

BAB 2

TINJAUAN PUSTAKA

2.1. SUMBER DAYA MANUSIA

Kualitas sumber daya manusia adalah salah satu faktor penting dalam meningkatkan produktivitas perusahaan. Oleh karena itu, perusahaan perlu melakukan evaluasi kinerja karyawan guna menilai keberhasilan atau kegagalan mereka dalam menjalankan tugas-tugasnya. Namun, dalam memberikan penghargaan kepada karyawan terbaik, seringkali terdapat kendala dalam memilih karyawan yang layak berdasarkan kriteria yang telah ditetapkan. Oleh karena itu, perusahaan membutuhkan sistem pendukung keputusan untuk membantu menentukan karyawan yang pantas menerima penghargaan (Abdul Koda et al. 2022).

2.2. DATA MINING

Data mining adalah suatu proses pengumpulan informasi dan data yang penting dalam jumlah yang besar atau *big data* (Sri Widaningsih 2019). Dalam proses ini sering kali memanfaatkan beberapa metode, seperti matematika, statistika dan pemanfaatan teknologi *artificial intelligence* (AI). Proses penambangan data terdiri dari beberapa tahapan dan teknik, dari adanya *cleansing* (pembersihan data), integrasi data, seleksi data dan data *transformation* hingga evaluasi pola dalam mendapatkan informasi dari data itu (Lestari et al. 2022). Selain itu fungsi *data mining* terbagi menjadi dua bagian, yakni deskriptif dan prediktif. Deskriptif sendiri berfungsi untuk memahami lebih jauh tentang data yang diamati, dengan melakukan sebuah proses diharap bisa mengetahui perilaku dai sebuah data tersebut (Lizsara et al. 2020). Data tersebut itulah yang nantinya dapat digunakan untuk mengetahui karakteristik dari data yang dimaksud. Sedangkan fungsi *predictive* adalah sebuah bagaimana sebuah proses nantinya akan menemukan pola tertentu dari suatu data, pola – pola tersebut dapat diketahui dari berbagai variabel yang ada pada data (Utomo and Mesran 2020). Ketika sudah menemukan pola maka pola yang didapat tersebut bisa digunakan untuk memprediksi variabel lain yang belum

diketahui nilai ataupun jenisnya. Selain dua fungsi tersebut *data mining* juga mempunyai fungsi seperti *characterization*, *discrimination*, *association*, *classification*, *clustering*, *outlier and trend analysis*, dan lain lain (Wibawa 2018).

2.3. KLASIFIKASI

. Klasifikasi adalah proses untuk mengidentifikasi sebuah objek data sebagai salah satu dari jenis yang telah dijelaskan sebelumnya. Metode pengelompokan klasifikasi didasarkan pada karakteristik yang dimiliki oleh suatu objek dan bisa dilakukan baik oleh manusia maupun dengan menggunakan teknologi (Sutisna and Yuniar 2023). Dalam klasifikasi, terdapat dua tugas utama, yakni membuat sebuah model sebagai bentuk prototipe yang akan disimpan sebagai referensi, dan menggunakan model tersebut untuk mengenali/mengklasifikasikan/melakukan prediksi terhadap objek data lain untuk menentukan kelas mana tempat objek data tersebut berada dalam model yang telah disimpan (Putro et al. 2020).

2.4. METODE NAÏVE BAYES

Model yang digunakan untuk perbandingan pertama adalah model *Naïve Bayes*, yang merupakan sebuah algoritma *machine learning* yang digunakan dalam berbagai jenis klasifikasi. Algoritma ini didasarkan pada rumus umum Teorema *Bayes*, yang menjadi dasar dari *Naïve Bayes* (Fatmawati 2016). Dalam *Naïve Bayes*, penambahan bobot pada setiap atribut dalam dataset menjadi dasar untuk meningkatkan akurasi klasifikasi. Hal ini tidak hanya melibatkan probabilitas setiap atribut, tetapi juga bobot dari setiap atributnya.

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

$$P(C_i) = \frac{\sum C_i}{n} \quad (2.1)$$

Keterangan:

$P(C_i)$: Probabilitas label kelas C_i

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

n : Jumlah total data latih

Menghitung nilai probabilitas tiap fitur.

$$P(x_k|C_i) = \frac{\sum x_k|C_i}{\sum C_i} \quad (2.2)$$

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

Menghitung nilai probabilitas tiap kelas pada tiap data.

