

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Tinjauan Pustaka**

##### **2.1.1 Data Mining**

Menurut Ardi Ramdani dan rekan-rekan (2021), data mining dapat didefinisikan Sebagai tahapan dalam memperoleh informasi berharga dari basis pengetahuan, yang dikombinasikan dengan keterampilan dan keahlian analisis tingkat lanjut. Proses ini sangat krusial dalam mengungkap pola dan tren yang tersembunyi. Data yang diperoleh dapat dianalisis menggunakan beragam algoritma data mining, yang memainkan peran penting dalam mengintegrasikan teknologi dan penelitian.

Proses ini mencakup identifikasi aturan asosiasi dan klasifikasi, serta deteksi yang bekerja dengan kategorisasi untuk menangani hasil yang beragam, yakni buruk, sedang, dan baik. Data mining menjadi metode yang sangat berharga berkat kemampuannya mengubah sejumlah besar data menjadi informasi dan pengetahuan yang bermanfaat. Wawasan dan informasi yang diperoleh melalui proses ini dapat diterapkan dalam berbagai bidang, seperti analisis pasar, deteksi penipuan dan loyalitas pelanggan, pengendalian produksi, serta penelitian di bidang eksplorasi ilmu pengetahuan.

Menurut (Surahman & Hayati, 2023) Penambangan data merupakan Tahapan analisis dan evaluasi sejumlah data yang digunakan untuk mengidentifikasi kaitan-kaitan yang tidak terduga. Proses ini bertujuan menyajikan Informasi yang disajikan secara jelas dan berguna bagi pemiliknya. Dalam penambangan data, informasi

diambil dan diidentifikasi untuk menghasilkan wawasan yang dapat ditindaklanjuti, menggunakan berbagai teknik seperti statistik, matematika, kecerdasan buatan, dan pembelajaran mesin yang diperoleh dari kumpulan data besar.

### **2.1.2 Bantuan Sosial**

Bantuan sosial merujuk pada dukungan yang diberikan oleh pemerintah daerah, yang bisa berupa uang maupun barang, kepada individu, keluarga, kelompok, atau masyarakat. Bantuan ini tidak bersifat tetap dan diberikan secara selektif. Tujuan dari pendekatan ini adalah untuk melindungi masyarakat dari potensi risiko sosial yang mungkin muncul. (Silvanie, dkk 2021).

Bantuan sosial adalah bentuk dukungan yang diberikan dalam bentuk uang, barang, atau jasa kepada individu, keluarga, kelompok, atau masyarakat yang berada dalam kondisi miskin, tidak mampu, dan/atau rentan secara sosial. Tujuannya adalah untuk membantu mereka memenuhi kebutuhan dasar dalam kehidupan sehari-hari. (Mastur Alfitri & Rusda, 2023).

### **2.1.3 Konsep Dasar Website**

#### **2.1.3.1 HTML (HyperText Markup Language)**

HTML (*HyperText Markup Language*) merupakan salah satu bahasa yang digunakan untuk mengembangkan situs web, termasuk dalam Sistem Kearsipan Pelayanan Publik ini. Dengan menggunakan tag-tag tertentu, HTML menyusun kode-kode yang kemudian diinterpretasikan oleh browser, sehingga halaman bisa ditampilkan dengan tata letak yang diinginkan. Bahasa HTML berfungsi untuk membangun struktur dasar halaman web secara teratur sebelum melanjutkan ke tahap desain dan pengembangan fungsionalitas. Selain itu, HTML biasanya

dipadukan dengan bahasa pemrograman CSS untuk menciptakan tampilan yang lebih menarik.

#### **2.1.3.2 PHP (HyperText Preprocessor)**

PHP merupakan bahasa pemrograman yang mendukung HTML dan memungkinkan pembuatan aplikasi dinamis untuk pengolahan dan pemrosesan data. Sintaksis yang disediakan dijalankan sepenuhnya di server, sementara hasil akhirnya yang akan dikirim ke browser. Dengan kata lain, PHP berfungsi sebagai bahasa skrip yang diletakkan dan diproses di server. Hasil pemrosesan tersebut kemudian dikirim ke klien, di mana pengguna akan mengaksesnya Melalui browser, PHP dikenal sebagai bahasa skrip yang terintegrasi dengan tag HTML dan dijalankan di sisi server, serupa dengan Active Server Pages (ASP) atau Java Server Pages (JSP) yang digunakan untuk membuat halaman web dinamis. Selain itu, PHP juga merupakan perangkat lunak dengan lisensi sumber terbuka. (Hermiati, 2021).

#### **2.1.3.3 Javascript**

JavaScript diperkenalkan pertama kali oleh Netscape pada tahun 1995. Di awal kemunculannya, bahasa pemrograman ini dikenal dengan nama "LiveScript" dan dirancang untuk digunakan dalam peramban Netscape Navigator 2. JavaScript terdiri dari serangkaian kode yang dapat dieksekusi dalam dokumen HTML. Seiring perkembangan sejarah internet, JavaScript tumbuh menjadi bahasa skrip utama dalam pengembangan web. Lebih dari itu, JavaScript juga memungkinkan penggunaan perintah event, yang tidak hanya memperluas kemampuan HTML, tetapi juga memberikan fleksibilitas dan interaktivitas yang lebih besar dalam pembuatan aplikasi web. (Christian, 2024).

