *K-Medoids* memiliki nilai *Silhouette Coefficient* (SC) yang lebih besar dibandingkan *Density Based Spatial Clustering of Application with Noise*, yakni masing-masing sebesar 0,635 dan 0,544 sehingga dalam hal ini metode *K-Medoids* memiliki kinerja yang lebih baik.

Kata Kunci : Analisis Klaster, DBSCAN, *K-Medoids* , *Silhouette Coefficient.*

ABSTRACT

Cluster analysis is an analysis that aims to group objects in the data based on their characteristics. The grouping of results from cluster analysis can be said to be good if each cluster is homogeneous; in this case, it can be validated using the silhouette coefficient test. However, the presence of outliers in the data can affect invalid grouping results. To avoid this, robust methods against outliers are used, such as K-Medoids and Density Based Spatial Clustering of Application with Noise. K-Medoids is a method that uses representative objects (medoids) as the center of the cluster, while Density Based Spatial Clustering of Application with Noise is a method that uses the principle of density in grouping objects in the data. The purpose of this study is to compare the results and performance of the two methods using the silhouette coefficient test. The data in this study is data on human development indicators in South Sulawesi Province in 2021. From the results of the analysis, the K-Medoids method produces 2 groups, namely the districts/cities group that has moderate human development indicators, consisting of 21 districts/cities, and the group with high indicators, consisting of 3 districts/cities, while Density Based Spatial Clustering of Application with Noise produces 1 group that has the same characteristics, consisting of 19 districts/cities, and the remaining 5 districts/cities are identified as noise. Based on the silhouette coefficient test, K-Medoids has a greater Silhouette Coefficient (SC) value than Density Based Spatial Clustering of Application with Noise, namely 0.635 and 0.544, respectively, so that in this case the K-Medoids method has better performance.

Keywords : *Cluster Analysis*, DBSCAN, *K-Medoids* , *Silhouette Coefficient.*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Indeks Pembangunan Manusia (IPM) adalah suatu indeks yang menjelaskan bagaimana penduduk di sebuah wilayah dapat mengakses hasil dari sebuah pembangunan untuk memperoleh pendidikan, kesehatan maupun pendapatan. Menurut Badan Pusat Statistik (BPS), terdapat 4 indikator dalam mengukur IPM, yaitu Umur Harapan Hidup (UHH), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS) dan Pengeluaran Per Kapita Disesuaikan (BPS, 2021).

Di Indonesia, menurut data yang dirilis oleh BPS pada 15 November 2021, Provinsi yang memiliki poin IPM tertinggi adalah Provinsi DKI Jakarta, sedangkan Provinsi yang memiliki IPM terendah adalah Papua yang artinya, secara kualitas, kehidupan masyarakat di DKI Jakarta lebih baik dibandingkan kualitas kehidupan masyarakat di Papua. Karena ukuran dari kualitas pembangunan manusia secara statistik diukur melalui IPM, maka peningkatan poin IPM di suatu wilayah menjadi hal yang terpenting untuk mengukur kualitas hidup masyarakat, tak terkecuali di Provinsi Sulawesi Selatan. Untuk itu, dalam membantu kinerja pemerintah setempat dalam meningkatkan poin IPM di Provinsi Sulawesi Selatan, terlebih dahulu mengidentifikasi kabupaten/kota yang memiliki kedekatan berdasarkan indikatornya melalui analisis klaster.

Analisis klaster mempunyai tujuan, yaitu mengelompokkan objek-objek yang memiliki kedekatan dan memisahkan objek-objek yang tidak memiliki kedekatan, dalam hal ini kedekatan tersebut dapat diukur melalui jarak antar objek. Suatu objek dapat dikatakan dekat, jika objek tersebut memiliki jarak terkecil diantara objek lainnya (Everitt dkk., 2011). Dalam mengelompokkan objek-objek tersebut, terdapat dua metode, yaitu metode hierarki dan non-hierarki. Menurut Santoso (2010), metode hierarki adalah metode yang mengelompokkan objek yang memiliki kedekatan, kemudian dilanjutkan pada kedekatan objek lainnya sampai membentuk klaster tingkatan (hierarki) antar objek. Sedangkan dalam metode non-hierarki, banyak *k* klaster yang akan terbentuk ditentukan sebelum proses klaster.

