

**ANALISIS PERBANDINGAN *ALGORITMA K-MEANS* DAN
HIERARCHIAL CLUSTERING UNTUK PENGELOMPOKAN DATA
PENDUDUK INDEKS PEMBANGUNAN MANUSIA PADA
KECAMATAN PERCUT SEI TUAN**

SKRIPSI

Oleh:

NAMA : ZATRIA MUHAMMAD

NPM : 198160071



PROGRAM STUDI TEKNIK INFOMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 13/9/24

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

**ANALISIS PERBANDINGAN ALGORITMA K-MEANS DAN
HIERARCHIAL CLUSTERING UNTUK PENGELOMPOKAN
DATA PENDUDUK INDEKS PEMBANGUNAN MANUSIA
PADA KECAMATAN PERCUT b SEI TUAN**

Skripsi

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana di fakultas teknik

Universitas Medan Area

Oleh:

ZATRIA MUHAMMAD

198160071

PROGRAM STUDI TEKNIK INFOMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 13/9/24

11

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah

3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area
Access From (repository.uma.ac.id)13/9/24

LEMBAR PENGESAHAN

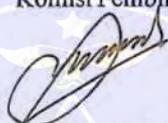
Judul Skripsi : Analisis Perbandingan *Algoritma K-means* dan *Hierarchical Clustering* Untuk Pengelompokan Data Penduduk Indeks Pembangunan Manusia Pada Kecamatan Percut Sei Tuan

Nama : Zatria Muhammad

NPM : 198160018

Fakultas : Teknik Informatika

Disetujui Oleh:
Komisi Pembimbing



Andre Hasudungan Lubis, S.Ti, M.Sc

Pembimbing

Diketahui:



Dr. Eng. Supriatno, ST, MT

Dekan Fakultas Teknik



Rizki Mulya, Kom, M.Kom

Ketua Program Studi

Tanggal Lulus: 5 April 2024



UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 13/9/24
IV

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

HALAMAN PERNYATAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri, adapun bagian – bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain setelah tulis sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulis ilmiah

Saya bersedia menerima saksi pencabutan gelar akademis yang saya peroleh dan saksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari di temukan adanya plagiat dalam skripsi ini

Medan, 5 April 2024



Zatria Muhammad

HALAMAN PERYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR/SKRIPSI/TESIS KEPENTINGAN AKADEMIS

Sebagai sibtas akademik Universitas Medan Area, Saya bertanda tangan dibawah ini:

Nama : Zatria Muhammad
Npm : 198160071
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

Demi Pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Univeristas Medan Area Hak Bebas Royalti Nonekslutif (Non-exclusive Royalty-Free Right) atas karya ilmiah saya yang berjudul :

ANALISIS PERBANDINGAN ALGORITMA K-MEANS DAN HIERARCHIAL CLUSTERING UNTUK PENGELOMPOKAN DATA PENDUDUK INDEKS PEMBANGUNAN MANUSIA PADA KECAMATAN PERCUT SEI TUAN

(Studi Kasus:kantor camat percut sei tuan)

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneklusif ini Universitas Medan Area berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pengkalan data (Databases), merawat, dan memmublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagi pemilik Hak Cipta Demikan pernyataan ini saya buat dengan sebenarnya.

Dibuat di :
Medan , 5 April 2024
Yang Menyatakan



Zatria Muhammad

ABSTRAK

Tingginya pertumbuhan memperoleh jumlah anggota setiap cluster k means dan hierarchical clustering. Mendapatkan hasil clustering pada k-means berupa gambar dengan bentuk 4 cluster Mendapatkan hasil clustering pada hierarchical clustering berupa gambar dendrogram. Menemukan hasil perhitungan dari algoritma dan hierarchical dengan menentukan hasil cluster. Analisi hasil perbandingan algoritma k-means dengan hierarchical telah berasis memperoleh nilai cluster rendah, cluster sedang, cluster tinggi. Hasil cluster dari setiap algoritma pada k-means memiliki 4 klaster terdiri dari: Klaster 0: 175 anggota, Klaster 1: 212 anggota, Klaster 2: 281 anggota, Klaster 3: 262 anggota. Hasil cluster dari setiap algoritma pada hierarchical memiliki 4 klaster terdiri dari: Jumlah Anggota Setiap Cluster: {0: 440, 1: 241, 2: 120, 3: 199}. K-means dapat pemilihan jumlah kluster sebelumnya (k), yang bisa menjadi tantangan terutama jika tidak ada pengetahuan awal tentang struktur data. Hierarchical clustering Tidak memerlukan jumlah kluster yang ditentukan sebelumnya. Pemilihan kluster dapat dilakukan dengan mengamati dendrogram.

Kata kunci : Indeks Pembangunan Manusia; Clustering;Data Mining; Algoritma K-means dan hierarchical.

ABSTRACT

The high growth of population must be followed by the country's economy which is essentially useful to be shown as the achievement of public welfare in economic development as an indicator of success through economic growth (Pratiwi & Indrajaya, 2019). In the research process got the results of the preprocessing stage which had 2 attributes such as: age gender and income. Obtaining the number of members of each cluster k means and hierarchical clustering. Getting clustering results on k-means in the form of images with the form of 4 clusters getting clustering results on hierarchical clustering in the form of dendrogram images. Finding the calculation results of the algorithm and hierarchical by determining the cluster results. Analyze the results of the comparison of the k-means algorithm with hierarchical clustering to obtain the value of low cluster, medium cluster, high cluster. The cluster results of each algorithm on k-means had 4 clusters consisting of: Cluster 0: 175 members, Cluster 1: 212 members, Cluster 2: 281 members, Cluster 3: 262 members. The cluster results of each algorithm on hierarchical had 4 clusters consisting of: Number of Members of Each Cluster: {0: 440, 1: 241, 2: 120, 3: 199}. K-means could pre-select the number of clusters (k), which could be challenging especially if there was no prior knowledge of the data structure. Hierarchical clustering did not require a predetermined number of clusters. Cluster selection could be done by observing the dendrogram.

Keywords: Human Development Index; Clustering; Data Mining; K-means and hierarchical algorithms.

RIWAYAT HIDUP

Penulis dilahirkan di Medan pada tanggal 10 Maret 2001 dari Bapak Zulpadli dan Ibu Indrawati. Penulis merupakan anak ke-empat dari enam bersaudara dan memiliki adik perempuan dan abang serta kakak perempuan. Penulis pertama kali mengenyam pendidikan di bangku SD madrasah hidayah pada tahun 2007 dan lulus pada tahun 2013. Kemudian penulis melanjutkan pendidikan ke jenjang SMP pada tahun 2013 di SMP Negeri 23 Medan dan lulus pada tahun 2016. Pada tahun yang sama, penulis melanjutkan ke jenjang selanjutnya yaitu di SMK Dwiwarna Medan dan lulus pada tahun 2019. Pada bulan September tahun 2019, penulis melanjutkan pendidikan di bangku kuliah dan terdaftar sebagai mahasiswa Fakultas Teknik dengan mengambil Program Studi Teknik Informatika di Universitas Medan Area.



KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa, atas berkat dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan skripsi yang berjudul “Analisis Perbandingan Algoritma K-means dan Hierarchical Clustering dalam Mengelompokkan Data Penduduk Pada Indeks Pembangunan Manusia Pada Kecamatan Percut Sei Tuan”. Skripsi ini merupakan salah satu syarat untuk menyelesaikan pendidikan untuk mencapai gelar sarjana di Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area.

Pada kesempatan ini, penulis mengucapkan banyak terima kasih juga kepada pihak-pihak yang telah memberikan banyak dukungan serta arahan sehingga penulis bisa menyelesaikan penelitian ini dengan baik, untuk itu penulis menyampaikan ucapan terimakasih kepada:

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc. selaku Rektor Universitas Medan Area.
2. Dr.Eng. Supriatno, ST, MT selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom, M.Kom selaku Kepala Program Studi Teknik Informatika Universitas Medan Area.
4. Bapak Andre Hasudungan Lubis, S.Ti, M.Sc, selaku Dosen pembimbing yang telah membantu penulis sehingga penulis dapat menyelesaikan skripsi ini.
5. Orang tua penulis yaitu Bapak Zulpadli dan Ibu Indrawati yang telah mendoakan tiada henti dan memberikan semangat serta membantu penulis sehingga penulis dapat menyelesaikan skripsi ini dengan sebaik baiknya.

6. Seluruh Dosen dan Staf Program Studi Teknik Informatika Universitas Medan Area.
7. Seluruh teman-teman yang sudah memberikan dukungannya selama penulisan proposal skripsi ini, khususnya teman-teman Teknik Informatika angkatan 2019.
8. Seluruh pihak yang tidak dapat disebutkan satu persatu yang membantu dalam menyelesaikan skripsi ini.
9. Teruntuk jodoh yang saat ini masih belum di ketahui keberadaannya entah di bumi bagian mana dan sedang menggenggam tangan siapa. Percalahlah kamu adalah salah satu alasan penulis untuk segera menyelesaikan skripsi ini, agar kelak kamu bangga terhadap penulis yang telah melewati hari-hari sulitnya sendirian. Mungkin saat ini bukan waktu yang tepat untuk bertemu, tapi penulis berharap kelak kita segera di pertemukan dengan versi terbaik kita masin- masing.

Penulis menyadari bahwa penelitian ini masih memiliki kekurangan, oleh karena itu kritik dan saran yang bersifat membangun sangat penulis harapkan demi kesempurnaan penelitian ini. Penulis berharap tugas penelitian ini dapat bermanfaat baik kalangan pendidikan maupun masyarakat. Akhir kata penulis ucapkan terima kasih.

