

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori Dasar**

Teori-teori yang menjadi dasar penelitian ini bersumber dari jurnal-jurnal berisikan informasi yang relevan dan sebagai kutipan untuk pendukung dalam upaya penelitian, dan berikut beberapa teori yang dipakai pada penelitian ini:

#### **2.2 KDD (*Knowledge Discovery In Database*)**

Data Mining pula di kenal dengan *Knowledge Discovery In Database* (KDD) yang merujuk pada pemanfaatan data masa lalu yang bertujuan mendeteksi pola, keteraturan, atau korelasi dari koleksi data yang luas, hasil analisis data mining memiliki potensi untuk meningkatkan proses pengambilan keputusan di masa depan (Siahaan & Fauzi, 2023)

Karena volume informasi yang semakin besar data mining sering dikenal dengan alat penting pada manajemen informasi, *Knowledge Discovery in Database Process* (KDD) proses metode data mining untuk menemukan informasi berharga, pola yang ada di dalam data, yang melibatkan algoritma untuk menemukan pola di dalam data (Simanjuntak et al., 2022). *Knowledge Discovery in Database Process* (KDD) sebagai proses penggunaan Data Mining dalam mencari berbagai informasi – informasi yang berharga, tujuan pada proses KDD untuk menemukan potensi data yang di dapat dari database yang dipelajari melalui pola yang dianalisis dan juga divisualisasikan agar mudah dipahami oleh pengguna (Borman & Wati, 2020).

### 2.3 Data Mining

Data Mining merujuk pada kumpulan metode atau teknik yang digunakan untuk mengidentifikasi pola yang tidak diketahui pada data yang telah di kumpulkan. Dengan adanya Data Mining kita dapat menemukan pengetahuan yang tersembunyi di database yang mungkin belum diketahui sebelumnya (Felicia Watratan et al., 2020). Data Mining membantu dalam fase perencanaan dengan menyediakan informasi yang tepat dan relevan dalam memprediksi berdasarkan pola dan pengalaman dari masa sebelumnya untuk kondisi pada saat ini (Entini A & Handoko, 2023).

Data Mining sebuah kegiatan untuk mengidentifikasi beragam jenis data, Data Mining proses yang dinamis dan interaktif yang digunakan untuk menghasilkan pola baru yang menarik dan sangat bermanfaat (Amalia, 2020). Model proses pada Data Mining biasanya sempurna sehingga bisa digeneralisasikan untuk kepentingan masa depan, Data Mining memanfaatkan kecerdasan buatan, statistik, matematika, dan *machine learning* dalam mengekstraksi dan mengenali informasi penting yang terkait dari berbagai sumber data besar (Sudarto & Handoko, 2023).

Data Mining proses megurutkan sejumlah besar data untuk menemukan pola dan hubungan yang bisa membantu untuk pemecah masalah bisnis dengan analisis data, teknik dan alat yang digunakan untuk penambangan data sangat membantu dalam memprediksi tren masa depan dan membuat keputusan yang akurat (Zulfikar et al., 2023).

Menurut (Idris, 2019) pengelompokan dalam Data Mining dapat dilakukan sebagai berikut:

1. Deskripsi: Peneliti biasanya mencari cara untuk mendeskripsikan dan menggambarkan pola serta tren tersembunyi dalam data.
2. Estimasi: Estimasi mempunyai kesamaan dengan klasifikasi, namun berfokus pada variabel numerik dari pada kategori.
3. Prediksi: Prediksi mempunyai kesamaan antara klasifikasi dan estimasi yang dimana hasil prediksi ini menggambarkan situasi yang belum terjadi.
4. Klasifikasi: Klasifikasi yang melibatkan variabel bersifat kategori, seperti mengelompokkan pendapat ke dalam tiga kategori, yaitu pendapatan tinggi, sedang, dan rendah.
5. *Clustering*: Pengelompokan record, observasi, atau memperhatikan kasus berdasarkan kesamaan nya.
6. Asosiasi: Mengidentifikasi atribut yang berhubungan dari berbagai peristiwa muncul dalam satu waktu.

#### **2.4 Metode *Naïve Bayes***

*Naïve Bayes* algoritma *Machine Learning* yang sederhana tetapi efektif, dengan mengklasifikasikan data berdasarkan asumsi dari setiap fitur data independent satu sama lain dalam pengaruh kelasnya, *Naïve Bayes* menghitung probabilitas data yang termasuk dalam kelas tertentu berdasarkan probabilitas fiturnya (Yulita et al., 2021).