Density Based Spatial Klaster*ing of Application with Noise* (DBSCAN) adalah pendekatan yang hierarki dalam menentuan *k* buah klaster. Metode DBSCAN di desain oleh Ester dkk. pada tahun 1996 untuk mengelompokkan objek-objek pada data spasial yang basisnya menggunakan kepadatan dari data dengan dua parameter, yaitu *Epsilon* dan *MinPoints*. Kedua parameter tersebut menjadi sebuah acuan dari objek yang termasuk dalam sebuah kepadatan atau objek yang termasuk dalam *noise* atau pencilan yang menjadikan metode DBSCAN *robust* terhadap pencilan (Id dkk., 2017).

K-Medoids juga merupakan metode yang *robust* terhadap pencilan, akan tetapi algoritma *K-Medoids* termasuk ke dalam metode non-hierarki karna tidak mengikuti proses yang hierarki dalam menentukan *k* buah klaster. Metode *K-Medoids* mendasari terbentuknya sebuah klaster dari objek yang representatif yang menjadi solusi dari permasalahan adanya pencilan dalam metode *K-Means* (Suyanto, 2017).

Hasil dari pengelompokan yang baik, memiliki kesamaan yang tinggi antar anggota klaster dan ketidaksamaan yang tinggi antara klaster yang satu dan klaster yang lain yang dapat diukur menggunakan uji *Silhouette Coefficient*. Jika nilai dari *Silhoette Coefficient* semakin mendekati 1, maka semakin baik pengelompokan yang terbentuk dan jika nilai *Silhouette Coefficient* lebih kecil atau sama dengan 0.25, maka hasil pengelompokan dianggap buruk (Kaufman dan Rousseeauw, 2005).

Penelitian sebelumnya terkait DBSCAN oleh Putri dkk. (2020) yang membandingkan metode DBSCAN dan *K-Means* meggunakan nilai *Silhouette Coefficient*, didapatkan hasil bahwa DBSCAN lebih baik dalam mengelompokkan status desa di Jawa Tengah Tahun 2020. Perbandingan terhadap DBSCAN juga dilakukan oleh Valarmathy dan Krishnaveni (2019) dalam penelitianya yang berjudul "*Performance Evaluation and Comparison of Clustering Algorithms used in Educational Data Mining*". Penelitian tersebut membandingkan 8 metode pengelompokan, yaitu *Expectation Maximimization*, CLOPE, DBSCAN, *Filtered Cluster*, *Farthest First*, COWEB, *K-Means* dana CLARA. Hasil dari peneltian tersebut menunjukkan DBSCAN memiliki kinerja yang lebih baik dibandingkan 7 metode lainya. Penelitian terkait *K-Medoids* juga telah dikaji oleh Raja (2020)

dalam merumuskan masalah dalam penelitianya, yaitu membandingkan metode *Centroid Linkage* dan *K-Medoids* dalam mengelompokkan kabupaten/kota di Sulawesi Selatan berdasarkan Indikator pendidikan dan perbandingan tersebut menggunakan nilai simpangan baku yang memiliki rasio terkecil. Hasilnya didapatkan bahwa tidak ada perbedaan antara *Centroid Linkage* dan *K-Medoids* karena menghasilkan nilai rasio simpangan baku yang sama.

Berdasarkan beberapa penelitian yang telah dilakukan dan uraian latar belakang tersebut, peneliti tertarik mengangkat judul "**Perbandingan Kinerja *K-Medoids* dan *Density Based Spatial Clustering of Application with Noise* menggunakan Uji *Silhoette Coefficient***".

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka dirumuskan masalah penelitian yang ingin dilakukan, yaitu :

1. Bagaimana hasil pengelompokan kabupaten/kota berdasarkan indikator pembangunan manusia di Provinsi Sulawesi Selatan menggunakan metode *K-Medoids* dan DBSCAN ?
2. Bagaimana perbandingan kinerja metode *K-Medoids* dan DBSCAN dalam pengelompokan kabupaten/kota berdasarkan indikator pembangunan manusia di Provinsi Sulawesi Selatan dengan uji *Silhouette Coefficient* ?

1.3 Batasan Masalah

Masalah dalam penelitian ini dibatasi, yaitu menggunakan jarak *Euclidean*, mengatasi multikolinearitas menggunakan metode Analisis Komponen Utama (AKU) dengan kriteria proporsi komulatif keragaman Komponen Utama (KU) berada pada interval 80%-90%, serta mengukur kinerja hasil klaster menggunakan uji *Silhouette Coefficient*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dirumuskan diatas maka tujuan penelitian yang ingin dicapai, yaitu :

1. Memperoleh hasil pengelompokan kabupaten/kota berdasarkan indikator pembangunan manusia di Provinsi Sulawesi Selatan menggunakan metode *K-Medoids* dan DBSCAN.

