BAB II

TINJAUAN PUSTAKA

2.1 Kepuasan Pelanggan

Kepuasan pelanggan merujuk pada sejauh mana perasaan atau penilaian positif yang dirasakan oleh pelanggan setelah membandingkan harapan mereka dengan pengalaman yang diperoleh dari hasil penyediaan produk maupun layanan oleh entitas perusahaan. Jika pengalaman yang dirasakan pelanggan melebihi atau sesuai dengan ekspektasi mereka, pelanggan akan merasa puas. Sebaliknya, apabila pengalaman tersebut tidak memenuhi harapan, ketidakpuasan akan muncul. Oleh karena itu, Agar tetap bersaing di pasar, perusahaan harus memperhatikan kepuasan pelanggan (Oktamala & Zuraidah, 2021). Kepuasan pelanggan memegang peranan penting dalam kesuksesan jangka panjang sebuah perusahaan, karena dapat berdampak pada loyalitas pelanggan, citra perusahaan, dan profitabilitas.

2.2 Data Mining

Data mining mulai dikenal pada dekade 1990-an sebagai teknik yang efisien untuk menggali informasi penting, untuk mengidentifikasi relasi data dan mengelompokkan data ke pada satu hingga beberapa kategori oleh sebab itu objek kategori memiliki kesamaan yang jelas (Mai et al., 2022). Data mining yakni proses pembelajaran yang melibatkan pengumpulan, pembersihan, pengolahan, analisis, dan penemuan informasi berharga dari data (Elisa Sinaga et al., 2022). Data mining yaitu proses penyaringan atau pencarian data yang menggunakan kumpulan *big data* melalui banyak tahapan guna memperoleh informasi penting dari data (Asyuti & Setyawan, 2023). Di dalam ranah ilmu komputer, data mining untuk mengeksplorasi data guna menemukan pola tersembunyi yang menghasilkan pengetahuan baru dalam kumpulan data (Tarigan et al., 2022).

Beberapa pengertian data mining tersebut dapat di simpulkan data mining merupakan suatu metode yang berkembang sejak tahun 1990-an untuk mengekstrak

pola dan informasi berharga dari kumpulan data besar melalui serangkaian tahapan, seperti pengumpulan, pembersihan, pengolahan, dan analisis. Sasaran utama dari data mining adalah untuk menemukan hubungan antar data, mengelompokkan data ke dalam cluster yang memiliki kesamaan tinggi, serta mengeksplorasi pola tersembunyi guna menghasilkan pengetahuan baru. Dalam beberapa pengertian data mining, melibatkan beberapa data *training* dan data *testing* dengan memilih algoritme dalam pengelolaan data dapat diterapkan dalam penelitian ini dan memberikan hasil sebuah prediksi.

Dalam pengumpulan informasi, dan pengolahan data yang di uji, data mining menerapkan berbagai teknik, antara lain:

- a. *Association Discovery*, Proses pada data mining bertujuan mengidentifikasi hubungan atau pola menarik antara variabel pada suatu kumpulan data. Teknik ini umum digunakan menemukan asosiasi yang sering muncul di dalam data.
- b. *Clustering*, merupakan teknik untuk mengelompokkan data berdasarkan kriteria atau karakteristik yang dimiliki oleh setiap data (Mahasiswa & Samudra, 2023).
- c. *Sequential Discovery*, merupakan teknik untuk mengidentifikasi pola di antara peristiwa yang terjadi dalam periode waktu tertentu. Metode ini digunakan untuk menemukan pola komoditas yang berulang, dengan fokus pada kebiasaan yang memiliki potensi pada masa depan.
- d. *Classification*, merupakan teknik yang digunakan untuk mengklasifikasikan data ke dalam kategori atau kelas tertentu berdasarkan pola yang ditemukan dalam data. Bertujuan untuk memprediksi kelas atau label dari data berdasarkan data yang sudah diketahui atau dilabeli sebelumnya.
- e. *Neural Network*, merupakan metode khusus untuk mengidentifikasi pola dan mengontrol tren. Inti dari proses ini adalah meniru cara kerja sistem saraf manusia

