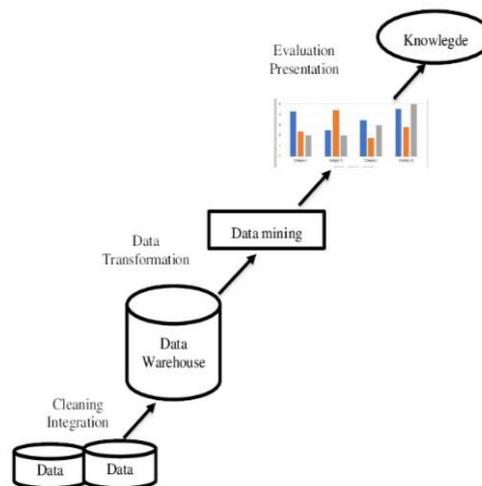


## BAB II KERANGKA TEORITIS

### A. Landasan Teori

#### 1. Pengertian Data Mining

Menurut (Turban 1995, 3), Data mining adalah proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam database besar. Serangkaian proses dalam data mining terdapat beberapa tahap (Builolo, 2020, pp. 2–3), yaitu:



**Gambar 2.1 Tahap Data Mining**

**(Sumber: Builolo, 2020)**

- (1) Data, yang harus dipersiapkan pertama kali dalam proses KDD adalah data. Data yang digunakan adalah data yang sudah terpisah dengan data operasional;
- (2) Data Selection, melakukan pemilihan data yang meliputi pembuatan kumpulan data target, penentuan variabel, pemilihan sampel data dan penyimpanan data pada sebuah berkas;
- (3) Pre-processing/Cleaning, proses cleaning meliputi pembuangan duplikasi data, perbaikan data yang inkonsisten, perbaikan kesalahan data, proses memperkaya data dengan menambah informasi lain yang relevan;
- (4) Data Transformation, data yang sudah tersedia digunakan untuk proses KDD dan diubah terlebih dahulu sesuai dengan algoritma/metode/teknik yang dipergunakan dalam data mining;

- (5) Data mining, tahap proses utama dalam penggalian dan pencarian pengetahuan dan informasi yang bermanfaat dengan menggunakan algoritma/metode/teknik tertentu sesuai dengan pengetahuan atau informasi yang dicari;
- (6) Interpretation/Evaluation, hasil dari proses data mining ditampilkan kedalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan seperti informasi ditampilkan dalam bentuk grafik, pohon keputusan ataupun dalam bentuk rule;
- (7) Knowledge, memperoleh pengetahuan atau informasi yang bermanfaat dan mudah dimengerti yang dihasilkan diimplementasikan sesuai dengan manfaat/kegunaan pengetahuan atau informasi tersebut.

## **2. Pengertian Klasifikasi**

Menurut (Prasetyo 2013, 45), klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Klasifikasi dapat didefinisikan secara detail sebagai pekerjaan yang melakukan pelatihan atau pembelajaran terhadap fungsi target / yang memetakan setiap set atribut (fitur)  $x$  ke satu dan sejumlah label kelas  $y$  yang tersedia. Pekerjaan pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori.

Klasifikasi merupakan penggalian data yang menetapkan item dalam koleksi untuk menargetkan kelas tertentu (Indah Werdiningsih, Barry Nuqoba and Muhammadun, 2020, p. 19). Klasifikasi dimulai dengan mengumpulkan data di mana data tersebut terdapat kelas yang sudah diketahui. Pada klasifikasi bayesian secara statistik mengacu pada teorema bayes untuk melakukan prediksi probabilitas keanggotaan suatu class. Klasifikasi dapat sebagai suatu kegiatan penataan pengetahuan secara universal kedalam berapa susunan sistematis

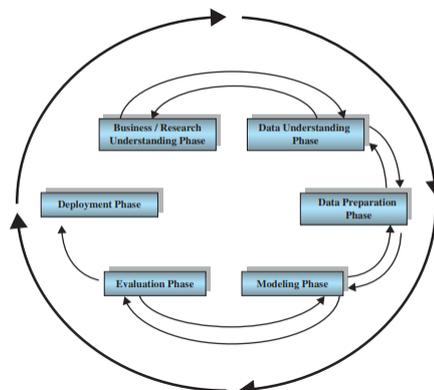
## **3. Pengertian Database**

Menurut (Connolly 2010, 54-66) database merupakan suatu pembagian kumpulan data yang berisi secara logika, dan keterangan dari masing-masing data yang didesain. Untuk mendapatkan informasi yang dibutuhkan sebuah organisasi. Database system yaitu kumpulan program aplikasi yang berinteraksi dengan basis data bersama dengan Database Management System (DBMS) dan basis data itu sendiri, sedangkan Database Management System (DBMS) yaitu sistem perangkat lunak yang memungkinkan pengguna untuk mendefinisikan, membuat, memelihara dan kontrol ke akses database.