2. Memperoleh perbandingan kinerja metode *K-Medoids* dan DBSCAN dalam pengelompokan kabupaten/kota berdasarkan indikator pembangunan manusia di Provinsi Sulawesi Selatan dengan uji *Silhouette Coefficient*.

1.5 Manfaat Penelitian

Hasil penelitian kali ini diharapkan dapat memberikan sebuah perspektif baru dalam memandang karakteristik indikator pembangunan manusia di Provinsi Sulawesi Selatan serta menambah wawasan memilih metode dalam analisis klaster yang *robust* terhadap pencilan.

BAB II
TINJAUAN PUSTAKA

2.1 Analisis Klaster

Analisis klaster adalah salah suatu analisis multivariat yang termasuk dalam metode interdepedensi, yaitu tidak terdapat satu pun variabel yang diartikan sebagai suatu variabel bebas atau variabel terikat (Bahroni dkk., 2016). Analisis klaster merupakan teknik analisis dalam menentukan sebuah kelompok dari beberapa individu atau objek berdasarkan kemiripan antara objek yang satu dengan objek yang lain. Secara umum, setidaknya terdapat 3 tahapan dalam analisis klaster, yaitu menghitung seberapa dekat objek yang satu dengan yang lain, kemudian dilanjutkan pada tahap proses pengelompokan dan terakhir mendeskripsikan setiap kelompok yang terbentuk (Hair dkk., 2010). Menurut Santoso (2010), ada 2 ciri-ciri yang terdapat pada klaster, yaitu:

1. Homogenitas (kesamaan) yang tinggi antar anggota dalam satu klaster.
2. Heterogenitas (perbedaan) yang tinggi antar klaster yang satu dengan klaster yang lain.

Terdapat dua metode dalam analisis klaster, yaitu metode hierarki dan metode non-hierarki. Metode hierarki adalah metode pengelompokan yang dimulai dari mengumpulkan kesamaan objek yang terdekat dari dua atau lebih objek dan dilanjutkan dengan objek lain yang memiliki kedekatan kedua dan seterusnya sehingga membentuk sebuah hierarki (tingkatan) antara objek yang mirip dan objek yang tidak mirip. Lain halnya dengan metode non-hierarki, yaitu metode pengelompokan yang dimulai dari penentuan jumlah klaster kemudian dilanjutkan dalam proses klaster yang tidak hierarki (Santoso, 2010).

2.1.1 Jarak Antar Objek

Mengetahui seberapa dekat atau sebarapa jauh suatu objek yang satu dengan objek yang lain, adalah tujuan utama dalam mengidentifikasi objek yang disajikan oleh data ke dalam klaster. Dalam menunjukkan kedekatan antar objek, dapat dilihat melalui ukuran kuantitatif yang sering digunakan yaitu ketidakmiripan (*dissimilarity*), kemiripan (*similarity*) atau biasa disebut sebagai kedekatan (*porximity*), dan jarak (*distance*). Objek dikatakan memiliki kedekatan, jika jarak

antara objek tersebut kecil dan kemiripan diantara objek tersebut besar (Everitt dkk., 2011).

Ukuran jarak yang digunakan untuk melihat kedekatan antar objek adalah menggunakan jarak *Euclidean* melalui Persamaan (2.1) (Kaufman dan Rausseeuw, 2005) sebagai berikut:

$$d(i,j) = \sqrt{\sum_{p=1}^{n} \left(x_{ip} - x_{jp}\right)^2} \tag{2.1}$$

keterangan :

n	:	banyak variabel
$d(i,j)$	:	jarak antar objek ke-*i* dan objek ke-*j*
x_{ip}	:	data dari subjek ke-*i* dan variabel ke-*p*
x_{jp}	:	data dari subjek ke-*j* dan variabel ke-*p*

2.1.2 Uji Multikolinearitas

Menurut Hair dkk. (2010), Multikolinearitas merupakan pelanggaran asumsi di dalam analisis klaster. Multikolinearitas dapat diartikan sebagai adanya hubungan atau korelasi antara variabel (Yamin, 2011). Salah satu teknik dalam mengetahui adanya multikolinearitas adalah menguji koefisien korelasi dari masing-masing variabel dengan hipotesis :

H_0: Tidak terdapat korelasi yang signifikan
H_1: Terdapat korelasi yang signifikan

Terima H_0, jika nilai dari $t_{hitung} \leq t_{tabel}$ dan Terima H_1, jika nilai dari $t_{hitung} > t_{tabel}$ dengan rumus t_{hitung} (Obilor dan Arnadi, 2018)

$$t_{hitung(i,j)} = \rho_{ij} \sqrt{\frac{n-2}{1-\rho_{ij}^2}} \tag{2.2}$$

degan ρ_{ij} adalah koefisien korelasi dari variabel ke-*i* dengan variabel ke-*j* yang dihitung melalui Persamaan (2.12).

