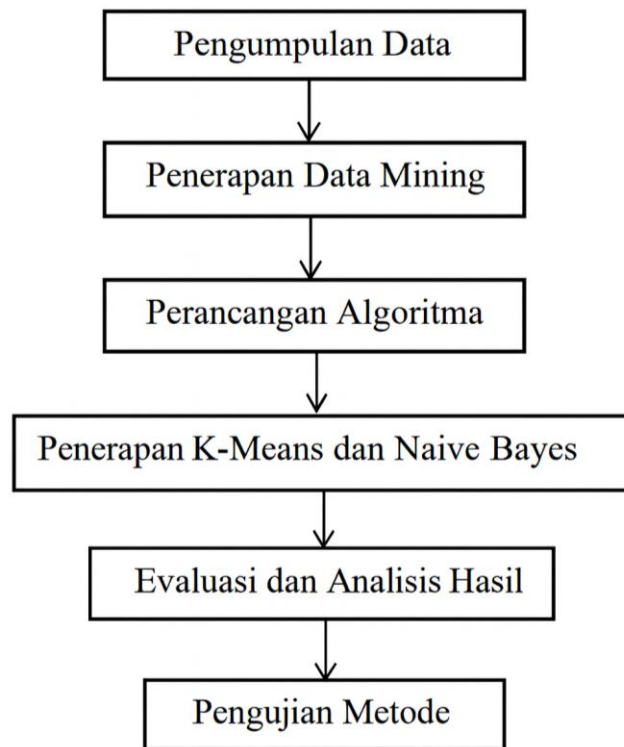


BAB III
METODOLOGI PENELITIAN

3.1. Kerangka Kerja Penelitian

Kerangka Kerja Penelitian merupakan tahapan – tahapan yang dilakukan oleh peneliti dalam menyelesaikan penelitian. Adapun langkah – langkah yang dilakukan dalam penelitian ini.



Gambar 3.1 Kerangka Kerja Penelitian

Dari Gambar 3.1 di atas terlihat jelas bahwa penelitian ini dilakukan secara bertahap dan sistematis, berikut ini adalah penjelasan dari gambar kerangka kerja di atas

Tahap-tahap sebagai berikut:

1. Pengumpulan Data

Tahap pertama dalam penelitian ini adalah pengumpulan data yang bertujuan untuk memperoleh informasi yang relevan terkait prestasi akademik Bahasa Inggris siswa SMA swasta Byangkari 2 Rantau Prapat.

Data dikumpulkan melalui pihak sekolah, wali kelas, serta penyebaran kuesioner kepada siswa. Variabel yang dihimpun mencakup nilai Bahasa Inggris, kehadiran, kursus bahasa inggris, minat belajar. Seluruh variabel ini dipilih karena memiliki keterkaitan dengan capaian akademik siswa dan menjadi dasar penting dalam proses analisis berbasis data mining. Data yang berhasil dikumpulkan kemudian menjadi fondasi utama dalam membangun model pengelompokan menggunakan algoritma K-Means dan model klasifikasi menggunakan Naive Bayes.

2. Setelah seluruh data terkumpul, langkah berikutnya adalah menerapkan teknik data mining melalui proses *preprocessing* untuk memastikan kualitas data yang akan dianalisis. Tahap ini meliputi pembersihan data dari nilai kosong atau duplikasi, normalisasi agar seluruh variabel berada dalam rentang yang seragam, serta transformasi data agar setiap atribut dapat diproses oleh algoritma K-Means dan Naive Bayes. Penerapan data mining pada tahap ini bertujuan untuk mempersiapkan data sehingga dapat menghasilkan pola dan informasi yang akurat. Proses ini sangat penting karena kualitas data secara langsung mempengaruhi kualitas hasil analisis yang akan dihasilkan pada tahap selanjutnya.

3. Perancangan Model K-Means dan Naive Bayes

Tahap ketiga merupakan proses perancangan model analisis yang melibatkan dua algoritma utama, yaitu K-Means dan Naive Bayes. Pertama, algoritma K-Means digunakan untuk mengelompokkan siswa berdasarkan

kemiripan nilai akademik dan faktor pendukung lainnya sehingga terbentuk cluster seperti kelompok prestasi tinggi, sedang, dan rendah.

Pengelompokan ini membantu pemahaman mengenai pola umum prestasi akademik di sekolah. Setelah cluster terbentuk, algoritma Naive Bayes digunakan untuk menganalisis lebih dalam faktor-faktor yang berpengaruh terhadap prestasi pada setiap cluster. Model Naive Bayes dibangun dengan membagi dataset menjadi data training dan data testing untuk melatih dan menguji kemampuan model dalam melakukan klasifikasi. Tahap ini menjadi inti dari penelitian karena menghasilkan model analisis yang dapat digunakan untuk memprediksi kategori prestasi siswa berdasarkan variabel pendukung.

Model-model ini kemudian diuji menggunakan dataset yang dibagi menjadi Data Training dan Data Testing.

- a. Data Training digunakan untuk melatih model agar bisa memprediksi faktor prestasi akademik siswa atau klasifikasi berdasarkan variabel yang ada.

- b. Data Testing digunakan untuk menguji kemampuan model dalam memprediksi atau mengklasifikasikan data yang belum pernah dilatih sebelumnya.

4. Evaluasi Model

Model yang telah dibangun selanjutnya dievaluasi untuk mengukur keefektifan dan akurasinya. Evaluasi algoritma K-Means dilakukan dengan meninjau kualitas cluster yang terbentuk, keseragaman data dalam cluster,

serta jarak antar centroid. Sementara itu, evaluasi Naive Bayes dilakukan menggunakan metrik seperti akurasi, precision, recall, dan confusion matrix untuk mengetahui kemampuan model dalam mengklasifikasikan data secara tepat. Hasil evaluasi memberikan gambaran mengenai keunggulan dan keterbatasan masing-masing metode, sekaligus memastikan bahwa model yang dihasilkan dapat dipercaya dan layak digunakan dalam proses pengambilan keputusan terkait prestasi siswa.

5. Pengambilan Keputusan Akademik

Tahap terakhir dalam penelitian ini adalah pemanfaatan hasil analisis untuk mendukung pengambilan keputusan strategis di sekolah. Informasi yang dihasilkan dari pengelompokan K-Means dan klasifikasi Naive Bayes dapat membantu sekolah dalam memahami faktor-faktor dominan yang memengaruhi prestasi Bahasa Inggris siswa. Selain itu, hasil analisis dapat digunakan untuk menentukan kebijakan akademik seperti penyusunan strategi pembelajaran yang lebih efektif, pemberian

bimbingan khusus kepada siswa yang berada dalam kelompok berisiko rendah, atau peningkatan dukungan akademik berdasarkan karakteristik tiap cluster. Dengan demikian, implementasi metode gabungan ini dapat memberikan dasar pertimbangan yang objektif dan berbasis data bagi sekolah dalam meningkatkan mutu pembelajaran Bahasa Inggris.

3.2. Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan pada SMA swasta Bhayangkari 2 Rantau Prapat. Sekolah ini dipilih karena memiliki karakteristik siswa yang beragam serta variasi prestasi akademik Bahasa Inggris yang cukup signifikan sehingga relevan untuk dianalisis menggunakan metode K-Means dan Naive Bayes. Penelitian dijadwalkan dilaksanakan pada bulan Januari hingga Maret 2026, dengan fokus utama pada pengumpulan data nilai Bahasa Inggris, kehadiran, kursus bahasa Inggris, dan minat belajar. Data yang terkumpul kemudian dianalisis untuk mengidentifikasi faktor-faktor yang memengaruhi prestasi akademik siswa melalui penerapan algoritma K-Means dan Naive Bayes.

