

CLASSIFICATION OF READING INTEREST USING THE NAÏVE BAYES CLASSIFIER METHOD

KLASIFIKASI MINAT BACA MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Muhammad Hafiz¹, Elvia Budianita², Alwis Nazir³, Siska Kurnia Gusti⁴

Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri
Sultan Syarif Kasim Riau^{1,2,3,4}

2050116357@students.uin-suska.ac.id¹, elvia.budianita@uin-suska.ac.id², alwis.nazir@uin-suska.ac.id³, siskakurniagusti@uin-suska.ac.id⁴

ABSTRACT

Reading interest is an important factor in supporting students' academic achievement; however, its assessment is often subjective. This study aims to classify the reading interest level of students at MTsN 1 Payakumbuh using the Naïve Bayes Classifier method. The input attributes consist of 20 Likert-scale questionnaire items representing students' reading habits, frequency, motivation, and reading preferences. The dataset comprises 911 student responses categorized into three classes: high (329 records), moderate (501 records), and low (81 records). Model evaluation was conducted using three training-testing data split scenarios, namely 90:10, 80:20, and 70:30. Performance evaluation using a confusion matrix shows that the 90:10 split achieved an accuracy of 96.74%, the 80:20 split achieved 97.81%, and the 70:30 split achieved 98.18%. These results indicate that the Naïve Bayes Classifier demonstrates high and consistent performance in classifying students' reading interest based on questionnaire data.

Keywords: Classification, Naïve Bayes Classifier, Reading Interest, Likert Scale.

ABSTRAK

Minat baca merupakan faktor penting dalam mendukung keberhasilan akademik siswa, namun pengukurannya masih sering dilakukan secara subjektif. Penelitian ini bertujuan untuk mengklasifikasikan tingkat minat baca siswa MTsN 1 Payakumbuh menggunakan metode Naïve Bayes Classifier. Atribut yang digunakan sebagai input berupa 20 butir pernyataan kuesioner skala Likert yang merepresentasikan kebiasaan, frekuensi, motivasi, serta preferensi membaca siswa. Data penelitian diperoleh dari 911 responden yang dikelompokkan ke dalam tiga kelas tingkat minat baca, yaitu tinggi (329 data), sedang (501 data), dan rendah (81 data). Pengujian model dilakukan menggunakan tiga skema pembagian data latih dan data uji, yaitu 90:10, 80:20, dan 70:30. Evaluasi performa model menggunakan confusion matrix menunjukkan bahwa skema 90:10 menghasilkan akurasi sebesar 96,74%, skema 80:20 sebesar 97,81%, dan skema 70:30 sebesar 98,18%. Hasil tersebut menunjukkan bahwa metode Naïve Bayes Classifier memiliki performa yang sangat baik dan konsisten dalam mengklasifikasikan tingkat minat baca siswa berdasarkan data kuesioner.

Kata kunci: Klasifikasi, Minat baca, Naïve Bayes Classifier, Skala Likert.

PENDAHULUAN

Membaca bagi sebagian masyarakat di Indonesia masih dianggap kurang menarik dan bahkan sering diabaikan, padahal aktivitas ini memiliki banyak manfaat seperti meningkatkan pengetahuan, memperkaya kosakata, dan melatih kemampuan berpikir kritis [1]. Membaca berperan penting dalam kehidupan manusia, karena melalui membaca seseorang dapat memperoleh

pengetahuan, informasi, dan pengalaman. Aktivitas membaca juga merupakan proses yang bersifat kritis dan kreatif, di mana pembaca berusaha memahami makna teks yang dibaca. Dengan demikian, membaca menjadi salah satu sarana utama untuk menambah wawasan serta memperluas pengetahuan [2].

Minat baca merupakan aspek penting dalam pembentukan budaya literasi di masyarakat. Minat adalah

suatu dorongan atau keinginan kuat dari dalam diri individu untuk melakukan aktivitas secara sukarela tanpa adanya tekanan, serta dapat dilakukan dalam berbagai kesempatan. Melaksanakan kegiatan berdasarkan minat tidaklah mudah, karena diperlukan usaha dan pengorbanan tertentu agar minat tersebut dapat tumbuh terhadap suatu aktivitas [3]. Menurut (Dandi et al. 2022), minat baca adalah kecenderungan hati atau perhatian yang besar terhadap aktivitas membaca sehingga seseorang terdorong untuk melakukannya dengan sukarela. Minat baca tidak hanya sebatas kemampuan teknis dalam memahami huruf dan kata, tetapi juga berkaitan dengan motivasi intrinsik yang mendorong seseorang untuk memperoleh informasi dan pengetahuan melalui kegiatan membaca [4]. Minat baca lahir dari keyakinan bahwa membaca memberikan manfaat langsung bagi perkembangan diri, baik secara akademik maupun nonakademik. Seseorang yang memiliki minat baca tinggi umumnya lebih percaya bahwa kegiatan membaca dapat memperluas wawasan, meningkatkan keterampilan berpikir kritis, serta membantu dalam memecahkan masalah kehidupan sehari-hari [5].

Di Indonesia, rendahnya minat baca telah menjadi isu serius yang berdampak pada kualitas sumber daya manusia. Rendahnya minat baca siswa banyak dipengaruhi oleh kurangnya kebiasaan membaca sejak dini, keterbatasan akses terhadap buku, serta rendahnya peran orang tua dan sekolah dalam menumbuhkan budaya literasi [6]. Berdasarkan data, indeks minat baca nasional hanya mencapai 0,01, jauh tertinggal dibandingkan negara maju yang rata-rata memiliki indeks antara 0,45 hingga 0,62 [7]. Laporan Programme for International Student Assessment (PISA) tahun 2018

menunjukkan bahwa Indonesia berada pada peringkat ke-72 dari 78 negara dalam literasi membaca [8]. Sedangkan hasil PISA 2022 memperlihatkan bahwa peringkat Indonesia pada bidang membaca mengalami peningkatan, yaitu naik lima posisi dibandingkan tahun 2018 [9]. Namun demikian, skor absolut membaca yang diperoleh Indonesia, yaitu sekitar 358 poin, masih berada di bawah rata-rata OECD [10].

