

BAB III

METODELOGI PENELITIAN

3.1. Waktu dan Tempat Penelitian

3.1.1 Waktu Penelitian

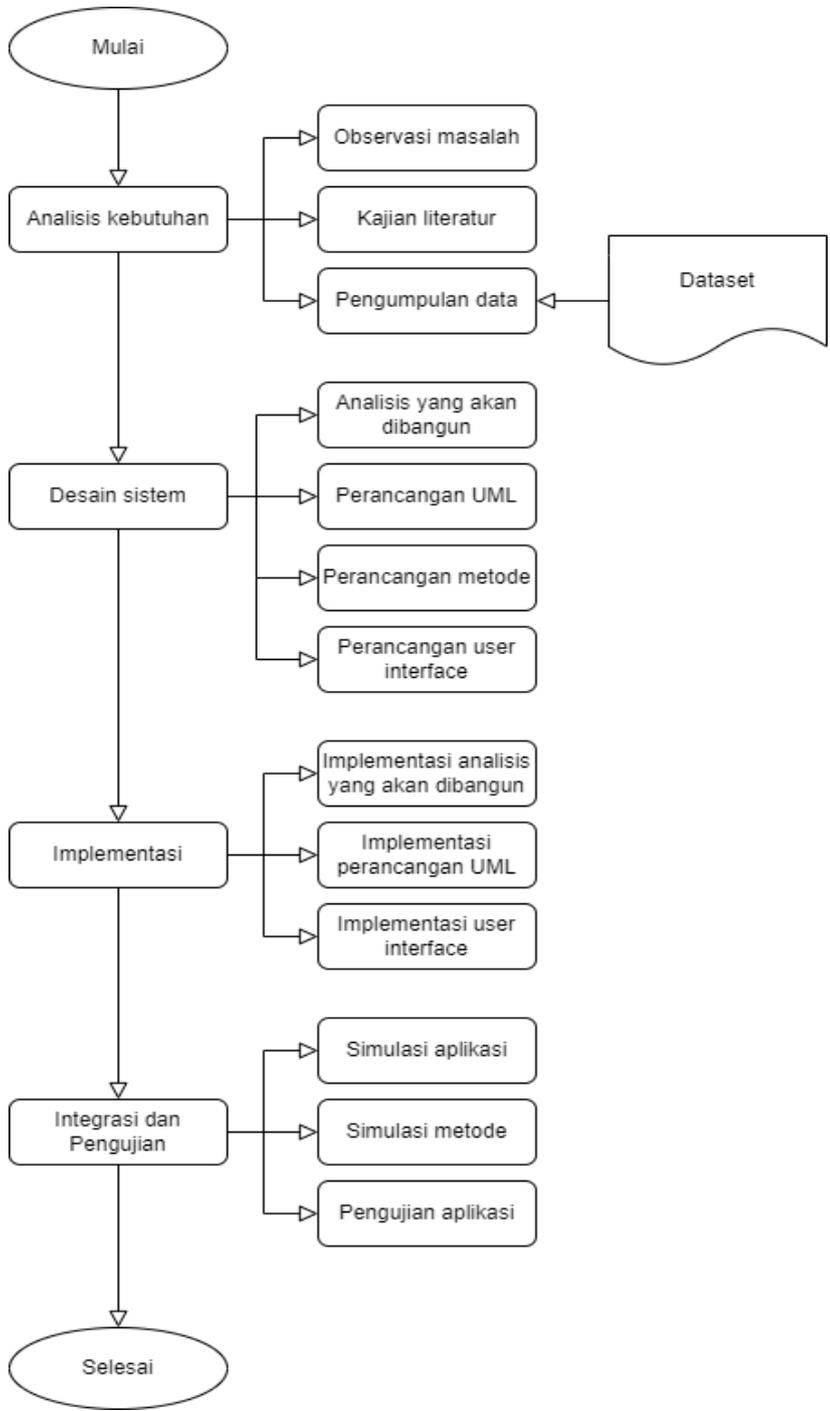
Penelitian proposal tesis ini dilakukan dari bulan Oktober 2023 sampai dengan Maret 2024 dan akan selesai pada bulan Agustus 2024 guna memperoleh data dan informasi yang akurat berhubungan dengan tema penulisan proposal tesis.

