

## **BAB II**

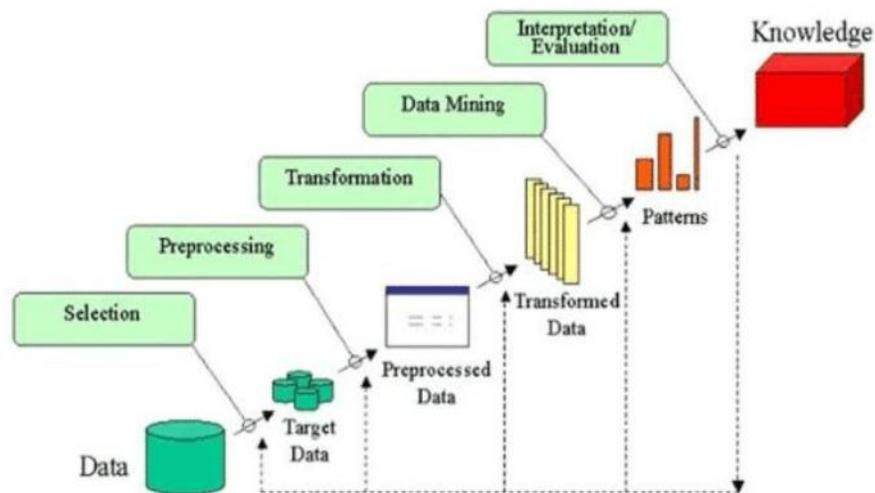
### **LANDASAN TEORI**

#### **2.1. Knowledge Discovery in Database**

Knowledge Discovery in Databases (KDD) merujuk pada proses ekstraksi informasi atau pola berharga yang sebelumnya tidak diketahui dari volume data yang besar. Ini mencakup serangkaian langkah, termasuk pemilihan data, pra-pemrosesan, transformasi, penambangan data, dan interpretasi hasil. Tujuan utama KDD adalah mengungkap pengetahuan dan wawasan tersembunyi yang dapat berkontribusi pada pengambilan keputusan yang lebih baik dan pemahaman tren mendasar dalam data. Ini melibatkan aplikasi berbagai teknik, seperti pembelajaran mesin, analisis statistik, dan visualisasi data, untuk mengidentifikasi pola, asosiasi, dan tren dalam kumpulan data yang kompleks. KDD memainkan peran penting dalam bidang seperti bisnis, kesehatan, keuangan, dan penelitian ilmiah, di mana kemampuan untuk mengekstrak pengetahuan yang bermakna dari jumlah data yang besar dapat menghasilkan strategi, prediksi, dan penemuan yang lebih baik.

Selain itu, KDD juga melibatkan evaluasi dan validasi hasil untuk memastikan keandalan dan kegunaan informasi yang ditemukan. Proses ini sering melibatkan iterasi dan penyesuaian terhadap model atau metode analisis data guna meningkatkan akurasi dan relevansi temuan. Selain itu, aspek etika dan privasi juga menjadi perhatian penting dalam KDD, mengingat sifat sensitif dari beberapa data yang diolah. Penerapan KDD tidak hanya membutuhkan keahlian dalam teknik analisis

data, tetapi juga pemahaman yang mendalam terhadap konteks domain yang bersangkutan. Dengan terus berkembangnya teknologi dan pertumbuhan data yang cepat, KDD menjadi semakin krusial dalam membantu organisasi dan peneliti menggali nilai dari data untuk membuat keputusan yang lebih informasional dan cerdas.



*Tabel 2. 1. Knowledge Discovery in Database*

Sumber Gambar: <https://www.researchgate.net>

Interpretation/Evaluation : Evaluasi hasil data mining adalah langkah kritis untuk menilai keberhasilan dan relevansi penemuan yang telah diidentifikasi.

Data Mining : Data mining melibatkan penggunaan teknik analisis yang canggih untuk mengungkap pola dan hubungan

dalam data yang dapat mendukung pengambilan keputusan yang lebih baik.