Contoh perhitungan metode *Naïve Bayes* :

Tabel 2.1 Data Training Karyawan PT. Sumatera

Tanggung Jawab	Komunikasi	Disiplin	Sikap	Keahlian	Kelas
A	B	A	B	B	Cukup
B	A	B	A	A	Baik
A	B	A	B	B	Cukup
B	A	B	A	A	Baik
A	B	A	B	B	Cukup
B	A	B	A	A	Baik

Pada **Tabel 2.1** akan dilakukan proses perhitungan probabilitas pada setiap kelas dengan mengambil salah satu bagian dari data *training* untuk dijadikan sebagai data testing pada **Tabel 2.2**.

Tabel 2.2 Contoh Data Testing Karyawan PT. Sumatera

Tanggung Jawab	Komunikasi	Disiplin	Sikap	Keahlian	Kelas
B	A	B	B	B	?

Pertanyaan :

Dari data testing di atas, coba klasifikasikan termasuk ke dalam kelas baik atau cukup ?

Perhitungan :

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

Keterangan:

$P(C_i)$: Probabilitas label kelas C_i

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

n : Jumlah total data latih

Label Kelas 1 :

$$P(1S) = \frac{\sum 1}{n} = \frac{3}{6} = 0,5$$

Label Kelas 2 :

$$P(2S) = \frac{\sum 2}{n} = \frac{3}{6} = 0,5$$

Menghitung Probabilitas Awal :

Pada proses menghitung Probabilitas awal akan menggunakan rumus pada (2.2). Pada tahap ini akan dihitung dengan rumus tersebut dan diimplementasikan ke semua atribut.

Tabel 2.3 Jumlah Data Fitur Tanggung Jawab

Tanggung Jawab	Baik	Cukup	Total
1	1	1	2
2	3	1	4
Total	4	2	6

Tabel 2.4 Nilai Probabilitas awal fitur tanggung jawab

Tanggung Jawab	Baik	Cukup
A	0,25	0,5
B	0,75	0,5

Tabel 2.5 Jumlah data fitur komunikasi

Komunikasi	Baik	Cukup	Total
A	3	0	3
B	1	2	3
Total	4	2	6

Tabel 2.6 Nilai Probabilitas awal fitur komunikasi

Komunikasi	Baik	Cukup
A	0,75	0
B	0,25	1

Tabel 2.7 Jumlah data fitur disiplin

Disiplin	Baik	Cukup	Total
A	1	2	3
B	3	0	3
Total	4	2	6

Tabel 2.8 Nilai Probabilitas awal fitur disiplin

Disiplin	Baik	Cukup
A	0,25	1
B	0,75	0

Tabel 2.9 Jumlah data fitur sikap

Sikap	Baik	Cukup	Total
A	3	0	3
B	1	2	3
Total	4	2	6

Tabel 2.10 Nilai Probabilitas awal fitur sikap

Sikap	Baik	Cukup
A	0,75	0
B	0,25	1

Tabel 2.11 Jumlah data fitur keahlian

Keahlian	Baik	Cukup	Total
A	3	0	3
B	1	2	3
Total	4	2	6

Tabel 2.12 Nilai Probabilitas awal fitur keahlian

Keahlian	Baik	Cukup
A	0,75	0
B	0,25	1

Setelah melakukan perhitungan pada probabilitas awal dengan menggunakan data pada **Tabel 2.2** untuk setiap atribut, akan dilakukan perhitungan probabilitas akhir pada semua atribut yang telah dihitung pada probabilitas awal.

Menghitung nilai probabilitas akhir tiap fitur

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

Keterangan:
 $P(C_i)$: Probabilitas label kelas C_i
 $\sum C_i$: Jumlah data dengan label kelas C_i
 n : Jumlah total data latih

Menghitung Probabilitas Akhir :

Label Kelas 1 :

$$P(1S) = \frac{\sum 1}{n} = \frac{3}{6} = 0,5$$

Label Kelas 2 :

$$P(2S) = \frac{\sum 2}{n} = \frac{3}{6} = 0,5$$

Pada tahap ini dilakukan proses perhitungan nilai probabilitas akhir berdasarkan nilai probabilitas awal yang sebelumnya sudah dilakukan proses perhitungan.

