



PENERAPAN ALGORITMA NAÏVE BAYES UNTUK PREDIKSI PENYAKIT DEPRESI PADA MAHASISWA

Adie Wahyudi Oktavia Gama^{1*}, Agustini Degni Melsy Grren¹, I Gusti Ngurah Darma Paramartha¹, Gede Humaswara Prathama¹, Ni Made Widnyani², Md. Wira Putra Dananjaya²

¹Fakultas Teknik dan Informatika, Universitas Pendidikan Nasional, Jl. Bedugul No.39, Sidakarya, Denpasar Selatan, Denpasar, Bali 80224, Indonesia

²Fakultas Bisnis, Sosial, Teknologi dan Humaniora, Universitas Bali Internasional, Gg. Jeruk No.9A, Tonja, Denpasar Timur, Denpasar, Bali 80234, Indonesia

*adiewahyudi@undiknas.ac.id

ABSTRAK

Depresi merupakan salah satu masalah kesehatan mental yang cukup serius di kalangan mahasiswa, namun seringkali luput dari perhatian karena adanya stigma sosial dan keterbatasan akses terhadap layanan psikologis. Penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes dalam memprediksi kondisi depresi pada mahasiswa berdasarkan berbagai faktor, seperti tekanan akademik, durasi tidur, kebiasaan makan, stres keuangan, dan riwayat keluarga dengan gangguan mental. Model dibangun menggunakan 502 data yang diperoleh dari platform Kaggle, melalui tahapan preprocessing data, transformasi, klasifikasi menggunakan Gaussian Naïve Bayes, serta evaluasi menggunakan confusion matrix. Proses implementasi dilakukan di Google Colab dengan memanfaatkan library scikit-learn. Hasil evaluasi menunjukkan kinerja model yang sangat baik dengan akurasi sebesar 97%, precision 96%, recall 98%, dan F1-score 97%. Temuan ini menunjukkan bahwa algoritma Naïve Bayes dapat digunakan secara efektif sebagai alat bantu skrining awal depresi yang bersifat anonim dan efisien, serta berpotensi mendukung peningkatan kesadaran dan intervensi kesehatan mental di lingkungan mahasiswa.

Kata kunci: depresi; kesehatan mental; machine learning; mahasiswa; naïve bayes; prediksi

APPLICATION OF THE NAÏVE BAYES ALGORITHM FOR PREDICTING DEPRESSION IN UNIVERSITY STUDENTS

ABSTRACT

Depression is one of the most serious mental health problems among college students, but it often goes unnoticed due to social stigma and limited access to psychological services. This study aims to apply the Naïve Bayes algorithm to predict depression in college students based on various factors, such as academic pressure, sleep duration, eating habits, financial stress, and family history of mental disorders. The model was built using 502 data obtained from the Kaggle platform, through the stages of data preprocessing, transformation, classification using Gaussian Naïve Bayes, and evaluation using a confusion matrix. The implementation process was carried out in Google Colab using the scikit-learn library. The evaluation results showed very good model performance with an accuracy of 97%, precision of 96%, recall of 98%, and F1-score of 97%. These findings indicate that the Naïve Bayes algorithm can be used effectively as an anonymous and efficient early screening tool for depression and has the potential to support increased awareness and mental health interventions in the college student environment.

Keywords: depression; machine learning; mental health; naïve bayes; prediction; students

PENDAHULUAN

Depresi merupakan salah satu gangguan kesehatan mental yang ditandai dengan perasaan tertekan, hilangnya minat atau kesenangan, penurunan energi, perasaan bersalah atau rendah diri, kesulitan tidur, berkurangnya nafsu makan, kelelahan, dan kesulitan berkonsentrasi. Kondisi ini sering kali menjadi kronis dan berulang, serta dapat secara signifikan mengganggu kemampuan seseorang dalam menjalani aktivitas sehari-hari. Menurut Diagnostic and Statistical Manual of Mental Disorders edisi ke-5 (DSM-5), seseorang dapat dikategorikan mengalami depresi apabila dalam kurun waktu minimal dua minggu mengalami lima atau lebih dari gejala (American Psychiatric Association, 2022). Dalam kasus yang paling parah, depresi dapat memicu pikiran untuk bunuh diri. Dalam upaya global untuk mencapai 17 Tujuan Pembangunan Berkelanjutan (SDGs) yang ditetapkan oleh Perserikatan Bangsa-Bangsa, kesehatan mental menjadi salah satu aspek penting yang perlu diperhatikan. Khususnya, SDGs 3 yang berfokus pada "Kesehatan yang Baik dan Kesejahteraan" mencakup target untuk meningkatkan kesehatan mental dan mencegah bunuh diri. Salah satu masalah kesehatan mental yang signifikan dan sering diabaikan adalah depresi (Diva Yuda et al., 2024). Menurut data dari Organisasi Kesehatan Dunia (WHO), memprediksikan bahwa pada tahun 2020 depresi menjadi salah satu penyakit mental yang banyak dialami dan menjadi penyebab kedua terbesar kematian setelah serangan jantung.

Berdasarkan penelitian yang dilakukan sebelumnya, sekitar 25% mahasiswa di Indonesia diketahui mengalami gejala depresi dengan tingkat keparahan sedang hingga berat. Kondisi ini memberikan dampak yang nyata terhadap penurunan motivasi belajar, terganggunya pola tidur, serta munculnya masalah dalam hubungan sosial (Ridha Dwiki Putri et al., 2024). Di berbagai belahan dunia depresi merupakan salah satu faktor utama yang menyebabkan masalah kesehatan dan gangguan fungsi pada mahasiswa. Di Indonesia, prevalensi gangguan mental emosional pada penduduk usia ≥ 18 mencapai 9,8% pada tahun 2018 (Nurabsharina & Kosasih, 2020). Angka ini menunjukkan bahwa masalah kesehatan mental, termasuk depresi, menjadi tantangan serius yang perlu ditangani. Namun, diagnosis dini dan penanganan depresi pada mahasiswa seringkali terhambat oleh berbagai faktor, seperti stigma sosial, kurangnya kesadaran, dan terbatasnya akses ke layanan kesehatan mental yang memadai. Hal ini semakin diperparah oleh kurangnya tenaga profesional kesehatan mental, terutama di daerah-daerah terpencil Indonesia (Hasbie et al., 2023). Dalam konteks ini, pemanfaatan teknologi informasi dapat menjadi solusi inovatif untuk mengatasi kesenjangan dalam diagnosis dan penanganan penyakit depresi. Machine Learning merupakan salah satu pendekatan yang menjanjikan untuk menyediakan akses yang lebih luas ke alat skrining dan diagnosis awal depresi.

Fenomena penyakit depresi di Indonesia semakin mengkhawatirkan, dengan banyak kasus yang tidak terdeteksi dan tidak mendapatkan penanganan yang tepat. Orang yang mengalami penyakit depresi cenderung memiliki toleransi stres yang rendah dan kerentanan yang tinggi terhadap tekanan. Fase ini sering disebut sebagai periode "Storm and Stress" karena perubahan emosional yang drastis, yang membuat masa ini penuh gejolak dan rentan terhadap konflik (Zuriah & Wardono, 2023). Menurut sebuah artikel kesehatan online (KlikDokter, 2022), salah satu faktor utama yang menyebabkan hal ini adalah rasa malu dan stigma sosial yang masih melekat pada masalah kesehatan mental, khususnya di kalangan mahasiswa. Banyak mahasiswa yang mengalami gejala depresi merasa enggan untuk mencari bantuan profesional, seperti psikiater atau psikolog, karena takut dianggap lemah atau gila oleh lingkungan sekitarnya. Akibatnya, mereka cenderung memendam masalah mereka, yang dapat memperparah kondisi mental dan berdampak negatif pada berbagai aspek kehidupan, termasuk prestasi akademik dan hubungan sosial. Disisi lain, akses ke layanan kesehatan mental yang terjangkau dan mudah diakses masih terbatas, terutama di daerah-daerah terpencil di Indonesia. Keterbatasan jumlah tenaga profesional kesehatan mental, seperti psikiater dan psikolog, juga menjadi tantangan tersendiri dalam upaya penanganan penyakit depresi secara komprehensif. Dalam konteks ini, prediksi penyakit depresi menggunakan algoritma Naïve Bayes menjadi solusi inovatif yang potensial. Model ini dapat

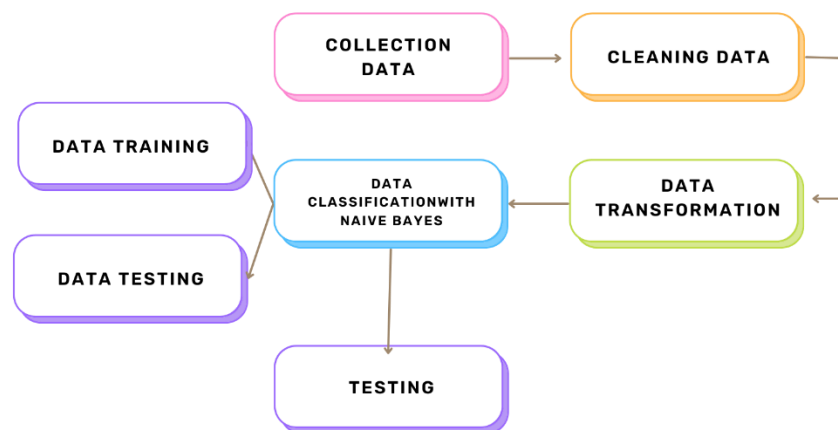
menjembatani kesenjangan antara kebutuhan prediksi awal dan keengganan orang untuk berkonsultasi langsung dengan profesional kesehatan mental.