*Naïve bayes* metode klasifikasi yang bergantung kepada prinsip probabilitas dan statistic yang pertama kali di ajukan oleh ilmuwan Inggris Thomas Bayes,

algoritma ini memanfaatkan Teorema Bayes dalam memproses informasi dari peristiwa sebelumnya lalu diperkirakan ke peristiwa masa yang akan datang (Leonila & Ineke Parkereng, 2023). Rumus yang di pakai dalam *Naïve Bayes* sebagai berikut:

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

**Rumus 2. 1** *Naïve Bayes*

Keterangan:

X : Data kelasnya yang belum diketahui

H : Hipotesis data kelas

P(H) : Peluang hipotesis H

P(X) : Peluang sampel data (X)

P(X|H): Peluang data sampel X pada kondisi hipotesis H

P(H|X): Peluang hipotesis H pada kondisi X

*Naïve Bayes* lebih menekankan proses pembelajaran daripada estimasi probabilitas, strukturnya yang sederhana membuatnya sangat efektif dan mudah dipahami dan pengerjaannya yang cepat, *Naïve Bayes* mengasumsikan nilai yang diinput bergantung pada nilai probabilitas pada nilai atribut lainnya yang dikenal dengan peluang bersyarat (Riyanah & Fatmawati, 2021).

## 2.5 Objek Penelitian

Objek penelitian di ambil dari sejarah Panti Asuhan Al-Ikhlas yang didirikan pada tahun 2007 – sekarang beralamat di Kavling Pelopor blok d65, Sagulung, Sei Lekop yang sudah terdaftar di Dinas Sosial Kota Batam berawal dari 17 orang anak sampai 40 orang anak yang awal mulanya Panti Asuhan Al-Ikhlas berdiri dari rumah kontrakan sebagai tempat tinggal sehingga dihibahkan tanah untuk mendirikan tempat tinggal yang lebih layak seperti sekarang.

Setiap harinya di Panti Asuhan Al-Ikhlas dipenuhi dengan beragam kegiatan, seperti anak-anak yang belajar di sekolah umum, menambah keimanan dengan beribadah dan mengaji Al-Quran, dan juga berkontribusi dalam kegiatan bersih – bersih dalam menjaga kebersihan lingkungan.

## 2.6 *Software RapidMiner*

RapidMiner yang sebelumnya dikenal sebagai *YALE (Yet Another Learning Environment)* alat yang sering digunakan dalam eksplorasi data karena dapat berjalan di berbagai sistem operasi, *Software RapidMiner* program perangkat lunak yang dikembangkan oleh *Ralf Klittenberg* tahun 2001 dan dapat diakses secara publik (Juanda Saputra & Izman Herdiansyah, 2022). *Kdnuggets* menyatakan bahwa RapidMiner salah satu program data mining terbaik dengan antarmuka grafis (*GUI*) *RapidMiner* pengguna dapat mendesain alur kerja secara menyeluruh dengan *GUI* menghasilkan *file XML* yang menunjukkan langkah – langkah analisis data yang akan dilakukan pengguna, dan *RapidMiner* dapat secara langsung membaca dan mengeksekusi *file XML* tersebut (Adrian & Suarna, 2023).

## 2.7 Penelitian Terdahulu

Penelitian ini dilakukan dengan mengacu pada penelitian terdahulu yang telah dilakukan, dengan tujuan untuk menemukan kesamaan dan gagasan baru dalam bidang yang sama, penelitian sebelumnya berfungsi sebagai referensi tambahan yang berharga bagi peneliti. Berikut beberapa penelitian terdahulu yang berkaitan dengan subjek yang akan diteliti oleh peneliti:

No	Judul Penelitian	Peneliti & Tahun Terbit	Hasil Penelitian
1	Implementasi algoritma <i>Naïve Bayes</i> untuk memprediksi pengaruh media sosial terhadap semangat belajar mahasiswa	(Satria & Hermanto, 2022, Vol 10, Jurnal Informasi dan Komputer)	Berdasarkan hasil pengujian menunjukkan bahwa algoritma <i>Naïve Bayes</i> memiliki tingkat akurasi sebesar 72%, ketepatan kelas 77,8%, dan hasil nilai AUC sebesar 0,768. Hasil ini menunjukkan bahwa model yang telah dibuat berhasil sehingga algoritma <i>Naïve Bayes</i> dapat membuat keputusan tentang bagaimana penggunaan media sosial memengaruhi keinginan mahasiswa untuk belajar lebih meningkat.

