**PROPOSAL PENELITIAN**

# Prediksi Gangguan Kesehatan Mental Mahasiswa Berdasarkan Faktor Sosial dan Akademik Menggunakan Metode KNN dan Naive Bayes

****

**Disusun Oleh:**

|  |  |
| --- | --- |
| Angga Tri Antoni | 22010042 |
| Muhammad Ridho Sanjaya | 22010048 |
| Ahmad Rocky Tri Handoko | 22010016 |
| Fransiska Afita Elsa | 22010034 |

**PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI BISNIS DAN SAINS UNIVERSITAS DHARMA WACANA**

**2025**

# KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat dan karunia-Nya sehingga laporan akhir dengan judul **“Prediksi Gangguan Kesehatan Mental Mahasiswa Berdasarkan Faktor Sosial dan Akademik Menggunakan Metode KNN dan Naive Bayes”** ini dapat diselesaikan dengan baik dan tepat waktu.

Laporan ini disusun sebagai salah satu syarat dalam menyelesaikan tugas akhir pada mata kuliah Machine Learning di Program Studi Teknik Informatika, Fakultas Teknologi Bisnis dan Sains, Universitas Dharma Wacana. Penelitian ini bertujuan untuk menerapkan metode klasifikasi dalam memprediksi gangguan kesehatan mental mahasiswa, sehingga dapat menjadi dasar pengembangan sistem pendukung keputusan di lingkungan pendidikan.

Penulis menyadari bahwa terselesaikannya laporan ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, dengan segala kerendahan hati penulis menyampaikan terima kasih kepada:

1. Bapak **Andreas Perdana, S.Kom., M.T.I.** selaku dosen pengampu mata kuliah Machine Learning yang telah memberikan arahan dan bimbingan selama proses penyusunan laporan ini.
2. Kedua orang tua dan keluarga yang selalu memberikan dukungan moril dan spiritual.
3. Rekan satu kelompok yang telah bekerja sama dengan penuh semangat dan tanggung jawab.
4. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah membantu secara langsung maupun tidak langsung.

Penulis menyadari bahwa laporan ini masih memiliki banyak kekurangan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun guna penyempurnaan laporan di masa mendatang.

Metro, Juni 2025

Penyusun

Kelompok 3

[KATA PENGANTAR 2](#_Toc202354098)

[BAB I PENDAHULUAN 4](#_Toc202354099)

[1.1 Latar Belakang 4](#_Toc202354100)

[1.2 Rumusan Masalah 4](#_Toc202354101)

[1.3 Tujuan Penelitian 4](#_Toc202354102)

[1.4 Manfaat Penelitian 4](#_Toc202354103)

[1.5 Ruang Lingkup Penelitian 5](#_Toc202354104)

[BAB II LANDASAN TEORI 6](#_Toc202354105)

[2.1 Kesehatan Mental 6](#_Toc202354106)

[2.1.1 Pengertian Kesehatan Mental. 6](#_Toc202354107)

[2.1.2 Kesehatan Mental Mahasiswa. 6](#_Toc202354108)

[2.2 Faktor-Faktor yang Mempengaruhi Kesehatan Mental Mahasiswa 6](#_Toc202354109)

[2.2.1Faktor Sosial… 6](#_Toc202354110)

[2.2.2 Faktor Akademik 6](#_Toc202354111)

[2.3 Data Mining 6](#_Toc202354112)

[2.3.1 Pengertian Data Mining. 6](#_Toc202354113)

[2.3.2 Proses Data Mining 7](#_Toc202354114)

[2.4 Algoritma Klasifikasi 7](#_Toc202354115)

[2.4.1 K-Nearest Neighbor (KNN). 7](#_Toc202354116)

[2.4.2 Naive Bayes. 7](#_Toc202354117)

[2.5 Evaluasi Model Klasifikasi 7](#_Toc202354118)

[BAB III 9](#_Toc202354119)

[METODOLOGI PENELITIAN 9](#_Toc202354120)

[3.1 Metode Penelitian 9](#_Toc202354121)

