

Klasifikasi Gangguan Tidur Menggunakan Algoritma XGBoost dengan SMOTE dan Grid Search

Classification of Sleep Disorders Using the XGBoost Algorithm with SMOTE and Grid Search

**Moh. Indra Kholid Khoirusshofi^{1*}, Fida Maisa Hana²,
Taftazani Ghazi Pratama³**

^{1,2,3} *Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kudus*

*corr-author: 42022110006@std.umku.ac.id

ABSTRAK

Gangguan tidur merupakan permasalahan kesehatan yang memiliki prevalensi tinggi dan berdampak signifikan terhadap kualitas hidup, sehingga diperlukan sistem deteksi dini yang akurat dan efisien. Perkembangan *machine learning* membuka peluang untuk membangun model klasifikasi gangguan tidur berbasis data, namun ketidakseimbangan kelas pada *dataset* medis sering menjadi tantangan yang menurunkan performa model, khususnya pada kelas minoritas. Penelitian ini bertujuan membangun model klasifikasi gangguan tidur menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) dengan penerapan *Synthetic Minority Oversampling Technique* (SMOTE) dan optimasi hiperparameter menggunakan *Grid Search*. *Dataset* yang digunakan adalah *Sleep Health and Lifestyle Dataset* yang terdiri dari 374 data dengan tiga kelas target, yaitu *Insomnia*, *No Disorder*, dan *Sleep Apnea*. Penelitian ini menguji empat skenario model, yaitu XGBoost tanpa SMOTE dan *Grid Search*, XGBoost dengan SMOTE, XGBoost dengan *Grid Search*, serta kombinasi SMOTE dan *Grid Search*. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, *Confusion Matrix*, serta *ROC Curve*. Hasil penelitian menunjukkan bahwa penerapan SMOTE meningkatkan sensitivitas model terhadap kelas minoritas, sedangkan optimasi hiperparameter menggunakan *Grid Search* meningkatkan stabilitas dan akurasi model secara keseluruhan. Kombinasi SMOTE dan *Grid Search* menghasilkan performa terbaik dengan akurasi mencapai 97% serta nilai *precision*, *recall*, dan *F1-score* yang seimbang pada seluruh kelas. Selain itu, evaluasi menggunakan *ROC Curve* menunjukkan nilai AUC pada rentang 0,99 hingga 1,00, yang mengindikasikan kemampuan model yang sangat baik dalam membedakan setiap kelas. Hasil ini menunjukkan bahwa pendekatan yang diusulkan mampu meningkatkan performa klasifikasi gangguan tidur dan berpotensi menjadi alternatif dalam pengembangan sistem deteksi dini gangguan tidur.

Kata-kata kunci: Klasifikasi, Gangguan Tidur, XGBoost, SMOTE, *Grid Search*

ABSTRACT

Sleep disorders are a health problem with a high prevalence and a significant impact on quality of life, highlighting the need for accurate and efficient early detection systems. The advancement of machine learning offers opportunities to develop data-driven models for sleep disorders classification; however, class imbalance in medical datasets remains a major challenge that can degrade model performance, particularly for minority classes.

This study aims to develop a sleep disorders classification model using the Extreme Gradient Boosting (XGBoost) algorithm by incorporating the Synthetic Minority Oversampling Technique (SMOTE) and hyperparameter optimization through Grid Search. The dataset used in this study is the Sleep Health and Lifestyle Dataset, consisting of 374 samples with three target classes: Insomnia, No Disorder, and Sleep Apnea. Four experimental scenarios were evaluated, namely XGBoost without SMOTE and Grid Search, XGBoost with SMOTE, XGBoost with Grid Search, and the combination of SMOTE and Grid Search. Model performance was evaluated using accuracy, precision, recall, F1-score, Confusion Matrix and ROC Curve metrics. The results indicate that SMOTE improves model sensitivity toward minority classes, while hyperparameter optimization using Grid Search enhances overall model stability and accuracy. The combination of SMOTE and Grid Search achieved the best performance, with an accuracy of 97% and balanced precision, recall, and F1-score across all classes. These results indicate that the proposed approach is able to improve sleep disorders classification performance and may serve as an alternative in the development of early detection systems for sleep disorders.

Keywords: Classification, Sleep Disorders, XGBoost, SMOTE, Grid Search

PENDAHULUAN

Gangguan tidur merupakan permasalahan kesehatan yang semakin mendapat perhatian secara global karena prevalensinya yang tinggi dan dampaknya yang signifikan terhadap kualitas hidup. Insomnia menjadi salah satu gangguan tidur yang paling umum, dengan lebih dari 850 juta orang dewasa di dunia atau sekitar 16,2% populasi global mengalami insomnia klinis (Benjafield *et al.*, 2025). Dalam penelitian Nurmiati Amir di Indonesia, prevalensi insomnia diperkirakan mencapai sekitar 10%, atau sekitar 28 juta dari total 238 juta penduduk (Bili *et al.*, 2025). Sementara itu studi pada remaja menunjukkan angka gangguan tidur di atas 35%, baik di wilayah urban maupun suburban (Tyas, 2022). Tingginya angka tersebut menunjukkan bahwa gangguan tidur merupakan masalah kesehatan masyarakat yang memerlukan perhatian serius.

Gangguan tidur tidak hanya berdampak pada kondisi fisik dan psikologis individu, tetapi juga menimbulkan konsekuensi ekonomi yang signifikan. Kurangnya kualitas dan durasi tidur berhubungan dengan penurunan performa kognitif, meningkatnya risiko kecelakaan kerja dan lalu lintas, serta menurunnya produktivitas kerja (Khan and Al-Jahdali, 2023). Selain itu, tidur yang tidak optimal meningkatkan risiko penyakit degeneratif seperti hipertensi, penyakit kardiovaskular, dan diabetes melitus (Wulansih *et al.*, 2024). Kondisi ini mendorong kebutuhan akan sistem deteksi dini gangguan tidur yang efektif, cepat, dan mudah diakses oleh masyarakat.

Metode diagnosis gangguan tidur secara konvensional umumnya dilakukan melalui pemeriksaan klinis seperti polisomnografi dan evaluasi oleh tenaga medis spesialis. Meskipun memiliki tingkat akurasi yang tinggi, metode tersebut memiliki keterbatasan dari sisi biaya, waktu, dan ketersediaan tenaga ahli, terutama di daerah dengan akses layanan kesehatan yang terbatas (Palombini *et al.*, 2024). Seiring dengan perkembangan teknologi kecerdasan buatan, pendekatan *machine learning* memiliki kemampuan dalam menganalisis data secara otomatis dan mengidentifikasi pola kompleks. Berbagai penelitian telah membuktikan efektivitas pendekatan ini di bidang kesehatan, seperti pengembangan sistem pendukung keputusan medis, pengembangan *chatbot* kesehatan mental dan prediksi penyakit kronis (Wahyudin *et al.*, 2023; Afrisia *et al.*, 2024; Haq *et al.*, 2025).

Dalam konteks klasifikasi gangguan tidur, penelitian sebelumnya telah membandingkan berbagai algoritma *machine learning* dan menunjukkan bahwa algoritma *boosting* memiliki performa yang kompetitif. Penelitian oleh (Mawardi *et al.*, 2025) menunjukkan bahwa XGBoost memiliki performa yang lebih rendah dibandingkan CatBoost, khususnya pada klasifikasi kelas *Sleep Apnea*. Temuan tersebut mengindikasikan adanya tantangan dalam mengenali kelas tertentu pada kasus gangguan tidur, yang dalam banyak studi sering dikaitkan dengan ketidakseimbangan kelas. Selain itu, penelitian tersebut belum melakukan penerapan teknik penanganan ketidakseimbangan kelas maupun optimasi hiperparameter secara eksplisit, sehingga masih terbuka peluang untuk meningkatkan performa model.

Merujuk pada studi yang dilakukan oleh (Anshory *et al.*, 2025) menunjukkan bahwa teknik penanganan ketidakseimbangan kelas dan optimasi hiperparameter dapat secara signifikan meningkatkan kinerja model klasifikasi gangguan tidur. Mereka mengembangkan model berbasis AdaBoost yang dikombinasikan dengan SMOTE dan Optuna pada *Sleep Health and Lifestyle Dataset*, dan berhasil mencapai akurasi 90,6% dengan *F1-score* yang tinggi untuk kelas *Insomnia* dan *Sleep Apnea*. Pendekatan tersebut memperlihatkan bahwa strategi penggabungan teknik *oversampling* dan optimasi hiperparameter mampu meningkatkan performa model, khususnya pada kelas minoritas, dan sekaligus memberikan indikasi bahwa masalah yang dialami XGBoost pada penelitian (Mawardi *et al.*, 2025) berpotensi diperbaiki melalui kombinasi metode serupa.

