

BAB II

KAJIAN PUSTAKA

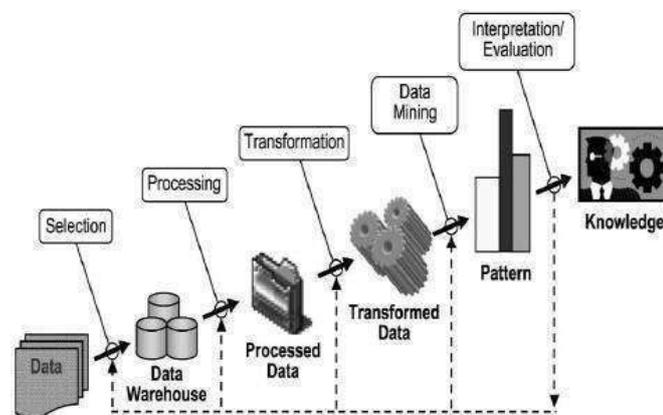
2.1 Teori Dasar

Landasan teori dalam penelitian ini akan membahas berbagai teori yang berkaitan dengan topic penelitian yang diajukan. Pembahasan teori-teori ini sangat penting sebagai alat untuk meningkatkan keakuratan metode penelitian yang dilakukan. Pembahasan akan mencakup definisi dan istilah-istilah yang akan dikaji dalam penelitian ini antara lain : *Knowledge Discovery in Database (KDD)*, data mining, algoritma *Naïve Bayes*, aplikasi *GetContact*, dan RapidMiner.

2.2 *Knowledge Discovery in Database*

Penggalian data dalam basis data, atau *Knowledge Discovery in Database (KDD)*, dikenal sebagai teknik yang kompleks yang digunakan untuk mengekstraksi informasi implisit yang sebelumnya tidak diketahui dan berpotensi digunakan sebagai data. Proses ini didasarkan pada prinsip bahwa informasi dapat ditemukan dari berbagai sumber yang terintegrasi dalam sebuah basis data. Data yang diperoleh dari basis data akan diproses terlebih dahulu menjadi format standar yang dapat diimplementasikan oleh berbagai algoritma data mining, yang akan menghasilkan pola sehingga diperoleh pengetahuan yang bermanfaat (Alghifari & Juardi, 2021)

Istilah data mining dan *knowledge discovery in databases* (KDD) sering dianggap sama dan digunakan untuk menjelaskan proses menemukan informasi tersembunyi dalam basis data yang besar (Liatas et al., 2024). Namun, kedua istilah ini sebenarnya memiliki konsep yang berbeda tetapi saling terkait. Data mining merupakan salah satu tahap dalam proses KDD (Alghifari & Juardi, 2021). Secara umum, tahapan dalam KDD adalah sebagai berikut:



Gambar 2. 1 Tahapan Proses *Knowledge Discovery in Database*
Sumber: (Walhidayat et al., 2021)

1. *Data Selection*

Sebelum memulai tahapan penggalian informasi dari KDD, perlu dilakukan pemilihan data dari kumpulan data operasional. Data yang dipilih untuk proses penambangan disimpan dalam file terpisah yang berbeda dari basis data operasional.

2. *Data Cleaning*

Sebelum memulai proses penggalian data di KDD, penting untuk melakukan prapemrosesan data. Prapemrosesan ini melibatkan tugas-tugas seperti menghilangkan entri duplikat, memverifikasi konsistensi data, dan memperbaiki kesalahan seperti kesalahan ketik. Selanjutnya proses

pengayaan dilakukan untuk melengkapi data dengan informasi tambahan yang relevan yang diperlukan untuk KDD, yang mungkin termasuk sumber data eksternal.

3. *Data Transformation*

Data Transformation ini serangkaian proses mengubah data yang telah dipilih sehingga sesuai untuk proses data mining. Proses *coding* dalam KDD adalah proses kreatif yang sangat tergantung pada jenis atau pola informasi yang dicari dalam basis data.

4. *Data Mining*

Proses data mining melibatkan pencarian pola atau informasi menarik di dalam kumpulan data tertentu menggunakan berbagai teknik atau metode. Ragam teknik, metode, atau algoritma yang digunakan dalam data mining sangat beragam, dan pemilihan yang sesuai tergantung pada tujuan dan proses keseluruhan dari proses penambangan data.

