

BAB II

TINJAUAN PUSTAKA

2.1 Teori/Konsep Terkait

Subbab ini akan memberikan penjelasan secara jelas mengenai konsep-konsep dasar yang akan digunakan dalam melakukan penyusunan tugas akhir ini, dengan tujuan agar dapat dipahami.

2.1.1 Data Mining

Data Mining adalah ekstraksi pola secara sistematis dari kumpulan data besar. Data Mining mencakup berbagai tahapan termasuk seleksi, pembersihan, integrasi, transformasi, identifikasi pola, validasi model, dan tampilan pengetahuan, yang semuanya berkontribusi pada proses penemuan pengetahuan secara keseluruhan[8].

2.1.2 Clustering

Pengelompokan objek atau variabel yang memiliki karakteristik yang mirip disebut *clustering*. Aplikasi *clustering* dalam analisis data, pengenalan pola, pemrosesan gambar, pengambilan informasi, dan pencitraan medis sangat penting, jadi banyak metode komputasi telah digunakan untuk mempelajari proses ini. Teknik clustering yang paling sering digunakan adalah Hierarchical Clustering dan K-Means[9].

2.1.3 Hierarchical Clustering

Hierarchical Clustering adalah metode yang digunakan untuk mengorganisasikan data dengan cara menggabungkan objek-objek yang letaknya berdekatan satu sama lain, sehingga menghasilkan struktur yang hierarkis. Selanjutnya, dilanjutkan dengan objek lain yang memiliki celah lain. Dan seterusnya, terus berlanjut hingga kelompok-kelompok tersebut diorganisasikan ke dalam struktur hierarki seperti pohon, dengan tingkatan berbeda yang mewakili derajat kesamaan yang berbeda-beda. Pada akhirnya, semua objek akan berkumpul menjadi satu cluster dengan cara

yang logis. Dendrogram berguna untuk menjelaskan proses hierarki[10].

2.1.4 Agglomerative Hierarchical Clustering (AHC)

Agglomerative Hierarchical Clustering merupakan algoritma pengelompokan bottom-up yang menggabungkan sekumpulan titik data ke dalam cluster. *Agglomerative Hierarchical Clustering* mencakup beberapa pendekatan seperti *single linkage*, *complete linkage*, dan *average linkage*[11].

2.1.5 Euclidean Distance

Euclidean Distance adalah metode pemetaan objek yang menghitung jarak antara dua titik untuk mengidentifikasi kesamaan dalam analisis cluster. Jarak Euclidean dapat dihitung secara kuantitatif menggunakan rumus ini[12]:

$$D(x, y) = \sqrt{\sum_{k=1}^n (X_k - Y_k)^2} \quad [12]$$

Dimana:

$D(x, y)$ = Jarak Euclidean objek data ke-I dan objek data ke-j

n = banyak parameter yang digunakan

X_k = objek data ke-k

Y_k = objek data ke-y

2.1.6 Linkage

1. Metode *Single Linkage*

Pendekatan *Single Linkage* melibatkan pengelompokan berbagai hal berdasarkan jarak terdekat atau terkecilnya. Tahapannya melibatkan pengelompokan dua objek yang paling dekat satu sama lain ke dalam cluster awal. Proses ini diulangi hingga semua objek menyatu menjadi satu cluster. Sebagaimana dikemukakan oleh Johnson dan Wichern, pengelompokan dimulai dengan mengidentifikasi jarak terendah pada himpunan $D = \{d_{ij}\}$ dan menggabungkan objek-objek yang bersesuaian, dilambangkan dengan U dan V, untuk membentuk sebuah cluster (UV). Sebuah cluster baru terbentuk antara (UV) dan cluster lain, seperti W[13].

$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \quad [13]$$

Keterangan:

d_{UW} = jarak terdekat dari cluster U dan W

d_{VW} = jarak terdekat dari cluster V dan W

2. Metode *Average Linkage*

Average linkage merupakan clustering yang menghitung jarak rata-rata antara pasangan objek dalam satu *cluster* dan semua objek di *cluster* lain. *Average linkage* menghitung jarak antara dua *cluster* dengan mengambil rata-rata jarak yang diukur di dalam setiap cluster. Metode *Average linkage* dapat dihitung sebagai berikut[14]:

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)}N_W} \quad [14]$$

dimana:

d_{ik} = jarak antar objek ke-i dalam cluster (UV) dan objek k dalam cluster w.

