

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

1	Pendahuluan	3
1.1	Latar Belakang	3
1.2	Rumusan Masalah	3
1.3	Tujuan Penelitian	4
1.4	Batasan Masalah	4
2	Tinjauan Pustaka	5
3	Metode dan Model	6
4	Metode Evaluasi	7
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
5	Hasil Evaluasi Model	11
5.1	Analisis Kuantitatif	21
5.2	Pembahasan Hasil	22
6	Kesimpulan	23

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 Skor F1 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. **Pembagian data:** Dataset dibagi dengan rasio 80% untuk *training* dan 20% untuk *testing*.

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

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

{DecisionTree, RandomForest, KNN, ..., LightGBM}.

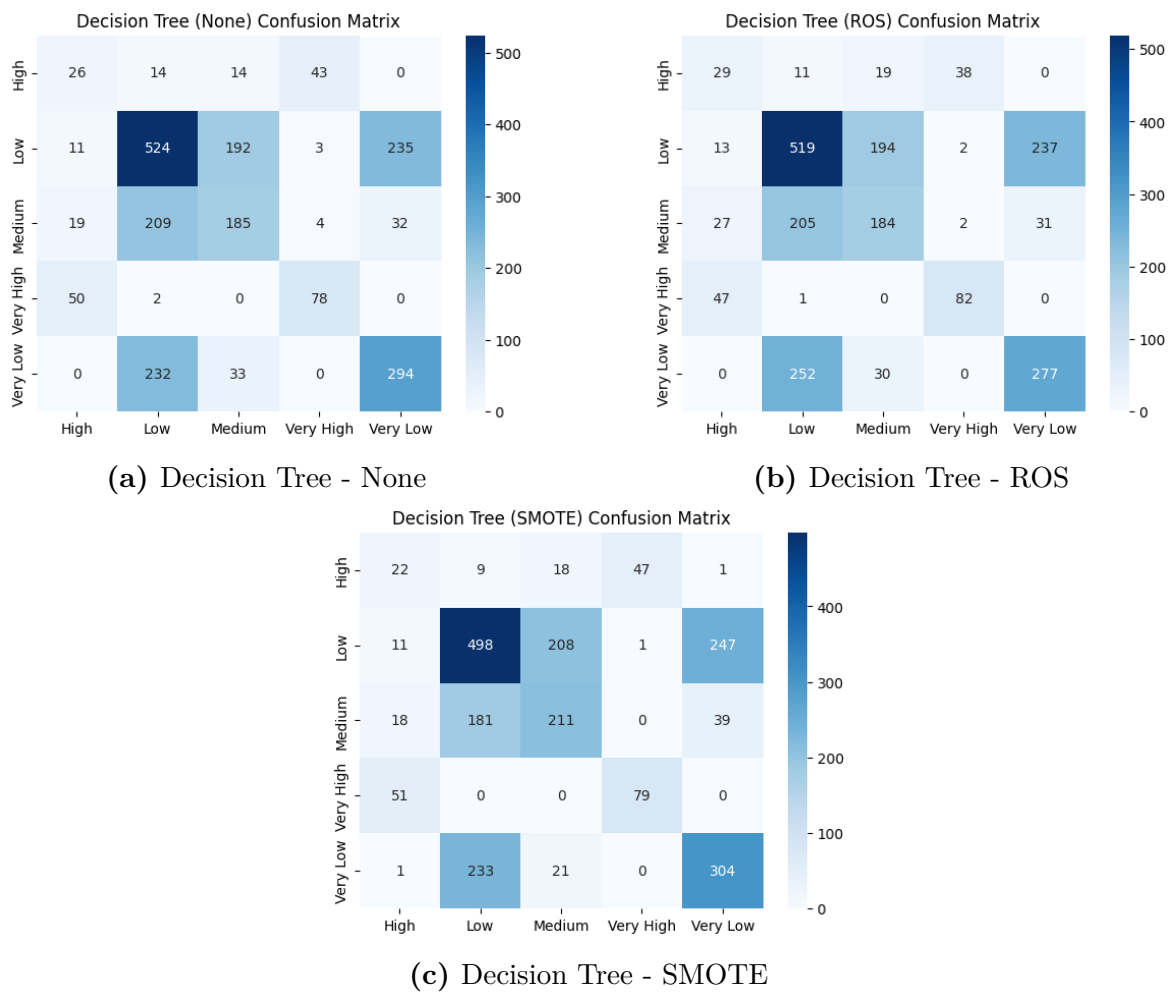
5. **Evaluasi:** Jalankan prediksi pada test set, hitung metrik (Rumus 3–6), dan simpan:

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

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

5 Hasil Evaluasi Model

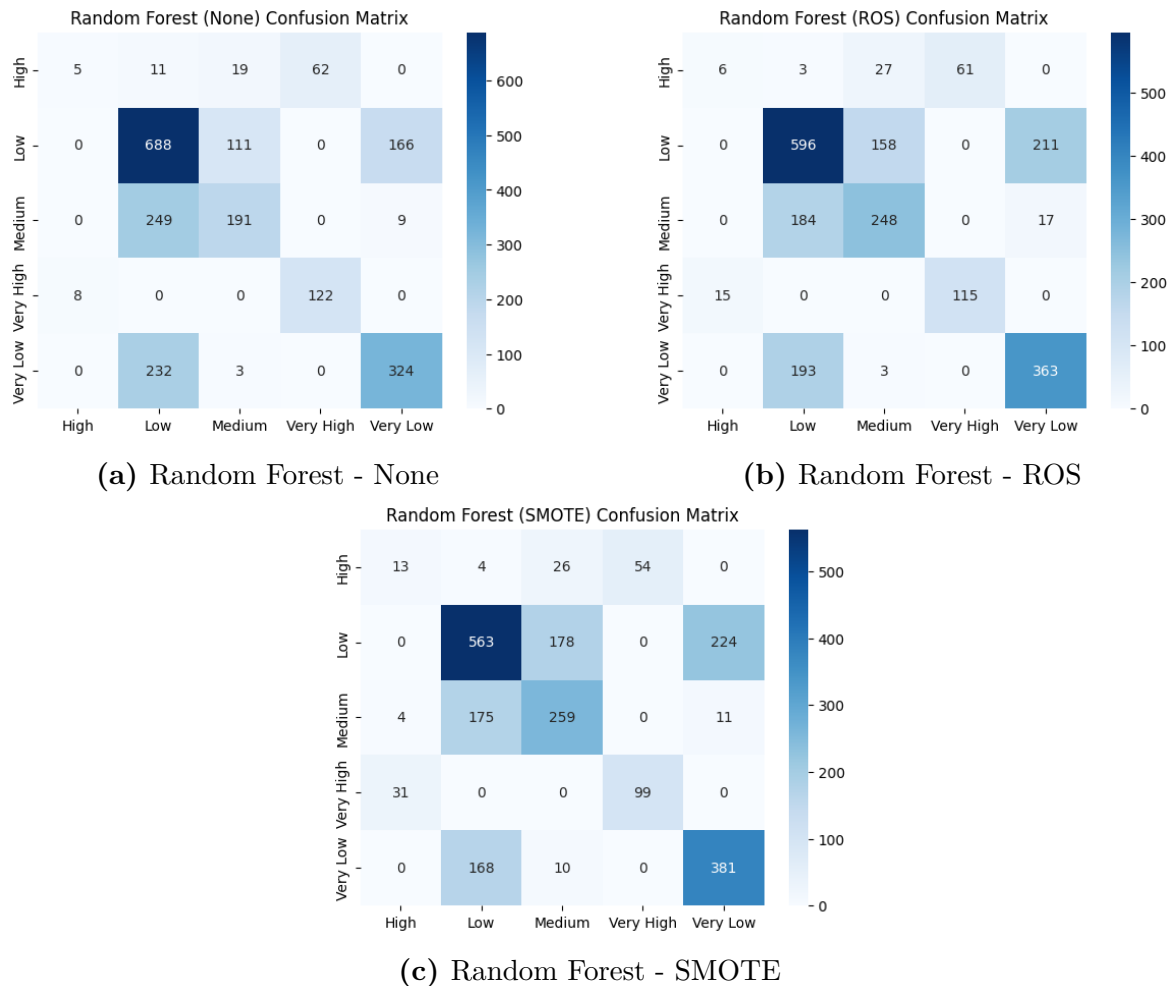
Hasil evaluasi performa dari sepuluh model klasifikasi: Decision Tree, Random Forest, KNN, Logistic Regression, Gaussian Naive Bayes, SVM, ANN, CatBoost, XGBoost, dan LightGBM dilakukan dalam tiga skenario penanganan ketidakseimbangan kelas: tanpa penyeimbangan data, dengan teknik *Synthetic Minority 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 Decision Tree tunggal pada data aslinya (GAMBAR 1.a) menunjukkan performa yang paling tidak stabil, dengan tingkat kebingungan yang tinggi dan tersebar di antara beberapa kelas. Tidak hanya salah mengklasifikasikan kelas "Medium" (209 kasus) dan "Very Low" (232 kasus) sebagai "Low", model ini juga keliru memprediksi 235 sampel "Low" sebagai "Very Low". Berbeda dengan model-model lainnya, implementasi teknik balancing baik ROS (GAMBAR 1.b) maupun SMOTE (GAMBAR 1.c) secara umum gagal memberikan perbaikan yang berarti. ROS hampir tidak