2.3 Klasifikasi

Klasifikasi yakni suatu metode untuk mengenali pola atau hubungan tertentu guna menerangkan atau memisahkan berbagai kelompok atau konsep data. Tujuannya adalah untuk memperkirakan kategori suatu objek yang sebelumnya belum memiliki label (Febriani & Sulistiani, 2021). Proses ini umumnya melibatkan pelatihan model menggunakan data yang telah dilabeli sebelumnya, sehingga model tersebut dapat mempelajari pola atau hubungan antara fitur-fitur data dan kelas yang sesuai. Setelah pelatihan, model yang telah dibangun untuk mengklasifikasikan objek baru yang tidak memiliki label, berdasarkan kesamaan karakteristik atau fitur yang ada.

2.4 *Naïve Bayes Classifier*

Metode Naïve Bayes Classifier bekerja berdasarkan probabilitas dengan memanfaatkan teorema *Bayes*, dimana setiap fitur dianggap tidak saling mempengaruhi. (Prasetyo et al., 2022a). Metode *Naïve Bayes* untuk mengelompokkan data pada kategori-kategori tertentu. Kinerja dari algoritma ini dievaluasi berdasarkan tingkat akurasi prediksi yang dihasilkannya (Anwar, 2022). Algoritma *Naïve Bayes* dijalankan melalui beberapa tahap, yaitu:

1. Probabilitas A Priori (*Prior Probability*)

Menggambarkan probabilitas awal suatu kelas sebelum mempertimbangkan data fitur. Dihitung berdasarkan frekuensi kemunculan kelas dalam dataset pelatihan.

$$P(C) = \frac{\text{Jumlah data yang termasuk kelas } C}{\text{Jumlah total data}} \quad (2.1)$$

Keterangan:

- a) $P(C)$ adalah probabilitas awal (prior) dari kelas C.
- b) C adalah kelas yang sedang dianalisis.
- c) Jumlah data yang termasuk kelas C adalah jumlah contoh data yang masuk ke dalam kelas C.

- d) Jumlah total data adalah jumlah keseluruhan contoh data dalam dataset pelatihan.

2. Probabilitas Kondisional (*Likelihood*)

Adalah probabilitas fitur X_i diberikan kelas C . Untuk mengukur kemungkinan setiap fitur muncul di dalam kelas tertentu.

$$P(X_i|C) = \frac{\text{Jumlah fitur } X_i, \text{ dalam kelas } C}{\text{Jumlah total data dalam kelas } C} \quad (2.2)$$

Keterangan:

- $P(X_i|C)$ adalah probabilitas fitur X_i diberikan kelas C .
- X_i adalah fitur ke- i dalam dataset (misalnya, harga produk, rating, atau fitur lainnya).
- Jumlah fitur X_i dalam kelas C adalah berapa kali fitur X_i muncul di dalam kelas C .
- Jumlah total data dalam kelas C adalah jumlah total contoh data yang ada dalam kelas C .

3. Teorema Bayes (*Posterior Probability*)

Teorema Bayes digunakan untuk menghitung probabilitas posteriori, yaitu probabilitas suatu kelas C diberikan data fitur $X_1, X_2, \dots, X_n$. *Naïve Bayes* menggunakan asumsi independensi antar fitur, sehingga dapat menghitung probabilitas posteriori dengan mengalikan probabilitas setiap fitur yang diberikan kelas.

$$P(C|X_1, X_2, \dots, X_n) \propto \prod_{i=1}^n P(X_i|C) \quad (2.3)$$

Keterangan:

- a) $P(C|X_1, X_2, \dots, X_3)$ adalah probabilitas posteriori kelas C diberikan fitur $X_1, X_2, \dots, X_3$
- b) $P(C)$ adalah probabilitas prior kelas C (seperti yang dihitung di rumus pertama).
- c) $P(X_1|C)$ adalah probabilitas fitur X_1 diberikan kelas C (seperti yang dihitung di rumus kedua).
- d) $\prod_{i=1}^n P(X_i|C)$ adalah perkalian dari probabilitas kondisi untuk semua fitur $X_1, X_2, \dots, X_3$
- e) Tanda $\propto$ menunjukkan bahwa probabilitas posteriori proporsional dengan produk dari prior dan likelihood.

