

BAB I

PENDAHULUAN

1.1 Latar Belakang

Stunting (kerdil) adalah kondisi di mana balita memiliki tinggi atau panjang badan yang lebih rendah dibandingkan dengan anak seusianya. Kondisi ini diukur menggunakan indikator panjang atau tinggi badan yang berada di bawah dua standar deviasi dari median standar pertumbuhan anak menurut WHO. *Stunting* pada balita merupakan masalah gizi kronis yang timbul dari berbagai faktor, termasuk kondisi sosial ekonomi, status gizi ibu saat hamil, penyakit yang dialami bayi, serta kurangnya asupan gizi pada masa bayi. Anak yang mengalami *stunting* berisiko menghadapi hambatan dalam mencapai perkembangan fisik dan kognitif yang optimal di masa mendatang (Harjanto et al., 2021).

Kabupaten Kuningan merupakan wilayah administratif yang terletak di bagian paling timur Provinsi Jawa Barat, berbatasan langsung dengan wilayah Provinsi Jawa Tengah. Secara geografis, kabupaten ini juga berbatasan dengan beberapa wilayah lain, yaitu Kabupaten Ciamis di bagian selatan, Kabupaten Majalengka di sisi barat, dan Kabupaten Cirebon di sebelah utara. Secara astronomis, Kabupaten Kuningan terletak pada koordinat 06⁰47'2" hingga 07⁰12'2" Lintang Selatan dan 108⁰23'2" hingga 108⁰47'2" Bujur Timur. Dengan luas wilayah mencapai 1 194,09 km² Kabupaten Kuningan mencakup sekitar 2,52 % dari total luas Provinsi Jawa Barat yang memiliki wilayah seluas kurang lebih ($\pm$ 44 357,00 km²) (BPS Kabupaten Kuningan, 2024).

Kabupaten Kuningan merupakan salah satu wilayah di Indonesia dengan angka prevalensi balita *stunting* yang tinggi. Kondisi ini membuat Kabupaten Kuningan masuk dalam daftar 100 Kabupaten/Kota prioritas untuk penanganan *stunting* di tingkat nasional pada tahun 2018 (Muhammad Farhan Waliyudin et al., 2022). Pada tahun 2024, berdasarkan data dari data status gizi balita (*stunting*) berdasarkan indeks tb/u menurut kecamatan dan puskesmas,

tercatat sebanyak 6.115 balita di Kabupaten Kuningan mengalami stunting (Dinas Kesehatan Kabupaten Kuningan, 2024). Pemerintah Kabupaten Kuningan telah melaksanakan beberapa program untuk menangani stunting, termasuk Program Bapak Asuh Anak Stunting yang melibatkan TNI/Kodim 0615 Kuningan dan pengembangan potensi bahan pangan lokal. Namun, beberapa intervensi gizi sensitif di luar bidang kesehatan tidak sepenuhnya efektif (BAPPEDA Kabupaten Kuningan, 2023). Oleh karena itu, diperlukan upaya pemetaan wilayah stunting untuk mengidentifikasi daerah-daerah dengan prevalensi stunting yang tinggi secara lebih spesifik. Melalui pemetaan ini, dapat diperoleh informasi yang akurat mengenai distribusi kasus stunting di setiap kecamatan. Informasi tersebut akan menjadi dasar untuk menyusun strategi penanganan yang lebih terarah dan efektif, termasuk prioritas alokasi sumber daya, intervensi gizi, dan program edukasi yang sesuai dengan kebutuhan setiap wilayah.

Data mining dianggap sebagai hasil dari perkembangan alami dalam teknologi informasi. Tahap awal pengumpulan data dan pembuatan basis data memainkan peran penting dalam menciptakan mekanisme yang lebih canggih untuk penyimpanan data, pengambilan informasi, pemrosesan kueri, dan pengelolaan transaksi di kemudian hari. Seiring waktu, pengelolaan informasi dan pengetahuan berkembang melalui berbagai fungsi utama, seperti pengumpulan data, pengolahan informasi, pengelolaan pengetahuan, dan analisis data tingkat lanjut. Salah satu arsitektur repositori data yang semakin berkembang adalah data warehouse, yang mengintegrasikan berbagai sumber data heterogen dalam satu skema terpadu di lokasi yang sama untuk mendukung proses pengambilan keputusan (Lal & Singh, 2022).

Density-Based Spatial Clustering of Applications (DBSCAN) adalah algoritma yang termasuk dalam kategori clustering berbasis kepadatan, di mana proses pembentukan *cluster* ditentukan berdasarkan kepadatan atau kedekatan jarak antara satu objek dengan objek lainnya dalam dataset (Yanto et al., 2024). *DBSCAN* merupakan algoritma pengelompokan berbasis densitas yang tidak memerlukan penentuan jumlah *cluster* di awal, serta mampu

mengidentifikasi *cluster* dengan berbagai bentuk dan ukuran, sambil mengabaikan *noise* atau data *outlier* yang tidak relevan. Proses pembentukan *cluster* dimulai dari titik pertama yang diinput dalam data (Yanto et al., 2024).