Berdasarkan hasil wawancara dengan Ibu Eli Elvita, selaku Wakil Kepala Kurikulum MTsN 1 Payakumbuh, diketahui bahwa minat baca siswa di kelas yang beliau ampu saat ini cukup tinggi, terutama dalam hal literasi digital. Banyak siswa yang justru lebih memahami materi dari sumber digital dibandingkan dari guru. Namun demikian, minat membaca terhadap buku tercetak masih tergolong rendah, terlihat dari jaranganya siswa berkunjung ke perpustakaan.

Menurut beliau, faktor utama yang memengaruhi tinggi atau rendahnya minat baca siswa sangat bergantung pada kebutuhan. Semakin besar kebutuhan siswa terhadap suatu informasi, maka semakin tinggi pula minat mereka untuk membaca. Dalam upaya menumbuhkan minat baca, berbagai langkah telah dilakukan, di antaranya memperkaya bahan bacaan, memperbolehkan siswa membawa telepon seluler sesuai kebutuhan, serta menyediakan kelas digital dan pojok baca digital. Beliau juga menilai peran perpustakaan sekolah penting dalam meningkatkan minat baca, terutama dengan memanfaatkan teknologi, seperti penggunaan e-book. Lebih lanjut, Ibu Eli menekankan bahwa minat baca memiliki peran yang sangat penting dalam pencapaian akademik siswa. Dengan semakin banyak membaca, wawasan siswa akan semakin luas, tidak hanya terbatas pada materi yang

diberikan guru. Adapun tantangan utama dalam mendorong budaya membaca di sekolah adalah kembali pada minat masing-masing siswa. Selain itu, pengaruh dunia digital, khususnya permainan daring, juga menjadi hambatan dalam meningkatkan minat baca siswa.

Hasil wawancara pada siswa menunjukkan bahwa minat baca siswa bervariasi; sebagian menyukai bacaan hiburan, sebagian lainnya lebih tertarik pada artikel atau jurnal ilmiah, dan ada yang hanya membaca saat menjelang ujian. Frekuensi membaca juga beragam, mulai dari jarang hingga setiap hari. Secara umum, siswa menganggap membaca penting untuk menambah wawasan, namun minat mereka dipengaruhi oleh rasa ingin tahu, jenis bacaan, dan media yang digunakan. Sebagian besar lebih memilih bacaan digital karena praktis, sementara sebagian lainnya menyukai buku cetak karena dianggap lebih bermakna.

Klasifikasi merupakan proses mengelompokkan data ke dalam beberapa kategori atau kelas berdasarkan atribut atau karakteristik tertentu dengan tujuan menemukan pola atau hubungan dalam data yang dapat digunakan untuk mendukung pengambilan keputusan atau melakukan prediksi terhadap data baru [11]. Naïve Bayes Classifier merupakan salah satu algoritma yang terdapat dalam teknik klasifikasi. Naïve Bayes Classifier adalah metode klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi bahwa setiap atribut saling independen. Metode ini sederhana namun mampu menghasilkan klasifikasi yang akurat, bahkan pada data berukuran besar [12]. Dengan kata lain, algoritma ini mengasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada

hubungannya dengan ciri dari kelas lainnya [13].

Permasalahan utama dalam penelitian ini adalah minimnya pendekatan berbasis teknologi yang digunakan untuk menganalisis dan mengidentifikasi tingkat minat baca siswa secara akurat. Selama ini, penilaian terhadap minat baca umumnya masih dilakukan secara manual melalui observasi dan interpretasi subjektif dari hasil kuesioner, sehingga berpotensi menimbulkan ketidaktepatan dalam pengelompokan tingkat minat baca siswa. Kondisi tersebut menunjukkan perlunya penerapan metode analisis berbasis data yang mampu memberikan hasil klasifikasi yang lebih objektif dan terukur.

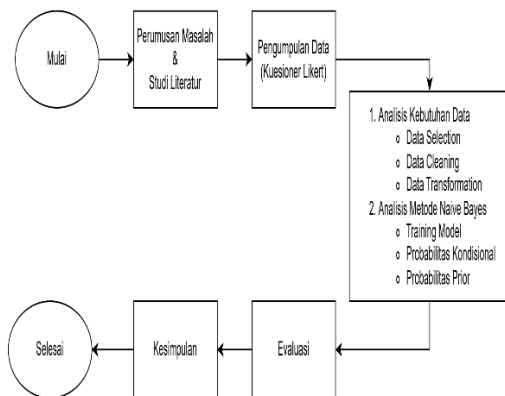
Meskipun beberapa penelitian sebelumnya telah menerapkan algoritma klasifikasi dalam menganalisis perilaku belajar atau minat siswa, namun penerapannya menggunakan data historis perpustakaan, bukan data kuesioner yang mencerminkan aspek psikologis serta perilaku membaca siswa [14]. Selain itu, validasi empiris terhadap hasil klasifikasi pada penelitian terdahulu juga belum optimal, sehingga akurasi dan keandalan model masih perlu ditingkatkan.

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk mengetahui tingkat minat baca siswa MTsN 1 Payakumbuh dengan menggunakan metode Naïve Bayes Classifier. Metode ini diterapkan untuk mengklasifikasikan siswa ke dalam kategori tingkat minat baca—tinggi, sedang, dan rendah—berdasarkan data kuesioner yang diperoleh. Hasil penelitian diharapkan dapat memberikan gambaran pola minat baca siswa serta menunjukkan potensi penerapan algoritma klasifikasi dalam

mendukung analisis pendidikan berbasis data.

