

**ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK
MEMPREDIKSI HASIL BELAJAR MAHASISWA
MENGUNAKAN METODE NAÏVE BAYES PADA FAKULTAS
TEKNIK PRODI TEKNOLOGI INFORMASI UNIVERSITAS
MUHAMMADIYAH PALEMBANG**



TESIS

MUHAMMAD FEBRIAN

232420022

ENTERPRISE IT INFRASTRUCTURE

PROGRAM STUDI TEKNIK INFORMATIKA – S2

PROGRAM PASCASARJANA

UNIVERSITAS BINA DARMA PALEMBANG

2026

**ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK
MEMPREDIKSI HASIL BELAJAR MAHASISWA
MENGUNAKAN METODE NAÏVE BAYES PADA FAKULTAS
TEKNIK PRODI TEKNOLOGI INFORMASI UNIVERSITAS
MUHAMMADIYAH PALEMBANG**



**Tesis ini diajukan sebagai salah satu syarat
untuk memperoleh gelar**

MAGISTER KOMPUTER

Muhammad Febrian

232420022

ENTERPRISE IT INFRASTRUCTURE

PROGRAM STUDI TEKNIK INFORMATIKA S2

PROGRAM PASCASARJANA

UNIVERSITAS BINA DARMA

2026

Halaman Pengesahan Pembimbing Tesis

Judul Tesis: ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK
MEMPREDIKSI HASIL BELAJAR MAHASISWA MENGGUNAKAN
METODE NAIVE BAYES PADA FAKULTAS TEKNIK PRODI
TEKNOLOGI INFORMASI UNIVERSITAS MUHAMMADIYAH
PALEMBANG

Oleh Muhammad Febrian , NIM 232420022 Tesis ini telah disetujui dan disahkan oleh
Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE IT
INFRASTRUCTURE, Program Pascasarjana Universitas Bina Darma pada 05 Maret 2026
dan telah dinyatakan LULUS.

Palembang, 05 Maret 2026
Mengetahui,
Program Studi Teknik Informatika – S2
Universitas Bina Darma

Ketua,



Dr. Usman Ependi, S.Kom., M.Kom.

Pembimbing,



Dr. Tata Sutabri, S.Kom., M.M.S.I

Halaman Pengesahan Penguji Tesis

Judul Tesis: ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK
MEMPREDIKSI HASIL BELAJAR MAHASISWA MENGGUNAKAN
METODE NAIVE BAYES PADA FAKULTAS TEKNIK PRODI
TEKNOLOGI INFORMASI UNIVERSITAS MUHAMMADIYAH
PALEMBANG

Oleh Muhammad Febrian , NIM 232420022 Tesis ini telah disetujui dan disahkan oleh
Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE IT
INFRASTRUCTURE, Program Pascasarjana Universitas Bina Darma pada 05 Maret 2026
dan telah dinyatakan LULUS.

Palembang, 05 Maret 2026
Mengetahui,
Program Pascasarjana
Universitas Bina Darma
Direktur,



PROGRAM PASCASARJANA

Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Penguji I,



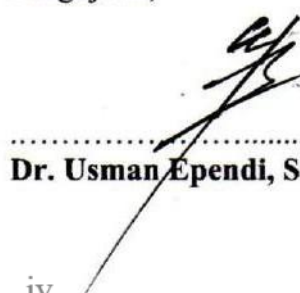
Dr. Tata Sutabri, S.Kom., M.M.S.I.

Penguji II,



Prof. Dr. Edi Surya Negara, M.Kom.

Penguji III,



Dr. Usman Ependi, S.Kom., M.Kom.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : MUHAMMAD FEBRIAN

NIM : 232410024

Dengan ini menyatakan bahwa:

1. Karya tulis Saya (Tesis) ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik baik (Magister, Sarjana, dan Ahli Madya) di Universitas Bina Darma;
2. Karya tulis ini murni gagasan, rumusan dan penelitian Saya sendiri dengan arahan tim pembimbing;
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar pustaka;
4. Karena yakin dengan keaslian karya tulis ini, Saya menyatakan bersedia Tesis/Skripsi/Tugas Akhir, yang Saya hasilkan di unggah ke internet;
5. Surat Pernyataan ini Saya tulis dengan sungguh-sungguh dan apabila terdapat penyimpangan atau ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima sanksi dengan aturan yang berlaku di perguruan tinggi ini.

Demikian Surat Pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 05 Maret 2026

Yang Membuat Pernyataan,



MUHAMMAD FEBRIAN

NIM : 232410024

ABSTRAK

Penelitian ini bertujuan menganalisis dan mengimplementasikan metode data mining menggunakan algoritma *Naïve Bayes* dalam memprediksi hasil belajar mahasiswa pada Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang. Latar belakang penelitian ini didasarkan pada pemanfaatan data akademik yang belum optimal dalam mendukung pengambilan keputusan strategis di lingkungan perguruan tinggi. Data yang digunakan berasal dari Sistem Informasi Akademik (SIKAD) serta hasil kuesioner mahasiswa, dengan total dataset sekitar 200 mahasiswa. Variabel yang dianalisis meliputi nilai mata kuliah, Indeks Prestasi Kumulatif (IPK), tingkat kehadiran, dan jumlah SKS. Proses penelitian dilakukan melalui tahapan pengumpulan data, preprocessing (data cleaning, transformasi, dan reduksi), pembagian data training dan testing dengan rasio 80:20, pemodelan menggunakan algoritma *Naïve Bayes*, serta evaluasi menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan F1-score. Hasil penelitian menunjukkan bahwa model *Naïve Bayes* mampu memberikan performa klasifikasi yang baik dengan tingkat akurasi sebesar 82%, *precision* 85%, *recall* 82%, dan F1-score 83%. Model lebih akurat dalam mengklasifikasikan mahasiswa kategori “Sangat Aktif” dibandingkan kategori “Kurang Aktif”. Temuan ini menunjukkan bahwa metode *Naïve Bayes* efektif digunakan untuk memprediksi hasil belajar mahasiswa dan dapat dimanfaatkan sebagai sistem pendukung keputusan dalam mengidentifikasi mahasiswa yang memerlukan perhatian akademik lebih lanjut.

Kata Kunci : *Data Mining*, *Naïve Bayes*, Hasil Belajar Mahasiswa, Prediksi Akademik, *Confusion Matrix*, Sistem Pendukung Keputusan

ABSTRACT

This study aims to analyze and implement data mining methods using the Naïve Bayes algorithm in predicting student learning outcomes at the Faculty of Engineering, Information Technology Study Program, University of Muhammadiyah Palembang. The background of this research is based on the use of academic data that is not optimal in supporting strategic decision-making in the university environment. The data used came from the Academic Information System (SIKAD) and the results of student questionnaires, with a total dataset of around 200 students. The variables analyzed included course scores, Cumulative Grade Point Average (GPA), attendance rate, and number of credits. The research process was carried out through the stages of data collection, preprocessing (data cleaning, transformation, and reduction), data sharing training and testing with an 80:20 ratio, modeling using the Naïve Bayes algorithm, and evaluation using confusion matrix, accuracy, precision, recall, and F1-score. The results showed that the Naïve Bayes model was able to provide good classification performance with an accuracy level of 82%, accuracy of 85%, recall of 82%, and an F1-score of 83%. The model is more accurate in classifying students in the "Very Active" category than in the "Less Active" category. These findings show that the Naïve Bayes method is effectively used to predict student learning outcomes and can be used as a decision support system in identifying students who need further academic attention.

Keywords : Data Mining, Naïve Bayes, Student Learning Outcomes, Academic Predictions, Confusion Matrix, Decision Support Systems

MOTTO DAN PERSEMBAHAN

MOTTO :

1. Sukses tidak datang dari apa yang diberikan oleh orang lain, Melainkan datang dari keyakinan dan kerja keras diri kita sendiri.
2. Usaha Keras itu tak akan mengkhianati Hasil.
3. Semua impian kita dapat menjadi nyata, jika kita memiliki keberanian untuk mengejanya.
4. Jangan menunda – nunda waktu anda biar pekerjaan anda bisa lancar.
5. Gunakan waktu anda dengan baik.
6. Tetap Semangat Walaupun badai selalu datang pada waktunya.
7. Kejarlah pendidikan setinggi langit.

PERSEMBAHAN

- Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya yang tak terhingga.
- Baginda Nabi Muhammad Shallallahu alaihi wasallam.
- Kedua Orang Tua saya yang selalu support Dan Memotivasiku.
- Saudara – Saudaraku yang selalu memotivasiku.
- Dosen Pembimbing Dr. Tata Sutabri, S.Kom., M.M.S.I. terima kasih atas bimbingannya selama ini.
- Teman-teman seangkatan 2023
- Almamater Universitas Bina Darma.
- Terima kasih kepada teman teman ku Magister Teknik Informatika, Terima kasih kepada Mamang ku Karnadi, S.Kom., M.Kom turut mensupport saya untuk menyelesaikan Laporan tesis saya.
- Pada Teman dan Rekan-rekan kerja yang mendukung dan membantu Penulis penyelesaian S2 ini.

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas berkat rahmatnya dan karunia-Nya skripsi saya dapat selesai dengan baik untuk memenuhi salah satu syarat untuk mendapatkan gelar Megister Komputer di Pascasarjana Universitas Bina Darma Palembang

Dalam penulisan tesis ini, tentunya masih jauh dari sempurna. Hal ini dikarenakan keterbatasannya pengetahuan yang dimiliki. Oleh karena itu dalam rangka melengkapi kesempurnaan dari penulis skripsi ini diharapkan adanya saran dan kririk yang diberikan bersifat membangun.

Pada kesempatan yang baik ini, tak lupa penulis menghaturkan terima kasih kepada semua pihak yang telah memberi bimbingan, pengarahan, nasehat dan pemikiran dalam penulisan skripsi ini, terutama kepada :

1. Dr. Sunda Ariana, M.Pd., M.M. selaku Rektor Universitas Bina Darma Palembang.
2. Prof. Dr. Ir. Achmad Syarifudin, M.Sc. selaku Direktur Pascasarjana Universitas Bina Darma Palembang
3. Dr. Usman Ependi, S.Kom.,M.Kom ketua Program Studi Megister Teknik Informatika.
4. Dr. Tata Sutabri, S.Kom., M.M, S.I selaku dosen pembimbing yang telah memberikan bimbingan dalam penulisan Tesis.
5. Prof Dr. Edi Surya Negara, S.Kom., M.Kom selaku penguji satu saya yang telah memberikan masukan dan saran dalam penulisan tesis.
6. Segenap Dosen Magister Teknik Informatika Universitas Bina Darma Palembang yang telah memberikan ilmu pengetahuan dan bimbingannya.
7. Orang tua tercinta, terima kasih atas doanya dan serta cinta yang telah diberikan.
8. Terima kasih kepada Mamang ku Karnadi, S.Kom., M.Kom turut mensupport saya untuk menyelesaikan Laporan tesis saya.
9. Keluarga besar Universitas Bina Darma Palembang khususnya teman-teman seperjuangan pada Prodi Magister teknik informatika atas semua

dukungan, semangat, serta kerja samanya.

10. Teman – teman Magister Teknik Informatika angkatan 2023 tercinta
11. Pada Teman dan Rekan-rekan kerja yang mendukung dan membantu Penulis menyelesaikan S2 ini.
12. Penulis menyadari penulisan tesis ini tidak luput dari berbagai kekurangan. Penulis mengharapkan saran dan kritik demi kesempurnaan dan perbaikannya sehingga akhirnya laporan skripsi ini dapat memberikan manfaat bagi bidang pendidikan dan bisa dikembangkan lebih lanjut lagi. Aamiin.

Akhir kata dari penulisan menyadari bahwa penulisan tesis ini belum mencapai kesempurnaan dalam arti yang sebenarnya. Namun penulis berharap penulisan tesis ini dapat bermanfaat bagi penulis sendiri dan para pembaca pada umumnya.

Palembang,

Muhammad Febrian

DAFTAR ISI

HALAMAN DEPAN	i
HALAMAN DEPAN TESIS	ii
HALAMAN PENGESAHAN PEMBIMBING TESIS	iii
HALAMAN PENGESAHAN UJIAN TESIS....	iv
SURAT PERNYATAAN	v
ABSTRAK.....	vi
<i>ABSTRACT</i>	vii
MOTTO DAN PERSEMBAHAN	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi masalah.....	3
1.3 Perumusan Masalah.....	4
1.4 Tujuan penelitian	4
1.5 Kebaruan	4
1.6 Manfaat penelitian	6
1.7 Pembatasan masalah.....	6
1.8 Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	9
2.1 Tinjauan Pustaka.....	9
2.1.1 <i>Data Maining</i>	9
2.2 Educational Data Maining.....	11

2.3	Hasil Belajar Mahasiswa	12
2.4	<i>Naïve Bayes Classifier</i>	13
2.5	Confusion Matrix	14
2.6	Penelitian Terdahulu.....	16
2.7	Kerangka kerja yang digunakan	21
2.7.1	Studi literatur.....	21
2.7.2	Identifikasi masalah dan tujuan.....	21
2.7.3	Pengumpulan data	21
2.7.4	Implementasi algoritma naïve bayes	21
2.7.5	Evaluasi Kinerja Sistem	21
2.7.6	Analisis dan Pembahasan Hasil.....	22
2.8	Google colabs.....	22
BAB III METODOLOGI.....		24
3.1	Metode Penelitian.....	24
3.2	Waktu dan tempat penelitian	24
3.2.1	Tempat penelitian	24
3.2.2	Waktu penelitian	25
3.3	Prosedur Penelitian.....	25
3.3.1	Pengumpulan data	26
3.3.2	Preprocessing Data	26
3.3.3	Pemodelan (Modeling)	27
3.3.4	Evaluasi Model.....	28
3.3.5	Analisis dan Interpretasi Hasil	28
3.4	Variabel Input Berdasarkan Data Kuesioner	28
3.4.1	Tingkat Kehadiran Mahasiswa	29
3.4.2	Distribusi Responden Berdasarkan Jenis Kelamin.....	29
3.4.3	Indeks Prestasi Kumulatif (IPK) Mahasiswa.....	30

BAB IV HASIL DAN PEMBAHASAN	32
4.1 Hasil	32
4.2 Implementasi Naïve Bayes	33
4.3 Preprocessing Data	34
4.3.1 Data Cleaning (Pembersihan Data)	34
4.3.2 Data Transformation	35
4.4 Preprocessing Data	38
4.4.1 Import Data	39
4.4.2 Seleksi Atribut	39
4.4.3 Data cleaning	39
4.4.4 Transformasi Data	40
4.4.5 Encoding Data	41
4.4.6 Pembentukan Dataset Akhir	41
4.5 Evaluasi Kinerja Model Berdasarkan Hasil Kuesioner	41
4.5.1 Tingkat Kehadiran Mahasiswa	42
4.5.2 Distribusi IPK Responden	43
4.5.3 Tingkat Motivasi Belajar Mahasiswa	44
4.5.4 Persepsi Mahasiswa terhadap Penggunaan Data Akademik untuk Prediksi.....	44
4.5.5 Keterkaitan Hasil Kuesioner dengan Kinerja Model	45
4.6 Confusion Matrix	46
4.7 Hasil Feature Importanc	48
4.8 Performa Model.....	49
4.9 Distribusi Kelas	51
4.10 5-Fold Cross-Validation	53
4.11 Hasil Naïve Bayes	54
4.12 Hasil Distribusi Akurasi dan Akurasi per Eksperimen	56

4.13	Distribusi Setiap Fitur Input	58
BAB IV PENUTUP.....		61
5.1	Kesimpulan.....	61
5.2	Saran.....	61
DAFTAR PUSTAKA.....		62
LAMPIRAN		64



DAFTAR GAMBAR

Gambar 2.1 Proses data mining atau <i>Knowledge Discovery from Data</i>	11
Gambar 4.1 <i>Confusion Matrix</i>	46
Gambar 4.2 <i>Feature Importance</i>	48
Gambar 4.3 Performa Model	50
Gambar 4.4 Distribusi Kelas.....	52
Gambar 4.5 <i>5-Fold Cross-Validation</i>	53
Gambar 4.6 Hasil <i>Naïve Bayes</i>	55
Gambar 4.7 Hasil Distribusi akurasi dan akurasi eksperimen	57
Gambar 4.8 Distribusi Setiap Fitur Input.....	59

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	15
Tabel 2.2 Akurasi	15
Tabel 2.3 Presisi	15
Tabel 2.4 Recall	16
Tabel 2.5 Penelitian terdahulu.....	16
Tabel 3.1 Waktu Penelitian	25
Tabel 3.2 Tingkat kehadiran mahasiswa	29
Tabel 3.3 Distribusi Responden Berdasarkan Jenis Kelamin	30
Tabel 3.4 Distribusi IPK Mahasiswa	30
Tabel 4.1 Ringkas distribusi data	33
Tabel 4.2 Nilai Akademik	35
Tabel 4.3 Transformasi Kehadiran Mahasiswa	36
Tabel 4.4 Encoding Skala Likert.....	37
Tabel 4.5 Encoding Tingkat Aktivitas.....	37
Tabel 4.6 Transformasi Data.....	40
Tabel 4.7 Transformasi IPK.....	40
Tabel 4.8 Kategori Motivasi	41
Tabel 4.9 Tingkat Kehadiran Mahasiswa	42
Tabel 4.10 Distribusi IPK Mahasiswa	43
Tabel 4.11 Tingkat Motivasi Belajar Mahasiswa	44
Tabel 4.12 Persepsi Mahasiswa terhadap Penggunaan Data Akademik	45

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital, banyak aspek kehidupan telah berubah, termasuk pendidikan tinggi. Mengadopsi sistem informasi pembelajaran dan manajemen yang terkomputerisasi telah menghasilkan akumulasi besar data akademik. Pada umumnya, data yang mencakup rekam jejak akademik siswa, termasuk nilai mata kuliah, tingkat kehadiran, Indeks Prestasi Kumulatif (IPK), dan interaksi dalam platform pembelajaran elektronik, hanya disimpan sebagai catatan administratif atau dilaporkan secara berkala. Fenomena ini menunjukkan bahwa ada perbedaan antara mendapatkan banyak data dan mendapatkan pengetahuan strategis, yang dapat membantu pengambilan keputusan di perguruan tinggi. Padahal, pola-pola, korelasi, dan informasi penting lainnya yang dapat dioptimalkan tersembunyi di balik akumulasi data.

Teori data mining muncul sebagai hasil dari kebutuhan untuk mengubah data mentah menjadi pengetahuan yang berguna. Data mining adalah kumpulan prosedur yang digunakan untuk mengekstraksi pola-pola menarik dan informasi baru dari kumpulan data yang sangat besar (Carudin et al., 2024). Dalam proses ini, teknik dari berbagai bidang digabungkan, seperti statistika, kecerdasan buatan, dan pembelajaran mesin. Setelah penggunaan data mining di bidang pendidikan, ada bidang baru yang disebut Educational Data Mining (EDM). EDM berfokus pada metode untuk memahami perilaku belajar siswa, memprediksi prestasi akademik mereka, meningkatkan efektivitas pembelajaran, dan memberikan saran kepada manajemen sekolah (Witten et al., 2016).

Fakultas Teknik Universitas Muhammadiyah Palembang, Program Studi Teknologi Informasi, selalu berusaha untuk meningkatkan proses dan hasil pendidikan. Hasil belajar siswa, yang merupakan ukuran utama keberhasilan proses pendidikan, menjadi fokus utama. Selain pencapaian IPK, perkembangan

kompetensi kognitif, afektif, dan psikomotorik selama studi ditunjukkan sebagai hasil belajar. Hasil belajar siswa dipengaruhi oleh banyak faktor kompleks yang saling berhubungan. Faktor-faktor ini dapat dibagi menjadi kategori internal, seperti motivasi belajar, minat, dan gaya belajar; kategori eksternal, seperti metode pengajaran guru, dukungan keluarga, dan lingkungan sosial; dan kategori pendukung teknologi, seperti ketersediaan informasi dan pemanfaatan media pembelajaran digital (Bloom, 1956).

Salah satu langkah strategis yang harus diambil adalah mengidentifikasi komponen utama yang memengaruhi hasil belajar siswa dan membuat model yang dapat memprediksi potensi akademik mereka. Ini memungkinkan lembaga untuk melakukan intervensi dini yang tepat sasaran kepada siswa yang diprediksi akan menurunkan prestasi. Untuk mencapai tujuan tersebut, pendekatan Knowledge Discovery in Database (KDD) menawarkan kerangka kerja sistematis yang dimulai dengan pemilihan data, pemrosesan awal (preprocessing), transformasi, dan interpretasi pola.

Naive Bayes Classifier adalah salah satu algoritma data mining yang paling banyak digunakan untuk tugas klasifikasi dan prediksi. Teorema Bayes adalah dasar algoritma ini, yang menggunakan data yang telah diamati untuk menghitung kemungkinan suatu kejadian akan terjadi di masa depan. Metode ini memiliki banyak keunggulan, terutama karena sangat sederhana untuk dikomputerisasi, menganggap bahwa ada independensi antar fitur yang kuat (naif), dan dapat menghasilkan akurasi yang kompetitif. Keunggulan ini terutama berlaku untuk dataset dengan struktur yang sederhana dan ukuran menengah (Rish, 2001; Handayani & Pribadi, 2015). Naive Bayes telah terbukti berguna dalam akademik untuk memprediksi hasil studi, kinerja akademik, dan kelulusan siswa (Rahman & Susanto, 2020; Putri & Nugroho, 2021). Terlepas dari fakta bahwa algoritma Naive Bayes telah banyak digunakan dalam penelitian prediksi akademik, hanya ada beberapa penelitian yang secara khusus melihat bagaimana algoritma ini dapat diterapkan pada mahasiswa Fakultas Teknik Program Studi Teknologi Informasi di Universitas Muhammadiyah Palembang.

Pola dan faktor determinan yang berbeda dapat dihasilkan oleh institusi, kurikulum, dan demografi siswa.

Oleh karena itu, sebuah kajian empiris yang komprehensif diperlukan untuk menganalisis dan menerapkan metode data mining, terutama algoritma Naive Bayes, untuk memprediksi hasil belajar mahasiswa dalam konteks tertentu. Dengan menggunakan data awal yang dikumpulkan melalui kuesioner, serta data dari Sistem Informasi Akademik (SIKAD), penelitian ini bertujuan untuk mengisi kekosongan tersebut. Faktor-faktor non-akademik seperti motivasi untuk belajar, tingkat keaktifan, dan persepsi dukungan lingkungan dapat digali melalui data kuesioner. Faktor-faktor ini mungkin tidak tercatat dalam basis data formal. Diharapkan model prediksi yang lebih luas dan akurat dapat dibuat dengan menggabungkan kedua sumber data ini. Hasil analisis tidak hanya akan menghasilkan model klasifikasi dengan metrik evaluasi seperti akurasi, presisi, dan recall, tetapi juga akan mengidentifikasi faktor yang paling penting melalui analisis fitur (feature importance). Penelitian ini diharapkan dapat memberikan rekomendasi berbasis data kepada fakultas dan pengelola program studi untuk meningkatkan kualitas pembelajaran dan untuk membuat sistem pendukung keputusan akademik yang lebih strategis dan responsif.

1.2 Identifikasi masalah

Penelitian ini berangkat dari sejumlah permasalahan yang dihadapi, di antaranya:

A. Pengelolaan data akademik

Data akademik yang tersedia seringkali hanya digunakan sebagai arsip tanpa digunakan untuk analisis lebih lanjut untuk mendukung pengambilan keputusan strategis. Akibatnya, potensi data untuk memberikan wawasan berharga menjadi kurang optimal.

B. Kualitas pendidikan

Meningkatkan kualitas pendidikan, terutama dalam program studi teknologi informasi universitas muhammadiyah Palembang, membutuhkan pendekatan berbasis data.

C. Metode analisis yang efisien

Tidak ada metode analisis prediksi yang sistematis dan dapat diandalkan yang digunakan untuk menemukan pola penting dalam data akademik. Akibatnya, komponen penting yang memengaruhi hasil belajar siswa belum diungkapkan secara menyeluruh.

1.3 Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang diangkat dalam penelitian ini adalah sebagai berikut Bagaimana penerapan metode Naïve Bayes dalam analisis data akademik mahasiswa pada Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, Faktor-faktor apa saja yang signifikan dalam memprediksi hasil belajar mahasiswa, dan Sejauh mana tingkat akurasi prediksi hasil belajar mahasiswa menggunakan metode Naïve Bayes.

1.4 Tujuan penelitian

Berdasarkan rumusan masalah yang telah dibuat, maka tujuan dari penelitian ini ialah Melakukan analisis data akademik mahasiswa di Program Studi Teknologi Informasi Fakultas Teknik Universitas Muhammadiyah Palembang untuk mengidentifikasi pola-pola yang mempengaruhi hasil belajar mereka, Menggunakan metode data mining, khususnya algoritma Naive Bayes, dalam proses klasifikasi atau prediksi hasil belajar mahasiswa (seperti IPK atau status kelulusan), Membangun model prediksi hasil belajar mahasiswa yang berbasis pada algoritma Naive Bayes yang dapat membahas bagaimana hasil belajar mereka dipengaruhi oleh berbagai faktor, Mengevaluasi kinerja metode Naive Bayes dalam memprediksi hasil belajar siswa berdasarkan metrik evaluasi seperti akurasi, presisi, dan recall, dan Berdasarkan hasil analisis data, memberikan saran strategis untuk pengelola program studi atau fakultas untuk meningkatkan kualitas pembelajaran.

1.5 Kebaruan

Kebaruan yang ada pada penelitian ini meliputi beberapa aspek yang belum banyak diimplementasikan dalam penelitian sebelumnya yaitu sebagai berikut:

A. Konteks Spesifik

Penelitian ini difokuskan pada Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, yang belum banyak dilakukan kajian serupa dengan pendekatan data mining khususnya metode Naïve Bayes. Hasil penelitian akan memberikan model prediksi yang sesuai dengan karakteristik mahasiswa di lingkungan universitas ini.

B. Pemanfaatan Data Akademik Mahasiswa

Data yang digunakan mencakup berbagai variabel akademik mahasiswa (misalnya nilai mata kuliah, kehadiran, aktivitas akademik, dan faktor lainnya). Pengolahan data dilakukan melalui tahapan data preprocessing untuk menghasilkan dataset yang lebih bersih dan representatif dalam mendukung proses klasifikasi.

C. Implementasi Algoritma Naïve Bayes

Walaupun metode Naïve Bayes telah banyak digunakan dalam penelitian prediksi akademik, penerapannya dalam lingkup spesifik mahasiswa Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang masih jarang dilakukan, Penelitian ini akan menunjukkan sejauh mana akurasi, presisi, dan recall metode Naïve Bayes dalam memprediksi hasil belajar mahasiswa di konteks lokal.

D. Pengembangan Sistem Pendukung Keputusan

Hasil analisis tidak hanya berhenti pada perhitungan manual, melainkan diimplementasikan dalam bentuk sistem aplikasi sederhana yang dapat membantu pihak fakultas atau program studi dalam melakukan pemantauan dan evaluasi hasil belajar mahasiswa secara lebih efektif.

E. Manfaat Praktis

Penelitian ini memberikan kontribusi praktis berupa alat prediksi berbasis data mining yang dapat mendukung dosen dan pihak fakultas dalam melakukan intervensi dini terhadap mahasiswa yang berpotensi mengalami penurunan prestasi akademik dan Kebaruan juga terletak pada pengintegrasian analisis data

mining dengan implementasi sistem, sehingga tidak hanya menghasilkan model teoritis, tetapi juga solusi yang dapat digunakan langsung.

1.6 Manfaat penelitian

Manfaat-manfaat yang diperoleh dari hasil penelitian ini dapat dilihat dari beberapa sudut pandang, yaitu bagi penulis, pengguna sistem, dan peneliti lain.

- A. Bagi institusi pendidikan : Memberikan wawasan baru dalam pengelolaan data akademik untuk mendukung pengambilan keputusan strategis.
- B. Bagi mahasiswa : Memberikan gambaran tentang faktor-faktor yang dapat meningkatkan keberhasilan akademik.
- C. Bagi peneliti lain: Menjadi referensi bagi penelitian lebih lanjut di bidang penerapan data mining dalam pendidikan.

1.7 Pembatasan masalah

Ruang lingkup penelitian ini dibatasi pada beberapa elemen berikut agar penelitian ini dapat dilakukan dengan tepat dan mencapai tujuan yang telah ditetapkan:

- A. Mahasiswa Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang adalah subjek penelitian ini.
- B. Penelitian ini menggunakan data dari Sistem Informasi Akademik (SIKAD) dan kuesioner yang diberikan kepada siswa, yang mencakup tingkat kehadiran, nilai mata kuliah, Indeks Prestasi Kumulatif (IPK), dan jumlah SKS.
- C. Untuk memprediksi hasil belajar siswa, penelitian ini tidak membandingkan metode klasifikasi Naïve Bayes dengan algoritma klasifikasi lainnya dalam proses pemodelan.
- D. Proses penelitian terbatas pada fase data mining, yang mencakup pengumpulan data, preprocessing, pemodelan menggunakan algoritma Naive Bayes, dan evaluasi model menggunakan confusion matrix dan metrik evaluasi seperti akurasi, ketepatan, recall, dan skor F1.

Hasil penelitian fokus pada analisis prediksi hasil belajar siswa berdasarkan data akademik yang tersedia, serta mengevaluasi kinerja model dalam klasifikasi data.

1.8 Sistematika Penulisan

Penulisan Sistematis skripsi ini diartikan sebagai pedoman ataupun garis besar penulisan laporan penelitian yang dilakukan serta secara jelas mendeskripsikan isi laporan penelitian guna menjalin hubungan antara bab pertama sampai akhir.

Sistem penulisan laporan tesis ini terdiri atas:

BAB I PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan serta kaidah penelitian, metode penelitian yang digunakan, metode pengumpulan data, metode analisis data, metode pengujian serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab kedua membahas tentang landasan teori, yaitu teori-teori umum serta khusus guna membantu penulisan skripsi ini. Masalah yang meliputi di dalamnya merupakan pembahasan mengenai metode *Neve bayes*, *Algoritma C4.5*

BAB III METODOLOGI

Bab ini membahas tahapan analisis dan perancangan sistem pendukung keputusan yang digunakan untuk memprediksi hasil belajar mahasiswa. Analisis dilakukan terhadap kebutuhan fungsional dan non-fungsional sistem, serta terhadap data yang dibutuhkan dalam proses prediksi. Selanjutnya, dilakukan perancangan sistem yang mencakup perancangan alur proses, struktur data, serta penerapan algoritma *Naïve Bayes* sebagai inti dari metode klasifikasi. Perancangan ini bertujuan untuk

membangun sistem yang mampu memberikan hasil prediksi yang akurat dan dapat dimanfaatkan oleh pihak akademik dalam proses pengambilan Keputusan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan pembahasan terhadap sistem pendukung keputusan yang dibangun menggunakan algoritma *Naïve Bayes* dalam memprediksi hasil belajar mahasiswa. Pembahasan mencakup proses pengolahan data, implementasi metode, hasil prediksi, serta evaluasi kinerja sistem. Analisis dilakukan berdasarkan data mahasiswa Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, dengan tujuan untuk mengetahui sejauh mana sistem dapat memberikan prediksi yang akurat dan membantu pihak akademik dalam pengambilan keputusan terkait pembinaan dan pemantauan mahasiswa.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang dilakukan, serta saran-saran yang dapat dikemukakan untuk penelitian lebih lanjut atau untuk penerapan model analisis sentimen berbasis aspek.

DAFTAR PUSTAKA Bagian ini mencantumkan semua sumber referensi yang digunakan dalam penulisan penelitian, termasuk buku, artikel jurnal, dan sumber lainnya yang relevan dengan topik penelitian.

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 Data Mining

Data mining adalah salah satu cabang ilmu komputer yang relatif baru, muncul sekitar tahun 90-an. Orang masih memperdebatkan bidang mana yang paling cocok untuk data mining karena mencakup database, kecerdasan buatan, statistik, dll. Ada yang mengatakan bahwa data mining hanyalah pengajaran mesin atau analisis statistik yang dilakukan di database. Data mining adalah proses ekstraksi atau penggalian data yang belum diketahui sebelumnya atau pola yang menarik dari kumpulan data yang besar. Tujuan utama dari proses ini adalah untuk menemukan pola yang tidak terlihat sebelumnya, yang dapat memberikan wawasan yang berharga dan mendukung pengambilan keputusan yang lebih baik (Carudin et al., 2024).

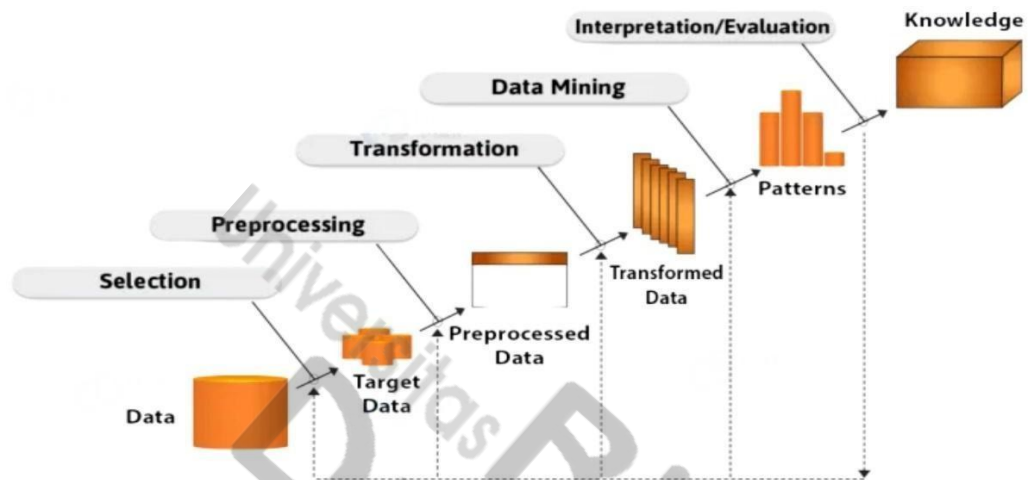
Data mining adalah salah satu teknik untuk menggali atau "menambang" pengetahuan dari kumpulan data yang sangat besar. Data mining adalah proses meninjau kumpulan data untuk menemukan hubungan yang tidak terduga dan meringkas data dengan cara yang berbeda dari sebelumnya sehingga pemilik data dapat memahaminya (Larose, 2005). Seperti yang dinyatakan oleh Turban et al. (2011), istilah "data mining" digunakan untuk menggambarkan penemuan pengetahuan yang ditemukan di dalam database. Proses ini adalah proses yang mengekstraksi dan mengidentifikasi informasi yang relevan dan berguna dari berbagai database yang sangat besar dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Tujuan data mining biasanya adalah untuk mengolah data dari database yang sangat besar. Dari data ini dilakukan pencarian pola atau trend. Selanjutnya, hasil pengolahan data mining dapat digunakan untuk pengambilan keputusan dan analisis yang diperlukan.

Ada beberapa alasan mengapa ilmu data mining sangat dibutuhkan saat

ini, salah satunya adalah fakta bahwa banyak data perusahaan atau organisasi hanya tersimpan di dalam database dan tidak dianalisis lebih lanjut untuk digunakan untuk kemajuan bisnis. Selain itu, perkembangan internet yang sangat pesat berdampak positif pada kemudahan akses data melalui berbagai perangkat hardware dan software yang memiliki kapasitas dan daya komputasi yang luar biasa. Tekanan kompetisi untuk meningkatkan pangsa pasar dan keuntungan semakin meningkat. Akibatnya, cara lain untuk menggali informasi dari data perusahaan atau organisasi diperlukan.

Algoritma data mining biasanya digunakan untuk ukuran data yang besar, tetapi beberapa jenis algoritma juga dapat bekerja dengan data yang lebih kecil. Data mining sederhana menggunakan kumpulan data dengan struktur di mana baris menunjukkan kasus-kasus individu (juga disebut sebagai pengamatan atau contoh) dan kolom menunjukkan atribut atau variabel dari kasus kasus. Jenis data yang digunakan tergantung pada algoritma yang akan digunakan. Data mining adalah proses yang menggunakan satu atau lebih teknik pembelajaran komputer untuk secara otomatis menganalisa dan mengekstraksi pengetahuan. Menurut Hermawati (2013), data mining juga merupakan proses mengekstraksi informasi atau sesuatu yang menarik atau penting dari data database, sehingga menghasilkan informasi yang sangat berharga.

Data mining merupakan salah satu proses inti dalam *Knowledge Discovery from Data* atau KDD. Banyak orang menyebut *data mining* sebagai sinonim dari KDD (Han *et al*, 2012). Proses KDD dapat diilustrasikan sebagai berikut:



Gambar 2.1 Proses data mining atau *Knowledge Discovery from Data*
(Sumber: Han et al, 2012)

Data mining adalah teknik untuk menemukan dan mendeskripsikan pola-pola yang ada dalam data sebagai sebuah alat untuk membantu menjelaskan data tersebut dan membuat prakiraan dari data itu (Witten & Eibe Frank dalam Mauriza, Ahmad dan Yusuf Sulisty, 2014).

Sedangkan menurut Turban dan Liang dalam Mujib Ridwan dkk (2013) menjelaskan bahwa data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar.