#### **2.1.3.4 Bootstrap**

Bootstrap merupakan framework yang mengintegrasikan HTML, CSS, dan JavaScript, yang dirancang untuk mempermudah dan mempercepat pengembangan situs web responsif. Framework ini dikembangkan oleh Mark Otto dan Jacob Thornton dari Twitter pada tahun 2011, framework open source ini sebelumnya dikenal dengan sebutan Twitter Blueprint. Sejak itu, popularitas Bootstrap melesat dengan lebih dari 27% situs web di seluruh dunia memilihnya sebagai solusi pengembangan mereka. Keberhasilan ini tidak terlepas dari konsistensi yang ditawarkan Bootstrap dibandingkan dengan framework lain pada masa itu. Salah satu keunggulan Bootstrap adalah kemampuannya untuk mengeliminasi kebutuhan pengkodean komponen situs web dari nol, berkat kumpulan file CSS dan JavaScript yang disusun dalam bentuk kelas yang dapat digunakan oleh para programmer. (Damanik, dkk., 2022).

#### **2.1.3.5 JQuery**

Menggunakan jQuery memiliki keuntungan jika dibandingkan dengan JavaScript standar, jQuery menyederhanakan penulisan kode JavaScript dengan memanfaatkan berbagai fungsi yang disediakan. JavaScript sendiri adalah bahasa skrip yang beroperasi di sisi klien atau peramban, sehingga dapat meningkatkan interaktivitas situs web. jQuery pertama kali diluncurkan pada tahun 2006 oleh John Resig dan sejak itu telah meraih popularitas yang luar biasa. Banyak situs web, termasuk yang terkenal seperti Google, Amazon, Twitter, ESPN, dan lainnya, telah mengadopsinya dalam pengembangan mereka. (Bengkulu, dkk., 2021).

#### **2.1.4 CSS (*Cascading Style Sheets*)**

CSS (*Cascading Style Sheets*) adalah bahasa stylesheet yang digunakan untuk mengatur tampilan dokumen yang ditulis dalam HTML. Dengan CSS, Anda dapat mengontrol berbagai elemen, seperti ukuran gambar, warna teks, warna tabel, dan masih banyak lagi. (Asrurin, 2021).

#### **2.1.5 Chrome**

Google Chrome adalah peramban web lintas platform yang dikembangkan oleh Google. Peramban ini pertama kali diperkenalkan pada tahun 2008 untuk sistem operasi Microsoft Windows, Chrome kemudian diperluas ke berbagai sistem operasi lainnya, termasuk Android, iOS, Linux, dan macOS, sehingga menjadi peramban default pada platform-platform tersebut. Selain itu, Chrome juga berperan sebagai Komponen utama dalam Chrome OS yang berperan sebagai platform untuk aplikasi web.

#### **2.1.6 Visual Studio Code**

Visual Studio Code, atau VCode, adalah editor kode sumber yang dikembangkan oleh Microsoft dan dapat digunakan di sistem operasi Linux, macOS, dan Windows, VCode menawarkan berbagai fitur canggih, seperti penyorotan sintaksis, penyelesaian kode otomatis, refactoring, audit, dan integrasi dengan Git. Kode sumber Visual Studio Code dapat diakses di repositori GitHub di bawah Lisensi MIT (Kode - OSS), sementara versi biner yang dirilis oleh Microsoft memiliki lisensi terpisah dan termasuk dalam kategori perangkat lunak berpemilik. Visual Studio Code pertama kali diluncurkan oleh Microsoft pada 29 April 2015, dalam acara konferensi Build 2015, dan versinya yang pratinjau segera menyusul setelah itu.

### **2.1.7 Database**

Basis data dapat diartikan sebagai kumpulan file yang saling terhubung atau kumpulan rekaman yang terorganisir dan saling terkait, yang dirancang untuk menyimpan data serta hubungan antar file. Dalam basis data, terdapat entitas, atribut, dan hubungan. Entitas merujuk pada berbagai objek yang ada dalam basis data, seperti orang, tempat, benda, konsep, atau peristiwa merupakan contoh entitas, sementara atribut merujuk pada ciri atau karakteristik yang menggambarkan aspek-aspek dari entitas tersebut disimpan. Sementara itu, hubungan menggambarkan keterkaitan antara satu entitas dengan entitas lainnya. (Setiawan & Fatimah, 2021).

#### **2.1.7.1 PostgresSQL**

PostgreSQL merupakan sistem manajemen basis data yang menggabungkan konsep relasional dan berorientasi objek. (ORDBMS) sumber terbuka yang dirancang untuk menangani berbagai jenis beban kerja, dari aplikasi kecil hingga besar dan kompleks. Dikenal karena kepatuhan terhadap standar SQL dan ekstensibilitasnya, PostgreSQL mendukung fitur transaksional penuh (ACID) yang memastikan konsistensi dan keandalan data. Selain itu, PostgreSQL mendukung tipe data berorientasi objek dan berbagai ekstensi seperti TimescaleDB untuk data seri waktu dan PostGIS untuk data spasial, memungkinkan fleksibilitas dalam menangani berbagai jenis data dan aplikasi. PostgreSQL juga menyediakan fitur keamanan yang kuat dan dukungan replikasi, termasuk replikasi streaming dan ekstensi seperti Citus untuk distribusi data dan kueri ke beberapa node, memastikan skalabilitas dan ketersediaan data yang tinggi. Sebagai perangkat lunak bebas dan open-source, PostgreSQL terus berkembang dengan kontribusi dari komunitas

pengembang global, menjadikannya salah satu pilihan utama untuk sistem manajemen basis data di berbagai industri.