Ketidakseimbangan kelas merupakan tantangan umum dalam *dataset* medis dan dapat menurunkan sensitivitas model terhadap kelas dengan signifikansi klinis tinggi (Cartus *et al.*, 2023). *Synthetic Minority Oversampling Technique* (SMOTE) terbukti efektif dalam meningkatkan representasi kelas minoritas dan memperbaiki performa klasifikasi pada berbagai studi. Penerapan SMOTE pada pemantauan kualitas Industrial IoT *assembly* efektif meningkatkan kinerja klasifikasi pada data kualitas produksi yang sangat tidak seimbang (Han *et al.*, 2024). Dalam penelitian XGBoost dengan SMOTE, peneliti berhasil menghasilkan *F1-score* hingga 0,94 dan *ROC-AUC* 0,98 pada *dataset Spambase* (Siagian *et al.*, 2025). Sedangkan penelitian lain melaporkan bahwa penambahan SMOTE pada model XGBoost secara signifikan meningkatkan *recall* kelas sirosis yang semula di bawah 2% menjadi di atas 70% tanpa mengorbankan akurasi secara keseluruhan (Syukron *et al.*, 2020). Secara umum, temuan-temuan tersebut menegaskan bahwa SMOTE efektif meningkatkan sensitivitas model terhadap kelas minoritas pada berbagai domain dengan karakteristik data tidak seimbang.

Selain aspek keseimbangan data, performa XGBoost juga sangat dipengaruhi oleh konfigurasi hiperparameter, sehingga optimasi menggunakan *Grid Search* diperlukan untuk menghasilkan kinerja model yang lebih optimal. Beberapa penelitian sebelumnya menunjukkan bahwa penerapan *Grid Search* mampu meningkatkan performa model secara signifikan. Dalam penelitian (Sugihartono *et al.*, 2025) menunjukkan bahwa optimasi XGBoost dengan *GridSearchCV* dapat meningkatkan akurasi 0,9930 dan ROC AUC 0,9997 serta menurunkan false negative 1,2 kasus. Pada penelitian lain, *GridSearchCV* terbukti meningkatkan kinerja akurasi model SVM menjadikannya model terbaik dengan akurasi awal 88,46% menjadi 94,05% (Pratama *et al.*, 2026). Oleh karena itu, penelitian ini mengusulkan penerapan kombinasi SMOTE dan *Grid Search* pada algoritma XGBoost untuk meningkatkan performa klasifikasi gangguan tidur. Pendekatan yang diusulkan diharapkan mampu menghasilkan model klasifikasi yang lebih akurat, seimbang, dan *robust*, serta berkontribusi pada pengembangan sistem deteksi dini gangguan tidur berbasis *machine learning*.

Meskipun berbagai penelitian telah menerapkan algoritma XGBoost, teknik SMOTE, maupun optimasi hiperparameter seperti *Grid Search* secara terpisah maupun

terbatas dalam kombinasi tertentu, penelitian yang secara sistematis mengkaji kontribusi masing-masing pendekatan dalam konteks klasifikasi gangguan tidur masih terbatas. Kebaruan penelitian ini difokuskan pada evaluasi komparatif empat skenario model XGBoost dalam klasifikasi gangguan tidur, yaitu tanpa perlakuan khusus, dengan SMOTE, dengan *Grid Search*, serta kombinasi keduanya. Pendekatan ini memberikan analisis yang lebih sistematis terhadap kontribusi masing-masing metode, terutama dalam meningkatkan kinerja pada kelas minoritas

METODE PENELITIAN

Seluruh rangkaian penelitian ini dilaksanakan menggunakan Google Colab sebagai lingkungan utama untuk pemrosesan data, analisis, dan implementasi algoritma. Berikut tahapan penelitian yang disajikan pada Gambar 1. Secara umum, tahapan penelitian ini terdiri dari beberapa langkah utama yang dilakukan secara berurutan, yaitu dimulai dari proses pengumpulan data dengan mengumpulkan *dataset* dari Kaggle, kemudian dilanjutkan dengan analisis eksplorasi data (EDA) untuk memahami karakteristik data. Selanjutnya dilakukan pra-pemrosesan data yang meliputi pembersihan data, transformasi fitur, dan normalisasi. Setelah itu, data dibagi menjadi data latih dan data uji (*data splitting*) dengan rasio tertentu. Untuk mengatasi ketidakseimbangan kelas, diterapkan teknik SMOTE pada data latih. Tahap berikutnya adalah pelatihan model menggunakan algoritma XGBoost, yang kemudian dioptimasi melalui optimasi hiperparameter menggunakan *Grid Search*. Terakhir, model dievaluasi menggunakan berbagai metrik seperti *accuracy*, *precision*, *recall*, *F1-score*, *Confusion Matrix*, serta *ROC Curve* untuk memastikan performa dan kemampuan generalisasi model.

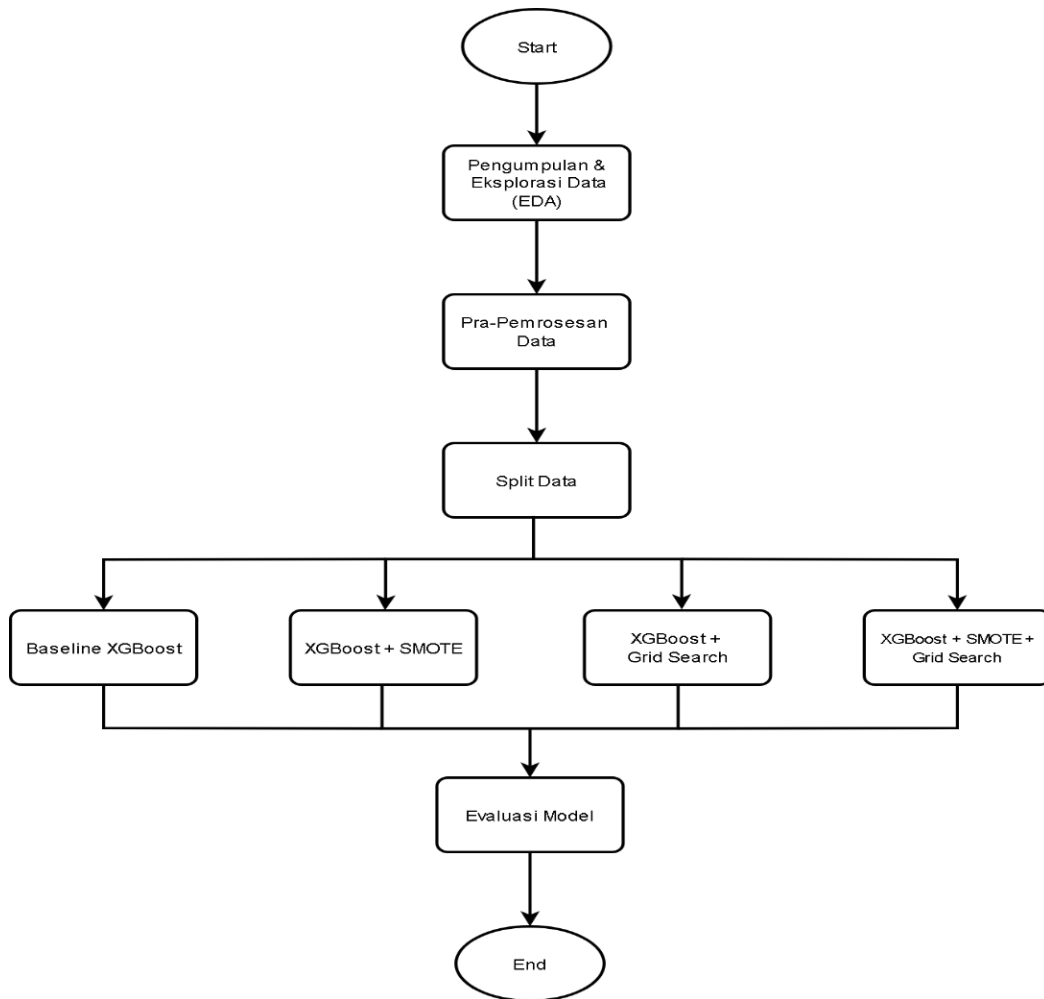
1. Pengumpulan Data dan Eksplorasi Data (EDA)

Dataset yang digunakan dalam penelitian ini adalah *Sleep Health and Lifestyle Dataset* yang diperoleh dari Kaggle dan dapat diakses melalui <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>. *Dataset* ini dipublikasikan pada tahun 2023 dan terdiri dari 374 data dengan 13 variabel seperti pada Tabel 1.

Setelah *dataset* diperoleh, dilakukan proses eksplorasi data (analisis eksplorasi data /EDA) untuk memahami karakteristik data secara menyeluruh. Tahap ini meliputi analisis statistik deskriptif, pemeriksaan keberadaan *missing value* dan data duplikat, serta analisis distribusi kelas pada variabel target. Pada *dataset* asli, kelas target dituliskan sebagai *None*. Namun, dalam penelitian ini istilah tersebut diubah menjadi *No Disorder* untuk menghindari ambiguitas dengan nilai kosong (*missing value*) serta untuk memberikan makna yang lebih jelas secara klinis. Proses EDA bertujuan memastikan data telah dipahami dengan baik sebelum masuk ke tahap pra-pemrosesan data dan pemodelan.

