

SKRIPSI

**ANALISIS PREDIKSI KELULUSAN MAHASISWA DENGAN
METODE *NAIVE BAYES CLASSIFIER***

**(Studi Kasus : Program Studi Diploma III Teknologi Bank Darah
di Poltekkes BSI Yogyakarta)**



MASUNGGING DWI CAHYO

NIM : 205411150

**PROGRAM STUDI INFORMATIKA
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS TEKNOLOGI
DIGITAL INDONESIA YOGYAKARTA**

2024

SKRIPSI

**ANALISIS PREDIKSI KELULUSAN MAHASISWA DENGAN
METODE *NAIVE BAYES CLASSIFIER***

**(Studi Kasus : Program Studi Diploma III Teknologi Bank Darah
di Poltekkes BSI Yogyakarta)**

Diajukan sebagai salah satu syarat untuk menyelesaikan studi



Disusun Oleh

MASUNGGING DWI CAHYO

NIM : 205411150

**PROGRAM STUDI INFORMATIKA
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

UJIAN SKRIPSI

Judul : Analisis Prediksi Kelulusan Mahasiswa Dengan Metode *Naive Bayes Classifier* (Studi Kasus : Program Studi Diploma III Teknologi Bank Darah di Poltekkes BSI Yogyakarta).

Nama : Masungging Dwi Cahyo

NIM : 205411150

Program Studi : Informatika

Program : Sarjana

Semester : Genap

Tahun Akademik : 2023/2024

Telah diperiksa dan disetujui untuk diujikan di hadapan Dewan Penguji Skripsi

Yogyakarta, 31 Juli 2024

Dosen Pembimbing

Sri Redjeki, S.Si., M.Kom., Ph.D.

NIDN : 0521047401

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS PREDIKSI KELULUSAN MAHASISWA DENGAN METODE
NAIVE BAYES CLASSIFIER**

**(Studi Kasus : Program Studi Diploma III Teknologi Bank Darah di
Poltekkes BSI Yogyakarta)**

**Telah dipertahankan di depan Dewan Penguji Skripsi dan dinyatakan
diterima untuk memenuhi sebagian persyaratan guna memperoleh Gelar**

**Sarjana Komputer
Program Studi Informatika
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta**

Yogyakarta, 31 Juli 2024

Dewan Penguji	NIDN	Tanda Tangan
1. Ariesta Damayanti, S.Kom., M.Cs.	0020047801
2. Maria Mediatrix Sebatubun, S.Kom., M.Eng.	0514089101
3. Sri Redjeki, S.Si., M.Kom., Ph.D.	0521047401

Mengetahui

Ketua Program Studi Informatika

Dini Fakta Sari, S.T., M.T.

NIDN : 0507108401

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa naskah skripsi ini belum pernah diajukan untuk memperoleh gelar Sarjana Komputer di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, Juli 2024

Masungging Dwi Cahyo

NIM : 205411150

HALAMAN PERSEMBAHAN

Untuk ibunda dan ayahanda serta keluarga tercinta sebagai tanda bakti, hormat, dan rasa terima kasih yang tiada terhingga kupersembahkan karya kecil ini yang telah memberikan kasih sayang, segala dukungan, dan cinta kasih yang tiada terhingga yang tiada mungkin dapat kubalas hanya dengan selebar kertas yang bertuliskan kata cinta dalam kata persembahan. Semoga ini menjadi langkah awal untuk membuat Ibu dan Ayah bahagia dan bangga memiliki anak sepertiku, walau aku sadar bahwa selama ini aku belum bisa membahagiakanmu.

Skripsi ini aku persembahkan pula untuk orang yang saya cintai sehingga engkau sangat istimewa bagiku. Terima kasih atas dukungan, kebaikan, perhatian dan semangat yang selalu kamu curahkan untukku.

Kupersembahkan pula karya ini untuk orang-orang yang tidak mengenaliku dari aku yang sesungguhnya peduli kepadamu. Walau engkau mungkin tidak mengenalku tapi kenallah hasil karyaku ini, maka engkau akan selalu mengingatkanmu.

HALAMAN MOTTO

“Jika kamu tidak sanggup menahan lelahnya belajar, maka kamu akan menanggung perihnya kebodohan.”

(Imam Syafi'i)

“Bencana akibat kebodohan adalah sebesar-besarnya musibah seorang manusia.”

(Imam Ghazali)

“Pendidikan mempunyai akar yang pahit, tapi buahnya manis.”

(Aristotle Onassis)

“Hidup itu bukan untuk mengeluh, tapi hidup itu untuk bersyukur. Maka wujudkan syukur itu dengan amal yang manfaat.”

“Tetap semangat dan bersyukur selalu.”

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas berkat dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan skripsi dengan judul "Analisis Prediksi Kelulusan Mahasiswa Dengan Metode *Naive Bayes Classifier* (Studi Kasus : Program Studi Diploma III Teknologi Bank Darah di Poltekkes BSI Yogyakarta)". Naskah ini disusun sebagai salah satu persyaratan untuk memperoleh gelar Sarjana Komputer dari Universitas Teknologi Digital Indonesia.

Dalam proses penyusunan skripsi ini, penulis banyak mendapatkan bimbingan, dukungan, dan arahan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Ibu Sri Redjeki, S.Si., M.Kom., Ph.D., selaku Rektor Universitas Teknologi Digital Indonesia.
2. Bapak Dr. Bambang Purnomosidi DP, S.E., Akt., S.Kom., MMSI, selaku Dekan Fakultas Teknologi Informasi Universitas Teknologi Digital Indonesia.
3. Ibu Dini Fakta Sari, S.T., M.T., selaku Ketua Program Studi Informatika Universitas Teknologi Digital Indonesia.
4. Ibu Sri Redjeki, S.Si., M.Kom., Ph.D., selaku dosen pembimbing skripsi yang telah membimbing, mengarahkan, memberikan semangat, dan mendoakan penulis dalam menyelesaikan skripsi ini.
5. Ibu Ariesta Damayanti, S.Kom., M.Cs. dan Ibu Maria Mediatrice Sebatubun, S.Kom., M.Eng., selaku dosen penguji. Terima kasih atas waktu, masukan dan arahnya.
6. Orang tua dan keluarga tercinta yang telah memberikan bantuan baik secara material maupun moril, serta senantiasa memberikan kasih sayang dan semangat kepada penulis

7. Semua pihak yang tidak dapat penulis sebutkan satu persatu, yang telah memberikan bantuan, dukungan, dan doanya kepada penulis dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan adanya kritik dan saran yang membangun dari pembaca, guna menyempurnakan kajian dan penulisan skripsi ini.

Yogyakarta, Juli 2024

Penulis

Masungging Dwi Cahyo

DAFTAR ISI

HALAMAN COVER	i
HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN UJIAN SKRIPSI	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
INTISARI	xvii
ABSTRACT	xviii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Ruang Lingkup	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	7
BAB 2 TINJAUAN PUSTAKA DAN DASAR TEORI	8
2.1 Tinjauan Pustaka	8
2.2 Dasar Teori	11

2.2.1	<i>Data Mining</i>	11
2.2.2	Praproses Data.....	13
2.2.3	Pembersihan Data.....	13
2.2.4	Pembagian Data	14
2.2.5	Klasifikasi	15
2.2.6	<i>Naive Bayes</i>	16
2.2.7	<i>Confusion Matrix</i>	17
2.2.8	<i>Use Case Diagram</i>	20
2.2.9	<i>Sequence Diagram</i>	20
2.2.10	<i>Website</i>	21
BAB III METODE PENELITIAN		22
3.1	Analisis Kebutuhan	22
3.1.1	Kebutuhan Perangkat Keras	22
3.1.2	Kebutuhan Perangkat Lunak	23
3.1.3	Kebutuhan Masukan (<i>Input</i>)	23
3.1.4	Kebutuhan Proses.....	25
3.1.5	Kebutuhan Keluaran (<i>Output</i>).....	38
3.2	Perancangan Sistem.....	38
3.2.1	<i>Use Case Diagram</i>	39
3.2.2	<i>Sequence Diagram</i>	39
3.3	Perancangan Antarmuka.....	40
3.4.1	Rancangan <i>Input</i>	40
3.4.2	Rancangan <i>Output</i>	42
BAB IV IMPLEMENTASI DAN PEMBAHASAN		43
4.1	Implementasi dan Uji Coba Sistem	43

4.1.1	Halaman <i>Login</i>	43
4.1.2	Halaman Utama.....	45
4.2	Implementasi Sistem	46
4.2.1	Pengujian akurasi sistem	54
4.2.2	Evaluasi	55
4.3	Pembahasan	57
BAB V PENUTUP		59
5.1	KESIMPULAN	59
5.2	SARAN	60
DAFTAR PUSTAKA		62
LAMPIRAN.....		65

DAFTAR TABEL

Tabel 2. 1. Struktur <i>Confusion Matrix</i> 2x2	18
Tabel 2. 2. Struktur <i>Confusion Matrix</i> 4x4	19
Tabel 3. 1. Data Demografis Mahasiswa.....	24
Tabel 3. 2. Data Akademik Mahasiswa.....	24
Tabel 3. 3. Jumlah Label Kelas Pada <i>Data Training</i>	30
Tabel 3. 4. Jumlah Atribut Pada Setiap Kelas Dalam <i>Data Training</i>	31
Tabel 3. 5. Probabilitas Atribut Jenis Kelamin Terhadap Setiap Label Kelas	32
Tabel 3. 6. Sampel Data Uji Yang Tidak Diketahui Label Kelasnya.....	33
Tabel 3. 7. <i>Confusion Matrix</i>	35
Tabel 3. 8. Label Kelas.....	38

DAFTAR GAMBAR

Gambar 3. 1. Diagram Alur Proses Penelitian.....	25
Gambar 3. 2. Data Mentah Sebelum Diolah.....	26
Gambar 3. 3. Data Setelah Digabungkan	27
Gambar 3. 4. Data Bersih	27
Gambar 3. 5. Data Siap Pakai.....	28
Gambar 3. 6. <i>Data Training</i>	29
Gambar 3. 7. <i>Data Testing</i>	29
Gambar 3. 8. <i>Use Case Diagram</i>	39
Gambar 3. 9. <i>Sequence Diagram</i>	40
Gambar 3. 10. Desain Antarmuka Halaman <i>Login</i>	41
Gambar 3. 11. Desain Antarmuka Halaman Prediksi.....	41
Gambar 3. 12. Desain Antarmuka Halaman Hasil Prediksi	42
Gambar 4. 1. Tampilan Halaman <i>Login</i>	43
Gambar 4. 2. Tampilan Gagal <i>Login</i>	44
Gambar 4. 3. Route Halaman <i>Login</i>	44
Gambar 4. 4. Tampilan Halaman Utama.....	46
Gambar 4. 5. Pengisian Semua Kolom Input Dengan Data Uji	46
Gambar 4. 6. Route Prediksi <i>Naive Bayes</i>	47
Gambar 4. 7. Kode Program Bagian Impor Modul Dari <i>Library</i> Eksternal.....	48
Gambar 4. 8. Kode Program Bagian Praproses Data	51
Gambar 4. 9. Kode Program Bagian Evaluasi Model	53
Gambar 4. 10. Hasil Prediksi Dari Data Uji	54
Gambar 4. 11. <i>Confusion Matrix</i>	55

DAFTAR LAMPIRAN

Lampiran 1. Tabel Probabilitas Atribut Asal Daerah Terhadap Setiap Label Kelas	65
Lampiran 2. Tabel Probabilitas Atribut IPS 1 Terhadap Setiap Label Kelas.....	65
Lampiran 3. Tabel Probabilitas Atribut IPS 2 Terhadap Setiap Label Kelas.....	66
Lampiran 4. Tabel Probabilitas Atribut IPS 3 Terhadap Setiap Label Kelas.....	66
Lampiran 5. Tabel Probabilitas Atribut IPS 4 Terhadap Setiap Label Kelas.....	67
Lampiran 6. Tabel Probabilitas Atribut Serologi Golongan Darah I Terhadap Setiap Label Kelas	67
Lampiran 7. Tabel Probabilitas Atribut Infeksi Menular Lewat Transfusi Darah I Terhadap Setiap Label Kelas.....	68
Lampiran 8. Tabel Probabilitas Atribut Serologi Golongan Darah II Terhadap Setiap Label Kelas	69
Lampiran 9. Tabel Probabilitas Atribut Infeksi Menular Lewat Transfusi Darah II Terhadap Setiap Label Kelas	70
Lampiran 10. Tabel Probabilitas Atribut Penyadapan Darah Terhadap Setiap Label Kelas	70
Lampiran 11. Tabel Probabilitas Atribut Serologi Golongan Darah III Terhadap Setiap Label Kelas.....	71
Lampiran 12. Tabel Probabilitas Atribut Infeksi Menular Lewat Transfusi Darah III Terhadap Setiap Label Kelas	72
Lampiran 13. Tabel Probabilitas Atribut Komponen Darah Terhadap Setiap Label Kelas	73
Lampiran 14. Tabel Probabilitas Atribut Serologi Golongan Darah IV Terhadap Setiap Label Kelas.....	73
Lampiran 15. Tabel Probabilitas Atribut Infeksi Menular Lewat Transfusi Darah IV Terhadap Setiap Label Kelas	74
Lampiran 16. Kode Program <i>login.html</i>	75
Lampiran 17. Kode Program <i>index.html</i>	75

Lampiran 18. Kode Program <i>hasil.html</i>	78
Lampiran 19. Kode Program <i>confusion_matrix.html</i>	79
Lampiran 20. Kode Program <i>style_1.css</i>	79
Lampiran 21. Kode Program <i>style_2.css</i>	80
Lampiran 22. Kode Program <i>style_3.css</i>	83
Lampiran 23. Kode Program <i>style_4.css</i>	84
Lampiran 24. Kode Program <i>app.py</i>	85
Lampiran 25. Kode Program <i>nbc.py</i>	87
Lampiran 26. Database <i>db_users</i>	89
Lampiran 27. Kriteria Kelulusan Ujian Sidang / Pendadaran	90
Lampiran 28. Keputusan Hasil Ujian Pendadaran	90
Lampiran 29. Catatan Perbaikan	91

INTISARI

Perguruan tinggi berperan penting dalam mencetak sumber daya manusia yang berkualitas, dan salah satu indikator kualitas pendidikan tinggi adalah tingkat kelulusan mahasiswa. Semakin banyak mahasiswa yang lulus tepat waktu, semakin baik kinerja perguruan tinggi tersebut.

Penelitian ini menggunakan data mahasiswa Program Studi Diploma III Teknologi Bank Darah di Poltekkes BSI Yogyakarta angkatan tahun 2020/2021, 2021/2022, dan 2022/2023. Atribut yang digunakan meliputi jenis kelamin, asal daerah, Indeks Prestasi Semester (IPS) 1 hingga 4, dan nilai mata kuliah kompetensi.

Hasil penelitian menunjukkan bahwa model *Naive Bayes Classifier* memiliki akurasi yang tinggi dalam memprediksi kelulusan mahasiswa, yaitu sebesar 94%. Berdasarkan hasil analisis, nilai F1-Score menunjukkan keseimbangan yang cukup baik antara presisi dan recall, yaitu sebesar 49%. Dengan demikian, model ini dapat digunakan secara efektif sebagai alat bantu dalam mengevaluasi dan meningkatkan kualitas pendidikan di Politeknik Kesehatan Bhakti Setya Indonesia (Poltekkes BSI) Yogyakarta.

Kata Kunci : Evaluasi, Kelulusan, *Naive Bayes*, Prediksi

ABSTRACT

Higher education plays a crucial role in producing high-quality human resources, and one indicator of the quality of higher education is the student graduation rate. The higher the number of students who graduate on time, the better the performance of the higher education institution.

This research used student data from the Diploma III Blood Bank Technology study program at Poltekkes BSI Yogyakarta for the 2020/2021, 2021/2022, and 2022/2023 academic years. The attributes used include gender, place of origin, Semester Grade Point Average (GPA) from 1 to 4, and competency course grades.

The research findings show that the Naive Bayes Classifier model has a high accuracy in predicting student graduation, which is 94%. Based on the analysis, the F1-Score indicates a fairly good balance between precision and recall, which is 49%. Thus, this model can be effectively used as a tool to evaluate and improve the quality of education at the Bhakti Setya Indonesia Health Polytechnic (Poltekkes BSI) Yogyakarta.

Keywords : *Evaluation, Graduation, Naive Bayes, Prediction*

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perguruan tinggi berperan sebagai penyelenggara pendidikan akademik bagi mahasiswa (Setiyani et al., 2020). Perguruan tinggi dituntut untuk menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga mampu menghasilkan sumber daya manusia yang berilmu, cakap, kreatif, dan berkompeten (Nathan & Scobell, 2012). Politeknik Kesehatan Bhakti Setya Indonesia merupakan salah satu lembaga pendidikan tinggi yang berperan penting dalam membentuk sumber daya manusia yang berkualitas, terutama dalam bidang kesehatan. Salah satu program studi yang tersedia di Politeknik Kesehatan Bhakti Setya Indonesia adalah Program Studi Diploma III Teknologi Bank Darah. Pendidikan Diploma III Teknologi Bank Darah bertujuan untuk menghasilkan tenaga kesehatan sebagai teknisi pelayanan darah profesional dengan sebutan Ahli Madya Kesehatan (Kemenkes RI, 2018).

Lulusan Program Studi Diploma III Teknologi Bank Darah diharapkan dapat berperan sebagai teknisi pelayanan darah, pelaksana pengujian mutu produk darah, serta edukator layanan darah. Artinya, mereka harus mampu berkomunikasi, memberikan informasi, dan edukasi terkait pelayanan darah (Riawati, 2020). Salah satu aspek pengukuran kualitas perguruan tinggi dalam keberhasilan penyelenggaraan pendidikan tinggi adalah tingkat kelulusan mahasiswa. Menurut Broto Legowo & Indiarso, semakin banyak mahasiswa yang lulus tepat waktu,

semakin baik pula kinerja perguruan tinggi tersebut. Tingkat kelulusan mahasiswa menjadi salah satu kriteria penilaian akreditasi bagi suatu perguruan tinggi atau program studi (Setiyani et al., 2020).

Sebagai bagian dari upaya meningkatkan mutu pendidikan, Politeknik Kesehatan Bhakti Setya Indonesia telah menetapkan standar kelulusan bagi mahasiswa, yaitu lulus tepat waktu dalam enam semester atau tiga tahun dengan nilai IPK (Indeks Prestasi Kumulatif) minimal 3,00. Namun, data dari Program Studi Teknologi Bank Darah dalam tiga tahun terakhir menunjukkan bahwa 11% mahasiswa tidak berhasil lulus sesuai standar tersebut. Apabila situasi ini terus berlanjut tanpa penanganan yang tepat, maka dapat berdampak pada penurunan kualitas lulusan serta akreditasi program studi.

Terpenuhinya standar kelulusan dan tingkat kelulusan yang tinggi untuk mahasiswa Program Studi Teknologi Bank Darah di Politeknik Kesehatan Bhakti Setya Indonesia sangatlah penting. Hal ini menunjukkan bahwa kualitas sumber daya manusia yang dihasilkan sesuai dengan standar industri kesehatan, terutama dalam bidang pelayanan darah. Selain itu, dapat meningkatkan reputasi institusi, menarik minat calon mahasiswa, mendukung pemenuhan kebutuhan tenaga kerja, dan memperkuat hubungan dengan industri atau instansi kesehatan. Ada berbagai faktor yang dapat mempengaruhi tingkat kelulusan mahasiswa atau tidak terpenuhinya standar kelulusan, antara lain faktor akademik, motivasi, dan faktor lain yang dapat mempengaruhi perilaku belajar mahasiswa seperti kondisi lingkungan atau tempat tinggal mahasiswa.

Oleh karena itu, diperlukan pendekatan yang inovatif untuk meningkatkan peluang kelulusan mahasiswa pada Program Studi Teknologi Bank Darah di Politeknik Kesehatan Bhakti Setya Indonesia agar dapat memenuhi atau bahkan melampaui standar kelulusan yang telah ditetapkan. Salah satu solusinya adalah mengembangkan sebuah aplikasi berbasis web yang memanfaatkan metode penggalian data (*data mining*), khususnya dengan menerapkan teknik klasifikasi seperti *Naive Bayes Classifier* (NBC). Metode *Naive Bayes* dipilih karena dapat mengklasifikasikan data dengan tipe yang beragam (kategorikal dan numerik). Sehingga tidak diperlukan transformasi data sebelumnya, serta tidak memerlukan dataset yang terlalu besar. Selain itu, *Naive Bayes* merupakan metode probabilitas statistik yang sederhana namun menghasilkan akurasi tinggi (Salmu & Solichin, 2017).