### **2.1.8 Aplikasi**

Aplikasi merupakan unit perangkat lunak yang dikembangkan untuk memenuhi kebutuhan aktivitas tertentu, seperti perdagangan, permainan, layanan komunitas, periklanan, atau proses-proses lain yang biasa dilakukan oleh manusia, disebut aplikasi perangkat lunak. (Bengkulu, dkk., 2021).

### **2.1.9 Decision Support System**

#### **2.1.8.1 Definisi Decision Support System (DSS)**

Sistem Pendukung Keputusan (DSS) merupakan sebuah sistem informasi yang dirancang untuk membantu dalam proses pengambilan keputusan dengan menyediakan data dan informasi yang relevan dengan menyediakan berbagai alat, model, serta data yang relevan. Tujuan utama dari DSS adalah membantu para pengambil keputusan dalam menganalisis situasi yang kompleks sehingga memungkinkan pengguna untuk membuat keputusan yang lebih tepat dan efektif. DSS dapat digunakan untuk menganalisis data historis, memprediksi tren di masa depan, mengevaluasi berbagai alternatif, serta menyediakan informasi yang diperlukan dalam memecahkan masalah atau mencapai tujuan tertentu. Dengan menggunakan DSS, pengguna dapat dengan cepat mengakses dan mengolah informasi, yang pada akhirnya memungkinkan penggunanya untuk membuat keputusan dengan cara yang lebih efisien dan efektif.

#### **2.1.8.2 Karakteristik *Decision Support System* (DSS)**

Beberapa karakteristik umum dari Sistem Pendukung Keputusan (DSS):



- 1) Interaktif: DSS memungkinkan pengguna untuk berinteraksi langsung dengan sistem, memeriksa data, mengubah parameter, dan melihat hasil analisis secara real-time.
- 2) Dukungan Multi-level: DSS dapat digunakan di berbagai tingkatan dalam organisasi, dari tingkat operasional hingga manajerial dan eksekutif. Langkah ini diambil dengan memberikan informasi yang sesuai dan relevan dengan kebutuhan pengguna.
- 3) Analisis Berbasis Model: DSS menggunakan model matematika dan statistik untuk menganalisis data dan membuat proyeksi atau simulasi untuk mendukung proses pengambilan keputusan.
- 4) Fleksibel: DSS dirancang untuk mudah disesuaikan dengan perubahan kebutuhan pengguna dan lingkungan bisnis. Ini dapat memperhitungkan variabilitas dalam data dan proses pengambilan keputusan.
- 5) Integrasi Data: DSS mampu menggabungkan data dari berbagai sumber, baik yang berasal dari dalam organisasi maupun dari sumber eksternal, untuk mendukung pengambilan keputusan yang lebih baik. Untuk memberikan pandangan yang holistik terhadap situasi yang sedang dianalisis.
- 6) Penggunaan Pengetahuan: DSS dapat memanfaatkan pengetahuan yang tersimpan dalam bentuk aturan bisnis, pengalaman, atau keahlian pakar untuk membantu dalam proses pengambilan keputusan.
- 7) Responsif terhadap Kecepatan: DSS dirancang untuk memberikan respons yang cepat terhadap permintaan pengguna, sehingga pengguna dapat menerima informasi yang diperlukan dalam waktu singkat.



- 8) Dukungan Grafis: DSS sering dilengkapi dengan antarmuka grafis yang intuitif dan visualisasi data yang membantu pengguna memahami informasi dengan lebih baik.
- 9) Dukungan Kolaborasi: Beberapa DSS dapat mendukung kolaborasi antara pengguna, yang memberi kesempatan bagi mereka untuk menyampaikan informasi, ide, dan pemikiran dalam proses pengambilan keputusan.
- 10) Keamanan: DSS menyediakan kontrol akses yang ketat untuk melindungi data sistem ini juga dirancang untuk menjaga kerahasiaan data, memastikan bahwa hanya pengguna yang berwenang yang dapat mengakses informasi tertentu dengan tingkat keamanan yang tinggi.