- Transformation : Transformasi data mencakup proses perubahan format atau struktur data agar sesuai dengan kebutuhan analisis atau pemodelan yang akan dilakukan.
- Preprocessing : Pra-pemrosesan melibatkan langkah-langkah untuk membersihkan dan mempersiapkan data sebelum proses analisis, sehingga memastikan keakuratan dan kualitas data.
- Selection : Pemilihan data melibatkan proses penentuan data yang paling relevan dan signifikan untuk dimasukkan ke dalam analisis lebih lanjut.

### **2.1.1. Data Mining**

Data mining adalah proses eksplorasi dan penemuan pola tersembunyi atau informasi yang berharga dalam suatu set data yang besar dan kompleks [4]. Tujuannya adalah untuk mengidentifikasi hubungan yang belum diketahui, tren, dan pola yang dapat memberikan wawasan strategis untuk pengambilan keputusan. Dalam proses ini, teknik-teknik statistika, matematika, dan kecerdasan buatan digunakan untuk menggali informasi yang relevan dari data yang sangat luas. Data mining tidak hanya terbatas pada penggalian data historis, tetapi juga mencakup kemampuan prediktif untuk membantu organisasi mengantisipasi dan merespons perubahan pasar atau kondisi lingkungan [5]. Dengan menggunakan berbagai

algoritma seperti decision trees, clustering, dan association rule mining, data mining menjadi alat yang sangat berharga untuk mendukung analisis bisnis, perencanaan strategis, dan pengambilan keputusan yang lebih efektif.

Data mining, sebagai bagian integral dari proses KDD, merupakan suatu pendekatan yang mendasari dalam menggali pola dan pengetahuan dari kumpulan data yang besar. Melibatkan teknik-teknik seperti clustering, association rule mining, dan classification, data mining memungkinkan identifikasi hubungan yang kompleks dan pola tersembunyi dalam data. Dengan menganalisis data secara mendalam, data mining dapat memberikan wawasan yang berharga dan mendukung pengambilan keputusan yang lebih tepat. Proses ini tidak hanya membantu dalam mengidentifikasi tren dan pola masa lalu, tetapi juga memungkinkan prediksi yang lebih baik terhadap perilaku masa depan. Data mining menjadi alat yang sangat berharga dalam berbagai industri, membantu perusahaan dan organisasi untuk meningkatkan efisiensi, mengoptimalkan strategi pemasaran, dan mengidentifikasi peluang baru. Meskipun memberikan manfaat yang besar, implementasi data mining juga menantang karena kompleksitas data yang terus berkembang dan pentingnya menjaga keamanan serta privasi informasi.

### **2.1.2. Database dan Data Processing**

Database dan pemrosesan data merupakan fondasi yang krusial dalam ekosistem informasi modern. Sebuah database adalah struktur penyimpanan sistematis yang memungkinkan pengorganisasian, penataan, dan akses efisien

terhadap data. Database menyediakan kerangka kerja yang terstruktur untuk menyimpan dan mengelola informasi, memastikan konsistensi, integritas, dan keamanan data. Di sisi lain, pemrosesan data melibatkan serangkaian langkah untuk mengumpulkan, mentransformasi, dan menganalisis data. Proses ini dapat melibatkan teknik-teknik seperti penggabungan data, pembersihan data, dan penghitungan statistik. Kombinasi database yang solid dengan pemrosesan data yang cermat memungkinkan organisasi untuk mengoptimalkan pengelolaan informasi, mendukung pengambilan keputusan yang informasional, dan memberikan dasar yang kuat bagi pengembangan dan inovasi lebih lanjut dalam dunia teknologi informasi.

### **2.1.3. Visualisation**

Visualisasi adalah suatu teknik penting dalam mewujudkan pemahaman yang mendalam dari data kompleks. Dengan menggambarkan informasi dalam bentuk grafis, visualisasi memfasilitasi interpretasi yang lebih cepat dan efektif, memungkinkan orang untuk menangkap pola, hubungan, dan tren dengan lebih mudah. Dalam konteks analisis data, visualisasi memainkan peran penting dalam menyajikan temuan analitis secara intuitif, yang dapat mendukung pengambilan keputusan yang tepat. Berbagai jenis visualisasi, seperti diagram batang, grafik garis, peta panas, dan visualisasi 3D, dapat disesuaikan dengan jenis data dan tujuan analisis. Dengan menggunakan alat visualisasi yang tepat, informasi yang kompleks dapat diubah menjadi narasi yang jelas dan dapat dipahami, memungkinkan para

profesional dari berbagai disiplin ilmu untuk menjelajahi, memahami, dan berkomunikasi mengenai data dengan lebih efektif.