N_{UV} = jumlah item pada cluster UV.

N_W = jumlah item pada cluster W.

3. Metode *Complete Linkage*

Metode *Complete Linkage*, disebut juga *linkage maksimum*, memastikan jarak antara dua *cluster* dengan menghitung jarak terjauh antara setiap anggota cluster yang satu cluster kedua. Langkah pertama adalah mencari nilai terkecil pada matriks D_{ij} , Selanjutnya objek-objek yang sesuai yaitu I dan J dikelompokkan menjadi satu sehingga membentuk cluster (IJ)[13]. Selanjutnya jarak antara cluster (IJ) dengan cluster K lainnya dihitung dengan menggunakan metode berikut:

$$d_{(IJ)K} = \max(d_{IK}, d_{JK}) \quad [13]$$

Keterangan:

d_{IK} = jarak terjauh dari cluster I dan K

d_{JK} = jarak terjauh dari cluster J dan K

2.1.7 Cophenetic Correlation Coefficient

Koefisien korelasi cophenetic umumnya digunakan dalam algoritma *Agglomerative Hierarchical Clustering* untuk menentukan kuantitas cluster yang dihasilkan.. Koefisien korelasi cophenetic berkisar antara 0 hingga 1. Nilai yang lebih besar atau mendekati 1 menunjukkan bahwa pendekatan korelasi yang digunakan cocok untuk clustering[12].

Berikut adalah rumus untuk perhitungan *cophenetic correlation coefficient*:

$$c = \frac{\sum_{i < j} (x(i,j) - \bar{x})(t(i,j) - t)}{[\sum_{i < j} (x(i,j) - \bar{x})^2] \sum_{i < j} (t(i,j) - t)^2} \quad [12]$$

Keterangan:

c = nilai cophenetic correlation coefficient

$x(i,j)$ = Euclidean distance antara cluster ke i dan j

$t(i,j)$ = dendogram dari cluster i dan j

\bar{x} = rata-rata dari $x(i,j)$

2.1.8 Python

Python adalah bahasa pemrograman berorientasi objek yang menggunakan templat skrip, terkadang dikenal sebagai bahasa skrip. Python adalah bahasa pemrograman sumber terbuka yang memungkinkan konfigurasi dan distribusi tidak terbatas. Ia mampu mengembangkan beragam aplikasi perangkat lunak dan kompatibel dengan banyak sistem operasi. Selain itu, perangkat lunak ini memiliki kode sumber terintegrasi, profiler, debugger, dan antarmuka yang melayani antarmuka, fungsi sistem, GUI, dan database[8].

2.1.9 Kelahiran

Kelahiran terdiri dari tiga tahap. Pembukaan jalan, keluarnya janin, dan keluarnya plasenta. Kematian selalu ada selama proses ini. "Kelahiran" dalam kamus kesehatan berarti ekspulsi lengkap atau ekstraksi hasil fertilisasi dari ibu, terlepas dari berapa lama kehamilan berlangsung. Ketika tali pusar telah dipotong atau tetap ada diplasenta, ibu harus tetap bernafas

atau menunjukkan tanda-tanda kehidupan seperti detak jantung, denyut nadi, atau gerakan otot setelah pemisahan[15].

2.1.10 Usia Produktif

Menurut World Health Organization (WHO), penduduk usia produktif adalah mereka yang berusia antara 15 dan 59 tahun. Namun, di Indonesia, istilah "penduduk usia produktif" sama dengan istilah "penduduk usia siap kerja." [16]. Dalam penelitian ini, kami mengambil usia produktif pada ibu dengan rentang usia 20 tahun sampai 39 tahun. Menurut website Badan Pusat Statistika Provinsi DKI Jakarta jumlah ibu dengan usia produktif dari tahun 2018-2021 adalah 8.910.381 Jiwa.

