

## **BAB 2**

### **TINJAUAN PUSTAKA**

#### **2.1 DATA MINING**

*Data mining* adalah proses pengumpulan informasi dan data penting dalam jumlah besar atau *big data*. Proses ini sering kali menggunakan berbagai metode, seperti matematika, statistika, dan teknologi kecerdasan buatan (*AI*). Proses penambangan data melibatkan beberapa tahapan dan teknik, mulai dari pembersihan data (*cleansing*), integrasi data, seleksi data, dan transformasi data hingga evaluasi pola untuk memperoleh informasi dari data tersebut. (Lestari et al. 2022).

*Data Mining* terbagi menjadi dua kategori, yaitu deskriptif dan prediktif. Fungsi deskriptif bertujuan untuk memahami lebih dalam data yang diamati. Dengan melalui proses tertentu, diharapkan dapat mengungkap karakteristik dari data tersebut. (Lizsara et al. 2020). Data tersebut nantinya dapat digunakan untuk memahami karakteristik yang dimaksud. Sementara itu, fungsi prediktif adalah sebuah proses yang bertujuan menemukan pola-pola tertentu dalam data. Pola-pola ini dapat diidentifikasi melalui berbagai variabel yang terdapat dalam data tersebut. Ketika sudah menemukan pola maka pola yang didapat tersebut bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya. (Utomo and Mesran 2020).

*Data mining* adalah metode, teknik, kecerdasan buatan, dan pembelajaran mesin yang diekstraksi untuk menghasilkan pengetahuan dan informasi berguna yang tersimpan dalam *database* besar. Secara prinsip, *data mining* mewarisi banyak aspek dan teknik dari berbagai bidang ilmu. *Data mining* bukanlah konsep baru, karena memiliki akar dalam berbagai disiplin ilmu tersebut (Nasir 2021).

#### **2.2 KLASIFIKASI**

Klasifikasi adalah proses analisis data yang digunakan untuk menetapkan label kelas pada sampel data yang ingin diidentifikasi. Klasifikasi memerlukan data input beserta hasilnya untuk menemukan pola hubungan dari data yang telah ditentukan

sebelumnya, dengan tujuan meningkatkan keakuratan hasil yang diperoleh dari data yang ada (Srirahayu and Pribadie 2023).

Dalam klasifikasi, terdapat dua tugas utama, pertama membuat model sebagai prototipe yang akan disimpan sebagai referensi, dan kedua, menggunakan model tersebut untuk mengenali, mengklasifikasikan, atau memprediksi objek data lainnya, sehingga dapat menentukan kelas yang sesuai untuk objek data tersebut berdasarkan model yang telah disimpan (Putro et al. 2020).

### 2.3 METODE NAÏVE BAYES

Metode klasifikasi *Naïve Bayes* digunakan untuk menganalisis kelompok data melalui perhitungan probabilitas. Prinsip utama dari *Naïve Bayes* melibatkan penerapan teorema Bayes, sebuah konsep statistik untuk mengestimasi probabilitas. Pengklasifikasi *Naïve Bayes* menghitung probabilitas setiap kelas berdasarkan atribut yang ada dan menentukan kelas yang paling optimal. Tujuannya adalah untuk menghitung dan mencari nilai probabilitas tertinggi sehingga data uji dapat diklasifikasikan ke dalam kategori yang sesuai (Agustiranti et al. 2024).

Berikut ini merupakan langkah-langkah dalam menentukan klasifikasi menggunakan metode *Naïve Bayes*:

1. Menghitung nilai probabilitas tiap kelas pada tiap data

$$P(C_i) = \frac{\sum C_i}{n} \quad (0.1)$$

Keterangan:

$P(C_i)$  : Probabilitas label kelas  $C_i$

$\sum C_i$  : Jumlah data dengan label kelas  $C_i$

$n$  : Jumlah total data latih

2. Menghitung nilai probabilitas awal tiap fitur.

$$P(x_k|C_i) = \frac{\sum x_k|C_i}{\sum C_i} \quad (0.2)$$

Keterangan:

$P(x_k|C_i)$  : Probabilitas fitur  $x_k$  dengan label kelas  $C_i$

$\sum x_k|C_i$  : Jumlah data fitur  $x_k$  dengan label kelas  $C_i$

$\sum C_i$  : Jumlah data dengan label kelas  $C_i$

3. Menghitung nilai probabilitas akhir tiap data menggunakan persamaan 2.2.

## 2.4 SUMBER DAYA MANUSIA

Kualitas sumber daya manusia adalah aspek penting dalam meningkatkan produktivitas perusahaan. Oleh karena itu, perusahaan perlu melakukan penilaian kinerja pegawai untuk menilai sejauh mana mereka berhasil atau gagal dalam melaksanakan tugas-tugas mereka. Namun, dalam memberikan penghargaan kepada pegawai terbaik, sering kali muncul kesulitan dalam memilih pegawai yang sesuai dengan kriteria yang telah ditetapkan. Oleh sebab itu, perusahaan membutuhkan sistem pendukung keputusan untuk membantu menentukan pegawai yang pantas menerima penghargaan. (Abdul Koda et al. 2022).