Asumsi *Naïve Bayes*: Dalam algoritma *Naïve Bayes*, diasumsikan bahwa seluruh fitur X_1 hingga X_3 bersifat independen satu sama lain, sehingga bisa mengalikan probabilitas kondisional masing-masing fitur tanpa memperhatikan hubungan antar fitur.

4. Klasifikasi (Prediksi Kelas)

Setelah menghitung probabilitas posteriori pada setiap kelas, Kelas dengan kemungkinan kemunculan tertinggi dipilih sebagai hasil prediksi untuk data yang diberikan. Probabilitas posterior yang paling besar menunjukkan bahwa kelas tersebut merupakan prediksi yang paling memungkinkan.

$$\hat{C} = \arg \max_c P(C) \prod_{i=1}^n P(X_i|C) \quad (2.4)$$

Keterangan:

- a) $\hat{C}$ adalah kelas yang diprediksi.

- b) $\arg \max_c$ berarti kita memilih kelas C yang menghasilkan nilai probabilitas tertinggi.
- c) $P(C) \prod_{i=1}^n P(X_i|C)$ adalah hasil perhitungan posteriori untuk kelas C yang diberikan fitur $X_1, X_2, \dots, X_3$

2.5 Confusion Matrix

digunakan menggambarkan dan mengevaluasi kesesuaian suatu algoritma *classification*. (A'yunnisa et al., 2022). Tabel *Confusion Matrix* memberikan ilustrasi mengenai sejauh mana model mampu mengklasifikasikan data ke pada kategori yang tepat.

Untuk klasifikasi biner, Struktur *Confusion Matrix* disusun dari 4 elemen penting berikut:

1. *True Positive (TP)*: Total data aktual *positive* yang berhasil diidentifikasi secara akurat sebagai positif oleh model.
2. *True Negative (TN)*: Menunjukkan total data yang actual negatif dan diklasifikasikan secara akurat sebagai negatif.
3. *False Positive (FP)*: Menggambarkan total data yang seharusnya negatif, namun diartikan positif (dikenal juga sebagai kesalahan tipe I).
4. *False Negative (FN)*: Adalah total data yang sebenarnya positif, tetapi diartika sebagai negatif (disebut juga kesalahan tipe II).

Tabel *Confusion Matrix* dapat dilihat pada tabel berikut:

Tabel 2.1 *Confusion Matrix*

	Prediksi Positif (P)	Prediksi Negatif (N)
Aktual Positif (P)	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Aktual Negatif (N)	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Pada tabel 2.1 Misalkan memiliki model digunakan untuk memprediksi apakah pelanggan adalah puas atau tidak puas (klasifikasi biner: puas vs tidak puas).

- a) *True Positive (TP)*: Pelanggan yang sebenarnya puas dan prediksi klasifikasi dengan benar sebagai puas.
- b) *True Negative (TN)*: Pelanggan sebenarnya tidak puas dan diklasifikasikan dengan tepat sebagai tidak puas.
- c) *False Positive (FP)*: Pelanggan yang actual tidak puas tetapi prediksi menunjukkan sebagai puas.
- d) *False Negative (FN)*: Pelanggan yang actual puas tetapi prediksi menunjukkan sebagai tidak puas.