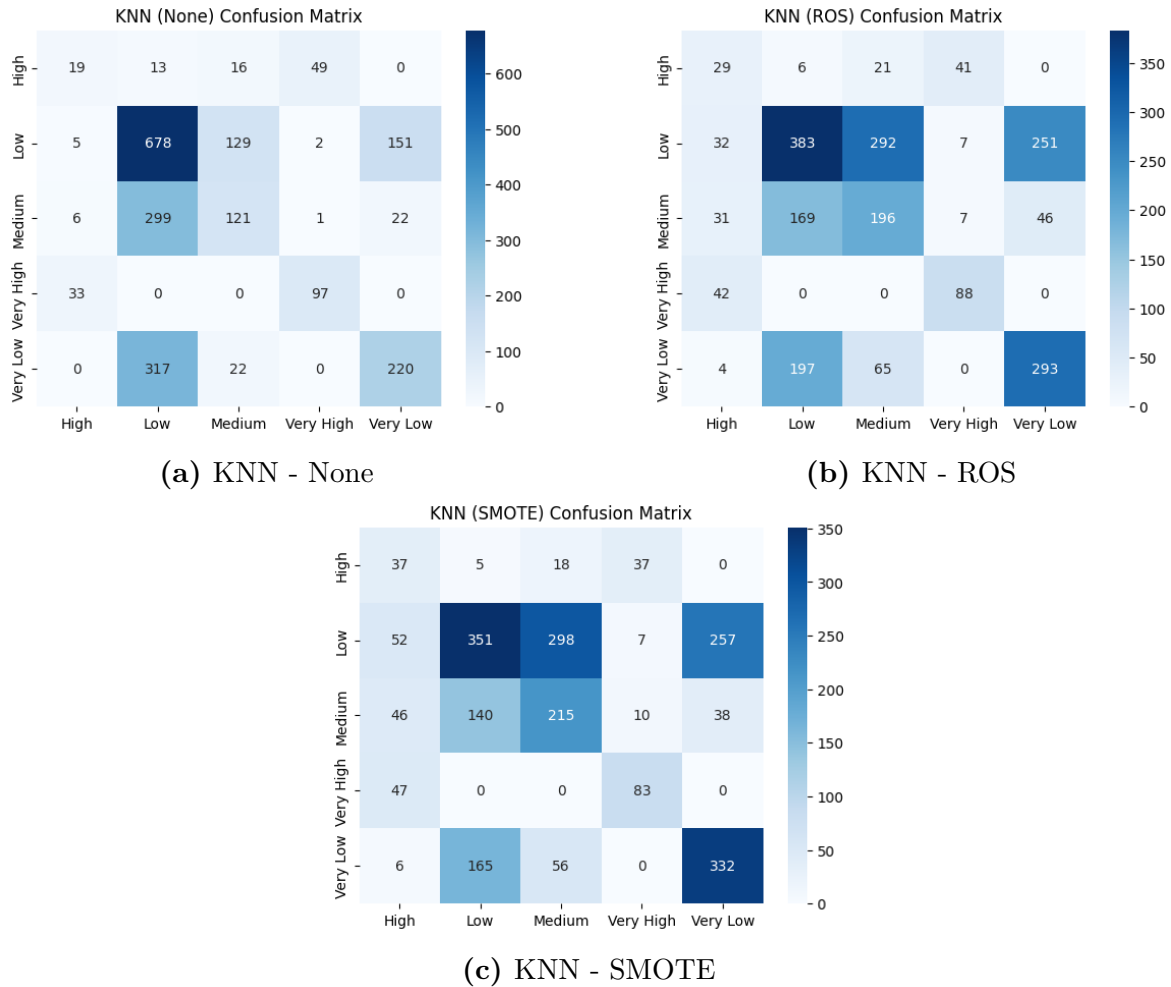
mengubah matriks kebingungan, sementara SMOTE memberikan sedikit peningkatan pada prediksi benar kelas "Medium" (211), namun dengan mengorbankan peningkatan kebingungan antara kelas "Low" dan "Very Low". Kegagalan ini mengindikasikan bahwa untuk model yang rentan terhadap overfitting seperti Decision Tree tunggal, teknik data balancing saja tidak cukup untuk membentuk batas keputusan yang baik.



GAMBAR 2. Confusion matrix untuk Random Forest.

