

# **Final Project Pembelajaran Mesin**

Determinasi dan Penerapan Model Machine Learning untuk Deteksi  
Dini Tingkat Kecemasan Berbasis Gaya Hidup

## **Kelas E Kelompok 8**

5025231030 – I Gusti Ngurah Arya Sudewa

5025231087 – Justin Chow

5025231113 – Reynandriel Pramas Thandya

Dosen Pengampu: Prof. Dr. Eng. Chastine Fatichah, S.Kom., M.Kom.

Institut Teknologi Sepuluh Nopember

Semester Genap 2024/2025

## **Ringkasan**

Gangguan kecemasan meningkat pesat seiring tekanan akademik, tuntutan pekerjaan, dan eksposur media sosial, namun penerapan machine learning untuk deteksi dini berdasarkan gaya hidup masih terbatas. Penelitian ini membandingkan sepuluh algoritma Decision Tree, Random Forest, KNN, Logistic Regression, Gaussian Naive Bayes, SVM, ANN, CatBoost, XGBoost, dan LightGBM dengan tiga teknik penyeimbangan data (tanpa balancing, SMOTE, dan Random Oversampling) menggunakan dataset `enhanced_anxiety_dataset.csv`. Hasil menunjukkan bahwa model ensemble pohon, khususnya Random Forest dipadukan ROS, mencapai akurasi 0,88 dan Macro F1 0,87, unggul atas kombinasi lain. Temuan ini menekankan pentingnya penyeimbangan data untuk meningkatkan ketepatan prediksi kecemasan berbasis gaya hidup.

## Daftar Isi

<b>1</b>	<b>Pendahuluan</b>	<b>3</b>
1.1	Latar Belakang . . . . .	3
1.2	Rumusan Masalah . . . . .	3
1.3	Tujuan Penelitian . . . . .	4
1.4	Batasan Masalah . . . . .	4
<b>2</b>	<b>Tinjauan Pustaka</b>	<b>5</b>
<b>3</b>	<b>Metode dan Model</b>	<b>6</b>
<b>4</b>	<b>Metode Evaluasi</b>	<b>7</b>
4.1	Deskripsi dan Analisis Dataset . . . . .	7
4.1.1	Sumber dan Karakteristik Data . . . . .	7
4.1.2	Deskripsi Variabel . . . . .	7
4.1.3	Transformasi Variabel Target . . . . .	8
4.1.4	Exploratory Data Analysis . . . . .	8
4.2	Desain Eksperimen . . . . .	9
4.3	Strategi Penyeimbangan Kelas . . . . .	9
4.4	Metrik Evaluasi . . . . .	9
4.5	Prosedur Eksperimen . . . . .	10
<b>5</b>	<b>Hasil Evaluasi Model</b>	<b>10</b>
5.1	Analisis Kuantitatif . . . . .	21
5.2	Pembahasan Hasil . . . . .	22
<b>6</b>	<b>Kesimpulan</b>	<b>23</b>

## Daftar Gambar

1	Confusion matrix untuk Decision Tree. . . . .	11
2	Confusion matrix untuk Random Forest. . . . .	12
3	Confusion matrix untuk KNN. . . . .	13
4	Confusion matrix untuk Logistic Regression. . . . .	14
5	Confusion matrix untuk Gaussian Naive Bayes. . . . .	15
6	Confusion matrix untuk Support Vector Machine. . . . .	16
7	Confusion matrix untuk Artificial Neural Network. . . . .	17
8	Confusion matrix untuk CatBoost. . . . .	18
9	Confusion matrix untuk XGBoost. . . . .	19
10	Confusion matrix untuk LightGBM. . . . .	20
11	Perbandingan akurasi model terhadap teknik balancing . . . . .	21
12	Perbandingan Macro F1-Score model terhadap teknik balancing . . . . .	22

## Daftar Tabel

1	Akurasi dan Macro F1-Score . . . . .	21
---	--------------------------------------	----

# 1 Pendahuluan

## 1.1 Latar Belakang

Prevalensi gangguan kesehatan mental, khususnya kecemasan, menunjukkan peningkatan yang signifikan di berbagai negara, termasuk Indonesia [7, 8]. Data WHO dan survei lokal mengindikasikan bahwa tekanan akademik, tuntutan pekerjaan, serta eksposur media sosial berperan besar dalam lonjakan kasus ini [9]. Di Indonesia, studi oleh Khoirunissa & Sukartini (2020) melaporkan bahwa sebagian besar remaja dan dewasa muda masih mengalami berbagai hambatan akses layanan kesehatan mental [8].

Faktor gaya hidup seperti kualitas tidur, tingkat aktivitas fisik, pola konsumsi kafein, dan asupan nutrisi telah terbukti berkorelasi erat dengan tingkat kecemasan dan depresi. Hautekiet et al. (2022) menemukan bahwa pola hidup sehat meningkatkan kesejahteraan mental dan menurunkan marker stres pada populasi usia menua [1]. Grajek et al. (2022) menegaskan bahwa konsumsi makanan kaya antioksidan dan lemak tak jenuh dapat memperbaiki mood dan menurunkan gejala depresi [2]. Basiri et al. (2023) bahkan menyoroti interaksi kompleks antara diabetes, nutrisi, kecemasan, dan depresi, yang menunjukkan pentingnya pendekatan holistik dalam intervensi kesehatan mental [3].

Penelitian lokal juga mendukung temuan internasional tersebut. Martia & Salman (2022) melaporkan bahwa remaja dengan pola tidur teratur dan aktivitas fisik terjadwal memiliki tingkat kecemasan lebih rendah dibandingkan kelompok kontrol [6]. Wilanda et al. (2024) menunjukkan bahwa pekerja kantoran di Jawa Barat yang mengadopsi gaya hidup sehat termasuk diet seimbang dan olahraga ringan mengalami peningkatan skor kesejahteraan mental [5]. Lebih lanjut, Farika et al. (2024) menekankan pentingnya program promosi kesehatan mental berbasis komunitas untuk remaja di kawasan perkotaan [4].

Meskipun demikian, adopsi teknik machine learning untuk memprediksi tingkat kecemasan berdasarkan variabel gaya hidup masih sangat terbatas, khususnya di konteks Indonesia. Berangkat dari celah inilah, penelitian ini bermaksud membandingkan sepuluh algoritma klasifikasi Decision Tree, Random Forest, KNN, Logistic Regression, Gaussian Naive Bayes, SVM, ANN, CatBoost, XGBoost, dan LightGBM dengan tiga teknik balancing (tanpa balancing, SMOTE, dan ROS) untuk membangun model prediksi kecemasan yang cepat, akurat, dan kontekstual [10].

## 1.2 Rumusan Masalah

Penelitian ini dirancang untuk menjawab serangkaian pertanyaan spesifik mengenai penerapan ML untuk deteksi kecemasan:

1. Di antara sepuluh model *machine learning* yang diuji, manakah yang menunjukkan performa terbaik dalam memprediksi kategori tingkat kecemasan berdasarkan data gaya hidup?

2. Bagaimana dampak signifikan dari penerapan teknik penanganan data tidak seimbang (*RandomOverSampler*, *SMOTE*, dan *Tanpa balancing*) terhadap metrik evaluasi seperti Accuracy dan F1-Score?
3. Kombinasi spesifik antara model algoritma dan teknik penyeimbangan data manakah yang menghasilkan solusi paling optimal dan dapat diandalkan untuk deteksi dini kecemasan?

### 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menerapkan dan mengevaluasi secara komprehensif sepuluh algoritma ML untuk tugas klasifikasi multi-kelas tingkat kecemasan.
2. Menganalisis secara kuantitatif dampak dari tiga metode penyeimbangan data yang berbeda terhadap performa setiap model.
3. Menentukan kombinasi model dan metode terbaik sebagai rekomendasi untuk pengembangan prototipe sistem deteksi kecemasan dini.

### 1.4 Batasan Masalah

Ruang lingkup penelitian ini dibatasi pada:

1. Sumber data tunggal yaitu dataset "enhanced\_anxiety\_dataset.csv" dari Kaggle.
2. Fitur yang dianalisis adalah yang tersedia dalam dataset tersebut tanpa rekayasa fitur ekstensif.
3. Evaluasi model didasarkan murni pada metrik performa kuantitatif. Interpretasi klinis dari hasil model berada di luar cakupan penelitian ini.

## 2 Tinjauan Pustaka

Berbagai studi terkini menunjukkan bahwa gaya hidup termasuk pola tidur, aktivitas fisik, konsumsi kafein, alkohol, dan aspek sosial berperan penting dalam mempengaruhi kesehatan mental, khususnya kecemasan. Namun hanya sedikit yang mengombinasikannya dengan pendekatan *machine learning* (ML) sebagai sistem deteksi dini. Kami merangkum beberapa penelitian utama di bawah ini.

Penelitian longitudinal menunjukkan bahwa pola hidup sehat terutama kombinasi antara aktivitas fisik dan diet berkualitas secara signifikan berkorelasi dengan kesejahteraan mental, termasuk mengurangi gejala kecemasan pada lansia dan pekerja kantoran di Indonesia [1, 5]. Grajek *et al.* (2022) menyatakan bahwa nutrisi secara kritis mempengaruhi mood dan kecemasan [2]. Sementara Basiri *et al.* (2023) menambahkan bahwa kondisi komorbiditas seperti diabetes dan pola makan buruk dapat memperburuk kecemasan dan depresi [3]. Selain itu, penelitian Martia & Salm-an (2022) dan Farika *et al.* (2024) memperjelas bahwa kualitas tidur dan intervensi berbasis komunitas sangat berperan dalam menurunkan kecemasan remaja [6, 4].

Dalam ranah ML, sebuah tinjauan sistematis [11] menyimpulkan bahwa algoritma Support Vector Machine (SVM) dan Logistic Regression mampu memprediksi stres dan kecemasan pada mahasiswa dengan metrik presisi dan akurasi yang tinggi. Selanjutnya, analisis menunjukkan bahwa model Random Forest mencapai akurasi hingga 91% dalam memprediksi kecemasan pada remaja, dengan variabel latar belakang keluarga dan konsumsi alkohol sebagai kontributor utama [12]. Di domain wearable, Abd-Alrazaq *et al.* (2023) menemukan bahwa penggunaan perangkat seperti Actiwatch yang dikombinasikan dengan algoritma Random Forest dan SVM dapat mendeteksi kecemasan berdasarkan data aktivitas fisik, pola tidur, dan detak jantung [13]. Sementara itu, Zhang *et al.* (2025) mengembangkan kerangka anomaly detection berbasis LSTM autoencoder untuk memantau lonjakan gejala kecemasan dan depresi dari data wearable, dengan adjusted F1-score mencapai 0,80 [14].