Salah satu solusi yang dapat diterapkan jika pelanggaran asumsi multikolinearitas terjadi, yaitu menghilangkan korelasi antar variabel dengan menerapkan metode Analisis Komponen Utama (AKU) (Hair dkk., 2010).

2.2 Analisis Komponen Utama

Analisis Komponen Utama (AKU) adalah metode yang fokusnya menjelaskan struktur dari varians-kovarians dari sekumpulan variabel yang memiliki hubungan linear dengan tujuan untuk mereduksi jumlah variabel pada data. Pada dasarnya, metode AKU mereduksi jumlah variabel dengan mentransformasi variabel-variabel lama tersebut dengan menghilangkan korelasi antar variabel sehingga menghasilkan variabel-variabel baru yang tidak berkorelasi satu sama lain, yang disebut sebagai Komponen Utama (KU) (Johnson dan Wichern, 2007).

Misalkan terdapat sebuah dataset yang terdiri dari *n* objek dan *p* variabel dapat dinyatakan dalam sebuah matriks X, yaitu

$$\mathbf{X}_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{2.3}$$

Kombinasi linear dari KU yang terbentuk terhadap variabel asalnya dapat dituliskan sebagai berikut (Johnson dan Wichern, 2007):

$$\text{KU}_j = e_{1j}Z_1 + e_{2j}Z_2 + \cdots + e_{pj}Z_p, j = 1,2,\ldots,p \tag{2.4}$$

keterangan :

KU_j	:	Komponen Utama ke-*j*
Z_p	:	Variabel ke-*p* yang terstandarisasi
$e_{p,j}$	:	Vektor eigen dari variabel ke-*p* Komponen Utama ke-*j* yang ternormalisasi

2.2.1 Nilai Eigen dan Vektor Eigen

Misalkan matriks $\mathbf{A}$ berukuran $n \times n$ dan $\mathbf{A}\boldsymbol{x}$ merupakan kelipatan skalar dari $\boldsymbol{x}$, maka vektor tak nol dari $\boldsymbol{x}$ disebut sebagai vektor eigen dari $\mathbf{A}$ dengan persamaan

$$\mathbf{A}\boldsymbol{x} = \lambda\boldsymbol{x} \tag{2.5}$$

dengan $\boldsymbol{x}$ adalah vektor eigen dari $\mathbf{A}$ yang bersesuaian dengan λ yang adalah nilai eigen dari $\mathbf{A}$ diperoleh melalui Persamaan (2.5) yang dapat ditulis sebagai berikut :

$$(\mathbf{A} - \lambda\mathbf{I})\boldsymbol{x} = 0 \tag{2.6}$$

Persamaan (2.6) mempunyai penyelesaian tak nol (non *trivial*) jika memenuhi Persamaan (2.7).

$$|\mathbf{A} - \lambda\mathbf{I}| = 0 \tag{2.7}$$

Untuk setiap $\boldsymbol{x}$ dari **A**, $\boldsymbol{e} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ adalah vektor eigen dari **A** yang bersesuaian dengan λ yang sama (Johnson dan Wichern, 2007).

2.2.2 Matriks Varians-Kovarians dan Matriks Korelasi

Menurut Johnson dan Wichern (2007), matriks Varians-Kovarians dapat dibentuk dalam sebuah matriks yang dinotasikan sebagai **Σ** pada Persamaan (2.10)

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu} \tag{2.8}$$

$$cov(\mathbf{X}) = \boldsymbol{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \tag{2.9}$$

$$= E\left(\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} [X_1 - \mu_1 \quad X_2 - \mu_2 \quad \cdots \quad X_p - \mu_p] \right)$$

$$= E \begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) & \cdots & E(X_1 - \mu_1)(X_p - \mu_p) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 & \cdots & E(X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p - \mu_p)(X_1 - \mu_1) & E(X_p - \mu_p)(X_2 - \mu_2) & \cdots & E(X_p - \mu_p)^2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_{p \times p} = cov(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \tag{2.10}$$

karena $E(X_i - \mu_i)(X_j - u_j) = E(X_j - u_j)\ (X_i - \mu_i)$, maka $\sigma_{ij} = \sigma_{ji}$, sehingga Persamaan (2.10) dapat dituliskan sebagai Persamaan (2.11)

$$\boldsymbol{\Sigma}_{p \times p} = cov(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \tag{2.11}$$