Penelitian terdahulu yang telah dilakukan oleh Taufik Dwi Harjanto, Arie Vatesia, dan Ruvita Faurina (2021) dalam judul “Analisis Penetapan Skala Prioritas Penanganan Balita Stunting Menggunakan Metode DBSCAN Clustering” membahas penerapan algoritma DBSCAN untuk menentukan skala prioritas dalam penanganan balita stunting. Penelitian ini menggunakan indikator seperti jumlah dokter, perawat, ahli gizi, posyandu, balita penerima ASI eksklusif, balita yang memiliki kartu KIA/KMS, dan ketersediaan sarana air bersih. Dengan parameter DBSCAN epsilon sebesar 114 dan MinPts 2, hasil pengelompokan menghasilkan tiga cluster prioritas yang dievaluasi menggunakan Silhouette Coefficient sebesar 0,5157, menandakan hasil pengelompokan yang layak (Harjanto et al., 2021).

Penelitian serupa dilakukan oleh Ihsan Bagus Fahad Arafat dkk. (2023) dalam jurnal “Clustering Gempabumi di Wilayah Regional VII Menggunakan Pendekatan DBSCAN”, yang menerapkan DBSCAN untuk memetakan data gempa bumi di Jawa Tengah, Yogyakarta, dan Jawa Timur berdasarkan koordinat, magnitudo, dan kedalaman gempa. Hasil analisis menghasilkan 2 hingga 6 cluster dengan satu noise dan nilai Silhouette Coefficient terbaik sebesar 0,499, menunjukkan pengelompokan yang cukup baik untuk mendukung mitigasi bencana (Bagus et al., 2023).

Sementara itu, Esra Kristiani Sihite dkk., dalam jurnal “Pembangunan WebGIS untuk Penderita Gizi Buruk di Kota Medan Berdasarkan Hasil Clustering Algoritma DBSCAN”, menggunakan DBSCAN untuk memetakan data gizi buruk dari 21 kecamatan di Kota Medan. Dengan epsilon 0,5 dan MinPts 2, penelitian ini menghasilkan tiga cluster dengan tingkat kasus rendah, sedang, dan tinggi yang divalidasi menggunakan Silhouette Index (0,5414) dan Dunn Index (0,5124). Sistem WebGIS yang dihasilkan memungkinkan pemantauan data kesehatan yang interaktif dengan tingkat keberhasilan pengujian Blackbox Testing mencapai 100%, membuktikan efektivitas sistem

dalam mendukung pengambilan keputusan di bidang kesehatan masyarakat (Sihite et al., 2024).

Untuk mengatasi masalah *stunting*, salah satu strategi yang dapat dilakukan adalah melakukan pemetaan wilayah untuk mengidentifikasi daerah prioritas berdasarkan tingkat prevalensi *stunting*. Pemetaan ini bertujuan untuk mendukung pengambilan keputusan yang lebih tepat dalam penentuan intervensi dan alokasi sumber daya, seperti distribusi tenaga kesehatan, peningkatan jumlah posyandu, penyediaan akses air bersih, dan pemberian ASI eksklusif. Salah satu pendekatan yang dapat digunakan adalah teknik klasterisasi spasial menggunakan algoritma *DBSCAN*, yang dapat mengelompokkan wilayah berdasarkan tingkat prevalensi *stunting* dan fasilitas kesehatan. Dengan metode ini, wilayah dengan tingkat *stunting* tertinggi dapat diprioritaskan dalam program intervensi dan pencegahan.

Berdasarkan uraian yang telah dipaparkan tersebut, maka peneliti tertarik untuk melakukan penelitian dalam bentuk skripsi dengan menentukan judul yaitu **“PENERAPAN ALGORITMA DBSCAN UNTUK PEMETAAN WILAYAH STUNTING DI KABUPATEN KUNINGAN”**.

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang sebelumnya, dapat diidentifikasi beberapa permasalahan yang muncul, antara lain:

1. Kabupaten Kuningan memiliki angka prevalensi balita *stunting* yang signifikan, dengan 6.115 kasus tercatat pada tahun 2024. Hal ini menjadikan Kuningan salah satu daerah dengan tingkat *stunting* tertinggi di Indonesia dan masuk ke dalam 100 kabupaten/kota prioritas penanganan *stunting*.
2. Tidak adanya pengelompokan wilayah *stunting* yang memadai di Kabupaten Kuningan menyebabkan kesulitan dalam memahami persebaran kasus di setiap kecamatan dan puskesmas. Hal ini

menghambat proses penentuan wilayah prioritas untuk intervensi yang lebih terfokus.

3. Tidak adanya pemetaan yang komprehensif mengenai wilayah dengan prevalensi *stunting* tinggi menyulitkan pengambil kebijakan untuk memahami pola dan distribusi kasus, sehingga alokasi sumber daya menjadi kurang optimal.

1.3 Rumusan Masalah

Dari identifikasi masalah yang telah diuraikan pada poin di atas, maka rumusan masalah dalam penelitian ini yaitu:

1. Bagaimana merancang dan membangun aplikasi berbasis data untuk mendukung pemetaan wilayah dengan prevalensi *stunting* tinggi di Kabupaten Kuningan?
2. Bagaimana penggunaan algoritma *DBSCAN* dalam mengelompokkan wilayah berdasarkan tingkat *stunting* untuk mendukung prioritas intervensi?

1.4 Batasan Masalah

Adapun batasan dalam penelitian ini meliputi beberapa aspek berikut:

1. Penelitian ini mengambil lokasi studi kasus di Kabupaten Kuningan
2. Metode algoritma *DBSCAN* digunakan untuk menganalisis dan mengelompokkan wilayah berdasarkan tingkat prevalensi *stunting* dan fasilitas kesehatan.
3. Data yang akan dianalisis adalah data status gizi balita (*stunting*) berdasarkan indeks TB/U dan data fasilitas kesehatan dari tahun 2024.

