

Predicting Depression Levels Among Final-Year Students Using CatBoost Algorithm

Prediksi Tingkat Depresi Mahasiswa Tingkat Akhir Menggunakan Algoritma CatBoost

Shintya Preity Aleng, Bernad J. D. Sitompul, Oktavian A. Lantang

Dept. of Electrical Engineering, Sam Ratulangi University Manado, Kampus Bahu St., 95115, Indonesia

e-mails : shintyaaleng026@student.unsrat.ac.id, bernadjdsitompul@unsrat.ac.id,
oktavian_lantang@unsrat.ac.id

Received: June 27th, 2025 revised: July 26th, 2025; accepted: August 25th, 2025

Abstract — Depression among final-year students completing their thesis is a serious issue that often goes unnoticed, yet it affects mental health and academic productivity. Academic pressure, deadlines, workload, and lack of social support contribute to the risk of depression. This study aims to build a predictive model for depression levels in final-year students based on the five categories from the PHQ-9, namely: Not Depressed, Mild, Moderate, Severe, and Very Severe, using the CatBoost algorithm. Data were collected through an online questionnaire from 450 respondents, consisting of nine PHQ-9 items as the depression level label, and additional questions on psychological factors (academic burden, deadlines, sleep quality, and social support) as well as demographic data (age, gender, study program, and semester) as predictive features. After selection, 440 valid responses were obtained. The preprocessing stages included handling invalid values, detecting outliers, renaming columns, and stratified data splitting into training and testing sets (70:30). Evaluation results show that the CatBoost algorithm outperformed Random Forest and Support Vector Machine (SVM), achieving an accuracy of 88%, macro F1-score of 89%, weighted F1-score of 88%, macro AUC of 97.87%, and weighted AUC of 97.44%. The model demonstrates strong potential as an effective early detection tool for depression among final-year students with high accuracy and efficiency.

Key words — CatBoost; Depression; Final-Year Students; Machine Learning; Prediction; Thesis.

Abstrak — Depresi pada mahasiswa tingkat akhir yang menyusun skripsi merupakan isu serius yang sering tidak disadari, namun berdampak pada kesehatan mental dan produktivitas akademik. Tekanan akademik, tenggat waktu, beban tugas, serta kurangnya dukungan sosial menjadi faktor yang berkontribusi terhadap risiko depresi. Penelitian ini bertujuan membangun model prediksi tingkat depresi mahasiswa akhir berdasarkan lima kategori dari PHQ-9, yaitu Tidak Depresi, Ringan, Sedang, Berat, dan Sangat Berat, menggunakan algoritma CatBoost. Data dikumpulkan melalui kuesioner daring dari 450 responden, yang mencakup sembilan item skala PHQ-9 sebagai label tingkat depresi, serta pertanyaan tambahan mengenai faktor psikologis (beban akademik, tenggat waktu, kualitas tidur, dan dukungan sosial) dan data demografis (usia, jenis kelamin, program studi, semester) sebagai fitur prediktor. Setelah proses seleksi, diperoleh 440 data valid. Tahapan prapemrosesan mencakup penanganan nilai tidak valid, deteksi outlier, pengubahan nama kolom, dan pembagian data menjadi data latih dan data uji secara stratifikasi (70:30). Hasil evaluasi menunjukkan bahwa algoritma CatBoost memberikan performa terbaik dibandingkan Random Forest dan Support Vector Machine (SVM), dengan akurasi 88%, macro F1-score 89%,

weighted F1-score 88%, serta AUC macro 97,87% dan weighted 97,44%. Model ini menunjukkan potensi sebagai alat bantu deteksi dini depresi pada mahasiswa tingkat akhir secara akurat dan efisien.

Kata kunci — CatBoost; Depresi; Machine Learning; Mahasiswa Akhir; Prediksi; Skripsi.

I. PENDAHULUAN

Depresi adalah gangguan kesehatan mental yang ditandai dengan perasaan sedih yang mendalam, kehilangan minat dalam aktivitas sehari-hari, serta gangguan pada pikiran, perasaan, dan perilaku individu. Menurut *World Health Organization* (WHO), depresi dapat mempengaruhi kualitas hidup seseorang, menyebabkan penurunan energi, kesulitan tidur, dan perasaan tidak berharga. Depresi dapat terjadi pada berbagai kelompok usia, namun pada mahasiswa, terutama yang berada di tingkat akhir, tekanan akademik sering menjadi faktor pendorong utama. Depresi pada mahasiswa tingkat akhir seringkali menjadi isu yang serius, terutama selama proses penyelesaian skripsi.

Sebanyak 48,9% mahasiswa tingkat akhir mengalami gejala depresi, yang didominasi oleh ketakutan untuk tidak dapat memenuhi tuntutan akademik [1]. Mahasiswa tingkat akhir seringkali tidak menyadari atau kurang peka terhadap stres dan depresi yang mereka alami, meskipun menghadapi tuntutan akademik yang berat, khususnya dalam proses penyelesaian tugas akhir. Hal ini dapat mempengaruhi kesejahteraan psikologis mereka dan berpotensi memperburuk kondisi mental jika tidak segera ditangani [2], [3]. Mahasiswa yang menghadapi beban tugas akademik yang tinggi cenderung mengalami tingkat stres dan depresi yang lebih tinggi [4]. Selain itu, kurangnya dukungan sosial dari teman maupun keluarga diketahui turut berkontribusi terhadap kondisi depresi mahasiswa [5].

Seiring perkembangan teknologi, pendekatan berbasis *machine learning* digunakan untuk mengidentifikasi kondisi mental secara lebih akurat. Salah satu algoritma yang menonjol adalah CatBoost, yaitu metode *gradient boosting* yang dioptimalkan untuk data kategorikal. CatBoost mengatasi permasalahan umum yang terdapat pada algoritma boosting lainnya, seperti *target leakage* dan *prediction shift*, melalui pendekatan *ordered boosting* [6]. Sejumlah studi menunjukkan bahwa CatBoost memiliki kinerja unggul dalam berbagai tugas

klasifikasi, termasuk dalam mengidentifikasi depresi menggunakan data survei berskala besar dengan akurasi dan AUC yang tinggi [7]. Selain itu, performa CatBoost juga terbukti lebih baik dibandingkan berbagai algoritma lainnya dalam konteks klasifikasi prediktif [8].

Namun, masih terbatas penelitian yang secara khusus menerapkan CatBoost untuk mengklasifikasikan tingkat keparahan depresi pada mahasiswa tingkat akhir, kelompok yang memiliki karakteristik stres psikologis tersendiri, seperti tekanan akademik, gangguan kualitas tidur, dan rendahnya dukungan sosial. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan membangun model klasifikasi multikelas berbasis CatBoost, menggunakan data survei yang mencakup variabel psikologis dan demografis, untuk mendukung identifikasi tingkat depresi secara lebih akurat dan tepat sasaran.

A. Depresi

Depresi adalah gangguan mental yang ditandai dengan suasana hati yang terus-menerus sedih, kehilangan minat, atau ketidakmampuan menikmati aktivitas sehari-hari dalam waktu yang cukup lama. Gangguan ini dapat dialami oleh siapa saja, dari anak-anak hingga lansia, dan bila tidak ditangani dengan baik, dapat mengganggu kehidupan sehari-hari serta meningkatkan risiko bunuh diri [9].

B. Prediksi

Prediksi adalah suatu proses yang terstruktur untuk memperkirakan kemungkinan terjadinya suatu peristiwa di masa depan dengan menggunakan informasi yang diperoleh dari data masa lalu dan saat ini. Tujuan dari prediksi adalah untuk meminimalkan kesalahan, yaitu selisih antara hasil yang diprediksi dan kenyataan yang terjadi. Meskipun prediksi tidak selalu menjamin hasil yang pasti mengenai apa yang akan terjadi di masa depan, proses ini berusaha untuk memberikan estimasi yang paling mendekati kenyataan [10].

C. Machine Learning

Machine learning (ML) adalah cabang dari kecerdasan buatan (*AI*) yang memungkinkan komputer untuk belajar dan membuat keputusan tanpa pemrograman eksplisit. Dengan memanfaatkan algoritma dan model statistik, *ML* dapat mengenali pola dalam data dan menghasilkan pengetahuan baru berdasarkan data pelatihan (*training data*). Proses ini sangat bergantung pada kemampuan untuk menganalisis data dan mengadaptasi model matematis untuk menyelesaikan berbagai tugas. Dengan potensi besar dalam inovasi teknologi, *ML* terus berkembang dan menjadi kunci dalam menciptakan sistem cerdas yang dapat belajar dari pengalaman untuk memberikan solusi yang lebih baik di masa depan [11].

