

Prediksi Multi Label Komorbiditas Kesehatan Mental Remaja berdasarkan Faktor Psikososial menggunakan XGBoost dan Explainable AI

Multi-Label Prediction of Adolescent Mental Health Comorbidity based on Psychosocial Factors using XGBoost and Explainable AI

¹M Dian Fauzi*, ²Ferian Fauzi Abdulloh

^{1,2}Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

^{1,2}Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta, Indonesia 55283

*e-mail: dianfauzi16@students.amikom.ac.id, ferian@amikom.ac.id

(received: 19 June 2026, revised: 2 September 2026, accepted: 3 September 2026)

Abstrak

Kesehatan mental remaja merupakan isu penting yang dapat memengaruhi kualitas hidup dan prestasi akademik. Pendekatan skrining berbasis Depression, Anxiety, and Stress Scales (DASS-42) umumnya memprediksi setiap kondisi secara terpisah, sehingga kurang mampu merepresentasikan komorbiditas yang sering terjadi antara depresi, kecemasan, dan stres. Penelitian ini bertujuan mengembangkan model prediksi risiko kesehatan mental remaja berbasis klasifikasi *multi-label* yang mampu menangkap hubungan antar-label sekaligus mengatasi ketidakseimbangan kelas. Model yang diusulkan mengintegrasikan *eXtreme Gradient Boosting* (XGBoost) dengan arsitektur Classifier Chains, sedangkan ketidakseimbangan data ditangani menggunakan *Multi-Label Synthetic Minority Over-Sampling Technique* (MLSMOTE). Dataset yang digunakan merupakan data yang diperoleh dari repositori DASS-19 pada platform Kaggle yang berbasis instrumen kuesioner DASS-42 dan telah melalui proses prapemrosesan sehingga menghasilkan 10.315 responden dengan 15 fitur psikososial sebagai prediktor. Hasil eksperimen menunjukkan bahwa model XGBoost Classifier Chains menghasilkan kinerja terbaik dengan Macro F1-Score sebesar 0,8774, Hamming Loss sebesar 0,1894, dan Exact Match sebesar 0,6413 pada data uji. Analisis menggunakan *SHapley Additive exPlanations* (SHAP) menunjukkan bahwa beberapa fitur kepribadian memiliki kontribusi dominan terhadap prediksi, serta mengonfirmasi adanya dependensi antar-label yang berhasil dimanfaatkan oleh arsitektur *Classifier Chains*. Hasil penelitian menunjukkan bahwa pendekatan yang diusulkan berpotensi digunakan sebagai pendukung skrining dini berbasis *machine learning* yang transparan untuk membantu mengidentifikasi risiko kesehatan mental pada remaja.

Kata kunci: klasifikasi, DASS-42, kesehatan mental, *multi-label*, SHAP, XGBoost.

Abstract

Adolescent mental health is an important issue that can affect quality of life and academic performance. Screening approaches based on the Depression, Anxiety, and Stress Scales (DASS-42) typically predict each condition separately, making them less capable of representing the comorbidities that frequently occur among depression, anxiety, and stress. This study aims to develop a multi-label classification model for predicting adolescent mental health risks that can capture inter-label dependencies while addressing class imbalance. The proposed model integrates *eXtreme Gradient Boosting* (XGBoost) with a Classifier Chains architecture, while data imbalance is addressed using the *Multi-Label Synthetic Minority Over-sampling Technique* (MLSMOTE). The dataset was obtained from the DASS-19 repository on the Kaggle platform, based on the DASS-42 questionnaire instrument, and underwent preprocessing, resulting in 10,315 respondents with 15 psychosocial features as predictors. Experimental results show that the XGBoost Classifier Chains model achieved the best performance, with a Macro F1-Score of 0.8774, Hamming Loss of 0.1894, and Exact Match of 0.6413 on the test data. Analysis using *SHapley Additive exPlanations* (SHAP) revealed that several personality-related features made dominant contributions to the predictions and confirmed the presence of inter-label dependencies effectively exploited by the Classifier Chains

<http://sistemasi.ftik.unisi.ac.id>

architecture. These findings indicate that the proposed approach has the potential to support transparent machine learning-based early screening and assist in identifying mental health risks among adolescents.

Keywords: classification, DASS-42, mental health, multi-label, SHAP, XGBoost.

1 Pendahuluan

Kesehatan mental merupakan aspek fundamental dari kesejahteraan manusia yang sering kali terabaikan, terutama pada fase perkembangan remaja. Data nasional menunjukkan lonjakan signifikan kasus gangguan psikologis, di mana jutaan remaja dilaporkan mengalami depresi, kecemasan, dan stres yang mengganggu fungsi kognitif serta interaksi sosial [1]. Tekanan akademik yang berat, dinamika keluarga, hingga disrupsi gaya hidup pascapandemi menjadi pemicu utama kemunduran emosional yang apabila tidak tertangani, dapat mengarah pada penurunan produktivitas hingga risiko fatal seperti bunuh diri [2], [3], [4]. Dalam praktik psikologi klinis, instrumen *Depression, Anxiety, and Stress Scales* (DASS-42) telah diakui secara luas dan digunakan untuk mengukur tingkat keparahan gangguan emosional [5], [6]. Namun, pemanfaatan instrumen ini secara konvensional yang mengandalkan agregasi skor linier sering kali gagal menangkap kompleksitas komorbiditas, yakni kondisi di mana seorang remaja mengalami depresi, kecemasan, dan stres secara bersamaan dengan gejala yang saling tumpang tindih [7], [8].

Penerapan *Machine Learning* (ML) telah membuka peluang besar dalam memprediksi kondisi kesehatan mental secara lebih presisi [6], [9]. Sayangnya, mayoritas penelitian terdahulu mengadopsi pendekatan klasifikasi target tunggal (*single-label classification*), yang memodelkan masing-masing gangguan secara terpisah sehingga mengabaikan korelasi probabilitas antar-gejala [5], [10]. Untuk merepresentasikan kompleksitas gejala yang sesungguhnya, diperlukan pendekatan *Multi-Label Classification* (MLC) yang mampu memprediksi lebih dari satu gangguan secara simultan [11], [12]. Selain masalah dependensi label, tantangan krusial lainnya adalah ketidakseimbangan kelas (*class imbalance*) yang ekstrem pada data psikometri, di mana kombinasi gejala komorbid tertentu jauh lebih jarang ditemui daripada kasus tunggal. Ketimpangan ini berpotensi menyebabkan bias mayoritas (*majority class bias*) sehingga model cenderung mengabaikan responden dengan pola gejala yang kompleks [11], [13]. Oleh karena itu, strategi penanganan ketidakseimbangan kelas yang mampu mempertahankan struktur korelasi antar-label menjadi prasyarat mutlak dalam pipeline pemodelan.

