

Optimasi Prediksi Risiko Gangguan Kesehatan Mental Menggunakan Hybrid Method: Naïve Bayes dan Relief-F

Chairunnisyah Widi Pratiwi¹, P.P.P.A.N.W.Fikrul Ilmi R.H.Zer²

¹Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

²Program Studi Manajemen Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

E-Mail : chairunnisyah.w.prtw@gmail.com¹ , fikrul@amiktunasbangsa.ac.id²

Article Info

Article history:

Received Jun 15, 2025

Revised Jul 14, 2025

Accepted Aug 02, 2025

Kata Kunci:

Kesehatan Mental

Naive Bayes

Relief-F

Prediksi Risiko

Machine Learning

Keywords:

Mental Health

Naïve Bayes

Relief-F

Risk Prediction

Machine Learning

ABSTRAK

Masalah gangguan kesehatan mental terus mengalami peningkatan secara global dan menjadi isu yang mendesak untuk ditangani, terutama di negara berkembang. Penelitian ini bertujuan untuk mengoptimasi prediksi risiko gangguan kesehatan mental dengan menerapkan pendekatan hybrid yang menggabungkan algoritma *Naive Bayes* dan teknik seleksi fitur *Relief-F*. Dataset yang digunakan berasal dari kompetisi data publik "Mental Health Data Competition" di Kaggle, dengan lebih dari 140.000 data yang telah melalui tahapan *preprocessing*, *encoding*, dan normalisasi. Metode *Relief-F* digunakan untuk menyeleksi fitur yang paling relevan, lalu fitur tersebut dijadikan input untuk algoritma *Naive Bayes* dalam membangun model prediksi. Hasil evaluasi model menunjukkan akurasi sebesar 90,35% dengan F1-score rata-rata 91%, yang menandakan performa tinggi dalam mengklasifikasikan individu berisiko depresi. Penelitian ini memberikan kontribusi signifikan dalam mendukung deteksi dini gangguan mental melalui pendekatan klasifikasi yang efisien dan akurat.

ABSTRACT

Mental health disorders continue to increase globally and have become a pressing issue, especially in developing countries. This study aims to optimize the prediction of mental health disorder risks by applying a hybrid approach that combines the Naive Bayes algorithm with the Relief-F feature selection technique. The dataset used was sourced from the public "Mental Health Data Competition" on Kaggle, comprising over 140,000 records that underwent preprocessing, encoding, and normalization. The Relief-F method was employed to select the most relevant features, which were then used as input for the Naive Bayes algorithm to build the prediction model. The model evaluation yielded an accuracy of 90.35% and an average F1-score of 91%, indicating strong performance in classifying individuals at risk of depression. This research contributes significantly to supporting early detection of mental health disorders through an efficient and accurate classification approach.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

Chairunnisyah Widi Pratiwi,

Program Studi Sistem Informasi, STIKOM Tunas Bangsa,

Jl. Jend. Sudirman Blok A No. 1,2&3, Pematangsiantar, Indonesia,

Email: chairunnisyah.w.prtw@gmail.com

1. PENDAHULUAN

Kesehatan mental merupakan bagian penting dari kesejahteraan manusia, tetapi sering kali tidak mendapatkan perhatian yang cukup, padahal berperan besar dalam menentukan kualitas hidup seseorang. Aspek mental dan fisik memiliki hubungan erat yang tidak bisa dipisahkan. Berdasarkan data dari *World Health Organization* (WHO), Hampir satu miliar orang di seluruh dunia menderita gangguan kesehatan mental,

dengan depresi dan kecemasan sebagai dua jenis gangguan yang paling banyak dialami (Khairan & Habib, 2024)(Ramdani et al., 2024). Ketika seseorang mengalami gangguan kesehatan mental, mereka mungkin mengalami kesulitan dalam berkonsentrasi, perubahan suasana hati yang drastis, serta kesulitan dalam mengendalikan emosi, yang berpotensi menyebabkan perilaku destruktif (Rahmawati et al., 2025)(Mayasari et al., 2024). Prevalensi gangguan kesehatan mental terus meningkat dari tahun ke tahun (Reni Handayani et al., 2021). Pengalaman hidup yang ekstrem, seperti kehilangan orang terkasih, stres berat, dan trauma psikologis, sering kali menjadi pemicu utama masalah kesehatan mental (Sari & Susmiatin, 2023). Di era globalisasi dan industri maju ini, masyarakat harus mampu menghadapi berbagai tantangan yang ada (Bali & Hajriyah, 2020). Berbagai tekanan dalam kehidupan modern juga menjadi faktor pemicu masalah kesehatan mental (Wijaya & Rachmat, 2025). Tekanan psikologis dapat didefinisikan sebagai kondisi penderitaan emosional yang ditandai dengan kombinasi gejala depresi (Gebremedhin et al., 2020). Masalah yang ada semakin buruk karena kurangnya pemahaman tentang kesehatan mental (Ferdian et al., 2024). Kesadaran masyarakat akan pentingnya kesehatan mental masih rendah, khususnya di negara berkembang seperti Indonesia. Stigma terhadap gangguan mental, pemahaman yang terbatas, serta akses yang sulit terhadap layanan kesehatan mental sering kali menjadi hambatan bagi individu untuk mendapatkan dukungan yang mereka perlukan (Lisma Nanda, 2025). Maka dari itu, deteksi dini dan edukasi sangat penting untuk meningkatkan kewaspadaan terhadap gangguan kesehatan mental yang bisa membahayakan individu (Purwanti et al., 2025).