Kajian khusus terhadap remaja oleh Frontiers (2024) mengidentifikasi variabel seperti kesepian, pola tidur akhir pekan, merokok, dan konsumsi makanan cepat saji sebagai prediktor Generalized Anxiety Disorder (GAD). XGBoost menunjukkan sensitivitas tinggi dalam feature selection, serta memanfaatkan SMOTE untuk meningkatkan performa dalam mengatasi ketidakseimbangan kelas [15]. Selain itu, Mardini *et al.* (2025) memvalidasi model ML berbasis data klinis dan determinan sosial di kalangan remaja, menggunakan XGBoost dan SHAP untuk interpretabilitas, mendemonstrasikan bahwa determinan sosial secara signifikan mempengaruhi prediksi kecemasan [16].

Ringkasan kontribusi terkait:

- Kombinasi gaya hidup dan data sensor wearable memungkinkan deteksi dini gejala kecemasan/depresi [13, 14].
- Algoritma tree-based seperti Random Forest dan XGBoost unggul dalam mendeteksi variabel prediktor utama dan menyelesaikan masalah kelas tidak seimbang dengan SMOTE [12, 15].

- Interpretabilitas model (misalnya SHAP atau LIME) membantu memetakan kontribusi masing-masing faktor gaya hidup dan sosial terhadap risiko kecemasan [16].

Berangkat dari temuan-temuan tersebut, penelitian ini menggunakan dataset gaya hidup dan kecemasan dari Kaggle (2019–2024) [10], serta menerapkan sepuluh algoritma ML dan tiga metode balancing (None, ROS, SMOTE) untuk menguji kombinasi fitur gaya hidup (termasuk sleep–stress interaction, caffeine usage, dan diet  $\times$  alcohol) dalam mendeteksi kategori kecemasan. Fokus penelitian ini adalah memberikan analisis komprehensif yang menggabungkan aspek teknik ML dan metode balancing untuk deteksi dini kecemasan berbasis gaya hidup.

### 3 Metode dan Model

Penelitian ini menggunakan sepuluh algoritma pembelajaran terawasi (*supervised learning*) yang umum digunakan dalam klasifikasi. Masing-masing model memiliki pendekatan dan karakteristik berbeda dalam mempelajari pola dari data. Tujuannya adalah untuk memprediksi kemungkinan seseorang mengalami gangguan kecemasan berdasarkan berbagai indikator gaya hidup dan kesehatan.

Model **Decision Tree** bekerja dengan membagi data secara rekursif berdasarkan fitur-fitur yang memberikan informasi klasifikasi paling besar, sehingga menghasilkan struktur pohon yang mudah diinterpretasikan. **Random Forest** adalah ansambel dari banyak pohon keputusan yang digabungkan untuk mengurangi variansi dan meningkatkan akurasi prediksi.

**K-Nearest Neighbors (KNN)** mengklasifikasikan suatu sampel berdasarkan mayoritas label dari beberapa tetangga terdekatnya di ruang fitur. **Logistic Regression** merupakan model linier yang memodelkan probabilitas suatu kejadian dengan fungsi logistik, dan cocok untuk klasifikasi biner atau multikelas dengan asumsi hubungan linier antara fitur dan logit.

**Gaussian Naive Bayes** menggunakan prinsip teorema Bayes dengan asumsi independensi antar fitur dan distribusi Gaussian, serta sangat efisien pada dataset dengan fitur banyak. **Support Vector Machine (SVM)** mencari hyperplane optimal yang memisahkan kelas-kelas secara maksimal dalam ruang dimensi tinggi, dan efektif untuk data dengan margin yang jelas.

**Artificial Neural Network (ANN)** meniru struktur otak manusia melalui jaringan neuron buatan dan dapat mempelajari hubungan kompleks non-linier antara fitur dan target. Model ini memanfaatkan beberapa lapisan tersembunyi yang dioptimasi melalui proses pembelajaran mendalam.

Tiga model lainnya merupakan bagian dari keluarga *gradient boosting* yang sangat kompetitif, yaitu **CatBoost**, **XGBoost**, dan **LightGBM**. Ketiganya menggunakan pendekatan bertahap untuk memperbaiki prediksi dari model sebelumnya, dan unggul dalam menangani fitur kategorikal, data besar, serta memiliki kemampuan generalisasi yang baik. Model-model ini juga dikenal karena efisiensi dan kecepatan proses

pelatihannya.

Seluruh model dibandingkan untuk menilai mana yang paling efektif dalam mengklasifikasikan kondisi mental berdasarkan fitur gaya hidup dan kesehatan individu.

## 4 Metode Evaluasi

### 4.1 Deskripsi dan Analisis Dataset

Tahapan awal dalam penelitian ini adalah pemahaman data. Dataset yang digunakan menjadi dasar bagi seluruh proses pemodelan dan evaluasi yang dilakukan.

#### 4.1.1 Sumber dan Karakteristik Data

Penelitian ini memanfaatkan dataset “Mental Health and Lifestyle Habits 2019–2024” dari Kaggle [10], yang memuat 18 variabel demografis dan gaya hidup. Setiap variabel memiliki peran spesifik dalam memprediksi tingkat kecemasan:

#### 4.1.2 Deskripsi Variabel

- **Age:** Umur responden dalam tahun; memengaruhi pola coping dan risiko kecemasan.
- **Gender:** Jenis kelamin (L/P); perbedaan biologis dan sosial dapat memengaruhi prevalensi kecemasan.
- **Occupation:** Kategori pekerjaan; stres kerja berbeda antar jenis profesi.
- **Sleep Hours:** Rata-rata jam tidur per hari; gangguan tidur sering terkait dengan kecemasan.
- **Physical Activity (hrs/week):** Jam aktivitas fisik mingguan; olahraga berpengaruh positif pada kesehatan mental.
- **Caffeine Intake (mg/day):** Asupan kafein harian; konsumsi tinggi berpotensi memicu kecemasan.
- **Alcohol Consumption (drinks/week):** Jumlah minuman beralkohol per minggu; konsumsi berlebih dapat memperburuk gejala.
- **Smoking:** Kebiasaan merokok (Ya/Tidak); nikotin memiliki efek stimulasi pada sistem saraf.
- **Family History of Anxiety:** Riwayat kecemasan dalam keluarga (Ya/Tidak); faktor genetik dan lingkungan.
- **Heart Rate (bpm)** dan **Breathing Rate (breaths/min):** Parameter fisiologis yang mencerminkan respons stres akut.
- **Sweating Level (1–5)** dan **Dizziness:** Gejala somatik kecemasan seperti keringat berlebih dan pusing.
- **Medication:** Penggunaan obat psikotropika (Ya/Tidak); dapat mengindikasikan riwayat gangguan mental.
- **Therapy Sessions (per month):** Jumlah sesi terapi; semakin banyak sesi mencerminkan tingkat dukungan atau keparahan.



- **Recent Major Life Event:** Peristiwa besar terakhir (Ya/Tidak); stresor eksternal memicu kecemasan.
- **Diet Quality (1–10):** Skor pola makan sehat; nutrisi memengaruhi mood dan fungsi kognitif.
- **Stress Level (1–10):** Skor stres subjektif yang kemudian dikategorikan untuk analisis klasifikasi.

#### 4.1.3 Transformasi Variabel Target

Variabel target asli dalam dataset ini adalah `Anxiety_Level`, sebuah metrik dengan rentang skor dari 1 hingga 10. Untuk final project ini, dilakukan sebuah langkah pra-pemrosesan pada variabel target untuk menyederhanakan masalah klasifikasi dan menciptakan kelas-kelas yang lebih dapat diinterpretasikan secara klinis, 10 tingkatan skor tersebut diagregasi menjadi **lima kategori tingkat kecemasan** yang definitif. Kelima kategori ini adalah:

1. Very Low
2. Low
3. Medium
4. High
5. Very High

Variabel kategorikal baru inilah yang selanjutnya digunakan sebagai variabel target (`Anxiety_Category`) dalam seluruh proses pelatihan dan evaluasi model klasifikasi pada penelitian ini.

#### 4.1.4 Exploratory Data Analysis

Sebelum membangun model, dilakukan *Exploratory Data Analysis* (EDA) secara menyeluruh untuk memahami karakteristik data:

- **Ringkasan Statistik:** Analisis deskriptif (mean, median, modus, standar deviasi, kuartil) untuk memetakan sebaran setiap fitur numerik membantu mendeteksi outlier dan skala variabel.
- **Inspeksi dan Imputasi Nilai Hilang:** Persentase dan pola NaN diidentifikasi. Nilai kosong pada fitur numerik diisi dengan rata-rata, sedangkan fitur kategorikal diisi dengan modus untuk mempertahankan distribusi asli.
- **Analisis Korelasi:** Matriks korelasi  $\text{corr}(X_i, X_j)$  dihitung untuk seluruh pasangan fitur numerik, kemudian divisualisasikan sebagai heatmap agar mudah mengidentifikasi fitur yang sangat berkorelasi (redundan) dan memutuskan apakah perlu reduksi dimensi.

- **Distribusi Kategori:** Histogram dan bar-plot dibuat untuk memeriksa keseimbangan kelas target (`Anxiety_Category`) serta variabel kategorikal utama (misalnya `Stress Level`, `Occupation`), sehingga dapat memprediksi potensi masalah imbalance.

Temuan EDA ini menjadi dasar pemilihan teknik preprocessing.

## 4.2 Desain Eksperimen

Desain eksperimental mengikuti skema hold-out sederhana namun robust:

- **Training set (80%):** Digunakan untuk fitting model. Proporsi ini dianggap cukup untuk menangkap pola kompleks dalam data.
- **Test set (20%):** Disimpan terpisah dan tidak disentuh selama pelatihan, sehingga evaluasi mencerminkan kinerja pada data baru.

Pembagian data dilakukan secara acak dengan parameter `random_state=42` untuk reproducibility. Meskipun hold-out tidak memanfaatkan semua data untuk validasi silang, metode ini cepat dan mudah diterapkan pada dataset besar.

## 4.3 Strategi Penyeimbangan Kelas

Ketidakseimbangan kelas target dapat menurunkan kemampuan model dalam mendeteksi kelas minoritas. Oleh karena itu, dipilih tiga strategi yang hanya diterapkan pada data *training*:

1. **None (Baseline).**
2. **ROS (Random Over-Sampling).** Menambah jumlah sampel minoritas dengan duplikasi acak sehingga:

$$|C_{\text{minor}}| \approx |C_{\text{major}}|. \quad (1)$$

Meskipun sederhana, risiko overfitting pada sampel duplikat perlu diwaspadai.

3. **SMOTE (Synthetic Minority Over-sampling).** Menghasilkan sampel sintetis antara tetangga minoritas:

$$\mathbf{x}_{\text{baru}} = \mathbf{x}_i + \lambda(\mathbf{x}_{\text{nn}} - \mathbf{x}_i), \quad \lambda \sim \mathcal{U}(0, 1) \quad (2)$$

sehingga menciptakan variasi baru dan umumnya mengurangi overfitting dibandingkan ROS.