Dengan memanfaatkan teknologi informasi, Machine Learning ini dapat menyediakan akses yang lebih luas dan mudah bagi orang yang mengalami penyakit depresi untuk melakukan skrining awal depresi secara mandiri dan anonim, serta mengurangi rasa malu dan stigma yang mungkin dirasakan. Selain itu, model ini juga dapat membantu dalam proses edukasi dan peningkatan kesadaran tentang kesehatan mental di berbagai kalangan, yang merupakan langkah penting dalam normalisasi diskusi tentang kesehatan mental. Untuk memastikan keakuratan model dalam mendiagnosis depresi, Algoritma yang digunakan yaitu Naïve Bayes dengan pengujian yang dilakukan menggunakan Confusion Matrix. Confusion Matrix merupakan salah satu metode pengujian model yang berbentuk tabel. Pengujiannya adalah dengan mengukur kinerja suatu set data (Devi Nurhayati & Widayani, 2021). Metode yang dikenal sebagai Confusion Matrix digunakan untuk mengukur kinerja atau tingkat akurasi dalam proses klasifikasi. Pengujian ini melibatkan pengukuran True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN), yang bertujuan untuk mengevaluasi performa model dalam hal akurasi, presisi, dan sensitivitas (Gama et al., 2025). Dengan pengujian ini, dapat ditentukan sejauh mana machine learning mampu memberikan hasil prediksi yang benar dan mengurangi tingkat kesalahan dalam klasifikasi depresi.

Penelitian ini bertujuan untuk mengembangkan model prediksi tingkat penyakit depresi menggunakan algoritma Naïve Bayes. Dengan meningkatnya kasus penyakit depresi pada mahasiswa, diperlukan metode yang dapat membantu dalam mengidentifikasi risiko penyakit depresi berdasarkan gejala yang dialami. Naïve Bayes dipilih karena kemampuannya dalam melakukan klasifikasi berbasis probabilitas yang sederhana namun efektif dalam menangani dataset dengan fitur yang saling independen. Dalam penelitian ini, data yang digunakan mencakup berbagai gejala depresi, seperti gangguan tidur, perasaan sedih berkepanjangan, hilangnya minat terhadap aktivitas, perubahan pola makan, serta faktor risiko lainnya. Model Naïve Bayes akan dilatih menggunakan dataset yang telah diklasifikasikan berdasarkan tingkat depresi, kemudian diuji untuk menilai akurasinya dalam memprediksi kondisi mahasiswa berdasarkan gejala yang diberikan. Untuk mengukur performa model, dilakukan pengujian menggunakan Confusion Matrix, yang mencakup akurasi, presisi, recall, dan F1-score guna menilai seberapa baik algoritma Naïve Bayes dalam mengklasifikasikan tingkat penyakit depresi. Hasil penelitian ini diharapkan dapat memberikan gambaran tentang efektivitas Naïve Bayes dalam prediksi penyakit depresi serta mengidentifikasi faktor-faktor yang paling berpengaruh dalam menentukan tingkat depresi seseorang. Dengan adanya model prediksi ini, diharapkan dapat menjadi referensi bagi penelitian lebih lanjut dalam bidang kesehatan mental dan penerapan teknologi dalam analisis prediktif.

METODE

Penelitian ini dilakukan kurang lebih selama 4 bulan dari bulan Januari – April 2025 dengan mengikuti alur sesuai Gambar 1. Tools utama yang digunakan yaitu Google Colab dalam mengolah data, membangun model, dan melakukan prediksi penyakit depresi. Google Colab dipilih sebagai tools dalam penelitian ini karena menyediakan akses gratis ke komputasi berbasis cloud dengan dukungan Central Processing Unit (CPU), memungkinkan eksekusi machine learning lebih cepat tanpa perlu instalasi perangkat lunak tambahan, serta mendukung kolaborasi dan integrasi langsung dengan pustaka populer seperti Scikit-learn dan TensorFlow, sehingga mempermudah proses pengolahan data dan implementasi algoritma. Rangkaian kegiatan ini memastikan bahwa penelitian berjalan secara sistematis dan menghasilkan model prediksi yang optimal.



Gambar 1. Alur Penelitian

Berdasarkan Gambar 1, diagram alur dalam gambar ini menggambarkan tahapan utama dalam proses klasifikasi data menggunakan algoritma Naïve Bayes, dengan dataset yang diperoleh dari Kaggle. Langkah pertama adalah Pengumpulan Data (Collection Data), di mana dataset yang didapat dari Kaggle diunduh dan dipersiapkan untuk dianalisis. Alasan memilih dataset dari Kaggle adalah karena praktis dan telah tersedia dalam format yang rapi, sehingga tidak perlu melakukan pengumpulan data secara langsung. Selain itu, dataset di Kaggle telah melalui proses validasi oleh komunitas pengguna, sehingga lebih terpercaya dan dapat diandalkan. Proses pengumpulan data dimulai dengan mencari dataset yang sesuai dengan topik penelitian, yaitu faktor-faktor yang mempengaruhi tingkat depresi di kalangan mahasiswa. Setelah data terkumpul, dilakukan proses pembersihan data (Cleaning Data) untuk memastikan bahwa tidak ada data yang tidak valid, duplikat, atau tidak relevan. Langkah ini bertujuan untuk meningkatkan kualitas data agar hasil analisis lebih akurat. Pada penelitian ini, data yang digunakan diperoleh dari situs Kaggle dengan nama dataset "*Depression Student Dataset*" dengan link <https://www.kaggle.com/datasets/ikynahidwin/depression-student-dataset>. Format dataset yang diperoleh berupa CSV, yang dimana dataset ini berisi data mahasiswa yang berkaitan dengan faktor-faktor risiko depresi. Pemilihan dataset ini bukan tanpa alasan, karena data tersebut sudah memiliki struktur yang jelas dan variabel-variabel yang memang relevan untuk analisis depresi di kalangan mahasiswa. Adapun rincian atribut indikator yang dianalisis seperti pada Tabel 1 berikut.

Tabel 1.

Indikator dalam Dataset

No	Indikator Atribut	Jumlah	Tipe Data	Tahun
1	Gender	502	Karakter	2024
2	Age	502	Numerik	2024
3	Academic Pressure	502	Numerik	2024
4	Study Satisfaction	502	Numerik	2024
5	Sleep Duration	502	Kategorikal	2024
6	Dietary Habits	502	Kategorikal	2024
7	Have you ever had suicidal thoughts ?	502	Kategorikal	2024
8	Study Hours	502	Numerik	2024
9	Financial Stress	502	Numerik	2024
10	Family History of Mental Illness	502	Kategorikal	2024
11	Depression	502	Kategorikal	2024

Data yang telah dibersihkan kemudian masuk ke tahap transformasi, dimana data dikonversi ke dalam format yang sesuai agar dapat digunakan dalam model Naïve Bayes. Transformasi ini mencakup normalisasi data, pengkodean nilai kategoris, atau teknik lainnya yang memastikan bahwa data siap untuk dianalisis dengan metode klasifikasi. Dalam proses klasifikasi, dataset dibagi menjadi dua bagian, yaitu

data training dan data testing. Pada penelitian ini menggunakan perbandingan 80% data digunakan untuk training dan 20% untuk testing, pemilihan variasi perbandingan ini dilakukan berdasarkan studi literatur dari penelitian sebelumnya, yang menunjukkan bahwa penggunaan rasio yang berbeda dapat mempengaruhi hasil akurasi model. Data pelatihan digunakan untuk melatih model Naïve Bayes agar mampu mengenali pola dalam data, sedangkan data pengujian berfungsi untuk mengukur seberapa baik model tersebut dalam memprediksi gejala depresi pada data yang belum pernah dilihat sebelumnya. Tahap klasifikasi dilakukan dengan algoritma Naïve Bayes. Pada tahap ini, model akan mengolah data dan menentukan kemungkinan seseorang mengalami depresi berdasarkan parameter yang telah ditentukan. Algoritma ini bekerja dengan prinsip Teorema Bayes, yang menyatakan bahwa (Gama & Wardhiana, 2023):

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \dots\dots\dots (1)$$

Di mana:

- P(A|B) = Peluang kejadian *A* dengan syarat *B*, dalam model ini dapat diartikan dengan peluang *A* jika diketahui kejadian jenis penyakit *B*.
- P(B|A) = Peluang *evidence B* jika diketahui hipotesis *A*.
- P(A) = Probabilitas hipotesis *A* tanpa memandang *evidence* manapun.
- P(B) = Peluang *evidence B*.