#### 4. Tahapan CRISP-DM

Menurut (Larose 2014, 4-6), CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah metodologi yang banyak digunakan untuk pengembangan proyek data mining. Terdapat siklus hidup CRISP-DM yang terdiri dari enam fase dan penjelasannya sebagai berikut;

- (1) *Business Understanding Phase* (Fase Pemahaman Bisnis) merupakan tahapan dilakukan dengan menentukan tujuan proyek dan kebutuhan dalam lingkup unit penelitian atau bisnis secara keseluruhan, serta menerjemahkan tujuan dan batasan menjadi formula dari masalah data mining, dan menciptakan strategi awal untuk mencapai tujuan;
- (2) *Data Understanding Phase* (Fase Pemahaman Data) merupakan tahapan dilakukan pengumpulan data yang kemudian dianalisis untuk mendapatkan informasi tentang pola dan kualitas data;
- (3) *Data Preparation Phase* (Fase Persiapan Data) merupakan tahapan menganalisis dan menyiapkan data dengan menyesuaikan kasus pada beberapa variabel yang diperlukan untuk tahapan pemodelan;
- (4) *Modelling Phase* (Fase Pemodelan) merupakan tahapan yang dilakukan pemilihan dan pengaplikasian teknik yang menyesuaikan pemodelan dengan kasus serta permasalahan;
- (5) *Evaluation Phase* (Fase Pengujian) merupakan tahapan di mana model yang dirancang dievaluasi untuk mengidentifikasi masalah penting yang terkait dengan bisnis atau penelitian dan untuk mengevaluasi apakah model tersebut telah mencapai tujuan bisnisnya atau tidak.
- (6) *Deployment Phase* (Fase Penyebaran) merupakan tahapan di mana model data mining digunakan dengan melibatkan penerapan proses data mining pada perusahaan atau pembuatan laporan tentang hasilnya.



**Gambar 2. 1 Alur CRISP-DM**  
(Sumber: Larose, 2005)

## 5. Metode Naïve Bayes

Menurut (Olson 2008, 102) menjelaskan Naïve Bayes untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Naïve Bayes untuk setiap kelas keputusan menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “master” tabel keputusan. Metode ini menggunakan probabilitas bersyarat sebagai dasarnya.

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Keterangan:

X = data dengan class yang belum diketahui

H = hipotesis data merupakan suatu class spesifik

P(H|X) = probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) = probabilitas hipotesis H (prior probabilitas)

P(X) = probabilitas X

P(X|H) = probabilitas X berdasarkan kondisi pada hipotesis H

Berikut contoh kasus kondisi cuaca dan berolahraga dengan algoritma naïve bayes menurut (Hari Suparwito *et al.*, 2023, pp. 118–121), dengan kasus menghitung probabilitas berolahraga dan tidak berolahraga dengan mengklasifikasikan apakah akan berolahraga atau tidak tergantung pada kondisi cuaca.

- (1) Menentukan dataset penelitian, untuk dataset yang dipilih yaitu data cuaca dan data apakah akan berolahraga atau tidak;

Cuaca	Olahraga
Cerah	Tidak
Cerah	Tidak
Mendung	Ya
Hujan	Ya
Hujan	Ya
Hujan	Tidak

Cuaca	Olahraga
Mendung	Ya
Cerah	Tidak
Cerah	Ya
Hujan	Ya
Cerah	Ya
Mendung	Ya
Mendung	Ya
Hujan	Tidak

- (2) Menyederhanakan perhitungan sebelum (prior) dan sesudahnya (posterior) akan digunakan tabel frekuensi. Tabel Frekuensi menunjukkan seberapa sering label muncul untuk setiap fitur;

Cuaca	Tidak	Ya
Mendung	0	4
Cerah	2	3
Hujan	3	2
Total	5	9

Dari tabel di atas, dihitung probabilitas sebelumnya (prior) sebagai berikut:

Cuaca	Tidak	Ya	P(Cuaca)
Mendung	0	4	$4/14 = 0.29$
Cerah	2	3	$5/14 = 0.36$
Hujan	3	2	$5/14 = 0.36$
Total	5	9	
	$5/14 = 0.36$	$9/14 = 0.64$	

Kemudian juga dihitung probabilitas posteriornya, sebagai berikut:

Cuaca	Tidak	Ya	Probabilitas Posterior Tidak	Probabilitas Posterior Ya
Mendung	0	4	$0/5 = 0$	$4/9 = 0.44$
Cerah	2	3	$2/5 = 0.4$	$3/9 = 0.33$
Hujan	3	2	$3/5 = 0.5$	$2/9 = 0.22$
Total	5	9		

- (3) Klasifikasi setiap nilai pengujian atau input, dengan menghitung probabilitas berolahraga ketika cuaca mendung.

(a) Probabilitas Berolahraga

Rumus Naïve Bayes nya adalah:

$$P(\text{Ya Mendung}) = (P(\text{Mendung} | \text{Ya}) P(\text{Ya})) / P(\text{Mendung})$$

Maka dihitung dulu probabilitas sebelumnya (prior):

$$P(\text{Mendung}) = 4/14 = 0.29 \quad P(\text{Ya}) = 9/14 = 0.64$$

Kemudian dihitung probabilitas posterior nya:

$$P(\text{Mendung} | \text{Ya}) = 4/9 = 0.44$$

Setelah mendapatkan probabilitas prior dan posterior nya, maka dengan mudah dapat dihitung probabilitas akan berolahraga jika cuaca mendung:

$$P(\text{Ya Mendung}) = (0.44 \cdot 0.64) / 0.29 = 0.98$$

(b) Probabilitas Tidak Berolahraga

Seperti sebelumnya, dapat dihitung juga probabilitas tidak berolahraga ketika cuaca mendung. Dihitung dulu probabilitas priornya:

$$P(\text{Mendung}) = 4/14 = 0.29 \quad P(\text{Tidak}) = 5/14 = 0.36$$

Kemudian dihitung probabilitas posteriornya:

$$P(\text{Mendung} | \text{Tidak}) = 0/5 = 0$$

(4) Setelah mendapatkan probabilitas prior dan posterior-nya, maka dengan mudah dapat dihitung probabilitas tidak akan berolahraga jika cuaca mendung:

$$P(\text{Tidak} | \text{Mendung}) = (0 \cdot 0.36) / 0.29 = 0$$

Dari kedua perhitungan di atas, maka terlihat bahwa probabilitas berolahraga mempunyai nilai yang tertinggi sehingga dapat disimpulkan bahwa jika cuaca mendung, tetap akan berolahraga.

