BAB II
TINJAUAN LITERATUR

2.1 Landasan Teori

1. Minat Baca

   Minat baca adalah ketertarikan seseorang terhadap kegiatan membaca yang dilakukan secara sadar dan berkelanjutan. Minat baca berkaitan dengan kemauan, ketertarikan, kebiasaan, dan motivasi seseorang dalam melakukan aktivitas membaca. Minat baca yang baik dapat mendorong seseorang untuk lebih aktif memperoleh informasi, memperluas wawasan, serta mengembangkan pengetahuan. Rendahnya minat baca dapat dipengaruhi oleh berbagai faktor, baik dari dalam diri maupun dari lingkungan sekitar. Hasil kajian mengenai rendahnya minat baca di Indonesia menunjukkan bahwa faktor internal meliputi kurangnya motivasi, kebiasaan membaca yang rendah, serta kurangnya tujuan membaca, sedangkan faktor eksternal meliputi keterbatasan sarana dan prasarana, waktu, teknologi, kondisi ekonomi, dan lingkungan (Anidi & Anlianna, 2024).

   Minat baca juga dipengaruhi oleh faktor internal dan eksternal yang saling berkaitan. Faktor internal antara lain kemampuan membaca, motivasi, kebiasaan membaca, dan kemauan untuk membaca. Sementara itu, faktor eksternal dapat berupa kesesuaian bahan bacaan, lingkungan keluarga, lingkungan sekolah, penggunaan telepon seluler, dan pengaruh teman sebaya. Dengan demikian, minat baca tidak hanya ditentukan oleh

kemampuan membaca, tetapi juga oleh kondisi yang mendukung seseorang dalam melakukan aktivitas membaca (Yuniar & Pamujo, 2024).

Dalam penelitian ini, minat baca digunakan sebagai variabel target yang diklasifikasikan ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi. Pengelompokan tersebut digunakan untuk memberikan gambaran tingkat minat baca responden berdasarkan indikator yang diperoleh dari kuesioner. Penggunaan kuesioner untuk mengukur karakteristik minat baca juga diterapkan dalam penelitian Hafiz et al (2026), yang menggunakan 20 butir pernyataan skala Likert untuk merepresentasikan kebiasaan, frekuensi, motivasi, dan preferensi membaca.

2. Literasi dan Peran Komunitas Literasi

Literasi merupakan kemampuan seseorang dalam memperoleh, memahami, menggunakan, dan memanfaatkan informasi dalam kehidupan sehari-hari. Literasi tidak hanya berkaitan dengan pendidikan formal, tetapi juga dapat berkembang melalui lingkungan sosial, kegiatan masyarakat, dan berbagai aktivitas yang mendukung pembelajaran sepanjang hayat. Dalam konteks minat baca, kegiatan literasi dapat menjadi sarana untuk membangun kebiasaan membaca serta meningkatkan keterlibatan masyarakat terhadap bahan bacaan dan sumber informasi.Komunitas literasi merupakan salah satu bentuk lingkungan sosial yang dapat mendukung pengembangan budaya membaca. Komunitas literasi dapat menyediakan ruang untuk membaca, berdiskusi, berbagi pengetahuan, memperoleh sumber belajar, dan melakukan kegiatan literasi lainnya. Penelitian Herwina & Nurul

(2022) menunjukkan bahwa kampung literasi dapat berfungsi sebagai sumber belajar dan sumber informasi masyarakat, sementara pengelola berperan sebagai motivator dan pembimbing yang mendorong peningkatan minat baca. Penelitian mengenai komunitas literasi juga menunjukkan bahwa faktor lingkungan, fasilitas, komunitas, serta media digital dapat berperan dalam pembentukan minat baca. Pada anggota komunitas literasi, lingkungan keluarga, teman, fasilitas, komunitas, media digital, serta perasaan senang dan manfaat yang diperoleh dari membaca menjadi faktor yang berkaitan dengan terbentuknya minat baca (Islam et al., 2024).