2. Pra-Pemrosesan Data

Setelah proses eksplorasi data selesai, dilakukan tahap pra-pemrosesan data untuk memastikan data siap digunakan dalam pelatihan model. Tahap awal diawali dengan menghapus fitur yang tidak relevan, seperti kolom indeks atau identitas data, karena tidak memberikan kontribusi terhadap proses klasifikasi. Selanjutnya, fitur *Blood Pressure* yang semula berbentuk *string* dengan format "*sistolik/diastolik*" dipisahkan menjadi dua nilai numerik, kemudian dikategorikan ke dalam kelas Normal, *Prehypertension*, dan *Hypertension* melalui proses *binning*.



Gambar 1. Tahapan penelitian

Tabel 1. Deskripsi dataset

No.	Atribut	Keterangan
1.	<i>Person ID</i>	Kode unik yang ditetapkan untuk mengidentifikasi setiap peserta.
2.	<i>Gender</i>	Jenis kelamin biologis individu (Laki-laki atau Perempuan).
3.	<i>Age</i>	Usia seseorang dinyatakan dalam tahun.
4.	<i>Occupation</i>	Jenis pekerjaan atau peran profesional yang dipegang oleh individu.
5.	<i>Sleep Duration</i>	Jumlah total waktu tidur individu setiap hari, diukur dalam jam.
6.	<i>Quality of Sleep</i>	Indeks kualitas tidur yang dinilai sendiri menggunakan skala penilaian yang berkisar antara 1 hingga 10.
7.	<i>Physical Activity Level</i>	Durasi aktivitas fisik seseorang per hari dalam hitungan menit.
8.	<i>Stress Level</i>	Ukuran intensitas stres yang dilaporkan sendiri, dinilai menggunakan skala penilaian yang berkisar antara 1 hingga 10.
9.	<i>BMI Category</i>	Kategori Indeks Massa Tubuh (BMI) seseorang, termasuk Berat Badan Kurang, Berat Badan Normal, atau Kelebihan Berat Badan.
10.	<i>Blood Pressure</i>	Tekanan darah individu ditampilkan sebagai nilai sistolik versus diastolik.
11.	<i>Heart Rate</i>	Jumlah detak jantung per menit saat istirahat.
12.	<i>Daily Steps</i>	Total langkah yang diambil seseorang dalam sehari.
13.	<i>Sleep Disorder</i>	Menunjukkan apakah individu tersebut mempunyai gangguan tidur, dan jika iya, jenisnya (<i>Insomnia, No Disorder, Sleep Apnea</i>).

Fitur-fitur kategorikal seperti *Gender*, *Occupation*, *BMI Category*, *BP Category*, dan *Sleep Disorder* kemudian dikonversi ke dalam bentuk numerik menggunakan teknik label *encoding*. Selanjutnya, fitur numerik seperti *Age*, *Sleep Duration*, *Quality of Sleep*, *Physical Activity Level*, *Stress Level*, *Heart Rate*, dan *Daily Steps* dinormalisasi menggunakan metode *Min–Max Scaling* untuk menyamakan rentang nilai dan mencegah dominasi fitur tertentu dalam proses pembelajaran model.

3. *Split Data*

Dataset yang telah melalui tahap pra-pemrosesan selanjutnya dibagi menjadi data latih dan data uji menggunakan metode *train–test split* dengan rasio 80:20. Pembagian data dilakukan sesuai kelompok masing-masing untuk menjaga proporsi kelas pada variabel target yang tidak seimbang. Tujuan pembagian ini adalah agar model dapat dievaluasi menggunakan data yang tidak pernah dilihat sebelumnya, sehingga hasil evaluasi mencerminkan kemampuan generalisasi model secara objektif. Tahap pembagian data ini menjadi dasar penerapan beberapa skenario eksperimen yang diuji dalam penelitian.

4. Pembuatan Model

Berdasarkan ukuran dataset yang relatif terbatas, penelitian ini menerapkan teknik validasi untuk memastikan kemampuan generalisasi model serta meminimalkan risiko *overfitting*. Validasi dilakukan menggunakan *k-fold cross validation* pada seluruh skenario pemodelan guna memperoleh estimasi performa yang lebih stabil. Selain itu, untuk menghindari *data leakage*, teknik SMOTE hanya diterapkan pada data latih, sedangkan evaluasi model dilakukan menggunakan data uji yang tidak pernah dilihat sebelumnya. Dengan pendekatan ini, model yang dihasilkan diharapkan lebih *robust* dan mampu melakukan generalisasi dengan baik pada data baru. Selain itu, dilakukan perbandingan antara performa pada data latih dan data uji untuk mengidentifikasi potensi *overfitting* pada model.

a. Model *Baseline XGBoost*

Pada skenario pertama, model klasifikasi dibangun menggunakan algoritma XGBoost sebagai model *baseline* tanpa penerapan teknik penyeimbangan kelas maupun optimasi hiperparameter, sehingga model menggunakan parameter bawaan. Model dikonfigurasi dengan *objective = multi:softprob* untuk klasifikasi multikelas dengan tiga kelas target, serta menggunakan *eval_metric = mlogloss*, sementara parameter lainnya menggunakan pengaturan default dari pustaka XGBoost dengan *random_state = 42* untuk menjaga konsistensi hasil eksperimen. Untuk memastikan kestabilan performa model, dilakukan evaluasi menggunakan teknik *Stratified k-fold cross validation* ($k = 5$) pada data latih dengan menjaga distribusi kelas pada setiap fold. Selanjutnya, model dilatih menggunakan data latih dan diuji pada data uji untuk mengukur kemampuan generalisasi. Model *baseline* ini digunakan sebagai acuan untuk membandingkan pengaruh penerapan teknik penanganan ketidakseimbangan kelas dan optimasi hiperparameter pada skenario berikutnya, meskipun pada kondisi ini model berpotensi lebih berpihak pada kelas mayoritas sehingga performa klasifikasi pada kelas minoritas cenderung kurang optimal.

b. Model XGBoost dengan SMOTE

Pada skenario kedua, setelah dilakukan pembagian data, diterapkan penanganan ketidakseimbangan kelas pada data latih menggunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE) sebelum proses pelatihan model. SMOTE dikonfigurasi dengan parameter *sampling_strategy = {0:120, 2:120}*, sehingga jumlah data pada kelas minoritas, yaitu *Insomnia* dan *Sleep Apnea*, masing-masing ditingkatkan menjadi 120 data tanpa mengubah jumlah kelas mayoritas. Parameter *k_neighbors = 3*

digunakan untuk membentuk sampel sintetis berdasarkan tiga tetangga terdekat dalam ruang fitur, serta *random_state* = 42 untuk menjaga konsistensi eksperimen. Proses *oversampling* dilakukan hanya pada data latih untuk menghindari *data leakage*, sementara data uji tetap menggunakan distribusi asli. Selanjutnya, model klasifikasi dibangun menggunakan algoritma XGBoost dengan konfigurasi *objective* = *multi:softmax*, *num_class* = 3, dan *eval_metric* = *mlogloss*, dengan parameter lainnya menggunakan pengaturan bawaan. Untuk memastikan kestabilan performa model, digunakan teknik *Stratified k-fold cross validation* ($k = 5$) dengan pendekatan *pipeline*, di mana proses SMOTE diterapkan secara dinamis pada setiap *fold* data latih tanpa memengaruhi data validasi. Model kemudian dilatih menggunakan data latih yang telah diseimbangkan dan dievaluasi pada data uji untuk mengukur kemampuan generalisasi, sehingga diharapkan mampu meningkatkan performa pada kelas minoritas tanpa mengorbankan kinerja keseluruhan model.