Beberapa penelitian mengemukakan bahwa faktor-faktor eksternal seperti kemiskinan, kekerasan, diskriminasi, dan ketidaksetaraan sosial memainkan peran signifikan dalam meningkatkan risiko individu mengalami gangguan kesehatan mental. Peneliti lain menyatakan Stres bisa dipicu oleh berbagai faktor, seperti manajemen waktu yang buruk, masalah keuangan, kurang tidur, aktivitas sosial yang berlebihan, serta kewajiban dan tanggung jawab tertentu. Dari Penelitian lain menyatakan penggunaan media sosial yang intens dapat meningkatkan kecemasan dan depresi. Serta lingkungan keluarga yang penuh konflik, kurangnya perhatian emosional, atau pola komunikasi yang buruk dapat memperburuk kondisi kesehatan mental seseorang. Penelitian yang telah dilakukan menunjukkan bahwa gangguan kesehatan mental dapat disebabkan oleh beberapa faktor, termasuk kondisi ekonomi, diskriminasi, manajemen waktu, kualitas tidur, aktivitas sosial, jumlah tanggung jawab, penggunaan media sosial, dan lingkungan keluarga.

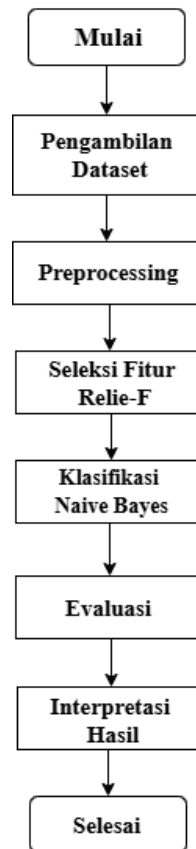
World Health Organization (WHO) menyatakan bahwa kesehatan mental tidak hanya berarti terbebas dari penyakit, tetapi juga mencakup kondisi fisik, psikologis, dan sosial yang seimbang. Meskipun kesehatan mental merupakan aspek krusial dalam kesehatan masyarakat, perhatian terhadapnya masih lebih rendah dibandingkan dengan kesehatan fisik di banyak negara (Halik et al., 2024). Berdasarkan survei yang dilakukan oleh WHO, gangguan kesehatan mental berkontribusi terhadap lebih dari 12% dari seluruh penyakit secara global, dengan 450 juta individu mengalami berbagai kondisi. Secara lebih spesifik, 4,4% dari populasi dunia mengalami depresi, sedangkan 3,6% mengalami gangguan kecemasan, menunjukkan bahwa masalah kesehatan mental semakin umum dan perlu mendapatkan penanganan serius (Kartika et al., 2025). Untuk Dalam proses seleksi fitur yang berhubungan dengan faktor penyebab gangguan kesehatan mental, Salah satu pendekatan yang bisa digunakan adalah metode *Relief-F*. Teknik ini memungkinkan penilaian serta pemilihan fitur atau variabel yang paling relevan terhadap target. Dalam konteks ini, metode *Relief-F* yang diterapkan dalam *Naïve Bayes* dapat membantu memilih fitur paling signifikan berdasarkan faktor penyebab gangguan kesehatan mental. Metode *Naïve Bayes* sendiri merupakan teknik klasifikasi data yang menggunakan pendekatan probabilitas dan statistik untuk memprediksi kemungkinan masa depan berdasarkan pola yang telah ditemukan sebelumnya, sehingga dikenal dengan Teorema Bayes (Fatah et al., 2025).

Penelitian ini bertujuan untuk melakukan Optimasi terhadap akurasi prediksi risiko gangguan kesehatan mental dengan menerapkan metode *hybrid* yang menggabungkan algoritma *Naïve Bayes* dan teknik seleksi fitur *Relief-F*, sehingga mampu mengidentifikasi variabel-variabel paling relevan dan meningkatkan performa model klasifikasi. Diharapkan, analisis ini dapat berkontribusi dalam mendukung deteksi dini dan penanganan masalah kesehatan mental yang lebih tepat.