## 4.4 Metrik Evaluasi

Untuk menilai kinerja model klasifikasi secara komprehensif, digunakan metrik berikut: **Accuracy:** Proporsi prediksi yang benar terhadap total sampel:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

**Precision:** Kemampuan model dalam menghasilkan prediksi positif yang benar:

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

**Recall:** Kemampuan model dalam menemukan seluruh kasus positif:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

**F1 Score (Makro):** Rata-rata harmonis antara precision dan recall yang dihitung per kelas, lalu dirata-rata:

$$F1_i = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}, \quad F1_{\text{Macro}} = \frac{1}{K} \sum_{i=1}^K F1_i \quad (6)$$

Metrik ini sangat berguna ketika distribusi kelas tidak merata.

**Confusion Matrix:** Matriks berukuran  $K \times K$  yang mencatat distribusi prediksi model untuk setiap kombinasi kelas aktual dan kelas yang diprediksi. Nilai diagonal menyatakan prediksi benar, sedangkan nilai non-diagonal menunjukkan kesalahan klasifikasi.

## 4.5 Prosedur Eksperimen

Langkah-langkah eksperimen dirinci sebagai berikut:

1. **Preprocessing:**

- Imputasi nilai hilang (mean/modus).
- Encoding kategorikal (`LabelEncoder`).
- Normalisasi numerik (`StandardScaler`).

2. **Penyeimbangan Kelas:** Terapkan ROS atau SMOTE pada training set sesuai konfigurasi.

3. **Pelatihan Model:** Latih sepuluh model klasifikasi:

$$\{\text{DecisionTree}, \text{RandomForest}, \text{KNN}, \dots, \text{LightGBM}\}.$$

4. **Evaluasi:** Jalankan prediksi pada test set, hitung metrik (Rumus 3–6), dan simpan:

- `model_metrics.csv`
- Visualisasi heatmap dan grafik perbandingan metrik di folder `plots/`

5. **Analisis Komparatif:** Bandingkan performa model dan strategi balancing, identifikasi trade-off antara akurasi dan sensitivitas terhadap kelas minoritas.

## 5 Hasil Evaluasi Model

Hasil evaluasi performa dari sepuluh model klasifikasi: Decision Tree, Random Forest, KNN, Logistic Regression, Gaussian Naive Bayes, SVM, ANN, CatBoost, XGBoost.

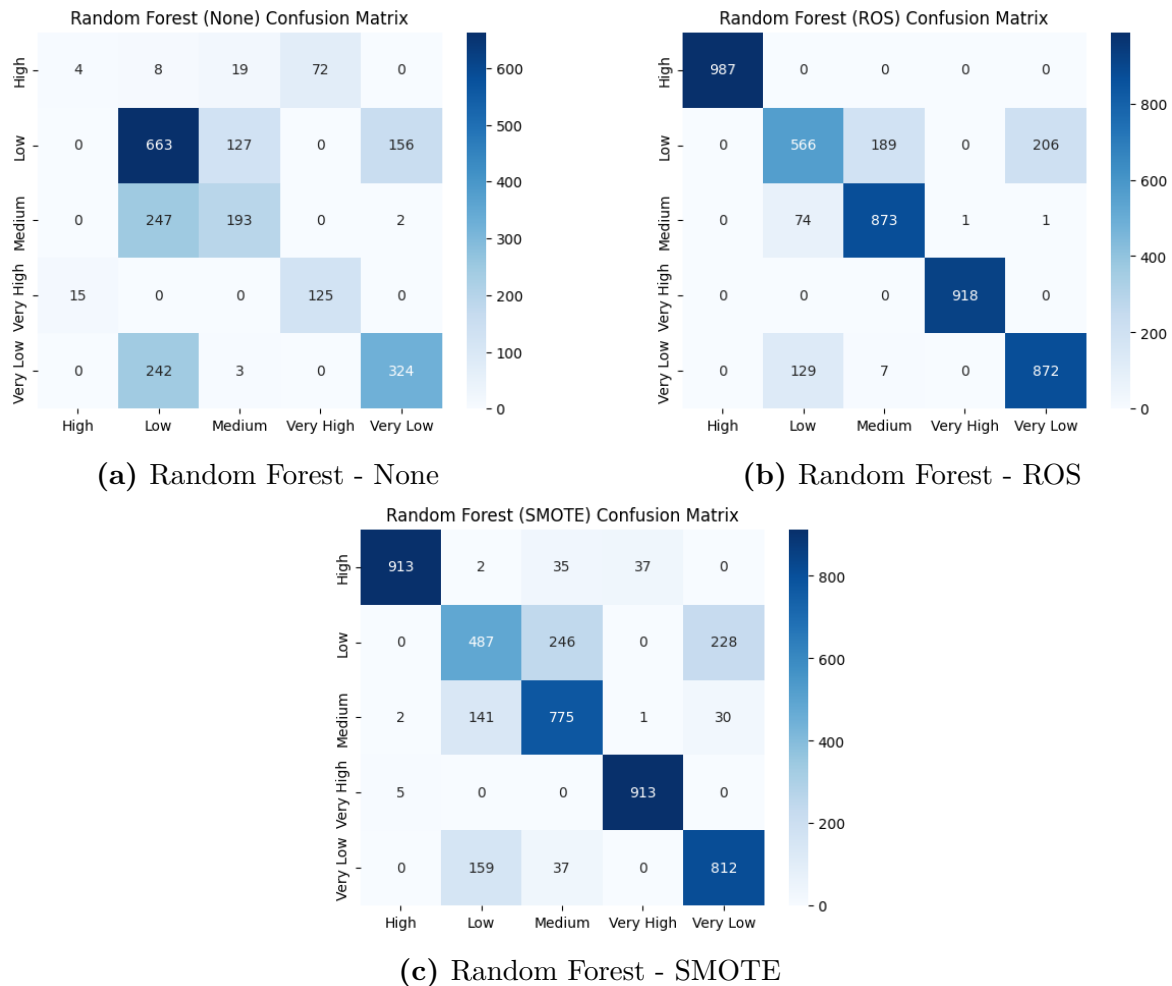
ost, dan LightGBM dilakukan dalam tiga skenario penanganan ketidakseimbangan kelas: tanpa penyeimbangan data, dengan teknik *Synthetic Minority Over-sampling Technique* (SMOTE), dan dengan *Random Oversampling* (ROS). Kinerja setiap model pertama-tama dianalisis secara kualitatif melalui confusion matrix untuk memahami pola prediksi dan jenis kesalahan yang dibuat pada masing-masing kelas.



**GAMBAR 1.** Confusion matrix untuk Decision Tree.

Analisis confusion matrix untuk model Decision Tree mengilustrasikan sensitivitasnya yang tinggi terhadap distribusi data. Pada skenario tanpa balancing (gambar a), model menunjukkan performa yang tidak konsisten, dengan kegagalan signifikan pada kelas-kelas minoritas seperti "High" (hanya 37 prediksi benar) dan "Very High" (90 prediksi benar). Setelah penerapan teknik balancing, baik ROS (gambar b) maupun SMOTE (gambar c) memberikan peningkatan performa yang substansial. Teknik ROS menghasilkan model dengan akurasi per kelas yang sangat tinggi, yang secara dramatis meningkatkan prediksi benar untuk "High" menjadi 987 dan "Very High" menjadi 915, dengan kesalahan klasifikasi yang minimal. Di sisi lain, SMOTE, meskipun mencatatkan nilai diagonal yang lebih rendah (misalnya, 825 untuk "High"), menunjukkan karakteristik yang berbeda. Kerentanan Decision Tree terhadap overfitting menjadi

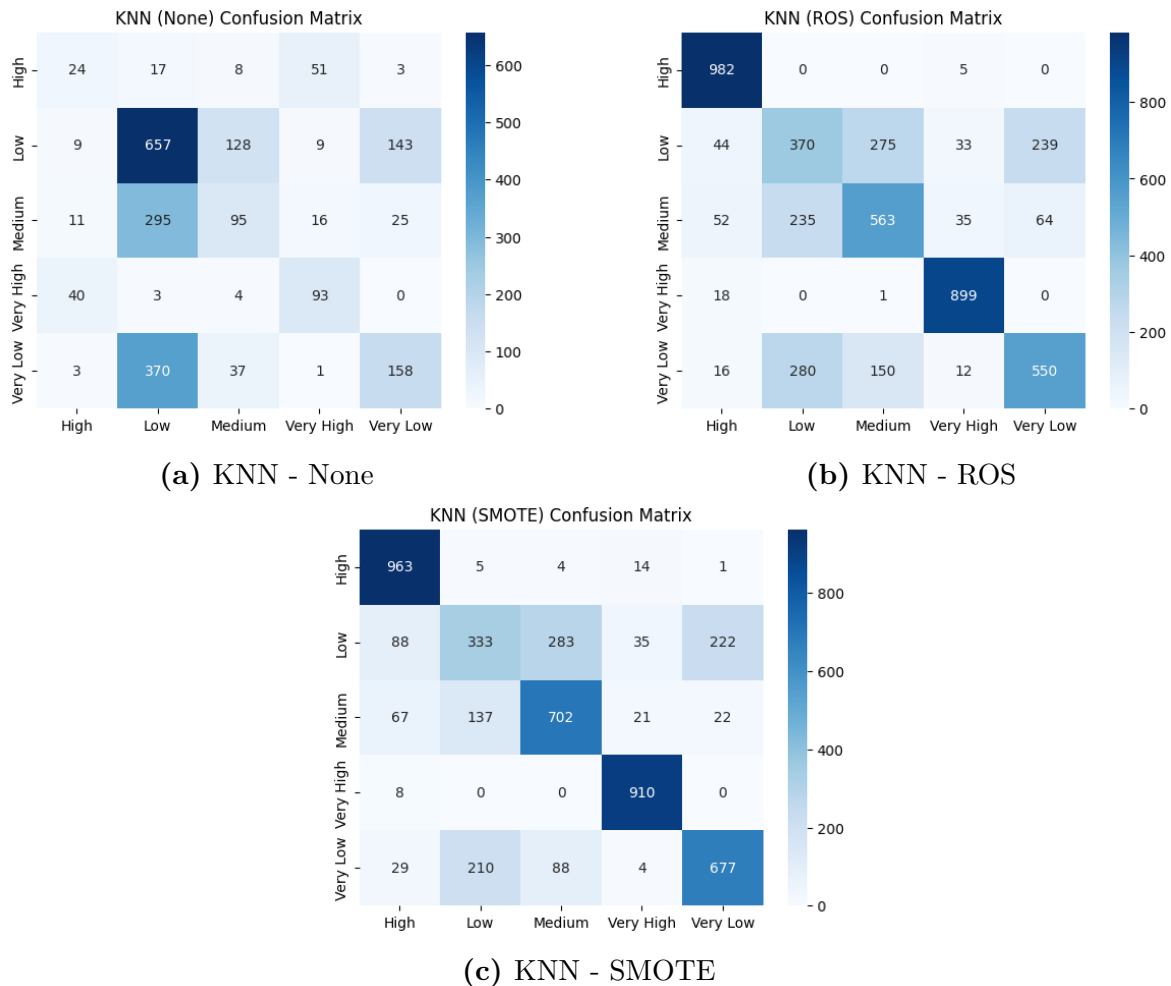
sangat nyata dalam perbandingan ini. Pendekatan duplikasi oleh ROS berisiko menciptakan aturan yang terlalu kaku dan terspesialisasi, sedangkan SMOTE mendorong pembentukan batas keputusan yang lebih umum, meskipun ini menghasilkan tingkat kebingungan yang lebih tinggi antar kelas yang berdekatan, seperti 242 sampel "Medium" yang salah diklasifikasikan sebagai "Low".