Hubungan linier antara variabel *random* X_i dan X_k adalah koefisien korelasi populasi ρ_{ij} yang dinyatakan dalam varians σ_{ii} dan kovarians σ_{ij} pada Persamaan (2.12)

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \tag{2.12}$$

Matriks koefisien korelasi adalah matriks simetris berukuran $p \times p$ pada Persamaan (2.13) berikut:

$$\boldsymbol{\rho}_{p\times p} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{11}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$\boldsymbol{\rho}_{p\times p} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \tag{2.13}$$

Matriks korelasi pada Persamaan (2.13) dapat menerangkan keeratan hubungan antar variabel yang nilainya berkisar antara $-1 \leq 0 \leq 1$. Jika nilai koefisienya sama dengan 1, maka mengindikasikan korelasi positif sempurna, sedangkan jika sama dengan -1, maka mengindikasikan korelasi negatif yang sempurna dan jika sama dengan 0, maka mengindikasikan tidak adanya korelasi (Hair dkk., 2010).

2.2.3 Algoritma Pembentukan Komponen Utama

Algoritma pembentukan KU dari metode AKU adalah sebagai berikut (Johnson dan Wichern, 2017) :

1. Menghitung matriks korelasi $\boldsymbol{\rho}$ dengan Persamaan (2.13)
2. Menghitung nilai eigen dari matriks korelasi $\boldsymbol{\rho}$ menggunakan rumus sebagai berikut:

$$|\boldsymbol{\rho} - \lambda\mathbf{I}| = \mathbf{0} \tag{2.14}$$

3. Menghitung vektor eigen dari matriks korelasi $\boldsymbol{\rho}$ dengan rumus sebagai berikut:

$$(\boldsymbol{\rho} - \lambda\mathbf{I})\boldsymbol{x} = \mathbf{0} \tag{2.15}$$

4. Menormalisasi vektor eigen dengan rumus

$$\boldsymbol{e} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \tag{2.16}$$

5. Membentuk KU menggunakan Persamaan (2.4)
6. Menghitung proporsi keragaman dengan rumus.

$$\frac{\lambda_{\mathrm{i}}}{\sum_{i=1}^{p}\lambda_{\mathrm{i}}} \tag{2.17}$$

7. Menghitung proporsi keragaman komulatif yang dapat dijelaskan oleh KU ke-*j* menggunakan rumus sebagai berikut

$$\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{p}\lambda_i} \times 100\%, k \leq p \tag{2.18}$$

Banyaknya KU yang terbentuk dapat ditentukan melalui kriteria proporsi komulatif keragaman KU, yaitu proporsi keragaman yang dapat menjelaskan data variabel asal berada pada interval 80%-90% yang dihitung melalui Persamaan (2.18) (Johnson dan Wichern, 2007).

2.3 Pencilan

Pencilan adalah sebuah observasi/objek pada data yang memiliki karakteristik yang sangat berbeda. Dalam analisis klaster, keberadaan pencilan dapat menyebabkan jarak antara kelompok pada data berubah, sehingga dapat mengubah kelompok atau klaster yang optimal. Keberadaan pencilan dapat diidentifikasi melalui jarak mahalanobis yang dapat dihitung dengan Persamaan (2.19)

$$d(\boldsymbol{X}, \boldsymbol{\mu}) = \sqrt{(\boldsymbol{X}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})} \tag{2.19}$$

dengan $d(\boldsymbol{X}, \boldsymbol{\mu})$ adalah jarak dari data ke rata-rata pada ruang dimensi tertentu. Nilai dari $d(\boldsymbol{X}, \boldsymbol{\mu})$ signifikan teridentifikasi sebagai pencilan, jika nilainya lebih besar dari nilai dari tabel statistik *chi-square* $\left(\chi^2_{\alpha:df}\right)$, dengan df adalah derajat bebas, yaitu jumlah variabel yang terlibat dikurang satu (Ghorbani, 2019).

2.4 *K-Medoids*

K-Medoids atau biasa disebut *Partitioning Around Method* (PAM) adalah salah satu metode analisis klaster yang termasuk dalam pendekatan non-hierarki yang telah ditentukan sebelumnya sebanyak *k* buah klaster lalu kemudian dilanjutkan menentukan objek yang representatif (*medoids*) disetiap *k*. Prinsipnya adalah meminimalkan jumlah ketidaksamaan (*dissimiliarity)* antara masing-masing objek dengan objek yang representatif (Suyanto, 2017).