c. Model XGBoost dengan Grid Search

Pada skenario ketiga, setelah dilakukan pembagian data, dilakukan optimasi hiperparameter pada algoritma XGBoost menggunakan metode *Grid Search* sebelum proses pelatihan model. Proses optimasi dilakukan pada data latih dengan mengevaluasi berbagai kombinasi parameter, meliputi *n_estimators* (100, 200, 300), *max_depth* (3, 5, 7), *learning_rate* (0.01, 0.05, 0.1), serta parameter *subsample* dan *colsample_bytree* masing-masing sebesar 0.8 dan 1.0. Setiap kombinasi parameter diuji menggunakan teknik *Stratified k-fold cross validation* ($k = 5$) dengan metrik evaluasi *accuracy* sebagai dasar pemilihan konfigurasi terbaik. Pendekatan ini bertujuan untuk memperoleh parameter optimal yang mampu meningkatkan performa model serta mengurangi risiko *overfitting*. Setelah diperoleh kombinasi hiperparameter terbaik (*best_estimator_*), model XGBoost dilatih menggunakan data latih dan selanjutnya dievaluasi pada data uji untuk mengukur kemampuan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

d. Model XGBoost dengan SMOTE dan Grid Search

Pada skenario keempat, setelah dilakukan pembagian data, diterapkan kombinasi penanganan ketidakseimbangan kelas menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) dan optimasi hiperparameter menggunakan metode *Grid Search*. Proses ini diimplementasikan menggunakan pendekatan *pipeline* yang mengintegrasikan SMOTE dan algoritma XGBoost dalam satu alur pemrosesan. SMOTE dikonfigurasi dengan parameter *sampling_strategy* = {0:120, 2:120}, *k_neighbors* = 5, dan *random_state* = 42, serta diterapkan hanya pada data latih untuk menyeimbangkan distribusi kelas tanpa memengaruhi data uji. Selanjutnya, dilakukan optimasi hiperparameter menggunakan *Grid Search* dengan teknik *Stratified k-fold cross validation* ($k = 5$) dan metrik evaluasi *accuracy*. Ruang pencarian hiperparameter meliputi *n_estimators* (100, 200, 300), *max_depth* (3, 5, 7), *learning_rate* (0.01, 0.05, 0.1), serta parameter *subsample* dan *colsample_bytree* masing-masing sebesar 0.8 dan 1.0. Setiap kombinasi parameter dievaluasi untuk memperoleh konfigurasi terbaik berdasarkan performa validasi silang. Setelah diperoleh kombinasi hiperparameter optimal (*best_estimator_*), model XGBoost dilatih menggunakan data latih yang telah diseimbangkan secara dinamis dan selanjutnya dievaluasi pada data uji untuk mengukur kemampuan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

5. Evaluasi Model

Tahap terakhir dalam penelitian ini adalah mengevaluasi kinerja model menggunakan *data testing*. Pada fase ini, digunakan sejumlah metrik evaluasi seperti

accuracy, *precision*, *recall*, *F1-score*, *Confusion Matrix*, serta *ROC Curve (Receiver Operating Characteristic)* untuk mengukur kemampuan model dalam membedakan setiap kelas secara lebih komprehensif.

Accuracy digunakan untuk mengukur tingkat ketepatan prediksi model terhadap keseluruhan data dan dihitung melalui rumus (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision menunjukkan proporsi prediksi positif yang benar dibandingkan dengan seluruh data yang diprediksi sebagai positif, sesuai dengan rumus (2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Sementara itu, *Recall* menggambarkan proporsi data positif yang berhasil dikenali model dari seluruh data yang sebenarnya termasuk kategori positif, sebagaimana pada rumus (3).

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

Adapun *F1-score* merupakan nilai gabungan antara *Precision* dan *Recall* yang memberikan gambaran keseimbangan performa keduanya, dihitung berdasarkan rumus (4).

$$F1 = \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Penelitian ini juga menggunakan *Confusion Matrix* untuk mengevaluasi kinerja model pada masing-masing kelas. *Confusion Matrix* merepresentasikan jumlah prediksi yang benar dan salah untuk setiap kelas dalam bentuk tabel, sehingga memudahkan analisis pola kesalahan klasifikasi, misalnya ketika kelas *Insomnia* atau *Sleep Apnea* salah diprediksi sebagai tidak mengalami gangguan tidur. Melalui *Confusion Matrix*, peneliti dapat mengamati secara lebih rinci distribusi *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* bagi setiap kelas, yang kemudian menjadi dasar dalam menafsirkan nilai *precision*, *recall*, dan *F1-score*.

Selain itu, penelitian ini juga menggunakan *ROC Curve (Receiver Operating Characteristic)* untuk mengevaluasi kemampuan model dalam membedakan setiap kelas pada berbagai nilai ambang (*threshold*). *ROC Curve* merupakan grafik yang menggambarkan hubungan antara *True Positive Rate (TPR)* atau *recall* dengan *False Positive Rate (FPR)*. Nilai TPR dihitung sebagai $TP/(TP+FN)$, sedangkan FPR dihitung sebagai $FP/(FP+TN)$.

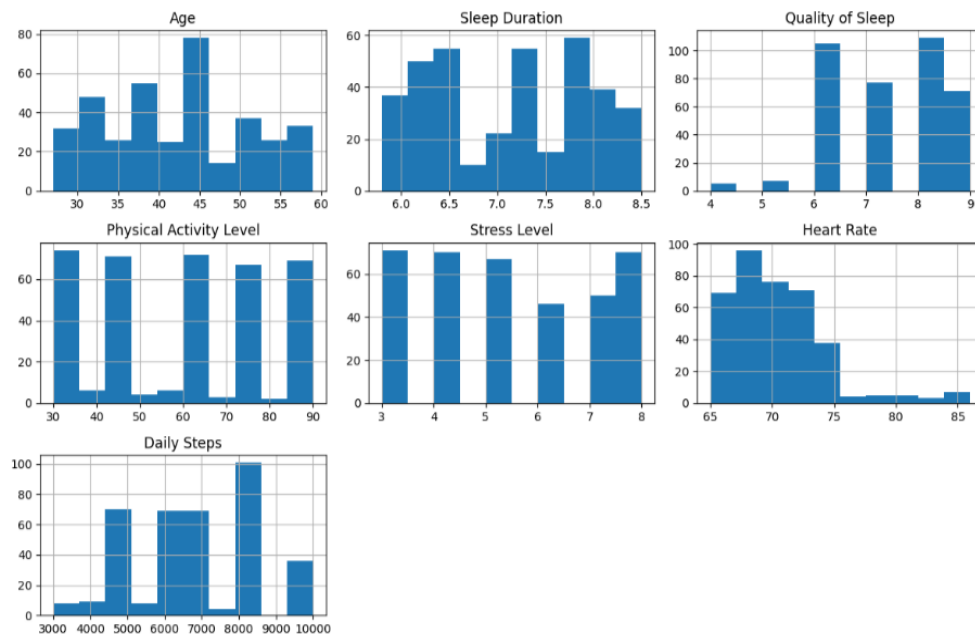
Kinerja model pada *ROC Curve* umumnya diukur menggunakan nilai *Area Under Curve (AUC)*, yang merepresentasikan kemampuan model dalam membedakan kelas secara keseluruhan. Nilai AUC berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan performa model yang sangat baik dalam membedakan kelas, sedangkan nilai mendekati 0,5 menunjukkan performa yang setara dengan tebakan acak. Dengan demikian, *ROC Curve* dan AUC digunakan sebagai metrik tambahan untuk memastikan bahwa model tidak hanya memiliki akurasi tinggi, tetapi juga memiliki kemampuan diskriminasi yang baik terhadap setiap kelas.

HASIL DAN PEMBAHASAN

1. Eksplorasi Data

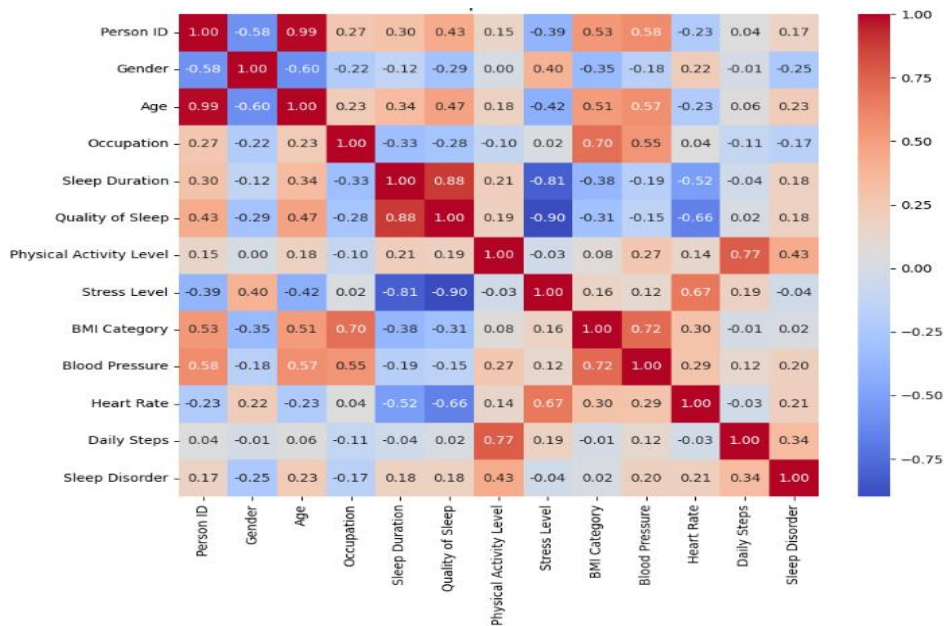
Gambaran umum data menunjukkan bahwa responden dalam *dataset* memiliki rata-rata usia sebesar 42,18 tahun, dengan rentang usia antara 27 hingga 59 tahun. Rata-rata durasi tidur responden adalah 7,13 jam per hari, sedangkan nilai rata-rata *Quality of Sleep*

sebesar 7,31 mengindikasikan kualitas tidur pada tingkat sedang hingga baik. Selain itu, rata-rata tingkat aktivitas fisik tercatat sebesar 59,17, tingkat stres sebesar 5,38, denyut jantung rata-rata 70,17 bpm, serta jumlah langkah harian sekitar 6.816 langkah, yang mencerminkan keberagaman karakteristik gaya hidup responden.