Dari *Confusion Matrix*, kita bisa menjumlahkan metrik evaluasi model klasifikasi, antara lain:

Akurasi:
$$\frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.5)$$

Recall:
$$\frac{TP}{TP + FN} \times 100\% \quad (2.6)$$

Precision:
$$\frac{TP}{TP + FP} \times 100\% \quad (2.7)$$

2.6 Review Artikel

Pencarian Sebagai penguatan topik penelitian, penulis melakukan analisis dari beberapa riset penelitian sebelumnya yang di tampilkan pada tabel dibawah.

Tabel 2.2 Penelitian Terkait

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
1.	(Prasetyo et al., 2022). Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter	<i>Naive Bayes</i>	bagaimana menganalisis sentimen masyarakat pada kualitas layanan	Hasil penelitian menunjukkan bahwa pengujian menggunakan 17 data menghasilkan akurasi

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
	Dengan Menggunakan Metode <i>Naïve Bayes</i> . OKTAL : Jurnal Ilmu Komputer dan Sains Volume 1, No. 06,		PT PLN di provinsi Banten dari data tweet di Twitter.	dengan hasil akurasi lebih dari 80%.
2.	(Rahman & Sutanto, 2023b). Data Mining Untuk Memprediksi Tingkat Kepuasan Konsumen Gojek Menggunakan Algoritma <i>Naive Bayes</i> . Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi Volume 18 Nomor 01	<i>Naïve Bayes</i>	pentingnya mengevaluasi tingkat kepuasan konsumen terhadap layanan Gojek, sebuah aplikasi transportasi yang sangat populer karena kepraktisan dan kecepatannya.	Perhitungan akurasi penerapan metode <i>Naive Bayes</i> menghasilkan nilai akurasi sebesar 88,9% dalam mengklasifikasikan kepuasan pelanggan..
3.	(Amanda et al., 2022). Analisis Kepuasan Konsumen terhadap Pelayanan Store Ms Glow Menggunakan Metode <i>Naïve Bayes</i> . Hello World Jurnal Ilmu Komputer – Vol. 1 No. 3.	<i>Naïve Bayes</i>	pentingnya evaluasi kualitas pelayanan di Store Wanda MS Glow, sebuah usaha jasa di bidang perawatan kecantikan, yang semakin menghadapi persaingan ketat dengan banyaknya pesaing di	Hasil penelitian menunjukkan bahwa dari 100 responden, 12 pelanggan tidak puas dan 13 pelanggan yang puas, dengan analisis menggunakan metode klasifikasi <i>naive Bayes</i> memperoleh hasil akurasi sebesar 88.00%

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
			industri yang sama.	
4.	(Delaga Prasetya et al., 2023). Prediksi Tingkat Kepuasan Pelanggan Terhadap Kualitas Suatu Toko Menggunakan Metode <i>Naïve Bayes</i> . Jurnal METHODIKA, Vol. 9 No. 2 SEPTEMBER 2023 P-ISSN : 2442-7861 E-ISSN : 2614-3143	<i>Naïve Bayes</i>	bagaimana memprediksi tingkat kepuasan pelanggan terhadap kualitas sebuah toko, yang merupakan faktor penting dalam menjaga loyalitas dan menarik lebih banyak pelanggan.	Hasil penelitian ini menyajikan bahwa penggunaan algoritma <i>Naïve Bayes</i> pada studi kasus, memperoleh nilai akurasi yang cukup baik, yang dibuktikan dengan capaian akurasi prediksi sebesar 80%.
5.	(Fajar & Dwi Bhakti, 2024). Klasifikasi Penilaian Kinerja Karyawan Menggunakan Algoritme <i>Naïve Bayes</i> (Studi Kasus PT. As Sabar Sukses Berkah). Kohesi: Jurnal Multidisiplin Saintek Volume 4 No 2 Tahun 2024	<i>Naïve Bayes</i>	pentingnya pengelolaan kinerja karyawan dalam sebuah perusahaan, yang merupakan Kualitas sumber daya manusia memegang peranan vital bagi kesuksesan organisasi	Hasil penelitian mengindikasikan bahwa nilai akurasi klasifikasi diperoleh dari algoritma <i>Naïve Bayes</i> tinggi, yaitu mencapai 90%.
6.	(Wulan & Basri, 2024). Analisis Sentimen Terhadap Layanan Nasabah Bank Menggunakan	<i>Naïve Bayes</i>	pentingnya layanan terhadap pelanggan di sektor perbankan,	Hasil penelitian menunjukkan bahwa metode <i>Naïve Bayes</i> menghasilkan luaran yang efektif dengan akurasi 76.32%.