4. Sistem yang dirancang berbasis Web untuk Dinas Kesehatan Kabupaten Kuningan.
5. Bahasa pemrograman yang digunakan yaitu *PHP*, *Python* dan *MySQL*.
6. Sistem mengelola data prevalensi stunting di setiap kecamatan dan fasilitas kesehatan untuk mendukung penentuan prioritas penanganan.
7. *Output* sistem hanya berupa desain dan laporan prioritas wilayah stunting.
8. Pemerintah Kabupaten Kuningan dapat mengelola data analisis prevalensi stunting dan laporan pemetaan wilayah stunting berdasarkan hasil klasterisasi.

1.5 Tujuan Penelitian

Berdasarkan permasalahan yang telah diidentifikasi sebelumnya, penelitian ini bertujuan untuk:

1. Membuat aplikasi pemetaan wilayah untuk mengidentifikasi daerah dengan prevalensi *stunting* tinggi di Kabupaten Kuningan berbasis Web.
2. Mengimplementasikan algoritma *DBSCAN* untuk menganalisis distribusi kasus *stunting* berdasarkan data fasilitas kesehatan dan prevalensi stunting di Kabupaten Kuningan.
3. Membantu Dinas Kesehatan Kabupaten Kuningan dalam memetakan wilayah prioritas penanganan stunting berdasarkan tingkat prevalensi dan jumlah fasilitas kesehatan.
4. Mendukung pengambilan keputusan terkait alokasi sumber daya, intervensi gizi, dan program edukasi untuk penanganan *stunting* yang lebih efektif.

5. Meningkatkan efektivitas dan efisiensi strategi penanganan stunting di Kabupaten Kuningan melalui pemetaan yang berbasis web.

1.6 Manfaat Penelitian

1.6.1 Manfaat Teoritis

1. Pengetahuan yang telah diperoleh selama studi di Fakultas Ilmu Komputer Universitas Kuningan dapat diterapkan dalam melakukan penelitian yang ilmiah dan sistematis terkait pemetaan wilayah stunting di Kabupaten Kuningan.
2. Dapat memperluas wawasan, pengetahuan, dan informasi mengenai penerapan teknik klasterisasi menggunakan algoritma *DBSCAN* untuk menganalisis distribusi kasus stunting dan mendukung strategi penanganan yang lebih efektif.

1.6.2 Manfaat Praktis

Dapat mempermudah Dinas Kesehatan dalam memperoleh informasi tentang wilayah dengan prevalensi *stunting* yang tinggi, membantu dalam merancang program intervensi yang lebih terfokus, serta menentukan prioritas alokasi sumber daya untuk daerah-daerah yang membutuhkan penanganan segera. Sistem ini juga memberikan dukungan dalam merancang strategi penanganan yang lebih efektif dan efisien, serta membantu dalam meningkatkan kualitas program edukasi dan intervensi terkait stunting di Kabupaten Kuningan.

1.7 Pertanyaan Penelitian

Dari identifikasi masalah yang telah diuraikan sebelumnya maka terdapat pertanyaan penelitian yaitu:

1. Apakah dengan adanya aplikasi berbasis Web dapat mempermudah pengambilan keputusan terkait dengan strategi penanganan *stunting*, alokasi sumber daya, dan prioritas intervensi di Kabupaten Kuningan?

2. Apakah algoritma DBSCAN dapat diimplementasikan untuk menganalisis distribusi kasus *stunting* dan mengelompokkan wilayah dengan prevalensi *stunting* tinggi di Kabupaten Kuningan?

1.8 Hipotesis Penelitian

Berdasarkan hasil analisis di atas, peneliti mengajukan hipotesis penelitian yaitu dengan dibuatnya aplikasi pemetaan wilayah *stunting* menggunakan algoritma *DBSCAN*, aplikasi ini dapat digunakan untuk menganalisis distribusi kasus *stunting* dan mempermudah pengambilan keputusan dalam menentukan prioritas intervensi, alokasi sumber daya, dan strategi penanganan *stunting* di Kabupaten Kuningan.

1.9 Metodologi Penelitian

Agar penelitian yang dilakukan lebih akurat dan terstruktur, diperlukan penerapan metodologi penelitian yang tepat. Beberapa pendekatan metodologis yang dapat digunakan antara lain:

1.9.1 Metode Pengumpulan Data

Metode pengumpulan data membahas tentang cara memperoleh data yang dibutuhkan dalam penelitian yang akan dilaksanakan, terdapat beberapa metode yang akan digunakan, antara lain:

a. Observasi

Observasi dilakukan secara langsung di Dinas Kesehatan Kabupaten Kuningan untuk mengamati dan mencatat data terkait prevalensi *stunting* di berbagai kecamatan. Data yang diperoleh melalui observasi ini akan mencakup informasi mengenai tingkat *stunting*, fasilitas kesehatan, akses air bersih, dan faktor-faktor lain yang mempengaruhi status gizi balita di wilayah tersebut.

b. Wawancara

Pengumpulan data dilakukan melalui wawancara dengan pihak terkait, seperti petugas kesehatan dari Dinas Kesehatan Kabupaten Kuningan, untuk mendapatkan informasi mendalam

mengenai kondisi stunting di tiap kecamatan. Wawancara ini bertujuan untuk menggali data yang tidak hanya tertera dalam dokumen tetapi juga berdasarkan pengalaman langsung di lapangan.

c. Studi Pustaka

Metode studi pustaka digunakan untuk mengumpulkan informasi dari buku, jurnal, artikel, dan sumber terpercaya lainnya yang berkaitan dengan topik *stunting* dan pemetaan wilayah. Sumber-sumber ini akan dijadikan sebagai dasar teori untuk mendukung analisis data dan sebagai referensi untuk merancang strategi penanganan *stunting* yang lebih efektif.