Gambar 2. Distribusi variabel numerik

Distribusi variabel numerik divisualisasikan menggunakan histogram seperti ditunjukkan pada Gambar 2. Hasil visualisasi menunjukkan bahwa sebagian besar variabel memiliki distribusi yang relatif terkonsentrasi pada rentang nilai tertentu. Variabel *Sleep Duration* dan *Quality of Sleep* cenderung berada pada kisaran nilai menengah, sedangkan *Heart Rate* terdistribusi pada rentang normal sekitar 65–75 bpm. Variabel *Daily Steps* menunjukkan penyebaran yang lebih luas dibandingkan variabel lainnya, yang mengindikasikan variasi tingkat aktivitas harian responden. Secara umum, tidak ditemukan pola distribusi yang ekstrem, sehingga data dinilai cukup representatif untuk tahap pemodelan.



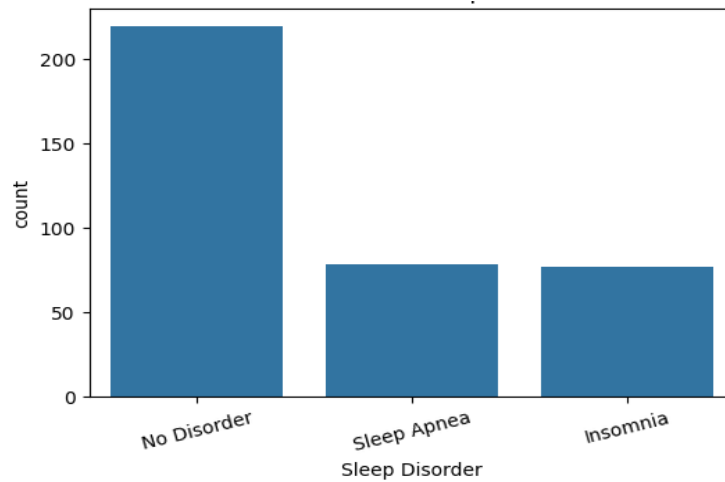
Gambar 3. Heatmap korelasi fitur

Analisis korelasi antar fitur dilakukan menggunakan heatmap seperti ditunjukkan pada Gambar 3. Hasil visualisasi menunjukkan adanya hubungan yang cukup kuat antar beberapa variabel. *Sleep Duration* memiliki korelasi positif yang tinggi dengan *Quality of Sleep*, yang mengindikasikan bahwa durasi tidur yang lebih lama cenderung diikuti dengan kualitas tidur yang lebih baik. Sebaliknya, *Stress Level* menunjukkan korelasi negatif yang kuat terhadap *Quality of Sleep*, yang menunjukkan bahwa peningkatan tingkat stres dapat menurunkan kualitas tidur. Selain itu, *Physical Activity Level* memiliki korelasi positif dengan *Daily Steps*, yang mencerminkan konsistensi antara aktivitas fisik dan jumlah langkah harian. Secara umum, tidak terdapat multikolinearitas yang sangat tinggi antar fitur, sehingga seluruh variabel masih relevan untuk digunakan dalam pemodelan.

Seluruh variabel dalam *dataset* memiliki jumlah data lengkap sebanyak 374 entri tanpa adanya *missing value* maupun data duplikat. Dengan kondisi tersebut, *dataset* dapat langsung digunakan pada tahap pra-pemrosesan dan pemodelan.

Pada analisis distribusi kelas menunjukkan bahwa distribusi kelas tidak seimbang seperti dalam Gambar 4, di mana kelas *No Disorder* mendominasi *dataset* dengan jumlah 219 data, sedangkan kelas *Insomnia* dan *Sleep Apnea* masing-masing memiliki 77 dan 78 data. Ketidakseimbangan ini menunjukkan bahwa proporsi kelas minoritas relatif jauh lebih kecil dibandingkan kelas mayoritas.

Gambar 4 menunjukkan distribusi kelas pada variabel target memiliki kondisi yang tidak seimbang dan berpotensi memengaruhi kinerja model dalam mendeteksi kelas minoritas; oleh karena itu, pada penelitian ini diterapkan teknik SMOTE sebelum proses pelatihan model. Secara keseluruhan, *dataset* memiliki kualitas data yang baik dengan data lengkap, tanpa nilai hilang maupun duplikasi, sehingga layak digunakan pada tahap pra-pemrosesan dan pemodelan klasifikasi gangguan tidur.



Gambar 4. Distribusi kelas *sleep disorder*

2. Pra-Pemrosesan Data

Tahap pra-pemrosesan menghasilkan *dataset* yang telah siap digunakan untuk proses pemodelan. Pada tahap ini, fitur-fitur yang tidak relevan seperti kolom indeks atau identitas data berhasil dihapus sehingga hanya atribut yang berkontribusi terhadap proses klasifikasi yang dipertahankan. Selain itu, fitur *Blood Pressure* yang semula berbentuk data *string* telah ditransformasikan melalui pemisahan nilai *sistolik* dan *diastolik*, kemudian dikategorikan ke dalam kelas *Normal*, *Prehypertension*, dan *Hypertension*. Hasil kategorisasi ini direpresentasikan dalam fitur *BP Category* berbentuk numerik dapat dilihat pada Gambar 5.

	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Heart Rate	Daily Steps	Sleep Disorder	BP Category
0	1	0.00000	9	0.111111	0.4	0.2	0.6	3	0.571429	0.171429	1	2
1	1	0.03125	1	0.148148	0.4	0.5	1.0	0	0.476190	1.000000	1	2
2	1	0.03125	1	0.148148	0.4	0.5	1.0	0	0.476190	1.000000	1	2
3	1	0.03125	6	0.037037	0.0	0.0	1.0	2	0.952381	0.000000	2	0
4	1	0.03125	6	0.037037	0.0	0.0	1.0	2	0.952381	0.000000	2	0

Gambar 5. Ringkasan struktur data

Pada Gambar 5 diperlihatkan seluruh fitur kategorikal, yaitu *Gender*, *Occupation*, *BMI Category*, *BP Category*, dan *Sleep Disorder*, telah berhasil dikonversi ke dalam bentuk numerik menggunakan teknik label *encoding*. Proses ini memungkinkan data kategorikal dapat diproses oleh algoritma XGBoost yang membutuhkan input numerik. Sementara itu, fitur numerik seperti *Age*, *Sleep Duration*, *Quality of Sleep*, *Physical Activity Level*, *Stress Level*, *Heart Rate*, dan *Daily Steps* telah dinormalisasi menggunakan metode *Min-Max Scaling*, sehingga seluruh nilai berada pada rentang 0 hingga 1.

Dataset hasil pra-pemrosesan terdiri dari 11 fitur numerik dan satu variabel target, yaitu *Sleep Disorder*, yang seluruhnya berada dalam format numerik sesuai Gambar 3. Kondisi ini menunjukkan bahwa data telah memenuhi prasyarat untuk digunakan pada tahap selanjutnya.

3. Pembuatan Model

a. Model *Baseline XGBoost* (Tabel 2)

Tabel 2. Hasil evaluasi kinerja model tanpa *Baseline* XGBoost

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Insomnia	0.78	0.93	0.85
<i>No Disorder</i>	1.00	0.93	0.96
<i>Sleep apnea</i>	0.94	0.94	0.94
<i>Accuracy</i>			0.93

Pada Tabel 2 menampilkan hasil evaluasi model *baseline* XGBoost yang menunjukkan akurasi sebesar 0,93, dengan performa terbaik pada kelas mayoritas (*No Disorder*) yang memperoleh nilai *F1-score* sebesar 0,96. Sebaliknya, performa pada kelas minoritas, khususnya *Insomnia* (*F1-score* 0,85), masih lebih rendah dibandingkan kelas lainnya. Perbedaan ini mengindikasikan adanya bias model terhadap kelas mayoritas yang disebabkan oleh distribusi data yang tidak seimbang.