2. METODE PENELITIAN

2.1 Tahap Penelitian

Penelitian ini dilakukan dengan tahapan-tahapan yang dilakukan agar dapat berjalan dengan baik dan terstruktur. Berikut ini tahap penelitian yang dilakukan oleh penulis, yang dapat dilihat pada Gambar 1 :



Gambar 1. Tahap Penelitian

Penelitian ini memakai pendekatan kuantitatif dengan metode eksperimen untuk membangun dan menguji sebuah model prediksi risiko gangguan kesehatan mental. Tahapan penelitian ini bisa dilihat pada Gambar 1. Tahapan ini bertujuan untuk melakukan prediksi tingkat risiko gangguan kesehatan mental menggunakan metode *hybrid ReliefF* dan algoritma *Naïve Bayes*. Proses dimulai dengan pengambilan dataset melalui website Kaggle yang telah dipisahkan menjadi data *train*, *test*, dan *sample_submission*. Selanjutnya, dilakukan tahap *preprocessing* yang mencakup *data cleaning* dan normalisasi data menggunakan metode *Min-Max Scaling* untuk menyeragamkan skala antar fitur. Setelahnya dilakukan penentuan label klasifikasi Target. Proses dilanjutkan dengan seleksi fitur menggunakan algoritma *ReliefF* untuk memilih fitur-fitur yang paling berpengaruh dalam mengelompokkan risiko gangguan kesehatan mental. Fitur-fitur terpilih kemudian digunakan sebagai input untuk proses pengelompokan menggunakan algoritma *Naïve Bayes*. Selanjutnya, dilakukan evaluasi model dengan mengukur akurasi klasifikasi untuk menilai performa algoritma terhadap data yang telah diproses. Interpretasi hasil yang diperoleh, menjadi dasar dalam merumuskan kesimpulan serta rekomendasi.

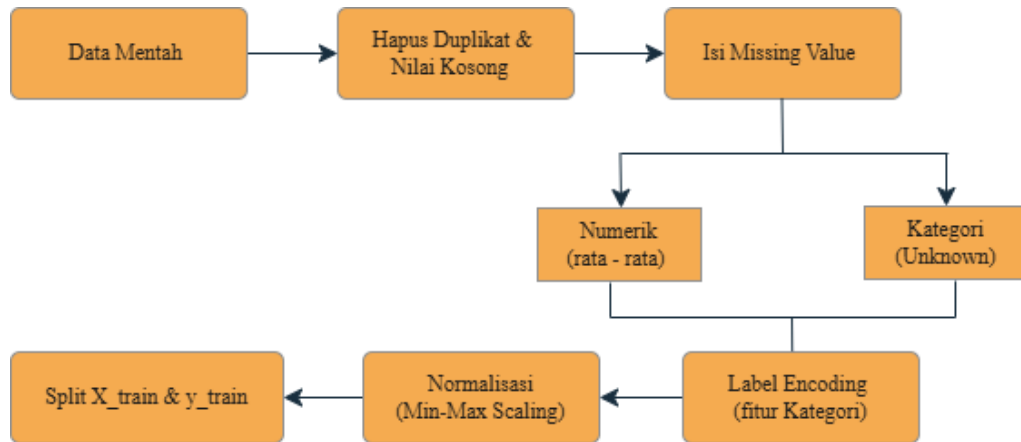
2.2 Sumber Data

Data yang digunakan bersumber dari dataset publik yang tersedia di Kaggle dengan judul *Mental Health Data Competition*. Dataset ini berisi 14.070 data yang memuat informasi mengenai usia, jenis kelamin, status pernikahan, status pekerjaan, tingkat pendidikan terakhir, frekuensi penggunaan media sosial, kualitas tidur, jumlah tanggung jawab pribadi, frekuensi interaksi sosial, kemampuan manajemen waktu, kondisi ekonomi, dukungan keluarga, dan pengalaman diskriminasi, depresi menjadi variabel targetnya. Populasi dalam penelitian ini mencakup individu dari beragam latar belakang sosial dan ekonomi yang berpartisipasi dalam survei kesehatan mental. Teknik sampling yang digunakan bersifat non-probabilistik karena data merupakan data sekunder yang telah tersedia secara daring. Pengumpulan data dilakukan dengan mengunduh file dalam format csv, yang kemudian dianalisis menggunakan perangkat lunak *Jupyter* dan *Python*.

2.3 Pra-Pemrosesan Data

Gambar 2 merupakan alur Tahapan *pra-pemrosesan data*. Tahapan ini dilakukan untuk memastikan kualitas data sebelum dilakukan tahap pemodelan (Rais et al., 2025). Dalam penelitian ini mencakup pembersihan data dengan menghapus nilai kosong dan duplikat, serta menangani data kategorikal

menggunakan teknik *encoding*. Fitur kategorikal diubah menjadi format numerik menggunakan teknik *Label Encoding*. Selanjutnya data numerik dinormalisasi menggunakan metode *Min-Max* untuk menjaga konsistensi skala antar fitur.