## 2.5 CONFUSION MATRIX

Setelah hasil klasifikasi diperoleh, langkah berikutnya adalah menguji hasil klasifikasi tersebut. Data uji dihitung menggunakan metode yang dipilih sesuai perhitungan dalam sistem ini, dengan menggunakan fitur dari klasifikasi data latih. Hasil pengujian ini akan dievaluasi performanya menggunakan confusion matrix. Pada prinsipnya, *confusion matrix* menyajikan informasi yang membandingkan hasil klasifikasi yang dihasilkan oleh sistem dengan hasil klasifikasi yang seharusnya. Tujuan dari evaluasi performa ini adalah untuk mengukur keefektifan prediksi yang telah diperoleh dalam perhitungan ini. (Dakhi and Aji 2022).

**Tabel 2. 1** Tabel *Confusion Matrix*

<i>Three Class Prediction</i>		<i>Predicted Class</i>		
		1	2	3
<i>Actual Class</i>	1	P11	P12	P13
	2	P21	P22	P23
	3	P31	P32	P33

$$\text{True Positive 1 (TP1)} = P11$$

$$\text{True Positive 2 (TP2)} = P22$$

$$\text{True Positive 3 (TP3)} = P33$$

$$\text{True Negative 1 (TN1)} = P22+P23+P32+P33$$

$$\text{True Negative 2 (TN2)} = P11+P13+P31+P33$$

$$\text{True Negative 3 (TN3)} = P11+P12+P21+P22$$

$$\text{False Positive 1 (FP1)} = P21+P31$$

$$\text{False Positive 2 (FP2)} = P12+P32$$

$$\text{False Positive 3 (FP3)} = P13+P23$$

$$\text{False Negative 1 (FN1)} = P12+P13$$

$$\text{False Negative 2 (FN2)} = P21+P23$$

$$\text{False Negative 3 (FN3)} = P31+P32$$

$$\text{Accuracy} = \frac{P11+P22+P33}{P11+P12+P13+P21+P22+P23+P31+P32+P33} \quad (0.3)$$

$$\text{Precision 1} = \frac{TP1}{TP1+FP1} \quad (0.4)$$

$$\text{Precision 2} = \frac{TP2}{TP2+FP2} \quad (0.5)$$

$$\text{Precision 3} = \frac{TP3}{TP3+FP3} \quad (0.6)$$

$$\text{Recall 1} = \frac{TP1}{TP1+FN1} \quad (0.7)$$

$$\text{Recall 2} = \frac{TP2}{TP2+FN2} \quad (0.8)$$

$$\text{Recall } 3 = \frac{TP3}{TP3+FN3} \quad (0.9)$$

## 2.6 PENILAIAN KINERJA PEGAWAI

Kinerja adalah bentuk dari kemampuan dalam bentuk hasil nyata dan *output* kerja yang dicapai oleh pegawai dalam menjalankan tugas dan tanggung jawab dari perusahaan atau instansi. Kinerja dipengaruhi oleh berbagai faktor, termasuk kemampuan individu dan lingkungan organisasi. Salah satu strategi untuk meningkatkan kinerja adalah melalui penilaian kinerja pegawai. Hasil penilaian ini dapat digunakan sebagai kesempatan bagi pegawai dan instansi untuk mengevaluasi kemampuan dan potensi mereka. Oleh karena itu, evaluasi kinerja menjadi alat penting untuk meningkatkan kualitas kerja di perusahaan atau instansi. Namun, penentuan pegawai teladan seharusnya tidak hanya didasarkan pada penilaian subyektif, tetapi juga pada hasil kinerja yang dicapai. Sayangnya, masih banyak instansi atau perusahaan yang melakukan penilaian kinerja secara tidak seimbang dan subjektif, sehingga menurunkan kualitas evaluasi. Untuk mengatasi masalah ini, diperlukan solusi untuk menghindari ketidakakuratan dan spekulasi dalam penilaian kinerja (Amida and Kristiana 2019).

## 2.7 PENELITIAN SEBELUMNYA

Berikut beberapa referensi pembelajaran dari beberapa contoh kasus yang hampir serupa dengan permasalahan yang dihadapi, antara lain

**Tabel 2. 2** Tabel penelitian terdahulu

Landasan Literatur	Metode yang digunakan	Masalah	Hasil penelitian
"Penerapan Metode Naïve Bayes Untuk Menentukan Penilaian Kinerja	Metode Naïve Bayes	Membahas penerapan metode Naïve Bayes dalam dalam	Metode Naïve Bayes dapat digunakan dalam proses