Pada kondisi ini, XGBoost cenderung lebih sering memprediksi kelas dengan jumlah data lebih besar karena algoritma berusaha meminimalkan kesalahan secara global. Akibatnya, pola pada kelas minoritas tidak dipelajari secara optimal, sehingga beberapa kasus *Insomnia* dan *Sleep Apnea* berpotensi salah diklasifikasikan sebagai *No Disorder*. Hal ini menunjukkan bahwa meskipun akurasi keseluruhan tergolong tinggi, model belum memiliki sensitivitas yang baik terhadap kelas minoritas, yang justru memiliki signifikansi klinis lebih penting. Temuan ini sejalan dengan konsep *imbalanced learning* yang menyatakan bahwa pada *dataset* dengan distribusi kelas tidak seimbang, model klasifikasi cenderung bias terhadap kelas mayoritas. Kondisi ini dapat menyebabkan nilai akurasi terlihat tinggi secara keseluruhan, meskipun kemampuan model dalam mendeteksi kelas minoritas sebenarnya rendah (Cartus *et al.*, 2023).

b. Model dengan SMOTE & tanpa *Grid Search* (Tabel 3)

Tabel 3. Hasil evaluasi kinerja model dengan SMOTE & tanpa *Grid Search*

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Insomnia	0.82	0.93	0.88
<i>No Disorder</i>	1.00	0.93	0.96
<i>Sleep apnea</i>	0.88	0.94	0.91
<i>Accuracy</i>			0.93

Hasil penerapan SMOTE pada model XGBoost dapat dilihat pada Tabel 3 di atas yang menunjukkan peningkatan performa pada kelas minoritas, terutama pada nilai *recall* dan *F1-score*, yang mengindikasikan bahwa model menjadi lebih mampu mengenali kasus *Insomnia* dan *Sleep Apnea* dibandingkan dengan model *baseline*. Meskipun akurasi keseluruhan tidak mengalami peningkatan yang signifikan, terjadi perbaikan pada keseimbangan performa antar kelas. Kondisi ini menunjukkan bahwa peningkatan performa tidak hanya diukur dari akurasi, tetapi juga dari kemampuan model dalam mendeteksi kelas minoritas yang memiliki signifikansi klinis.

Peningkatan tersebut disebabkan oleh perubahan distribusi data yang sebelumnya tidak seimbang menjadi lebih proporsional. Pada kondisi awal, jumlah data kelas minoritas yang lebih sedikit menyebabkan model cenderung bias terhadap kelas mayoritas dan kurang mampu menangkap pola pada kelas minoritas. Teknik SMOTE mengatasi permasalahan ini dengan menghasilkan sampel sintesis berdasarkan kedekatan antar data dalam ruang fitur, sehingga memperkaya representasi kelas minoritas dan membantu model membentuk *decision boundary* yang lebih optimal. Fenomena ini sejalan dengan konsep *imbalanced learning* yang menyatakan bahwa

ketidakseimbangan kelas merupakan tantangan umum dalam *dataset* medis dan dapat menurunkan sensitivitas model terhadap kelas dengan signifikansi klinis tinggi (Cartus *et al.*, 2023).

Secara empiris, efektivitas SMOTE telah dibuktikan dalam berbagai penelitian. Penelitian oleh (Han *et al.*, 2024) menunjukkan bahwa penerapan metode SMOTE yang dikombinasikan dengan XGBoost mampu meningkatkan performa klasifikasi pada data Industrial IoT yang tidak seimbang, sementara (Siagian *et al.*, 2025) melaporkan bahwa kombinasi SMOTE dan XGBoost mampu mencapai *F1-score* 0,94 dan ROC-AUC 0,98. Selain itu, (Syukron *et al.*, 2020) menemukan bahwa SMOTE dapat meningkatkan *recall* kelas minoritas secara signifikan tanpa menurunkan akurasi. Dengan demikian, hasil penelitian ini konsisten dengan studi sebelumnya yang menegaskan bahwa SMOTE efektif dalam meningkatkan sensitivitas model pada *dataset* tidak seimbang.

c. Model dengan *Grid Search* & tanpa SMOTE

Tabel 4. Hasil evaluasi kinerja model dengan *Grid Search* tanpa SMOTE

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Insomnia	0.93	0.87	0.90
<i>No Disorder</i>	1.00	1.00	1.00
<i>Sleep apnea</i>	0.88	0.94	0.91
<i>Accuracy</i>			0.96

Tabel 4 memperlihatkan hasil penerapan *Grid Search* pada model XGBoost yang menghasilkan peningkatan performa dibandingkan model *baseline*, ditunjukkan dengan kenaikan akurasi menjadi 0,96 serta perbaikan nilai *F1-score* pada kelas Insomnia. Hal ini menunjukkan bahwa konfigurasi hiperparameter memiliki peran penting dalam meningkatkan kinerja model, bahkan tanpa dilakukan penyeimbangan data.

Peningkatan tersebut disebabkan oleh kemampuan *Grid Search* dalam menemukan kombinasi parameter optimal yang sesuai dengan karakteristik *dataset*. Parameter seperti *max_depth*, *learning_rate*, dan *n_estimators* berpengaruh langsung terhadap kompleksitas model dan proses pembelajaran. Dengan konfigurasi yang tepat, model mampu menangkap pola non-linear secara lebih efektif sehingga menghasilkan prediksi yang lebih akurat dan stabil.

Namun demikian, performa model masih didominasi oleh kelas mayoritas, yang terlihat dari nilai *F1-score* sempurna pada kelas *No Disorder*. Hal ini menunjukkan bahwa optimasi hiperparameter saja belum cukup untuk mengatasi ketidakseimbangan kelas. Temuan ini menegaskan bahwa optimasi parameter dan penanganan distribusi data memiliki peran yang berbeda namun saling melengkapi, sebagaimana juga ditunjukkan dalam penelitian sebelumnya (Sugihartono *et al.*, 2025)

d. Model dengan SMOTE & *Grid Search* (Tabel 5)

Tabel 5. Hasil evaluasi kinerja model dengan SMOTE & *Grid Search*

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Insomnia	0.93	0.93	0.93
<i>No Disorder</i>	1.00	1.00	1.00
<i>Sleep apnea</i>	0.94	0.94	0.94
<i>Accuracy</i>			0.97

Penerapan kombinasi SMOTE dan *Grid Search* menghasilkan performa terbaik dibandingkan seluruh skenario yang diuji, sebagaimana ditunjukkan pada Tabel 5, dengan akurasi mencapai 0,97 serta nilai *precision*, *recall*, dan *F1-score* yang relatif seimbang pada seluruh kelas. Peningkatan ini menunjukkan bahwa model tidak hanya mampu mencapai akurasi tinggi, tetapi juga memiliki kemampuan klasifikasi yang lebih merata antara kelas mayoritas dan minoritas. Hal ini menandakan bahwa model telah mampu mengatasi dua permasalahan utama dalam klasifikasi data medis, yaitu ketidakseimbangan kelas dan konfigurasi parameter yang belum optimal.

Hasil optimasi menunjukkan bahwa kombinasi parameter terbaik yang diperoleh adalah $max_depth = 5$, $learning_rate = 0,01$, $n_estimators = 300$, $subsample = 0,8$, dan $colsample_bytree = 0,8$.

Selain itu, hasil validasi silang (*k-fold cross validation*) menunjukkan nilai rata-rata akurasi sebesar 88,63%, yang lebih rendah dibandingkan hasil pengujian pada data uji sebesar 97,33%. Perbedaan ini menunjukkan bahwa model memiliki performa yang lebih tinggi pada data uji, namun secara umum masih mampu melakukan generalisasi dengan baik. Variasi ini dapat disebabkan oleh ukuran *dataset* yang relatif kecil, sehingga distribusi data pada setiap *fold* dan data uji dapat memengaruhi hasil evaluasi. Meskipun terdapat perbedaan tersebut, nilai performa yang diperoleh tetap berada pada kategori tinggi, sehingga model tidak menunjukkan indikasi *overfitting* yang signifikan.

Kinerja optimal tersebut terjadi karena adanya efek gabungan antara SMOTE dan *Grid Search*. SMOTE berperan dalam memperbaiki distribusi data dengan meningkatkan representasi kelas minoritas, sehingga model memiliki informasi yang cukup untuk mempelajari pola *Insomnia* dan *Sleep Apnea* secara lebih baik. Sementara itu, *Grid Search* mengoptimalkan hiperparameter XGBoost seperti *learning_rate*, *max_depth*, dan *n_estimators*, sehingga model dapat menyesuaikan kompleksitasnya dengan karakteristik data yang telah diseimbangkan. Kombinasi ini memungkinkan model *membentuk decision boundary* yang lebih akurat dan tidak bias terhadap kelas tertentu.