Tabel 2.13 Jumlah data fitur tanggung jawab

Tanggung Jawab	Baik	Cukup	Total
A	0	3	3
B	3	0	3
Total	3	3	6

Tabel 2.14 Nilai probabilitas akhir fitur tanggung jawab

Tanggung Jawab	Baik	Cukup
A	0	1
B	1	0

Tabel 2.15 Jumlah data fitur komunikasi

Komunikasi	Baik	Cukup	Total
A	3	0	3
B	0	3	3
Total	3	3	6

Tabel 2.16 Nilai probabilitas akhir fitur komunikasi

Komunikasi	Baik	Cukup
A	1	0
B	0	1

Tabel 2.17 Jumlah data fitur disiplin

Disiplin	Baik	Cukup	Total
A	0	3	3
B	3	0	3
Total	3	3	6

Tabel 2.18 Nilai probabilitas akhir fitur disiplin

Disiplin	Baik	Cukup
A	0	1
B	1	0

Tabel 2.19 Jumlah data fitur sikap

Sikap	Baik	Cukup	Total
A	3	0	3
B	0	3	3
Total	3	3	6

Tabel 2.20 Nilai probabilitas akhir fitur sikap

Sikap	Baik	Cukup
A	1	0
B	0	1

Tabel 2.21 Jumlah data fitur keahlisan

Keahlian	Baik	Cukup	Total
A	3	0	3
B	1	2	3
Total	4	2	6

Tabel 2.22 Nilai probabilitas akhir keahlian

Keahlian	Baik	Cukup
A	0,75	0
B	0,25	1

Pada proses ini setelah melakukan semua perhitungan probabilitas akhir pada semua atribut dilakukan proses perhitungan nilai akhir untuk setiap kelas dengan perhitungan seperti berikut :

$$P(1|X) = 0,5 \times 0,667 \times 1 \times 1 \times 1 \times 0 \times 0,25$$

$$P(1|X) = 0,0001$$

$$P(2|X) = 0,5 \times 0,333 \times 0 \times 0 \times 0 \times 0 \times 0,25$$

$$P(2|X) = 0$$

Tabel 2.23 Hasil klasifikasi akhir

Tanggung Jawab	Komunikasi	Disiplin	Sikap	Keahlian	Kelas
B	A	B	B	B	?

Dari proses perhitungan sampai klasifikasi mendapatkan hasil bahwa dari kriteria kinerja karyawan tersebut masuk kedalam kelas Baik.

2.5. *CONFUSION MATRIX*

Setelah hasil klasifikasi didapatkan, proses selanjutnya adalah melakukan pengujian dari hasil klasifikasi. Data uji kemudian dihitung dengan metode yang dipilih pada perhitungan yang ada pada sistem ini, dengan fitur dari klasifikasi data latih. Hasil dari pengujian ini akan dilakukan evaluasi performa menggunakan *confusion matrix*. Secara prinsip, *confusion matrix* berisi informasi yang membandingkan hasil klasifikasi yang dihasilkan oleh sistem dengan hasil klasifikasi yang seharusnya. Evaluasi performa ini bertujuan untuk mengukur keefektifan dari hasil prediksi yang telah didapatkan pada perhitungan ini (Dakhi and Aji 2022).

Tabel 2.24 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

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

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

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

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

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

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

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

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

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

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

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

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

$$\text{Accuracy} = \frac{P11+P22+P33}{P11+P12+P13+P21+P22+P23+P31+P32+P33} \quad (2.3)$$

$$\text{Precision 1} = \frac{TP1}{TP1+FP1} \quad (2.4)$$

$$\text{Precision 2} = \frac{TP2}{TP2+FP2} \quad (2.5)$$

$$\text{Precision 3} = \frac{TP3}{TP3+FP3} \quad (2.6)$$

$$\text{Recall 1} = \frac{TP1}{TP1+FN1} \quad (2.7)$$

$$\text{Recall 2} = \frac{TP2}{TP2+FN2} \quad (2.8)$$

$$\text{Recall 3} = \frac{TP3}{TP3+FN3} \quad (2.9)$$

2.6. METODE WATERFALL

Metode *waterfall* adalah salah satu model atau pendekatan pengembangan perangkat lunak yang bersifat *linier* atau berurutan, di mana setiap tahapannya harus diselesaikan secara berurutan dan tahapan selanjutnya baru dimulai setelah tahapan sebelumnya selesai (Nur 2019).