D. Categorical Boosting (Boosting)

Categorical Boosting (CatBoost) merupakan algoritma pembelajaran mesin berbasis *gradient boosting* yang dikembangkan oleh Yandex. CatBoost adalah implementasi dari *Gradient Boosting on Decision Trees (GBDT)* yang

menggabungkan teknik *gradient boosting* dengan pohon keputusan untuk menghasilkan model yang canggih dalam berbagai aplikasi dengan data terstruktur [12]. Algoritma *machine learning* seperti *gradient boosting* dan *decision tree* menghasilkan akurasi yang tinggi, dengan CatBoost yang dirancang khusus untuk menangani fitur kategorikal. CatBoost lebih cepat dibandingkan algoritma *boosting* lainnya karena menggunakan pohon simetris dalam prosesnya. Dengan adanya perubahan data dari waktu ke waktu, CatBoost terbukti sebagai algoritma yang paling kompeten, efektif dalam bekerja dengan dataset besar, dan memiliki kebutuhan latensi yang rendah [13].

E. Evaluasi

Evaluasi performa model dilakukan menggunakan berbagai metrik evaluasi yang umum digunakan dalam klasifikasi, seperti *confusion matrix*, *accuracy*, *precision*, *recall*, *f1-score*, serta ROC-AUC.

Confusion matrix adalah alat yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan menyajikan perbandingan antara hasil prediksi model dan nilai aktual dari data [14].

Accuracy adalah metrik yang digunakan untuk mengukur seberapa sering model membuat prediksi yang benar [15].

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

Precision mengukur sejauh mana model akurat dalam memprediksi kelas positif [15].

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

Recall sangat penting karena mengukur sejauh mana model dapat mendeteksi seluruh kasus positif yang sebenarnya ada dalam data [15].

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

F1-score adalah metrik yang menggabungkan presisi dan recall dalam satu nilai, dengan cara menghitung rata-rata harmonik dari keduanya [16].

$$F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Kurva ROC (*Receiver Operating Characteristic*) menunjukkan hubungan antara *True Positive Rate (TPR)* dan *False Positive Rate (FPR)* pada berbagai nilai threshold. AUC (*Area Under the Curve*) menggambarkan kemampuan model membedakan antar kelas, di mana nilai mendekati 1 menunjukkan performa sangat baik, dan nilai mendekati 0,5 setara dengan tebakan acak.

F. Penelitian Terkait

Perbandingan performa algoritma CatBoost dengan berbagai metode pembelajaran mesin lainnya, seperti Logistic Regression, Random Forest, AdaBoost, Decision Tree, Neural Network, Gradient Boosting, dan XGBoost, dalam dua konteks prediktif, yaitu persetujuan pinjaman dan promosi pegawai. Pada kasus persetujuan pinjaman, CatBoost menunjukkan performa tertinggi dengan akurasi 73,2%, F1-score 0,75 dan precision 0,83, mengungguli semua algoritma pembanding. Sementara itu, dalam prediksi promosi pegawai, CatBoost

kembali menempati peringkat teratas bersama *XGBoost*, masing-masing mencatat akurasi 94% dan F1-score 0,95 [8].

Perbandingan performa lima algoritma pembelajaran mesin yaitu Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, dan CatBoost dalam mengidentifikasi depresi berdasarkan data survei NHANES tahun 2018 yang melibatkan 5533 partisipan dewasa. Hasil penelitian menunjukkan bahwa CatBoost memiliki kinerja terbaik dengan akurasi 91,67% dan AUC 0.967. Penelitian ini menyimpulkan bahwa algoritma CatBoost efektif dalam memprediksi risiko depresi, terutama karena kemampuannya mengolah variabel numerik dan kategorikal secara efisien, serta berpotensi menjadi alat bantu skrining yang akurat berbasis data survei berskala besar [7].

Model klasifikasi kondisi kesehatan mental mahasiswa menggunakan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes berdasarkan 100 data survei. KNN menghasilkan akurasi 85% untuk kecemasan, 80% untuk depresi, dan 70% untuk serangan panik, sedangkan Naive Bayes mencatat akurasi 70–75%. Pendekatan ini masih menggunakan algoritma dasar dan belum memanfaatkan model ensemble yang lebih kuat seperti CatBoost [17].

Model prediksi tingkat obesitas menggunakan algoritma CatBoost dengan data dari 2.111 individu yang berasal dari Meksiko, Peru, dan Kolombia. Dataset tersebut mencakup informasi demografi, gaya hidup, dan kondisi kesehatan seperti umur, tinggi badan, berat badan, kebiasaan makan, aktivitas fisik, serta riwayat keluarga terkait obesitas. Model CatBoost dilatih dan dibandingkan dengan algoritma lain seperti Logistic Regression, KNN, Naive Bayes, dan Random Forest. CatBoost menghasilkan akurasi tertinggi sebesar 95,98%, dengan precision 96,08%, recall 95,98%, dan F1-score 96,00%. Analisis *feature importance* mengidentifikasi berat badan, tinggi badan, dan jenis kelamin sebagai fitur yang paling memengaruhi prediksi tingkat obesitas [18].

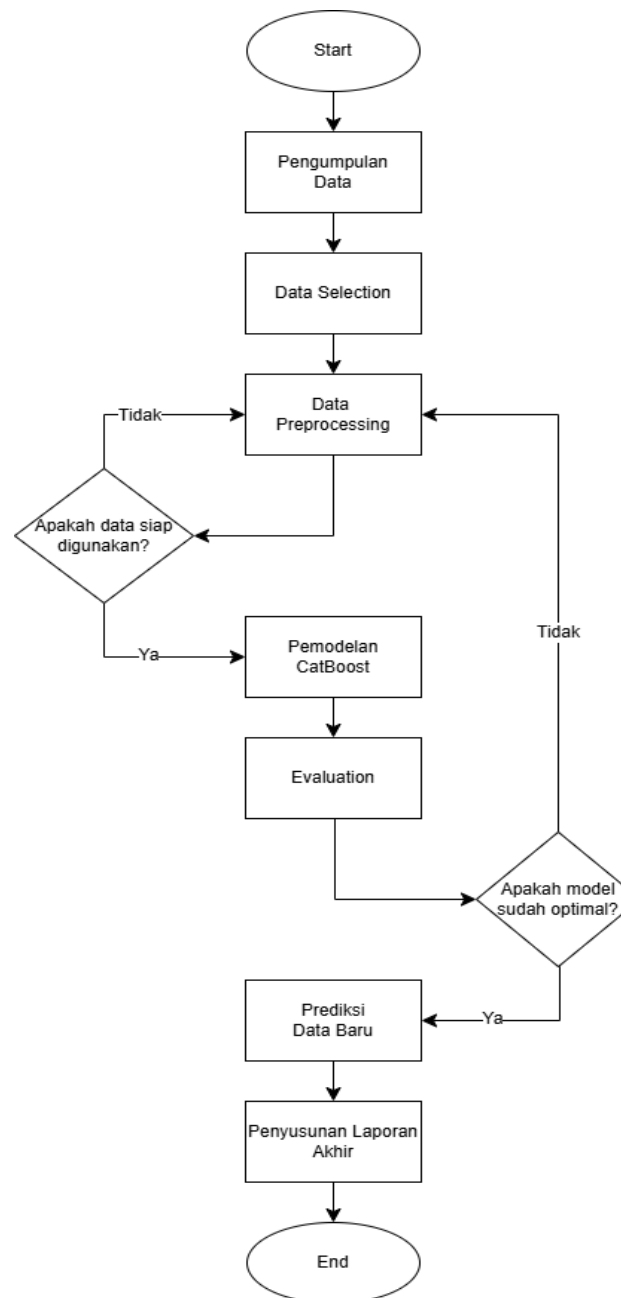
Model klasifikasi multi-kelas menggunakan CatBoost untuk memprediksi tingkat depresi pada pasien penyakit jaringan ikat (CTDs) seperti lupus, *rheumatoid arthritis*, dan *Sjögren's syndrome*. Dataset yang digunakan mencakup 480 pasien yang dikumpulkan secara retrospektif dari Nanjing First Hospital antara tahun 2019–2023. Data yang dikumpulkan mencakup faktor demografis, klinis, psikologis, serta hasil laboratorium. Tingkat depresi diukur dengan kuesioner PHQ-9. Model CatBoost menghasilkan F1-score sebesar 0.879 untuk kategori tidak depresi, 0.627 untuk depresi ringan, dan 0.588 untuk depresi sedang–berat. Model ini juga dianalisis menggunakan SHAP untuk mengevaluasi kontribusi masing-masing fitur terhadap prediksi [19].