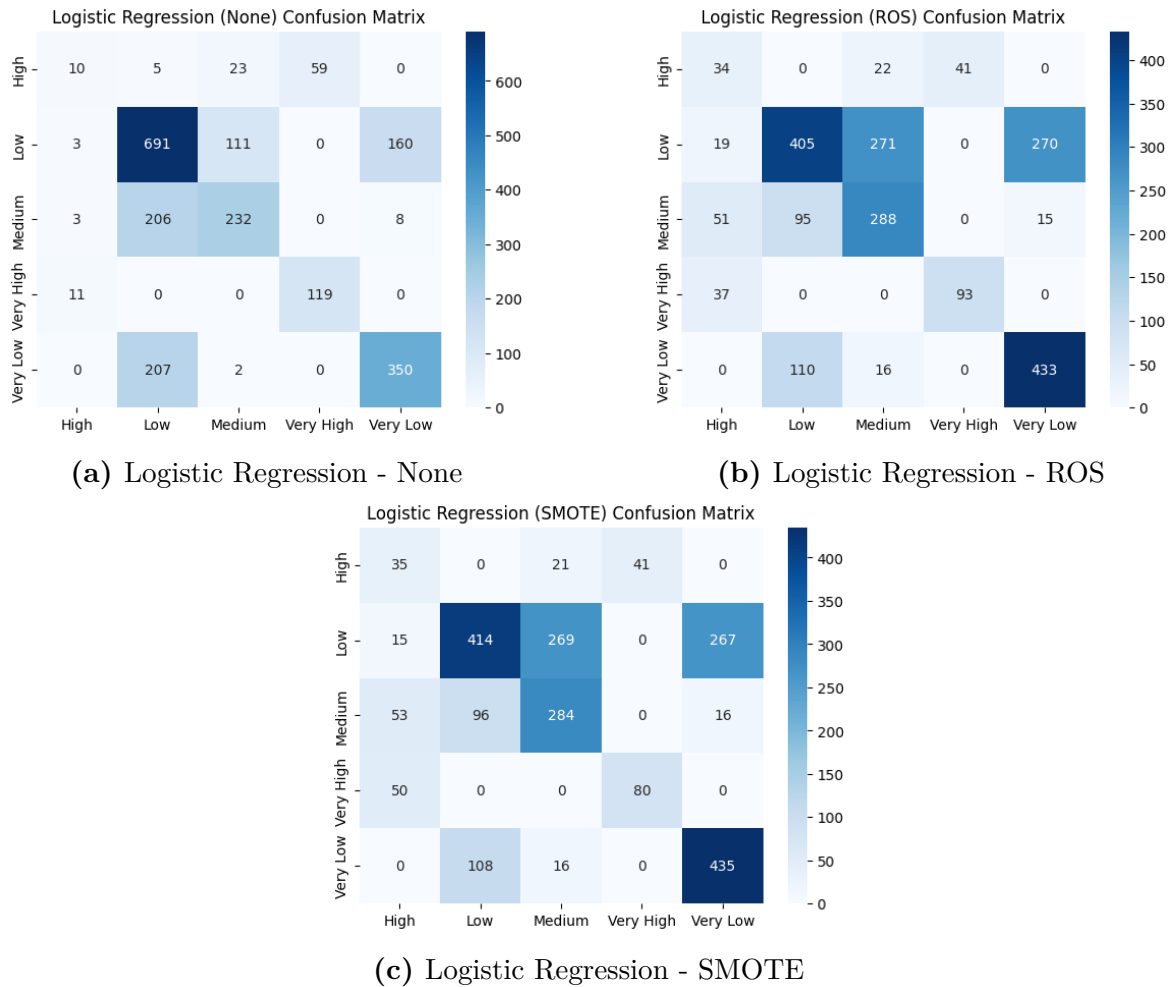
Model Random Forest pada kondisi data tidak seimbang (GAMBAR 2.a) menunjukkan kelemahan yang umum dijumpai, dengan bias kuat terhadap kelas "Low". Hal ini terlihat dari 249 kasus "Medium" dan 232 kasus "Very Low" yang salah diklasifikasikan sebagai "Low", serta performa yang sangat minim pada kelas "High". Penerapan teknik balancing, baik ROS (GAMBAR 2.b) maupun SMOTE (GAMBAR 2.c), berhasil memperbaiki masalah ini dengan menunjukkan pertukaran kinerja yang efektif. Di antara keduanya, SMOTE terlihat sedikit lebih unggul dalam membentuk model yang lebih tergeneralisasi. SMOTE lebih berhasil menekan misklasifikasi "Medium" menjadi "Low" (175 kasus) dan meningkatkan prediksi benar untuk kelas "Medium" (259) serta "Very Low" (381), melampaui hasil dari ROS. Performa superior dari SMOTE ini datang dengan trade-off yang wajar, yaitu penurunan akurasi pada kelas "Low"

yang sedikit lebih besar (563 benar) dibandingkan ROS (596 benar), mengindikasikan bahwa pembuatan sampel sintetis oleh SMOTE paling efektif untuk model ini.



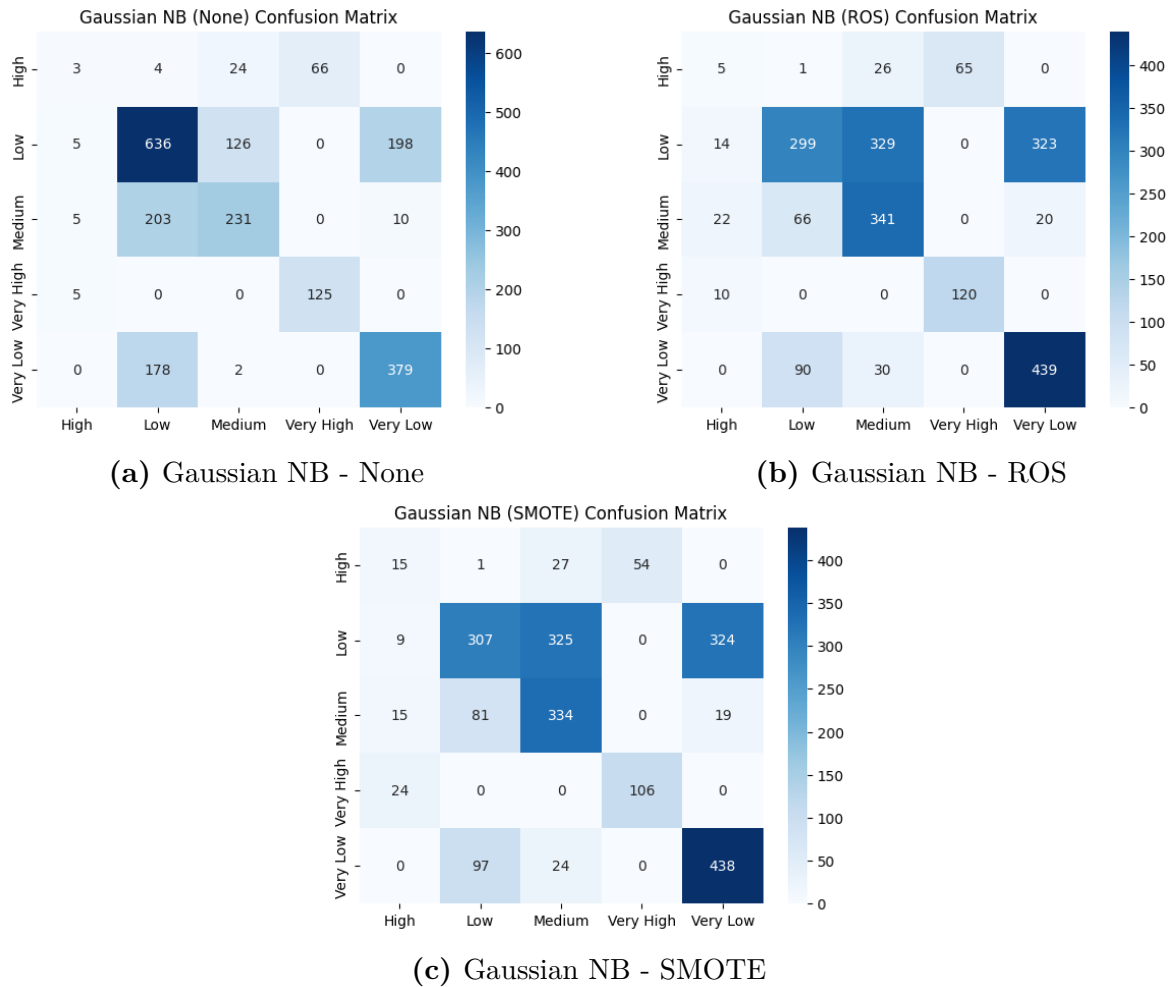
GAMBAR 3. Confusion matrix untuk KNN.

Model K-Nearest Neighbors (KNN) pada data aslinya (GAMBAR 3.a) menunjukkan salah satu bias terkuat, dengan performa yang sangat buruk pada kelas "Medium" dan "Very Low". Hal ini dibuktikan oleh jumlah misklasifikasi yang ekstrem ke kelas "Low", di mana 299 sampel "Medium" dan 317 sampel "Very Low" salah diprediksi. Implementasi teknik balancing, baik ROS (GAMBAR 3.b) maupun SMOTE (GAMBAR 3.c), kembali menunjukkan fenomena "koreksi berlebih" di mana masalah awal teratasi namun timbul kebingungan baru. Di antara keduanya, SMOTE lebih agresif dalam meningkatkan performa kelas minoritas, mencapai prediksi benar yang lebih tinggi untuk "Medium" (215) dan "Very Low" (332) dibandingkan ROS (196 dan 293). Akan tetapi, trade-off-nya adalah penurunan performa yang lebih tajam pada kelas "Low", dengan hanya 351 prediksi benar, lebih rendah dari ROS (383). Ini mengindikasikan bahwa untuk KNN, SMOTE menciptakan batas keputusan yang lebih radikal untuk menolong kelas minoritas, namun dengan ongkos kebingungan yang lebih besar pada kelas mayoritas.