Setiap fase memiliki masukan dan keluaran yang terperinci dan prosesnya berurutan serta tidak berulang. Pendekatan ini sangat sesuai untuk pengembangan perangkat lunak yang stabil, membutuhkan sedikit perubahan kebutuhan, dan memiliki spesifikasi yang jelas. Di bawah ini terdapat ilustrasi dari pendekatan model air terjun (*waterfall*).



Gambar 2.1 Ilustrasi metode *Waterfall*

2.7. BLACK BOX TESTING

Black Box Testing adalah sebuah teknik dalam pengujian perangkat lunak, dilakukan tanpa pengetahuan rinci tentang cara kerja kode program di dalamnya. Metode ini memperlakukan sistem atau program sebagai kotak hitam di mana input dimasukkan dan output yang dihasilkan dievaluasi, tanpa memperhatikan bagaimana proses program berjalan di dalamnya. (Hidayat and Muttaqin 2020).

Setiap fase memiliki masukan dan keluaran yang terperinci dan prosesnya berurutan serta tidak berulang. Pendekatan ini sangat sesuai untuk pengembangan perangkat lunak yang stabil, membutuhkan sedikit perubahan kebutuhan, dan memiliki spesifikasi yang jelas. Di bawah ini terdapat ilustrasi dari pendekatan model *waterfall*.

2.8. PENELITIAN SEBELUMNYA

Berikut Beberapa referensi pembelajaran dari beberapa contoh kasus yang hampir sama dengan permasalahan yang dihadapi, antara lain :

1. (Sudrajat 2022) dengan judul “Penerapan Metode Naïve Bayes Untuk Menentukan Penilaian Kinerja Karyawan PT.Sinergi Guna Solusindo”.

Penelitian tersebut membahas tentang penerapan metode naïve bayes untuk penilaian kinerja karyawan yang ada di perusahaan. Hasil penelitian tersebut menyebutkan bahwa didapatkan hasil prediksi menentukan karyawan tetap dengan cepat dan akurat, dari pengujian yang dilakukan dengan membandingkan data training dengan data testing menggunakan aplikasi pendukung Rapid Miner didapat tingkat akurasi sebesar 94.00% (Sudrajat 2022).

2. (Purba et al. 2019) dengan judul “Implementasi Data Mining Untuk Pengelompokan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering”.

Penelitian tersebut membahas tentang Implementasi Data Mining Untuk Pengelompokan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering dan Metode K-Mean Clustering dapat digunakan untuk melakukan pengelompokan dan prediksi karyawan yang berpotensi PHK ke dalam 2 kelompok, yaitu memiliki potensi PHK dan tidak memiliki potensi PHK (Purba et al. 2019).

3. (Abdul Koda et al. 2022) dengan judul “Penentuan Bonus Karyawan Dengan Menggunakan Algoritma K-Nearest Neighbor”.

Penelitian tersebut membahas tentang Prediksi penjualan produk Unilever menggunakan metode *K-Nearest Neighbor*, Dari penelitian tersebut dihasilkan bahwa Berdasarkan hasil perhitungan data mining menggunakan teknik klasifikasi dan algoritma K-Penelitian untuk menentukan Bonus Karyawan melibatkan faktor-faktor seperti Lamanya Masa Kerja, Kehadiran yang baik, Loyalitas terhadap perusahaan, dan Hasil Pekerjaan yang dicapai. Dalam penelitian ini, ditemukan bahwa tingkat akurasi hasil penilaian mencapai 86,81% (Abdul Koda et al. 2022).

4. (Widaningsih 2019) dengan judul “Perbandingan metode data mining untuk prediksi nilai dan waktu kelulusan mahasiswa program studi Teknik informatika dengan algoritma, Naïve Bayes, KNN, dan SVM”.

Penelitian tersebut membahas tentang hasil perbandingan terlihat bahwa algoritma *Naïve bayes* memiliki nilai yang paling baik untuk semua kategori performansi dibandingkan dengan algoritma lainnya. Untuk nilai accuracy dan AUC nilai terbesar adalah yang terbaik, sedangkan untuk error adalah nilai yang terkecil. Nilai AUC untuk Naïve bayes dan C4.5 termasuk ke dalam kategori “baik”, sedangkan untuk algoritma SVM dan KNN termasuk ke dalam kategori “cukup” (Widaningsih 2019).