**GAMBAR 2.** Confusion matrix untuk Random Forest.

Model Random Forest, yang merupakan ensemble dari Decision Tree, menunjukkan ketahanan yang lebih baik namun tetap mendapat manfaat signifikan dari penyeimbangan data. Pada skenario tanpa balancing (gambar a), model menunjukkan kelemahan fatal dalam mengenali kelas minoritas, dengan hanya 4 prediksi benar untuk kelas "High" dan 125 untuk "Very High", yang mengindikasikan bias yang kuat meskipun menggunakan mekanisme ensemble. Setelah data diseimbangkan, baik ROS (gambar b) maupun SMOTE (gambar c) menghasilkan peningkatan performa yang sangat besar. ROS kembali menunjukkan matriks yang sangat bersih dengan akurasi per kelas yang sangat tinggi, seperti 987 prediksi benar untuk "High" dan 873 untuk "Medium". SMOTE, di sisi lain, mencapai nilai yang sedikit lebih rendah pada beberapa kelas (misalnya, 913 untuk "High" dan 775 untuk "Medium"). Perbedaan ini

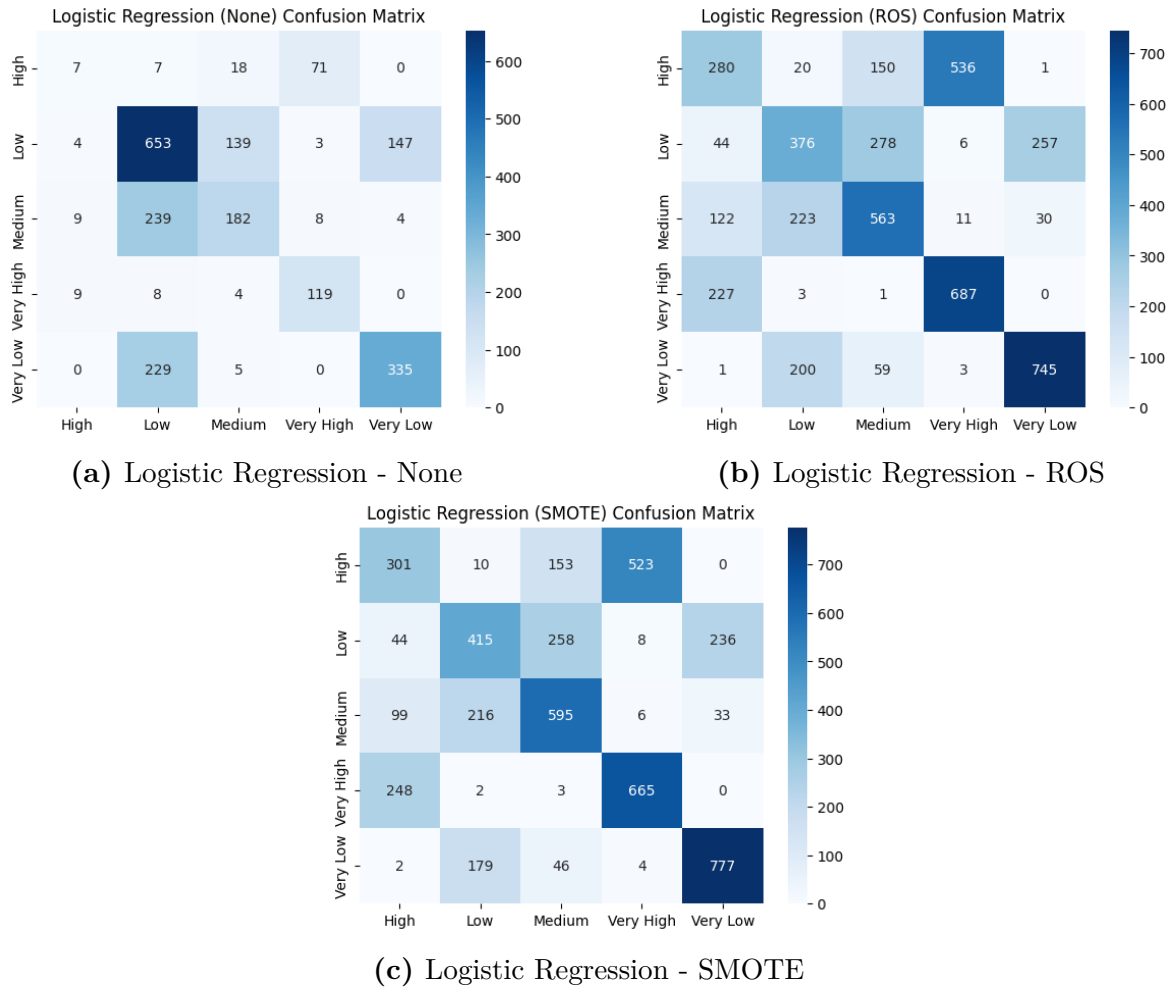
mengindikasikan bahwa sementara ROS memaksimalkan akurasi pada sampel-sampel yang ada, pendekatan SMOTE dalam menciptakan data sintetis mendorong Random Forest untuk membentuk batas keputusan yang lebih tergeneralisasi, yang terlihat dari adanya tingkat kebingungan yang lebih tinggi antar kelas yang berdekatan, seperti 141 sampel "Medium" yang salah diklasifikasikan sebagai "Low".



**GAMBAR 3.** Confusion matrix untuk KNN.

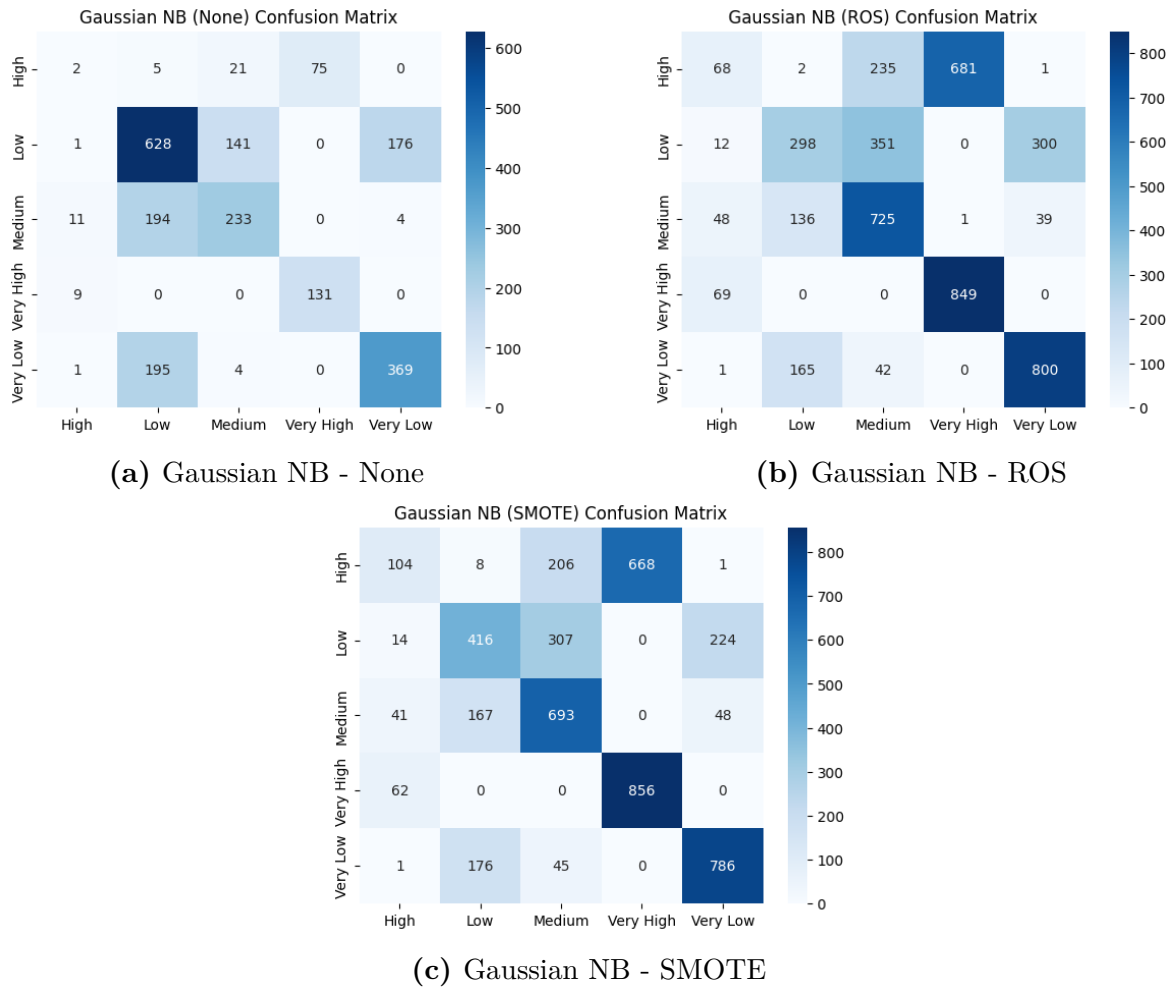
Evaluasi terhadap model K-Nearest Neighbors (KNN) memperlihatkan sensitivitasnya yang tinggi terhadap distribusi kelas, yang merupakan karakteristik umum dari algoritma berbasis jarak. Pada kondisi tanpa balancing (gambar a), performa model sangat tidak seimbang, dengan hanya 24 prediksi benar untuk kelas "High" dan 95 untuk "Medium", serta kecenderungan kuat untuk salah mengklasifikasikan sampel sebagai "Low". Setelah penerapan teknik balancing, baik ROS (gambar b) maupun SMOTE (gambar c), terjadi peningkatan performa yang signifikan dalam pengenalan kelas-kelas minoritas. Walaupun kedua teknik berhasil meningkatkan jumlah prediksi benar, SMOTE menunjukkan keunggulan yang jelas. Hal ini terlihat dari kemampuannya memprediksi kelas "Medium" secara lebih akurat dengan 702 prediksi benar dibandingkan 563 pada ROS, serta 677 prediksi benar untuk "Very Low" dibandingkan 550

pada ROS. Sebagai algoritma berbasis jarak, KNN mendapat manfaat lebih dari cara SMOTE menciptakan lingkungan data (neighborhoods) yang lebih padat dan kohesif untuk kelas minoritas melalui sampel sintesis, yang memungkinkan model membentuk batas keputusan yang lebih akurat daripada hanya mengandalkan duplikasi titik data.