2	Implementasi Data Mining untuk memprediksi kelulusan siswa dengan metode <i>Naïve Bayes</i>	(Sinaga & Handoko, 2021, Vol.04, Jurnal <i>Comasie</i> )	<p>Dengan adanya penelitian ini Data Mining dilibatkan dengan 210 data siswa yang dianalisis dengan metode <i>Naïve Bayes</i> lalu diperoleh hasil jumlah lulus siswa dan yang tidak lulus dari jurusan farmasi pada SMK Putra Jaya School. Data Mining mempermudah dan mempercepat proses pengolahan data kelulusan siswa serta memiliki peran yang sangat penting dalam mengumpulkan berbagai informasi mengenai nilai kelulusan siswa.</p>

3	<p><i>Sentiment analysis on social media against public policy using multinomial Naïve Bayes</i></p>	<p>(Zulfikar et al., 2023, <i>Journal of informatics</i>)</p>	<p><i>The research focused on the patterns generated by the Naive Bayes algorithm in the classification of social media users' tweets. The results showed that this algorithm has a high accuracy rate, reaching 90.25%, and revealed how it is very effective in identifying feelings in text, distinguishing between positive and negative opinions. This suggests that the Naive Bayes algorithm is the right choice to the research objective, namely to classify data based on public interest from samples or training data.</i></p>
4	<p><i>Crime Analysis for Multistate Network using Naive Bayes Classifier</i></p>	<p>(Jangra et al., 2019, <i>Internasional Journal of</i></p>	<p><i>This analysis's primary goal is to assist police officers and detectives in recognizing and comprehending specific and</i></p>



		<p><i>Computer Science and Mobile Computing)</i></p>	<p><i>pressing criminal concerns. One of the main goals of this study is to identify the patterns that now exist in relation to the behaviors of offenders in order to predict all potential future crime events. An innovative method of crime prediction is used in this study. This method is predicated on the naïve bayes and KNN techniques. Prediction outcomes are analyzed using execution time and accuracy. The Naïve Bayes classifier is used in this study to predict offending.</i></p>
--	--	--	---

5	<p>Penerapan Data Mining dalam klasifikasi data anggota kopdit sejahtera Bandar Lampung dengan Algoritma <i>Naïve Bayes</i></p>	<p>(Borman &amp; Wati, 2020)</p>	<p>Penelitian ini memberikan kesimpulan bahwa Algoritma <i>Naïve Bayes</i> secara efektif digunakan untuk mengklasifikasikan data tentang kelalaian pinjaman anggota Koperasi Kredit Sejahtera. Klasifikasi menggunakan <i>WEKA</i> dan <i>RapidMiner</i> menunjukkan tingkat akurasi sebesar 70,33% dengan <i>recall</i> sebesar 70,33% dan presisi 100%. Hasil ini menunjukkan semakin banyak data pengujian yang digunakan semakin besar juga nilai akurasi yang dihasilkan.</p>
---	---	----------------------------------	---

6	Implementasi Data Mining dengan Algoritma Naïve Bayes untuk klasifikasi kelayakan penerima bantuan	(Entini A & Handoko, 2023, Jurnal <i>Comasie</i> )	Naive Bayes adalah algoritma yang digunakan dalam data mining, termasuk dalam teknik klasifikasi yang memanfaatkan probabilitas dan statistik untuk memperkirakan atau memprediksi kemungkinan kejadian berdasarkan data sebelumnya. Algoritma ini bekerja dengan dua kelas: layak dan tidak layak. Hasil klasifikasi yang dihasilkan akan mendukung proses pengambilan keputusan terkait penentuan penerima bantuan sembako. Berdasarkan pengujian manual dan penggunaan perangkat lunak RapidMiner, algoritma ini menunjukkan akurasi sebesar 80%.
7	Implementasi Algoritma <i>Naïve</i>	(Felicia Watratan et	Hasil penelitian ini menunjukkan bahwa <i>Naïve</i>

	<p><i>Bayes</i> untuk memprediksi tingkat penyebaran <i>Covid-19</i> di Indonesia</p>	<p>al., 2020, <i>JACOST</i>)</p>	<p><i>Bayes</i> berhasil memprediksi jumlah kasus <i>Covid-19</i> per provinsi dengan tingkat keakuratan sebesar 48,4848% dengan menggunakan data kasus <i>Covid-19</i> per provinsi sebagai data pelatihan, 16 dari 33 data yang diuji berhasil diklasifikasikan dengan metode ini.</p>
8	<p>Penerapan Data Mining untuk Memprediksi Hasil Kelulusan Siswa Menggunakan Metode <i>Naïve Bayes</i></p>	<p>(Amalia, 2020, JUISI)</p>	<p>Validasi dan evaluasi menunjukkan <i>Naïve Bayes</i> memiliki Tingkat akurasi yang signifikan dengan mencapai 82,00%. Dan data dari kolektif hasil ujian sekolah berstandar nasional untuk tahun ajaran 2018/2019 di Kalimantan Tengah, khususnya di Kotawaringin Timus, memberikan kontribusi yang</p>