#### 2.1.9 Confusion Matrix

Manurut (Khotimah & Utami, 2022) Confusion Matrix adalah sebuah alat yang digunakan untuk menilai kinerja model klasifikasi dengan cara membandingkan hasil prediksi model terhadap data aktual yang sebenarnya yang telah dibangun. Dalam matriks ini, hasil prediksi kelas dibandingkan dengan kelas data aktual. Data yang diperoleh kemudian digunakan untuk menghitung berbagai metrik evaluasi model seperti akurasi, presisi, recall, dan f-score. Hasil evaluasi yang diperoleh dari matriks kebingungan dapat disajikan dalam bentuk tabel berikut:

Tabel 2. 1 Confusion Matrix

Data Aktual	Data Prediksi		
	TRUE	FALSE	TOTAL
TRUE	TP	FN	P

FALSE	FP	TN	N
TOTAL	P	N'	P+N

Catatan:

- TP (True Positive) adalah data positif yang diklasifikasikan dengan benar.
- TN (True Negative) adalah data negatif yang diklasifikasikan dengan benar.
- FP (False Positive) adalah data negatif yang salah diklasifikasikan sebagai positif.
- FN (False Negative) adalah data positif yang salah diklasifikasikan sebagai negatif.

#### 2.1.10 Naïve Bayes

Klasifikasi Naif Bayes, atau sering disebut juga sebagai Klasifikasi Bayesian, merupakan teknik statistik untuk mengklasifikasikan data dengan memprediksi probabilitas keanggotaan suatu kategori. Metode ini berlandaskan pada teorema Bayes, yang memungkinkan proses klasifikasi dengan hasil yang setara dengan pohon keputusan dan jaringan saraf. Salah satu keunggulan Klasifikasi Naif Bayes adalah kinerjanya yang sangat baik dan cepat saat diterapkan pada basis data besar. Algoritma ini mampu mengukur probabilitas keanggotaan kelompok dengan efektif. Secara keseluruhan, Klasifikasi Naif Bayes terbukti menawarkan tingkat akurasi dan kecepatan yang sangat mengesankan, terutama ketika digunakan pada data dalam jumlah besar. (Surahman & Hayati, 2023).

Menurut (Firdaus & Setiadi, 2023.) Naive Bayes merupakan salah satu teknik statistik yang digunakan untuk penalaran induktif dalam masalah klasifikasi. Metode ini mengandalkan probabilitas yang diperoleh melalui teorema Bayes. Sebelum mengaplikasikan teorema ini dalam konteks klasifikasi pada penambahan data, penting untuk memahami terlebih dahulu konsep dasar dan definisi dari teorema Bayes itu sendiri.

Algoritma Naive Bayes dapat dijelaskan dengan persamaan berikut:

$$P(H|X) = P(X|H)P(H) / P(X)$$

Di mana:

- X merujuk pada data yang kelasnya belum diketahui (label target).
- H adalah kelas hipotetis yang diinginkan untuk diuji.
- $P(H|X)$  adalah probabilitas posterior, yaitu kemungkinan kelas H setelah mempertimbangkan data X.
- $P(X|H)$  adalah likelihood, yang menunjukkan probabilitas data X terjadi jika H benar.
- $P(H)$  adalah prior, yaitu probabilitas kelas H sebelum data X diamati.
- $P(X)$  adalah evidence, yaitu probabilitas keseluruhan data X.

Dengan menggunakan persamaan ini, Naive Bayes menghitung probabilitas untuk setiap kelas dan memilih kelas dengan probabilitas tertinggi sebagai prediksi.

### 2.1.11 C5.0

Algoritma C5.0, yang merupakan pengembangan dari pohon keputusan, adalah salah satu metode klasifikasi dalam pembelajaran mesin yang digunakan untuk membangun pohon keputusan. Merupakan versi terbaru dari algoritma C4.5 yang dikembangkan oleh Ross Quinlan, C5.0 menawarkan sejumlah perbaikan signifikan, memberikan efisiensi dan kinerja yang lebih baik dibandingkan dengan versi sebelumnya.

Pohon keputusan sendiri adalah struktur hierarkis yang digunakan untuk menggambarkan proses pengambilan keputusan dan konsekuensi yang mungkin terjadi melalui serangkaian aturan keputusan. Algoritma C5.0 bekerja dengan membangun pohon keputusan optimal untuk dataset yang ada dengan membagi data berdasarkan fitur-fitur relevan yang tersedia.

Menurut (Firdaus & Setiadi, 2023.) Pohon keputusan (decision tree) adalah algoritma yang menggunakan struktur seperti pohon untuk menganalisis data dan membuat keputusan berdasarkan informasi yang ada pada data tersebut. Menurut Ye dalam Sutoyo, tujuan utama dari algoritma ini adalah untuk mengidentifikasi pola klasifikasi dan prediksi dari sebuah dataset, serta menggambarkan hubungan antara variabel atribut  $x$  dengan variabel target  $y$  dalam bentuk pohon

Hafizan dan Putri menjelaskan bahwa pohon keputusan dapat digunakan untuk mengelompokkan data yang besar menjadi kelompok-kelompok lebih kecil dengan mengikuti aturan-aturan keputusan tertentu. Algoritma ini bekerja dengan

menggunakan konsep entropi dan gain yang menjadi bagian penting dalam perhitungannya. Berikut penjelasan mengenai kedua konsep tersebut:

a. Entropi

Entropi adalah ukuran yang menunjukkan tingkat ketidakpastian atau keragaman keputusan terkait dengan nilai atribut dalam data. Semakin tinggi nilai entropi, semakin besar ketidakpastian keputusan yang ada, yang akan memengaruhi perhitungan gain. Rumus untuk menghitung entropi adalah sebagai berikut:

$$\text{Entropi}(S) = - \sum_{i=1}^n p_i \cdot \log_2 p_i$$

Keterangan:

- $S$  = himpunan kasus
- $n$  = jumlah atribut  $A$
- $p_i$  = proporsi  $S_i$  terhadap  $S$ , yang dapat dihitung dengan  $\frac{|S_i|}{|S|}$
- $\log_2$  = logaritma basis 2

Entropi digunakan dalam pembangunan pohon keputusan untuk menentukan apakah suatu atribut memiliki cabang yang banyak atau tidak. Jika entropi suatu atribut adalah 0, berarti atribut tersebut tidak memerlukan cabang lebih lanjut. Sebaliknya, jika nilai entropi mencapai maksimum (1), atribut tersebut akan membutuhkan banyak cabang karena tingkat ketidakpastian data yang tinggi.

b. Gain (Information Gain)

Gain, atau yang dikenal juga dengan information gain, adalah metode yang digunakan untuk memilih atribut pada setiap node dalam pohon keputusan. Fungsi dari gain adalah untuk mengukur sejauh mana atribut dapat memengaruhi keputusan serta seberapa efektif variabel tersebut dalam mengklasifikasikan data ke dalam kelas-kelas yang berbeda. Rumus untuk menghitung gain adalah sebagai berikut:

$$\text{Gain}(S,A) = \text{Entropi}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot \text{Entropi}(S_i)$$

Keterangan:

- S merujuk pada himpunan kasus
- A adalah atribut
- n adalah jumlah atribut A
- $|S_i|/|S|$  menunjukkan jumlah kasus pada partisi ke-i
- $|S|$  adalah jumlah kasus dalam himpunan S

Setio, dkk. menjelaskan bahwa dalam algoritma pohon keputusan, nilai gain digunakan untuk menentukan atribut mana yang akan dipilih sebagai simpul dalam pohon keputusan. Atribut dengan nilai gain tertinggi akan dipilih sebagai simpul utama dalam pohon keputusan tersebut.

### 2.1.12 Klasifikasi

Klasifikasi dalam data mining merupakan proses pengelompokan objek ke dalam kategori atau kelas yang telah ditentukan sebelumnya. Proses ini sering

dimanfaatkan untuk memprediksi kelas tertentu. Ada beberapa komponen utama dalam proses klasifikasi, yaitu:

1. Kelas: Variabel dependen yang menjadi fokus dari klasifikasi.
2. Prediktor: Variabel independen yang berfungsi sebagai properti dari data yang akan dikategorikan.
3. Dataset Pelatihan: Sekelompok data yang mencakup variabel prediktor (independen) serta atribut yang sudah dikategorikan, yang digunakan untuk melatih model klasifikasi.
4. Dataset Pengujian: Data baru yang digunakan untuk menguji dan mengevaluasi kinerja model klasifikasi yang telah dilatih, dengan tujuan mengukur akurasi dari model tersebut.

Dengan memahami komponen-komponen ini, kita dapat lebih mudah melakukan klasifikasi dan mendapatkan hasil yang akurat. (Surahman & Hayati, 2023).

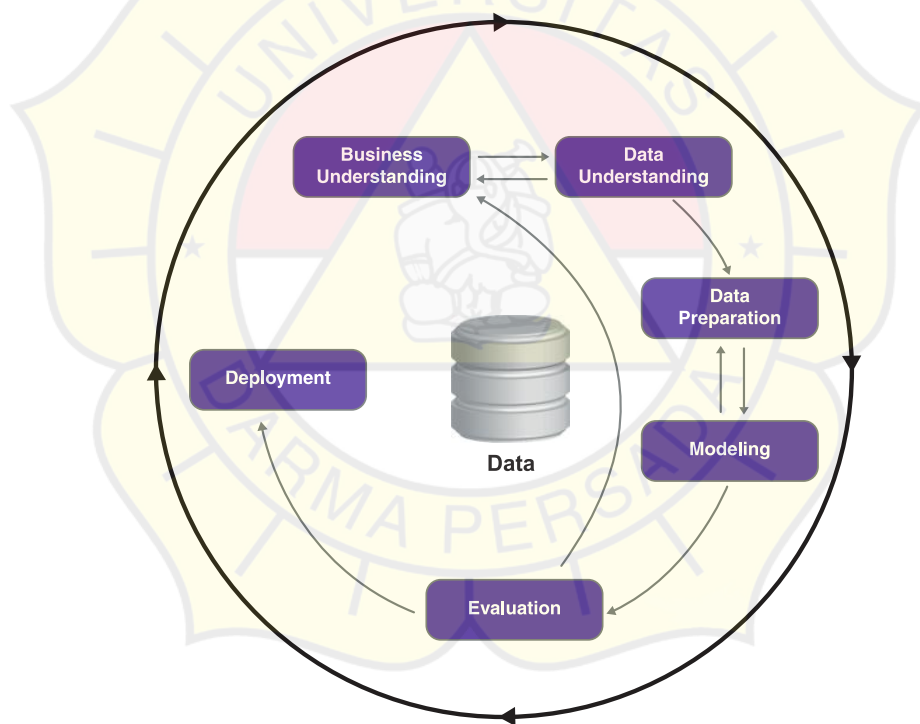
#### **2.1.13 CRISP-DM**

Menurut (Fadillah, 2015) Metode CRISP-DM adalah kerangka kerja standar dalam penambangan data yang dikembangkan oleh tiga pionir di bidang ini, yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), dan NCR. Pengembangan metode ini dilakukan lebih lanjut melalui berbagai lokakarya antara tahun 1997 hingga 1999, di mana lebih dari 300 organisasi memberikan kontribusi terhadap proses pemodelan. Akhirnya, pada tahun 1999, versi CRISP-DM 1.0 resmi dirilis.