Dari sisi pemodelan, *eXtreme Gradient Boosting* (XGBoost) telah banyak digunakan sebagai algoritma *ensemble* yang efektif untuk menangani data tabular dengan interaksi non-linier yang rumit [14]. Dengan mengintegrasikan XGBoost dan strategi *Classifier Chains*, model dapat mempelajari hierarki dependensi antar-label kesehatan mental secara sekuensial, sehingga komorbiditas dapat dimodelkan secara lebih realistis [12], [15]. Meskipun mampu memberikan akurasi prediksi yang tinggi, algoritma kompleks seperti XGBoost sering kali bersifat *black-box*, sehingga menyulitkan tenaga profesional kesehatan untuk memahami dasar logika prediksi yang dihasilkan. Dalam konteks sistem pendukung keputusan awal, transparansi model merupakan syarat mutlak agar hasil prediksi dapat memberikan gambaran awal yang tepat. Penerapan *Explainable AI* (XAI) melalui metode *SHapley Additive exPlanations* (SHAP) menjadi pendekatan yang esensial untuk mengungkap kontribusi relatif setiap fitur prediktor, sehingga proses deteksi dini dapat didukung oleh interpretasi yang jelas dan dapat dipertanggungjawabkan [2], [3], [16].

Meskipun penelitian terdahulu telah mengeksplorasi masing-masing komponen secara terpisah, terdapat dua celah penelitian yang belum terjawab secara integratif. Pertama, seluruh studi yang menggunakan instrumen DASS-42 dalam kerangka *machine learning* masih memformulasikan prediksi sebagai klasifikasi *single-label*, sehingga gagal merepresentasikan realitas gejala komorbid [5], [6]. Berdasarkan telaah literatur yang dilakukan, masih terbatas penelitian yang mengintegrasikan strategi *Classifier Chains* pada XGBoost dengan penanganan ketidakseimbangan kelas *multi-label* (MLSMOTE) secara simultan untuk data psikometri remaja [12], [13], [15]. Kedua, meskipun XAI berbasis SHAP telah digunakan pada prediksi kesehatan mental *single-label* [2], [3], [16], penerapannya dalam kerangka *multi-label* psikometri yang mampu mengidentifikasi faktor risiko dan faktor pelindung secara per-label masih sangat jarang dikaji, sehingga tenaga profesional tidak memiliki dasar eksplanatori yang memadai untuk menindaklanjuti hasil skrining komorbiditas.

Berdasarkan celah tersebut, penelitian ini bertujuan membangun arsitektur *machine learning* untuk mendeteksi dini risiko komorbiditas depresi, kecemasan, dan stres pada remaja berdasarkan faktor psikososial secara *multi-label* dan transparan. Arsitektur yang diusulkan mengintegrasikan XGBoost berbasis *Classifier Chains* untuk pemodelan dependensi antar-label, penanganan ketidakseimbangan kelas menggunakan pendekatan *oversampling* berbasis (MLSMOTE) guna mempertahankan integritas data, serta SHAP untuk interpretasi transparan secara global dan lokal. Seluruh model dievaluasi menggunakan ambang batas probabilitas standar (0,5) pada *hold-out test set* untuk mensimulasikan kondisi skrining di dunia nyata secara objektif dan terukur.

2 Tinjauan Literatur

Berbagai penelitian telah mengimplementasikan *machine learning* untuk memprediksi kondisi kesehatan mental berdasarkan instrumen psikometri. Sulak dan Koklu mengeksplorasi metode *data mining* konvensional seperti *Naive Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* untuk menganalisis skala DASS-42 pada mahasiswa [5], sementara Nisa dan Rosyadi mengandalkan algoritma *Decision Tree* dan *Naive Bayes* untuk prediksi serupa [10], dan Patil serta Deshpande menggunakan *Decision Tree* dan *Random Forest* dengan hasil yang sebanding [6]. Demikian pula, Shinde dan Ghotkar [9] serta Juwariyah dkk. [17] mengonfirmasi bahwa faktor demografis memainkan peran penting dalam prediksi kesehatan mental berbasis *machine learning*. Namun, seluruh penelitian tersebut memiliki satu kelemahan mendasar: mereka memformulasikan masalah psikiatri sebagai klasifikasi target tunggal (*single-label*), sehingga mengabaikan korelasi empiris bahwa gejala depresi, kecemasan, dan stres sering kali beririsan secara komorbid [7], [8]. Febriani dkk. [11] mencoba mengatasi irisan ini dengan melakukan klasifikasi *multi-label* pada kasus *cyberpsychology* remaja menggunakan SVM, namun model mereka mengalami penurunan performa drastis akibat kegagalan menangani ketidakseimbangan kelas (*imbalanced data*) yang ekstrem pada label minoritas.

Tantangan kedua dalam prediksi berbantuan kuesioner medis adalah besarnya ruang dimensi fitur mentah yang berpotensi memicu multikolinearitas. Beberapa penelitian sebelumnya berupaya mengatasinya dengan menerapkan algoritma seleksi fitur, seperti pendekatan berbasis *wrapper* menggunakan *Powershap* [16] atau statistik *univariate SelectKBest* [18]. Berbeda dengan pendekatan tersebut, penelitian ini mengadopsi strategi penyaringan berbasis pengetahuan domain (*domain-driven filtering*) pada tahap prapemrosesan untuk mengeliminasi variabel yang tidak relevan dengan konteks demografis remaja Indonesia. Hasilnya, diperoleh subset 15 fitur psikososial inti yang selanjutnya dipertahankan seluruhnya dalam proses pemodelan. Pendekatan ini dipilih dengan pertimbangan bahwa algoritma *ensemble* modern seperti XGBoost memiliki mekanisme regularisasi inherent dan pembobotan fitur (*feature importance*) yang kuat, sehingga mampu menangani multikolinearitas secara mandiri tanpa perlu mereduksi fitur secara paksa yang dapat mengorbankan keutuhan interpretasi psikososial [14].