Medan, 29 Mei 2024

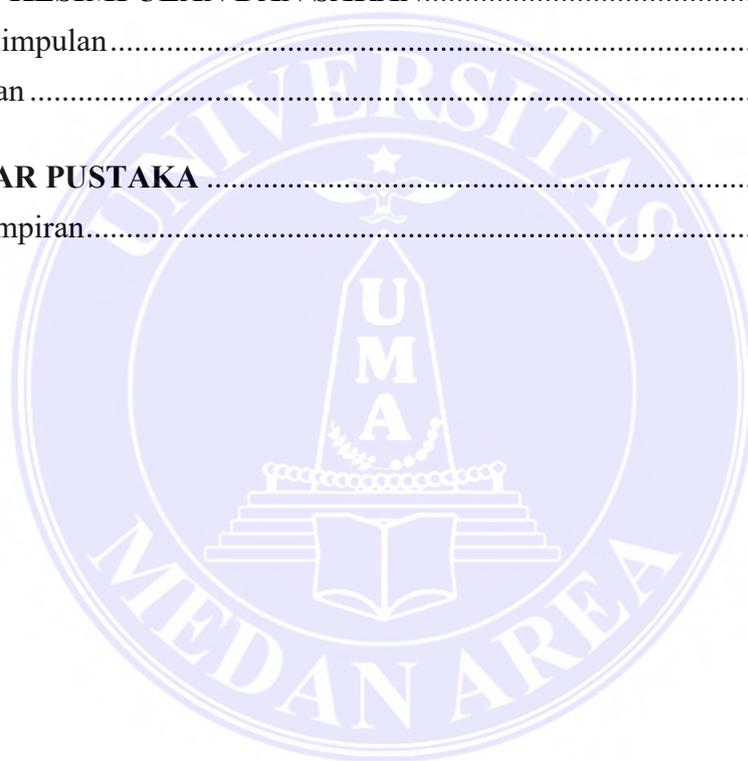
Penulis,

Zatria Muhammad
19816007 1

DAFTAR ISI

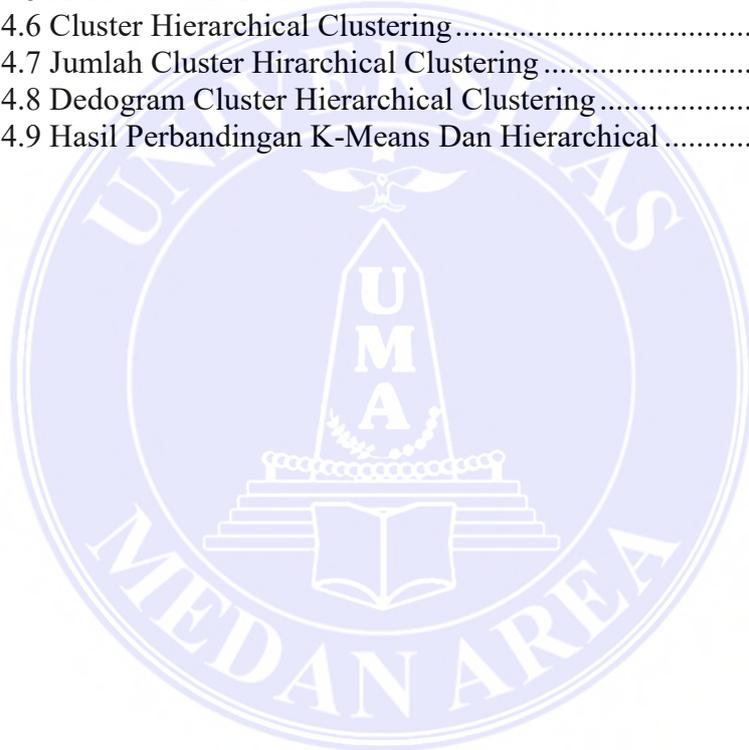
LEMBAR PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAN	iii
HALAMAN PERYATAAN PERSETUJUAN PUBLIKASI	v
ABSTRAK	vii
RIWAYAT HIDUP	viii
KATA PENGANTAR	ix
DAFTAR ISI	xv
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
1.6 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	9
2.1. Indeks Pembangunan Manusia	9
2.2. Data Mining	12
2.3 Clustering	15
2.3.1 Tujuan Clustering	16
2.3.2. Proses Clustering	16
2.3.3. Identifikasi Hubungan	17
2.4. <i>K-Means</i>	17
2.5. Hierarchical Clustering	19
2.6 Davis Bouldin Indeks	22
2.7. Penelitian Terdahulu	23
BAB III METODE PENELITIAN	27
3.1 Lokasi Penelitian	27
3.2 Tahapan penelitian	28
3.3 Pengumpulan Data	29
3.4 <i>Preprocessing</i>	29

3.4.1 Atibut Data.....	30
3.5. Algoritma <i>k-means</i>	32
3.5 Algoritma <i>hierarchical clustering</i>	34
BAB IV HASIL DAN PEMBAHASAN	37
4.1 Hasil.....	37
4.1.1 Pengumpulan Data.....	37
4.1.2 Tahap Preprocessing	38
4.2. Pembahasan	37
BAB V KESIMPULAN DAN SARAN	47
5.1 Kesimpulan.....	49
5.2 Saran	50
DAFTAR PUSTAKA	50
6.1 Lampiran.....	56



DAFTAR GAMBAR

Gambar 2.1.Tahapan Proses Knowledge Discovery in Database	15
Gambar 3.1 Lokasi Penelitian	28
Gambar 3.2 Diagram dalam tahapan penelitian	29
Gambar 3.3 Tahapan algoritma K-Means	36
Gambar 3.4 Tahapan algoritma Hierarchial clustering	38
Gambar 3.5 Menghitung Jarak Menggunakan Dedogram	36
Gambar 4.1 Grafik Metode Elbow	39
Gambar 4.2 Peprosesing Data	40
Gambar 4.3 Anggota Cluster	41
Gambar 4.4 Jumlah Cluster K-Means	41
Gambar 4.5 Hasil Cluster K-Means	42
Gambar 4.6 Cluster Hierarchical Clustering	43
Gambar 4.7 Jumlah Cluster Hirarchical Clustering	43
Gambar 4.8 Dedogram Cluster Hierarchical Clustering	44
Gambar 4.9 Hasil Perbandingan K-Means Dan Hierarchical	45



DAFTAR TABEL

Tabel 1.1 Data Penduduk Kecamatan Percut Sei Tuan di Tahun 2019-2020.....	3
Tabel 2.1 Penelitian Terdahulu	25
Tabel 3.1 Variabel.....	31
Tabel 3.2 Data	32
Tabel 3.3 Menghitung Jarak Menggunakan Dedogram.....	36
Tabel 4.1 Data Penduduk	40
Tabel 4.2 Hasil jumlah anggota cluster k-means	42
Tabel 4.3 Hasil jumlah anggota cluster Hierarchical clustering	44



BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia adalah suatu berkembangnya negara sebagai penduduk dalam jumlah yang terus bertambah dalam cukup besar pasar yang dapat melihat suatu segi input tenaga kerja. Jumlah populasi yang besar dapat memacu pertumbuhan ekonomi melalui kontribusinya sebagai penyedia tenaga kerja. Namun, bertambahnya jumlah orang yang bekerja tanpa menambah jam kerjanya akan memperbesar kemungkinan munculnya masalah ketenagakerjaan, seperti meningkatnya kemungkinan kejahatan, penggunaan narkoba, dan fenomena sosial ekonomi di masyarakat umum (Mahroji & Nurkhasanah, 2019). Tingginya pertumbuhan jumlah penduduk harus diikuti dengan ekonomi negara yang pada hakikatnya berguna untuk di tunjukkan sebagai tercapainya kesejahteraan masyarakat dalam pembangunan perekonomian sebagai indikator keberhasilan melalui pertumbuhan ekonomi (Pratiwi & Indrajaya, 2019).

Indeks Pembangunan Manusia (IPM) merupakan indikator penting dalam mengukur kualitas proyek konstruksi yang berpusat pada manusia. IPM yang pertama kali disebutkan oleh United Nations Development Programme (UNDP) pada tahun 1990 merupakan data strategis yang digunakan untuk menentukan ruang lingkup kegiatan pemerintah. IPM mengklarifikasi tingkat akses kepentingan hubungan dengan hasil implementasi pendapatan, kesehatan,

pendidikan, dan inisiatif lainnya. Tujuan pengelompokan IPM adalah untuk data penduduk dalam beberapa kelompok yang memiliki karakteristik IPM tingkat tinggi. (Azzahra & Wijayanto, 2022). IPM merupakan faktor yang berguna dalam peningkatan nilai kinerja pembangunan sosial ekonomi, sehingga bertujuan untuk meningkatkan pengetahuan penduduk agar lebih sejahtera. IPM mampu menjelaskan level kesejahteraan untuk menunjukkan kemampuan dalam menggunakan dan memanfaatkan sumber pertumbuhan ekonomi, dengan jumlah pendapatan perkapita, angka harapan hidup, pendidikan (Sabilla & Sumarsono, 2022).

Sehingga konsep dalam pembangunan manusia dianalisis dan diketahui dari sudut manusianya, tidak hanya dari pertumbuhan ekonominya. Penduduk dipandang sebagai tujuan yang ada, sedangkan proses pembangunan dipandang sebagai strategi utama untuk mencapai tujuan yang telah ditetapkan. Produktivitas, Pemerataan, Kestinambungan, dan Pemberdayaan adalah empat faktor yang harus diperhatikan untuk memastikan tercapainya tujuan pembangunan.(Mongan,2019).

Kecamatan Percut Sei Tuan, yang terletak di Kabupaten Deli Serdang, memiliki batas administratif dengan beberapa Kecamatan di Kota Medan. Selain itu, wilayahnya juga berbatasan dengan Kecamatan Labuhan Deli dan Kecamatan Batang Kuis. Kecamatan Percut sei tuan salah satu penduduknya cukup dan banyak terdapat jumlah kelurahan yang ada di percut sei tuan yang memiliki 20 kelurahan dengan jumlah penduduk seperti ada pada Tabel 1.1 di bawah. Struktur organisasi Pemerintah Kecamatan Percut Sei Tuan sesuai Peraturan Daerah Kabupaten Deli Serdang. Nomor 1 Tahun 2014, Tanggal 05 Maret 2014 Serdang,

tentang Perubahan atas Peraturan Daerah No. 5 Tahun 2007.

Tabel 1.1 Data Penduduk Kecamatan Percut Sei Tuan di Tahun 2019-2020

No	Desa/Kelurahan	Luas(km ²)	Penduduk (jiwa)		
			Jumlah	Persentase	Kepadatan per km ²
1	Amplas	3,10	11 125	2,19	3 271
2	Kenangan	1,27	20 372	5,84	21 274
3	Tembung	5,35	20 372	13,20	11 425
4	Sumber Rejo Timur	4,16	28 328	6,46	7 190
5	Sei Rotan	5,16	30 624	6,61	5 932
6	Bandar Kalippa	18,48	35 665	9,06	2 269
7	Bandar Khalippa	7,25	41 530	9,99	6 379
8	Medan Estate	6,90	12 123	4,00	2 684
9	Laut Dendang	1,70	15 048	3,89	10 601
10	Sampali	23,93	25 061	7,20	1 394
11	Bandar Setia	3,50	25 474	5,34	7 057
12	Kolam	5,98	17 679	3,76	2 913
13	Saentis	24,00	17 947	4,21	813
14	Cinta Rakyat	1,48	14 432	3,34	10 434
15	Cinta Damai	11,76	5 008	1,23	486
16	Pematang Lalang	20,10	1 497	0,40	93
17	Percut	10,63	14 640	3,48	1 516
18	Tanjung Rejo	19,00	10 273	2,42	591
19	Tanjung Selamat	16,32	5 470	1,36	385
20	Kenangan baru	0,72	21 186	6,00	38 607
Percut Sei Tuan 2020		190,79	402 468	100	2 426
2019		190,79	462 936	100	2 381

Bertambahnya Pertambahan jumlah penduduk dalam suatu kota atau kabupaten yang tidak ditangani dengan baik dan sistematis dapat menyebabkan dampak negatif, termasuk peningkatan tindakan tidak aman, tingkat pencemaran lingkungan yang meningkat, berkurangnya lahan hijau karena digunakan untuk perkembangan pemukiman, dampak terhadap pertumbuhan ekonomi, dan berbagai dampak negatif lainnya yang berdampak pada perekonomian (Sonang dkk., 2019). Sehingga dibutuhkan pengelompokan IPM untuk mengambil tingkatan pembangunan IPM yang ada di daerah tersebut.

Data mining dapat digunakan untuk pengelompokan (*clustering*) data yang digunakan untuk mengelola *knowledge*. *Clustering* yang umum di kenal terdiri *hierachical clustering* dan *partitioning*. Metode *hierarchical clustering*

sendiri terdiri dari *complete linkage clustering*, *single linkage clustering*, *average linkage clustering* dan *centroid linkage clustering*. Sedangkan metode sendiri terdiri dari *k-means* dan *fuzzy k-means*. Dalam proses mengatasi permasalahan sehari-hari yang ada, pemerintah daerah dan masyarakat bekerja sama, dan diharapkan dapat menciptakan lapangan kerja baru dan tumbuhnya inisiatif ekonomi (pertumbuhan ekonomi) di wilayah tersebut. (Safira dkk., 2020).

Terdapat dua metode dalam *clustering* adalah proses memindahkan satu objek ke sekelompok objek terkait dengan karakteristik terkait. *K-Means* adalah satu-satunya teknik pengelompokan yang dapat digunakan dalam masalah tersebut di atas. Sebagai metode pengelompokan data *non-hierarkis* yang mengelompokkan data ke dalam satu atau lebih kelompok. Dalam satu kluster, data yang memiliki karakteristik identik digabungkan dengan data yang memiliki karakteristik berbeda pada *cluster* yang berbeda. (Imantika dkk., 2019).