Prinsip dalam memilih objek yang representatif dalam sebuah klaster yang menjadikan *K-Medoids* merupakan pengembangan dari metode *K-Means* yang dapat mengatasi salah satu kelemahannya, yaitu sensitif terhadap pencilan. Penyebabnya ialah dalam metode *K-Means* menggunakan rata-rata terpusat (*centroid*) sebagai pusat dari sebuah klaster sehingga jika terdapat pencilan pada data, maka akan mengubah distribusi dari data yang akan berdampak pada rata-rata terpusat (Suyanto, 2017).

Menurut Han dan Kamber (2006), adapun langkah-langkah dalam analisis metode *K-Medoids* adalah sebagai berikut:

1. Menentukan jumlah *k* buah klaster dan menentukan secara acak objek yang menjadi representatif (*medoids*) sebanyak *k*.
2. Menghitung jarak *Euclidean* antara masing-masing objek dengan objek representatif (*medoids*) melalui rumus

$$d(x_{ip}, o_{mp}) = \sqrt{\sum_{p=1}^{n} (x_{ip} - o_{mp})^2} \qquad (2.20)$$

keterangan :

$d(x_{ip}, o_{mp})$	:	jarak antara objek *x* ke-*i* pada variabel ke-*p* dengan objek *medoids* ke-*m* variabel ke-*p*.
x_{ip}	:	objek *x* ke-*i* pada variabel ke-*p*.
o_{mp}	:	objek *medoids* ke-*m* variabel ke-*p*.

3. Mengidentifikasi setiap objek ke dalam sebuah klaster yang sesuai berdasarkan jarak terdekat dari *medoids*.
4. Menghitung fungsi objektif, yaitu jumlah jarak terdekat dari *medoids* untuk setiap objek.
5. Memilih objek secara acak yang tidak representatif (non-*medoids*) sebanyak *k*.
6. Menghitung jarak *Euclidean* antara masing-masing objek dengan objek yang tidak representatif (non-*medoids*) melalui rumus

$$d(x_{ip}, o_{hp}) = \sqrt{\sum_{p=1}^{n} (x_{ip} - o_{hp})^2} \qquad (2.21)$$

keterangan :

$d(x_{ip}, o_{hp})$: jarak antara objek *x* ke-*i* pada variabel ke-*p* dengan objek non-*medoids* ke-*h* variabel ke-*p.*

o_{hp} : objek non-*medoids* ke-*h* variabel ke-*p.*

7. Menentukan setiap objek ke dalam sebuah klaster yang sesuai berdasarkan jarak terdekat dari non-*medoids* dan menghitung fungsi objektif untuk non-*medoids*.
8. Mengubah *medoids* menjadi non-*medoids* yang representatif jika nilai fungsi objektif *medoids* $>$ fungsi objektif non-*medoids*, sedangkan jika nilai fungsi objektif *medoids* $<$ fungsi objektif non-*medoids* maka yang diubah adalah non-*medoids*.
9. Melakukan langkah 5-8 sampai *medoids* tidak mengalami perubahan.

2.5 *Density Based Spatial Clustering of Application with Noise*

Density Based Spatial Clustering of Application with Noise (DBSCAN) adalah salah satu algoritma pengelompokan analisis klaster yang berdasarkan kepadatan (*density*) pada data (Eko, 2012). Menurut Nagpal dan Mann (2011) DBSCAN merupakan salah satu metode analisis klaster yang membangun area berdasarkan kepadatan yang terkoneksi (*density connected*). Konsep kepadatan yaang diinginkan adalah jika objek berada dalam radius $Epsilon(\epsilon)$ dan jumlahnya $\geq MinPoints$. *Epsilon*(ϵ) pada DBSCAN merepresentasikan nilai dari ambang batas antara tetangga yang dipertimbangkan sebagai kepadatan dan *MinPoints* adalah banyaknya objek yang terdapat dalam radius *Epsilon*(ϵ) (Hasler dkk., 2019). Terdapat tiga macam status dari setiap objek didalam konsep kepadatan yaitu objek yang menjadi inti (*core*), objek yang menjadi batas (*border*), dan objek yang menjadi *noise* (Nur, 2017).