Berdasarkan pengertian tersebut, maka dapat disimpulkan bahwa data mining merupakan serangkaian proses ekstraksi dan identifikasi dengan melalui tahap perhitungan untuk mendapatkan informasi baru berdasarkan data yang telah ada.

2.2 Educational Data Maining

Bidang penelitian yang disebut *Educational Data Mining* (EDM) berfokus pada metode untuk mengekstrak informasi bermanfaat dari data yang dibuat di lingkungan pendidikan. Data ini mungkin berasal dari sistem manajemen pembelajaran (LMS), ujian, nilai tugas, kehadiran, aktivitas diskusi, dan log interaksi siswa dengan materi pembelajaran.

Tujuan EDM adalah untuk mendapatkan pemahaman tentang perilaku belajar, memprediksi kinerja siswa, dan memberikan saran untuk meningkatkan efektivitas pembelajaran. EDM biasanya menggunakan teknik data mining seperti klasifikasi, regresi, clustering, asosiasi, dan analisis deret waktu yang disesuaikan dengan pendidikan.

2.3 Hasil Belajar Mahasiswa

Hasil belajar merupakan perubahan perilaku yang diperoleh mahasiswa setelah melalui proses pembelajaran, baik berupa aspek pengetahuan (kognitif), keterampilan (psikomotor), maupun sikap (afektif). Hasil belajar menjadi indikator utama untuk menilai keberhasilan suatu proses pendidikan.

Menurut Bloom (1956), hasil belajar dapat diklasifikasikan ke dalam tiga ranah, yaitu:

1. Ranah Kognitif → berkaitan dengan kemampuan berpikir, memahami, mengingat, menganalisis, mengevaluasi, dan mencipta.
2. Ranah Afektif → mencakup sikap, nilai, motivasi, serta penerimaan terhadap pembelajaran.
3. Ranah Psikomotor → berkaitan dengan keterampilan fisik, koordinasi, dan kemampuan melakukan suatu tindakan.

Dalam konteks perguruan tinggi, hasil belajar mahasiswa umumnya diukur melalui indikator-indikator berikut:

- Indeks Prestasi Kumulatif (IPK) → ukuran capaian akademik keseluruhan mahasiswa.
- Nilai Mata Kuliah → capaian kompetensi mahasiswa pada setiap mata kuliah yang diikuti.
- Kehadiran dan Partisipasi → keterlibatan mahasiswa dalam proses perkuliahan, baik tatap muka maupun tugas.
- Kompetensi Keterampilan → keterampilan yang diperoleh sesuai bidang

studi, misalnya kemampuan menguasai perangkat lunak, pemrograman, atau analisis data pada bidang Teknologi Informasi.

Hasil belajar mahasiswa dipengaruhi oleh berbagai faktor, antara lain:

1. Faktor Internal → motivasi belajar, minat, gaya belajar, kemampuan kognitif, dan kondisi psikologis.
2. Faktor Eksternal → lingkungan belajar, metode pengajaran dosen, fasilitas kampus, serta dukungan keluarga atau sosial.
3. Faktor Pendukung Teknologi → ketersediaan akses informasi, media pembelajaran digital, serta penerapan teknologi pendidikan.

Dengan demikian, hasil belajar mahasiswa tidak hanya sekadar nilai akademik, melainkan mencerminkan perkembangan pengetahuan, keterampilan, dan sikap yang diperoleh selama menempuh pendidikan. Dalam penelitian ini, hasil belajar mahasiswa yang digunakan sebagai data adalah nilai akademik dan data pendukung lain yang tersedia pada sistem akademik Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang.

2.4 Naïve Bayes Classifier

Teori Bayesian, yang dikenal sebagai teori sekita yang diciptakan oleh Thomas Bayes pada tahun 1950, sering ditemukan dalam studi ilmu statistika yang berbasis pada teorema atau aturan Bayes dan menyatakan bahwa kondisi probabilitas suatu kejadian (hipotesis) bergantung pada kejadian lain (bukti). Pada dasarnya, teorema ini mengatakan bahwa kita dapat memprediksi apa yang akan terjadi di masa depan (Moriesta et al., 2017). Dengan menggunakan perhitungan probabilitas, pengklasifikasi Bayes dapat memprediksi jumlah keanggotaan kelas yang akan masuk ke dalam kelas tertentu dari data tuple (Handayani & Pribadi, 2015).

Naive Bayes digunakan untuk pengenalan pola dan klasifikasi yang berada dibawah variasi yang berbeda dalam pola pengklasifikasi untuk probabilitas dan kemungkinan dasar (Luo, 2021). *Naive Bayes* menggunakan model fitur independen dengan asumsi independensi yang kuat (naif) (Harijanto et al., 2018).

Secara umum teori *Naive Bayes* ditulis dalam bentuk:

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \quad (\text{Rumus 4})$$

Keterangan :

X = Data dengan *class* yang belum diketahui.

H = Hipotesis data merupakan suatu *class* spesifik.

$P(H|X)$ = Probabilitas hipotesis H berdasar kondisi X (pos teriori probabilitas).

$P(H)$ = Probabilitas hipotesis H (priorprobabilitas).

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H. $P(X)$ = Probabilitas X.

Rumus diatas menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik sampel secara global (disebut juga *evidence*) (Anggraini et al., 2019).

2.5 Confusion Matrix

Confusion matrix adalah sebuah tabel yang menyajikan informasi mengenai jumlah data uji yang diklasifikasikan dengan benar maupun yang salah oleh model. Tabel ini memungkinkan perbandingan langsung antara hasil klasifikasi yang dihasilkan oleh model dan label sebenarnya dari data uji. Dengan demikian, *confusion matrix* memberikan gambaran yang jelas mengenai akurasi dan kesalahan klasifikasi, serta dapat digunakan untuk menghitung metrik evaluasi lainnya seperti *precision*, *recall*, *f1-score*, dan *accuracy* secara lebih rinci.

Tabel 2.1 Confusion Matrix

Hasil Prediksi	Nilai Aktual	
	Positif	Negatif
Positif	TP	FP
Negatif	EN	TN

Tabel 2.1 Merupakan tabel *confusion matrix* *True Positive* (TP) adalah jumlah data dengan nilai aktual kelas dan nilai prediksi kelas positif oleh model klasifikasi, *False Positive* (FP) adalah jumlah nilai aktual kelas negatif dan nilai prediksi kelas positif oleh klasifikasi, *False Negative* (FN) adalah jumlah nilai aktual kelas positif dan nilai prediksi kelas negatif oleh model klasifikasi, *True Negative* (TN) adalah jumlah nilai aktual kelas negatif dan nilai prediksi kelas negatif oleh model klasifikasi.

Akurasi adalah gambaran seberapa akurat hasil klasifikasi model yang dibuat, berikut adalah rumus untuk mendapatkan nilai akurasi.

Tabel 2.2 Akurasi

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

Persamaan merupakan nilai akurasi yang dihasilkan dari nilai *true positive* (TP) ditambah dengan nilai *true negative* (TN) dibagi dengan penjumlahan dari nilai *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Presisi adalah tingkat akurasi nilai prediksi model dalam memprediksi kelas positif dan benar. Nilai ini merupakan hasil dari nilai prediksi positif dibandingkan dengan jumlah nilai positif yang diklasifikasikan oleh model. Nilai presisi dapat diperoleh dengan rumus berikut.

Tabel 2.3 Presisi

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

Persamaan merupakan nilai presisi yang bisa diperoleh dari perbandingan nilai *True positive* (TP) dengan jumlah dari nilai *true positive* (TP) dan *false positive* (FP).

Recall adalah jumlah nilai *true positive* (TP) dibandingkan dengan jumlah nilai aktual positif. Persamaan adalah rumus yang digunakan untuk mendapatkan nilai *recall*.

Tabel 2.4 Recall

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

2.6 Penelitian Terdahulu

Penelitian mengenai untuk memprediksi hasil belajar mahasiswa menggunakan metode naive bayes pada perguruan tinggi . Berikut adalah beberapa penelitian terdahulu dan hasil penelitiannya pada Tabel 2.1 sebagai berikut:

Tabel 2.5 Penelitian terdahulu

No	Peneliti (Tahun)	Judul	Metode	Hasil	Relevansi
1	Rahman & Susanto (2020)	Penerapan Algoritma Naïve Bayes untuk Prediksi Kelulusan Mahasiswa	Naïve Bayes	Akurasi 82%	Dasar penggunaan NB untuk prediksi kelulusan
2	Putri & Nugroho (2021)	Implementasi Data Mining dengan Naïve Bayes dalam Prediksi	Naïve Bayes	Akurasi 80–85%	Efektivitas atribut akademik

		Prestasi Akademik Mahasiswa			
3	Hidayat & Lestari (2019)	Analisis Prediksi Nilai Mahasiswa Menggunakan C4.5 dan Naïve Bayes	NB & C4.5	NB unggul dataset kecil (78%)	Pembanding NB dan C4.5
4	Sari & Pratama (2022)	Prediksi Kinerja Akademik Mahasiswa Menggunakan Naïve Bayes Classifier	Naïve Bayes	Akurasi 84%	NB efektif untuk klasifikasi prestasi
5	Wibowo (2020)	Data Mining untuk Prediksi Nilai Mahasiswa dengan KNN dan Naïve Bayes	KNN & NB	NB lebih cepat, KNN lebih akurat	NB unggul dalam efisiensi
6	Handayani (2021)	Prediksi Hasil Studi Mahasiswa Menggunakan Naïve Bayes	Naïve Bayes	Akurasi 81%	Mendukung penerapan NB di akademik

7	Yusuf & Anwar (2020)	Analisis Prestasi Akademik Menggunakan Naïve Bayes	Naïve Bayes	Akurasi 79%	Prediksi hasil belajar berbasis data
8	Astuti (2021)	Penerapan Data Mining untuk Prediksi Kelulusan Mahasiswa	Naïve Bayes	Akurasi 83%	NB terbukti konsisten di prediksi
9	Santoso (2019)	Sistem Prediksi IPK Mahasiswa	NB & Decision Tree	NB akurasi lebih stabil	NB sesuai untuk dataset akademik
10	Kurniawan (2022)	Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes	Naïve Bayes	Akurasi 82.5%	Penerapan NB di kelulusan
11	Jannah (2020)	Algoritma Naïve Bayes untuk Prediksi Nilai Matakuliah	Naïve Bayes	Akurasi 80%	NB sesuai prediksi per mata kuliah
12	Nugraha (2021)	Penerapan Data Mining untuk Prediksi Hasil	Naïve Bayes	Akurasi 81%	Relevan dengan prediksi hasil belajar

		Belajar Mahasiswa			
13	Wahyudi (2019)	Perbandingan Naïve Bayes dan KNN untuk Prediksi Akademik	NB & KNN	NB lebih cepat, KNN lebih akurat	Kelebihan NB pada kecepatan
14	Susilowati (2022)	Prediksi Nilai Mahasiswa Menggunakan Data Mining	Naïve Bayes	Akurasi 83%	NB akurat untuk prediksi nilai
15	Fadillah (2021)	Implementasi Naïve Bayes dalam Prediksi Hasil Belajar	Naïve Bayes	Akurasi 82%	Mendukung penelitian ini
16	Gunawan (2020)	Prediksi Keberhasilan Akademik Mahasiswa	Naïve Bayes	Akurasi 81%	NB terbukti konsisten
17	Maharani (2021)	Penerapan NB untuk Prediksi Prestasi Akademik	Naïve Bayes	Akurasi 84%	Relevansi tinggi dengan penelitian
18	Arifin (2019)	Data Mining untuk Analisis IPK Mahasiswa	Naïve Bayes	Akurasi 78%	NB dapat prediksi IPK

19	Setiawan (2021)	Implementasi Algoritma Naïve Bayes pada Prediksi Nilai	Naïve Bayes	Akurasi 83%	NB efektif dalam prediksi nilai
20	Amelia (2022)	Prediksi Kelulusan Mahasiswa dengan Naïve Bayes	Naïve Bayes	Akurasi 82%	Sesuai dengan tujuan penelitian
21	Rahim (2020)	Analisis Data Mining untuk Prestasi Mahasiswa	Naïve Bayes	Akurasi 80%	NB cocok untuk analisis akademik
22	Dewi (2021)	Prediksi Hasil Belajar dengan Naïve Bayes	Naïve Bayes	Akurasi 81%	NB digunakan pada data pendidikan
23	Syahrul (2019)	Penerapan Naïve Bayes dalam Sistem Akademik	Naïve Bayes	Akurasi 82%	Mendukung implementasi di kampus
24	Nisa (2021)	Sistem Prediksi Prestasi Mahasiswa	Naïve Bayes	Akurasi 83%	NB sebagai pendukung keputusan
25	Putra (2022)	Prediksi Nilai Akhir Mahasiswa	Naïve Bayes	Akurasi 84%	Sangat relevan dengan penelitian ini

		dengan Naïve Bayes			
--	--	-----------------------	--	--	--

2.7 Kerangka kerja yang digunakan

Dalam kerangka kerja penelitian ini, tahapan sistematis yang dilakukan peneliti, mulai dari pengumpulan data hingga analisis hasil, digambarkan sebagai berikut:

2.7.1 Studi literatur

Mengumpulkan dan mempelajari teori yang relevan. Ini termasuk gagasan data mining, algoritma Naïve Bayes, dan penelitian terdahulu tentang memprediksi hasil belajar siswa.

2.7.2 Identifikasi masalah dan tujuan

Menentukan masalah utama, yaitu bagaimana menggunakan algoritma Naïve Bayes untuk memprediksi hasil belajar siswa dan merumuskan tujuan penelitian.

2.7.3 Pengumpulan data

Data diperoleh melalui observasi, wawancara, dan dokumentasi dari sistem akademik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, yang mencakup nilai mata kuliah, IPK, dan kehadiran mahasiswa.

2.7.4 Implementasi algoritma naïve bayes

Membangun model klasifikasi dengan menggunakan algoritma Naïve Bayes. Tahap ini mencakup pembagian data menjadi data latih (training) dan data uji (testing), pelatihan model, serta penerapan pada data uji.

2.7.5 Evaluasi Kinerja Sistem

Mengukur performa model menggunakan metode evaluasi seperti confusion matrix, akurasi, presisi, recall, dan F1-score.

2.7.6 Analisis dan Pembahasan Hasil

Membahas hasil prediksi dan evaluasi kinerja algoritma Naïve Bayes, serta menginterpretasikan faktor-faktor yang berpengaruh terhadap hasil belajar mahasiswa.

2.8 Google colabs

Google Colab (singkatan dari *Google Colaboratory*) merupakan sebuah platform berbasis *cloud computing* yang disediakan oleh Google. Platform ini memungkinkan pengguna untuk mengeksekusi kode *Python* dalam lingkungan berbasis *cloud* tanpa perlu menginstal atau mengatur lingkungan lokal mereka sendiri. *Google Colab* sering digunakan oleh para ilmuwan data, peneliti, dan pengembang untuk melakukan berbagai jenis pekerjaan, termasuk pemrosesan data, pengembangan model kecerdasan buatan, analisis data, dan pelatihan model pembelajaran mesin (*machine learning*). Kemampuannya dalam mendukung pemrosesan data skala besar serta integrasinya dengan berbagai pustaka *Python* menjadikannya alat yang sangat relevan dalam aktivitas penelitian dan pengembangan berbasis data.

Beberapa fitur utama dari *Google Colab* adalah :

1. *Python di Cloud* = kita dapat menulis dan mengeksekusi kode *Python* langsung di browser web tanpa perlu menginstal *Python* atau pustaka di komputer lokal.
2. *Gratis Google Colab* = Layanan ini disediakan secara gratis oleh Google. Sehingga sangat ekonomis dan dapat diakses oleh siapa saja tanpa menggunakannya biaya.
3. *GPU Gratis Google Colab* = Menyediakan akses gratis ke GPU (*Graphics Processing Unit*) yang sangat berguna dalam melatih model pembelajaran mesin yang memerlukan daya komputasi yang tinggi.
4. *Akses ke Penyimpanan Google Drive* = Pengguna dapat mengakses dan menyimpan file langsung di Google Drive, yang memudahkan berbagi dan menyimpan pekerjaan.

5. Notebook Interaktif *Google Colab* = Menggunakan format "notebook" yang memungkinkan untuk menggabungkan kode, teks, gambar, dan hasil dalam satu dokumen interaktif. Ini sangat berguna untuk dokumentasi dan berbagi analisis.
6. Pustaka Tersedia di *Google Colab* = Menyediakan banyak pustaka umum seperti *NumPy*, *pandas*, *TensorFlow*, *PyTorch*, dan sebagainya. Kita juga dapat mengimpor pustaka-pustaka ini ke dalam lingkungan *Colab* dengan mudah.
7. Kerjasama Tim = Pengguna dapat berbagi notebook *Colab* dengan anggota tim dan berkolaborasi dalam waktu nyata.
8. Fleksibel dan Mudah Digunakan = *Google Colab* merupakan alat yang fleksibel, mudah digunakan dan ramah bagi pemula. namun tetap cukup kuat untuk mendukung kebutuhan komputasi lanjutan dalam pengembangan model dan eksplorasi data. *Google Colab* adalah salah satu alat yang populer dalam komunitas ilmu data dan kecerdasan buatan karena kemudahan penggunaannya, akses ke sumber daya komputasi yang kuat, dan kemampuan untuk berkolaborasi dengan rekan tim.

BAB III

METODOLOGI

3.1 Metode Penelitian

Jenis penelitian ini adalah penelitian terapan (applied research) dengan pendekatan kuantitatif dan eksperimental komputasional. Penelitian terapan dipilih karena berfokus pada penerapan konsep dan metode *data mining* untuk menyelesaikan permasalahan nyata, yaitu memprediksi hasil belajar mahasiswa berdasarkan data akademik.

Pendekatan kuantitatif digunakan karena penelitian ini melibatkan pengolahan data numerik, penerapan algoritma statistik (Naive Bayes), serta pengukuran kinerja model prediksi melalui nilai akurasi, presisi, recall, dan f1-score.

Selain itu, penelitian ini juga bersifat eksperimental komputasional, karena dilakukan proses percobaan penerapan metode Naive Bayes terhadap dataset mahasiswa dengan tujuan memperoleh hasil terbaik dari model prediksi yang dibangun.

3.2 Waktu dan tempat penelitian

3.2.1 Tempat penelitian

Penelitian ini dilaksanakan di Fakultas Teknik, Program Studi Teknologi Informasi, Universitas Muhammadiyah Palembang. Pemilihan lokasi penelitian didasarkan pada beberapa pertimbangan, yaitu:

- Program Studi Teknologi Informasi memiliki sistem informasi akademik (SIKAD) yang terstruktur dan terdokumentasi dengan baik.
- Tersedianya data akademik mahasiswa yang lengkap dan relevan untuk keperluan analisis data mining.
- Kesesuaian objek penelitian dengan fokus penelitian, yaitu prediksi hasil belajar mahasiswa berbasis algoritma Naïve Bayes

- Dukungan dari pihak fakultas dalam proses pengumpulan data dan validasi informasi akademik.

Data yang digunakan dalam penelitian ini diperoleh dari Sistem Informasi Akademik (SIKAD) serta hasil penyebaran kuesioner kepada mahasiswa aktif Program Studi Teknologi Informasi.

3.2.2 Waktu penelitian

Penelitian ini dilaksanakan selama kurang lebih empat bulan, dimulai pada bulan Oktober 2025 hingga Januari 2026. Rincian tahapan waktu penelitian adalah sebagai berikut:

Tabel 3.1 Waktu Penelitian

No	Kegiatan	Waktu
1	Pengumpulan data akademik dan kuesioner	Oktober 2025
2	Preprocessing dan validasi data	November 2025
3	Pembangunan dan pelatihan model Naïve Bayes	Desember 2025
4	Evaluasi model dan analisis hasil	Januari 2026
5	Penyusunan laporan tesis	Januari 2026

Durasi waktu tersebut disesuaikan dengan kebutuhan penelitian mulai dari pengumpulan data hingga penyusunan laporan akhir.

3.3 Prosedur Penelitian

Prosedur penelitian ini disusun secara sistematis untuk menjamin bahwa proses analisis dan penerapan metode data mining dengan algoritma Naïve Bayes dalam memprediksi hasil belajar mahasiswa dilakukan secara terstruktur, objektif, dan dapat dipertanggungjawabkan secara ilmiah. Tahapan penelitian mencakup mulai dari pengumpulan data hingga analisis hasil model klasifikasi. Penelitian ini terdiri dari lima tahapan utama, yaitu: pengumpulan data, preprocessing data, pemodelan, evaluasi model, dan analisis hasil.

3.3.1 Pengumpulan data

Tahap pertama dalam penelitian ini adalah pengumpulan data, yang bertujuan memperoleh dataset relevan dan sesuai kebutuhan analisis. Data yang digunakan terdiri dari data sekunder dan data primer. Data sekunder diperoleh dari Sistem Informasi Akademik (SIKAD) Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang. Data ini mencakup nilai mata kuliah, Indeks Prestasi Kumulatif (IPK), persentase kehadiran, jumlah SKS, dan status akademik mahasiswa. Dataset terdiri dari sekitar 200 mahasiswa yang telah menempuh minimal dua semester dan memiliki data akademik lengkap. Selain itu, data primer diperoleh melalui penyebaran kuesioner kepada 92 mahasiswa aktif. Kuesioner ini bertujuan untuk mendapatkan informasi tambahan terkait motivasi belajar, tingkat kehadiran, serta persepsi mahasiswa terhadap penerapan data mining dalam dunia akademik. Data primer digunakan sebagai pendukung dalam validasi dan interpretasi hasil model. Sebelum digunakan dalam analisis, seluruh data yang diperoleh diperiksa kelengkapan dan konsistensinya untuk memastikan tidak ada data ganda maupun kesalahan pencatatan.

3.3.2 Preprocessing Data

Setelah data terkumpul, dilakukan tahap preprocessing untuk menyiapkan data agar layak digunakan dalam proses pemodelan. Tahap ini sangat penting karena kualitas data akan memengaruhi kinerja model klasifikasi.

- Data Cleaning

Pada tahap ini dilakukan penghapusan data duplikat berdasarkan Nomor Induk Mahasiswa (NIM). Selain itu, dilakukan pemeriksaan terhadap nilai yang hilang (missing value). Jika ditemukan data yang tidak lengkap pada atribut utama seperti nilai atau kehadiran, data tersebut akan dihapus atau diperbaiki sesuai dengan kebijakan pengolahan data. Validasi rentang nilai juga dilakukan untuk memastikan bahwa nilai akademik berada dalam rentang 0–100 dan IPK berada dalam rentang 0,00–4,00.

- Data transformation

Nilai numerik seperti nilai mata kuliah dan persentase kehadiran diubah menjadi bentuk kategorikal melalui proses diskritisasi. Transformasi ini bertujuan menyesuaikan data dengan karakteristik algoritma Naïve Bayes, yang bekerja lebih optimal pada data kategorikal. Nilai akademik diklasifikasikan ke dalam kategori A, B, C, D, dan E, sedangkan kehadiran dikelompokkan menjadi kategori sangat tinggi, tinggi, sedang, dan rendah.

- Feature Selection

Pada tahap ini dilakukan pemilihan atribut yang paling relevan terhadap variabel target (hasil belajar). Variabel yang digunakan dalam model meliputi IPK, nilai akademik, kehadiran, dan jumlah SKS. Proses ini bertujuan mengurangi kompleksitas model sekaligus meningkatkan akurasi klasifikasi.

- Data Splitting

Dataset yang telah diproses kemudian dibagi menjadi dua bagian, yaitu 80% sebagai data training dan 20% sebagai data testing. Pembagian ini dilakukan menggunakan metode stratified sampling untuk memastikan distribusi kelas tetap seimbang pada kedua subset data.

3.3.3 Pemodelan (Modeling)

Tahap pemodelan merupakan inti dari penelitian ini, di mana dilakukan pembangunan model klasifikasi menggunakan algoritma Naïve Bayes. Model dibangun dengan menghitung probabilitas prior untuk setiap kelas hasil belajar (Baik, Cukup, dan Kurang), kemudian menghitung probabilitas likelihood dari setiap atribut terhadap masing-masing kelas. Berdasarkan perhitungan tersebut, dihitung probabilitas posterior untuk menentukan kelas dengan probabilitas tertinggi. Proses pemodelan dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka machine learning di platform Google Colab. Model dilatih menggunakan data training, kemudian diuji menggunakan data testing untuk mengevaluasi performanya dalam memprediksi hasil belajar mahasiswa.

3.3.4 Evaluasi Model

Setelah model dibangun, dilakukan evaluasi untuk mengukur tingkat kinerjanya. Evaluasi dilakukan menggunakan metode confusion matrix yang membandingkan kelas aktual dengan kelas prediksi. Dari confusion matrix tersebut dihitung beberapa metrik evaluasi, yaitu:

- Accuracy : untuk mengukur tingkat ketepatan klasifikasi secara keseluruhan.
- Precision : untuk mengukur ketepatan model dalam memprediksi kelas tertentu.
- Recall : untuk mengukur kemampuan model dalam mendeteksi data yang benar-benar termasuk dalam kelas tertentu.
- F1-Score : sebagai ukuran keseimbangan antara precision dan recall.

Evaluasi ini bertujuan untuk mengetahui sejauh mana model mampu memprediksi hasil belajar mahasiswa secara akurat.

3.3.5 Analisis dan Interpretasi Hasil

Tahap terakhir adalah analisis dan interpretasi hasil model. Pada tahap ini dilakukan pembahasan terhadap hasil evaluasi model, distribusi kelas, serta pengaruh masing-masing variabel terhadap hasil belajar mahasiswa. Hasil prediksi kemudian dibandingkan dengan kondisi nyata berdasarkan data kuesioner untuk menilai kesesuaian antara model dan kondisi akademik mahasiswa. Analisis ini juga digunakan untuk memberikan rekomendasi kepada pihak fakultas dalam melakukan intervensi dini terhadap mahasiswa yang berpotensi mengalami penurunan prestasi akademik.

3.4 Variabel Input Berdasarkan Data Kuesioner

Penelitian ini menggunakan data yang dikumpulkan dari Sistem Informasi Akademik (SIKAD) dan kuesioner yang dibagikan kepada mahasiswa Program Studi Teknologi Informasi. Sebuah kuesioner disebarluaskan secara online dan berhasil dikumpulkan dari 123 siswa yang menjawab. Data ini digunakan untuk

mengidentifikasi variabel seperti tingkat kehadiran, capaian IPK, dan karakteristik responden.

3.4.1 Tingkat Kehadiran Mahasiswa

Tingkat kehadiran mahasiswa menggambarkan tingkat kedisiplinan mahasiswa dalam mengikuti kegiatan perkuliahan. Kehadiran menjadi salah satu faktor penting yang dapat mempengaruhi keberhasilan belajar mahasiswa.

Tabel 3.2 Tingkat Kehadiran Mahasiswa

Kategori Kehadiran	Jumlah	Persentase
Selalu hadir	87	70,73%
Sering hadir ($\geq 75\%$)	26	21,14%
Kadang hadir (50–74%)	6	4,88%
Jarang hadir ($< 50\%$)	4	3,25%
Total	123	100%

Berdasarkan Tabel 3.2 dapat diketahui bahwa sebagian besar mahasiswa memiliki tingkat kehadiran yang tinggi. Sebanyak 87 mahasiswa (70,73%) menyatakan selalu hadir dalam perkuliahan, sedangkan 26 mahasiswa (21,14%) berada pada kategori sering hadir. Sementara itu, hanya sebagian kecil mahasiswa yang berada pada kategori kadang hadir sebanyak 6 mahasiswa (4,88%) dan jarang hadir sebanyak 4 mahasiswa (3,25%). Hasil ini menunjukkan bahwa mayoritas mahasiswa Program Studi Teknologi Informasi memiliki kedisiplinan yang baik dalam mengikuti kegiatan perkuliahan, sehingga kehadiran dapat menjadi salah satu indikator penting dalam menganalisis hasil belajar mahasiswa.

3.4.2 Distribusi Responden Berdasarkan Jenis Kelamin

Distribusi responden berdasarkan jenis kelamin digunakan untuk mengetahui komposisi mahasiswa laki-laki dan perempuan yang menjadi responden dalam penelitian ini.

Tabel 3.3 Distribusi Responden Berdasarkan Jenis Kelamin

Jenis Kelamin	Jumlah	Persentase
Laki-laki	77	62,60%
Perempuan	46	37,40%
Total	123	100%

Berdasarkan Tabel 3.3 dapat diketahui bahwa responden penelitian didominasi oleh mahasiswa laki-laki sebanyak 77 orang (62,60%), sedangkan mahasiswa perempuan sebanyak 46 orang (37,40%). Distribusi ini menunjukkan bahwa jumlah mahasiswa laki-laki lebih besar dibandingkan mahasiswa perempuan. Kondisi ini juga umum terjadi pada program studi yang berkaitan dengan bidang teknologi informasi.

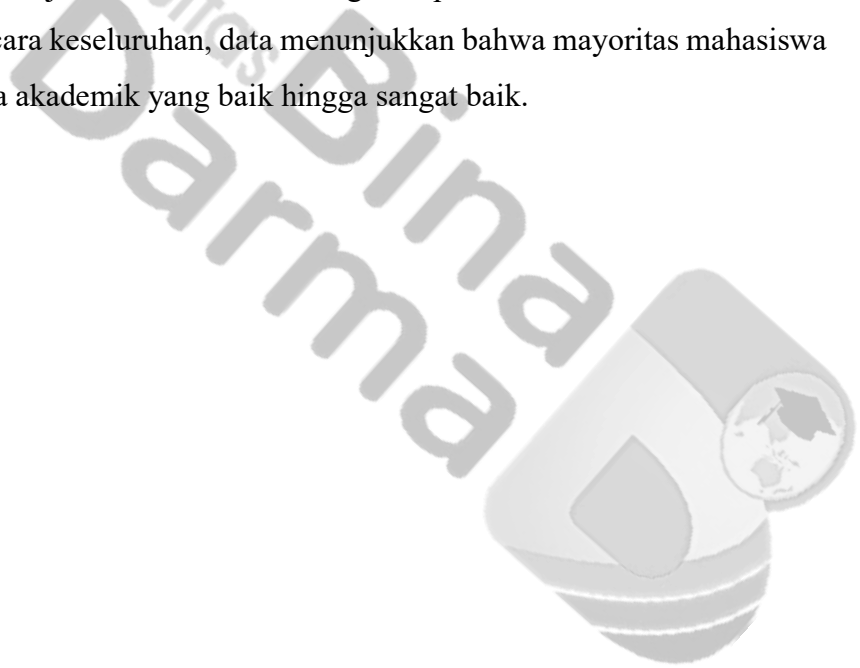
3.4.3 Indeks Prestasi Kumulatif (IPK) Mahasiswa

Indeks Prestasi Kumulatif (IPK) merupakan indikator utama yang digunakan untuk menilai prestasi akademik mahasiswa selama menempuh pendidikan di perguruan tinggi.

Tabel 3.4 Distribusi IPK Mahasiswa

Rentang IPK	Jumlah	Persentase
≥ 3.50	60	48,78%
3.00 – 3.49	55	44,72%
2.50 – 2.99	7	5,69%
Total	123	100%

Berdasarkan Tabel 3.4 dapat diketahui bahwa sebagian besar mahasiswa memiliki capaian akademik yang baik. Sebanyak 60 mahasiswa (48,78%) memiliki $IPK \geq 3,50$ yang termasuk dalam kategori prestasi akademik sangat baik. Selanjutnya terdapat 55 mahasiswa (44,72%) yang memiliki IPK pada rentang 3,00–3,49, yang juga menunjukkan capaian akademik yang baik. Sementara itu hanya 7 mahasiswa (5,69%) yang memiliki IPK pada rentang 2,50–2,99, yang menunjukkan bahwa jumlah mahasiswa dengan capaian akademik relatif rendah sangat sedikit. Secara keseluruhan, data menunjukkan bahwa mayoritas mahasiswa memiliki performa akademik yang baik hingga sangat baik.



BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

Studi ini dilakukan di Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang. Tujuan dari penelitian ini adalah untuk menganalisis dan menerapkan teknik data mining yang menggunakan algoritma Naive Bayes untuk memprediksi hasil belajar mahasiswa. Data yang digunakan dalam penelitian ini berasal dari dua sumber utama: Sistem Informasi Akademik (SIKAD) dan data primer yang diberikan kepada siswa melalui kuesioner. Data akademik yang digunakan meliputi nilai mata kuliah, tingkat kehadiran, Indeks Prestasi Kumulatif (IPK), dan jumlah SKS yang diambil siswa. Dataset yang digunakan dalam penelitian ini terdiri dari sekitar 200 data mahasiswa dari berbagai angkatan Program Studi Teknologi Informasi. Dalam proses klasifikasi, setiap data siswa diwakili dalam sejumlah karakteristik. Nilai akademik, IPK, tingkat kehadiran, jumlah SKS, dan label hasil belajar mahasiswa adalah semua atribut yang termasuk dalam kategori ini. Diharapkan model klasifikasi yang mampu memprediksi kategori hasil belajar siswa secara akurat akan dihasilkan melalui proses analisis melalui beberapa tahapan utama. Tahapan-tahap ini termasuk preprocessing data, pembagian data untuk pelatihan dan pengujian, dan evaluasi model menggunakan matriks keacauan.

Dataset penelitian ini diperoleh dari hasil kuesioner mahasiswa Program Studi Teknologi Informasi dengan jumlah sebanyak 123 responden. Distribusi data menunjukkan bahwa mayoritas mahasiswa memiliki tingkat kehadiran yang tinggi serta capaian akademik yang baik. Pada aspek kehadiran, sebanyak 87 responden (70,73%) menyatakan selalu hadir dalam perkuliahan, 26 responden (21,14%) sering hadir, 6 responden (4,88%) kadang hadir, dan 4 responden (3,25%) jarang hadir.

Sementara itu, pada aspek Indeks Prestasi Kumulatif (IPK), dari 122 responden yang mengisi data IPK, sebanyak 60 mahasiswa (49,18%) berada pada

kategori $\geq 3,50$, 55 mahasiswa (45,08%) berada pada kategori 3,00–3,49, dan 7 mahasiswa (5,74%) berada pada kategori 2,50–2,99. Dengan demikian, distribusi data dalam dataset menunjukkan bahwa sebagian besar responden memiliki karakteristik akademik yang baik dan layak digunakan sebagai dasar dalam proses analisis serta pemodelan prediksi hasil belajar mahasiswa.

Tabel 4.1 Ringkas distribusi data

Variabel	Kategori	Jumlah	Persentase
Kehadiran	Selalu hadir	87	70,73%
	Sering hadir	26	21,14%
	Kadang hadir	6	4,88%
	Jarang hadir	4	3,25%
IPK	$\geq 3,50$	60	49,18%
	3,00–3,49	55	45,08%
	2,50–2,99	7	5,74%

4.2 Implementasi Naïve Bayes

Langkah-langkah implementasi algoritma Naïve Bayes dalam penelitian ini dijelaskan sebagai berikut:

1. Data Preparation

Data yang telah melalui tahap *preprocessing* dibagi menjadi dua bagian:

- 80% data digunakan untuk *training*
- 20% data digunakan untuk *testing*

2. Training Model Naïve Bayes

Sistem menghitung probabilitas prior ($P(\text{Class})$) dan probabilitas bersyarat ($P(X|\text{Class})$) untuk setiap variabel.

4.3 Preprocessing Data

Preprocessing data merupakan tahap penting dalam proses data mining karena memastikan bahwa data yang digunakan sudah bersih, terstruktur, dan siap diolah oleh algoritma. Pada penelitian ini, preprocessing dilakukan melalui beberapa langkah, yaitu *data cleaning*, *data transformation*, *data reduction*, dan *data splitting*. Berikut uraian lengkapnya :

4.3.1 Data Cleaning (Pembersihan Data)

Pada tahap ini, data diperiksa untuk memastikan tidak adanya nilai yang hilang, data ganda, maupun nilai yang tidak sesuai. Beberapa proses yang dilakukan:

1. Menghapus data duplikat

Data mahasiswa yang muncul lebih dari satu kali (duplikat) dihapus agar tidak memengaruhi akurasi model.

2. Penanganan Missing Values

Jika terdapat nilai kosong pada kolom Tugas, UTS, UAS, atau Kehadiran, maka data tersebut dihapus. Pertimbangan ini diambil karena persentase missing values kecil sehingga tidak memengaruhi jumlah sampel secara signifikan.

3. Validasi format dan range nilai

Nilai Tugas, UTS, dan UAS harus berada pada rentang 0–100, dan Kehadiran harus berada pada rentang 0–100%. Data yang tidak sesuai dilakukan koreksi atau dihapus sesuai kebijakan preprocessing.

4.3.2 Data Transformation

Transformasi data merupakan proses mengubah format data asli ke dalam bentuk yang lebih sesuai untuk kebutuhan pemodelan dalam data mining. Pada penelitian ini, transformasi dilakukan pada variabel nilai akademik dan kehadiran mahasiswa agar mudah diolah oleh algoritma Naïve Bayes yang bekerja dengan format data terklasifikasi atau terkodekan.