METODE PENELITIAN

Metodologi penelitian adalah cara atau langkah sistematis yang digunakan untuk mengumpulkan, menganalisis, dan mengolah data guna mencapai tujuan penelitian. Adapun metodologi dalam penelitian ini adalah sebagai berikut :



Gambar 1. Metode Penelitian

Perumusan masalah

Bertujuan untuk merumuskan pokok permasalahan yang akan diteliti secara jelas, terarah, dan sistematis. Dalam hal ini yaitu bagaimana metode Naïve Bayes Classifier dapat mengidentifikasi Tingkat Minat Baca Siswa. Sebelum perumusan masalah dilakukan, peneliti terlebih dahulu melakukan studi literatur dengan menelaah penelitian-penelitian terdahulu yang berkaitan dengan minat baca dan penerapan metode klasifikasi dalam bidang pendidikan. Hasil studi literatur tersebut menjadi dasar dalam merumuskan permasalahan penelitian serta menentukan metode Naïve Bayes Classifier sebagai pendekatan yang digunakan.

Pengumpulan data

Pada tahap ini, data yang dikumpulkan bertujuan untuk mengukur tingkat minat baca siswa. Instrumen

penelitian yang digunakan berupa kuesioner dengan 20 item pernyataan berbasis skala Likert 1–5, yang telah melalui proses validasi oleh Dr. H. Kusnedi, M.Pd.. Kuesioner ini diberikan kepada 911 siswa sebagai responden penelitian. Skor yang diperoleh kemudian dianalisis secara kuantitatif dengan mengelompokkan hasil ke dalam tiga kategori tingkat minat baca, yaitu tinggi, sedang, dan rendah, sesuai dengan interval penilaian yang telah ditetapkan.

Analisis Kebutuhan Data

Analisa kebutuhan data dilakukan untuk menganalisa dataset yang dibutuhkan. Dataset yang telah dikumpulkan berasal dari responden yaitu siswa-siswi MTsN 1 Payakumbuh.

1. Data Selection

Pada tahap ini dilakukan proses seleksi data guna memastikan bahwa data yang digunakan sesuai dengan kebutuhan penelitian. Dataset yang diperoleh berasal dari data minat baca siswa yang dikumpulkan melalui kuesioner yang telah divalidasi oleh Dr. H. Kusnedi, M.Pd. Kuesioner tersebut memuat beberapa atribut, yaitu atribut identitas responden yang terdiri dari nama, kelas, dan jenis kelamin, serta 20 butir pernyataan kuesioner yang merepresentasikan aspek minat baca siswa. Dalam proses seleksi data, atribut identitas responden seperti nama, kelas, dan jenis kelamin tidak digunakan dalam analisis karena tidak berkaitan langsung dengan tujuan klasifikasi tingkat minat baca. Oleh karena itu, dataset yang digunakan dalam penelitian ini hanya mencakup 20 atribut berupa butir pernyataan kuesioner yang selanjutnya digunakan sebagai fitur input dalam pemodelan menggunakan metode Naïve Bayes Classifier.

2. Data Cleaning

Tahap merupakan proses untuk mengatasi missing value, data duplikat, data yang tidak konsisten, dan outlier yang ditemukan.

3. Data Transformation

Data transformation dibutuhkan untuk menyesuaikan data ke dalam bentuk yang dapat diolah oleh algoritma Naïve Bayes Classifier. Transformasi dilakukan dengan mengubah setiap pilihan jawaban pada skala Likert menjadi nilai numerik sehingga mempermudah proses perhitungan probabilitas dan pemodelan klasifikasi. Tahap ini memastikan bahwa data siap digunakan pada proses data mining dan analisis lebih lanjut. Berikut tabel yang menunjukkan transformasi data :

Tabel 1. Transformasi Data

Kuesioner	Keterangan	Bobot Nilai
SS	Sangat Setuju	5
S	Setuju	4
N	Netral	3
TS	Tidak Setuju	2
STS	Sangat Tidak Setuju	1

Analisis Metode Naïve Bayes Classifier

Naïve Bayes Classifier adalah metode klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes untuk menentukan kemungkinan suatu data termasuk dalam kelas tertentu. Metode ini mengasumsikan bahwa setiap atribut bersifat independen satu sama lain. Keunggulan utama Naïve Bayes Classifier adalah kesederhanaannya serta kemampuannya memberikan hasil klasifikasi yang akurat, bahkan dengan jumlah data yang besar [12].

Algoritma ini beroperasi dengan menganalisis berbagai kombinasi serta frekuensi nilai yang terdapat dalam basis data. Pendekatan ini didasari oleh

Teorema Bayes, yang mengasumsikan bahwa setiap atribut dalam data bersifat independen dan tidak memiliki keterkaitan satu sama lain. Naïve Bayes dikembangkan berdasarkan teori yang diperkenalkan oleh ilmuwan asal Inggris, Thomas Bayes, yang memungkinkan perhitungan probabilitas suatu peristiwa berdasarkan data atau pengalaman sebelumnya [15]. Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)}$$

Keterangan :

X = Kelas data yang belum diketahui

H = Hipotesa data X adalah kelas spesifik

P(H|X) = Kemungkinan Hipotesa H berdasarkan keadaan X (posteriori prob.)

P(H) = Kemungkinan Hipotesa H (prior prob.)

P(X|H) = Kemungkinan X berdasarkan keadaan tersebut

P(X) = Kemungkinan dari X

Pemodelan Algoritma Naïve Bayes Classifier

1. Menghitung Probabilitas Prior

Probabilitas prior menunjukkan kemungkinan awal suatu kelas sebelum melihat fitur lainnya.

$$P(Y) = (\text{Jumlah data dalam kelas } Y) / N$$

Keterangan :

P(Y) : Probabilitas prior dari kelas Y

N : Total jumlah data

2. Menghitung Probabilitas Kondisi

Menghitung probabilitas setiap fitur pada setiap kelas.

$$P(X_i|Y) = \frac{\text{Jumlah data dengan fitur } X \text{ dalam kelas } Y}{\text{Total jumlah data dalam kelas } Y}$$