Langkah terakhir dalam proses ini adalah pengujian model, yang bertujuan untuk mengevaluasi performa model dalam memprediksi depresi. Evaluasi ini mencakup empat metrik utama, yaitu akurasi, precision, recall dan F1 Score. Keempat metrik ini digunakan untuk menilai seberapa efektif model dalam melakukan klasifikasi. Metrik evaluasi dapat dihitung untuk menilai performa model, antara lain:

- a. Accuracy : Mengukur sejauh mana model dapat memprediksi dengan benar secara keseluruhan.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (2)$$

- b. Precision : Mengukur seberapa tepat model dalam mengklasifikasikan kelas positif, yaitu proporsi prediksi positif yang benar.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (3)$$

- c. Recall : Mengukur kemampuan model dalam mendeteksi semua data positif yang benar.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (4)$$

- d. F1-Score : Merupakan rata-rata harmonik antara Precision dan Recall, yang digunakan untuk menyeimbangkan keduanya.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \dots\dots\dots (5)$$

Hasil dari pengujian ini akan digunakan untuk menilai tingkat akurasi model dan menentukan apakah perlu dilakukan optimasi lebih lanjut. Dengan proses yang sistematis ini, diharapkan model Naïve Bayes dapat memberikan prediksi depresi yang lebih akurat dan bermanfaat dalam mendukung deteksi dini gangguan kesehatan mental. Survei kesehatan mental nasional menunjukkan bahwa satu dari tiga mahasiswa di Indonesia mengalami masalah kesehatan mental, namun hanya sebagian kecil yang mendapatkan bantuan profesional (Tyas et al., 2023). Menurut sebuah survei yang dilaporkan oleh Detik.com (2024) orang yang mengalami penyakit depresi sering merasa terisolasi dan tidak memiliki akses yang memadai untuk mendapatkan dukungan psikologis, baik karena keterbatasan layanan kesehatan mental maupun rasa malu akibat stigma (Rully Desthian Pahlephi, 2022).

HASIL DAN PEMBAHASAN

Implementasi Pengumpulan Data

Tabel 1 berisikan indikator Gender, Age, Academic Pressure, Study Satisfaction, Sleep Duration, Dietary Habits, Have you ever had suicidal thoughts?, Study Hours, Financial Stress, Family History of Mental Illness dan Depression yang dijadikan atribut, lalu banyaknya data yang digunakan pada dataset tersebut sebanyak 502 data, lalu pada tipe data masing-masing atribut terdiri dari karakter, numerik dan kategorikal. Atribut Depression ini berperan sebagai label target yang akan diprediksi oleh model sedangkan untuk atribut lainnya berperan sebagai indikator dalam pengolahan data. Adapun dataset yang telah dikumpulkan sebagai berikut.

	Gender	Age	Academic Pressure	Study Satisfaction	Sleep Duration	Dietary Habits	Have you ever had suicidal thoughts ?	Study Hours	Financial Stress	Family History of Mental Illness	Depression
0	Male	28	2.0	4.0	7-8 hours	Moderate	Yes	9	2	Yes	No
1	Male	28	4.0	5.0	5-6 hours	Healthy	Yes	7	1	Yes	No
2	Male	25	1.0	3.0	5-6 hours	Unhealthy	Yes	10	4	No	Yes
3	Male	23	1.0	4.0	More than 8 hours	Unhealthy	Yes	7	2	Yes	No
4	Female	31	1.0	5.0	More than 8 hours	Healthy	Yes	4	2	Yes	No
...
497	Male	26	5.0	2.0	More than 8 hours	Unhealthy	No	8	3	No	Yes
498	Male	24	2.0	1.0	Less than 5 hours	Unhealthy	Yes	8	5	No	Yes
499	Female	23	3.0	5.0	5-6 hours	Healthy	No	1	5	Yes	No
500	Male	33	4.0	4.0	More than 8 hours	Healthy	No	8	1	Yes	No
501	Male	18	5.0	3.0	More than 8 hours	Unhealthy	No	6	2	Yes	Yes

502 rows × 11 columns

Gambar 2. Dataset Depresi Mahasiswa

Gambar 2 berisikan data sampel yang diambil dari 502 data yang dianalisis dan diolah, akan tetapi pada data di atas masih dalam bentuk mentah, artinya belum dilakukan transformasi data ke bentuk numerik sesuai dengan kebutuhan algoritma Naïve Bayes. Proses pengumpulan data ini menjadi tahap awal sebelum dataset dapat diolah lebih lanjut melalui tahap cleaning, transformasi, dan klasifikasi menggunakan Google Colab. Platform google colab ini sangat praktis karena tidak memerlukan instalasi perangkat lunak tambahan dengan sistem operasi yang digunakan adalah Ubuntu 22.04.4 LTS, dengan prosesor Intel(R) Xeon(R) CPU @ 2.20GHz, dan kapasitas RAM sebesar 12,7 GB. Setelah data terkumpul dan dikonfirmasi kesesuaiannya, maka dilanjutkan ke tahapan preprocessing untuk menyiapkan data agar siap dianalisis menggunakan algoritma Naïve Bayes.

Implementasi Cleaning Data

Setelah proses pengumpulan data selesai dilakukan, tahapan selanjutnya dalam penelitian ini adalah melakukan pembersihan data (*data cleaning*). Tahap ini bertujuan untuk memastikan bahwa dataset yang digunakan dalam penelitian benar-benar bersih, tidak terdapat kesalahan, serta siap diolah menggunakan algoritma Naïve Bayes. Proses pembersihan data diawali dengan pengecekan terhadap data kosong atau *missing value*. Pengecekan ini penting dilakukan untuk memastikan bahwa seluruh atribut memiliki nilai yang lengkap, karena data kosong dapat mempengaruhi akurasi hasil prediksi. *Cleaning Data* dilakukan di Google Colab dengan menggunakan *library* Pandas. Berikut hasil dari *Cleaning Data*.

```

▶ df.info()

df.isnull().sum()

df.dropna(inplace=True)

df.drop_duplicates(inplace=True)

↔ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 502 entries, 0 to 501
Data columns (total 11 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Gender                                     502 non-null    object
1   Age                                         502 non-null    int64
2   Academic Pressure                          502 non-null    float64
3   Study Satisfaction                         502 non-null    float64
4   Sleep Duration                             502 non-null    object
5   Dietary Habits                             502 non-null    object
6   Have you ever had suicidal thoughts ?     502 non-null    object
7   Study Hours                                502 non-null    int64
8   Financial Stress                           502 non-null    int64
9   Family History of Mental Illness          502 non-null    object
10  Depression                                 502 non-null    object
dtypes: float64(2), int64(3), object(6)
memory usage: 43.3+ KB

```

Gambar 3. Proses Cleaning Data

Berdasarkan hasil pengukuran yang dilakukan menggunakan google colab berdasarkan Gambar 3, ditemukan bahwa tidak ada data kosong dalam dataset yang digunakan, pengecekan ini dilakukan menggunakan fungsi `.isnull()`. Seluruh baris data telah terisi dengan lengkap, sehingga tidak diperlukan proses imputasi atau penghapusan data pada tahap ini. Setelah memastikan tidak ada data kosong, dilakukan pula pengecekan terhadap data ganda atau duplikat. Data duplikat dapat menyebabkan hasil analisis menjadi bias karena adanya pengulangan informasi yang sama. Oleh karena itu, pengecekan ini dilakukan menggunakan fungsi `.duplicated()`. Dari hasil pengecekan, tidak ditemukan adanya data yang terduplikasi di dalam dataset. Hal ini menunjukkan bahwa setiap data mahasiswa bersifat unik dan tidak terjadi pengulangan data. Selain itu, dilakukan juga pengecekan tipe data pada masing-masing atribut. Dataset yang diunduh masih memiliki beberapa atribut dengan tipe data karakter (*string*), seperti *Gender*, *Dietary Habits*, *Have You Ever Had Suicidal Thoughts*, *Family History Of Mental Illness* dan *Depression*. Pengecekan dilakukan untuk memastikan bahwa tipe data tersebut sesuai dan nantinya bisa diolah lebih lanjut. Tipe data tersebut akan dikonversi menjadi bentuk numerik di tahap transformasi data, namun pada tahap cleaning ini cukup dipastikan bahwa data sudah rapi dan sesuai format.