Dalam konteks masyarakat Kota Bengkulu, pembinaan komunitas Taman Bacaan Masyarakat melalui literasi digital juga menunjukkan adanya peningkatan minat baca dan keterampilan literasi digital masyarakat. Hasil tersebut menunjukkan bahwa kegiatan pembinaan dan pendampingan dapat menjadi bagian penting dalam mendukung pemanfaatan fasilitas literasi oleh masyarakat (Syahril & Astria, 2025).

3. Data Mining

Data mining merupakan proses untuk memperoleh pola, pengetahuan, atau informasi yang bermanfaat dari sekumpulan data melalui penerapan teknik analisis tertentu. Data mining digunakan dalam berbagai bidang untuk mengolah data menjadi informasi yang dapat mendukung analisis dan pengambilan keputusan. Salah satu kajian mengenai data mining menjelaskan bahwa data mining digunakan untuk mengekstraksi informasi dari data dan menghasilkan pengetahuan yang dapat dimanfaatkan sesuai

kebutuhan (Srirahayu & Pribadie, 2023). Data mining memiliki berbagai teknik yang dapat digunakan sesuai dengan tujuan analisis. Salah satu teknik yang banyak digunakan adalah klasifikasi, yaitu pengelompokan data berdasarkan karakteristik tertentu menggunakan algoritma pembelajaran mesin. Tinjauan mengenai algoritma data mining menunjukkan bahwa klasifikasi merupakan salah satu teknik utama dalam pengolahan data dan dapat diterapkan menggunakan berbagai algoritma, seperti Decision Tree, Support Vector Machine, Neural Network, K-Nearest Neighbor, dan algoritma klasifikasi lainnya (Ramadhani & Rosyid, 2022). Dalam penelitian ini, data mining digunakan untuk mengolah data hasil kuesioner mengenai aktivitas literasi masyarakat secara sistematis. Tahapan pengolahan tersebut diarahkan untuk memperoleh pola yang dapat digunakan dalam proses klasifikasi tingkat minat baca masyarakat. Dengan demikian, data mining berfungsi sebagai pendekatan analisis yang mendukung pemanfaatan data kuesioner menjadi informasi yang lebih terstuktur.

4. Klasifikasi

   Klasifikasi merupakan salah satu teknik dalam data mining dan machine learning yang digunakan untuk mengelompokkan suatu data ke dalam kelas atau kategori tertentu berdasarkan karakteristik yang dimilikinya. Proses klasifikasi menggunakan data yang telah memiliki label atau kelas sebagai dasar pembentukan model, kemudian model tersebut digunakan untuk memprediksi kelas data baru. Klasifikasi dapat digunakan pada berbagai

jenis permasalahan dan dapat diterapkan dengan berbagai algoritma machine learning (Ramadhani & Rosyid, 2022).

Dalam penerapannya, klasifikasi umumnya melibatkan data training untuk membangun model dan data testing untuk mengetahui kemampuan model dalam melakukan prediksi terhadap data yang belum digunakan selama proses pelatihan. Penelitian klasifikasi pada berbagai bidang menunjukkan bahwa performa model dapat dievaluasi dengan membandingkan hasil prediksi terhadap kelas aktual. Penggunaan confusion matrix dan metrik evaluasi seperti accuracy, precision, recall, dan F1-score merupakan salah satu pendekatan yang umum digunakan untuk menilai hasil klasifikasi.