3.1.2 Tempat Penelitian

Penelitian dilakukan di Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Surya Intan Kotabumi yang berlokasi di Jl. Ibrahim Syarif No. 107 Kotabumi Lampung Utara. Sekolah Tinggi Manajemen Informatika dan Komputer Surya Intan Kotabumi yang berkedudukan di Kotabumi didirikan pada tanggal 10 Juli 2000 oleh Yayasan Puncak Surya Intan. Berdasarkan surat keputusan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 1671D10/2000 Tentang Status Terdaftar Sekolah Tinggi Manajemen Informatika dan Komputer. Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Surya Intan Kotabumi Memiliki Program Pendidikan Strata Satu (S1) yaitu Sistem Informasi (SI) dan Teknik Informatika (TI), dan Program Pendidikan Diploma Tiga (D3) yaitu Manajemen Informatika (MI).

3.2 Metode Penelitian

Penelitian ini dilakukan secara sistematis, berikut ini langkah-langkahnya :



Gambar 3.1 Metodologi Penelitian

3.3 Analisis Kebutuhan Yang Digunakan

Dalam penelitian ini teknik data mining direkomendasikan untuk digunakan dalam evaluasi tanggal kelulusan siswa.

A. Observasi.

Pada titik ini, penulis harus menilai kesulitannya berpartisipasi dalam identifikasi masalah, menetapkan tujuan dan memanfaatkannya penelitian mereka. Ini dilakukan secara bertahap mengidentifikasi masalah dengan memeriksa situasi dan situasi saat ini adalah fokus utama masalah dalam penelitian ini. Mengingat masalah yang dicapai oleh penelitian ini dan hal-hal yang dapat diperoleh para ilmuwan dan orang-orang jika tidak, itu dilakukan dalam fase target penelitian dan keunggulan ilmiah.

B. Kajian literatur.

Tinjauan literatur yang sesuai akan dilakukan hal ini berkaitan dengan topik penelitian.

C. Pengumpulan data.

Informasi tentang siswa dikumpulkan studi ini langsung dari program studi S-1 teknik informatika.

3.4 Desain Sistem

Perancangan sistem dalam penelitian ini, tidak hanya tampilannya saja, tapi apa arti desain dalam Metode ini merupakan rancangan sistematis yang meliputi:

A. Analisis proses yang akan dibangun

Gambaran analisis tersebut alur yang nantinya akan dimasukkan dalam aplikasi prediksi kelulusan Mahasiswa. Nilai akurasi algoritma K-Nearest Neighbor dan Naive Bayes juga dapat dibandingkan dalam Aplikasi ini.

B. Perancangan Menggunakan RapidMiner

Proses merancang dan membangun alur kerja (workflow) analisis data menggunakan platform RapidMiner. Dalam konteks ini, "perancangan" mengacu pada langkah-langkah yang diambil untuk mengatur, memproses, dan menganalisis data menggunakan alat-alat dan fungsi yang tersedia dalam RapidMiner.

C. Perancangan Algoritma K-NN dan Naive Bayes

Ada metode, K-NN dan Naive Bayes, dalam aplikasi yang akan dibangun untuk memecahkan permasalahan yang telah dianalisis dalam sistem yang disarankan.

3.5 Proses Implementasi

Pendekatan ini melibatkan simulasi algoritma KNN dan Naive Bayes untuk menentukan apakah metode yang digunakan dapat diimplementasikan pada aplikasi yang akan dibangun. Teknik ini diimplementasikan dengan mengintegrasikan metode-algoritma yang digunakan pada aplikasi.