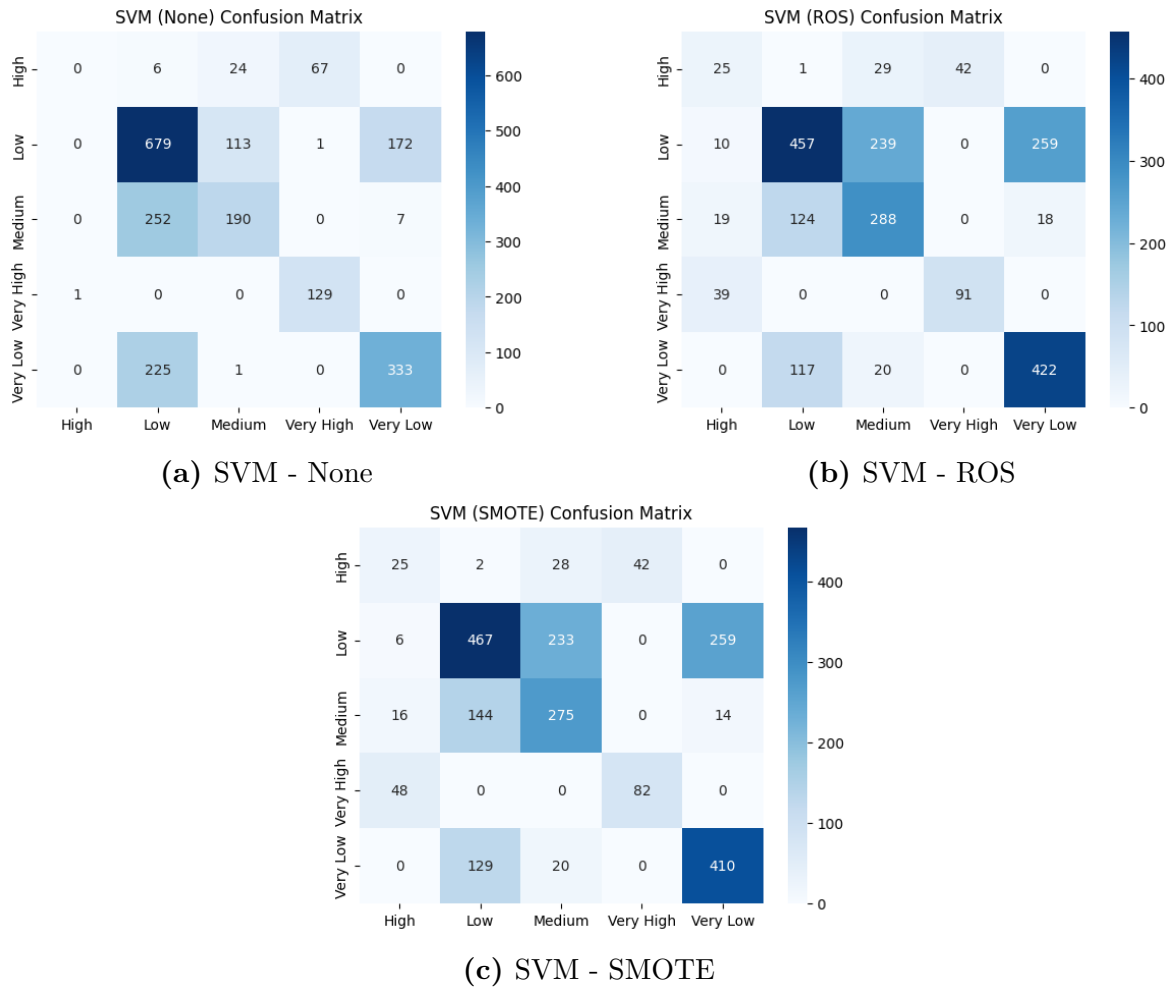
GAMBAR 4. Confusion matrix untuk Logistic Regression.

Model Logistic Regression pada data aslinya (GAMBAR 4.a) menampilkan pola kelemahan yang umum, yaitu bias yang kuat terhadap kelas "Low". Hal ini mengakibatkan tingginya jumlah misklasifikasi dari kelas "Medium" (206 kasus) dan "Very Low" (207 kasus) yang salah diprediksi sebagai "Low". Penerapan teknik balancing ROS (GAMBAR 4.b) dan SMOTE (GAMBAR 4.c) juga menunjukkan fenomena "koreksi berlebih", di mana keduanya menghasilkan pola yang nyaris identik. Meskipun berhasil memperbaiki masalah awal dan meningkatkan performa pada kelas minoritas seperti "High" (dari 10 menjadi sekitar 35 prediksi benar), keduanya menciptakan kebingungan baru. Kelas "Low" kini banyak disalahklasifikasikan sebagai "Medium" (kurang lebih 270 kasus) dan "Very Low" (sekitar 270 kasus), yang membuat akurasi pada kelas "Low" menurun drastis dari 691 menjadi sekitar 410. Hal ini menegaskan bahwa untuk model ini, teknik balancing hanya memindahkan titik kebingungan utama, bukan menyelesaikannya secara tuntas.



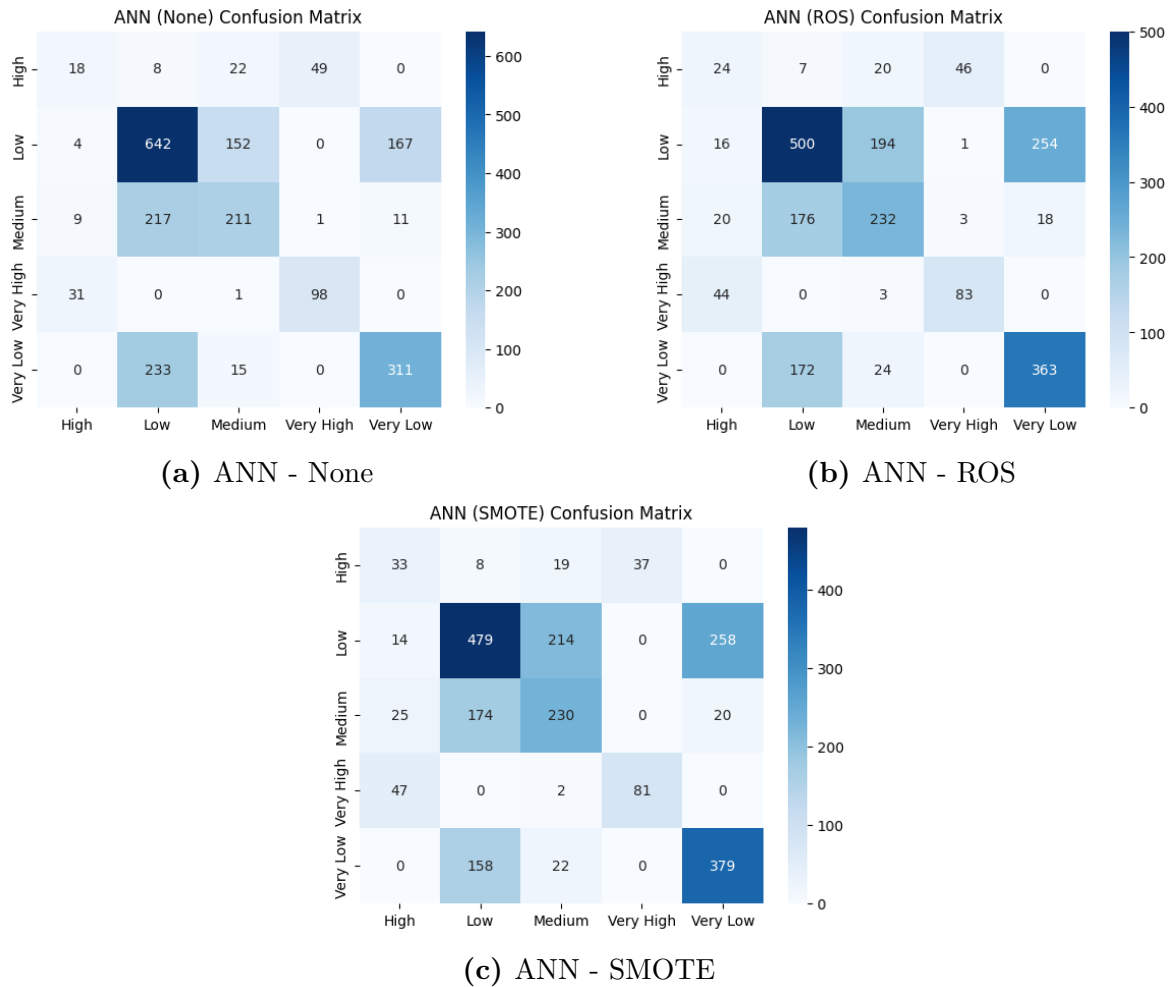
GAMBAR 5. Confusion matrix untuk Gaussian Naive Bayes.