Meskipun berbagai studi sebelumnya telah menunjukkan efektivitas algoritma CatBoost dalam tugas klasifikasi, penelitian-penelitian tersebut belum secara spesifik diterapkan pada konteks mahasiswa tingkat akhir yang menghadapi tekanan akademik selama proses penyusunan skripsi. Berbeda dari pendekatan sebelumnya, penelitian ini membangun kontribusi dengan menerapkan algoritma CatBoost dalam klasifikasi multikelas tingkat depresi pada mahasiswa tingkat akhir, menggunakan data survei aktual yang relevan. Dengan pendekatan ini, diharapkan model yang dihasilkan mampu memberikan hasil klasifikasi yang lebih akurat dan aplikatif

sesuai dengan kondisi nyata yang dihadapi oleh mahasiswa pada fase akhir studinya.

II. METODE

Penelitian ini mengikuti tahapan terstruktur berdasarkan alur kerja data mining. Proses dimulai dengan pengumpulan data melalui kuesioner daring kepada mahasiswa tingkat akhir, dilanjutkan dengan seleksi, preprocessing, dan transformasi data. Tahap utama mencakup pembangunan model klasifikasi menggunakan algoritma CatBoost dan evaluasi performa model menggunakan berbagai metrik. Seluruh tahapan disusun secara sistematis agar proses penelitian berjalan terarah dan



Gambar 1. Tahapan Penelitian

menghasilkan model prediksi yang akurat. Adapun gambaran umum tahapan penelitian dapat dilihat pada Gambar 1.

A. Pengumpulan Data

Data dikumpulkan melalui kuesioner daring berbasis Google Form, terdiri dari tiga bagian : (1) data demografis responden, seperti jenis kelamin, usia, program studi, dan semester; (2) kuesioner PHQ-9 (*Patient Health Questionnaire-9*) untuk mengukur gejala utama depresi; (3) pertanyaan tambahan mengenai faktor psikologis seperti beban akademik, stres tenggat, durasi dan kualitas tidur, serta dukungan sosial yang disusun melalui konsultasi dengan pihak yang memiliki kompetensi psikologi terapan dan akademik.

Kuesioner PHQ-9 terdiri dari sembilan pertanyaan yang mengukur gejala depresi dalam dua minggu terakhir, dengan skala respons 0 hingga 3. Skor total kemudian diklasifikasikan ke dalam lima kategori tingkat depresi : Tidak Depresi (0–4), Ringan (5–9), Sedang (10–14), Berat (15–19), dan Sangat Berat (20–27).

Kuesioner disebarakan secara daring kepada mahasiswa tingkat akhir Universitas Sam Ratulangi secara sukarela dan anonim.

Penelitian ini disusun dengan memperhatikan prinsip-prinsip etika penelitian yang berlaku. Sebelum mengisi kuesioner, responden terlebih dahulu membaca bagian deskripsi yang memuat informasi tentang tujuan penelitian, manfaat yang diharapkan, sifat partisipasi yang sukarela, jaminan anonimitas dan kerahasiaan data, serta hak partisipan untuk menghentikan partisipasi kapan saja tanpa konsekuensi. Partisipasi hanya dilanjutkan apabila responden memberikan persetujuan secara sadar setelah memahami informasi tersebut. Peneliti menegaskan bahwa seluruh data yang dikumpulkan tidak mengandung informasi pribadi yang dapat mengidentifikasi individu secara langsung dan hanya digunakan untuk kepentingan akademik.

Untuk menjaga integritas dan validitas data serta mengurangi potensi *social desirability bias*, peneliti secara eksplisit menambahkan penjelasan bahwa tidak ada jawaban yang benar atau salah, dan setiap pertanyaan harus dijawab secara jujur berdasarkan kondisi pribadi masing-masing. Karena pengisian dilakukan secara daring dan anonim, responden memiliki ruang privasi yang lebih besar untuk menjawab secara terbuka dan apa adanya.

B. Data Selection

Pada tahap *data selection*, dipilih data yang relevan dari survei online terkait tingkat depresi mahasiswa akhir. Kolom yang tidak berkontribusi, seperti Timestamp, dihapus. Responden yang terdeteksi menjawab secara tidak serius berdasarkan pola jawaban ekstrem, rendahnya variasi, atau ketidakkonsistenan logis, juga dieliminasi untuk memastikan kualitas dan representativitas data dalam proses analisis.

C. Data Preprocessing

Tahap *preprocessing* diawali dengan pemeriksaan data duplikat dan *missing values*, yang hasilnya menunjukkan tidak

ada data ganda maupun nilai kosong. Penanganan nilai tidak valid dilakukan pada kolom Usia dengan mengonversi input non-numerik menjadi NaN, lalu diimputasi menggunakan modus. Outlier diperiksa menggunakan metode IQR dan batas logis berdasarkan skala kuesioner; nilai-nilai di luar rentang valid, seperti isian semester yang tidak sesuai, diganti dengan modus. Selanjutnya, nama kolom disederhanakan untuk meningkatkan keterbacaan dan konsistensi, misalnya 'Usia (isi dengan angka, cth: 21)' menjadi 'Usia', dan pertanyaan PHQ-9 diubah menjadi PHQ_1 hingga PHQ_9. Setelah pembersihan, data dibagi menjadi 70% data latih dan 30% data uji menggunakan *stratified split* agar proporsi kelas target tetap seimbang.

D. Transformation

Tahap transformasi data difokuskan pada variabel target 'Tingkat_Depresi' yang semula berupa label kategorikal. Karena CatBoost memerlukan target dalam format numerik untuk klasifikasi multikelas, label dikonversi menggunakan LabelEncoder dari `sklearn.preprocessing`. Hasil encoding adalah yakni Tidak Depresi = 0, Ringan = 1, Sedang = 2, Berat = 3, dan Sangat Berat = 4. Transformasi ini memungkinkan model memproses target dengan format yang sesuai untuk evaluasi dan prediksi.

E. Pemodelan CatBoost

Model prediksi tingkat depresi mahasiswa akhir dibangun menggunakan algoritma `CatBoostClassifier`, yang mampu menangani fitur numerik dan kategorikal secara langsung tanpa encoding manual. Parameter model diatur untuk mengontrol kompleksitas dan mencegah overfitting, dengan nilai optimal seperti *depth* = 7, *iterations* = 1000, dan *learning_rate* = 0.03.

Untuk meningkatkan performa, dilakukan *hyperparameter tuning* menggunakan Optuna sebanyak 50 percobaan (*n_trials* = 50), menghasilkan konfigurasi terbaik dengan *iterations* = 1365, *depth* = 8, dan *learning_rate* = 0.0312. Proses ini menggunakan akurasi pada data uji sebagai metrik evaluasi.

Analisis *feature importance* menunjukkan bahwa fitur PHQ-9 mendominasi kontribusi terhadap prediksi, khususnya PHQ_1 (sedih/murung) dan PHQ_2 (kehilangan minat).

F. Evaluasi Model

Model dievaluasi menggunakan metrik klasifikasi multikelas, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*, serta ROC-AUC. Evaluasi dilakukan terhadap data uji untuk menilai performa model dalam memprediksi tingkat depresi mahasiswa. Confusion matrix digunakan untuk memvisualisasikan distribusi prediksi benar dan salah di setiap kelas. Berikut perhitungan metrik evaluasi klasifikasi berdasarkan *confusion matrix* yang diperoleh dari hasil prediksi model terhadap data uji. Model melakukan klasifikasi ke dalam lima kelas, yaitu Tidak Depresi, Ringan, Sedang, Berat, dan Sangat Berat, yang ditentukan berdasarkan hasil pengelompokan skor PHQ-9.