Penelitian sebelumnya oleh Nurul Khasanah, dkk (Khasanah et al., 2022) berjudul “Prediksi Kelulusan Mahasiswa Dengan Metode *Naive Bayes*” menggunakan 379 data (303 *data training*, 76 *data testing*) dengan atribut nama, status mahasiswa, status perkawinan, IPS, IPK, dan status kelulusan. Hasil menunjukkan akurasi 88,16%, *precision* 93,62%, dan *recall* 88% (*good classification*). Penelitian lain oleh Neni Purwati dan Agnes Dwi Januanti (Purwati & Januanti, 2021) berjudul “Prediksi Tingkat Kelulusan Mahasiswa Dengan Algoritma *Naive Bayes*” menggunakan 500 data (80% *training*, 20% *testing*) dengan atribut jurusan, biaya, jenis kelamin, beasiswa, jumlah dan nilai rata-rata mata kuliah semester 1 dan 2. Hasil penelitian menunjukkan akurasi 95%, presisi 95,16%, *recall* 95%, dan *F1-Score* 95%.

Penelitian ini memprediksi kelulusan mahasiswa Program Studi Teknologi Bank Darah di Politeknik Kesehatan Bhakti Setya Indonesia menggunakan variabel penting seperti jenis kelamin, asal daerah, Indeks Prestasi Semester (IPS) dari semester 1 sampai 4, dan mata kuliah kompetensi. Aplikasi yang dikembangkan dengan metode *Naive Bayes Classifier* diharapkan dapat memberikan prediksi yang akurat sebagai alat pengambilan keputusan terkait kelulusan. Tujuan penelitian ini adalah menganalisis prediksi kelulusan untuk mengembangkan strategi pembelajaran yang efektif dalam meningkatkan tingkat kelulusan mahasiswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas diperoleh rumusan masalah sebagai berikut :

1. Bagaimana mengembangkan sebuah aplikasi prediksi kelulusan mahasiswa Program Studi Diploma III Teknologi Bank Darah di Politeknik Kesehatan Bhakti Setya Indonesia berbasis web menggunakan metode penggalian data (*data mining*), khususnya dengan menerapkan teknik klasifikasi seperti *Naive Bayes Classifier* (NBC).
2. Bagaimana aplikasi tersebut dapat membantu meningkatkan kualitas pembelajaran dan meningkatkan tingkat kelulusan mahasiswa sesuai standar yang ditetapkan oleh Politeknik Kesehatan Bhakti Setya Indonesia, serta memberikan dukungan kepada mahasiswa yang memerlukan bantuan tambahan.

1.3 Ruang Lingkup

Ruang lingkup dari penelitian ini adalah :

1. Metode data mining yang akan diimplementasikan dalam pembuatan aplikasi prediksi kelulusan mahasiswa adalah metode klasifikasi *Naive Bayes Classifier* (NBC).
2. Data yang digunakan berasal dari 247 mahasiswa Program Studi DIII Teknologi Bank Darah di Politeknik Kesehatan Bhakti Setya Indonesia yang lulus dalam tiga tahun terakhir (2020/2021, 2021/2022 dan 2022/2023). Data tersebut dibagi menjadi data latih (*training data*) sebanyak 197 data (80%) dan data uji (*testing data*) sebanyak 50 data (20%).
3. Variabel prediktor yang akan digunakan untuk memprediksi kelulusan mahasiswa meliputi :
 - Jenis kelamin
 - Asal daerah
 - Indeks Prestasi Semester (IPS) dari semester 1 sampai 4.
 - Nilai Mata Kuliah Kompetensi, yaitu meliputi Serologi Golongan Darah I, Serologi Golongan Darah II, Serologi Golongan Darah III, Serologi Golongan Darah IV, Infeksi Menular Lewat Transfusi Darah I, Infeksi Menular Lewat Transfusi Darah II, Infeksi Menular Lewat Transfusi Darah III, Infeksi Menular Lewat Transfusi Darah IV, Penyadapan Darah, dan Komponen Darah.
4. Melakukan prediksi kelulusan mahasiswa berdasarkan variabel prediktor yang ditetapkan.

5. Variabel target atau kelas dibagi menjadi empat kategori, yaitu :
 - Label kelas = 1 (Jika mahasiswa lulus dalam 3 tahun dan $IPK \geq 3,00$)
 - Label kelas = 2 (Jika mahasiswa lulus dalam 3 tahun dan $IPK < 3,00$)
 - Label kelas = 3 (Jika mahasiswa lulus lebih dari 3 tahun dan $IPK \geq 3,00$)
 - Label kelas = 4 (Jika mahasiswa lulus lebih dari 3 tahun dan $IPK < 3,00$)
6. Performa dan akurasi model aplikasi prediksi akan diukur menggunakan metrik seperti akurasi, persisi, *recall*, atau *F1-score*.
7. Mengevaluasi efektivitas aplikasi dalam memprediksi kelulusan mahasiswa berdasarkan hasil yang diperoleh dari performa pengukuran model aplikasi untuk memastikan keakuratan dan kegunaannya.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Mengembangkan model aplikasi prediksi kelulusan mahasiswa Program Studi DIII Teknologi Bank Darah di Politeknik Kesehatan Bhakti Setya Indonesia berbasis web menggunakan metode *Naive Bayes Classifier* (NBC).
2. Mengukur performa dan akurasi model prediksi kelulusan menggunakan metrik seperti akurasi, persisi, *recall*, atau *F1-score*.
3. Mengevaluasi akurasi yang dihasilkan sistem dengan pengukuran akurasi secara manual menggunakan metrik.
4. Mengevaluasi efektivitas aplikasi prediksi dalam membantu pengambilan keputusan terkait strategi pembelajaran dan dukungan kepada mahasiswa.
5. Memberikan rekomendasi yang konkret dan terukur kepada Politeknik Kesehatan Bhakti Setya Indonesia untuk meningkatkan kualitas pembelajaran

dan tingkat kelulusan mahasiswa Program Studi Diploma III Teknologi Bank Darah, berdasarkan temuan dan hasil evaluasi dari penelitian ini.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

1. Bagi perguruan tinggi atau Program Studi :
 - Memberikan kontribusi dalam pembuatan kebijakan yang lebih efektif untuk meningkatkan tingkat kelulusan mahasiswa.
 - Meningkatkan mutu pendidikan di Program Studi Teknologi Bank Darah dengan menyediakan informasi prediksi kelulusan yang akurat.
2. Bagi mahasiswa :
 - Memberikan perkiraan yang lebih akurat terkait lama masa studi yang dibutuhkan.
 - Membantu pemahaman terhadap standar nilai kelulusan dan memberikan arahan yang lebih jelas dalam upaya mencapainya.

BAB 2

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Penelitian ini membahas tentang bagaimana hasil prediksi kelulusan mahasiswa dengan metode klasifikasi *Naive Bayes Classifier*. Berikut penelitian terdahulu yang pernah dilakukan dengan menerapkan metode klasifikasi *Naive Bayes Classifier* :

Penelitian yang dilakukan oleh Kemal Refta Diska dan Khairi Budayawan (Diska & Budayawan, 2023), dengan judul “Sistem Informasi Prediksi Kelulusan Menggunakan Metode *Naive Bayes* (Studi Kasus: Prodi Pendidikan Teknik Informatika)”. Data diuji menggunakan data mahasiswa PTI angkatan 2014 dan 2015 sebagai *data training* berjumlah 94 data, dan data mahasiswa angkatan 2016 sebagai *data testing* berjumlah 46 data. Atribut yang digunakan dalam penelitian ini adalah IPS semester 1 hingga 6, dan total sks. Hasil dari penelitian yang dilakukan menyatakan bahwa dari 46 *data testing* memperoleh hasil *accuracy* 82,61%, *precision* 91,66%, dan *recall* 61,11%.

Penelitian yang dilakukan oleh Firman Azhar Riyadi dan Donny Avianto (Riyadi & Avianto, 2020), dengan judul “Implementasi Metode *Naive Bayes* Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu Prodi Informatika (Studi Kasus Universitas Teknologi Yogyakarta)”. Penelitian ini menggunakan data mahasiswa angkatan 2014 sebanyak 200 data dengan atribut kelamin, sks1, sks2, sks3, sks4, ipk1, ipk2, ipk3, dan ipk4. Penelitian yang dilakukan yaitu dengan uji coba

perbandingan data uji sebanyak 60% dan data latih sebanyak 40% maka didapatkan akurasi 91,86%.

Penelitian yang dilakukan oleh Nurul Khasanah, dkk (Khasanah et al., 2022), dengan judul “Prediksi Kelulusan Mahasiswa Dengan Metode *Naive Bayes*”. Penelitian ini menggunakan sebanyak 379 data, dengan metode *Naive Bayes*, dengan rincian *data training* 303 data dan *data testing* 76 data. Atribut yang digunakan nama, status mahasiswa, status perkawinan, IPS, IPK, dan status kelulusan. Dengan tahapan identifikasi masalah, pengumpulan data, *data cleaning*, *data transformation* (dibagi menjadi *data training* dan *data tesing*), klasifikasi dengan NBC, validasi, evaluasi dan hasil. Hasil penelitian yang diperoleh yaitu akurasi = 88,16%, *precision* = 93,62% dan *recall* = 88%.

Penelitian yang dilakukan oleh Supardi Salmu dan Achmad Solichin (Salmu & Solichin, 2017), dengan judul “Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan *Naive Bayes* (Studi Kasus : UIN Syarif Hidayatullah Jakarta)”. Atribut yang digunakan pada penelitian ini yaitu meliputi jenis kelamin, jenis seleksi, pendapatan ayah, pendidikan ibu, IP semester 1 sampai 4, dan SKS semester 1 sampai 4. Penelitian ini menunjukkan hasil akurasi *Naive Bayes* sebesar 80,72%, presisi 28,13%, *recall* 28,57%, dan *F1-Score* 32,53%. Dengan rincian *data training* sebanyak 1662 data dan *data testing* sebanyak 587 data.

Penelitian yang dilakukan oleh M. Riski Qisthiano, dkk (Qisthiano et al., 2021), dengan judul “Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Dengan Metode *Naive Bayes*”. Atribut yang digunakan dalam penelitian ini adalah Jurusan, Perguruan Tinggi, Jenis Kelas, Nilai IP

Sementara dari semester 1 hingga 4, tahun lulus, dan angkatan kuliah. Selanjutnya pembagian dataset dibagi 70% untuk *data training* dan 30% sebagai *data testing*. Penelitian ini menguji proses algoritma *Naive Bayes* menggunakan K-Fold Validation. Hasil penelitian ini merupakan akurasi dari model prediksi yang dilakukan, dimana hasil akurasi yang didapatkan adalah 0.8103 atau sebesar 81,03%.

Penelitian yang dilakukan oleh Neni Purwati dan Agnes Dwi Januanti (Purwati & Januanti, 2021), dengan judul “Prediksi Tingkat Kelulusan Mahasiswa Dengan Algoritma *Naive Bayes*”. Atribut yang digunakan dalam penelitian ini yaitu meliputi jurusan, biaya, jenis kelamin, beasiswa, jumlah mata kuliah yang disetujui semester pertama, nilai rata-rata semester pertama, jumlah mata kuliah yang disetujui semester 2, dan nilai rata-rata semester 2. Menggunakan dataset sebanyak 500 data, kemudian dibagi menjadi 80% *data training* dan 20% *data testing*. Penelitian yang dilakukan mendapatkan hasil akurasi sebesar 95%, presisi 95,16%, *recall* 95%, dan *F1-Score* 95%.

Penelitian yang dilakukan oleh Muhammad Gunawan, Muhammad Zarlis, dan Roslina (Gunawan et al., 2021), dengan judul “Analisis Komparasi Algoritma *Naive Bayes* dan *k-Nearest Neighbor* Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu”. Data yang digunakan dalam penelitian ini adalah Data Alumni Mahasiswa Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara Tahun Angkatan 2015. Data mahasiswa yang digunakan adalah 100 data mahasiswa. Data penelitian yang diambil memiliki atribut *input* indeks prestasi semester (IPS) 1 hingga 5 dan atribut output ialah Kelulusan. Penelitian ini

menghasilkan kesimpulan bahwa algoritma *Naive Bayes* memiliki tingkat akurasi yang sama dengan algoritma KNN dalam memprediksi kelulusan mahasiswa program studi Pendidikan Kedokteran yaitu sebesar 90 %. Hasil ini diperoleh dari analisis yang dilakukan dengan aplikasi Weka terhadap 30 *data testing* dan 100 *data training*.

Penelitian yang dilakukan oleh Sidik Rahmatullah dan Ema Utami (Rahmatullah, 2019), dengan judul “Prediksi Tingkat Kelulusan Tepat Waktu Dengan Metode *Naive Bayes* Dan *K-Nearest Neighbor*”. Atribut-atribut data yang digunakan adalah NPM, jenis kelamin, IPS 1-5, konsentrasi, asal tinggal, jumlah sks, status pekerjaan, dan status kelulusan. Uji coba dilakukan dengan menggunakan data lulusan mahasiswa S1 Sistem informasi STMIK Dian Cipta Cendikia Kotabumi sebanyak 600 data untuk *training* dan 180 data untuk *testing*. Hasil uji coba menunjukkan bahwa dengan menggunakan *Naive Bayes* menghasilkan akurasi sebesar 85%, sedangkan menggunakan algoritma *K-nearest neighbor* menghasilkan akurasi sebesar 68,89 %.

2.2 Dasar Teori

2.2.1 Data Mining

Data mining merupakan proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar (Qisthiano et al., 2021). *Data mining* sering disebut dengan *Knowledge Discovery in Database* atau disingkat dengan KDD, yaitu merupakan proses pengumpulan, analisis, dan penggunaan data historis untuk menelusuri data yang ada guna membangun model yang dapat mengenali pola data lain yang berukuran besar. *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 (Widaningsih, 2019).

Secara garis besar, *data mining* adalah proses pencarian dan analisis pada suatu kumpulan data (*database*) untuk menemukan pola yang menarik dengan tujuan mengekstrak informasi dan pengetahuan yang akurat serta potensial, sehingga dapat dipahami dan berguna dalam pengambilan keputusan (Setiyani et al., 2020). Menurut (Mustafa et al., 2018) tahap – tahap *data mining* adalah sebagai berikut :

1. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data yang tidak relevan.

2. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.

3. Seleksi data (*data selection*)

Seleksi data merupakan proses penyeleksian data. Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

4. Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.

5. Proses *mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (*pattren evaluation*)

Untuk mengidentifikasi pola – pola menarik ke dalam *knowledge based* yang ditemukan.

7. Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.2.2 Praproses Data

Praproses data melibatkan beberapa langkah penting untuk memastikan data yang digunakan dalam pengembangan model prediksi adalah data yang bersih, lengkap, dan sesuai. Praproses merupakan sebuah tahap awal yang harus dilakukan pada *data mining* untuk mempersiapkan data mentah sebelum dilakukan proses lain. Praproses data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem. Tujuan praproses adalah untuk mendapatkan hasil yang lebih akurat, pengurangan waktu perhitungan untuk *large scale problem*, dan membuat nilai data menjadi lebih kecil tanpa merubah informasi didalamnya (Nasution et al., 2019).

2.2.3 Pembersihan Data

Pembersihan data (*Data Cleaning*) adalah tahap awal dalam proses *Knowledge Discovery in Databases* (KDD) yang melibatkan penghapusan atau perbaikan data yang tidak lengkap, salah, atau tidak konsisten. *Data cleaning* adalah

proses mendeteksi, memperbaiki atau bahkan menghapus catatan, tabel, dan database yang salah atau tidak akurat (Baiq Nurul Azmi et al., 2023). Pada tahap ini, akan dilakukan proses penghapusan data yang tidak relevan, duplikat, atau yang memiliki banyak nilai kosong. Atribut-atribut yang nilainya kosong seluruhnya atau sebagian besar kosong akan dihilangkan untuk mendapatkan atribut yang relevan dan valid. Tujuannya adalah untuk memastikan bahwa data yang digunakan adalah data yang valid, dapat diandalkan, dan bebas dari *noise* seperti *missing value*, inkonsistensi, dan redundansi.

Proses penggantian *missing value* dilakukan dengan mengganti data yang hilang atau tidak sesuai dengan atribut yang ada. Menurut Amien dkk, *missing value* dapat diganti dengan nilai *median* atau *mean* dari masing-masing atribut, dengan tujuan untuk mengoptimalkan hasil yang didapatkan (Amien et al., 2023). Sedangkan inkonsistensi data terjadi ketika ada data yang redundan. Data redundan adalah menumpuknya data-data yang sama yang tidak dibutuhkan di dalam database. Salah satu cara untuk menyelesaikan masalah inkonsistensi data adalah dengan melakukan eliminasi atau menghapus beberapa data yang tidak konsisten pada database (Sanjaya & Sulistyono, 2015).

2.2.4 Pembagian Data

Pembagian data adalah proses membagi dataset menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk mengevaluasi performa model. Pada penelitian ini akan dilakukan pembagian data dengan proporsi 80% untuk data latih dan 20% untuk data uji. Pembagian data ini bertujuan untuk memastikan

bahwa model dapat memprediksi dengan akurat pada data baru atau data yang belum pernah dilihat sebelumnya. Pembagian jumlah data latih dan data uji adalah salah satu faktor yang menentukan akurasi, sehingga kesalahan dalam menentukan komposisi kedua tipe data tersebut akan mempengaruhi nilai akurasi dan presisi yang diperoleh (Baiq Nurul Azmi et al., 2023).

2.2.5 Klasifikasi

Klasifikasi merupakan proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Haditsah, 2018). Klasifikasi banyak digunakan untuk memprediksi kelas pada label tertentu, yaitu dengan mengklasifikasikan data atau membangun model berdasarkan *training set* dan nilai – nilai atau label kelas yang digunakan dalam mengklasifikasikan atribut tertentu (Tangkelayuk, 2022).

Dalam proses klasifikasi, terjadi identifikasi kelompok dari suatu obyek berdasarkan kesamaan fitur tertentu, dimana setiap kelompok telah terbentuk melalui suatu proses tertentu. Proses klasifikasi biasanya dibagi menjadi dua fase yaitu fase *learning* dan fase *test*. Pada fase *learning* sebagian data yang telah diketahui kelas datanya diumpangkan untuk membentuk model perkiraan. Kemudian pada fase *test* model yang sudah dibentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi model. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi.

2.2.6 Naive Bayes

Naive Bayes Classifier yaitu salah satu metode *machine learning* yang menggunakan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memperkirakan probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes (Gunawan et al., 2021). Teorema tersebut dikombinasikan dengan *naive* yang mana diasumsikan keadaan antar atribut saling bebas. Jadi didalam penerapannya, algoritma *Naive Bayes Classifier* tidak ada hubungan antara satu atribut dengan atribut yang lain, atau dengan kata lain satu atribut tidak berpengaruh dengan atribut yang lain, sekalipun mungkin atribut tersebut saling berhubungan (Sigid Widodo et al., 2023).

Teorema bayes merupakan dasar dari *naive bayes classifier*. *Naive Bayes Classifier* berfungsi menghitung dan mencari nilai probabilitas paling tinggi untuk mengklasifikasikan sebuah data uji dengan kategori yang tepat. Teknik prediksi probabilitas yang sederhana didasarkan pada penerapan teorema bayes yang secara umum dinyatakan sebagai berikut :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.1)$$

Keterangan :

X = Data dengan kelas yang belum diketahui

H = Hipotesis data X merupakan suatu label kelas tertentu

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

$P(X/H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H (Probabilitas *Likelihood*)

$P(H)$ = Probabilitas hipotesis H (Probabilitas Prior)

$P(X)$ = Probabilitas X (Probabilitas *Evidence*)

Berikut tahapan dari algoritma *Naive Bayes Classifier* :

1. Mengitung probabilitas prior ($P(H)$), yaitu menentukan probabilitas prior dari setiap kelas H.
2. Menghitung probabilitas *likelihood* ($P(X|H)$), yaitu menghitung probabilitas bahwa data X muncul dalam setiap kelas H.
3. Menghitung probabilitas *evidence* ($P(X)$), yaitu menghitung probabilitas dari data yang diamati secara keseluruhan, tanpa memperhatikan kelasnya.
4. Menghitung probabilitas posterior ($P(H|X)$), yaitu dengan cara menggunakan rumus diatas untuk menghitung probabilitas bahwa hipotesis H terjadi berdasarkan data X yang diamati.
5. Prediksi kelas, yaitu menentukan kelas dengan probabilitas posterior tertinggi.

2.2.7 *Confusion Matrix*

Confusion matrix merupakan alat yang sangat penting dalam evaluasi kinerja model klasifikasi dalam *machine learning*. Matriks ini menyajikan jumlah prediksi yang benar dan salah yang dibuat oleh model dibandingkan dengan hasil sebenarnya. *Confusion matrix* memberikan gambaran yang lebih detail tentang kesalahan klasifikasi yang dilakukan oleh model. *Confusion matrix* merupakan matriks yang menampilkan prediksi klasifikasi dan klasifikasi yang aktual, serta

digunakan untuk memperoleh nilai *precision*, *recall*, dan *accuracy* (Rahayu et al., 2021). Berikut *confusion matrix* dengan dua label kelas berbentuk tabel 2x2 yang memberikan informasi tentang *true positive*, *false positive*, *true negative*, dan *false negative*.