## 6. Pemodelan Prototype pada SDLC

Menurut (Raymond McLeod 2007, 188), metode merupakan cara untuk melakukan sesuatu, sehingga SDLC sebuah proses logika yang digunakan oleh seorang system analyst untuk mengembangkan sebuah sistem informasi yang melibatkan requirements, validation, training dan pemilik sistem. SDLC identik dengan teknik pengembangan sistem waterfall, karena tahapannya menurun dari

atas kebawah. Masalah akan didefinisikan dalam tahapan-tahapan perencanaan dan analisis. Solusi-solusi alternatif diidentifikasi dan di evaluasi dalam tahap desain.



**Gambar 2. 2 SDLC Konsep**  
(Sumber: Raymond Mcleod, 2007)

## 7. Pengertian Sistem Pendukung Keputusan

Decision Support System (DSS) bukanlah alat atau tools yang digunakan untuk mengambil keputusan, melainkan sistem pendukung keputusan merupakan sebuah sistem yang akan dapat membantu untuk mengambil Keputusan (Sharda et al., 2021). Dilengkapi dengan sebuah informasi melalui data yang telah diproses dengan secara relevan, dalam mengambil keputusan untuk sebuah permasalahan dengan akurat serta efisien waktu. Decision Support System memiliki tujuan dalam menyiapkan sebuah informasi, menimbang, membimbing serta memberikan prediksi serta mengarahkan kepada pengguna informasi agar dapat melakukan pengambilan keputusan dengan lebih baik

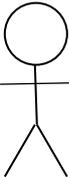
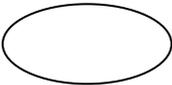
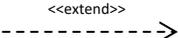
Menurut Rahman,A dalam (Diana, 2018, p. 19) menyatakan bahwa DSS merupakan implementasi teori-teori pengambilan keputusan yang telah diperkenalkan oleh ilmu-ilmu seperti operation research dan management science. Dahulu untuk mencari penyelesaian masalah yang dihadapi harus dilakukan perhitungan iterasi secara manual, namun dengan menggunakan computer untuk menyelesaikan suatu persoalan dengan waktu relatif singkat. Sistem pendukung keputusan dapat memecahkan masalah semi terstruktur dengan memberikan informasi atau saran mengenai keputusan tertentu, informasi tersebut dapat diberikan dalam bentuk laporan berkala, laporan khusus maupun model matematis. Model tersebut juga mempunyai kemampuan untuk memberikan saran dalam tingkat yang bervariasi.

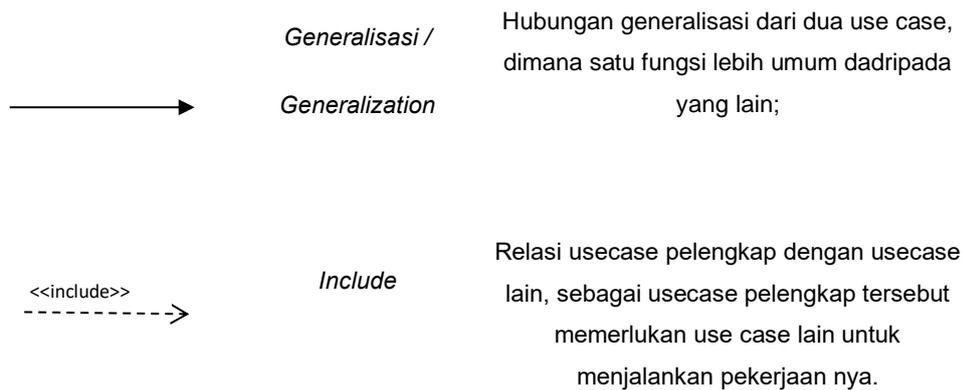
## 8. Pemrograman

Python pada dasarnya memiliki style yang serupa dengan pseudocode, perbedaannya hanyalah Python dapat dijalankan di computer dan menampilkan hasil (Dr. Asrul Huda 2020, 7). Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi. Python umumnya digunakan sebagai bahasa skrip meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa skrip. Dengan menggunakan bahasa pemrograman Python memungkinkan penggunaannya untuk ditulis secara ringkas, jelas, dan mudah dibaca.

## 9. UML (Unified Modeling Language)

UML (Unified Modeling Language) adalah alat bantu yang sudah menjadi standar dalam dunia pengembangan sistem perangkat lunak berorientasi objek (Anardani 2019, 13-14). UML menjadi bahasa yang handal dalam memvisualisasi rancangan sitem perangkat lunak. UML memungkinkan para pengembang sistem membuat blue print dalam bentuk yang baku dan mudah dimengerti sehingga bisa hasil rancangan bisa dikomunikasikan dengan pihak lain.