CRISP-DM, yang merupakan singkatan dari Cross Industry Standard Process for Data Mining, adalah sebuah metode standar yang digunakan dalam proyek



penambangan data untuk mengarahkan dan mengelola berbagai tahap analisis data. Kerangka kerja ini terdiri dari serangkaian fase yang terorganisir secara sistematis untuk membantu pengelolaan proyek penambangan data secara efektif. CRISP-DM membantu para profesional untuk mengorganisir langkah-langkah yang terlibat dalam menambang data, mulai dari pemahaman masalah bisnis hingga penerapan solusi yang ditemukan kembali ke dalam lingkungan bisnis. Kerangka kerja ini mencakup enam tahapan utama: memahami masalah bisnis, memahami pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi, termasuk pengenalan gambar atau klasifikasi gambar.



Gambar 2. 1 Gambar Proses CRISP DM

#### 1. Pemahaman bisnis (*Business Understanding*)

Tahap pertama ini berfokus pada pemahaman yang mendalam mengenai masalah bisnis yang ingin diselesaikan atau tujuan yang ingin dicapai melalui proyek penambangan data. Ini mencakup mengidentifikasi tujuan

bisnis, kebutuhan pemangku kepentingan, serta konsep solusi yang diinginkan.

2. Pemahaman data (*Data Understanding*)

Tahapan kedua adalah memahami data yang tersedia untuk proyek. Ini melibatkan pengumpulan data, pemahaman karakteristik data, serta evaluasi kualitas data.

3. Persiapan data (*Data Preparation*)

Tahapan ketiga yaitu melibatkan persiapan data untuk analisis lebih lanjut. Ini bisa mencakup membersihkan data, mengintegrasikan data dari berbagai sumber, dan mengubah format data agar sesuai dengan kebutuhan analisis.

4. Pemodelan (*Modeling*)

Sepanjang tahapan ini melibatkan pengembangan model atau teknik analitis untuk mengekstraksi informasi atau pola dari data. Ini bisa mencakup penerapan algoritma machine learning, statistik, atau teknik analisis data lainnya.

5. Evaluasi (*Evaluation*)

Setelah model dibangun, tahap selanjutnya adalah evaluasi, yang berfokus pada menilai hasil dari model yang telah dibuat. Proses ini melibatkan penerapan metrik evaluasi untuk mengukur efektivitas kinerja model dan memastikan bahwa solusi yang dihasilkan sesuai dengan tujuan bisnis yang telah ditetapkan pada tahap sebelumnya.

6. Deployment (*Penyebaran*)

Pada tahapan terakhir adalah implementasi solusi yang ditemukan kembali ke dalam lingkungan bisnis. Ini melibatkan penerapan model atau solusi

yang dihasilkan ke dalam operasi bisnis yang sesungguhnya, serta memastikan bahwa solusi tersebut dapat digunakan secara efektif oleh pemangku kepentingan yang relevan.

## 2.1.14 Permodelan UML

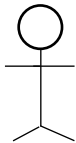
### 2.1.14.1 UML

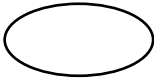

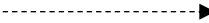
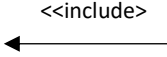

Bahasa Pemodelan Pengembangan Sistem, yang juga dikenal sebagai Bahasa Pemodelan Terpadu, merupakan bahasa standar yang sangat populer di industri. Bahasa ini digunakan untuk mengidentifikasi kebutuhan, melakukan analisis dan perancangan, serta menggambarkan arsitektur dalam konteks pemrograman berorientasi objek.

### 2.1.14.2 Use Case Diagram

Diagram use case adalah representasi visual yang menggambarkan interaksi antara aktor (pengguna) dan sistem. Diagram ini menggambarkan bagaimana fungsionalitas sistem dilihat dari sudut pandang pengguna. Use case merujuk pada serangkaian skenario atau langkah-langkah yang menjelaskan bagaimana aktor berinteraksi dengan sistem untuk mencapai tujuan tertentu. Dalam diagram use case, hubungan antara aktor dan use case, serta interaksi antar use case, digambarkan dengan jelas.

Tabel 2. 2Simbol-simbol Diagram Use Case


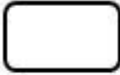

Simbol	Keterangan
	Aktor: Mewakili peran individu, sistem lain, atau perangkat dalam proses komunikasi terkait dengan penggunaan tertentu.