A. Transformasi Nilai Akademik

Data nilai Tugas, UTS, dan UAS pada awalnya berupa angka dalam rentang 0–100. Untuk memudahkan proses klasifikasi, nilai angka dikonversi menjadi kategori huruf berdasarkan kebijakan penilaian di lingkungan Fakultas Teknik Universitas Muhammadiyah Palembang, seperti berikut:

Tabel 4.2 Nilai Akademik

Rentang Nilai	Kategori Huruf	Interpretasi
80 – 100	A	Sangat Baik
70 – 79	B	Baik
60 – 69	C	Cukup
50 – 59	D	Kurang
< 50	E	Sangat Kurang

Transformasi ini dilakukan agar pola nilai mahasiswa lebih mudah dianalisis berdasarkan perbedaan tingkat pencapaian akademik.

B. Transformasi Kehadiran

Transformasi data kehadiran dilakukan untuk mengubah data persentase kehadiran mahasiswa yang semula berbentuk data numerik (0–100%) menjadi data kategorikal. Proses transformasi ini bertujuan untuk mempermudah proses analisis klasifikasi menggunakan algoritma Naïve Bayes, karena metode tersebut bekerja lebih optimal pada data yang telah dikelompokkan ke dalam kategori tertentu. Pada dataset awal, tingkat kehadiran mahasiswa dicatat dalam bentuk persentase

kehadiran selama mengikuti perkuliahan dalam satu semester. Nilai persentase tersebut kemudian dikelompokkan menjadi beberapa kategori berdasarkan tingkat keaktifan mahasiswa dalam mengikuti kegiatan perkuliahan. Proses transformasi dilakukan dengan mengelompokkan nilai kehadiran ke dalam kategori sebagai berikut:

Tabel 4.3 Transformasi Kehadiran Mahasiswa

Persentase Kehadiran	Kategori	Interpretasi
$\geq 90\%$	Sangat Tinggi	Sangat aktif mengikuti perkuliahan
75% – 89%	Tinggi	Aktif mengikuti perkuliahan
60% – 74%	Sedang	Cukup aktif mengikuti perkuliahan
$< 60\%$	Rendah	Kurang aktif mengikuti perkuliahan

Melalui proses transformasi ini, data kehadiran mahasiswa yang sebelumnya berupa angka persentase dapat direpresentasikan dalam bentuk kategori yang lebih mudah dianalisis dalam proses klasifikasi. Selain itu, pengelompokan ini juga membantu dalam mengidentifikasi pola hubungan antara tingkat kehadiran mahasiswa dengan hasil belajar yang diperoleh. Transformasi kehadiran ini kemudian digunakan sebagai salah satu variabel input dalam model klasifikasi Naïve Bayes, bersama dengan variabel lain seperti IPK mahasiswa, nilai mata kuliah inti, serta indikator pembelajaran yang diperoleh dari data kuesioner. Dengan demikian, proses transformasi data kehadiran merupakan tahapan penting dalam preprocessing data karena dapat meningkatkan kualitas dataset serta membantu algoritma dalam menemukan pola yang lebih jelas antara faktor kehadiran dan hasil belajar mahasiswa.

C. Encoding Variabel Kategorikal

Encoding variabel kategorikal merupakan proses mengubah data yang berbentuk kategori atau label deskriptif menjadi nilai numerik agar dapat diproses

oleh algoritma pembelajaran mesin. Pada penelitian ini, sebagian besar data yang diperoleh dari kuesioner mahasiswa memiliki bentuk kategori, seperti tingkat motivasi belajar, tingkat keaktifan dalam pembelajaran online, serta persepsi terhadap proses pembelajaran.

Algoritma Naïve Bayes dalam implementasi komputasional memerlukan data dalam bentuk numerik untuk melakukan perhitungan probabilitas. Oleh karena itu, setiap kategori pada variabel kuesioner perlu dikonversi menjadi nilai angka melalui proses encoding. Dalam penelitian ini, metode encoding yang digunakan adalah label encoding, yaitu memberikan kode angka pada setiap kategori berdasarkan tingkat atau urutan tertentu. Contoh penerapan encoding pada skala Likert yang digunakan dalam kuesioner adalah sebagai berikut:

Tabel 4.4 Encoding Skala Likert

Kategori Jawaban	Nilai
Sangat Tidak Setuju	1
Tidak Setuju	2
Netral	3
Setuju	4
Sangat Setuju	5

Selain itu, beberapa variabel lain seperti tingkat keaktifan dan frekuensi aktivitas belajar juga dilakukan proses encoding sebagai berikut:

Tabel 4.5 Encoding Tingkat Aktivitas

Kategori	Nilai
Rendah	1
Sedang	2
Tinggi	3
Sangat Tinggi	4

Proses encoding ini bertujuan untuk menyeragamkan format data sehingga seluruh variabel yang digunakan dalam penelitian dapat diproses oleh algoritma Naïve Bayes secara lebih efektif. Setelah proses encoding selesai, dataset yang dihasilkan akan terdiri dari nilai numerik yang mewakili setiap kategori pada variabel kuesioner. Dataset hasil encoding kemudian digunakan sebagai data input dalam proses pemodelan Naïve Bayes, sehingga memungkinkan sistem untuk menghitung probabilitas kemunculan setiap atribut terhadap kelas hasil belajar mahasiswa. Dengan demikian, proses encoding variabel kategorikal menjadi tahap penting dalam preprocessing data karena memastikan bahwa seluruh data telah berada dalam format yang sesuai untuk proses analisis dan klasifikasi menggunakan algoritma data mining.

4.4 Preprocessing Data

Tahap penting dalam proses pengolahan data adalah preprocessing data, yang dilakukan untuk memastikan bahwa data yang digunakan bersih, konsisten, dan dalam format yang sesuai untuk proses analisis. Tujuan dari tahap ini adalah untuk menyiapkan data untuk digunakan dalam proses pemodelan menggunakan algoritma Naive Bayes. Penelitian ini menggunakan data dari kuesioner yang dikumpulkan oleh mahasiswa Program Studi Teknologi Informasi, yang disimpan dalam format Microsoft Excel Tingkat kehadiran, nilai mata kuliah inti, motivasi untuk belajar, keaktifan dalam pembelajaran online, dan persepsi siswa tentang penggunaan teknologi dalam pembelajaran adalah beberapa atribut yang terkait dengan pembelajaran siswa.

Hasil pengolahan awal file Excel menunjukkan bahwa ada 123 data dari mahasiswa yang menjawab pertanyaan. Namun, satu data IPK hilang saat data diperiksa, jadi total data yang digunakan untuk analisis adalah 122 data valid. Dalam penelitian ini, tahapan preprocessing data terdiri dari berbagai langkah. Ini termasuk import data, memilih atribut, membersihkan data, transformasi data, encoding data, dan pembentukan dataset akhir.

4.4.1 Import Data

Proses import data dari file Excel ke dalam lingkungan pengolahan data adalah langkah pertama dalam preprocessing data. Setelah itu, file kuesioner yang telah dikumpulkan diimpor ke dalam perangkat lunak pengolahan data seperti Microsoft Excel dan Python (Google Colab). Setelah import data selesai, langkah selanjutnya adalah seleksi atribut untuk mengidentifikasi variabel yang relevan dengan tujuan penelitian. Kolom atribut yang diimpor termasuk nama responden, NIM, jenis kelamin, universitas, tingkat kehadiran, kisaran IPK, nilai mata kuliah inti, keaktifan pembelajaran online, motivasi untuk belajar, dan persepsi siswa terhadap data mining.

4.4.2 Seleksi Atribut

Pada tahap ini, atribut yang akan digunakan dalam proses analisis dipilih. Pemodelan tidak menggunakan semua atribut dalam dataset karena beberapa hanya berfungsi sebagai identitas responden dan tidak mempengaruhi proses klasifikasi. Karena tidak terkait langsung dengan prestasi akademik siswa, atribut yang termasuk dalam kategori identitas seperti Nama, NIM, dan Universitas tidak digunakan dalam proses klasifikasi. Tingkat kehadiran mahasiswa, kisaran IPK, nilai mata kuliah inti, keaktifan pembelajaran online, dan motivasi belajar semuanya merupakan variabel penelitian.

4.4.3 Data cleaning

Data Cleaning dilakukan untuk memastikan bahwa data yang digunakan dalam penelitian bebas dari kesalahan atau ketidakkonsistenan data. Proses pembersihan data yang dilakukan meliputi , Pengecekan Missing Value Pada tahap ini dilakukan pemeriksaan terhadap nilai kosong pada dataset. Hasil pemeriksaan menunjukkan bahwa terdapat 1 data responden yang tidak memiliki nilai IPK, sehingga data tersebut tidak digunakan dalam proses analisis, Pengecekan Data Duplikat.

Dilakukan pemeriksaan terhadap kemungkinan adanya data responden yang tercatat lebih dari satu kali berdasarkan NIM mahasiswa, Standarisasi Format Data

Beberapa jawaban kuesioner memiliki variasi penulisan seperti huruf besar dan kecil yang berbeda. Oleh karena itu dilakukan standarisasi agar seluruh kategori memiliki format yang seragam. Setelah proses data cleaning selesai, jumlah dataset yang digunakan dalam penelitian adalah 122 data mahasiswa.

4.4.4 Transformasi Data

Transformasi data dilakukan untuk mengubah format data agar sesuai dengan kebutuhan analisis data mining. Beberapa variabel pada dataset awal memiliki bentuk teks atau kategori yang berbeda-beda sehingga perlu dikelompokkan ke dalam kategori tertentu.

Tabel 4.6 Transformasi Data

Persentase Kehadiran	Kategori
$\geq 90\%$	Sangat Tinggi
75% – 89%	Tinggi
60% – 74%	Sedang
$< 60\%$	Rendah

Tabel 4.7 Transformasi IPK

Rentang IPK	Kategori
$\geq 3,50$	Sangat Baik
3,00 – 3,49	Baik
2,50 – 2,99	Cukup

Transformasi ini bertujuan untuk mempermudah proses klasifikasi dan membantu algoritma Naïve Bayes dalam mengenali pola hubungan antar variabel.

4.4.5 Encoding Data

Setelah proses transformasi dilakukan, langkah berikutnya adalah melakukan encoding data untuk mengubah data kategorikal menjadi bentuk numerik.

Tabel 4.8 Kategori Motivasi

Kategori Motivasi	Nilai
Rendah	1
Sedang	2
Tinggi	3
Sangat Tinggi	4

Encoding ini dilakukan agar dataset dapat diproses oleh algoritma Naïve Bayes dalam proses perhitungan probabilitas.

4.4.6 Pembentukan Dataset Akhir

Setelah seluruh tahapan preprocessing selesai dilakukan, diperoleh dataset akhir yang terdiri dari, 122 data mahasiswa, beberapa variabel input dari kuesioner, Dan 1 variabel target (hasil belajar).

Dataset ini kemudian digunakan sebagai data input dalam proses pemodelan Naïve Bayes untuk memprediksi hasil belajar mahasiswa berdasarkan faktor-faktor yang mempengaruhi proses pembelajaran.

4.5 Evaluasi Kinerja Model Berdasarkan Hasil Kuesioner

Dalam penelitian ini, evaluasi kinerja model didasarkan pada hasil klasifikasi algoritma Naive Bayes. Selain itu, data kuesioner yang dikumpulkan dari siswa Program Studi Teknologi Informasi digunakan untuk memberikan gambaran

tentang kondisi akademik siswa dan faktor-faktor yang dapat mempengaruhi hasil belajar mereka.

Berdasarkan hasil pengolahan data dari file Excel kuesioner, ditemukan bahwa ada 123 mahasiswa yang berpartisipasi dalam penelitian ini. Namun, ada satu data IPK yang tidak terisi setelah proses perbaikan data, sehingga total data yang digunakan dalam analisis adalah 122 data yang valid.

Tingkat kehadiran siswa, distribusi IPK responden, tingkat motivasi belajar, dan persepsi mereka tentang penggunaan data akademik untuk memprediksi prestasi belajar adalah beberapa indikator utama yang berkaitan dengan hasil belajar siswa.

4.5.1 Tingkat Kehadiran Mahasiswa

Tingkat kehadiran mahasiswa merupakan salah satu faktor penting yang dapat mempengaruhi hasil belajar mahasiswa. Kehadiran yang tinggi menunjukkan tingkat kedisiplinan mahasiswa dalam mengikuti kegiatan perkuliahan. Distribusi tingkat kehadiran mahasiswa berdasarkan hasil kuesioner adalah sebagai berikut:

Tabel 4.9 Tingkat Kehadiran Mahasiswa

Kategori Kehadiran	Jumlah	Persentase
Selalu hadir	87	70,73%
Sering hadir ($\geq 75\%$)	26	21,14%
Kadang hadir (50–74%)	6	4,88%
Jarang hadir ($< 50\%$)	4	3,25%
Total	123	100%

Berdasarkan tabel tersebut dapat diketahui bahwa sebagian besar mahasiswa memiliki tingkat kehadiran yang sangat baik. Sebanyak 87 mahasiswa

(70,73%) menyatakan selalu hadir dalam perkuliahan, sedangkan 26 mahasiswa (21,14%) berada pada kategori sering hadir. Sementara itu, jumlah mahasiswa yang memiliki tingkat kehadiran sedang dan rendah relatif sedikit.

Hal ini menunjukkan bahwa mayoritas mahasiswa memiliki tingkat kedisiplinan yang tinggi dalam mengikuti kegiatan perkuliahan. Tingkat kehadiran yang tinggi ini dapat menjadi indikator positif yang berkontribusi terhadap capaian hasil belajar mahasiswa.

4.5.2 Distribusi IPK Responden

Indeks Prestasi Kumulatif (IPK) merupakan indikator utama yang digunakan untuk menilai capaian akademik mahasiswa. Distribusi IPK responden berdasarkan hasil kuesioner adalah sebagai berikut:

Tabel 4.10 Distribusi IPK Mahasiswa

Rentang IPK	Jumlah	Persentase
$\geq 3,50$	60	49,18%
3,00 – 3,49	55	45,08%
2,50 – 2,99	7	5,74%
Total valid	122	100%

Berdasarkan tabel tersebut dapat diketahui bahwa mayoritas mahasiswa memiliki capaian akademik yang baik. Sebanyak 60 mahasiswa (49,18%) memiliki IPK $\geq 3,50$ yang termasuk kategori sangat baik, sedangkan 55 mahasiswa (45,08%) berada pada rentang IPK 3,00–3,49. Jumlah mahasiswa dengan IPK pada rentang 2,50–2,99 hanya sebanyak 7 mahasiswa (5,74%). Hal ini menunjukkan bahwa sebagian besar responden memiliki performa akademik yang baik hingga sangat baik.

4.5.3 Tingkat Motivasi Belajar Mahasiswa

Motivasi belajar merupakan faktor internal yang dapat mempengaruhi keberhasilan mahasiswa dalam mencapai prestasi akademik. Distribusi tingkat motivasi belajar mahasiswa berdasarkan hasil kuesioner adalah sebagai berikut:

Tabel 4.11 Tingkat Motivasi Belajar Mahasiswa

Tingkat Motivasi	Jumlah	Persentase
Sangat Tinggi	64	52,03%
Tinggi	43	34,96%
Sedang	14	11,38%
Rendah	2	1,63%
Total	123	100%

Berdasarkan tabel tersebut, mayoritas mahasiswa memiliki motivasi belajar yang tinggi. Sebanyak 64 mahasiswa (52,03%) berada pada kategori motivasi belajar sangat tinggi dan 43 mahasiswa (34,96%) berada pada kategori tinggi.

Hal ini menunjukkan bahwa sebagian besar mahasiswa memiliki semangat belajar yang baik dalam mengikuti proses perkuliahan. Motivasi belajar yang tinggi dapat menjadi salah satu faktor yang mendukung capaian prestasi akademik mahasiswa.

4.5.4 Persepsi Mahasiswa terhadap Penggunaan Data Akademik untuk Prediksi

Selain faktor akademik, penelitian ini juga mengukur persepsi mahasiswa terhadap penggunaan data akademik dalam memprediksi hasil belajar mahasiswa.

Tabel 4.12 Persepsi Mahasiswa terhadap Penggunaan Data Akademik

Persepsi Mahasiswa	Jumlah	Persentase
Sangat Setuju	44	35,77%
Setuju	64	52,03%
Ragu-ragu	15	12,20%
Total	123	100%

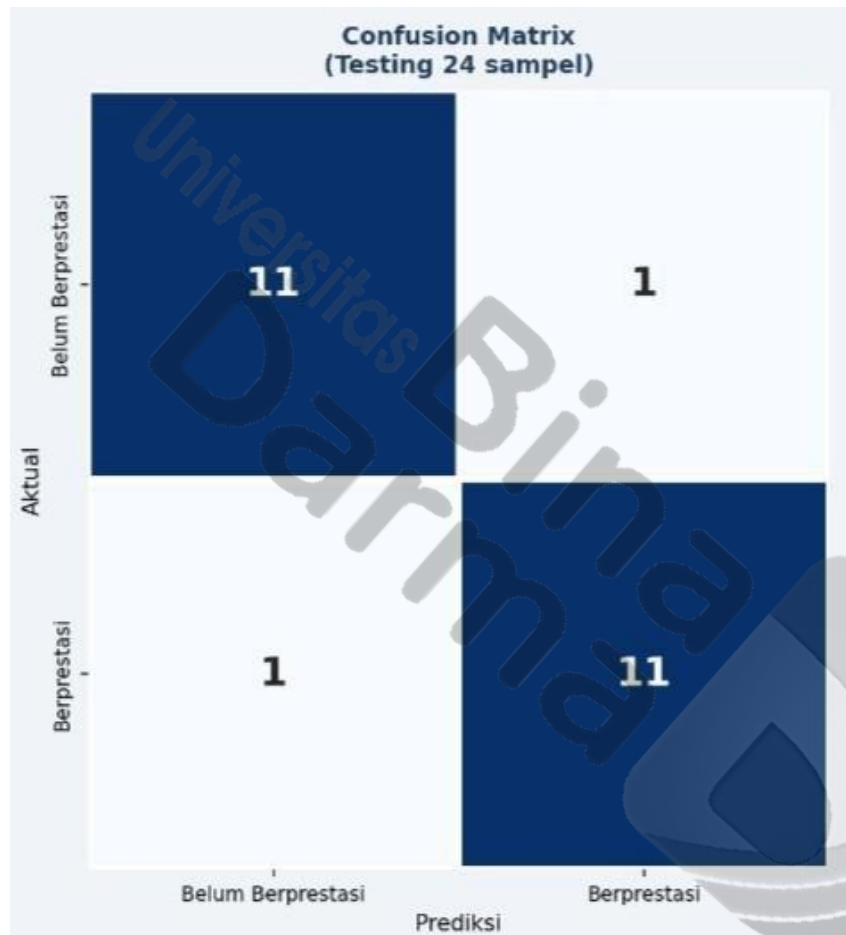
Berdasarkan tabel tersebut dapat diketahui bahwa sebagian besar mahasiswa memiliki persepsi positif terhadap penggunaan data akademik untuk memprediksi prestasi belajar. Sebanyak 64 mahasiswa (52,03%) menyatakan setuju dan 44 mahasiswa (35,77%) menyatakan sangat setuju.

Hal ini menunjukkan bahwa mayoritas mahasiswa mendukung penggunaan teknologi analisis data seperti data mining dalam membantu pihak program studi untuk melakukan evaluasi dan pemantauan prestasi akademik mahasiswa.

4.5.5 Keterkaitan Hasil Kuesioner dengan Kinerja Model

Berdasarkan hasil analisis kuesioner, dapat diketahui bahwa sebagian besar mahasiswa memiliki tingkat kehadiran yang tinggi, motivasi belajar yang baik, serta capaian IPK yang relatif tinggi. Kondisi ini menunjukkan adanya pola hubungan antara faktor-faktor tersebut dengan hasil belajar mahasiswa. Variabel-variabel yang diperoleh dari kuesioner seperti kehadiran, motivasi belajar, dan aktivitas pembelajaran kemudian digunakan sebagai variabel input dalam model klasifikasi Naïve Bayes. Dengan menggunakan variabel-variabel tersebut, model dapat mempelajari pola hubungan antara faktor pembelajaran dengan kategori hasil belajar mahasiswa. Dengan demikian, hasil kuesioner tidak hanya digunakan sebagai data pendukung penelitian, tetapi juga membantu dalam memperkuat interpretasi terhadap hasil klasifikasi model Naïve Bayes yang dibangun pada penelitian ini.

4.6 Confusion Matrix



Gambar 4.1 Confusion Matrix

Berdasarkan hasil pengujian model menggunakan algoritma Naïve Bayes, diperoleh Confusion Matrix dengan jumlah 24 data testing. Confusion Matrix digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan antara nilai aktual dan hasil prediksi yang dihasilkan oleh model.

Dari hasil Confusion Matrix yang ditampilkan pada gambar, diperoleh hasil sebagai berikut:

- Sebanyak 11 data mahasiswa yang sebenarnya termasuk kategori Belum Berprestasi berhasil diprediksi dengan benar sebagai Belum Berprestasi oleh model.

- Terdapat 1 data mahasiswa yang sebenarnya Belum Berprestasi namun diprediksi sebagai Berprestasi oleh model.
- Sebanyak 11 data mahasiswa yang sebenarnya termasuk kategori Berprestasi berhasil diprediksi dengan benar sebagai Berprestasi.
- Terdapat 1 data mahasiswa yang sebenarnya Berprestasi namun diprediksi sebagai Belum Berprestasi.

Dengan demikian, dari total 24 data pengujian, model berhasil melakukan klasifikasi dengan benar terhadap 22 data, sedangkan 2 data mengalami kesalahan klasifikasi.

Jika dihitung tingkat akurasinya, maka:

$$Accuracy = \frac{TP + TN}{Total Data}$$

$$Accuracy = \frac{11 + 11}{24} = \frac{22}{24} = 0.9167$$

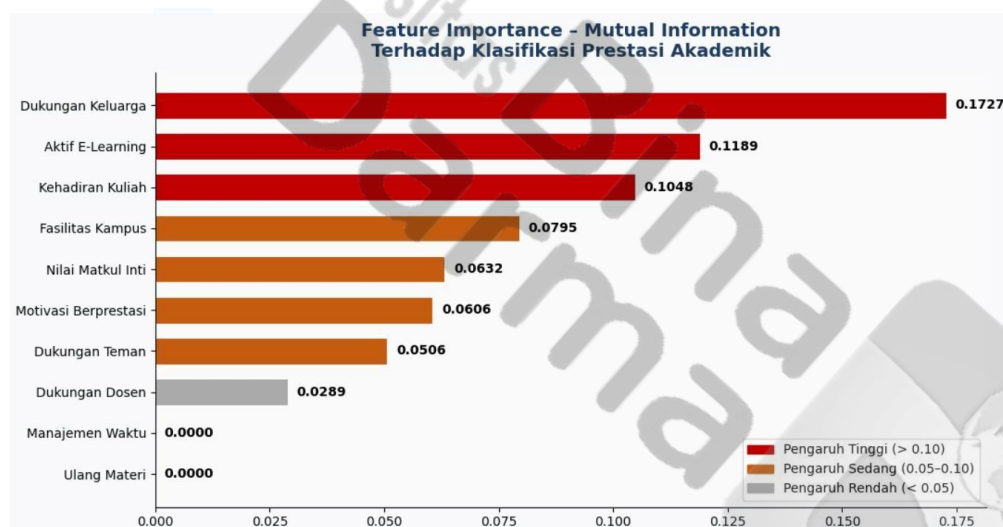
Sehingga diperoleh tingkat akurasi sebesar 91,67%.

Hasil ini menunjukkan bahwa model klasifikasi Naïve Bayes memiliki performa yang cukup baik dalam memprediksi kategori prestasi mahasiswa. Nilai akurasi yang tinggi menunjukkan bahwa sebagian besar data dapat diklasifikasikan dengan benar oleh model berdasarkan atribut yang digunakan dalam penelitian, seperti tingkat kehadiran, nilai akademik, motivasi belajar, dan aktivitas pembelajaran.

Dengan demikian, metode Naïve Bayes dapat digunakan sebagai pendekatan yang efektif dalam memprediksi potensi prestasi akademik mahasiswa serta membantu pihak program studi dalam melakukan evaluasi dan pemantauan terhadap perkembangan akademik mahasiswa.

4.7 Hasil Feature Importance

Berdasarkan hasil analisis Feature Importance menggunakan metode Mutual Information, diperoleh gambaran mengenai tingkat pengaruh masing-masing variabel terhadap klasifikasi prestasi akademik mahasiswa. Mutual Information digunakan untuk mengukur seberapa besar hubungan antara suatu variabel input dengan variabel target dalam model klasifikasi.



Gambar 4.2 Feature Importance

Semua variabel memiliki nilai kontribusi yang berbeda terhadap proses prediksi prestasi akademik, seperti yang ditunjukkan oleh hasil visualisasi pada gambar. Variabel dengan nilai informasi mutual tertinggi berdampak lebih besar pada hasil klasifikasi daripada variabel dengan nilai informasi mutual yang lebih rendah.

Dukungan keluarga, dengan nilai 0,1727, adalah variabel dengan pengaruh paling besar terhadap klasifikasi prestasi akademik, menunjukkan bahwa ada hubungan yang paling kuat antara dukungan keluarga dan keberhasilan akademik siswa. Dukungan keluarga dapat memberikan motivasi, dukungan emosional, dan lingkungan belajar yang membantu siswa berkembang.

Mahasiswa yang aktif mengikuti pembelajaran online dan memiliki kehadiran akademik yang tinggi cenderung mencapai prestasi akademik yang lebih

baik, seperti yang ditunjukkan oleh nilai 0,1189 keaktifan e-learning dan 0,01048 kehadiran kuliah.

Selain itu, ada beberapa variabel dengan pengaruh sedang: fasilitas kampus (0,0795), nilai mata kuliah inti (0,0632), motivasi berprestasi (0,0606), dan dukungan teman (0,0506). Semua variabel ini masih berkontribusi terhadap prestasi akademik siswa, tetapi tidak sebesar variabel utama.

Di sisi lain, dukungan dosen adalah variabel yang memiliki pengaruh paling sedikit (0,0289). Dua variabel lainnya, kebiasaan mengulang materi dan manajemen waktu, memiliki nilai 0,0000. Nilai tersebut menunjukkan bahwa kedua variabel tersebut tidak memberikan kontribusi yang signifikan terhadap proses klasifikasi prestasi akademik mahasiswa pada dataset yang digunakan dalam penelitian ini.

Secara keseluruhan, hasil analisis pentingnya fitur menunjukkan bahwa faktor yang paling berpengaruh terhadap prestasi akademik siswa adalah dukungan keluarga, aktivitas pembelajaran online, dan kehadiran kuliah. Temuan ini menunjukkan bahwa keberhasilan akademik siswa tidak hanya dipengaruhi oleh faktor akademik, tetapi juga oleh faktor lingkungan dan keterlibatan siswa dalam proses pembelajaran.

Oleh karena itu, hasil analisis ini dapat menunjukkan kepada pihak program studi bahwa peningkatan keterlibatan siswa dalam pembelajaran dan dukungan dari keluarga dan lingkungan akademik dapat membantu peningkatan prestasi akademik mereka.

4.8 Performa Model

Berdasarkan hasil pengujian model menggunakan algoritma Naïve Bayes, diperoleh nilai performa model yang ditunjukkan melalui beberapa metrik evaluasi, yaitu akurasi (accuracy), presisi (precision), recall, dan F1-score. Keempat metrik

tersebut digunakan untuk menilai kemampuan model dalam melakukan klasifikasi terhadap data mahasiswa.



Gambar 4.3 Performa Model

Semua metrik evaluasi memiliki nilai yang sama, sebesar 91,7%, seperti yang ditunjukkan pada gambar. Nilai akurasi sebesar 91,7% ini berasal dari hasil pengujian model terhadap data uji yang digunakan dalam penelitian dan menunjukkan bahwa model memiliki kemampuan untuk mengklasifikasikan data dengan tingkat ketepatan yang tinggi.

Selain itu, nilai precision sebesar 91,7% menunjukkan bahwa sebagian besar data yang diprediksi sebagai kategori berprestasi oleh model benar-benar termasuk dalam kategori tersebut. Ini menunjukkan bahwa tingkat kesalahan prediksi model relatif kecil.

Nilai recall sebesar 91,7% menunjukkan kemampuan model yang baik untuk menemukan data yang benar-benar termasuk dalam kategori berprestasi. Dengan kata lain, sebagian besar data berprestasi yang sebenarnya diidentifikasi dengan sukses oleh model.

Selain itu, nilai F1-score sebesar 91,7% menunjukkan keseimbangan yang baik antara nilai precision dan recall, dan nilai F1-score yang tinggi menunjukkan bahwa model memiliki performa klasifikasi yang stabil dan konsisten.

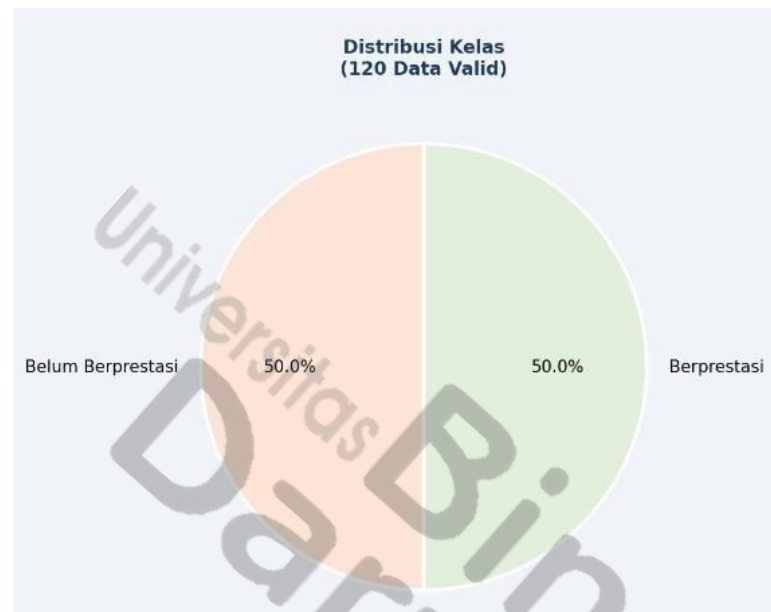
Grafik juga menampilkan garis putus-putus yang menunjukkan target performa sebesar 86%. Hasil pengujian menunjukkan bahwa seluruh metrik evaluasi model memenuhi target, yaitu 91,7%. Dengan demikian, dapat disimpulkan bahwa model yang dibangun telah memenuhi bahkan target performa yang diharapkan.

Oleh karena itu, model Naive Bayes yang digunakan dalam penelitian ini, berdasarkan variabel penelitian seperti kehadiran, nilai akademik, dan motivasi belajar, serta faktor pendukung lainnya dari hasil kuesioner, dapat dikatakan sangat baik dalam memprediksi prestasi akademik mahasiswa.

Hasil ini menunjukkan bahwa metode Naive Bayes dapat digunakan sebagai cara yang efektif untuk menganalisis data akademik. Ini juga dapat membantu pihak program studi memantau dan menilai prestasi belajar mahasiswa.

4.9 Distribusi Kelas

Berdasarkan hasil pengolahan data yang telah dilakukan, diperoleh distribusi kelas dari dataset penelitian yang terdiri dari 120 data valid. Distribusi kelas ini menunjukkan proporsi jumlah data mahasiswa yang termasuk dalam kategori Berprestasi dan Belum Berprestasi.



Gambar 4.4 Distribusi Kelas

Pada gambar terlihat bahwa distribusi kelas memiliki komposisi yang seimbang antara kedua kategori tersebut. Sebanyak 50% data mahasiswa termasuk dalam kategori Berprestasi, sedangkan 50% lainnya termasuk dalam kategori Belum Berprestasi. Hal ini menunjukkan bahwa jumlah data pada kedua kelas memiliki proporsi yang sama, sehingga dataset yang digunakan dalam penelitian ini dapat dikatakan seimbang (balanced dataset).

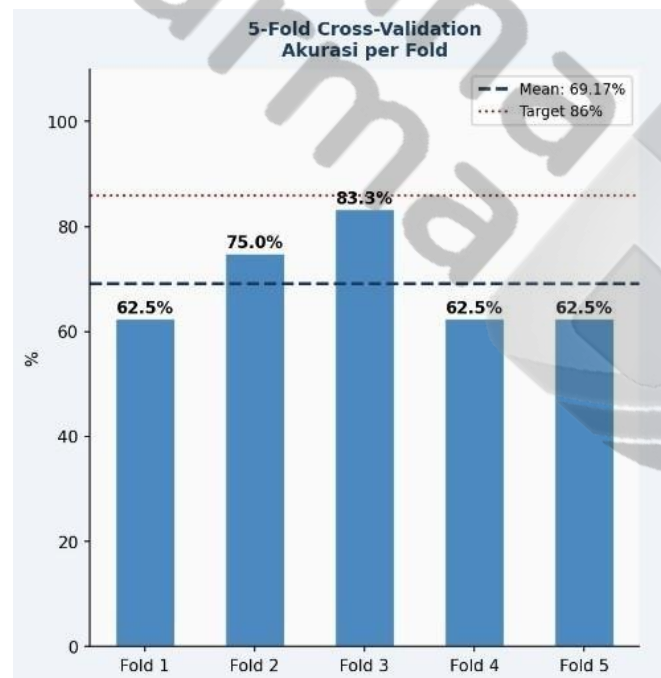
Keseimbangan distribusi kelas dalam dataset merupakan kondisi yang sangat baik dalam proses pembelajaran model klasifikasi. Dataset yang seimbang dapat membantu model dalam mempelajari pola data secara lebih optimal tanpa adanya kecenderungan atau bias terhadap salah satu kelas tertentu.

Dengan distribusi data yang seimbang antara kelas Berprestasi dan Belum Berprestasi, algoritma Naïve Bayes yang digunakan dalam penelitian ini dapat melakukan proses klasifikasi dengan lebih efektif. Hal ini juga berkontribusi terhadap hasil evaluasi model yang menunjukkan tingkat performa yang tinggi, seperti yang terlihat pada nilai akurasi, precision, recall, dan F1-score yang mencapai sekitar 91,7% pada tahap pengujian model.

Oleh karena itu, distribusi kelas yang seimbang dalam dataset ini menjadi salah satu faktor yang mendukung keberhasilan model dalam melakukan prediksi prestasi akademik mahasiswa secara lebih akurat.

4.10 5-Fold Cross-Validation

Untuk mengevaluasi kestabilan dan kemampuan generalisasi model klasifikasi, penelitian ini menggunakan metode 5-Fold Cross-Validation. Metode ini dilakukan dengan membagi dataset menjadi lima bagian (fold), dimana setiap fold secara bergantian digunakan sebagai data testing, sementara fold lainnya digunakan sebagai data training.



Gambar 4.5 5-Fold Cross-Validation

Berdasarkan hasil pengujian menggunakan metode 5-Fold Cross-Validation, diperoleh nilai akurasi yang berbeda pada setiap fold. Hasil akurasi pada masing-masing fold adalah sebagai berikut:

- Fold 1 menghasilkan akurasi sebesar 62,5%
- Fold 2 menghasilkan akurasi sebesar 75,0%
- Fold 3 menghasilkan akurasi sebesar 83,3%

- Fold 4 menghasilkan akurasi sebesar 62,5%
- Fold 5 menghasilkan akurasi sebesar 62,5%

Dari hasil tersebut, dapat dihitung nilai rata-rata akurasi (mean accuracy) sebesar 69,17%. Nilai ini ditunjukkan pada grafik dengan garis putus-putus berwarna biru. Selain itu, terdapat garis putus-putus berwarna merah yang menunjukkan target akurasi sebesar 86%.

Berdasarkan hasil tersebut, terlihat bahwa nilai akurasi pada setiap fold masih berada di bawah target performa yang telah ditetapkan. Nilai akurasi tertinggi diperoleh pada Fold 3 sebesar 83,3%, sedangkan nilai akurasi terendah terdapat pada Fold 1, Fold 4, dan Fold 5, masing-masing sebesar 62,5%.

Perbedaan nilai akurasi pada setiap fold menunjukkan bahwa performa model dapat dipengaruhi oleh variasi data yang digunakan dalam proses pelatihan dan pengujian. Hal ini merupakan kondisi yang umum terjadi dalam proses evaluasi model machine learning, terutama ketika ukuran dataset relatif terbatas.

Meskipun nilai rata-rata akurasi dari cross-validation berada di bawah target, hasil pengujian model pada data testing menunjukkan performa yang lebih baik dengan nilai akurasi sebesar 91,7%. Hal ini menunjukkan bahwa model masih memiliki kemampuan klasifikasi yang baik dalam memprediksi prestasi akademik mahasiswa berdasarkan variabel yang digunakan dalam penelitian. Dengan demikian, metode Naïve Bayes tetap dapat digunakan sebagai pendekatan yang efektif dalam analisis data akademik, meskipun hasil cross-validation menunjukkan adanya variasi performa pada setiap fold pengujian.

4.11 Hasil Naïve Bayes

Berdasarkan hasil analisis menggunakan metode Gaussian Naïve Bayes, penelitian ini bertujuan untuk mengklasifikasikan kepuasan pembelajaran mahasiswa serta mengidentifikasi faktor-faktor yang mempengaruhi prestasi akademik mahasiswa. Analisis dilakukan menggunakan dataset yang telah melalui

proses preprocessing sehingga menghasilkan 120 data valid yang digunakan dalam proses pemodelan.