**GAMBAR 4.** Confusion matrix untuk Logistic Regression.

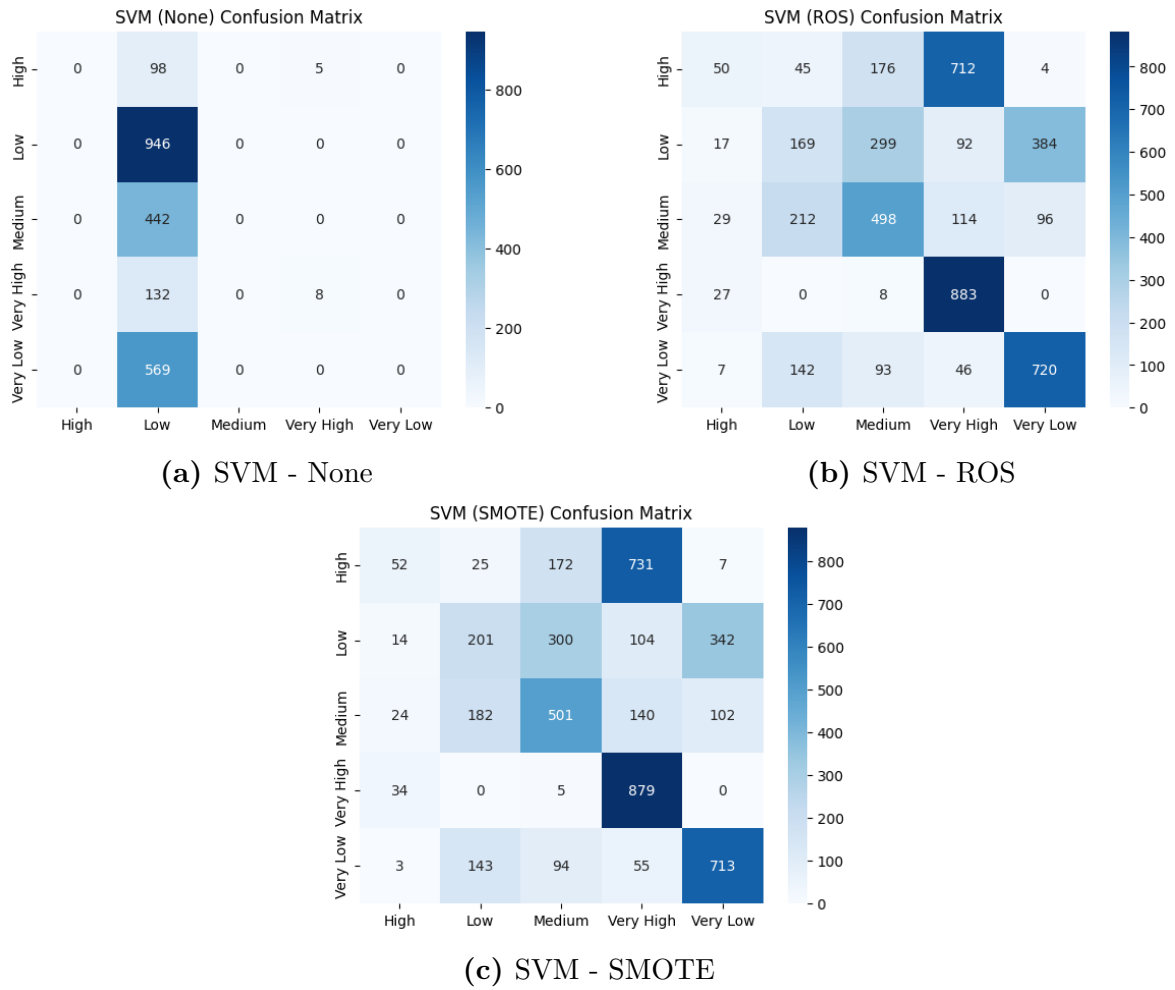
Analisis pada model Regresi Logistik mengungkapkan bagaimana penyeimbangan data tidak hanya memperbaiki performa, tetapi juga dapat menunjukkan kelemahan dari algoritma ini. Pada kondisi tanpa balancing (gambar a), model ini hampir sepenuhnya gagal, yang secara jelas ditunjukkan oleh hanya 7 prediksi benar untuk kelas "High". Setelah data diseimbangkan, kedua teknik menyingkap kelemahan fundamental dari model ini. Meskipun SMOTE (gambar c) menunjukkan hasil yang sedikit lebih unggul daripada ROS (gambar b) dengan meningkatkan prediksi "High" yang benar menjadi 301 dari 280, keduanya tetap tidak mampu mengatasi kebingungan tinggi antara kelas "High" dan "Very High", dengan 523 sampel "High" salah diklasifikasikan sebagai "Very High" pada skenario SMOTE. Kegigihan kesalahan ini mengindikasikan bahwa batas keputusan linear yang dibentuk oleh Regresi Logistik tidak cukup kompleks untuk memisahkan kedua kelas tersebut secara efektif.



**GAMBAR 5.** Confusion matrix untuk Gaussian Naive Bayes.

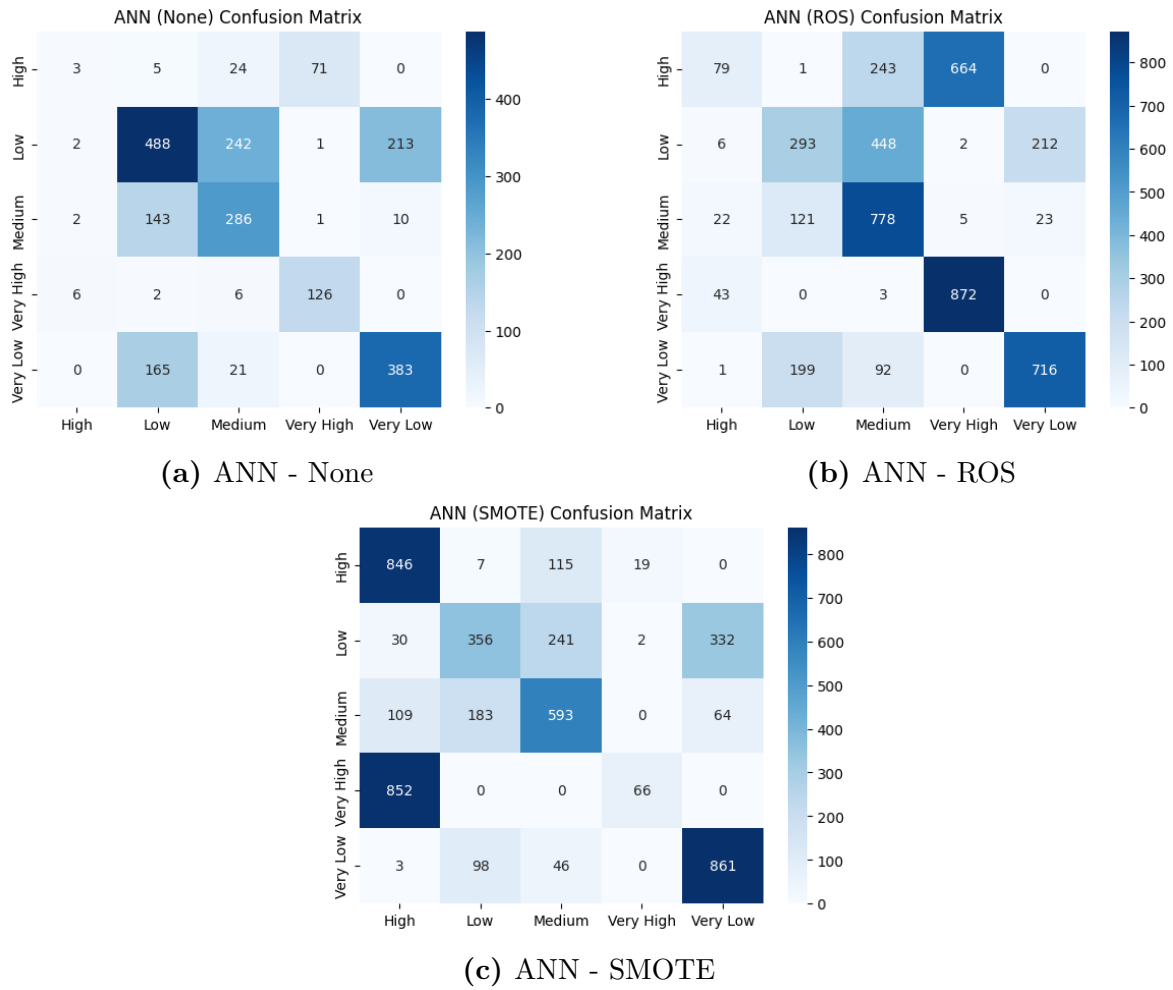
Model Gaussian Naive Bayes (NB) memperlihatkan keterbatasan paling signifikan di antara semua model yang diuji. Pada kondisi tanpa balancing (gambar a), model ini hampir tidak memiliki kemampuan prediktif, yang secara dramatis ditunjukkan dengan hanya 2 prediksi benar untuk kelas "High". Meskipun penerapan teknik balancing memberikan perbaikan, performa model secara umum tetap rendah dan justru menyoroti kelemahan fundamentalnya. Teknik SMOTE (gambar c) menunjukkan hasil yang lebih baik daripada ROS (gambar b), terutama dalam meningkatkan prediksi benar untuk kelas "High" dari 68 menjadi 104. Namun, kedua teknik tersebut tetap tidak mampu mengatasi masalah utama model ini. Tingkat kebingungan yang sangat tinggi tetap ada, terutama antara kelas "High" dan "Very High", di mana pada skenario SMOTE pun masih terdapat 668 sampel yang salah diklasifikasikan. Performa yang rendah secara keseluruhan ini mengindikasikan bahwa asumsi dasar Gaussian NB kemungkinan besar tidak terpenuhi, sehingga model probabilistik sederhana ini tidak mampu memodelkan hubungan antar variabel yang kompleks secara akurat.