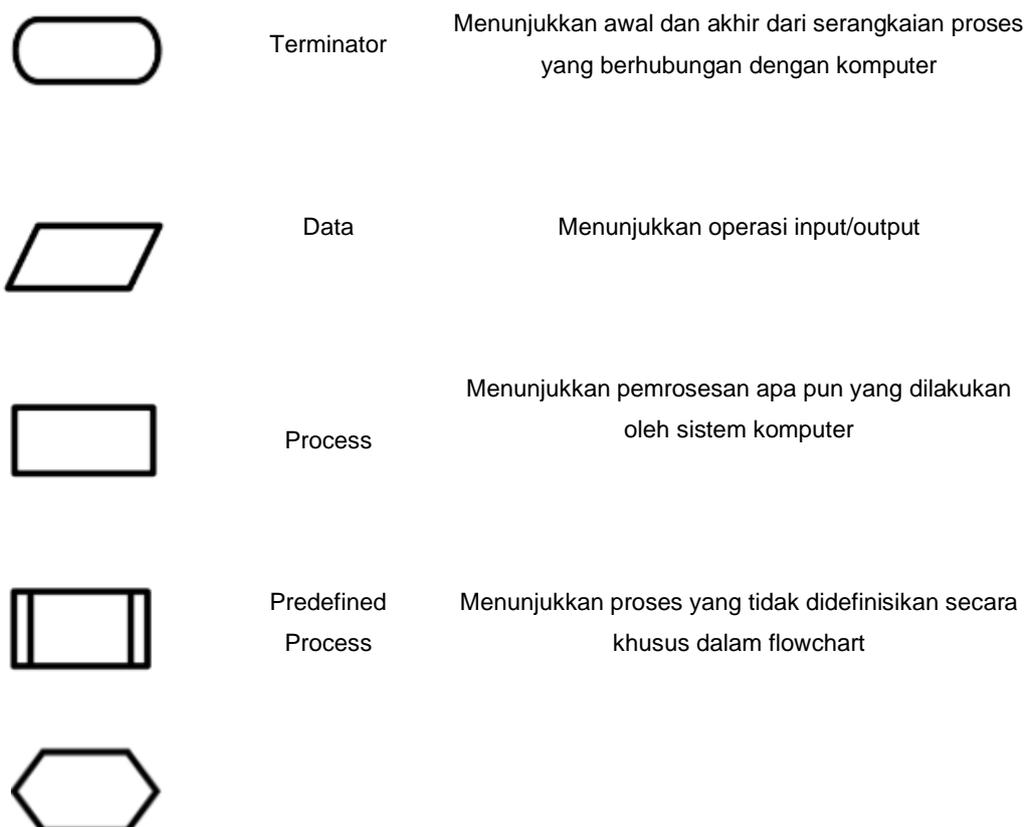
	Actor	Orang yang digunakan untuk menjelaskan siapa saja pengguna yang terdapat pada system;
	Use Case	Aktivitas yang dilakukan oleh pengguna sistem, biasanya menggunakan awalan kata kerja;
	Asosiasi/ Association	Relasi yang digunakan untuk interaksi antar use case dan actor;
	<i>Ekstensi/</i> <i>Extend</i>	Relasi yang digunakan untuk interaksi use case tambahan dengan use case;



**Gambar 2. 3 UML**  
**(Sumber: Anardani, 2019)**

### 10. Pengertian Flowchart

Flowchart dapat digunakan untuk menunjukkan urutan langkah untuk melakukan pekerjaan apa pun. Serangkaian operasi sederhana yang melibatkan penerimaan input, melakukan operasi aritmatika pada input, dan menunjukkannya kepada pengguna menunjukkan struktur logika urutan suatu program (Khairunnisa 2023, 89-90).



	Preparation	Memberi nilai awal dan akhir
	Arrow	Menghubungkan simbol-simbol
	Dokumen	Digunakan ketika input berasal dari dokumen dan output menuju dokumen
	Decision	Menunjukkan titik mana pun dalam proses di mana keputusan harus dibuat untuk menentukan tindakan lebih lanjut
	Connector	Menghubungkan bagian-bagian dari flowchart yang dilanjutkan pada halaman yang sama
	Off Page Connector	Menghubungkan bagian-bagian dari diagram alur yang dilanjutkan ke halaman terpisah
	Delay	Menunjukkan penundaan atau menunggu dalam proses untuk mendapatkan input dari beberapa proses lain

**Gambar 2. 4 Flowchart**  
**(Sumber: Khairunnisa, 2023)**

### 11. Pengertian Blackbox

Black box testing tidak berkonsentrasi pada perilaku internal dan struktur program tetapi black box testing adalah salah satu strategi pengujian penting dalam software testing yang berkonsentrasi menemukan keadaan bagaimana program tidak berjalan sesuai dengan spesifikasinya. Selanjutnya pada tahun 1988 Kaner mengatakan bahwa black box testing di gunakan untuk pengujian eksplorasi (Wicaksono 2021).

### 12. Pengertian PSSUQ

(Wahyuningrum, 2023, p. 54) Post Study System Usability Questionnaire (PSSUQ) adalah salah satu metode kuesioner, dimana metode PSSUQ pada awal

munculnya, PSSUQ terdiri dari 18 item pernyataan. Kemudian karena satu dari lima karakteristik belum tercakup, maka ditambahkan satu pernyataan tambahan menjadi 19 pada PSSUQ versi kedua. Setelah beberapa tahun penggunaan PSSUQ versi kedua, analisis item mengidentifikasi adanya realibilitas yang relatif kecil pada tiga pertanyaan, sehingga jumlah item-nya kembali berkurang menjadi 16. Berikut adalah nilai-nilai pertanyaan pada PSSUQ :

- (a) Overall yaitu rata-rata tanggapan untuk pernyataan 1 hingga 16.
- (b) System Quality (SysQual) merupakan faktor yang mengevaluasi kualitas sistem, dalam hal ini dihitung dari rata-rata pernyataan 1 hingga 6.
- (c) Information Quality (InfoQual) merupakan faktor yang digunakan untuk mengevaluasi kualitas informasi rata-rata dari pernyataan 7 hingga 12.
- (d) Interface Quality (IntQual) merupakan faktor yang digunakan untuk mengevaluasi kualitas antar muka rata-rata pernyataan 13 hingga 16.