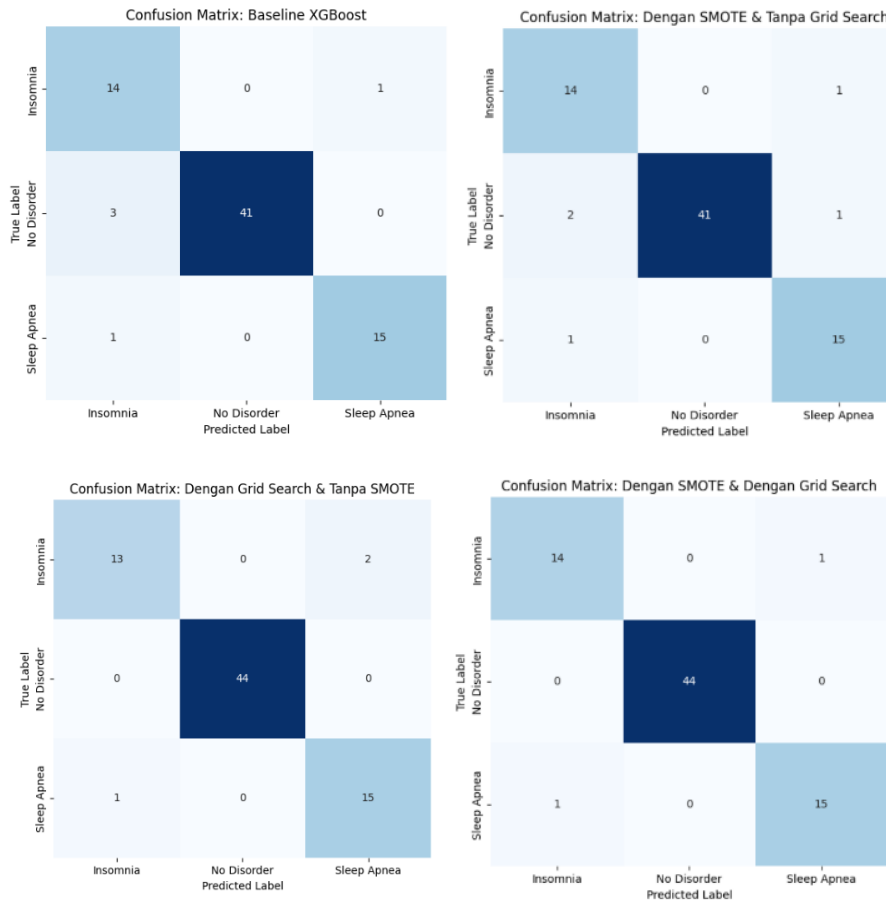
Selain itu, model tidak lagi cenderung mengklasifikasikan kasus minoritas sebagai *No Disorder*, yang sebelumnya menjadi kelemahan utama pada model *baseline*. Temuan ini konsisten dengan penelitian sebelumnya yang menyatakan bahwa integrasi teknik penyeimbangan data dan optimasi hiperparameter mampu meningkatkan performa model secara signifikan pada *dataset* tidak seimbang (Syukron *et al.*, 2020; Siagian *et al.*, 2025). Dengan demikian, kombinasi SMOTE dan *Grid Search* terbukti menjadi pendekatan yang efektif dalam menghasilkan model klasifikasi yang lebih akurat, stabil, dan tidak bias.

Pada Gambar 6 menampilkan visualisasi *Confusion Matrix* pada empat skenario eksperimen dalam kerangka klasifikasi prediktif, yaitu: (1) *Baseline* XGBoost, (2) XGBoost dengan SMOTE, (3) XGBoost dengan *Grid Search*, dan (4) XGBoost dengan SMOTE dan *Grid Search*. Pada keempat skenario tersebut, kelas *No Disorder* secara konsisten menunjukkan performa klasifikasi yang sangat baik, dengan 41 prediksi benar pada model awal dan meningkat menjadi 44 prediksi benar tanpa kesalahan setelah penerapan *Grid Search*.

Sebaliknya, peningkatan performa pada kelas *Insomnia* dan *Sleep Apnea* lebih terlihat setelah penerapan optimasi hiperparameter. Pada model awal, kelas *No Disorder* masih memiliki 3 kesalahan klasifikasi, setelah menambahkan *Grid Search* kesalahan tersebut berkurang menjadi 0. Kombinasi SMOTE dan *Grid Search* menghasilkan performa paling stabil dengan 14 prediksi benar pada *Insomnia*, 44 pada *No Disorder*, dan 15 pada *Sleep Apnea*, sehingga kesalahan klasifikasi menjadi minimal. Secara keseluruhan, hasil ini menunjukkan bahwa penyeimbangan data dan optimasi

hiperparameter memberikan kontribusi dalam meningkatkan kestabilan dan ketepatan model pada seluruh kelas.

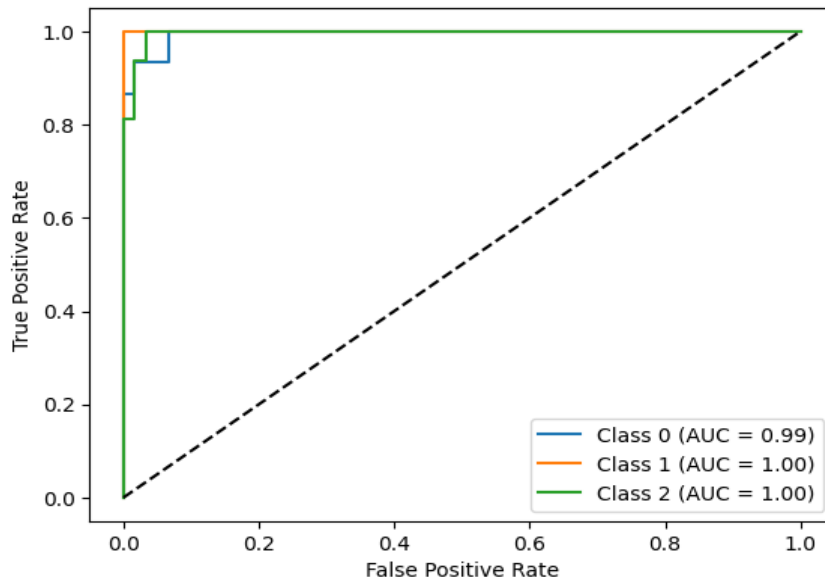
e. Analisis Confusion Matrix (Gambar 6)



Gambar 6. Confusion Matrix keempat skenario

f. Evaluasi ROC Curve (Gambar 7)

Untuk memastikan bahwa model tidak mengalami overfitting, dilakukan evaluasi tambahan menggunakan *ROC Curve* pada model terbaik, yaitu skenario keempat (XGBoost dengan SMOTE dan *Grid Search*), seperti ditunjukkan pada Gambar 7. Hasil menunjukkan bahwa nilai *Area Under Curve* (AUC) untuk masing-masing kelas berada pada rentang 0,99 hingga 1,00, yang mengindikasikan kemampuan model yang sangat baik dalam membedakan setiap kelas gangguan tidur.



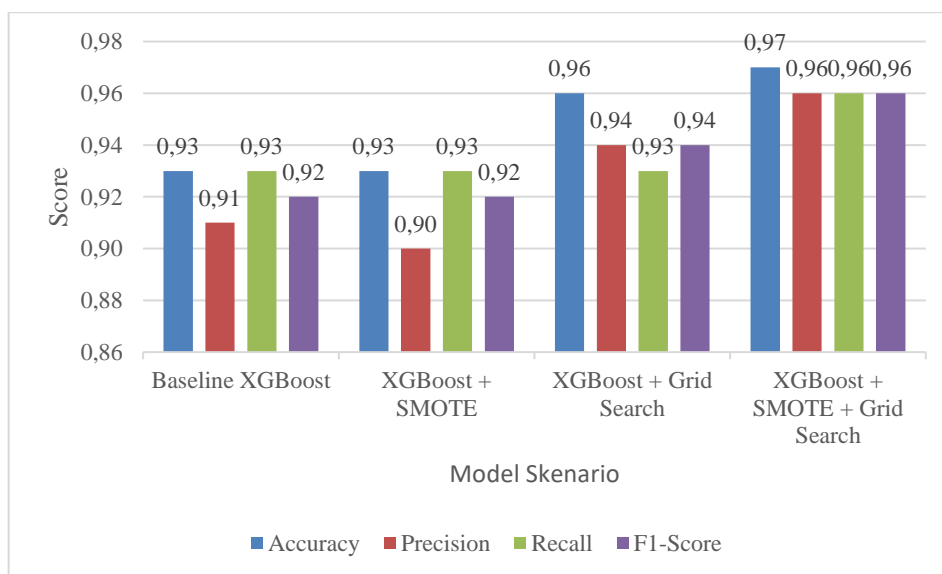
Gambar 7. ROC Curve klasifikasi multikelas

Kurva ROC yang mendekati sudut kiri atas menunjukkan bahwa model memiliki tingkat *true positive rate* yang tinggi dan *false positive rate* yang rendah. Hal ini menandakan bahwa model tidak hanya memiliki akurasi yang tinggi, tetapi juga mampu melakukan klasifikasi secara konsisten pada berbagai *threshold*.

Meskipun demikian, nilai AUC yang sangat tinggi perlu diinterpretasikan secara hati-hati mengingat ukuran *dataset* yang relatif kecil.

4. Pembahasan Hasil Evaluasi

Dalam penelitian ini diuji empat skenario pemodelan berbasis XGBoost, yaitu Model 1 tanpa SMOTE dan *Grid Search*, Model 2 dengan SMOTE tanpa *Grid Search*, Model 3 dengan *Grid Search* tanpa SMOTE, serta Model 4 dengan kombinasi SMOTE dan *Grid Search*, sebagaimana dirangkum pada Gambar 8.