3.3. Populasi dan Sampel

Populasi penelitian adalah seluruh siswa kelas X IPA SMA swasta Bhayangkari 2 Rantau Prapat. Sampel ditentukan Siswa kelas X Menggunakan teknik purposive sampling, yaitu siswa yang memiliki data lengkap mencakup nilai Bahasa Inggris, kehadiran, kursus bahasa Inggris, dan minat belajar.

3.4. Variabel Penelitian

Data yang digunakan dalam penelitian ini diperoleh dari dua sumber utama, yaitu data primer dan data sekunder. Data nilai Bahasa Inggris (X1) dan kehadiran siswa (X2) diperoleh dari dokumen resmi sekolah, berupa nilai rapor dan data absensi siswa. Sementara itu, data kursus Bahasa Inggris (X3) dan minat belajar siswa (X4) dikumpulkan melalui kuesioner yang diberikan kepada siswa. Penggunaan dua sumber data ini bertujuan untuk memperoleh data yang objektif serta melengkapi informasi yang tidak tersedia dalam administrasi sekolah. Variabel penelitian dalam penerapan algoritma K-Means dan Naïve Bayes untuk mengetahui prestasi nilai prestasi akademik bahasa inggris siswa di SMA Swasta Bhayangkari 2 Rantau Prapat siswa kelas X angkatan 2025/2026 yang terdiri dari:

1. Nilai Bahasa Inggris (X1)
2. Kehadiran(X2)
3. Kursus Bahasa Inggris (X3)
4. Minat Belajar (X4)
5. Prestasi (Y)

3.5 Langkah – Langkah Penerapan Algoritma K-Means

Algoritma K-Means digunakan untuk mengelompokkan prestasi akademik siswa berdasarkan nilai Bahasa Inggris ke dalam beberapa klaster, yaitu rendah, sedang, tinggi. Pengelompokan ini bertujuan untuk mengidentifikasi pola kemampuan akademik siswa secara objektif berdasarkan kemiripan data nilai. Metode ini termasuk dalam teknik unsupervised learning, karena tidak memerlukan label kelas di awal.

Tabel 3.1 Data Prestasi Akademik Siswa

No	Nama Siswa	Nilai B. Inggris	Kehadiran%	Kursus	Minat	Prestasi
1	AB**** TA*****	Rendah	Rendah	Ya	Sedang	Rendah
2	CD**** AS**** TO**	Sedang	Rendah	Tidak	Tinggi	Sedang
3	EF** ME*****	Rendah	Sedang	Tidak	Rendah	Rendah
4	G**** E**	Rendah	Rendah	Tidak	Sedang	Rendah
5	IJ** EV**	Sedang	Tinggi	Ya	Sedang	Sedang
6	KL** HE**** LU***	Sedang	Rendah	Tidak	Sedang	Sedang
7	MN*** MA*****	Rendah	Tinggi	Tidak	Tinggi	Rendah
8	OP*** KH*****	Sedang	Sedang	Tidak	Sedang	Sedang
9	QR** PU*****	Tinggi	Tinggi	Tidak	Tinggi	Tinggi
10	ST***** RA*	Rendah	Sedang	Tidak	Sedang	Rendah
11	UV*** TH***	Rendah	Sedang	Tidak	Rendah	Rendah
12	WX*** AF*****	Rendah	Rendah	Ya	Rendah	Rendah
13	YZ** VI***	Rendah	Sedang	Tidak	Sedang	Rendah
14	AB***** PA**	Rendah	Rendah	Tidak	Tinggi	Rendah
15	CD**** PU***	Rendah	Sedang	Ya	Rendah	Rendah
16	EF***** WI***	Sedang	Sedang	Ya	Tinggi	Sedang
17	GH*** YO****	Tinggi	Sedang	Ya	Tinggi	Tinggi
18	IJ*** HU*****	Tinggi	Tinggi	Ya	Sedang	Tinggi
19	KL** UL***	Rendah	Tinggi	Tidak	Rendah	Rendah
20	MN* SY*****	Sedang	Tinggi	Tidak	Sedang	Sedang
21	OP**** SA**	Sedang	Sedang	Tidak	Tinggi	Sedang
22	QR*****	Tinggi	Sedang	Ya	Sedang	Tinggi
23	ST*****	Rendah	Rendah	Ya	Tinggi	Rendah
24	UV*** SU**	Sedang	Sedang	Ya	Rendah	Sedang
25	WX*** DO***	Sedang	Rendah	Ya	Rendah	Sedang
26	YZ**** SE*****	Rendah	Tinggi	Ya	Sedang	Rendah
27	AB**** AU***	Rendah	Tinggi	Ya	Sedang	Rendah
28	CD** MA****	Tinggi	Tinggi	Tidak	Sedang	Tinggi
29	EF***** LO****	Rendah	Sedang	Ya	Tinggi	Rendah
30	LI**** VI****	Sedang	Sedang	Ya	Sedang	Sedang
31	IJ***** WA****	Tinggi	Sedang	Tidak	Tinggi	Tinggi
32	KL**** TH****	Sedang	Tinggi	Ya	Rendah	Sedang
33	MN**** JE***	Rendah	Tinggi	Ya	Rendah	Rendah
34	OP**** HA*****	Rendah	Sedang	Ya	Rendah	Rendah

35	QR*** SI*****	Rendah	Rendah	Ya	Sedang	Rendah
36	ST***** JA****	Sedang	Tinggi	Ya	Tinggi	Sedang
37	UV*** TA*****	Tinggi	Tinggi	Ya	Rendah	Tinggi
38	WX**** MA*****	Tinggi	Rendah	Ya	Tinggi	Tinggi
39	VZ** TA*****	Sedang	Rendah	Tidak	Tinggi	Sedang
40	VB** SA*****	Rendah	Sedang	Tidak	Tinggi	Rendah
41	ZD** KI*****	Tinggi	Rendah	Tidak	Tinggi	Tinggi
42	ZF** GI*****	Tinggi	Sedang	Tidak	Sedang	Tinggi
43	ZH** TY*****	Sedang	Tinggi	Tidak	Tinggi	Sedang
44	VJ** KU*****	Rendah	Tinggi	Tidak	Sedang	Rendah
45	KL** LM*****	Sedang	Rendah	Ya	Sedang	Sedang
46	VN*** DO*****	Tinggi	Sedang	Ya	Rendah	Tinggi
47	VP** LU*****	Tinggi	Tinggi	Ya	Sedang	Tinggi
48	ZR** MO*****	Tinggi	Sedang	Ya	Tinggi	Tinggi
49	ZT** NO*****	Tinggi	Sedang	Tidak	Rendah	Tinggi
50	ZV** AN*****	Tinggi	Tinggi	Ya	Tinggi	Tinggi

Pada Tabel 3.1 Data Prestasi Akademik Siswa dilakukan transformasi data. Transformasi data dilakukan untuk mengonversi data hasil kuesioner menjadi data numerik. Sehingga dapat dihitung menggunakan algoritma K-Means Clustering, memenuhi syarat perhitungan jarak Euclidean, dan Menjaga konsistensi matematis dan metodologis. Secara konseptual, tahap ini merupakan bagian dari preprocessing data dalam data mining. Nilai Bahasa Inggris dikategorikan menjadi 1(rendah) jika nilainya ≤ 65 , 2(sedang) jika berada pada rentang 66–79, dan 3 (tinggi) jika ≥ 80 . Kehadiran dibagi menjadi 1(rendah) apabila persentase kehadiran $\leq 70\%$, 2 (sedang) jika 71–85%, dan 3 (tinggi) jika lebih dari 85%. Kursus Bahasa Inggris menunjukkan apakah siswa mengikuti kursus, dengan kategori tidak (1) dan ya (2). Minat Belajar terdiri atas tiga tingkat, yaitu 1(rendah), 2(sedang), dan

3(tinggi).Prestasi diklasifikasikan ke dalam tiga kategori, yaitu prestasi rendah, sedang, dan tinggi.