#### **2.1.4. Statistik**

Statistik adalah cabang ilmu matematika yang berfokus pada pengumpulan, analisis, interpretasi, presentasi, dan pengorganisasian data. Tujuan utama statistik adalah untuk menyajikan pola dan tren yang ada dalam data, sehingga dapat diambil kesimpulan atau membuat prediksi berdasarkan informasi yang ada. Dengan menggunakan metode statistik, kita dapat membuat generalisasi yang dapat diterapkan pada populasi lebih luas berdasarkan sampel data yang diambil. Statistik memiliki peran penting di berbagai bidang, termasuk ilmu sosial, ilmu alam, ekonomi, kedokteran, dan bisnis. Melalui konsep-konsep seperti mean, median, modus, deviasi standar, dan uji hipotesis, statistik memberikan alat yang kuat untuk mendukung pengambilan keputusan, penelitian, dan evaluasi kinerja, menjadikannya instrumen kunci dalam menghadapi dan menginterpretasi keragaman data dalam konteks dunia nyata.

#### **2.1.5. Pattern Recognition**

Pattern recognition, atau pengenalan pola, merupakan cabang ilmu yang berfokus pada pengembangan teknik dan algoritma untuk mengenali pola dalam data. Tujuan utama dari pattern recognition adalah untuk mengidentifikasi struktur atau karakteristik tertentu dalam suatu himpunan data, sehingga sistem dapat

mengklasifikasikan atau memberikan makna pada informasi tersebut. Metode pattern recognition mencakup berbagai teknik seperti pengenalan wajah, pengenalan suara, pengenalan tulisan tangan, dan pengenalan pola visual dalam gambar atau video. Dengan memanfaatkan pendekatan machine learning dan kecerdasan buatan, pattern recognition memungkinkan komputer untuk belajar dari data, mengekstrak fitur yang relevan, dan membuat prediksi atau keputusan berdasarkan pola yang telah diidentifikasi. Dalam konteks teknologi modern, pengenalan pola memiliki aplikasi luas, dari sistem keamanan hingga pengolahan citra medis, menjadikannya bidang penelitian dan pengembangan yang terus berkembang.

## **2.2. Model Clustering**

Model clustering dengan menggunakan metode K-Means adalah salah satu teknik populer dalam dunia data mining untuk mengelompokkan data berdasarkan kemiripan fitur atau atributnya [6]. Algoritma K-Means bekerja dengan cara membagi dataset menjadi beberapa kluster, di mana setiap kluster memiliki pusat yang merepresentasikan rata-rata atribut dari anggota kluster tersebut [7]. Prosesnya melibatkan inisialisasi pusat kluster, kemudian memperbarui pusat kluster dan mengelompokkan data berdasarkan jarak terdekat hingga iterasi konvergen [8].

Keunggulan K-Means terletak pada kemampuannya mengatasi data dengan jumlah besar dan menentukan kluster tanpa memerlukan label kelas sebelumnya [9]. Hasil dari K-Means memberikan insight tentang sejauh mana data tersebar dalam beberapa kelompok yang signifikan. Namun, penggunaan K-Means memiliki

beberapa catatan. Algoritma ini sangat sensitif terhadap inisialisasi awal pusat kluster, dan jumlah kluster (K) harus ditentukan sebelumnya.

Penerapan K-Means dalam data mining dapat ditemukan di berbagai bidang, seperti pengelompokan pelanggan berdasarkan perilaku belanja, analisis genetika untuk mengelompokkan pola genetik, atau analisis citra medis untuk mengelompokkan gambar berdasarkan karakteristik tertentu [10]. Dengan menggunakan K-Means, para peneliti dan praktisi dapat menggali struktur dalam data, mengidentifikasi kelompok yang tersembunyi, dan mengambil langkah-langkah berdasarkan wawasan yang dihasilkan, menjadikan metode ini sebagai alat yang efektif dalam penambangan data dan analisis kelompok.