Tabel 2. 1. Struktur *Confusion Matrix 2x2*

Prediksi Aktual	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Dimana :

- *True Positive* (TP) : Jumlah prediksi positif yang benar (model memprediksi positif dan sebenarnya positif).
- *True Negative* (TN) : Jumlah prediksi negatif yang benar (model memprediksi negatif dan sebenarnya negatif).
- *False Positive* (FP) : Jumlah prediksi positif yang salah (model memprediksi positif tetapi sebenarnya negatif).
- *False Negative* (FN) : Jumlah prediksi negatif yang salah (model memprediksi negatif tetapi sebenarnya positif).

Berikut adalah rumus-rumus untuk mengukur performa model klasifikasi menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score* :

1. Akurasi (*accuracy*), yaitu untuk mengukur seberapa sering model klasifikasi membuat prediksi yang benar dari semua prediksi yang dibuat.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.2)$$

2. Presisi (*precision*), yaitu untuk mengukur seberapa banyak prediksi positif yang benar dari semua prediksi positif yang dibuat oleh model.

$$Precision = \frac{TP}{TP + FP} \quad (2.3)$$

3. *Recall*, yaitu untuk mengukur seberapa banyak kasus yang sebenarnya positif yang benar diprediksi oleh model.

$$Recall = \frac{TP}{TP + FN} \quad (2.4)$$

4. *F1-Score*, yaitu kombinasi dari presisi dan *recall* untuk memberikan keseimbangan antara kedua metrik.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.5)$$

Jika penelitian ini menggunakan 4 label kelas, maka *confusion matrix* akan menjadi matriks 4x4. Setiap baris mewakili jumlah *instance* kelas sebenarnya, dan setiap kolom mewakili jumlah *instance* kelas yang diprediksi. Bentuk tabel *confusion matrix* 4x4 adalah sebagai berikut:

Tabel 2. 2. Struktur *Confusion Matrix* 4x4

Prediksi Aktual	A	B	C	D
A	TP	FN	FN	FN
B	FP	TP	FN	FN
C	FP	FP	TP	FN
D	FP	FP	FP	TP

2.2.8 *Use Case Diagram*

Use case diagram atau diagram *use case* adalah diagram untuk memodelkan perilaku suatu sistem yang akan dirancang dengan menggambarkan interaksi antara satu atau lebih aktor yang akan menggunakan sistem. *Use case diagram* terdiri dari sebuah aktor dan interaksi yang dilakukannya, aktor tersebut dapat berupa manusia, perangkat keras, sistem lain, ataupun yang berinteraksi dengan sistem (Kurniawan & Syarifuddin, 2020). Antara *use case* dengan aktor atau dengan *use case* terdapat beberapa *links* hubungan *include*, *extend*, *generalization*, dan lain-lain (Setiyani, 2021). *Use Case Diagram* adalah salah satu diagram yang digunakan dalam perancangan sistem untuk menggambarkan hubungan interaksi antara pengguna (aktor) dengan aplikasi yang sedang dirancang (sistem). Diagram ini dapat mendeskripsikan jenis-jenis interaksi yang terjadi antara pengguna dan sistem, serta mendokumentasikan kebutuhan fungsional yang harus dipenuhi oleh sistem. Dengan *use case diagram*, maka dapat memvisualisasikan dan memahami alur kerja aplikasi yang dibuat dari sudut pandang pengguna.

2.2.9 *Sequence Diagram*

Diagram sekuen (*Sequence Diagram*) mendeskripsikan perilaku objek pada *use case* dengan menjelaskan alur waktu hidup dari objek dan pesan atau *message* yang diterima dan dikirim antar objek. *Sequence Diagram* adalah diagram yang digunakan untuk menggambarkan urutan interaksi antar objek dalam sistem berdasarkan urutan waktu. Diagram ini menunjukkan bagaimana objek-objek saling berinteraksi melalui pertukaran pesan (*message*) dalam suatu skenario. *Sequence diagram* memiliki dimensi vertikal yang mewakili urutan waktu, dan dimensi

horizontal yang mewakili objek-objek yang terlibat dalam interaksi tersebut. Dengan *sequence diagram*, maka dapat memvisualisasikan dan memahami alur proses bisnis serta kolaborasi antar komponen sistem secara lebih rinci.

2.2.10 Website

Website adalah kumpulan halaman web yang saling terkait dan dapat diakses melalui Internet. Halaman-halaman tersebut dapat berisi teks, gambar, video, dan elemen interaktif lainnya yang disajikan menggunakan bahasa markup seperti HTML (*HyperText Markup Language*). Struktur dasar sebuah *website* terdiri dari berbagai elemen, seperti halaman beranda (*homepage*), halaman-halaman konten, *navigasi* (menu), *footer*, dan lain-lain. Halaman-halaman ini saling terhubung dengan tautan atau *hyperlink*.

Menurut Tangkudung dkk, *website* merupakan kumpulan halaman web yang berhubungan antara satu dengan lainnya, halaman pertama sebuah *website* adalah *home page*, sedangkan halaman demi halamannya secara mandiri disebut *web page*, dengan kata lain *website* adalah situs yang dapat diakses dan dilihat oleh para pengguna internet di seluruh dunia (Tangkudung et al., 2019). Aplikasi *website* dibagi menjadi dua jenis, yaitu *website* statis dan *website* dinamis. *Website* statis merupakan *website* yang isinya jarang berubah atau tetap, serta memiliki sifat satu arah dan tidak interaktif. Sedangkan *website* dinamis merupakan *website* yang isinya sering berubah, serta memiliki sifat dua arah dan interaktif.

BAB III

METODE PENELITIAN

3.1 Analisis Kebutuhan

Analisis kebutuhan merupakan tahap awal di mana dilakukan pengumpulan data dan informasi yang akan digunakan sebagai dasar penelitian. Penelitian ini menggunakan metode kuantitatif dengan jenis eksperimental, yaitu menguji kebenaran hipotesis melalui analisis statistik dengan melibatkan investigasi dan manipulasi variabel-variabel terkait dengan masalah penelitian. Metode eksperimental adalah pendekatan yang melibatkan pengujian langsung terhadap alat atau intervensi yang dirancang dan diimplementasikan dalam penelitian.

Analisis kebutuhan ini mencakup kebutuhan perangkat keras (*hardware*), kebutuhan perangkat lunak (*software*), kebutuhan masukan (*input*), kebutuhan proses, dan kebutuhan keluaran (*output*) sistem.

3.1.1 Kebutuhan Perangkat Keras

Perangkat keras (*hardware*) yang digunakan untuk membangun aplikasi prediksi kelulusan mahasiswa adalah laptop dengan spesifikasi berikut :

1. Processor Intel Core i3
2. Harddisk 256 GB
3. Memori RAM 4 GB
4. Mouse logitech

3.1.2 Kebutuhan Perangkat Lunak

Perangkat lunak atau *software* merupakan program yang ditanamkan dalam sebuah perangkat keras (*Hardware*). Berikut adalah perangkat lunak (*software*) dan bahasa pemrograman yang digunakan dalam proses pembangunan sistem dalam penelitian ini :

1. Sistem operasi : Windows 11 home 64 bit
2. *Text editor* (Visual Studio Code)
3. Bahasa pemrograman Python
4. *HyperText Mark up Language* (HTML)
5. *Cascading Style Sheet* (CSS)
6. *Web framework* flask
7. *Database* MySQL
8. *XAMPP Server*

3.1.3 Kebutuhan Masukan (*Input*)

Kebutuhan masukan mencakup semua data dan informasi yang diperlukan sistem untuk melakukan prediksi kelulusan mahasiswa. Berikut adalah data masukan yang digunakan dalam sistem prediksi kelulusan mahasiswa Program Studi Teknologi Bank Darah di Politeknik Kesehatan Bhakti Setya Indonesia :

1. Data Demografis

Data demografis adalah informasi yang menggambarkan karakteristik suatu populasi. Data demografis yang digunakan dalam penelitian ini meliputi atribut yang dapat menggambarkan latar belakang mahasiswa.

Tabel 3. 1. Data Demografis Mahasiswa

No	Nama Data	Penjelasan
1.	Jenis Kelamin	Merupakan data jenis kelamin mahasiswa (laki-laki/perempuan).
2.	Asal Daerah	Merupakan informasi tentang asal daerah atau tempat tinggal mahasiswa (DIY/luar DIY).

2. Data Akademik

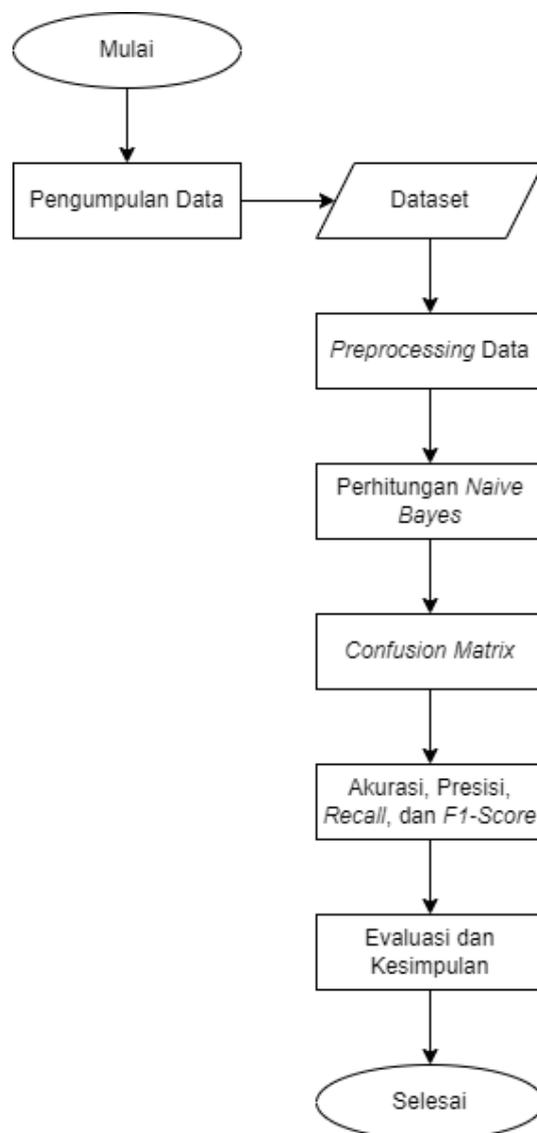
Data akademik adalah informasi yang berkaitan dengan prestasi dan kegiatan akademik mahasiswa. Data ini penting dalam proses analisis dan prediksi kelulusan karena mencerminkan kinerja akademik mahasiswa selama masa studinya.

Tabel 3. 2. Data Akademik Mahasiswa

No	Nama Data	Penjelasan
1.	Nilai Indeks Prestasi Semester (IPS)	Merupakan data IP dari setiap semester mahasiswa (IPK Semester 1 sampai 4).
2.	Nilai Mata Kuliah Kompetensi	Merupakan nilai dari setiap mata kuliah kompetensi yang wajib diambil oleh mahasiswa dan wajib lulus nilainya (Serologi Golongan Darah I, Serologi Golongan Darah II, Serologi Golongan Darah III, Serologi Golongan Darah IV, Infeksi Menular Lewat Transfusi Darah I, Infeksi Menular Lewat Transfusi Darah II, Infeksi Menular Lewat Transfusi Darah III, Infeksi Menular Lewat Transfusi Darah IV, Penyesapan Darah, dan Komponen Darah).

3.1.4 Kebutuhan Proses

Penelitian ini menggunakan metode *Naive Bayes Classifier* untuk memprediksi kelulusan mahasiswa pada Program Studi Diploma III Teknologi Bank Darah di Poltekkes BSI Yogyakarta. Secara garis besar, alur penelitian ini dapat dibagi menjadi beberapa tahap, yang digambarkan dalam diagram alur proses penelitian.



Gambar 3. 1. Diagram Alur Proses Penelitian

1. Pengumpulan Data

a. Studi Literatur

Dalam tahap studi literatur, penulis mengumpulkan data dan informasi yang relevan dengan penelitian, meliputi kajian teoritis tentang metode *Naive Bayes Classifier*, penelitian sejenis, serta mempelajari metode yang akan digunakan.

b. Data Sekunder

Sumber data dalam penelitian ini adalah data sekunder yang diperoleh dari basis data akademik Politeknik Kesehatan Bhakti Setya Indonesia. Penulis mengumpulkan data mahasiswa Program Studi Diploma III Teknologi Bank Darah yang lulus dalam tiga tahun terakhir, yaitu tahun akademik 2020/2021, 2021/2022, dan 2022/2023.

A	B	C	D	E	F	G	H	I	J
No	NIM	Nama	Program Studi	Status	IPS	Jumlah sks	IPK	Jumlah sks	Biaya Kuliah
1	18114027	FENDI	D3 Teknologi Bank Darah	Aktif	3.18	22	3.18	22	NaN
2	18114059	NIA AURALIA FIKRIYAH	D3 Teknologi Bank Darah	Aktif	3.86	22	3.86	22	NaN
3	18114033	IKA NOVI SAVITRI	D3 Teknologi Bank Darah	Aktif	3.73	22	3.73	22	NaN
4	18114075	SHENDY SEPTIYANI PUTRI	D3 Teknologi Bank Darah	Aktif	3.5	22	3.50	22	NaN
5	18114048	MELINDA PUDYA WIDURI	D3 Teknologi Bank Darah	Aktif	3.64	22	3.64	22	NaN
6	18114045	MARIO GRENFRIIS NDABUKE	D3 Teknologi Bank Darah	Aktif	2.45	22	2.45	22	NaN
7	18114028	FRISILIA GRYAS TRIES YULIANTI	D3 Teknologi Bank Darah	Aktif	3.32	22	3.32	22	NaN
8	18114081	SUNARIA	D3 Teknologi Bank Darah	Aktif	3.05	22	3.05	22	NaN
9	18114008	ALBAH INAYAH KOROMPOT	D3 Teknologi Bank Darah	Aktif	2.64	22	2.64	22	NaN
10	18114001	Abdul Hakim	D3 Teknologi Bank Darah	Aktif	3.41	22	3.41	22	NaN
11	18114085	WIDYA AYU LESTARI	D3 Teknologi Bank Darah	Aktif	3.73	22	3.73	22	NaN
12	18114043	MARIA NOVIATI ROMAN	D3 Teknologi Bank Darah	Aktif	3.77	22	3.77	22	NaN
13	18114077	SINTIA SARI	D3 Teknologi Bank Darah	Aktif	3.5	22	3.50	22	NaN

Gambar 3. 2. Data Mentah Sebelum Diolah

Setelah memperoleh data mentah, langkah selanjutnya adalah mengolah dan menggabungkan data tersebut, menghasilkan 258 data mahasiswa. Namun, data ini masih perlu dibersihkan, sehingga selanjutnya akan dilakukan praproses data.

No	NIM	Nama	Program Studi	JK	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPK LULUS	
245	244	20114075	NUR ZAHRA AFIFAH	D3 Teknologi Bank Darah	Wanita	3.45	3.24	3.37	3.33	2.97	3.58	3.11
246	245	20114076	SULJSTIAWATI ABDULLAH	D3 Teknologi Bank Darah	Wanita	3.18	2.72	3.33	3.04	2.74	3.83	3.12
247	246	20114077	SRIRAHMA DUMAI	D3 Teknologi Bank Darah	Wanita	3.06	2.85	3.38	2.98	3.35	3.92	3.52
248	247	20114078	MUFIDA RAHMAH	D3 Teknologi Bank Darah	Wanita	2.86	2.86	3.70	3.34	3.13	3.25	3.17
249	248	20114079	FATNI SYAHGITA	D3 Teknologi Bank Darah	Wanita	3.44	3.44	3.40	3.20	3.29	3.33	3.33
250	249	20114080	LUH RIRIN AULIANA W.	D3 Teknologi Bank Darah	Wanita	3.20	2.51	3.45	3.37	3.17	3.92	3.38
251	250	20114081	NOVIANA	D3 Teknologi Bank Darah	Wanita	3.22	3.09	3.43	3.10	0.00	0	0.00
252	251	20114082	ANDI NHITA PATURUNGI	D3 Teknologi Bank Darah	Wanita	3.47	3.33	0.00	0.00	3.10	3.92	3.29
253	252	20114083	ACHMAD SAIF	D3 Teknologi Bank Darah	Pria	3.07	1.23	3.38	3.29	0.42	0.00	0.00
254	253	20114084	NUR RAHYANI RASYID	D3 Teknologi Bank Darah	Wanita	3.17	3.05	1.84	0.99	3.42	3.62	0.00
255	254	20114085	SEPTIAN YUDIYANSYAH OLLI	D3 Teknologi Bank Darah	Pria	2.86	0.64	2.93	3.28	2.81	3.67	3.06
256	255	20114086	WA SARIANI	D3 Teknologi Bank Darah	Wanita	2.60	2.69	2.98	3.12	2.90	3.42	3.06
257	256	20114087	HARNIYANTI ODE	D3 Teknologi Bank Darah	Wanita	3.00	2.56	3.34	2.94	2.85	3.75	3.11
258	257	20114088	SUHARJO	D3 Teknologi Bank Darah	Pria	2.68	2.80	3.40	3.04	2.99	3.75	3.10
259	258	20114089	ISKANDAR ALI TUASIKAL	D3 Teknologi Bank Darah	Pria	2.74	2.84	3.20	3.04	2.90	3.01	0.00

Gambar 3. 3. Data Setelah Digabungkan

2. Prapemrosesan Data

a. Pembersihan Data

Tahap pembersihan data meliputi proses penghapusan atau perbaikan data yang tidak valid, duplikat, atau tidak lengkap. Juga dilakukan penanganan data kosong (*missing values*) dengan mengisi nilainya menggunakan median dari masing-masing atribut, serta menghapus data yang banyak memiliki nilai kosong.