Model Gaussian Naive Bayes (GNB) pada data aslinya (GAMBAR 5.a) menunjukkan kelemahan klasik berupa bias terhadap kelas "Low", di mana kelas "Medium" (203 kasus) dan "Very Low" (178 kasus) seringkali salah diklasifikasikan sebagai "Low". Uniknya, implementasi teknik balancing baik ROS (GAMBAR 5.b) maupun SMOTE (GAMBAR 5.c) justru memperkenalkan masalah baru yang signifikan. Meskipun keduanya berhasil mengurangi misklasifikasi awal ke kelas "Low", terjadi sebuah "koreksi berlebih" yang parah. Model yang telah diseimbangkan kini secara masif salah mengklasifikasikan sampel asli "Low" sebagai "Medium" (329 kasus pada ROS, 325 pada SMOTE) dan "Very Low" (323 kasus pada ROS, 324 pada SMOTE). Hal ini menyebabkan performa pada kelas "Low" anjlok drastis dari 636 menjadi sekitar 300 prediksi benar, yang menunjukkan bahwa untuk GNB, teknik balancing hanya menggeser letak kesalahan klasifikasi secara ekstrem alih-alih memperbaikinya.



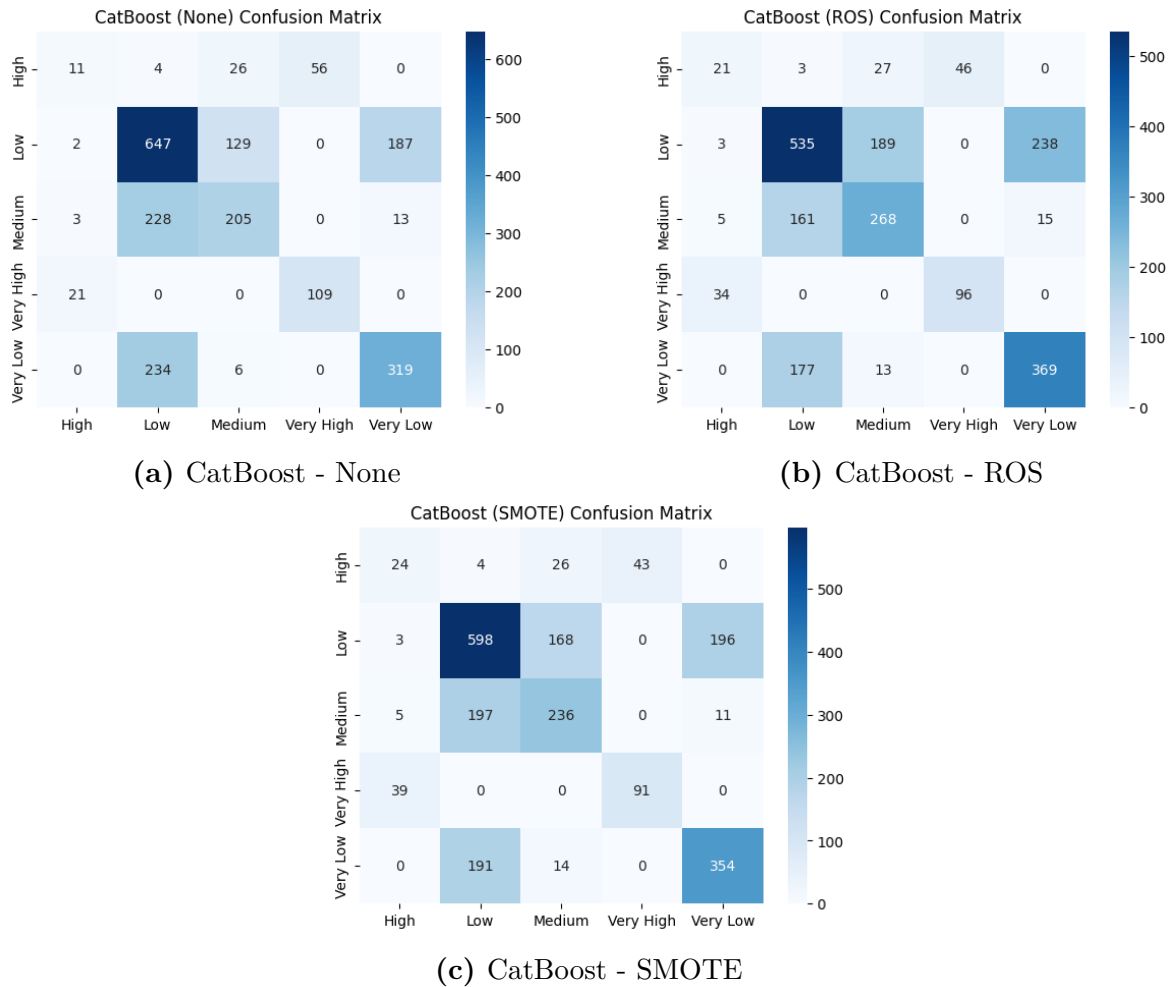
GAMBAR 6. Confusion matrix untuk Support Vector Machine.

Model Support Vector Machine (SVM) pada data aslinya (GAMBAR 6.a) menunjukkan kelemahan paling signifikan, dengan kegagalan total untuk mengidentifikasi kelas "High" (0 prediksi benar) dan bias yang ekstrem terhadap kelas "Low". Hal ini terlihat dari besarnya misklasifikasi pada kelas "Medium" (252 kasus) dan "Very Low" (225 kasus) yang keliru diprediksi sebagai "Low". Implementasi teknik balancing, baik ROS (GAMBAR 6.b) maupun SMOTE (GAMBAR 6.c), memberikan perbaikan yang sangat dramatis dan menghasilkan performa yang nyaris identik. Kedua teknik berhasil membuat model mampu mengenali kelas "High" (25 prediksi benar) dan secara signifikan meningkatkan akurasi untuk kelas "Medium" dan "Very Low". Perbedaan utama terletak pada detail kecil: ROS sedikit lebih unggul dalam mengurangi misklasifikasi "Medium" menjadi "Low" (124 kasus) dibandingkan SMOTE (144 kasus), sementara SMOTE mempertahankan performa yang sedikit lebih baik pada kelas mayoritas "Low" (467 prediksi benar) dibandingkan ROS (457 prediksi benar). Hal ini menunjukkan bahwa untuk SVM, kedua metode balancing sama-sama efektif dalam mengatasi bias data yang parah.