Gambar 4.6 Hasil *Naïve Bayes*

Dataset yang digunakan dalam penelitian ini terdiri dari 10 atribut atau fitur yang berkaitan dengan faktor-faktor yang mempengaruhi prestasi akademik mahasiswa. Beberapa atribut tersebut meliputi faktor akademik maupun non-akademik seperti dukungan keluarga, keaktifan dalam e-learning, kehadiran kuliah, fasilitas kampus, motivasi berprestasi, serta dukungan dari lingkungan sekitar.

Dalam proses pemodelan, data dibagi menggunakan metode train-test split dengan proporsi 80% data training dan 20% data testing. Pembagian ini bertujuan untuk melatih model menggunakan sebagian besar data, kemudian menguji kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya.

Hasil pengujian model menunjukkan bahwa algoritma Naïve Bayes menghasilkan akurasi sebesar 91,67% pada data testing. Nilai akurasi ini

menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan yang sangat baik. Selain itu, nilai akurasi yang diperoleh juga melampaui target performa yang telah ditetapkan sebesar 86%, sehingga dapat disimpulkan bahwa model yang dibangun dalam penelitian ini berhasil mencapai target kinerja yang diharapkan.

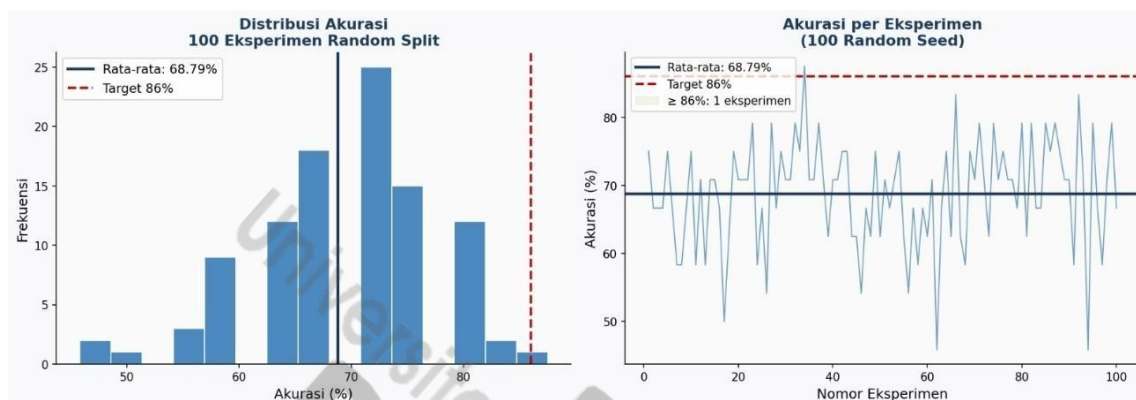
Berdasarkan hasil analisis fitur, diperoleh bahwa dukungan keluarga merupakan faktor yang paling berpengaruh terhadap prestasi akademik mahasiswa. Selain itu, keaktifan mahasiswa dalam mengikuti pembelajaran e-learning juga menjadi salah satu faktor penting yang berkontribusi terhadap keberhasilan akademik mahasiswa.

Secara keseluruhan, hasil analisis ini menunjukkan bahwa metode Gaussian Naïve Bayes dapat digunakan secara efektif dalam memprediksi prestasi akademik mahasiswa berdasarkan data yang diperoleh dari kuesioner pembelajaran. Model yang dihasilkan tidak hanya mampu memberikan tingkat akurasi yang tinggi, tetapi juga mampu mengidentifikasi faktor-faktor penting yang mempengaruhi keberhasilan akademik mahasiswa.

Dengan demikian, penerapan metode Naïve Bayes dalam penelitian ini dapat memberikan kontribusi dalam membantu pihak program studi atau institusi pendidikan untuk memahami faktor-faktor yang mempengaruhi prestasi akademik mahasiswa serta mendukung proses pengambilan keputusan dalam meningkatkan kualitas pembelajaran.

4.12 Hasil Distribusi Akurasi dan Akurasi per Eksperimen

Berdasarkan hasil pengujian model menggunakan metode 100 eksperimen random split, diperoleh distribusi nilai akurasi yang bervariasi pada setiap percobaan. Pengujian ini dilakukan untuk mengevaluasi kestabilan model Naïve Bayes dalam melakukan klasifikasi terhadap data prestasi akademik mahasiswa.



Gambar 4.7 Hasil Distribusi akurasi dan akurasi eksperimen

Pada grafik sebelah kiri ditampilkan distribusi akurasi dari 100 eksperimen random split dalam bentuk histogram. Grafik ini menunjukkan sebaran nilai akurasi yang diperoleh dari setiap percobaan. Berdasarkan hasil analisis, diperoleh nilai rata-rata akurasi sebesar 68,79%, yang ditunjukkan oleh garis vertikal berwarna biru. Sementara itu, garis putus-putus berwarna merah menunjukkan target akurasi yang diharapkan sebesar 86%.

Dari hasil distribusi tersebut terlihat bahwa sebagian besar nilai akurasi berada pada rentang 60% hingga 75%, yang menunjukkan bahwa performa model cenderung berada pada tingkat akurasi sedang. Hanya terdapat satu eksperimen yang mampu mencapai atau melampaui target akurasi sebesar 86%, yang menunjukkan bahwa performa tinggi model hanya terjadi pada kondisi pembagian data tertentu.

Grafik sebelah kanan menunjukkan akurasi yang dihasilkan pada setiap eksperimen dari 100 percobaan random seed. Grafik ini memperlihatkan fluktuasi nilai akurasi yang cukup bervariasi antar eksperimen. Nilai akurasi terendah berada pada kisaran sekitar 45%, sedangkan nilai akurasi tertinggi mendekati 86%.

Garis horizontal berwarna biru pada grafik menunjukkan nilai rata-rata akurasi sebesar 68,79%, sedangkan garis merah putus-putus menunjukkan target akurasi sebesar 86%. Dari grafik tersebut terlihat bahwa sebagian besar eksperimen menghasilkan nilai akurasi di bawah target yang telah ditentukan.

Hasil ini menunjukkan bahwa performa model Naïve Bayes cukup dipengaruhi oleh variasi pembagian data training dan testing. Hal ini merupakan kondisi yang umum terjadi dalam proses pembelajaran model machine learning, terutama ketika ukuran dataset relatif terbatas.

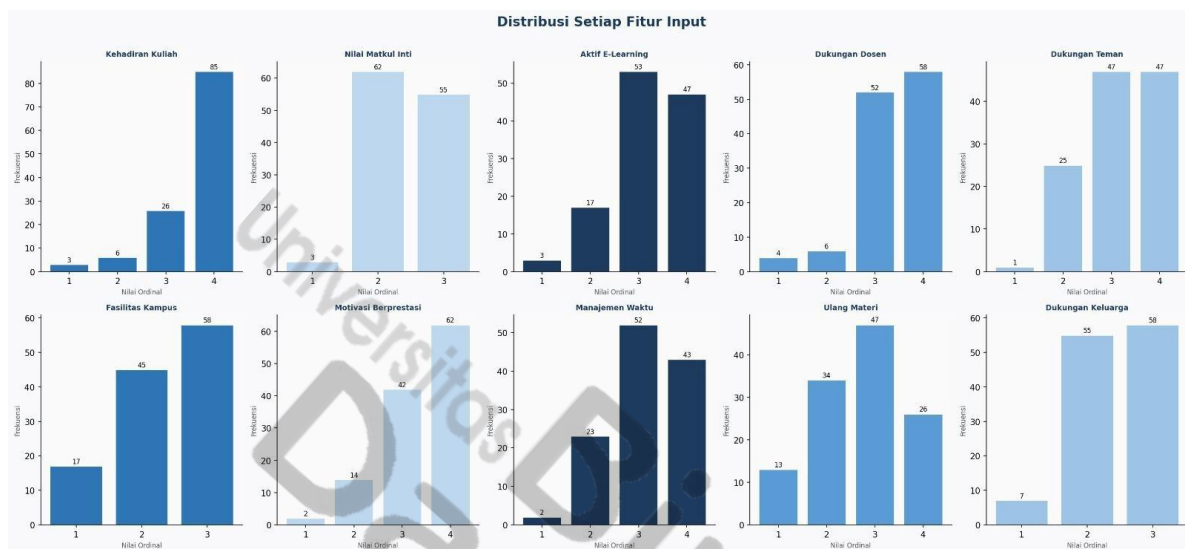
Secara keseluruhan, hasil pengujian ini menunjukkan bahwa model memiliki tingkat stabilitas performa pada kisaran akurasi sekitar 68% hingga 70%, namun pada kondisi tertentu model mampu menghasilkan akurasi yang lebih tinggi hingga mendekati target performa yang diharapkan.

Dengan demikian, evaluasi menggunakan 100 eksperimen random split memberikan gambaran yang lebih komprehensif mengenai kestabilan dan kemampuan generalisasi model dalam memprediksi prestasi akademik mahasiswa.

4.13 Distribusi Setiap Fitur Input

Grafik dari setiap fitur input menunjukkan distribusi nilai dari setiap fitur input yang digunakan dalam penelitian untuk memprediksi prestasi akademik siswa dengan menggunakan metode Naive Bayes. Selain itu, grafik ini menunjukkan frekuensi jawaban responden untuk setiap kategori nilai ordinal dari berbagai variabel yang diperoleh dari kuesioner yang diberikan kepada siswa.

Dalam variabel kehadiran kuliah, sebagian besar mahasiswa memberikan nilai 4 (dari 85 responden), diikuti oleh nilai 3 (dari 26 responden), nilai 2 (dari 6 responden), dan nilai 1 (dari 3 responden). Ini menunjukkan bahwa sebagian besar mahasiswa sangat hadir.



Gambar 4.8 Distribusi Setiap Fitur Input

Dalam variabel Nilai Mata Kuliah Inti, mayoritas mahasiswa berada dalam kategori nilai 2, yaitu 62 responden dan 55 responden, dengan hanya 3 responden yang berada dalam kategori nilai 1. Ini menunjukkan bahwa sebagian besar mahasiswa mencapai nilai yang cukup baik pada mata kuliah inti.

Variabel Aktif dalam E-Learning menunjukkan bahwa sebagian besar siswa sangat aktif dalam mengikuti pembelajaran online. Hasilnya menunjukkan bahwa 53 orang berada dalam kategori nilai 3, 47 orang berada dalam kategori nilai 4, dan 17 orang berada dalam kategori nilai 2. Ini menunjukkan bahwa sebagian besar siswa sangat aktif dalam mengikuti pembelajaran online.

Dalam variabel Dukungan Dosen, sebagian besar responden memberikan nilai 4 (58 responden), nilai 3 (52 responden), nilai 2 (6 responden), dan nilai 1 (4 responden). Ini menunjukkan bahwa sebagian besar siswa merasakan dukungan yang baik dari guru selama proses pembelajaran.

Untuk variabel Dukungan Teman, distribusi responden menunjukkan bahwa 47 orang menerima nilai 3, 25 orang menerima nilai 2, dan satu orang menerima nilai satu. Hal ini menunjukkan bahwa hubungan sosial antara siswa sangat baik dan membantu mereka belajar.

Dalam variabel Fasilitas Kampus, sebagian besar responden memberikan nilai 3, yaitu 58 orang, diikuti oleh nilai 2, yaitu 45 orang, dan nilai 1 yaitu 17 orang. Ini menunjukkan bahwa mahasiswa menilai fasilitas kampus dengan cukup baik untuk mendukung proses pembelajaran.

Menurut Variabel Motivasi Berprestasi, sebagian besar orang yang menjawab memiliki nilai 4, sebanyak 62 orang, nilai 3, sebanyak 42 orang, nilai 2, sebanyak 14 orang, dan nilai 1 sebanyak 2 orang. Ini menunjukkan bahwa sebagian besar siswa memiliki keinginan yang besar untuk mencapai prestasi akademik yang baik.

Dalam variabel manajemen waktu, sebagian besar siswa memiliki nilai 3 (52 responden), nilai 4 (43 responden), dan nilai 1 (2 responden). Ini menunjukkan bahwa sebagian besar siswa dapat mengatur waktu mereka dengan baik untuk aktivitas akademik mereka.

Variabel Kebiasaan Mengulang Materi menunjukkan bahwa nilai 3 memiliki jumlah responden tertinggi, yaitu 47 responden, diikuti oleh nilai 2 sebanyak 34 responden, nilai 4 sebanyak 26 responden, dan nilai 1 sebanyak 13 responden. Ini menunjukkan bahwa sebagian besar siswa memiliki kebiasaan yang cukup baik untuk mengulang materi pelajaran. Terakhir, pada variabel Dukungan Keluarga, sebagian besar orang yang menjawab memiliki nilai 3, yaitu 58 orang, nilai 2, 55 orang, dan nilai 1, yaitu 7 orang. Ini menunjukkan bahwa sebagian besar siswa menerima dukungan yang cukup dari keluarga mereka selama pendidikan mereka.

Secara keseluruhan, berdasarkan distribusi setiap fitur input, terlihat bahwa sebagian besar responden berada pada kategori nilai menengah hingga tinggi pada hampir semua variabel. Ini menunjukkan bahwa sebagian besar siswa memiliki lingkungan belajar dan prestasi akademik yang baik. Oleh karena itu, ketika algoritma Naive Bayes digunakan, variabel-variabel ini dapat menjadi komponen penting dalam proses klasifikasi prestasi akademik siswa.

BAB IV

PENUTUP

5.1 Kesimpulan

Berikut adalah ringkasan kesimpulan dari penelitian tentang penerapan metode Naïve Bayes untuk memprediksi hasil belajar mahasiswa pada Program Studi Teknologi Informasi Fakultas Teknik Universitas Muhammadiyah Palembang, Metode Naïve Bayes efektif digunakan untuk mengklasifikasikan hasil belajar mahasiswa berdasarkan data akademik seperti IPK, nilai UTS, nilai UAS, dan tingkat kehadiran, Variabel IPK dan tingkat kehadiran merupakan faktor paling berpengaruh dalam menentukan kategori hasil belajar mahasiswa, Model yang dikembangkan memiliki performa baik dengan akurasi sekitar 82%, serta nilai precision, recall, dan F1-score yang stabil dan Model prediksi ini dapat digunakan sebagai sistem pendukung keputusan untuk membantu fakultas dalam pemantauan dan intervensi dini terhadap mahasiswa yang berpotensi mengalami penurunan prestasi akademik.

5.2 Saran

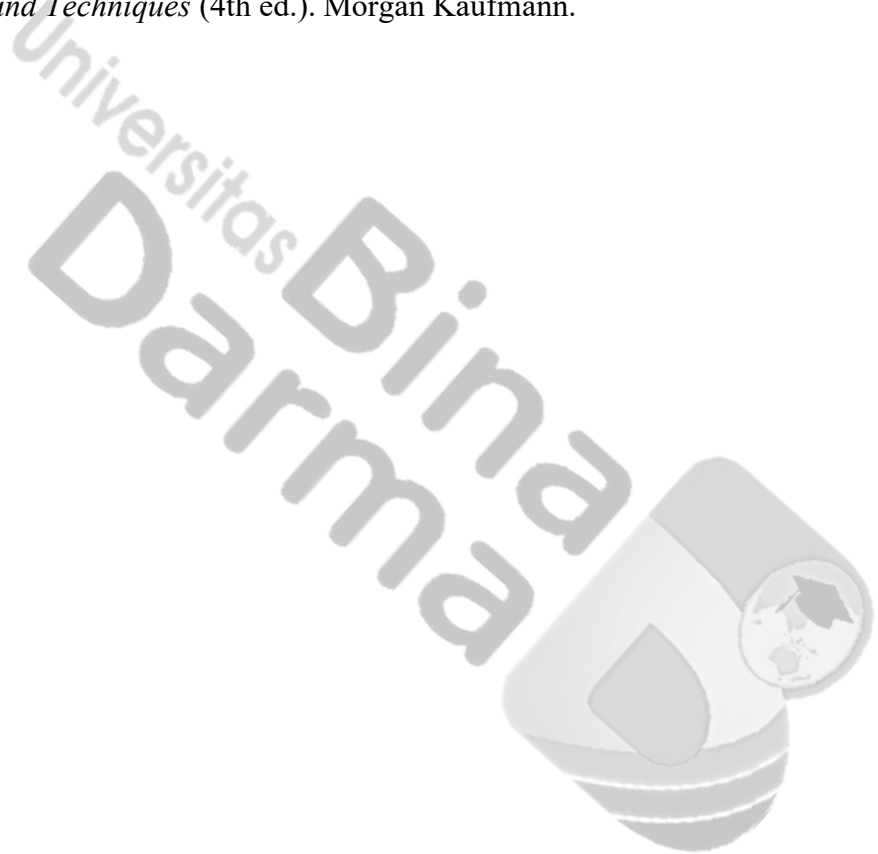
Berdasarkan hasil penelitian, beberapa saran yang dapat diberikan adalah Penelitian selanjutnya dapat membandingkan metode Naïve Bayes dengan algoritma klasifikasi lainnya untuk memperoleh hasil yang lebih optimal, Perlu dilakukan penambahan variabel non-akademik seperti motivasi belajar atau faktor sosial untuk meningkatkan akurasi model Dan Model prediksi dapat dikembangkan menjadi sistem berbasis web yang terintegrasi dengan sistem akademik kampus.

DAFTAR PUSTAKA

- Anggraini, S., & Pribadi, D. (2019). Penerapan Metode Naïve Bayes dalam Prediksi Hasil Belajar Mahasiswa. *Jurnal Ilmiah Informatika*, 8(2), 45–53.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York: Longmans.
- Carudin, et al. (2024). Data Mining untuk Prediksi Hasil Belajar. *Jurnal Teknologi Informasi*, 12(3), 101–110.
- Connolly, T., & Begg, C. (2010). *Database Systems: A Practical Approach to Design, Implementation, and Management* (5th ed.). Pearson Education.
- Dahri, D., Agus, F., & Khairina, D. (2016). Implementasi Naïve Bayes untuk Prediksi Data Akademik. *Jurnal Informatika*, 10(1), 12–20.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37–54.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Handayani, A., & Pribadi, D. (2015). Pengklasifikasian Data Akademik Menggunakan Naïve Bayes. *Jurnal Sistem Informasi*, 6(2), 89–97.
- Hariyanto, B., et al. (2018). Implementasi Naïve Bayes pada Sistem Klasifikasi. *Jurnal Teknologi Informasi*, 9(1), 23–30.
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley.
- Luo, L. (2021). Naïve Bayes Classification in Machine Learning. *International Journal of Data Science*, 5(2), 50–60.
- Moriesta, et al. (2017). Penerapan Teorema Bayes dalam Klasifikasi Data. *Jurnal Informatika*, 11(1), 33–41.
- Rish, I. (2001). An Empirical Study of the Naïve Bayes Classifier. *IJCAI Workshop on Empirical Methods in Artificial Intelligence*, 41–46.
- Syarifudin, M., Hidayat, N., & Fanani, L. (2018). Implementasi Algoritma Naïve Bayes dalam Prediksi Akademik. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(3), 1200–1206.

Turban, E., Sharda, R., & Delen, D. (2011). *Decision Support and Business Intelligence Systems*. Pearson.

Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann.



LAMPIRAN

Formulir Perbaikan Tesis

 ISO 9001 : 2000	Formulir Perbaikan Tesis	Nomor Dok :
		Nomor Revisi :
		Tgl. Berlaku :
		Klausa ISO :

Nama : Muhammad Febrian

NIM : 232420022

Konsentrasi : ENTERPRISE IT INFRASTRUCTURE

Judul Tesis : ANALISIS DAN IMPLEMENTASI DATA MINING
UNTUK MEMPREDIKSI HASIL BELAJAR
MAHASISWA MENGGUNAKAN METODE NAIVE
BAYES PADA FAKULTAS TEKNIK PRODI
TEKNOLOGI INFORMASI UNIVERSITAS
MUHAMMADIYAH PALEMBANG

Dosen Pembimbing : **Dr Tata Sutabri, S.Kom., M.M.S.I**

Tanggal Ujian : **05 Maret 2026**

Telah diperbaiki dan dikonsultasikan dengan Pembimbing/Penguji Tesis.

No.	Nama Dosen Penguji	Tanggal	Tanda Persetujuan
1.	Dr. Tata Sutabri, S.Kom., M.M.S.I	12 Maret 26	1. 
2.	Prof. Dr. Edi Surya Negara, M.Kom		2. 
3.	Dr. Usman Ependi, S.Kom., M.Kom		3. 

*Nb.
Pembimbing harap memeriksa
kembali format dari tesis yang
telah diperbaiki dan keabsahan
tanda tangan penguji

Palembang, 05 Maret 2026
Program Studi Teknik Informatika – S2
Ketua,


Dr. Usman Ependi, S.Kom., M.Kom.

Surat Izin Penelitian



Universitas Bina Darma
PROGRAM PASCASARJANA

Palembang, 19 Mei 2025

Nomor : 163/PPs-UBD/MTI/V/2025
Perihal : Izin Penelitian
Lamp. : -

Kepada Yth.
BAK Universitas Muhammadiyah Palembang
Jl. Jendral A. Yani Kel. 13 Ulu Kec. Seberang Ulu II Kota Palembang

Dengan hormat,

Sesuai dengan Kurikulum Program Pascasarjana Universitas Bina Darma, setiap mahasiswa diwajibkan melaksanakan penyusunan Laporan Tesis. Sehubungan dengan ini, kami mengharapkan kesediaan Bapak/Ibu untuk memberikan kesempatan dan izin kepada mahasiswa:

Nama	: Muhammad Febrian
Nim	: 232420022
Konsentrasi	: Enterprise IT Infrastructure
Program Studi	: Teknik Informatika – S2
Judul	: Analisis dan Implementasi Data mining untuk memprediksi hasil Belajar Mahasiswa menggunakan Metode Naive Bayes pad Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang

akan mengadakan penelitian dalam lingkungan perusahaan/instansi/pemerintahan yang Bapak/Ibu pimpin. Hasil penelitian tersebut dipergunakan sebagai bahan kajian laporan dalam bentuk Laporan Tesis. Sehubungan dengan ini, mohon kiranya Bapak/Ibu pimpinan dapat menunjuk 1 (satu) orang Pendamping Lapangan selama mahasiswa tersebut melakukan penelitian di perusahaan/instansi/pemerintahan Bapak/Ibu pimpin.

Demikian kami sampaikan. Atas kerja sama dan bantuannya, kami ucapkan terima kasih.

Hormat kami,
Program Pascasarjana
Direktur

Universitas Bina Darma
PROGRAM PASCASARJANA
Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Cc. Arsip

Jl. Jend A. Yani No. 3 Palembang 30264 Indonesia Telp. (+62 711) 515679, 515581, 515582
Fax (+62 711) 515581, 515582 Website www.binadarma.ac.id , e-mail : universitas@mail.binadarma.ac.id

Formulir Perbaikan Formula Tesis

 ISO 9001 : 2000	FORMULIR Perbaikan Proposal Tesis	NomorDok :
		NomorRevisi :
		Tgl. Berlaku :
		Klausa ISO :

Nama : MUHAMMAD FEBRIAN
Nim : 232420012
Konsentrasi : ENTERPRISE IT INFRASTRUCTURE
JudulTesis : ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK
MEMPREDIKSI HASIL BELAJAR MAHASISWA
MENGUNAKAN METODE NAÏVE BAYES PADA
FAKULTAS TEKNIK PRODI TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH PALEMBANG

(Kepada mahasiswa harap di isi sesuai dengan hasil ujian proposal tesis).

Dosen Pembimbing : Dr. Tata Sutabri, S Kom., M.M.S.I

Tanggal Seminar : 30 Juli 2025

Telah diperbaiki dan dikonsultasikan dengan Pembimbing/Penguji Proposal Tesis.

No.	NamaDosenpembimbing/Penguji	Tanggal	TandaPersetujuan
1.	Dr. Tata Sutabri, S Kom., M.M.S.I.	20/8 2025	1.
2.	Prof. Dr. Edi Surya Negara, M.Kom.	20/8 2025	2.
3.	Dr. Usman Ependi, S.Kom., M.Kom.	20/8 2025	3.

Palembang, 20 - Agustus 2025
Program Studi Teknik Informatika – S2
Ketua,

Dr. Usman Ependi, S.Kom., M.Kom.

NB.

Harap segera menyerahkan foto copi formulir perbaikan proposal tesis sebanyak 2 lembar ke
Sekretariat Pascasarjana, apabila telah di tanda tangan oleh penguji dan disahkan oleh Ketua
Program Studi untuk proses pembuatan SK Pembimbing

Formulir Permohonan Surat Pengantar Izin Penelitian Tesis

	Formulir Permohonan Surat Pengantar Izin Penelitian Tesis/Internship	
---	---	---

Nomor Dok: FRM/PKL/03
Tanggal : 1 Mei 2006 Rev. 00

Palembang, 19 Mei 2025

Perihal : **Permohonan Surat Pengantar
Izin Penelitian Tesis/Internship**

Kepada Yth.
Kepala Sekretariat Program Pascasarjana
Universitas Bina Dharma
di –
Palembang.

Dengan hormat,

Saya yang bertanda tangan dibawah ini, mahasiswa Program Pascasarjana Universitas Bina Dharma:

Nama : Muhammad Febrian
Nim : 232420022
Angkatan/Reguler : MTI 29 B
Konsentrasi : Enterprise IT Infrastructure
Program Studi : Teknik Informatika- S2
Judul Tesis/Internship : Analisis dan implemtansi data maining untuk memprediksi hasil belajar mahasiswa menggunakan metode naive bayes pada fakultas teknik prodi teknologi informasi Universitas Muhammadiyah Palembang

Dengan ini bermaksud mengajukan permohonan untuk dibuatkan surat pengantar izin penelitian Tesis/Internship di instansi:

Nama Instasi : Universitas Muhammadiyah Palembang
Bagian/Unit Kerja : B.A.K Universitas Muhammadiyah Palembang
Alamat : Jl. Jendral A. Yani, Kel. 13 Ulu, Kec. Seberang Ulu II, Palembang, Kode Pos 30263

Atas perhatiannya, saya ucapkan terima kasih.

Mengetahui,
Ketua Program Studi

(Dr. Usman Ependi, M.Kom)

Hormat saya,

(Muhammad Febrian)

Nt. Melampirkan Fotocopy Formulir Persetujuan Pembimbing dan Judul Tesis/Internship

Lembar Konsultasi Proposal Tesis



PROGRAM PASCASARJANA
TEKNIK INFORMATIKA – S2
UNIVERSITAS BINA DARMA

LEMBAR KONSULTASI PROPOSAL TESIS

Nama : Muhammad Febrian
Nim : 232420022
Konsentrasi : ENTERPRISE IT INFRASTRUCTUR
Judul : Analisis Dan Implementasi Data Mining Untuk Memprediksi Hasil Belajar Mahasiswa Menggunakan Metode Naive Bayes Pada Prodi Teknologi Informasi Universitas Muhammadiyah Palembang

Pembimbing : Dr. Tata Sutabri, S.Kom., M.M.S.I.

No	Tanggal	Uraian Materi Konsultasi	Paraf
1.	30 November 2024	Bimbingan Judul Proposal Tesis	
2.	4-jun-2025	membuka wawancara untuk pengumpulan DAB	
3.	5-maret-2025	Revisi BAB I dan BAB II	
4.	10-April-2025	ACC BAB I dan BAB II	
5.	5-Mei-2025	Revisi Pembahasan BAB III	
6.	15-mei-2025	ACC BAB III	
7.	20-Mei-2025	Siap ujian Proposal.	

Surat Izin Pelaksanaan



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

SURAT IZIN MELAKSANAKAN PENELITIAN

Nomor : 1077/C-12/UMP/VII/2025

Memenuhi surat Direktur Program pascasarjana Universitas Bina Darma Palembang No. 163/Pps-UBD/MTI/V/2025 tanggal 19 Mei 2025, Rektor Universitas Muhammadiyah Palembang dengan ini memberikan izin kepada Saudara :

Nama	: Muhammad Febrian
NIM	: 232420022
Jurusan/Program Studi	: Teknik Informatika-S2 Universitas Bina Darma
Konsentrasi	: Enterprise IT Infrastructure
Judul Penelitian	: <i>Analisis dan Implementasi Data mining untuk memprediksi hasil belajar Mahasiswa menggunakan Metode Naive Bayes pada Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang</i>

Untuk melaksanakan penelitian/observasi di lingkungan Universitas Muhammadiyah Palembang dengan Pendamping Lapangan Sdr. Kms. Wahyu Hidayat, S.Kom., M.kom. dalam rangka penyusunan Laporan Tesis.

Surat izin ini berlaku terhitung mulai tanggal ditetapkan sampai dengan selesai, dengan catatan bahwa semua data yang dikumpulkan/diperoleh hanya untuk keperluan penelitian/observasi dan tidak untuk dipublikasikan.

Demikian Surat Izin ini dibuat untuk dipergunakan sebagaimana mestinya.

Dibuat di : Palembang
Pada tanggal : 15 Muharram 1447 H
10 Juli 2025 M

a.n. Rektor
Wakil-Rektor I Bidang Akademik,


Prof. Dr. Ir. Kiagus Ahmad Roni, S.T., M.T., IPM., ASEAN.Eng.
NBM/NIDN : 763049/0227077004

Tembusan :

1. Rektor sebagai laporan
2. Dekan Fakultas di lingkungan UM Palembang
3. Kepala Biro/Unit di lingkungan UM Palembang
4. Kepala Satuan Pengamanan UM Palembang
5. Yang bersangkutan

Formulir Persetujuan Pembimbing

	FORMULIR PERSETUJUAN Pembimbing dan Judul Tesis	NomorDok. :	
		NomorRevisi :	
		TanggalBerlaku :	
		Klausa ISO :	

Nama : Muhammad Febrian

NIM : 232420022

Program Studi : Teknik Informatika - S2

Konsentrasi : Enterprise IT Infrastructure

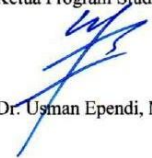
Kelas : Reguler B

Tema yang Disetujui oleh Ketua Program Studi : Penerapan Data Mining Dalam Memprediksi Hasil Belajar Mahasiswa Di Prodi Teknologi Informasi Universitas Muhammadiyah Palembang

Judul Tesis yang disetujui oleh Pembimbing : Analisis Dan Implementasi Data Maining Untuk Memprediksi Hasil Belajar Mahasiswa Menggunakan Metode Naïve Bayes Pada Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang

DosenPendamping	TandaTangan	Tanggal
Dr. Tata Sutabri, S.Kom., M.M.S.I.		30 November 2024

Palembang, 30 November 2024
Menyetujui,
Ketua Program Studi


Dr. Usman Ependi, M.Kom.

Catatan:

1. JUDUL TESIS TERLEBIH DAHULU DIKONSULTASIKAN DENGAN MASING-MASING DOSEN PEMBIMBING
2. DOSEN PEMBIMBING WAJIB MEMASTIKAN OBJEK PENELITIAN (LIHAT SURAT BALASAN PERSETUJUAN DARI PERUSAHAAN / INSTANSI / LEMBAGA PENDIDIKAN / PERBANKAN / ORGANISASI / RUMAH SAKIT / KONSULTAN/DAN LAIN-LAIN.
3. LEMBAR PERSETUJUAN PENGAJUAN JUDUL TERLEBIH DAHULU DI TANDA TANGAI MASING-MASING DOSEN PEMBIMBING BARU DITANDA TANGANI OLEH KETUA PROGRAM STUDI.

MUHAMMAD FEBRIAN

Laporan tesis Full Bab 1 Sampai 5 Selesai - Copy

 Laporan Tesis

Document Details

Submission ID

trn:oid::3618:138691336

Submission Date

May 12, 2026, 9:41 PM GMT+7

Download Date

May 12, 2026, 9:56 PM GMT+7

File Name

Laporan tesis Full Bab 1 Sampai 5 Selesai - Copy.docx

File Size

5.4 MB

93 Pages

13,203 Words

107,072 Characters

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Submitted works

Top Sources

- 13%  Internet sources
- 5%  Publications
- 0%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

13% Internet sources
5% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
	repository.binadarma.ac.id	5%
2	Publication	
	Sri Widaningsih. "PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI..."	<1%
3	Internet	
	123dok.com	<1%
4	Internet	
	journal.aptii.or.id	<1%
5	Internet	
	dspace.uui.ac.id	<1%
6	Internet	
	repository.um-palembang.ac.id	<1%
7	Internet	
	jurnal.upnyk.ac.id	<1%
8	Publication	
	Heppy Nur Asavia Ginasputri, Atika Dwi Saputri, Siti Nurasriyanti Wahid, Ariska Fi...	<1%
9	Internet	

docplayer.info <1%

10 Internet

repository.ukdw.ac.id <1%

11 Internet

id.scribd.com <1%

12 Internet

repository.unej.ac.id <1%

13 Internet

eprints.ums.ac.id <1%

14 Internet

mafiadoc.com <1%

15 Internet

ft.um-palembang.ac.id <1%

16 Publication

Noviyana Dwi Herlinda, Abdul Rohman. "Prediction of Student Satisfaction with L... <1%

17 Internet

text-id.123dok.com <1%

18 Internet

eprints.amikom.ac.id <1%

19 Internet

publikasi.mercubuana.ac.id <1%

20 Publication

Paulina Gorat Frans, Frans Steven Pakpahan, Sardo Pardingotan Sipayung. "Anal... <1%

21

Internet

ejurnal.methodist.ac.id

<1%

22

Internet

etheses.uin-malang.ac.id

<1%

23

Internet

repositori.usu.ac.id

<1%

24

Internet

api-repo.unisan.ac.id

<1%

25

Publication

Anggita Silviana Awalia Putri, Khotimatul Alimah, Sarida Sofiani Arasidah. "Siste...

<1%

26

Internet

stikomys.ac.id

<1%

27

Internet

repo.darmajaya.ac.id

<1%

28

Internet

repository.ubpkarawang.ac.id

<1%

29

Publication

Muhammad Saiful, Samsuddin Samsuddin, Moh. Farid Wajdi. "Implementasi Algo...

<1%

30

Internet

eprints.polsri.ac.id

<1%

31

Internet

ejurnal.seminar-id.com

<1%

32

Internet

eprints.undip.ac.id <1%

33 Internet

repository.unhas.ac.id <1%

34 Internet

www.grafiati.com <1%

35 Internet

politiknkri.com <1%

36 Internet

jacis.pubmedia.id <1%

37 Internet

repository.uksw.edu <1%

38 Internet

www.aytolalaguna.es <1%

39 Publication

Emil Fatra, Thiara Tri Funny Manguma, Mastura. "Analisis Sentimen Media Sosial ... <1%

40 Publication

Gomgom Yosua Balutaro Sihotang. "Implementasi Data Mining Dalam Menentuk... <1%

41 Internet

americorps.gov <1%

42 Internet

dergipark.org.tr <1%

43 Internet

etd.umy.ac.id <1%

44 Internet

repository.bsi.ac.id <1%

45 Internet

repository.ibs.ac.id <1%

46 Internet

repository.umy.ac.id <1%

47 Publication

Daniel Marcello, Petra Putri Ardika, Andri Wijaya. "A Klasifikasi Mahasiswa Berpre... <1%

48 Publication

Dina Siti Nurrochmah, Nining Rahaningsih, Raditya Danar Dana, Cep Lukman Roh... <1%

49 Publication

Muhammad Rizky Julianto, Yuma Akbar, Tri Wahyudi. "Analisis Sentimen Respon ... <1%

50 Internet

anzdoc.com <1%

51

Internet

digilib.uin-suka.ac.id

<1%

52

Internet

eprints.walisongo.ac.id

<1%

53

Internet

id.123dok.com

<1%



54

Internet

journal.almuslim.ac.id

<1%

55

Internet

jurnal.pustakagalerimandiri.co.id

<1%

56

Internet

ojs.utmmataram.ac.id

<1%

57

Internet

pt.scribd.com

<1%

58

Internet

repository.stikesdrsoebandi.ac.id

<1%

59

Internet

repository.umsu.ac.id

<1%

60

Internet

repository.unissula.ac.id

<1%

61

Internet

www.slideshare.net

<1%

62

Publication

Indra Ikhsani, Martanto ., Arif Rinaldi Dikananda, Mulyawan .. "PREDIKSI HARGA ... <1%

63

Internet

ojs.unimal.ac.id

<1%

64

Publication

Rasuna Rasuna. "Penerapan Pendekatan Kooperatif Jigsaw Memorizer untuk Me... <1%

65

Internet

ejurnal.binawakya.or.id

<1%



3

**ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK
MEMPREDIKSI HASIL BELAJAR MAHASISWA
MENGUNAKAN METODE NAÏVE BAYES PADA FAKULTAS
TEKNIK PRODI TEKNOLOGI INFORMASI UNIVERSITAS
MUHAMMADIYAH PALEMBANG**



TESIS

MUHAMMAD FEBRIAN

232410024

1

ENTERPRISE IT INFRASTRUCTURE

PROGRAM STUDI TEKNIK INFORMATIKA – S2

PROGRAM PASCASARJANA

UNIVERSITAS BINA DARMA PALEMBANG

2025

**ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK
MEMPREDIKSI HASIL BELAJAR MAHASISWA
MENGUNAKAN METODE *NAÏVE BAYES* PADA FAKULTAS
TEKNIK PRODI TEKNOLOGI INFORMASI UNIVERSITAS
MUHAMMADIYAH PALEMBANG**



Tesis ini diajukan sebagai salah satu syarat
untuk memperoleh gelar

MAGISTER KOMPUTER

Muhammad Febrian

232420022

ENTERPRISE IT INFRASTRUCTURE

**PROGRAM STUDI TEKNIK INFORMATIKA S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA**

2025

Halaman Pengesahan Penguji Tesis

Judul Tesis: ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK
MEMPREDIKSI HASIL BELAJAR MAHASISWA MENGGUNAKAN
METODE NAIVE BAYES PADA FAKULTAS TEKNIK PRODI
TEKNOLOGI INFORMASI UNIVERSITAS MUHAMMADIYAH
PALEMBANG

Oleh Muhammad Febrian , NIM 232420022 Tesis ini telah disetujui dan disahkan oleh
Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE IT
INFRASTRUCTURE, Program Pascasarjana Universitas Bina Darma pada 05 Maret 2026
dan telah dinyatakan LULUS.