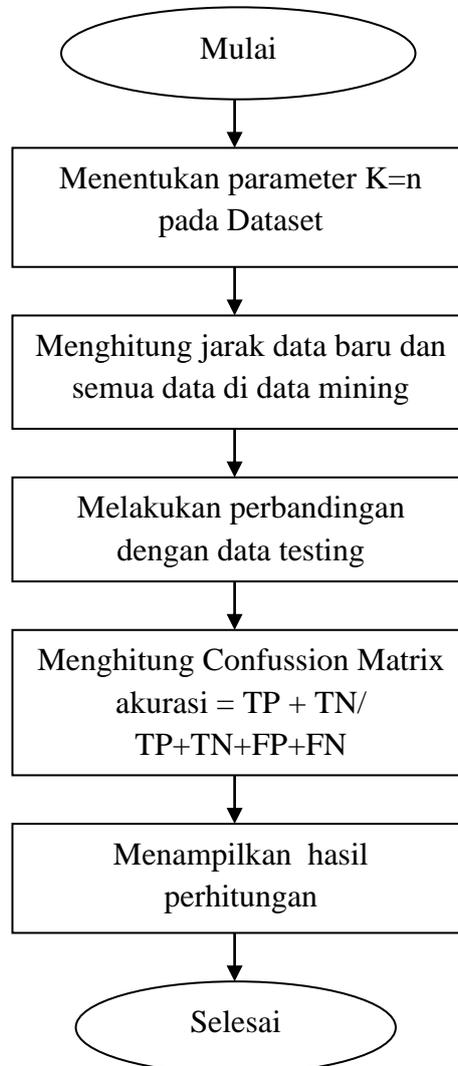
3.6 Integrasi dan Pengujian

Setelah semua modul dibuat dan diuji pada tahap implementasi, sistem integrasikan menjadi satu kesatuan pada tahap ini. Setelah menyelesaikan proses integrasi, sistem diperiksa dan diuji untuk kemungkinan masalah dan kesalahan sistem.

3.7 Penerapan Metode Yang Digunakan

Langkah ini terdiri dari simulasi algoritma digunakan dalam penelitian untuk menentukan apakah metode tersebut dapat diterapkan benar atau salah Algoritma yang digunakan adalah K-NN dan *Naive Bayes* diikuti oleh klasifikasi data dan perbandingan nilai akurasi menggunakan algoritma K-NN dan *Naive Bayes*, dengan menggunakan variable IPS, Waktu kelulusan, Jenis kelamin dan Absensi.

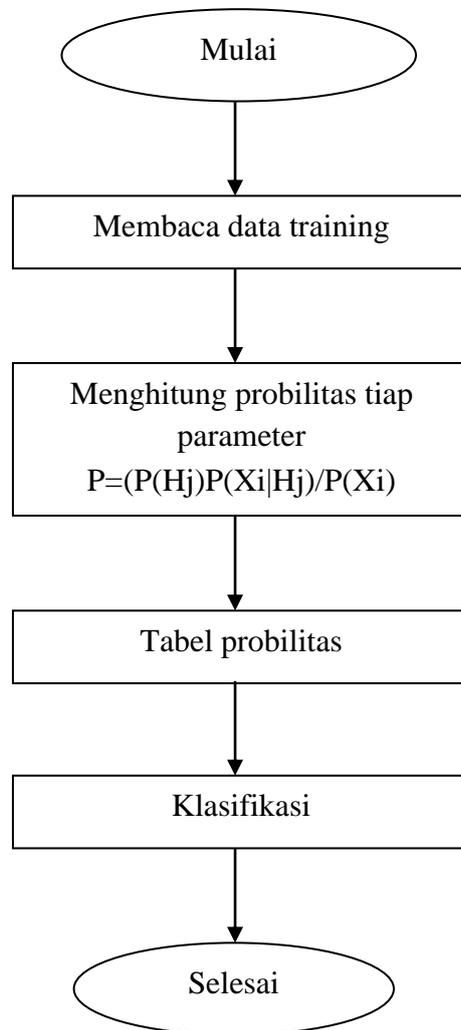
3.8 Algoritma *K-Nearest Neighbor*



Gambar 3. 2 Metode *K-Nearest Neighbor*

Implementasi pendekatan K-NN akan berlangsung di penelitian ini mencari nilai parameter K. Jarak antara data baru dan semua data dalam kumpulan maka pendidikan dipertimbangkan. Setelah itu, perbandingan dibuat dengan data percobaan dan matriks kebingungan dihasilkan. Setelah itu, aplikasi menunjukkan data prediksi.