### **2.3. Algoritma K-Means**

Algoritma K-Means adalah salah satu teknik clustering yang paling umum digunakan dalam data mining dan analisis data [11]. Tujuannya adalah mengelompokkan data menjadi K kluster, di mana setiap kluster memiliki pusatnya sendiri yang merepresentasikan nilai rata-rata dari anggota kluster tersebut [12]. Algoritma ini memanfaatkan pendekatan iteratif yang melibatkan tahap inisialisasi, pengelompokan, dan pembaruan pusat kluster hingga mencapai konvergensi. Proses dimulai dengan inisialisasi K pusat kluster secara acak. Selanjutnya, setiap data akan diatributkan ke kluster yang memiliki pusat terdekat. Setelah atribusi selesai, pusat kluster diperbarui berdasarkan rata-rata data di dalam kluster. Iterasi ini terus berlanjut hingga pusat kluster tidak berubah secara signifikan atau sejumlah iterasi

tertentu telah tercapai. Untuk pengolahan data pada metode K-Means, penulis menggunakan rumus jarak (Euclidean Distance). Untuk rumus nya yaitu sebagai berikut.

$$d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$$

Di mana untuk penjelasannya yaitu sebagai berikut.

“d” adalah jarak Euclidean

$(x_1, y_1)$  adalah koordinat titik pertama

$(x_2, y_2)$  adalah koordinat titik kedua.

Salah satu keunggulan K-Means adalah efisiensinya dalam menangani dataset yang besar. Selain itu, karena tidak memerlukan informasi label kelas sebelumnya, K-Means sering digunakan dalam pengelompokan data tanpa pengawasan. Meskipun demikian, pemilihan jumlah kluster (K) dapat menjadi tantangan, dan inisialisasi pusat kluster yang buruk dapat menghasilkan solusi yang tidak optimal. Penerapan algoritma K-Means dapat ditemui dalam berbagai bidang, termasuk segmentasi pelanggan, analisis citra, pengelompokan teks, dan pemrosesan sinyal. Kendati sederhana, K-Means tetap menjadi salah satu algoritma clustering yang paling populer dan efektif dalam analisis data.

## **2.4. Alat Bantu Program/Tools Pendukung**

### **2.4.1. Orange**

Orange merupakan sebuah perangkat lunak sains data yang populer dan berdaya guna, dirancang untuk mempermudah analisis data, visualisasi, dan pemodelan.

Dirilis sebagai perangkat lunak open-source, Orange menyediakan antarmuka pengguna grafis yang intuitif, memungkinkan para pengguna, termasuk mereka yang tidak memiliki latar belakang teknis yang kuat, untuk menjalankan berbagai analisis data tanpa perlu menulis kode. Software ini menawarkan beragam alat analisis data, mulai dari eksplorasi data hingga pembangunan model machine learning, dan memungkinkan pengguna untuk menggabungkan dan memanipulasi data dengan mudah.

Salah satu fitur unggulan Orange adalah kemampuannya dalam menyajikan visualisasi yang informatif dan dapat dimengerti, membantu pengguna untuk memahami struktur dan pola dalam data mereka. Orange juga memiliki koleksi modul atau komponen yang dapat digunakan untuk berbagai tugas, termasuk clustering, klasifikasi, regresi, dan analisis teks. Kelebihan lainnya adalah fleksibilitasnya yang tinggi, memungkinkan pengguna untuk mengonfigurasi dan menyesuaikan alur kerja analisis data sesuai kebutuhan mereka.

## **2.5. Puskesmas**

Pusat Kesehatan Masyarakat (Puskesmas) adalah lembaga kesehatan yang berfungsi sebagai pusat pelayanan kesehatan primer di tingkat masyarakat, terutama di tingkat kecamatan. Puskesmas memiliki peran strategis dalam memberikan layanan kesehatan yang menyeluruh dan merata kepada penduduk di wilayahnya. Salah satu fokus utama Puskesmas adalah upaya promotif dan preventif, seperti penyuluhan kesehatan, imunisasi, dan pemantauan kesehatan ibu dan anak. Puskesmas juga

menyediakan layanan kesehatan kuratif dengan memberikan pengobatan dasar serta merujuk pasien yang membutuhkan perawatan lanjutan ke fasilitas kesehatan yang lebih tinggi.