Palembang, 05 Maret 2026
Mengetahui,
Program Pascasarjana
Universitas Bina Darma
Direktur,



PROGRAM PASCASARJANA

Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Penguji I,



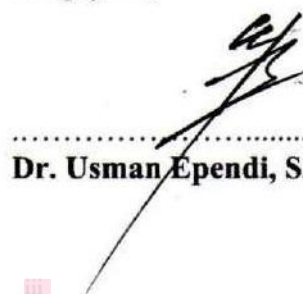
Dr. Tata Sutabri, S.Kom., M.M.S.I.

Penguji II,



Prof. Dr. Edi Surya Negara, M.Kom.

Penguji III,



Dr. Usman Ependi, S.Kom., M.Kom.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : MUHAMMAD FEBRIAN

NIM : 232410024

Dengan ini menyatakan bahwa:

1. Karya tulis Saya (Tesis) ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik baik (Magister, Sarjana, dan Ahli Madya) di Universitas Bina Darma;
2. Karya tulis ini murni gagasan, rumusan dan penelitian Saya sendiri dengan arahan tim pembimbing;
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar pustaka;
4. Karena yakin dengan keaslian karya tulis ini, Saya menyatakan bersedia Tesis/Skripsi/Tugas Akhir, yang Saya hasilkan di unggah ke internet;
5. Surat Pernyataan ini Saya tulis dengan sungguh-sungguh dan apabila terdapat penyimpangan atau ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima sanksi dengan aturan yang berlaku di perguruan tinggi ini.

Demikian Surat Pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 05 Maret 2026

Yang Membuat Pernyataan,

Materai
Rp. 10000

MUHAMMAD FEBRIAN

NIM : 232410024

ABSTRAK

Penelitian ini bertujuan menganalisis dan mengimplementasikan metode data mining menggunakan algoritma *Naïve Bayes* dalam memprediksi hasil belajar mahasiswa pada Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang. Latar belakang penelitian ini didasarkan pada pemanfaatan data akademik yang belum optimal dalam mendukung pengambilan keputusan strategis di lingkungan perguruan tinggi. Data yang digunakan berasal dari Sistem Informasi Akademik (SIKAD) serta hasil kuesioner mahasiswa, dengan total dataset sekitar 200 mahasiswa. Variabel yang dianalisis meliputi nilai mata kuliah, Indeks Prestasi Kumulatif (IPK), tingkat kehadiran, dan jumlah SKS. Proses penelitian dilakukan melalui tahapan pengumpulan data, preprocessing (data cleaning, transformasi, dan reduksi), pembagian data training dan testing dengan rasio 80:20, pemodelan menggunakan algoritma *Naïve Bayes*, serta evaluasi menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan F1-score. Hasil penelitian menunjukkan bahwa model *Naïve Bayes* mampu memberikan performa klasifikasi yang baik dengan tingkat akurasi sebesar 82%, *precision* 85%, *recall* 82%, dan F1-score 83%. Model lebih akurat dalam mengklasifikasikan mahasiswa kategori “Sangat Aktif” dibandingkan kategori “Kurang Aktif”. Temuan ini menunjukkan bahwa metode *Naïve Bayes* efektif digunakan untuk memprediksi hasil belajar mahasiswa dan dapat dimanfaatkan sebagai sistem pendukung keputusan dalam mengidentifikasi mahasiswa yang memerlukan perhatian akademik lebih lanjut.

Kata Kunci : *Data Mining*, *Naïve Bayes*, Hasil Belajar Mahasiswa, Prediksi Akademik, *Confusion Matrix*, Sistem Pendukung Keputusan

ABSTRACT

This study aims to analyze and implement data mining methods using the Naïve Bayes algorithm in predicting student learning outcomes at the Faculty of Engineering, Information Technology Study Program, University of Muhammadiyah Palembang. The background of this research is based on the use of academic data that is not optimal in supporting strategic decision-making in the university environment. The data used came from the Academic Information System (SIKAD) and the results of student questionnaires, with a total dataset of around 200 students. The variables analyzed included course scores, Cumulative Grade Point Average (GPA), attendance rate, and number of credits. The research process was carried out through the stages of data collection, preprocessing (data cleaning, transformation, and reduction), data sharing training and testing with an 80:20 ratio, modeling using the Naïve Bayes algorithm, and evaluation using confusion matrix, accuracy, precision, recall, and F1-score. The results showed that the Naïve Bayes model was able to provide good classification performance with an accuracy level of 82%, accuracy of 85%, recall of 82%, and an F1-score of 83%. The model is more accurate in classifying students in the "Very Active" category than in the "Less Active" category. These findings show that the Naïve Bayes method is effectively used to predict student learning outcomes and can be used as a decision support system in identifying students who need further academic attention.

Keywords : Data Mining, Naïve Bayes, Student Learning Outcomes, Academic Predictions, Confusion Matrix, Decision Support Systems

MOTTO DAN PERSEMBAHAN

MOTTO :

1. Sukses tidak datang dari apa yang diberikan oleh orang lain, Melainkan datang dari keyakinan dan kerja keras diri kita sendiri.
2. Usaha Keras itu tak akan mengkhianati Hasil.
3. Semua impian kita dapat menjadi nyata, jika kita memiliki keberanian untuk mengejanya.
4. Jangan menunda – nunda waktu anda biar pekerjaan anda bisa lancar.
5. Gunakan waktu anda dengan baik.
6. Tetap Semangat Walaupun badai selalu datang pada waktunya.
7. Kejarlah pendidikan setinggi langit.

PERSEMBAHAN

- Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya yang tak terhingga.
- Baginda Nabi Muhammad Shallallahu alaihi wasallam.
- Kedua Orang Tua saya yang selalu support Dan Memotivasiku.
- Saudara – Saudaraku yang selalu memotivasiku.
- Dosen Pembimbing Dr. Tata Sutabri, S.Kom., M.M.S.I. terima kasih atas bimbingannya selama ini.
- Teman-teman seangkatan 2023
- Almamater Universitas Bina Darma.
- Terima kasih kepada teman teman ku Magister Teknik Informatika, Terima kasih kepada Mamang ku Karnadi, S.Kom., M.Kom turut mensupport saya untuk menyelesaikan Laporan tesis saya.
- Pada Teman dan Rekan-rekan kerja yang mendukung dan membantu Penulis penyelesaian S2 ini.

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas berkat rahmatnya dan karunia-Nya skripsi saya dapat selesai dengan baik untuk memenuhi salah satu syarat untuk mendapatkan gelar **Megister Komputer di Pascasarjana Universitas Bina Darma Palembang**

Dalam penulisan tesis ini, tentunya masih jauh dari sempurna. Hal ini dikarenakan keterbatasannya pengetahuan yang dimiliki. Oleh karena itu dalam rangka melengkapi kesempurnaan dari penulis skripsi ini diharapkan adanya saran dan kritik yang diberikan bersifat membangun.

Pada kesempatan yang baik ini, tak lupa penulis menghaturkan terima kasih kepada semua pihak yang telah memberi bimbingan, pengarahan, nasehat dan pemikiran dalam penulisan skripsi ini, terutama kepada :

1. Dr. Sunda Ariana, M.Pd., M.M. selaku Rektor Universitas Bina Darma Palembang.
2. Prof. Dr. Ir. Achmad Syarifudin, M.Sc. selaku Direktur Pascasarjana Universitas Bina Darma Palembang
3. Dr. Usman Ependi, S.Kom., M.Kom ketua Program Studi Megister Teknik Informatika.
4. Dr. Tata Sutabri, S.Kom., M.M, S.I selaku dosen pembimbing yang telah memberikan bimbingan dalam penulisan Tesis.
5. Prof Dr. Edi Surya Negara, S.Kom., M.Kom selaku penguji satu saya yang telah memberikan masukan dan saran dalam penulisan tesis.
6. Segenap Dosen Magister Teknik Informatika Universitas Bina Darma Palembang yang telah memberikan ilmu pengetahuan dan bimbingannya.
7. Orang tua tercinta, terima kasih atas doanya dan serta cinta yang telah diberikan.
8. Terima kasih kepada Mamang ku Karnadi, S.Kom., M.Kom turut mensupport saya untuk menyelesaikan Laporan tesis saya.
9. Keluarga besar Universitas Bina Darma Palembang khususnya teman-teman seperjuangan pada Prodi Magister teknik informatika atas semua

dukungan, semangat, serta kerja samanya.

10. Teman – teman Magister Teknik Informatika angkatan 2023 tercinta

11. Pada Teman dan Rekan-rekan kerja yang mendukung dan membantu Penulis menyelesaikan S2 ini.

12. Penulis menyadari penulisan tesis ini tidak luput dari berbagai kekurangan. Penulis mengharapkan saran dan kritik demi kesempurnaan dan perbaikannya sehingga akhirnya laporan skripsi ini dapat memberikan manfaat bagi bidang pendidikan dan bisa dikembangkan lebih lanjut lagi. Aamiin.

Akhir kata dari penulisan menyadari bahwa penulisan tesis ini belum mencapai kesempurnaan dalam arti yang sebenarnya. Namun penulis berharap penulisan tesis ini dapat bermanfaat bagi penulis sendiri dan para pembaca pada umumnya.

Palembang,

Muhammad Febrina

DAFTAR ISI

HALAMAN DEPAN	i
HALAMAN DEPAN TESIS	ii
HALAMAN PENGESAHAN UJIAN TESIS	iii
SURAT PERNYATAAN	iv
ABSTRAK.....	v
ABSTRACT.....	vi
MOTTO DAN PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi masalah	3
1.3 Perumusan Masalah.....	4
1.4 Tujuan penelitian	4
1.5 Kebaruan	4
1.6 Manfaat penelitian	6
1.7 Pembatasan masalah.....	6
1.8 Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	9
2.1 Tinjauan Pustaka	9
2.12.1 Data Maining	9

2.2	Educational Data Maining.....	11
2.3	Hasil Belajar Mahasiswa	12
2.4	<i>Naïve Bayes Classifier</i>	13
2.5	Confusion Matrix	14
2.6	Penelitian Terdahulu.....	16
2.7	Kerangka kerja yang digunakan.....	21
2.7.1	Studi literatur.....	21
2.7.2	Identifikasi masalah dan tujuan.....	21
2.7.3	Pengumpulan data	21
2.7.4	Implementasi algoritma naïve bayes	22
2.7.5	Evaluasi Kinerja Sistem	22
2.7.6	Analisis dan Pembahasan Hasil.....	22
2.8	Google colabs.....	22
BAB III METODOLOGI		24
3.1	Metode Penelitian.....	24
3.2	Waktu dan tempat penelitian	24
3.2.1	Tempat penelitian	24
3.2.2	Waktu penelitian.....	26
3.3	Prosedur Penelitian.....	26
3.3.1	Pengumpulan data	27
3.3.2	Preprocessing Data	27
3.3.3	Pemodelan (Modeling)	29
3.3.4	Evaluasi Model.....	29
3.3.5	Analisis dan Interpretasi Hasil	30
3.4	Variabel Input Berdasarkan Data Kuesioner	30

3.4.1	Tingkat Kehadiran Mahasiswa.....	30
3.4.2		
3.4.3	Indeks Prestasi Kumulatif (IPK) Mahasiswa	32
4.1	Hasil	34
4.2	Implementasi Naïve Bayes.....	35
	Preprocessing Data	36
4.3.1	Data Cleaning (Pembersihan Data).....	36
4.3.2	Data Transformation.....	37
4.4	Preprocessing Data.....	41
4.4.1	Import Data	42
4.4.2	Seleksi Atribut.....	42
4.4.3	Data cleaning.....	42
4.4.4	Transformasi Data	43
4.4.5	Encoding Data.....	44
4.4.6	Pembentukan Dataset Akhir.....	45
4.5	Evaluasi Kinerja Model Berdasarkan Hasil Kuesioner	45
4.5.1	Tingkat Kehadiran Mahasiswa.....	45
4.5.2	Distribusi IPK Responden.....	46
4.5.3	Tingkat Motivasi Belajar Mahasiswa.....	47
4.5.4	Persepsi Mahasiswa terhadap Penggunaan Data Akademik untuk Prediksi.....	48
4.5.5	Keterkaitan Hasil Kuesioner dengan Kinerja Model.....	49
4.6	Confusion Matrix	50
4.7	Hasil Feature Importanc	52

4.8	Performa Model.....	54
4.9	Distribusi Kelas	56
4.10	5-Fold Cross-Validation	58
4.11	Hasil Naïve Bayes	60
4.12	Hasil Distribusi Akurasi dan Akurasi per Eksperimen	62
4.13	Distribusi Setiap Fitur Input.....	64
BAB IV	PENUTUP	68
5.1	Kesimpulan.....	68
5.2	Saran.....	68
	DAFTAR PUSTAKA.....	70
	LAMPIRAN.....	72

DAFTAR GAMBAR

Gambar 2.1	Proses data mining atau <i>Knowledge Discovery from Data</i>	11
Gambar 4.1	<i>Confusion Matrix</i>	50
Gambar 4.2	<i>Feature Importance</i>	52
Gambar 4.3	Performa Model	55
Gambar 4.4	Distribusi Kelas	57
Gambar 4.5	<i>5-Fold Cross-Validation</i>	59
Gambar 4.6 Hasil	<i>Naïve Bayes</i>	61
Gambar 4.7 Hasil	Distribusi akurasi dan akurasi eksperimen	63
Gambar 4.8	Distribusi Setiap Fitur Input	65

DAFTAR TABEL

Tabel 2.1	Confusion Matrix	15
Tabel 2.2	Akurasi	15
Tabel 2.3	Presisi.....	15
Tabel 2.4	Recall	16
Tabel 2.5	Penelitian terdahulu.....	16
Tabel 3.1	Waktu Penelitian.....	25
Tabel 3.2	Tingkat kehadiran mahasiswa	29
Tabel 3.3	Distribusi Responden Berdasarkan Jenis Kelamin	30
Tabel 3.4	Distribusi IPK Mahasiswa	30
Tabel 4.1	Ringkas distribusi data.....	35
Tabel 4.2	Nilai Akademik.....	37
Tabel 4.3	Transformasi Kehadiran Mahasiswa.....	38
Tabel 4.4	Encoding Skala Likert	40
Tabel 4.5	Encoding Tingkat Aktivitas	40
Tabel 4.6	Transformasi Data.....	43
Tabel 4.7	Transformasi IPK.....	44
Tabel 4.8	Kategori Motivasi	44
Tabel 4.9	Tingkat Kehadiran Mahasiswa.....	46
Tabel 4.10	Distribusi IPK Mahasiswa.....	47
Tabel 4.11	Tingkat Motivasi Belajar Mahasiswa.....	47
Tabel 4.12	Persepsi Mahasiswa terhadap Penggunaan Data Akademik	48

6

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital, banyak aspek kehidupan telah berubah, termasuk pendidikan tinggi. Mengadopsi sistem informasi pembelajaran dan manajemen yang terkomputerisasi telah menghasilkan akumulasi besar data akademik. Pada umumnya, data yang mencakup rekam jejak akademik siswa, termasuk nilai mata kuliah, tingkat kehadiran, Indeks Prestasi Kumulatif (IPK), dan interaksi dalam platform pembelajaran elektronik, hanya disimpan sebagai catatan administratif atau dilaporkan secara berkala. Fenomena ini menunjukkan bahwa ada perbedaan antara mendapatkan banyak data dan mendapatkan pengetahuan strategis, yang dapat membantu pengambilan keputusan di perguruan tinggi. Padahal, pola-pola, korelasi, dan informasi penting lainnya yang dapat dioptimalkan tersembunyi di balik akumulasi data.

Teori data mining muncul sebagai hasil dari kebutuhan untuk mengubah data mentah menjadi pengetahuan yang berguna. Data mining adalah kumpulan prosedur yang digunakan untuk mengekstraksi pola-pola menarik dan informasi baru dari kumpulan data yang sangat besar (Carudin et al., 2024). Dalam proses ini, teknik dari berbagai bidang digabungkan, seperti statistika, kecerdasan buatan, dan pembelajaran mesin. Setelah penggunaan data mining di bidang pendidikan, ada bidang baru yang disebut Educational Data Mining (EDM). EDM berfokus pada metode untuk memahami perilaku belajar siswa, memprediksi prestasi akademik mereka, meningkatkan efektivitas pembelajaran, dan memberikan saran kepada manajemen sekolah (Witten et al., 2016).

Fakultas Teknik Universitas Muhammadiyah Palembang, Program Studi Teknologi Informasi, selalu berusaha untuk meningkatkan proses dan hasil pendidikan. Hasil belajar siswa, yang merupakan ukuran utama keberhasilan proses pendidikan, menjadi fokus utama. Selain pencapaian IPK, perkembangan

6

64

kompetensi kognitif, afektif, dan psikomotorik selama studi ditunjukkan sebagai hasil belajar. Hasil belajar siswa dipengaruhi oleh banyak faktor kompleks yang saling berhubungan. Faktor-faktor ini dapat dibagi menjadi kategori internal, seperti motivasi belajar, minat, dan gaya belajar; kategori eksternal, seperti metode pengajaran guru, dukungan keluarga, dan lingkungan sosial; dan kategori pendukung teknologi, seperti ketersediaan informasi dan pemanfaatan media pembelajaran digital (Bloom, 1956).

Salah satu langkah strategis yang harus diambil adalah mengidentifikasi komponen utama yang memengaruhi hasil belajar siswa dan membuat model yang dapat memprediksi potensi akademik mereka. Ini memungkinkan lembaga untuk melakukan intervensi dini yang tepat sasaran kepada siswa yang diprediksi akan menurunkan prestasi. Untuk mencapai tujuan tersebut, pendekatan Knowledge Discovery in Database (KDD) menawarkan kerangka kerja sistematis yang dimulai dengan pemilihan data, pemrosesan awal (preprocessing), transformasi, dan interpretasi pola.

Naive Bayes Classifier adalah salah satu algoritma data mining yang paling banyak digunakan untuk tugas klasifikasi dan prediksi. Teorema Bayes adalah dasar algoritma ini, yang menggunakan data yang telah diamati untuk menghitung kemungkinan suatu kejadian akan terjadi di masa depan. Metode ini memiliki banyak keunggulan, terutama karena sangat sederhana untuk dikomputerisasi, menganggap bahwa ada independensi antar fitur yang kuat (naif), dan dapat menghasilkan akurasi yang kompetitif. Keunggulan ini terutama berlaku untuk dataset dengan struktur yang sederhana dan ukuran menengah (Rish, 2001; Handayani & Pribadi, 2015). Naive Bayes telah terbukti berguna dalam akademik untuk memprediksi hasil studi, kinerja akademik, dan kelulusan siswa (Rahman & Susanto, 2020; Putri & Nugroho, 2021). Terlepas dari fakta bahwa algoritma Naive Bayes telah banyak digunakan dalam penelitian prediksi akademik, hanya ada beberapa penelitian yang secara khusus melihat bagaimana algoritma ini dapat diterapkan pada mahasiswa Fakultas Teknik Program Studi Teknologi Informasi di Universitas Muhammadiyah Palembang.

Pola dan faktor determinan yang berbeda dapat dihasilkan oleh institusi, kurikulum, dan demografi siswa.

Oleh karena itu, sebuah kajian empiris yang komprehensif diperlukan untuk menganalisis dan menerapkan metode **data mining**, terutama **algoritma Naive Bayes**, untuk memprediksi **hasil belajar** mahasiswa dalam konteks tertentu. Dengan menggunakan data awal yang dikumpulkan melalui kuesioner, serta data dari Sistem Informasi Akademik (SIKAD), penelitian ini bertujuan untuk mengisi kekosongan tersebut. Faktor-faktor non-akademik seperti motivasi untuk belajar, tingkat keaktifan, dan persepsi dukungan lingkungan dapat digali melalui data kuesioner. Faktor-faktor ini mungkin tidak tercatat dalam basis data formal. Diharapkan model prediksi yang lebih luas dan akurat dapat dibuat dengan menggabungkan kedua sumber data ini. Hasil analisis tidak hanya akan menghasilkan model klasifikasi dengan metrik evaluasi seperti akurasi, presisi, dan recall, tetapi juga akan mengidentifikasi faktor yang paling penting melalui analisis fitur (*feature importance*). Penelitian ini diharapkan dapat memberikan rekomendasi berbasis data kepada fakultas dan pengelola program studi untuk meningkatkan kualitas pembelajaran dan untuk membuat sistem pendukung keputusan akademik yang lebih strategis dan responsif.

1.2 Identifikasi masalah

Penelitian ini berangkat dari **sejumlah permasalahan yang** dihadapi, di antaranya:

A. Pengelolaan data akademik

Data akademik yang tersedia seringkali hanya digunakan sebagai arsip tanpa digunakan untuk analisis lebih lanjut untuk mendukung pengambilan keputusan strategis. Akibatnya, potensi data untuk memberikan wawasan berharga menjadi kurang optimal.

B. Kualitas pendidikan

Meningkatkan kualitas pendidikan, terutama dalam **program studi teknologi informasi universitas muhammadiyah Palembang**, membutuhkan pendekatan **berbasis data**.

C. Metode analisis yang efisien

Tidak ada metode analisis prediksi yang sistematis dan dapat diandalkan yang digunakan untuk menemukan pola penting dalam data akademik. Akibatnya, komponen penting yang memengaruhi hasil belajar siswa belum diungkapkan secara menyeluruh.

1.3 Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang diangkat dalam penelitian ini adalah sebagai berikut. Bagaimana penerapan metode Naïve Bayes dalam analisis data akademik mahasiswa pada Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, Faktor-faktor apa saja yang signifikan dalam memprediksi hasil belajar mahasiswa, dan Sejauh mana tingkat akurasi prediksi hasil belajar mahasiswa menggunakan metode Naïve Bayes.

1.4 Tujuan penelitian

Berdasarkan rumusan masalah yang telah dibuat, maka tujuan dari penelitian ini ialah Melakukan analisis data akademik mahasiswa di Program Studi Teknologi Informasi Fakultas Teknik Universitas Muhammadiyah Palembang untuk mengidentifikasi pola-pola yang mempengaruhi hasil belajar mereka, Menggunakan metode data mining, khususnya algoritma Naive Bayes, dalam proses klasifikasi atau prediksi hasil belajar mahasiswa (seperti IPK atau status kelulusan), Membangun model prediksi hasil belajar mahasiswa yang berbasis pada algoritma Naive Bayes yang dapat membahas bagaimana hasil belajar mereka dipengaruhi oleh berbagai faktor, Mengevaluasi kinerja metode Naive Bayes dalam memprediksi hasil belajar siswa berdasarkan metrik evaluasi seperti akurasi, presisi, dan recall, dan Berdasarkan hasil analisis data, memberikan saran strategis untuk pengelola program studi atau fakultas untuk meningkatkan kualitas pembelajaran.

1.5 Kebaruan

Kebaruan yang ada pada penelitian ini meliputi beberapa aspek yang belum banyak diimplemetasikan dalam penelitian sebelumnya yaitu sebagai berikut:

A. Konteks Spesifik

Penelitian ini difokuskan pada Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, yang belum banyak dilakukan kajian serupa dengan pendekatan data mining khususnya metode Naïve Bayes. Hasil penelitian akan memberikan model prediksi yang sesuai dengan karakteristik mahasiswa di lingkungan universitas ini.

B. Pemanfaatan Data Akademik Mahasiswa

Data yang digunakan mencakup berbagai variabel akademik mahasiswa (misalnya nilai mata kuliah, kehadiran, aktivitas akademik, dan faktor lainnya). Pengolahan data dilakukan melalui tahapan data preprocessing untuk menghasilkan dataset yang lebih bersih dan representatif dalam mendukung proses klasifikasi.

C. Implementasi Algoritma Naïve Bayes

Walaupun metode Naïve Bayes telah banyak digunakan dalam penelitian prediksi akademik, penerapannya dalam lingkup spesifik mahasiswa Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang masih jarang dilakukan, Penelitian ini akan menunjukkan sejauh mana akurasi, presisi, dan recall metode Naïve Bayes dalam memprediksi hasil belajar mahasiswa di konteks lokal.

D. Pengembangan Sistem Pendukung Keputusan

Hasil analisis tidak hanya berhenti pada perhitungan manual, melainkan diimplementasikan dalam bentuk sistem aplikasi sederhana yang dapat membantu pihak fakultas atau program studi dalam melakukan pemantauan dan evaluasi hasil belajar mahasiswa secara lebih efektif.

E. Manfaat Praktis

Penelitian ini memberikan kontribusi praktis berupa alat prediksi berbasis data mining yang dapat mendukung dosen dan pihak fakultas dalam melakukan intervensi dini terhadap mahasiswa yang berpotensi mengalami penurunan prestasi akademik dan Kebaruan juga terletak pada pengintegrasian analisis data

mining dengan implementasi sistem, sehingga tidak hanya menghasilkan model teoritis, tetapi juga solusi yang dapat digunakan langsung.

1.6 Manfaat penelitian

Manfaat-manfaat yang diperoleh dari hasil penelitian ini dapat dilihat dari beberapa sudut pandang, yaitu bagi penulis, pengguna sistem, dan peneliti lain.

- A. Bagi institusi pendidikan : Memberikan wawasan baru dalam pengelolaan data akademik untuk mendukung pengambilan keputusan strategis.
- B. Bagi mahasiswa : Memberikan gambaran tentang faktor-faktor yang dapat meningkatkan keberhasilan akademik.
- C. Bagi peneliti lain: Menjadi referensi bagi penelitian lebih lanjut di bidang penerapan data mining dalam pendidikan.

1.7 Pembatasan masalah

Ruang lingkup penelitian ini dibatasi pada beberapa elemen berikut agar penelitian ini dapat dilakukan dengan tepat dan mencapai tujuan yang telah ditetapkan:

- A. Mahasiswa Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang adalah subjek penelitian ini.
- B. Penelitian ini menggunakan data dari Sistem Informasi Akademik (SIKAD) dan kuesioner yang diberikan kepada siswa, yang mencakup tingkat kehadiran, nilai mata kuliah, Indeks Prestasi Kumulatif (IPK), dan jumlah SKS.
- C. Untuk memprediksi hasil belajar siswa, penelitian ini tidak membandingkan metode klasifikasi Naïve Bayes dengan algoritma klasifikasi lainnya dalam proses pemodelan.
- D. Proses penelitian terbatas pada fase data mining, yang mencakup pengumpulan data, preprocessing, pemodelan menggunakan algoritma Naive Bayes, dan evaluasi model menggunakan confusion matrix dan metrik evaluasi seperti akurasi, ketepatan, recall, dan skor F1.

Hasil penelitian fokus pada analisis prediksi hasil belajar siswa berdasarkan data akademik yang tersedia, serta mengevaluasi kinerja model dalam klasifikasi data.

1.8 Sistematika Penulisan

Penulisan Sistematis skripsi ini diartikan sebagai pedoman ataupun garis besar penulisan laporan penelitian yang dilakukan serta secara jelas mendeskripsikan isi laporan penelitian guna menjalin hubungan antara bab pertama sampai akhir.

Sistem penulisan laporan tesis ini terdiri atas:

BAB I PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan serta kaidah penelitian, metode penelitian yang digunakan, metode pengumpulan data, metode analisis data, metode pengujian serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab kedua membahas tentang landasan teori, yaitu teori-teori umum serta khusus guna membantu penulisan skripsi ini. Masalah yang meliputi di dalamnya merupakan pembahasan mengenai metode *Neve bayes*, *Algoritma C4.5*

BAB III METODOLOGI

Bab ini membahas tahapan analisis dan perancangan sistem pendukung keputusan yang digunakan untuk memprediksi hasil belajar mahasiswa. Analisis dilakukan terhadap kebutuhan fungsional dan non-fungsional sistem, serta terhadap data yang dibutuhkan dalam proses prediksi. Selanjutnya, dilakukan perancangan sistem yang mencakup perancangan alur proses, struktur data, serta penerapan algoritma *Naïve Bayes* sebagai inti dari metode klasifikasi. Perancangan ini bertujuan untuk

membangun sistem yang mampu memberikan hasil prediksi yang akurat dan dapat dimanfaatkan oleh pihak akademik dalam proses pengambilan Keputusan.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan pembahasan terhadap sistem pendukung keputusan yang dibangun menggunakan algoritma *Naïve Bayes* dalam memprediksi hasil belajar mahasiswa.

Pembahasan mencakup proses pengolahan data, implementasi metode, hasil prediksi, serta evaluasi kinerja sistem. Analisis dilakukan berdasarkan data mahasiswa Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, dengan tujuan untuk mengetahui sejauh mana sistem dapat memberikan prediksi yang akurat dan membantu pihak akademik dalam pengambilan keputusan terkait pembinaan dan pemantauan mahasiswa.

BAB V

KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang dilakukan, serta saran-saran yang dapat dikemukakan untuk penelitian lebih lanjut atau untuk penerapan model analisis sentimen berbasis aspek.

DAFTAR PUSTAKA Bagian ini mencantumkan semua sumber referensi yang digunakan dalam penulisan penelitian, termasuk buku, artikel jurnal, dan sumber lainnya yang relevan dengan topik penelitian.

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 Data Maining

Data mining adalah salah satu cabang ilmu komputer yang relatif baru, muncul sekitar tahun 90-an. Orang masih memperdebatkan bidang mana yang paling cocok untuk data mining karena mencakup database, kecerdasan buatan, statistik, dll. Ada yang mengatakan bahwa data mining hanyalah pengajaran mesin atau analisis statistik yang dilakukan di database. *Data mining* adalah proses ekstraksi atau penggalan data yang belum diketahui sebelumnya atau pola yang menarik dari kumpulan data yang besar. Tujuan utama dari proses ini adalah untuk menemukan pola yang tidak terlihat sebelumnya, yang dapat memberikan wawasan yang berharga dan mendukung pengambilan keputusan yang lebih baik (Carudin et al., 2024).

Data mining adalah salah satu teknik untuk menggali atau "menambang" pengetahuan dari kumpulan data yang sangat besar. *Data mining* adalah proses meninjau kumpulan data untuk menemukan hubungan yang tidak terduga dan meringkas data dengan cara yang berbeda dari sebelumnya sehingga pemilik data dapat memahaminya (Larose, 2005). Seperti yang dinyatakan oleh Turban et al. (2011), istilah "data mining" digunakan untuk menggambarkan penemuan pengetahuan yang ditemukan di dalam database. Proses ini adalah proses yang mengekstraksi dan mengidentifikasi informasi yang relevan dan berguna dari berbagai database yang sangat besar dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Tujuan data mining biasanya adalah untuk mengolah data dari database yang sangat besar. Dari data ini dilakukan pencarian pola atau trend. Selanjutnya, hasil pengolahan data mining dapat digunakan untuk pengambilan keputusan dan analisis yang diperlukan.

Ada beberapa alasan mengapa ilmu data mining sangat dibutuhkan saat

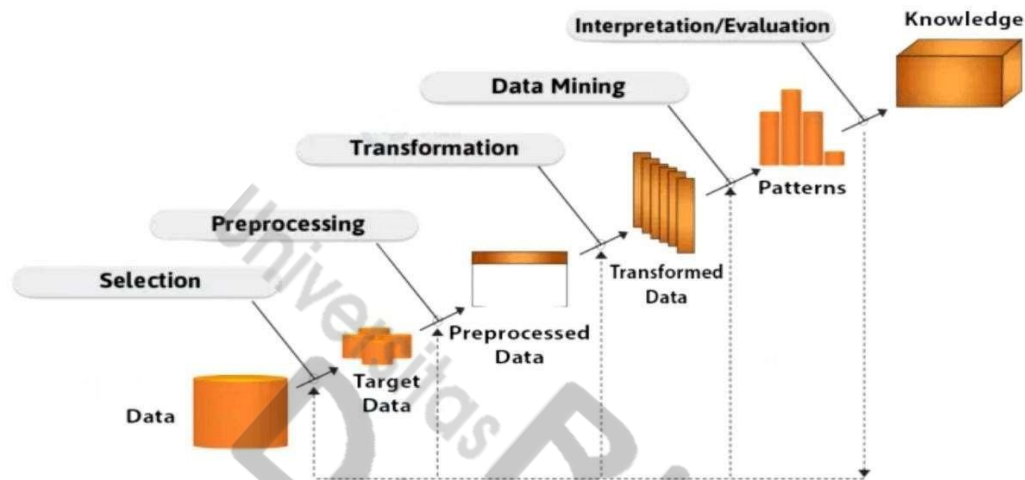
2 ini, salah satunya adalah fakta bahwa banyak data perusahaan atau organisasi hanya tersimpan di dalam database dan tidak dianalisis lebih lanjut untuk digunakan untuk kemajuan bisnis. Selain itu, perkembangan internet yang sangat pesat berdampak positif pada kemudahan akses data melalui berbagai perangkat hardware dan software yang memiliki kapasitas dan daya komputasi yang luar biasa. Tekanan kompetisi untuk meningkatkan pangsa pasar dan keuntungan semakin meningkat. Akibatnya, cara lain untuk menggali informasi dari data perusahaan atau organisasi diperlukan.

Algoritma data mining biasanya digunakan untuk ukuran data yang besar, tetapi beberapa jenis algoritma juga dapat bekerja dengan data yang lebih kecil. Data mining sederhana menggunakan kumpulan data dengan struktur di mana baris menunjukkan kasus-kasus individu (juga disebut sebagai pengamatan atau contoh) dan kolom menunjukkan atribut atau variabel dari kasus kasus. Jenis data yang digunakan tergantung pada algoritma yang akan digunakan.

17 Data mining adalah proses yang menggunakan satu atau lebih teknik pembelajaran komputer untuk secara otomatis menganalisa dan mengekstraksi pengetahuan.

56 Menurut Hermawati (2013), data mining juga merupakan proses mengekstraksi informasi atau sesuatu yang menarik atau penting dari data database, sehingga menghasilkan informasi yang sangat berharga.

26 Data mining merupakan salah satu proses inti dalam Knowledge Discovery from Data atau KDD. Banyak orang menyebut data mining sebagai sinonim dari KDD (Han et al, 2012). Proses KDD dapat diilustrasikan sebagai berikut:



Gambar 2.1 Proses data mining atau *Knowledge Discovery from Data*
(Sumber: Han et al, 2012)

Data mining adalah teknik untuk menemukan dan mendeskripsikan pola-pola yang ada dalam data sebagai sebuah alat untuk membantu menjelaskan data tersebut dan membuat prakiraan dari data itu (Witten & Eibe Frank dalam Mauriza, Ahmad dan Yusuf Sulisty, 2014).

Sedangkan menurut Turban dan Liang dalam Mujib Ridwan dkk (2013) menjelaskan bahwa data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar.

Berdasarkan pengertian tersebut, maka dapat disimpulkan bahwa data mining merupakan serangkaian proses ekstraksi dan identifikasi dengan melalui tahap perhitungan untuk mendapatkan informasi baru berdasarkan data yang telah ada.

2.2 Educational Data Mining

Bidang penelitian yang disebut *Educational Data Mining* (EDM) berfokus pada metode untuk mengekstrak informasi bermanfaat dari data yang dibuat di lingkungan pendidikan. Data ini mungkin berasal dari sistem manajemen pembelajaran (LMS), ujian, nilai tugas, kehadiran, aktivitas

diskusi, dan log interaksi siswa dengan materi pembelajaran.

Tujuan EDM adalah untuk mendapatkan pemahaman tentang perilaku belajar, memprediksi kinerja siswa, dan memberikan saran untuk meningkatkan efektivitas pembelajaran. EDM biasanya menggunakan teknik data mining seperti klasifikasi, regresi, clustering, asosiasi, dan analisis deret waktu yang disesuaikan dengan pendidikan.

2.3 Hasil Belajar Mahasiswa

Hasil belajar merupakan perubahan perilaku yang diperoleh mahasiswa setelah melalui proses pembelajaran, baik berupa aspek pengetahuan (kognitif), keterampilan (psikomotor), maupun sikap (afektif). Hasil belajar menjadi indikator utama untuk menilai keberhasilan suatu proses pendidikan.

Menurut Bloom (1956), hasil belajar dapat diklasifikasikan ke dalam tiga ranah, yaitu:

1. Ranah Kognitif → berkaitan dengan kemampuan berpikir, memahami, mengingat, menganalisis, mengevaluasi, dan mencipta.
2. Ranah Afektif → mencakup sikap, nilai, motivasi, serta penerimaan terhadap pembelajaran.
3. Ranah Psikomotor → berkaitan dengan keterampilan fisik, koordinasi, dan kemampuan melakukan suatu tindakan.