NO	NIM	NAMA	PROGRAM STUDI	JENIS KELAMIN	ASAL DAERAH	IPS_1	IPS_2	IPS_3	IPS_4	IPS_5
242	241	20114083	ACHMAD SAIF	Pria	LUAR DIY	2.86	0.64	1.84	0.99	0.00
243	242	20114084	NUR RAHYANI RASYID	Wanita	LUAR DIY	2.60	2.69	2.93	3.28	3.42
244	243	20114085	SEPTIAN YUDIYANSYAH OLLI	Pria	LUAR DIY	3.00	2.56	2.98	3.12	2.90
245	244	20114086	WA SARIANI	Wanita	LUAR DIY	2.68	2.80	3.34	2.94	2.90
246	245	20114087	HARNIYANTI ODE	Wanita	LUAR DIY	2.74	2.84	3.40	3.04	2.90
247	246	20114088	SUHARJO	Pria	LUAR DIY	2.47	2.55	3.20	3.04	2.90
248	247	20114089	ISKANDAR ALI TUASIKAL	Pria	LUAR DIY	2.55	0.36	1.26	2.83	2.90

Gambar 3. 4. Data Bersih

b. Seleksi Data

Selanjutnya, tahap seleksi data atau seleksi fitur (*feature selection*) dilakukan untuk memilih variabel-variabel yang paling berpengaruh dalam proses prediksi. Pada tahap ini, atribut yang tidak digunakan akan dihapus, seperti NO, NIM, NAMA, PROGRAM STUDI, IPS_5, IPS_6, IPK LULUS, dan NILAI KARYA TULIS.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	JENIS KELAMIN	ASAL DAERAH	IPS_1	IPS_2	IPS_3	IPS_4	SEROLOGI GOLONGAN DARAH I	INFEKSI MENULAR LEWAT TRANSFUSI DARAH I	SEROLOGI GOLONGAN DARAH II	INFEKSI MENULAR LEWAT TRANSFUSI DARAH II	PENYADAPAN DARAH	SEROLOGI GOLONGAN DARAH III	INFEKSI MENULAR LEWAT TRANSFUSI DARAH III	KOMPONEN DARAH	SEROLOGI GOLONGAN DARAH IV	INFEKSI MENULAR LEWAT TRANSFUSI DARAH IV	LABEL KELAS
1																	
2	Wanita	DI YOGYAKARTA	3.50	2.93	3.43	2.98	C	B	B	A	A	B	B	B	C	A	3
3	Pria	LUAR DIY	3.30	3.50	2.94	2.86	B	C	B	B	A	A	B	B	B	B	3
4	Pria	LUAR DIY	2.09	0.65	1.53	2.98	B	B	B	B	A	B	B	B	B	A	1
5	Wanita	LUAR DIY	1.32	0.91	1.95	1.96	B	B	B	B	A	B	B	B	B	A	1
6	Wanita	LUAR DIY	3.32	2.76	2.42	2.00	C	C	C	B	B	C	B	B	C	A	4
7	Pria	LUAR DIY	3.41	3.09	3.38	1.58	C	B	B	B	A	B	C	B	B	A	1
8	Wanita	LUAR DIY	2.86	2.86	2.88	1.60	C	B	B	B	B	C	C	B	B	B	2
9	Wanita	LUAR DIY	3.36	3.00	3.81	2.16	B	B	A	B	A	B	A	B	A	A	1

Gambar 3. 5. Data Siap Pakai

c. Pembagian Data

Tahap terakhir adalah pembagian data, yang dilakukan dengan membagi data menjadi data latih (*training data*) sebanyak 197 data (80%) dan data uji (*testing data*) sebanyak 50 data (20%). Pada data uji, label kelas akan disembunyikan sehingga model dapat memprediksi berdasarkan atribut-atribut yang tersedia.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	JENIS KELAMIN	ASAL DAERAH	IPS_1	IPS_2	IPS_3	IPS_4	SEROLOGI GOLONGAN DARAH I	INFEKSI MENULAR LEWAT TRANSFUSI DARAH I	SEROLOGI GOLONGAN DARAH II	INFEKSI MENULAR LEWAT TRANSFUSI DARAH II	PENYADAPAN DARAH	SEROLOGI GOLONGAN DARAH III	INFEKSI MENULAR LEWAT TRANSFUSI DARAH III	KOMPONEN DARAH	SEROLOGI GOLONGAN DARAH IV	INFEKSI MENULAR LEWAT TRANSFUSI DARAH IV	LABEL KELAS
192	Pria	LUAR DIY	3.53	3.26	3.45	3.22	C	B	B	B	A	B	B	B	B	B	1
193	Wanita	LUAR DIY	3.86	3.50	3.67	2.53	B	C	B	A	B	B	A	A	A	A	1
194	Wanita	LUAR DIY	3.91	3.00	4.00	3.84	B	A	A	A	A	A	A	A	B	A	1
195	Wanita	LUAR DIY	3.45	3.14	3.52	1.80	C	B	B	B	B	C	B	B	B	A	1
196	Pria	LUAR DIY	2.73	2.18	3.71	2.86	C	C	B	A	A	B	B	B	B	A	1
197	Wanita	LUAR DIY	3.36	2.97	3.46	3.14	B	B	B	B	A	B	B	B	C	B	1
198	Wanita	LUAR DIY	3.45	2.73	3.71	3.63	B	B	B	A	A	B	A	A	B	A	1

Gambar 3. 6. Data Training

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	JENIS KELAMIN	ASAL DAERAH	IPS_1	IPS_2	IPS_3	IPS_4	SEROLOGI GOLONGAN DARAH I	INFEKSI MENULAR LEWAT TRANSFUSI DARAH I	SEROLOGI GOLONGAN DARAH II	INFEKSI MENULAR LEWAT TRANSFUSI DARAH II	PENYADAPAN DARAH	SEROLOGI GOLONGAN DARAH III	INFEKSI MENULAR LEWAT TRANSFUSI DARAH III	KOMPONEN DARAH	SEROLOGI GOLONGAN DARAH IV	INFEKSI MENULAR LEWAT TRANSFUSI DARAH IV	LABEL KELAS
45	Wanita	LUAR DIY	3.66	3.52	3.56	0.79	B	B	B	B	A	B	B	B	B	A	
46	Pria	LUAR DIY	3.00	3.00	3.04	1.86	B	C	C	B	B	B	C	C	B	A	
47	Wanita	LUAR DIY	3.86	3.09	3.90	3.63	B	B	B	A	A	B	A	A	A	A	
48	Wanita	LUAR DIY	3.82	3.55	3.52	2.32	A	B	A	B	B	B	B	B	A	A	
49	Wanita	LUAR DIY	3.24	2.69	3.40	2.77	B	C	B	B	A	B	B	C	C	B	
50	Wanita	LUAR DIY	1.77	2.00	1.79	2.09	B	C	C	C	C	C	B	A	B	C	
51	Pria	LUAR DIY	2.47	2.55	3.20	3.04	B	B	B	B	A	B	B	B	B	C	

Gambar 3. 7. Data Testing

3. Perhitungan Naive Bayes Classifier

Pada tahap ini, dilakukan perhitungan dengan menggunakan metode *Naive Bayes Classifier* untuk memprediksi kelulusan mahasiswa berdasarkan data yang telah dikumpulkan dan diproses.

a. Menghitung Probabilitas Prior (P(H))

Probabilitas prior dihitung dari proporsi kemunculan kelas pada data latih, dengan membagi jumlah data per kelas dengan total data.

Tabel 3. 3. Jumlah Label Kelas Pada *Data Training*

Label Kelas	Jumlah
1	176
2	11
3	4
4	6
Total	197

Probabilitas prior untuk masing-masing kelas adalah sebagai berikut :

$$\text{➤ } P(\text{Label Kelas}="1") = \frac{176}{197} = 0,893$$

$$\text{➤ } P(\text{Label Kelas}="2") = \frac{11}{197} = 0,056$$

$$\text{➤ } P(\text{Label Kelas}="3") = \frac{4}{197} = 0,02$$

$$\text{➤ } P(\text{Label Kelas}="4") = \frac{6}{197} = 0,031$$

b. Menghitung *Likelihood* ($P(X | H)$)

Likelihood adalah probabilitas bersyarat yang menunjukkan peluang nilai atribut (X) pada setiap kelas target (H). *Likelihood* dihitung untuk memprediksi kelas target berdasarkan nilai atribut. Hasilnya dapat dilihat pada Tabel 3.8 yang menampilkan Probabilitas Atribut Jenis Kelamin Terhadap Setiap Label Kelas.

Tabel 3. 4. Jumlah Atribut Pada Setiap Kelas Dalam *Data Training*

Atribut		Label Kelas				Total
		1	2	3	4	
Jenis Kelamin	Laki-laki	41	7	2	2	52
	Perempuan	135	4	2	4	145
Asal Daerah	DIY	14	0	1	2	17
	Luar DIY	164	11	3	5	180
IPS 1	≥ 3.00	142	1	4	1	148
	< 3.00	34	10	0	5	49
IPS 2	≥ 3.00	102	0	1	0	103
	< 3.00	74	11	3	6	94
IPS 3	≥ 3.00	153	6	2	0	161
	< 3.00	23	5	2	6	36
IPS 4	≥ 3.00	96	1	1	0	98
	< 3.00	80	10	3	6	99
Serologi Golongan Darah I	A	22	1	0	0	23
	B	127	2	1	0	130
	C	27	8	3	6	44
	D	0	0	0	0	0
Infeksi Menular Lewat Transfusi Darah I	A	35	0	0	0	35
	B	124	4	2	2	132
	C	17	7	2	4	30
	D	0	0	0	0	0
Serologi Golongan Darah II	A	26	0	0	0	26
	B	134	6	3	2	145
	C	16	5	1	4	26
	D	0	0	0	0	0
Infeksi Menular Lewat Transfusi Darah II	A	31	0	1	0	32
	B	129	2	2	5	138
	C	16	9	1	1	27
	D	0	0	0	0	0
Penyadapan Darah	A	132	5	3	0	140
	B	41	5	0	2	48
	C	3	1	1	4	9
	D	0	0	0	0	0
Serologi Golongan Darah III	A	37	1	1	0	39
	B	116	6	2	2	126
	C	23	4	1	4	32
	D	0	0	0	0	0

Atribut		Label Kelas				Total
		1	2	3	4	
Infeksi Menular Lewat Transfusi Darah III	A	49	0	0	0	49
	B	102	4	4	2	112
	C	25	7	0	4	36
	D	0	0	0	0	0
Komponen Darah	A	57	1	1	0	59
	B	102	6	2	1	111
	C	17	4	1	5	27
	D	0	0	0	0	0
Serologi Golongan Darah IV	A	42	0	0	0	42
	B	93	6	3	0	102
	C	41	5	1	6	53
	D	0	0	0	0	0
Infeksi Menular Lewat Transfusi Darah IV	A	123	4	3	2	132
	B	45	5	1	3	54
	C	8	2	0	1	11
	D	0	0	0	0	0

Tabel 3. 5. Probabilitas Atribut Jenis Kelamin Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Laki-laki Kelas 1)	$\frac{41}{176} = 0,233$
P(Laki-laki Kelas 2)	$\frac{7}{11} = 0,636$
P(Laki-laki Kelas 3)	$\frac{2}{4} = 0,5$
P(Laki-laki Kelas 4)	$\frac{2}{6} = 0,333$
P(Perempuan Kelas 1)	$\frac{135}{176} = 0,767$
P(Perempuan Kelas 2)	$\frac{4}{11} = 0,367$
P(Perempuan Kelas 3)	$\frac{2}{4} = 0,5$
P(Perempuan Kelas 4)	$\frac{4}{6} = 0,667$

Untuk probabilitas atribut terhadap setiap label kelas yang lainnya, selengkapnya dapat dilihat pada bagian lampiran.

c. Probabilitas Posterior ($P(H | X)$)

Tahap terakhir dalam proses klasifikasi adalah menghitung probabilitas posterior, yaitu proses untuk menentukan probabilitas akhir dari setiap kelas target berdasarkan probabilitas prior dan probabilitas bersyarat (*likelihood*) yang telah dihitung sebelumnya.

Tabel 3. 6. Sampel Data Uji Yang Tidak Diketahui Label Kelasnya

No	Atribut	Nilai
1.	Jenis kelamin	Laki-laki
2.	Asal Daerah	Luar DIY
3.	Indeks Prestasi Semester 1	3.14
4.	Indeks Prestasi Semester 2	2.09
5.	Indeks Prestasi Semester 3	3.33
6.	Indeks Prestasi Semester 4	3.00
7.	Serologi Golongan Darah I	C
8.	Infeksi Menular Lewat Transfusi Darah I	C
9.	Serologi Golongan Darah II	C
10.	Infeksi Menular Lewat Transfusi Darah II	C
11.	Penyadapan Darah	A
12.	Serologi Golongan Darah III	B
13.	Infeksi Menular Lewat Transfusi Darah III	B
14.	Komponen Darah	B
15.	Serologi Golongan Darah IV	B
16.	Infeksi Menular Lewat Transfusi Darah IV	A

Probabilitas posterior menunjukkan kemungkinan seorang mahasiswa dengan karakteristik data (X) untuk masuk ke dalam masing-masing label kelas (1, 2, 3, atau 4). Data mahasiswa akan diklasifikasikan ke dalam kelas dengan probabilitas posterior yang tertinggi. Berikut perhitungan probabilitas posterior untuk masing-masing label kelas :

- $P(\text{Label Kelas}="1" | X) = P(\text{Label Kelas}="1") \times P(X | \text{Label Kelas}="1")$
 $= 0,893 \times 0,233 \times 0,932 \times 0,807 \times 0,42 \times 0,869 \times$
 $0,545 \times 0,534 \times 0,097 \times 0,091 \times 0,091 \times 0,75 \times$
 $0,659 \times 0,579 \times 0,579 \times 0,528 \times 0,699$
 $= 8,16 \times 10^{-7}$
- $P(\text{Label Kelas}="2" | X) = P(\text{Label Kelas}="2") \times P(X | \text{Label Kelas}="2")$
 $= 0,056 \times 0,636 \times 1 \times 0,091 \times 1 \times 0,545 \times 0,091 \times$
 $0,727 \times 0,636 \times 0,455 \times 0,818 \times 0,455 \times 0,545 \times$
 $0,364 \times 0,545 \times 0,545 \times 0,364$
 $= 2,7 \times 10^{-7}$
- $P(\text{Label Kelas}="3" | X) = P(\text{Label Kelas}="3") \times P(X | \text{Label Kelas}="3")$
 $= 0,02 \times 0,5 \times 0,75 \times 1 \times 0,75 \times 0,5 \times 0,25 \times 0,75 \times$
 $0,5 \times 0,25 \times 0,25 \times 0,75 \times 0,5 \times 1 \times 0,5 \times 0,75 \times$
 $0,75$
 $= 1,73 \times 10^{-6}$
- $P(\text{Label Kelas}="4" | X) = P(\text{Label Kelas}="4") \times P(X | \text{Label Kelas}="4")$
 $= 0,031 \times 0,333 \times 0,833 \times 0,167 \times 1 \times 0 \times 0 \times 1 \times$
 $0,667 \times 0,667 \times 0,167 \times 0 \times 0,333 \times 0,333 \times 0,167$
 $\times 0 \times 0,333$
 $= 0$

Berdasarkan nilai probabilitas posterior tertinggi, data mahasiswa (X) paling mungkin masuk ke dalam **Label Kelas 3**.

4. Evaluasi Model

Tahap terakhir dalam pengembangan model adalah evaluasi model prediksi dengan menggunakan matriks konfusi (*confusion matrix*) untuk menghitung akurasi, presisi, *recall*, dan *F1-score* dari model yang dibangun.

Tabel 3. 7. Confusion Matrix

Prediksi Aktual	Kelas 1	Kelas 1	Kelas 1	Kelas 1
Kelas 1	46	0	0	0
Kelas 2	3	0	0	0
Kelas 3	0	0	1	0
Kelas 4	0	0	0	0

a. Akurasi

Akurasi menunjukkan kemampuan model dalam memprediksi data dengan benar secara keseluruhan. Akurasi dihitung dengan membagi jumlah prediksi benar (*true positive* dan *true negative*) dengan total jumlah data.

$$\text{Akurasi} = \frac{TP_1 + TP_2 + TP_3 + TP_4}{\text{Total Instance}} = \frac{46 + 0 + 1 + 0}{50} = 0,94$$

b. Presisi

Presisi menunjukkan tingkat ketepatan model dalam memprediksi data positif. Presisi dihitung dengan membagi jumlah *true positive* dengan jumlah *true positive* dan *false positive*.

$$\text{➤ Presisi Kelas 1} = \frac{TP_1}{TP_1 + FP_1} = \frac{46}{46 + 3 + 0 + 0} = 0,94$$

$$\text{➤ Presisi Kelas 2} = \frac{TP_2}{TP_2 + FP_2} = \frac{0}{0 + 0 + 0 + 0} = 0$$

$$\begin{aligned} \text{➤ } \textit{Presisi Kelas 3} &= \frac{TP_3}{TP_3 + FP_3} = \frac{1}{0 + 0 + 1 + 0} = 1 \\ \text{➤ } \textit{Presisi Kelas 4} &= \frac{TP_4}{TP_4 + FP_4} = \frac{0}{0 + 0 + 0 + 0} = 0 \end{aligned}$$

c. *Recall*

Recall menunjukkan tingkat kelengkapan model dalam mendeteksi data positif.

Recall dihitung dengan membagi jumlah *true positive* dengan jumlah *true positive* dan *false negative*.

$$\begin{aligned} \text{➤ } \textit{Recall Kelas 1} &= \frac{TP_1}{TP_1 + FN_1} = \frac{46}{46 + 0 + 0 + 0} = 1 \\ \text{➤ } \textit{Recall Kelas 2} &= \frac{TP_2}{TP_2 + FN_2} = \frac{0}{3 + 0 + 0 + 0} = 0 \\ \text{➤ } \textit{Recall Kelas 3} &= \frac{TP_3}{TP_3 + FN_3} = \frac{1}{0 + 0 + 1 + 0} = 1 \\ \text{➤ } \textit{Recall Kelas 4} &= \frac{TP_4}{TP_4 + FN_4} = \frac{0}{0 + 0 + 0 + 0} = 0 \end{aligned}$$

d. *F1-Score*

F1-Score adalah rata-rata harmonis antara presisi dan *recall*, yang memberikan penilaian seimbang terhadap kedua metrik tersebut.

$$\begin{aligned} \text{➤ } \textit{F1 - Score Kelas 1} &= 2 \times \frac{\textit{Presisi}_{Kelas 1} \times \textit{Recall}_{Kelas 1}}{\textit{Presisi}_{Kelas 1} + \textit{Recall}_{Kelas 1}} \\ &= 2 \times \frac{0,94 \times 1}{0,94 + 1} = 0,97 \\ \text{➤ } \textit{F1 - Score Kelas 2} &= 2 \times \frac{\textit{Presisi}_{Kelas 2} \times \textit{Recall}_{Kelas 2}}{\textit{Presisi}_{Kelas 2} + \textit{Recall}_{Kelas 2}} \\ &= 2 \times \frac{0 \times 0}{0 + 0} = 0 \\ \text{➤ } \textit{F1 - Score Kelas 3} &= 2 \times \frac{\textit{Presisi}_{Kelas 3} \times \textit{Recall}_{Kelas 3}}{\textit{Presisi}_{Kelas 3} + \textit{Recall}_{Kelas 3}} \end{aligned}$$

$$= 2 \times \frac{1 \times 1}{1 + 1} = 1$$

$$\text{➤ } F1 - \text{Score Kelas 4} = 2 \times \frac{\text{Presisi}_{\text{Kelas 4}} \times \text{Recall}_{\text{Kelas 4}}}{\text{Presisi}_{\text{Kelas 4}} + \text{Recall}_{\text{Kelas 4}}}$$

$$= 2 \times \frac{0 \times 0}{0 + 0} = 0$$

F1-Score Rata-rata :

$$= \frac{F1 - \text{Score}_{\text{Kelas 1}} + F1 - \text{Score}_{\text{Kelas 2}} + F1 - \text{Score}_{\text{Kelas 3}} + F1 - \text{Score}_{\text{Kelas 4}}}{4}$$

$$= \frac{0,97 + 0 + 1 + 0}{4}$$

$$= 0,49$$

5. Evaluasi model

Analisis hasil evaluasi model bertujuan untuk menganalisis dan menarik kesimpulan dari hasil evaluasi model, serta memberikan rekomendasi untuk perbaikan atau pengembangan model selanjutnya.

- a. Hasil perhitungan akurasi model adalah 0,94 atau 94%, menunjukkan bahwa model secara keseluruhan mampu memprediksi kelulusan mahasiswa dengan sangat baik, dengan tingkat kesalahan prediksi 6%.
- b. Pada hasil uji presisi, model memiliki presisi yang sangat baik untuk Kelas 1 (94%) dan Kelas 3 (100%), namun presisi untuk Kelas 2 dan Kelas 4 adalah 0, artinya model tidak dapat memprediksi kelas-kelas tersebut dengan benar.
- c. Dari sisi uji *recall*, model memiliki *recall* yang sangat baik untuk Kelas 1 dan Kelas 3, yaitu masing-masing 100%. Tetapi *recall* untuk Kelas 2 dan Kelas 4 adalah 0, menunjukkan bahwa model tidak dapat mendeteksi kelas-kelas tersebut dengan tepat.

- d. Nilai rata-rata *F1-Score* dari semua kelas adalah 0,49 atau 49%, yang merupakan rata-rata harmonis antara presisi dan *recall*. Nilai ini menunjukkan bahwa secara keseluruhan, performa model masih dapat ditingkatkan, terutama dalam memprediksi Kelas 2 dan Kelas 4.
- e. Secara umum, model memiliki performa yang sangat baik dalam memprediksi Kelas 1 dan Kelas 3, namun masih lemah dalam memprediksi Kelas 2 dan Kelas 4.

3.1.5 Kebutuhan Keluaran (*Output*)

Keluaran dari aplikasi prediksi kelulusan mahasiswa adalah informasi hasil prediksi kelulusan mahasiswa berdasarkan label kelas (1, 2, 3, atau 4). Label kelas ditentukan berdasarkan *input* data yang diberikan, dengan rincian sebagai berikut :

Tabel 3. 8. Label Kelas

No	Label kelas	Keterangan
1.	1	Lulus dalam 3 tahun dan $IPK \geq 3,00$
2.	2	Lulus dalam 3 tahun dan $IPK < 3,00$
3.	3	Lulus lebih dari 3 tahun dan $IPK \geq 3,00$
4.	4	Lulus lebih dari 3 tahun dan $IPK < 3,00$

Laporan prediksi dapat digunakan oleh pihak akademik untuk evaluasi dan perbaikan proses pembelajaran.

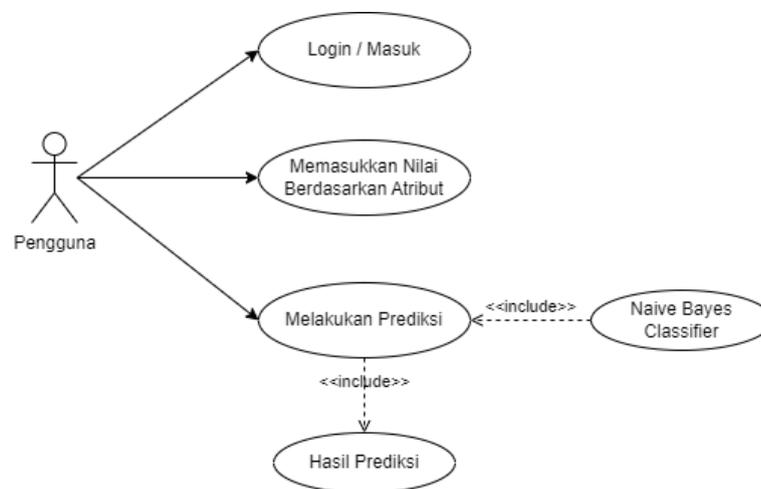
3.2 Perancangan Sistem

Perancangan adalah tahap penting dalam pembuatan aplikasi agar sesuai harapan. Perancangan sistem dilakukan untuk mempermudah pembangunan dan

mendokumentasikan sistem. Dokumentasi perancangan berguna untuk pemeliharaan dan pengembangan aplikasi di masa depan.