	Kasus penggunaan: Abstraksi serta interaksi antara sistem dan aktor.
	Asosiasi: Merupakan penggambaran abstrak mengenai hubungan antara aktor dan kasus penggunaan.
	Generalitas: Menetapkan spesialisasi aktor yang akan berkontribusi dalam kasus penggunaan.
	Menunjukkan bahwa suatu kasus penggunaan sepenuhnya mencerminkan fungsionalitas dari kasus penggunaan lainnya.
	Menunjukkan bahwa sebuah kasus penggunaan dapat dianggap sebagai perluasan fungsional dari kasus penggunaan lainnya ketika sebuah kondisi tertentu telah terpenuhi.

#### 2.1.14.3 Activity Diagram

Diagram aktivitas merupakan sebuah diagram alir yang memvisualisasikan alur kerja atau proses dalam suatu sistem. Diagram ini menggambarkan urutan langkah atau aktivitas yang terjadi, serta keputusan dan kondisi yang mempengaruhi aliran aktivitas tersebut. Umumnya, diagram aktivitas digunakan untuk merepresentasikan perilaku tingkat tinggi dari suatu sistem atau proses bisnis.

Tabel 2. 3 Simbol-simbol Activity Diagram

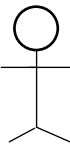
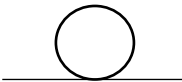
Simbol	Nama	Keterangan
	Status awal	Sebuah diagram aktivitas memiliki sebuah status awal.
	Aktivitas	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
	Percabangan / Decision	Percabangan dimana ada pilihan aktivitas yang lebih dari satu.

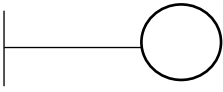
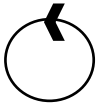


	Penggabungan / Join	Penggabungan dimana yang mana lebih dari satu aktivitas lalu digabungkan jadi satu.
	Status Akhir	Status akhir yang dilakukan sistem, sebuah diagram aktivitas memiliki sebuah status akhir
	Swimlane	Swimlane memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi.

#### 2.1.14.4 Sequence Diagram

Diagram sekuens adalah jenis diagram interaksi yang menunjukkan bagaimana objek-objek dalam suatu sistem saling berkomunikasi dalam konteks skenario tertentu. Diagram ini menyajikan urutan pesan atau pemanggilan metode yang dikirimkan antar objek dalam rentang waktu tertentu. Diagram sekuens sangat berguna untuk memodelkan interaksi dinamis antar objek dan menggambarkan aliran kontrol yang terjadi dari satu objek ke objek lainnya dalam sistem.

Tabel 2. 4 Simbol Sequence Diagram

NO	Gambar	Nama	Keterangan
1		Actor	Menunjukkan individu yang berinteraksi dengan sistem
2		Entity Class	Menampilkan hubungan yang akan terjadi

3		Boundary Class	Mengilustrasikan gambaran dari form
4		Control Class	Menjelaskan keterkaitan antara batas sistem dan tabel
5		A focus of Control & A Life Line	Menjelaskan titik awal dan akhir komunikasi
6		A message	Menggambarkan proses pengiriman pesan

## 2.2 Kajian Peneliti Terdahulu

### 1. Hasil Penelitian (Fatmawati, 2024)

pada penelitian (Fatmawati, 2024) berjudul **“Implementasi Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sosial”** Dalam penelitian ini, peneliti berusaha menangani masalah penentuan penerima bantuan sosial yang tepat. Dengan memanfaatkan metode Naïve Bayes, tujuan dari penelitian ini adalah untuk mengembangkan model prediksi yang dapat memastikan bahwa bantuan sosial disalurkan kepada pihak yang benar-benar membutuhkan. Penelitian ini bertujuan untuk meningkatkan efisiensi dan keadilan dalam proses penyaluran bantuan sosial. Yaitu dengan pendekatan berbasis metode Naïve Bayes, peneliti merancang model yang dimana mampu memprediksi kelayakan penerima bantuan sosial dengan akurasi tinggi, berdasarkan informasi yang tersedia.

Penelitian ini memanfaatkan metode Naïve Bayes yang sangat sesuai untuk mengatasi masalah klasifikasi terstruktur dalam kumpulan data yang kompleks.

Untuk mempermudah proses analisis data, penelitian ini juga menggunakan perangkat lunak RapidMiner, karena pendekatan manual dapat memakan waktu yang cukup lama. Fokus utama dari penelitian ini adalah mengklasifikasikan individu yang memenuhi syarat atau tidak memenuhi syarat untuk menerima bantuan sosial, berdasarkan kriteria tertentu seperti tingkat pendapatan dan status perkawinan.

Salah satu hasil utama dari penelitian ini adalah tingkat akurasi pengujian Klasifikasi yang dilakukan dengan menggunakan algoritma Naïve Bayes berhasil mencapai tingkat akurasi sebesar 86,36%, Hasil ini menunjukkan bahwa model tersebut mampu memprediksi kelayakan penerima bantuan sosial dengan efektif, sehingga berkontribusi pada peningkatan efisiensi dan keadilan dalam distribusi bantuan sosial.

Namun, penelitian ini memiliki kelemahan, terutama terkait dengan terbatasnya jumlah kriteria yang digunakan. Penambahan kriteria dapat berpotensi meningkatkan akurasi prediksi. Selain itu, jumlah data yang digunakan dalam proses data mining juga memengaruhi hasil pengujian, yang tentunya berperan penting dalam menentukan kelas pada data uji.