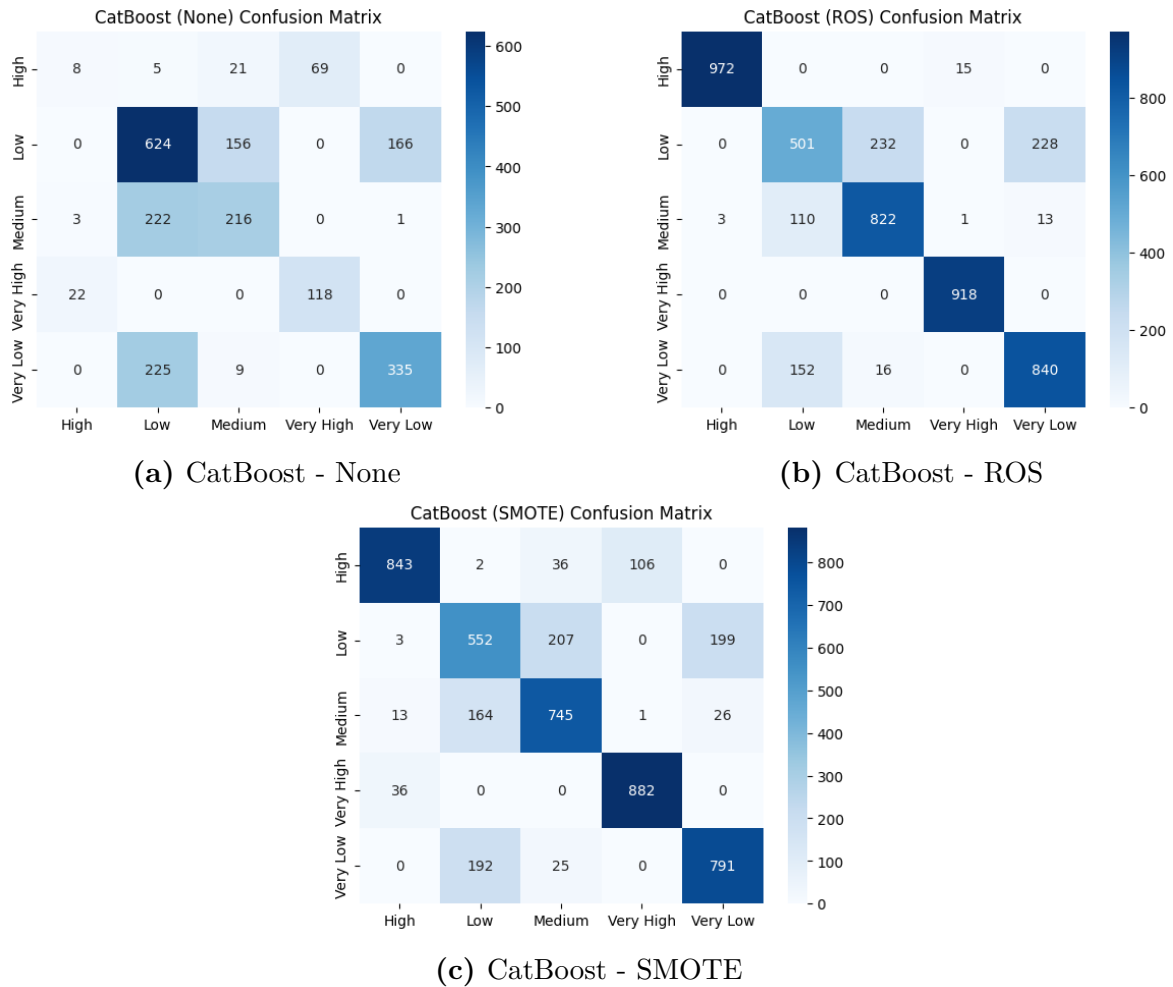
**GAMBAR 6.** Confusion matrix untuk Support Vector Machine.

Model Support Vector Machine (SVM) menunjukkan ketergantungan paling ekstrem terhadap keseimbangan data dibandingkan model lainnya. Pada skenario tanpa balancing (gambar a), model mengalami kegagalan total, di mana ia hanya memprediksi satu kelas ("Low") untuk hampir semua sampel dan sama sekali tidak dapat mengidentifikasi kelas "High", "Medium", dan "Very Low" (masing-masing 0 prediksi benar). Fenomena ini menunjukkan ketidakmampuan SVM untuk menemukan hyper-plane pemisah yang memadai pada data yang sangat tidak seimbang. Setelah penerapan balancing, baik ROS (gambar b) maupun SMOTE (gambar c), model mulai dapat membedakan antar kelas, namun dengan kelemahan signifikan yang terungkap. Kedua teknik menunjukkan bahwa SVM sangat kesulitan membedakan kelas "High" dan "Very High"; pada skenario SMOTE, sebanyak 731 sampel "High" salah diklasifikasikan sebagai "Very High". Kesulitan yang persisten dalam membedakan level kecemasan yang berdekatan namun krusial ini mengindikasikan bahwa struktur data pada penelitian ini kemungkinan tidak cocok untuk pendekatan klasifikasi berbasis margin yang digunakan oleh SVM.



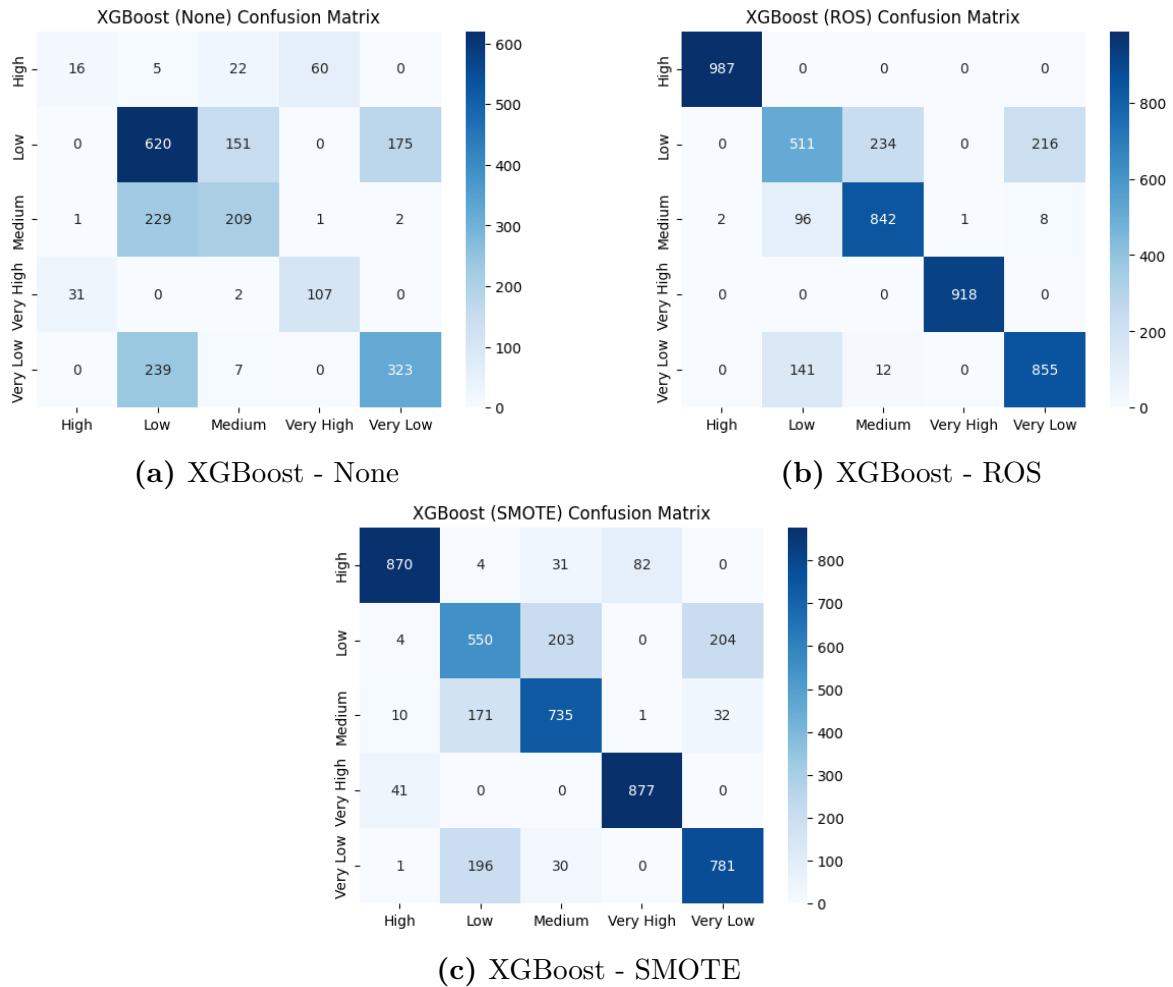
**GAMBAR 7.** Confusion matrix untuk Artificial Neural Network.

Model Artificial Neural Network (ANN) menunjukkan sensitivitas yang tinggi terhadap ketidakseimbangan kelas, sekaligus memperlihatkan bagaimana teknik balancing yang berbeda dapat menghasilkan pola kesalahan yang unik. Pada kondisi tanpa balancing (gambar a), model ANN memiliki kinerja yang sangat buruk, dengan hanya 3 prediksi benar untuk kelas "High", yang menunjukkan bias kuat terhadap kelas mayoritas. Setelah penerapan ROS (gambar b), muncul pola kesalahan baru yang signifikan: model menjadi sangat bingung antara kelas "High" dan "Very High", di mana 664 sampel "High" salah diklasifikasikan sebagai "Very High". Uniknya, teknik SMOTE (gambar c) menghasilkan pola kesalahan yang berkebalikan. Model menjadi sangat akurat dalam mengenali kelas "High" (846 prediksi benar), namun dengan "mengorbankan" kelas "Very High" yang kini hampir seluruhnya salah diklasifikasikan sebagai "High" (852 kesalahan). Pola kesalahan yang drastis dan berkebalikan ini mengindikasikan bahwa model ANN pada konfigurasi ini sangat sensitif terhadap metode penyeimbangan data, dan kemampuannya untuk membedakan kelas "High" dan "Very High" sangat bergantung pada distribusi data latih yang dihasilkan oleh ROS atau SMOTE.