Berdasarkan jumlah kelas, klasifikasi dapat dibedakan menjadi klasifikasi biner dan klasifikasi multikelas. Klasifikasi biner memiliki dua kelas, sedangkan klasifikasi multikelas memiliki lebih dari dua kelas. Penelitian ini termasuk klasifikasi multikelas karena tingkat minat baca dibagi menjadi tiga kelas, yaitu rendah, sedang, dan tinggi**.**

5. Algoritma *Naïve Bayes*

   a. Pengertian *Naïve Bayes*

      Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes untuk menentukan kelas suatu data berdasarkan karakteristik yang dimilikinya. Algoritma ini banyak digunakan karena proses perhitungannya relatif sederhana dan dapat diterapkan pada berbagai permasalahan klasifikasi. Penerapan *Naïve Bayes* pada data kuesioner telah dilakukan oleh

Erinsyah et al (2024), yang menggunakan kuesioner berskala Likert dan *Naïve Bayes* untuk membantu proses evaluasi dan klasifikasi data.

b. Teorema Bayes

Dasar perhitungan algoritma *Naïve Bayes* menggunakan Teorema Bayes untuk memperoleh probabilitas suatu kelas berdasarkan data yang diamati. Secara umum, persamaan Teorema Bayes dapat dituliskan sebagai berikut:

$$P(C|X) = [P(X|C) \times P(C)] / P(X)$$

Keterangan:

P(C|X) = probabilitas suatu kelas C berdasarkan data X (*posterior*)

P(X|C) = probabilitas data X pada kelas C (*likelihood*)

P(C) = probabilitas awal kelas C (*prior*)

P(X) = probabilitas data X (*evidence*)

Dalam proses klasifikasi, model menghitung probabilitas untuk setiap kelas kemudian memilih kelas dengan nilai probabilitas posterior tertinggi sebagai hasil prediksi.

c. Asumsi *Naïve Bayes*

Nama “Naïve” berasal dari asumsi bahwa setiap fitur dianggap memiliki independensi bersyarat terhadap fitur lainnya ketika kelas telah diketahui. Meskipun asumsi tersebut bersifat sederhana, *Naïve Bayes* tetap banyak diterapkan dalam berbagai masalah klasifikasi karena proses perhitungannya efisien dan dapat menghasilkan

probabilitas untuk setiap kelas. Penerapannya pada berbagai masalah klasifikasi menunjukkan bahwa *Naïve Bayes* dapat digunakan sebagai model yang relatif sederhana dengan proses pelatihan yang efisien (Rieuwpassa et al., 2024).

d. Proses Klasifikasi *Naïve Bayes*

Secara umum, proses klasifikasi *Naïve Bayes* dilakukan melalui beberapa langkah, yaitu menghitung probabilitas prior setiap kelas, menghitung probabilitas likelihood dari setiap fitur terhadap masing-masing kelas, menghitung probabilitas posterior, kemudian menentukan kelas dengan probabilitas terbesar sebagai hasil klasifikasi. Penerapan *Naïve Bayes* pada data kuesioner minat baca juga menunjukkan penggunaan algoritma tersebut untuk menentukan kategori tingkat minat baca (Utari & Ulfah, 2021).

6. Evaluasi Kinerja Model

Evaluasi kinerja model merupakan tahap yang digunakan untuk mengetahui kemampuan model dalam menghasilkan prediksi yang sesuai dengan kelas aktual. Pada penelitian ini, evaluasi model dilakukan menggunakan confusion matrix, accuracy, precision, recall, F1-score, dan 10-fold cross validation. Penggunaan beberapa metrik tersebut diperlukan agar kinerja model dapat dilihat tidak hanya dari ketepatan secara keseluruhan, tetapi juga dari kemampuan model dalam mengenali masing-masing kelas. Penelitian klasifikasi pada berbagai bidang menunjukkan penggunaan

confusion matrix beserta accuracy, precision, recall, dan F1-score sebagai ukuran performa model.

a) Confusion Matrix

Confusion matrix merupakan tabel yang digunakan untuk membandingkan kelas aktual dengan kelas hasil prediksi model. Confusion matrix menunjukkan jumlah prediksi yang benar dan salah pada setiap kelas. Pada klasifikasi biner, komponen confusion matrix terdiri dari True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Pada klasifikasi multikelas, komponen tersebut dapat dianalisis untuk setiap kelas menggunakan pendekatan *one-versus-rest*.Confusion matrix dapat digunakan sebagai dasar untuk menghitung accuracy, precision, recall, dan F1-score. Selain memberikan informasi mengenai jumlah prediksi yang benar, confusion matrix juga dapat digunakan untuk mengetahui kelas yang paling sering mengalami kesalahan klasifikasi.