5. *Interpretation / Evaluation*

Pola informasi yang muncul dari proses data mining harus disajikan dengan cara yang dapat dipahami dengan mudah oleh pihak yang berkepentingan. Langkah ini merupakan bagian dari proses KDD yang disebut interpretasi. Pada tahap ini, dilakukan evaluasi terhadap apakah pola atau informasi yang ditemukan sesuai dengan fakta atau hipotesis yang telah ada sebelumnya.

Penemuan Pengetahuan dalam Database atau *Knowledge Discovery in Databases*.

2.3 Data Mining

Data mining merujuk pada teknik atau cara yang digunakan untuk melakukan ekstraksi informasi pada sebuah data, tujuan diimplementasikan data mining adalah untuk memahami pola atau *pattern* yang terdapat di dalam data tersebut, seperti mengetahui perilaku pelanggan sesuai dengan kebiasaan konsumsi, data perilaku ini dapat digunakan untuk melakukan prediksi permintaan pelanggan dimasa depan. Dalam penelitian ini digunakan untuk mengekstrak informasi pengguna aplikasi *GetContact* terhadap keamanan panggilan dan pesan (Kusuma & Ali, 2024).

Berdasarkan penjelasan di atas dapat disimpulkan bahwa data mining lebih dari sekedar mengekstraksi data. Data mining melibatkan metodologi yang sangat sistematis dan terstruktur. Dalam proses data mining, data historis dianalisis dengan cermat, dan transformasi data dilakukan jika diperlukan. Berbagai teknik kemudian diterapkan untuk menemukan pola, keteraturan, atau hubungan yang mungkin tersembunyi di dalam kumpulan data.

Menurut Mardi (2019) dijelaskan bahwa data mining dikelompokan berdasarkan tugas yang dapat dilakukan seperti deskripsi, estimasi, prediksi, klasifikasi, pengklusteran dan asosiasi.

2.4 Algoritma *Naïve Bayes*

Naïve bayes dikenal sebagai metode klasifikasi probabilistic yang mudah. Metode ini menghitung probabilitas dengan meringkas frekuensi dan kombinasi nilai dalam kumpulan data yang tersedia. Metode ini bergantung pada asumsi sederhana bahwa nilai atribut secara kondisional tidak bergantung satu sama lain

ketika diberikan nilai output. Dengan kata lain, probabilitas gabungan dari pengamatan nilai *output* ditentukan dengan mengalikan probabilitas dari masing-masing atribut. Metode klasifikasi *Naïve Bayes* terkenal dengan formula yang sederhana dan mudah digunakan, namun tetap memiliki akurasi yang tinggi dibandingkan dengan metode lainnya. *Naïve Bayes* termasuk dalam kategori algoritma klasifikasi yang menggunakan pendekatan probabilitas dan statistic. Dinamai dari nama ilmuwan *Inggris* Thomas Bayes, metode ini bertujuan untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa lalu dan sering disebut sebagai Teorema Bayes. Teorema ini ditetapkan secara “naïf”, dengan asumsi bahwa atribut-atributnya tidak bergantung satu sama lain. Dalam konteks klasifikasi *Naïve Bayes*, diasumsikan bahwa ada atau tidaknya sebuah fitur di satu kelas tidak berhubungan dengan fitur-fitur di kelas lainnya. (Kusuma & Ali, 2024).

Menggunakan *Naïve Bayes*, dalam pengklasifikasi data dapat dilakukan dengan akurat dengan mempertimbangkan pola dan hubungan antarvariabel yang teridentifikasi dari data sebelumnya. Algoritma ini menjadi kunci penting dalam analisis prediktif, pengambilan keputusan, dan aplikasi di berbagai bidang, termasuk analisis risiko, pengelolaan keuangan, serta pengenalan pola dalam teks dan gambar (Hant & Hendry, 2022).

Berikut ini adalah rumus persamaan dari teorema *Bayes*:

$$P(A|B) = \frac{(P(B|A)) + (P(A))}{(P(B))}$$

Rumus 2. 1 *Naïve Bayes*.