1) Confusion Matrix

Tabel 1. Confusion Matrix Tiap Kelas

		Predicted				
		Tidak Depresi	Ringan	Sedang	Berat	Sangat Berat
Actual	Tidak Depresi	12	0	0	0	0
	Ringan	1	19	2	0	0
	Sedang	0	1	32	5	0
	Berat	0	0	4	24	3
	Sangat Berat	0	0	0	0	29

Sedang	$2 \times \frac{0.842 \times 0.842}{0.842 + 0.842}$	0.842
Berat	$2 \times \frac{0.827 \times 0.774}{0.827 + 0.774}$	0.799
Sangat Berat	$2 \times \frac{0.906 \times 1.00}{0.906 + 1.00}$	0.951

5) Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

2) Precision

$$Precision = \frac{TP}{TP+FP}$$

Tabel 3. Perhitungan Precision per Kategori Depresi

Kategori	Perhitungan Precision	Nilai
Tidak Depresi	$\frac{12}{12+1} = \frac{12}{13}$	0.923
Ringan	$\frac{19}{19+1} = \frac{19}{20}$	0.95
Sedang	$\frac{32}{32+6} = \frac{32}{38}$	0.842
Berat	$\frac{24}{24+5} = \frac{24}{29}$	0.827
Sangat Berat	$\frac{29}{29+3} = \frac{29}{32}$	0.906

3) Recall

$$Recall = \frac{TP}{TP+FN}$$

Tabel 4. Perhitungan Recall per Kategori Depresi

Kategori	Perhitungan Recall	Nilai
Tidak Depresi	$\frac{12}{12+0} = \frac{12}{12}$	1.00
Ringan	$\frac{19}{19+3} = \frac{19}{22}$	0.864
Sedang	$\frac{32}{32+6} = \frac{32}{38}$	0.842
Berat	$\frac{24}{24+7} = \frac{24}{31}$	0.774
Sangat Berat	$\frac{29}{29+0} = \frac{29}{29}$	1.00

4) F1-score

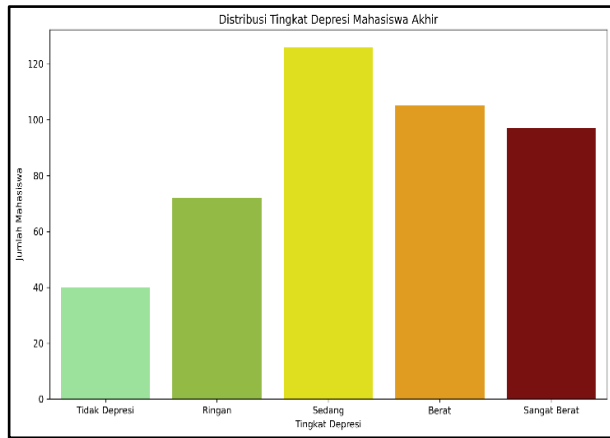
$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Tabel 5. Perhitungan F1-score per Kategori Depresi

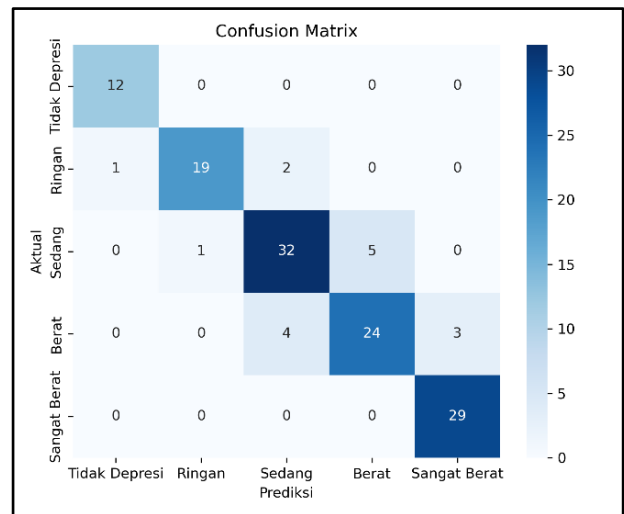
Kategori	Perhitungan F1-score	Nilai
Tidak Depresi	$2 \times \frac{0.923 \times 1.00}{0.923 + 1.00}$	0.96
Ringan	$2 \times \frac{0.95 \times 0.864}{0.95 + 0.864}$	0.905

Tabel 2. Fitur Dataset

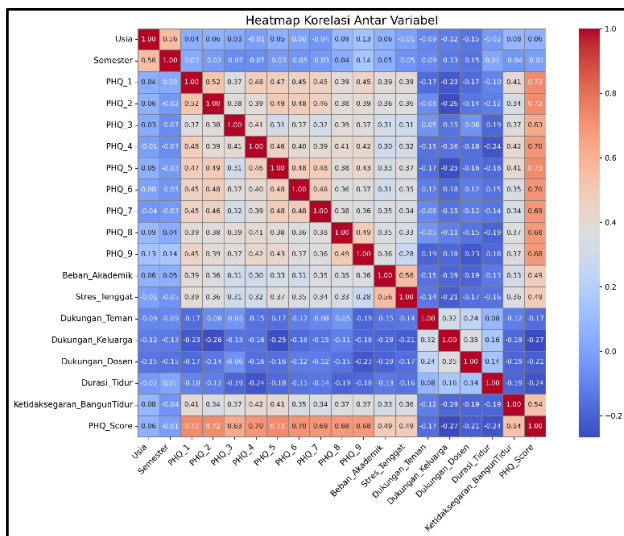
Fitur	Tipe Data	Deskripsi
Jenis Kelamin	Object	Laki-laki / Perempuan.
Usia	Integer	Usia mahasiswa.
Program Studi	Object	Program studi mahasiswa.
Semester	Integer	Semester mahasiswa saat ini.
PHQ_1	Integer	Perasaan sedih, murung, atau putus asa.
PHQ_2	Integer	Kehilangan minat atau kesenangan dalam aktivitas.
PHQ_3	Integer	Kesulitan tidur atau tidur berlebihan.
PHQ_4	Integer	Merasa lelah atau kekurangan energi.
PHQ_5	Integer	Merasa kurang percaya diri atau merasa bersalah.
PHQ_6	Integer	Kesulitan berkonsentrasi.
PHQ_7	Integer	Merasa lambat atau terlalu gelisah.
PHQ_8	Integer	Perubahan nafsu makan.
PHQ_9	Integer	Pikiran untuk menyakiti diri atau bunuh diri.
Beban_Akademik	Integer	Kewalahan dengan beban akademik saat menyusun skripsi.
Stres_Tenggat	Integer	Merasa stres karena tenggat waktu skripsi.
Dukungan_Teman	Integer	Besarnya dukungan dari teman selama pengerjaan skripsi.
Dukungan_Keluarga	Integer	Besarnya dukungan dari keluarga selama pengerjaan skripsi.
Dukungan_Dosen	Integer	Besarnya dukungan dari dosen pembimbing selama pengerjaan skripsi.
Durasi Tidur	Integer	Jumlah jam tidur rata-rata.
Ketidaksegaran_Bangun Tidur	Integer	Merasa tidak segar saat bangun tidur.



Gambar 3. Distribusi Tingkat Depresi



Gambar 4. Confusion Matrix



Gambar 2. Hasil Correlation Matrix dengan Heatmap

Tabel 6. Parameter yang Digunakan Model

No.	Parameter	Nilai
1.	depth	7
2.	iterations	1000
3.	learning_rate	0.03
4.	loss_function	'MultiClass'
5.	early_stopping_rounds	50
6.	random_state	42
7.	verbose	100
8.	cat_features	cat_features

Dataset kemudian dibagi menggunakan metode stratified split menjadi 70% data latih (308 data) dan 30% data uji (132 data), dengan mempertahankan proporsi distribusi kelas agar tetap seimbang. Tabel 5. menunjukkan fitur-fitur yang terdapat dalam dataset penelitian ini

B. Exploratory Data Analysis (EDA)

Tahap Exploratory Data Analysis (EDA) dilakukan untuk memahami pola dan distribusi data setelah pembersihan dan rekayasa fitur. Karena tidak ditemukan nilai kosong, outlier ekstrem, maupun duplikasi signifikan, EDA difokuskan pada analisis data bersih. Hasil EDA memberikan gambaran awal tentang karakteristik demografis dan psikologis responden, serta hubungan antar variabel yang relevan terhadap tingkat depresi, sehingga mendukung proses pemodelan selanjutnya.

1) Tingkat Depresi

Berdasarkan distribusi tingkat depresi pada Gambar 2, mayoritas mahasiswa akhir mengalami depresi dalam berbagai tingkat keparahan. Kategori terbanyak adalah Depresi Sedang (lebih dari 120 mahasiswa), diikuti oleh Depresi Berat (lebih dari 100 mahasiswa) dan Depresi Sangat Berat (sekitar 97 mahasiswa), yang mencerminkan tekanan psikologis yang signifikan selama proses penyelesaian skripsi. Sementara itu, Depresi Ringan dialami oleh sekitar 70 mahasiswa, dan hanya sekitar 40 mahasiswa yang berada pada kategori Tidak Depresi.

$$Accuracy = \frac{12 + 19 + 32 + 24 + 29}{132} = \frac{116}{132} = 0.878$$

III. HASIL DAN PEMBAHASAN

A. Dataset

Dataset dalam penelitian ini diperoleh dari survei terhadap 450 mahasiswa tingkat akhir di Universitas Sam Ratulangi. Setiap entri mewakili satu responden dengan 21 kolom, mencakup data demografis, skor PHQ-9, serta faktor psikologis seperti beban akademik, stres tenggat, dukungan sosial, durasi dan kualitas tidur. Setelah menghapus kolom yang tidak relevan seperti *timestamp*, sebanyak 20 kolom digunakan dalam proses pemodelan.