Gambar 2. *Processing Data*

2.4 Seleksi Fitur *Relief-F*

Relief-F adalah pengembangan dari algoritma *Relief* yang pertama kali diperkenalkan oleh Kira dan Rendell pada tahun 1992 (Musthafa Rajabi et al., 2023). Metode *Relief-F* memiliki tingkat efisiensi yang tinggi dan mampu bekerja dengan berbagai karakteristik tipe data tanpa batasan tertentu (Musthafa Rajabi et al., 2023). Algoritma *Relief-F* diimplementasikan pada data pelatihan untuk mengevaluasi relevansi setiap fitur terhadap variabel target.

2.5 *Naïve Bayes (Modeling)*

Proses klasifikasi menggunakan *Naïve Bayes* dilakukan dengan memanfaatkan data latih yang telah dibagi sebelumnya menggunakan metode *k-fold cross validation*. Selama tahap pelatihan dan pengujian data, setiap karakter dievaluasi satu per satu berdasarkan fitur yang tersedia dalam dataset.

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

$P(C|X)$ = Probabilitas kelas C terhadap data X

$P(X|C)$ = Probabilitas data X terhadap kelas C

$P(C)$ = Probabilitas awal kelas C

$P(X)$ = Probabilitas awal data

2.6 Evaluasi Model

Kedua model dievaluasi menggunakan metrik klasifikasi standar, yaitu Akurasi (*Accuracy*), Presisi (*Precision*), Recall (*Recall/Sensitivity*), *F1-Score*, dan *Area Under the Receiver Operating Characteristic Curve (AUC-ROC)*. Matriks Konfusi (*Confusion Matrix*) juga digunakan untuk analisis detail kinerja klasifikasi.

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

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Precision = \frac{TP+TN}{FP+FN} \quad (4)$$

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

2.7 Python

Penelitian ini mengadopsi *Python* sebagai bahasa pemrograman utama dalam pengembangan perangkat lunak. Keunggulan *Python* terletak pada fleksibilitasnya serta ekosistem pustaka yang luas, yang memungkinkan pengembangan berbagai modul penting, termasuk algoritma pengolahan data dan antarmuka pengguna grafis. Pemilihan *Python* didasarkan pada kemampuannya dalam mempercepat proses prototipe serta integrasi yang optimal dengan berbagai teknologi pendukung komputasi ilmiah dan analisis data, aspek yang sangat vital dalam mencapai tujuan penelitian ini.

3. HASIL DAN PEMBAHASAN

3.1 Sumber Data

Dataset yang digunakan merupakan dataset publik bernama *Mental Health Data Competition* yang diambil dari website resmi *Kaggle*. *Dataset* berisi informasi Variabel seperti usia, jenis kelamin, tekanan akademik, pola tidur, serta status kesehatan mental individu. Total data yang digunakan untuk pelatihan berjumlah 1.016 baris (*data train*) dan 1.000+ baris (*data test*). Untuk perhitungan *dataset* Kesehatan mental tidak dilakukan penghitungan secara manual dikarenakan jumlah data yang banyak sehingga perhitungan dilakukan dengan menggunakan Bahasa pemrograman *python 3*, Gambar. 3 dan 4 menampilkan 5 teratas isi dari *data train* dan *data test*.

id	Gender	Age	Academic Pressure	Sleep Duration	Depression	
0	0	Female	49.0	NaN	More than 8 hours	0
1	1	Male	26.0	NaN	Less than 5 hours	1
2	2	Male	33.0	5.0	5-6 hours	1
3	3	Male	22.0	NaN	Less than 5 hours	1
4	4	Female	30.0	NaN	5-6 hours	0

Gambar 3. Data Train

id	Gender	Age	Academic Pressure	Sleep Duration	
0	140700	Male	53.0	NaN	Less than 5 hours
1	140701	Female	58.0	NaN	Less than 5 hours
2	140702	Male	53.0	NaN	7-8 hours
3	140703	Female	23.0	5.0	More than 8 hours
4	140704	Male	47.0	NaN	7-8 hours

Gambar 4. Data Test

Untuk melakukan perhitungan dengan menggunakan *python 3*, Langkah pertama yang harus dilakukan adalah meng-*import library* yang dibutuhkan. *Library* ini berisi fungsi yang membantu dalam proses perhitungan dan analisis data. Daftar *library* yang digunakan dalam penelitian ini dapat dilihat pada Gambar 5 berikut :