Keterangan:

- B : Data dengan kelas yang belum diketahui.
- A : Hipotesa data B merupakan suatu kelas spesifik.
- $P(A)$: Probabilitas hipotesa A (*prior probability* / probabilitas awal).
- $P(B)$: Probabilitas B.
- $P(B|A)$: Probabilitas hipotesa B berdasarkan kondisi A.
- $P(A|B)$: Probabilitas hipotesa A berdasarkan kondisi B (*Posterior probability* / probabilitas akhir)

Dalam metode ini, setiap atribut memberikan kontribusinya sendiri dalam proses pengambilan keputusan, dengan bobot atribut yang sama pentingnya dan setiap atribut dianggap tidak terkait satu sama lain. Jika terdapat k atribut yang saling bebas (independen), nilai probabilitas dapat dihitung sebagai berikut. Klasifikasi dengan menggunakan *classifier Naïve Bayes* yang mengasumsikan nilai atribut yang terpisah atau independen dapat dijelaskan melalui rumus berikut (Gustientiedina et al., 2020):

$$P(X_1, \dots, x_k | C) = P(X_1 | C) \times P(X_2 | C) \times \dots \times P(X_k | C)$$

Rumus 2. 2 Naive Bayes untuk Probabilitas Gabungan Atribut.

Keterangan:

- $P(X_1, \dots, x_k | C)$: Probabilitas kelas C menghasilkan *intence* X.
- $P(X_1 | C)$: Probabilitas kelas C menghasilkan nilai yang diamati untuk atribut ke-1 (X_1).
- $P(X_2 | C)$: Probabilitas kelas C menghasilkan nilai yang diamati untuk atribut ke-2 (X_2).

Untuk menghitung nilai akhir kelas menggunakan rumus:

$$C_m = \operatorname{argmax} P(X|C)$$

Rumus 2. 3 Menentukan Kelas Maksimum.

Keterangan:

C_{map} : Hipotesa nilai tertinggi
 $\operatorname{argmax}_{c \in C}$: Nilai rata-rata dari setiap kelas.

Adapun alur proses algoritma *Naïve Bayes* adalah sebagai berikut (Watratan *et al.*, 2020):

1. Baca data *training*
2. Hitung jumlah dan probabilitas, namun apabila data numerik maka:
 1. Cari nilai *mean* dan standar deviasi dari masing-masing parameter yang merupakan data numerik.
 2. Cari nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
 3. Mendapatkan nilai dalam tabel mean, standart deviasi dan probabilitas.
4. Dapatkan solusi.

Menurut (Watratan, 2020) penggunaan algoritma *Naïve Bayes* memiliki beberapa kelebihan seoerti dapat digunakan untuk jenis data kuantitatif maupun data kualitatif, *Naïve Bayes* tidak memerlukan jumlah data yang besar, tidak perlu melakukan proses *training* yang terlalu banyak, kemampuan dalam menangani nilai yang hilang, perhitungan yang cepat dan efisien, mudah dimengerti dan dibuat, bisa digunakan untuk klasifikasi masalah biner atau *multiclass*.

Akan tetapi, penggunaan algoritma *Naïve Bayes* bukan tanpa kekurangan. Adapun kekurangan dari algortima ini adalah apabila probabilitas kondisionalnya nol, maka probabilitas prediksi juga akan bernilai nol, asumsi bahwa masing-masing variabel independent membuat berkurangnya akurasi, karena biasanya ada korelasi antara varibel yang satu dengan yang lain, keakuratan tidak bisa diukur menggunakan satu probabilitas saja. Butuh bukti-bukti lain untuk membuktikannya, serta untuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya. Keberhasilan sangat bergantung pada pengetahuan awal tersebut.

2.5 *Software Pendukung*

Penelitian ini dilakukan dengan menggunakan bantuan dari perangkat lunak pendukung yaitu RapidMiner, salah satu lunak *opensource* yang menyediakan solusi untuk analisis data mining, *text mining*, dan analisis prediktif. Dengan berbagai teknik deskriptif dan prediktif, RapidMiner memberikan wawasan kepada pengguna untuk membantu mereka membuat keputusan yang optimal. Perangkat ini menawarkan lebih dari 500 operator data mining, termasuk operator untuk input, output, preprocessing data, dan visualisasi. RapidMiner dapat berfungsi secara mandiri untuk analisis data maupun diintegrasikan ke dalam produk lain. Ditulis dalam bahasa Java, perangkat lunak ini dapat dijalankan di berbagai sistem operasi (Musfiroh et al., 2021). Versi RapidMiner yang digunakan dalam penelitian ini adalah versi 10.0.0.