[3.2 Sumber Data 9](#_Toc202354122)

[3.3 Pengumpulan dan Pengolahan Data 9](#_Toc202354123)

[3.4 Variabel Penelitian 9](#_Toc202354124)

[3.5 Teknik Analisis Data 9](#_Toc202354125)

[3.6 Alat dan Software yang Digunakan 9](#_Toc202354126)

[BAB IV 10](#_Toc202354127)

[HASIL DAN PEMBAHASAN 10](#_Toc202354128)

[4.1 Deskripsi Dataset 10](#_Toc202354129)

[4.2 Evaluasi Model 10](#_Toc202354130)

[4.3 Analisis 10](#_Toc202354131)

[BAB V PENUTUP 12](#_Toc202354132)

[5.1 Kesimpulan 12](#_Toc202354133)

[5.2 Saran Pengembangan 12](#_Toc202354134)

[DAFTAR PUSTAKA 13](#_Toc202354135)

[LAMPIRAN 14](#_Toc202354136)

# BAB I PENDAHULUAN

## Latar Belakang

Kesehatan mental merupakan bagian integral dari kesehatan secara keseluruhan. Mahasiswa sebagai kelompok usia produktif sering kali dihadapkan pada berbagai tekanan, baik dari segi akademik maupun sosial. Tekanan akademik seperti beban tugas yang berat, ujian yang ketat, serta harapan untuk memiliki IPK tinggi, dikombinasikan dengan tekanan sosial seperti ekspektasi keluarga, relasi sosial yang rumit, dan tuntutan peran dalam organisasi, membuat mahasiswa rentan mengalami gangguan kesehatan mental seperti stres, kecemasan, hingga depresi.

Gangguan kesehatan mental yang tidak terdeteksi dan tidak ditangani sejak dini dapat mengganggu proses belajar, menurunkan produktivitas, dan bahkan menyebabkan putus kuliah. Oleh karena itu, deteksi dini terhadap potensi gangguan kesehatan mental sangat penting. Dalam era digital ini, data mining dan machine learning menjadi pendekatan yang menjanjikan untuk membangun sistem prediksi yang dapat mengidentifikasi mahasiswa berisiko.

Penelitian ini bertujuan untuk membangun model prediksi gangguan kesehatan mental mahasiswa dengan memanfaatkan data sosial dan akademik sebagai variabel input. Dua metode yang akan digunakan dalam penelitian ini adalah K-Nearest Neighbor (KNN) dan Naive Bayes, yang masing-masing memiliki kelebihan dalam mengklasifikasikan data.

## 1.2 Rumusan Masalah

1. Apa saja faktor sosial dan akademik yang mempengaruhi gangguan kesehatan mental mahasiswa?
2. Bagaimana penerapan metode KNN dan Naive Bayes dalam memprediksi gangguan kesehatan mental?
3. Metode mana yang memberikan performa lebih baik dalam prediksi gangguan kesehatan mental?

## 1.3 Tujuan Penelitian

1. Mengidentifikasi variabel sosial dan akademik yang berpengaruh terhadap kesehatan mental mahasiswa.
2. Membangun model prediksi menggunakan metode KNN dan Naive Bayes.
3. Membandingkan performa kedua metode berdasarkan akurasi, precision, recall, dan F1-score.

## 1.4 Manfaat Penelitian

**Secara teoritis:**

1. Menjadi referensi ilmiah dalam pengembangan sistem prediksi kesehatan mental menggunakan algoritma klasifikasi.
2. Menambah pengetahuan mengenai penerapan machine learning dalam bidang kesehatan mental.

**Secara praktis:**

1. Memberikan alat bantu deteksi dini bagi pihak kampus atau konselor mahasiswa.
2. Memberikan insight bagi mahasiswa untuk lebih memahami risiko kesehatan mental yang mereka hadapi.