Algoritma *k-means* dapat digunakan untuk mengelolah data penduduk. Salah satunya adalah penelitian yang dilakukan oleh (Nasution et al., 2020), Penerapan Algoritma *K-Means* dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi. Dari hasil pengelompokan diperoleh sebanyak 8 cluster tinggi dan 26 cluster rendah. Diharapkan penelitian juga bisa memberikan masukan setiap pemerintah agar dapat memberikan perhatian lebihnya kepada penduduk. Algoritma *Agglomerative Hierarchical Clustering* dapat digunakan untuk mengkluster data penduduk. Salah satunya adalah penelitian yang dilakukan oleh. (Dani dkk., 2019). Penerapan *Hierarchical Clustering*

Metode *Agglomerative* runtun waktu pada data, Berdasarkan hasil analisis, setiap jarak ukuran kemiripan dengan hasil yang terbaik untuk mengelompokkan Kabupaten/Kota di Provinsi Kalimantan Timur adalah menggunakan jarak berbasis autocorrelation (ACF) dengan nilai koefisien korelasi cophenetic sebesar 0,99. Algoritma optimal untuk pengelompokkan adalah algoritma average linkage, dipilih karena memiliki nilai koefisien korelasi cophenetic yang tinggi dibandingkan dengan algoritma lainnya. Hasilnya menghasilkan dua klaster yang representatif berdasarkan koefisien silhouette (Dani dkk., 2019).

Dari berbagai kelebihan yang dimiliki algoritma *K-Means* dan *Hierarchial Clustering*, penulis ingin mengetahui hasil pengklasteran IPM dari penduduk oleh kedua algoritma tersebut. Sehingga dapat dibuktikan algoritma mana yang terbaik dan cocok pada data penduduk kecamatan Percut Sei Tuan. Karena dari pada itu, penelitian ini akandituangkan dengan bentuk tugas akhir yang berjudul “**Analisis Perbandingan Algoritma K-means Dan Hierarchial Clustering Untuk Pengelompokan Data Penduduk Pada Kantor Camat (Studi Kasus kantor camat percut sei tuan)**”.

1.2 Rumusan Masalah

Bagaimana perbandingan algoritma *k-means* dengan algoritma *hierarchical* dengan mengelompokkan penduduk Percut Sei Tuan pada Indeks Pembangunan Manusia.

1.3 Batasan Masalah

1. Data penelitian ini dapat digunakan dalam proses pengambilan secara langsung dari kantor camat percut sei tuan

2. Proses penelitian ini menggunakan Algoritma *k-means* dan *hierarchical clustering*
3. Atribut yang digunakan Umur dan pendapatan,
4. Sampel data penduduk sejumlah 1000 data

1.4 Tujuan Penelitian

Beberapa tujuan dari penelitian ini dilakukan antara lain:

1. metode terbaik untuk mengetahui pengelompokan penduduk indeks pembangunan manusia berdasarkan indikator IPM dengan perbandingan 2 metode yaitu *k-means* dan *hierarchical clustering*.
2. Menguji perbandingan Algoritma k-means dengan algoritma *hierarchical clustering* untuk di lakukan penelitian, apakah dua algoritma tersebut dapat digunakan untuk melihat hasil dalam sebuah perbandingan pengelompokan penduduk pada indeks pembangunan manusia pada kecamatan percut sei tuan.
3. Untuk Mengelompokkan data indeks pembangunan manusia di percut sei tuan.
4. Mampu mengelompokkan daerah-daerah ke dalam beberapa kelompok, di mana setiap kelompok memiliki tingkat keseragaman karakteristik IPM yang tinggi di dalamnya, sementara terdapat keberagaman tingkat karakteristik IPM antara kelompok-kelompok tersebut..

1.5 Manfaat Penelitian

1. Sebagai bahan evaluasi kinerja pemerintah dalam pembangunan kualitas hidup manusia.

2. Dapat digunakan sebagai bahan perencanaan kebijakan pemerintah selanjutnya
3. Dapat membandingkan beberapa metode agar diperoleh metode yang paling tepat untuk clustering indeks pembangunan manusia dan hasil cluster
4. Wilayah-wilayah dalam kelompok yang sama memiliki kemiripan yang tinggi pada Indeks Pembangunan Manusia, dapat diterapkan pada wilayah dalam suatu kelompok dapat dijadikan acuan kebijakan wilayah lain dalam kelompok yang sama.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini ditunjukkan kepada pembaca untuk lebih mudah memahami isi dari penelitian ini. Berikut sistematika penulisan penelitian ini secara garis besar, yaitu:

BAB I PENDAHULUAN

Pada bagian pengantar, akan dibahas mengenai konteks atau latar belakang penelitian ini, yang bertujuan untuk menjelaskan mengapa penelitian ini dilakukan. Selanjutnya, akan dibahas rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian.

BAB II TINJAUAN PUSTAKA

Pada tinjauan pustaka membahas hal – hal yang mendasar berisi teori – teori yang berkaitan dengan analisis perbandingan algoritma k-means dengan hierarchical clustering untuk pengelompokan data penduduk indeks dalam pembangunan manusia pada kecamatan percut sei tuan dan hal- hal yang mendukung dalam melakukan penelitian analisis *clustering*.

BAB III METODE PENELITIAN

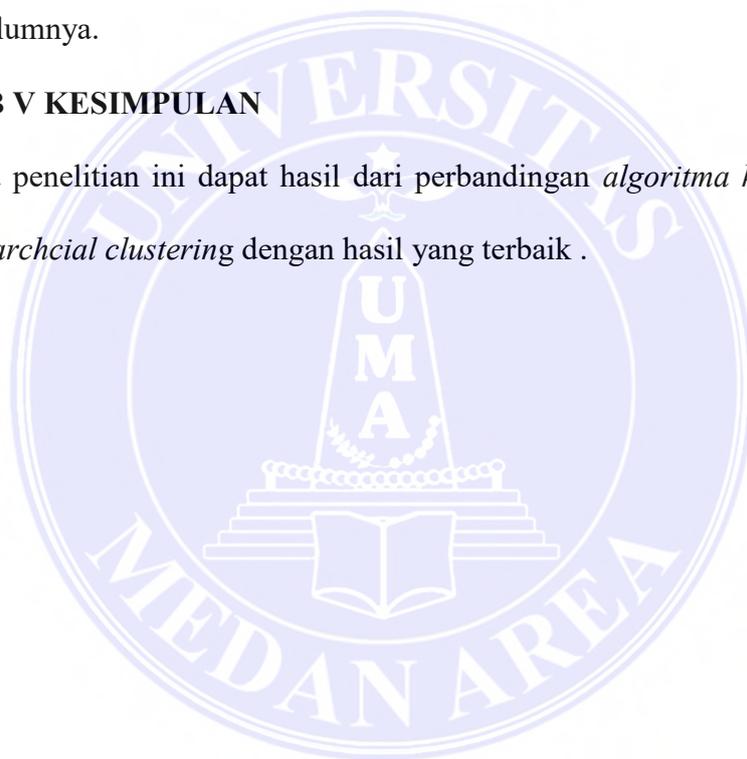
Pada bagian ini membahas tentang sumber data dan tahapan yang dilakukan dalam melakukan penelitian ini serta gambaran umum sistem yang akan dikerjakan.

BAB IV HASIL DAN PEMBAHASAN

Pada bagian ini untuk mempersiapkan dan menerapkan. Pada tahap ini penulis akan menjabarkan analisis hasil yang akan dikemukakan dalam analisis dengan hasil data yang terbaik, detail tentang keputusan *final* dari bab sebelumnya.

BAB V KESIMPULAN

Pada penelitian ini dapat hasil dari perbandingan *algoritma k-means* dengan *hierarchcial clustering* dengan hasil yang terbaik .



BAB II

TINJAUAN PUSTAKA

2.1. Indeks Pembangunan Manusia

Kesehatan, tingkat pendidikan, dan indikator ekonomi merupakan salah satu indikator untuk mengetahui pembangunan ekonomi yang mengukur taraf kualitas fisik dan non fisik penduduk. Karena itu, manusia adalah satu-satunya kekayaan bangsa sesungguhnya. Pembangunan sumber daya manusia baik secara fisik maupun non fisik menggunakan makna sebagai tolak ukur kinerja penduduk. Pertumbuhan ekonomi dan pembangunan manusia saling terkait dan sebagian besar berkontribusi satu sama lain. Tingkat pembangunan manusia yang sangat menentukan kemampuan penduduk dalam menyerap dan mengelola sumber-sumber pertumbuhan ekonomi, baik kaitannya dengan teknologi atau melawan kelembagaan sebagai sarana penting untuk mencapai pertumbuhan (Utami, 2020).

Indeks Pembangunan Manusia berguna untuk membandingkan kegiatan konstruksi manusia baik yang dilakukan di luar negeri maupun di dalam negeri. Pembangunan manusia sangat penting karena jika suatu daerah tidak memiliki Sumber Daya Alam (SDA) yang berpotensi bermanfaat, masih dapat dibangun dan ditingkatkan dengan menggunakan Sumber Daya Manusia (SDM). Jadi, dalam membangun suatu wilayah tertentu, sumber daya manusia sangat penting. Indonesia, negara yang saat ini sedang berkembang, terus berupaya meningkatkan kesejahteraan warganya dengan mendorong pembangunan ekonomi yang bermanfaat bagi penduduk remaja maupun dewasa. (Kiha dkk., 2021).

Indeks Pembangunan Manusia (IPM) melakukan perbandingan standar hidup, standar pendidikan, dan norma budaya untuk setiap negara di dunia. IPM digunakan untuk mengklasifikasikan apakah suatu negara adalah negara yang matang, berkembang, atau terbatas serta untuk mengukur dampak kebijakan ekonomi terhadap kualitas hidup. Dan jika ada tolak keraguan tentang suatu wilayah, apakah itu wilayah, negara, atau bahkan seluruh wilayah, perlu dicatat bahwa dengan IPM yang tinggi, keberhasilan program untuk membangun wilayah itu dapat diperkirakan. Untuk menciptakan IPM dengan standar kualitas yang tinggi, pemerintah harus membuat program dan fasilitas yang dirancang khusus untuk tujuan tersebut. Namun, mereka juga harus bekerja dengan masyarakat umum untuk memastikan bahwa program tersebut dapat berjalan. (Widodo dkk., 2020).

Adapun atribut yang digunakan menurut IPM pada data penelitian yang ada beberapa penjelasan dari atribut data tersebut yang terdiri dari :

1. Umur

Mengukur hidup yang berkualitas yang panjang umur dan sehat, pilihan untuk mempunyai akses terhadap sumber daya agar di butuhkan yang dapat hidup secara layak. Sehingga pembangunan manusia mampu menciptakan yang lingkungan dari mungkin bagi rakyat untuk menikmati umur panjang. Karena oleh itu manusia harus di posisikan sebagai potensi kekayaan bangsa (Patadang dkk., 2021).

2. Kesehatan

Merupakan hal yang penting dalam mewujudkan pelayanan kesehatan yang dapat dijangkau oleh lapisan masyarakat disetiap wilayah

di kecamatan, pendistribusian fasilitas kesehatan yang merata disetiap wilayah di kecamatan percut sei tuan yang mampu menjangkau dari segi biaya maupun penempatan. Hal lain yang berhubungan dengan kesehatan yang perlu diperhatikan adalah tenaga kesehatan atau obat-obatan. Jumlah dokter harus ditambah untuk meningkatkan ketersediaan pelayanan kesehatan masyarakat, serta untuk menyediakan obat tenaga dengan insentif yang diperlukan untuk mengobati penduduk kota kecil secara efektif. (Marlin dkk., 2022)

3. Pendidikan

Indeks Pembangunan Manusia (IPM) suatu keinginan yang mengukur pembangunan sosio-ekonomi suatu negara, yang mengkombinasikan keinginan di bidang pendidikan. Dapat membangun sarana dan sistem pendidikan yang baik dalam hal pendidikan mutlak yang dibutuhkan. Pengeluaran pembangunan dari sektor konstruksi dapat disetujui untuk penyediaan infrastruktur pendidikan dan penyampaian materi pendidikan secara tepat waktu kepada seluruh warga negara Indonesia. (Mahuze dkk., 2022).