3.2.1 Use Case Diagram

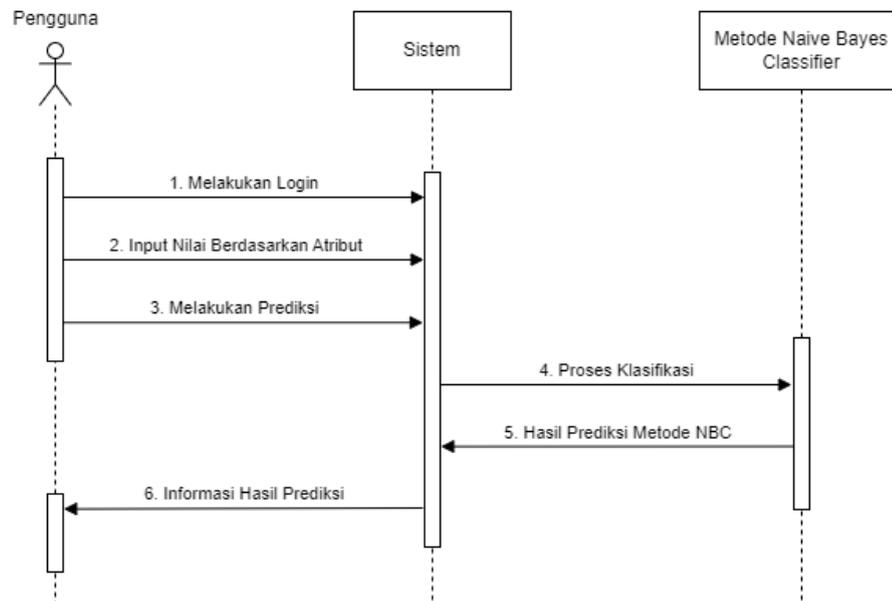
Berikut ini adalah *use case diagram* yang menggambarkan hubungan interaksi antara pengguna (aktor), yaitu admin program studi dengan aplikasi yang sedang dirancang (sistem).



Gambar 3. 8. Use Case Diagram

3.2.2 Sequence Diagram

Berikut ini adalah diagram sekuen (*sequence diagram*) yang digunakan untuk menggambarkan urutan interaksi antar objek dalam sistem berdasarkan urutan waktu.



Gambar 3. 9. *Sequence Diagram*

3.3 Perancangan Antarmuka

Perancangan antarmuka pengguna (*user interface*) merupakan tahapan penting dalam pengembangan aplikasi sebelum implementasi program. Tahap ini mencakup perencanaan dan pembuatan desain tampilan halaman *website* untuk memberikan gambaran jelas tentang bagaimana pengguna akan berinteraksi dengan aplikasi. Antarmuka pengguna dirancang dengan memperhatikan aspek kemudahan penggunaan (*usability*) agar memudahkan navigasi dan interaksi pengguna.

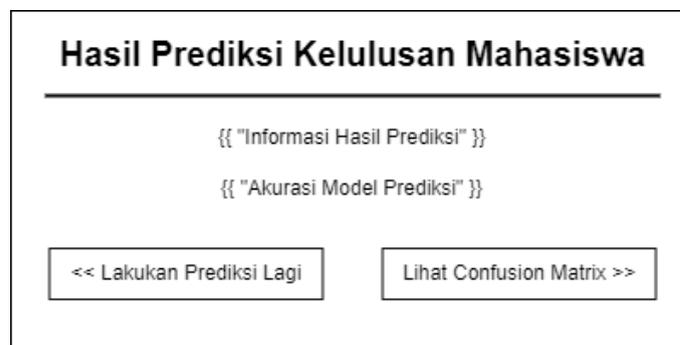
3.4.1 Rancangan *Input*

Rancangan antarmuka *input* menyediakan sarana bagi pengguna untuk memasukkan data atau informasi yang diperlukan sebelum melakukan prediksi kelulusan mahasiswa. Berikut rincian perancangan halaman antarmuka *input* :

1. Halaman *Login*

3.4.2 Rancangan *Output*

Rancangan output mencakup antarmuka yang menampilkan hasil prediksi kelulusan mahasiswa setelah proses analisis selesai. Tampilan output akan menunjukkan informasi hasil prediksi berdasarkan data yang dimasukkan pengguna. Tujuannya adalah menyediakan informasi dengan jelas dan mudah dipahami oleh pengguna mengenai perkiraan kemungkinan kelulusan mahasiswa.



Gambar 3. 12. Desain Antarmuka Halaman Hasil Prediksi

BAB IV

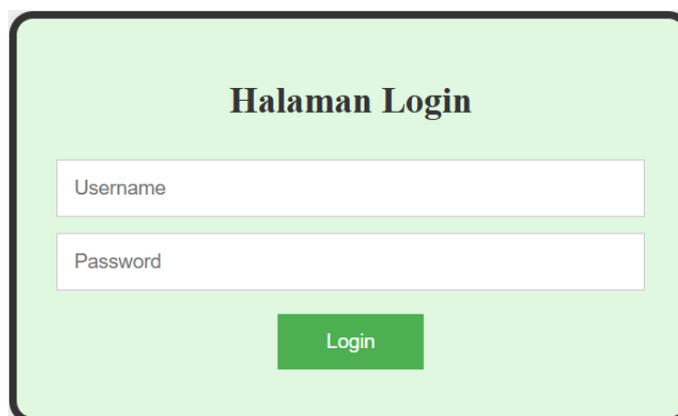
IMPLEMENTASI DAN PEMBAHASAN

4.1 Implementasi dan Uji Coba Sistem

Tahap implementasi dan uji coba sistem dilakukan untuk memverifikasi bahwa aplikasi yang sedang dikembangkan dapat dijalankan dan berfungsi dengan baik sesuai yang diharapkan. Pada tahap ini, aplikasi akan dilakukan pengujian lebih rinci terkait dengan hasil yang diperoleh dan manfaatnya bagi pengguna.

4.1.1 Halaman *Login*

Halaman *login* merupakan tampilan pertama saat aplikasi dijalankan. Sebelum memasuki halaman utama untuk melakukan prediksi, pengguna harus melakukan *login* dengan menggunakan *username* = “admin” dan *password* = “12345”, kemudian mengklik *login*, maka sistem akan memverifikasi dan memberikan akses ke halaman utama. Apabila *login* berhasil, maka pengguna dapat melakukan prediksi kelulusan mahasiswa.



The image shows a login form with a light green background. At the top, the text "Halaman Login" is centered. Below it are two input fields: "Username" and "Password". At the bottom center is a green button labeled "Login".

Gambar 4. 1. Tampilan Halaman *Login*

Jika pengguna salah menginputkan *username* atau *password*, maka sistem akan menampilkan peringatan “**Username atau password salah !**”.



Gambar 4. 2. Tampilan Gagal *Login*

```

1 # Routes
2 @app.route('/', methods=['GET', 'POST'])
3 def login():
4     error = None
5     if request.method == "POST":
6         username = request.form["username"]
7         password = request.form["password"]
8
9         cur = mysql.connection.cursor()
10        cur.execute("SELECT * FROM users WHERE username = %s AND password = %s", (username, password))
11        user = cur.fetchone()
12        cur.close()
13
14        if user:
15            return redirect(url_for('home')) # Mengarahkan pengguna ke halaman home setelah login sukses
16        else:
17            error = "Username atau password salah !"
18
19        return render_template("login.html", error=error)
20

```

Gambar 4. 3. Route Halaman *Login*

Penjelasan kode program untuk route halaman *login* :

- Baris 5-7 merupakan fungsi untuk mengambil nilai *username* dan *password* yang dikirimkan oleh pengguna.

- Baris 9-12 merupakan fungsi untuk melakukan koneksi ke *database* MySQL, kemudian akan mengeksekusi *query SQL* untuk mencari pengguna dengan *username* dan *password* yang cocok di tabel *users*.
- Baris 14-15 merupakan fungsi untuk mengarahkan pengguna ke halaman utama atau *home*, jika *username* dan *password* ditemukan di *database*.
- Baris 16-17 merupakan fungsi untuk menampilkan pesan *error* di halaman *login* jika *username* dan *password* yang dimasukkan salah atau tidak terdapat di *database*.

4.1.2 Halaman Utama

Halaman Utama merupakan halaman inti untuk melakukan prediksi. Pada halaman ini, pengguna wajib memasukkan nilai pada semua kolom *input* sesuai dengan nama atributnya. Untuk atribut jenis kelamin, pengguna dapat memilih antara laki-laki atau perempuan. Untuk atribut asal daerah, pengguna dapat memilih Yogyakarta atau luar Yogyakarta. Untuk atribut Indeks Prestasi Semester 1 hingga 4, pengguna harus menginputkan nilai dengan tipe numerik atau angka. Untuk atribut matakuliah kompetensi, pengguna dapat memilih nilai A, B, C, D, atau E.

Sebelum melakukan prediksi, pastikan semua kolom *input* telah terisi dengan benar. Pengguna dapat mengklik tombol "Prediksi NBC" untuk memulai proses klasifikasi menggunakan metode *Naive Bayes*. Informasi hasil prediksi yang didapatkan bergantung pada nilai-nilai yang dimasukkan oleh pengguna.

Gambar 4. 4. Tampilan Halaman Utama

4.2 Implementasi Sistem

Pada bagian ini dijelaskan implementasi algoritma *Naive Bayes Classifier* untuk memprediksi kelulusan mahasiswa. Tahap pertama adalah mengisi semua kolom input data uji dengan benar dan lengkap, agar sistem dapat melakukan prediksi. Berikut gambar halaman utama dengan data uji yang telah diisi :

Gambar 4. 5. Pengisian Semua Kolom Input Dengan Data Uji

```

1 # Route untuk melakukan prediksi dengan metode nbc
2 @app.route('/predict_nbc', methods=['POST'])
3 def prediction_nbc():
4     # Ambil data dari form input
5     JENIS_KELAMIN = int(request.form['JENIS KELAMIN'])
6     ASAL_DAERAH = int(request.form['ASAL DAERAH'])
7     IPS_1 = float(request.form['IPS_1'])
8     IPS_2 = float(request.form['IPS_2'])
9     IPS_3 = float(request.form['IPS_3'])
10    IPS_4 = float(request.form['IPS_4'])
11    SEROLOGI_GOLONGAN_DARAH_I = int(request.form['SEROLOGI GOLONGAN DARAH I'])
12    SEROLOGI_GOLONGAN_DARAH_II = int(request.form['SEROLOGI GOLONGAN DARAH II'])
13    SEROLOGI_GOLONGAN_DARAH_III = int(request.form['SEROLOGI GOLONGAN DARAH III'])
14    SEROLOGI_GOLONGAN_DARAH_IV = int(request.form['SEROLOGI GOLONGAN DARAH IV'])
15    INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_I = int(request.form['INFEKSI MENULAR LEWAT TRANSFUSI DARAH I'])
16    INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_II = int(request.form['INFEKSI MENULAR LEWAT TRANSFUSI DARAH II'])
17    INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_III = int(request.form['INFEKSI MENULAR LEWAT TRANSFUSI DARAH III'])
18    INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_IV = int(request.form['INFEKSI MENULAR LEWAT TRANSFUSI DARAH IV'])
19    PENYADAPAN_DARAH = int(request.form['PENYADAPAN DARAH'])
20    KOMPONEN_DARAH = int(request.form['KOMPONEN DARAH'])
21
22    # Buat data test dari input form
23    data_test = [[JENIS_KELAMIN, ASAL_DAERAH, IPS_1, IPS_2, IPS_3, IPS_4, SEROLOGI_GOLONGAN_DARAH_I, SEROLOGI_GOLONGAN_DARAH_II,
24                SEROLOGI_GOLONGAN_DARAH_III, SEROLOGI_GOLONGAN_DARAH_IV, INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_I, INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_II,
25                INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_III, INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_IV, PENYADAPAN_DARAH, KOMPONEN_DARAH]]
26
27    # Dapatkan prediksi dan akurasi menggunakan NBC
28    result = predict_nbc(data_test)
29    accuracy = get_accuracy_nbc()
30
31    # Format akurasi dengan dua digit di belakang koma
32    formatted_accuracy = f"{accuracy:.2f}"
33
34    # Render hasil prediksi dan akurasi ke hasil.html
35    return render_template('hasil.html', hasil=result, akurasi=formatted_accuracy, matrix_route=url_for('confusion_matrix_nbc'))
36

```

Gambar 4. 6. Route Prediksi *Naive Bayes*

Penjelasan kode program untuk route prediksi dengan *Naive Bayes* :

- Baris 5-20 merupakan fungsi untuk mengambil nilai-nilai data yang dikirimkan dari *form input*, kemudian menyimpan nilai tersebut dalam variabel dengan tipe data yang sesuai, seperti *integer* atau *float*.
- Baris 23-25 merupakan fungsi untuk menyusun data input yang diterima menjadi sebuah *list*, yang selanjutnya akan digunakan sebagai data uji (*data test*).
- Baris 28 merupakan fungsi untuk melakukan prediksi menggunakan data yang telah diberikan.
- Baris 29 merupakan fungsi untuk menghitung nilai akurasi dari model *Naive Bayes Classifier* (NBC).
- Baris 32 merupakan fungsi untuk memformat nilai akurasi menjadi dua digit di belakang koma.

- Baris 35 merupakan fungsi untuk mengirimkan hasil prediksi dan akurasi ke halaman `result.html` agar dapat ditampilkan kepada pengguna, serta menyediakan *route* untuk melihat *confusion matrix*.

Berikut ini merupakan bagian-bagian dari kode program `nbc.py` yang merepresentasikan implementasi algoritma *Naive Bayes Classifier* untuk klasifikasi data :

```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
```

Gambar 4. 7. Kode Program Bagian Impor Modul Dari *Library* Eksternal

Penjelasan :

1. *import pandas as pd*

- Digunakan untuk mengimpor *library* `pandas`, yang merupakan salah satu *library* Python paling populer untuk manipulasi dan analisis data.
- Dengan mengimpor `pandas`, maka dapat menggunakan berbagai fungsi dan struktur data seperti *DataFrame* dan *Series* untuk bekerja dengan data.

2. *from sklearn.decomposition import PCA*

- Digunakan untuk mengimpor kelas `PCA` (*Principal Component Analysis*) dari modul `sklearn.decomposition`.

- PCA adalah teknik reduksi dimensi yang dapat digunakan untuk mengidentifikasi fitur-fitur penting dalam dataset dengan mengurangi jumlah fitur tanpa menghilangkan banyak informasi.

3. *from sklearn.model_selection import train_test_split*

- Digunakan untuk mengimpor fungsi *train_test_split* dari modul *sklearn.model_selection*.
- *train_test_split* digunakan untuk membagi dataset menjadi subset pelatihan (*training*) dan subset pengujian (*testing*), yang berguna untuk mengevaluasi kinerja model *machine learning*.

4. *from sklearn.naive_bayes import GaussianNB*

- Digunakan untuk mengimpor kelas *GaussianNB* dari modul *sklearn.naive_bayes*.
- *GaussianNB* adalah salah satu algoritma klasifikasi *Naive Bayes* yang mengasumsikan bahwa fitur-fitur dalam dataset terdistribusi normal.

5. *from sklearn.metrics import accuracy_score*

- Digunakan untuk mengimpor fungsi *accuracy_score* dari modul *sklearn.metrics*.
- *accuracy_score* digunakan untuk menghitung akurasi model klasifikasi dengan membandingkan prediksi model dengan label sebenarnya.

6. *from sklearn.preprocessing import StandardScaler*

- Digunakan untuk mengimpor kelas *StandardScaler* dari modul *sklearn.preprocessing*.

- *StandardScaler* digunakan untuk melakukan standardisasi fitur-fitur dalam dataset, dengan mengurangi rata-rata dan membagi dengan standar deviasi. Ini berguna untuk meningkatkan kinerja algoritma *machine learning*.

7. *from sklearn.impute import SimpleImputer*

- Fungsi ini digunakan untuk mengimpor kelas *SimpleImputer* dari modul *sklearn.impute*.
- *SimpleImputer* digunakan untuk menangani *missing values* dalam dataset dengan mengisi nilai yang hilang dengan strateginya, seperti rata-rata, median, atau konstantanya.

```
# Load dataset dari file Excel
dataset = pd.read_excel("DATASET.xlsx",
sheet_name='DATASHEET')

# Menghapus kolom yang tidak diperlukan dari dataset
df = dataset.drop(['NO', 'NIM', 'NAMA', 'PROGRAM STUDI',
'IPS_5', 'IPS_6', 'IPK LULUS', 'NILAI KARYA TULIS'], axis=1)

# Memisahkan fitur (x) dan label (y) dari dataset
x = df.iloc[:, :-1].values # Matriks fitur (semua kolom
kecuali kolom terakhir)

y = df['LABEL KELAS']      # Label (kolom 'LABEL KELAS')

# Inisialisasi objek Standard Scaler untuk normalisasi fitur
scaler = StandardScaler()

# Menerapkan standar scaler pada fitur
x_scaled = scaler.fit_transform(x)

# Mengganti nilai NaN dengan rata-rata kolom
imputer = SimpleImputer(strategy='mean')
x_imputed = imputer.fit_transform(x_scaled)

# Mengaplikasikan PCA dengan 5 komponen utama
pca = PCA(n_components=5)
pca_dataset = pca.fit_transform(x_imputed)
```

```

# Membagi dataset menjadi data pelatihan 80% dan data
pengujian 20%
x_train, x_test, y_train, y_test =
train_test_split(pca_dataset, y, test_size=0.20,
random_state=42)

# Membuat objek Gaussian Naive Bayes (NBC)
nbc = GaussianNB()

# Melatih model NBC dengan data pelatihan
model = nbc.fit(x_train, y_train)

```

Gambar 4. 8. Kode Program Bagian Praproses Data

Penjelasan kode program :

1. *Load* dataset dari file Excel

Menggunakan *pd.read_excel()* untuk membaca data dari file Excel dengan nama "DATASET.xlsx" dan mengambil sheet bernama "DATASHEET".

Kemudian dataset yang dibaca disimpan dalam variabel dataset.

2. Menghapus kolom yang tidak diperlukan dari dataset

Menggunakan *df.drop()* untuk menghapus beberapa kolom yang tidak digunakan, seperti "NO", "NIM", "NAMA", dll. Kemudian hasil dari penghapusan kolom disimpan dalam variabel df.

3. Memisahkan fitur (x) dan label (y) dari dataset

Fitur atau variabel *independen* disimpan dalam variabel x menggunakan *df.iloc[:, :-1].values*. Label atau variabel *dependen* disimpan dalam variabel y menggunakan *df['LABEL KELAS']*.

4. Normalisasi fitur menggunakan *Standard Scaler*

Membuat objek *StandardScaler* dan menyimpannya dalam variabel *scaler*. Kemudian menerapkan standar normalisasi pada fitur x menggunakan *scaler.fit_transform()*, lalu hasilnya disimpan dalam variabel *x_scaled*.

5. Mengisi nilai *NaN* dengan rata-rata kolom

Membuat objek *SimpleImputer* dengan strategi pengisian nilai *NaN* menggunakan rata-rata kolom. Yaitu dengan menerapkan *SimpleImputer* pada *x_scaled* dan hasilnya disimpan dalam variabel *x_imputed*.

6. Menerapkan PCA dengan 5 komponen utama

Membuat objek PCA dengan 5 komponen utama. Menerapkan PCA pada *x_imputed* dan menyimpan hasilnya dalam variabel *pca_dataset*.

7. Membagi dataset menjadi data pelatihan dan pengujian

Menggunakan *train_test_split()* untuk membagi *pca_dataset* dan y menjadi data pelatihan (x_{train} , y_{train}) dan data pengujian (x_{test} , y_{test}) dengan rasio 80:20.