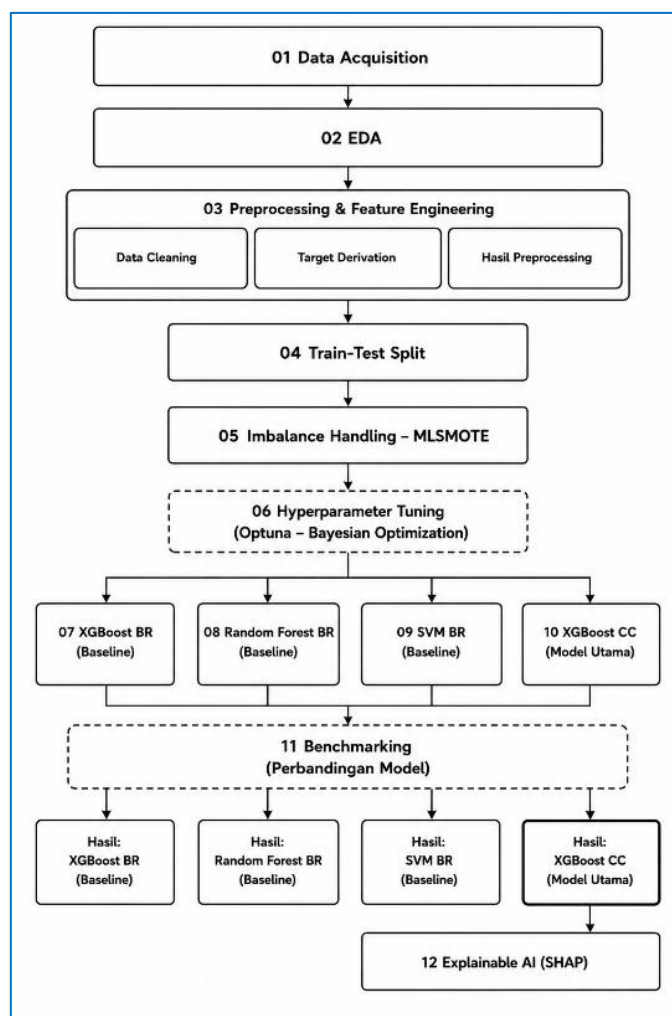
Tantangan terakhir yang menghambat adopsi *machine learning* dalam domain asesmen kesehatan emosional adalah kurangnya transparansi dari model *black-box*. Zhao [2] dan Zulfa dkk. [3] telah memelopori penggunaan kombinasi XGBoost dan *Explainable AI* (SHAP) untuk mengungkap faktor-faktor yang paling memicu risiko depresi dan berkorelasi dengan prestasi akademik mahasiswa. Serupa dengan itu, Bozorgmehr dan Weltermann memanfaatkan SHAP untuk mengekstraksi faktor pelindung (*protective factors*) dari stres kronis pada data *cross-sectional* berskala nasional [16]. Pendekatan interpretasi berbasis *game theory* ini terbukti menggeser kedudukan AI dari sekadar mesin prediktor menjadi pendukung keputusan skrining berbasis *machine learning* yang transparan [3], [16]. Namun, seluruh studi interpretabilitas tersebut masih beroperasi dalam kerangka *single-label*. Belum ada yang mengeksplorasi bagaimana SHAP dapat menjelaskan prediksi komorbiditas dalam konteks *multi-label* psikometri secara per-label, yang sangat dibutuhkan agar tenaga profesional dapat memberikan visualisasi eksplanatori tambahan bagi praktisi untuk memahami pola risiko secara lebih terukur.

Sintesis literatur di atas menunjukkan adanya *research gap* yang tajam: belum ada studi yang membangun kerangka kerja terpadu untuk prediksi komorbiditas kesehatan mental remaja secara *multi-label* sekaligus menyertakan interpretasi klinis yang utuh. Artikel ini mengisi celah tersebut dengan menawarkan *novelty* berupa: (1) integrasi arsitektur *Classifier Chains* pada XGBoost dengan

penanganan ketidakseimbangan kelas menggunakan murni MLSTMOTE untuk menangkap dependensi komorbiditas depresi-kecemasan-stres secara simultan tanpa mengorbankan data riil, (2) pemanfaatan seluruh fitur psikososial (tanpa seleksi fitur) untuk menjaga kelengkapan konteks psikososial dalam interpretasi SHAP, dan (3) evaluasi objektif menggunakan ambang batas probabilitas standar (0,5) untuk mensimulasikan kondisi deteksi dini di dunia nyata secara adil dan terukur.

3 Metode Penelitian

Bab ini menguraikan tahapan dan metodologi komputasional yang dirancang untuk membangun sistem prediksi kesehatan mental remaja secara *multi-label*. Seluruh alur penelitian diilustrasikan pada **Gambar 1**, dimulai dari akuisis data hingga interpretasi model menggunakan (*Explainable AI*).



Gambar 1 Alur penelitian

3.1 Pengumpulan Data

Penelitian ini menggunakan dataset DASS-19 yang diperoleh dari platform Kaggle [19]. Penamaan tersebut merujuk pada identitas repositori, bukan varian instrumen. Dataset merupakan hasil survei daring berskala besar yang dikumpulkan pada periode 2017–2019 menggunakan instrumen Depression, Anxiety, and Stress Scales (DASS-42). Dataset mentah terdiri dari 39.775 responden dan 172 atribut yang mencakup item DASS-42, indikator kepribadian TIPI, Vocabulary Check List (VCL), atribut demografi, serta metadata sistem. Dataset ini dipilih karena menyediakan informasi psikologis dan psikososial yang relevan untuk membangun model prediksi kesehatan mental remaja.

3.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) dilakukan untuk memahami karakteristik dataset serta mengidentifikasi potensi permasalahan yang dapat memengaruhi proses pemodelan. Analisis meliputi pemeriksaan *missing value*, distribusi data, deteksi outlier, distribusi label target, dan hubungan antar-label (*label co-occurrence*). Hasil EDA menunjukkan adanya *missing value* pada beberapa atribut demografi, nilai usia yang tidak realistis akibat kesalahan pengisian data, serta ketidakseimbangan distribusi label pada ketiga target kesehatan mental. Selain itu, analisis *co-occurrence* memperlihatkan adanya irisan yang cukup tinggi antara depresi, kecemasan, dan stres, yang mengindikasikan keberadaan pola komorbiditas pada data. Lebih lanjut, analisis distribusi kombinasi labelset menunjukkan bahwa pola (0,0,0) yakni responden yang tidak berisiko pada ketiga kondisi secara bersamaan merupakan kombinasi yang paling jarang ditemui pada subset data latih. Kondisi ini merupakan konsekuensi dari karakteristik *self-selection* bias pada pengisian kuesioner daring, di mana individu yang secara sukarela mengisi instrumen psikologi cenderung merupakan mereka yang sedang mengalami tekanan emosional, serta tingginya ko-okurensi antar ketiga kondisi pada populasi remaja usia 10–18 tahun. Temuan tersebut menjadi dasar dalam proses pembersihan data, penanganan ketidakseimbangan kelas menggunakan MLSTMOTE, serta penerapan pendekatan klasifikasi multi-label pada penelitian ini.

3.3 Prapemrosesan Data dan Feature Engineering

Prapemrosesan data dilakukan untuk mentransformasi dataset mentah ke dalam format yang siap dipelajari oleh algoritma machine learning. Tahapan ini meliputi pembersihan data (*data cleaning*), seleksi fitur berbasis domain, dan penurunan target (*target derivation*) berdasarkan domain *knowledge* psikologi.