4. Pendapatan

Pendapatan di Indonesia adalah hasil dari tingkat remunerasi negara yang sangat rendah untuk memenuhi kebutuhan dasar. Berbagai upaya telah dilakukan untuk mengurangi tingkat ketimpangan di suatu wilayah, dengan pembangunan ekonomi sebagai yang utama. Pendapatan adalah isu penting lainnya untuk pembangunan bangsa di setiap negara. Ketimpangan Pendapatan terhubung dengan distribusi pendapatan yang

tersedia bagi penduduk di negara manapun. Tingkat produktivitas yang meningkat berarti bahwa distribusi pendapatan masyarakat umum tertinggal (Febriyani & Anis, 2021).

2.2. Data Mining

Data mining adalah suatu proses dalam menemukan pola, korelasi, dan tren terbaru yang sedang berkembang dan yang memiliki makna dengan memilih – memilih data dalam jumlah yang cukup besar yang tersimpan di dalam repositor (Nabila dkk., 2021). Tujuan utama dari data mining yaitu untuk menggali, menambang, untuk menemukan pengetahuan dari data yang telah didapat. Data mining adalah salah satu langkah dalam (KDD) *Knowledge discovery in database* (Kharis & Zili, 2022). *Knowledge discovery in database* (KDD) merupakan mengidentifikasi pola yang ada di dalam data dan keseluruhan proses untuk mencari data, dimana pola yang di dapatkan memiliki sifat baru, sah, dapat dimengerti dan bermanfaat.

Menurut (Kharis & Zili, 2022) proses *knowledge discovery in database* (KDD) secara garis besar diantaranya :

1. *Data Integration* (Integrasi data)
2. *Data Cleaning* (Pembersihan data)
3. *Data Transformation* (Transformasi data)
4. *Data Selection* (Pemilihan data)
5. *Data Mining* (Penggalian data)
6. *Pattern Evaluation* (Evaluasi pola)
7. *Knowledge presentation* (Penyajian pengetahuan)

Setiap algoritma/metode/teknik tersebut mempunyai fungsi dan tujuan yang berbeda-beda. Berikut pengelompokan *data mining* berdasarkan fungsi tujuan (Buulolo, 2020):

- Deskripsi

Secara umum, peneliti mempertimbangkan cara untuk mendeskripsikan pola dan tren dalam data yang dimaksudkan untuk mengidentifikasi dan memahami pola berulang dalam pola dan mengubah pola tersebut menjadi alat yang dapat digunakan untuk menyederhanakan berbagai aktivitas. Misalnya, di toko belanja, pelanggan sering membeli produk B dan C secara bersamaan dan bersamaan. Namun, manajemen supermarket dapat mengubah daftar produk dengan menempatkan B dan C di lokasi yang sama atau berdekatan, sehingga kecil kemungkinan pelanggan akan memutuskan untuk membeli produk tersebut lagi. Satu-satunya algoritma yang disebutkan dalam deskripsi adalah algoritma apriori.

- Estimasi

Definisi dari estimasi adalah perkiraan/prediksi, dapat mirip dengan estimasi dan klasifikasi, kecuali variabel yang tujuan kurang lebih kearah numerik dari pada kategori. Contohnya perkiraan pendapatan seorang sales penjualan produk tertentu berdasarkan lama bekerja, perkiraan pendapatan perusahaan pada bulan tertentu dan sebagainya. Algoritma yang terdapat dalam estimasi adalah regresi linear sederhana, regresi linear berganda dan lain-lain.

- **Prediksi**

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi dimasa depan). secara umum prediksi hampir sama dengan klasifikasi. Salah satu fungsi *data mining* dapat di gunakan adalah untuk memprediksi nilai dari hasil prediksi yang di gunakan pada masa yang akan datang berdasarkan data-data sebelumnya.

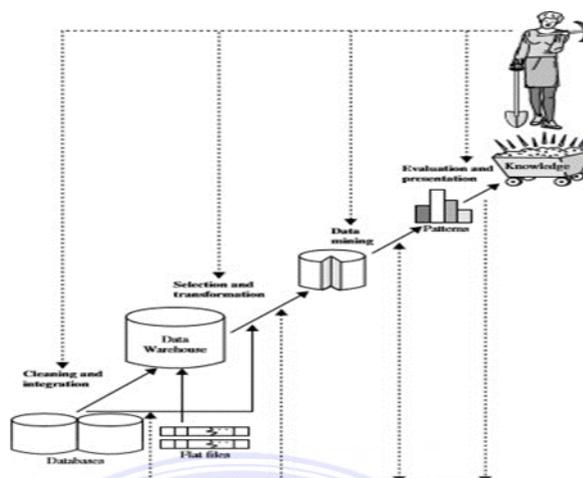
- *Klasifikasi*

Dalam mengklasifikasikan variabel, tujuannya bersifat kategoris. Sebagai contoh, kita dapat mengelompokkan pendapatan menjadi tiga kategori, yakni tinggi, sedang, dan rendah.

- *Clustering*

Clustering lebih berfokus pada pengelompokan kasus ke dalam kelas yang menunjukkan kemiripan. Dalam clustering, perbedaannya dengan klasifikasi terletak pada penggunaan variabel keputusan atau target. Sebagai contoh, kita dapat mengelompokkan keluarga yang memenuhi syarat dan yang tidak memenuhi syarat untuk mendapatkan program Keluarga Harapan (PKH) berdasarkan jumlah pendapatan mereka.

Salah satu teknik yang paling penting untuk analisis data adalah analisis klaster, juga dikenal sebagai pengelompokan. Tujuan dari metode analisis data clustering adalah untuk mengelompokkan data yang memiliki kesamaan karakteristik dalam satu wilayah tertentu. *K-Means* adalah satu-satunya deskriptor yang digunakan dalam pengembangan metode clustering, dan 6*- kdd menurut (Utomo & Mesran, 2020) :



Gambar 2.1. Tahapan Proses Knowledge Discovery in Database (KDD) (Utomo & Mesran, 2020) .

2.3 Clustering

Clustering tujuan teknik data mining tertentu adalah untuk mendistribusikan data (objek) ke beberapa cluster atau kelompok sehingga objek yang mirip harus menjadi bagian dari cluster yang sama dengan objek yang berbeda. Menurut skenario kasus terbaik, tujuan utama pengelompokan adalah untuk memaksimalkan homogenitas internal di dalam sebuah kluster dan heterogenitas eksternal di antara kluster serupa. (Adha dkk., 2021).

Clustering adalah teknik untuk mengatur data yang sering digunakan dalam data mining atau analisis data. Proses menempatkan satu set objek data tertentu ke dalam kantong yang disebut sebagai cluster disebut clustering. Oleh karena itu, metode pengelompokan yang dimaksud sangat berguna untuk menemukan kelompok yang tidak mudah diidentifikasi dalam data. Ada contoh di mana pengelompokan menerapkan karakteristik yang identik ke satu tempat dan data yang identik menghasilkan penerimaan sekelompok objek dengan

karakteristik yang identik. Centroid model yang digunakan dalam penerapan *K-Means* digunakan untuk mengklasifikasikan atau mengelompokkan data menggunakan beberapa atribut yang berbeda satu sama lain (Parlambang & Fauziah, 2020).

2.3.1 Tujuan Clustering

Tujuan utama analisis kelompok adalah mengklasifikasi objek pada penduduk setiap desa di kecamatan percut sei tuan seperti penduduk desa percut, penduduk desa tembung, penduduk desa bandar khalipa produk, dll ke dalam kelompok-kelompok yang relatif homogen didasarkan pada setiap kelompok variabel yang sedang digunakan untuk melakukan penelitian. Subyek penelitian ini adalah seluruh Kecamatan percut sei tuan. Oleh karena itu, tujuan penggunaan dengan tujuan analisis dalam penelitian ini adalah preservasi data. Dengan menggunakan analisis klaster, kecamatan akan dipindahkan secara klaster berdasarkan beberapa variabel, seperti yang terjadi pada contoh ini..

2.3.2. Proses Clustering

Proses clustering saat ini dilakukan secara manual karena membutuhkan banyak waktu, tenaga, dan kesabaran. Pada penelitian ini, tujuannya adalah agar data Clustering lebih mudah digunakan agar dapat sebagai referensi dalam kampanye promosi produk Telkomsel secara nasional. Metode yang digunakan untuk mengatasi permasalahan tersebut adalah *clustering* dengan algoritma *K-Means*. Algoritma K-Means adalah alat untuk memecah data dalam jumlah besar menjadi bagian yang lebih kecil dan lebih mudah diatur. Data dari penelitian dikompilasi di sini. (Handoko dk., 2020).

2.3.3. Identifikasi Hubungan

Identifikasi Hubungan. Dengan bantuan cluster yang ditentukan dan struktur yang memperhitungkan data yang digunakan untuk mewakili cluster, subjek memiliki kesempatan untuk menjelaskan hubungan antara dua pengamatan yang belum tentu sama dengan pengamatan individu. Jika analisis seperti analisis diskriminan digunakan untuk mengidentifikasi jaringan secara empiris, atau jika suatu kelompok menggunakan metodologi yang lebih ketat, struktur yang diturunkan dari analisis klaster kemungkinan akan menunjukkan jaringan atau persamaan dan perbedaan yang sebelumnya tidak dikenal.

2.4. K-Means

Means adalah sebuah teknik *Clustering* yang umum digunakan dalam partisi. Algoritma *K-Means* adalah model Centroid. Model Centroid adalah model yang menggunakan Centroid untuk membuat *Cluster*. Pusat galaksi dikenal sebagai centroid. Centroid berdasarkan nol. Untuk menghitung jarak sebuah data objek terhadap Centroid, gunakan Centroid. Setiap data objek dianggap sebagai bagian dari cluster jika memiliki hubungan dengan Centroid *Cluster* saat ini. Algoritma *K-Means* dapat digambarkan sebagai algoritma pembelajaran yang sangat berguna untuk menangani permasalahan pengelompokan tertentu dengan tujuan meminimalkan bahaya ganda. (Hajar ddk., 2020)

K-Means salah satu metode tunggal untuk data kelompok yang dapat menampilkan data kelompok pada data yang hadir dalam bentuk dua kelompok atau lebih. Metode ini melibatkan pengelompokan data sedemikian rupa sehingga data dengan karakter yang sama akan ditempatkan pada kelompok yang sama dan data dengan karakter yang berbeda akan ditempatkan pada

kelompok yang juga berbeda. K-means adalah algoritma yang paling kuat dan paling sering digunakan untuk clustering data. *Geometri Euclidean* dapat digunakan saat melakukan perhitungan jarak ke-I (x_i) di lokasi gugus ke-k (c_k), yang disebut "dik". (Adha dkk., 2021).

$$d_{ik} = \sqrt{\sum_j^m (X_{ik} - C_{ij})^2} \quad (2.2)$$

Keterangan :

d_{ik} :jarak objek ke-i pada pusat kelompok ke-k

X_{ik} :nilai objek ke-i pada variabel k

C_{ij} :pusat kelompok ke-i pada variabel j

m :jumlah variabel yang di gunakan

j menyatakan objek, menyatakan pusat kelompok

i menyatakan keanggotaan kelompok

K-Means merupakan teknik *clustering* yang diperoleh dari sebuah dataset dengan cara menghitung jarak dari setiap titik ke pusat cluster secara iterative. Algoritma *K-Means* memiliki beberapa aturan yaitu (Prastiwi et al., 2022):

1. Total cluster yang dibutuhkan
2. Jenis atribut adalah numeric. Proses untuk clustering menggunakan algoritma *k-means* adalah sebagai berikut :

- 1 Tentukan jumlah k yang diinginkan untuk total kluster yang akan dibentuk.
- 2 Mulailah dengan memilih nilai centroid kluster awal berdasarkan jumlah k.
- 3 Gunakan rumus jarak *Euclidean* untuk menghitung jarak antara setiap data masukan dan setiap centroid, dengan tujuan menemukan jarak terdekat

antara setiap titik data dan centroid.