## B. Tinjauan Pustaka

Berikut ini adalah asumsi dan keterbatasan pada penelitian yang dilakukan :

- (1) **Menurut penelitian** (Arifin, Handoko and Efendi, 2022), **dengan judul “ Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan ”**, Program Keluarga Harapan (PKH) merupakan program pemerintah dalam bentuk tunai untuk Rumah Tangga Sangat Miskin (RTSM) yang kualifikasinya terpenuhi terkait dengan upaya peningkatan kualitas manusia tersebut bukan dalam bidang pendidikan tetapi juga kesehatan; dalam jangka pendek Program PKH diperhitungkan bisa mengurangi biaya pengeluaran keluarga miskin serta mengurangi kemiskinan dalam jangka panjang; untuk menerima Program PKH) pemerintah sudah menetapkan beberapa kriteria, diantaranya penghasilan, status kepemilikan rumah, ukuran rumah, tipe lantai, atap, dinding, serta jenis sumber air; adapun cara untuk melakukan penyelesaian terhadap kriteria yang sudah ditetapkan yaitu dengan memanfaatkan teknik data mining melalui Metode Naïve Bayes.
- (2) **Menurut penelitian** (Hakam *et al.*, 2020) **dengan judul “ Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan ”**, lokasi bisnis memainkan peran penting dalam penjualan. Lokasi usaha di perkotaan menjadikan penjual lebih mudah mendistribusikan kegiatan untuk orang-orang; Kegiatan distribusi erat kaitannya dengan kegiatan penjualan; Apabila terdapat transaksi penjualan, klasifikasinya potensial dan tidak potensial pelanggan akan dibutuhkan; Salah satu metode yang dapat digunakan untuk klasifikasi adalah data mining; Satu dari

data mining yang paling sering digunakan untuk klasifikasi adalah metode Naive Bayes.

- (3) Menurut penelitian (Taufik *et al.*, 2023) dengan judul “ **Penerapan Metode Naive Bayes Untuk Mendiagnosa Penyakit ISPA Pada Anak** ”, ISPA (Infeksi Saluran Pernafasan Akut) adalah suatu penyakit gangguan saluran pernapasan yang dapat menimbulkan berbagai spektrum penyakit mulai dari penyakit tanpa gejala, infeksi ringan sampai penyakit yang parah dan mematikan akibat faktor lingkungan; Hasil yang diperoleh adalah terciptanya sistem pakar yang dapat mendiagnosa gangguan tidur pada remaja berbasis website yang dapat digunakan masyarakat dimana pun dan kapan pun selama terhubung dengan jaringan internet; Sistem pakar untuk mendiagnosa Penyakit ISPA pada anak dirancang melalui proses yang diawali dengan memperoleh pengetahuan pakar, kemudian merancang basis data sesuai data yang telah diperoleh; Sistem pakar untuk mendiagnosa Penyakit ISPA pada anak dapat diimplementasikan di dunia medis untuk dapat digunakan dalam pendiagnosaan Penyakit ISPA pada anak dengan terlebih dahulu terkoneksi dengan akses internet.
- (4) Menurut penelitian (Endah Widya Ningsih, 2020) dengan judul “ **Penerapan Algoritma Naïve Bayes Dalam Penentuan Kelayakan Penerima Kartu Jakarta Pintar Plus** ”, Kelayakan penerima Kartu Jakarta Pintar Plus masih belum tepat sasaran dikarenakan penerimaan yang masih bersifat subjektif. Oleh karenanya, peneliti membuat penelitian ini dengan menggunakan data yang sudah ada sebelumnya atau disebut dengan data training; Peneliti menggunakan metode Naïve Bayes dengan 7 atribut yang saling independen untuk menentukan kelayakan; Peneliti juga menggunakan aplikasi pendukung Rapidmined 5.3 untuk menguji keakurasian metode Naïve Bayes. Pengujian dilakukan dengan 134 data training dan 34 data testing; Metode algoritma Naïve Bayes memanfaatkan data training untuk menghasilkan probabilitas setiap kriteria untuk class yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi suatu kondisi berdasarkan proses klasifikasi yang dilakukan oleh metode Naïve Bayes itu sendiri.
- (5) Menurut (Adam Cahya Armadananto, Ahmad Bagus Setiawan, 2023) dengan judul “ **Klasifikasi Keluarga Penerima Bantuan Sosial Menggunakan Metode Naïve Bayes** ”, kemiskinan mempengaruhi kualitas sumber daya manusia suatu negara seperti pendidikan, kesehatan dan gizi; pada penelitian ini menggunakan metode Naïve Bayes dengan menggunakan sistem pengkalsifikasian ini dapat membantu perangkat desa dalam melakukan klasifikasi Keluarga Penerima Manfaat Bantuan Sosial yang sesuai dengan keadaan lapangan; hasil dari

pengujian terhadap metode Naïve Bayes berhasil diterapkan dengan melakukan klasifikasi data dengan sistem yang sudah dirancang maka didapatkan tingkat akurasi yang baik dari hasil perhitungan sistem ini dibuat dengan tampilan user friendly agar dapat memudahkan penggunaanya; dengan hasil persentase tinggi perhitungan menggunakan metode naïve bayes, maka hasil tersebut akan dapat memaksimalkan para petugas untuk memaksimalkan penyaluran bantuan sosial agar tepat sasaran.