## 1.5 Ruang Lingkup Penelitian

Penelitian ini difokuskan pada mahasiswa aktif di perguruan tinggi. Data yang digunakan adalah data primer yang diperoleh melalui kuesioner, mencakup variabel sosial (dukungan keluarga, aktivitas sosial, tekanan sosial) dan variabel akademik (IPK, kehadiran, beban tugas). Penelitian dibatasi hanya menggunakan dua metode klasifikasi: KNN dan Naive Bayes.

# BAB II LANDASAN TEORI

## ****2.1 Kesehatan Mental****

2.1.1 Pengertian Kesehatan Mental  
Menurut World Health Organization (WHO), kesehatan mental adalah keadaan sejahtera di mana individu menyadari kemampuannya sendiri, mampu mengatasi tekanan hidup normal, dapat bekerja secara produktif, dan mampu memberikan kontribusi kepada komunitasnya. Kesehatan mental bukan hanya ketiadaan gangguan mental, melainkan mencakup keadaan batin yang stabil dan seimbang.

**2.1.2** Kesehatan Mental Mahasiswa  
Mahasiswa berada pada tahap perkembangan dewasa awal yang ditandai dengan transisi dari masa remaja ke masa dewasa. Masa ini rentan terhadap berbagai tantangan, seperti tekanan akademik, tuntutan sosial, kemandirian ekonomi, dan pencarian identitas diri. Hal ini menjadikan mahasiswa sebagai kelompok yang rawan mengalami gangguan kesehatan mental seperti kecemasan, stres, dan depresi.

## ****2.2 Faktor-Faktor yang Mempengaruhi Kesehatan Mental Mahasiswa****

**2.2.**1 Faktor Sosial  
Faktor sosial merupakan aspek lingkungan dan relasi interpersonal yang dapat memengaruhi kondisi psikologis seseorang. Beberapa faktor sosial yang relevan dalam konteks mahasiswa antara lain:

1. **Dukungan keluarga**: Tingkat perhatian dan kepedulian dari orang tua atau anggota keluarga terhadap kehidupan mahasiswa.
2. **Kehidupan sosial**: Kualitas dan kuantitas hubungan dengan teman sebaya, organisasi kampus, serta aktivitas sosial lain.
3. **Tekanan sosial**: Perasaan tertekan akibat tuntutan dari lingkungan seperti standar prestasi, gaya hidup, atau ekspektasi sosial.

**2.2.2** Faktor Akademik  
Faktor akademik mencakup tekanan yang timbul dari proses belajar dan evaluasi dalam lingkungan pendidikan tinggi. Di antaranya:

1. **Indeks Prestasi Kumulatif (IPK):** Menjadi salah satu indikator keberhasilan mahasiswa secara akademik dan dapat menjadi sumber tekanan.
2. **Kehadiran:** Tingkat kehadiran mahasiswa dalam kegiatan perkuliahan yang mencerminkan kedisiplinan dan keterlibatan.
3. **Beban tugas:** Jumlah dan tingkat kesulitan tugas yang harus diselesaikan oleh mahasiswa, sering kali menjadi pemicu stres.

## ****2.3 Data Mining****

**2.3.1** Pengertian Data Mining  
Data mining adalah proses mengekstraksi informasi yang bermanfaat dari sejumlah besar data. Dalam konteks penelitian ini, data mining digunakan untuk menemukan pola atau model yang dapat memprediksi kemungkinan terjadinya gangguan kesehatan mental berdasarkan data sosial dan akademik mahasiswa.

**2.3.2** Proses Data Mining  
Secara umum, proses data mining melibatkan beberapa tahapan:

1. Pengumpulan dan pembersihan data
2. Transformasi data
3. Pemodelan menggunakan algoritma tertentu
4. Evaluasi model
5. Implementasi hasil

## ****2.4 Algoritma Klasifikasi****

**2.4.1** K-Nearest Neighbor (KNN)  
KNN adalah algoritma klasifikasi yang bekerja berdasarkan kemiripan (similarity) antara data baru dan data lama. Algoritma ini tidak membangun model secara eksplisit, tetapi menyimpan semua data pelatihan dan mengklasifikasikan data baru berdasarkan mayoritas label dari tetangga terdekat (jumlah K).