Dari total data, 10 responden dieliminasi karena tidak memenuhi kriteria validitas, sehingga 440 data digunakan untuk pemodelan. Label target ditentukan dari skor total PHQ-9 dan diklasifikasikan ke dalam lima kategori tingkat depresi: Tidak Depresi (0), Ringan (1), Sedang (2), Berat (3), dan Sangat Berat (4), membentuk kasus klasifikasi multikelas.

Ini menegaskan bahwa sebagian besar mahasiswa tingkat akhir tidak sepenuhnya terbebas dari tekanan mental, dan bahkan banyak di antaranya berada dalam kondisi yang dapat mengganggu konsentrasi, motivasi, serta produktivitas akademik. Tingginya proporsi responden dalam kategori depresi sedang hingga sangat berat menunjukkan urgensi akan perhatian psikologis dan intervensi pendukung yang lebih sistematis di lingkungan perguruan tinggi.

2) Korelasi Antar Variabel

Heatmap menunjukkan korelasi antar variabel dalam dataset, termasuk usia, semester, beban akademik, stres tenggat, dukungan sosial (teman, keluarga, dosen), durasi tidur, ketidaknyamanan saat bangun tidur, serta item-item gejala depresi (PHQ_1 hingga PHQ_9). Pada Gambar 3., terlihat bahwa item-item PHQ-9 yang paling menonjol yang memiliki korelasi positif yang tinggi satu sama lain, serta dengan skor total PHQ, yang mencerminkan konsistensi dalam pengukuran gejala depresi. Usia dan semester memiliki korelasi positif moderat. Beban akademik dan stres tenggat menunjukkan korelasi positif dengan gejala depresi, sedangkan dukungan sosial dan durasi tidur menunjukkan korelasi negatif, menandakan efek protektif terhadap depresi. Sebaliknya, ketidaknyamanan saat bangun tidur berkorelasi positif dengan gejala depresi. Skor total PHQ menunjukkan korelasi sangat kuat dengan seluruh item PHQ, mencerminkan tingkat keparahan depresi secara keseluruhan.

C. Pemodelan CatBoost

Pelatihan model dimulai dengan pengaturan beberapa parameter utama yang mengontrol kompleksitas model, kecepatan belajar, jumlah iterasi, dan fungsi loss yang sesuai untuk klasifikasi multikelas. Mekanisme early stopping digunakan untuk menghentikan pelatihan lebih awal apabila model tidak menunjukkan perbaikan pada data validasi selama beberapa iterasi berturut-turut, sehingga menghindari overfitting dan menghemat waktu komputasi. Selain itu, pengaturan tingkat verbose memungkinkan pemantauan proses pelatihan secara berkala. Model dilatih menggunakan data latih, sekaligus menyertakan parameter `cat_features` untuk memberi tahu algoritma fitur mana yang berupa kategori.

1) Hyperparameter Tuning Menggunakan Optuna

Tabel 7. menyajikan hasil pemilihan hyperparameter terbaik yang diperoleh melalui proses optimasi menggunakan Optuna dengan jumlah percobaan sebanyak 50 kali ($n_trials = 50$). Parameter-parameter ini digunakan untuk membangun model CatBoost dengan kinerja optimal dalam memprediksi tingkat depresi mahasiswa akhir.

2) Analisis Feature Importance

Berdasarkan hasil Feature Importance pada Tabel 8., gejala depresi dari PHQ-9 mendominasi kontribusi terhadap model prediksi. Fitur ‘PHQ_1’ (sedih/murung) dan ‘PHQ_2’ (kurang minat) menjadi prediktor utama dengan skor tertinggi (10.29 dan 9.98), menunjukkan bahwa gejala inti depresi sangat menentukan klasifikasi tingkat depresi. Di luar gejala, ‘Program Studi’ menempati peringkat ketiga (9.15), mengindikasikan pengaruh bidang akademik terhadap kesehatan mental mahasiswa.

Tabel 7. Hasil Hyperparameter Terbaik dari Optuna

No.	Parameter	Nilai
1.	iterations	1365
2.	depth	8
3.	learning_rate	0.0312
4.	l2_leaf_reg	5.5340
5.	random_strength	0.7494
6.	bagging_temperature	0.6474
7.	border_count	132

Tabel 8. Feature Importance

No.	Feature	Importance
1.	PHQ_1	10.290361
2.	PHQ_2	9.985478
3.	Program Studi	9.157297
4.	PHQ_5	8.632822
5.	PHQ_6	8.411613
6.	PHQ_3	8.372773
7.	PHQ_9	8.240717
8.	PHQ_8	7.987224
9.	PHQ_7	7.660854
10.	PHQ_4	6.718332
11.	Dukungan Keluarga	2.336415
12.	Durasi Tidur	1.914967
13.	Stres Tenggat	1.805654
14.	Dukungan Dosen	1.783893
15.	Ketidaksegaran BangunTidur	1.766001
16.	Usia	1.515575
17.	Beban Akademik	1.071248
18.	Dukungan Teman	1.017254
19.	Semester	0.880959
20.	Jenis Kelamin	0.450564

Fitur lain dari PHQ-9 seperti ‘PHQ_5’, ‘PHQ_6’, dan ‘PHQ_9’ turut memberi kontribusi tinggi, menggarisbawahi pentingnya indikator fisik dan emosional. Dari faktor non-gejala, ‘Dukungan Keluarga’ (2.33), ‘Durasi Tidur’ (1.91), dan ‘Stres Tenggat’ (1.80) memiliki pengaruh yang cukup besar, disusul oleh ‘Dukungan Dosen’ dan ‘Ketidaksegaran BangunTidur’. Faktor demografis seperti ‘Usia’, ‘Semester’, dan ‘Jenis Kelamin’ menunjukkan pengaruh yang lebih rendah, namun tetap relevan sebagai pelengkap prediksi.

D. Evaluasi Model

1) Confusion Matrix

Analisis confusion matrix dilakukan untuk mengevaluasi performa klasifikasi model CatBoost pada masing-masing kelas tingkat depresi. Hasilnya menunjukkan bahwa model mampu membedakan kelima kategori depresi dengan baik, dengan jumlah prediksi benar (*true positive*) yang tinggi di setiap kelas. Untuk kelas Tidak Depresi, model berhasil mengklasifikasikan 12 data dengan benar tanpa adanya *false positive* maupun *false negative*, menunjukkan akurasi sempurna pada kategori ini. Pada kelas Ringan, terdapat 19 prediksi benar, hanya dengan 1 kasus *false positive* dan 3 *false negative*. Kinerja model juga cukup baik pada kelas Sedang, dengan 32 data berhasil diprediksi secara akurat, disertai 6 kesalahan pada masing-masing sisi *false positive* dan *false negative*. Kelas Berat

Tabel 9. Hasil Metrik Evaluasi Baseline Model

Class	Precision	Recall	F1-score	Support	AUC
Tidak Depresi	0.92	1.00	0.96	12	0.9993
Ringan	0.95	0.86	0.90	22	0.9802
Sedang	0.84	0.84	0.84	38	0.9608
Berat	0.83	0.77	0.80	31	0.9582
Sangat Berat	0.91	1.00	0.95	29	0.9950
Accuracy			88%	132	
Macro Average	0.89	0.90	0.89	132	0.9787
Weighted Average	0.88	0.88	0.88	132	0.9744