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
	Teknik Klasifikasi <i>Naïve Bayes</i> . Jurnal Kecerdasan Buatan dan Teknologi Informasi (2024) 3(2) 68-74		yang memiliki pengaruh besar pada kepuasan serta kesetiaan nasabah. Salah satu tantangan utama yang dihadapi oleh sektor perbankan adalah mengukur tingkat kepuasan nasabah secara tepat.	
7.	(Damanik et al., 2021) .Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma <i>Naïve Bayes</i> . Jurnal Sistem Informasi dan Teknologi (2021) 88-94	<i>Naïve Bayes</i>	tantangan yang dihadapi dalam memprediksi kepuasan mahasiswa pada pendidikan <i>online</i> selama COVID-19.	Hasil pengujian dengan data <i>training</i> berjumlah 80 serta data <i>testing</i> berjumlah 30 menunjukkan akurasi, dengan akurasi mencapai 100%, serta nilai <i>Precision</i> dan <i>Recall</i> masing-masing sebesar 100%.
8.	(Siska et al., 2023b) .Implementasi Metode <i>Naïve Bayes</i> Pada Prediksi Penyakit Seliak. KOPERTIP Jurnal Ilmiah Manajemen Informatika dan Komputer (2023) 7(1) 8-13	<i>Naïve Bayes</i>	Kesulitan dalam mendiagnosis penyakit seliak (Celiac Disease), yang memiliki gambaran klinis yang bervariasi dan profil pasien yang berbeda-	Penelitian ini menyajikan bahwa metode <i>Naïve Bayes</i> terbukti ampuh dalam memprediksi penyakit seliak, dengan akurasi tinggi. yaitu 93,03%.

No	Landasan Literatur	Metode Penelitian	Permasalahan	Hasil Penelitian
			beda, sehingga mempersulit diagnosis yang akurat.	
9.	(Cahyo, 2023) .Metode <i>Naïve Bayes</i> Untuk Klasifikasi Masa Studi Sarjana. Teknologipintar.org Volume 3 (4), 2023	<i>Naïve Bayes</i>	ketidakpastian mahasiswa mengenai kelulusan tepat waktu atau terlambat, yang dapat mempengaruhi akreditasi perguruan tinggi serta dampak lainnya bagi mahasiswa dan institusi.	Hasil penelitian <i>Naive Bayes</i> dapat diimplementasikan secara efektif untuk memprediksi kelulusan mahasiswa. Berdasarkan data alumni yang sudah lulus dengan akurasi 95% untuk k=5 (5-fold cross-validation) 94% untuk k=10 (10-fold cross-validation)
10.	(Sinaga et al., 2022) .Penerapan Algoritma <i>Naive Bayes</i> untuk Klasifikasi Prediksi Penerimaan Siswa Baru. Journal of Machine Learning and Data Analytics (MALDA) Volume 1, No. 1, Januari 2022 Page: 55-64	<i>Naïve Bayes</i>	kesulitan memprediksi jumlah penerimaan siswa baru pada sekolah-sekolah SMK swasta, khususnya di SMK Swasta Anak Bangsa.	Proses prediksi menggunakan <i>Naive Bayes</i> terbukti lebih cepat dan cukup akurat dalam memprediksi siswa baru yang diterima di SMK Swasta Anak Bangsa, serta dapat menghasilkan nilai yang cukup tinggi dengan tingkat akurasi prediksi sebesar 76,67%