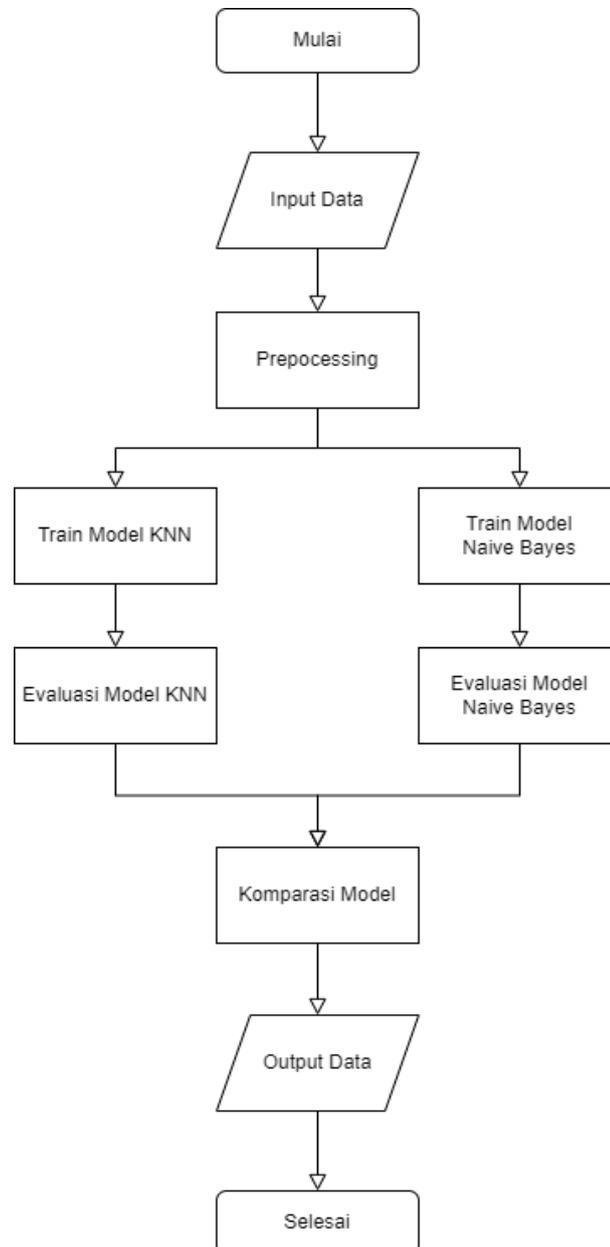
3.9 Algoritma *Native Bayes*



Gambar 3. 3 Metode *Native Bayes*

Pendekatan *Native Bayes* yang digunakan dalam penelitian ini, membaca data pelatihan terlebih dahulu lebih awal Setelah itu, pemilihan masing-masing parameter dihitung dan kemudian menghitung cosinus kesamaan. Akibatnya, akurasi dibandingkan nanti Menggunakan pendekatan K-NN.

3.10 Komparasi Algoritma KNN dan *Naïve Bayes*



Gambar 3. 4 Komparasi KNN dan Naive Bayes

Komparasi membantu kita memahami mana metode yang lebih sesuai atau efektif dalam menangani prediksi kelulusan mahasiswa berdasarkan karakteristik dan kebutuhan spesifik dari data yang digunakan. Itu juga membantu pengambilan keputusan dalam memilih model untuk diterapkan dalam kasus tertentu.

3.11 Contoh Perhitungan Algoritma KNN

Untuk menghitung prediksi kelulusan mahasiswa menggunakan algoritma KNN dengan variable nilai IPK Semester 1 sampai 4, absensi, waktu kelulusan dan jenis kelamin, kita akan menggunakan rumus *Euclidean Distance* sebagai matrik jarak. Menggunakan data latih dan data uji.

Data Latih (*Training Data*)

No	IPK1	IPK2	IPK3	IPK4	Absensi	Waktu Kelulusan	Jenis Kelamin	Kelulusan
1	3,5	3,6	3,7	3,8	90	4	Laki-laki	Lulus
2	3,2	3,4	3,5	3,6	80	4	Perempuan	Lulus
3	2,8	3,0	2,9	3,2	75	5	Laki-laki	Tidak Lulus
4	3,9	3,8	3,9	4,0	95	4	Perempuan	Lulus

Data Uji (*Testing Data*)

No	IPK1	IPK2	IPK3	IPK4	Absensi	Waktu Kelulusan	Jenis Kelamin
5	3,0	3,2	3,1	3,4	85	5	Laki-laki

Langkah-langkah perhitungan:

1. Menentukan Nilai K (Jumlah Terdekat) : K= 3.

2. Menghitung jarak :

- Jarak anatar data uji dan data latih ke-1 :

$$\sqrt{(3.0 - 3.5)^2 + (3.2 - 3.6)^2 + (3.1 - 3.7)^2 + (3.4 - 3.8)^2 + (85 - 90)^2 + (5 - 4)^2 + (0 - 1)^2}$$

- Jarak antara data uji dan data latih ke-2 :

$$\sqrt{(3.0 - 3.2)^2 + (3.2 - 3.4)^2 + (3.1 - 3.5)^2 + (3.4 - 3.6)^2 + (85 - 80)^2 + (5 - 4)^2 + (0 - 0)^2}$$

- Jarak antara data uji dan data latih ke-3 :

$$\sqrt{(3.0 - 2.8)^2 + (3.2 - 3.0)^2 + (3.1 - 2.9)^2 + (3.4 - 3.2)^2 + (85 - 75)^2 + (5 - 5)^2 + (0 - 1)^2}$$

3. Menentukan Tetangga terdekat : Pilih K tetangga terdekat berdasarkan jarak terkecil

- Tetangga Tetangga terdekat : Data latih ke-1, ke-2, dan ke-3.