3. Normalisasi (Laplace Smoothing)

Pada penelitian ini, proses normalisasi menggunakan metode *Laplace Smoothing* hanya diterapkan pada perhitungan manual untuk

kasus klasifikasi tingkat minat baca. Normalisasi tersebut digunakan guna menghindari probabilitas nol pada saat perhitungan manual *Naïve Bayes Classifier*. Namun, pada implementasi menggunakan *Python*, *Laplace Smoothing* tidak lagi diterapkan secara terpisah karena pustaka atau fungsi yang digunakan telah memiliki mekanisme penanganan data tersendiri. Dengan demikian, tahapan normalisasi hanya muncul pada perhitungan manual dan tidak digunakan dalam proses komputasi di *Python*.

$$P(x_i | Y) = \frac{f + 1}{n + k}$$

Keterangan :

f = jumlah kemunculan fitur x_i pada kelas Y

n = total jumlah data pada kelas Y

k = jumlah kategori/fitur unik

Tambahan “+1” digunakan untuk menghindari probabilitas nol

4. Menghitung Nilai Prediksi

Nilai prediksi kelas target dari data uji dihitung berdasarkan nilai probabilitas fitur yang telah dihitung sebelumnya.

$$P(Y|X) \propto P(Y) \times \prod_{i=1}^n P(X_i|Y)$$

Keterangan :

$P(C|X)$: Probabilitas kelas Y berdasarkan data X adalah nilai yang ingin diprediksi atau diketahui setelah memperoleh data X .

\propto : Proporsional

$\prod_{i=1}^n P(X_i|Y)$: Hasil perkalian dari probabilitas kondisional dari setiap atribut X_i terhadap kelas Y , untuk semua atribut dalam data X .

Skala Likert

Skala Likert merupakan skala psikometrik yang sering dipakai dalam kuesioner dan menjadi salah satu skala yang paling umum digunakan dalam

penelitian berbentuk survei [16]. Pada penelitian ini, penilaian minat baca dilakukan menggunakan 20 butir pernyataan dengan lima pilihan jawaban, yaitu Sangat Setuju (SS), Setuju (S), Netral (N), Tidak Setuju (TS), dan Sangat Tidak Setuju (STS). Setiap pilihan jawaban diberikan bobot nilai dengan rentang 1 sampai 5. Penentuan kelas tingkat minat baca dilakukan menggunakan metode interval. Metode ini dilakukan dengan menghitung panjang interval kelas berdasarkan selisih antara skor maksimum dan skor minimum, kemudian dibagi dengan jumlah kelas yang ditentukan [17]. Kuesioner terdiri dari 20 butir pernyataan dengan skala Likert 1–5, sehingga skor maksimum dan minimum dapat dihitung sebagai berikut:

Skor Maksimum = $20 \times 5 = 100$

Skor Minimum = $20 \times 1 = 20$

Jumlah kelas tingkat minat baca ditetapkan sebanyak 3 kelas, yaitu rendah, sedang, dan tinggi. Panjang interval kelas dihitung menggunakan rumus [17] :

Interval =

$$\frac{\text{Skor Maksimum} - \text{Skor Minimum}}{\text{Jumlah Kelas}}$$

(6)

$$\text{Interval} = \frac{100 - 20}{3} = \frac{80 - 20}{3} = 26,67$$

Berdasarkan perhitungan tersebut, tingkat minat baca diklasifikasikan ke dalam tiga kategori, yaitu tinggi, sedang, dan rendah. dapat dilihat pada tabel 2:

Tabel 2. Skala Likert

Skor	Keterangan
20 – 46	Minat Baca Rendah
47 – 73	Minat Baca Sedang
74 – 100	Minat Baca Tinggi

Selanjutnya, seluruh skor responden dikelompokkan ke dalam interval kelas yang telah ditentukan. Hasil pengelompokan menunjukkan bahwa dari total 911 siswa, terdapat 81 siswa yang termasuk dalam kategori minat baca rendah, 501 siswa berada pada kategori minat baca sedang, dan 329 siswa berada pada kategori minat baca tinggi.

Confusion Matrix

Pengujian dalam penelitian ini menggunakan Confusion Matrix sebagai metode untuk mengukur akurasi model, dengan cara membandingkan hasil prediksi klasifikasi dengan data aktual guna mengetahui tingkat ketepatan model dalam mengelompokkan data.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Keterangan :

TP (True Positive) : Jumlah data positif yang diprediksi benar

TN (True Negative) : Jumlah data negatif yang diprediksi benar

FP (False Positive) : Jumlah data negatif yang diprediksi sebagai positif

FN (False Negative) : Jumlah data positif yang diprediksi sebagai negatif

All Precision = Precision

$A+B+C/\text{Jumlah kelas}$ (8)

All Recall = Recall $A+B+C/\text{Jumlah Kelas}$

Kesimpulan

Kesimpulan diperoleh berdasarkan hasil perhitungan menggunakan algoritma *Naïve Bayes Classifier* yang menghasilkan nilai akurasi.

HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil penelitian yang diperoleh dari

penerapan algoritma *Naïve Bayes Classifier* terhadap data minat baca siswa. Analisis dilakukan untuk mengelompokkan responden ke dalam tiga kategori, yaitu minat baca tinggi, sedang, dan rendah berdasarkan skor kuesioner. Selain itu, pada bab ini juga dipaparkan hasil perhitungan akurasi model sebagai gambaran tingkat ketepatan algoritma *Naïve Bayes* dalam melakukan klasifikasi. Penyajian hasil meliputi proses pengolahan data, perhitungan probabilitas, hasil klasifikasi, serta evaluasi performa model yang kemudian dianalisis lebih lanjut pada subbagian pembahasan.

Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan di MTsN 1 Payakumbuh dengan melibatkan peserta didik kelas VII dan VIII. Jumlah responden keseluruhan adalah 911 siswa. Instrumen yang digunakan berupa kuesioner yang terdiri dari 20 butir pernyataan terkait tingkat minat baca. Setiap pernyataan diukur menggunakan skala Likert dengan lima kategori jawaban, yaitu Sangat Setuju, Setuju, Netral, Tidak Setuju, dan Sangat Tidak Setuju. Kuesioner ini disusun untuk mengukur kecenderungan minat baca siswa secara kuantitatif sehingga data yang diperoleh dapat dianalisis menggunakan metode *Naïve Bayes Classifier*. Berdasarkan hasil pengelompokan skor kuesioner, data yang telah di peroleh diklasifikasikan ke dalam tiga kelas tingkat minat baca, yaitu minat baca tinggi sebanyak 329 siswa, minat baca sedang sebanyak 501 siswa, dan minat baca rendah sebanyak 81 siswa. Adapun parameter pernyataan untuk 20 butir kuesioner ditampilkan pada tabel 3 sebagai berikut :

Tabel 3. Kuisisioner Minat Baca

No	Pertanyaan	SS	S	N	TS	STS
----	------------	----	---	---	----	-----

1.	Saya senang membaca buku di waktu luang.
2.	Saya memiliki koleksi buku pribadi di rumah.
3.	Saya membaca buku secara rutin setiap hari.
4.	Saya merasa membaca buku adalah kegiatan yang menyenangkan
5.	Saya merasa membaca buku dapat menambah wawasan saya.
6.	Saya membaca buku meskipun tidak ada tugas dari guru.
7.	Saya senang berada di perpustakaan untuk membaca buku.
8.	Saya suka membaca buku cerita, novel, atau komik.
9.	Saya suka membaca buku pelajaran untuk memperdalam materi.
10.	Saya membaca berita atau artikel di internet atau media cetak.
11.	Saya lebih suka membaca daripada bermain media sosial.
12.	Saya mengikuti kegiatan literasi atau membaca bersama di sekolah.
13.	Saya antusias jika ada lomba atau kegiatan berkaitan dengan membaca.
14.	Saya sering membeli buku dengan uang saku saya sendiri.
15.	Orang tua saya mendorong dan mendukung saya untuk rajin membaca.
16.	Saya senang jika diberi hadiah berupa buku.
17.	Saya mampu menyelesaikan satu buku dalam beberapa hari.
18.	Saya membaca buku karena saya ingin tahu isi ceritanya, bukan karena disuruh.
19.	Saya terbiasa membawa buku bacaan ke sekolah.
20.	Saya tertarik mengunjungi toko buku atau pameran buku.

Pre Processing Data

Tahap pre-processing data dilakukan untuk mempersiapkan data sebelum memasuki proses analisis menggunakan algoritma Naïve Bayes Classifier. Pada penelitian ini, tidak seluruh tahapan pre-processing diperlukan karena karakteristik data yang diperoleh relatif bersih dan terstruktur.

a. Data Selection

Pada tahap data *selection*, dilakukan pemilihan atribut yang relevan dengan kebutuhan analisis menggunakan algoritma *Naïve Bayes Classifier*. Data awal yang diperoleh dari kuesioner mencakup beberapa atribut identitas seperti nama, kelas, dan jenis kelamin. Atribut identitas tersebut tidak digunakan dalam proses klasifikasi minat baca sehingga dihilangkan agar analisis berfokus pada

variabel yang berkaitan langsung dengan tujuan penelitian. Oleh karena itu, data yang digunakan adalah 20 butir pernyataan skala Likert yang digunakan

sebagai fitur dalam proses pemodelan. Struktur data sebelum dan sesudah proses seleksi data dilakukan ditunjukkan pada Tabel 4 dan 5.

Tabel 4. Data sebelum seleksi

No.	Nama	Kelas	Jenis Kelamin	1. Saya senang membaca buku di waktu luang.	2. Saya memiliki koleksi buku pribadi di rumah.	3. Saya membaca buku secara rutin setiap hari.	...	20. Saya tertarik mengunjungi toko buku atau pameran buku.
1	Muhammad Dheandra Pratama	Delapan (VIII)	Laki-laki	Netral	Tidak Setuju	Tidak Setuju	...	Setuju
2	Rafki Assadiqi	Delapan (VIII)	Laki-laki	Netral	Netral	Netral	...	Sangat Setuju
3	Muhammad Farid	Tujuh (VII)	Laki-laki	Sangat Setuju	Sangat Setuju	Sangat Setuju	...	Sangat Setuju
...			
911	Khafka Hertyu Trifayza	Delapan (VIII)	Perempuan	Sangat Setuju	Setuju	Setuju	...	Setuju

Tabel 5. Data Setelah Seleksi

No.	1. Saya senang membaca buku di waktu luang.	2. Saya memiliki koleksi buku pribadi di rumah.	3. Saya membaca buku secara rutin setiap hari.	...	20. Saya tertarik mengunjungi toko buku atau pameran buku.
1	Netral	Tidak Setuju	Tidak Setuju	...	Setuju
2	Netral	Netral	Netral	...	Sangat Setuju
3	Sangat Setuju	Sangat Setuju	Sangat Setuju	...	Sangat Setuju
...
911	Sangat Setuju	Setuju	Setuju	...	Setuju

b. Data Cleaning

Pada tahap data cleaning, tidak ditemukan adanya data yang hilang (missing values), duplikasi, atau ketidaksesuaian format. Setiap respon tercatat dengan baik sesuai pilihan jawaban skala Likert.

c. Data Transformation

Tahap transformasi data menjadi bagian penting dalam penelitian ini.

Transformasi dilakukan untuk mengubah kategori jawaban skala Likert menjadi bentuk numerik yang dapat diproses oleh algoritma Naïve Bayes Classifier. Setiap jawaban dikonversi ke dalam skor angka sehingga mempermudah tahap perhitungan probabilitas dan pemodelan klasifikasi pada proses data mining. Tabel 6 dan 7 menunjukkan hasil dari sebelum dan setelah transformasi data :

Tabel 6. Data Sebelum di Transformasi

No.	1. Saya senang membaca buku di waktu luang.	2. Saya memiliki koleksi buku pribadi di rumah.	3. Saya membaca buku secara rutin setiap hari.	...	20. Saya tertarik mengunjungi toko buku atau pameran buku.	Kelas
1	Netral	Tidak Setuju	Tidak Setuju	...	Setuju	Sedang
2	Netral	Netral	Netral	...	Sangat Setuju	Sedang
3	Sangat Setuju	Sangat Setuju	Sangat Setuju	...	Sangat Setuju	Tinggi
...
911	Sangat Setuju	Setuju	Setuju	...	Setuju	Tinggi