2.1.11 Microsoft Power BI

Salah satu program intelijen bisnis atau alat analitik bisnis adalah Microsoft Power BI, yang dapat meningkatkan wawasan organisasi. Microsoft Power BI memungkinkan analisis adhoc, mempermudah persiapan data, dan terhubung ke ratusan sumber data. Hasil laporan dapat ditampilkan di perangkat mobile dan web[17]. Power BI adalah gabungan dari berbagai jenis perangkat lunak, aplikasi, dan connector. Pengguna dapat memanfaatkan platform ini untuk menghasilkan laporan dan visualisasi, serta menggunakan referensi untuk tujuan analisis dan pengambilan keputusan. Power BI mengubah data kompleks menjadi data visual yang dinamis dan mudah dipahami[18].

2.2 Penelitian Terdahulu

Berikut adalah penelitian terdahulu yang dapat digunakan sebagai bahan referensi dalam penulisan ini.

Tabel 2.1 Literatur Review

No	Judul Artikel, Tahun Artikel, Penulis	Permasalahan	Metode Penelitian	Hasil Penelitian
1	IMPLEMENTASI ALGORITMA K-MEDOIDS CLUSTERING DALAM PENGELOMPOKAN ANGKA KELAHIRAN DI KOTA PEMANTANG SIANTAR, 2020, Sari Dewi Purba, Fitri Satria Tambunan, Heru Syahputra Saragih, Ilham Sudahri Damanik, Irfan Okprana, Harly Sistem Informasi, Jurusan Tunas bangsa Pematangsiantar, Stikom[7]	<ul style="list-style-type: none"> • angka kelahiran yang tinggi menjadi permasalahan utama kependudukan indonesia 	K-Medoids Clustering	<ul style="list-style-type: none"> • Sebanyak 43 desa diidentifikasi sebagai cluster rendah, dan 10 desa lainnya diidentifikasi sebagai cluster rendah berdasarkan perhitungan K-Medoids. • Untuk meningkatkan akurasi, proses implementasi menggunakan aplikasi RapidMiner 5.3 digunakan.
2	FAKTOR-FAKTOR YANG MEMPENGARUHI JUMLAH KELAHIRAN DI PROVINSI SUMATERA BARAT DENGAN MENGGUNAKAN ANALISIS FAKTOR, 2018, Febrina, Resti[4]	<ul style="list-style-type: none"> • Salah satu masalah yang belum teratasi di Sumatera Barat adalah kelahiran. • Di Sumatera Barat, angka kelahiran 	Analisis Faktor	<ul style="list-style-type: none"> • Hasil analisis data menunjukkan bahwa tiga faktor yang memengaruhi jumlah kelahiran di Provinsi Sumatera Barat adalah wanita dengan

		<p>sangat tinggi, dengan rata-rata seorang ibu melahirkan tiga hingga empat anak..</p> <ul style="list-style-type: none"> • Jumlah kelahiran yang meningkat, meningkatkan jumlah penduduk, yang berdampak pada kesejahteraan masyarakat. 		<p>pendidikan rendah atau sarjana, wanita yang menikah pada usia remaja akhir, wanita yang menikah di bawah umur, dan wanita yang pernah menggunakan alat kontrasepsi.</p>
3	<p>IMPLEMENTASI DATA MINING UNTUK PREDIKSI ANGKA KELAHIRAN BAYI PADA DESA SIBOLANGIT MENGGUNAKAN MULTI REGRESI, 2019, Y. Shara, I. Santoso, R. Kustini[2]</p>	<ul style="list-style-type: none"> • Angka kelahiran bayi di Desa Sibolangit terus meningkat setiap tahunnya. • Meskipun ada beberapa peningkatan, masalah status gizi kurang pada bayi di Desa Sibolangit masih belum diperbaiki. 	<p>Metode Multi Regresi</p>	<ul style="list-style-type: none"> • Penelitian ini menghasilkan data yang dapat digunakan sebagai acuan untuk menentukan jumlah kelahiran bayi. Kinerja dan waktu proses pengolahan data menjadi lebih efektif dan efisien, dan hasil prediksi jumlah kelahiran bayi