1.9.2 Metode Pengembangan Sistem

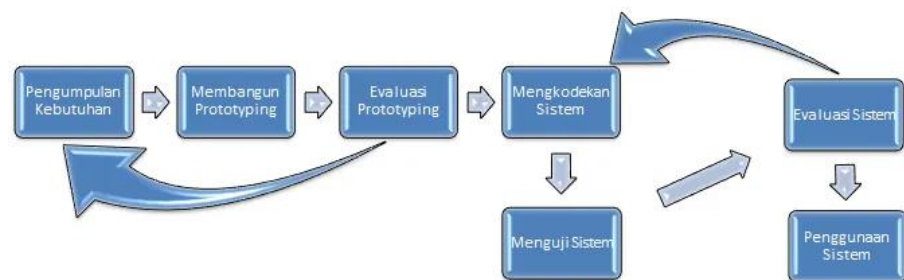
Pengembangan system penelitian ini menggunakan metode *prototyping* untuk membangun model perangkat lunak. *Prototype* yang dibuat merupakan versi awal dari sistem yang bertujuan untuk menyajikan konsep, menguji desain, mengidentifikasi berbagai potensi masalah, dan menyediakan solusi untuk memperbaikinya. Proses ini memungkinkan pengujian berulang dan penyempurnaan sistem hingga mencapai fungsionalitas yang diinginkan.

Metode *prototyping* digunakan dalam sistem ini untuk membantu pengguna memahami langkah-langkah kerja sistem yang dirancang, sehingga memastikan sistem dapat berfungsi sesuai kebutuhan. Tujuan dari metode *prototyping* dalam penelitian ini adalah untuk membangun representasi awal dari model aplikasi yang akan dikembangkan. Prosesnya dimulai dengan pembuatan mockup sebagai gambaran awal desain aplikasi, yang kemudian dievaluasi oleh pengguna. Hasil evaluasi ini digunakan sebagai panduan bagi pengembang perangkat lunak dalam membangun aplikasi yang sebenarnya (Fridayanthie et al., 2021).

Menurut S. Kosasi pada Jurnal Teknik Informatika dan Sistem Informasi, pendekatan throw-away prototyping merupakan metode di mana spesifikasi kebutuhan sistem telah ditentukan secara rinci sejak awal, sehingga pendekatan ini mampu meminimalkan risiko terjadinya ketidaksesuaian antara sistem yang dibangun dengan kebutuhan yang sebenarnya (Khalida et al., 2020).

Kelebihan metode *prototyping* dalam pengembangan perangkat lunak meliputi komunikasi yang efektif antara pengguna dan pengembang, memungkinkan pengembang memahami kebutuhan dengan lebih baik. Pengguna dapat terlibat secara aktif dalam proses pengembangan, sehingga membantu mempercepat waktu pengerjaan sistem. Selain itu, karena pengguna memiliki gambaran yang jelas tentang hasil akhir yang diinginkan, implementasi menjadi lebih mudah. Klien juga dapat melihat representasi awal dari sistem melalui *prototype* yang dibuat (Zailani et al., 2020).

Tahapan *prototype* dapat dilihat pada gambar dibawah ini.



Gambar 1. 1 Metode Prototype (Zuhri et al., n.d.)

1. Pengumpulan Kebutuhan

Pada tahap awal, langkah yang dilakukan adalah mengidentifikasi permasalahan dan mengumpulkan data dari Dinas Kesehatan

Kabupaten Kuningan, khususnya terkait prevalensi stunting. Hal ini dilakukan melalui observasi untuk mendapatkan data balita stunting, fasilitas kesehatan, dan faktor-faktor lain seperti ketersediaan air bersih serta akses layanan kesehatan. Selain itu, wawancara dilakukan dengan pihak Dinas Kesehatan dan petugas terkait untuk memperoleh informasi mendalam mengenai kendala yang dihadapi, seperti kurangnya alokasi sumber daya yang tepat sasaran, keterbatasan fasilitas kesehatan di wilayah tertentu, dan tingginya angka stunting di beberapa kecamatan. Informasi yang diperoleh kemudian diolah untuk menentukan kebutuhan sistem yang akan dibangun, seperti pemetaan wilayah stunting berbasis data untuk mengidentifikasi prioritas intervensi yang lebih efektif.

2. Membangun *Prototype*

Tahap selanjutnya adalah membangun prototype dengan merancang model awal yang berfokus pada pembuatan sistem pemetaan wilayah stunting. Misalnya, merancang fitur untuk input data prevalensi stunting dan fasilitas kesehatan, serta output berupa visualisasi peta wilayah dengan klaster prioritas penanganan.

3. Evaluasi *Prototype*

Pada tahap ini, dilakukan evaluasi untuk memastikan apakah prototipe yang telah dibuat sudah sesuai dengan kebutuhan. Jika prototipe dinilai sesuai, proses dapat dilanjutkan ke tahap berikutnya. Namun, jika terdapat ketidaksesuaian, langkah sebelumnya akan diulang untuk melakukan perbaikan.