3.3.1 Data Cleaning

Proses pembersihan data dilakukan untuk menyaring data mentah agar layak digunakan dalam proses pemodelan. Tahapan prapemrosesan data secara sistematis beserta deskripsinya dirangkum pada Tabel 1.

Tabel 1 Tahapan prapemrosesan

Tahap	Deskripsi
Filtering usia	Mempertahankan responden usia 10–18 tahun
Penghapusan metadata	Menghapus kolom posisi soal, waktu pengerjaan, dan log sistem
Penghapusan VCL	Menghapus 16 fitur Vocabulary Check List
Penghapusan fitur demografi	Menghapus atribut yang tidak relevan atau memiliki kualitas data rendah
Penanganan missing value	Menghapus kolom dengan missing value tinggi

Setelah proses pembersihan data, jumlah responden berkurang dari 39.775 menjadi 10.315 data. Dataset hasil prapemrosesan selanjutnya digunakan pada tahap pembentukan label target dan pemodelan.

3.3.2 Penurunan Target Berbasis Domain Knowledge (Target Derivation)

Label target dibentuk berdasarkan pedoman interpretasi DASS-42 [5], [8]. Skor depresi, kecemasan, dan stres dihitung dengan menjumlahkan item pada masing-masing subskala, kemudian dikonversi menjadi label biner menggunakan ambang batas kategori sedang ke atas sebagai kondisi berisiko (1) dan kategori normal hingga ringan sebagai kondisi tidak berisiko (0). Proses ini menghasilkan tiga label target, yaitu *risk_depression*, *risk_anxiety*, dan *risk_stress*. Untuk menghindari *data leakage*, seluruh item DASS-42 dihapus setelah proses pembentukan target selesai.

3.4 Pemisahan Data Latih dan Data Uji

Untuk mengevaluasi kemampuan generalisasi model terhadap data baru, dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) dengan rasio 80:20. Pembagian ini menggunakan metode stratified *hold-out split* berdasarkan kombinasi label untuk memastikan proporsi setiap pola komorbiditas tetap seimbang di kedua subset. Nilai *random state* ditetapkan secara tetap untuk menjamin reproduktibilitas eksperimen. Data uji selanjutnya dikunci dan tidak digunakan dalam

proses pelatihan, validasi, maupun optimasi parameter untuk menghindari kebocoran data (*data leakage*).

3.5 MLSMOTE Balancing

Ketidakseimbangan kelas pada data latih ditangani menggunakan *Multi-Label Synthetic Minority Over-Sampling Technique* (MLSMOTE) [13] dengan parameter dan mekanisme sebagai berikut:

1. Nilai Tetangga Terdekat: Algoritme MLSMOTE menggunakan parameter $k = 5$ (*k-neighbors*) untuk menentukan ruang tetangga terdekat dari sampel minoritas acuan (*seed sample*).
2. Proses Oversampling: Fitur sintetis baru dibuat melalui interpolasi linier acak antara sampel acuan dengan salah satu tetangganya yang dipilih secara acak, menggunakan bobot distribusi seragam antara nilai 0 hingga 1.
3. Penentuan Label Target: Penelitian ini menggunakan pewarisan langsung (*direct inheritance*) di mana label sampel sintetis baru sepenuhnya disamakan dengan label sampel acuan asli. Langkah ini diambil untuk menjaga konsistensi hubungan komorbiditas klinis yang realistis.
4. Strategi Penyeimbangan: Berdasarkan hasil analisis distribusi labelset pada tahap EDA, kombinasi (0,0,0) atau kondisi tidak berisiko pada ketiga label secara bersamaan terbukti menjadi pola labelset yang paling jarang ditemukan pada subset latih, sebuah kondisi yang tidak lazim namun dapat dijelaskan oleh *self-selection* bias pada pengisian kuesioner daring berbasis DASS-42. Oleh karena itu, oversampling difokuskan pada labelset minoritas tersebut. Proses ini hanya diterapkan pada data latih, sedangkan data uji tetap dipertahankan pada kondisi aslinya untuk menjaga objektivitas evaluasi.

3.6 Model Development

3.6.1 Arsitektur Classifier Chains dengan XGBoost

Model yang diusulkan mengintegrasikan XGBoost dengan arsitektur Classifier Chains. XGBoost digunakan sebagai algoritma dasar karena kemampuannya dalam memodelkan hubungan nonlinier pada data tabular, sedangkan Classifier Chains memungkinkan pemanfaatan dependensi antar-label dengan menjadikan prediksi label sebelumnya sebagai fitur tambahan pada proses prediksi label berikutnya, dengan urutan *chain* yang ditetapkan sebagai *risk_depression* → *risk_anxiety* → *risk_stress* berdasarkan hubungan irisan gejala DASS-42 yang saling memengaruhi satu sama lain [7], [8]. Untuk menjamin validitas evaluasi, arsitektur ini dirancang bebas dari kebocoran data (*data leakage*) melalui pemisahan mekanisme pada dua fase sesuai dengan protokol standar pemodelan sekuensial standar [12], [15]:

1. Fase Pelatihan (*Training*): Model mempelajari hubungan dependensi sekuensial menggunakan label aktual (*ground truth*) yang hanya berasal dari data latih [12]. Tidak ada informasi dari data uji yang terlibat dalam fase ini.
2. Fase Pengujian (*Testing/Inference*): Saat mengevaluasi data uji, model tidak menggunakan label aktual data uji sama sekali sebagai fitur masukan. Sebagai gantinya, fitur tambahan yang diumpangkan ke klasifikasi berikutnya adalah nilai prediksi sekuensial (yaitu *pred_depression* untuk memprediksi kecemasan, dan *pred_anxiety* bersama *pred_depression* untuk memprediksi stres) yang dihasilkan secara mandiri oleh model pada tahap rantai sebelumnya [15]. Dengan demikian, integritas *testing set* tetap terjaga secara objektif tanpa adanya kebocoran informasi (*train-test leakage*).

3.6.2 Hyperparameter Tuning dan Cross Validation

Optimasi hyperparameter dilakukan menggunakan *Optuna* berbasis *Tree-structured Parzen Estimator* (TPE) [18], [20]. Proses pencarian parameter dievaluasi menggunakan *Multilabel Stratified 5-Fold Cross Validation* dengan *Macro F1-Score* sebagai fungsi objektif. Pendekatan ini digunakan untuk memperoleh konfigurasi parameter yang stabil dan memiliki kemampuan generalisasi yang baik.

3.7 Benchmarking Model

Untuk mengevaluasi efektivitas model yang diusulkan, dilakukan analisis perbandingan (*benchmarking*) terhadap tiga model baseline yang dilatih menggunakan data dan fitur prediktor yang sama. Daftar arsitektur model pembanding beserta model usulan dirangkum pada Tabel 2.