				menjadi lebih tepat dan akurat. .
4	<p>KLASTERISASI MENGGUNAKAN AGGLOMERATIVE HIERARCHICAL CLUSTERING UNTUK MEMODELKAN WILAYAH BANJIR, 2022, R. Praktiko, N. Damastuti[12]</p>	<ul style="list-style-type: none"> Selama musim hujan, banjir adalah bencana yang paling sering terjadi di provinsi Jawa Timur. Provinsi Jawa Timur mengalami 574 bencana banjir dari 2014 hingga 2015, menurut catatan Badan Penanggulangan Bencana (BNPB). Bencana banjir dapat disebabkan oleh banyak faktor, salah satunya adalah jumlah informasi yang lambat dikumpulkan. 	<p>Agglomerative Hierarchical Clustering (AHC)</p>	<ul style="list-style-type: none"> Hasil penelitian ini divisualisasikan dalam format SIG. Berdasarkan hasil uji kelompok optimal dengan metode elbow, provinsi Jawa Timur dibagi menjadi tiga kelompok dengan karakteristik rendah, sedang, dan tinggi. Hasil uji performa cluster menggunakan cophenetic correlation coefficient menunjukkan bahwa metode average linkage memberikan

				<p>solusi cluster yang lebih baik dibandingkan dengan metode AHC lainnya, dengan skor 0,92.</p> <ul style="list-style-type: none"> • Dengan menggunakan 950 data, pemodelan dampak banjir di wilayah jawa timur yang menggunakan agglomerative hierarchical clustering metode average linkage menunjukkan akurasi dengan cophenetic correlation coefficient sebesar 0,92488.
5	ANALISIS CLUSTER DENGAN ALGORITMA K-MEANS, FUZZY C-MEANS DAN HIERARCHICAL CLUSTERING (Studi Kasus: Indeks	<ul style="list-style-type: none"> • Pembangunan IPM tidak merata di setiap Kabupaten atau Kota. 	K-Means, Fuzzy C-Means dan Hierarchical Clustering.	<ul style="list-style-type: none"> • Hasil penelitian ini menunjukkan bahwa metode K-Means dan Hierarchical memiliki

	<p>Pembangunan Manusia tahun 2019), 2021, R. Rahmati, A. Wijayanto[10]</p>		<p>2 jumlah cluster optimal, dan metode Fuzzy C-Means memiliki 3 jumlah cluster optimal.</p> <ul style="list-style-type: none"> • Dalam penelitian ini, metode K-Means adalah metode terbaik setelah membandingkan nilai validasi dari ketiga metode Indeks • Nilai indeks Calinski Harabasz metode Fuzzy C-Means adalah 152,9 dengan cluster optimum 3 sedangkan metode K-Means memiliki nilai indeks yang lebih tinggi, yaitu 153,95. Namun, metode AGNES memiliki nilai indeks Calinski Harabasz yang lebih rendah daripada
--	--	---	--

				metode K-Means, yaitu 143,12.
6	ANALISIS PERTUMBUHAN DAN KEPADATAN PENDUDUK TERHADAP PERTUMBUHAN EKONOMI, 2021, D. Yunianto Badan Pusat Statistik Penajam Paser Utara, K. Timur[3]	<ul style="list-style-type: none"> Dengan banyaknya penduduk yang dipengaruhi oleh tinggi rendahnya laju pertumbuhan penduduk, pertumbuhan penduduk Samarinda dari tahun 2007 hingga 2017 terjadi secara fluktuatif, dengan pertumbuhan sebesar 3,28 persen pada tahun 2007 dan penurunan sebesar 2,22 persen pada tahun 2017. Dengan demikian, pertumbuhan penduduk Samarinda akan terus meningkat seiring dengan peningkatan populasi. 	Regresi Linier Berganda	<ul style="list-style-type: none"> Berdasarkan hasil data regresi dalam penelitian ini, dengan kenaikan sebesar 1%, kepadatan penduduk akan menurunkan pertumbuhan ekonomi Samarinda sebesar 22,333 persen. Pertumbuhan penduduk berdampak positif dan signifikan pada pertumbuhan ekonomi Samarinda. Dengan kata lain, pertumbuhan penduduk akan mendorong pertumbuhan ekonomi Samarinda. Kepadatan penduduk berdampak negatif