Selain itu, Puskesmas memiliki peran penting dalam kegiatan pemantauan dan pengendalian penyakit, termasuk program-program pengendalian penyakit menular dan non-menular. Puskesmas juga menjadi basis untuk kegiatan-kegiatan penelitian dan surveilans kesehatan di tingkat kecamatan. Faktanya, Puskesmas tidak hanya menyediakan layanan kesehatan fisik, tetapi juga mendukung kesehatan mental dan sosial masyarakat dengan mengintegrasikan berbagai aspek kesehatan. Pentingnya peran Puskesmas dalam sistem kesehatan nasional juga tercermin dalam upayanya untuk meningkatkan aksesibilitas layanan kesehatan, memberikan pelayanan yang berkualitas, dan melibatkan masyarakat dalam pengambilan keputusan terkait kesehatan. Dengan demikian, Puskesmas menjadi tulang punggung dalam mencapai tujuan pembangunan kesehatan dan meningkatkan kualitas hidup masyarakat di tingkat lokal.

## **2.6. Balita Penderita Gizi Buruk**

Balita yang mengalami gizi buruk merupakan suatu kondisi serius yang dapat memiliki dampak jangka panjang pada kesehatan dan perkembangan anak. Gizi buruk pada balita biasanya terjadi akibat asupan nutrisi yang tidak mencukupi, baik dari segi kuantitas maupun kualitas. Pada usia 0-5 tahun, fase ini sangat kritis dalam perkembangan fisik dan mental anak, sehingga gizi buruk dapat menghambat

pertumbuhan dan mengakibatkan masalah kesehatan yang berpotensi fatal. Balita yang mengalami gizi buruk biasanya menunjukkan gejala seperti berat badan di bawah standar, pertumbuhan yang lambat, dan penurunan energi serta ketahanan tubuh. Kekurangan nutrisi pada masa ini juga dapat memengaruhi perkembangan otak, berdampak pada kemampuan kognitif dan kecerdasan anak. Anak-anak yang menderita gizi buruk memiliki risiko tinggi terhadap penyakit infeksi karena sistem kekebalan tubuh yang melemah.

Pentingnya pencegahan dan penanganan gizi buruk pada balita melibatkan upaya bersama antara keluarga, masyarakat, dan pelayanan kesehatan. Program-program pendidikan gizi, pemberian makanan bergizi, serta akses mudah ke layanan kesehatan merupakan langkah-langkah krusial untuk mencegah dan mengatasi gizi buruk pada balita. Dukungan psikososial dan pendampingan bagi orang tua juga penting guna memastikan penerapan praktik-praktik gizi yang sehat dan memberikan perawatan yang memadai untuk pemulihan balita yang mengalami gizi buruk.

## 2.7. Metodologi Penelitian

### 2.5.1. Penelitian Terdahulu

*Tabel 2. 2. Peneliti Terdahulu*

Referensi Penelitian	1
Judul	Review on the Application of Machine Learning Algorithms in the Sequence

	Data Mining of DNA
Nama	Aimin Yang <sup>1</sup> , Wel Zhang <sup>1</sup> , Jiahao Wang <sup>1</sup> , Ke Yang <sup>2</sup> , Yang Han <sup>1*1</sup> , and Limin Zhang <sup>3*1</sup>
Tahun	2020
Hasil	Penerapan data mining dalam penelitian tentang penyelarasan sekuens DNA, klasifikasi sekuens DNA, pengelompokan sekuens DNA, dan penambangan pola DNA telah menjadi krusial dalam memahami kompleksitas informasi genetik. Dengan memanfaatkan teknik data mining seperti algoritma penyelarasan, klasifikasi, dan pengelompokan, peneliti dapat mengidentifikasi hubungan antar sekuens DNA, mengklasifikasikan mereka ke dalam kelompok yang relevan, serta mengeksplorasi pola-pola genetik yang dapat memberikan wawasan mendalam

	tentang fungsionalitas dan evolusi genom. Pendekatan ini memberikan kontribusi penting dalam bidang biologi molekuler, membuka pintu untuk penemuan baru terkait penyakit genetik, hubungan filogenetik, dan adaptasi evolusioner [13].
Referensi Penelitian	2
Judul	State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing
Nama	Michal Krassowski <sup>1</sup> , Vivek Das <sup>2</sup> , Sangram K. Sahu <sup>3</sup> , and Biswapriya B. Misra <sup>4</sup>
Tahun	2020
Hasil	Penerapan data mining untuk mengintegrasikan dua atau lebih kumpulan data omics telah menjadi strategi yang efektif dalam menganalisis, memvisualisasikan, dan menginterpretasi