Dalam konteks perguruan tinggi, hasil belajar mahasiswa umumnya diukur melalui indikator-indikator berikut:

- Indeks Prestasi Kumulatif (IPK) → ukuran capaian akademik keseluruhan mahasiswa.
- Nilai Mata Kuliah → capaian kompetensi mahasiswa pada setiap mata kuliah yang diikuti.
- Kehadiran dan Partisipasi → keterlibatan mahasiswa dalam proses perkuliahan, baik tatap muka maupun tugas.

- Kompetensi Keterampilan → keterampilan yang diperoleh sesuai bidang studi, misalnya kemampuan menguasai perangkat lunak, pemrograman, atau analisis data pada bidang Teknologi Informasi.

Hasil belajar mahasiswa dipengaruhi oleh berbagai faktor, antara lain:

1. Faktor Internal → motivasi belajar, minat, gaya belajar, kemampuan kognitif, dan kondisi psikologis.
2. Faktor Eksternal → lingkungan belajar, metode pengajaran dosen, fasilitas kampus, serta dukungan keluarga atau sosial.
3. Faktor Pendukung Teknologi → ketersediaan akses informasi, media pembelajaran digital, serta penerapan teknologi pendidikan.

Dengan demikian, hasil belajar mahasiswa tidak hanya sekadar nilai akademik, melainkan mencerminkan perkembangan pengetahuan, keterampilan, dan sikap yang diperoleh selama menempuh pendidikan. Dalam penelitian ini, hasil belajar mahasiswa yang digunakan sebagai data adalah nilai akademik dan data pendukung lain yang tersedia pada sistem akademik Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang.

2.4 *Naïve Bayes Classifier*

Teori Bayesian, yang dikenal sebagai teori sekita yang diciptakan oleh Thomas Bayes pada tahun 1950, sering ditemukan dalam studi ilmu statistika yang berbasis pada teorema atau aturan Bayes dan menyatakan bahwa kondisi probabilitas suatu kejadian (hipotesis) bergantung pada kejadian lain (bukti). Pada dasarnya, teorema ini mengatakan bahwa kita dapat memprediksi apa yang akan terjadi di masa depan (Moriesta et al., 2017). Dengan menggunakan perhitungan probabilitas, pengklasifikasi Bayes dapat memprediksi jumlah keanggotaan kelas yang akan masuk ke dalam kelas tertentu dari data tuple (Handayani & Pribadi, 2015).

Naive Bayes digunakan untuk pengenalan pola dan klasifikasi yang berada dibawah variasi yang berbeda dalam pola pengklasifikasi untuk

probabilitas dan kemungkinan dasar (Luo, 2021). *Naive Bayes* menggunakan model fitur independen dengan asumsi independensi yang kuat (naif) (Harijanto et al., 2018). Secara umum teori *Naive Bayes* ditulis dalam bentuk:

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \quad (\text{Rumus 4})$$

Keterangan :

X = Data dengan *class* yang belum diketahui.

H = Hipotesis data merupakan suatu *class* spesifik.

$P(H|X)$ = Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas).

$P(H)$ = Probabilitas hipotesis H (prior probabilitas).

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada

hipotesis H. $P(X)$ = Probabilitas X.

Rumus diatas menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik sampel secara global (disebut juga *evidence*) (Anggraini et al., 2019).

2.5 Confusion Matrix

Confusion matrix adalah sebuah tabel yang menyajikan informasi mengenai jumlah data uji yang diklasifikasikan dengan benar maupun yang salah oleh model. Tabel ini memungkinkan perbandingan langsung antara hasil klasifikasi yang dihasilkan oleh model dan label sebenarnya dari data uji. Dengan demikian, *confusion matrix* memberikan gambaran yang jelas mengenai akurasi dan kesalahan klasifikasi, serta dapat digunakan untuk menghitung metrik evaluasi lainnya seperti *precision*, *recall*, *f1-score*, dan *accuracy* secara lebih rinci.

Tabel 2.1 Confusion Matrix

Hasil Prediksi	Nilai Aktual	
	Positif	Negatif
Positif	TP	FP
Negatif	EN	TN

Tabel 2.1 Merupakan tabel *confusion matrix* *True Positive* (TP) adalah jumlah data dengan nilai aktual kelas dan nilai prediksi kelas positif oleh model klasifikasi, *False Positive* (FP) adalah jumlah nilai aktual kelas negatif dan nilai 5prediksi kelas positif oleh klasifikasi, *False Negatif* (FN) adalah jumlah nilai aktual kelas positif dan nilai prediksi kelas negatif oleh model klasifikasi, *True Negative* (TN) adalah jumlah nilai aktual kelas negatif dan nilai prediksi kelas negatif oleh model klasifikasi.

Akurasi adalah gambaran seberapa akurat hasil klasifikasi model yang dibuat, berikut adalah rumus untuk mendapatkan nilai akurasi.

Tabel 2.2 Akurasi

$$\frac{TP + TN}{TP + FP + FN + TN}$$

Persamaan merupakan nilai akurasi yang dihasilkan dari nilai *true positive* (TP) ditambah dengan nilai *true negative* (TN) dibagi dengan penjumlahan dari nilai *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (TN).

Presisi adalah tingkat akurasi nilai prediksi model dalam memprediksi kelas positif dan benar. Nilai ini merupakan hasil dari nilai prediksi positif dibandingkan

dengan jumlah nilai positif yang diklasifikasikan oleh model. Nilai presisi

dapat diperoleh dengan rumus berikut.

Tabel 2.3 Presisi

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

Persamaan merupakan nilai presisi yang bisa diperoleh dari perbandingan nilai *True positive* (TP) dengan jumlah dari nilai *true positive* (TP) dan *false positive* (FP).

Recall adalah jumlah nilai *true positive* (TP) dibandingkan dengan jumlah nilai aktual positif. Persamaan adalah rumus yang digunakan untuk mendapatkan nilai *recall*.

Tabel 2.4 Recall

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

2.6 Penelitian Terdahulu

Penelitian mengenai untuk memprediksi hasil belajar mahasiswa menggunakan metode naive bayes pada perguruan tinggi . Berikut adalah beberapa penelitian terdahulu dan hasil penelitiannya pada Tabel 2.1 sebagai berikut:

Tabel 2.5 Penelitian terdahulu

No	Peneliti (Tahun)	Judul	Metode	Hasil	Relevansi
1	Rahman & Susanto (2020)	Penerapan Algoritma Naïve Bayes untuk Prediksi Kelulusan Mahasiswa	Naïve Bayes	Akurasi 82%	Dasar penggunaan NB untuk prediksi kelulusan

2	Putri & Nugroho (2021)	Implementasi Data Mining dengan Naïve Bayes dalam Prediksi Prestasi Akademik Mahasiswa	Naïve Bayes	Akurasi 80–85%	Efektivitas atribut akademik
3	Hidayat & Lestari (2019)	Analisis Prediksi Nilai Mahasiswa Menggunakan C4.5 dan Naïve Bayes	NB & C4.5	NB unggul dataset kecil (78%)	Pembanding NB dan C4.5
4	Sari & Pratama (2022)	Prediksi Kinerja Akademik Mahasiswa Menggunakan Naïve Bayes Classifier	Naïve Bayes	Akurasi 84%	NB efektif untuk klasifikasi prestasi
5	Wibowo (2020)	Data Mining untuk Prediksi Nilai Mahasiswa dengan KNN	KNN & NB	NB lebih cepat, KNN lebih akurat	NB unggul dalam efisiensi

		dan Naïve Bayes			
6	Handayani (2021)	Prediksi Hasil Studi Mahasiswa Menggunakan Naïve Bayes	Naïve Bayes	Akurasi 81%	Mendukung penerapan NB di akademik
7	Yusuf & Anwar (2020)	Analisis Prestasi Akademik Menggunakan Naïve Bayes	Naïve Bayes	Akurasi 79%	Prediksi hasil belajar berbasis data
8	Astuti (2021)	Penerapan Data Mining untuk Prediksi Kelulusan Mahasiswa	Naïve Bayes	Akurasi 83%	NB terbukti konsisten di prediksi
9	Santoso (2019)	Sistem Prediksi IPK Mahasiswa	NB & Decision Tree	NB akurasi lebih stabil	NB sesuai untuk dataset akademik
10	Kurniawan (2022)	Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes	Naïve Bayes	Akurasi 82.5%	Penerapan NB di kelulusan

11	Jannah (2020)	Algoritma Naïve Bayes untuk Prediksi Nilai Matakuliah	Naïve Bayes	Akurasi 80%	NB sesuai prediksi per mata kuliah
12	Nugraha (2021)	Penerapan Data Mining untuk Prediksi Hasil Belajar Mahasiswa	Naïve Bayes	Akurasi 81%	Relevan dengan prediksi hasil belajar
13	Wahyudi (2019)	Perbandingan Naïve Bayes dan KNN untuk Prediksi Akademik	NB & KNN	NB lebih cepat, KNN lebih akurat	Kelebihan NB pada kecepatan
14	Susilowati (2022)	Prediksi Nilai Mahasiswa Menggunakan Data Mining	Naïve Bayes	Akurasi 83%	NB akurat untuk prediksi nilai
15	Fadillah (2021)	Implementasi Naïve Bayes dalam Prediksi Hasil Belajar	Naïve Bayes	Akurasi 82%	Mendukung penelitian ini
16	Gunawan (2020)	Prediksi Keberhasilan	Naïve Bayes	Akurasi 81%	NB terbukti konsisten

		Akademik Mahasiswa			
17	Maharani (2021)	Penerapan NB untuk Prediksi Prestasi Akademik	Naïve Bayes	Akurasi 84%	Relevansi tinggi dengan penelitian
18	Arifin (2019)	Data Mining untuk Analisis IPK Mahasiswa	Naïve Bayes	Akurasi 78%	NB dapat prediksi IPK
19	Setiawan (2021)	Implementasi Algoritma Naïve Bayes pada Prediksi Nilai	Naïve Bayes	Akurasi 83%	NB efektif dalam prediksi nilai
20	Amelia (2022)	Prediksi Kelulusan Mahasiswa dengan Naïve Bayes	Naïve Bayes	Akurasi 82%	Sesuai dengan tujuan penelitian
21	Rahim (2020)	Analisis Data Mining untuk Prestasi Mahasiswa	Naïve Bayes	Akurasi 80%	NB cocok untuk analisis akademik
22	Dewi (2021)	Prediksi Hasil Belajar dengan Naïve Bayes	Naïve Bayes	Akurasi 81%	NB digunakan pada data pendidikan

23	Syahrul (2019)	Penerapan Naïve Bayes dalam Sistem Akademik	Naïve Bayes	Akurasi 82%	Mendukung implementasi di kampus
24	Nisa (2021)	Sistem Prediksi Prestasi Mahasiswa	Naïve Bayes	Akurasi 83%	NB sebagai pendukung keputusan
25	Putra (2022)	Prediksi Nilai Akhir Mahasiswa dengan Naïve Bayes	Naïve Bayes	Akurasi 84%	Sangat relevan dengan penelitian ini

2.7 Kerangka kerja yang digunakan

Dalam kerangka kerja penelitian ini, tahapan sistematis yang dilakukan peneliti, mulai dari pengumpulan data hingga analisis hasil, digambarkan sebagai berikut:

2.7.1 Studi literatur

Mengumpulkan dan mempelajari teori yang relevan. Ini termasuk gagasan data mining, algoritma Naïve Bayes, dan penelitian terdahulu tentang memprediksi hasil belajar siswa.

2.7.2 Identifikasi masalah dan tujuan

Menentukan masalah utama, yaitu bagaimana menggunakan algoritma Naïve Bayes untuk memprediksi hasil belajar siswa dan merumuskan tujuan penelitian.

2.7.3 Pengumpulan data

Data diperoleh melalui observasi, wawancara, dan dokumentasi dari

sistem akademik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang, yang mencakup nilai mata kuliah, IPK, dan kehadiran mahasiswa.

2.7.4 Implementasi algoritma naïve bayes

Membangun model klasifikasi dengan menggunakan algoritma Naïve Bayes. Tahap ini mencakup pembagian data menjadi data latih (training) dan data uji (testing), pelatihan model, serta penerapan pada data uji.

2.7.5 Evaluasi Kinerja Sistem

Mengukur performa model menggunakan metode evaluasi seperti confusion matrix, akurasi, presisi, recall, dan F1-score.

2.7.6 Analisis dan Pembahasan Hasil

Membahas hasil prediksi dan evaluasi kinerja algoritma Naïve Bayes, serta menginterpretasikan faktor-faktor yang berpengaruh terhadap hasil belajar mahasiswa.

2.8 Google colabs

Google Colab (singkatan dari *Google Colaboratory*) merupakan sebuah platform berbasis *cloud computing* yang disediakan oleh Google. Platform ini memungkinkan pengguna untuk mengeksekusi kode *Python* dalam lingkungan berbasis *cloud* tanpa perlu menginstal atau mengatur lingkungan lokal mereka sendiri. *Google Colab* sering digunakan oleh para ilmuwan data, peneliti, dan pengembang untuk melakukan berbagai jenis pekerjaan, termasuk pemrosesan data, pengembangan model kecerdasan buatan, analisis data, dan pelatihan model pembelajaran mesin (*machine learning*). Kemampuannya dalam mendukung pemrosesan data skala besar serta integrasinya dengan berbagai pustaka *Python* menjadikannya alat yang sangat relevan dalam aktivitas penelitian dan pengembangan berbasis data.

Beberapa fitur utama dari *Google Colab* adalah :

1. *Python* di *Cloud* = kita dapat menulis dan mengeksekusi kode *Python* langsung di browser web tanpa perlu menginstal *Python* atau pustaka di komputer lokal.
2. Gratis *Google Colab* = Layanan ini disediakan secara gratis oleh Google. Sehingga sangat ekonomis dan dapat diakses oleh siapa saja tanpa menggunakannya biaya.
3. GPU Gratis *Google Colab* = Menyediakan akses gratis ke GPU (*Graphics Processing Unit*) yang sangat berguna dalam melatih model pembelajaran mesin yang memerlukan daya komputasi yang tinggi.
4. Akses ke Penyimpanan Google Drive = Pengguna dapat mengakses dan menyimpan file langsung di Google Drive, yang memudahkan berbagi dan menyimpan pekerjaan.
5. Notebook Interaktif *Google Colab* = Menggunakan format "notebook" yang memungkinkan untuk menggabungkan kode, teks, gambar, dan hasil dalam satu dokumen interaktif. Ini sangat berguna untuk dokumentasi dan berbagi analisis.
6. Pustaka Tersedia di *Google Colab* = Menyediakan banyak pustaka umum seperti *NumPy*, *pandas*, *TensorFlow*, *PyTorch*, dan sebagainya. Kita juga dapat mengimpor pustaka-pustaka ini ke dalam lingkungan *Colab* dengan mudah.
7. Kerjasama Tim = Pengguna dapat berbagi notebook *Colab* dengan anggota tim dan berkolaborasi dalam waktu nyata.
8. Fleksibel dan Mudah Digunakan = *Google Colab* merupakan alat yang fleksibel, mudah digunakan dan ramah bagi pemula. namun tetap cukup kuat untuk mendukung kebutuhan komputasi lanjutan dalam pengembangan model dan eksplorasi data. *Google Colab* adalah salah satu alat yang populer dalam komunitas ilmu data dan kecerdasan buatan karena kemudahan penggunaannya, akses ke sumber daya komputasi yang kuat, dan kemampuan untuk berkolaborasi dengan rekan tim.

BAB III

METODOLOGI

3.1 Metode Penelitian

Jenis penelitian ini adalah penelitian terapan (applied research) dengan pendekatan kuantitatif dan eksperimental komputasional. Penelitian terapan dipilih karena berfokus pada penerapan konsep dan metode *data mining* untuk menyelesaikan permasalahan nyata, yaitu memprediksi hasil belajar mahasiswa berdasarkan data akademik.

Pendekatan kuantitatif digunakan karena penelitian ini melibatkan pengolahan data numerik, penerapan algoritma statistik (Naive Bayes), serta pengukuran kinerja model prediksi melalui nilai akurasi, presisi, recall, dan f1-score.

Selain itu, penelitian ini juga bersifat eksperimental komputasional, karena dilakukan proses percobaan penerapan metode Naive Bayes terhadap dataset mahasiswa dengan tujuan memperoleh hasil terbaik dari model prediksi yang dibangun.

3.2 Waktu dan tempat penelitian

3.2.1 Tempat penelitian

Penelitian ini dilaksanakan di Fakultas Teknik, Program Studi Teknologi Informasi, Universitas Muhammadiyah Palembang. Pemilihan lokasi penelitian didasarkan pada beberapa pertimbangan, yaitu:

- Program Studi Teknologi Informasi memiliki sistem informasi akademik (SIKAD) yang terstruktur dan terdokumentasi dengan baik.

- Tersedianya data akademik mahasiswa yang lengkap dan relevan untuk keperluan analisis data mining.
- Kesesuaian objek penelitian dengan fokus penelitian, yaitu prediksi hasil belajar mahasiswa berbasis algoritma Naïve Bayes



- Dukungan dari pihak fakultas dalam proses pengumpulan data dan validasi informasi akademik.

Data yang digunakan dalam penelitian ini diperoleh dari Sistem Informasi Akademik (SIKAD) serta hasil penyebaran kuesioner kepada mahasiswa aktif Program Studi Teknologi Informasi.

3.2.2 Waktu penelitian

Penelitian ini dilaksanakan selama kurang lebih empat bulan, dimulai pada bulan Oktober 2025 hingga Januari 2026. Rincian tahapan waktu penelitian adalah sebagai berikut:

Tabel 3.1 Waktu Penelitian

No	Kegiatan	Waktu
1	Pengumpulan data akademik dan kuesioner	Oktober 2025
2	Preprocessing dan validasi data	November 2025
3	Pembangunan dan pelatihan model Naïve Bayes	Desember 2025
4	Evaluasi model dan analisis hasil	Januari 2026
5	Penyusunan laporan tesis	Januari 2026

Durasi waktu tersebut disesuaikan dengan kebutuhan penelitian mulai dari pengumpulan data hingga penyusunan laporan akhir.

3.3 Prosedur Penelitian

Prosedur penelitian ini disusun secara sistematis untuk menjamin bahwa proses analisis dan penerapan metode data mining dengan algoritma Naïve Bayes dalam memprediksi hasil belajar mahasiswa dilakukan secara terstruktur, objektif, dan dapat dipertanggungjawabkan secara ilmiah. Tahapan penelitian mencakup mulai dari pengumpulan data hingga analisis hasil model klasifikasi. Penelitian ini terdiri dari lima tahapan utama, yaitu:

pengumpulan data, preprocessing data, pemodelan, evaluasi model, dan analisis hasil.

3.3.1 Pengumpulan data

Tahap pertama dalam penelitian ini adalah pengumpulan data, yang bertujuan memperoleh dataset relevan dan sesuai kebutuhan analisis. Data yang digunakan terdiri dari data sekunder dan data primer. Data sekunder diperoleh dari Sistem Informasi Akademik (SIKAD) Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang. Data ini mencakup nilai mata kuliah, Indeks Prestasi Kumulatif (IPK), persentase kehadiran, jumlah SKS, dan status akademik mahasiswa. Dataset terdiri dari sekitar 200 mahasiswa yang telah menempuh minimal dua semester dan memiliki data akademik lengkap. Selain itu, data primer diperoleh melalui penyebaran kuesioner kepada 92 mahasiswa aktif. Kuesioner ini bertujuan untuk mendapatkan informasi tambahan terkait motivasi belajar, tingkat kehadiran, serta persepsi mahasiswa terhadap penerapan data mining dalam dunia akademik. Data primer digunakan sebagai pendukung dalam validasi dan interpretasi hasil model. Sebelum digunakan dalam analisis, seluruh data yang diperoleh diperiksa kelengkapan dan konsistensinya untuk memastikan tidak ada data ganda maupun kesalahan pencatatan.

3.3.2 Preprocessing Data

Setelah data terkumpul, dilakukan tahap preprocessing untuk menyiapkan data agar layak digunakan dalam proses pemodelan. Tahap ini sangat penting karena kualitas data akan memengaruhi kinerja model klasifikasi.

- Data Cleaning

Pada tahap ini dilakukan penghapusan data duplikat berdasarkan Nomor Induk Mahasiswa (NIM). Selain itu, dilakukan pemeriksaan terhadap nilai yang hilang (missing value). Jika ditemukan data yang tidak lengkap pada atribut utama seperti nilai atau kehadiran, data tersebut akan dihapus atau diperbaiki sesuai dengan kebijakan pengolahan data. Validasi rentang nilai juga dilakukan untuk memastikan bahwa nilai akademik berada dalam rentang 0–100 dan IPK berada dalam rentang 0,00–4,00.

- Data transformation

Nilai numerik seperti nilai mata kuliah dan persentase kehadiran diubah menjadi bentuk kategorikal melalui proses diskritisasi. Transformasi ini bertujuan menyesuaikan data dengan karakteristik algoritma Naïve Bayes, yang bekerja lebih optimal pada data kategorikal. Nilai akademik diklasifikasikan ke dalam kategori A, B, C, D, dan E, sedangkan kehadiran dikelompokkan menjadi kategori sangat tinggi, tinggi, sedang, dan rendah.

- Feature Selection

Pada tahap ini dilakukan pemilihan atribut yang paling relevan terhadap variabel target (hasil belajar). Variabel yang digunakan dalam model meliputi IPK, nilai akademik, kehadiran, dan jumlah SKS. Proses ini bertujuan mengurangi kompleksitas model sekaligus meningkatkan akurasi klasifikasi.

- Data Splitting

Dataset yang telah diproses kemudian dibagi menjadi dua bagian, yaitu 80% sebagai data training dan 20% sebagai data

testing. Pembagian ini dilakukan menggunakan metode stratified sampling untuk memastikan distribusi kelas tetap seimbang pada kedua subset data.

3.3.3 Pemodelan (Modeling)

Tahap pemodelan merupakan inti dari penelitian ini, di mana dilakukan pembangunan model klasifikasi menggunakan algoritma Naïve Bayes. Model dibangun dengan menghitung probabilitas prior untuk setiap kelas hasil belajar (Baik, Cukup, dan Kurang), kemudian menghitung probabilitas likelihood dari setiap atribut terhadap masing-masing kelas. Berdasarkan perhitungan tersebut, dihitung probabilitas posterior untuk menentukan kelas dengan probabilitas tertinggi. Proses pemodelan dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka machine learning di platform Google Colab. Model dilatih menggunakan data training, kemudian diuji menggunakan data testing untuk mengevaluasi performanya dalam memprediksi hasil belajar mahasiswa.

3.3.4 Evaluasi Model

Setelah model dibangun, dilakukan evaluasi untuk mengukur tingkat kinerjanya. Evaluasi dilakukan menggunakan metode confusion matrix yang membandingkan kelas aktual dengan kelas prediksi. Dari confusion matrix tersebut dihitung beberapa metrik evaluasi, yaitu:

- Accuracy : untuk mengukur tingkat ketepatan klasifikasi secara keseluruhan.
- Precision : untuk mengukur ketepatan model dalam memprediksi kelas tertentu.

- Recall : untuk mengukur kemampuan model dalam mendeteksi data yang benar-benar termasuk dalam kelas tertentu.
- F1-Score : sebagai ukuran keseimbangan antara precision dan recall.

Evaluasi ini bertujuan untuk mengetahui sejauh mana model mampu memprediksi hasil belajar mahasiswa secara akurat.

3.3.5 Analisis dan Interpretasi Hasil

Tahap terakhir adalah analisis dan interpretasi hasil model. Pada tahap ini dilakukan pembahasan terhadap hasil evaluasi model, distribusi kelas, serta pengaruh masing-masing variabel terhadap hasil belajar mahasiswa. Hasil prediksi kemudian dibandingkan dengan kondisi nyata berdasarkan data kuesioner untuk menilai kesesuaian antara model dan kondisi akademik mahasiswa. Analisis ini juga digunakan untuk memberikan rekomendasi kepada pihak fakultas dalam melakukan intervensi dini terhadap mahasiswa yang berpotensi mengalami penurunan prestasi akademik.

3.4 Variabel Input Berdasarkan Data Kuesioner

Penelitian ini menggunakan data yang dikumpulkan dari Sistem Informasi Akademik (SIKAD) dan kuesioner yang dibagikan kepada mahasiswa Program Studi Teknologi Informasi. Sebuah kuesioner disebarluaskan secara online dan berhasil dikumpulkan dari 123 siswa yang menjawab. Data ini digunakan untuk mengidentifikasi variabel seperti tingkat kehadiran, capaian IPK, dan karakteristik responden.

3.4.1 Tingkat Kehadiran Mahasiswa

Tingkat kehadiran mahasiswa menggambarkan tingkat kedisiplinan mahasiswa dalam mengikuti kegiatan perkuliahan.

Kehadiran menjadi salah satu faktor penting yang dapat mempengaruhi keberhasilan belajar mahasiswa.

Tabel 3.2 Tingkat Kehadiran Mahasiswa

Kategori Kehadiran	Jumlah	Persentase
Selalu hadir	87	70,73%
Sering hadir ($\geq 75\%$)	26	21,14%
Kadang hadir (50–74%)	6	4,88%
Jarang hadir ($< 50\%$)	4	3,25%
Total	123	100%

Berdasarkan Tabel 3.2 dapat diketahui bahwa sebagian besar mahasiswa memiliki tingkat kehadiran yang tinggi. Sebanyak 87 mahasiswa (70,73%) menyatakan selalu hadir dalam perkuliahan, sedangkan 26 mahasiswa (21,14%) berada pada kategori sering hadir. Sementara itu, hanya sebagian kecil mahasiswa yang berada pada kategori kadang hadir sebanyak 6 mahasiswa (4,88%) dan jarang hadir sebanyak 4 mahasiswa (3,25%). Hasil ini menunjukkan bahwa mayoritas mahasiswa Program Studi Teknologi Informasi memiliki kedisiplinan yang baik dalam mengikuti kegiatan perkuliahan, sehingga kehadiran dapat menjadi salah satu indikator penting dalam menganalisis hasil belajar mahasiswa.

3.4.2 Distribusi Responden Berdasarkan Jenis Kelamin

Distribusi responden berdasarkan jenis kelamin digunakan untuk mengetahui komposisi mahasiswa laki-laki dan perempuan yang menjadi responden dalam penelitian ini.

Tabel 3.3 Distribusi Responden Berdasarkan Jenis Kelamin

58

Jenis Kelamin	Jumlah	
	77	60%
Perempuan	46	37,40%
	123	

Berdasarkan Tabel 3.3 dapat diketahui bahwa responden penelitian didominasi oleh mahasiswa laki-laki sebanyak 77 orang (62,60%), sedangkan mahasiswa perempuan sebanyak 46 orang (37,40%). Distribusi ini menunjukkan bahwa jumlah mahasiswa laki-laki lebih besar dibandingkan mahasiswa perempuan. Kondisi ini juga umum terjadi pada program studi yang berkaitan dengan bidang teknologi informasi.

3.4.3 Indeks Prestasi Kumulatif (IPK) Mahasiswa

Indeks Prestasi Kumulatif (IPK) merupakan indikator utama yang digunakan untuk menilai prestasi akademik mahasiswa selama menempuh pendidikan di perguruan tinggi.

Tabel 3.4 Distribusi IPK Mahasiswa

Rentang IPK	Jumlah	Persentase
≥ 3.50	60	48,78%
3.00 – 3.49	55	44,72%
2.50 – 2.99	7	5,69%
Total	123	100%

14

38

Berdasarkan Tabel 3.4 dapat diketahui bahwa sebagian besar mahasiswa memiliki capaian akademik yang baik. Sebanyak 60 mahasiswa (48,78%) memiliki $IPK \geq 3,50$ yang termasuk dalam kategori prestasi akademik sangat baik. Selanjutnya terdapat 55 mahasiswa (44,72%) yang memiliki IPK pada rentang 3,00–3,49, yang juga menunjukkan capaian akademik yang baik. Sementara itu hanya 7 mahasiswa (5,69%) yang memiliki IPK pada rentang 2,50–2,99, yang menunjukkan bahwa jumlah mahasiswa dengan capaian akademik relatif rendah sangat sedikit. Secara keseluruhan, data menunjukkan bahwa mayoritas mahasiswa memiliki performa akademik yang baik hingga sangat baik.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

Studi ini dilakukan di Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang. Tujuan dari penelitian ini adalah untuk menganalisis dan menerapkan teknik data mining yang menggunakan algoritma Naive Bayes untuk memprediksi hasil belajar mahasiswa. Data yang digunakan dalam penelitian ini berasal dari dua sumber utama: Sistem Informasi Akademik (SIKAD) dan data primer yang diberikan kepada siswa melalui kuesioner. Data akademik yang digunakan meliputi nilai mata kuliah, tingkat kehadiran, Indeks Prestasi Kumulatif (IPK), dan jumlah SKS yang diambil siswa. Dataset yang digunakan dalam penelitian ini terdiri dari sekitar 200 data mahasiswa dari berbagai angkatan Program Studi Teknologi Informasi. Dalam proses klasifikasi, setiap data siswa diwakili dalam sejumlah karakteristik. Nilai akademik, IPK, tingkat kehadiran, jumlah SKS, dan label hasil belajar mahasiswa adalah semua atribut yang termasuk dalam kategori ini. Diharapkan model klasifikasi yang mampu memprediksi kategori hasil belajar siswa secara akurat akan dihasilkan melalui proses analisis melalui beberapa tahapan utama. Tahapan-tahap ini termasuk preprocessing data, pembagian data untuk pelatihan dan pengujian, dan evaluasi model menggunakan matriks keakuratan.

Dataset penelitian ini diperoleh dari hasil kuesioner mahasiswa Program Studi Teknologi Informasi dengan jumlah sebanyak 123 responden. Distribusi data menunjukkan bahwa mayoritas mahasiswa memiliki tingkat kehadiran yang tinggi serta

capaian akademik yang baik. Pada aspek kehadiran, sebanyak 87 responden (70,73%) menyatakan selalu hadir dalam perkuliahan, 26 responden (21,14%) sering hadir, 6 responden (4,88%) kadang hadir, dan 4 responden (3,25%) jarang hadir.

Sementara itu, pada aspek Indeks Prestasi Kumulatif (IPK), dari 122 responden yang mengisi data IPK, sebanyak 60 mahasiswa (49,18%) berada pada kategori $\geq 3,50$, 55 mahasiswa (45,08%) berada pada kategori 3,00–3,49, dan 7 mahasiswa (5,74%) berada pada kategori 2,50–2,99. Dengan demikian, distribusi data dalam dataset menunjukkan bahwa sebagian besar responden memiliki karakteristik akademik yang baik dan layak digunakan sebagai dasar dalam proses analisis serta pemodelan prediksi hasil belajar mahasiswa.

Tabel 4.1 Ringkas distribusi data

Variabel	Kategori	Jumlah	Persentase
Kehadiran	Selalu hadir	87	70,73%
	Sering hadir	26	21,14%
	Kadang hadir	6	4,88%
	Jarang hadir	4	3,25%
IPK	$\geq 3,50$	60	49,18%
	3,00–3,49	55	45,08%
	2,50–2,99	7	5,74%

4.2 Implementasi Naïve Bayes

Langkah-langkah implementasi algoritma Naïve Bayes dalam penelitian ini dijelaskan sebagai berikut:

1. Data Preparation

Data yang telah melalui tahap *preprocessing* dibagi menjadi dua bagian:

- o 80% data digunakan untuk *training*
- o 20% data digunakan untuk *testing*

2. Training Model Naïve Bayes

Sistem menghitung probabilitas prior ($P(\text{Class})$) dan probabilitas bersyarat ($P(X|\text{Class})$) untuk setiap variabel.

4.3 Preprocessing Data

Preprocessing data merupakan tahap penting dalam proses data mining karena memastikan bahwa data yang digunakan sudah bersih, terstruktur, dan siap diolah oleh algoritma. Pada penelitian ini, preprocessing dilakukan melalui beberapa langkah, yaitu *data cleaning*, *data transformation*, *data reduction*, dan *data splitting*. Berikut uraian lengkapnya :

4.3.1 Data Cleaning (Pembersihan Data)

Pada tahap ini, data diperiksa untuk memastikan tidak adanya nilai yang hilang, data ganda, maupun nilai yang tidak sesuai. Beberapa proses yang dilakukan:

1. Menghapus data duplikat

Data mahasiswa yang muncul lebih dari satu kali (duplikat) dihapus agar tidak memengaruhi akurasi model.

2. Penanganan Missing Values

Jika terdapat nilai kosong pada kolom Tugas, UTS, UAS,

atau Kehadiran, maka data tersebut dihapus, Pertimbangan ini diambil karena persentase missing values kecil sehingga tidak memengaruhi jumlah sampel secara signifikan.

3. Validasi format dan range nilai

Nilai Tugas, UTS, dan UAS harus berada pada rentang 0–100, dan Kehadiran harus berada pada rentang 0–100%. Data yang tidak sesuai dilakukan koreksi atau dihapus sesuai kebijakan preprocessing.

4.3.2 Data Transformation

Transformasi data merupakan proses mengubah format data asli ke dalam bentuk yang lebih sesuai untuk kebutuhan pemodelan dalam data mining. Pada penelitian ini, transformasi dilakukan pada variabel nilai akademik dan kehadiran mahasiswa agar mudah diolah oleh algoritma Naïve Bayes yang bekerja dengan format data terklasifikasi atau terkodekan.

A. Transformasi Nilai Akademik

Data nilai Tugas, UTS, dan UAS pada awalnya berupa angka dalam rentang 0–100. Untuk memudahkan proses klasifikasi, nilai angka dikonversi menjadi kategori huruf berdasarkan kebijakan penilaian di lingkungan Fakultas Teknik Universitas Muhammadiyah Palembang, seperti berikut:

Tabel 4.2 Nilai Akademik

Rentang iiiNilai	Kategori iiiHuruf	Interpretasi
80 iii– iii100	A	Sangat iiiBaik
70 – 79	B	Baik
60 iii– iii69	C	Cukup

50 – 59	D	Kurang
< 50	E	Sangat Kurang

Transformasi ini dilakukan agar pola nilai mahasiswa lebih mudah dianalisis berdasarkan perbedaan tingkat pencapaian akademik.

B. Transformasi Kehadiran

Transformasi data kehadiran dilakukan untuk mengubah data persentase kehadiran mahasiswa yang semula berbentuk data numerik (0–100%) menjadi data kategorikal. Proses transformasi ini bertujuan untuk mempermudah proses analisis klasifikasi menggunakan algoritma Naïve Bayes, karena metode tersebut bekerja lebih optimal pada data yang telah dikelompokkan ke dalam kategori tertentu. Pada dataset awal, tingkat kehadiran mahasiswa dicatat dalam bentuk persentase kehadiran selama mengikuti perkuliahan dalam satu semester. Nilai persentase tersebut kemudian dikelompokkan menjadi beberapa kategori berdasarkan tingkat keaktifan mahasiswa dalam mengikuti kegiatan perkuliahan. Proses transformasi dilakukan dengan mengelompokkan nilai kehadiran ke dalam kategori sebagai berikut:

Tabel 4.3 Transformasi Kehadiran Mahasiswa

Persentase Kehadiran	Kategori	Interpretasi
$\geq 90\%$	Sangat Tinggi	Sangat aktif mengikuti perkuliahan
75% – 89%	Tinggi	Aktif mengikuti perkuliahan
60% – 74%	Sedang	Cukup aktif mengikuti perkuliahan

< 60%	Rendah	Kurang aktif mengikuti perkuliahan
-------	--------	------------------------------------

Melalui proses transformasi ini, data kehadiran mahasiswa yang sebelumnya berupa angka persentase dapat direpresentasikan dalam bentuk kategori yang lebih mudah dianalisis dalam proses klasifikasi. Selain itu, pengelompokan ini juga membantu dalam mengidentifikasi pola hubungan antara tingkat kehadiran mahasiswa dengan hasil belajar yang diperoleh. Transformasi kehadiran ini kemudian digunakan sebagai salah satu variabel input dalam model klasifikasi Naïve Bayes, bersama dengan variabel lain seperti IPK mahasiswa, nilai mata kuliah inti, serta indikator pembelajaran yang diperoleh dari data kuesioner. Dengan demikian, proses transformasi data kehadiran merupakan tahapan penting dalam preprocessing data karena dapat meningkatkan kualitas dataset serta membantu algoritma dalam menemukan pola yang lebih jelas antara faktor kehadiran dan hasil belajar mahasiswa.

C. Encoding Variabel Kategorikal

Encoding variabel kategorikal merupakan proses mengubah data yang berbentuk kategori atau label deskriptif menjadi nilai numerik agar dapat diproses oleh algoritma pembelajaran mesin. Pada penelitian ini, sebagian besar data yang diperoleh dari kuesioner mahasiswa memiliki bentuk kategori, seperti tingkat motivasi belajar, tingkat keaktifan dalam pembelajaran online, serta persepsi terhadap proses pembelajaran.