4. Mengkodekan Sistem

Pada tahap ini, dilakukan perancangan, pembangunan, dan penerapan aplikasi web sesuai kebutuhan yang telah ditentukan, kemudian diterjemahkan ke dalam bahasa pemrograman yang sesuai.

5. Menguji Sistem

Proses pengujian sistem dilakukan setelah tahap pengkodean selesai. Pada tahap ini, pengujian difokuskan pada dua aspek utama: logika internal sistem untuk memastikan semua pernyataan telah diuji, serta fungsionalitas eksternal untuk mendeteksi kesalahan dan memastikan sistem sesuai dengan proses bisnis yang telah dirancang. Dalam penelitian ini, pengujian sistem akan dilakukan menggunakan metode *Black Box Testing* dan *White Box Testing*.

6. Evaluasi Sistem

Tahap ini melibatkan evaluasi terhadap semua langkah yang telah dilakukan sebelumnya. Jika hasilnya sesuai dengan yang diharapkan, proses dapat dilanjutkan ke tahap berikutnya. Namun, jika ditemukan ketidaksesuaian atau perlu revisi, proses akan kembali ke tahap pertama atau kedua untuk dilakukan perbaikan.

7. Penggunaan Sistem

Setelah sistem selesai diimplementasikan dan diuji, langkah selanjutnya adalah melakukan pemeliharaan sistem (*maintenance*) untuk memastikan sistem tetap berjalan dengan baik dan berfungsi sebagaimana mestinya.

1.9.3 Metode Penyelesaian Masalah

Metode penyelesaian masalah yang digunakan dalam penelitian ini diantaranya:

a. Data Mining

Data mining adalah proses menggali informasi, pola-pola yang relevan, atau data penting dari sebuah database untuk menghasilkan informasi yang bernilai tinggi (Salman et al., 2022). *Data mining* adalah istilah yang merujuk pada proses menemukan pengetahuan dari *database*. Proses ini melibatkan ekstraksi dan identifikasi informasi yang relevan serta pengetahuan terkait dari kumpulan data yang sangat besar, dengan memanfaatkan teknik

statistik, matematika, kecerdasan buatan, dan pembelajaran mesin (Mirantika et al., n.d.).

Menurut Gartner Group, *data mining* adalah proses menganalisis kumpulan data berskala besar yang tersimpan dalam sistem penyimpanan, dengan menggunakan metode pengenalan pola seperti teknik statistik dan matematika untuk mengidentifikasi hubungan, pola, dan tren yang signifikan. (Kusrini; Luthfi, 2009).

“*Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Kusrini; Luthfi, 2009)”.

Dari definisi-definisi yang telah disampaikan, menurut Kusrini dan Emha Taufiq Luthfi *data mining* adalah:

1. *Data mining* adalah proses otomatis yang dilakukan pada data yang telah tersedia.
2. Data yang diproses merupakan kumpulan data berukuran sangat besar.
3. Tujuan dari *data mining* adalah menemukan hubungan atau pola yang berpotensi memberikan informasi yang bermanfaat.

Data mining dibagi menjadi beberapa kelompok yaitu:

1. Deskripsi

Pola dalam data seringkali sulit dipahami, sehingga peneliti dan analis berusaha mencari cara untuk menjelaskan pola tersebut. Penjelasan pola ini bertujuan untuk memberikan pemahaman yang lebih sederhana dan mudah dimengerti.

2. Estimasi

Estimasi memiliki kemiripan dengan klasifikasi, tetapi variabel target dalam estimasi lebih berfokus pada data numerik. Sebagai contoh, estimasi dilakukan untuk memprediksi nilai indeks prestasi kumulatif (IPK) mahasiswa program

pascasarjana berdasarkan nilai IPK mereka saat menempuh program sarjana.

3. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi, tetapi berbeda karena hasil nilainya tidak tersedia atau tidak terjadi saat ini, melainkan akan terjadi di masa depan. Contohnya adalah memprediksi harga minyak untuk dua bulan ke depan.

4. Klasifikasi

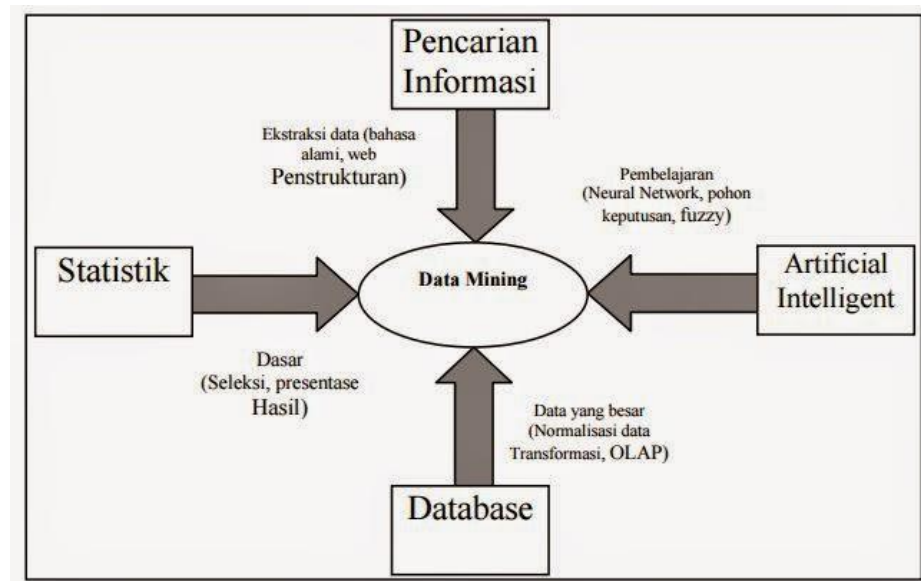
Dalam klasifikasi, variabel target cenderung berupa kategori atau bersifat kategorikal. Contohnya adalah klasifikasi ulasan produk atau layanan yang dibagi menjadi tiga kategori: ulasan positif, ulasan negatif, dan ulasan netral.