b) Accuracy

Accuracy merupakan ukuran yang menunjukkan proporsi seluruh data yang berhasil diprediksi dengan benar dibandingkan dengan seluruh data yang diuji.Rumus accuracy pada klasifikasi multiclass dapat dituliskan sebagai:

$$\text{Accuracy} = \text{Jumlah prediksi benar} / \text{Jumlah seluruh data}$$

atau:

$$\text{Accuracy} = \Sigma \text{ nilai diagonal confusion matrix} / N$$

Keterangan:

Σ nilai diagonal confusion matrix = jumlah prediksi yang benar

N = jumlah seluruh data testing

Pada penelitian ini, accuracy digunakan untuk mengetahui tingkat ketepatan keseluruhan model Gaussian *Naïve Bayes* dalam mengklasifikasikan tingkat minat baca masyarakat ke dalam kategori rendah, sedang, dan tinggi.

c) Precision

Precision merupakan ukuran yang menunjukkan tingkat ketepatan prediksi suatu kelas. Precision menjawab pertanyaan mengenai seberapa banyak data yang diprediksi sebagai suatu kelas benar-benar berasal dari kelas tersebut.

Rumus precision:

$$\text{Precision} = TP / (TP + FP)$$

Keterangan:

TP = jumlah data yang diprediksi benar sebagai kelas tertentu

FP = jumlah data yang diprediksi sebagai kelas tertentu tetapi sebenarnya termasuk kelas lain

Pada klasifikasi multiclass, precision dapat dihitung untuk masing-masing kelas melalui pendekatan *one-versus-rest*. Nilai precision yang tinggi menunjukkan bahwa prediksi positif pada kelas tertentu memiliki tingkat ketepatan yang tinggi.

d) Recall

Recall merupakan metrik evaluasi yang digunakan untuk mengetahui kemampuan model dalam menemukan atau mengenali seluruh data yang sebenarnya termasuk dalam suatu kelas. Recall juga disebut sebagai sensitivity.

Rumus recall adalah:

$$Recall = TP / (TP + FN)$$

Keterangan:

TP = jumlah data positif yang diprediksi benar

FN = jumlah data yang sebenarnya positif tetapi diprediksi sebagai kelas lain

Nilai recall yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar data yang sebenarnya termasuk dalam kelas tertentu. Dalam penelitian ini, recall digunakan untuk mengetahui kemampuan algoritma Gaussian *Naïve Bayes* dalam mengenali responden yang sebenarnya termasuk kategori rendah, sedang, maupun tinggi.

e) F1-Score

F1-score merupakan metrik evaluasi yang digunakan untuk mengetahui keseimbangan antara precision dan recall. F1-score memiliki nilai yang tinggi apabila model memiliki precision dan recall yang sama-sama baik. Metrik ini berguna terutama ketika distribusi kelas dalam dataset tidak seimbang.

Rumus F1-score adalah:

F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

Keterangan:

Precision = ketepatan prediksi positif

Recall = kemampuan model menemukan data yang sebenarnya positif

Dalam penelitian ini, F1-score digunakan sebagai salah satu ukuran untuk mengevaluasi keseimbangan kinerja model dalam mengklasifikasikan tiga kategori tingkat minat baca. Penggunaan F1-score juga penting karena distribusi data penelitian dapat berbeda antar kategori.

f) Cross Validation

Cross validation merupakan teknik evaluasi model yang digunakan untuk mengetahui kemampuan generalisasi dan kestabilan model terhadap data yang digunakan. Teknik ini dilakukan dengan membagi dataset menjadi beberapa bagian atau *fold*. Model kemudian dilatih dan diuji secara bergantian menggunakan bagian data yang berbeda.Pada penelitian ini digunakan 10-fold cross validation. Dataset dibagi menjadi sepuluh bagian dengan ukuran yang relatif sama. Pada setiap iterasi, sembilan bagian digunakan sebagai data training dan satu bagian digunakan sebagai data validasi. Proses tersebut dilakukan sebanyak sepuluh kali sehingga setiap bagian dataset memperoleh kesempatan menjadi data validasi.