Implementasi Transformasi Data

Setelah proses *cleaning data* selesai dilakukan, tahap selanjutnya adalah transformasi data. Transformasi data merupakan langkah penting agar data yang semula berbentuk karakter atau kategori dapat diubah menjadi bentuk numerik, sehingga dapat diproses oleh algoritma Naïve Bayes. Pada penelitian ini, proses transformasi data dilakukan sepenuhnya menggunakan Google Colab, dengan memanfaatkan *library*. Dataset yang digunakan memiliki beberapa atribut dengan tipe data kategorikal seperti *Gender*, *Dietary Habits*, *Have You Ever Had Suicidal Thoughts*, *Family History Of Mental Illness*, dan *Depression*. Atribut-atribut ini awalnya memiliki nilai dalam bentuk teks, seperti “Laki-laki”, “Perempuan”, “Tidak Sehat”, “Sedang”, “Sehat”, “Ya”, dan “Tidak”. Untuk keperluan pengolahan data menggunakan algoritma Naïve Bayes, nilai-nilai tersebut perlu dikonversi ke bentuk numerik dengan sebagai berikut.

```
[ ] df['Gender'] = df['Gender'].map({'Male': 1, 'Female': 0})

df['Sleep Duration'] = df['Sleep Duration'].map({
    '5-6 hours':0,
    '7-8 hours':1,
    'Less than 5 hours':2,
    'More than 8 hours':3
})

df['Dietary Habits'] = df['Dietary Habits'].map({
    'Unhealthy': 0,
    'Moderate': 1,
    'Healthy': 2
})

df['Have you ever had suicidal thoughts ?'] = df['Have you ever had suicidal thoughts ?'].map({
    'No': 0,
    'Yes': 1
})

df['Family History of Mental Illness'] = df['Family History of Mental Illness'].map({
    'No': 0,
    'Yes': 1
})

df['Depression'] = df['Depression'].map({
    'No': 0,
    'Yes': 1
})

df.head(502)
```

Gambar 4. Proses Transformasi Data

Berdasarkan Gambar 4, proses transformasi diawali dengan mengubah atribut *Gender*. Nilai “Laki-laki” dikodekan menjadi 1, sedangkan “Perempuan” dikodekan menjadi 0. Hal ini dilakukan dengan fungsi `.map()` di Google Colab, sehingga setiap data pada kolom *Gender* terkonversi secara otomatis sesuai ketentuan tersebut. Atribut berikutnya adalah *Sleep Duration*, yang semula berbentuk angka (numerik), namun perlu diklasifikasikan ulang ke dalam kategori numerik diskrit. Kriteria klasifikasinya adalah sebagai berikut: durasi tidur kurang dari 5 jam per hari dikodekan menjadi 0, durasi tidur antara 5 hingga 6 jam menjadi 1, durasi 7 hingga 8 jam menjadi 2, dan durasi lebih dari 8 jam menjadi 3. Konversi ini dilakukan menggunakan fungsi `.apply()` dengan bantuan fungsi *custom* pada Python. Selanjutnya, atribut *Dietary Habits* yang memiliki nilai “Tidak Sehat”, “Sedang”, dan “Sehat” juga dikonversi menjadi angka dengan skema Tidak Sehat = 0, Sedang = 1, dan Sehat = 2. Proses ini dilakukan agar algoritma dapat mengenali pola data numerik dalam proses klasifikasi. Atribut *Have You Ever Had Suicidal Thoughts* dan *Family History Of Mental Illness* yang memiliki pilihan jawaban “Ya” dan “Tidak”, masing-masing diubah menjadi 1 untuk “Ya” dan 0 untuk “Tidak”. Transformasi ini bersifat biner dan dilakukan secara langsung dengan fungsi `.map()` agar konsisten dengan kebutuhan model. Terakhir, atribut *Depression* yang menjadi label target penelitian juga dikonversi dari nilai “Ya” dan “Tidak” menjadi 1 dan 0. Atribut ini merupakan variabel yang akan diprediksi oleh model Naïve Bayes berdasarkan data fitur yang telah diubah ke bentuk numerik. Adapun hasil dari transformasi data adalah sebagai berikut:

	Gender	Age	Academic Pressure	Study Satisfaction	Sleep Duration	Dietary Habits	Have you ever had suicidal thoughts ?	Study Hours	Financial Stress	Family History of Mental Illness	Depression
0	1	28	2.0	4.0	1	1	1	9	2	1	0
1	1	28	4.0	5.0	0	2	1	7	1	1	0
2	1	25	1.0	3.0	0	0	1	10	4	0	1
3	1	23	1.0	4.0	3	0	1	7	2	1	0
4	0	31	1.0	5.0	3	2	1	4	2	1	0
...
497	1	26	5.0	2.0	3	0	0	8	3	0	1
498	1	24	2.0	1.0	2	0	1	8	5	0	1
499	0	23	3.0	5.0	0	2	0	1	5	1	0
500	1	33	4.0	4.0	3	2	0	8	1	1	0
501	1	18	5.0	3.0	3	0	0	6	2	1	1

502 rows x 11 columns

Gambar 5. Hasil Transformasi Data

Pada Gambar 5 menunjukkan bahwa nilai-nilai yang semula berupa karakter telah berhasil diubah menjadi numerik sesuai dengan kriteria yang telah ditetapkan. Hasil transformasi ini menjadi dasar bagi proses klasifikasi yang akan dilakukan pada tahap berikutnya.

Implementasi Klasifikasi Data dengan Naïve Bayes

Setelah proses cleaning dan transformasi data selesai dilakukan, tahapan berikutnya dalam penelitian ini adalah klasifikasi data. Umumnya Naïve Bayes adalah algoritma machine learning yang banyak digunakan dalam klasifikasi, termasuk dalam konteks prediksi kesehatan mental seperti depres (Surya Wisnugraha et al., 2023). Klasifikasi dilakukan dengan tujuan untuk memprediksi apakah seorang mahasiswa mengalami depresi atau tidak, berdasarkan data atribut yang telah disiapkan. Proses ini dilakukan sepenuhnya menggunakan Google Colab, dengan memanfaatkan library scikit-learn sebagai alat bantu utama dalam pengolahan klasifikasi data. Naïve Bayes memiliki keunggulan dalam menangani dataset kompleks dengan banyak variabel independen, menjadikannya pilihan ideal untuk mendiagnosis gangguan kesehatan mental. Penelitian sebelumnya menunjukkan bahwa metode ini dapat memberikan hasil yang akurat dalam memprediksi tingkat kesehatan mental dengan akurasi mencapai 85% dengan menggunakan algoritma ini (Tengku Omri Wikana et al., 2024).

```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
train_size = len(X_train)
test_size = len(X_test)

labels = ['Data Training', 'Data Testing']
sizes = [train_size, test_size]
colors = ['#66b3ff', '#ff9999']

def absolute_value(val):
    total = sum(sizes)
    absolute = int(round(val * total / 100))
    return f'{absolute} data'

plt.figure(figsize=(6,6))
plt.pie(sizes, labels=labels, colors=colors, autopct=absolute_value, startangle=140)
plt.title('Proportion of Training and Testing Data Amount')
plt.axis('equal')
plt.show()

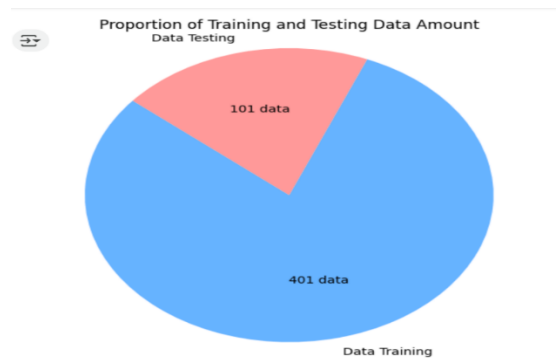
```

Gambar 6. Implementasi data training dan data testing

Pada Gambar 6 menunjukkan langkah awal sebelum membangun model klasifikasi adalah melakukan pembagian dataset menjadi dua bagian, yaitu data latih (training data) dan data uji (testing data). Data

latih digunakan untuk melatih model dalam mengenali pola hubungan antara atribut-atribut dengan kondisi depresi, sedangkan data uji digunakan untuk mengukur kemampuan model dalam memprediksi data baru yang belum pernah dikenali sebelumnya.