5. (Hozairi et al. 2021) dengan judul “Implementasi orange data mining untuk klasifikasi kelulusan mahasiswa dengan model k-nearest neighbor, decision tree serta naïve bayes”.

Penelitian tersebut membahas tentang menggunakan model K-Nearest Neighbor, Decision tree serta naïve bayes untuk mengklasifikasi status kelulusan mahasiswa Teknik informatika universitas Islam Madura diperoleh hasil bahwa kinerja naïve bayes lebih unggul dari k-nearest neighbor serta decision tree. Terbukti bahwa dari 35 data uji yang digunakan naïve bayes memiliki nilai akurasi 89%, presisi 88%, sedangkan k-nearest neighbor memiliki nilai akurasi 77%, presisi 76% dan decision tree memiliki nilai akurasi 74% dan presisi 84%. Kontribusi riset ini bisa digunakan oleh manajemen program studi Teknik informatika Universitas Islam Madura untuk mendeteksi sejak awal kondisi mahasiswa supaya kelulusannya tidak terlambat dan mempengaruhi nilai akreditasi program studi Teknik informatika Universitas Islam Madura (Alim 2021).

6. (Huda et al. 2020) dengan judul “Prediksi Penerimaan Pegawai Baru Dengan Metode *Naive Bayes*”.

Penelitian tersebut membahas tentang Penerimaan pegawai baru dengan metode *Naive Bayes*, maka diperoleh hasil persentase 100% untuk keakuratan prediksi, di mana dari 17 data pegawai baru yang diuji berhasil diklasifikasikan semuanya dengan benar (Huda et al. 2020).

7. (Damanik et al. 2021) dengan judul “ Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma *Naïve Bayes*.”

Penelitian tersebut membahas tentang memprediksi tingkatan kepuasan mahasiswa dalam pendidikan daring serta membagikan donasi pada akademi besar dalam mengambil kebijakan yang berhubungan dengan kenaikan mutu pendidikan secara daring. Informasi yang digunakan diperoleh dengan membagikan kuesioner kepada mahasiswa angkatan 2020/2021 sebanyak 110 mahasiswa (Damanik et al. 2021).

8. (Dewi et al. 2021) dengan judul “Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma *Naïve Bayes Classifier*”.

Penelitian tersebut membahas tentang . Dari pengujian di *RapidMiner* menghasilkan nilai akurasi sebesar 91,67% dan nilai ROC sebesar 0,979 yang artinya algoritma *Naïve Bayes Classifier* sangat baik digunakan pada penelitian ini. Setelah melakukan pengujian menggunakan software *RapidMiner* dan mendapatkan hasil pengujian, kemudian diimplementasikan ke dalam sebuah sistem menggunakan PHP dan MySQL yang dirancang untuk melakukan prediksi kenaikan jabatan (Dewi et al. 2021).

9. (Miharja 2021) dengan judul Penerapan Data Mining Penerimaan Karyawan Menggunakan Metode *Naive Bayes Classifier*.

Penelitian tersebut membahas tentang Seleksi tenaga kerja di perusahaan hitung dengan metode *Naive Bayes* dan akan menghasilkan sebuah keputusan layak atau tidaknya seorang tenaga kerja itu sendiri . Dengan adanya sistem ini memudahkan permasalahan dalam menentukan calon tenaga kerja yang diharapkan Oleh perusahaan dan Dapat mengklasifikasikan kelayakan calon tenaga kerja dan dapat dipertanggungjawabkan oleh pihak perusahaan (Miharja 2021).

10. (Fitriani et al. 2022) dengan judul “Implementasi Metode *Naive Bayes* Dalam Penyeleksian Karyawan untuk Penempatan Bagian Pemasaran”.

Penelitian tersebut membahas tentang penyeleksian karyawan untuk bagian Pemasaran dengan menggunakan metode *naive bayes*

menghasilkan nilai akurasi sebesar 87,22% dan nilai AUC sebesar 0,920 dengan tingkat diagnosa Excellent Classification. Sehingga didapat kesimpulan bahwa menggunakan metode naive bayes bisa menjadi metode yang baik untuk implementasi pada penyeleksian karyawan untuk penempatan bagian Pemasaran (Fitriani et al. 2022).