Tabel 3.2 Data Transformasi Prestasi Akademik Siswa

No	Nama Siswa	Nilai B. Inggris	Kehadiran%	Kursus	Minat
1	AB**** TA*****	1	1	2	2
2	CD**** AS***** TO****	2	1	1	3
3	EF** ME*****	1	2	1	1
4	G**** E**	1	1	1	2
5	IJ** EV**	2	3	2	2
6	KL** HE***** LU****	2	1	1	2
7	MN*** MA*****	1	3	1	3
8	OP*** KH*****	2	2	1	2
9	QR** PU*****	3	3	1	3
10	ST***** RA*	1	2	1	2
11	UV*** TH***	1	2	1	1
12	WX*** AF*****	1	1	2	1
13	YZ** VI***	1	2	1	2
14	AB***** PA**	1	1	1	3
15	CD**** PU***	1	2	2	1
16	EF***** WI***	2	2	2	3
17	GH*** YO*****	3	2	2	3
18	IJ*** HU*****	3	3	2	2
19	KL** UL***	1	3	1	1
20	MN* SY*****	2	3	1	2
21	OP***** SA**	2	2	1	3
22	QR*****	3	2	2	2
23	ST*****	1	1	2	3
24	UV*** SU**	2	2	2	1
25	WX*** DO***	2	1	2	1
26	YZ**** SE*****	1	3	2	2
27	AB**** AU***	1	2	2	2
28	CD** MA****	3	3	1	2
29	EF***** LO****	1	2	2	3
30	LI**** VI****	2	2	2	2

31	IJ***** WA****	3	2	1	3
32	KL**** TH****	2	3	2	1
33	MN**** JE****	1	3	2	1
34	OP**** HA*****	1	2	2	1
35	QR*** SI*****	1	1	2	2
36	ST***** JA****	2	3	2	3
37	UV*** TA*****	3	3	2	1
38	WX**** MA*****	3	1	2	3
39	VZ** TA*****	2	1	1	3
40	VB** SA*****	1	2	1	3
41	ZD** KI*****	3	1	1	3
42	ZF** GI*****	3	2	1	2
43	ZH** TY*****	2	3	1	3
44	VJ** KU*****	1	3	1	2
45	KL** LM*****	2	1	2	2
46	VN*** DO*****	3	2	2	1
47	VP** LU*****	3	3	2	2
48	ZR** MO*****	3	2	2	3
49	ZT** NO*****	3	2	1	1
50	ZV** AN*****	3	3	2	3

Berdasarkan tabel di atas, data sampel telah dikonversi sehingga dapat digunakan dalam proses perhitungan dengan metode algoritma K-Means Clustering. Tahap awal yang dilakukan adalah menetapkan jumlah cluster, yaitu sebanyak tiga cluster. Setelah jumlah cluster ditentukan, langkah selanjutnya adalah menentukan nilai centroid awal secara acak. Nilai centroid awal yang digunakan pada proses ini ditunjukkan pada Tabel 3.3 berikut.

Tabel 3.3 Centeroid Awal

N0	Nama Siswa	Nilai B. inggris	Kehadiran%	Kursus	Minat	
1	G**** E**	1	1	1	2	C1
2	LI**** VI****	2	2	2	2	C2
3	ZV** AN*****	3	3	2	3	C3

Setelah menentukan nilai Centeroid, langkah selanjutnya adalah menghitung jarak setiap data terhadap masing-masing centroid menggunakan rumus Euclidean Distance. Perhitungan jarak ini bertujuan untuk menentukan data tersebut termasuk ke dalam cluster yang sesuai. Proses perhitungannya dilakukan sebagai berikut, dengan hasil yang disajikan pada Tabel 3.4

Jarak ke C1

$$\sqrt{(1 - 1)^2 + (1 - 1)^2 + (2 - 1)^2 + (2 - 2)^2}$$

$$= 1$$

Jarak ke C2

$$\sqrt{(1 - 2)^2 + (1 - 2)^2 + (2 - 2)^2 + (2 - 2)^2}$$

$$= 1,73205$$

Jarak ke C3

$$\sqrt{(1 - 3)^2 + (1 - 3)^2 + (2 - 2)^2 + (2 - 3)^2}$$

$$= 3,6055513$$

Tabel 3.4 Hasil Perhitungan Iterasi Ke – 1

No	Nama Siswa	X1	X2	X3	X4	C1	C2	C3	Hasil Cluster
1	AB**** TA*****	1	1	2	2	1	1,73205	3,6055513	C1
2	CD**** AS**** TO***	2	1	1	3	1,7320508	1,73205	2,6457513	C2
3	EF** ME*****	1	2	1	1	1,4142136	2	3,7416574	C1
4	G**** E**	1	1	1	2	0	2	3,7416574	C1
5	IJ** EV**	2	3	2	2	2,6457513	1	1,7320508	C2
6	KL** HE**** LU***	2	1	1	2	1,4142136	1,41421	2,8284271	C1
7	MN*** MA*****	1	3	1	3	2,236068	2,23607	3	C1
8	OP*** KH*****	2	2	1	2	1,7320508	1	2,236068	C2
9	QR** PU*****	3	3	1	3	3,6055513	2,23607	1	C3
10	ST***** RA*	1	2	1	2	1	1,73205	3,3166248	C1
11	UV*** TH***	1	2	1	1	1,4142136	2	3,7416574	C1
12	WX*** AF****	1	1	2	1	1,4142136	2	4	C1
13	YZ** VI***	1	2	1	2	1	1,73205	3,3166248	C1
14	AB***** PA**	1	1	1	3	1	2,23607	3,6055513	C1
15	CD**** PU***	1	2	2	1	1,7320508	1,73205	3,6055513	C1
16	EF***** WI***	2	2	2	3	2,236068	1	1,7320508	C2
17	GH*** YO****	3	2	2	3	3,3166248	1,73205	1	C3
18	IJ*** HU*****	3	3	2	2	3,6055513	1,73205	1	C3
19	KL** UL***	1	3	1	1	2,236068	2,23607	3,6055513	C1
20	MN* SY*****	2	3	1	2	2,4494897	1,41421	2	C2
21	OP**** SA**	2	2	1	3	2	1,41421	2	C2
22	QR*****	3	2	2	2	3,1622777	1,41421	1,4142136	C3
23	ST*****	1	1	2	3	1,4142136	2	3,4641016	C1
24	UV*** SU**	2	2	2	1	2,236068	1	2,6457513	C2
25	WX*** DO****	2	1	2	1	2	1,41421	3,1622777	C2
26	YZ**** SE*****	1	3	2	2	2,236068	1,73205	3	C2
27	AB**** AU***	1	2	2	2	1,4142136	1,41421	3,1622777	C1
28	CD** MA****	3	3	1	2	3,4641016	2	1,4142136	C3
29	EF***** LO****	1	2	2	3	1,7320508	1,73205	3	C1
30	LI**** VI****	2	2	2	2	2	0	2	C2
31	IJ***** WA****	3	2	1	3	3,1622777	2	1,4142136	C3
32	KL**** TH****	2	3	2	1	2,8284271	1,41421	2,4494897	C2
33	MN***** JE***	1	3	2	1	2,4494897	2	3,4641016	C2
34	OP**** HA*****	1	2	2	1	1,7320508	1,73205	3,6055513	C1