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder, MinMaxScaler
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from skrebate import ReliefF
from IPython.display import display
import matplotlib.pyplot as plt
import seaborn as sns
```

Gambar 5. Library

3.2 Pra-Pemrosesan Data

Tahap pemrosesan data dilakukan untuk meningkatkan kualitas data sebelum dianalisis atau digunakan dalam model pembelajaran mesin. Pra-pemrosesan data dilakukan dengan menghapus nilai kosong dan duplikat dari data mentah. Kemudian, nilai yang hilang digantikan dengan rata-rata atau "Unknown" untuk menjaga konsistensi. Data kategorikal diubah menjadi bentuk numerik menggunakan Label Encoding. Setelah itu, semua fitur numerik dinormalisasi dengan *Min-Max Scaling* agar rentang nilainya seragam.

id	Name	Gender	Age	City	Working Professional or Student	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Family History of Mental Illness	Depression
0	Aaradhya	Female	49.0	Ludhiana	Working Professional	Chef	NaN	5.0	NaN	NaN	2.0	More than 8 hours	Healthy	BHM	No	1.0	2.0	No	0
1	Vivan	Male	26.0	Varanasi	Working Professional	Teacher	NaN	4.0	NaN	NaN	3.0	Less than 5 hours	Unhealthy	LLB	Yes	7.0	3.0	No	1
2	Yuvraj	Male	33.0	Visakhapatnam	Student	NaN	5.0	NaN	8.97	2.0	NaN	5-6 hours	Healthy	B.Pharm	Yes	3.0	1.0	No	1
3	Yuvraj	Male	22.0	Mumbai	Working Professional	Teacher	NaN	5.0	NaN	NaN	1.0	Less than 5 hours	Moderate	BBA	Yes	10.0	1.0	Yes	1
4	Rhea	Female	30.0	Kanpur	Working Professional	Business Analyst	NaN	1.0	NaN	NaN	1.0	5-6 hours	Unhealthy	BBA	Yes	9.0	4.0	Yes	0

Gambar 6. Data Train Awal

Gambar 6 merupakan isi kondisi awal data dengan berbagai jenis data (numerik dan kategorikal) serta keberadaan nilai-nilai yang hilang (*missing values*) seperti "NaN" dari dataset pelatihan (data *train*) yang menampilkan lima baris teratas dari dataset tersebut, memberikan gambaran tentang format dan jenis data sebelum transformasi.

id	Name	Gender	Age	City	Working Professional or Student	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Family History of Mental Illness	Depression
0	11	0	49.0	50	1	10	3.142273	5.000000	7.658636	2.94494	2.000000	29	7	33	0	1.0	2.0	0	0
1	407	1	26.0	93	1	55	3.142273	4.000000	7.658636	2.94494	3.000000	27	20	63	1	7.0	3.0	0	1
2	417	1	33.0	97	0	59	5.0000000	2.998998	8.970000	2.000000	2.974404	15	7	21	1	3.0	1.0	0	1
3	417	1	22.0	64	1	55	3.142273	5.0000000	7.658636	2.94494	1.000000	27	15	28	1	10.0	1.0	1	1
4	286	0	30.0	37	1	9	3.142273	1.0000000	7.658636	2.94494	1.000000	15	20	28	1	9.0	4.0	1	0

Gambar 7. Data Setelah Transformasi

Dari hasil Gambar 7, data telah melewati tahapan *preprocessing* awal, Tahap ini meliputi penghapusan duplikat, pengisian nilai kosong (dengan rata-rata untuk numerik dan "Unknown" untuk kategorikal), serta *encoding* fitur kategorikal menjadi numerik. Data yang telah diubah dilakukan pemisahan fitur dan target seperti pada Gambar 8 :

```

X_train = train.drop(columns=['Depression'], errors='ignore')
y_train = train['Depression']

print(" X_train shape:", X_train.shape)
print(" y_train shape:", y_train.shape)