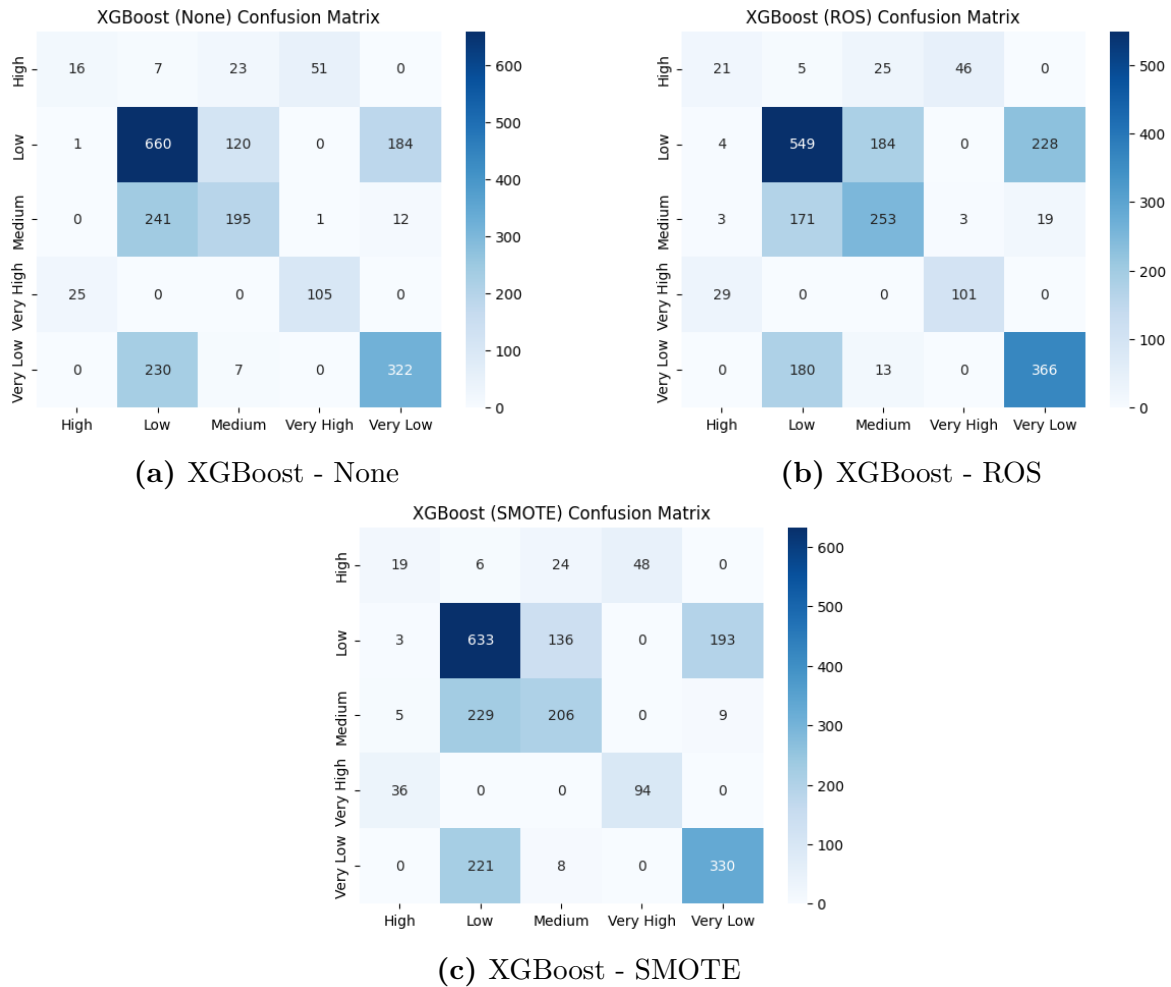
GAMBAR 7. Confusion matrix untuk Artificial Neural Network.

Model Artificial Neural Network (ANN) menunjukkan pola kinerja yang serupa dengan model lainnya saat diuji pada data tidak seimbang. Pada kondisi awal (GAMBAR 7.a), model ANN menampilkan bias yang jelas terhadap kelas "Low", yang dibuktikan oleh tingginya jumlah misklasifikasi dari kelas "Medium" (217 kasus) dan "Very Low" (233 kasus) yang salah diprediksi sebagai "Low". Penggunaan teknik data balancing, yaitu ROS (GAMBAR 7.b) dan SMOTE (GAMBAR 7.c), berhasil meningkatkan performa model secara signifikan, terutama dalam mengenali kelas-kelas minoritas. Secara rinci, SMOTE menunjukkan keunggulan dalam meningkatkan prediksi benar untuk kelas "High" (33) dan "Very Low" (379), sedikit melampaui ROS (24 dan 363). Namun, peningkatan ini datang dengan konsekuensi yang lebih besar, di mana performa SMOTE pada kelas "Low" (479 prediksi benar) menjadi yang terendah di antara ketiganya. Hal ini mengindikasikan bahwa untuk model ANN ini, SMOTE lebih agresif dalam membantu model mempelajari kelas minoritas, tetapi trade-off-nya adalah penurunan akurasi yang lebih nyata pada kelas mayoritas dibandingkan dengan pendekatan ROS.



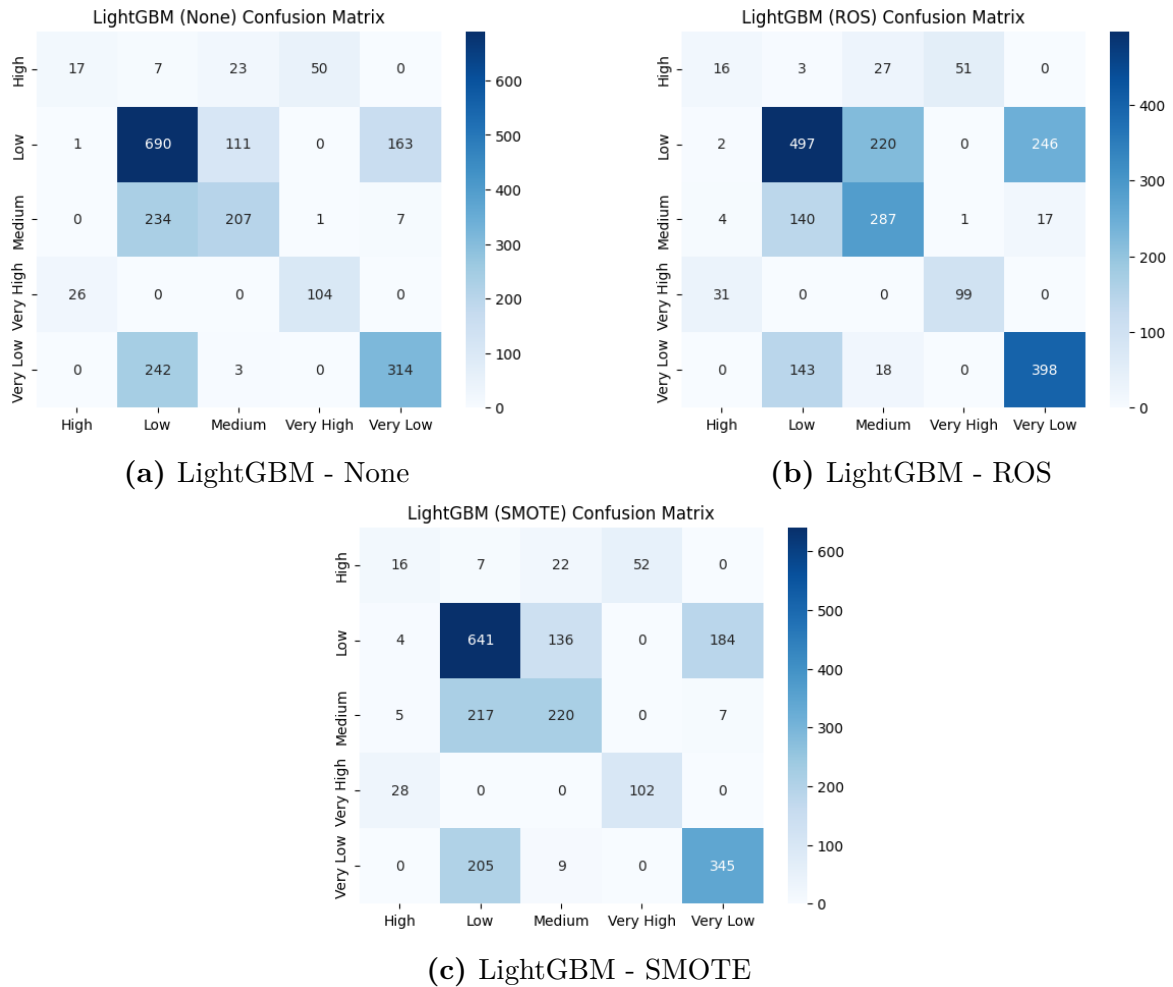
GAMBAR 8. Confusion matrix untuk CatBoost.