Tabel 2 Baseline model

Model	Arsitektur
-------	------------

<http://sistemasi.ftik.unisi.ac.id>

XGBoost	Classifier Chain (Proposed Model)
XGBoost	Binary Relevance (Baseline)
Random Forest	Binary Relevance (Baseline)
Support Vector Machine	Binary Relevance (Baseline)

Model baseline dipilih menggunakan pendekatan tradisional Binary Relevance (BR) [12] untuk memberikan perbandingan yang adil terhadap model usulan Classifier Chains (CC) [15]. Seluruh model diuji dan dievaluasi secara objektif menggunakan ambang batas probabilitas (*threshold*) sebesar 0,5 pada subset data uji guna menjamin validitas perbandingan kinerja.

3.8 Evaluasi Model

Evaluasi dilakukan menggunakan Macro F1-Score, Hamming Loss, Exact Match, dan Macro ROC-AUC. Macro F1-Score digunakan untuk mengukur keseimbangan presisi dan recall pada setiap label, Hamming Loss untuk mengukur kesalahan prediksi rata-rata, Exact Match untuk menilai ketepatan prediksi seluruh kombinasi label secara simultan, sedangkan Macro ROC-AUC digunakan untuk mengevaluasi kemampuan diskriminasi model secara keseluruhan.

3.9 Explainable AI Menggunakan SHAP

Interpretasi model dilakukan menggunakan SHapley Additive exPlanations (SHAP) [3], [16]. Analisis dilakukan pada tingkat global menggunakan *summary plot* untuk mengidentifikasi fitur yang paling berpengaruh terhadap prediksi, serta pada tingkat lokal menggunakan *waterfall plot* untuk menjelaskan kontribusi setiap fitur terhadap prediksi individu. Pendekatan ini digunakan untuk meningkatkan transparansi model dan mendukung interpretasi hasil prediksi.

4 Hasil dan Pembahasan

Penelitian ini menyajikan hasil dari serangkaian eksperimen *machine learning* yang dirancang untuk mendeteksi dini risiko komorbiditas kesehatan mental remaja secara *multi-label*. Pembahasan akan diuraikan secara sistematis dimulai dari hasil prapemrosesan data, penanganan ketidakseimbangan kelas, evaluasi komparatif model, hingga interpretasi transparansi menggunakan *Explainable AI*.

4.1 Hasil Prapemrosesan dan Karakteristik Data

Tahap prapemrosesan dilakukan untuk menghasilkan dataset yang sesuai dengan kebutuhan pemodelan klasifikasi multi-label serta mengurangi potensi bias selama proses pelatihan model. Dataset awal yang diperoleh dari Kaggle terdiri atas 39.775 responden dan 172 atribut yang mencakup item DASS-42, indikator kepribadian TIPI, Vocabulary Check List (VCL), atribut demografi, serta metadata sistem. Berdasarkan hasil eksplorasi awal, ditemukan beberapa permasalahan data yang berpotensi memengaruhi kualitas pemodelan, di antaranya keberadaan missing value pada atribut tertentu, nilai usia yang tidak sesuai dengan populasi penelitian, serta ketidakseimbangan distribusi label target.

Analisis distribusi labelset pada data latih awal (8.252 data) mengungkap temuan yang menarik: kombinasi labelset (0,0,0) yang merepresentasikan kondisi tidak berisiko pada ketiga dimensi justru merupakan pola yang paling jarang ditemui dibandingkan kombinasi labelset berisiko lainnya. Fenomena ini dapat dijelaskan oleh dua faktor utama. Pertama, karakteristik *self-selection* bias pada pengumpulan data daring, di mana individu yang secara sukarela berpartisipasi dalam survei psikologi cenderung adalah mereka yang sedang mengalami tekanan emosional. Kedua, tingginya tingkat ko-okurensi antar ketiga kondisi pada populasi remaja yang teridentifikasi pada tahap EDA, sehingga sebagian besar responden cenderung masuk ke dalam kombinasi labelset berisiko pada lebih dari satu dimensi secara bersamaan. Kondisi distribusi yang terbalik dari asumsi umum ini menjadi justifikasi empiris penerapan MLSMOTE yang difokuskan pada penyeimbangan labelset (0,0,0) sebagai kelas minoritas.

Proses pembersihan data dilakukan secara bertahap melalui penyaringan responden pada rentang usia 10–18 tahun untuk menyesuaikan fokus penelitian pada kelompok remaja. Selanjutnya dilakukan penghapusan metadata sistem, eliminasi fitur Vocabulary Check List (VCL), penghapusan atribut demografi yang tidak relevan atau memiliki kualitas data rendah, serta penanganan missing value. Setelah proses pembersihan selesai, dilakukan pembentukan tiga label target yaitu *risk_depression*,

risk_anxiety, dan *risk_stress* berdasarkan pedoman interpretasi DASS-42 menggunakan ambang batas kategori sedang ke atas sebagai kondisi berisiko. Untuk mencegah terjadinya *data leakage*, seluruh item DASS-42 yang digunakan dalam pembentukan label kemudian dihapus dari dataset sebelum proses pelatihan model.

Hasil prapemrosesan menunjukkan bahwa jumlah data berkurang dari 39.775 menjadi 10.315 responden dengan 15 fitur psikososial yang dipertahankan sebagai prediktor. Fitur tersebut terdiri atas sepuluh indikator kepribadian TIPI dan lima atribut demografi yang dianggap masih relevan terhadap karakteristik remaja. Ringkasan transformasi dataset disajikan pada Tabel 3.

Tabel 3 Ringkasan transformasi dataset

Tahap	Jumlah Data	Jumlah Fitur	Keterangan
Dataset Awal	39.775	172	Data mentah DASS-19 dari Kaggle.
Setelah prapemrosesan	10.315	15	Setelah filtering usia, penghapusan metadata, VCL, missing value, dan pembentukan target.
Data latih awal	8.252	15	Hasil <i>stratified hold-out split</i> (80%).
Setelah MLSMOTE	14.580	15	Penambahan 6.328 data sintetis.
Data uji	2.063	15	Tidak dilakukan oversampling.

Dataset hasil prapemrosesan kemudian dibagi menggunakan pendekatan *stratified hold-out split* dengan rasio 80:20 untuk menjaga proporsi kombinasi label pada data latih dan data uji. Hasil pembagian menghasilkan 8.252 data latih dan 2.063 data uji. Sebelum proses pemodelan dilakukan, distribusi label pada data latih dianalisis dan menunjukkan adanya ketidakseimbangan kelas yang cukup tinggi sehingga berpotensi menyebabkan model lebih dominan mempelajari pola mayoritas. Oleh karena itu, penanganan ketidakseimbangan dilakukan menggunakan metode MLSMOTE yang diterapkan hanya pada data latih. Proses oversampling menghasilkan penambahan 6.328 sampel sintetis sehingga jumlah data latih dari 8.252 meningkat menjadi 14.580 data, sedangkan data uji dipertahankan dalam kondisi asli untuk menjaga objektivitas evaluasi model.