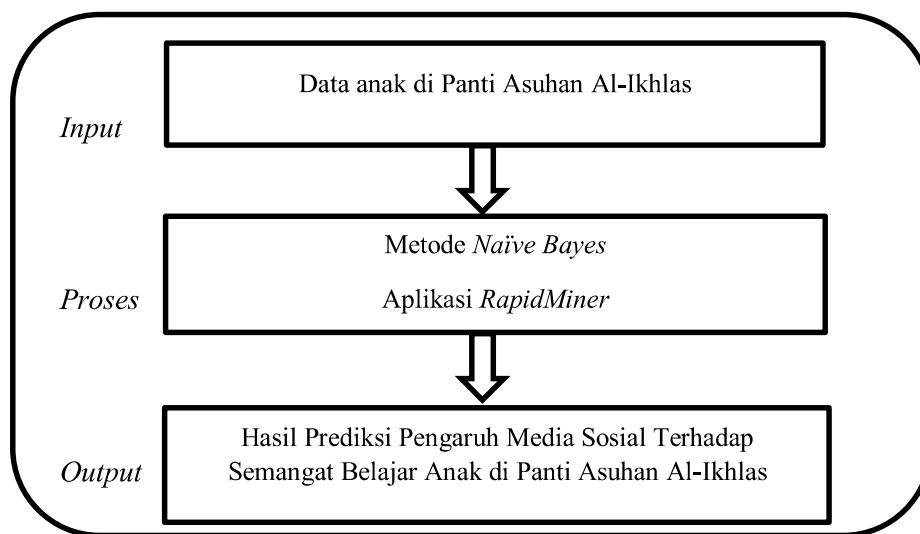
			<p>sangat relevan dalam penelitian.</p> <p>Oleh karena itu, metode <i>Naïve Bayes</i> dapat dianggap sebagai pendekatan yang efektif dan efisien dalam menentukan kelulusan siswa.</p>
9	<p>Implementasi Data Mining dengan Algoritma <i>Naïve Bayes</i> untuk memprediksi angka kelahiran</p>	<p>(Idris, 2019, Jurnal Pelita Informatika)</p>	<p>Hasil Analisa dan perancangan pada penelitian ini telah menarik kesimpulan bahwa Data Mining dengan Algoritma <i>Naïve Bayes</i> dapat diterapkan dengan efektif di Kantor Kepala Desa Lalang. Penerapannya memberikan kemudahan kepada pihak administrasi dengan melakukan evaluasi dan juga penentuan angka kelahiran, dengan pemilihan data training yang tepat karena dapat memengaruhi hasil pengujian. Oleh karena itu, tingkat akurasi yang tinggi atau</p>

			rendah dapat dipengaruhi oleh pemilihan data training yang sesuai dan tepat.
10	Klasifikasi Anak Berpotensi Putus Sekolah dengan Metode Naïve Bayes di Kabupaten Manokwari	(Leonila & Ineke Pakereng, 2023, J-SAKTI )	Penelitian ini menghasilkan Tingkat akurasi prediksi sebesar 70%, dengan Tingkat presisi kelas 88,89%. Metode ini memiliki keterbatasan dalam memprediksi anak-anak yang mungkin atau sangat mungkin putus sekolah dengan Tingkat presisi dan <i>recall</i> kelas hanya 0,00%.

## 2.8 Kerangka Pemikiran

Kerangka pemikiran secara garis besar mengenai alur logika struktur yang membantu mengorganisir dan mengarahkan penelitian atau pemikiran dalam suatu karya ilmiah atau proyek.

Data yang digunakan diambil dari data anak di Panti Asuhan Al-Ikhlas sebagai *input* di dalam penelitian ini, lalu di proses dengan Metode *Naïve Bayes* dan diuji menggunakan aplikasi *RapidMiner* sehingga dapat menghasilkan *output* hasil prediksi pengaruh media sosial terhadap semangat belajar anak di Panti Asuhan Al-Ikhlas. Berikut gambaran kerangka pemikiran dalam penelitian ini:



**Gambar 2. 1** Kerangka Pemikiran