RapidMiner menyediakan antarmuka grafis (*Graphic User Interface*) yang memudahkan pengguna dalam merancang alur kerja analisis. Antarmuka ini menghasilkan berkas XML (*Extensible Markup Language*) yang memetakan proses analisis yang diinginkan pengguna untuk diterapkan pada data. Berkas XML ini kemudian dapat dibaca oleh RapidMiner untuk menjalankan analisis secara otomatis.

2.6. **Objek Penelitian**

Objek penelitian yang digunakan dalam penelitian ini adalah aplikasi *GetContact*, Aplikasi ini digunakan sebagai “Penyekat *Spam*” dan “Maklumat Panggilan” untuk pengguna. *GetContact* menapis panggilan mengganggu dan hanya membenarkan orang yang pengguna pilih untuk berkomunikasi dengan

pengguna lainnya. Pengguna dapat mengenal panggilan diterima lewat nomor-nomor yang tidak didaftarkan dalam kontak. *GetContact* memberitahu pengguna jika pengguna mendapatkan yang tidak diinginkan. Hal ini bertujuan melindungi pengguna dari panggilan-panggilan robot, jurujual telepon, dan panggilan-panggilan *scam*. Aplikasi ini telah memiliki jutaan pengguna di seluruh dunia termasuk Indonesia dan lebih dari 2 juta pengguna telah memberikan ulasan terhadap penggunaan aplikasi ini di *Google Play Store* (Hanifah et al., 2022)

GetContact sangat tepat untuk dijadikan objek penelitian karena berdasarkan hasil riset, salah satu fitur dari aplikasi ini yang dapat mengakses kontak-kontak yang berada dalam perangkat pengguna merupakan salah satu bentuk pelanggaran hukum karena hak privasi konsumen (kerahasiaan data) dilanggar, meski begitu aplikasi ini juga memberikan dampak positif karena pengguna dapat dengan mudah melakukan pengecekan terhadap nomor yang melakukan panggilan, SMS, atau pesan *whatsapp*. Selain itu aplikasi ini dapat memblokir telpon spam kepada pengguna (Santiri et al., 2022).

2.7 Penelitian Terdahulu

Penelitian mengenai data mining, *knowledge discovery in database* (KDD) maupun penggunaan algoritma *Naïve Bayes* telah banyak dilakukan sebelumnya, untuk memberikan gambaran dan perbandingan mengenai penelitian ini, beberapa penelitian tersebut akan dipaparkan di bawah ini.

Tabel 2. 1 Penelitian Terdahulu

No	Judul Penelitian	Peneliti	Tahun	Hasil Penelitian
1	<i>Application of Knowledge</i>	(Llatas et al., 2024)	2024	Penelitian ini membahas tentang penerapan KDD untuk

No	Judul Penelitian	Peneliti	Tahun	Hasil Penelitian
	<i>Discovery in Databases (KDD) to environmental, economic, and social indicators used in BIM workflow to support sustainable design</i>			menganalisis indikator lingkungan, ekonomi, dan sosial dalam desain bangunan guna mendukung Pembangunan berkelanjutan. KDD digunakan untuk mengidentifikasi korelasi antara variabel-variabel tersebut dan menyajikan empat scenario untuk membantu pengambilan keputusan yang sesuai.
2	Prediksi Dampak Pembelajaran <i>Hybrid Learning</i> Menggunakan <i>Naïve Bayes</i>	(Yusnida Lase et al., 2023)	2023	Penelitian ini dilakukan dengan tujuan memprediksi dampak pembelajaran hibrida pada mahasiswa Politeknik Negeri Medan. Data yang diperoleh kemudian akan diprediksi dengan menggunakan teknik penambangan data dan algoritma <i>Naïve Bayes</i> . Hasil penelitian ini menunjukkan bahwa algoritma <i>Naïve Bayes</i> mampu melakukan prediksi terhadap dampak pembelajaran hibrida mencapai 92% dengan total data uji yang digunakan adalah 40%.
3	Klasifikasi ke-Yaan Layanan Akademik di	(Suparyanto, 2022)	2022	Penelitian ini dilakukan dengan tujuan untuk mengklasifikasi keYaan mahasiswa terhadap