Algoritma Naïve Bayes dalam implementasi komputasional memerlukan data dalam bentuk numerik untuk melakukan perhitungan probabilitas. Oleh karena itu, setiap kategori pada

variabel kuesioner perlu dikonversi menjadi nilai angka melalui proses encoding. Dalam penelitian ini, metode encoding yang digunakan adalah label encoding, yaitu memberikan kode angka pada setiap kategori berdasarkan tingkat atau urutan tertentu. Contoh penerapan encoding pada skala Likert yang digunakan dalam kuesioner adalah sebagai berikut:

Tabel 4.4 Encoding Skala Likert

Kategori	Nilai
Sangat Tidak Setuju	1
Tidak Setuju	2
	3
Setuju	4
Sangat Setuju	5

Selain itu, beberapa variabel lain seperti tingkat keaktifan dan frekuensi aktivitas belajar juga dilakukan proses encoding sebagai berikut:

Tabel 4.5 Encoding Tingkat Aktivitas

Kategori	Nilai
Rendah	1
Sedang	2
Tinggi	3
Sangat Tinggi	4

Proses encoding ini bertujuan untuk menyeragamkan format data sehingga seluruh variabel yang digunakan dalam penelitian dapat diproses oleh algoritma Naïve Bayes secara lebih efektif. Setelah proses encoding selesai, dataset yang dihasilkan akan

terdiri dari nilai numerik yang mewakili setiap kategori pada variabel kuesioner. Dataset hasil encoding kemudian digunakan sebagai data input dalam proses pemodelan Naïve Bayes, sehingga memungkinkan sistem untuk menghitung probabilitas kemunculan setiap atribut terhadap kelas hasil belajar mahasiswa. Dengan demikian, proses encoding variabel kategorikal menjadi tahap penting dalam preprocessing data karena memastikan bahwa seluruh data telah berada dalam format yang sesuai untuk proses analisis dan klasifikasi menggunakan algoritma data mining.

4.4 Preprocessing Data

Tahap penting dalam proses pengolahan data adalah preprocessing data, yang dilakukan untuk memastikan bahwa data yang digunakan bersih, konsisten, dan dalam format yang sesuai untuk proses analisis. Tujuan dari tahap ini adalah untuk menyiapkan data untuk digunakan dalam proses pemodelan menggunakan algoritma Naive Bayes. Penelitian ini menggunakan data dari kuesioner yang dikumpulkan oleh mahasiswa Program Studi Teknologi Informasi, yang disimpan dalam format Microsoft Excel Tingkat kehadiran, nilai mata kuliah inti, motivasi untuk belajar, keaktifan dalam pembelajaran online, dan persepsi siswa tentang penggunaan teknologi dalam pembelajaran adalah beberapa atribut yang terkait dengan pembelajaran siswa.

Hasil pengolahan awal file Excel menunjukkan bahwa ada 123 data dari mahasiswa yang menjawab pertanyaan. Namun, satu data IPK hilang saat data diperiksa, jadi total data yang digunakan untuk analisis adalah 122 data valid. Dalam penelitian ini, tahapan preprocessing data terdiri dari berbagai langkah. Ini termasuk import data, memilih atribut, membersihkan

data, transformasi data, encoding data, dan pembentukan dataset akhir.

4.4.1 Import Data

Proses import data dari file Excel ke dalam lingkungan pengolahan data adalah langkah pertama dalam preprocessing data. Setelah itu, file kuesioner yang telah dikumpulkan diimpor ke dalam perangkat lunak pengolahan data seperti Microsoft Excel dan Python (Google Colab). Setelah import data selesai, langkah selanjutnya adalah seleksi atribut untuk mengidentifikasi variabel yang relevan dengan tujuan penelitian. Kolom atribut yang diimpor termasuk nama responden, NIM, jenis kelamin, universitas, tingkat kehadiran, kisaran IPK, nilai mata kuliah inti, keaktifan pembelajaran online, motivasi untuk belajar, dan persepsi siswa terhadap data mining.

4.4.2 Seleksi Atribut

Pada tahap ini, atribut yang akan digunakan dalam proses analisis dipilih. Pemodelan tidak menggunakan semua atribut dalam dataset karena beberapa hanya berfungsi sebagai identitas responden dan tidak mempengaruhi proses klasifikasi. Karena tidak terkait langsung dengan prestasi akademik siswa, atribut yang termasuk dalam kategori identitas seperti Nama, NIM, dan Universitas tidak digunakan dalam proses klasifikasi. Tingkat kehadiran mahasiswa, kisaran IPK, nilai mata kuliah inti, keaktifan pembelajaran online, dan motivasi belajar semuanya merupakan variabel penelitian.

4.4.3 Data cleaning

Data Cleaning dilakukan untuk memastikan bahwa data yang digunakan dalam penelitian bebas dari kesalahan atau

ketidakkonsistenan data. Proses pembersihan data yang dilakukan meliputi , Pengecekan Missing Value Pada tahap ini dilakukan pemeriksaan terhadap nilai kosong pada dataset. Hasil pemeriksaan menunjukkan bahwa terdapat 1 data responden yang tidak memiliki nilai IPK, sehingga data tersebut tidak digunakan dalam proses analisis, Pengecekan Data Duplikat.

Dilakukan pemeriksaan terhadap kemungkinan adanya data responden yang tercatat lebih dari satu kali berdasarkan NIM mahasiswa, Standarisasi Format Data Beberapa jawaban kuesioner memiliki variasi penulisan seperti huruf besar dan kecil yang berbeda. Oleh karena itu dilakukan standarisasi agar seluruh kategori memiliki format yang seragam. Setelah proses data cleaning selesai, jumlah dataset yang digunakan dalam penelitian adalah 122 data mahasiswa.

4.4.4 Transformasi Data

Transformasi data dilakukan untuk mengubah format data agar sesuai dengan kebutuhan analisis data mining. Beberapa variabel pada dataset awal memiliki bentuk teks atau kategori yang berbeda-beda sehingga perlu dikelompokkan ke dalam kategori tertentu.

Tabel 4.6 Transformasi Data

Persentase Kehadiran	Kategori
$\geq 90\%$	Sangat Tinggi
75% – 89%	Tinggi
60% – 74%	Sedang

< 60%	Rendah
-------	--------

Tabel 4.7 Transformasi IPK

Rentang iiiIPK	Kategori
$\geq 3,50$	Sangat iiiBaik
3,00 – 3,49	Baik
2,50 iii– iii2,99	Cukup

Transformasi ini bertujuan untuk mempermudah proses klasifikasi dan membantu algoritma Naïve Bayes dalam mengenali pola hubungan antar variabel.

4.4.5 Encoding Data

Setelah proses transformasi dilakukan, langkah berikutnya adalah melakukan encoding data untuk mengubah data kategorikal menjadi bentuk numerik.

Tabel 4.8 Kategori Motivasi

Kategori iiiMotivasi	Nilai
Rendah	1
Sedang	2
Tinggi	3
Sangat Tinggi	4

Encoding ini dilakukan agar dataset dapat diproses oleh algoritma Naïve Bayes dalam proses perhitungan probabilitas.

4.4.6 Pembentukan Dataset Akhir

Setelah seluruh tahapan preprocessing selesai dilakukan, diperoleh dataset akhir yang terdiri dari, 122 data mahasiswa, beberapa variabel input dari kuesioner, Dan 1 variabel target (hasil belajar).

Dataset ini kemudian digunakan sebagai data input dalam proses pemodelan Naïve Bayes untuk memprediksi hasil belajar mahasiswa berdasarkan faktor-faktor yang mempengaruhi proses pembelajaran.

4.5 Evaluasi Kinerja Model Berdasarkan Hasil Kuesioner

Dalam penelitian ini, evaluasi kinerja model didasarkan pada hasil klasifikasi algoritma Naive Bayes. Selain itu, data kuesioner yang dikumpulkan dari siswa Program Studi Teknologi Informasi digunakan untuk memberikan gambaran tentang kondisi akademik siswa dan faktor-faktor yang dapat mempengaruhi hasil belajar mereka.

Berdasarkan hasil pengolahan data dari file Excel kuesioner, ditemukan bahwa ada 123 mahasiswa yang berpartisipasi dalam penelitian ini. Namun, ada satu data IPK yang tidak terisi setelah proses perbaikan data, sehingga total data yang digunakan dalam analisis adalah 122 data yang valid.

Tingkat kehadiran siswa, distribusi IPK responden, tingkat motivasi belajar, dan persepsi mereka tentang penggunaan data akademik untuk memprediksi prestasi belajar adalah beberapa indikator utama yang berkaitan dengan hasil belajar siswa.

4.5.1 Tingkat Kehadiran Mahasiswa

Tingkat kehadiran mahasiswa merupakan salah satu faktor penting yang dapat mempengaruhi hasil belajar mahasiswa.

Kehadiran yang tinggi menunjukkan tingkat kedisiplinan mahasiswa dalam mengikuti kegiatan perkuliahan. Distribusi tingkat kehadiran mahasiswa berdasarkan hasil kuesioner adalah sebagai berikut:

Tabel 4.9 Tingkat Kehadiran Mahasiswa

Kategori Kehadiran	Jumlah	Persentase
Selalu hadir	87	70,73%
Sering hadir ($\geq 75\%$)	26	21,14%
Kadang hadir (50–74%)	6	4,88%
Jarang hadir ($< 50\%$)	4	3,25%
Total	123	100%

Berdasarkan tabel tersebut dapat diketahui bahwa sebagian besar mahasiswa memiliki tingkat kehadiran yang sangat baik. Sebanyak 87 mahasiswa (70,73%) menyatakan selalu hadir dalam perkuliahan, sedangkan 26 mahasiswa (21,14%) berada pada kategori sering hadir. Sementara itu, jumlah mahasiswa yang memiliki tingkat kehadiran sedang dan rendah relatif sedikit.

Hal ini menunjukkan bahwa mayoritas mahasiswa memiliki tingkat kedisiplinan yang tinggi dalam mengikuti kegiatan perkuliahan. Tingkat kehadiran yang tinggi ini dapat menjadi indikator positif yang berkontribusi terhadap capaian hasil belajar mahasiswa.

4.5.2 Distribusi IPK Responden

Indeks Prestasi Kumulatif (IPK) merupakan indikator utama yang digunakan untuk menilai capaian akademik mahasiswa.

Distribusi IPK responden berdasarkan hasil kuesioner adalah sebagai berikut:

Tabel 4.10 Distribusi IPK Mahasiswa

Rentang ⁱⁱⁱ IPK	Jumlah	Persentase
$\geq$ ⁱⁱⁱ 3,50	60	49,18%
3,00 – 3,49	55	45,08%
2,50 ⁱⁱⁱ – ⁱⁱⁱ 2,99	7	5,74%
Total valid	122	100%

Berdasarkan tabel tersebut dapat diketahui bahwa mayoritas mahasiswa memiliki capaian akademik yang baik. Sebanyak 60 mahasiswa (49,18%) memiliki IPK $\geq$ 3,50 yang termasuk kategori sangat baik, sedangkan 55 mahasiswa (45,08%) berada pada rentang IPK 3,00–3,49. Jumlah mahasiswa dengan IPK pada rentang 2,50–2,99 hanya sebanyak 7 mahasiswa (5,74%). Hal ini menunjukkan bahwa sebagian besar responden memiliki performa akademik yang baik hingga sangat baik.

4.5.3 Tingkat Motivasi Belajar Mahasiswa

Motivasi belajar merupakan faktor internal yang dapat mempengaruhi keberhasilan mahasiswa dalam mencapai prestasi akademik. Distribusi tingkat motivasi belajar mahasiswa berdasarkan hasil kuesioner adalah sebagai berikut:

Tabel 4.11 Tingkat Motivasi Belajar Mahasiswa

Tingkat ⁱⁱⁱ Motivasi	Jumlah	Persentase
---------------------------------	--------	------------

Sangat iiiTinggi	64	52,03%
Tinggi	43	34,96%
Sedang	14	11,38%
Rendah	2	1,63%
Total	123	100%

Berdasarkan tabel tersebut, mayoritas mahasiswa memiliki motivasi belajar yang tinggi. Sebanyak 64 mahasiswa (52,03%) berada pada kategori motivasi belajar sangat tinggi dan 43 mahasiswa (34,96%) berada pada kategori tinggi.

Hal ini menunjukkan bahwa sebagian besar mahasiswa memiliki semangat belajar yang baik dalam mengikuti proses perkuliahan. Motivasi belajar yang tinggi dapat menjadi salah satu faktor yang mendukung capaian prestasi akademik mahasiswa.

4.5.4 Persepsi Mahasiswa terhadap Penggunaan Data Akademik untuk Prediksi

Selain faktor akademik, penelitian ini juga mengukur persepsi mahasiswa terhadap penggunaan data akademik dalam memprediksi hasil belajar mahasiswa.

Tabel 4.12 Persepsi Mahasiswa terhadap Penggunaan Data Akademik

Persepsi iiiMahasiswa	Jumlah	Persentase
Sangat iiiSetuju	44	35,77%

Setuju	64	52,03%
Ragu-ragu	15	12,20%
Total	123	100%

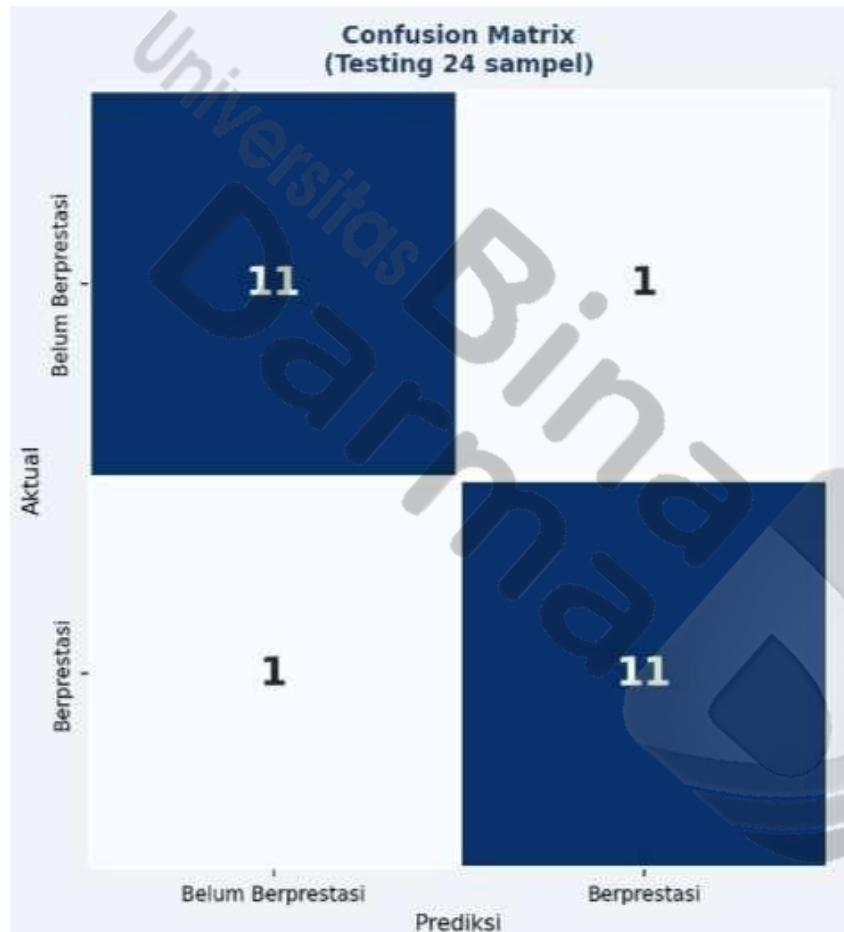
Berdasarkan tabel tersebut dapat diketahui bahwa sebagian besar mahasiswa memiliki persepsi positif terhadap penggunaan data akademik untuk memprediksi prestasi belajar. Sebanyak 64 mahasiswa (52,03%) menyatakan setuju dan 44 mahasiswa (35,77%) menyatakan sangat setuju.

Hal ini menunjukkan bahwa mayoritas mahasiswa mendukung penggunaan teknologi analisis data seperti data mining dalam membantu pihak program studi untuk melakukan evaluasi dan pemantauan prestasi akademik mahasiswa.

4.5.5 Keterkaitan Hasil Kuesioner dengan Kinerja Model

Berdasarkan hasil analisis kuesioner, dapat diketahui bahwa sebagian besar mahasiswa memiliki tingkat kehadiran yang tinggi, motivasi belajar yang baik, serta capaian IPK yang relatif tinggi. Kondisi ini menunjukkan adanya pola hubungan antara faktor-faktor tersebut dengan hasil belajar mahasiswa. Variabel-variabel yang diperoleh dari kuesioner seperti kehadiran, motivasi belajar, dan aktivitas pembelajaran kemudian digunakan sebagai variabel input dalam model klasifikasi Naïve Bayes. Dengan menggunakan variabel-variabel tersebut, model dapat mempelajari pola hubungan antara faktor pembelajaran dengan kategori hasil belajar mahasiswa. Dengan demikian, hasil kuesioner tidak hanya digunakan sebagai data pendukung penelitian, tetapi juga membantu dalam memperkuat interpretasi terhadap hasil klasifikasi model Naïve Bayes yang dibangun pada penelitian ini.

4.6 Confusion Matrix



Gambar 4.1 Confusion Matrix

Berdasarkan hasil pengujian model menggunakan algoritma Naïve Bayes, diperoleh Confusion Matrix dengan jumlah 24 data testing. Confusion Matrix digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan antara nilai aktual dan hasil prediksi yang dihasilkan oleh model.

Dari hasil Confusion Matrix yang ditampilkan pada gambar, diperoleh hasil sebagai berikut:

- Sebanyak 11 data mahasiswa yang sebenarnya termasuk kategori Belum Berprestasi berhasil diprediksi dengan benar sebagai Belum Berprestasi oleh model.
- Terdapat 1 data mahasiswa yang sebenarnya Belum Berprestasi namun diprediksi sebagai Berprestasi oleh model.
- Sebanyak 11 data mahasiswa yang sebenarnya termasuk kategori Berprestasi berhasil diprediksi dengan benar sebagai Berprestasi.
- Terdapat 1 data mahasiswa yang sebenarnya Berprestasi namun diprediksi sebagai Belum Berprestasi.

Dengan demikian, dari total 24 data pengujian, model berhasil melakukan klasifikasi dengan benar terhadap 22 data, sedangkan 2 data mengalami kesalahan klasifikasi.

Jika dihitung tingkat akurasi, maka:

$$Accuracy = \frac{TP + TN}{Total \ Data}$$

$$Accuracy = \frac{11 + 11}{24} = \frac{22}{24} = 0.9167$$

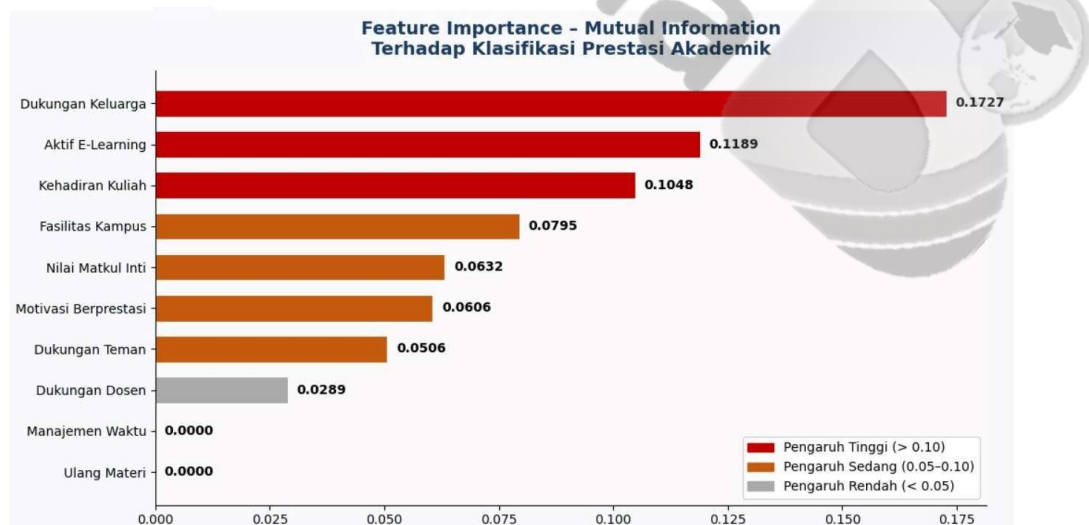
Sehingga diperoleh tingkat akurasi sebesar 91,67%.

Hasil ini menunjukkan bahwa model klasifikasi Naïve Bayes memiliki performa yang cukup baik dalam memprediksi kategori prestasi mahasiswa. Nilai akurasi yang tinggi menunjukkan bahwa sebagian besar data dapat diklasifikasikan dengan benar oleh model berdasarkan atribut yang digunakan dalam penelitian, seperti tingkat kehadiran, nilai akademik, motivasi belajar, dan aktivitas pembelajaran.

Dengan demikian, metode Naïve Bayes dapat digunakan sebagai pendekatan yang efektif dalam memprediksi potensi prestasi akademik mahasiswa serta membantu pihak program studi dalam melakukan evaluasi dan pemantauan terhadap perkembangan akademik mahasiswa.

4.7 Hasil Feature Importanc

Berdasarkan hasil analisis Feature Importance menggunakan metode Mutual Information, diperoleh gambaran mengenai tingkat pengaruh masing-masing variabel terhadap klasifikasi prestasi akademik mahasiswa. Mutual Information digunakan untuk mengukur seberapa besar hubungan antara suatu variabel input



Gambar 4.2 Feature Importance

dengan variabel target dalam model klasifikasi.

Semua variabel memiliki nilai kontribusi yang berbeda terhadap proses prediksi prestasi akademik, seperti yang ditunjukkan oleh hasil visualisasi pada gambar. Variabel dengan nilai informasi mutual tertinggi berdampak lebih besar pada

hasil klasifikasi daripada variabel dengan nilai informasi mutual yang lebih rendah.

Dukungan keluarga, dengan nilai 0,1727, adalah variabel dengan pengaruh paling besar terhadap klasifikasi prestasi akademik, menunjukkan bahwa ada hubungan yang paling kuat antara dukungan keluarga dan keberhasilan akademik siswa. Dukungan keluarga dapat memberikan motivasi, dukungan emosional, dan lingkungan belajar yang membantu siswa berkembang.

Mahasiswa yang aktif mengikuti pembelajaran online dan memiliki kehadiran akademik yang tinggi cenderung mencapai prestasi akademik yang lebih baik, seperti yang ditunjukkan oleh nilai 0,1189 keaktifan e-learning dan 0,01048 kehadiran kuliah.

Selain itu, ada beberapa variabel dengan pengaruh sedang: fasilitas kampus (0,0795), nilai mata kuliah inti (0,0632), motivasi berprestasi (0,0606), dan dukungan teman (0,0506). Semua variabel ini masih berkontribusi terhadap prestasi akademik siswa, tetapi tidak sebesar variabel utama.

Di sisi lain, dukungan dosen adalah variabel yang memiliki pengaruh paling sedikit (0,0289). Dua variabel lainnya, kebiasaan mengulang materi dan manajemen waktu, memiliki nilai 0,0000. Nilai tersebut menunjukkan bahwa kedua variabel tersebut tidak memberikan kontribusi yang signifikan terhadap proses klasifikasi prestasi akademik mahasiswa pada dataset yang digunakan dalam penelitian ini.

Secara keseluruhan, hasil analisis pentingnya fitur menunjukkan bahwa faktor yang paling berpengaruh terhadap prestasi akademik siswa adalah dukungan keluarga, aktivitas

pembelajaran online, dan kehadiran kuliah. Temuan ini menunjukkan bahwa keberhasilan akademik siswa tidak hanya dipengaruhi oleh faktor akademik, tetapi juga oleh faktor lingkungan dan keterlibatan siswa dalam proses pembelajaran.

Oleh karena itu, hasil analisis ini dapat menunjukkan kepada pihak program studi bahwa peningkatan keterlibatan siswa dalam pembelajaran dan dukungan dari keluarga dan lingkungan akademik dapat membantu peningkatan prestasi akademik mereka

4.8 Performa Model

Berdasarkan hasil pengujian model menggunakan algoritma Naïve Bayes, diperoleh nilai performa model yang ditunjukkan melalui beberapa metrik evaluasi, yaitu akurasi (accuracy), presisi (precision), recall, dan F1-score. Keempat metrik tersebut

digunakan untuk menilai kemampuan model dalam melakukan klasifikasi terhadap data mahasiswa.



Gambar 4.3 Performa Model

Semua metrik evaluasi memiliki nilai yang sama, sebesar 91,7%, seperti yang ditunjukkan pada gambar. Nilai akurasi sebesar 91,7% ini berasal dari hasil pengujian model terhadap data uji yang digunakan dalam penelitian dan menunjukkan bahwa model memiliki kemampuan untuk mengklasifikasikan data dengan tingkat ketepatan yang tinggi.

Selain itu, nilai precision sebesar 91,7% menunjukkan bahwa sebagian besar data yang diprediksi sebagai kategori berprestasi oleh model benar-benar termasuk dalam kategori tersebut. Ini menunjukkan bahwa tingkat kesalahan prediksi model relatif kecil.

Nilai recall sebesar 91,7% menunjukkan kemampuan model yang baik untuk menemukan data yang benar-benar termasuk dalam kategori berprestasi. Dengan kata lain, sebagian besar data berprestasi yang sebenarnya diidentifikasi dengan sukses oleh model.

Selain itu, nilai F1-score sebesar 91,7% menunjukkan keseimbangan yang baik antara nilai precision dan recall, dan nilai F1-score yang tinggi menunjukkan bahwa model memiliki performa klasifikasi yang stabil dan konsisten.

Grafik juga menampilkan garis putus-putus yang menunjukkan target performa sebesar 86%. Hasil pengujian menunjukkan bahwa seluruh metrik evaluasi model memenuhi target, yaitu 91,7%. Dengan demikian, dapat disimpulkan bahwa model yang dibangun telah memenuhi bahkan target performa yang diharapkan.

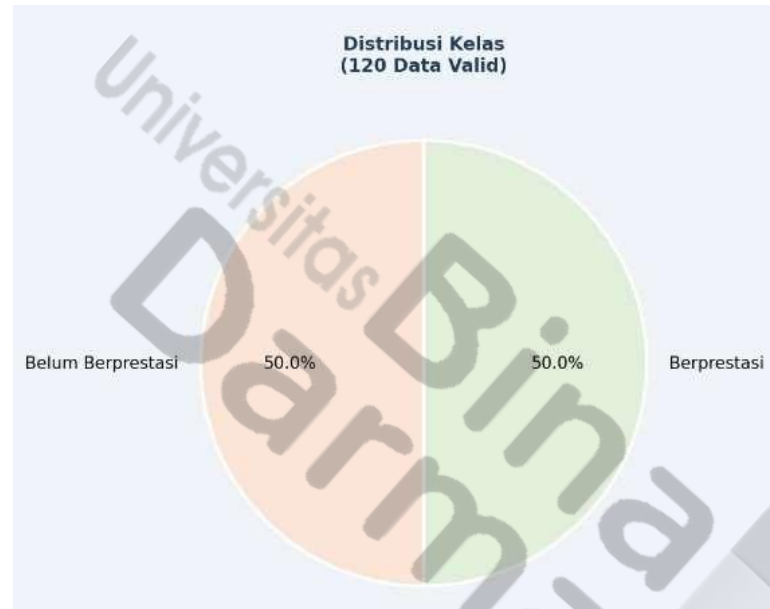
Oleh karena itu, model Naive Bayes yang digunakan dalam penelitian ini, berdasarkan variabel penelitian seperti kehadiran, nilai akademik, dan motivasi belajar, serta faktor pendukung lainnya dari hasil kuesioner, dapat dikatakan sangat baik dalam memprediksi prestasi akademik mahasiswa.

Hasil ini menunjukkan bahwa metode Naive Bayes dapat digunakan sebagai cara yang efektif untuk menganalisis data akademik. Ini juga dapat membantu pihak program studi memantau dan menilai prestasi belajar mahasiswa.

4.9 Distribusi Kelas

Berdasarkan hasil pengolahan data yang telah dilakukan, diperoleh distribusi kelas dari dataset penelitian yang terdiri dari 120 data valid. Distribusi kelas ini menunjukkan proporsi jumlah

data mahasiswa yang termasuk dalam kategori Berprestasi dan Belum Berprestasi.



Gambar 4.4 Distribusi Kelas

Pada gambar terlihat bahwa distribusi kelas memiliki komposisi yang seimbang antara kedua kategori tersebut. Sebanyak 50% data mahasiswa termasuk dalam kategori Berprestasi, sedangkan 50% lainnya termasuk dalam kategori Belum Berprestasi. Hal ini menunjukkan bahwa jumlah data pada kedua kelas memiliki proporsi yang sama, sehingga dataset yang digunakan dalam penelitian ini dapat dikatakan seimbang (balanced dataset).

Keseimbangan distribusi kelas dalam dataset merupakan kondisi yang sangat baik dalam proses pembelajaran model klasifikasi. Dataset yang seimbang dapat membantu model dalam mempelajari pola data secara lebih optimal tanpa adanya kecenderungan atau bias terhadap salah satu kelas tertentu.

Dengan distribusi data yang seimbang antara kelas Berprestasi dan Belum Berprestasi, algoritma Naïve Bayes yang digunakan dalam penelitian ini dapat melakukan proses klasifikasi dengan lebih efektif. Hal ini juga berkontribusi terhadap hasil evaluasi model yang menunjukkan tingkat performa yang tinggi, seperti yang terlihat pada nilai akurasi, precision, recall, dan F1-score yang mencapai sekitar 91,7% pada tahap pengujian model.

Oleh karena itu, distribusi kelas yang seimbang dalam dataset ini menjadi salah satu faktor yang mendukung keberhasilan model dalam melakukan prediksi prestasi akademik mahasiswa secara lebih akurat.

4.10 5-Fold Cross-Validation

Untuk mengevaluasi kestabilan dan kemampuan generalisasi model klasifikasi, penelitian ini menggunakan metode 5-Fold Cross-Validation. Metode ini dilakukan dengan membagi dataset menjadi lima bagian (fold), dimana setiap fold secara bergantian digunakan sebagai data testing, sementara fold lainnya digunakan sebagai data training.



Gambar 4.5 5-Fold Cross-Validation

Berdasarkan hasil pengujian menggunakan metode 5-Fold Cross-Validation, diperoleh nilai akurasi yang berbeda pada setiap fold. Hasil akurasi pada masing-masing fold adalah sebagai berikut:

- Fold 1 menghasilkan akurasi sebesar 62,5%
- Fold 2 menghasilkan akurasi sebesar 75,0%
- Fold 3 menghasilkan akurasi sebesar 83,3%
- Fold 4 menghasilkan akurasi sebesar 62,5%
- Fold 5 menghasilkan akurasi sebesar 62,5%

Dari hasil tersebut, dapat dihitung nilai rata-rata akurasi (mean accuracy) sebesar 69,17%. Nilai ini ditunjukkan pada grafik dengan garis putus-putus berwarna biru. Selain itu, terdapat garis putus-putus berwarna merah yang menunjukkan target akurasi sebesar 86%.

Berdasarkan hasil tersebut, terlihat bahwa nilai akurasi pada setiap fold masih berada di bawah target performa yang telah ditetapkan. Nilai akurasi tertinggi diperoleh pada Fold 3 sebesar 83,3%, sedangkan nilai akurasi terendah terdapat pada Fold 1, Fold 4, dan Fold 5, masing-masing sebesar 62,5%.

Perbedaan nilai akurasi pada setiap fold menunjukkan bahwa performa model dapat dipengaruhi oleh variasi data yang digunakan dalam proses pelatihan dan pengujian. Hal ini merupakan kondisi yang umum terjadi dalam proses evaluasi model machine learning, terutama ketika ukuran dataset relatif terbatas.

Meskipun nilai rata-rata akurasi dari cross-validation berada di bawah target, hasil pengujian model pada data testing menunjukkan performa yang lebih baik dengan nilai akurasi sebesar 91,7%. Hal ini menunjukkan bahwa model masih memiliki kemampuan klasifikasi yang baik dalam memprediksi prestasi akademik mahasiswa berdasarkan variabel yang digunakan dalam penelitian. Dengan demikian, metode Naïve Bayes tetap dapat digunakan sebagai pendekatan yang efektif dalam analisis data akademik, meskipun hasil cross-validation menunjukkan adanya variasi performa pada setiap fold pengujian.

4.11 Hasil Naïve Bayes

Berdasarkan hasil analisis menggunakan metode Gaussian Naïve Bayes, penelitian ini bertujuan untuk mengklasifikasikan kepuasan pembelajaran mahasiswa serta mengidentifikasi faktor-faktor yang mempengaruhi prestasi akademik mahasiswa. Analisis dilakukan menggunakan dataset yang telah melalui proses preprocessing sehingga menghasilkan 120 data valid yang digunakan dalam proses pemodelan.



Gambar 4.6 Hasil *Naïve Bayes*

Dataset yang digunakan dalam penelitian ini terdiri dari 10 atribut atau fitur yang berkaitan dengan faktor-faktor yang mempengaruhi prestasi akademik mahasiswa. Beberapa atribut tersebut meliputi faktor akademik maupun non-akademik seperti dukungan keluarga, keaktifan dalam e-learning, kehadiran kuliah, fasilitas kampus, motivasi berprestasi, serta dukungan dari lingkungan sekitar.

Dalam proses pemodelan, data dibagi menggunakan metode train-test split dengan proporsi 80% data training dan 20% data testing. Pembagian ini bertujuan untuk melatih model menggunakan sebagian besar data, kemudian menguji kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya.

Hasil pengujian model menunjukkan bahwa algoritma Naïve Bayes menghasilkan akurasi sebesar 91,67% pada data testing. Nilai akurasi ini menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan yang sangat baik. Selain itu, nilai akurasi yang diperoleh juga melampaui target performa yang telah ditetapkan sebesar 86%, sehingga dapat disimpulkan bahwa model yang dibangun dalam penelitian ini berhasil mencapai target kinerja yang diharapkan.

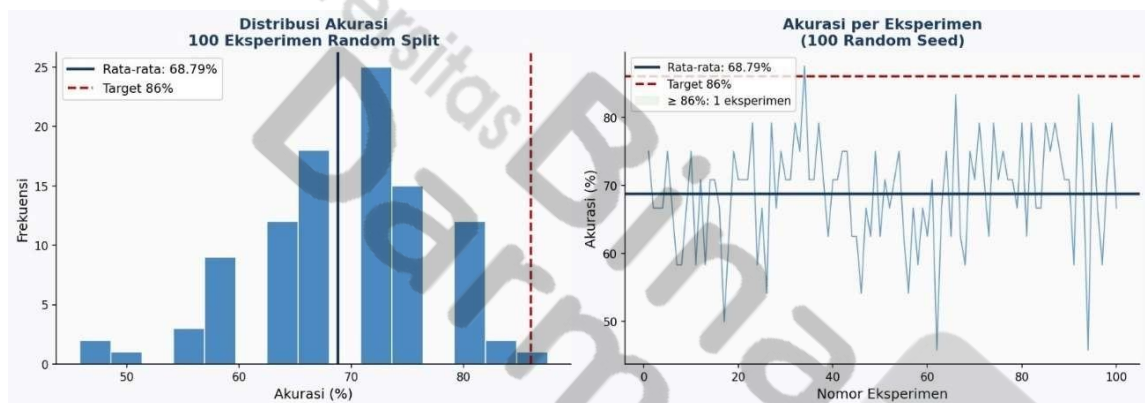
Berdasarkan hasil analisis fitur, diperoleh bahwa dukungan keluarga merupakan faktor yang paling berpengaruh terhadap prestasi akademik mahasiswa. Selain itu, keaktifan mahasiswa dalam mengikuti pembelajaran e-learning juga menjadi salah satu faktor penting yang berkontribusi terhadap keberhasilan akademik mahasiswa.

Secara keseluruhan, hasil analisis ini menunjukkan bahwa metode Gaussian Naïve Bayes dapat digunakan secara efektif dalam memprediksi prestasi akademik mahasiswa berdasarkan data yang diperoleh dari kuesioner pembelajaran. Model yang dihasilkan tidak hanya mampu memberikan tingkat akurasi yang tinggi, tetapi juga mampu mengidentifikasi faktor-faktor penting yang mempengaruhi keberhasilan akademik mahasiswa.

Dengan demikian, penerapan metode Naïve Bayes dalam penelitian ini dapat memberikan kontribusi dalam membantu pihak program studi atau institusi pendidikan untuk memahami faktor-faktor yang mempengaruhi prestasi akademik mahasiswa serta mendukung proses pengambilan keputusan dalam meningkatkan kualitas pembelajaran.

4.12 Hasil Distribusi Akurasi dan Akurasi per Eksperimen

Berdasarkan hasil pengujian model menggunakan metode 100 eksperimen random split, diperoleh distribusi nilai akurasi yang bervariasi pada setiap percobaan. Pengujian ini dilakukan untuk mengevaluasi kestabilan model Naïve Bayes dalam melakukan klasifikasi terhadap data prestasi akademik mahasiswa.



Gambar 4.7 Hasil Distribusi akurasi dan akurasi eksperimen

Pada grafik sebelah kiri ditampilkan distribusi akurasi dari 100 eksperimen random split dalam bentuk histogram. Grafik ini menunjukkan sebaran nilai akurasi yang diperoleh dari setiap percobaan. Berdasarkan hasil analisis, diperoleh nilai rata-rata akurasi sebesar 68,79%, yang ditunjukkan oleh garis vertikal berwarna biru. Sementara itu, garis putus-putus berwarna merah menunjukkan target akurasi yang diharapkan sebesar 86%.