4. Voting atau penentuan kelas : Hitung frekuensi masing-masing kelas diantara tetangga terdekat. Kelas dengan frekuensi tertinggi menjadi prediksi kelas.

- Kelas mayoritas : “Lulus”

Sebagai hasilnya, berdasarkan prediksi KNN dengan K=3, data uji (No.5) diprediksi “Lulus”. Ini karena mayoritas dari 3 tetangga terdekat adalah mahasiswa yang lulus.

3.12 Contoh perhitungan algoritma Naïve Bayes

Berdasarkan data latih dibawah ini :

Data latih (Data Training)

No	IPK1	IPK2	IPK3	IPK4	Absensi	Waktu Kelulusan	Jenis Kelamin	Kelulusan
1	3,5	3,6	3,7	3,8	90	4	Laki-laki	Lulus
2	3,2	3,4	3,5	3,6	80	4	Perempuan	Lulus
3	2,8	3,0	2,9	3,2	75	5	Laki-laki	Tidak Lulus
4	3,9	3,8	3,9	4,0	95	4	Perempuan	Lulus
5	2,5	2,6	2,7	2,8	75	5	Laki-laki	Tidak Lulus

Maka dapat kita hitung dengan algoritma naïve bayes dengan langkah-langkah sebagai berikut :

1. Probilitas Prior

- $P(\text{Lulus}) = \frac{2}{5} = 0,4$

- $P(\text{Tidak lulus}) = \frac{3}{5} = 0,6$

2. Probilitas likelihood untuk setiap fitur :

- Probabilitas Likelihood untuk IPK_Semester1:

$$P(\text{IPK_Semester1}=3.5|\text{Ya})=5/2=0.4$$

$$P(\text{IPK_Semester1}=3.5|\text{Tidak})=5/1=0.2$$

$$P(\text{IPK_Semester1}=3.2|\text{Ya})=5/1=0.2$$

- Probabilitas Likelihood untuk IPK_Semester2:

$$P(\text{IPK_Semester2}=3.6|\text{Ya})=5/1=0.2$$

$$P(\text{IPK_Semester2}=3.1|\text{Tidak})=5/1=0.2$$

$$P(\text{IPK_Semester2}=2.9|\text{Ya})=5/0=0.0$$

- Probabilitas Likelihood untuk IPK_Semester3:

$$P(\text{IPK_Semester3}=3.7|\text{Ya})=5/1=0.2$$

$$P(\text{IPK_Semester3}=3.0|\text{Tidak})=5/1=0.2$$

$$P(\text{IPK_Semester3}=3.1|\text{Ya})=5/0=0.0$$

- Probabilitas Likelihood untuk IPK_Semester4:

$$P(\text{IPK_Semester4}=3.8|\text{Ya})=5/2=0.4$$

$$P(\text{IPK_Semester4}=3.3|\text{Tidak})=5/1=0.2$$

$$P(\text{IPK_Semester4}=4.1|\text{Ya})=5/0=0.0$$

- Probabilitas Likelihood untuk Absensi:

$$P(\text{Absensi}=90|\text{Ya})=5/1=0.2$$

$$P(\text{Absensi}=85|\text{Tidak})=5/1=0.2$$

$$P(\text{Absensi}=80|\text{Ya})=5/0=0.0$$

- Probabilitas Likelihood untuk Waktu_Kelulusan:

$$P(\text{Waktu_Kelulusan}=4|\text{Ya})=5/3=0.6$$

$$P(\text{Waktu_Kelulusan}=4|\text{Tidak})=5/1=0.2$$

$$P(\text{Waktu_Kelulusan}=5|Y_a)=50=0.0$$

- Probabilitas Likelihood untuk Jenis_Kelamin:

$$P(\text{Jenis_Kelamin}=L|Y_a)=52=0.4$$

$$P(\text{Jenis_Kelamin}=P|\text{Tidak})=52=0.4$$

$$P(\text{Jenis_Kelamin}=L|Y_a)=52=0.4$$

3. Probabilitas Posterior:

- Probabilitas Posterior untuk Kelulusan (Ya):