No	Judul Penelitian	Peneliti	Tahun	Hasil Penelitian
	STMIK El Rahma Menggunakan Metode Algoritma <i>Naïve Bayes</i>			pelayanan akademik di STMIK El Rahma dengan menerapkan teknik penambangan data dan metode <i>Naïve Bayes</i> sebagai <i>classifier</i> . Berdasarkan beberapa indikator hasil dari penelitian ini menunjukkan bahwa algoritma <i>Naïve Bayes</i> mampu melakukan klasifikasi terhadap keYaan layanan akademik di STMIK El Rahma sebesar 84.21%.
4	<i>Data Mining Technique Using Naïve Bayes Algorithm to Predict Shopee Cosumer Satisfication Among Millenial Generation</i>	(Hant & Hendry, 2022)	2022	Penelitian ini dilakukan dengan tujuan melakukan prediksi terhadap keYaan pengguna aplikasi <i>shopee</i> terutama yang berasal dari generasi milenial. Prediksi akan dilakukan dengan menerapkan teknik <i>data mining</i> dan menggunakan algoritma <i>Naïve Bayes</i> . Hasil penelitian dengan menggunakan 14 atribut menunjukkan bahwa diperoleh beberapa atribut masukan yang menciptakan hubungan sebab akibat ketika melakukan klasifikasi konsumen yang Ya atau Tidak. Hasil klasifikasi menunjukkan bahwa algoritma <i>Naïve Bayes</i> cocok digunakan

No	Judul Penelitian	Peneliti	Tahun	Hasil Penelitian
				untuk melakukan klasifikasi berlandaskan pada akurasi yang diperoleh sebesar 89.65%.
5	Analisis Ke-Ya-an Konsumen terhadap Pelayanan Store MS Glow Menggunakan Metode <i>Naïve Bayes</i>	(Amanda et al., 2022)	2022	Penelitian ini dilakukan dengan tujuan untuk mengukur Tingkat keYaan konsumen Store Wanda MS Glow melalui kuisiner. Data yang telah dikumpulkan akan dilakukana analisis lebih lanjut menggunakan algoritma <i>Naïve Bayes</i> untuk menentukan tingkat keYaan pelanggan. Hasil penelitian ini menunjukkan akurasi <i>Naïve Bayes</i> dalam melakukan klasifikasi mencapai 88% sehingga dapat digunakan untuk mengklasifikasi data ulasan keYaan pengguna yang belum dilatih
6	Implementasi Algoritma <i>Naïve Bayes</i> untuk Memprediksi Tingkat Penyebaran Covid-19 di Indonesia	Alfina Felicia Watratan, Arwini Puspita. B, Dikwan Moeis	2020	Penelitian ini dilakukan dengan tujuan untuk melakukan prediksi terhadap tingkat penyebaran Covid-19 di Indonesia. Algoritma <i>Naïve Bayes</i> dipilih untuk melakukan prediksi penyebaran Covid-19 di Indonesia. Hasil dari penelitian ini menunjukkan bahwa <i>Naïve</i>