Hasil evaluasi dari setiap fold kemudian dirata-ratakan untuk memperoleh nilai kinerja model secara keseluruhan.

Rumus rata-rata hasil cross validation adalah:

CV Score = (Score1 + Score2 + Score3 + ... + Score10) / 10

Keterangan:

Score1 sampai Score10 = nilai evaluasi pada masing-masing fold.

Penggunaan 10-fold cross validation pada penelitian ini bertujuan untuk mengetahui apakah performa algoritma *Naïve Bayes* tetap konsisten ketika model diuji pada pembagian data yang berbeda. Semakin konsisten hasil evaluasi pada setiap fold, semakin baik kemampuan generalisasi model terhadap data yang belum pernah digunakan sebelumnya.

2.2 Penelitian Terkait

Tabel 1.2 Penelitian Terdahulu yang Relevan

Peneliti& Tahun	Judul Penelitian	Metode/ Algoritma	Objek Penelitian	Hasil Utama	Relevansi
Utari & Ulfah (2021)	Penerapan *Naïve Bayes* untuk Prediksi Minat Baca	Data Mining *Naïve Bayes*	Pengunjung Perpustakaan GOR Pajajaran	*Naïve Bayes* digunakan untuk memprediksi minat baca berdasarkan usia dan menghasilkan accuracy	Menjadi dasar penerapan *Naïve Bayes* pada permasalahan minat baca

Peneliti& Tahun	Judul Penelitian	Metode/ Algoritma	Objek Penelitian	Hasil Utama	Relevansi
				80% pada pengujian yang dilaporkan	
Hafiz et al (2026)	Klasifikasi Minat Baca Mengguna kan *Naïve Bayes* Classifier	*Naïve Bayes* Classifier	911 responden siswa MTsN 1 Payakumb uh, 20 butir kuesioner Likert	Tiga kelas minat baca; accuracy 96,74% pada 90:10, 97,81% pada 80:20, dan 98,18% pada 70:30	Penelitian yang paling dekat karena sama-sama menggunakan Naïve Bayes, kuesioner Likert, dan tiga kelas minat baca
Dewi et al (2025)	Classificat ion of Book Borrower Patterns Based on Profession Using the *Naïve Bayes* Algorithm	Gaussian *Naïve Bayes* dan Multinomi al Naïve Bayes	4.476 transaksi peminjam an buku Dinas Perpustak aan Sidoarjo	Gaussian *Naïve Bayes* menghasilk an accuracy tertinggi sebesar 97% pada skenario dataset random	Mendukung penerapan Gaussian *Naïve Bayes* pada data perpustakaan dan variabel yang berkaitan dengan minat baca.
Murlena & Syahindra (2024)	Applicatio n of the *Naïve*	Naïve Bayes	Data kunjungan dan	*Naïve Bayes* digunakan	Sangat relevan karena sama-

Peneliti& Tahun	Judul Penelitian	Metode/ Algoritma	Objek Penelitian	Hasil Utama	Relevansi
	Bayes Algorithm in Classifying the Reading Interests of Regional Library Visitors		peminjaman buku perpustakaan daerah	untuk mengklasifikasikan pola minat baca pengunjung perpustakaan	sama melakukan klasifikasi minat baca menggunakan Naïve Bayes, tetapi objeknya pengunjung perpustakaan
Utari & Ulfah (2021)	Implementasi Data Mining untuk Mengetahui Minat Baca Peserta Didik Menggunakan Naives Bayes pada Perpustakaan SMP Negeri 2	Data Mining, Naïve Bayes	Data peminjaman buku, jenis buku, dan frekuensi kunjungan perpustakaan	*Naïve Bayes* memperoleh accuracy sebesar 80% dalam mengklasifikasikan minat baca peserta didik	Relevan karena sama-sama mengklasifikasikan minat baca, tetapi objek penelitian berada pada lingkungan pendidikan formal.