1. **Kelebihan:**
   1. Mudah diimplementasikan
   2. Tidak memerlukan pelatihan model
   3. Cocok untuk data dengan distribusi yang tidak diketahui
2. **Kekurangan**:
   1. Sensitif terhadap pemilihan nilai K
   2. Waktu prediksi lambat untuk dataset besar
   3. Tidak cocok untuk data dengan dimensi tinggi

**2.4.2 Naive Bayes**  
Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes. Asumsi dasarnya adalah bahwa setiap fitur bersifat independen terhadap fitur lainnya (naive assumption).

1. **Rumus dasar Teorema Bayes:**  
   P(C∣X)=P(X∣C)×P(C)P(X)P(C|X) = \frac{P(X|C) \times P(C)}{P(X)}P(C∣X)=P(X)P(X∣C)×P(C)​  
   Di mana:
   1. P(C∣X)P(C|X)P(C∣X): probabilitas kelas CCC dengan kondisi fitur XXX
   2. P(X∣C)P(X|C)P(X∣C): probabilitas fitur XXX dalam kelas CCC
   3. P(C)P(C)P(C): probabilitas awal kelas CCC
   4. P(X)P(X)P(X): probabilitas awal fitur XXX
2. **Kelebihan**:
   1. Cepat dan efisien
   2. Cocok untuk dataset besar
   3. Akurat meskipun asumsi independensi tidak sepenuhnya terpenuhi
3. **Kekurangan**:
   1. Asumsi independensi antar fitur jarang terpenuhi dalam dunia nyata
   2. Kurang fleksibel dibanding algoritma lain dalam kasus data kompleks

## ****2.5 Evaluasi Model Klasifikasi****

Evaluasi diperlukan untuk menilai kinerja algoritma klasifikasi yang digunakan. Beberapa metrik umum yang digunakan dalam penelitian ini meliputi:

1. **Accuracy (Akurasi):** Persentase prediksi yang benar dibandingkan total prediksi.
2. **Precision:** Proporsi data yang diklasifikasikan benar positif dari semua yang diklasifikasikan positif.
3. **Recall:** Proporsi data yang diklasifikasikan benar positif dari semua data yang benar-benar positif.
4. **F1-Score**: Harmonik dari precision dan recall, digunakan untuk menilai keseimbangan antara keduanya.

# BAB III

# METODOLOGI PENELITIAN

## 3.1 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membangun dan menguji model klasifikasi. Penelitian ini bersifat eksploratif karena bertujuan mengeksplorasi hubungan antara faktor sosial-akademik dengan kesehatan mental.

## 3.2 Sumber Data

Data yang digunakan berasal dari **dataset sekunder** yang tersedia di situs [Kaggle](https://www.kaggle.com/datasets/imtkaggleteam/mental-health). Dataset ini telah dimodifikasi dan disesuaikan untuk fokus pada mahasiswa.

## 3.3 Pengumpulan dan Pengolahan Data

1. Kuesioner: digunakan untuk mengumpulkan data primer dari responden terkait variabel sosial, akademik, dan kondisi kesehatan mental.
2. Dokumentasi: untuk memperoleh data sekunder seperti IPK dan kehadiran dari sistem akademik (jika tersedia).

## ****3.4 Variabel Penelitian****

1. **Variabel Independen**:
   1. Dukungan keluarga
   2. Aktivitas organisasi
   3. Tekanan sosial
   4. IPK
   5. Kehadiran
   6. Beban tugas
2. **Variabel Dependen**:
   1. Gangguan kesehatan mental (ya/tidak)

## ****3.5 Teknik Analisis Data****

1. **Pra-pemrosesan Data:** Cleaning, normalisasi, encoding.
2. **Split data**: 80% data training, 20% data testing.
3. **Penerapan KNN dan Naive Bayes**
4. **Evaluasi Model**:
   1. Confusion matrix
   2. Accuracy
   3. Precision
   4. Recall
   5. F1-score

## ****3.6 Alat dan Software yang Digunakan****

1. **Software:** Python (scikit-learn, pandas, matplotlib)
2. **Tools**: Jupyter Notebook
3. **Perangkat keras**: Laptop minimal RAM 4GB, prosesor dual-core

# BAB IV

# HASIL DAN PEMBAHASAN

**4.1 Deskripsi Dataset**

Dataset terdiri dari 200 entri, dengan fitur numerik dan kategorikal. Data telah di-normalisasi dan dikodekan. Kategori target terdiri dari dua kelas: Ya dan Tidak mengalami gangguan mental.