Tabel 7. Data Setelah di Transformasi

No.	Pert. 1	Pert. 2	Pert. 3	...	Pert. 20	Kelas
1	3	2	2	...	4	Sedang
2	3	3	3	...	5	Sedang
3	5	5	5	...	5	Tinggi
...
911	5	4	4	...	4	Tinggi

Pemodelan Naïve Bayes Classifier**a. Nilai probabilitas kelas**

Perhitungan probabilitas kelas (*prior probability*) dilakukan untuk mengetahui proporsi masing-masing kategori minat baca sebelum mempertimbangkan nilai atribut. Dengan total 911 responden, terdapat 329 siswa pada kategori minat baca tinggi, 501 siswa pada kategori sedang, dan 81 siswa pada kategori rendah. Probabilitas kelas dihitung menggunakan rumus:

$$P(Y) = \frac{\text{Jumlah data dalam kelas } Y}{N}$$

Hasil perhitungan menunjukkan bahwa probabilitas kelas Tinggi adalah :

$$P(\text{Tinggi}) = \frac{329}{911} = 0,3611$$

Probabilitas kelas Sedang adalah :

$$P(\text{Sedang}) = \frac{501}{911} = 0,5499$$

Sedangkan probabilitas kelas Rendah adalah :

$$P(\text{Rendah}) = \frac{81}{911} = 0,0889$$

Nilai probabilitas ini menjadi dasar dalam proses klasifikasi Naïve Bayes, di mana setiap kategori minat baca diberi peluang awal yang mencerminkan distribusi data nyata. Probabilitas ini selanjutnya dikombinasikan dengan probabilitas bersyarat dari masing-masing atribut untuk menghitung nilai posterior pada tahap pemodelan.

b. Probabilitas kondisional

Setelah menghitung probabilitas kelas, langkah selanjutnya adalah menghitung probabilitas kondisional untuk setiap atribut. Pada penelitian ini terdapat 20 atribut berupa butir pernyataan skala Likert, masing-masing memiliki nilai 1 sampai 5. Perhitungan probabilitas bersyarat dilakukan untuk menentukan seberapa besar peluang suatu nilai atribut muncul pada kategori kelas tertentu (tinggi, sedang, atau rendah). Pada tahap perhitungan probabilitas kondisional, penelitian ini menggunakan teknik *Laplace Smoothing* untuk mengatasi kemungkinan munculnya nilai frekuensi nol pada suatu atribut. Metode ini menambahkan konstanta 1 pada frekuensi masing-masing atribut agar seluruh probabilitas tetap terdefinisi.

Probabilitas kondisional :

$$P(X_i|Y) = \frac{\text{Jumlah data dengan fitur } X_i \text{ dalam kelas } Y}{\text{Total jumlah data dalam kelas } Y}$$

Laplace Smoothing :

$$P(X_i|Y) = \frac{\text{Jumlah data dengan fitur } X_i \text{ dalam kelas } Y + 1}{\text{Total jumlah data dalam kelas } Y + k}$$

Substitusi nilai :

$$P(X = 1|\text{Rendah}) = \frac{21 + 1}{71 + 5} = 0,2895$$

$$P(X=2|\text{Rendah}) = 0,6711$$

$$P(X=3|\text{Rendah}) = 0,0132$$

$$P(X=4|\text{Rendah}) = 0,0132$$

$$P(X=5|\text{Rendah}) = 0,0132$$

Selanjutnya dilanjutkan dengan menghitung seluruh probabilitas kondisional pada data latih yang telah ditetapkan.

c. Nilai prediksi

Pada tahap perhitungan nilai prediksi, algoritma Naïve Bayes menentukan kelas minat baca dengan mengombinasikan probabilitas prior dan probabilitas kondisi dari setiap atribut. Probabilitas prior diperoleh dari proporsi jumlah data pada masing-masing kelas, yaitu Tinggi, Sedang, dan Rendah. Selanjutnya, untuk setiap responden, seluruh nilai atribut (20 butir pernyataan) dibandingkan dengan probabilitas kondisi pada setiap kelas. Kondisi tersebut merupakan probabilitas kemunculan suatu nilai pada kelas tertentu yang sebelumnya telah dihitung, termasuk menggunakan Laplace Smoothing untuk mengatasi nilai frekuensi nol.

Proses prediksi dilakukan dengan mengalikan probabilitas prior dengan hasil perkalian semua probabilitas kondisi berdasarkan nilai jawaban responden. Secara matematis, nilai prediksi dihitung menggunakan persamaan:

$$P(Y|X) \propto P(Y) \times \prod_{i=1}^n P(X_i|Y)$$

Substitusi nilai :

$$\begin{aligned} P(\text{Rendah} | \text{Data ke-1}) &= \\ 0,013 \times 0,013 \times 0,013 \times 0,013 \times \\ 0,013 \times 0,013 \times 0,013 \times 0,026 \times \\ 0,0132 \times 0,0132 \times 0,0132 \times 0,0132 \\ \times 0,0132 \times 0,0132 \times 0,0132 \times \\ 0,0132 \times 0,0132 \times 0,0132 \times 0,0132 \\ \times 0,0132 \\ &= 4,83903\text{E-}38 \end{aligned}$$

$$P(\text{Sedang} | \text{Data ke-1}) = 2,31151\text{E-}23$$

$$P(\text{Tinggi} | \text{Data ke-1}) = 3,84969\text{E-}09$$

Setelah dilanjutkan dengan menghitung seluruh nilai prediksi pada data latih yang telah ditetapkan, dilakukan perhitungan manual menggunakan algoritma Naïve Bayes yang menghasilkan nilai probabilitas dan prediksi kelas. Tahapan ini dijadikan dasar dalam penyusunan