		<ul style="list-style-type: none"> ● Persebaran penduduk dan kepadatan penduduk yang tidak merata. ● Padatan penduduk Samarinda terus meningkat setiap tahunnya. Samarinda memiliki luas 11.603,94 km², dengan 16 orang tinggal di setiap 1 km². Namun, pada tahun 2017, 24 orang tinggal di setiap km². Jumlah penduduk yang terus meningkat setiap tahun memengaruhi peningkatan kepadatan penduduk tersebut. 		<p>dan signifikan terhadap pertumbuhan ekonomi Samarinda. Dengan kata lain, jika kepadatan penduduk meningkat, pertumbuhan ekonomi Samarinda akan menurun.</p>
--	--	--	--	--

7	<p>PENERAPAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING BERBASIS SINGLE LINKAGE UNTUK PENGELOMPOKAN JUDUL SKRIPSI, 2019, I. Irmayansyah, S. Khaosaroh[19]</p>	<ul style="list-style-type: none"> • Semakin banyak mahasiswa yang mengambil penelitian yang hampir identik dari segi tema, objek, atau metode dengan penelitian sebelumnya karena meningkatnya jumlah penelitian skripsi. 	<p>Hierarchical Agglomerative Clustering</p>	<ul style="list-style-type: none"> • Hasil penelitian menunjukkan bahwa proses perhitungan jarak antar data menghasilkan jarak terdekat sebesar 1, yang memungkinkan untuk mengetahui urutan kesesuaian metode pada subjek yang dipilih dengan hasil perhitungan. Hasil perhitungan kemudian diterapkan pada sistem pengajuan judul Skripsi. • Uji sistem pengajuan judul skripsi menunjukkan presentase tingkat kelayakan sebesar 82 persen.
---	---	---	--	---

8	<p>PENERAPAN HIERARCHICAL CLUSTERING METODE AGGLOMERATIVE PADA DATA RUNTUN WAKTU, 2019, Andrea Tri Rian DanI, Sri Wahyuningsih, Nanda Arista Rizki[20]</p>	<ul style="list-style-type: none"> • Meskipun Provinsi Kaltim memiliki sumber daya alam yang luar biasa, sebagian besar belum dimanfaatkan. Jumlah penduduk yang tinggi merupakan salah satu penyebab kemiskinan. • Jika jumlah penduduk terus meningkat tanpa mengikuti perkembangan faktor-faktor lainnya, tentu akan muncul masalah baru, seperti lapangan kerja yang kurang dan tingkat kemakmuran yang rendah. 	<p>Hierarchical Clustering Metode Agglomerative</p>	<ul style="list-style-type: none"> • Hasil penelitian menunjukkan bahwa jarak yang didasarkan pada autocorrelation based distance (ACF), dengan koefisien korelasi cophenetic sebesar 0,99, adalah jarak dengan pengukuran kemiripan terbaik untuk mengelompokkan Kabupaten dan Kota di Provinsi Kalimantan Timur. • Dengan jumlah cluster yang representatif berdasarkan koefisien silhouette sebesar 0,68, algoritma Average Linkage adalah algoritma
---	--	---	---	---

				pengelompokan terbaik.
9	ANALISA PERBANDINGAN METODE HIERARCHICAL DAN K-MEANS DALAM CLUSTERING DATA TERHADAP PENJUALAN JAJANSAMAVIVI, 2020, H. Setiawan, K, Karamitta, K.Hosea et al.[8]	<ul style="list-style-type: none"> • Untuk mempermudah penjual toko online jajansamavivi dalam menganalisa penjualan 	Metode Hierarchical dan K-Means Clustering.	<ul style="list-style-type: none"> • Penelitian ini didapatkan hasil toko online Jajansamavivi membagi penjualan menjadi tiga kelompok: kelompok penjualan rendah, kelompok penjualan menengah, dan kelompok penjualan tinggi. Tujuan dari penelitian ini adalah untuk meningkatkan penjualan kelompok penjualan rendah dengan menawarkan promosi atau diskon harga. • Penelitian ini menghasilkan Perbandingan penggunaan metode hierarchial dan k-