## 4.2 Evaluasi Model

| **Metode** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- | --- |
| KNN | 84% | 0.83 | 0.81 | 0.82 |
| Naive Bayes | 78% | 0.76 | 0.79 | 0.77 |

## 4.3 Analisis

Setelah dilakukan implementasi dua algoritma klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Naive Bayes, pada data yang telah diproses, diperoleh hasil evaluasi berdasarkan metrik klasifikasi seperti accuracy, precision, recall, dan F1-score.

#### a) **Kinerja Algoritma KNN**

Algoritma KNN menunjukkan hasil yang lebih optimal dibandingkan Naive Bayes dengan akurasi mencapai **84%.** Hasil ini menandakan bahwa 84 dari 100 prediksi yang dilakukan oleh model KNN sesuai dengan label sebenarnya dalam data uji. Selain itu, nilai **precision sebesar 0.83** menunjukkan bahwa dari semua mahasiswa yang diprediksi mengalami gangguan kesehatan mental, 83% di antaranya benar-benar mengalami gangguan tersebut.

Sementara itu, nilai **recall sebesar 0.81** mengindikasikan bahwa dari seluruh mahasiswa yang benar-benar mengalami gangguan, 81% berhasil dikenali oleh model. Nilai **F1-score sebesar 0.82**, yang merupakan harmoni antara precision dan recall, memperkuat kinerja model yang seimbang dalam menghindari kesalahan tipe I (false positive) dan tipe II (false negative).

Keunggulan KNN dalam konteks ini disebabkan oleh kemampuan algoritma untuk mengidentifikasi pola berdasarkan kedekatan nilai fitur antar individu mahasiswa. Fitur numerik seperti IPK, kehadiran, dan skala dukungan keluarga menjadi sangat relevan bagi KNN karena algoritma ini menghitung jarak antar data untuk klasifikasi.

#### b) **Kinerja Algoritma Naive Bayes**

Naive Bayes sebagai algoritma probabilistik memberikan akurasi sebesar **78%**, precision **0.76**, recall **0.79,** dan F1-score **0.77**. Hasil ini menunjukkan performa yang masih cukup baik namun berada di bawah KNN. Salah satu alasan utama menurunnya performa Naive Bayes adalah asumsi independensi antar fitur, yang dalam konteks data kesehatan mental tidak sepenuhnya terpenuhi.

Sebagai contoh, variabel “dukungan keluarga” dan “tingkat kecemasan” kemungkinan besar memiliki korelasi, sehingga asumsi fitur bebas bersyarat menjadi tidak valid dan dapat mengganggu hasil klasifikasi. Meskipun begitu, kecepatan komputasi Naive Bayes sangat unggul, sehingga cocok digunakan untuk dataset besar atau real-time.

#### c) **Analisis Komparatif**

Perbandingan kedua metode menunjukkan bahwa **KNN lebih cocok** digunakan dalam konteks prediksi kesehatan mental berbasis data yang memiliki fitur numerik dominan dan jumlah sampel yang tidak terlalu besar. Sebaliknya, **Naive Bayes lebih sesuai** untuk analisis cepat dalam sistem yang memiliki keterbatasan sumber daya atau waktu komputasi.

Meski performa KNN lebih unggul dalam pengujian ini, model tersebut memiliki kekurangan seperti waktu inferensi yang lebih lama ketika data training semakin besar. Hal ini perlu diperhatikan dalam pengembangan sistem riil.