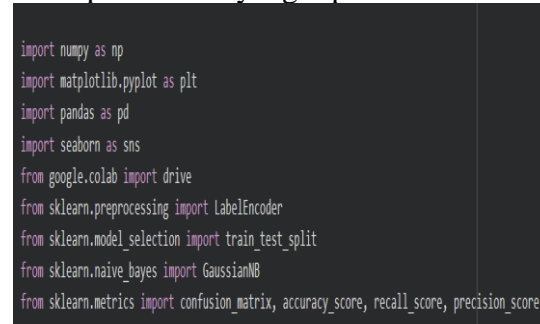
model prediksi minat baca. Selanjutnya, proses perhitungan dijalankan pada Google Colab menggunakan Python di untuk membangun model prediksi yang dapat digunakan dalam proses klasifikasi data.

d. Implementasi

Implementasi metode Naïve Bayes Classifier dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python. Dataset kemudian dibagi menjadi data latih dan data uji dengan beberapa skema pembagian, yaitu 90% data latih : 10% data uji, 80% data latih : 20% data uji, dan 70% data latih : 30% data uji. Proses klasifikasi dilakukan menggunakan algoritma Naïve Bayes, sedangkan evaluasi performa model dilakukan menggunakan confusion matrix untuk memperoleh nilai akurasi, recall, dan precision.

Langkah-langkah implementasi metode menggunakan Python adalah sebagai berikut :

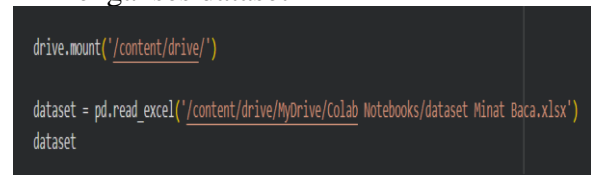
1. Import modul yang diperlukan



```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from google.colab import drive
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import confusion_matrix, accuracy_score, recall_score, precision_score
```

Gambar 2. Import Library

2. Mount Google Drive untuk mengakses dataset



```
drive.mount('/content/drive/')

dataset = pd.read_excel('/content/drive/MyDrive/Colab Notebooks/dataset Minat Baca.xlsx')
dataset
```

Gambar 3. Mount Google Drive untuk mengakses dataset

3. Memisahkan dataset menjadi fitur (x) dan variabel target (y)

```
x = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
```

Gambar 4. Membagi fitur dan kelas label

4. Pembagian data latih dan data uji

```
x_train, x_test, y_train, y_test = train_test
```

Gambar 5. Membagi data latih dan data uji

5. Membuat dan melatih model Naïve Bayes

```
classifier = GaussianNB()
classifier.fit(x_train, y_train)
```

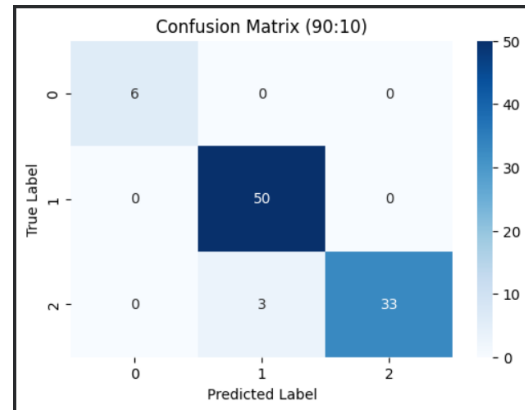
Gambar 6. Model Naïve Bayes

e. Evaluasi

Evaluasi kinerja model dilakukan dengan menggunakan confusion matrix untuk melihat tingkat ketepatan hasil klasifikasi. Berdasarkan hasil pengujian, seluruh data yang berjumlah 911 sampel berhasil diklasifikasikan dengan benar ke dalam kelas masing-masing, yaitu rendah, sedang, dan tinggi.

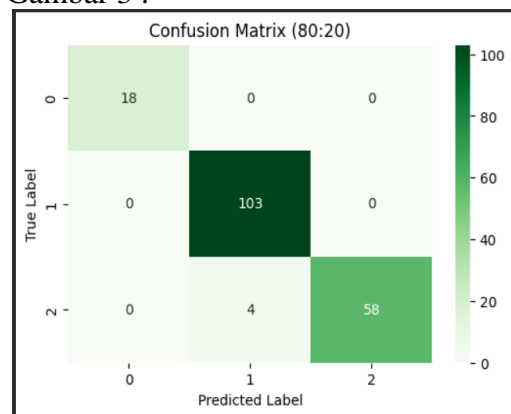
Pengujian model Naïve Bayes Classifier dilakukan menggunakan tiga skema pembagian data, yaitu 90:10, 80:20, dan 70:30.

Pada skema 90:10 digunakan 819 data latih dan 92 data uji dengan distribusi kelas uji terdiri dari 6 kelas rendah, 50 kelas sedang, dan 36 kelas tinggi, yang menghasilkan akurasi sebesar 96,74%, recall 97,22%, dan precision 98,11%. Hasil evaluasi Confusion Matrix dengan pembagian data 90:10 dapat dilihat pada Gambar 2 :



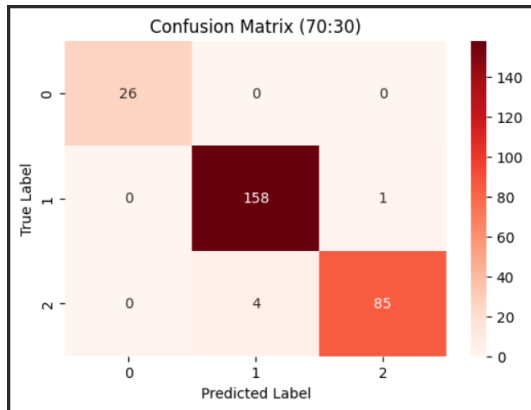
Gambar 7. Confusion Matrix 90:10

Pada skema 80:20 menggunakan 728 data latih dan 183 data uji dengan distribusi 18 kelas rendah, 103 kelas sedang, dan 62 kelas tinggi, serta menghasilkan akurasi 97,81%, recall 97,85%, dan precision 98,75%. Hasil evaluasi Confusion Matrix dengan pembagian data 80:20 dapat dilihat pada Gambar 3 :