35	QR*** SI*****	1	1	2	2	1	1,73205	3,6055513	C1
36	ST***** JA****	2	3	2	3	2,8284271	1,41421	1,4142136	C2
37	UV*** TA*****	3	3	2	1	3,7416574	2	2	C3
38	WX**** MA*****	3	1	2	3	3,1622777	2	2	C2
39	VZ** TA*****	2	1	1	3	1,7320508	1,73205	2,6457513	C2
40	VB** SA*****	1	2	1	3	1,4142136	2	3,1622777	C1
41	ZD** KI*****	3	1	1	3	3	2,23607	2,236068	C2
42	ZF** GI*****	3	2	1	2	3	1,73205	1,7320508	C3
43	ZH** TY*****	2	3	1	3	2,6457513	1,73205	1,7320508	C2
44	VJ** KU*****	1	3	1	2	2	2	3,1622777	C1
45	KL** LM*****	2	1	2	2	1,7320508	1	2,6457513	C2
46	VN*** DO*****	3	2	2	1	3,3166248	1,73205	2,236068	C2
47	VP** LU*****	3	3	2	2	3,6055513	1,73205	1	C3
48	ZR** MO*****	3	2	2	3	3,3166248	1,73205	1	C3
49	ZT** NO*****	3	2	1	1	3,1622777	2	2,4494897	C2
50	ZV** AN*****	3	3	2	3	3,7416574	2	0	C3

Setelah proses pada iterasi 1, maka selanjutnya dilakukan proses terhadap iterasi 2. Pada iterasi 2 proses yang dilakukan sama halnya dengan iterasi 1, namun untuk nilai centeroid didapatkan berdasarkan nilai rata – rata dari pengelompokan data berdasarkan hasil cluster. Hasilnya dapat dilihat pada Tabel 3.5 sebagai berikut:

Tabel 3.5 Nilai Centeroid Baru Iterasi Ke - 2

X1	X2	X3	X4	
1,125	1,5625	1,375	1,9375	C1
2,0385	2,19231	1,6154	2,0385	C2
3	2,625	1,625	2,625	C3

Selanjutnya dengan proses yang sama seperti iterasi 1, dapat dilihat hasil dari iterasi ke -2 pada Tabel 3.6 dibawah ini:

Tabel 3.6 Hasil Perhitungan Iterasi Ke- 2

No	Nama Siswa	C1	C2	C3	Hasil Cluster
1	AB**** TA*****	0,8615	1,9308	3,3048	C1
2	CD**** AS**** TO****	1,76556	1,6516	2,6309	C2
3	EF** ME*****	1,11453	1,9108	3,0285	C1
4	G**** E**	0,70156	1,9896	3,3424	C1
5	IJ** EV**	1,99805	0,8971	1,5562	C2
6	KL** HE**** LU****	1,41145	1,3434	2,2742	C2
7	MN*** MA*****	1,83499	2,0279	3,2283	C1
8	OP*** KH*****	1,36645	0,6482	1,7093	C2
9	QR** PU*****	3,21981	1,9506	1,5562	C3
10	ST***** RA*	0,60596	1,6044	2,987	C1
11	UV*** TH****	1,11453	1,9108	3,0285	C1
12	WX*** AF*****	1,27169	2,192	3,3424	C1
13	YZ** VI****	0,60596	1,6044	2,987	C1
14	AB***** PA**	1,27169	2,2095	3,5947	C1
15	CD**** PU****	1,22155	1,8494	2,987	C1
16	EF***** WI****	1,80061	1,0547	2,1028	C2
17	GH*** YO****	2,95672	1,72	1,5562	C3
18	IJ*** HU*****	3,08094	1,6281	0,6495	C3
19	KL** UL****	1,76556	2,0655	2,987	C1
20	MN* SY*****	1,93447	1,0176	1,6346	C2
21	OP**** SA**	1,72979	1,159	2,1615	C2
22	QR*****	2,75993	1,4267	0,8197	C3
23	ST*****	1,36645	2,1566	3,5598	C1
24	UV*** SU**	1,72979	1,1253	1,7093	C2
25	WX*** DO****	1,76556	1,6282	2,2742	C2
26	YZ**** SE*****	1,57867	1,7201	2,902	C1
27	AB**** AU****	0,78561	1,5308	2,9448	C1
28	CD** MA****	3,0401	1,6975	0,8197	C3
29	EF***** LO****	1,31992	1,8073	3,2283	C1
30	LI**** VI****	1,45506	0,4351	1,6346	C2
31	IJ***** WA****	2,91414	1,7859	1,6346	C3
32	KL**** TH****	2,20617	1,3717	1,6346	C2
33	MN**** JE****	1,83499	2,0089	2,9448	C1
34	OP**** HA*****	1,22155	1,8494	2,987	C1

35	QR*** SI*****	0,8615	1,9308	3,3048	C1
36	ST***** JA****	2,26212	1,3145	2,0425	C2
37	UV*** TA*****	3,21981	1,9307	0,8197	C3
38	WX**** MA*****	2,97778	2,084	2,1615	C2
39	VZ** TA*****	1,76556	1,6516	2,6309	C2
40	VB** SA*****	1,22155	1,8701	3,2668	C1
41	ZD** KI*****	2,9355	2,1387	2,2185	C2
42	ZF** GI*****	2,71426	1,5054	0,9601	C3
43	ZH** TY*****	2,20617	1,3995	2,1028	C2
44	VJ** KU*****	1,49739	1,7859	2,9448	C1
45	KL** LM*****	1,49739	1,2546	2,2185	C2
46	VN*** DO*****	2,91414	1,7642	0,9601	C3
47	VP** LU*****	3,08094	1,6281	0,6495	C3
48	ZR** MO*****	2,95672	1,72	1,5562	C3
49	ZT** NO*****	2,87092	1,8284	1,0825	C3
50	ZV** AN*****	3,2584	1,8905	1,4737	C3

Setelah proses pada iterasi 2, maka selanjutnya dilakukan proses terhadap iterasi 3. Pada iterasi 3 proses yang dilakukan sama halnya dengan iterasi 2, namun untuk nilai centeroid didapatkan berdasarkan nilai rata – rata dari pengelompokan data berdasarkan hasil cluster. Hasilnya dapat dilihat pada Tabel 3.7 sebagai berikut:

Tabel 3.7 Nilai Centeroid Baru Iterasi Ke – 3

X1	X2	X3	X4	
1	1,95	1,5	1,9	C1
2,1176	1,88235	1,5294	2,2941	C2
3	2,46154	1,6154	2,3077	C3

Selanjutnya dengan proses yang sama seperti iterasi 2, dapat dilihat hasil dari iterasi ke -3 pada Tabel 3.8 dibawah ini:

Tabel 3.8 Hasil Perhitungan Iterasi Ke – 3

No	Nama Siswa	C1	C2	C3	Hasil Cluster
1	AB**** TA*****	1,07819	1,9991	3,2216	C1
2	CD**** AS***** TO***	2,08866	1,6617	2,2347	C2
3	EF** ME*****	1,03078	1,6967	3,2096	C1
4	G**** E**	1,07819	1,9088	3,2572	C1
5	IJ** EV**	1,83371	1,2589	1,5914	C2
6	KL** HE***** LU***	1,77834	1,2589	2,147	C2
7	MN*** MA*****	1,60078	2,0283	3,0246	C1
8	OP*** KH*****	1,50416	0,9056	1,639	C2
9	QR** PU*****	3,25	1,7813	1,0714	C3
10	ST***** RA*	0,51235	2,1135	2,9473	C1
11	UV*** TH***	1,03078	2,1135	3,2096	C1
12	WX*** AF*****	1,40089	1,8933	3,4632	C1
13	YZ** VI***	0,51235	1,814	2,9473	C1
14	AB***** PA**	1,53704	2,2873	3,3157	C1
15	CD**** PU***	1,03078	1,7977	3,1735	C1
16	EF***** WI***	1,86078	0,8725	1,6853	C2
17	GH*** YO*****	3,07612	1,3708	0,9166	C3
18	IJ*** HU*****	3,05982	2,1685	0,7298	C3
19	KL** UL***	1,47054	2,0283	3,2216	C1
20	MN* SY*****	1,83371	1,4337	1,6623	C2
21	OP***** SA**	1,86078	0,639	1,7524	C2
22	QR*****	2,87446	1,5135	0,675	C3
23	ST*****	1,53704	2,2744	3,2807	C1
24	UV*** SU**	1,75	1,392	2,0177	C2
25	WX***DO***	1,9906	1,0555	2,4483	C2
26	YZ**** SE*****	1,16726	2,0138	2,9211	C1
27	AB**** AU***	0,51235	1,6793	2,9079	C1
28	CD** MA****	3,05982	1,8934	0,8737	C3
29	EF***** LO****	1,20934	1,6793	2,9733	C1
30	LI**** VI****	1,50416	0,8725	1,567	C2

31	IJ***** WA****	3,07612	1,8778	1,0349	C3
32	KL**** TH*****	2,04022	1,7813	2,0366	C2
33	MN**** JE****	1,47054	2,3756	3,1856	C1
34	OP**** HA*****	1,03078	1,6793	3,1735	C1
35	QR*** SI*****	1,07819	1,9991	3,2216	C1
36	ST***** JA****	2,136	1,7813	1,708	C3
37	UV*** TA*****	3,18787	1,8778	1,4656	C3
38	WX**** MA*****	3,21908	1,748	1,6623	C3
39	VZ** TA*****	2,08866	1,2589	2,2347	C2
40	VB** SA*****	1,20934	1,814	3,0118	C1
41	ZD** KI*****	3,21908	1,6439	1,7303	C2
42	ZF** GI*****	2,87446	1,5328	0,8285	C3
43	ZH** TY*****	2,136	1,282	1,7742	C2
44	VJ** KU*****	1,16726	2,0283	2,9603	C1
45	KL** LM*****	1,77834	1,6439	2,0925	C2
46	VN**** DO*****	3,0104	1,3708	1,4391	C2
47	VP** LU*****	3,05982	1,8778	0,7298	C3
48	ZR** MO*****	3,07612	1,862	0,9166	C3
49	ZT** NO*****	3,0104	1,5328	1,5172	C3
50	ZV** AN*****	3,25	2,8794	0,9577	C3

Setelah proses pada iterasi 3, maka selanjutnya dilakukan proses terhadap iterasi 4. Pada iterasi 4 proses yang dilakukan sama halnya dengan iterasi 3, namun untuk nilai centeroid didapatkan berdasarkan nilai rata – rata dari pengelompokan data berdasarkan hasil cluster. Hasilnya dapat dilihat pada Tabel 3.9 sebagai berikut:

Tabel 3.9 Nilai Centeroid Baru Iterasi Ke- 4

X1	X2	X3	X4	
1	1,95	1,5	1,9	C1
2,125	2	1,5	2,125	C2
2,9286	2,42857	1,6429	2,3571	C3

Selanjutnya dengan proses yang sama seperti iterasi 3, dapat dilihat hasil dari iterasi ke 4 pada Tabel 3.10 dibawah ini:

Tabel 3. 10. Hasil Perhitungan Iterasi Ke - 4

No	Nama Siswa	C1	C2	C3	Hasil Cluster
1	AB**** TA*****	1,07819	1,5959	3,12005	C1
2	CD**** AS**** TO***	2,08866	1,6724	2,1429	C2
3	EF** ME*****	1,03078	1,6724	3,1429	C1
4	G**** E**	1,07819	1,5959	3,1655	C1
5	IJ** EV**	1,83371	1,4307	1,5186	C2
6	KL** HE**** LU***	1,77834	1,4307	2,0751	C2
7	MN*** MA*****	1,60078	1,8157	2,9312	C1
8	OP*** KH*****	1,50416	1,0232	1,5649	C2
9	QR** PU*****	3,25	2,5094	1,0785	C3
10	ST***** RA*	0,51235	1,2437	2,8571	C1
11	UV*** TH***	1,03078	1,6724	3,1429	C1
12	WX*** AF*****	1,40089	1,9486	3,3836	C1
13	YZ** VI***	0,51235	1,2437	2,8571	C1
14	AB***** PA**	1,53704	1,8157	3,2103	C1
15	CD**** PU***	1,03078	1,6724	3,0971	C1
16	EF***** WI***	1,86078	1,3405	1,5649	C2
17	GH*** YO*****	3,07612	2,3015	0,8571	C3
18	IJ*** HU*****	3,05982	2,3552	0,7693	C3
19	KL** UL***	1,47054	1,9486	3,1655	C1
20	MN* SY*****	1,83371	1,4307	1,6099	C2
21	OP***** SA**	1,86078	1,3405	1,6537	C2
22	QR*****	2,87446	2,1323	0,6701	C3
23	ST*****	1,53704	1,8157	3,1655	C1
24	UV*** SU**	1,75	1,5155	1,9691	C2
25	WX*** DO***	1,9906	1,8157	2,3947	C2
26	YZ**** SE*****	1,16726	1,5959	2,832	C1
27	AB**** AU***	0,51235	1,2437	2,8067	C1
28	CD** MA*****	3,05982	2,3552	0,9368	C3
29	EF***** LO*****	1,20934	1,5155	2,8571	C1
30	LI**** VI*****	1,50416	1,0232	1,4708	C2

31	IJ***** WA*****	3,07612	2,3015	1,0102	C3
32	KL**** TH*****	2,04022	1,8157	2,0051	C2
33	MN**** JE****	1,47054	1,9486	3,12	C1
34	OP**** HA*****	1,03078	1,6724	3,0971	C1
35	QR*** SI*****	1,07819	1,5959	3,12	C1
36	ST***** JA*****	2,136	1,6724	1,6099	C3
37	UV*** TA*****	3,18787	2,6071	1,5186	C3
38	WX**** MA*****	3,21908	2,5094	1,6099	C3
39	VZ** TA*****	2,08866	1,6724	2,1429	C2
40	VB** SA*****	1,20934	1,5155	2,9067	C1
41	ZD** KI*****	3,21908	2,5094	2,6985	C2
42	ZF** GI*****	2,87446	2,1323	0,8571	C3
43	ZH** TY*****	2,136	1,6724	1,6963	C2
44	VJ** KU*****	1,16726	1,5959	2,882	C1
45	KL** LM*****	1,77834	1,4307	2,0051	C2
46	VN**** DO*****	3,0104	2,4077	2,7834	C2
47	VP** LU*****	3,05982	2,3552	0,7693	C3
48	ZR** MO*****	3,07612	2,3015	0,8571	C3
49	ZT** NO*****	3,0104	2,4077	1,5649	C3
50	ZV** AN*****	3,25	2,5094	0,9368	C3

Pada proses perhitungan ini, hasil perhitungan yang sama berada di iterasi ke -3 dan iterasi ke -4, hal itu dapat dilihat dari hasil Cluster tiap data tidak berpindah lagi, maka iterasi selanjutnya tidak perlu dilakukan. Hasil dari perhitungan Clustering ini menunjukkan bahwa ada 21 siswa dengan kategori nilai tinggi yang berada di Cluster 1 (C1), 15 siswa dengan kategori nilai sedang yang berada di Cluster 2 (C2), dan 14 siswa dengan kategori rendah yang berada di Cluster 3 (C3). Hasil dari pengelompokan atau hasil dari clustering dapat dilihat pada tabel dibawah ini:

Tabel 3.11 Hasil Cluster 1 (Tinggi)

No	Nama Siswa
1	AB**** TA*****
2	EF** ME*****
3	G**** E**
4	MN*** MA*****
5	ST***** RA*
6	UV*** TH***
7	WX*** AF*****
8	YZ** VI***
9	AB***** PA**
10	CD**** PU***
11	KL** UL***
12	ST*****
13	YZ**** SE*****
14	AB**** AU***
15	YZ**** SE*****
16	AB**** AU***
17	EF***** LO*****
18	MN***** JE***
19	OP**** HA*****
20	QR*** SI*****
21	VB** SA*****

Tabel 3.12 Hasil Cluster 2 (Sedang)

No	Nama Siswa
1	CD*** AS* TO***
2	IJ** EV**
3	KL** HE*****
4	OP*** KH*****
5	EF***** WI***
6	MN* SY*****
7	OP***** SA**
8	UV*** SU**
9	WX***DO***
10	LI**** VI*****
11	KL**** TH*****

12	VZ** TA*****
13	ZD** KI*****
14	ZH** TY*****
15	KL** LM*****

Tabel 3.13 Hasil Cluster 3 (Rendah)

No	Nama Siswa
1	QR** PU*****
2	GH*** YO*****
3	IJ*** HU*****
4	QR*****
5	CD** MA*****
6	IJ***** WA*****
7	ST***** JA*****
8	UV*** TA*****
9	WX***** MA*****
10	ZF***** GI***
11	VP** LU***
12	ZR***** MO*****
13	ZT***** NO***
14	ZV*** AN***

Hasil pengelompokan berdasarkan nilai centroid menunjukkan bahwa Cluster 1 (C1) termasuk dalam kategori rendah, yang menggambarkan siswa dengan tingkat Prestasi rendah. Selanjutnya, Cluster 2 (C2) berada pada kategori sedang, yang menunjukkan tingkat prestasi sedang. Selanjutnya, Cluster 3 (C3) berada pada kategori tinggi, yang menggambarkan siswa dengan tingkat prestasi tinggi. Proses iterasi dihentikan karena tidak terjadi perubahan yang signifikan pada nilai centroid antar iterasi, sehingga hasil klasterisasi dinyatakan telah mencapai kondisi konvergen. Adapun hasil pengelompokan masing-masing cluster disajikan pada Tabel 3.11, Tabel 3.12, dan Tabel 3.13.

3.6 . Penerapan Algoritma Naive Bayes

Algoritma Naive Bayes merupakan metode klasifikasi yang digunakan untuk memprediksi atau menentukan kelas prestasi akademik siswa berdasarkan sejumlah atribut penilaian yang telah ditentukan. Pada penelitian ini, algoritma Naive Bayes digunakan untuk mengklasifikasikan prestasi akademik Bahasa Inggris siswa ke dalam kategori tertentu, seperti rendah, sedang, dan tinggi, berdasarkan data hasil pengelompokan sebelumnya. Proses klasifikasi dilakukan dengan menghitung probabilitas suatu data siswa termasuk ke dalam kelas prestasi tertentu, kemudian memilih kelas dengan nilai probabilitas tertinggi sebagai hasil akhir.

1. Pengumpulan Data:

Dataset yang digunakan sudah terbagi menjadi Training Data dan Testing Data (80:20), yang masing-masing memiliki fitur dari Nilai(X1), kehadiran (X2), Kursus (X3), Minat (X4). dan label yang mengindikasikan prestasi (Rendah, Sedang, Tinggi).

2. Fitur yang digunakan dalam model adalah variabel dari nilai, kehadiran, kursus, dan minat.
3. Target (Label) Target yang ingin diprediksi adalah Label (Rendah, Sedang, Tinggi), yang merupakan prestasi akademik siswa.

Tabel 3.14. Data Atribut Siswa untuk Klasifikasi Prestasi Akademik

No	Nama Siswa	Nilai	Kehadiran%	Kursus	Minat	Label
1	AB**** TA*****	1	1	2	2	Rendah
2	CD**** AS**** TO***	2	1	1	3	Sedang
3	EF** ME*****	1	2	1	1	Rendah
4	G**** E**	1	1	1	2	Rendah

5	IJ** EV**	2	3	2	2	Sedang
6	KL** HE**** LU***	2	1	1	2	Sedang
7	MN*** MA*****	1	3	1	3	Rendah
8	OP*** KH*****	2	2	1	2	Sedang
9	QR** PU*****	3	3	1	3	Tinggi
10	ST***** RA*	1	2	1	2	Rendah
11	UV*** TH***	1	2	1	1	Rendah
12	WX*** AF*****	1	1	2	1	Rendah
13	YZ** VI***	1	2	1	2	Rendah
14	AB***** PA**	1	1	1	3	Rendah
15	CD**** PU***	1	2	2	1	Rendah
16	EF***** WI***	2	2	2	3	Sedang
17	GH*** YO*****	3	2	2	3	Tinggi
18	IJ*** HU*****	3	3	2	2	Tinggi
19	KL** UL***	1	3	1	1	Rendah
20	MN* SY*****	2	3	1	2	Sedang
21	OP**** SA**	2	2	1	3	Sedang
22	QR*****	3	2	2	2	Tinggi
23	ST*****	1	1	2	3	Rendah
24	UV*** SU**	2	2	2	1	Sedang
25	WX*** DO***	2	1	2	1	Sedang
26	YZ**** SE*****	1	3	2	2	Rendah
27	AB**** AU***	1	2	2	2	Rendah
28	CD** MA*****	3	3	1	2	Tinggi
29	EF***** LO****	1	2	2	3	Rendah
30	LI**** VI*****	2	2	2	2	Sedang
31	IJ***** WA****	3	2	1	3	Tinggi
32	KL**** TH****	2	3	2	1	Sedang
33	MN**** JE***	1	3	2	1	Rendah
34	OP**** HA*****	1	2	2	1	Rendah
35	QR*** SI*****	1	1	2	2	Rendah
36	ST***** JA****	2	3	2	3	Tinggi
37	UV*** TA*****	3	3	2	1	Tinggi
38	WX**** MA*****	3	1	2	3	Tinggi
39	VZ** TA*****	2	1	1	3	Sedang
40	VB** SA*****	1	2	1	3	Rendah

Pada Tabel 3.14 menyajikan data atribut siswa yang digunakan sebagai data latih dalam proses klasifikasi prestasi akademik Bahasa Inggris menggunakan algoritma Naive Bayes. Data terdiri dari 40 siswa, dengan masing-masing siswa direpresentasikan oleh lima atribut penilaian, yaitu nilai Bahasa Inggris, tingkat kehadiran, keikutsertaan kursus, minat belajar, serta kategori Label sebagai kelas target. Setiap atribut dinyatakan dalam bentuk data kategorik bernilai 1, 2, dan 3, yang merepresentasikan kategori rendah, sedang, dan tinggi. Berdasarkan data yang tersedia, jumlah siswa dengan prestasi kategori rendah (1) sebanyak 20 siswa, kategori sedang (2) sebanyak 10 siswa, dan kategori tinggi (3) sebanyak 10 siswa.