	<p>informasi biologis kompleks. Dengan menggabungkan data omics seperti genomics, transcriptomics, proteomics, dan metabolomics, teknik data mining memungkinkan identifikasi pola hubungan yang lebih holistik antara berbagai tingkat ekspresi gen dan molekul biologis. Analisis ini tidak hanya meningkatkan pemahaman tentang regulasi genetik dan jalur biologis, tetapi juga memberikan wawasan yang lebih mendalam tentang mekanisme yang mendasari proses biologis kompleks. Pendekatan ini memungkinkan peneliti untuk membuat hubungan yang lebih komprehensif antara variabel-variabel omics yang berbeda, membantu mengidentifikasi signatur biologis yang penting, dan merinci interaksi kompleks di tingkat sistem dalam studi biologi molekuler [14].</p>
--	---

Referensi Penelitian	3
Judul	Brief introduction of medical database and data mining technology in big data era
Nama	Jin Yang <sup>1,2</sup> Yuanjie Li <sup>3</sup> Qingqing Liu <sup>1,2</sup> Li Li <sup>1</sup> Aozi Feng <sup>1</sup> Tianyi Wang <sup>4,5</sup> Shuai Zheng <sup>4</sup> Anding Xu <sup>6</sup> Jun Lyu <sup>1,2</sup>
Tahun	2020
Hasil	Data mining memainkan peran krusial dalam membantu peneliti klinis memahami dan menerapkan teknologi database secara efektif. Dengan memanfaatkan teknik data mining, peneliti dapat mengeksplorasi dataset klinis yang besar dan kompleks untuk mengidentifikasi pola, tren, dan hubungan yang mungkin sulit diakses secara manual. Analisis data mining memungkinkan penemuan informasi yang tersembunyi, mendukung pengambilan keputusan yang informasional, dan

	<p>memfasilitasi pemahaman yang lebih mendalam tentang faktor-faktor yang mempengaruhi kesehatan pasien. Selain itu, penerapan data mining dalam konteks teknologi database memungkinkan integrasi data dari berbagai sumber, memfasilitasi interoperabilitas, dan menciptakan landasan untuk pengembangan model prediktif atau analisis risiko yang lebih canggih. Dengan demikian, data mining tidak hanya meningkatkan efisiensi penelitian klinis, tetapi juga mendukung implementasi teknologi database dengan cara yang menghasilkan informasi klinis yang lebih berarti dan relevan [15].</p>
Referensi Penelitian	4
Judul	Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan

Nama	Ari Sulistiyawati <sup>1,*</sup> , Eko Supriyanto <sup>2</sup>
Tahun	2021
Hasil	<p>Penerapan metode k-means dalam menentukan kelas unggulan siswa telah menjadi pendekatan yang efektif untuk analisis prestasi akademik. Dengan menggunakan k-means, data prestasi siswa seperti nilai ujian, keterlibatan dalam kegiatan ekstrakurikuler, dan aspek lainnya dapat dikelompokkan secara objektif ke dalam kategori yang relevan. Proses ini memungkinkan identifikasi kelompok siswa yang memiliki karakteristik serupa dalam hal prestasi, memungkinkan sekolah untuk mengidentifikasi dan memberdayakan siswa berpotensi tinggi. Pendekatan ini juga dapat membantu dalam pengembangan strategi pendidikan yang disesuaikan dengan kebutuhan masing-</p>