Pada penelitian ini, dataset dibagi dengan komposisi 80% sebagai data latih dan 20% sebagai data uji. Pembagian ini dilakukan secara acak namun tetap terkontrol dengan menggunakan parameter `random_state=42` agar hasilnya konsisten dan dapat direproduksi. Adapun hasil dari pembagian data sebagai berikut.



Gambar 7. Proporsi Klasifikasi Data

Berdasarkan diagram pie yang ditampilkan pada gambar 7, terlihat bahwa dari total 502 data, sebanyak 401 data digunakan untuk proses pelatihan (training) dan 101 data sisanya digunakan untuk pengujian (testing). Pembagian ini sesuai dengan parameter `test_size=0.2` yang berarti 20% dari data dijadikan data uji, sementara 80% sisanya untuk melatih model. Visualisasi ini memperjelas bahwa porsi terbesar memang diarahkan ke data training karena model membutuhkan lebih banyak data agar bisa belajar pola dengan lebih baik sebelum diuji. Melalui pembagian ini, proses evaluasi performa model menjadi lebih jelas karena model diuji pada data yang belum pernah dilihat sebelumnya.

```

▶ train_data = pd.concat([X_train, y_train], axis=1)
  test_data = pd.concat([X_test, y_test], axis=1)

  print("Data Training")
  display(train_data)

  print("\nData Testing")
  display(test_data)

```

Gambar 8. Penggabungan Data Training dan Testing Secara Lengkap

Pada bagian gambar 8 ini, *data training* dan *data testing* yang sebelumnya telah dipisahkan, digabung kembali antara fitur (X) dan label target (y) supaya bisa ditampilkan dalam bentuk tabel yang utuh. Proses penggabungan ini dilakukan secara horizontal, artinya kolom-kolom dari data fitur dan data target disatukan dalam satu data frame, sehingga setiap baris data bisa langsung menunjukkan fitur dan labelnya. Setelah digabung, data ditampilkan menggunakan fungsi `display()` agar kita bisa melihat dan

mengecek isi dari *data training* dan *data testing* secara lengkap dengan lebih mudah. Adapun hasil dari proses penggabungan dari *data training* dan *data testing* sebagai berikut.

Data Training												
Gender	Age	Academic Pressure	Study Satisfaction	Sleep Duration	Dietary Habits	Have you ever had suicidal thoughts ?	Study Hours	Financial Stress	Family History of Mental Illness	Depression		
423	0	28	4.0	2.0	3	1	0	0	3	0	0	
19	1	29	4.0	3.0	3	0	1	1	3	0	1	
323	1	33	4.0	1.0	2	1	0	4	5	0	1	
333	0	34	2.0	5.0	0	2	0	11	4	0	0	
56	1	29	3.0	1.0	3	1	1	4	1	1	1	
...	
106	0	25	2.0	2.0	3	1	0	11	2	1	0	
270	1	28	5.0	4.0	3	1	1	6	4	0	1	
348	1	19	3.0	1.0	2	0	0	10	4	0	1	
435	0	32	3.0	1.0	1	1	1	7	3	1	1	
102	0	20	2.0	5.0	3	0	1	4	2	0	0	

401 rows × 11 columns

Gambar 9. Sampel Data Training 80%

Data Testing												
Gender	Age	Academic Pressure	Study Satisfaction	Sleep Duration	Dietary Habits	Have you ever had suicidal thoughts ?	Study Hours	Financial Stress	Family History of Mental Illness	Depression		
268	0	33	4.0	3.0	1	2	0	2	2	0	0	
73	1	26	3.0	4.0	3	0	1	11	5	0	1	
289	0	28	3.0	2.0	2	2	1	10	3	1	1	
155	0	29	3.0	2.0	2	1	0	5	2	1	0	
104	1	19	2.0	5.0	0	2	0	9	1	1	0	
...	
86	0	32	5.0	3.0	1	2	1	3	3	0	1	
75	0	31	1.0	3.0	1	0	0	10	1	1	0	
440	0	32	3.0	4.0	0	1	1	12	1	0	0	
15	1	28	5.0	3.0	0	2	1	8	3	1	1	
250	1	20	3.0	3.0	3	2	1	6	5	1	1	

101 rows × 11 columns

Gambar 10. Sampel Data Testing 20%

Gambar 9 dan gambar 10 merupakan sampel data dari hasil klasifikasi dengan menggunakan perbandingan rasio 80:20 dimana 80% untuk training dan 20% untuk testing. Berdasarkan rasio tersebut, didapatkan jumlah data sebanyak 401 sebagai data training dan 101 sebagai data testing. Setelah data selesai melalui tahap cleaning, transformasi, dan klasifikasi berupa pembagian data latih dan data uji, tahap selanjutnya adalah implementasi algoritma Naïve Bayes untuk memproses data yang telah disiapkan. Implementasi ini bertujuan untuk memprediksi apakah seorang mahasiswa mengalami depresi atau tidak, berdasarkan pola hubungan antara atribut-atribut yang sudah ada di dalam dataset.

```
[13] import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

result_df = X_test.copy()
result_df['Actual'] = y_test.values
result_df['Prediction'] = y_pred

attributes = ['Gender', 'Sleep Duration', 'Dietary Habits',
             'Have you ever had suicidal thoughts?',
             'Family History of Mental Illness']

for attr in attributes:
    plt.figure(figsize=(9, 5))

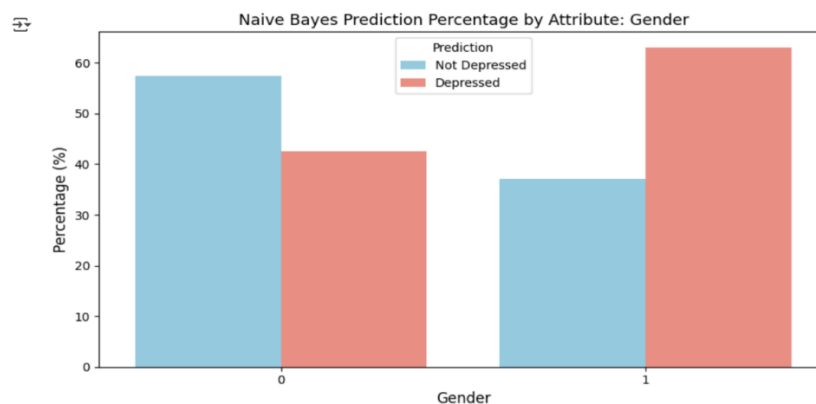
    count_df = result_df.groupby(attr)['Prediction'].value_counts(normalize=True).rename('Percentage').reset_index()
    count_df['Percentage'] = count_df['Percentage'] * 100
    count_df['Prediction'] = count_df['Prediction'].map({0: 'Not Depressed', 1: 'Depressed'})

    sns.barplot(x=attr, y='Percentage', hue='Prediction', data=count_df,
               palette={'Not Depressed': 'skyblue', 'Depressed': 'salmon'})

    plt.title(f'Naive Bayes Prediction Percentage by Attribute: {attr}', fontsize=13)
    plt.ylabel('Percentage (%)', fontsize=12)
    plt.xlabel(attr, fontsize=12)
    plt.legend(title='Prediction')
    plt.tight_layout()
    plt.show()
```

Gambar 11. Visualisasi Hasil Prediksi Model Naive Bayes Berdasarkan Atribut Kategori