Tabel 3.15. Data Uji Siswa untuk Klasifikasi Prestasi Akademik

N0	Nama siswa	X1	X2	X3	X4	Label
41	ZD** KI*****	3	1	1	3	Tinggi
42	ZF** GI*****	3	2	1	2	Tinggi
43	ZH** TY*****	2	3	1	3	Tinggi
44	VJ** KU*****	1	3	1	2	Rendah
45	KL** LM*****	2	1	2	2	Sedang
46	VN*** DO*****	3	2	2	1	Tinggi
47	VP** LU*****	3	3	2	2	Tinggi
48	ZR** MO*****	3	2	2	3	Tinggi
49	ZT** NO*****	3	2	1	1	Tinggi
50	ZV** AN*****	3	3	2	3	Tinggi

Pada Tabel 3.15 menampilkan data uji yang terdiri dari 10 siswa dengan nomor urut 41 hingga 50 yang digunakan dalam proses klasifikasi prestasi akademik Bahasa Inggris menggunakan algoritma Naive Bayes. Berdasarkan data tersebut, diperoleh hasil bahwa 8 siswa (80%) termasuk dalam kategori prestasi tinggi, 1 siswa (10%) berada pada kategori sedang, dan 1 siswa (10%) berada

pada kategori rendah. Atribut nilai Bahasa Inggris didominasi oleh kategori tinggi dengan 7 siswa memiliki nilai kategori 3, sementara atribut kehadiran dan minat belajar menunjukkan variasi yang relatif seimbang. Data uji ini digunakan untuk menguji kemampuan model Naive Bayes dalam memprediksi prestasi akademik siswa secara akurat dan konsisten berdasarkan pola data yang telah dipelajari sebelumnya.

Selanjutnya menentukan Probabilitas Prior untuk masing masing kelas (prestasi rendah, sedang dan tinggi) Ini dihitung berdasarkan distribusi kelas dalam dataset pelatihan:

$$P(C_k) = \frac{\text{Jumlah data pada kelas } C_k}{\text{Jumlah seluruh data}}$$

Keterangan

$P(C_k)$ = prior probability untuk kelas ke-k

Jumlah data pada kelas C_k = frekuensi kemunculan kelas tersebut

Jumlah seluruh data = total data latih.

Tabel 3.16 Probabilitas Prior Kelas

Probabilitas Prior	
Label	Nilai
Rendah	0,4
Sedang	0,32
Tinggi	0,28

Tabel 3.16 menyajikan nilai probabilitas prior pada algoritma Naive Bayes untuk tiga kelas prestasi akademik, yaitu rendah, sedang, dan tinggi. Berdasarkan data yang ditampilkan, probabilitas prior untuk kelas prestasi rendah sebesar 0,40 atau 40%, kelas sedang sebesar 0,32 atau 32%, dan kelas tinggi sebesar 0,28 atau

28%. Nilai probabilitas ini diperoleh dari proporsi jumlah data latih pada masing-masing kelas prestasi. Probabilitas prior digunakan sebagai dasar awal dalam perhitungan Naive Bayes untuk menentukan probabilitas posterior, sehingga berpengaruh langsung terhadap hasil klasifikasi prestasi akademik siswa.

Selanjutnya adalah melakukan perhitungan likelihood untuk setiap fitur yang digunakan dalam penelitian, yaitu X1, X2, X3, dan X4. Perhitungan likelihood bertujuan untuk mengetahui peluang kemunculan masing-masing nilai atribut pada setiap kelas prestasi akademik. Nilai likelihood diperoleh berdasarkan frekuensi kemunculan data pada setiap fitur terhadap masing-masing kelas, yang selanjutnya digunakan dalam proses perhitungan probabilitas posterior pada algoritma Naive Bayes.

Tabel 3.17 Probabilitas Likelihood Setiap Fitur Kelas Prestasi Akademik

likelihood untuk setiap fitur kelas				
Featur	Value	Rendah	Sedang	Tinggi
Nilai B. Inggris	1	1	0	0
Nilai B. Inggris	2	0	0,875	0,071428571
Nilai B. Inggris	3	0	0,125	0,928571429
Kehadiran%	1	0,3	0,375	0,071428571
Kehadiran%	2	0,45	0,375	0,428571429
Kehadiran%	3	0,25	0,25	0,5
Kursus	1	0,5	0,5	0,357142857
Kursus	2	0,5	0,5	0,642857143
Kursus	3	0	0	0
Minat Belajar	1	0,35	0,25	0,142857143
Minat Belajar	2	0,4	0,375	0,357142857
Minat Belajar	3	0,25	0,375	0,5

Pada tabel diatas menyajikan hasil perhitungan probabilitas likelihood untuk setiap fitur terhadap kelas prestasi akademik, yaitu rendah, sedang, dan tinggi, yang digunakan dalam algoritma Naive Bayes. Fitur yang dianalisis

meliputi nilai Bahasa Inggris, persentase kehadiran, keikutsertaan kursus, dan minat belajar, dengan masing-masing fitur memiliki beberapa nilai kategorik. Berdasarkan tabel tersebut, dapat diketahui bahwa pada fitur nilai Bahasa Inggris, kategori nilai 3 memiliki probabilitas tertinggi pada kelas prestasi tinggi, yaitu sebesar 0,93, sedangkan nilai 2 lebih dominan pada kelas sedang dengan probabilitas 0,88. Pada fitur kehadiran, nilai kategori 3 menunjukkan probabilitas tertinggi pada kelas tinggi sebesar 0,50. Fitur kursus menunjukkan bahwa nilai 2 memiliki kecenderungan lebih besar pada prestasi tinggi dengan probabilitas 0,64. Sementara itu, pada fitur minat belajar, kategori 3 memiliki probabilitas tertinggi pada kelas prestasi tinggi sebesar 0,50. Nilai-nilai likelihood ini digunakan sebagai komponen utama dalam perhitungan probabilitas posterior untuk menentukan klasifikasi prestasi akademik siswa secara akurat.

Kemudian Terapkan Teorema Bayes untuk menghitung probabilitas posterior untuk masing-masing kelas. Formula umum:

$$P(C | X) = \frac{P(C) \prod P(x_i | C)}{P(X)}$$

Di mana:

$P(C | X)$ adalah probabilitas posterior kelas C diberikan data X ,

$P(C)$ adalah prior probabilitas kelas, $P(x_i | C)$ adalah likelihood dari fitur x

$P(X)$ adalah normalisasi yang tidak mempengaruhi perbandingan antar kelas.

Tabel 3.18 Hasil Perhitungan Probabilitas Posterior pada 10 Data Testing

Tabel Posterior Dengan 10 Data Testing						
label	prior	Nilai	Kehadiran %	Kursus	Minat Belajar	Posterior
		3	1	1	3	
Rendah	0,4	0	0,3	0,5	0,25	0
Sedan	0,32	0,125	0,375	0,5	0,375	0,0028125
Tinggi	0,28	0,928571429	0,071428571	0,357142857	0,5	0,003316327
		3	2	1	2	
Rendah	0,4	0	0,45	0,5	0,4	0
Sedan	0,32	0,125	0,375	0,5	0,375	0,0028125
Tinggi	0,28	0,928571429	0,428571429	0,357142857	0,357142857	0,014212828
		2	3	1	3	
Rendah	0,4	0	0,25	0,5	0,25	0
Sedan	0,32	0,875	0,25	0,5	0,375	0,013125
Tinggi	0,28	0,071428571	0,5	0,357142857	0,5	0,001785714
		1	3	1	2	
Rendah	0,4	1	0,25	0,5	0,4	0,02
Sedan	0,32	0	0,25	0,5	0,375	0
Tinggi	0,28	0	0,5	0,357142857	0,357142857	0
		2	1	2	2	
Rendah	0,4	0	0,3	0,5	0,4	0
Sedan	0,32	0,875	0,375	0,5	0,375	0,0196875
Tinggi	0,28	0,071428571	0,071428571	0,642857143	0,357142857	0,000327988
		3	2	2	1	
label	prior	3	2	2	1	Posterior