Peneliti& Tahun	Judul Penelitian	Metode/ Algoritma	Objek Penelitian	Hasil Utama	Relevansi
	Palembang				

Berdasarkan penelitian terdahulu, dapat diketahui bahwa algoritma *Naïve Bayes* telah banyak diterapkan pada permasalahan klasifikasi yang berkaitan dengan minat baca maupun data berbasis kuesioner. Utari & Ulfah (2021) menerapkan *Naïve Bayes* untuk memprediksi minat baca pengunjung perpustakaan berdasarkan karakteristik usia dan data peminjaman buku. Murlena & Syahindra (2024) juga menerapkan *Naïve Bayes* untuk mengklasifikasikan minat baca pengunjung perpustakaan menggunakan data kunjungan dan peminjaman buku. Sementara itu, Utari & Ulfah (2021) menggunakan *Naïve Bayes* untuk mengetahui minat baca peserta didik pada perpustakaan sekolah dan memperoleh accuracy sebesar 80%.

Penelitian Erinsyah et al (2024) menunjukkan bahwa *Naïve Bayes* juga dapat diterapkan pada data kuesioner berbasis skala Likert, sehingga mendukung penggunaan data kuesioner sebagai sumber data klasifikasi dalam penelitian ini. Penelitian Hafiz et al (2026) memiliki tingkat kesamaan paling tinggi dengan penelitian ini karena menggunakan 20 butir kuesioner skala Likert untuk mengklasifikasikan minat baca ke dalam tiga kategori, yaitu tinggi, sedang, dan rendah. Penelitian tersebut menggunakan 911 responden dan menghasilkan accuracy terbaik sebesar 98,18% pada pembagian data 70:30.

Selain *Naïve Bayes* secara umum, penelitian Dewi et al (2025) membandingkan Gaussian *Naïve Bayes* dan Multinomial *Naïve Bayes* dalam mengklasifikasikan

pola peminjaman buku. Hasil penelitian menunjukkan bahwa Gaussian *Naïve Bayes* memperoleh accuracy tertinggi sebesar 97% pada skenario dataset random. Penelitian Nugroho et al. (2023) juga menunjukkan penggunaan Gaussian *Naïve Bayes* bersama K-Fold Cross Validation, termasuk 10-fold cross validation, dalam mengevaluasi performa model.

Meskipun penelitian terdahulu telah menunjukkan bahwa *Naïve Bayes* dapat digunakan untuk mengklasifikasikan minat baca, objek dan konteks penelitian masih berbeda dengan penelitian ini. Penelitian Utari & Ulfah (2021), Murlena dan Syahindra (2024), serta Nadira dan Sutabri (2024) berfokus pada perpustakaan atau lingkungan pendidikan. Sementara itu, Hafiz et al (2026) berfokus pada siswa MTsN 1 Payakumbuh.

Penelitian ini memiliki perbedaan dengan penelitian sebelumnya karena berfokus pada masyarakat Bengkulu yang mengikuti Program Komunitas Literasi Aksaraya Semesta, bukan pada siswa maupun pengguna perpustakaan. Selain itu, penelitian ini menggunakan 10 fitur aktivitas literasi sebagai variabel input, mengklasifikasikan tingkat minat baca ke dalam tiga kategori, serta menerapkan *Naïve Bayes* dengan evaluasi menggunakan confusion matrix, accuracy, precision, recall, F1-score, dan 10-fold cross validation. Dengan demikian, penelitian ini berupaya memperluas penerapan algoritma *Naïve Bayes* dari konteks perpustakaan dan pendidikan formal ke konteks masyarakat pada program literasi berbasis komunitas nonformal, khususnya di Bengkulu.