Gambar 8. Confusion Matrix 80:20

Sementara itu, pada skema 70:30 digunakan 637 data latih dan 274 data uji dengan distribusi kelas uji sebanyak 26 kelas rendah, 159 kelas sedang, dan 89 kelas tinggi, yang menghasilkan performa terbaik dengan akurasi 98,18%, recall 98,29%, dan precision 98,79%. Hasil evaluasi Confusion Matrix dengan pembagian data 70:30 dapat dilihat pada Gambar 4 :



Gambar 9. Confusion Matrix 70:30

Hasil ini menunjukkan bahwa metode Naïve Bayes Classifier memiliki performa yang konsisten dan sangat baik dalam mengklasifikasikan tingkat minat baca siswa

SIMPULAN

Penelitian ini menunjukkan bahwa metode Naïve Bayes Classifier dapat digunakan untuk mengklasifikasikan tingkat minat baca siswa MTsN 1 Payakumbuh berdasarkan 911 data responden yang diperoleh dari 20 butir pernyataan kuesioner. Hasil pengujian menggunakan confusion matrix menunjukkan akurasi sebesar 96,74% pada pembagian data 90:10, 97,81% pada pembagian data 80:20, dan 98,18% pada pembagian data 70:30. Hasil tersebut menunjukkan bahwa metode Naïve Bayes Classifier berpotensi digunakan sebagai pendekatan dalam analisis minat baca siswa serta mendukung evaluasi pendidikan dan pengembangan penelitian selanjutnya.

DAFTAR PUSTAKA

[1] A. Albani, "Upaya Membangun Minat Membaca Melalui Program Beraksi (Berugak Literasi) Di Smp Islam Musthofa Kamal," *J. Tadbir J. Islam. Educ. Manag.*, vol. 5, no. 1, pp. 24–35, 2021, doi: 10.51700/attadbir.v5i1.122.

[2] N. Nurhaliza, U. Usman, and S. Sultan, "Minat Baca Siswa SMP: Faktor Latar Belakang Ekonomi dan Pendidikan Keluarga (Middle School Students Reading Interests: Factors of Economic Background and Family Education)," *Indones. Lang. Educ. Lit.*, vol. 7, no. 2, p. 323, 2022, doi: 10.24235/ileal.v7i2.10208.

[3] N. Sari, "Jurnal Edukatif," *J. Edukatif*, vol. 3, no. 1, pp. 190–196, 2025, [Online]. Available: file:///C:/Users/User/Downloads/190-196+Ninang+Sari.pdf

[4] S. Dandi, Misdalina, and Novianti, "Analisis Faktor Penyebab Rendahnya Minat Baca pada Siswa Kelas 5 SD Negeri 4 Tanjung Lago," *J. Pendidik. dan Konseling*, vol. 4, no. 4, pp. 1404–1409, 2022, [Online]. Available: <https://core.ac.uk/download/pdf/322599509.pdf>

[5] E. D. Afriani, S. Masfuah, and M. Roysa, "Analisis Minat Baca Siswa Kelas V Sekolah Dasar Dalam Pembelajaran Daring," *J. Prasasti Ilmu*, vol. 1, no. 3, pp. 21–27, 2021, doi: 10.24176/jpi.v1i3.6648.

[6] Zulfan Efendi, Wahyu Nur Hisyam, and Andhita Risko Faristiana, "Kurangnya Minat Baca Buku Kalangan Mahasiswa," *Student Sci. Creat. J.*, vol. 1, no. 4, pp. 382–398, 2023, doi: 10.55606/sscj-amik.v1i4.1676.

[7] N. Agustina, I. S. Ramdhani, and Enawar, "Analisis Gerakan Literasi Pojok Baca Terhadap Minat Baca Kelas 4 SDN Bojong 04," *Al-Irsyad*, vol. 105, no. 2, p. 79, 2022, [Online]. Available: <https://core.ac.uk/download/pdf/322599509.pdf>

- [8] P. D. Lestari, E. Herlina, A. N. Putri, and S. F. Giwangsa, "Pengaruh Gerakan Literasi terhadap Kemampuan Literasi Membaca Siswa Sekolah Dasar," *J. Basicedu*, vol. 7, no. 6, pp. 4003–4009, 2023, doi: 10.31004/basicedu.v7i6.6633.
- [9] Polibatam, "8d270284110535bc9e7f8a210cc83399ccafdaf7 @ www.polibatam.ac.id," 2023. [Online]. Available: <https://www.polibatam.ac.id/peringkat-indonesia-pada-pisa-2022-naik-5-6-posisi-dibanding-2018/>
- [10] OECD, "Indonesia Student Performance." [Online]. Available: <https://gpseducation.oecd.org/CountryProfile?plotter=h5&primaryCountry=IDN&treshold=5&topic=PI>
- [11] I. G. I. Sudipa, I. G. M. Darmawiguna, I. M. Dendi, and M. Sanjaya, *Buku ajar data mining*, no. January. 2024.
- [12] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 452–458, 2023, doi: 10.36040/jati.v7i1.6303.
- [13] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [14] M. Imron, "PENERAPAN DATA MINING ALGORITMA NAIVES BAYES DAN PART UNTUK MENGETAHUI MINAT BACA MAHASISWA DI PERPUSTAKAAN STMIK AMIKOM PURWOKERTO," 2017.
- [15] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *J. Inform.*, vol. 8, no. 2, pp. 111–122, 2021, doi: 10.31294/ji.v8i2.10494.
- [16] S. Syofian, T. Setiyaningsih, N. Syamsiah, T. Informatika, F. Teknik, and U. D. Persada, "Otomatisasi metode penelitian skala likert berbasis web," no. November, pp. 1–8, 2015.
- [17] A. A. Santika, T. H. Saragih, and M. Muliadi, "Penerapan Skala Likert pada Klasifikasi Tingkat Kepuasan Pelanggan Agen Brilink Menggunakan Random Forest," *J. Sist. dan Teknol. Inf.*, vol. 11, no. 3, p. 405, 2023, doi: 10.26418/justin.v11i3.62086.