Dari hasil distribusi tersebut terlihat bahwa sebagian besar nilai akurasi berada pada rentang 60% hingga 75%, yang menunjukkan bahwa performa model cenderung berada pada tingkat akurasi sedang. Hanya terdapat satu eksperimen yang mampu mencapai atau melampaui target akurasi sebesar 86%, yang menunjukkan bahwa performa tinggi model hanya terjadi pada kondisi pembagian data tertentu.

Grafik sebelah kanan menunjukkan akurasi yang dihasilkan pada setiap eksperimen dari 100 percobaan random seed. Grafik ini memperlihatkan fluktuasi nilai akurasi yang cukup bervariasi antar eksperimen. Nilai akurasi terendah berada pada kisaran sekitar 45%, sedangkan nilai akurasi tertinggi mendekati 86%.

Garis horizontal berwarna biru pada grafik menunjukkan nilai rata-rata akurasi sebesar 68,79%, sedangkan garis merah putus-putus menunjukkan target akurasi sebesar 86%. Dari grafik tersebut terlihat bahwa sebagian besar eksperimen menghasilkan nilai akurasi di bawah target yang telah ditentukan.

Hasil ini menunjukkan bahwa performa model Naïve Bayes cukup dipengaruhi oleh variasi pembagian data training dan testing. Hal ini merupakan kondisi yang umum terjadi dalam proses pembelajaran model machine learning, terutama ketika ukuran dataset relatif terbatas.

Secara keseluruhan, hasil pengujian ini menunjukkan bahwa model memiliki tingkat stabilitas performa pada kisaran akurasi sekitar 68% hingga 70%, namun pada kondisi tertentu model mampu menghasilkan akurasi yang lebih tinggi hingga mendekati target performa yang diharapkan.

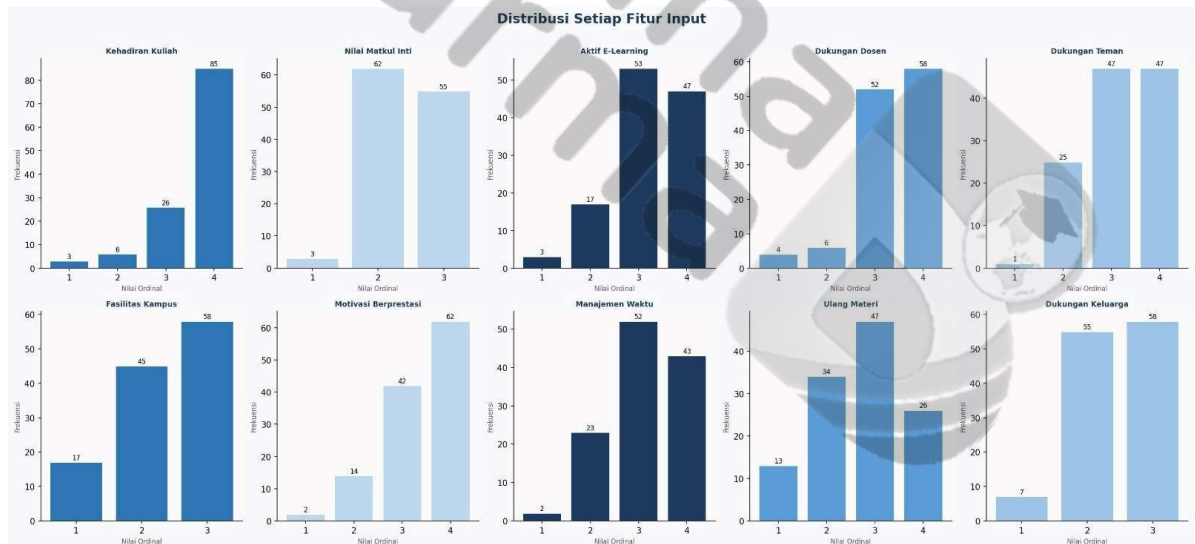
Dengan demikian, evaluasi menggunakan 100 eksperimen random split memberikan gambaran yang lebih komprehensif mengenai kestabilan dan kemampuan generalisasi model dalam memprediksi prestasi akademik mahasiswa.

4.13 Distribusi Setiap Fitur Input

Grafik dari setiap fitur input menunjukkan distribusi nilai dari setiap fitur input yang digunakan dalam penelitian untuk memprediksi prestasi akademik siswa dengan menggunakan metode

Naive Bayes. Selain itu, grafik ini menunjukkan frekuensi jawaban responden untuk setiap kategori nilai ordinal dari berbagai variabel yang diperoleh dari kuesioner yang diberikan kepada siswa.

Dalam variabel kehadiran kuliah, sebagian besar mahasiswa memberikan nilai 4 (dari 85 responden), diikuti oleh nilai 3 (dari 26 responden), nilai 2 (dari 6 responden), dan nilai 1 (dari 3 responden). Ini menunjukkan bahwa sebagian besar mahasiswa sangat hadir.



Gambar 4.8 Distribusi Setiap Fitur Input

Dalam variabel Nilai Mata Kuliah Inti, mayoritas mahasiswa berada dalam kategori nilai 2, yaitu 62 responden dan 55 responden, dengan hanya 3 responden yang berada dalam kategori nilai 1. Ini menunjukkan bahwa sebagian besar mahasiswa mencapai nilai yang cukup baik pada mata kuliah inti.

Variabel Aktif dalam E-Learning menunjukkan bahwa sebagian besar siswa sangat aktif dalam mengikuti pembelajaran

online. Hasilnya menunjukkan bahwa 53 orang berada dalam kategori nilai 3, 47 orang berada dalam kategori nilai 4, dan 17 orang berada dalam kategori nilai 2. Ini menunjukkan bahwa sebagian besar siswa sangat aktif dalam mengikuti pembelajaran online.

Dalam variabel Dukungan Dosen, sebagian besar responden memberikan nilai 4 (58 responden), nilai 3 (52 responden), nilai 2 (6 responden), dan nilai 1 (4 responden). Ini menunjukkan bahwa sebagian besar siswa merasakan dukungan yang baik dari guru selama proses pembelajaran.

Untuk variabel Dukungan Teman, distribusi responden menunjukkan bahwa 47 orang menerima nilai 3, 25 orang menerima nilai 2, dan satu orang menerima nilai satu. Hal ini menunjukkan bahwa hubungan sosial antara siswa sangat baik dan membantu mereka belajar.

Dalam variabel Fasilitas Kampus, sebagian besar responden memberikan nilai 3, yaitu 58 orang, diikuti oleh nilai 2, yaitu 45 orang, dan nilai 1 yaitu 17 orang. Ini menunjukkan bahwa mahasiswa menilai fasilitas kampus dengan cukup baik untuk mendukung proses pembelajaran.

Menurut Variabel Motivasi Berprestasi, sebagian besar orang yang menjawab memiliki nilai 4, sebanyak 62 orang, nilai 3, sebanyak 42 orang, nilai 2, sebanyak 14 orang, dan nilai 1 sebanyak 2 orang. Ini menunjukkan bahwa sebagian besar siswa memiliki keinginan yang besar untuk mencapai prestasi akademik yang baik.

Dalam variabel manajemen waktu, sebagian besar siswa memiliki nilai 3 (52 responden), nilai 4 (43 responden), dan

nilai 1 (2 responden). Ini menunjukkan bahwa sebagian besar siswa dapat mengatur waktu mereka dengan baik untuk aktivitas akademik mereka.

Variabel Kebiasaan Mengulang Materi menunjukkan bahwa nilai 3 memiliki jumlah responden tertinggi, yaitu 47 responden, diikuti oleh nilai 2 sebanyak 34 responden, nilai 4 sebanyak 26 responden, dan nilai 1 sebanyak 13 responden. Ini menunjukkan bahwa sebagian besar siswa memiliki kebiasaan yang cukup baik untuk mengulang materi pelajaran. Terakhir, pada variabel Dukungan Keluarga, sebagian besar orang yang menjawab memiliki nilai 3, yaitu 58 orang, nilai 2, 55 orang, dan nilai 1, yaitu 7 orang. Ini menunjukkan bahwa sebagian besar siswa menerima dukungan yang cukup dari keluarga mereka selama pendidikan mereka.

Secara keseluruhan, berdasarkan distribusi setiap fitur input, terlihat bahwa sebagian besar responden berada pada kategori nilai menengah hingga tinggi pada hampir semua variabel. Ini menunjukkan bahwa sebagian besar siswa memiliki lingkungan belajar dan prestasi akademik yang baik. Oleh karena itu, ketika algoritma Naive Bayes digunakan, variabel-variabel ini dapat menjadi komponen penting dalam proses klasifikasi prestasi akademik siswa.

BAB IV

PENUTUP

5.1 Kesimpulan

Berikut adalah ringkasan kesimpulan dari penelitian tentang penerapan metode Naïve Bayes untuk memprediksi hasil belajar mahasiswa pada Program Studi Teknologi Informasi Fakultas Teknik Universitas Muhammadiyah Palembang, Metode Naïve Bayes efektif digunakan untuk mengklasifikasikan hasil belajar mahasiswa berdasarkan data akademik seperti IPK, nilai UTS, nilai UAS, dan tingkat kehadiran, Variabel IPK dan tingkat kehadiran merupakan faktor paling berpengaruh dalam menentukan kategori hasil belajar mahasiswa, Model yang dikembangkan memiliki performa baik dengan akurasi sekitar 82%, serta nilai precision, recall, dan F1-score yang stabil dan Model prediksi ini dapat digunakan sebagai sistem pendukung keputusan untuk membantu fakultas dalam pemantauan dan intervensi dini terhadap mahasiswa yang berpotensi mengalami penurunan prestasi akademik.

5.2 Saran

Berdasarkan hasil penelitian, beberapa saran yang dapat diberikan adalah Penelitian selanjutnya dapat membandingkan metode Naïve Bayes dengan algoritma klasifikasi lainnya untuk memperoleh hasil yang lebih optimal, Perlu dilakukan penambahan variabel non-akademik seperti motivasi belajar atau faktor sosial untuk meningkatkan akurasi model Dan Model prediksi dapat dikembangkan menjadi sistem berbasis web yang terintegrasi dengan sistem akademik kampus.



DAFTAR PUSTAKA

- Anggraini, S., & Pribadi, D. (2019). Penerapan Metode Naïve Bayes dalam Prediksi Hasil Belajar Mahasiswa. *Jurnal Ilmiah Informatika*, 8(2), 45–53.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York: Longmans.
- Carudin, et al. (2024). Data Mining untuk Prediksi Hasil Belajar. *Jurnal Teknologi Informasi*, 12(3), 101–110.
- Connolly, T., & Begg, C. (2010). *Database Systems: A Practical Approach to Design, Implementation, and Management* (5th ed.). Pearson Education.
- Dahri, D., Agus, F., & Khairina, D. (2016). Implementasi Naïve Bayes untuk Prediksi Data Akademik. *Jurnal Informatika*, 10(1), 12–20.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37–54.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Handayani, A., & Pribadi, D. (2015). Pengklasifikasian Data Akademik Menggunakan Naïve Bayes. *Jurnal Sistem Informasi*, 6(2), 89–97.
- Harijanto, B., et al. (2018). Implementasi Naïve Bayes pada Sistem Klasifikasi. *Jurnal Teknologi Informasi*, 9(1), 23–30.
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley.
- Luo, L. (2021). Naïve Bayes Classification in Machine Learning. *International Journal of Data Science*, 5(2), 50–60.
- Moriesta, et al. (2017). Penerapan Teorema Bayes dalam Klasifikasi Data. *Jurnal Informatika*, 11(1), 33–41.

- Rish, I. (2001). An Empirical Study of the Naïve Bayes Classifier. *IJCAI Workshop on Empirical Methods in Artificial Intelligence*, 41–46.
- Syarifudin, M., Hidayat, N., & Fanani, L. (2018). Implementasi Algoritma Naïve Bayes dalam Prediksi Akademik. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(3), 1200–1206.
- Turban, E., Sharda, R., & Delen, D. (2011). *Decision Support and Business Intelligence Systems*. Pearson.
- Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann.

LAMPIRAN

Formulir Perbaikan Tesis

 ISO 9001 : 2000	Formulir Perbaikan Tesis	Nomor Dok :
		Nomor Revisi :
		Tgl. Berlaku :
		Klausa ISO :

Nama : Muhammad Febrian
 NIM : 232420022
 Konsentrasi : ENTERPRISE IT INFRASTRUCTURE
 Judul Tesis : ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI HASIL BELAJAR MAHASISWA MENGGUNAKAN METODE NAIVE BAYES PADA FAKULTAS TEKNIK PRODI TEKNOLOGI INFORMASI UNIVERSITAS MUHAMMADIYAH PALEMBANG
 Dosen Pembimbing : Dr Tata Sutabri, S.Kom., M.M.S.I
 Tanggal Ujian : 05 Maret 2026

Telah diperbaiki dan dikonsultasikan dengan Pembimbing/Penguji Tesis.

No.	Nama Dosen Penguji	Tanggal	Tanda Persetujuan
1.	Dr. Tata Sutabri, S.Kom., M.M.S.I	12 Maret 26	1. 
2.	Prof. Dr. Edi Surya Negara, M.Kom		2. 
3.	Dr. Usman Ependi, S.Kom., M.Kom		3. 

*Nb.
 Pembimbing harap memeriksa kembali format dari tesis yang telah diperbaiki dan keabsahan tanda tangan penguji

Palembang, 05 Maret 2026
 Program Studi Teknik Informatika – S2
 Ketua,


 Magister Teknik Informatika.
Dr. Usman Ependi, S.Kom., M.Kom.

Surat Izin Penelitian



Universitas Bina Darma
PROGRAM PASCASARJANA

Palembang, 19 Mei 2025

Nomor : 163/PPs-UBD/MTI/V/2025
Perihal : Izin Penelitian
Lamp. : -

Kepada Yth.
BAK Universitas Muhammadiyah Palembang
Jl. Jendral A. Yani Kel. 13 Ulu Kec. Seberang Ulu II Kota Palembang

Dengan hormat,

Sesuai dengan Kurikulum Program Pascasarjana Universitas Bina Darma, setiap mahasiswa diwajibkan melaksanakan penyusunan Laporan Tesis. Sehubungan dengan ini, kami mengharapkan kesediaan Bapak/Ibu untuk memberikan kesempatan dan izin kepada mahasiswa:

Nama	: Muhammad Febrian
Nim	: 232420022
Konsentrasi	: Enterprise IT Infrastructure
Program Studi	: Teknik Informatika – S2
Judul	: Analisis dan Implementasi Data mining untuk memprediksi hasil Belajar Mahasiswa menggunakan Metode Naïve Bayes pad Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang

akan mengadakan penelitian dalam lingkungan perusahaan/instansi/pemerintahan yang Bapak/Ibu pimpin. Hasil penelitian tersebut dipergunakan sebagai bahan kajian laporan dalam bentuk Laporan Tesis. Sehubungan dengan ini, mohon kiranya Bapak/Ibu pimpinan dapat menunjuk 1 (satu) orang Pendamping Lapangan selama mahasiswa tersebut melakukan penelitian di perusahaan/instansi/pemerintahan Bapak/Ibu pimpin.

Demikian kami sampaikan. Atas kerja sama dan bantuannya, kami ucapkan terima kasih.

Hormat kami,
Program Pascasarjana
Direktur



Universitas Bina Darma
PROGRAM PASCASARJANA
Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Cc. Arsip

Jl. Jend A. Yani No. 3 Palembang 30264 Indonesia Telp. (+62 711) 515679, 515581, 515582
Fax (+62 711) 515581, 515582 Website www.binadarma.ac.id , e-mail : universitas@mail.binadarma.ac.id

Formulir Perbaikan Formula Tesis

 ISO 9001 : 2000	FORMULIR Perbaikan Proposal Tesis	NomorDok :	
		NomorRevisi :	
		Tgl. Berlaku :	
		Klausa ISO :	

Nama : MUHAMMAD FEBRIAN
Nim : 23242002
Konsentrasi : ENTERPRISE IT INFRASTRUCTURE
JudulTesis : ANALISIS DAN IMPLEMENTASI DATA MINING UNTUK
MEMPREDIKSI HASIL BELAJAR MAHASISWA
MENGUNAKAN METODE NAÏVE BAYES PADA
FAKULTAS TEKNIK PRODI TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH PALEMBANG

(Kepada mahasiswa harap di isi sesuai dengan hasil ujian proposal tesis).

Dosen Pembimbing : Dr. Tata Sutabri, S Kom., M.M.S.I

Tanggal Seminar : 30 Juli 2025

Telah diperbaiki dan dikonsultasikan dengan Pembimbing/Penguji Proposal Tesis.

No.	NamaDosenpembimbing/Penguji	Tanggal	TandaPersetujuan
1.	Dr. Tata Sutabri, S Kom., M.M.S.I.	20/8 2025	1.
2.	Prof. Dr. Edi Surya Negara, M.Kom.	20/8 2025	2.
3.	Dr. Usman Ependi, S.Kom., M.Kom.	20/8 2025	3.

Palembang, 20 - Agustus 2025
Program Studi Teknik Informatika – S2
Ketua,

Dr. Usman Ependi, S.Kom., M.Kom.

NB.

Harap segera menyerahkan foto copi formulir perbaikan proposal tesis sebanyak 2 lembar ke
Sekretariat Pascasarjana, apabila telah di tanda tangan oleh penguji dan disahkan oleh Ketua
Program Studi untuk proses pembuatan SK Pembimbing

Formulir Permohonan Surat Pengantar Izin Penelitian Tesis

	Formulir Permohonan Surat Pengantar Izin Penelitian Tesis/Internship	
---	---	---

Nomor Dok: FRM/PKL/03
Tanggal : 1 Mei 2006 Rev. 00

Palembang, 19 Mei 2025

Perihal : **Permohonan Surat Pengantar
Izin Penelitian Tesis/Internship**

Kepada Yth.
Kepala Sekretariat Program Pascasarjana
Universitas Bina Dharma
di –
Palembang.

Dengan hormat,

Saya yang bertanda tangan dibawah ini, mahasiswa Program Pascasarjana Universitas Bina Dharma:

Nama : Muhammad Febrian
Nim : 232420022
Angkatan/Reguler : MTI 29 B
Konsentrasi : Enterprise IT Infrastructure
Program Studi : Teknik Informatika- S2
Judul Tesis/Internship : Analisis dan implemtansi data maining untuk memprediksi hasil belajar mahasiswa menggunakan metode naive bayes pada fakultas teknik prodi teknologi informasi Universitas Muhammadiyah Palembang

Dengan ini bermaksud mengajukan permohonan untuk dibuatkan surat pengantar izin penelitian Tesis/Internship di instansi:

Nama Instasi : Universitas Muhammadiyah Palembang
Bagian/Unit Kerja : B.A.K Universitas Muhammadiyah Palembang
Alamat : Jl. Jendral A. Yani, Kel. 13 Ulu, Kec. Seberang Ulu II, Palembang, Kode Pos 30263

Atas perhatiannya, saya ucapkan terima kasih.

Mengetahui,
Ketua Program Studi

(Dr. Usman Ependi, M.Kom)

Hormat saya,

(Muhammad Febrian)

Nt. Melampirkan Fotocopy Formulir Persetujuan Pembimbing dan Judul Tesis/Internship

Lembar Konsultasi Proposal Tesis



**PROGRAM PASCASARJANA
TEKNIK INFORMATIKA – S2
UNIVERSITAS BINA DARMA**

LEMBAR KONSULTASI PROPOSAL TESIS

Nama : Muhammad Febrian
Nim : 232420022
Konsentrasi : ENTERPRISE IT INFRASTRUCTUR
Judul : Analisis Dan Implementasi Data Mining Untuk Memprediksi Hasil Belajar Mahasiswa Menggunakan Metode Naive Bayes Pada Prodi Teknologi Informasi Universitas Muhammadiyah Palembang

Pembimbing : Dr. Tata Sutabri, S.Kom., M.M.S.I.

No	Tanggal	Uraian Materi Konsultasi	Paraf
1.	30 November 2024	Bimbingan Judul Proposal Tesis	
2.	4-jan-2025	membuka wawasan untuk pengisian DAB	
3.	5-maret-2025	Revisi BAB I dan BAB II	
4.	10-April-2025	ACC BAB I dan BAB II	
5.	5-Mei-2025	Revisi Pembahasan BAB III	
6.	15-mei-2025	ACC BAB III	
7.	20-Mei-2025	Siap ujian Proposal.	

Surat Izin Pelaksanaan



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

SURAT IZIN MELAKSANAKAN PENELITIAN
Nomor : 1077/C-12/UMP/VII/2025

Memenuhi surat Direktur Program pascasarjana Universitas Bina Darma Palembang No. 163/Pps-UBD/MTI/V/2025 tanggal 19 Mei 2025, Rektor Universitas Muhammadiyah Palembang dengan ini memberikan izin kepada Saudara :

Nama	: Muhammad Febrian
NIM	: 232420022
Jurusan/Program Studi	: Teknik Informatika-S2 Universitas Bina Darma
Konsentrasi	: Enterprise IT Infrastructure
Judul Penelitian	: <i>Analisis dan Implementasi Data mining untuk memprediksi hasil belajar Mahasiswa menggunakan Metode Naive Bayes pada Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang</i>

Untuk melaksanakan penelitian/observasi di lingkungan Universitas Muhammadiyah Palembang dengan Pendamping Lapangan Sdr. Kms. Wahyu Hidayat, S.Kom., M.kom. dalam rangka penyusunan Laporan Tesis.

Surat izin ini berlaku terhitung mulai tanggal ditetapkan sampai dengan selesai, dengan catatan bahwa semua data yang dikumpulkan/diperoleh hanya untuk keperluan penelitian/observasi dan tidak untuk dipublikasikan.

Demikian Surat Izin ini dibuat untuk dipergunakan sebagaimana mestinya.

Dibuat di : Palembang
Pada tanggal : 15 Muharram 1447 H
10 Juli 2025 M
a.n. Rektor
Wakil-Rektor I Bidang Akademik,


Prof. Dr. Ir. Kiagus Ahmad Roni, S.T., M.T., IPM., ASEAN.Eng.
NBM/NIDN : 763049/0227077004

Tembusan :

1. Rektor sebagai laporan
2. Dekan Fakultas di lingkungan UM Palembang
3. Kepala Biro/Unit di lingkungan UM Palembang
4. Kepala Satuan Pengamanan UM Palembang
5. Yang bersangkutan

(Pascasarjana Program Studi : Hukum, Manajemen, Pend. Biologi & Teknik Kimia)
Fakultas : Teknik, Ekonomi & Bisnis, Keguruan & Ilmu Pendidikan, Pertanian, Hukum, Agama Islam & Kedokteran.
www.um-palembang.ac.id | facebook.com/UMPalembang | umpcenter | umpalembang

Formulir Persetujuan Pembimbing

	FORMULIR PERSETUJUAN Pembimbing dan Judul Tesis	NomorDok. :	
		NomorRevisi :	
		TanggalBerlaku :	
		Klausa ISO :	

Nama : Muhammad Febrian
 NIM : 232420022
 Program Studi : Teknik Informatika - S2
 Konsentrasi : Enterprise IT Infrastructure
 Kelas : Reguler B
 Tema yang Disetujui oleh Ketua Program Studi : Penerapan Data Mining Dalam Memprediksi Hasil Belajar Mahasiswa Di Prodi Teknologi Informasi Universitas Muhammadiyah Palembang
 Judul Tesis yang disetujui oleh Pembimbing : Analisis Dan Implementasi Data Maining Untuk Memprediksi Hasil Belajar Mahasiswa Menggunakan Metode Naïve Bayes Pada Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang

DosenPendamping	TandaTangan	Tanggal
Dr. Tata Sutabri, S.Kom., M.M.S.I.		30 November 2024

Palembang, 30 November 2024
 Menyetujui,
 Ketua Program Studi

Dr. Usman Ependi, M.Kom.

Catatan:

- JUDUL TESIS TERLEBIH DAHULU DIKONSULTASIKAN DENGAN MASING-MASING DOSEN PEMBIMBING
- DOSEN PEMBIMBING WAJIB MEMASTIKAN OBJEK PENELITIAN (LIHAT SURAT BALASAN PERSETUJUAN DARI PERUSAHAAN / INSTANSI / LEMBAGA PENDIDIKAN / PERBANKAN / ORGANISASI / RUMAH SAKIT / KONSULTAN/DAN LAIN-LAIN.
- LEMBAR PERSETUJUAN PENGAJUAN JUDUL TERLEBIH DAHULU DI TANDA TANGAI MASING-MASING DOSEN PEMBIMBING BARU DITANDA TANGANI OLEH KETUA PROGRAM STUDI.

Studi Kasus Program Studi Teknologi Informasi: Metode Naive Bayes untuk Memprediksi Hasil Belajar Mahasiswa

Muhammad Febrian^{1*}; Tata Sutabri²; Edi Surya Negara^{3*}, usman endi^{4*}.

Program Studi Magister Teknik Informatika¹²³⁴
Universitas Bina Darma, Palembang, sumatra Selatan¹²³⁴
Binadarma.ac.id¹
Author's E-mail^{1*}

¹ e-mail: febrianmuhammad311099@gmail.com

² e-mail: tata.sutabri@binadarma.ac.id

³ e-mail: edi.suryanegara@binadarma.ac.id

⁴ e-mail: usman.ependi@binadarma.ac.id

(*) Corresponding Author

(Responsible for the Quality of Paper Content)



The creation is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License.

Abstract— This research aims to analyze and implement data mining methods using the Naïve Bayes algorithm to predict student learning outcomes in the Faculty of Engineering, Information Technology Study Program at Muhammadiyah University Palembang. The background of this research is based on the suboptimal utilization of academic data in supporting strategic decision-making in the higher education environment. The data used comes from the Academic Information System (SIKAD) and student questionnaire results, with a total dataset of around 200 students. The variables analyzed include course grades, Cumulative Grade Point Average (CGPA), attendance rate, and the number of credit hours. The research process was carried out thru stages of data collection, preprocessing (data cleaning, transformation, and reduction), splitting the data into training and testing sets with an 80:20 ratio, modeling using the Naïve Bayes algorithm, and evaluation using confusion matrix, accuracy, precision, recall, and F1-score. The research results show that the Naïve Bayes model is capable of providing good classification performance with an accuracy rate of 82%, precision of 85%, recall of 82%, and F1-score of 83%. The model is more accurate in classifying students in the "Very Active" category compared to the "Less Active" category. These findings indicate that the Naïve Bayes method is effective in predicting students' academic performance and can be utilized as a decision support system in identifying students who require further academic attention.

Keywords: Data Mining, Naïve Bayes, Student Learning Outcomes, Academic Prediction, Confusion Matrix, Decision Support System

Intisari— Penelitian ini bertujuan menganalisis dan mengimplementasikan metode data mining menggunakan algoritma Naïve Bayes dalam memprediksi hasil belajar mahasiswa pada Fakultas Teknik Program Studi Teknologi Informasi Universitas Muhammadiyah Palembang. Latar belakang penelitian ini didasarkan pada pemanfaatan data akademik yang belum optimal dalam mendukung pengambilan keputusan strategis di lingkungan perguruan tinggi. Data yang digunakan berasal dari Sistem Informasi Akademik (SIKAD) serta hasil kuesioner mahasiswa, dengan total dataset sekitar 200 mahasiswa. Variabel yang dianalisis meliputi nilai mata kuliah, Indeks Prestasi Kumulatif (IPK), tingkat kehadiran, dan jumlah SKS. Proses penelitian dilakukan melalui tahapan pengumpulan data, preprocessing (data cleaning, transformasi, dan reduksi), pembagian data training dan testing dengan rasio 80:20, pemodelan menggunakan algoritma Naïve Bayes, serta evaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Naïve Bayes mampu memberikan performa klasifikasi yang baik dengan tingkat akurasi sebesar 82%, precision 85%,



recall 82%, dan F1-score 83%. Model lebih akurat dalam mengklasifikasikan mahasiswa kategori “Sangat Aktif” dibandingkan kategori “Kurang Aktif”. Temuan ini menunjukkan bahwa metode Naïve Bayes efektif digunakan untuk memprediksi hasil belajar mahasiswa dan dapat dimanfaatkan sebagai sistem pendukung keputusan dalam mengidentifikasi mahasiswa yang memerlukan perhatian akademik lebih lanjut.

Kata Kunci : *Data Mining, Naïve Bayes, Hasil Belajar Mahasiswa, Prediksi Akademik, Confusion Matrix, Sistem Pendukung Keputusan*

INTRODUCTION

In the current digital era, data-based technology has become increasingly sophisticated, creating a need to analyze the knowledge derived from collected data. Due to the large number of data, information that is frequently expensive cannot be obtained in a straightforward manner [1]. Because of this, there is a need for a method to extract important information that has been developed, and one effective method is data mining. The primary goal of educational institutions, such as Universitas Muhammadiyah Palembang's Program Studi Teknologi Informasi, is to improve the quality of education that can be achieved by mastery of the subject [2]. A number of factors, such as kehadiran, nilai tugas, ujian, and akademik latar, affect the results of mahasiswa learning [3]. In other words, academic data that is frequently used is just used as an arsip and somewhat optimized for strategic analysis to support keputusan development. Furthermore, with systematic analysis, the data can be used to identify important components that affect academic achievement and predict learning outcomes in the future. This process of gaining knowledge from databases is known as Knowledge Discovery in Databases (KDD), where data mining is one of its initial steps [4]. Numerous studies have demonstrated the effectiveness of the classification algorithm in educational contexts. According to Rahman & Susanto's (2020) research, kelulusan mahasiswa using Naïve Bayes has an accuracy of 82% [5]. Additional research by Putri & Nugroho (2021) confirms the effectiveness of academic attributes in predicting mahasiswa performance using the same algorithm, with an accuracy level of 80–85% [6]. According to Hidayat & Lestari (2019), they compare the algorithms C4.5 and Naïve Bayes and conclude that the latter is more suitable for small datasets [7]. Ultimately, these studies provide strong evidence that the Naïve Bayes method, with its accuracy and ability to provide good results, has a great deal of potential for use in academic prediction [8]. Although the Naïve Bayes method has been widely used, this study offers new information in a specific context, namely the Faculty of Technology Program Study of Information

Technology at Universitas Muhammadiyah Palembang, which has not yet been widely discussed. In addition, this research goes beyond just developing a model; it also incorporates preliminary data from the researcher to validate non-academic factors such learning motivation and ensuring that the variable is relevant to the condition of the mahasiswa. The main goal of this study is to analyze and apply the Naïve Bayes algorithm to predict student learning outcomes, identify significant factors that affect them, and evaluate the model's performance.

MATERIALS AND METHODS

In the current digital era, data-based technology has become increasingly sophisticated, creating a need to analyze the knowledge derived from collected data. Due to the large number of data, information that is frequently expensive cannot be obtained in a straightforward manner [1]. Because of this, there is a need for a method to extract important information that has been developed, and one effective method is data mining. The primary goal of educational institutions, such as Universitas Muhammadiyah Palembang's Program Studi Teknologi Informasi, is to improve the quality of education that can be achieved by mastery of the subject [2]. A number of factors, such as kehadiran, nilai tugas, ujian, and akademik latar, affect the results of mahasiswa learning [3]. In other words, academic data that is frequently used is just used as an arsip and somewhat optimized for strategic analysis to support keputusan development. Furthermore, with systematic analysis, the data can be used to identify important components that affect academic achievement and predict learning outcomes in the future. This process of gaining knowledge from databases is known as Knowledge Discovery in Databases (KDD), where data mining is one of its initial steps [4]. Numerous studies have demonstrated the effectiveness of the classification algorithm in educational contexts. According to Rahman & Susanto's (2020) research, kelulusan mahasiswa using Naïve Bayes has an accuracy of 82% [5]. Additional research by Putri & Nugroho (2021) confirms the effectiveness of academic attributes in predicting mahasiswa performance



using the same algorithm, with an accuracy level of 80–85% [6]. According to Hidayat & Lestari (2019), they compare the algorithms C4.5 and Naïve Bayes and conclude that the latter is more suitable for small datasets [7]. Ultimately, these studies provide strong evidence that the Naïve Bayes method, with its accuracy and ability to provide good results, has a great deal of potential for use in academic prediction [8]. Although the Naïve Bayes method has been widely used, this study offers new information in a specific context, namely the Faculty of Technology Program Study of Information Technology at Universitas Muhammadiyah Palembang, which has not yet been widely discussed. In addition, this research goes beyond just developing a model; it also incorporates preliminary data from the researcher to validate non-academic factors such learning motivation and ensuring that the variable is relevant to the condition of the mahasiswa. The main goal of this study is to analyze and apply the Naïve Bayes algorithm to predict student learning outcomes, identify significant factors that affect them, and evaluate the model's performance.

RESULTS AND DISCUSSION

This study successfully applies the Naïve Bayes algorithm to predict learning outcomes. The model is built using 160 latih data and evaluated using 40 uji data.

1. Kinerja Model Evaluation using Confusion Matrix
Kinerja model is evaluated using a confusion matrix that compares the actual and predicted values of the uji data for the two primary categories of "Sangat Aktif" and "Kurang Aktif." The results show that out of 40 uji data, the model successfully classified 23 data as benar and 5 data as salah.

Tabel 1. Confusion Matrix Hasil Prediksi Naïve Bayes

Aktual \ Prediksi	Kurang Aktif	Sangat Aktif
Kurang Aktif	1 (TN)	2 (FP)
Sangat Aktif	3 (FN)	22 (TP)
<i>Sumber: (Hasil Penelitian, 2026)</i>		

True Positive (TP) = 22: "Sangat Aktif" mahasiswa is predicted by the model.

Model benar memprediksi mahasiswa "Kurang Aktif": True Negative (TN) = 1.

False Positive (FP) = 2: The model suggests that "Kurang Aktif" is a "Sangat Aktif" mahasiswa.

False Negative (FN) = 3: The model predicts "Sangat Aktif" students as "Kurang Aktif" students.

To obtain more clear explanations, the Normalized Confusion Matrix is used. The results show that only 33% of the data in the "Kurang Aktif" category can be predicted by the model, whereas 67% of the data are classified as "Sangat Aktif". Conversely, in the "Sangat Aktif" class, the model indicates very good performance, with a prediction accuracy of 88% and just 12% salah. This indicates that the "Sangat Aktif" model is more accurate than the "Kurang Aktif" model in terms of mahasiswa characteristics. This discrepancy may be caused by the distribution of data that was used for the "Sangat Aktif" class, making the model less than ideal for teaching minoritas [9].

2. Model Performance Based on Evaluation Metrics

Based on the confusion matrix's values, an evaluation method is used to improve the model's overall performance. The results are shown in Table 2 and illustrated in Figure 1.

Table 2: Performance Evaluation of the Naïve Bayes Model

Metrik	Rumus	Hasil
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	82%
Precision	$TP / (TP+FP)$	85%
Recall	$TP / (TP+FN)$	82%
F1-Score	$2 \times (Precision \times Recall) / (Precision + Recall)$	83%

CONCLUSION

This study successfully applies the Naïve Bayes method to predict the learning outcomes of students in the Program Studying Information Technology at Universitas Muhammadiyah Palembang. The developed model indicates good performance with an accuracy threshold of 82%, precision of 85%, recall of 82%, and F1-score of 83%. This study confirms that the Naïve Bayes method is effective for classifying learning outcomes based on academic data such as IPK, UTS/UAS, and kehadiran tingkat. The longer analysis shows that the model is more accurate in predicting the behavior of students in the "Sangat Aktif" category compared to "Kurang Aktif" category. It is also supported by kuesioner data that indicates the level of motivation and kehadiran among students. This study offers a practical contribution in the form of a prediksi model that

may be used as a system of keputusan awal pendukung for faculty members to carry out dini intervention and pemantauan against mahasiswa who may experience academic difficulties. Subsequent research might compare the Naïve Bayes algorithm with other classification techniques, such as Support Vector Machine or Random Forest, and emphasize the addition of more non-academic variables to improve model accuracy.

REFERENCE

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012.
- [2] B. S. Bloom, *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York: Longmans, 1956.
- [3] S. Anggraini and D. Pribadi, "Penerapan Metode Naïve Bayes dalam Prediksi Hasil Belajar Mahasiswa," *J. Ilm. Inform.*, vol. 8, no. 2, pp. 45–53, 2019.
- [4] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Mag.*, vol. 17, no. 3, pp. 37–54, 1996.
- [5] Rahman and Susanto, "Penerapan Algoritma Naïve Bayes untuk Prediksi Kelulusan Mahasiswa," *J. Teknol. Inf.*, vol. 9, no. 1, pp. 12–20, 2020.
- [6] Putri and Nugroho, "Implementasi Data Mining dengan Naïve Bayes dalam Prediksi Prestasi Akademik Mahasiswa," *J. Inform.*, vol. 11, no. 2, pp. 55–63, 2021.
- [7] Hidayat and Lestari, "Analisis Prediksi Nilai Mahasiswa Menggunakan C4.5 dan Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 4, pp. 1200–1206, 2019.
- [8] I. Rish, "An Empirical Study of the Naïve Bayes Classifier," in *IJCAI Workshop on Empirical Methods in Artificial Intelligence*, 2001, pp. 41–46.
- [9] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Morgan Kaufmann, 2016.
- [10] Kurniawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes," *J. Nas. Teknol. dan Sist. Inf.*, vol. 8, no. 1, pp. 22–30, 2022.