$$\begin{aligned} P(Y_a|\text{data}) &\propto P(Y_a) \times P(\text{IPK_Semester1}=3.5 | Y_a) \times P(\text{IPK_Semester2}=3.6 | \\ &Y_a) \times P(\text{IPK_Semester3}=3.7 | Y_a) \times P(\text{IPK_Semester4}=3.8 | Y_a) \times P(\text{Absensi} \\ &=90 | Y_a) \times P(\text{Waktu_Kelulusan}=4 | Y_a) \times P(\text{Jenis_Kelamin}=L | Y_a) \\ &= \propto 0.5 \times 0.4 \times 0.2 \times 0.2 \times 0.4 \times 0.2 \times 0.6 \times 0.4 \end{aligned}$$

$$= \propto 0.001536$$

- Probabilitas Posterior untuk Tidak Lulus (Tidak):

$$\begin{aligned} P(\text{Tidak}|\text{data}) &\propto P(\text{Tidak}) \times P(\text{IPK_Semester1}=3.5 | \text{Tidak}) \times P(\text{IPK_Semester} \\ &2=3.6 | \text{Tidak}) \times P(\text{IPK_Semester3}=3.7 | \text{Tidak}) \times P(\text{IPK_Semester4}=3.8 | \\ &\text{Tidak}) \times P(\text{Absensi}=90 | \text{Tidak}) \times P(\text{Waktu_Kelulusan}=4 | \text{Tidak}) \times P \\ &(\text{Jenis_Kelamin}=L | \text{Tidak}) \end{aligned}$$

$$= \propto 0.5 \times 0.2 \times 0.2 \times 0.2 \times 0.2 \times 0.2 \times 0.2 \times 0.4$$

$$= \propto 0.000256$$

Setelah menghitung kedua nilai ini, dapat dibandingkan dan memilih kategori dengan probabilitas tertinggi sebagai prediksi. Dalam kasus ini, kita membandingkan $P(Y_a|\text{data})$ dan $P(\text{Tidak}|\text{data})$ untuk menentukan apakah

mahasiswa diprediksi lulus atau tidak. Dalam contoh ini, nilai $P(Ya|data)$ lebih tinggi, sehingga kita dapat memprediksi bahwa mahasiswa tersebut lulus.

4. Normalisasi Probabilitas

Setelah menghitung probabilitas posterior untuk setiap kategori (Ya dan Tidak), selanjutnya perlu melakukan normalisasi sehingga total probabilitas untuk kedua kategori sama dengan 1.

- Normalisasi Probabilitas untuk Kelulusan (Ya):

$$P(Ya|data) = \frac{P(Ya|Data)}{P(Ya|Data) + P(Tidak|Data)}$$

$$\frac{0,001536}{0,001536 + 0,000256}$$

$$\frac{0,001536}{0,001792}$$

$$\approx 0,858$$

- Normalisasi Probabilitas untuk Tidak Lulus (Tidak):

$$P(Tidak|data) = \frac{P(Tidak|Data)}{P(Ya|Data) + P(Tidak|Data)}$$

$$\frac{0,000256}{0,001536 + 0,000256}$$

$$\frac{0,000256}{0,001792}$$

$$\approx 0,142$$

Setelah normalisasi, kita mendapatkan probabilitas posterior yang dinormalisasi untuk setiap kategori. Dalam hal ini, probabilitas posterior untuk "Ya" adalah sekitar 0.858, sedangkan probabilitas posterior untuk "Tidak" adalah sekitar

0.142. Oleh karena itu, kita dapat memprediksi bahwa mahasiswa tersebut lebih mungkin lulus berdasarkan data yang diberikan.

5. Prediksi

Dengan menghitung normalisasi probabilitas, kita dapat membuat prediksi berdasarkan nilai tertinggi antara probabilitas posterior untuk "Ya" dan "Tidak". Dalam hal ini, kita memilih kategori dengan probabilitas tertinggi sebagai prediksi. Berdasarkan normalisasi probabilitas sebelumnya:

- Probabilitas normalisasi untuk Kelulusan (Ya): ≈ 0.858
- Probabilitas normalisasi untuk Tidak Lulus (Tidak): ≈ 0.142

Karena probabilitas normalisasi untuk "Ya" lebih tinggi, kita memprediksi bahwa mahasiswa tersebut akan lulus.

Jadi, prediksi berdasarkan data yang diberikan adalah bahwa mahasiswa tersebut diperkirakan lulus.