5. *Clustering*

Clustering adalah proses pengelompokan data untuk membentuk kelompok atau kategori yang memiliki kesamaan dalam karakteristik tertentu. Contohnya, pengelompokan pelanggan untuk menentukan target pemasaran produk di suatu perusahaan.

6. Asosiasi

Asosiasi adalah teknik dalam data mining yang digunakan untuk mengidentifikasi atribut yang muncul secara bersamaan. Dalam konteks bisnis, teknik ini sering disebut sebagai analisis keranjang belanja atau *Market Basket Analysis*. Contohnya, menemukan produk-produk di toko yang sering dibeli bersama dan produk yang jarang atau tidak pernah dibeli bersamaan.



Gambar 1. 2 Bidang Ilmu Data Mining (Kusrini; Luthfi, 2009)

Gambar 2 menggambarkan bahwa *data mining* memiliki akar yang berasal dari berbagai disiplin ilmu, seperti kecerdasan buatan (*artificial intelligence*), pembelajaran mesin (*machine learning*), statistika, basis data (*database*), serta pengambilan informasi (*information retrieval*). Bidang *Data mining* bukanlah hal yang baru. *Data mining* sulit untuk didefinisikan karena mewarisi banyak konsep dan metode dari bidang yang lebih terdahulu.

b. Algoritma DBSCAN

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) adalah salah satu metode yang mempelopori teknik pengelompokan berbasis kepadatan, yang juga dikenal dengan istilah *density-based clustering*. *DBSCAN* adalah metode *clustering* yang membentuk area berdasarkan kepadatan yang saling terhubung (*density connected*). Setiap objek dalam radius tertentu (*cluster*) harus memiliki setidaknya sejumlah data minimum. Objek yang tidak termasuk dalam *cluster* dianggap

sebagai *noise* (Harjanto et al., 2021). Ada beberapa komponen atau istilah yang terdapat di algoritma *DBSCAN* (Harjanto et al., 2021):

- 1) *Eps*: Merupakan jumlah titik yang ada dalam radius tertentu.
- 2) *Noise Point*: Titik yang terletak di luar area kepadatan atau *Eps*.
- 3) *Border Point*: Titik yang berada di batas *cluster*, memiliki kurang dari *MinPts* dalam *Eps*, tetapi masih berada dalam area sekitar titik inti.
- 4) *MinPts*: Jumlah tetangga terdekat yang digunakan untuk menentukan lingkungan lokal suatu objek.
- 5) *Core Point*: Titik yang terletak di dalam inti *cluster*.

Adapun langkah-langkah Algoritma *DBSCAN* secara umum adalah sebagai berikut:

1. Normalisasi data untuk menyamakan skala kedua variabel (persentase dan fasilitas kesehatan) ke *range* [0, 1].data. Normalisasi dilakukan dengan menggunakan rumus *Min-Max* Sebagai berikut :

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \dots\dots\dots(1)$$

2. Menentukan nilai *epsilon* (ϵ) dan *MinPts* sangat penting karena nilai *epsilon* yang optimal dapat mempengaruhi kualitas *cluster* yang dihasilkan. Penentuan nilai *epsilon* yang tepat dilakukan dengan menggunakan metode jarak *KNN* (*k-nearest neighbors*) atau k-tetangga terdekat (Karmenova et al., 2022). Proses utama dalam perhitungan nilai epsilon optimal dengan metode *KNN* adalah menghitung jarak dalam matriks titik ke *KNN*. Jarak rata-rata antara setiap titik dan *KNN* dihitung, dan nilai k diatur sesuai dengan nilai *MinPts*. Jarak k

kemudian diplot dalam kurva yang menunjukkan kenaikan, di mana perubahan tajam pada kurva menunjukkan jarak k yang optimal. Salah satu persamaan yang digunakan dalam perhitungan jarak KNN adalah *Euclidean Distance*, dengan rumus sebagai berikut:

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots(2)$$

di mana:

x_i : titik x ke i

y_i : titik y ke i

3. Melakukan pengelompokan (*clustering*) dengan menggunakan *epsilon* (ϵ) dan *MinPts* sebagai parameter masukan.
4. Melakukan pengujian klaster menggunakan *Silhouette Coefficient* dengan rumus sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \dots\dots\dots(3)$$

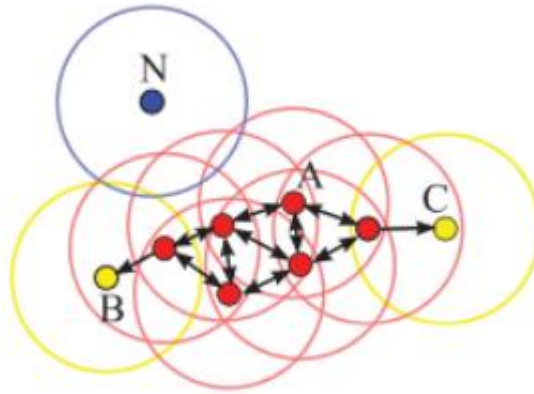
di mana:

si : nilai *Silhouette Coefficient*

$a(i)$: rata-rata jarak antara titik I dengan semua titik dalam *cluster A* (*cluster* tempat titik I berada)

$b(i)$: rata-rata jarak antara titik I dengan semua titik dalam *cluster* selain A .