8. Membuat dan melatih model *Gaussian Naive Bayes*

Membuat objek *GaussianNB()* dan menyimpannya dalam variabel *nbc*. Melatih model dengan data pelatihan menggunakan *nbc.fit()* dan menyimpan model dalam variabel *model*.

```
def predict(data_test):
    # Normalisasi dan menerapkan PCA pada data uji
    pca_test = pca.transform(scaler.transform(data_test))

    # Membuat prediksi menggunakan model yang sudah dilatih
    y_pred = model.predict(pca_test)

    # Mengembalikan hasil prediksi dalam bentuk deskriptif
    if y_pred == 1:
```

```

        return "Mahasiswa lulus dalam 3 tahun dengan IPK >
3,00"
    elif y_pred == 2:
        return "Mahasiswa lulus dalam 3 tahun dengan IPK <
3,00"
    elif y_pred == 3:
        return "Mahasiswa lulus lebih dari 3 tahun dengan
IPK > 3,00"
    else:
        return "Mahasiswa lulus lebih dari 3 tahun dengan
IPK < 3,00"

def get_accuracy():
    # Membuat prediksi pada data pengujian
    y_prediction = model.predict(x_test)

    # Menghitung dan mengembalikan akurasi model
    return accuracy_score(y_test, y_prediction)

def get_y_true():
    # Mengembalikan nilai sebenarnya dari data pengujian
    return y_test

def get_y_pred():
    # Membuat prediksi pada data pengujian
    return model.predict(x_test)

```

Gambar 4. 9. Kode Program Bagian Evaluasi Model

Penjelasan kode program :

1. Fungsi *predict(data_test)*

Ini adalah fungsi untuk membuat prediksi menggunakan model yang telah dilatih sebelumnya. Dalam fungsi ini, data uji dinormalisasi dan ditransformasi menggunakan PCA yang sama seperti pada data pelatihan. Kemudian, model yang telah dilatih sebelumnya digunakan untuk membuat prediksi pada data uji. Hasil prediksi dikembalikan dalam bentuk deskriptif berdasarkan kelas yang diprediksi.

2. Fungsi *get_accuracy()*

Ini adalah fungsi untuk menghitung akurasi model pada data pengujian. Dalam fungsi ini, prediksi dibuat pada data pengujian menggunakan model yang telah dilatih. Kemudian, akurasi model dihitung menggunakan *accuracy_score* dari *library sklearn.metrics* dan nilai akurasi tersebut dikembalikan.

3. Fungsi *get_y_true()*

Ini adalah fungsi untuk mengembalikan nilai sebenarnya (*ground truth*) dari data pengujian.

4. Fungsi *get_y_pred()*

Ini adalah fungsi untuk mengembalikan hasil prediksi yang dibuat pada data pengujian.

4.2.1 Pengujian akurasi sistem

Pengujian akurasi sistem dilakukan untuk mengetahui seberapa akurat sistem dalam memprediksi kelulusan mahasiswa. Pengujian ini dilakukan dengan membandingkan hasil prediksi sistem menggunakan algoritma *Naive Bayes Classifier* dengan perhitungan secara manual menggunakan *confusion matrix*. Berikut hasil akurasi yang dihasilkan oleh sistem dan informasi prediksi data uji :



The image shows a light blue rectangular box with a black border. At the top, the text "HASIL PREDIKSI KELULUSAN MAHASISWA" is written in red, bold, uppercase letters. Below this, a horizontal line separates the title from the content. The content consists of two lines of black text: "Mahasiswa lulus dalam 3 tahun dengan IPK $\geq 3,00$ " and "Akurasi model : 0.94". At the bottom of the box, there are two blue buttons with white text. The left button says "<< Lakukan Prediksi Lagi" and the right button says "Lihat Confusion Matrix >>".

Gambar 4. 10. Hasil Prediksi Dari Data Uji

Berdasarkan pengujian yang telah dilakukan, sistem mampu mencapai akurasi sebesar 0,94 atau 94%. Hasil ini menunjukkan bahwa sistem mampu memprediksi kelulusan mahasiswa dengan tingkat ketepatan yang cukup tinggi.

4.2.2 Evaluasi

Setelah melakukan pengujian akurasi sistem, tahap selanjutnya adalah mengevaluasi hasil prediksi. Evaluasi dilakukan dengan tujuan untuk mengetahui kinerja dan efektivitas model prediksi yang telah dibangun. Metrik evaluasi yang dihitung meliputi akurasi, presisi, *recall*, dan *F1-score*. Perhitungan metrik evaluasi ini menggunakan data yang diperoleh dari *confusion matrix* :

Confusion Matrix				
	Prediksi Kelas 1	Prediksi Kelas 2	Prediksi Kelas 3	Prediksi Kelas 4
Kelas Aktual 1	46	0	0	0
Kelas Aktual 2	3	0	0	0
Kelas Aktual 3	0	0	1	0
Kelas Aktual 4	0	0	0	0

[Kembali ke Home](#)

Gambar 4. 11. *Confusion Matrix*

1. Akurasi

Akurasi digunakan untuk menghitung presentase jumlah prediksi yang benar dari seluruh data uji. Akurasi dihitung dengan membandingkan jumlah prediksi

yang benar dengan total jumlah data uji. Nilai akurasi yang diperoleh dari hasil perhitungan yang telah dilakukan adalah sebesar 0.94 atau 94%.

2. Presisi

Presisi digunakan untuk menghitung seberapa tepat sistem dalam memprediksi kelas positif. Presisi dihitung dengan membandingkan jumlah prediksi positif yang benar dengan total jumlah prediksi positif. Nilai presisi Kelas 1 sebesar 0,94 atau 94%, presisi Kelas 2 sebesar 0, presisi Kelas 3 sebesar 1 atau 100%, dan presisi Kelas 4 sebesar 0.

3. Recall

Recall digunakan untuk menghitung seberapa banyak kelas positif yang berhasil diprediksi dengan benar. *Recall* dihitung dengan membandingkan jumlah prediksi positif yang benar dengan total jumlah data positif. Nilai *recall* Kelas 1 sebesar 1 atau 100%, *recall* Kelas 2 adalah 0, *recall* Kelas 3 sebesar 1 atau 100%, dan *recall* Kelas 4 adalah 0.

4. F1-Score

F1-Score digunakan untuk menghitung rata-rata harmonik antara presisi dan *recall*. Nilai rata-rata *F1-Score* yang diperoleh yaitu sebesar 0.49 atau 49%. Nilai ini menunjukkan bahwa sistem prediksi memiliki akurasi yang sedang dalam memprediksi kelulusan mahasiswa. Nilai *F1-Score* yang cenderung rendah mengindikasikan bahwa sistem belum dapat mencapai keseimbangan yang baik antara presisi dan *recall*.

4.3 Pembahasan

Hasil implementasi dan pengujian sistem menunjukkan bahwa sistem prediksi kelulusan mahasiswa dengan metode *Naive Bayes Classifier* dapat berjalan dengan baik dan memberikan hasil yang akurat. Sistem ini memiliki potensi untuk membantu pihak institusi dalam mengambil keputusan terkait upaya-upaya yang perlu dilakukan untuk meningkatkan kelulusan mahasiswa.

Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa model prediksi yang digunakan memiliki tingkat akurasi yang sangat baik, yaitu mencapai 94%. Hal ini berarti sistem mampu memprediksi kelulusan mahasiswa dengan sangat akurat dan hanya memiliki kesalahan sebesar 6%. Selain akurasi, juga dilakukan perhitungan nilai presisi, *recall*, dan *F1-score* untuk memberikan gambaran yang lebih menyeluruh tentang kinerja sistem.

Presisi model untuk Kelas 1 dan Kelas 3 sangat baik, masing-masing mencapai 94% dan 100%. Hal ini mengindikasikan bahwa model dapat memprediksi dengan tepat mahasiswa yang termasuk dalam kelas-kelas tersebut. Namun, presisi model untuk Kelas 2 dan Kelas 4 adalah 0, yaitu menunjukkan bahwa model belum dapat memprediksi dengan benar kelas-kelas ini. Rendahnya presisi juga kemungkinan menunjukkan bahwa model memiliki kesulitan dalam membedakan karakteristik mahasiswa di Kelas 2 dan Kelas 4, sehingga perlu dilakukan perbaikan pada model.

Dari sisi *recall*, model juga menunjukkan performa yang sangat baik untuk Kelas 1 dan Kelas 3, dengan nilai *recall* mencapai 100% untuk kedua kelas tersebut. Ini berarti model dapat mendeteksi dengan tepat mahasiswa yang termasuk dalam

Kelas 1 dan Kelas 3. Namun sayangnya, *recall* model untuk Kelas 2 dan Kelas 4 adalah 0, artinya model belum dapat mendeteksi dengan benar kelas-kelas ini. Rendahnya *recall* juga kemungkinan menunjukkan bahwa model memiliki kesulitan dalam mengenali karakteristik mahasiswa di Kelas 2 dan Kelas 4, sehingga perlu dilakukan perbaikan pada model.

Nilai rata-rata *F1-score* sistem yang diperoleh melalui perhitungan metrik yaitu sebesar 49%. *F1-Score* merupakan ukuran keseimbangan antara presisi dan *recall*, yang menggambarkan kinerja sistem secara keseluruhan. Nilai *F1-score* sebesar 49% termasuk dalam kategori sedang. Hal ini menunjukkan bahwa masih terdapat ruang untuk meningkatkan kinerja model agar dapat memprediksi dengan lebih baik secara keseluruhan.

Secara umum, hasil evaluasi menunjukkan bahwa sistem prediksi kelulusan mahasiswa yang telah diimplementasikan memiliki kinerja yang cukup baik. Model menunjukkan akurasi yang tinggi dalam memprediksi mahasiswa yang termasuk dalam Kelas 1 dan Kelas 3, namun masih memiliki kesulitan dalam memprediksi mahasiswa yang termasuk dalam Kelas 2 dan Kelas 4 dengan tepat. Hal ini ditunjukkan oleh nilai *F1-score* model secara keseluruhan yang berada pada kategori sedang, yaitu sekitar 49%. Nilai *F1-score* yang sedang mengindikasikan bahwa model memiliki keseimbangan yang cukup antara presisi (ketepatan prediksi) dan *recall* (kemampuan mendeteksi semua kelas dengan benar), namun masih terdapat ruang untuk peningkatan kinerja model agar dapat memprediksi dengan lebih akurat, terutama untuk Kelas 2 dan Kelas 4.

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai prediksi kelulusan mahasiswa dengan metode *Naive Bayes Classifier* pada Program Studi Diploma III Teknologi Bank Darah di Poltekkes BSI Yogyakarta, dapat diambil beberapa kesimpulan sebagai berikut :

1. Implementasi Metode *Naive Bayes Classifier*

Metode *Naive Bayes Classifier* telah berhasil diterapkan untuk memprediksi kelulusan mahasiswa. Hasil dari implementasi ini menunjukkan bahwa model yang dibangun mampu mengklasifikasikan mahasiswa ke dalam kategori label kelas yang tersedia berdasarkan data historis akademik mereka.

2. Akurasi Model

Model *Naive Bayes* yang diimplementasikan memiliki performa yang cukup baik secara umum, dengan akurasi yang memadai dalam memprediksi kelulusan mahasiswa. Model ini berhasil memprediksi dengan tepat mahasiswa yang termasuk dalam Kelas 1 dan Kelas 3, dengan nilai presisi yang tinggi, yaitu 94% untuk Kelas 1 dan 100% untuk Kelas 3. Dari sisi *recall*, model juga menunjukkan performa yang sangat baik untuk Kelas 1 dan Kelas 3, dengan nilai *recall* mencapai 100% untuk kedua kelas ini. Namun, model masih mengalami kesulitan dalam memprediksi dan mendeteksi mahasiswa yang

termasuk dalam Kelas 2 dan Kelas 4, dengan nilai presisi dan *recall* yang rendah, yaitu 0 untuk kedua kelas tersebut.

3. Evaluasi Model

Nilai *F1-score* keseluruhan dari model sebesar 49% menunjukkan bahwa keseimbangan antara presisi dan *recall* berada pada kategori sedang. Ini berarti bahwa sistem masih memiliki ruang untuk peningkatan, terutama dalam meningkatkan kemampuan deteksi pada Kelas 2 dan Kelas 4.

4. Kinerja Keseluruhan

Secara keseluruhan, sistem prediksi yang dibangun telah menunjukkan kinerja yang baik, terutama dalam memprediksi Kelas 1 (lulus dalam 3 tahun dengan $IPK \geq 3,00$) dan Kelas 3 (lulus lebih dari 3 tahun dengan $IPK \geq 3,00$). Namun, perlu adanya perbaikan lebih lanjut untuk meningkatkan prediksi pada Kelas 2 (lulus dalam 3 tahun dengan $IPK < 3,00$) dan Kelas 4 (lulus lebih dari 3 tahun dengan $IPK < 3,00$) agar hasil prediksi menjadi lebih akurat dan menyeluruh.

5.2 SARAN

Berdasarkan kesimpulan yang telah disampaikan, beberapa saran yang dapat diberikan untuk penelitian selanjutnya dan pengembangan sistem adalah sebagai berikut :

1. Penggunaan Data yang Lebih Beragam

Untuk meningkatkan akurasi prediksi, disarankan untuk menggunakan dataset yang lebih besar dan lebih beragam. Hal ini akan membantu model dalam mengenali pola yang lebih kompleks dan meningkatkan performa pada kelas-kelas yang saat ini memiliki presisi atau *recall* rendah.

2. Penerapan Metode Lain

Selain metode *Naive Bayes*, penelitian selanjutnya dapat mencoba metode klasifikasi lain seperti *Decision Tree*, *Random Forest*, atau *Support Vector Machine* (SVM) untuk dibandingkan kinerjanya. Kombinasi beberapa metode (*ensemble methods*) juga dapat dipertimbangkan untuk memperoleh hasil yang lebih akurat.

3. Peningkatan Fitur dan Atribut Data

Penelitian lebih lanjut dapat menambahkan fitur-fitur baru yang mungkin berpengaruh terhadap prediksi kelulusan, seperti data kehadiran, partisipasi dalam kegiatan ekstrakurikuler, dan faktor-faktor non-akademik lainnya.

4. Pengembangan Sistem Berbasis Web

Untuk mempermudah penggunaan oleh pihak akademik, disarankan untuk mengembangkan sistem prediksi kelulusan berbasis web yang *user-friendly* dan dapat diakses oleh dosen dan staf akademik untuk memantau dan memprediksi kelulusan mahasiswa secara *real-time*.

5. Evaluasi Berkelanjutan

Evaluasi model harus dilakukan secara berkala untuk memastikan bahwa model tetap akurat dan relevan seiring dengan perubahan kurikulum dan dinamika akademik lainnya. Pembaruan data dan *retraining* model secara berkala juga diperlukan untuk mempertahankan performa sistem.

DAFTAR PUSTAKA

- Amien, I. L. F., Astuti, W., & Lhaksamana, K. M. (2023). *Perbandingan Metode Naïve Bayes dan KNN (K-Nearest Neighbor) dalam Klasifikasi Penyakit Diabetes*. *E-Proceeding of Engineering*, 10(2), 1911–1920.
- Baiq Nurul Azmi, Arief Hermawan, & Donny Avianto. (2023). *Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver*. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 4(4), 281–290. <https://doi.org/10.35746/jtim.v4i4.298>
- Diska, K. R., & Budayawan, K. (2023). *Sistem Informasi Prediksi Kelulusan Menggunakan Metode Naïve Bayes Classifier (Studi Kasus: Prodi Pendidikan Teknik Informatika)*. *Jurnal Pendidikan Tambusai*, 7(1), 936–943. <https://doi.org/10.31004/jptam.v7i1.5375>
- Gunawan, M., Zarlis, M., & Roslina, R. (2021). *Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu*. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 513–523. <https://doi.org/10.30865/mib.v5i2.2925>
- Haditsah, A. (2018). *Klasifikasi Masyarakat Miskin menggunakan Metode Naïve Bayes*. *ILKOM Jurnal Ilmiah*, 10(2), 160–165.
- Kemenkes RI. (2018). *Kurikulum Inti Pendidikan Diploma III Teknologi Bank Darah*. Kementerian Kesehatan RI. 2019.
- Khasanah, N., Salim, A., Afni, N., Komarudin, R., & Maulana, Y. I. (2022). *Prediksi Kelulusan Mahasiswa Dengan Metode Naïve Bayes*. *Technologia: Jurnal Ilmiah*, 13(3), 207. <https://doi.org/10.31602/tji.v13i3.7312>
- Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). *Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier*. *Creative Information Technology Journal*, 4(2), 151. <https://doi.org/10.24076/citec.2017v4i2.106>
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). *Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN*. *Computer Engineering, Science and System Journal*, 4(1), 78.

<https://doi.org/10.24114/cess.v4i1.11458>

- Nathan, A. J., & Scobell, A. (2012). *Model Algoritma K-nearest Neighbor untuk memprediksi kelulusan mahasiswa*. *Foreign Affairs*, 91(5), 1–9.
- Purwati, N., & Januanti, A. D. (2021). *Prediksi Tingkat Kelulusan Mahasiswa dengan Algoritma Naïve Bayes* (Vol. 2, Issue 1).
- Qisthiano, M. R., Kurniawan, T. B., Negara, E. S., & Akbar, M. (2021). *Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes*. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(3), 987–994. <https://doi.org/10.30865/mib.v5i3.3030>
- Rahayu, W. I., Prianto, C., & Novia, E. A. (2021). *Perbandingan Algoritma K-Means dan Naive Bayes untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan pada PT. Pertamina (Persero)*. *Jurnal Teknik Informatika*, 13(2), 1–8.
- Rahmatullah, S. (2019). *Prediksi Tingkat Kelulusan Tepat Waktu Dengan Metode Naïve Bayes Dan K-Nearest Neighbor*. *Jurnal Informasi Dan Komputer*, 7(1), 7–16. <https://doi.org/10.35959/jik.v7i1.118>
- Riawati, D. (2020). *Analisis Nilai Praktikum Edukator Pelayanan Darah Terhadap Nilai Ujian Akhir Semester (UAS) Komunikasi Informasi Edukasi (KIE) Dan Promosi Kesehatan*. *Avicenna : Journal of Health Research*, 3(1), 72–79.
- Riyadi, F. A., & Avianto, D. (2020). *Implementasi Metode Naive Bayes Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu Prodi Informatika (Studi Kasus : Universitas Teknologi Yogyakarta)*. 1–9. <http://eprints.uty.ac.id/4863/>
- Salmu, S., & Solichin, A. (2017). *Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta*. *Seminar Nasional Multidisiplin Ilmu (SENMI) 2017, April*, 701–709.
- Sanjaya, B. A., & Sulistyono, S. (2015). *Big Data : Inkonsistensi Data dan Solusinya*. *Seminar Nasional Teknologi Informasi Dan Multimedia*, 6–8.
- Setiyani, L., Wahidin, M., Awaludin, D., & Purwani, S. (2020). *Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review*. *Faktor Exacta*, 13(1), 35–43. <https://doi.org/10.30998/faktorexacta.v13i1.5548>

- Sigid Widodo, A. Z. M., Pandu Kusuma, A., & Dwi Puspitasari, W. (2023). *Analisis Algoritma Naive Bayes Classifier (Nbc) Pada Klasifikasi Tingkat Minat Barang Di Toko Violet Cell*. JATI (Jurnal Mahasiswa Teknik Informatika), 7(1), 87–94. <https://doi.org/10.36040/jati.v7i1.5692>
- Tangkelayuk, A. (2022). *The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naive Bayes, dan Decision Tree*. Jurnal Teknik Informatika Dan Sistem Informasi, 9(2), 1109–1119. <https://doi.org/10.35957/jatisi.v9i2.2048>
- Tangkudung, I., Dako, R. R., & Dako, A. Y. (2019). *Evaluasi website menggunakan metode ISO/IEC 25010*. Seminar Nasional Teknologi, Sains Dan Humaniora 2019 (Semana TECH 2019), 1(1), 87–107. <http://jurnal.poligon.ac.id/index.php/semantech/article/view/463>
- Widaningsih, S. (2019). *Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C.45, Naive Bayes, KNN, Dan SVM*. Jurnal Tekno Insentif, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>

LAMPIRAN

Lampiran 1. Tabel Probabilitas Atribut Asal Daerah Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(DIY Kelas 1)	$\frac{14}{176} = 0,079$
P(DIY Kelas 2)	$\frac{0}{11} = 0$
P(DIY Kelas 3)	$\frac{1}{4} = 0,25$
P(DIY Kelas 4)	$\frac{2}{6} = 0,333$
P(Luar DIY Kelas 1)	$\frac{164}{176} = 0,932$
P(Luar DIY Kelas 2)	$\frac{11}{11} = 1$
P(Luar DIY Kelas 3)	$\frac{3}{4} = 0,75$
P(Luar DIY Kelas 4)	$\frac{5}{6} = 0,833$

Lampiran 2. Tabel Probabilitas Atribut IPS 1 Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(IPS 1 \geq 3,00 Kelas 1)	$\frac{142}{176} = 0,807$
P(IPS 1 \geq 3,00 Kelas 2)	$\frac{1}{11} = 0,091$
P(IPS 1 \geq 3,00 Kelas 3)	$\frac{4}{4} = 1$
P(IPS 1 \geq 3,00 Kelas 4)	$\frac{1}{6} = 0,167$
P(IPS 1 $<$ 3,00 Kelas 1)	$\frac{34}{176} = 0,193$
P(IPS 1 $<$ 3,00 Kelas 2)	$\frac{10}{11} = 0,909$
P(IPS 1 $<$ 3,00 Kelas 3)	$\frac{0}{4} = 0$