4.2 Hasil Evaluasi dan Perbandingan Model

Evaluasi dilakukan untuk mengukur efektivitas model usulan dalam memprediksi komorbiditas kesehatan mental remaja secara multi-label. Seluruh model dilatih menggunakan data hasil penanganan ketidakseimbangan kelas yang sama dan dievaluasi pada data uji asli menggunakan ambang probabilitas 0,5 untuk menjaga konsistensi perbandingan. Metrik evaluasi yang digunakan meliputi Macro F1-Score, Exact Match, Hamming Loss, dan Macro ROC-AUC untuk memberikan gambaran performa model dari berbagai aspek klasifikasi multi-label.

Sebelum evaluasi akhir dilakukan, seluruh model terlebih dahulu melalui proses optimasi hyperparameter menggunakan Optuna berbasis *Tree-structured Parzen Estimator* (TPE). Untuk menjaga konsistensi evaluasi dan meminimalkan bias pemilihan parameter, proses tuning dilakukan secara independen pada masing-masing model menggunakan *Multilabel Stratified 5-Fold Cross Validation* dengan Macro F1-Score sebagai fungsi objektif. Hasil validasi menunjukkan bahwa model berbasis XGBoost memiliki performa yang stabil selama proses optimasi, dengan nilai Macro F1 terbaik sebesar $0,8713 \pm 0,0017$ pada arsitektur Binary Relevance dan $0,8759 \pm 0,0017$ pada arsitektur Classifier Chain.

Hasil evaluasi akhir seluruh model pada data uji ditunjukkan pada Tabel 4.

Tabel 4 Perbandingan kinerja seluruh model

Model	Macro F1 Score	Exact Match	Hamming Loss	ROC-AUC
Random Forest (BR)	0.8748	0.6229	0.1897	0.8287
SVM (BR)	0.8251	0.5783	0.2414	0.8254
XGBoost (BR)	0.8719	0.6098	0.1949	0.8142
XGBoost (CC)	0.8774	0.6413	0.1894	0.8231

Secara umum, model berbasis ensemble menunjukkan performa yang lebih baik dibandingkan pendekatan Support Vector Machine. Model usulan, yaitu XGBoost Classifier Chain, memperoleh

nilai Macro F1-Score tertinggi sebesar 0,8774 dan Exact Match tertinggi sebesar 0,6413, serta menghasilkan Hamming Loss terendah sebesar 0,1894. Hasil tersebut menunjukkan bahwa model mampu menghasilkan keseimbangan prediksi antar-label sekaligus meningkatkan kemampuan dalam memprediksi kombinasi kondisi depresi, kecemasan, dan stres secara simultan.

Menariknya, Random Forest menghasilkan performa yang sangat kompetitif dengan Macro F1 sebesar 0,8748 dan Macro ROC-AUC sebesar 0,8287, bahkan sedikit lebih tinggi dibandingkan XGBoost *Classifier Chain* pada metrik ROC-AUC. Namun demikian, model usulan tetap menunjukkan performa keseluruhan yang lebih baik pada *Exact Match* dan *Hamming Loss*. Temuan ini mengindikasikan bahwa kemampuan diskriminasi probabilistik yang tinggi belum tentu menghasilkan ketepatan prediksi kombinasi label yang lebih baik dalam konteks klasifikasi multi-label.

Jika dibandingkan dengan XGBoost *Binary Relevance* yang menggunakan algoritma dasar yang sama, peningkatan performa pada arsitektur *Classifier Chain* menunjukkan bahwa pemanfaatan dependensi antar-label memberikan kontribusi terhadap peningkatan kualitas prediksi. Pendekatan *Binary Relevance* memperlakukan setiap label secara independen sehingga tidak memanfaatkan hubungan antar kondisi psikologis [12]. Sebaliknya, *Classifier Chain* menggunakan prediksi label sebelumnya sebagai fitur tambahan untuk membantu pembentukan keputusan pada label berikutnya. Peningkatan nilai Exact Match dari 0,6098 menjadi 0,6413 menunjukkan bahwa pendekatan tersebut lebih efektif dalam menangkap pola komorbiditas yang umum terjadi pada kesehatan mental remaja.

Untuk memperoleh gambaran yang lebih rinci mengenai kemampuan model pada masing-masing target prediksi, dilakukan evaluasi per-label menggunakan metrik precision, recall, dan F1-score. Hasil evaluasi per-label untuk model XGBoost Classifier Chain disajikan pada Tabel 5.

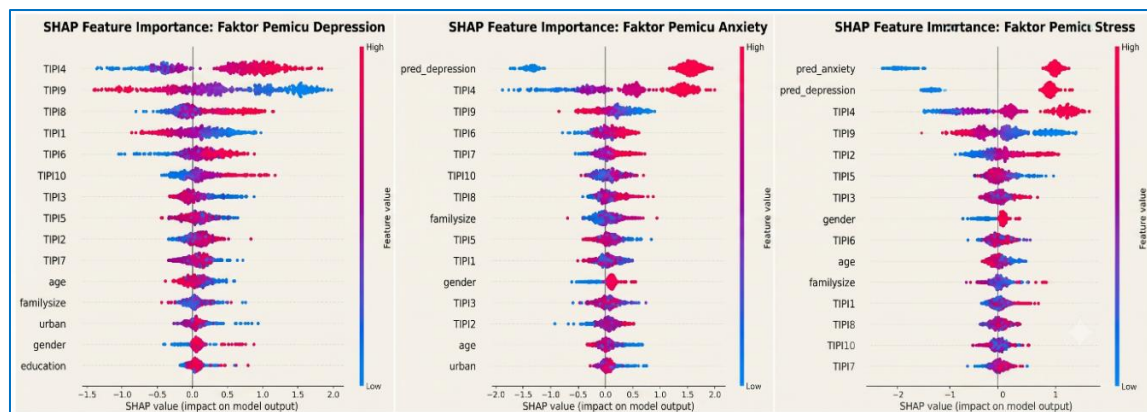
Tabel 5 Hasil evaluasi per-label model XGBoost classifier chain

Label	Precision	Recall	F1-Score
Depression	0.85	0.91	0.88
Anxiety	0.88	0.91	0.89
Stress	0.80	0.92	0.86

Evaluasi per-label menunjukkan bahwa model usulan menghasilkan performa yang relatif seimbang pada seluruh target, dengan nilai F1 sebesar 0,88 untuk depresi, 0,89 untuk kecemasan, dan 0,86 untuk stres. Label kecemasan memperoleh performa tertinggi dengan keseimbangan precision dan recall yang baik, sedangkan label stres menghasilkan nilai recall tertinggi meskipun memiliki precision yang relatif lebih rendah dibandingkan label lainnya. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam mengidentifikasi sebagian besar kasus positif tanpa mengalami penurunan performa yang signifikan pada label tertentu. Temuan ini memperkuat hasil evaluasi global bahwa pendekatan Classifier Chain mampu mempertahankan kualitas prediksi secara konsisten pada seluruh kondisi kesehatan mental yang diamati.