No	Judul Penelitian	Peneliti	Tahun	Hasil Penelitian
				<i>Bayes</i> hanya mampu melakukan klasifikasi dengan benar terhadap 16 dari 33 provinsi dengan keakuratan sebesar 48.48%.
7	<i>Analysis of the effect of the lecturer satisfaction with the Naïve Bayes Data Mining technique on institutional performance</i>	(Aisyah et al., 2021)	2020	Penelitian ini dilakukan dengan tujuan untuk melakukan analisis dari keYaan dosen terhadap kinerja institusi dengan menerapkan algoritma <i>Naïve Bayes</i> dan teknik penambangan data. Penelitian ini dilakukan dengan melakukan kuisisioner dengan beberapa indikator utama yaitu <i>readiness, compassion, reliability</i> dan <i>accountability</i> . Hasil dari penelitian ini menghasilkan akurasi algoritma <i>Naïve Bayes</i> sebesar 85% dan direkomendasikan untuk melakukan prediksi terhadap tingkat keYaan dosen berdasarkan hasil yang diperoleh dari <i>software RapidMiner</i>
8	Algoritma <i>Naïve Bayes</i> untuk Prediksi Pelayanan Perekaman <i>e-KTP</i>	(Apandi & Sugianto, 2019)	2019	Penelitian ini dilakukan untuk mengetahui keYaan Masyarakat terhadap pelayanan perekaman <i>e-KTP</i> di Kecamatan Batujajar dengan menerapkan algoritma <i>Naïve Bayes</i> dan <i>Decision Tree</i> .

No	Judul Penelitian	Peneliti	Tahun	Hasil Penelitian
				<p>Hasil dari penelitian ini menunjukkan bahwa algoritma <i>Naïve Bayes</i> menghasilkan performa yang lebih baik dibandingkan <i>Decision Tree</i> dengan performa <i>Naïve Bayes</i> akurasi 91.70%, dan <i>f1-measure</i> 93.92%. Sementara <i>Decision Tree</i> menghasilkan akurasi 65.90%, dan <i>f1-measure</i> 79,26% dengan nilai test signifikan T mencapai 0.000 yang artinya setiap indikator saling berkaitan.</p>
9	<p>Penerapan <i>Naïve Bayes</i> untuk Memprediksi Tingkat Ke-Ya-an Mahasiswa Terhadap Pelayanan Akademis</p>	<p>(Gustientiedina et al., 2020)</p>	<p>2020</p>	<p>Penelitian ini dilakukan dengan tujuan melakukan prediksi terhadap tingkat keYa-an mahasiswa STIKOM Pelita Indonesia Pekanbaru dan STIE Pelita Indonesia Pekanbaru terhadap pelayanan akademik di kedua kampus ini dengan menerapkan teknik penambangan data dan algoritma <i>Naïve Bayes</i>. Berdasarkan indikator yang telah ditetapkan hasil penelitian ini menunjukkan bahwa algoritma <i>Naïve Bayes</i> dapat memprediksi hingga</p>

No	Judul Penelitian	Peneliti	Tahun	Hasil Penelitian
				mencapai akurasi sebesar 96.71%.
10	Algoritma <i>Naïve Bayes</i> dalam Memprediksi KeYa-an Nasabah	(Sadewo et al., 2019)	2019	Penelitian ini dilakukan dengan tujuan melakukan prediksi keYa-an nasabah Bank BTN Pematangsiantar dengan menerapkan teknik <i>data mining</i> dan algoritma <i>Naïve Bayes</i> . Hasil dari penelitian ini menunjukkan bahwa algoritma <i>Naïve Bayes</i> mencapai akurasi sebesar 88% sehingga dapat disimpulkan bahwa algoritma <i>Naïve Bayes</i> efektif dalam melakukan prediksi terhadap keYa-an pelanggan.

Sumber: (Data Penelitian, 2024)

2.8 Kerangka Pemikiran

Kerangka penelitian adalah bentuk penelitian ini berfokus pada prediksi terhadap pengaruh penggunaan aplikasi *GetContact* terhadap keamanan berdasarkan data yang diperoleh dari kuisioner dari penghuni *Dormitory* Batamindo Blok Q17 Muka Kuning yang akan dijadikan input data metode *Naïve Bayes* dengan bantuan perangkat lunak RapidMiner. Proses ini akan membantu mengevaluasi tingkat pengaruh penggunaan aplikasi *GetContact* terhadap pengguna sehingga dapat menjadi wawasan yang bermanfaat untuk pengguna lain yang ingin menggunakan aplikasi tersebut sebagai langkah preventif keamanan. Dengan demikian kerangka pemikiran ini memberikan landasan penting dalam melakukan prediksi pengaruh penggunaan aplikasi *GetContact* terhadap aspek keamanan. Di bawah ini merupakan kerangka pemikiran yang digunakan dalam penelitian ini.