P(X H)	Probabilitas
P(IPS 1 < 3,00 Kelas 4)	$\frac{5}{6} = 0,833$

Lampiran 3. Tabel Probabilitas Atribut IPS 2 Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(IPS 2 \geq 3,00 Kelas 1)	$\frac{102}{176} = 0,58$
P(IPS 2 \geq 3,00 Kelas 2)	$\frac{0}{11} = 0$
P(IPS 2 \geq 3,00 Kelas 3)	$\frac{1}{4} = 0,25$
P(IPS 2 \geq 3,00 Kelas 4)	$\frac{0}{6} = 0$
P(IPS 2 < 3,00 Kelas 1)	$\frac{74}{176} = 0,42$
P(IPS 2 < 3,00 Kelas 2)	$\frac{11}{11} = 1$
P(IPS 2 < 3,00 Kelas 3)	$\frac{3}{4} = 0,75$
P(IPS 2 < 3,00 Kelas 4)	$\frac{6}{6} = 1$

Lampiran 4. Tabel Probabilitas Atribut IPS 3 Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(IPS 3 \geq 3,00 Kelas 1)	$\frac{153}{176} = 0,869$
P(IPS 3 \geq 3,00 Kelas 2)	$\frac{6}{11} = 0,545$
P(IPS 3 \geq 3,00 Kelas 3)	$\frac{2}{4} = 0,5$
P(IPS 3 \geq 3,00 Kelas 4)	$\frac{0}{6} = 0$
P(IPS 3 < 3,00 Kelas 1)	$\frac{23}{176} = 0,131$
P(IPS 3 < 3,00 Kelas 2)	$\frac{5}{11} = 0,455$

P(X H)	Probabilitas
P(IPS 3 < 3,00 Kelas 3)	$\frac{2}{4} = 0,5$
P(IPS 3 < 3,00 Kelas 4)	$\frac{6}{6} = 1$

Lampiran 5. Tabel Probabilitas Atribut IPS 4 Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(IPS 4 \geq 3,00 Kelas 1)	$\frac{96}{176} = 0,545$
P(IPS 4 \geq 3,00 Kelas 2)	$\frac{1}{11} = 0,091$
P(IPS 4 \geq 3,00 Kelas 3)	$\frac{1}{4} = 0,25$
P(IPS 4 \geq 3,00 Kelas 4)	$\frac{0}{6} = 0$
P(IPS 4 < 3,00 Kelas 1)	$\frac{80}{176} = 0,455$
P(IPS 4 < 3,00 Kelas 2)	$\frac{10}{11} = 0,909$
P(IPS 4 < 3,00 Kelas 3)	$\frac{3}{4} = 0,75$
P(IPS 4 < 3,00 Kelas 4)	$\frac{6}{6} = 1$

Lampiran 6. Tabel Probabilitas Atribut Serologi Golongan Darah I Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Serologi Golongan Darah I = A Kelas 1)	$\frac{22}{176} = 0,125$
P(Serologi Golongan Darah I = A Kelas 2)	$\frac{1}{11} = 0,091$
P(Serologi Golongan Darah I = A Kelas 3)	$\frac{0}{4} = 0$
P(Serologi Golongan Darah I = A Kelas 4)	$\frac{0}{6} = 0$

P(X H)	Probabilitas
P(Serologi Golongan Darah I = B Kelas 1)	$\frac{127}{176} = 0,722$
P(Serologi Golongan Darah I = B Kelas 2)	$\frac{2}{11} = 0,182$
P(Serologi Golongan Darah I = B Kelas 3)	$\frac{1}{4} = 0,25$
P(Serologi Golongan Darah I = B Kelas 4)	$\frac{0}{6} = 0$
P(Serologi Golongan Darah I = C Kelas 1)	$\frac{27}{176} = 0,534$
P(Serologi Golongan Darah I = C Kelas 2)	$\frac{8}{11} = 0,727$
P(Serologi Golongan Darah I = C Kelas 3)	$\frac{3}{4} = 0,75$
P(Serologi Golongan Darah I = C Kelas 4)	$\frac{6}{6} = 1$

Lampiran 7. Tabel Probabilitas Atribut Infeksi Menular Lewat Transfusi Darah I Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Infeksi Menular Lewat Transfusi Darah I = A Kelas 1)	$\frac{35}{176} = 0,199$
P(Infeksi Menular Lewat Transfusi Darah I = A Kelas 2)	$\frac{0}{11} = 0$
P(Infeksi Menular Lewat Transfusi Darah I = A Kelas 3)	$\frac{0}{4} = 0$
P(Infeksi Menular Lewat Transfusi Darah I = A Kelas 4)	$\frac{0}{6} = 0$
P(Infeksi Menular Lewat Transfusi Darah I = B Kelas 1)	$\frac{124}{176} = 0,705$
P(Infeksi Menular Lewat Transfusi Darah I = B Kelas 2)	$\frac{4}{11} = 0,367$
P(Infeksi Menular Lewat Transfusi Darah I = B Kelas 3)	$\frac{2}{4} = 0,5$
P(Infeksi Menular Lewat Transfusi Darah I = B Kelas 4)	$\frac{2}{6} = 0,333$
P(Infeksi Menular Lewat Transfusi Darah I = C Kelas 1)	$\frac{17}{176} = 0,097$

P(X H)	Probabilitas
P(Infeksi Menular Lewat Transfusi Darah I = C Kelas 2)	$\frac{7}{11} = 0,636$
P(Infeksi Menular Lewat Transfusi Darah I = C Kelas 3)	$\frac{2}{4} = 0,5$
P(Infeksi Menular Lewat Transfusi Darah I = C Kelas 4)	$\frac{4}{6} = 0,667$

Lampiran 8. Tabel Probabilitas Atribut Serologi Golongan Darah II Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Serologi Golongan Darah II = A Kelas 1)	$\frac{26}{176} = 0,148$
P(Serologi Golongan Darah II = A Kelas 2)	$\frac{0}{11} = 0$
P(Serologi Golongan Darah II = A Kelas 3)	$\frac{0}{4} = 0$
P(Serologi Golongan Darah II = A Kelas 4)	$\frac{0}{6} = 0$
P(Serologi Golongan Darah II = B Kelas 1)	$\frac{134}{176} = 0,761$
P(Serologi Golongan Darah II = B Kelas 2)	$\frac{6}{11} = 0,545$
P(Serologi Golongan Darah II = B Kelas 3)	$\frac{3}{4} = 0,75$
P(Serologi Golongan Darah II = B Kelas 4)	$\frac{2}{6} = 0,333$
P(Serologi Golongan Darah II = C Kelas 1)	$\frac{16}{176} = 0,091$
P(Serologi Golongan Darah II = C Kelas 2)	$\frac{5}{11} = 0,455$
P(Serologi Golongan Darah II = C Kelas 3)	$\frac{1}{4} = 0,25$
P(Serologi Golongan Darah II = C Kelas 4)	$\frac{4}{6} = 0,667$

Lampiran 9. Tabel Probabilitas Atribut Infeksi Menular Lewat Transfusi Darah II Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Infeksi Menular Lewat Transfusi Darah II = A Kelas 1)	$\frac{31}{176} = 0,176$
P(Infeksi Menular Lewat Transfusi Darah II = A Kelas 2)	$\frac{0}{11} = 0$
P(Infeksi Menular Lewat Transfusi Darah II = A Kelas 3)	$\frac{1}{4} = 0,25$
P(Infeksi Menular Lewat Transfusi Darah II = A Kelas 4)	$\frac{0}{6} = 0$
P(Infeksi Menular Lewat Transfusi Darah II = B Kelas 1)	$\frac{129}{176} = 0,733$
P(Infeksi Menular Lewat Transfusi Darah II = B Kelas 2)	$\frac{2}{11} = 0,182$
P(Infeksi Menular Lewat Transfusi Darah II = B Kelas 3)	$\frac{2}{4} = 0,5$
P(Infeksi Menular Lewat Transfusi Darah II = B Kelas 4)	$\frac{5}{6} = 0,833$
P(Infeksi Menular Lewat Transfusi Darah II = C Kelas 1)	$\frac{16}{176} = 0,091$
P(Infeksi Menular Lewat Transfusi Darah II = C Kelas 2)	$\frac{9}{11} = 0,818$
P(Infeksi Menular Lewat Transfusi Darah II = C Kelas 3)	$\frac{1}{4} = 0,25$
P(Infeksi Menular Lewat Transfusi Darah II = C Kelas 4)	$\frac{1}{6} = 0,167$

Lampiran 10. Tabel Probabilitas Atribut Penyadapan Darah Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Penyadapan Darah = A Kelas 1)	$\frac{132}{176} = 0,75$
P(Penyadapan Darah = A Kelas 2)	$\frac{5}{11} = 0,455$
P(Penyadapan Darah = A Kelas 3)	$\frac{3}{4} = 0,75$
P(Penyadapan Darah = A Kelas 4)	$\frac{0}{6} = 0$

P(X H)	Probabilitas
P(Penyadapan Darah = B Kelas 1)	$\frac{41}{176} = 0,233$
P(Penyadapan Darah = B Kelas 2)	$\frac{5}{11} = 0,455$
P(Penyadapan Darah = B Kelas 3)	$\frac{0}{4} = 0$
P(Penyadapan Darah = B Kelas 4)	$\frac{2}{6} = 0,333$
P(Penyadapan Darah = C Kelas 1)	$\frac{3}{176} = 0,017$
P(Penyadapan Darah = C Kelas 2)	$\frac{1}{11} = 0,091$
P(Penyadapan Darah = C Kelas 3)	$\frac{1}{4} = 0,25$
P(Penyadapan Darah = C Kelas 4)	$\frac{4}{6} = 0,667$

Lampiran 11. Tabel Probabilitas Atribut Serologi Golongan Darah III Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Serologi Golongan Darah III = A Kelas 1)	$\frac{37}{176} = 0,21$
P(Serologi Golongan Darah III = A Kelas 2)	$\frac{1}{11} = 0,091$
P(Serologi Golongan Darah III = A Kelas 3)	$\frac{1}{4} = 0,25$
P(Serologi Golongan Darah III = A Kelas 4)	$\frac{0}{6} = 0$
P(Serologi Golongan Darah III = B Kelas 1)	$\frac{116}{176} = 0,659$
P(Serologi Golongan Darah III = B Kelas 2)	$\frac{6}{11} = 0,545$
P(Serologi Golongan Darah III = B Kelas 3)	$\frac{2}{4} = 0,5$
P(Serologi Golongan Darah III = B Kelas 4)	$\frac{2}{6} = 0,333$
P(Serologi Golongan Darah III = C Kelas 1)	$\frac{23}{176} = 0,131$

P(X H)	Probabilitas
P(Serologi Golongan Darah III = C Kelas 2)	$\frac{4}{11} = 0,364$
P(Serologi Golongan Darah III = C Kelas 3)	$\frac{1}{4} = 0,25$
P(Serologi Golongan Darah III = C Kelas 4)	$\frac{4}{6} = 0,667$

Lampiran 12. Tabel Probabilitas Atribut Infeksi Menular Lewat Transfusi Darah III Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Infeksi Menular Lewat Transfusi Darah III = A Kelas 1)	$\frac{49}{176} = 0,278$
P(Infeksi Menular Lewat Transfusi Darah III = A Kelas 2)	$\frac{0}{11} = 0$
P(Infeksi Menular Lewat Transfusi Darah III = A Kelas 3)	$\frac{0}{4} = 0$
P(Infeksi Menular Lewat Transfusi Darah III = A Kelas 4)	$\frac{0}{6} = 0$
P(Infeksi Menular Lewat Transfusi Darah III = B Kelas 1)	$\frac{102}{176} = 0,579$
P(Infeksi Menular Lewat Transfusi Darah III = B Kelas 2)	$\frac{4}{11} = 0,364$
P(Infeksi Menular Lewat Transfusi Darah III = B Kelas 3)	$\frac{4}{4} = 1$
P(Infeksi Menular Lewat Transfusi Darah III = B Kelas 4)	$\frac{2}{6} = 0,333$
P(Infeksi Menular Lewat Transfusi Darah III = C Kelas 1)	$\frac{25}{176} = 0,142$
P(Infeksi Menular Lewat Transfusi Darah III = C Kelas 2)	$\frac{7}{11} = 0,636$
P(Infeksi Menular Lewat Transfusi Darah III = C Kelas 3)	$\frac{0}{4} = 0$
P(Infeksi Menular Lewat Transfusi Darah III = C Kelas 4)	$\frac{4}{6} = 0,667$

Lampiran 13. Tabel Probabilitas Atribut Komponen Darah Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Komponen Darah = A Kelas 1)	$\frac{57}{176} = 0,324$
P(Komponen Darah = A Kelas 2)	$\frac{1}{11} = 0,091$
P(Komponen Darah = A Kelas 3)	$\frac{1}{4} = 0,25$
P(Komponen Darah = A Kelas 4)	$\frac{0}{6} = 0$
P(Komponen Darah = B Kelas 1)	$\frac{102}{176} = 0,579$
P(Komponen Darah = B Kelas 2)	$\frac{6}{11} = 0,545$
P(Komponen Darah = B Kelas 3)	$\frac{2}{4} = 0,5$
P(Komponen Darah = B Kelas 4)	$\frac{1}{6} = 0,167$
P(Komponen Darah = C Kelas 1)	$\frac{17}{176} = 0,097$
P(Komponen Darah = C Kelas 2)	$\frac{4}{11} = 0,364$
P(Komponen Darah = C Kelas 3)	$\frac{1}{4} = 0,25$
P(Komponen Darah = C Kelas 4)	$\frac{5}{6} = 0,833$

Lampiran 14. Tabel Probabilitas Atribut Serologi Golongan Darah IV Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Serologi Golongan Darah IV = A Kelas 1)	$\frac{42}{176} = 0,239$
P(Serologi Golongan Darah IV = A Kelas 2)	$\frac{0}{11} = 0$
P(Serologi Golongan Darah IV = A Kelas 3)	$\frac{0}{4} = 0$
P(Serologi Golongan Darah IV = A Kelas 4)	$\frac{0}{6} = 0$

P(X H)	Probabilitas
P(Serologi Golongan Darah IV = B Kelas 1)	$\frac{93}{176} = 0,528$
P(Serologi Golongan Darah IV = B Kelas 2)	$\frac{6}{11} = 0,545$
P(Serologi Golongan Darah IV = B Kelas 3)	$\frac{3}{4} = 0,75$
P(Serologi Golongan Darah IV = B Kelas 4)	$\frac{0}{6} = 0$
P(Serologi Golongan Darah IV = C Kelas 1)	$\frac{41}{176} = 0,233$
P(Serologi Golongan Darah IV = C Kelas 2)	$\frac{5}{11} = 0,455$
P(Serologi Golongan Darah IV = C Kelas 3)	$\frac{1}{4} = 0,25$
P(Serologi Golongan Darah IV = C Kelas 4)	$\frac{6}{6} = 1$

Lampiran 15. Tabel Probabilitas Atribut Infeksi Menular Lewat Transfusi Darah IV Terhadap Setiap Label Kelas

P(X H)	Probabilitas
P(Infeksi Menular Lewat Transfusi Darah IV = A Kelas 1)	$\frac{123}{176} = 0,699$
P(Infeksi Menular Lewat Transfusi Darah IV = A Kelas 2)	$\frac{4}{11} = 0,364$
P(Infeksi Menular Lewat Transfusi Darah IV = A Kelas 3)	$\frac{3}{4} = 0,75$
P(Infeksi Menular Lewat Transfusi Darah IV = A Kelas 4)	$\frac{2}{6} = 0,333$
P(Infeksi Menular Lewat Transfusi Darah IV = B Kelas 1)	$\frac{45}{176} = 0,256$
P(Infeksi Menular Lewat Transfusi Darah IV = B Kelas 2)	$\frac{5}{11} = 0,455$
P(Infeksi Menular Lewat Transfusi Darah IV = B Kelas 3)	$\frac{1}{4} = 0,25$
P(Infeksi Menular Lewat Transfusi Darah IV = B Kelas 4)	$\frac{3}{6} = 0,5$
P(Infeksi Menular Lewat Transfusi Darah IV = C Kelas 1)	$\frac{8}{176} = 0,045$

P(X H)	Probabilitas
P(Infeksi Menular Lewat Transfusi Darah IV = C Kelas 2)	$\frac{2}{11} = 0,182$
P(Infeksi Menular Lewat Transfusi Darah IV = C Kelas 3)	$\frac{0}{4} = 0$
P(Infeksi Menular Lewat Transfusi Darah IV = C Kelas 4)	$\frac{1}{6} = 0,167$

Lampiran 16. Kode Program *login.html*

```

<!DOCTYPE html>
<html>
<head>
  <title>Login</title>
  <link rel="stylesheet" type="text/css" href="{{
url_for('static', filename='style_1.css') }}">
</head>
<body>
  <div class="container">
    <h2>Halaman Login</h2>
    {% if error %}
      <p style="color: red;">{{ error }}</p>
    {% endif %}
    <form method="post" action="/">
      <input type="text" name="username"
placeholder="Username" required><br>
      <input type="password" name="password"
placeholder="Password" required><br>
      <input type="submit" value="Login">
    </form>
  </div>
</body>
</html>

```

Lampiran 17. Kode Program *index.html*

```

<!DOCTYPE html>
<html>
  <head>
    <title>Home</title>
    <link rel="stylesheet" type="text/css" href="{{
url_for('static', filename='style_2.css') }}">
  </head>
  <body>
    <div class="container">
      <div class="header">
        <h2>APLIKASI PREDIKSI KELULUSAN MAHASISWA
TEKNOLOGI BANK DARAH</h2>

```

```

        </div>
        <hr>
        <form method="post">
            <div class="form-group">
                <label for="JENIS KELAMIN">Jenis
Kelamin</label>
                <select name="JENIS KELAMIN" required>
                    <option value="" disabled selected>--
Pilih Jenis Kelamin --</option>
                    <option value="1">Laki-laki</option>
                    <option value="2">Perempuan</option>
                </select>
            </div>
            <div class="form-group">
                <label for="ASAL DAERAH">Asal
Daerah</label>
                <select name="ASAL DAERAH" required>
                    <option value="" disabled selected>--
Pilih Asal Daerah --</option>
                    <option value="1">Yogyakarta</option>
                    <option value="2">Luar
Yogyakarta</option>
                </select>
            </div>
            <div class="form-group">
                <label for="IPS_1">Nilai Indeks Prestasi
Semester 1</label>
                <input type="text" name="IPS_1"
placeholder="Skala 0.00 sampai 4.00" required>
            </div>
            <div class="form-group">
                <label for="IPS_2">Nilai Indeks Prestasi
Semester 2</label>
                <input type="text" name="IPS_2"
placeholder="Skala 0.00 sampai 4.00" required>
            </div>
            <div class="form-group">
                <label for="IPS_3">Nilai Indeks Prestasi
Semester 3</label>
                <input type="text" name="IPS_3"
placeholder="Skala 0.00 sampai 4.00" required>
            </div>
            <div class="form-group">
                <label for="IPS_4">Nilai Indeks Prestasi
Semester 4</label>
                <input type="text" name="IPS_4"
placeholder="Skala 0.00 sampai 4.00" required>
            </div>
            <div class="form-group">
                <label for="SEROLOGI GOLONGAN DARAH
I">Nilai Serologi Golongan Darah I</label>
                <select name="SEROLOGI GOLONGAN DARAH I"
required>
                    <option value="" disabled selected>--
Pilih Nilai --</option>

```

```

        <option value="4">A</option>
        <option value="3">B</option>
        <option value="2">C</option>
        <option value="1">D</option>
        <option value="0">E</option>
    </select>
</div>
<div class="form-group">
    <label for="INFEKSI MENULAR LEWAT TRANSFUSI
DARAH II">Nilai Infeksi Menular Lewat Transfusi Darah
II</label>
        <select name="INFEKSI MENULAR LEWAT
TRANSFUSI DARAH II" required>
            <option value="" disabled selected>--
Pilih Nilai --</option>
                <option value="4">A</option>
                <option value="3">B</option>
                <option value="2">C</option>
                <option value="1">D</option>
                <option value="0">E</option>
            </select>
</div>
<div class="form-group">
    <label for="INFEKSI MENULAR LEWAT TRANSFUSI
DARAH III">Nilai Infeksi Menular Lewat Transfusi Darah
III</label>
        <select name="INFEKSI MENULAR LEWAT
TRANSFUSI DARAH III" required>
            <option value="" disabled selected>--
Pilih Nilai --</option>
                <option value="4">A</option>
                <option value="3">B</option>
                <option value="2">C</option>
                <option value="1">D</option>
                <option value="0">E</option>
            </select>
</div>
<div class="form-group">
    <label for="INFEKSI MENULAR LEWAT TRANSFUSI
DARAH IV">Nilai Infeksi Menular Lewat Transfusi Darah
IV</label>
        <select name="INFEKSI MENULAR LEWAT
TRANSFUSI DARAH IV" required>
            <option value="" disabled selected>--
Pilih Nilai --</option>
                <option value="4">A</option>
                <option value="3">B</option>
                <option value="2">C</option>
                <option value="1">D</option>
                <option value="0">E</option>
            </select>
</div>
<div class="form-group">
    <label for="PENYADAPAN DARAH">Nilai
Penyadapan Darah</label>