4.3 Interpretasi Model Menggunakan SHAP

Selain mengevaluasi performa prediktif, penelitian ini juga menekankan aspek interpretabilitas untuk memahami faktor yang memengaruhi keputusan model. Oleh karena itu, dilakukan analisis menggunakan SHapley Additive exPlanations (SHAP) pada model terbaik, yaitu XGBoost Classifier Chain. Analisis dilakukan pada 500 sampel data uji untuk memperoleh interpretasi global, serta satu sampel representatif untuk mengevaluasi perilaku prediksi pada tingkat individu.

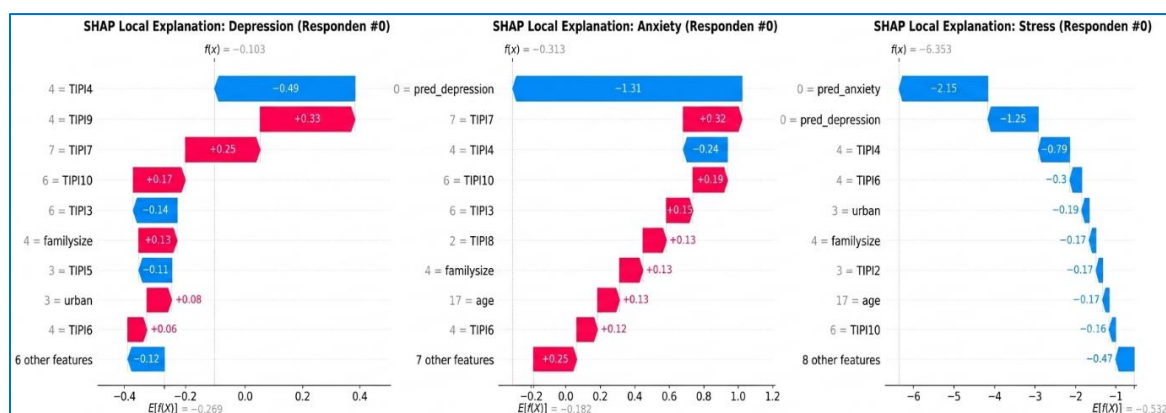


Gambar 2 Interpretasi global

Gambar 2. menunjukkan hasil *SHAP Summary Plot* yang merepresentasikan kontribusi fitur secara global terhadap masing-masing label. Pada prediksi depresi, fitur TIPI4 (*Anxious, easily upset*) dan TIPI9 (*calm, emotionally stable*) menjadi faktor yang paling dominan dalam memengaruhi keputusan model, diikuti oleh indikator kepribadian lainnya seperti TIPI8 (*Disorganized, careless*), TIPI1 (*extraverted, enthusiastic*), dan TIPI6 (*reserved, quiet*). Sementara itu, variabel demografi seperti usia, pendidikan, dan wilayah tempat tinggal menunjukkan kontribusi yang relatif kecil. Temuan ini mengindikasikan bahwa karakteristik psikologis memiliki pengaruh yang lebih besar dibandingkan atribut demografis dalam proses identifikasi risiko kesehatan mental.

Pola yang berbeda terlihat pada prediksi kecemasan dan stres. Pada label kecemasan, variabel *pred_depression* menjadi faktor dengan kontribusi tertinggi, sedangkan pada label stres kontribusi terbesar berasal dari *pred_anxiety* dan *pred_depression*. Dominasi kedua variabel tersebut menunjukkan bahwa prediksi antar-label digunakan secara aktif oleh model untuk membentuk keputusan berikutnya. Hasil ini memberikan indikasi bahwa kondisi depresi, kecemasan, dan stres tidak sepenuhnya berdiri sendiri, tetapi memiliki keterkaitan yang dapat dimanfaatkan oleh pendekatan *Classifier Chain*. Temuan analitik ini secara teoritis sejalan dengan literatur psikologi yang membuktikan bahwa gangguan depresi, kecemasan, dan stres sering terjadi secara bersamaan (ko-okurensi) dan menunjukkan tumpang tindih spektrum gejala pada skala DASS-42 [7], [8].

Untuk memperjelas mekanisme pengambilan keputusan model pada tingkat individu, dilakukan analisis menggunakan *SHAP Waterfall Plot* terhadap salah satu responden pada data uji sebagaimana ditunjukkan pada Gambar 3.



Gambar 3. Interpretasi local

Pada prediksi kecemasan, kontribusi terbesar berasal dari variabel *pred_depression*, sedangkan pada prediksi stres faktor dominan berasal dari *pred_anxiety* dan *pred_depression*. Pola tersebut konsisten dengan hasil interpretasi global dan menunjukkan bahwa keputusan model tidak hanya dipengaruhi oleh karakteristik responden, tetapi juga oleh hasil prediksi kondisi psikologis sebelumnya.

Secara keseluruhan, hasil interpretasi menggunakan SHAP menunjukkan bahwa peningkatan performa XGBoost Classifier Chain tidak hanya berasal dari kemampuan algoritma dasar, tetapi juga dari kemampuannya memanfaatkan hubungan antar-label secara eksplisit. Temuan ini memperkuat asumsi penelitian bahwa pendekatan multi-label lebih sesuai untuk memodelkan komorbiditas kesehatan mental dibandingkan pendekatan yang memperlakukan setiap kondisi secara independen.

5 Kesimpulan

Penelitian ini berhasil mengembangkan model klasifikasi multi-label untuk mengidentifikasi risiko komorbiditas kesehatan mental remaja menggunakan XGBoost Classifier Chain, MLSMOTE, dan optimasi Optuna. Menggunakan dataset sekunder DASS-19 dari platform Kaggle yang berbasis instrumen kuesioner DASS-42, model usulan menunjukkan performa terbaik dengan Macro F1-Score 0,8774, Exact Match 0,6413, dan Hamming Loss 0,1894, yang membuktikan bahwa pemodelan dependensi antar-label mampu meningkatkan akurasi prediksi kondisi depresi, kecemasan, dan stres secara simultan. Analisis SHAP lebih lanjut mengonfirmasi korelasi komorbiditas secara empiris, di mana prediksi label sebelumnya (depresi dan kecemasan) menjadi faktor dominan dalam memprediksi label berikutnya. Meskipun demikian, model ini dirancang murni sebagai sistem pendukung keputusan awal berbasis machine learning yang masih terbatas pada dataset internasional serta cakupan fitur psikososial tertentu. Penelitian selanjutnya diharapkan dapat menggunakan data yang lebih representatif untuk populasi remaja Indonesia serta mengeksplorasi metode penanganan ketidakseimbangan kelas dan pemodelan dependensi label yang lebih adaptif.