Meskipun dikenal sebagai algoritma boosting yang tangguh, CatBoost tetap menunjukkan kerentanan klasik terhadap data tidak seimbang (GAMBAR 4.a). Model awalnya menampilkan bias yang familier, di mana performanya terpusat pada kelas "Low" sehingga menyebabkan kelas "Medium" (228 kasus) dan "Very Low" (234 kasus) seringkali terserap dan salah diklasifikasikan. Intervensi menggunakan teknik balancing memberikan hasil yang kontras, di mana ROS (GAMBAR 4.b) dan SMOTE (GAMBAR 4.c) tampak mengambil filosofi perbaikan yang berbeda. ROS secara tegas menukar akurasi kelas "Low" untuk mendapatkan peningkatan signifikan pada kelas lain, terbukti dari naiknya prediksi benar untuk "Medium" (268) dan "Very Low" (369). Sebaliknya, SMOTE mengambil pendekatan yang lebih konservatif; peningkatannya pada kelas minoritas tidak seagresif ROS, namun ia berhasil mempertahankan kinerja yang lebih superior pada kelas "Low" (598) dibandingkan ROS (535). Pilihan di antara keduanya menjadi jelas: ROS lebih sesuai jika tujuannya adalah memaksimalkan pengenalan pada kelas minoritas, sementara SMOTE menjadi opsi yang lebih aman untuk memperbaiki model tanpa mengorbankan kinerja pada kelas mayoritas secara drastis.



GAMBAR 9. Confusion matrix untuk XGBoost.

Model XGBoost pada kondisi data tidak seimbang (GAMBAR 9.a) menunjukkan kelemahan umum dengan adanya bias kuat terhadap kelas "Low". Ini dibuktikan dari tingginya misklasifikasi kelas "Medium" (241 kasus) dan kelas "Very Low" (230 kasus) yang salah diprediksi sebagai "Low". Implementasi teknik balancing memberikan hasil yang sangat berbeda antara ROS (GAMBAR 9.b) dan SMOTE (GAMBAR 9.c), di mana hanya ROS yang berhasil menunjukkan pertukaran kinerja yang efektif. ROS secara signifikan lebih unggul, berhasil menekan misklasifikasi "Medium" menjadi "Low" hingga hanya 171 kasus. Sebaliknya, SMOTE hampir tidak memberikan perbaikan, dengan misklasifikasi yang sama masih berada di angka 229 kasus. Keberhasilan ROS ini datang dengan trade-off yang wajar, yaitu penurunan akurasi pada kelas "Low" menjadi 549 prediksi benar, sementara model SMOTE mempertahankan akurasi tinggi pada kelas "Low" (633 benar) justru karena ia gagal menyeimbangkan model secara efektif.



GAMBAR 10. Confusion matrix untuk LightGBM.

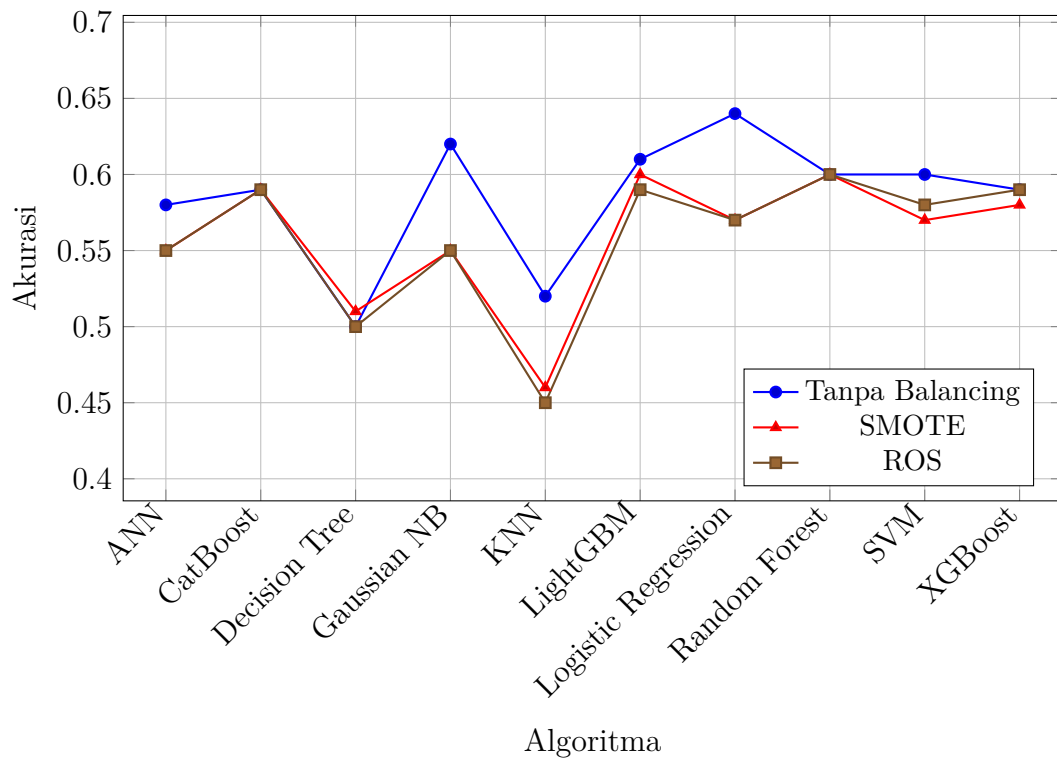
Model LightGBM, sebagai algoritma boosting yang efisien, menunjukkan perubahan performa yang signifikan ketika dihadapkan pada data yang telah diseimbangkan. Pada pengujian dengan data asli tanpa balancing (GAMBAR 10.a), model menunjukkan kelemahan berupa bias yang kuat terhadap kelas "Low", terbukti dari banyaknya sampel "Medium" (234) dan "Very Low" (242) yang salah diklasifikasikan. Implementasi teknik balancing, baik ROS (GAMBAR 10.b) maupun SMOTE (GAMBAR 10.c), secara substansial memperbaiki kapabilitas model tersebut. Secara spesifik, ROS secara drastis mengurangi kesalahan klasifikasi ini, namun dengan konsekuensi menurunnya performa pada kelas "Low" itu sendiri (497 prediksi benar). Di sisi lain, SMOTE menunjukkan sebuah pertukaran kinerja yang menghasilkan model lebih seimbang; meski perbaikannya tidak seagresif ROS, ia mampu mempertahankan akurasi yang jauh lebih tinggi pada kelas "Low" (641 prediksi benar), yang mengindikasikan pembentukan batas keputusan yang lebih generalis dan tidak terlalu mengorbankan kelas mayoritas.