```

```

<select name="PENYADAPAN DARAH" required>
  <option value="" disabled selected>--
Pilih Nilai --</option>
  <option value="4">A</option>
  <option value="3">B</option>
  <option value="2">C</option>
  <option value="1">D</option>
  <option value="0">E</option>
</select>
</div>
<div class="form-group">
  <label for="KOMPONEN DARAH">Nilai Komponen
Darah</label>
  <select name="KOMPONEN DARAH" required>
  <option value="" disabled selected>--
Pilih Nilai --</option>
    <option value="4">A</option>
    <option value="3">B</option>
    <option value="2">C</option>
    <option value="1">D</option>
    <option value="0">E</option>
  </select>
</div>
<div class="button-group">
  <button type="button" class="back-button"
onclick="location.href='/'">Keluar</button>
  <div class="prediction-buttons">
    <button type="submit"
formaction="/predict_nbc">Prediksi NBC</button>
  </div>
</div>
</form>
</div>
</body>
</html>

```

Lampiran 18. Kode Program *hasil.html*

```

<!DOCTYPE html>
<html>
<head>
  <title>Hasil Prediksi</title>
  <link rel="stylesheet" type="text/css" href="{{
    url_for('static', filename='style_3.css') }}">
</head>
<body>
  <div class="result-container">
    <h2>HASIL PREDIKSI KELULUSAN MAHASISWA</h2>
    <hr>
    <p>{{ hasil }}</p>
    <p>Akurasi model : {{ akurasi }}</p>
    <div class="button-container">
      <a href="/home"><< Lakukan Prediksi Lagi</a>

```

```

        <a href="{{ matrix_route }}">Lihat Confusion Matrix
        >></a>
    </div>
</div>
</body>
</html>

```

Lampiran 19. Kode Program *confusion_matrix.html*

```

<!DOCTYPE html>
<html>
<head>
    <title>Confusion Matrix</title>
    <link rel="stylesheet" type="text/css" href="{{
url_for('static', filename='style_4.css') }}">
</head>
<body>
    <div class="result-container">
        <h2>Confusion Matrix</h2>
        <table border="2">
            <tr>
                <td></td>
                <td>Prediksi Kelas 1</td>
                <td>Prediksi Kelas 2</td>
                <td>Prediksi Kelas 3</td>
                <td>Prediksi Kelas 4</td>
            </tr>
            {% for i in range(confusion_matrix.shape[0]) %}
            <tr>
                <td>Kelas Aktual {{ i+1 }}</td>
                {% for j in range(confusion_matrix.shape[1]) %}
                <td>{{ confusion_matrix[i][j] }}</td>
                {% endfor %}
            </tr>
            {% endfor %}
        </table>
        <a href="/home">Kembali ke Home</a>
    </div>
</body>
</html>

```

Lampiran 20. Kode Program *style_1.css*

```

body {
    background-color: #f2f2f2;
}

.container {
    width: 30%;
    text-align: center;
    padding: 20px 25px;
}

```

```

margin: auto;
margin-top: 50px;
background-color: #e0f7e0;
box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
border: 5px solid #333;
border-radius: 15px;
}

input[type="text"], input[type="password"],
input[type="email"], select {
width: 100%;
padding: 10px;
margin: 5px 0;
box-sizing: border-box;
border: 1px solid #ccc;
}

input[type="submit"] {
width: 25%;
padding: 10px;
margin: 10px 0;
box-sizing: border-box;
background-color: #4CAF50;
color: white;
border: none;
cursor: pointer;
}

a {
text-decoration: none;
color: blue;
margin: 10px;
}

h2, h3 {
color: #333;
}

.error {
color: red;
}

```

Lampiran 21. Kode Program *style_2.css*

```

body {
font-family: Arial, sans-serif;
background-color: #f0f0f0;
margin: 0;
padding: 0;
}

.container {
width: 90%;
}

```

```
text-align: left;
padding: 15px;
margin: 20px auto;
background:
  linear-gradient(to bottom, rgba(173, 216, 230, 0.7) 0%,
    rgba(173, 216, 230, 0.7) 15%, rgba(255, 250, 205, 0.7) 50%);
border: 5px solid #333;
box-shadow: 0 0 50px rgba(0, 0, 0, 0.897);
border-radius: 25px;
}

.header {
text-align: center;
margin-bottom: 20px;
}

h2 {
text-align: center;
margin: 10px auto;
color: #2532e7;
font-size: 25px;
}

hr {
border: none;
border-top: 5px solid black;
margin-top: -6px;
margin-bottom: 20px;
}

form {
display: flex;
flex-wrap: wrap;
justify-content: space-between;
}

.form-group {
display: flex;
align-items: center;
margin-bottom: 15px;
flex: 1 1 45%;
}

label {
flex: 1;
text-align: left;
margin-left: 20px;
font-weight: bold;
color: #555;
font-size: 14px;
}

input[type="text"],
select {
flex: 0.6;
```

```
padding: 8px;
border: 1.5px solid #ccc;
border-radius: 5px;
font-size: 14px;
text-align: center;
margin-left: 5px;
margin-right: 20px;
}

.button-group {
display: flex;
justify-content: space-between;
align-items: center;
margin-top: 15px;
width: 100%;
}

.back-button,
.prediction-buttons button {
padding: 10px;
font-size: 14px;
border: none;
border-radius: 5px;
cursor: pointer;
transition: background-color 0.3s ease;
}

.back-button {
width: 15%;
background-color: #dc3545;
color: #ffffff;
margin-left: 400px;
}

.back-button:hover {
background-color: #c82333;
}

.prediction-buttons {
display: flex;
gap: 3px;
justify-content: flex-end;
width: 100%;
}

.prediction-buttons button {
background-color: #007bff;
color: #ffffff;
margin-right: 400px;
}

.prediction-buttons button:hover {
background-color: #0056b3;
}
```

Lampiran 22. Kode Program *style_3.css*

```
body {
  font-family: Arial, sans-serif;
  background-color: #f0f0f0;
  margin: 0;
  padding: 0;
}

.result-container {
  width: 45%;
  margin: auto;
  margin-top: 70px;
  padding: 20px;
  background-color: #e0f7fa;
  border-radius: 15px;
  border: 5px solid #333;
  border-bottom-width: 7px;
}

h2 {
  text-align: center;
  margin-bottom: 20px;
  color: #e72525;
}

hr {
  border: none;
  border-top: 3px solid black;
}

p {
  font-size: 18px;
  text-align: center;
}

.button-container {
  display: flex;
  justify-content: center;
  gap: 50px;
  margin-top: 30px;
  margin-bottom: 20px;
}

a {
  display: inline-block;
  width: auto;
  padding: 10px 20px;
  background-color: #007bff;
  color: white;
  text-align: center;
  text-decoration: none;
  border-radius: 8px;
}

a:hover {
```

```
background-color: #0056b3;  
}
```

Lampiran 23. Kode Program *style_4.css*

```
body {  
    font-family: Arial, sans-serif;  
    background-color: #f0f0f0;  
    margin: 0;  
    padding: 0;  
}  
  
.result-container {  
    width: 45%;  
    margin: auto;  
    margin-top: 50px;  
    padding: 20px;  
    background-color: #f5fffa;  
    border-radius: 25px;  
    border: 5px solid #333;  
    border-bottom-width: 7px;  
}  
  
h2 {  
    text-align: center;  
    margin-bottom: 20px;  
    color: #2532e7;  
}  
  
hr {  
    border: none;  
    border-top: 3px solid black;  
}  
  
p {  
    font-size: 18px;  
    text-align: center;  
}  
  
a {  
    display: block;  
    width: 25%;  
    margin: auto;  
    margin-top: 30px;  
    margin-bottom: 20px;  
    padding: 10px;  
    background-color: #007BFF;  
    color: white;  
    text-align: center;  
    text-decoration: none;  
    border-radius: 8px;  
}
```

```

a:hover {
    background-color: #0056b3;
}

table {
    width: 90%;
    margin: 20px auto;
    border-collapse: collapse;
    text-align: center;
}

th, td {
    padding: 5px;
    border: 1px solid #333;
}

th {
    background-color: #007BFF;
    color: white;
}

```

Lampiran 24. Kode Program *app.py*

```

from flask import Flask, render_template, request, redirect, url_for
from flask_mysql import MySQL
from nbc import predict as predict_nbc, get_accuracy as get_accuracy_nbc, get_y_true as get_y_true_nbc, get_y_pred as get_y_pred_nbc
from sklearn.metrics import confusion_matrix
import numpy as np

app = Flask(__name__)

# Konfigurasi database
app.config['MYSQL_HOST'] = 'localhost'
app.config['MYSQL_USER'] = 'root'
app.config['MYSQL_PASSWORD'] = ''
app.config['MYSQL_DB'] = 'db_users'

mysql = MySQL(app)

def ensure_confusion_matrix_size(cm, num_classes=4):
    if cm.shape[0] < num_classes:
        new_cm = np.zeros((num_classes, num_classes), dtype=int)
        new_cm[:cm.shape[0], :cm.shape[1]] = cm
        return new_cm
    return cm

# Routes
@app.route('/', methods=['GET', 'POST'])
def login():

```

```

error = None
if request.method == "POST":
    username = request.form["username"]
    password = request.form["password"]

    cur = mysql.connection.cursor()
    cur.execute("SELECT * FROM users WHERE username = %s
AND password = %s", (username, password))
    user = cur.fetchone()
    cur.close()

    if user:
        return redirect(url_for('home')) # Mengarahkan
pengguna ke halaman home setelah login sukses
    else:
        error = "Username atau password salah !"

    return render_template("login.html", error=error)

@app.route("/home")
def home():
    return render_template("index.html")

# Route untuk melakukan prediksi dengan metode nbc
@app.route('/predict_nbc', methods=['POST'])
def prediction_nbc():
    # Ambil data dari form input
    JENIS_KELAMIN = int(request.form['JENIS KELAMIN'])
    ASAL_DAERAH = int(request.form['ASAL DAERAH'])
    IPS_1 = float(request.form['IPS_1'])
    IPS_2 = float(request.form['IPS_2'])
    IPS_3 = float(request.form['IPS_3'])
    IPS_4 = float(request.form['IPS_4'])
    SEROLOGI_GOLONGAN_DARAH_I = int(request.form['SEROLOGI
GOLONGAN DARAH I'])
    SEROLOGI_GOLONGAN_DARAH_II = int(request.form['SEROLOGI
GOLONGAN DARAH II'])
    SEROLOGI_GOLONGAN_DARAH_III = int(request.form['SEROLOGI
GOLONGAN DARAH III'])
    SEROLOGI_GOLONGAN_DARAH_IV = int(request.form['SEROLOGI
GOLONGAN DARAH IV'])
    INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_I =
int(request.form['INFEKSI MENULAR LEWAT TRANSFUSI DARAH I'])
    INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_II =
int(request.form['INFEKSI MENULAR LEWAT TRANSFUSI DARAH II'])
    INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_III =
int(request.form['INFEKSI MENULAR LEWAT TRANSFUSI DARAH III'])
    INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_IV =
int(request.form['INFEKSI MENULAR LEWAT TRANSFUSI DARAH IV'])
    PENYADAPAN_DARAH = int(request.form['PENYADAPAN DARAH'])
    KOMPONEN_DARAH = int(request.form['KOMPONEN DARAH'])

    # Buat data test dari input form

```

```

data_test = [[JENIS_KELAMIN, ASAL_DAERAH, IPS_1, IPS_2,
IPS_3, IPS_4, SEROLOGI_GOLONGAN_DARAH_I,
SEROLOGI_GOLONGAN_DARAH_II,
SEROLOGI_GOLONGAN_DARAH_III,
SEROLOGI_GOLONGAN_DARAH_IV,
INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_I,
INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_II,
INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_III,
INFEKSI_MENULAR_LEWAT_TRANSFUSI_DARAH_IV, PENYADAPAN_DARAH,
KOMPONEN_DARAH]]

# Dapatkan prediksi dan akurasi menggunakan NBC
result = predict_nbc(data_test)
accuracy = get_accuracy_nbc()

# Format akurasi dengan dua digit di belakang koma
formatted_accuracy = f"{accuracy:.2f}"

# Render hasil prediksi dan akurasi ke hasil.html
return render_template('hasil.html', hasil=result,
akurasi=formatted_accuracy,
matrix_route=url_for('confusion_matrix_nbc'))

# Route untuk menampilkan confusion matrix NBC
@app.route('/confusion_matrix_nbc')
def confusion_matrix_nbc():
    # Mendapatkan true labels dan predicted labels dari model
    NBC
    y_true = get_y_true_nbc()
    y_pred = get_y_pred_nbc()

    # Menghitung confusion matrix menggunakan sklearn
    cm = confusion_matrix(y_true, y_pred)
    cm = ensure_confusion_matrix_size(cm, num_classes=4)

    # Render template confusion_matrix.html dengan confusion
    matrix yang telah dihitung
    return render_template('confusion_matrix.html',
confusion_matrix=cm)

if __name__ == '__main__':
    app.run(debug=True)

```

Lampiran 25. Kode Program *nbc.py*

```

import pandas as pd
from sklearn.decomposition import PCA
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer

```

```

# Load dataset dari file Excel
dataset = pd.read_excel("DATASET.xlsx", sheet_name='DATASHEET')

# Menghapus kolom yang tidak diperlukan dari dataset
df = dataset.drop(['NO', 'NIM', 'NAMA', 'PROGRAM STUDI',
'IPS_5', 'IPS_6', 'IPK LULUS', 'NILAI KARYA TULIS'], axis=1)

# Memisahkan fitur (x) dan label (y) dari dataset
x = df.iloc[:, :-1].values # Matriks fitur (semua kolom
kecuali kolom terakhir)
y = df['LABEL KELAS']      # Label (kolom 'LABEL KELAS')

# Inisialisasi objek Standard Scaler untuk normalisasi fitur
scaler = StandardScaler()

# Menerapkan standar scaler pada fitur
x_scaled = scaler.fit_transform(x)

# Mengganti nilai NaN dengan rata-rata kolom
imputer = SimpleImputer(strategy='mean')
x_imputed = imputer.fit_transform(x_scaled)

# Mengaplikasikan PCA dengan 5 komponen utama
pca = PCA(n_components=5)
pca_dataset = pca.fit_transform(x_imputed)

# Membagi dataset menjadi data pelatihan 80% dan data pengujian
20%
x_train, x_test, y_train, y_test =
train_test_split(pca_dataset, y, test_size=0.20,
random_state=42)

# Membuat objek Gaussian Naive Bayes (NBC)
nbc = GaussianNB()

# Melatih model NBC dengan data pelatihan
model = nbc.fit(x_train, y_train)

def predict(data_test):
    # Normalisasi dan menerapkan PCA pada data uji
    pca_test = pca.transform(scaler.transform(data_test))

    # Membuat prediksi menggunakan model yang sudah dilatih
    y_pred = model.predict(pca_test)

    # Mengembalikan hasil prediksi dalam bentuk deskriptif
    if y_pred == 1:
        return "Mahasiswa lulus dalam 3 tahun dengan IPK >
3,00"
    elif y_pred == 2:
        return "Mahasiswa lulus dalam 3 tahun dengan IPK <
3,00"
    elif y_pred == 3:
        return "Mahasiswa lulus lebih dari 3 tahun dengan IPK >
3,00"

```

```

        else:
            return "Mahasiswa lulus lebih dari 3 tahun dengan IPK <
3,00"

def get_accuracy():
    # Membuat prediksi pada data pengujian
    y_prediction = model.predict(x_test)

    # Menghitung dan mengembalikan akurasi model
    return accuracy_score(y_test, y_prediction)

def get_y_true():
    # Mengembalikan nilai sebenarnya dari data pengujian
    return y_test

def get_y_pred():
    # Membuat prediksi pada data pengujian
    return model.predict(x_test)

```

Lampiran 26. Database *db_users*

Server: localhost:3306 » Database: db_users » Table: users

Showing rows 0 - 0 (1 total, Query took 0.0005 seconds.)

`SELECT * FROM `users``

Number of rows: 25 | Filter rows: Search this table

	id	username	password	ulangpass	email	nama	jekel
<input type="checkbox"/>	1	admin	12345		alexsander123@gmail.com	Alex Sander	Male

With selected: Edit Copy Delete Export

Lampiran 27. Kriteria Kelulusan Ujian Sidang / Pendaran

	A	B	C	D	E	F	G	H	I	J	K
1	PEMBERITAHUAN SEBELUM UJIAN : Pengumpulan akhir dokumen Tugas Akhir/Skripsi melewati batas akhir ganjil 2022/2023, mahasiswa harus menyelesaikan registrasi dan KRS semester berikutnya.										
2	KRITERIA KELULUSAN UJIAN SIDANG / PENDARAN										
3											
4											
5											
6											
7	1. Lulus ujian tanpa syarat, disebut kriteria 1.										
8	2. Lulus bersyarat, disebut kriteria 2, yaitu dengan sedikit perbaikan atau penyempurnaan text dan atau program dalam waktu maks sampai tanggal 30 September 2024 dan tidak ada ujian lagi. Jika dalam waktu yang ditentukan mahasiswa tersebut tidak dapat menyelesaikan, maka, mahasiswa yang bersangkutan dianggap tidak lulus ujian.										
9	3. Tidak lulus ujian sidang/pendadaran, disebut kriteria 3, dijelaskan, disarankan Ketua Tim Penguji untuk mempelajari ulang materi, merombak program, teks, atau mengganti judul.										
10	Ketentuan bagi peserta yang tidak lulus ujian sidang / pendadaran.										
11	1) Mahasiswa wajib menempuh ujian sidang/pendadaran ulang										
12	2) Kesempatan ujian sidang/pendadaran ulang hanya diberikan dalam rentang waktu maksimum 6 bulan, setelah ujian sidang/pendadaran										
13	3) Jika sampai batas waktu maksimum 6 bulan tersebut belum dapat diajukan/diselesaikan, maka calon peserta ujian dinyatakan sebagai mahasiswa peserta Skripsi TGA baru, dengan segala ketentuan yang berlaku bagi peserta baru										
14	4) Mahasiswa yang akan menempuh ujian sidang/pendadaran ulang ini diwajibkan membayar biaya ujian setara 2 SKS praktik, sesuai tahun angkatan										
15											
16											
17											
18											
19	Yogyakarta, _____										
20	Memahami dan bersedia										
21	Mematuhi peraturan di atas,										
22											
23											
24											
25											
26	Nama Mahasiswa										
27	MASUNGGING DWI CAHYO										
28											
29											
30											

Lampiran 28. Keputusan Hasil Ujian Pendaran

	A	B	C	D	E	F	G	H	I	J
1	 YAYASAN PENDIDIKAN WIDYA BAKTI YOGYAKARTA UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA 									
2	Jl. Raya Janti (Majapahit) No.143, Yogyakarta, 55198, Telp (0274) 486664, Website: www.utdi.ac.id , E-mail: info@utdi.ac.id									
3	-----									
4	KEPUTUSAN HASIL UJIAN PENDADARAN									
5										
6										
7										
8										
9										
10										
11										
12	Sesuai dengan hasil sidang pendadaran pada tanggal					maka				
13										
14	Nama Mahasiswa	MASUNGGING DWI CAHYO								
15	NIM / Program Studi	205411150 / Informatika								
16	Jenjang	S1								
17										
18		dinyatakan	LULUS	dengan kriteria	2					
19										
20										
21										
22	Ketua Penguji	Ariesta Damayanti, S.Kom., M.Cs.								
23										
24										
25										

Lampiran 29. Catatan Perbaikan

	A	B	C	D	E	F	G	H	I	J
1	 <p style="text-align: center;">YAYASAN PENDIDIKAN WIDYA BAKTI YOGYAKARTA UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA Jl. Raya Janti (Majapahit) No.143, Yogyakarta, 55198, Telp (0274) 486664, Website: www.utdi.ac.id , E-mail: info@utdi.ac.id</p> 									
2										
3										
4										
5										
6										
7										
8										
9										
10										
11	Hari, tanggal		: Rabu, 31 Juli 2024							
12	Waktu		: 13.00							
13	Nama		: MASUNGGING DWI CAHYO							
14	No. Mahasiswa / Prodi		: 205411150 / Informatika							
15		No	Hal yang harus diperbaiki						Pemberi Catatan	
16		1.	<p>Perkuat alasan menggunakan naive bayes dengan menambahkan penelitian2 sebelumnya di latar belakang.</p> <p>ruang lingkup poin 2 dan 7 dijadikan satu. bab 3 berisi metode dan rancangan penelitian yang dilakukan, sehingga tidak berisi teori2. teori ditulis di bab 2. penjelasan tentang notasi2 use case diagram, dan sequence diagram dihapus.</p>							
17	2.									
18	3.									
19	4.									
20										
21										
22										
23										