## 2. Hasil Penelitian (Lidysari dkk., 2022)

Pada Penelitian (Lidysari, dkk., 2022) berjudul **“Penerapan Data Mining Dalam Menentukan Kelayakan Penerima Bantuan Sosial Pemko Dengan Algoritma C4.5 (Kasus Kantor Kelurahan Martoba)”** Penelitian ini melibatkan pengolahan data dengan menerapkan teknik data mining, terutama melalui Penggunaan algoritma C4.5, yang merupakan metode klasifikasi berbasis teknik



pohon keputusan untuk menghasilkan model yang mudah dipahami dan memberikan tingkat akurasi yang cukup baik. Proses pengolahan data dilakukan melalui serangkaian langkah, termasuk pemilihan data, praproses, transformasi, penambangan data, dan interpretasi. Data yang digunakan dalam penelitian ini adalah data primer yang diperoleh langsung dari pengumpulan yang dilakukan oleh peneliti. Untuk analisis data, peneliti memanfaatkan Microsoft Excel dan RapidMiner sebagai alat bantu.

Tujuan utama penelitian ini adalah untuk merancang sistem yang dapat mendukung dalam proses membantu staf Kantor Desa Martoba dalam menentukan kelayakan penerima kesejahteraan PEMKO dengan cara yang lebih efektif dan efisien, serta meningkatkan akurasi dalam proses seleksi penerima bantuan kesejahteraan.

Metodologi yang digunakan dalam penelitian ini adalah metode data mining dengan memanfaatkan algoritma C4. 5, yaitu sebuah algoritma klasifikasi yang memanfaatkan teknik decision tree. Pendekatan ini memungkinkan pembuatan model yang mudah dipahami sekaligus menjamin tingkat akurasi yang memadai. Dalam proses pengolahan data, penelitian ini memanfaatkan Microsoft Excel untuk analisis awal dan RapidMiner untuk memperoleh hasil pengujian yang valid.

Temuan utama dari penelitian ini menunjukkan bahwa penerapan algoritma C4. 5 dapat meningkatkan tingkat akurasi dalam mengevaluasi kelayakan penerima bantuan sosial dari PEMKO. Selain itu, algoritma ini juga memberikan dukungan bagi staf kantor desa Martoba dalam menyeleksi penerima bantuan sosial dengan lebih efektif dan efisien.

Kelemahan pada penelitian ini adalah masih menggunakan C4.5 bukan C.5 yang Dimana C.5 sendiri mempunyai kelebihan dari C4.5 seperti:

- 1) Kecepatan dan Efisiensi
- 2) Penanganan Data yang Lebih Baik
- 3) Penggunaan Memori yang Lebih Efisien
- 4) Akurasi yang Lebih Tinggi

### 3. Hasil Penelitian (Surahman, 2023)

Pada Penelitian (Surahman, 2023) berjudul **“Implementasi Algoritma Naïve Bayes Untuk Prediksi Penerima Bantuan Sosial”** Dalam penelitian ini, peneliti membahas tujuan untuk melakukan prediksi penerima bantuan sosial di Desa Jatipancur dengan memanfaatkan algoritma Naive Bayes. Tujuan utama dari penelitian ini adalah untuk meningkatkan efektivitas program bantuan sosial sehingga dapat disalurkan dengan tepat kepada warga yang berhak. Dalam pengolahan data, peneliti memanfaatkan aplikasi RapidMiner untuk menerapkan algoritma Naive Bayes dalam memprediksi penerima bantuan sosial di desa tersebut.

Dengan adanya penelitian ini, diharapkan bahwa dana bantuan sosial ini dapat disalurkan secara akurat kepada masyarakat yang membutuhkan, sekaligus meningkatkan efisiensi penyaluran dana. Metodologi yang digunakan dalam penelitian ini meliputi observasi dan wawancara, serta proses Knowledge Discovery in Databases (KDD) sebagai langkah utama dalam analisis. Proses pengolahan data dalam penelitian ini terdiri dari beberapa langkah sebagai berikut

- 1) Pre-processing/Cleaning

- 2) Transformasi
- 3) Data mining
- 4) Evaluasi
- 5) Penyajian Hasil

Hasil utama dari penelitian ini menunjukkan bahwa penerapan algoritma Naïve Bayes berhasil meningkatkan akurasi dalam menentukan kelayakan penerima bantuan sosial, dengan tingkat akurasi mencapai 91,10% dengan class precision dan class recall sebesar 91.85% menggunakan algoritma Naive Bayes. serta dapat membantu Puskesmas Amanah desa jatipancur dalam proses seleksi penerima bantuan sosial dengan lebih efektif dan efisien.

Kelemahan dalam penelitian ini terletak pada keterbatasan data, di mana penelitian ini mengandalkan data sekunder yang mungkin tidak sepenuhnya menggambarkan kondisi yang aktual atau lengkap dari Puskesmas Amanah desa Jatipancur, yang mungkin memiliki keterbatasan dalam hal kualitas dan kelengkapan. Data sekunder ini mungkin tidak mencakup semua variabel atau faktor yang relevan dalam memprediksi penerima bantuan sosial.