Referensi

- [1] Center for Reproductive Health, University of Queensland, and Johns Hopkins Bloomberg School of Public Health, "Indonesia – National Adolescent Mental Health Survey (I-NAMHS): Laporan Penelitian," 2022.
- [2] G. Zhao, "Multi-Model Machine Learning Framework for Effective Depression Risk Management," *Association for Computing Machinery*, pp. 1081–1086, Oct. 2025, DOI: 10.1145/3785706.3785875.
- [3] E. Zulfa, H. Amir, R. Ginting, and Sudarno, "Analisis Korelasi Kesehatan Mental dan Indeks Prestasi Mahasiswa Jurusan Administrasi Niaga Politeknik Negeri Jakarta dengan Kombinasi Metode Xgboost dan Shap," *Jurnal Administrasi Profesional*, Vol. 05, No. 01, pp. 26–37, Jul. 2024, DOI: 10.32722/jap.v5i1.6923.
- [4] M. Rahma, M. Fikry, and Y. Afrillia, "Prediksi Kesehatan Mental Remaja berdasarkan Faktor Lingkungan Sekolah menggunakan Machine Learning," *Jurnal Informatika: Jurnal Pengembangan IT*, Vol. 10, No. 2, pp. 382–390, Apr. 2025, DOI: 10.30591/jpit.v10i2.8556.
- [5] S. A. Sulak and N. Koklu, "Analysis of Depression, Anxiety, Stress Scale (DASS-42) With Methods of Data Mining," *Eur. J. Educ.*, Vol. 59, No. 4, pp. 1–18, Sep. 2024, DOI: 10.1111/ejed.12778.
- [6] M. C. G. Patil and P. D. S. Deshpande, "Harnessing Machine Learning for Accurate Prediction of Mental Health Conditions based on DASS-42 scores," *Journal of Information Systems Engineering and Management*, Vol. 2025, No. 1s, pp. 2468–4376, Dec. 2024, DOI: 10.52783/jisem.v10i1s.231.
- [7] E. G. B. Rahmat, W. Wiharto, and U. Salamah, "Mental Health Detection Expert System Model based on DASS-42 using Fuzzy Inference System," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, Vol. 8, No. 1, pp. 304–323, Jan. 2026, DOI: 10.35882/jeeemi.v8i1.1443.
- [8] Sri Kusumadewi and Hepi Wahyuningsih, "Model Sistem Pendukung Keputusan Kelompok untuk Penilaian Gangguan Depresi, Kecemasan dan Stress berdasarkan DASS-42," *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 7, No. 2, pp. 219–228, Feb. 2020, DOI: 10.25126/jtiik.2020721052.
- [9] S. S. Shinde and A. S. Ghotkar, "From Questionnaires to Actionable Insights: Machine Learning for Mental Stress Detection," *Engineering, Technology and Applied Science Research*, Vol. 15, No. 6, pp. 29240–29250, Dec. 2025, DOI: 10.48084/etasr.13513.

- [10] C. Nisa, “Prediksi Kesehatan Mental Mahasiswa Universitas Yudharta Pasuruan dengan Klasifikasi *Decision Tree*,” *Jurnal Informatika dan Teknik Elektro Terapan*, Vol. 13, No. 3S1, pp. 925–933, Oct. 2025, DOI: 10.23960/jitet.v13i3S1.7925.
- [11] W. Febriani, S. E. Prastya, B. Sabella, and F. D. Marleny, “Analysis of the Impact of Violent Content on Social Media on Adolescent Cyberpsychology using Support Vector Machine and Random Forest,” *Journal of Applied Informatics and Computing (JAIC)*, Vol. 10, No. 1, pp. 707–711, Feb. 2026, DOI: 10.30871/jaic.v10i1.11415.
- [12] M.-L. Zhang, Y.-K. Li, X.-Y. Liu, and X. Geng, “Binary Relevance for Multi-Label Learning: an Overview,” *Front. Comput. SCI.*, Apr. 2018, DOI: doi.org/10.1007/s11704-017-7031-7.
- [13] P. Das and D. H. Mazumder, “Predicting Anatomical Therapeutic Chemical Drug Classes from 17 molecules Properties of Drugs by Multi-Label Binary Relevance Approach with MLSMOTE,” *ACM International Conference Proceeding Series*, Vol. 21, pp. 1–7, Dec. 2021, DOI: 10.1145/3512452.3512453.
- [14] N. Riduan and A. H. Muhammad, “Optimasi Model XGBoost dengan Genetic Algorithm untuk Prediksi Kesehatan Mental Siswa Sekolah Menengah berbasis *Machine Learning*,” *Jurnal Teknologi Informasi dan Komunikasi*, Vol. 3, No. 4, pp. 1427–1436, Oct. 2025, DOI: 10.35870/jtik.v9i4.4013.
- [15] H. Joe and H. G. Kim, “Multi-Label Classification with XGBoost for Metabolic Pathway Prediction,” *BMC Bioinformatics*, Vol. 25, No. 1, pp. 1–15, Dec. 2024, DOI: 10.1186/s12859-024-05666-0.
- [16] A. Bozorgmehr and B. Weltermann, “Prediction of Chronic Stress and Protective Factors in Adults: Development of an Interpretable Prediction Model based on XGBoost and SHAP using National Cross-sectional DEGS1 Data,” *JMIR AI*, Vol. 2, May 2023, DOI: 10.2196/41868.
- [17] S. Juwariyah, A. Hulvi, N. Riduan, and K. Kusrini, “Mengukur Faktor Demografi Psikologis: memprediksi Depresi, Kecemasan, dan Stres dengan menggunakan *Machine Learning*,” *Komputika: Jurnal Sistem Komputer*, Vol. 13, No. 2, pp. 149–156, Sep. 2024, DOI: 10.34010/komputika.v13i2.11793.
- [18] Ștefana Duță and A. E. Sultana, “Optimizing Depression Classification using Combined Datasets and Hyperparameter Tuning with Optuna,” *Sensors*, Vol. 25, No. 7, Apr. 2025, DOI: 10.3390/s25072083.
- [19] Y. Prajapati, “DASS-19 Depression, Anxiety & Stress Scales - 2019,” Kaggle. Accessed: May 29, 2026. [Online]. Available: <https://www.kaggle.com/datasets/yashpra1010/dass-19>
- [20] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” *Cornell University*, Jul. 2019, DOI: 10.48550/arXiv.1907.10902.