Tabel 10. Hasil Metrik Evaluasi Menggunakan Hyperparameter Tuning

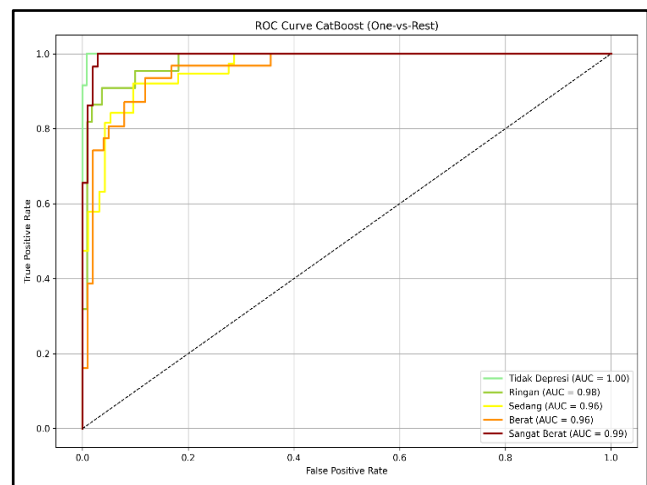
Class	Precision	Recall	F1-score	Support	AUC
Tidak Depresi	0.92	1.00	0.96	12	1.000
Ringan	0.95	0.86	0.90	22	0.9810
Sedang	0.87	0.87	0.87	38	0.9591
Berat	0.86	0.81	0.83	31	0.9598
Sangat Berat	0.91	1.00	0.95	29	0.9946
Accuracy			89%	132	
Macro Average	0.90	0.91	0.90	132	0.9789
Weighted Average	0.89	0.89	0.89	132	0.9744

menunjukkan performa yang seimbang, dengan 24 prediksi benar, 5 *false positive*, dan 7 *false negative*. Sementara itu, kelas Sangat Berat mencatat hasil yang sangat baik, dengan 29 prediksi benar, hanya 3 *false positive*, dan tanpa kesalahan *false negative*.

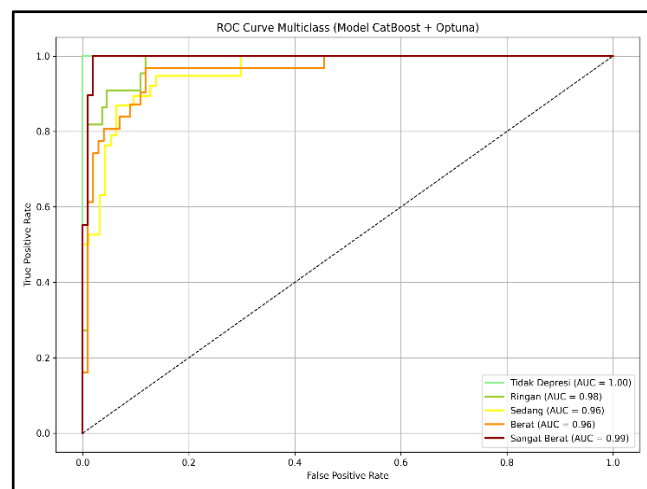
Temuan ini mengindikasikan bahwa model memiliki kemampuan klasifikasi yang stabil dan akurat dalam skenario multikelas, terutama pada kelas-kelas mayoritas. Sebagian besar kesalahan prediksi terjadi antar kelas yang berdekatan secara ordinal, yang masih dapat diterima dalam konteks analisis gejala depresi.

2) Hasil Metrik Evaluasi pada Baseline Model

Evaluasi performa model CatBoost dalam mengklasifikasikan tingkat depresi mahasiswa ditunjukkan pada Tabel 9, dengan metrik akurasi, precision, recall, F1-score, dan AUC. Model mencapai akurasi keseluruhan sebesar 88%, dengan performa terbaik pada kelas “Tidak Depresi” (F1-score 0.96) dan “Sangat Berat” (F1-score 0.95). Kelas “Ringan” juga menunjukkan hasil baik (F1-score 0.90), meskipun recall sedikit lebih rendah. Kinerja pada kelas “Sedang” relatif seimbang (F1-score 0.84), sementara performa terendah terdapat pada kelas “Berat” (F1-score 0.80), mengindikasikan adanya tantangan dalam mendeteksi kategori ini. Nilai macro average F1-score sebesar 0.89 dan rata-rata AUC sebesar 0.9787 menunjukkan bahwa model memiliki performa klasifikasi yang merata dan kemampuan diskriminatif yang sangat baik.



Gambar 5. Kurva ROC Baseline Model



Gambar 6. Kurva ROC menggunakan Hyperparameter Tuning

3) Hasil Metrik menggunakan Hyperparameter Tuning

Setelah dilakukan hyperparameter tuning, model CatBoost menunjukkan peningkatan akurasi dan kemampuan generalisasi pada data uji (132 data). Akurasi keseluruhan meningkat menjadi 89%, dengan nilai F1-score macro average sebesar 0.90 dan weighted average 0.89. Kinerja tertinggi dicapai pada kelas “Tidak Depresi” dan “Sangat Berat” dengan precision dan recall mencapai 1.00, serta nilai AUC sempurna (1.00). Kelas “Sedang” dan “Berat” juga menunjukkan peningkatan presisi dan keseimbangan deteksi. Rata-rata nilai AUC seluruh kelas mencapai 0.9789, menandakan kemampuan diskriminasi model yang sangat baik terhadap masing-masing tingkat depresi.

Dari hasil evaluasi sebelum dan sesudah penerapan hyperparameter tuning, dapat disimpulkan bahwa tuning pada model *CatBoost* memberikan peningkatan performa klasifikasi secara menyeluruh. Dengan kemampuan klasifikasi yang lebih konsisten dalam membedakan tingkat depresi, penerapan hyperparameter tuning dapat direkomendasikan dalam pemodelan prediksi tingkat depresi agar hasil klasifikasi menjadi lebih optimal, akurat, dan andal.

Tabel 11. Perbandingan Baseline Model Catboost dengan Algoritma Lain

Metrik	SVM	Random Forest	CatBoost	CatBoost – SVM	CatBoost – RF
Precision (Macro Average)	75%	79%	89%	14%	10%
Recall (Macro Average)	76%	80%	90%	14%	10%
F1-Score (Macro Average)	75%	79%	89%	14%	10%
AUC (Macro Average)	93.87%	96.72%	97.87%	4%	1.15%
Precision (Weighted Average)	73%	78%	88%	15%	10%
Recall (Weighted Average)	73%	78%	88%	15%	10%
F1-Score (Weighted Average)	72%	77%	88%	16%	11%
AUC (Weighted Average)	92.64%	95.99%	97.44%	4.8%	1.45%
Accuracy	73%	78%	88%	15%	10%

4) Kurva ROC

ROC Curve pada Gambar 5. menunjukkan performa model baseline CatBoost yang sudah sangat baik, dengan nilai AUC (Area Under Curve) berkisar antara 0.96 hingga 0.99 untuk masing-masing kelas tingkat depresi. Nilai ini mencerminkan kemampuan model dalam membedakan kelas-kelas depresi (ringan, sedang, berat) secara cukup akurat. Namun, setelah dilakukan tuning hyperparameter menggunakan Optuna, performa model meningkat secara signifikan sebagaimana terlihat pada Gambar 6. Model hasil tuning menunjukkan nilai AUC yang mencapai 1.00 pada hampir semua kelas, yang berarti model mampu memisahkan setiap tingkat depresi dengan sempurna. Peningkatan ini menunjukkan bahwa proses optimasi menggunakan Optuna berhasil memaksimalkan kinerja model dalam klasifikasi tingkat depresi mahasiswa akhir.

5) Perbandingan Algoritma

Hasil perbandingan tiga algoritma klasifikasi yakni Support Vector Machine, Random Forest, dan CatBoost, menunjukkan bahwa CatBoost memiliki performa paling unggul dalam memprediksi tingkat depresi mahasiswa akhir. Berdasarkan metrik evaluasi macro average, CatBoost mencatat precision sebesar 89%, recall 90%, dan f1-score 89%. Angka ini lebih tinggi dibandingkan dengan Random Forest (79% – 80%) dan SVM (75% – 76%). Begitu pula pada weighted average, CatBoost tetap mencatat skor tertinggi dengan precision, recall, dan f1-score sebesar 88%, menunjukkan kemampuannya dalam mempertahankan performa meskipun distribusi kelas tidak seimbang.

CatBoost juga mencatat nilai AUC tertinggi, yaitu 97.87% untuk macro average dan 97.44% untuk weighted average, yang menunjukkan kemampuan model dalam membedakan antar kelas depresi secara konsisten. Dari segi akurasi, CatBoost

Tabel 12. Perbandingan Efisiensi Komputasi Antar Algoritma

Algoritma	Waktu Latih (detik)	Waktu Prediksi (detik)	Penggunaan Memori (mb)
CatBoost	56.01	0.0064	0.01
SVM	1.10	0.0061	0.26
Random Forest	0.24	0.0147	0.91

Tabel 13. Hasil 5-Fold Cross-Validation Model CatBoost

Fold	Akurasi
1	86.36%
2	72.73%
3	94.32%
4	87.50%
5	77.27%
Rata-rata	83.64%

mencatat skor 88%, lebih tinggi dibandingkan Random Forest (78%) dan SVM (73%). Hasil ini menegaskan bahwa CatBoost tidak hanya unggul dari sisi akurasi, tetapi juga lebih stabil dan andal dalam memprediksi kelas-kelas depresi dengan distribusi data yang kompleks, menjadikannya pilihan terbaik dalam konteks penelitian ini.