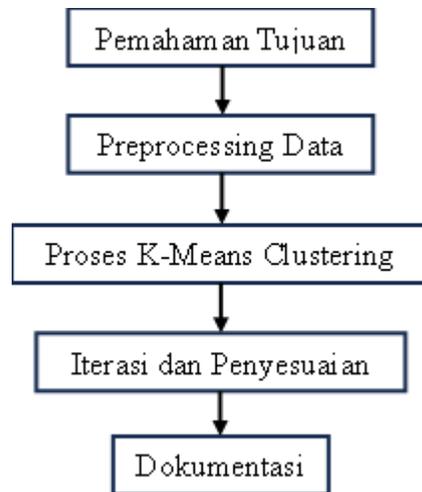
	<p>masing kelompok, memberikan dukungan tambahan kepada siswa yang memerlukan tantangan lebih besar atau bimbingan khusus. Dengan demikian, penerapan metode k-means untuk menentukan kelas unggulan siswa tidak hanya membantu meningkatkan pengelolaan pendidikan, tetapi juga mendukung pengembangan potensi individu siswa secara lebih efektif [16].</p>
Referensi Penelitian	5
Judul	PENERAPAN ALGORITMA K-MEANS DALAM MENENTUKAN TINGKAT PENYEBARAN PANDEMI COVID-19 DI INDONESIA
Nama	Nayuni Dwitri , Jose A Tampubolon , Sandi Prayoga, P.P.P.A.N.W Fikrul Ilmi R.H Zer, Dedy Hartama
Tahun	2020

Hasil	<p>Penerapan metode k-means dalam mengelompokkan data penyebaran COVID-19 telah membuktikan kegunaannya dalam menganalisis pola dan tren epidemiologi. Dengan menggunakan informasi seperti jumlah kasus harian, tingkat penularan, dan geografis, metode k-means dapat mengelompokkan wilayah atau negara ke dalam kategori dengan karakteristik penyebaran yang serupa. Ini memungkinkan peneliti dan pemerintah untuk mengidentifikasi kluster penyebaran, mengarahkan sumber daya kesehatan dengan lebih efisien, dan mengambil langkah-langkah pencegahan yang sesuai untuk setiap kelompok. Selain itu, analisis k-means dapat membantu dalam memahami dinamika penyebaran virus, memungkinkan pengambilan keputusan yang lebih tepat sasaran dalam</p>
-------	---

	<p>pengelolaan pandemi. Dengan demikian, metode k-means memberikan kontribusi penting dalam memahami dan merespons penyebaran COVID-19 secara lebih efektif [17].</p>
Referensi Penelitian	6
Judul	Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah
Nama	Castaka Agus Sugianto <sup>1</sup> , Ayu Hendrati Rahayu <sup>2</sup> , Aditia Gusman <sup>3</sup>
Tahun	2020
Hasil	<p>Penerapan metode K-Means dalam pengelompokan penyakit pasien di Puskesmas Cigugur Tengah membuktikan kebergunaannya dalam meningkatkan manajemen kesehatan masyarakat. Dengan menggunakan data medis seperti gejala penyakit, riwayat kesehatan, dan</p>

	<p>hasil pemeriksaan laboratorium, metode K-Means dapat mengelompokkan pasien ke dalam kategori penyakit yang serupa. Hal ini memungkinkan petugas kesehatan untuk dengan cepat mengidentifikasi pola penyebaran penyakit tertentu di wilayah tersebut, memfasilitasi respon yang lebih cepat dan efektif. Selain itu, pengelompokan pasien berdasarkan karakteristik penyakit mereka dapat membantu Puskesmas dalam perencanaan sumber daya dan penyusunan program pencegahan yang lebih terarah, dengan demikian meningkatkan upaya pencegahan dan pengendalian penyakit di tingkat komunitas. Dengan penerapan K-Means, Puskesmas dapat mengoptimalkan pelayanan kesehatan dan memberikan dampak positif pada kesejahteraan masyarakat Cigugur Tengah [18].</p>
--	--

### 2.5.2. Keerangka Kerja Penelitian



Gambar diatas merupakan desain Penelitian yang akan dilakukan, untuk penjelasan dari setiap alurnya yaitu sebagai berikut:

#### **1. Pemahaman Tujuan**

Pemahaman judul "Analisis Data Mining Clustering pada Balita yang Penderita Gizi Buruk di Puskesmas Sigambal menggunakan Metode K-Means" menjadi langkah awal yang penting untuk memahami fokus dan tujuan penelitian yang akan dibahas. Judul ini memberikan petunjuk bahwa penelitian akan berfokus pada analisis data mining clustering, khususnya terkait balita yang mengalami gizi buruk di Puskesmas Sigambal, dengan metode K-Means sebagai pendekatan utama.