Menurut Irving dkk. (2015) terdapat beberapa istilah umum dalam DBSCAN, yaitu sebagai berikut:

a. $N_\epsilon(p)$: untuk objek $p^* \in X$, lingkungan ϵ yang didefinisikan sebagai

$$N_\epsilon(p^*) = \{x \in X | dist(x, p^*) < \epsilon\}$$

dengan $\epsilon \in \mathbb{R}^+$, X adalah dataset dan *dist* adalah fungsi jarak.

b. *Core point* : sebuah objek $p^* \in X$ adalah sebuah titik inti jika

$$|N_\epsilon(p^*)| \geq MinPoints$$

dengan $MinPoints \in \mathbb{Z}^+$

c. *Directly density reachable* : sebuah objek $p^* \in N_\epsilon$ adalah kepadatan yang dicapai langsung dari $x \in X$ jika memenuhi

$$p^* \in N_\epsilon(x)$$

dengan x adalah objek *core point*.

d. *Density reachable* : sebuah objek $y \in X$ adalah kepadatan yang dicapai dari $x \in X$ jika ada sebuah rantai objek $p_1^*, p_2^*, \ldots, p_n^*$ dengan $y = p_1^*$ dan $x = p_n^*$ dan masing-masing $p_n^* \neq y$ adalah *directly density reachable* dari p_{n-1}^*.

e. *Density connected* : sebuah objek $p^* \in X$ adalah kepadatan yang terhubung ke $q^* \in X$ jika ada sebuah objek $x \in X$ sehingga p^*, q^* adalah *density reachable* dari x.

f. Klaster : *C* adalah sebuah kelompok dari X jika $C \subset X$ dan untuk setiap $p^*, q^* \in C, p^*$ dan q^* adalah *density connected*.

g. *Border point* : objek p^* adalah titik perbatasan, jika $p^* \in C$ dan p^* bukan *core point*.

h. *Noise point* : p^* adalah objek pencilan jika

$$p^* \notin C$$

2.5.1 Penentuan nilai *Epsilon* dan *MinPoints*

Untuk mencari nilai *Epsilon* yang optimal, dapat ditentukan dengan algoritma *K-Nearest Nieghbour* (KNN) melalui plot jarak KNN antara setiap objek dari tetangga terdekat yang terurut dan melihat *knee* (siku) dari kurva. Maksud dari penentuan siku dari kurva KNN adalah agar objek-objek yang terletak dalam sebuah klaster memiliki jarak KNN yang kecil sementara untuk objek-objek dianggap *noise* memiliki jarak KNN yang besar (Hasler dkk., 2019). Algoritma KNN adalah algoritma klasifikasi berdasarkan jarak terdekat dari objek yang satu dengan objek yang lain (Yustanti, 2012). Menurut Rahmah dan Sitanggung (2016), langkah-langkah KNN dalam menentukan parameter *Epsilon* yang optiimal adalah sebagai berikut:

1. Menentukan jumlah *k* tetangga terdekat.
2. Menghitung jarak antar objek dengan Persamaan (2.1).

3. Mengambil perhitungan jarak ke-*k* tetangga terdekat.
4. Mengurutkan dari terkecil hingga terbesar pengambilan perhitungan jarak untuk setiap objek.
5. Membuat kurva *k-dist* yang terbentuk berdasarkan perbandingan antara jarak terurut dan objek terurut.
6. Mengamati titik terjadinya siku atau perubahan jarak yang kritis dari kurva *k-dist* dan mengambil jarak tersebut yang sebagai parameter *Epsilon.*

Sedangkan dalam menentukan parameter *MinPoints*, menurut Hasler dkk. (2019), setidaknya adalah jumlah variabel dari sebuah dataset ditambah satu.

2.5.2 Algoritma DBSCAN

Menurut Khurin'in (2021) algoritma metode DBSCAN adalah sebagai berikut

1. Menentukan parameter *Epsilon* dan *MinPoints*.
2. Memilih objek p^* secara acak.
3. Menghitung jarak antara objek p^* dengan objek *x* yang lain dengan rumus.

$$d(x_{ij}, p_j) = \sqrt{\sum_{j=1}^{n} (x_{ij} - p_j)^2} \qquad (2.22)$$

keterangan :

$d(x_{ij}, p_j)$: jarak antara objek x ke-i pada variabel ke-j dengan objek *p* variabel ke-*j*.
x_{ij} : objek *x* ke-*i* pada variabel ke-*j*.
p_j : objek *p* variabel ke-*j*.