- 4 Kelompokkan setiap item data berdasarkan jarak minimum ke *centroid*.
- 5 Perbarui nilai pusat kluster (centroid).
- 6 Ulangi langkah 3 hingga 5 hingga anggota setiap kluster tetap tidak berubah. Ini merupakan iterasi untuk memastikan stabilitas kluster.

Berikut ini adalah rumus-rumus yang digunakan dalam perhitungan algoritma K-Means Clustering.

$$de = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (2.3)$$

Algoritma *K-Means Clustering* merupakan metode yang sederhana namun efektif dalam mengidentifikasi kluster dalam data, dengan langkah-langkah sebagai berikut:

- 1 Tetapkan nilai *k* sebagai jumlah *cluster* yang diinginkan dan tentukan pusat-pusat *cluster*.
- 2 Menghitung jarak antara setiap data dengan pusat *cluster* menggunakan rumus *Euclidean*.

2.5. Hierarchical Clustering

Menurut algoritma *clustering*, data akan dikelompokkan ke dalam *cluster* berdasarkan seberapa dekat satu set data terkait dengan yang lain. Tujuan *clustering* adalah untuk meminimalkan kesamaan antara anggota *cluster* yang berbeda sambil memaksimalkan kesamaan dalam satu *cluster*. *Hierarchical clustering* adalah algoritma pengelompokan yang populer dalam kategori ini. Satu-satunya algoritma pengelompokan yang dapat digunakan untuk mengelompokkan dokumen bersama-sama disebut *hierarchial clustering*. Dari

teknik *hierarchical clustering* seseorang dapat memperoleh kumpulan partisi yang berurutan, di mana yang terakhir berisi yang pertama.

a. *Cluster – cluster* yang mempunyai poin – poin individu. *Cluster – cluster* ini berada di level yang paling bawah.

b. Sebuah cluster yang didalamnya terdapat poin – poin yang dipunyai semua cluster didalamnya. *Single cluster* ini berada di level yang paling atas.

Hierarchical Clustering merupakan suatu teknik analisis kelompok yang bertujuan untuk membentuk hirarki kelompok data. Terdapat dua jenis strategi umum dalam pengelompokan, yaitu *Agglomerative (Bottom-Up)* dan *Devisive (Top-Down)*. Metode pembentukan *cluster* biasanya diklasifikasikan berdasarkan jenis struktur klaster yang dihasilkan. Secara umum, metode klastering dapat dibagi menjadi dua kategori, yakni *Non-Hierarchical Clustering* (klastering nonhirarkhis) dan *Hierarchical Clustering* (klastering hirarkhis). Dalam konteks ini, diterapkan metode *Hierarchical Clustering* untuk mengelompokkan berbagai kriteria data yang diperoleh melakukan prediksi (Pratama dkk., 2020).

Langkah – langkah Algoritma *Agglomerative Hierarchical Clustering* :

1. Hitung Matrik Jarak antar data.
2. Gabungkan dua kelompok terdekat berdasarkan parameter kedekatan yang ditentukan.
3. Perbarui Matrik Jarak antar data untuk merepresentasikan kedekatan diantara kelompok baru dan kelompok yang masih tersisa.
4. Ulangi langkah 2 dan 3 hingga hanya satu kelompok yang tersisa.

Setelah langkah di atas, maka terbentuklah Matrik Jarak, misal dengan

Manhattan Distance, yang ditunjukkan pada rumus:

$$D = \sum_{i=1}^N |b_i - a_i| \quad (2.4)$$

Keterangan :

D = Kriteria

b_i = Data 1

a_i = Data 2

Beberapa metode Pengelompokan Agglomerative Hierarchical :

1. *Single Linkage* (Jarak Terdekat) Metode ini sangat sesuai untuk digunakan dalam kasus pengelompokan yang tidak tergantung pada bentuk (shape independent clustering), karena kemampuannya dalam membentuk pola tertentu dari kluster, seperti yang ditunjukkan dalam rumus. :

$$d_{uv} = \min\{d_{uv}\}, d_{uv} \in D \quad (2,5)$$

Keterangan :

D = Kriteria

d_{uv} = Data

\in = Eliminasi

2. *Complete Linkage* (Jarak Terjauh) Metode ini sangat efektif dalam mengurangi *varians* di dalam kluster karena melibatkan *centroid* saat menggabungkan antar kluster. Selain itu, metode ini cocok untuk data yang memiliki outlier, sebagaimana diindikasikan dalam rumus:

$$d_{uv} = \max\{d_{uv}\}, d_{uv} \in D \quad (2.6)$$

Keterangan :

D = Kriteria

d_{uv} = Data

\in = Eliminasi

3. *Average Linkage* (Jarak Rata-Rata) Metode ini lebih efektif daripada metode hierarchical. Namun, dibandingkan dengan metode lain, seperti metode hierarchical yang digunakan dalam diskusi, metode ini harus dibayar dengan kecepatan komputasi yang jauh lebih lambat:

$$d_{uv} = Average\{d_{uv}\}, d_{uv} \in D \quad (2.7)$$

Keterangan :

D = Kriteria

d_{uv} = Data

\in = Eliminasi

2.6 Davis Bouldin Indeks

Davies Bouldin Index (DBI) diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 adalah metrik untuk mengevaluasi hasil algoritma clustering. *Davies-Bouldin Index (DBI)* adalah ukuran tertentu yang digunakan untuk menentukan jumlah kluster terbaik setelah proses pengklasteran selesai. Tujuan pendekatan DBI ini adalah untuk meminimalkan crosstalk antara dua kluster yang berdekatan dan memaksimumkannya antara dua kluster yang berdekatan. Nilai DBI yang lebih kecil yang digunakan (yang non-negatif) menghasilkan cluster yang lebih baik yang dihasilkan oleh algoritma pengelompokan K-mode yang sedang digunakan. Rumus Davies-Bouldin tercantum di bawah ini. (Az-zahra dkk., 2021)

$$DBI = \frac{1}{K} \times \sum_A^K =_1 R_a \quad (2.4)$$

Dengan

$$R_a = \max_{a \neq b} R_{ab} \text{ dan } R_{ab} = \frac{s_a + s_b}{d(v_a, v_b)} \quad (2.5)$$

Keterangan:

DBI = *Indeks Davies-bouldin*

K = Jumlah kluster

R_{ab} = Ukuran kemiripan antara kluster ke-a dan kluster ke-b

s_a = Ukuran dispersi kluster ke-a

s_b = Ukuran dispersi kluster ke-b

a = 1,2,.....,k

b = 1,2,.....,k

2.7. Penelitian Terdahulu

Berikut ada beberapa penelitian terdahulu yang menjadi referensi pada penelitian analisis perbandingan algoritma k-means dengan hierarchical clustering dalam penggolompokan data penduduk indeks pembangunan manusia kecamatan percut sei tuan yaitu:

Tabel 2.1 Penelitian Terdahulu

No.	Penulis (Tahun)	Topik	Hasil
1.	(Sikana & Wijayanto, 2021)	Analisa perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering	Hasil perhitungan menunjukkan, Metode ini memberikan Silhoutte Score sebesar 0,6291, Calinski-Harabasz Index Dengan nilai Silhoutte Score sebesar 0,5511, Calinski-Harabasz Index sebesar 1525,4007, dan Davies-Bouldin Index sebesar 0,5234, metode terbaik untuk mengelompokkan kabupaten/kota di Indonesia berdasarkan Indeks Pembangunan Manusia tahun 2019 adalah K-Means Clustering. Jumlah kluster optimum yang digunakan adalah 6, menunjukkan performa yang baik dalam evaluasi berbagai metrik tersebut.
2.	(Kamila dkk., 2019)	Perbandingan Algoritma K-	Pemrosesan menggunakan

		Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau	metode K-Means hanya memerlukan waktu rata-rata 1 detik, sementara pemrosesan data pada K-Medoids memakan waktu rata-rata 1 menit 38 detik di platform RapidMiner. Dalam hal nilai Davies-Bouldin Index (DBI), K-Means memiliki nilai lebih rendah dibandingkan dengan K-Medoids, yaitu 0,112 dibandingkan dengan 0,119, menunjukkan tingkat kompak dan separasi yang lebih baik dalam kluster K-Means. Hasil pengelompokan secara dominan menunjukkan bahwa agen Buana Listya Tama TBK, PT mendominasi, diikuti oleh agen Samudera Sarana Karunia, PT.
3.	(Yuliani & others, 2021)	Penerapan Data Mining untuk Mengeluster Data Penduduk Miskin Menggunakan Algoritma KMeans di Dusun Bagik Endep Sukamulia Timur	hasil dengan Cluster 1 berjumlah 18 penduduk dengan kriteria Penduduk ekonomi tinggi, Cluster 2 berjumlah 72 Penduduk dengan kriteria

			Penduduk ekonomi sedang, dan Cluster 3 berjumlah 110 penduduk dengan kriteria Penduduk ekonomi rendah.
4.	(Firdaus dkk., 2019)	Pengelompokan data persediaan obat menggunakan perbandingan metode k-means dengan hierarchical clustering single linkage	Hasil dari penelitian ini menunjukkan bahwa algoritma HCC Single linkage mampu memberikan hasil yang terbaik dengan validitas Sillhoutte Index (SI) sebesar 0.8629 sedangkan algoritma K-Means mendapatkan nilai validitas SI sebesar 0.8414
5.	Miranti Alysha Zulia Larasati, Nurul Anisa Sri Winarsih, Muhammad Syarifur (2020)	Penerapan Metode K-Means Clustering dalam Menganalisa Sentimen Masyarakat Terhadap K-Popers Pada Twitter	menghasilkan nilai Silhouette index 0.687974 yang artinya dataset tersebut digolongkan pada struktur yang baik