				means untuk mengelompokkan data penjualan di toko online Jajansamavivi dapat menghasilkan pembagian penjualan rendah, penjualan menengah, dan penjualan tinggi.
10	KOMPARASI ANALISIS K-MEDOIDS CLUSTERING DAN HIERARCHICAL CLUSTERING (Studi Kasus: Data Kriminalitas di Indonesia Tahun 2020), 2022, S.Faradilla[21]	<ul style="list-style-type: none"> • Kasus kriminalitas di Indonesia meningkat setiap tahun. • Jumlah kejadian kejahatan pada tahun 2018 sebanyak 294.281, menurut Badan Pusat Statistik, turun pada tahun 2019 sebanyak 269.324, dan pada tahun 2020 sebanyak 247.218, menurut Badan Pusat Statistik. Dari seluruh Polda/Provinsi, Polda Sumatera Utara 	K-Medoids Clustering dan Hierarchical Clustering	<ul style="list-style-type: none"> • Penelitian ini menyimpulkan, Dari berbagai jenis kejahatan yang ada di Indonesia, terdapat beberapa provinsi yang menonjol karena jumlah kasusnya. Kepolisian Sumatera Utara adalah provinsi dengan jumlah kasus tertinggi untuk masing-masing jenis kejahatan, yaitu kejahatan yang melibatkan nyawa,

		<p>memiliki jumlah kejadian tertinggi sebanyak 32.990, diikuti oleh Polda Metro Jaya (DKI Jakarta) sebanyak 1 26.585, dan Polda Jawa Timur sebanyak 17.642.</p> <ul style="list-style-type: none"> • Banyaknya tindakan kriminal atau kejahatan. 		<p>fisik, kemerdekaan, penggunaan kekerasan tanpa kekerasan, penggunaan kekerasan, narkoba, dan penipuan.</p> <ul style="list-style-type: none"> • Cluster 1 terdiri dari 28 provinsi dan cluster 2 terdiri dari 6 provinsi, menurut hasil cluster yang dihasilkan menggunakan algoritma k-medoids pada data kriminalitas Indonesia tahun 2020. • Cluster 1 terdiri dari 29 provinsi dan cluster 2 terdiri dari 5 provinsi, menurut hasil cluster yang dihasilkan menggunakan algoritma hierarchical
--	--	---	--	--

				<p>pada data kriminalitas Indonesia tahun 2020.</p> <ul style="list-style-type: none"> • Hasil perbandingan metode cluster menunjukkan bahwa algoritma clustering hierarki metode Ward adalah yang terbaik untuk data kriminalitas tahun 2020.
11	<p>PENERAPAN DATA MINING DALAM MENGELOMPOKKAN DATA RIWAYAT AKADEMIK SEBELUM KULIAH DAN DATA KELULUSAN MAHASISWA MENGGUNAKAN METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING, 2021, Banu Harli Trimulya Suandi As, Lisna Zahrotun[22]</p>	<ul style="list-style-type: none"> • Prosedur penerimaan mahasiswa baru di Fakultas Teknologi Industri Universitas Ahmad Dahlan Ketidakseimbangan yang signifikan antara jumlah siswa dan dosen terjadi karena banyak siswa yang masuk ke proses dan sedikit yang lulus tepat waktu. 	Hierarchical Clustering	<ul style="list-style-type: none"> • Dalam penelitian ini menghasilkan 158 data siswa yang direkomendasikan, yang semuanya berasal dari Pulau Jawa. Nilai matematika rata-rata untuk dataset Informatika, Industri, dan Elektro adalah lebih dari 80, dan