Gambar 8. Perbandingan evaluasi kinerja model

Hasil evaluasi pada Gambar 8 menunjukkan bahwa model XGBoost Tanpa SMOTE dan juga *Grid Search* masih melakukan kesalahan klasifikasi pada kelas insomnia dan *sleep apnea*, yang menurunkan nilai *recall* dan *F1-score* dibandingkan kelas *No Disorder*, meskipun akurasi keseluruhan tetap tinggi. Pola ini sejalan dengan (Mawardi *et al.*, 2025) yang melaporkan kesalahan klasifikasi XGBoost pada kelas *Sleep Apnea* sehingga menurunkan *recall* dan *F1-score*, mengindikasikan bahwa algoritma XGBoost tidak selalu optimal untuk kelas tertentu pada kasus gangguan tidur. Penerapan SMOTE pada penelitian ini terbukti meningkatkan sensitivitas model terhadap kelas *Insomnia* dan *Sleep Apnea*, tercermin dari peningkatan *recall* dan *F1-score* dibandingkan kondisi tanpa SMOTE, selaras dengan temuan (Anshory *et al.*, 2025) dan studi lain pada data tidak seimbang yang menunjukkan efektivitas SMOTE dalam meningkatkan performa kelas minoritas (Syukron *et al.*, 2020; Han *et al.*, 2024).

Selain penyeimbangan distribusi data, optimasi hiperparameter menggunakan *Grid Search* juga berperan penting dalam meningkatkan stabilitas dan akurasi XGBoost, karena setelah pencarian parameter terbaik nilai akurasi, *presisi*, *recall*, dan *F1-score* menjadi lebih seimbang antar kelas sesuai dengan penelitian (Sugihartono *et al.*, 2025). Kinerja terbaik diperoleh ketika SMOTE dan *Grid Search* diterapkan secara bersamaan, ditandai dengan akurasi sekitar 97% dan kesalahan klasifikasi yang minimal pada seluruh kelas, sehingga kedua teknik tersebut saling melengkapi: SMOTE meningkatkan kepekaan model terhadap pola *Insomnia* dan *Sleep Apnea*, sedangkan *Grid Search* memastikan konfigurasi parameter XGBoost berada pada titik optimal.

Meskipun model menghasilkan akurasi yang tinggi, yaitu sebesar 97%, perlu diperhatikan bahwa ukuran *dataset* yang relatif kecil berpotensi menyebabkan *overfitting*. Oleh karena itu, dilakukan evaluasi tambahan menggunakan *ROC Curve* dan validasi silang untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik. Hasil evaluasi menunjukkan bahwa nilai AUC berada pada kategori sangat baik, sementara hasil *cross validation* menunjukkan performa yang lebih rendah dibandingkan data uji. Perbedaan ini dapat disebabkan oleh variasi distribusi data pada *dataset* yang relatif kecil. Selain itu, penerapan SMOTE hanya pada data latih juga memastikan tidak terjadinya *data leakage*. Dengan demikian, model yang dihasilkan tidak menunjukkan indikasi *overfitting* yang signifikan dan tetap mampu melakukan generalisasi terhadap data baru.

KESIMPULAN

Berdasarkan hasil pengujian terhadap empat skenario model dalam klasifikasi gangguan tidur, penelitian ini menunjukkan bahwa penerapan SMOTE berkontribusi dalam meningkatkan sensitivitas model terhadap kelas minoritas, sedangkan optimasi hiperparameter menggunakan *Grid Search* mampu meningkatkan stabilitas dan akurasi model secara keseluruhan. Kombinasi XGBoost dengan SMOTE dan *Grid Search* menghasilkan kinerja terbaik dengan akurasi mencapai 97%, serta nilai *precision*, *recall*, dan *F1-score* yang seimbang pada seluruh kelas. Hasil ini juga diperkuat oleh evaluasi *ROC Curve* dan validasi silang yang menunjukkan bahwa model memiliki kemampuan generalisasi yang baik. Dengan demikian, pendekatan yang diusulkan terbukti efektif dalam meningkatkan performa klasifikasi gangguan tidur dan berpotensi menjadi dasar dalam pengembangan sistem deteksi dini berbasis *machine learning* yang lebih akurat dan *robust*.

DAFTAR PUSTAKA

Afrisia, S.P., Hana, F.M. and Wahyudin, W.C. (2024) 'Implementasi Metode Long Short Term Memory (LSTM) pada Chatbot Kesehatan Mental Mahasiswa', *Sainteks*,

- 21(2), p. 107. Available at: <https://doi.org/10.30595/sainteks.v21i2.23869>.
- Anshory, M.N. *et al.* (2025) 'Application of Adaboost Algorithm with SMOTE and Optuna Techniques in Sleep Disorder Classification', *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 7(2), pp. 415–426. Available at: <https://doi.org/10.35882/ijeemi.v7i2.99>.
- Benjafield, A. V. *et al.* (2025) 'Estimation of the global prevalence and burden of insomnia: a systematic literature review-based analysis', *Sleep Medicine Reviews*, 82. Available at: <https://doi.org/10.1016/j.smr.2025.102121>.
- Bili, B.E., Febriani, H. and Sunaryo (2025) 'Insomnia Pada Mahasiswa Stikes Wira Husada Yogyakarta', *Jurnal Kesehatan Masyarakat*, 18, pp. 11–20. Available at: <https://doi.org/10.47317/jkm.v18i1.738>.
- Cartus, A.R. *et al.* (2023) 'Outcome class imbalance and rare events: An underappreciated complication for overdose risk prediction modeling', *Addiction*, 118(6), pp. 1167–1176. Available at: <https://doi.org/10.1111/add.16133>.
- Han, Y., Wei, Z. and Huang, G. (2024) 'An imbalance data quality monitoring based on SMOTE-XGBOOST supported by edge computing', *Scientific Reports*, 14(1), pp. 1–13. Available at: <https://doi.org/10.1038/s41598-024-60600-x>.
- Haq, O.M., Ridwan, A. and Pratama, T.G. (2025) 'Analisis Perbandingan Kinerja Algoritma Naïve Bayes Dan KNN Untuk Memprediksi Penyakit Diabetes', *Progresif: Jurnal Ilmiah Komputer*, 21(1), p. 193. Available at: <https://doi.org/10.35889/progresif.v21i1.2424>.
- Khan, M.A. and Al-Jahdali, H. (2023) 'The consequences of sleep deprivation on cognitive performance', *Neurosciences*, 28(2), pp. 91–99. Available at: <https://doi.org/10.17712/nsj.2023.2.20220108>.
- Mawardi, A.B., Pradini, R.S. and Haris, M.S. (2025) 'Komparasi Algoritma Boosting untuk Prediksi Gangguan Tidur', 13(3). Available at: <https://doi.org/10.23960/jitet.v13i3.7281>.
- Palombini, L.D.O. *et al.* (2024) '2024 Position Statement on the Use of Different Diagnostic Methods for Sleep Disorders in Adults - Brazilian Sleep Association', *Sleep Science*, 17(4), pp. e476–e492. Available at: <https://doi.org/10.1055/s-0044-1800887>.
- Pratama, T.G. *et al.* (2026) 'Deteksi Kesehatan Mental Mahasiswa Muslim dengan Pendekatan Pendidikan Islam menggunakan Machine Learning', 15, pp. 144–156. Available at: <https://doi.org/10.32520/stmsi.v15i1.5732>.
- Siagian, N.A. *et al.* (2025) 'Integrating SMOTE with XGBoost for Robust Classification on Imbalanced Datasets: A Dual-Domain Evaluation', *Sinkron*, 9(3), pp. 1094–1107. Available at: <https://doi.org/10.33395/sinkron.v9i3.15029>.
- Sugihartono, T. *et al.* (2025) 'Optimalisasi Kinerja Extreme Gradient Boosting melalui Grid Search untuk Prediksi Risiko Penyakit Jantung', 5(10), pp. 3074–3084. Available at: <https://doi.org/10.52436/1.jpti.1346>.
- Syukron, M., Santoso, R. and Widiharih, T. (2020) 'Perbandingan Metode SMOTE Random Forest dan SMOTE XGBoost untuk Klasifikasi Tingkat Penyakit Hepatitis C pada Imbalance Class Data', 9, pp. 227–236. Available at: <https://doi.org/10.14710/j.gauss.9.3.227-236>.
- Tyas, L.W. (2022) 'Hubungan Depresi, Kecemasan dan Stres dengan Kejadian Insomnia pada Remaja Sekolah Menengah Atas Kota Surabaya', *Preventif: Jurnal Kesehatan Masyarakat*, 13(3), pp. 540–547. Available at:

<https://doi.org/10.22487/preventif.v13i3.405>.

- Wahyudin, W.C., Hana, F.M. and Prihandono, A. (2023) 'Prediksi Stunting pada Balita Di Rumah Sakit Kota Semarang Menggunakan Naive Bayes', 2019, pp. 32–36. Available at: <https://doi.org/10.26751/jikoma.v4i1.1792.g1066>.
- Wulansih, N.C. *et al.* (2024) 'Literature Review: The Impact of Poor Sleep Duration and Quality on Health in Productive Age', *Jurnal Kesehatan Poltekkes Kemenkes Ri Pangkalpinang.*, 12(1), pp. 71–82. Available at: <https://doi.org/10.32922/jkp.v12i1.932>.