BAB III

METODE PENELITIAN

3.1 Lokasi Penelitian

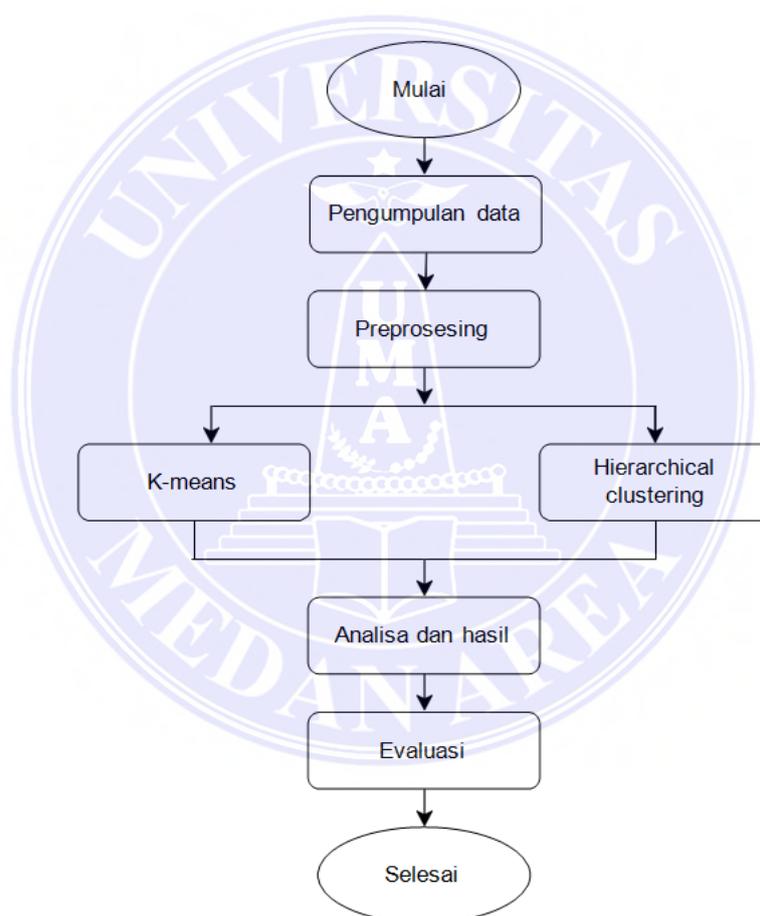
Percut Sei Tuan adalah satu-satunya desa di Provinsi Sumatera Utara Kabupaten Deli Serdang. Kecamatan Percut Sei Tuan memiliki sekitar 20 desa, termasuk Amplas, Bandar Khalipah, Bandar Klippa, Bandar Setia, Cinta Damai, Cinta Rakyat, Kenangan, Kenangan Baru, Kolam, Laut Dendang, Medan Estate, Pematang Lalang, Percut, Saentis, Sambirejo Timur, Sampali, Sei Rotan, Tanjung Rejo, Tanjung Selamat, dan Tembung. Pada tahun 2005, Balai Penyuluhan Pertanian Kecamatan mulai berfungsi sebagai tempat pelatihan untuk penyuluh dan pelaku utama, berikut lokasi peta penelitian yang dilakukan dapat dilihat pada Gambar 3.1.



Gambar 3.1 Lokasi Penelitian

3.2 Tahapan penelitian

Pada penelitian ini menggunakan prosedur kerja yang meliputi beberapa prosedur yang di mulai dengan melakukan persiapan penelitian, pengumpulan data, pengolahan data serta pengambilan keputusan data (analisa atau hasil). Untuk lebih jelas, berikut tahapan penelitian yang dilakukan dapat dilihat pada Gambar 3.2.



Gambar 3.2 Diagram dalam tahapan penelitian

3.3 Pengumpulan Data

Data yang dikumpulkan dalam penelitian ini didasarkan pada dokumen data penduduk yang setiap desa memiliki jumlah penduduk yang dihasilkan dari publikasi Statistik Indonesia dan diproses dari hasil Survei Sosial Ekonomi nasional/Susenas, Badan Pusat Statistik Nasional. Data yang digunakan dalam penelitian ini adalah data dari tahun 2021-2023. Pada penelitian ini, peneliti akan mengambil dataset dari kantor camat percut sei tuan atau sebuah situs website resmi yang telah mengeluarkan banyak dataset, yang dimana data tersebut bertujuan untuk dilakukan sebuah penelitian.

Sampel data yang akan digunakan pada penelitian perbandingan algoritma k-means dengan hierarchial clustering dalam pengelompokan penduduk kecamatan percut sei tuan meliputi termasuk jumlah penduduk Dengan jumlah sample data yang digunakan dalam penelitian ini adalah data penduduk dari kecamatan percut sei tuan. Dimana sebanyak 20 data penduduk di ambil sample data dan berapa jumlah penduduknya setiap desa. Data yang digunakan adalah jumlah penduduk setiap desa di kecamatan percut sei tuan.

3.4 Preprocessing

Preprocessing adalah Data preprocessing terbagi menjadi dua bagian utama, yaitu data cleaning dan data reduction. Data cleaning mencakup pemeriksaan data untuk menghindari redundansi dan inkonsistensi, serta menghilangkan variabel yang tidak dapat dihitung. Sementara itu, data reduction merupakan proses untuk menghapus data pada atribut yang kurang dominan, sehingga jumlah data dapat dikurangi, namun tetap menghasilkan data yang akurat.

3.4.1 Atribut Data

Pada penelitian ini, atribut yang digunakan adalah umur (A1), kesehatan (A2), pendidikan, dan pendapatan. Atribut umur dan pendapatan merupakan atribut yang bersifat numerikal, yang artinya data berupa angka atau nominal. Sedangkan atribut kesehatan dan pendidikan bersifat kategorikal, sehingga atribut tersebut perlu untuk di *action* agar menjadi dalam bentuk numerikal

Tabel 3.1 Variabel

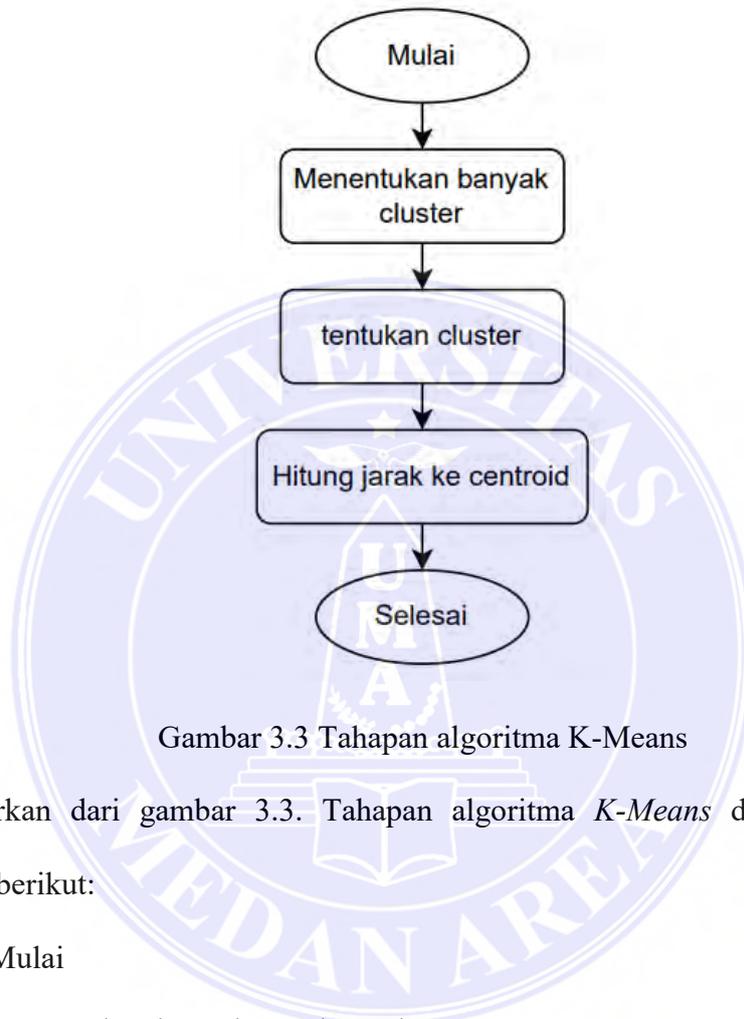
No	Atribut		Keterangan
1	K1	Umur	Numerikal (tahun)
2	K2	Pendapatan	Numerikal (Rupiah)
3	K3	Jenis Kelamin	Numerikal (angka)

Tabel 3.2 Data

No	Nama	Jenis Kelamin	Umur	Pendapatan
1	Ade kurniawan	L	25	5.000.000
2	Muhammad agus	L	48	2.500.000
3	Anggi fadilah	P	48	7.000.000
4	Bagas satria	L	35	1.500.000
....
997	Lala saputri	P	20	6.000.000
998	Maya sari	P	33	9.000.000
999	Omar simbolon	L	31	1.000.000
1000	Samuel sinaga	L	50	2.000.000

3.5. Algoritma *k-means*

Setelah data siap diolah, proses algoritma K-Means tertuang dalam flowchart dibawah ini



Gambar 3.3 Tahapan algoritma K-Means

Berdasarkan dari gambar 3.3. Tahapan algoritma *K-Means* dapat dijabarkan sebagai berikut:

- a. Mulai
- b. Menentukan banyaknya cluster k

Penentuan banyaknya klaster pada penelitian ini, klaster dibagi menjadi 4 atribut yaitu umur, kesehatan, pendidikan, pendapatan, dengan mengklaster serta menyesuaikan banyaknya klaster

- c. Menentukan cluster (titik centroid)

cluster terdiri dari C1, C2, C3:

di ambil data ke 5 sebagai pusat cluster ke-1 sebagai tingkat

pendapatan sedang

di ambil data ke 17 sebagai pusat cluster ke-2 sebagai tingkat

pendapatan rendah

di ambil data ke 9 sebagai pusat cluster ke-3 sebagai tingkat

pendapatan tinggi

d. Menghitung jarak data ke centroid

Tahap selanjutnya adalah menghitung jarak data ke titik centroid

dengan menggunakan Euclidean Distance atau jarak terdekat untuk

mendapatkan hasil data masuk ke klaster C1, C2, C3.

Iterasi pertama jarak data ke centroid C1

$$(d1,c1) = \sqrt{((25-32)^2 + (2-2)^2 + (5-2)^2 + (0.22-0.44)^2)}$$

$$= 7.62$$

$$(d2,c1) = \sqrt{((48-32)^2 + (1-2)^2 + (4-2)^2 + (0.17-0.22)^2)}$$

$$= 16.16$$

$$(d3,c1) = \sqrt{((48-32)^2 + (2-2)^2 + (5-2)^2 + (0.67-0.22)^2)}$$

$$= 16.28$$

$$(d4,c1) = \sqrt{((35-32)^2 + (2-2)^2 + (2-2)^2 + (0.06-0.22)^2)}$$

$$= 3.00$$

$$(d5,c1) = \sqrt{((32-32)^2 + (2-2)^2 + (2-2)^2 + (0-0.22)^2)}$$

$$= 0.22$$

(Seterusnya sampai n data).

Iterasi pertama jarak data ke centroid C2

$$(d1,c2) = \sqrt{((25-20)^2 + (2-2)^2 + (5-5)^2 + (0.44-0)^2)}$$

$$= 2.83$$

$$(d2,c2) = \sqrt{((48-20)^2 + (1-2)^2 + (4-5)^2 + (0.17-0)^2)}$$

$$= 2.24$$

$$(d3,c2) = \sqrt{((48-20)^2 + (2-2)^2 + (5-5)^2 + (0.67)^2)}$$

$$= 25.02$$

$$(d4,c2) = \sqrt{((35-20)^2 + (2-2)^2 + (2-5)^2 + (0.06-0)^2)}$$

$$= 44.19$$

$$(d5,c2) = \sqrt{((32-20)^2 + (2-2)^2 + (2-5)^2 + (0-0)^2)}$$

$$= 24.01$$

(Seterusnya sampai n data).