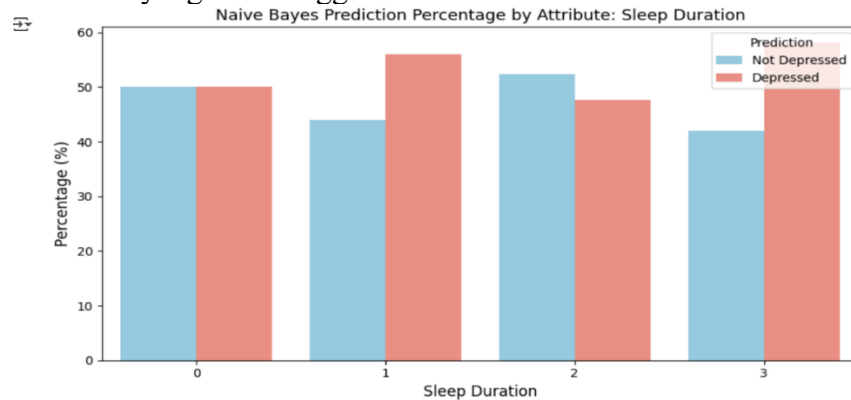
Dari Gambar 11, tahap pertama dalam implementasi ini adalah melatih model menggunakan data latih (X_{train} dan y_{train}). Pada proses ini, algoritma Naïve Bayes akan menghitung probabilitas kemunculan masing-masing fitur terhadap kelas target ($Depression = 1$ atau 0). Model mempelajari hubungan antara atribut seperti *Gender*, *Sleep Duration*, *Dietary Habits*, *Suicidal Thoughts*, dan *Family History* dengan kondisi depresi mahasiswa. Setelah model selesai dilatih, langkah berikutnya adalah melakukan prediksi terhadap data uji (X_{test}). Proses ini dilakukan menggunakan fungsi *predict()*, di mana model akan menentukan apakah data uji termasuk kategori depresi atau tidak berdasarkan probabilitas yang telah dihitung dari data latih. Hasil prediksi yang dihasilkan oleh model kemudian dibandingkan dengan label aktual dari data uji (y_{test}). Perbandingan ini akan digunakan untuk mengukur seberapa akurat model dalam melakukan prediksi, serta seberapa baik model mengenali pola depresi di data mahasiswa. Berikut merupakan hasil implementasi Algoritma Naïve Bayes di masing-masing atribut :



Gambar 12. Implementasi Terhadap Atribut Gender

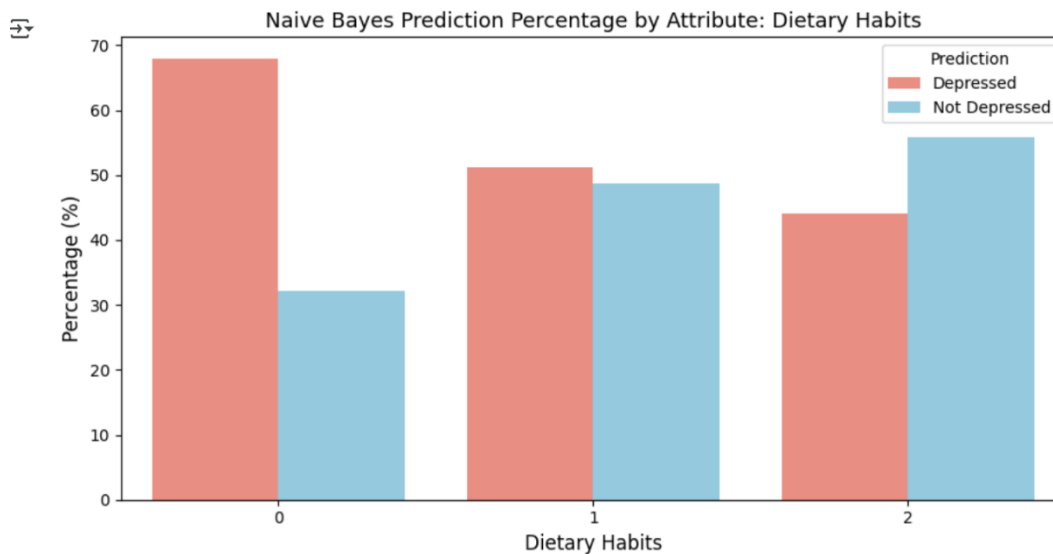
Pada Gambar 12 atribut *Gender* terlihat bahwa model memprediksi mahasiswa laki-laki (kategori 1) lebih banyak mengalami depresi dibandingkan tidak depresi. Sebaliknya, mahasiswa perempuan (kategori 0) lebih banyak diklasifikasikan sebagai tidak mengalami depresi. Hasil ini menunjukkan

bahwa dalam data yang digunakan, mahasiswa laki-laki cenderung memiliki kecenderungan lebih besar untuk diklasifikasikan sebagai depresi oleh model. Hal ini menarik, mengingat isu kesehatan mental sering kali lebih banyak dibahas pada perempuan, namun dalam konteks mahasiswa, laki-laki justru menunjukkan potensi risiko yang lebih tinggi menurut model.



Gambar 13. Implementasi Terhadap Atribut Sleep Duration

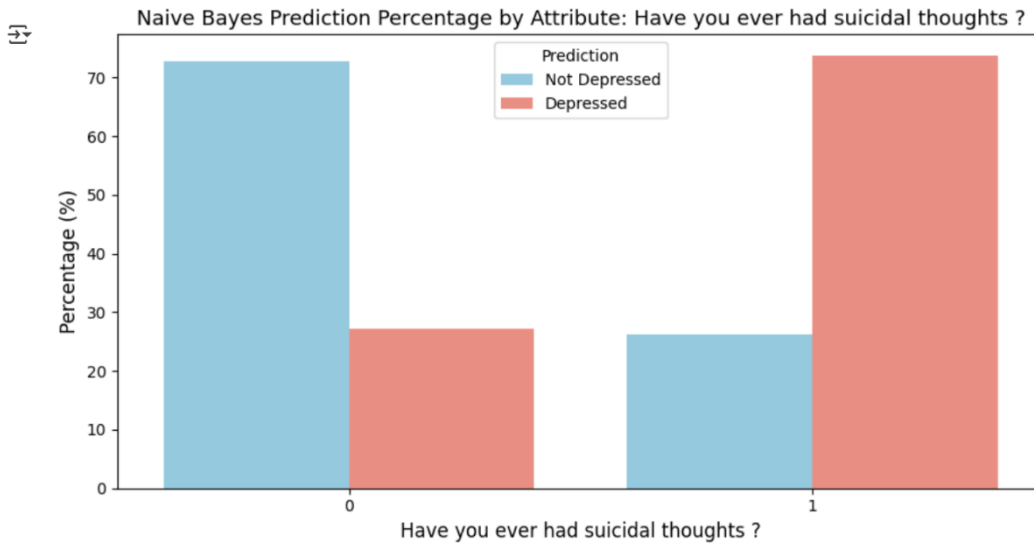
Selanjutnya pada gambar 13, pada atribut *Sleep Duration* atau durasi tidur, hasil visualisasi menunjukkan bahwa mahasiswa dengan durasi tidur ekstrem, baik kurang dari 5 jam (kategori 2) maupun lebih dari 8 jam (kategori 3) lebih banyak diprediksi mengalami depresi. Sementara itu, pada kategori durasi tidur yang dianggap ideal, yaitu 7–8 jam (kategori 1), model tetap memprediksi sebagian besar sebagai depresi, yang mungkin disebabkan oleh pengaruh variabel lain. Temuan ini mendukung literatur yang menyatakan bahwa gangguan pola tidur merupakan indikator penting dalam kesehatan mental, termasuk depresi.



Gambar 14. Implementasi Terhadap Atribut Dietary Habits

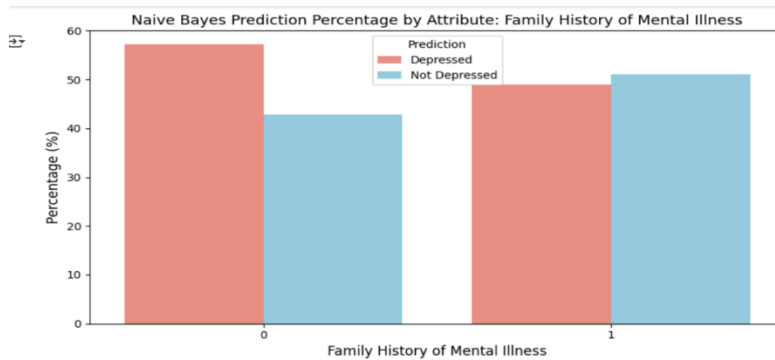
Atribut berikutnya yang divisualisasikan pada Gambar 14 adalah *Dietary Habits*, dimana atribut ini menggambarkan pola makan mahasiswa dalam kesehariannya, yang dikelompokkan dalam beberapa kategori. Berdasarkan hasil visualisasi, terlihat bahwa mahasiswa dengan pola makan tidak sehat atau tidak teratur (kategori 0 dan 1) cenderung diprediksi lebih banyak mengalami depresi dibandingkan tidak depresi. Sebaliknya, pada kategori mahasiswa dengan pola makan yang baik dan teratur (kategori 2 dan 3), proporsi prediksi tidak depresi oleh model lebih tinggi. Temuan ini sejalan dengan beberapa penelitian

terdahulu yang menunjukkan bahwa pola makan berhubungan erat dengan kesehatan mental. Mahasiswa yang tidak menjaga pola makan secara baik cenderung memiliki risiko lebih tinggi mengalami gangguan psikologis, termasuk stres dan depresi. Model Naive Bayes berhasil menangkap korelasi ini dari data, meskipun tentu saja masih ada kemungkinan kesalahan prediksi karena keterbatasan fitur atau pengaruh variabel lain.