- (6) Menurut (Dennis *et al.*, 2022), dengan judul “ **Penerapan Algoritma Naïve Bayes Untuk Pengelompokkan Predikat Peserta Uji Kemahiran Berbahasa Indonesia** ”, Uji Kemahiran Berbahasa Indonesia adalah uji kemahiran untuk mengukur kemahiran berbahasa seseorang dalam berkomunikasi dengan menggunakan bahasa Indonesia, baik penutur Indonesia maupun penutur asing; pemeringkatan UKBI memiliki 7 kategori pemeringkatan yang terdiri dari peringkat istimewa, sangat unggul, unggul, madya, semenjana, marginal, dan terbatas. Jumlah Peserta yang mengikuti UKBI di Balai Bahasa Provinsi Riau sudah lebih dari 1000 namun belum ada yang mengelola data tersebut menjadi sebuah pengetahuan baru; salah satu upaya yang bisa dilakukan dengan data tersebut adalah klasifikasi; Algoritma Naïve Bayes Classification merupakan salah satu algoritma klasifikasi yang sangat efektif (mendapatkan hasil yang tepat) dan efisien (proses penalaran dilakukan memanfaatkan input yang ada dengan cara yang relatif cepat).
- (7) Menurut penelitian (Surbakti, 2022), dengan judul “ **Implementasi Metode Naïve Bayes Untuk Diagnosa Penyakit Hati** ”, organ hati menjadi pelengkap bagi tubuh manusia sehingga bisa bekerja dengan baik didalam tubuh; penyakit gangguan hati ini sangat berbahaya apabila tidak dideteksi sejak dini dapat menyebabkan kronis dan berujung kematian; berdasarkan masalah diatas sehingga dibutuhkan suatu aplikasi berupa sistem pakar dalam mendiagnosa penyakit gangguan hati menggunakan metode naïve bayes; sistem ini dibangun dengan tujuan untuk membantu masyarakat umum guna untuk mendiagnosa penyakit gangguan hati secara dini, agar masyarakat umum untuk lebih cepat mengetahui jenis penyakit yang ada pada penyakit gangguan hati.
- (8) Menurut penelitian (Damuri *et al.*, 2021), dengan judul “ **Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako** ”, Kemiskinan adalah salah satu persoalan mendasar yang menjadi pusat perhatian pemerintah di suatu negara. Salah satu aspek penting untuk mendukung Strategi Penanggulangan Kemiskinan adalah tersedianya data kemiskinan yang akurat dan tepat sasaran; Naïve Bayes

merupakan salah satu metode yang dapat digunakan untuk mengklasifikasikan data; hasil klasifikasi yang dilakukan nantinya akan membantu pengelola bantuan untuk mengambil keputusan terkait klasifikasi penentuan penerima bantuan sembako. Prediksi penerima bantuan sembako yang digunakan terdapat dua kelas, yaitu layak dan tidak layak; data yang digunakan untuk prediksi yaitu data sampel dari desa XYZ sehingga penelitian ini menggunakan algoritma Naïve Bayes

- (9) Menurut penelitian** (Setiawan, Halilintar and Wahyuniar, 2021), **dengan judul “ Penerapan Metode Naive Bayes Untuk Klasifikasi Penentuan Penerima Bantuan PKH ”**, salah satu caranya yaitu dengan membuat sebuah program bantuan bernama Program dalam mengatasi kemiskinan di Indonesia; keluarga Harapan (PKH) dimana program bantuan ini diperuntukkan untuk rakyat miskin guna menanggulangi masalah kemiskinan di Indonesia; dengan menggunakan metode naive bayes dan laplace correction dimana dalam metode ini menggunakan kriteria yang telah ditentukan, kemudian proses perhitungan dimulai, setelah selesai perhitungan maka program akan menghasilkan data berupa keterangan penerima atau bukan penerima pada menu klasifikasi penerima PKH.
- (10) Menurut penelitian** (Cahyanti, Gata and Sarasati, 2021), **dengan judul “ Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit ”**, penyakit kanker kulit merupakan penyakit yang tumbuh di jaringan kulit dimana kondisi ini ditandai dengan perubahan pada kulit; dalam penelitian ini dilakukan komparasi dengan metode dari data mining untuk klasifikasi yaitu naïve bayes dan K-Nearest Neighbor untuk memprediksi tingkat keberhasilan immunotherapy dalam menyembuhkan penyakit kanker kulit; dalam proses pengujiannya, peneliti menggunakan aplikasi weka untuk mengolah data dan melakukan pengujian.

Tabel 2.7 Tinjauan Pustaka

No	Nama Peneliti	Judul Penelitian	Jurnal	Kontribusi
1	Andhini Asri Awaliyah Arifin , Wiwin Handoko, Zulfan Efendi	Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan	J-Com (Journal of Computer) Vol. 2 No. 1, Maret 2022, hlm. 21 - 26  <a href="https://jurnal.stmikroyal.ac.id/index.php/i-com/article/view/1577/812">https://jurnal.stmikroyal.ac.id/index.php/i-com/article/view/1577/812</a>	Kontribusi yang didapat adalah nilai precision untuk kelas positif 100%, untuk kelas negatif 77%, nilai recall untuk kelas positif 80%, untuk kelas negatif 100%, nilai f1-score untuk kelas positif 89%, untuk kelas negatif 87%, dan nilai akurasi 88%.
2	Hakam Febtadianran o Putro, Retno Tri Vulandari, Wawan LaksitoYuly Saptomo	Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan	Jurnal TIKomSiN, Vol. 8, No. 2, Oktober 2020  <a href="https://p3m.sinus.ac.id/jurnal/index.php/TIKomSiN/article/view/500/415">https://p3m.sinus.ac.id/jurnal/index.php/TIKomSiN/article/view/500/415</a>	Kontribusi yang didapat adalah hasil dalam menggunakan metode algoritma naïve bayes dengan Confusion Matrix menunjukkan nilai akurasi mencapai 92%, nilai presisi mencapai 100%, nilai recall mencapai 91%.
3	Purwadi, Faisal Taufik, Dicky Nofriansyah, Kamil Erwansyah, Luthfi Zulfahmi,	Penerapan Metode Naive Bayes Untuk Mendiagnosa Penyakit ISPA Pada Anak	Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Vol. 6 No 2 Juli 2023  <a href="https://ojs.trigunadhharma.ac.id/index.php/jsk/article/view/7948/2154">https://ojs.trigunadhharma.ac.id/index.php/jsk/article/view/7948/2154</a>	Kontribusi yang didapat adalah sistem pakar menggunakan metode naïve bayes untuk mendiagnosa Penyakit ISPA pada anak