<p><i>Karyawan PT.Sinergi Guna Solusindo”.</i> (Sudrajat 2022)</p>		<p>meningkatkan kinerja sumber daya manusia dengan cara melakukan penilaian prestasi kerja</p>	<p>penyeleksian untuk menentukan penilaian kinerja karyawan PT. Sinergi Guna Solusindo.</p>
<p><i>“Evaluasi Penilaian Kinerja Karyawan Dengan Metode Naïve Bayes”.</i> (Kenny et al. 2024b)</p>	<p><i>Metode Naïve Bayes</i></p>	<p>Membahas penerapan metode klasifikasi <i>Naïve Bayes</i> guna mencari keakuratan data dalam mengklasifikasi penilaian kinerja karyawan mengenai kompeten dan tidak kompeten seorang karyawan.</p>	<p>Metode <i>Naive Bayes</i> dapat menghasilkan nilai untuk mengklasifikasi kompeten dan tidak kompeten karyawan.</p>
<p><i>“Penentuan Bonus Karyawan Dengan Menggunakan Algoritma K-Nearest Neighbor”.</i>(Abdul Koda et al. 2022)</p>	<p><i>Algoritma K-Nearest Neighbor</i></p>	<p>Membahas transparansi proses penilaian yang dapat memberikan efek positif bagi peningkatan</p>	<p><i>Algoritma K-Nearest Neighbor</i> berhasil memberikan penilaian yang dapat menentukan Bonus Karyawan</p>

		motivasi kerja pegawai	
<p>“<i>Data Mining Menggunakan Algoritma Naïve Bayes Classifier Untuk Evaluasi Kinerja Karyawan</i>”. (Haya Pangestu 2023)</p>	<p><i>Algoritma Naïve Bayes Classifier</i></p>	<p>Menerapkan Data Mining untuk evaluasi kinerja karyawan menggunakan metode <i>Naïve Bayes Classifier</i></p>	<p>Metode <i>Naive Bayes</i> berhasil memberikan penilaian evaluasi kinerja karyawan untuk perpanjangan kontrak</p>
<p>“<i>Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naïve Bayes</i>”. (Hozairi et al. 2021)</p>	<p><i>K-Nearest Neighbor, Decision Tree Serta Naïve Bayes</i></p>	<p>Menerapkan aplikasi orange <i>data mining</i> dengan menggunakan model <i>K-Nearest Neighbor (K-NN)</i>, <i>Decision Tree</i> serta <i>Naive Bayes</i> untuk klasifikasi ketepatan kelulusan mahasiswa</p>	<p>Klasifikasi tingkat kelulusan mahasiswa Prodi Teknik Informatika Universitas Islam Madura merekomendasikan model <i>Naive Bayes</i> karena memiliki tingkat akurasi lebih baik dibanding <i>K-NN</i> dan <i>Decision Tree</i></p>
<p>“<i>Prediksi Penerimaan Pegawai Baru Dengan Metode</i></p>	<p><i>Metode Naive Bayes</i></p>	<p>Membahas penerapan metode klasifikasi <i>Naïve Bayes</i> untuk</p>	<p>Metode <i>Naive Bayes</i> berhasil memberikan penilaian untuk</p>

<i>Naive Bayes</i> ". (Huda et al. 2020)		penilaian pegawai baru	diterima sebagai pegawai baru
<i>"Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma Naive Bayes"</i> . (Damanik et al. 2021)	<i>Algoritma Naive Bayes</i>	Memprediksi tingkat kepuasan terhadap pembelajaran daring menggunakan Algoritma <i>Data Mining</i> metode <i>Naive Bayes</i>	Model Algoritma <i>Naive Bayes</i> bisa direkomendasikan untuk prediksi tingkat kepuasan dalam pembelajaran daring pada AMIK Tunas Bangsa Pematangsiantar
<i>"Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma Naive Bayes Classifier"</i> . (Dewi et al. 2021)	<i>Metode Algoritma Naive Bayes Classifier</i>	Menentukan layak tidaknya seorang karyawan yang telah dipromosikan untuk naik jabatan	Sistem pendukung keputusan yang dibangun ini dapat diaplikasikan pada PT. Busana Indah Global (BIG) sehingga mampu mempermudah dalam menentukan kelayakan kenaikan jabatan bagi karyawannya
<i>"Penerapan Data Mining Penerimaan Karyawan Menggunakan"</i>	<i>Metode Naive Bayes</i>	Penerapan algoritma <i>Naive Bayes</i> untuk menentukan calon	Dengan adanya sistem ini memudahkan dalam menentukan

<p><i>Metode Naive Bayes Classifier</i>". (Miharja 2021)</p>		<p>tenaga kerja yang sesuai dengan keinginan perusahaan</p>	<p>calon tenaga kerja yang diharapkan oleh perusahaan dan Dapat mengklarifikasikan kelayakan calon tenaga kerja</p>
<p><i>"Implementasi Metode Naive Bayes Dalam Penyeleksian Karyawan untuk Penempatan Bagian Pemasaran"</i>. (Fitriani et al. 2022)</p>	<p><i>Metode Naive Bayes</i></p>	<p>Metode klasifikasi data mining untuk memprediksi penyeleksian karyawan untuk bagian Pemasaran yaitu dengan mnggunakan metode <i>Naive Bayes</i></p>	<p>Metode <i>Naive Bayes</i> menjadi metode yang baik untuk implementasi pada penyeleksian karyawan untuk penempatan bagian Pemasaran</p>