		<ul style="list-style-type: none"> • Jumlah siswa yang menggunakan fasilitas kampus melebihi kapasitas, sehingga pendidikan tidak efektif. 		<p>untuk dataset Kimia adalah lebih dari 67.</p> <ul style="list-style-type: none"> • Data yang disarankan diperoleh dalam penelitian ini, dengan jumlah data masing-masing 43, 24, 19, dan 72. • Nilai Silhouette Coefisien untuk program studi masing-masing 0,868, 0,883, 0,879, dan 0,873. Ini adalah hasil yang sangat baik untuk metode Silhouette Coefisien.
12	PENERAPAN AGGLOMERATIVE HIERARCHICAL CLUSTERING UNTUK SEGMENTASI PELANGGAN, 2020, W. Widyawati, W. Saptomo, Y. Utami[23]	<ul style="list-style-type: none"> • Dengan munculnya bisnis baru, perusahaan harus memiliki strategi pemasaran yang tepat untuk memenuhi 	Agglomerative Hierarchical Clustering	<ul style="list-style-type: none"> • Hasil penelitian menggunakan koefisien Silhouette menunjukkan bahwa hasil dari tujuh cluster adalah yang

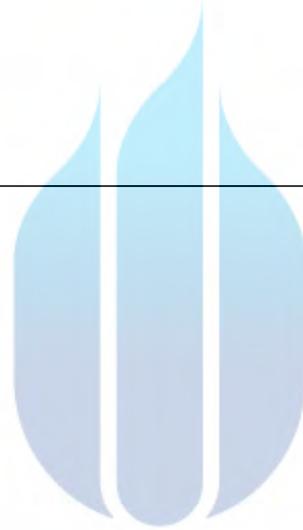
		<p>kebutuhan pelanggan mereka. Langkah pertama adalah memahami jenis pelanggan mereka dan membuat strategi pemasaran yang sesuai dengan mereka.</p>		<p>terbaik jika dibandingkan dengan hasil dari dua cluster hingga dua puluh cluster. Ini karena cluster - cluster ini memiliki nilai minus yang paling kecil.</p> <ul style="list-style-type: none"> • Untuk segmentasi pelanggan, tujuh cluster dibuat berdasarkan hasil koefisien Silhouette, dengan masing-masing cluster menggambarkan jenis pelanggan saat ini. • Menurut penelitian ini, cluster 1 memiliki 20 anggota, cluster 2 memiliki 43 anggota, cluster 3 memiliki 75 anggota, cluster 4 memiliki 158
--	--	---	--	--

				anggota, cluster 5 memiliki 9 anggota, cluster 6 memiliki 2 anggota, dan cluster 7 memiliki 1 anggota.
13	ANALISIS CLUSTER DENGAN MENGGUNAKAN METODE HIERARKI UNTUK PENGELOMPOKAN KECAMATAN DI KABUPATEN LANGKAT BERDASARKAN INDIKATOR KESEHATAN, 2019, N. Khairun[24]	<ul style="list-style-type: none"> • Tingkat kesehatan di Kabupaten Langkat menurun karena kurangnya kesadaran, kemauan, dan kemampuan untuk menjalani gaya hidup sehat. Ini ditunjukkan oleh banyaknya orang yang mengeluh tentang kesehatan mereka. 	Analisis Cluster dengan Metode Ward	<ul style="list-style-type: none"> • Hasil analisis cluster dengan metode ward diperoleh hasil 3 cluster kecamatan yang bergabung berdasarkan faktor yang mempengaruhi derajat kesehatan di Kabupaten Langkat, diantaranya angka kesakitan, penolong kelahiran dan angka harapan hidup, dimana 3 cluster tersebut terdiri dari cluster 1 beranggotakan 10 kecamatan dengan tingkat kesehatan, cluster 2

				beranggotakan 7 kecamatan, dan cluster 3 beranggotakan 6 kecamatan.
14	PEMETAAN KABUPATEN/KOTA DI JAWA TIMUR BERDASARKAN INDIKATOR KESEJAHTERAAN RAKYAT MENGGUNAKAN ANALISIS CLUSTER HIERARKI, 2022, I. Wahyuni, S.Wulandari[25]	<ul style="list-style-type: none"> • Belum ada pemetaan indikator kesejahteraan rakyat di Jawa Timur berdasarkan kabupaten atau kota. 	Analisis Cluster Hierarki	<ul style="list-style-type: none"> • Penelitian ini menghasilkan hasil analisis cluster hierarki. Dengan menggunakan indikator kesejahteraan rakyat, kelompok-kelompok di Jawa Timur dikelompokkan menjadi 4 kategori: sangat tinggi, tinggi, sedang, dan rendah. • Cluster 1 memiliki tingkat kesejahteraan masyarakat sedang, kcluster 2 rendah, cluster 3 tinggi, dan