No	Nama Peneliti	Judul Penelitian	Jurnal	Kontribusi
4	Endah Widya Ningsih, Hardiyani	Penerapan Algoritma Naïve Bayes Dalam Penentuan Kelayakan Penerima Kartu Jakarta Pintar Plus	Jurnal Teknik Komputer AMIK BSI Vol. 6 No.1 Januari 2020 <a href="https://ejournal.bsi.ac.id/ejurnal/index.php/itk/article/view/6680/pdf">https://ejournal.bsi.ac.id/ejurnal/index.php/itk/article/view/6680/pdf</a>	Kontribusi yang didapat adalah Metode Naïve Bayes dapat mempermudah dalam menentukan siapa saja siswa yang layak mendapatkan Kartu Jakarta Pintar Plus.
5	Adam Cahya Armadananto, Ahmad Bagus Setiawan, Danang	Klasifikasi Keluarga Penerima Bantuan Sosial Menggunakan Metode Naïve Bayes	INOTEK, Vol.7 Agustus 2023 <a href="https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/3447/2254">https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/3447/2254</a>	Kontribusi yang didapat adalah metode Naïve Bayes dengan menggunakan sistem pengkalsifikasian ini dapat membantu perangkat desa dalam melakukan klasifikasi Keluarga Penerima Manfaat Bantuan Sosial yang sesuai dengan keadaan lapangan.
6	Amat Damuri, Umbar Riyanto, Hengki Rusdianto, Mohammad Aminudin	Penerapan Algoritma Naïve Bayes Untuk Pengelompokan Predikat Peserta Uji Kemahiran Berbahasa Indonesia	Jurnal Media Informatika Budidarma Volume 6, Nomor 2, April 2022, Page 1183-1190 <a href="https://stmik-budidarma.ac.id/ejurnal/index.php/mib/article/view/3956/2674">https://stmik-budidarma.ac.id/ejurnal/index.php/mib/article/view/3956/2674</a>	Kontribusi yang didapat adalah hasil akurasi yang baik maka Algoritma Naive Bayes dikombinasikan dengan feature selection Adaboost dengan skema pengujian 70:30 dan 80:20. Hasil penelitian yang dilakukan menghasilkan nilai akurasi tertinggi yaitu 89% .

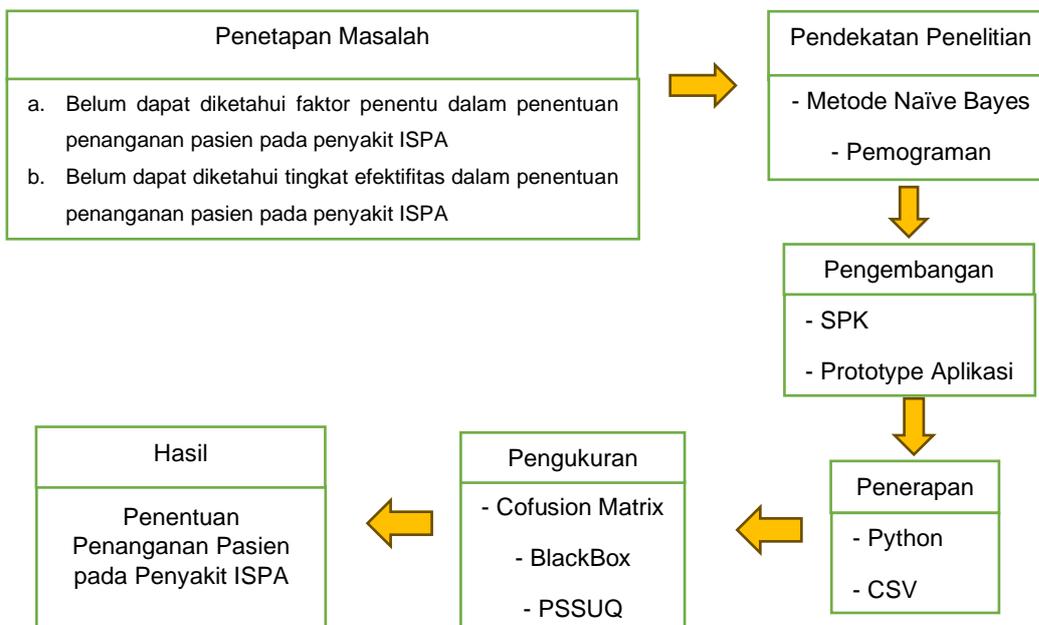
No	Nama Peneliti	Judul Penelitian	Jurnal	Kontribusi
7	Justrina Surbakti	Implementasi Metode Naïve Bayes Untuk Diagnosa Penyakit Hati	JIKOMSI. Vol.5 No.1 Maret 2022, pp 34-40  <a href="https://ejournal.sisfokomtek.org/index.php/jikom/article/view/243/229">https://ejournal.sisfokomtek.org/index.php/jikom/article/view/243/229</a>	Kontribusi yang didapat adalah membantu masyarakat umum guna untuk mendiagnosa penyakit gangguan hati secara dini, agar masyarakat umum untuk lebih cepat mengetahui jenis penyakit yang ada pada penyakit gangguan hati.
8	Amat Damuri, Umbar Riyanto, Hengki Rusdianto, Mohammad Aminudin	Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako	JURIKOM (Jurnal Riset Komputer), Vol. 8 No. 6, Desember 2021  <a href="https://www.ejournal.stmik-budidarma.ac.id/index.php/jurikom/article/view/3655/2403">https://www.ejournal.stmik-budidarma.ac.id/index.php/jurikom/article/view/3655/2403</a>	Kontribusi yang didapat adalah menggunakan algoritma Naïve Bayes, pada confusion matrix didapatkan akurasi dari 135 data training dengan 40 data testing dan tujuh atribut menghasilkan akurasi sebesar 86%, recall 85%, dan presisi 88%.
9	Dian Ade Setiawan, Risa Helilintar, Lilia Sinta Wahyuniar	Penerapan Metode Naive Bayes Untuk Klasifikasi Penentuan Penerima Bantuan PKH	Seminar Nasional Inovasi Teknologi UN PGRI Kediri, 24 Juli 2021  <a href="https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1137/738">https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1137/738</a>	Kontribusi yang didapat adalah memudahkan pemerintah dalam mengklasifikasikan penerima bantuan PKH serta untuk memudahkan masyarakat dalam pengecekan penerima bantuan PKH dengan menggunakan metode naive bayes.
10	F. Lia Dwi Cahyanti, Windu Gata, Fajar Sarasati	Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat	Jurnal Ilmiah Universitas Batanghari Jambi, 21(1), Februari 2021	Kontribusi yang didapat adalah untuk memprediksi tingkat keberhasilan immunotherapy dalam menyembuhkan penyakit kanker kulit dengan penerapan klasifikasi dari