# Tampilkan 5 baris pertama dari X dan y
print("\n ♦ X_train:")
display(X_train.head())

print("\n ♦ y_train:")
display(y_train.head())

```

Gambar 8. Pemisahan fitur dan Target

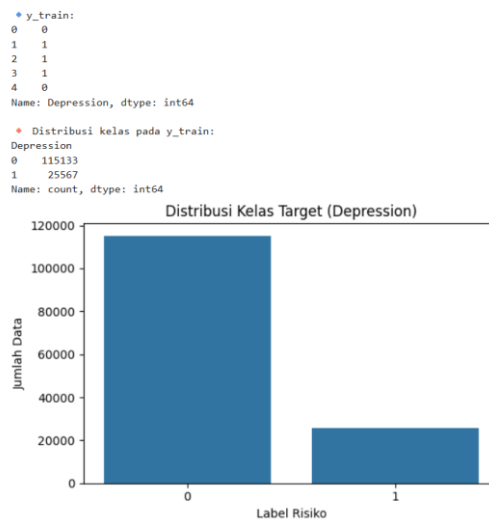
Setelah *preprocessing*, fitur dan target dipisahkan menjadi variabel x_{train} dan y_{train} , dengan hasil *shape*:

x_{train} shape : (140700, 19)
 y_{train} shape : (140700,)

id	Name	Gender	Age	City	Working Professional or Student	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Family History of Mental Illness
0	11	0	49.0	50	1	10	3.142273	5.000000	7.658636	2.94494	2.000000	29	7	33	0	1.0	2.0	0
1	407	1	26.0	93	1	55	3.142273	4.000000	7.658636	2.94494	3.000000	27	20	63	1	7.0	3.0	0
2	417	1	33.0	97	0	59	5.000000	2.998998	8.970000	2.000000	2.974404	15	7	21	1	3.0	1.0	0
3	417	1	22.0	64	1	55	3.142273	5.000000	7.658636	2.94494	1.000000	27	15	28	1	10.0	1.0	1
4	286	0	30.0	37	1	9	3.142273	1.000000	7.658636	2.94494	1.000000	15	20	28	1	9.0	4.0	1

Gambar 9. Data x train

Gambar 9 menampilkan isi dari data *train* yang saat ini hanya berisi fitur, tanpa adanya target di dalamnya.



Gambar 10. Distribusi Kelas Target

Gambar 10 merupakan distribusi kelas target yang dalam penelitian ini diberi label *Depression* (Depresi). Grafik batang menunjukkan jumlah data untuk dua kategori resiko 0 dan 1.

1. Kategori 0 : mempresentasikan kelas mayoritas dengan jumlah data sebanyak 115.133. Dalam konteks klasifikasi depresi, kategori ini kemungkinan besar mengindikasikan individu yang tidak mengalami depresi atau berisiko rendah.
2. Kategori 1 : Merepresentasikan kelas minoritas dengan jumlah data sebanyak 25.567. Kategori ini kemungkinan besar mengindikasikan individu yang mengalami depresi atau berisiko tinggi.

3.3 Seleksi Fitur dengan *Relief-F*

Dikarenakan algoritma *Relief-F* membutuhkan sumber daya komputasi yang besar, pada Gambar 8 seleksi fitur dilakukan dengan mengambil sampel 200 data dari dataset. Hasil seleksi fitur dengan *Relief-F* ini menvalidasi pemilihan variabel yang digunakan dalam model. Fitur-fitur yang teridentifikasi sebagai yang

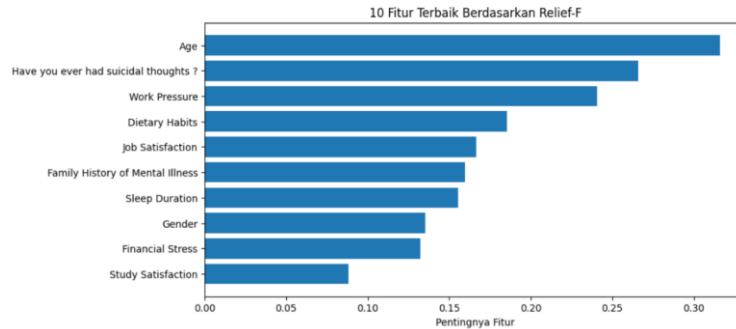
paling penting digunakan sebagai input untuk algoritma klasifikasi *Naïve Bayes*, seleksi ini dilakukan agar model berfokus pada atribut-atribut yang paling berpengaruh dalam memprediksi risiko gangguan kesehatan mental.

```
# Ambil 200 baris data untuk seleksi fitur
subset_X = X_train_scaled[:200]
subset_y = y_train[:200]

relief = ReliefF(n_features_to_select=10, n_neighbors=20)
relief.fit(subset_X, subset_y)

# Ambil fitur terbaik
selected_indices = relief.top_features_[:10]
selected_features = [X_train.columns[i] for i in selected_indices]
```

Gambar 11. Implementasi seleksi fitur *Relief-F*



Gambar 12. Seleksi fitur *Relief-F*

Pada Gambar 12 grafik batang horizontal menunjukkan 10 fitur teratas yang memiliki tingkat relevansi paling tinggi terhadap variabel target "*Depression*". Nilai pada sumbu x, yang diberi label "Pentingnya Fitur", merepresentasikan bobot atau skor relevansi yang diberikan oleh algoritma *Relief-F* untuk setiap fitur. Semakin panjang batang, menunjukkan semakin tinggi nilai bobot fitur tersebut, yang berarti fitur tersebut semakin penting dalam memprediksi risiko gangguan kesehatan mental.

3.4 Modeling dengan *Naïve Bayes*

Model dibangun menggunakan algoritma *Naïve Bayes Gaussian*, yang dilatih pada data *training* dengan 10 fitur hasil seleksi seperti pada gambar 14. Model ini menggunakan prinsip probabilistik dan asumsi independensi antar fitur untuk menentukan kelas risiko (*Depresi*) secara otomatis.