Iterasi pertama jarak data ke centroid C3

$$(d1,c3) = \sqrt{((25-20)^2 + (2-2)^2 + (5-5)^2 + (0.44-0.56)^2)}$$

$$= 5.00$$

$$(d2,c3) = \sqrt{((48-20)^2 + (1-2)^2 + (4-5)^2 + (0.17-0.56)^2)}$$

$$= 28.04$$

$$(d1,c3)_- = \sqrt{((48-20)^2 + (2-2)^2 + (5-5)^2 + (0.67-0.56)^2)}$$

$$= 28.00$$

$$(d4,c3) = \sqrt{((35-20)^2 + (2-2)^2 + (2-5)^2 + (0.06-0.56)^2)}$$

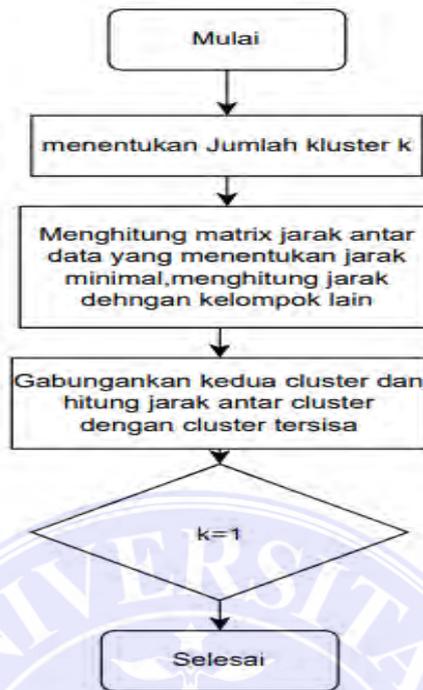
$$= 15.31$$

$$(d5,c3) = \sqrt{((32-20)^2 + (2-2)^2 + (2-5)^2 + (0-0.56)^2)}$$

$$= 12.38$$

3.5 Algoritma hierarchical clustering

Data sampel yang akan digunakan untuk perhitungan manual hierarchial clustering adalah 20% dari populasi yaitu sebanyak 20 data. Proses algoritma hierarchial clustering tertuang dalam flowchart dibawah ini:



Gambar 3.4 Tahapan Hierarchical clustering

Berdasarkan flowchart penelitian pada gambar 1 dapat dijelaskan sebagai berikut:

- a. Mulai
- b. Matriks Jarak antar data

$$(P1) = \sqrt{(25-48)+(2-1)^2+(4-5)+(0.44-0.17)}$$

$$(P2) = \sqrt{(48-48)+(1-2)^2+(4-5)+(0.17-0.67)}$$

$$(P3) = \sqrt{(48-35)+(2-2)+(2-2)+(0.67-0.06)}$$

$$(P4) = \sqrt{(32-35)+(2-2)^2+(2-2)+(0-0.06)}$$

$$(P5) = \sqrt{(32-60)+(2-1)^2+(2-1)+(0-0.11)}$$

$$(P7) = \sqrt{(60-21)+(1-1)^2+(1-5)+(0.11-1.00)}$$

$$(P8) = \sqrt{(21-20)+(1-2)^2+(5-4)+(1.00-0.22)}$$

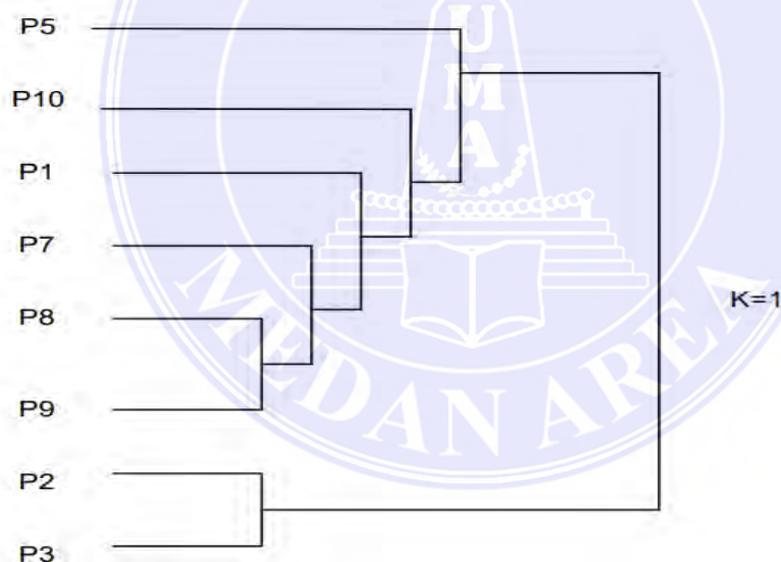
$$(P9) = \sqrt{(20-20)+(2-5)^2+(4-5)+(0.22-1.00)}$$

$$(P10) = \sqrt{(20-31)+(2-2)^2+(5-2)+(1.00-0.11)}$$

Tabel 3.3 Menghitung Jarak Cluster

	P1	P2	P3	P5	P7	P8	P9	P10
P1	0							
P2	23.04511142	0						
P3	23.00107351	1.5	0					
P4	10.44754682	13.1914	13.3557					
P5	7.628730619	16.1564	16.2925	0				
P6	35.24359674	12.3694	12.7007	28.0359117				
P7	4.160365606	27.0314	27.0206	14.071357	0			
P8	5.10385959	28.0179	28.0214	15.3587752	1.89867	0		
P9	5.030769521	28.0481	28.002	15.1336637	1.41421	1.26686	0	
P10	6.71648056	17.1465	17.2716	5.0884158	10.5257	11.1809	11.4364	0

c. Menggabungkan dari kedua kluster dan hitung jarak antar kluster dengan kluster tersisa. Penggabungan kluster diambil dari nilai kelompok paling minimum pada matriks sehingga tersisa hanya satu kluster tunggal ($K=1$).



Gambar 3.5 Menghitung Jarak Menggunakan Dedogram

d. Selesai

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dari pembahasan dan uji coba yang dilakukan, maka penelitian ini dapat disimpulkan:

1. analisis perbandingan algoritma *k-means* memiliki hasil 0.41 sedangkan hierarchical clustering 0.38 .
2. Hasil cluster dari setiap algoritma pada k-means memiliki 4 kluster terdiri dari:
Kluster 0: 175 anggota, Kluster 1: 212 anggota, Kluster 2: 281 anggota, Kluster : 262 anggota
3. Hasil cluster dari setiap algoritma pada hierarchial memiliki 4 kluster terdiri dari:
Jumlah Anggota Setiap Cluster: {0: 440, 1: 241, 2: 120, 3: 199}
5. K-means dapat pemilihan jumlah kluster sebelumnya (k), yang bisa menjadi tantangan terutama jika tidak ada pengetahuan awal tentang struktur data
4. Hierarchical clustering Tidak memerlukan jumlah kluster yang ditentukan sebelumnya. Pemilihan kluster dapat dilakukan dengan mengamati dendrogram.

5.2 Saran

Saran untuk mengembangkan pada penelitian ini dimasa yang akan datang sebagai berikut:

1. Tambah data pada suatu kluster untuk mempertimbangkan dampak penambahan data pada kluster agar hasil clustering makin akurat.
2. di buat aplikasi berbasis web untuk indeks pembangunan manusia (IPM) agar di situs web dapat memberikan sejumlah manfaat dalam konteks menyajikan informasi dan pemahaman yang lebih baik terkait perkembangan manusia di suatu wilayah atau negara.
3. penambahan atribut diartikan secara luas sebagai karakteristik atau kelompok-kelompok tertentu yang muncul melalui proses clustering di karenakan atribut yang digunakan sangat dikit.

DAFTAR PUSTAKA

- Adha, R., Nurhaliza, N., Sholeha, U., & Mustakim, M. (2021). Perbandingan algoritma DBSCAN dan k-means clustering untuk pengelompokan kasus Covid-19 di dunia. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, 18(2), 206–211.
- Az-zahra, A. A., Marsaoly, A. F., Lestyani, I. P., & Salsabila, R. (2021). Penerapan Algoritma K-Modes Clustering Dengan Validasi Davies Bouldin Index Pada Pengelompokan Tingkat Minat Belanja Online Di Provinsi Daerah Istimewa Yogyakarta. *Jurnal Matematika Dan Statistika Serta Aplikasinya Vol*, 9(1).
- Azzahra, A., & Wijayanto, A. W. (2022). Perbandingan Agglomerative Hierarchical dan K-Means dalam Pengelompokan Provinsi Berdasarkan Pelayanan Kesehatan Maternal. *SISTEMASI: Jurnal Sistem Informasi*, 11(2), 481–495.
- Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. Deepublish.
- Dani, A. T. R., Wahyuningsih, S., & Rizki, N. A. (2019). Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu. *Jambura Journal of Mathematics*, 1(2), 64–78.
- Febriyani, A., & Anis, A. (2021). Pengaruh Pertumbuhan Ekonomi, Investasi dan Indeks Pembangunan Manusia Terhadap Ketimpangan Distribusi Pendapatan di Indonesia. *Jurnal Kajian Ekonomi Dan Pembangunan*, 3(4), 9–16.
- Firdaus, R. D., Laksana, T. G., & Ramadhani, R. D. (2019). Pengelompokan Data Persediaan Obat Menggunakan Perbandingan Metode K-Means Dengan

- Hierarchical Clustering Single Linkage. *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, 2(1), 33–48.
- Hajar, S., Novany, A. A., Windarto, A. P., Wanto, A., & Irawan, E. (2020). Penerapan K-Means Clustering pada ekspor minyak kelapa sawit menurut negara tujuan. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 1(1), 314–318.
- Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(1), 76–88.
- Imantika, D., Bachtiar, F. A., & Rokhmawati, R. I. (2019). Penerapan metode k-means clustering dan analytical hierarchy process (ahp) untuk pengelompokan kinerja guru dan karyawan pada sma brawijaya smart school. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN, 2548, 964X*.
- Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119–125.
- Kharis, S. A. A., & Zili, A. H. A. (2022). Learning Analytics dan Educational Data Mining pada Data Pendidikan. *Jurnal riset pembelajaran matematika sekolah*, 6(1), 12–20.
- Kiha, E. K., SERAN, S., & LAU, H. T. (2021). Pengaruh jumlah penduduk,

pengangguran, dan kemiskinan terhadap indeks pembangunan manusia (ipm) di kabupaten belu. *Jurnal Ekonomi, Sosial & Humaniora*, 2(07), 60–84.

Mahroji, D., & Nurkhasanah, I. (2019). Pengaruh Indeks Pembangunan Manusia Terhadap Tingkat Pengangguran Di Provinsi Banten. *Jurnal Ekonomi-Qu*, 9(1).

Mahuze, W., Masinambow, V., & Lopian, A. (2022). Pengaruh pengeluaran pemerintah pada bidang pendidikan dan kesehatan terhadap indeks pembangunan manusia di kabupaten boven digoei. *Jurnal Berkala Ilmiah Efisiensi*, 22(3).

Marlin, S., Mus, A. R., & Junaid, A. (2022). Pengaruh Belanja Fungsi Pendidikan, Kesehatan, Perumahan dan Fasilitas Umum terhadap Indeks Pembangunan Manusia. *Jurnal Ekonomika*, 6(2), 210–224.

Mongan, J. J. S. (2019). Pengaruh pengeluaran pemerintah bidang pendidikan dan kesehatan terhadap indeks pembangunan manusia di Indonesia. *Indonesian Treasury Review: Jurnal Perbendaharaan, Keuangan Negara Dan Kebijakan Publik*, 4(2), 163–176.

Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100–108.

Nasution, I., Windarto, A. P., & Fauzan, M. (2020). Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi. *Building of Informatics, Technology and Science (BITS)*, 2(2), 76–83.

Parlambang, b., & fauziah, f. (2020). Implementasi algoritma k-means dalam

proses penilaian kuesioner kepada dosen guna mendukung kepuasan mahasiswa terhadap dosen. *Jurnal ilmiah teknologi dan rekayasa*, 25(2), 161–173.

Patadang, T. M., Rotinsulu, T. O., & Rorong, I. P. F. (2021). Pengaruh Pendapatan Asli Daerah (PAD), Dana Bagi Hasil (DBH), Dana Alokasi Umum (DAU), Dan Dana Alokasi Khusus (DAK) Terhadap Indeks Pembangunan Manusia (IPM) Minahasa Tenggara Dengan Pengeluaran Di Bidang Kesehatan Sebagai Variabel Intervening. *Jurnal Pembangunan Ekonomi Dan Keuangan Daerah*, 22(2), 132–149.

Prastiwi, H., Pricilia, J., & Rasywir, E. (2022). Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, 2(1), 141–148.