	r					
Rendah	0,4	0	0,45	0,5	0,35	0
Sedan	0,32	0,125	0,375	0,5	0,25	0,001875
Tinggi	0,28	0,9285714 29	0,42857142 9	0,6428571 43	0,14285714 3	0,0102332 36
label	prior	3	3	2	2	Posterior
Rendah	0,4	0	0,25	0,5	0,4	0
Sedan	0,32	0,125	0,25	0,5	0,375	0,001875
Tinggi	0,28	0,9285714 29	0,5	0,6428571 43	0,35714285 7	0,0298469 39
label	prior	3	2	2	3	Posterior
Rendah	0,4	0	0,45	0,5	0,25	0
Sedan	0,32	0,125	0,375	0,5	0,375	0,0028125
Tinggi	0,28	0,9285714 29	0,42857142 9	0,6428571 43	0,5	0,0358163 27
label	prior	3	2	1	1	Posterior
Rendah	0,4	0	0,45	0,5	0,35	0
Sedan	0,32	0,125	0,375	0,5	0,25	0,001875
Tinggi	0,28	0,9285714 29	0,42857142 9	0,3571428 57	0,14285714 3	0,0056851 31
label	prior	3	3	2	3	Posterior
Rendah	0,4	0	0,25	0,5	0,25	0
Sedan	0,32	0,125	0,25	0,5	0,375	0,001875
Tinggi	0,28	0,9285714 29	0,5	0,6428571 43	0,5	0,0417857 14

Pada tabel diatas menyajikan hasil perhitungan probabilitas posterior pada 10 data testing menggunakan algoritma Naive Bayes berdasarkan empat fitur,

yaitu nilai Bahasa Inggris, persentase kehadiran, keikutsertaan kursus, dan minat belajar, dengan tiga kelas prestasi akademik yaitu rendah, sedang, dan tinggi. Setiap data testing dihitung nilai posterior-nya dengan mengalikan probabilitas prior dan probabilitas likelihood dari masing-masing fitur. Berdasarkan hasil perhitungan, nilai posterior terbesar pada sebagian besar data testing berada pada kelas prestasi tinggi, yang menunjukkan bahwa mayoritas data uji memiliki kecenderungan prestasi akademik yang tinggi. Nilai posterior tertinggi yang diperoleh adalah sebesar 0,0418, sedangkan nilai posterior terendah bernilai 0, yang terjadi ketika salah satu probabilitas likelihood bernilai nol. Dari 10 data testing, sebanyak 8 data diklasifikasikan ke dalam kelas tinggi, 1 data ke dalam kelas sedang, dan 1 data ke dalam kelas rendah. Hasil ini menunjukkan bahwa model Naive Bayes mampu mengklasifikasikan data testing secara konsisten berdasarkan pola data latih yang digunakan.

3.6.1 Evaluasi Model

Setelah model diterapkan pada data testing, buat confusion matrix untuk menghitung True Positives (TP), False Positives (FP), True Negatives (TN), dan False Negatives (FN).

Metrik Evaluasi:

Akurasi: Proporsi prediksi yang benar terhadap jumlah total data.

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Total\ Data}$$

Presisi: Mengukur ketepatan model dalam mengklasifikasikan kelas positif.

$$Presisi = \frac{TP}{TP+FP}$$

Recall: Mengukur kemampuan model dalam menemukan semua data yang relevan

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: Kombinasi antara presisi dan recall untuk mendapatkan ukuran kinerja yang seimbang.

$$F1 - Score = 2X \frac{Presisi \times Recall}{Presisi + Recall}$$

Tabel 3.19 Perbandingan Label Asli dan Label Prediksi Hasil Klasifikasi Naïve Bayes pada Data Testing (Data Uji)

N0	Label asli	Label prediksi
41	Tinggi	Tinggi
42	Tinggi	Tinggi
43	Tinggi	Tinggi
44	Rendah	Rendah
45	Sedang	Sedang
46	Tinggi	Tinggi
47	Tinggi	Tinggi
48	Tinggi	Tinggi
49	Tinggi	Tinggi
50	Tinggi	Tinggi

Pada tabel diatas menunjukkan perbandingan antara label asli dan label prediksi pada 10 data testing siswa yang diklasifikasikan menggunakan algoritma Naive Bayes. Berdasarkan hasil pengujian, seluruh data testing menghasilkan label prediksi yang sesuai dengan label aslinya. Secara kuantitatif, hal ini menunjukkan bahwa tingkat akurasi klasifikasi mencapai 100%, dengan 10 data terklasifikasi benar dan tidak ditemukan kesalahan klasifikasi. Dari keseluruhan data testing tersebut, terdapat 8 siswa dengan prestasi akademik kategori tinggi, 1

siswa kategori sedang, dan 1 siswa kategori rendah, yang seluruhnya berhasil diprediksi dengan tepat. Hasil ini menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan prestasi akademik siswa secara akurat berdasarkan atribut yang digunakan.

Tabel 3.20 Confusion Matrix Hasil Klasifikasi Prestasi Akademik Siswa

Aktual \ Prediksi	Rendah	Sedang	Tinggi
Tinggi	0	0	8
Sedang	0	1	0
Rendah	1	0	0

Pada tabel diatas menyajikan confusion matrix yang menunjukkan perbandingan antara label aktual dan label prediksi pada hasil klasifikasi prestasi akademik siswa menggunakan algoritma Naive Bayes. Berdasarkan tabel tersebut, seluruh data testing berhasil diklasifikasikan dengan benar, di mana 8 data dengan label aktual tinggi diprediksi sebagai tinggi, 1 data dengan label sedang diprediksi sebagai sedang, dan 1 data dengan label rendah diprediksi sebagai rendah. Secara kuantitatif, hasil ini menunjukkan bahwa tidak terdapat kesalahan klasifikasi pada ketiga kelas prestasi akademik. Dengan demikian, nilai akurasi, precision, dan recall yang dihasilkan masing-masing mencapai 100%, yang menandakan bahwa model Naive Bayes memiliki performa yang sangat baik dalam mengklasifikasikan prestasi akademik siswa berdasarkan atribut yang digunakan.

Tabel 3.21 Hasil Evaluasi Kinerja Model Naive Bayes

Metrik	Nilai	persentase
Accuracy	1	100%
Precision	1	100%
Recall	1	100%
F1-Score	1	100%

Tabel diatas menyajikan hasil evaluasi kinerja algoritma Naive Bayes berdasarkan beberapa metrik pengujian, yaitu accuracy, precision, recall, dan F1-score. Berdasarkan data yang ditampilkan, seluruh metrik evaluasi memperoleh nilai 1 atau setara dengan 100%, yang menunjukkan bahwa model mampu mengklasifikasikan seluruh data testing dengan sangat baik. Secara kuantitatif, nilai accuracy sebesar 100% menandakan tidak adanya kesalahan klasifikasi pada 10 data uji. Nilai precision dan recall yang sama-sama mencapai 100% menunjukkan bahwa model memiliki kemampuan yang sangat tinggi dalam memprediksi kelas prestasi akademik secara tepat dan konsisten. Sementara itu, nilai F1-score sebesar 100% mengindikasikan keseimbangan yang optimal antara precision dan recall. Dengan distribusi data uji yang terdiri dari 1 kelas Rendah, 1 kelas Sedang, dan 8 kelas Tinggi, pencapaian metrik ini menunjukkan kinerja model yang sangat stabil di seluruh kategori.