5.1 Analisis Kuantitatif

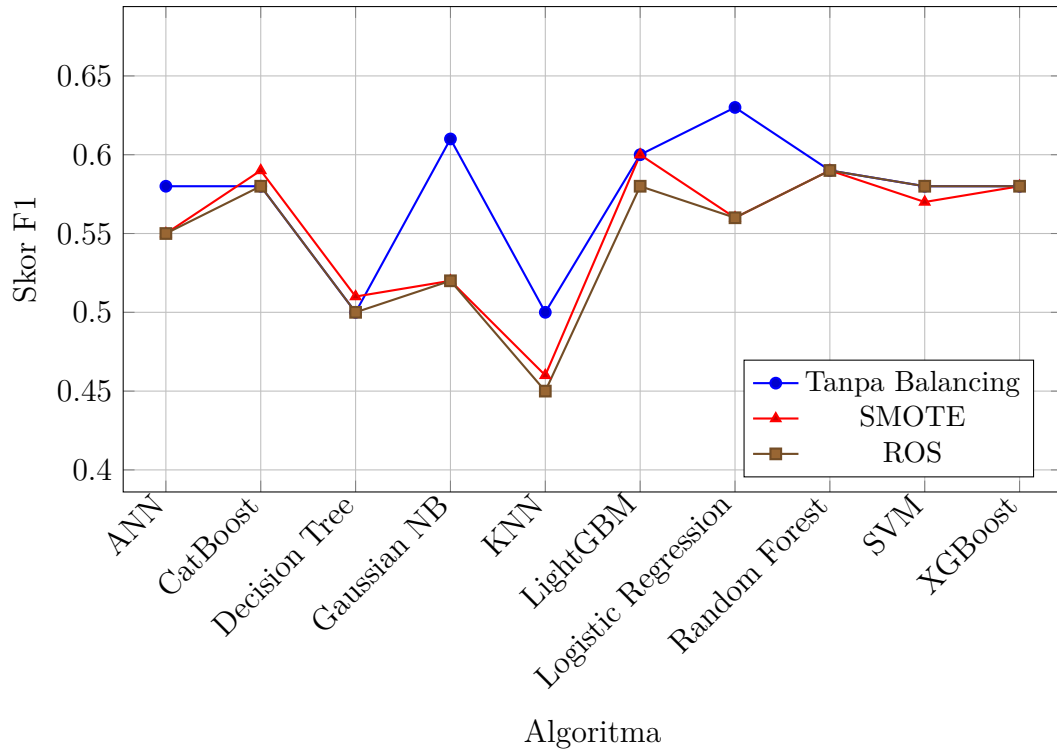
Untuk melengkapi analisis kualitatif dari confusion matrix, evaluasi kuantitatif dilakukan dengan membandingkan metrik performa utama di seluruh model dan teknik penyeimbangan. Tabel 1 merepresentasikan nilai Accuracy dan Macro F1-Score untuk 30 kombinasi eksperimen. Selanjutnya, Gambar 11 dan Gambar 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
ANN	0.58	0.58	0.55	0.55	0.55	0.55
CatBoost	0.59	0.58	0.59	0.59	0.59	0.58
Decision Tree	0.50	0.50	0.51	0.51	0.50	0.50
Gaussian NB	0.62	0.61	0.55	0.52	0.55	0.52
KNN	0.52	0.50	0.46	0.46	0.45	0.45
LightGBM	0.61	0.60	0.60	0.60	0.59	0.58
Logistic Regression	0.64	0.63	0.57	0.56	0.57	0.56
Random Forest	0.60	0.59	0.60	0.59	0.60	0.59
SVM	0.60	0.58	0.57	0.57	0.58	0.58
XGBoost	0.59	0.58	0.58	0.58	0.59	0.58



GAMBAR 11. Perbandingan akurasi model terhadap teknik balancing



GAMBAR 12. Perbandingan Skor F1 model terhadap teknik balancing

5.2 Pembahasan Hasil

Dari data yang disajikan pada Tabel 1 dan divisualisasikan pada Gambar 11 dan 12, dapat ditarik beberapa kesimpulan penting yang berlawanan dengan ekspektasi umum. Pertama, intervensi penyeimbangan data dengan teknik oversampling (SMOTE dan ROS) secara konsisten memberikan dampak negatif atau tidak signifikan pada hampir semua model yang diuji. Model-model seperti Logistic Regression, Gaussian NB, ANN, dan KNN mengalami penurunan performa yang cukup jelas baik pada metrik akurasi maupun Macro F1-Score setelah data diseimbangkan. Hal ini mengindikasikan bahwa untuk dataset ini, masalah utama bukanlah ketidakseimbangan kelas, melainkan karakteristik data atau keterbatasan fundamental dari beberapa model itu sendiri.

Kedua, model-model dengan performa terbaik secara konsisten ditemukan pada kondisi data asli tanpa penyeimbangan. Logistic Regression muncul sebagai model paling unggul secara keseluruhan, mencapai akurasi tertinggi (0.64) dan Macro F1-Score tertinggi (0.63) pada data asli. Diikuti oleh Gaussian NB (Akurasi 0.62, F1 0.61) dan beberapa model ensemble seperti LightGBM dan Random Forest. Fakta bahwa tidak ada satupun konfigurasi model dengan oversampling yang mampu melampaui performa model pada data asli menegaskan kembali bahwa penambahan sampel, baik sintesis maupun duplikasi, justru memperkenalkan noise yang mengganggu proses pembelajaran model.

Ketiga, saat membandingkan kedua teknik oversampling, tidak ada yang menunjukkan keunggulan jelas karena keduanya sama-sama menurunkan performa pada model-

model unggulan. Analisis pada confusion matrix di tahap sebelumnya menunjukkan bahwa teknik oversampling seringkali menyebabkan "koreksi berlebih", di mana model menjadi bingung dan salah mengklasifikasikan kelas mayoritas yang sebelumnya dapat diprediksi dengan baik. Penurunan skor metrik ini adalah bukti kuantitatif dari fenomena tersebut. Meskipun oversampling mungkin memperbaiki beberapa jenis misklasifikasi spesifik, perbaikan tersebut dibayar dengan penurunan kemampuan generalisasi model secara keseluruhan pada data uji.

Model-model ensemble seperti CatBoost, Random Forest, dan XGBoost menunjukkan resiliensi yang lebih baik terhadap oversampling, dengan penurunan performa yang tidak sedrastis model lainnya. Namun, mereka tetap tidak mendapatkan manfaat positif. Hal ini menyoroti bahwa arsitektur ensemble yang kuat sekalipun tidak dapat mengekstrak keuntungan dari data yang telah ditambah dengan sampel artifisial atau duplikat pada kasus ini. Kesimpulannya, untuk dataset ini, strategi yang lebih menjanjikan adalah melakukan pemilihan model dan tuning parameter pada data asli yang tidak seimbang, karena intervensi oversampling terbukti kontra-produktif.

6 Kesimpulan

Penelitian ini membandingkan sepuluh algoritma pembelajaran mesin untuk deteksi dini tingkat kecemasan berbasis gaya hidup, dengan investigasi khusus pada dampak teknik penyeimbangan kelas. Hasil menunjukkan bahwa, berlawanan dengan ekspektasi umum, performa model pada data asli tanpa balancing justru menjadi tolok ukur tertinggi. Model-model seperti Logistic Regression dan Gaussian NB menunjukkan akurasi dan Macro F1-Score yang solid dan menjadi yang teratas, mengindikasikan bahwa data asli sudah memiliki struktur yang cukup informatif untuk dipelajari.

Setelah penerapan teknik balancing, ditemukan bahwa metode oversampling (ROS dan SMOTE) secara konsisten terbukti kontra-produktif. Hampir seluruh model, termasuk kelompok ensemble pohon (Random Forest, CatBoost, XGBoost, LightGBM), mengalami penurunan performa setelah proses oversampling. Konfigurasi terbaik secara mutlak adalah model Logistic Regression pada data asli, yang mencapai akurasi 0.64 dan Macro F1-Score 0.63, melampaui semua kombinasi model yang datanya telah diseimbangkan.

Untuk penelitian selanjutnya, disarankan untuk mengalihkan fokus dari oversampling. Perlu dieksplorasi teknik penanganan ketidakseimbangan kelas yang berbeda, seperti metode undersampling (contohnya, Tomek Links atau NearMiss) atau pendekatan berbasis biaya (cost-sensitive learning). Selain itu, optimasi hyperparameter yang ekstensif pada model-model berkinerja terbaik di data asli, seperti Logistic Regression dan LightGBM, serta analisis kepentingan fitur secara mendalam, menjadi langkah yang lebih logis. Pendalaman ini diharapkan dapat memaksimalkan potensi informasi dari data asli dan membangun model yang lebih robust tanpa bergantung pada modifikasi data yang terbukti berisiko.

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.