				<p>cluster 4 sangat tinggi.</p> <ul style="list-style-type: none"> • Karena ketidaksetaraan yang lebih rendah dalam hal kesetaraan ekonomi, kesehatan, dan pendidikan, cluster 2 memiliki tingkat kesejahteraan rakyat rendah.
15	<p>HIERARCHICAL CLUSTERING BAHAN MENU DI KANTIN UNIVERSITAS UNTUK MENUNJANG IMPLEMENTASI SISTEM JAMINAN HALAL, 2021, Sucipto Sucipto, Mohammad Syarifuddin Al-Mubarak, Danang Triagus Setiyawan[26]</p>	<ul style="list-style-type: none"> • Dalam memenuhi kebutuhan pangan civitas akademika, kantin sangat diperlukan di perguruan tinggi. • Kantin yang meski harganya murah, perlu juga dijamin kehalalan menunya. 	<p>Hierarchical Clustering</p>	<ul style="list-style-type: none"> • Penelitian ini mendapatkan hasil yaitu, bahan dikelompokkan menjadi 2 cluster yakni bahan kritis dan bahan non kritis. • Hasil yang didapat menunjukkan pda kantin ada 262 menu dan 722 bahan. • Identifikasi bahan ada 28,95% bahan

				bersertifikat halal dan 71,05% bahan tidak bersertifikat halal yang terdiri atas 67,45% termasuk bahan non-kritis dan 3,55% bahan kritis.
--	--	--	--	---



UNIVERSITAS
MERCU BUANA

2.3 Analisis Literatur Review



Gambar 2.1 Grafik Literatur Review

Berdasarkan perbandingan dengan penelitian terdahulu yang ditulis oleh Fitri Sari Dewi Purba, Heru Satria Tambunan, Ilham Syahputra Saragih, Irfan Sudahri Damanik, Harly Okprana yang berjudul "IMPLEMENTASI ALGORITMA K-MEDOIDS CLUSTERING DALAM PENGELOMPOKAN ANGKA KELAHIRAN DI KOTA PEMANTANG SIANTAR" penelitian ini memperoleh 2 klaster, yaitu cluster rendah dan tinggi. Sumber data diperoleh dari Dinas Kependudukan dan Pencatatan Sipil Pemantangsiantar data yang diambil adalah data jumlah angka kelahiran pada tahun 2015-2019. Jumlah data yang digunakan terdiri dari 53 kelurahan yang dibagi menjadi 2 cluster yaitu rendah dan tinggi, dari hasil perhitungan k-medoids diperoleh sebanyak 43 kelurahan sebagai cluster rendah, 10 kelurahan sebagai cluster tinggi. Dari pemaparan penelitian sebelumnya didapati saran untuk melakukan perbandingan dari algoritma clustering yang lainnya, dan juga diharapkan menggunakan data yang baru dan lebih detail, oleh karena itu pada penelitian ini akan melakukan clustering menggunakan algoritma hierarchical clustering dengan data angka kelahiran DKI Jakarta dari tahun 2018-2023 dan dengan faktor berdasarkan usia produktif seorang ibu.

Adapun untuk keunggulan dari algoritma hierarchical clustering menurut Gupta dan Rastogi, Algoritma Hierarchical Clustering memiliki kelebihan sederhana, mudah dimengerti secara komprehensif, dan jumlah kelompok tidak mesti ditentukan dahulu. Pengembangan dari penelitian ini adalah menggunakan data angka kelahiran dari 2018-2023 yang diperoleh langsung dari sumber internal Pemerintah Provinsi Daerah Khusus Ibukota Jakarta yaitu Dinas Kependudukan Dan Pencatatan Sipil DKI Jakarta dengan menggunakan algoritma hierarchical clustering dan kami memilih faktor berdasarkan usia produktif seorang ibu dari banyaknya faktor-faktor yang menyebabkan angka kelahiran menjadi tinggi.