Temuan ini sejalan dengan hasil studi sebelumnya [7], dimana CatBoost menunjukkan performa tertinggi dalam mengidentifikasi depresi berdasarkan data survei NHANES, mengungguli algoritma lain seperti Logistic Regression, SVM, dan Random Forest. Meskipun penelitian tersebut tidak secara spesifik menargetkan mahasiswa akhir, kesamaan hasil menunjukkan bahwa CatBoost memiliki kinerja yang stabil pada domain kesehatan mental berbasis survei.

Keunggulan ini tidak lepas dari karakteristik algoritma CatBoost yang menggunakan pendekatan ordered boosting untuk menghindari target leakage dan mampu menangani fitur kategorikal secara langsung tanpa proses encoding tambahan [6]. Dalam konteks penelitian ini, performa model juga diperkuat oleh penambahan fitur-fitur yang relevan dengan kondisi mahasiswa akhir, seperti beban akademik, durasi tidur, kualitas tidur, dan dukungan sosial. Dengan demikian, CatBoost menjadi alternatif yang relevan dan menjanjikan untuk sistem skrining depresi pada mahasiswa tingkat akhir, khususnya dalam konteks penyusunan skripsi.

6) Analisis Efisiensi Komputasi

Selain mengevaluasi performa prediksi dari algoritma klasifikasi, penelitian ini juga mengukur efisiensi komputasi dari masing-masing model. Parameter yang dibandingkan mencakup waktu pelatihan (*training time*), waktu prediksi, dan penggunaan memori saat pelatihan. Hasil analisis disajikan pada Tabel 12.

Dari tabel 12, terlihat bahwa CatBoost membutuhkan waktu pelatihan yang lebih lama dibandingkan algoritma lain, yaitu sekitar 56 detik. Hal ini dapat dipahami mengingat CatBoost merupakan algoritma boosting yang secara internal melakukan iterasi dan optimasi yang kompleks untuk meningkatkan akurasi. Namun, CatBoost tetap menunjukkan efisiensi tinggi dalam waktu prediksi, hanya sekitar 0.0064 detik, serta

penggunaan memori yang sangat rendah (0.01 MB), menunjukkan keunggulannya dalam efisiensi eksekusi setelah model terlatih.

Sebaliknya, SVM memiliki waktu pelatihan yang relatif cepat (1.10 detik) dan waktu prediksi tercepat (0.0061 detik), namun penggunaan memorinya sedikit lebih tinggi dibanding CatBoost. Random Forest menjadi algoritma dengan waktu pelatihan tercepat (0.24 detik), namun memiliki waktu prediksi dan penggunaan memori yang lebih tinggi dibanding CatBoost dan SVM.

Secara keseluruhan, CatBoost memberikan keseimbangan antara performa prediktif dan efisiensi sumber daya, membuatnya ideal untuk diterapkan dalam sistem prediksi real-time berbasis web di lingkungan institusi pendidikan.

7) Validasi Model dengan K-Fold Cross-Validation

Sebagai langkah evaluasi tambahan untuk mengukur kestabilan dan kemampuan generalisasi model terhadap data yang tidak terlihat sebelumnya, dilakukan validasi menggunakan metode 5-Fold Cross-Validation. Teknik ini membagi dataset menjadi lima bagian (fold) yang seimbang, lalu melatih model pada empat fold dan menguji pada satu fold secara bergantian. Pendekatan ini bertujuan untuk mengurangi kemungkinan bias akibat pembagian data dan memberikan gambaran performa model yang lebih robust.

Hasil cross-validation pada Tabel 13 menunjukkan bahwa model CatBoost menghasilkan performa yang relatif stabil di setiap fold, dengan akurasi berkisar antara 72.73% hingga 94.32%. Rata-rata akurasi keseluruhan sebesar 83.64% menunjukkan bahwa model memiliki kemampuan prediktif yang baik serta tidak terlalu bergantung pada distribusi data tertentu, sehingga dapat diandalkan untuk diterapkan pada data baru yang tidak digunakan dalam proses pelatihan.

E. Implementasi Model dalam Aplikasi Web

Model prediksi tingkat depresi mahasiswa akhir berbasis CatBoostClassifier yang dioptimalkan dengan Optuna telah diterapkan dalam aplikasi web Flask. Aplikasi ini memungkinkan pengguna mengunggah data baru, menjalankan klasifikasi, dan melihat hasil prediksi secara langsung. Data yang digunakan berasal dari kuesioner Google Form terpisah dari data latih dan data uji, sehingga mencerminkan kasus nyata.

Model dan encoder disimpan dalam format.pkl untuk efisiensi prediksi. Aplikasi terdiri dari dua halaman: unggah file dan hasil prediksi. Setelah unggah, sistem menampilkan tabel klasifikasi lima kategori depresi (Tidak Depresi, Ringan, Sedang, Berat, dan Sangat Berat) dengan penanda warna, pratinjau data, rekap jumlah kategori, serta fitur unduh hasil. Desain ini memudahkan analisis cepat dan interaktif.

Dari sisi institusi pendidikan tinggi, sistem ini memiliki potensi *cost-benefit* yang tinggi. Biaya pengembangan dan implementasi tergolong rendah karena aplikasi bersifat ringan, tidak membutuhkan lisensi perangkat lunak komersial, dan dapat dijalankan secara lokal tanpa memerlukan infrastruktur server khusus. Dengan klasifikasi yang berjalan otomatis, sistem ini tidak menuntut keterlibatan langsung tenaga profesional secara terus-menerus, sehingga menghemat sumber daya manusia.

Di sisi manfaat, sistem ini berperan dalam mendeteksi risiko depresi mahasiswa secara lebih cepat dan sistematis. Hal ini memungkinkan institusi memberikan intervensi dini dan tepat sasaran, sekaligus mendukung upaya peningkatan retensi studi serta menciptakan lingkungan kampus yang lebih responsif terhadap isu kesehatan mental. Dengan mempertimbangkan efisiensi biaya dan dampak sosial yang signifikan, sistem ini layak diadopsi sebagai bagian dari layanan pendukung institusi terhadap mahasiswa.

IV. KESIMPULAN DAN SARAN

Penelitian ini telah berhasil membangun dan mengevaluasi model prediksi tingkat depresi pada mahasiswa tingkat akhir yang sedang menyusun skripsi, dengan memanfaatkan algoritma CatBoost. Dataset yang digunakan terdiri atas 440 responden valid, yang sebelumnya telah melalui proses seleksi dari total 450 data kuesioner daring. Model menggunakan sembilan item PHQ-9 sebagai label tingkat depresi, serta fitur tambahan berupa faktor psikologis (beban akademik, stres, dukungan sosial, kualitas tidur) dan data demografis (usia, jenis kelamin, program studi, semester). Dengan lima kelas untuk tingkat depresi yakni Tidak Depresi, Ringan, Sedang, Berat, dan Sangat Berat. Seluruh data telah melalui tahap prapemrosesan, termasuk penanganan nilai tidak valid, deteksi dan penanganan outlier dengan metode IQR dan batas logis, penyederhanaan nama kolom, serta pembagian data menggunakan stratified split 70:30 agar distribusi kelas tetap seimbang antara data latih dan data uji.

Model CatBoost telah dibandingkan dengan dua algoritma baseline, yaitu Random Forest dan Support Vector Machine (SVM). Evaluasi performa menunjukkan bahwa CatBoost memberikan hasil terbaik, dengan akurasi 88%, macro F1-score 89%, weighted F1-score 88%, serta AUC tertinggi sebesar 97,87% (macro average) dan 97,44% (weighted average). Berdasarkan confusion matrix dan nilai F1-score per kelas, model mampu mengklasifikasikan tingkat depresi dalam lima kategori secara efektif. Dan untuk menguji kestabilan prediksi, model telah divalidasi menggunakan metode 5-fold stratified cross-validation, yang menghasilkan rata-rata akurasi sebesar 83,64%. Hasil ini menunjukkan bahwa model memiliki konsistensi performa pada data yang tidak terlihat sebelumnya.