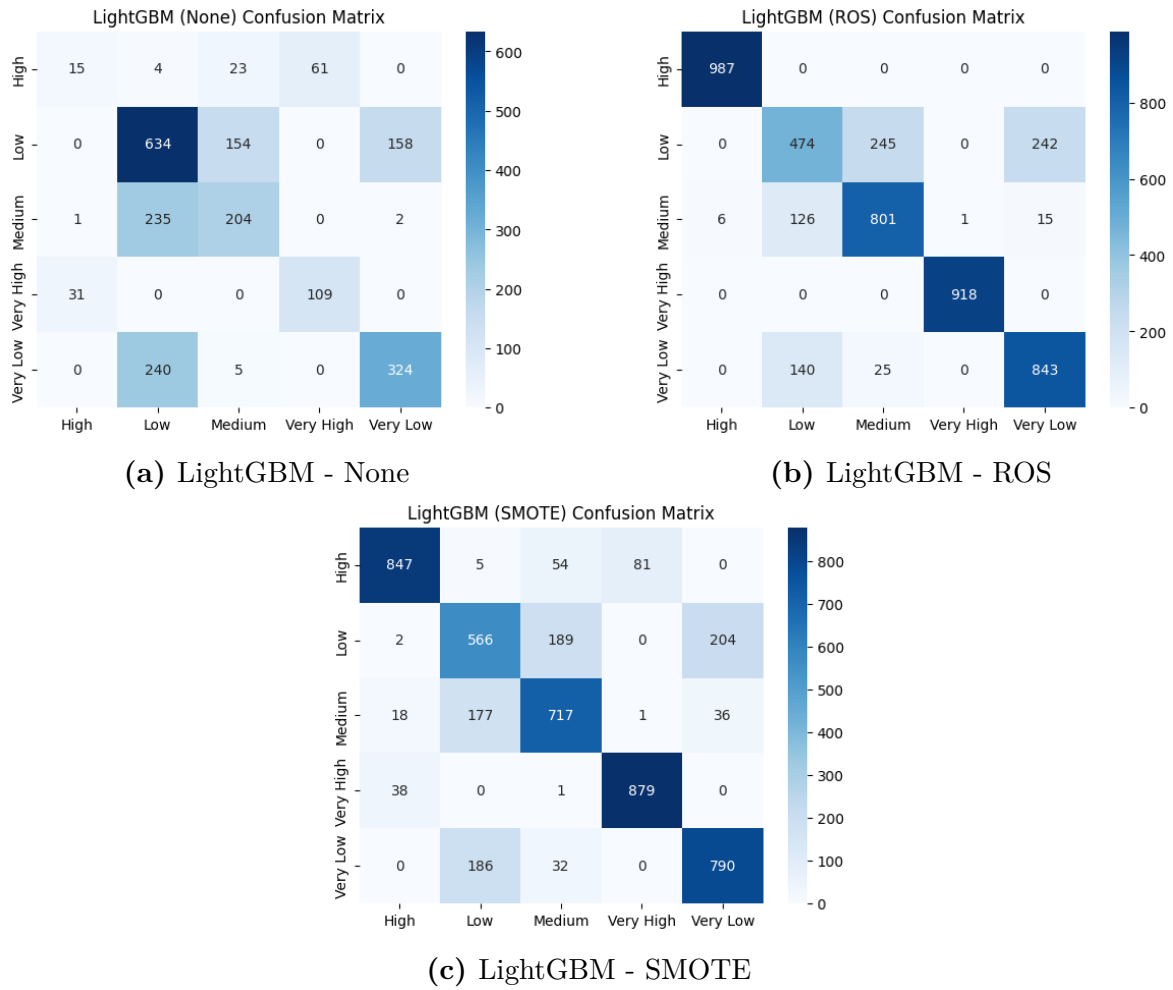
**GAMBAR 8.** Confusion matrix untuk CatBoost.

Model CatBoost, sebagai salah satu algoritma gradient boosting, menunjukkan peningkatan performa yang sangat besar setelah penanganan data tidak seimbang. Pada kondisi awal tanpa balancing (gambar a), kapabilitas model sangat terbatas, yang secara spesifik terlihat dari hanya 8 prediksi benar untuk kelas "High" dan 118 untuk kelas "Very High". Setelah aplikasi teknik balancing, baik ROS (gambar b) maupun SMOTE (gambar c) secara dramatis meningkatkan kemampuan prediksi model di semua kelas. Teknik ROS menghasilkan matriks dengan akurasi per kelas yang sangat tinggi, misalnya 972 prediksi benar untuk "High" dan 918 untuk "Very High". Di sisi lain, SMOTE, meskipun mencatatkan nilai diagonal yang sedikit lebih rendah pada beberapa kelas (misalnya 843 untuk "High" dan 745 untuk "Medium" dibandingkan 822 pada ROS), menunjukkan karakteristik pertukaran kinerja. Pendekatan SMOTE yang menghasilkan sampel sintesis mendorong CatBoost untuk membangun model yang lebih tergeneralisasi, yang terlihat dari adanya tingkat kebingungan yang lebih tinggi antar kelas yang berdekatan, seperti 164 sampel "Medium" yang salah diklasifikasikan sebagai "Low".



**GAMBAR 9.** Confusion matrix untuk XGBoost.

Model XGBoost, sebagai salah satu algoritma boosting terkemuka, menunjukkan peningkatan performa yang sangat signifikan ketika dikombinasikan dengan teknik penyeimbangan data. Pada kondisi tanpa balancing (gambar a), performa model sangat terbatas, yang secara jelas terilustrasi oleh kemampuannya yang sangat rendah dalam mengidentifikasi kelas "High", dengan hanya 16 prediksi yang benar. Setelah penerapan teknik balancing, kapabilitas model meningkat secara dramatis. Teknik ROS (gambar b) menghasilkan matriks yang secara visual sangat akurat, dengan nilai diagonal yang sangat tinggi, seperti 987 prediksi benar untuk kelas "High". Sementara itu, SMOTE (gambar c), meskipun mencatatkan nilai yang sedikit lebih rendah pada kelas tersebut (870 prediksi benar), menunjukkan pertukaran kinerja yang penting untuk model yang rentan overfitting seperti XGBoost. Pendekatan SMOTE yang menghasilkan data sintetis mendorong model untuk membangun batas keputusan yang lebih tergeneralisasi, yang terlihat dari adanya kebingungan yang lebih merata antar kelas yang berdekatan, misalnya 171 sampel "Medium" yang salah diklasifikasikan sebagai "Low".



**GAMBAR 10.** Confusion matrix untuk LightGBM.

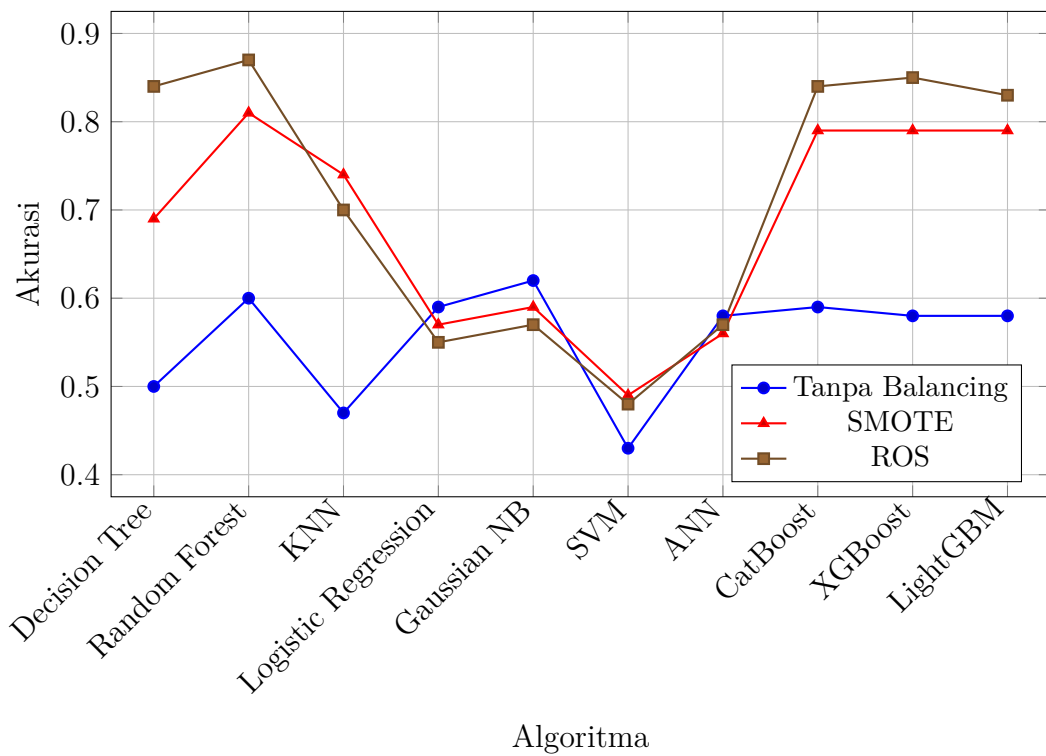
Model LightGBM, sebagai algoritma boosting yang efisien, menunjukkan pola peningkatan performa yang konsisten dengan model berbasis pohon lainnya ketika dihadapkan pada data yang diseimbangkan. Pada pengujian dengan data asli tanpa balancing (gambar a), model ini menunjukkan kelemahan signifikan, terbukti dari kemampuannya yang sangat minim dalam mengenali kelas "High" dengan hanya 15 prediksi benar. Implementasi teknik balancing, baik ROS (gambar b) maupun SMOTE (gambar c), secara substansial memperbaiki kapabilitas model tersebut. Secara spesifik, ROS menghasilkan model dengan akurasi per kelas yang sangat tinggi, yang secara dramatis meningkatkan prediksi benar untuk kelas "High" menjadi 987. Di sisi lain, SMOTE, meskipun mencatatkan nilai yang lebih rendah pada kategori tersebut (847 prediksi benar), menunjukkan sebuah pertukaran kinerja yang menghasilkan model yang lebih tergeneralisasi. Pendekatan SMOTE dalam menciptakan sampel sintetis terlihat menghasilkan kebingungan yang lebih merata antar kelas yang berdekatan, misalnya 177 sampel "Medium" yang salah diklasifikasikan sebagai "Low", yang mengindikasikan pembentukan batas keputusan yang lebih 'lunak' namun lebih robust secara keseluruhan.