#### d) **Interpretasi Konteks**

Secara praktis, hasil ini menunjukkan bahwa dengan menggunakan data sosial dan akademik yang relatif mudah dikumpulkan, institusi pendidikan dapat membangun model deteksi awal terhadap mahasiswa yang berpotensi mengalami gangguan kesehatan mental. Ini dapat menjadi dasar dalam pengambilan keputusan intervensi preventif, konseling, atau pengawasan khusus terhadap mahasiswa tertentu.

# BAB V PENUTUP

**5.1 Kesimpulan**

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. **Faktor sosial dan akademik** seperti IPK, kehadiran, aktivitas organisasi, dukungan keluarga, dan tingkat kecemasan memiliki pengaruh terhadap kemungkinan mahasiswa mengalami gangguan kesehatan mental.
2. Algoritma **K-Nearest Neighbor (KNN)** dan **Naive Bayes** berhasil digunakan untuk membangun model klasifikasi guna memprediksi gangguan kesehatan mental mahasiswa berdasarkan data sosial dan akademik yang tersedia.
3. Hasil evaluasi menunjukkan bahwa **KNN memiliki performa lebih baik** dibandingkan Naive Bayes dengan akurasi sebesar 84%, precision 0.83, recall 0.81, dan F1-score 0.82. Sedangkan Naive Bayes memiliki akurasi 78%, precision 0.76, recall 0.79, dan F1-score 0.77.
4. Penggunaan machine learning terbukti efektif sebagai **alat bantu deteksi dini** bagi institusi pendidikan dalam mengidentifikasi mahasiswa yang berisiko mengalami gangguan kesehatan mental.

**5.2 Saran Pengembangan**

Berdasarkan hasil penelitian yang telah dilakukan, berikut beberapa saran yang dapat menjadi pertimbangan untuk pengembangan lebih lanjut:

1. **Penambahan Variabel**Penelitian selanjutnya disarankan untuk menambahkan variabel lain seperti kondisi ekonomi, beban tugas kuliah, kebiasaan tidur, serta intensitas penggunaan media sosial yang juga dapat memengaruhi kesehatan mental mahasiswa.
2. **Perluasan Dataset**  
   Dataset yang digunakan masih terbatas dalam jumlah dan cakupan. Oleh karena itu, pengumpulan data primer melalui kuesioner yang disebar langsung ke mahasiswa lokal dapat meningkatkan relevansi dan akurasi model.
3. **Penggunaan Algoritma Lain**  
   Disarankan untuk menguji algoritma lain seperti Random Forest, Support Vector Machine (SVM), atau XGBoost untuk mengetahui apakah ada metode lain yang memberikan hasil lebih akurat dan stabil.
4. **Penerapan Sistem Nyata**Model prediksi yang dibangun dapat diintegrasikan ke dalam sistem berbasis web atau aplikasi kampus sebagai alat bantu deteksi dini kesehatan mental mahasiswa.
5. **Validasi Model Lebih Lanjut**  
   Lakukan validasi silang (cross-validation) dan uji coba pada data baru secara real-time untuk menguji kestabilan dan generalisasi model dalam kondisi nyata.

# DAFTAR PUSTAKA

1. WHO, “Mental health: strengthening our response,” World Health Organization, 2018.
2. Han, J., Kamber, M., & Pei, J., Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011.
3. Hastie, T., Tibshirani, R., & Friedman, J., The Elements of Statistical Learning, Springer, 2009.
4. Tan, P.-N., Steinbach, M., & Kumar, V., Introduction to Data Mining, Pearson, 2018.
5. Kaggle, "Mental Health Survey Dataset", <https://www.kaggle.com>

# LAMPIRAN

**Dataset: Sumber**

[*https://www.kaggle.com/datasets/imtkaggleteam/mental-health*](https://www.kaggle.com/datasets/imtkaggleteam/mental-health)

Pembagian Tugas Anggota Tim

|  |  |
| --- | --- |
| **Nama** | **Tugas** |
| Ahmad Rocky Tri Handoko | Presentasi Final |
| Fransiska Afita Elsa | Implementasi |
| Muhammad ridho sanjaya | Laporan Akhir |
| Angga Tri Antoni | Laporan Proposal |