Gambar 15. Implementasi Terhadap Atribut Have you ever had suicidal thoughts?

Pada gambar 15, atribut ini merupakan salah satu indikator yang sangat penting dalam deteksi awal gangguan depresi. Dari hasil visualisasi, terlihat bahwa mahasiswa yang menjawab “pernah” (kategori 1) memiliki tingkat prediksi depresi yang sangat tinggi oleh model. Sebagian besar dari mereka diklasifikasikan oleh model sebagai "Depresi", yang menunjukkan bahwa model mampu mengenali bahwa adanya pikiran bunuh diri merupakan indikator kuat terhadap kemungkinan seseorang mengalami depresi. Sementara itu, pada mahasiswa yang tidak pernah memiliki pikiran untuk bunuh diri (kategori 0), prediksi "tidak depresi" lebih mendominasi. Hasil ini menunjukkan bahwa model bekerja cukup akurat dalam mengasosiasikan fitur ini dengan label depresi, karena secara psikologis, adanya ide atau pikiran bunuh diri memang sangat erat kaitannya dengan tingkat keparahan depresi seseorang.

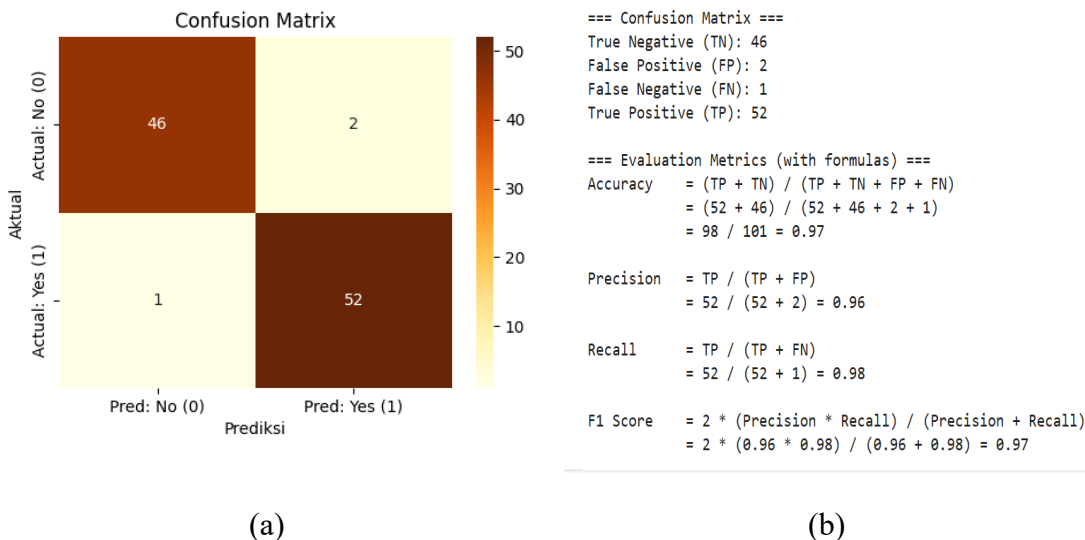


Gambar 16. Implementasi Terhadap Atribut Family History of Mental Illness

Gambar 16 adalah atribut *Family History of Mental Illness* atau riwayat keluarga terhadap penyakit mental. Hasil menunjukkan bahwa mahasiswa yang tidak memiliki riwayat keluarga (kategori 0) justru lebih banyak diklasifikasikan sebagai mengalami depresi. Sebaliknya, mahasiswa yang memiliki riwayat keluarga (kategori 1) lebih banyak diklasifikasikan sebagai tidak depresi. Temuan ini agak bertentangan dengan ekspektasi awal, karena secara teoritis, individu dengan riwayat keluarga penyakit mental lebih rentan terhadap depresi. Ketidaksesuaian ini kemungkinan besar disebabkan oleh ketidakseimbangan jumlah data pada masing-masing kategori, atau bisa juga karena adanya variabel perancu lain yang mempengaruhi hasil prediksi model. Dengan demikian, hasil dari implementasi algoritma Naïve Bayes pada data mahasiswa telah menunjukkan kemampuan yang cukup baik dalam memprediksi kondisi depresi. Model berhasil mempelajari pola-pola yang terdapat pada data latih dan mampu memberikan prediksi yang akurat terhadap data uji. Visualisasi hasil prediksi dan tabel perbandingan menunjukkan bahwa sebagian besar prediksi model sesuai dengan kondisi aktual mahasiswa. Hal ini menandakan bahwa algoritma Naïve Bayes dapat menjadi metode yang efektif dan efisien dalam membantu mengidentifikasi risiko depresi pada mahasiswa berdasarkan atribut-atribut yang dianalisis.

Evaluasi Performa Model

Setelah model Naïve Bayes berhasil dibangun dan diaplikasikan pada data uji, tahap selanjutnya adalah melakukan evaluasi terhadap kinerja model tersebut. Evaluasi ini bertujuan untuk mengukur seberapa baik model dalam memprediksi kondisi depresi pada mahasiswa berdasarkan data yang tersedia. Dengan menggunakan berbagai metrik evaluasi seperti confusion matrix, akurasi, precision, recall, dan F1 score, kita dapat memperoleh gambaran yang lebih komprehensif mengenai kelebihan dan keterbatasan model. Model Naïve Bayes diterapkan menggunakan *library scikit-learn* dengan varian GaussianNB. Model dilatih menggunakan data latih (X_train dan y_train) dan kemudian melakukan prediksi pada data uji (X_test). Hasil prediksi tersebut dibandingkan dengan label aktual (y_test) untuk menghasilkan confusion matrix yang menggambarkan jumlah prediksi benar dan salah secara rinci. Confusion matrix digunakan untuk memperoleh nilai precision, recall, dan accuracy. Nilai Confusion matrix biasanya ditunjukkan dalam satuan persen (%) (Sari et al., 2023). Dan berikut hasil dari evaluasi performa model.



Gambar 17. Confusion Matrix

True Positive adalah jumlah kasus yang benar-benar positif dan berhasil diprediksi sebagai positif oleh model. True Negative adalah jumlah kasus yang benar-benar negatif dan berhasil diprediksi sebagai negatif oleh model. Sementara itu, False Positive adalah jumlah kasus yang sebenarnya negatif tetapi diprediksi sebagai positif oleh model, dan False Negative adalah jumlah kasus yang sebenarnya positif

tetapi diprediksi sebagai negatif oleh model (Subarkah et al., 2022). Pada Gambar 18 (a) dan (b) Confusion matrix memberikan gambaran rinci mengenai prediksi model dibandingkan dengan kondisi aktual. Berdasarkan hasil pengujian, confusion matrix model ini menunjukkan nilai sebagai berikut:

- a. *True Negative* (TN) sebanyak 46, yaitu jumlah data yang diprediksi negatif (tidak depresi) dan memang benar tidak depresi.
- b. *False Positive* (FP) sebanyak 2, yaitu data yang diprediksi positif (depresi), namun sebenarnya negatif.
- c. *False Negative* (FN) sebanyak 1, yaitu data yang diprediksi negatif, namun sebenarnya positif (depresi).
- d. *True Positive* (TP) sebanyak 52, yaitu data yang diprediksi positif dan memang benar mengalami depresi.

Setelah dilakukan pengujian, model menunjukkan hasil yang sangat baik. Berdasarkan *confusion matrix*, model berhasil memprediksi dengan benar sebanyak 46 data yang tidak depresi (*True Negative*) dan 52 data yang benar-benar mengalami depresi (*True Positive*). Hanya ada sedikit kesalahan, yaitu 2 data salah diprediksi sebagai depresi (*False Positive*) dan 1 data yang mengalami depresi tapi tidak terdeteksi (*False Negative*). Dari hasil tersebut, diperoleh nilai akurasi sebesar 97%, presisi 96%, recall 98%, dan F1-score 97%. Nilai-nilai ini menunjukkan bahwa model mampu mengenali kondisi depresi dengan tingkat ketepatan yang tinggi. Rasio 80:20 yang digunakan dalam penelitian ini dianggap ideal karena memberikan keseimbangan antara proses pelatihan dan pengujian model. Dengan cara ini, hasil evaluasi model menjadi lebih akurat dan dapat mencerminkan performa sebenarnya. Dengan adanya evaluasi ini, performa model dapat diukur secara objektif dan menyeluruh. Hal ini tidak ditemukan dalam artikel penelitian sebelumnya, yang hanya menampilkan hasil prediksi dalam bentuk angka persentase untuk satu kasus siswa, tanpa menunjukkan seberapa akurat sistem jika digunakan secara umum. Selain itu, penelitian ini juga menyertakan visualisasi hasil klasifikasi terhadap beberapa atribut penting, seperti jenis kelamin, pola makan, durasi tidur, dan faktor-faktor lainnya, sehingga pembaca dapat memahami pola hubungan antar variabel dengan lebih jelas. Visualisasi ini tidak tersedia dalam penelitian terdahulu yang berbasis sistem pakar.