```
# ===== TRAINING NAIVE BAYES =====
model = GaussianNB()
model.fit(X_train_selected, y_train)
y_pred_train = model.predict(X_train_selected)
```

Gambar 13. *Training* dengan Model *Naïve Bayes*

4.5 Evaluasi Model

Berikut ini hasil evaluasi performa model klasifikasi *Naïve Bayes* terhadap data pelatihan pada Gambar 14 :

```
Evaluasi Model Naive Bayes
Akurasi: 0.903582089522388

Classification Report:
              precision    recall  f1-score   support

     0       0.95         0.93         0.94       115133
     1       0.71         0.78         0.75       25567

 accuracy          0.90       140700
 macro avg         0.83         0.86         0.84       140700
 weighted avg         0.91         0.90         0.91       140700
```

Gambar 14. Hasil Evaluasi Model

Gambar 14 menyajikan hasil evaluasi performa model klasifikasi *Naïve Bayes* terhadap data pelatihan. dalam bentuk *classification report*, yang menyediakan metrik evaluasi penting seperti *precision*, *recall*, *f1-*

score, dan *support* untuk setiap kelas (0 dan 1), serta akurasi keseluruhan, rata-rata makro (*macro avg*), dan rata-rata berbobot (*weighted avg*) dan hasilnya disajikan dan diringkas dalam Tabel 1.

Matrik Evaluasi	Nilai
Akurasi	90.35%
Precision	91%
Recall	90%
F1-Score	91%

Hasil evaluasi menunjukkan bahwa model mencapai akurasi keseluruhan 90.35%, artinya sebagian besar prediksinya benar. Untuk Kelas 0 (Tidak Depresi), model sangat baik dalam mengidentifikasi individu tanpa gejala depresi, dengan presisi 0.95, *recall* 0.93, dan *f1-score* 0.94. Sementara itu, untuk Kelas 1 (Depresi) yang merupakan kelas minoritas, model memperoleh presisi 0.71, *recall* 0.78, dan *f1-score* 0.75. Meskipun sedikit lebih rendah dari Kelas 0, nilai *f1-score* 0.75 tetap menunjukkan bahwa model cukup efektif dalam mengenali kasus depresi. Selain itu, *recall* yang lebih tinggi dibandingkan *precision* pada Kelas 1 menunjukkan bahwa model cenderung mampu mengidentifikasi sebagian besar individu yang benar-benar mengalami depresi.

4. KESIMPULAN

Penelitian ini membuktikan bahwa kombinasi algoritma *Naïve Bayes* dengan metode seleksi fitur *Relief-F* dapat meningkatkan akurasi dalam memprediksi risiko gangguan kesehatan mental. Model yang dikembangkan mencapai akurasi sebesar 90,35% dan menunjukkan kinerja yang baik dalam mengklasifikasikan individu yang berisiko mengalami depresi. Pendekatan ini efektif dalam mendukung deteksi dini dan memiliki potensi untuk dikembangkan lebih lanjut guna diterapkan dalam sistem skrining kesehatan mental di dunia nyata.

ACKNOWLEDGEMENTS

Terima kasih disampaikan kepada Allah SWT atas rahmatnya, dosen pembimbing, serta semua pihak yang telah membantu dalam penelitian ini.

REFERENCES

- A. Rahim, A. M., Ingrid Yanuar Risca Pratiwi, & Muhammad Ainul Fikri. (2023). Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Classifier. *Indonesian Journal of Computer Science*, 12(5), 2995–3011. <https://doi.org/10.33022/ijcs.v12i5.3413>
- Bali, M. M. E. I., & Hajriyah, H. B. (2020). Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) Sekolah Tinggi Islam Blambangan (STIB) Banyuwangi MOMENTUM Modernisasi Pendidikan Agama Islam Di Era Revolusi Industri 4.0. *Jurnal Sosial Dan Agama*, 09(1), 42–62. <https://doi.org/10.14421/jpai.2019.161-01>.
- Fatah, Z., Sains, F., Ibrahimy, U., Timur, S. J., Sains, F., Ibrahimy, U., & Timur, S. J. (2025). Klasifikasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa. *Jurnal Ilmiah Multidisiplin Ilmu*, 2(1), 29–37.