Model telah diimplementasikan dalam aplikasi web berbasis Flask yang menyediakan antarmuka interaktif untuk prediksi tingkat depresi. Aplikasi ini juga menggunakan dataset baru yang berbeda dari data latih dan uji, dan hasilnya menunjukkan bahwa model tetap mempertahankan performa prediktif yang baik terhadap data luar sampel. Hal ini memperkuat potensi penerapan model dalam deteksi dini kondisi psikologis mahasiswa.

Namun, penelitian ini memiliki batasan yaitu ruang lingkup data hanya berasal dari mahasiswa Universitas Sam Ratulangi, sehingga generalisasi hasil ke populasi mahasiswa secara nasional perlu dilakukan secara hati-hati. Dengan demikian, pengujian lebih lanjut terhadap data dari berbagai perguruan tinggi akan sangat bermanfaat untuk menilai robustness dan generalizability model secara lebih luas.

Untuk pengembangan selanjutnya, disarankan untuk

memperluas jumlah data dan keragaman karakteristik responden. Selain itu, penambahan fitur-fitur kategorikal yang lebih variatif dan relevan perlu dipertimbangkan untuk meningkatkan kapasitas prediktif model. Hal ini penting mengingat algoritma CatBoost memiliki keunggulan dalam mengolah data kategorikal secara langsung tanpa perlu proses one-hot encoding, sehingga mampu menangkap pola dari data tersebut secara lebih efektif dan efisien.

KUTIPAN

- [1] B. Siskia, U. Subroto, and M. Kurniawati, "Korelasi Resiliensi dan Depresi Mahasiswa Tingkat Akhir Universitas X," vol. 4, pp. 3576–3585, 2024.
- [2] E. Sitepu, Juliana Tampubolon, Sudianto Manulang, and Sisti Nadia Amalia, "Analisis Faktor-Faktor yang Berhubungan dengan Kejadian Stres Pada Mahasiswa Tingkat Akhir S1 Matematika di Universitas Negeri Medan," *Statistika*, vol. 24, no. 1, pp. 93–101, 2024, doi: 10.29313/statistika.v24i1.3257.
- [3] A. H. Z. Tarigan, Y. A. Appulembang, and I. P. Nugroho, "Pengaruh Stress Management Terhadap Resiliensi Mahasiswa Semester Akhir Di Palembang," *J. Bimbingan. Dan Konseling Ar-Rahman*, vol. 7, no. 1, p. 12, 2021, doi: 10.31602/jbkr.v7i1.4989.
- [4] B. B. Pamungkas, "Hubungan Level Stres Akademik Dengan Tingkat Kecemasan Pada Mahasiswa Selama Pandemi Covid-19 Di Indonesia," *Skripsi, Fak. Ilmu Kesehatan. Progr. Stud. S1 Keperawatan Univ. Muhammadiyah Surakarta*, 2021.
- [5] A. Putrama, "Sistem Pakar Diagnosa Depresi Mahasiswa Akhir Beserta Solusi Penanganannya Dengan Metode Certainty Factor," vol. 2, no. 1, pp. 41–49, 2020.
- [6] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: Unbiased boosting with categorical features," *Adv. Neural Inf. Process. Syst.*, no. Section 4, pp. 6638–6648, 2020.
- [7] J. Lu, "Identifying Depression Using Machine Learning," *Front. Comput. Intell. Syst.*, vol. 11, no. 1, pp. 64–69, 2025.
- [8] A. A. Ibrahim, R. L. Ridwan, M. M. Muhammed, R. O. Abdulaziz, and G. A. Saheed, "Comparison of the CatBoost Classifier with other Machine Learning Methods," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 11, pp. 738–748, 2020, doi: 10.14569/IJACSA.2020.0111190.
- [9] J. Chudri, D. Nazma, E. Istriana, Kartini, Junaidi, and V. Hendrilie, "Depresi pada Pekerja: Kenali Gejala dan Pencegahannya," vol. 2, no. 1, pp. 258–267, 2025, doi: <https://doi.org/10.25105/abdimastrimedika.v2i1.22007>.
- [10] D. B. Srisulistiowati, M. Khaerudin, and S. Rejeki, "Sistem Informasi Prediksi Penjualan Alat Tulis Kantor Dengan Metode Fp-Growth (Studi Kasus Toko Koperasi Sekolah Bina Mulia)," *J. Sist. Inf. Univ. Suryadarma*, vol. 8, no. 2, 2021, doi: 10.35968/jsi.v8i2.739.
- [11] R. K. Dinata and N. Hasdyna, *Machine Learning*. Unimal Press, 2020.
- [12] C. P. Ananda, "Machine Learning Untuk Prediksi Gaya Hidup Berdasarkan Socioeconomic Status (SES) Menggunakan Algoritma Catboost Studi Kasus: Mahasiswa UIN Jakarta," *Nucl. Phys.*, vol. 13, no. 1, pp. 104–116, 2023.
- [13] S. Barua, D. Gavandi, P. Sangle, L. Shinde, and J. Ramteke, "Swindle: Predicting the Probability of Loan Defaults using CatBoost Algorithm," *Proc. - 5th Int. Conf. Comput. Methodol. Commun. ICCMC 2021*, no. Iccmc, pp. 1710–1715, 2021, doi: 10.1109/ICCMC51019.2021.9418277.
- [14] H. F. Fadli and A. F. Hidayatullah, "Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM," *Univ. Islam Indones.*, vol. 2, no. No. 1, pp. 1–6, 2021.
- [15] A. R. Dani and I. Handayani, "Klasifikasi Motif Batik Yogyakarta Menggunakan Metode GLCM dan CNN," *J. Teknol. Terpadu*, vol. 10, no. 2, pp. 142–156, 2024, [Online]. Available: <https://journal.nurulfikri.ac.id/index.php/jtt>
- [16] J. C. Obi, "A comparative study of several classification metrics and their performances on data," *World J. Adv. Eng. Technol. Sci.*, vol. 8, no. 1, pp. 308–314, 2023, doi: 10.30574/wjaets.2023.8.1.0054.
- [17] N. Moningka, Raynold, M. Hafidurrohman, W. A. Tri R, and Kusriani, "Klasifikasi Mental Mahasiswa Menggunakan Metode Machine Learning," *J. Quancom*, vol. 1, no. 2, pp. 27–32, 2023, [Online]. Available: <https://www.kaggle.com/datasets/shariful07/student-mental->
- [18] A. Maulana, R. P. F. Afidh, N. B. Maulydia, G. M. Idroes, and S.

Rahimah, "Predicting Obesity Levels with High Accuracy: Insights from a CatBoost Machine Learning Model," *Infolitika J. Data Sci.*, vol. 2, no. 1, pp. 17–27, 2024, doi: 10.60084/ijds.v2i1.195.

- [19] L. Yang *et al.*, "Application of machine learning in depression risk prediction for connective tissue diseases," *Sci. Rep.*, vol. 15, no. 1, p. 1706, 2025, doi: 10.1038/s41598-025-85890-7.



Shintya Preity Aleng lahir di Amurang, Sulawesi Utara pada tanggal 20 September 2003. Penulis mengawali pendidikan formal di Taman Kanak-Kanak GMIM Yordan Lewet pada tahun 2007 hingga 2009. Pada jenjang sekolah dasar, penulis pertama kali menempuh pendidikan di SD Inpres Tumpaan 1 pada tahun 2009 hingga 2011, kemudian melanjutkan dan menyelesaikan pendidikan dasar di SD Negeri Wakan dari tahun 2011 hingga 2015. Pendidikan tingkat menengah pertama ditempuh di SMP Negeri 1 Amurang pada periode 2015 hingga 2018. Selanjutnya, penulis melanjutkan pendidikan menengah atas di SMA Negeri 1 Amurang dari tahun 2018 sampai 2021. Pada tahun 2021, penulis melanjutkan studi Strata-1 di Universitas Sam Ratulangi pada Program Studi Teknik Informatika, Jurusan Elektro, Fakultas Teknik. Selama masa perkuliahan, penulis menjadi anggota *Unsrat IT Community* dan pernah mengikuti Lomba Inovasi Digital Mahasiswa (LIDM) di Divisi Inovasi Pembelajaran Digital Pendidikan. Selain dalam kegiatan akademik dan organisasi, penulis juga pernah menjalani program magang di Kantor Pusat Bank SulutGo, tepatnya pada Divisi Operasional dan Layanan, sebagai bagian dari pengembangan keterampilan praktis dan pemahaman dunia kerja secara langsung.