5. Melakukan visualisasi *cluster* yang diperoleh.



Gambar 1. 3 Ilustrasi Model Cluster DBSCAN (Bagus et al., 2023)

Adapun contoh perhitungan Algoritma DBSCAN sebagai berikut:

1. Penentuan Nilai Epsilon dan MinPts

Tabel 1. 1 Contoh Dataset

Kecamatan	Prevalensi Stunting (%) (x)	Jumlah Fasilitas Kesehatan (y)	Prevalensi Stunting ternormalisasi	Jumlah Fasilitas Kesehatan ternormalisasi
A	30	5	0.333	0.500
B	45	3	0.833	0.167
C	20	8	0	1
D	50	2	1	0
E	35	4	0.5	0.333

a. Hitung menggunakan rumus *Euclidean Distance*:

$$d(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \dots\dots\dots(4)$$

1) Jarak $d(A,B)$

$$d(A,B) = \sqrt{(0.333 - 0.833)^2 + (0.5 - 0.167)^2}$$

$$d(A,B) = \sqrt{0.25 + 0.111}$$

$$d(A,B) = \sqrt{0.361} = 0.601$$

2) Jarak $d(A,C)$

$$d(A,C) = \sqrt{(0.333 - 0)^2 + (0.500 - 1)^2}$$

$$d(A,C) = \sqrt{0.111 + 0.25}$$

$$d(A,C) = \sqrt{0.361} = 0.601$$

3) Jarak $d(A,D)$

$$d(A,D) = \sqrt{(0.333 - 1)^2 + (0.5 - 0)^2}$$

$$d(A,D) = \sqrt{0.445 + 0.25}$$

$$d(A,D) = \sqrt{0.695} = 0.834$$

4) Jarak $d(A,E)$

$$d(A,E) = \sqrt{(0.333 - 0.5)^2 + (0.5 - 0.333)^2}$$

$$d(A,E) = \sqrt{0.028 + 0.028}$$

$$d(A,E) = \sqrt{0.056} = 0.236$$

5) Jarak $d(B,C)$

$$d(B,C) = \sqrt{(0.833 - 0)^2 + (0.167 - 1)^2}$$

$$d(B,C) = \sqrt{0.694 + 0.694}$$

$$d(B,C) = \sqrt{1.388} = 1.178$$

6) Jarak $d(B,D)$

$$d(B,D) = \sqrt{(0.833 - 1)^2 + (0.167 - 0)^2}$$

$$d(B,D) = \sqrt{0.028 + 0.028}$$

$$d(B,D) = \sqrt{0.056} = 0.236$$

7) Jarak $d(B,E)$

$$d(B,E) = \sqrt{(0.833 - 0.5)^2 + (0.167 - 0.333)^2}$$

$$d(B,E) = \sqrt{0.111 + 0.028}$$

$$d(B,E) = \sqrt{0.139} = 0.373$$

8) Jarak $d(C,D)$

$$d(C,D) = \sqrt{(0 - 50)^2 + (1 - 0)^2}$$

$$d(C,D) = \sqrt{1 + 1}$$

$$d(C,D) = \sqrt{2} = 1.414$$

9) Jarak $d(C,E)$

$$d(C,E) = \sqrt{(0 - 0.5)^2 + (1 - 0.333)^2}$$

$$d(C,E) = \sqrt{0.25 + 0.445}$$

$$d(C,E) = \sqrt{0.695} = 0.834$$

10) Jarak $d(D,E)$

$$d(D,E) = \sqrt{(1 - 0.5)^2 + (0 - 0.333)^2}$$

$$d(D,E) = \sqrt{0.25 + 0.111}$$

$$d(D,E) = \sqrt{0.361} = 0.601$$

11) Dari hasil perhitungan menghasilkan jarak ke tetangga terdekat.

Tabel 1. 2 Rekapitulasi Hasil Perhitungan

Pasangan Titik	Jarak (Euclidean Distance)
A,B	0.601
A,C	0.601
A,D	0.834
A,E	0.236
B,C	1.178
B,D	0.236
B,E	0.373
C,D	1.414

C,E	0.834
D,E	0.601

12) Untuk setiap titik, hitung jarak ke tetangga terdekat ($k = MinPts$). Misalkan $MinPts = 3$:

Tabel 1. 3 Jarak Ke Tetangga Terdekat

Kecamatan	Jarak ke Tetangga Terdekat
A	0.601, 0.601, 0.333
B	0.333, 0.601, 0.601
C	0.601, 0.745, 1.201
D	0.601, 0.745, 1.054
E	0.333, 0.333, 0.745

13) Ambil jarak ke tetangga ke-3 (karena $MinPts = 3$) untuk setiap titik:

Tabel 1. 4 Jarak ke-k=3

Kecamatan	Jarak ke-k = 3
A	0.601
B	0.601
C	1.201
D	1.054
E	0.745

14) Urutkan jarak dari kecil ke besar untuk membuat *kNN Distance Plot*:

{0.601, 0.601, 0.745, 1.054, 1.201}

15) Plotkan nilai-nilai jarak tersebut untuk mencari perubahan tajam (*elbow point*). Nilai *epsilon* (ϵ) yang optimal adalah jarak pada perubahan signifikan pertama. Dari data ini:

Lonjakan terjadi dari 0.745 ke 1.054.