- Ferdian, D., Hikmat, R., Zuqriefa, A. B., Ma'ruf, T. L. H., Noviana, M., Harahap, S. M. I., Sutanto, H., & Hutapea, M. R. (2024). Pengaruh Edukasi Kesehatan untuk Meningkatkan Pengetahuan tentang Kesehatan Mental. *MAHESA : Malahayati Health Student Journal*, 4(5), 2058–2067. <https://doi.org/10.33024/mahesa.v4i5.14585>
- Gebremedhin, H. T., Biffu, B. B., Lebessa, M. T., Weldeyes, A. Z., Gebru, T. T., & Petrucka, P. (2020). Prevalence and associated factors of psychological distress among secondary school students in Mekelle city, Tigray region, Ethiopia: Cross-sectional study. *Psychology Research and Behavior Management*, 13, 473–480. <https://doi.org/10.2147/PRBM.S252779>
- Halik, A., Aini, Z., Siagian, K., Jabal, U., Aceh, G., Sains, U., & Nyak, C. (2024). Program Counseling Prevention Kesehatan Mental Berbasis Sekolah: Kajian Literature Review. *Jurnal Al-Taujih*, 10(2), 154–169.
- Kartika, D. K., Shofiah, V., & Rajab, K. (2025). Psikoterapi Islam untuk Meningkatkan Kesehatan Mental. *Jurnal Ilmu Pendidikan Dan Psikologi*, 2, 78–85.
- Khairan, C. A., & Habib, M. S. (2024). Chatbot Ai Dalam Identifikasi Awal Gangguan Kesehatan Mental Di Indonesia : *Jurnal Empati*, 13, 498–508.
- Lisma Nanda, N. (2025). Peran Keluarga dalam Meningkatkan Kesadaran Tentang Kesehatan Mental. *Journal Innovation in Education*, 3(1), 54–56.
- Mayasari, A. C., Kertapati, Y., & Widayanti, D. M. (2024). Penyuluhan “Mengelola Kesehatan Mental Dan Kenakalan Remaja” Di Sma Muhammadiyah 3 Surabaya. *Jurnal Pengabdian Kesehatan*, 7(2), 477. <http://jpk.jurnal.stikeskendekiautamakudus.ac.id>

- Musthafa Rajabi, K., Witanti, W., & Yuniarti, R. (2023). Penerapan Algoritma K-Nearest Neighbor (KNN) Dengan Fitur Relief-F Dalam Penentuan Status Stunting. *INNOVATIVE: Journal Of Social Science Research*, 3(4), 3555–3568. <https://j-innovative.org/index.php/Innovative/article/view/3885/2801>
- Purwanti, L. E., Ernawati, H., Sukamto, F. I., & Verawati, M. (2025). Edukasi dan skrining sebagai upaya peningkatan kesehatan mental remaja di panti asuhan. *JMM (Jurnal Masyarakat Mandiri)*, 9(1), 6–12.
- Rahmawati, A., Agustini, A., Rizki Ramadhan, M., Malihatul Zulfaa, N., & Nisrina, H. (2025). Gangguan Kesehatan Mental Terhadap Performa Akademik Mahasiswa Tingkat Awal Astri. *Journal of Education and Learning*, 3(1), 1–15.
- Rais, A. N., Warjiyono, & Putra, J. L. (2025). Optimasi Prediksi Risiko Kredit Dengan Preprocessing Dan Hyperparameter Tuning. *JIKA (Jurnal Informatika)*, 9(1), 59–65. <https://doi.org/http://dx.doi.org/10.31000/jika.v9i1.12782>
- Ramdani, R. F., Herlambang, A. D., Falhadi, M. M., Fadilah, M. Z., Evan, C., Turnip, L., & Mulyana, A. (2024). Membangun Kesejahteraan Pikiran Untuk Kesehatan Mental Melalui Gaya Hidup Sehat Dan Olahraga. *Indo-MathEdu Intellectuals Journal*, 5(3), 2928–2936.
- Reni Handayani, A., Arifatus Sholihah, N., Lestari, A., Gustia Kesuma, E., & Rafi'ah. (2021). Edukasi Hukum dan Etika Terhadap Kesehatan Mental Responsif Gender Bagi Kader Kesehatan di Labuan Badas Sumbawa. *Jurnal Abdimas Indonesia*, 1(2), 26–32. <https://dmi-journals.org/jai/article/view/226>
- Sari, M. K., & Susmiatin, E. A. (2023). Deteksi Dini Kesehatan Mental Emosional pada Mahasiswa. *Jurnal Ilmiah STIKES Yarsi Mataram*, 13(1), 10–17. <https://doi.org/10.57267/jisym.v13i1.226>
- Wenny. (2024). Normalisasi Data Kependudukan Dengan Model Min Max Dan Algoritma K-Means Untuk Pengelompokan Tingkat Ekonomi Masyarakat. *Bulletin of Information System Research (BIOS)*, 2(2), 53–63. <https://doi.org/https://doi.org/10.62866/bios.v2i2.141>
- Wijaya, V., & Rachmat, N. (2025). Penerapan SMOTE dan Regresi Logistik Pada Website Skrining Awal Kesehatan Mental Mahasiswa 1,2. *Jurnal Algoritme*, 5(2), 121–130. <https://doi.org/10.35957/algoritme>.