No	Nama Peneliti	Judul Penelitian	Jurnal	Kontribusi
		Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit	<a href="http://ji.unbari.ac.id/index.php/ilmiah/article/view/1189/826">http://ji.unbari.ac.id/index.php/ilmiah/article/view/1189/826</a>	metode naïve bayes dan K-Nearest Neighbor

Berdasarkan penelitian-penelitian sebelumnya, maka penelitian ini dilakukan dengan permasalahan yang serupa serta metode penyelesaian yang sama yaitu Algoritma Naïve bayes. Namun terdapat perbedaan pada penelitian kali ini yang terdapat pada letak implementasi dimana pada penelitian ini menerapkan algoritma Naïve Bayes ke dalam prototype kontribusi yang dapat membantu Dinas Kesehatan dalam mengambil keputusan mengenai penentuan penanganan pasien pada penyakit ISPA. Dan variabel yang digunakan usia pasien, jenis kelamin, anamnesa, diagnosis, kode ICD X, kasus baru/lama, keterangan (status), unit pelayanan, jenis kunjungan.

### C. Kerangka Pemikiran

Berikut adalah kerangka pemikiran untuk memecahkan masalah penelitian ini yang digambarkan pada gambar berikut.



**Gambar 2. 5 Kerangka Pemikiran**

Pada gambar kerangka pemikiran dijelaskan sebagai berikut :

- (1) Penetapan masalah, yaitu terkait belum dapat diketahui faktor penentu dalam penentuan penanganan pasien pada penyakit ISPA dan belum dapat diketahui tingkat efektifitas dalam penentuan penanganan pasien pada penyakit ISPA;
- (2) Pendekatan, penelitian menggunakan Algoritma Naïve Bayes dan pemograman;
- (3) Pengembangan, menggunakan SPK (Sistem Pendukung Keputusan) dan prototype aplikasi;
- (4) Penerapan, menggunakan bahasa pemograman Python dan menggunakan CSV.
- (5) Pengukuran, menggunakan confusion matrix, BlackBox, dan PSSUQ.
- (6) Hasil, prototype yang menampilkan penentuan penanganan pasien pada penyakit ISPA dengan tepat dan efektif.

#### **D. Hipotesis Penelitian**

Metode Naive Bayes merupakan metode klasifikasi yang menggunakan probabilitas untuk memprediksi kelas suatu objek. Menurut penelitian sebelumnya, metode naïve bayes mampu mengklasifikasi data dengan mudah dan sederhana dengan menghitung probabilitas suatu data poin dalam kelas tertentu dengan akurat. Sebagai metode klasifikasi yang efisiensi dan sederhana, maka hal ini berkaitan dengan permasalahan pada penelitian ini yaitu penentuan penanganan pasien pada penyakit ISPA sehingga pasien dapat mendapatkan penanganan yang tepat dan cepat. Berdasarkan pemahaman diatas dan penelitian yang telah dilakukan oleh Setiawan, Halilintar and Wahyuniar, tahun 2021, tentang Penerapan Metode Naïve Bayes untuk Klasifikasi Penentuan Penerima Bantuan PKH, dengan metode Naïve Bayes membantu pemerintah dalam mengklasifikasi penerima bantuan PKH dan memudahkan pengecekan penerima bantuan PKH. Maka hipotesis dari penelitian ini dapat ditetapkan bahwa metode naive bayes memiliki potensi dalam menentukan penanganan pasien pada penyakit ISPA secara akurat dan efektif.