## 5.1 Analisis Kuantitatif

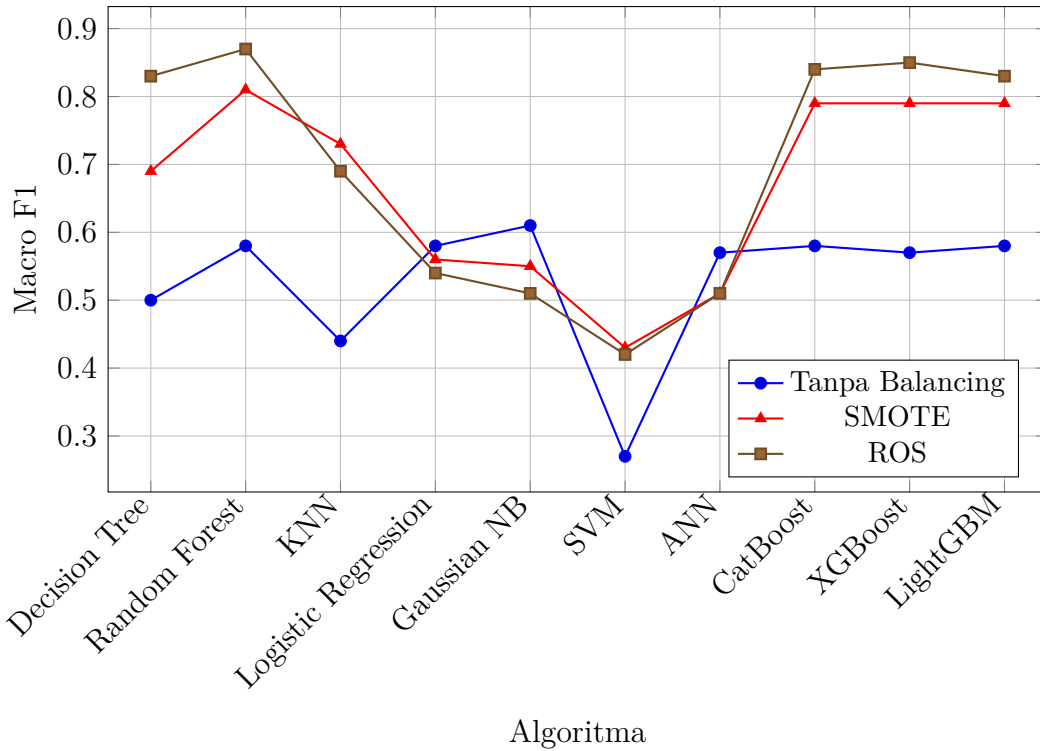
Untuk melengkapi analisis kualitatif dari confusion matrix, evaluasi kuantitatif dilakukan dengan membandingkan metrik performa utama di seluruh model dan teknik penyeimbangan. Tabel 1 merepresentasikan nilai Accuracy dan Macro F1-Score untuk 30 kombinasi eksperimen. Selanjutnya, Gambar 11 dan Gambar 12 memvisualisasikan perbandingan ini untuk memberikan gambaran yang jelas mengenai dampak signifikan dari setiap strategi penyeimbangan data.

**TABEL 1.** Akurasi dan Macro F1-Score

Model	Tanpa Balancing		SMOTE		ROS	
	Accuracy	Macro F1	Accuracy	Macro F1	Accuracy	Macro F1
Decision Tree	0.50	0.50	0.69	0.69	0.84	0.83
Random Forest	0.60	0.58	0.81	0.81	<b>0.87</b>	<b>0.87</b>
KNN	0.47	0.44	0.74	0.73	0.70	0.69
Logistic Regression	0.59	0.58	0.57	0.56	0.55	0.54
Gaussian NB	0.62	0.61	0.59	0.55	0.57	0.51
SVM	0.43	0.27	0.49	0.43	0.48	0.42
ANN	0.58	0.57	0.56	0.51	0.57	0.51
CatBoost	0.59	0.58	0.79	0.79	0.84	0.84
XGBoost	0.58	0.57	0.79	0.79	0.85	0.85
LightGBM	0.58	0.58	0.79	0.79	0.83	0.83



**GAMBAR 11.** Perbandingan akurasi model terhadap teknik balancing



**GAMBAR 12.** Perbandingan Macro F1-Score model terhadap teknik balancing

## 5.2 Pembahasan Hasil

Dari data yang disajikan pada Tabel 1 dan divisualisasikan pada Gambar 11 dan 12, beberapa temuan kunci dapat ditarik. Pertama, intervensi penyeimbangan data memberikan dampak yang sangat bervariasi tergantung pada jenis algoritma. Untuk model-model ensemble berbasis pohon dan KNN, penerapan ROS maupun SMOTE secara konsisten memberikan peningkatan performa yang dramatis. Dampak ini paling terlihat pada metrik Macro F1-Score, di mana model-model tersebut melonjak dari kisaran 0.50-0.60 ke atas 0.80. Namun, untuk model Logistic Regression, Gaussian NB, dan ANN, teknik oversampling justru cenderung menurunkan performa, mengindikasikan bahwa keterbatasan fundamental model lebih dominan daripada masalah ketidakseimbangan kelas.

Kedua, model ensemble berbasis pohon keputusan (Random Forest, XGBoost, CatBoost, LightGBM) secara konsisten menjadi kelompok model dengan performa terbaik setelah data diseimbangkan. Di antara semua kombinasi, Random Forest yang dipadukan dengan ROS muncul sebagai konfigurasi paling unggul, mencapai Accuracy 0.87 dan Macro F1-Score 0.87. Hal ini mengindikasikan bahwa arsitektur ensemble pada Random Forest, yang dirancang untuk mengurangi overfitting, bekerja sangat efektif ketika diberi data yang seimbang secara representatif oleh ROS.

Ketiga, saat membandingkan kedua teknik oversampling, ROS secara umum menghasilkan skor akurasi dan F1 yang lebih tinggi daripada SMOTE untuk model-model

berkinerja terbaik. Meskipun analisis pada confusion matrix menyiratkan bahwa SMOTE dapat mendorong generalisasi yang lebih baik dengan menciptakan sampel sintetis, metrik evaluasi pada set data uji ini menunjukkan bahwa duplikasi sampel sederhana oleh ROS sudah cukup untuk memungkinkan model-model ensemble mempelajari batas keputusan yang sangat efektif. Sebaliknya, model seperti KNN mendapat manfaat lebih besar dari SMOTE (0.74), yang konsisten dengan sifatnya sebagai algoritma berbasis jarak yang diuntungkan oleh pembentukan lingkungan data (neighborhoods) yang lebih padat.

Model-model yang lebih sederhana seperti Logistic Regression dan Gaussian Naive Bayes tidak menunjukkan peningkatan signifikan, bahkan mengalami penurunan kinerja dalam beberapa kasus setelah oversampling. Ini menyoroti bahwa keterbatasan fundamental mereka asumsi linearitas pada Logistic Regression dan asumsi independensi fitur pada Gaussian NB merupakan penghalang kinerja yang lebih besar daripada ketidakseimbangan kelas itu sendiri. Demikian pula, performa SVM dan ANN tetap berada di bawah model ensemble, menandakan bahwa dengan konfigurasi saat ini, keduanya kurang cocok untuk dataset ini dibandingkan dengan ensemble berbasis pohon.

## 6 Kesimpulan

Penelitian ini membandingkan sepuluh algoritma pembelajaran mesin untuk deteksi dini tingkat kecemasan berbasis gaya hidup, dengan perhatian khusus pada ketidakseimbangan kelas. Hasil menunjukkan bahwa tanpa balancing, performa model cenderung rendah dan tidak stabil, terutama pada metrik Macro F1-Score yang sensitif terhadap kelas minoritas.

Setelah penerapan teknik balancing, model ensemble pohon (Random Forest, CatBoost, XGBoost, LightGBM) secara konsisten mendominasi. Kombinasi Random Forest dengan Random Oversampling (ROS) menempati puncak performa mencapai akurasi 0,87 dan Macro F1-Score 0,87 sementara SMOTE juga memberikan peningkatan namun sedikit di bawah ROS.

Untuk penelitian selanjutnya, disarankan mengeksplorasi teknik balancing adaptif seperti ADASYN, menguji ensemble stacking atau arsitektur neural network dengan fine-tuning hyperparameter, serta melakukan analisis kepentingan fitur dan metrik per kelas. Pendalaman ini diharapkan memberikan wawasan yang lebih kaya tentang variabel gaya hidup mana yang paling berpengaruh dan meningkatkan generalisasi model pada populasi yang lebih luas.

## Pustaka

- [1] Hautekiet, P., *et al.* (2022). A healthy lifestyle is positively associated with mental health and well-being and core markers in ageing. *BMC Medicine*, 20(1), 328.
- [2] Grajek, M., *et al.* (2022). Nutrition and mental health: A review of current knowledge about the impact of diet on mental health. *Frontiers in Nutrition*, 9, 943998.



- [3] Basiri, R., *et al.* (2023). Exploring the interrelationships between diabetes, nutrition, anxiety, and depression: implications for treatment and prevention strategies. *Nutrients*, 15(19), 4226.
- [4] Farika, S. A., *et al.* (2024). Promosi Kesehatan tentang Pentingnya Menjaga Kesehatan Mental pada Remaja. *Jurnal Pengabdian Dan Pemberdayaan Kesehatan*, 1(1), 69–77.
- [5] Wilanda, A., *et al.* (2024). Pengaruh Gaya Hidup Sehat, Kualitas Tidur, dan Pola Makan terhadap Tingkat Kesehatan Mental pada Pekerja Kantoran di Jawa Barat. *Jurnal Multidisiplin West Science*, 3(01), 69–77.
- [6] Martia, E., & Salman, S. (2022). Pengaruh gaya hidup sehat terhadap kesehatan mental remaja. *Syntax Literate; Jurnal Ilmiah Indonesia*, 7(11), 16126–16132.
- [7] Sheldon, E., *et al.* (2021). Prevalence and risk factors for mental health problems in university undergraduate students: A systematic review with meta-analysis. *Journal of Affective Disorders*, 287, 282–292.
- [8] Khoirunissa, D. H., & Sukartini, N. M. (2020). Kesehatan Mental Sumber Daya Manusia Indonesia. *Jurnal Sains Sosio Humaniora*, 4(1), 241–258.
- [9] Walburg, F. S., *et al.* (2023). Effectiveness of a lifestyle intervention for people with a severe mental illness in Dutch outpatient mental health care: a randomized clinical trial. *JAMA Psychiatry*, 80(9), 886–894.
- [10] Kaggle (2024). Mental Health and Lifestyle Habits 2019-2024. <https://www.kaggle.com/datasets/atharvasoundankar/mental-health-and-lifestyle-habits-2019-2024/data>
- [11] Daza, A., Saboya, N., NecocheaChamorro, J. I., Zavaleta Ramos, K., & Vásquez Valencia, Y. D. R. (2023). Systematic review of machine learning techniques to predict anxiety and stress in college students. *Informatics in Medicine Unlocked*, 43, Article 101391. doi:10.1016/j.imu.2023.101391
- [12] Tabares Tabares, M., Vélez Álvarez, C., Bernal Salcedo, J., & Murillo Rendón, S. (2024). Anxiety in young people: Analysis from a machine learning model. *Acta Psychologica (Amsterdam)*, 248, 104410. doi:10.1016/j.actpsy.2024.104410
- [13] Abd-Alrazaq, A., *et al.* (2023). Wearable Artificial Intelligence for Detecting Anxiety: Systematic Review and Meta-Analysis. *J Med Internet Res*, 25, e48754. doi:10.2196/48754
- [14] Zhang, Y., *et al.* (2025). An Explainable Anomaly Detection Framework for Monitoring Depression and Anxiety Using Consumer Wearable Devices. *arXiv*. arXiv:2505.03039

- [15] Moon, Y., & Woo, H. (2024). Key risk factors of generalized anxiety disorder in adolescents: machine learning study. *Frontiers in Public Health*, 12, Article 1504739. doi:10.3389/fpubh.2024.1504739
- [16] Mardini, M. T., et al. (2025). Identifying Adolescent Depression and Anxiety Through Real-World Data and Social Determinants of Health. *JMIR Mental Health*, 12, e66665. doi:10.2196/66665 *JMIR Mental Health*, 12, e66665.