4. Menentukan objek *p* termasuk *core points* dengan kondisi :

$$|N_\epsilon(p)| \geq MinPoints$$

5. Mengulangi langkah 2-4 sampai semua objek telah diproses.
6. Membentuk klaster C dengan 2 kondisi kepadatan yaitu *density reachable* dan *density connected.*

2.6 Uji *Silhouette Coefficient*

Untuk mengukur seberapa baik atau buruk hasil dari klaster yang terbentuk maka dilakukan evaluasi. Evaluasi yang dilakukan menggunakan metode *Silhouette Coefficient* dengan langkah-langkah perhitungan *Slihouette Coefficient* adalah sebagai berikut (Handoyo, 2014):

1. Menghitung jarak rata-rata dari suatu objek, misalkan objek ke-*i* dengan semua objek lain yang berada dalam satu klaster dengan persamaan

$$a(i) = \frac{1}{n_A - 1} \sum_{j \in A, j \neq i}^{n_A} d(i,j) \tag{2.23}$$

dengan n_A banyaknya objek pada satu klaster A

2. Menghitung rata-rata jarak dari objek ke-i tersebut dengan semua objek pada klaster lainnya, misalkan C adalah semua klaster yang terbentuk selain A, maka

$$b(i) = \min_{C \neq A} \left(\frac{1}{n_C} \sum_{j \in C}^{n_C} d(i,j) \right) \tag{2.24}$$

dengan n_C banyaknya objek pada satu klaster C.

3. Menghitung nilai *silhouette* dengan Persamaan (2.23) berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i); b(i)\big)} \tag{2.25}$$

4. Menghitung nilai *Silhoutte Coefficient* (SC)

$$SC = \frac{1}{n} \sum_{i}^{n} s(i) \tag{2.26}$$

dengan nilai SC berada pada interval $-1 \leq SC \leq 1$.

Nilai SC dapat menjelaskan seberapa dekat kemiripan objek didalam klaster. Jika nilai SC mendekati angka 1, maka semakin baik hasil klaster*ing* dalam satu klaster. Sebaliknya jika nilai SC mendekati angka -1, maka semakin buruk hasil klaster*ing* dalam satu klaster (Kaufman dan Rousseeauw, 2005)

Tabel 2.1 Kriteria Silhouette Coefficient

Nilai SC	**Kriteria**
$0.7 < \text{SC} \leq 1$	Struktur kuat
$0.5 < \text{SC} \leq 0.7$	Struktur baik
$0.25 < \text{SC} \leq 0.5$	Struktur lemah
$0.25 \leq \text{SC}$	Struktur buruk

2.7 Indikator Pembangunan Manusia

Menurut BPS (2021), terdapat 4 indikator dalam mengukur IPM di suatu daerah yaitu :

1. Umur Harapan Hidup

   Umur Harapan Hidup (UHH) saat Lahir adalah rata-rata estimasi umur sesorang yang dapat ditempuh sejak orang tersebut lahir. UHH dapat menjelaskan derajat suatu kesehatan masyarakat. Dasar dari perhitungan UHH adalah menggunakan Angka Kematian Bayi (AKB) dengan pola model West Coale-demeny Trussel equation dan proyeksi AKB.

2. Harapan Lama Sekolah

   Harapan Lama Sekolah (HLS) adalah angka (dalam tahun) lama sekolah yang diharapkan pada anak umur tertentu di masa mendatang. HLS dapat menggambarkan suatu kondisi pembangunan sistem pendidikan di berbagai jenjang. Dasar dari perhitungan HLS adalah seseorang yang mengikuti program wajib belajar yang dihitung sejak usia 7 tahun ke atas. Ekspektasi dari HLS yaitu menjumlahkan seluruh peluang yang mungkin setiap variabel.

3. Rata-rata Lama Sekolah

   Rata-rata Lama Sekolah (RLS) adalah jumlah tahun yang digunakan seseorang dalam menjalani pendidikan formal yang diasumsikan jika nilai RLS tidak mengalam penurunan didalam kondisi yang normal di suatu wilayah. Dasar dari perhitungan RLS dihitung sejak usia 25 tahun ke atas. yang diasumsikan jika seseorang tersebut telah menyelesaikan proses pendidikan.

4. Pengeluaran Per Kapita Disesuaikan

   Pengeluaran per kapita disesuaikan (PPP) adalah nilai pengeluaran per kapita dan paritas daya beli seseorang. Rata-rata pengeluaran per kapita dihitung dari level provinsi sampai ke level kabupaten/kota. Perhitungan PPP menggunakan 96 komoditas, yaitu 66 komoditas makanan dan 30 sisanya adalah komoditas yang bukan makanan.

Pengukuran poin pencapaian dari IPM dari suatu wilayah, dapat dikategorikan menjadi beberapa kategori

Tabel 2.2 Kategorisasi Capaian IPM

IPM	Kategori	Jumlah Kabupaten/Kota
IPM $\geq$ 80	Sangat Tinggi	1
70 $\leq$ IPM$<$ 80	Tinggi	11
60 $\leq$ IPM$<$ 70	Sedang	12
IPM $<$ 60	Rendah	0