Istilah "Analisis Data Mining Clustering" menunjukkan bahwa penelitian ini akan menggunakan teknik data mining untuk menganalisis pola dan struktur data yang kompleks. Penerapan clustering pada data balita penderita gizi buruk menyoroti

tujuan untuk mengidentifikasi kelompok-kelompok homogen dalam dataset, mungkin untuk merinci pola-pola kesehatan atau karakteristik tertentu.

## ***2. Preprocessing Data***

Preprocessing data adalah tahapan kritis dalam analisis data yang melibatkan serangkaian langkah untuk memastikan bahwa data yang digunakan dalam model atau analisis memiliki kualitas yang baik. Proses ini melibatkan beberapa langkah esensial, pertama-tama adalah pemilihan data, di mana kita memilih subset dari dataset yang relevan untuk tujuan analisis tertentu. Setelah itu, pembersihan data menjadi aspek penting, di mana outlier, nilai yang hilang, atau kesalahan lainnya diidentifikasi dan diperbaiki. Pemilihan variabel kemudian dilakukan untuk mengidentifikasi fitur-fitur yang memiliki kontribusi signifikan terhadap hasil analisis. Normalisasi, sebuah teknik untuk membawa nilai-nilai dalam dataset ke dalam rentang standar, dan standarisasi, mengubah distribusi nilai menjadi distribusi standar dengan rata-rata nol dan deviasi standar satu, membantu memastikan konsistensi dan stabilitas dalam analisis. Keseluruhan, preprocessing data adalah fondasi yang diperlukan untuk memastikan bahwa data yang digunakan dalam analisis atau model memiliki kualitas yang memadai dan dapat menghasilkan hasil yang lebih akurat dan dapat diandalkan.

## ***3. Proses K-Means Clustering***

Proses K-Means Clustering merupakan suatu metode analisis data yang bertujuan untuk mengelompokkan observasi ke dalam beberapa kelompok atau

cluster berdasarkan kesamaan karakteristik. Langkah pertama dalam proses ini adalah pemilihan metode K-Means Clustering, di mana kita harus menentukan berapa banyak cluster yang diinginkan untuk membagi data. Setelah itu, dilakukan pemilihan jumlah kelompok (Cluster), yang dapat memengaruhi hasil clustering. Inisialisasi Cluster menjadi langkah berikutnya, di mana posisi awal centroid ditentukan secara acak atau dengan menggunakan metode tertentu. Proses clustering melibatkan pengelompokan observasi ke dalam cluster-cluster berdasarkan kedekatan dengan centroid terdekat. Evaluasi data clustering kemudian dilakukan untuk mengukur seberapa baik data telah terkelompok, dengan menggunakan metrik seperti inertia atau indeks validitas. Evaluasi dan variabel juga menjadi langkah penting untuk menilai kualitas hasil clustering dan memastikan pemilihan jumlah cluster yang tepat. Terakhir, interpretasi hasil dilakukan untuk memberikan pemahaman yang lebih dalam tentang makna dari setiap cluster yang terbentuk. Keseluruhan, proses K-Means Clustering melibatkan serangkaian tahapan yang saling terkait untuk menghasilkan kelompok-kelompok yang memiliki kemiripan karakteristik di dalam dataset.

#### ***4. Iterasi dan Penyesuaian***

Iterasi dan penyesuaian merupakan aspek kritis dalam analisis data mining clustering pada balita penderita gizi buruk di Puskesmas Sigambal menggunakan metode K-Means. Proses ini melibatkan serangkaian iterasi di mana kluster diperbarui berulang kali untuk mencapai konvergensi dan menghasilkan solusi yang

optimal. Iterasi dan penyesuaian berguna untuk memastikan bahwa algoritma K-Means menghasilkan kluster yang stabil dan mewakili pola yang sebenarnya dalam dataset.

## **5. *Dokumentasi***

Dokumentasi memegang peranan penting dalam analisis data mining clustering pada balita penderita gizi buruk di Puskesmas Sigambal menggunakan metode K-Means. Proses dokumentasi ini melibatkan pencatatan dan pengarsipan semua langkah-langkah, parameter, dan hasil analisis untuk memastikan transparansi, reproducibility, dan akuntabilitas dalam penelitian ini.