Pilih $\epsilon = 0.745$ sebagai nilai optimal karena mencakup sebagian besar hubungan antar titik dengan baik.

b. Menentukan tetangga dalam Radius $\epsilon = 0.745$:

1. Titik A: Tetangga: B (0.601), C (0.601), E (0.333).
Total Tetangga: 3 $\rightarrow$ *Core Point*.

2. Titik B: Tetangga: A (0.601), E (0.333), D (0.601).
Total Tetangga: 3 $\rightarrow$ *Core Point*.

3. Titik C: Tetangga: A (0.601), E (0.745). Total Tetangga: 2 (kurang dari $MinPts=3$) $\rightarrow$ *Border Point*.

4. Titik D: Tetangga: B (0.601), E (0.745). Total Tetangga: 2 (kurang dari $MinPts=3$) $\rightarrow$ *Border Point*.

5. Titik E: Tetangga: A (0.333), B (0.333), D (0.745).
Total Tetangga: 3 $\rightarrow$ *Core Point*.

c. Identifikasi *Core Points* dan *Border Points*

1. *Core Points*: Titik A, B, E : Memiliki $\geq MinPts$ tetangga.

2. *Border Points*: Titik C, D : Kurang dari $MinPts$, tetapi terhubung dengan *core point*.

d. Proses *Clustering*

- 1) Mulai dari Titik A : A adalah *core point*. Tetangga langsung: B, C, E . Gabungkan A, B, C, E ke dalam *Cluster 0*.
 - 2) Lanjut ke Titik B : B adalah *core point*. Tetangga langsung: A, E, D . Titik D bergabung ke dalam *Cluster 0*.
 - 3) Titik E : E adalah *core point*. Tetangga langsung: A, B, C . Semua tetangga sudah termasuk dalam *Cluster 0*.
 - 4) Titik C, D : Kedua titik adalah *border points*, tetapi sudah masuk dalam *Cluster 0*.
- e. Hasil Akhir:

Tabel 1. 5 Hasil Akhir Perhitungan

Kecamatan	Cluster
A	0
B	0
C	0
D	0
E	0

Kesimpulan: *Core Points*: A, B, E . *Border Points*: C, D .
Noise Points: Tidak ada. Semua titik membentuk satu *cluster* besar (*Cluster 0*) karena saling terhubung melalui *core points*.

- f. Hitung Rata-Rata Jarak Dalam *Cluster* ($a(i)$)

$$a(A) = \frac{d(A,B) + d(A,C) + d(A,E)}{3}$$

$$a(A) = \frac{0.601 + 0.601 + 0.333}{3} = 0.512$$

$$a(B) = \frac{d(B,A) + d(B,D) + d(B,E)}{3}$$

$$a(B) = \frac{0.601 + 0.601 + 0.333}{3} = 0.512$$

$$a(C) = \frac{d(C,A) + d(C,E)}{2}$$

$$a(C) = \frac{0.601 + 0.745}{2} = 0.673$$

$$a(D) = \frac{d(D,B) + d(D,E)}{2}$$

$$a(D) = \frac{0.601 + 0.745}{2} = 0.673$$

$$a(E) = \frac{d(E,A) + d(E,B) + d(E,C)}{3}$$

$$a(E) = \frac{0.333 + 0.333 + 0.745}{3} = 0.470$$

g. Hitung Rata-Rata Jarak Antar *Cluster* ($b(i)$)

Karena semua titik berada dalam satu *cluster*, $b(i)$ untuk setiap titik adalah tidak terhingga (∞), karena tidak ada *cluster* lain untuk dibandingkan. Oleh karena itu, kita langsung menyatakan bahwa: $b(i) = \infty$ untuk setiap titik i

h. Hitung *Silhouette Coefficient* $s(i)$

Untuk setiap titik i , $s(i)$ dihitung sebagai:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Karena $b(i) = \infty$ untuk semua titik, maka:

$$s(i) = \frac{\infty - a(i)}{\infty} = 1 \text{ untuk setiap titik } i$$

i. Rata-Rata *Silhouette Coefficient* (SC)

Karena semua nilai $s(i) = 1$, maka rata-rata *Silhouette Coefficient* (SC) untuk seluruh *dataset* adalah:

$$SC = \frac{1+1+1+1+1}{5} = 1$$

1.11 Sistematika Penelitian

BAB I : PENDAHULUAN

Pada Bab ini diuraikan mengenai Latar Belakang, Identifikasi Masalah, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Ruang Lingkup Pembahasan, dan Sistematika Penulisan.

BAB II : LANDASAN TEORITIS

Pada Bab ini membahas mengenai teori-teori yang melandasi penulisan Naskah Seminar Hasil Penelitian ini diantaranya pembuatan aplikasi, perancangan sistem seperti *Activity Diagram*, *Class Diagram*, dan *Sequence Diagram*.

BAB III : ANALISA DAN PERANCANGAN

Pada Bab ini membahas mengenai sistem yang sedang berjalan, fungsi dan kegunaan sistem yang sudah ada, dan sistem yang diusulkan.

BAB IV : IMPLEMENTASI DAN PENGUJIAN

Pada bab ini membahas mengenai perancangan program dan listing program yang akan dibangun.

BAB V : KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dan saran dari Naskah Seminar Hasil Penelitian serta saran untuk perbaikan aplikasi yang telah dibuat.