Pratama, A., Arisandi, D., Perdana, N. J., & others. (2020). Implementasi Metode Agglomerative Hierarchical Clustering Pada Website Pemilihan Tempat Futsal Studi Kasus Kota Depok. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 8(1), 1–6.

Pratiwi, N. P. A., & Indrajaya, I. G. B. (2019). Pengaruh Pertumbuhan Ekonomi dan Pengeluaran Pemerintah Terhadap Penyerapan Tenaga Kerja Serta Kesejahteraan Masyarakat di Provinsi Bali. *Buletin Studi Ekonomi*, 24(2).

Rohmah, A., Sembiring, F., & Erfina, A. (2021). Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: Smk Yaspim Gegerbitung). *Seminar Nasional Sistem Informasi Dan Manajemen Informatika Universitas Nusa Putra*, 1(01), 290–

298.

Sabilla, T. M., & Sumarsono, H. (2022). Pengaruh belanja pemerintah, pendapatan asli daerah, penanaman modal dalam negeri, indeks pembangunan manusia terhadap PDRB. *Forum Ekonomi*, 24(1), 54–64.

Safira, D., Mustakim, M., Lestari, E. D., Iffa, M., & Annisa, S. (2020). Pengelompokan Jumlah Penduduk Sumatera Barat Berdasarkan Angkatan Kerja Menggunakan Algoritma K-Means. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 6(1), 26–31.

Sikana, A. M., & Wijayanto, A. W. (2021). Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering. *Jurnal Ilmu Komputer*, 14(2), 66–78.

Sonang, S., Purba, A. T., & Pardede, F. O. I. (2019). Pengelompokan jumlah penduduk berdasarkan kategori usia dengan metode k-means. *Jurnal Tekinkom (Teknik Informasi Dan Komputer)*, 2(2), 166–172.

Utami, F. P. (2020). Pengaruh indeks pembangunan manusia (IPM), kemiskinan, pengangguran terhadap pertumbuhan ekonomi di Provinsi Aceh. *Jurnal Samudra Ekonomika*, 4(2), 101–113.

Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437–444.

Widodo, E., Mashita, S. N., & Prasetyowati, Y. G. (2020). Perbandingan Metode Average Linkage, Complete Linkage, dan Ward'S pada Pengelompokan

Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia. *Faktor Exacta*, 13(2), 81–87.

Yuliani, R., & others. (2021). Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur. *Infotek: Jurnal Informatika Dan Teknologi*, 4(1), 39–50.



LAMPIRAN

Lampiran SK Pembimbing

 **UNIVERSITAS MEDAN AREA**
FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax.(061) 7366998 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 756/FT.6/01.10/X/2023 23 Oktober 2023
Lamp : -
Hal : **Perubahan Judul Tugas Akhir**

Yth, Pembimbing Tugas Akhir
Andre Hasudungan Lubis, S. Ti, MSc
di
Tempat

Dengan hormat, Sehubungan dengan adanya perubahan judul tugas akhir maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa tersebut :

N a m a : Zatria Muhammad
N P M : 198160071
Jurusan : Teknik Informatika

Maka dengan hormat kami mengharapkan kesediaan saudara :

Andre Hasudungan Lubis, S. Ti, MSc (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :

"Analisis Perbandingan Algoritma K-Means dan Hierarchial Clustering untuk Pengelompokkan Data Penduduk dalam Indeks Pembangunan Manusia".

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.


Dr. Ratumanan, S. Kom, M. Kom



Lampiran surat pengambilan data

 **UNIVERSITAS MEDAN AREA**
FAKULTAS TEKNIK
Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 (061) 7366578, 7360168, 7364348, 7366781, Fax.(061) 7366998 Medan 20223
Kampus II : Jalan Setabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A (061) 8225602, Fax. (061) 8225331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 751 /FT.6/01.10/X/2023 23 Oktober 2023
Lamp : -
Hal : Penelitian Dan Pengambilan Data Tugas Akhir

Yth. Kepala Badan Kesatuan Bangsa dan Politik Kab. Deli Serdang
Jln. Karya Dharma No.4, Lubuk Pakam
Di
Deli Serdang

Dengan hormat,
Kami mohon kesediaan Bapak/Ibu berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	NAMA	NPM	PRODI
1	Zatria Muhammad	198160071	Teknik Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir pada perusahaan/Instansi yang Bapak/Ibu Pimpin.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan ilmiah dan Skripsi yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul penelitian :

Analisis Perbandingan Algoritma K-Means dan Hierarchical Clustering untuk Pengelompokan Data Penduduk pada Kantor Camat

atas perhatian dan kerja sama yang baik diucapkan terima kasih.

Dekan,

Syah, S. Kom, M. Kom

Tembusan :
1. Ka. BAMAI
2. Mahasiswa
3. File

Lampiran surat keterangan selesai penelitian

PEMERINTAH KABUPATEN DELI SERDANG
KECAMATAN PERCUT SEI TUAN
Alamat : Jl. Besar Tembung No. 22 Kode Pos – 20371
Telp. (061) 738008 Email : kecamatanpstuan@gmail.com website : www.percutsaeituankec.go.id

Percut Sei Tuan, 10 Desember 2023

Nomor : 070 /
Sifat : Biasa
Lampiran : -
Perihal : **SELESAI RISET**

Kepada Yth :
Dekan Universitas Medan Area
Fakultas Teknik
Di
Tempat

Sehubungan dengan Surat Izin Penelitian/Riset yang dilaksanakan di Kantor Camat Percut Sei Tuan, yang dilaksanakan oleh :

Nama : Zatria Muhammad
Alamat : Jalan Sempurna Dusun I Melati Desa Sambirejo Timur
NIM : 198160071
Pekerjaan : Mahasiswa
Jurusan : Teknik Informatika
Judul : Analisis Perbandingan Algoritma K-Means dan Hierarchical Clustering untuk Pengelompokan Data Penduduk Kantor Camat.
Daerah/Lokasi : Kantor Camat Percut Sei Tuan
Lama Penelitian : 1 (Satu) Bulan
Peserta : Sendiri
Penanggung Jawab : Dr. Rahmad Syah, S.Kom, M.kom

Bersama ini menerangkan bahwa Mahasiswa yang bersangkutan diatas benar telah selesai melaksanakan Riset di wilayah Kantor Camat Percut Sei Tuan.

Demikian surat ini disampaikan, untuk dapat dipergunakan sebagaimana mestinya.

PEMERINTAH KABUPATEN DELI SERDANG
KECAMATAN PERCUT SEI TUAN
SEKRETARIS
JANUARI, SE, M.S
Pangkat TK I
NIP.19780827 200502 1 004

Lampiran hasil plagiasi



Lampiran Source Code Google Colaboratory

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from sklearn.preprocessing import LabelEncoder
from sklearn.cluster import KMeans
from sklearn.datasets import make_blobs
from sklearn.cluster import AgglomerativeClustering
from scipy.cluster.hierarchy import dendrogram, linkage
import scipy.cluster.hierarchy as sch
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import silhouette_score
from google.colab import files

# Mengambil kolom 'Usia' dan 'Pendapatan'
dfKmeans = dfdata[['Usia', 'Jenis Kelamin', 'Pendapatan']]

# Menjalankan KMeans Clustering
Sum_of_squared_distances = []
K = range(1, 15)
for k in K:
    km = KMeans(n_clusters=k)
    km = km.fit(dfKmeans)
    Sum_of_squared_distances.append(km.inertia_)

# Plot Elbow Method
plt.plot(K, Sum_of_squared_distances, 'bx-')
plt.xlabel('k')
plt.ylabel('Sum_of_squared_distances')
plt.title('Elbow Method For Optimal k')
plt.show()

#Metrik keseluruhan Silhouette memberikan gambaran tentang
seberapa baik objek pada satu kluster sesuai dengan kluster
lainnya. Semakin tinggi nilai Silhouette, semakin baik
range_n_clusters = [2, 3, 4, 5]
for n_cluster in range_n_clusters:
    kmeans = KMeans(n_clusters = n_cluster, max_iter =1000)
    kmeans.fit(dfKmeans)
    labels = kmeans.predict(dfKmeans)
    from sklearn.metrics import silhouette_score
    silhouette_avg = silhouette_score(dfKmeans, labels)
    print(silhouette_avg)

```

```
#Nilainilai Silhouette yang Anda peroleh tampaknya
mengindikasikan bahwa model dengan 2 kluster memberikan hasil
yang paling baik,
#karena memiliki nilai Silhouette yang paling tinggi.
```

```
# Melakukan aktivasi algoritma K-Means
kmeans = KMeans(n_clusters=4)
kmeans.fit(dfKmeans)
```

```
data_c1, _ = make_blobs(n_samples=1000, centers=2,
random_state=42)
data_c2, _ = make_blobs(n_samples=1000, centers=2,
random_state=24)
kmeans = KMeans(n_clusters=4, random_state=42)
predicted_labels = kmeans.fit_predict(data)
```

```
for cluster_label in range(4):
    cluster_members = np.where(predicted_labels ==
cluster_label)[0]
    print(f"Klaster {cluster_label}: {len(cluster_members)}
anggota")
```

```
# Menghitung total cluster
total_cluster = len(np.unique(kmeans.labels_))
```

```
print(f"Total cluster: {total_cluster}")
```

```
plt.scatter(data[:, 0], data[:, 1], c=kmeans_labels,
cmap='viridis', s=30)
plt.title('K-Means Clustering')
plt.show()
```

```
# Memasukkan hasil label ke dalam suatu
data_clustered = pd.DataFrame(dfKmeans)
data_clustered['cluster'] = pd.DataFrame(kmeans.labels_)
data_clustered = data_clustered.sort_values('cluster')
data_clustered = np.asarray(data_clustered)
data_clustered
```

```
# Membuat similarity Matrix
from scipy.spatial.distance import cdist
dist = cdist(data_clustered[:,0:2],data_clustered[:,0:2],
metric='euclidean')
plt.imshow(dist)
plt.title("Pendapatan Similarity Matrix")
plt.show()
# Membentuk dendrogram dengan metode ward
```

```

dendrogram = sch.dendrogram(sch.linkage(dfkmeans,
method='ward'))
# Melakukan fitting hierarchical clustering
model = AgglomerativeClustering(n_clusters=4,
affinity='euclidean', linkage='ward')
model.fit(dfkmeans)
labels = model.labels_
labels

```

```

df['Cluster Hierarchical'] = pd.DataFrame(labels)
df.head()
# Membuat objek Hierarchical Clustering
hierarchical = AgglomerativeClustering(n_clusters=4 )
data, true_labels = make_blobs(n_samples=1000, centers=4,
random_state=42)
unique, counts = np.unique(labels, return_counts=True)
cluster_sizes = dict(zip(unique, counts))
print("Jumlah Anggota Setiap Cluster:", cluster_sizes)
dbscan = DBSCAN(eps=0.5, min_samples=5)
dbscan_labels = dbscan.fit_predict(data)

```

```

# Fungsi untuk mengevaluasi dan menampilkan hasil Silhouette
Score
def evaluate_silhouette_score(model, X, title):
    labels = model.fit_predict(X)
    silhouette_avg = silhouette_score(X, labels)
    print(f'{title} Silhouette Score: {silhouette_avg}')
    return silhouette_avg

```

```

kmeans = KMeans(n_clusters=4, random_state=42)
kmeans_silhouette = evaluate_silhouette_score(kmeans,
data_scaled, "K-Means")
agg_clustering = AgglomerativeClustering(n_clusters=4)
agg_silhouette = evaluate_silhouette_score(agg_clustering,
data_scaled, "Hierarchical")

```

```

plt.bar(labels, silhouette_scores, color=['blue', 'green'])
plt.title('Comparison of Silhouette Scores')
plt.xlabel('Clustering Algorithm')
plt.ylabel('Silhouette Score')
plt.show()

```