Dalam proses pengerjaan, terdapat beberapa kendala yang dihadapi. Salah satunya adalah pada tahap pra-pemrosesan data, di mana perlu dilakukan konversi nilai kategorik menjadi bentuk numerik agar bisa diproses oleh model. Sebuah studi mengembangkan model berbasis Naïve Bayes untuk memprediksi status depresi pada mahasiswa, dengan hasil akurasi mencapai 78,03% dan sensitivitas tinggi dalam mendeteksi kasus positif depresi. Hal ini membutuhkan ketelitian agar tidak terjadi kesalahan interpretasi. Selain itu, walaupun model Gaussian Naive Bayes cocok untuk data dengan distribusi normal, proses pengecekan kecocokan data dengan asumsi model tetap menjadi tantangan. Secara keseluruhan, dapat disimpulkan bahwa pendekatan yang digunakan dalam penelitian ini bersifat lebih modern dari penelitian lain yang sejenis, data-driven, dan didukung dengan evaluasi serta visualisasi yang memperkuat hasilnya. Ini menjadi salah satu keunggulan utama dari penelitian ini, karena tidak hanya fokus pada hasil akhir, tetapi juga memperhatikan proses dan kualitas model secara keseluruhan.

SIMPULAN

Penelitian ini berhasil menunjukkan bahwa algoritma Gaussian Naïve Bayes dapat digunakan secara efektif untuk memprediksi kondisi depresi pada mahasiswa. Dataset yang digunakan berisi berbagai atribut yang relevan, seperti durasi tidur, tekanan akademik, kebiasaan makan, hingga riwayat keluarga dengan gangguan mental. Semua atribut ini saling berkontribusi dalam membantu model mengenali pola yang mengarah pada kondisi depresi. Model yang dikembangkan melalui Google Colab dengan menggunakan pustaka scikit-learn berhasil mencapai performa yang sangat baik. Berdasarkan hasil evaluasi, model memperoleh akurasi sebesar 97%, precision sebesar 96%, recall sebesar 98%, dan F1-

score sebesar 97%. Ini menunjukkan bahwa model sangat akurat dalam mengklasifikasikan apakah seorang mahasiswa mengalami depresi atau tidak. Salah satu poin penting yang ditemukan dalam penelitian ini adalah bahwa atribut “pernah memiliki pikiran bunuh diri” dan “durasi tidur ekstrem” memiliki pengaruh besar dalam penentuan hasil klasifikasi. Model juga cukup andal dalam mengidentifikasi mahasiswa yang berada dalam kondisi rawan tanpa harus bergantung pada pendekatan medis tradisional. Secara keseluruhan, penerapan algoritma Naïve Bayes dalam penelitian ini tidak hanya menunjukkan keakuratan yang tinggi, tetapi juga membuktikan bahwa teknologi dan machine learning dapat berperan aktif dalam mendukung kesehatan mental, terutama pada kalangan mahasiswa yang selama ini cenderung menyembunyikan kondisi psikologis mereka karena stigma atau rasa malu.

DAFTAR PUSTAKA

- American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association Publishing. <https://doi.org/10.1176/appi.books.9780890425787>
- Devi Nurhayati, S., & Widayani, W. (2021). Sistem Rekomendasi Wisata Kuliner di Yogyakarta dengan Metode Item-Based Collaborative Filtering Yogyakarta Culinary Recommendation System with Item-Based Collaborative Filtering Method. In *JACIS : Journal Automation Computer Information System* (Vol. 1, Issue 2). <https://manganenakyog.my.id/>,
- Diva Yuda, T., Sarifah, F., Syaefanhas Mulyadi, D., Sri Rahma Ayu, W., Nadhir Fasya, N., Adityas Noerdiansyah, I., & Teknik Sipil, J. (2024). Tresna Diva Yuda: Analisis Kontribusi Perusahaan terhadap ... Analisis Kontribusi Perusahaan terhadap Sustainable Development Goals (SDGs). In *Indonesian Journal of Engineering and Technology (INAJET)* (Vol. 6, Issue 2). <https://journal.unesa.ac.id/index.php/inajet>
- Gama, A. W. O., Dennatan, M., Dharmayasa, I. G. N. P., Maw, M. M., Sugiana, I. P., & Suryanti, I. (2025). Identifying Key Factors Causing Flooding Using Machine Learning. *Journal of Applied Data Sciences*, 6(1), 115–130. <https://doi.org/10.47738/jads.v6i1.463>
- Gama, A. W. O., & Wardhiana, I. N. G. A. M. (2023). Naïve bayes on diagnostic expert system for menstrual disorders. In *Journal of Intelligent Decision Support System (IDSS)* (Vol. 6, Issue 2). <https://doi.org/https://doi.org/10.35335/idss.v6i2.130>
- Hasbie, R., Wahiddin, D., & Ratna Juwita, A. (2023). *Algoritma Certainty Factor Untuk Diagnosa Penyakit Depresi Pada Remaja. IV(1)*.
- Nurabsharina, A. P., & Kosasih, R. (2020). APLIKASI SISTEM PAKAR DIAGNOSIS TINGKAT DEPRESI PADA REMAJA BERBASIS ANDROID. *Jurnal Ilmiah Informatika Komputer*, 25(1), 76–85. <https://doi.org/10.35760/ik.2020.v25i1.2418>
- Ridha Dwiki Putri, D., Reza Fahlevi, M., Sadikin, M., Utami, R., Rizki Fajar Utomo, M., & Studi Rekayasa Perangkat Lunak, P. (2024). *Prediksi Tingkat Depresi Remaja Menggunakan Metode Naïve Bayes Classifier: Analisis Faktor Psikologis Dan Lingkungan*. 5(4), 2034–2043.
- Rully Desthian Pahlephi. (2022). Depresi Adalah: Penyebab, Gejala, dan Cara Mengatasinya. *Detik.Com*.

- Sari, L., Romadloni, A., Lityaningrum, R., & Hastuti, H. D. (2023). Implementation of LightGBM and Random Forest in Potential Customer Classification. *TIERS Information Technology Journal*, 4(1), 43–55. <https://doi.org/10.38043/tiers.v4i1.4355>
- Subarkah, P., Solikhatin, S. A., Darmayanti, I., Ikhsan, A. N., Hidayah, D. U., & Anjani, R. M. (2022). Prediction of Education Level in Population Data Using Naïve Bayes Algorithm. *TIERS Information Technology Journal*, 3(2), 69–75. <https://doi.org/10.38043/tiers.v3i2.3865>
- Surya Wisnugraha, W., Farida, I. N., & Widyadara, M. A. D. (2023). Implementasi Algoritma Naïve Bayes Dalam Menentukan Diagnosa Tingkat Depresi Mahasiswa Akhir Terhadap Pengerjaan Skripsi. In *Agustus* (Vol. 7). Online.
- Tyas, D. M., Pertiwi, A., & Nisa, V. Z. (2023). Sosialisasi Peningkatan Pemahaman Kesehatan Mental Pada Remaja. *Jurnal Pengabdian Masyarakat Bangsa*, 1(10). <https://jurnalpengabdianmasyarakatbangsa.com/index.php/jpmmba/article/view/567/458>
- Tengku Omri Wikana, Tioria Pasaribu, & Hotler Manurung. (2024). Penerapan Metode Naïve Bayes untuk Memprediksi Tingkat Kesehatan Mental Siswa Menjelang Akhir Masa Sekolah. *Saturnus : Jurnal Teknologi Dan Sistem Informasi*, 2(4), 296–306. <https://doi.org/10.61132/saturnus.v2i4.364>
- Zuriah, T. S., & Wardono, G. (2023). HUBUNGAN ANTARA ALEXITHYMIA DENGAN DEPRESI PADA REMAJA. *Blantika: Multidisciplinary Journal*, 2(1). <https://blantika.publikasiku.id/>