JURNAL ILMU KOMPUTER
KHARISMA .TECH

Jurnal KHARISMA Tech | ISSN:1907-2317 | e-ISSN: 2810-0344
https://jurnal.kharisma.ac.id/kharismatech
published by: Pusat Penelitian STMIK KHARISMA Makassar

Volume: 21 no. 02 – September 2026- hlm. 174-188

PEMODELAN *DECISION TREE* YANG INTERPRETATIF UNTUK KLASIFIKASI TINGKAT DEPRESI PADA MAHASISWA MENGGUNAKAN DATA PHQ-9

Oleh:

Munirah[1*], Sunardi[2], Abdul Fadlil[3]
[1,2,3] Program Doktor Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia 55191
[1]Program Studi Teknik Informatika, Universitas Muhammadiyah Ponorogo, Jawa Timur, Indonesia 63471
e-mail: [1]2536083030@webmail.uad.ac.id, [2]sunardi@mti.uad.ac.id, [3]fadlil@mti.uad.ac.id

Abstrak: *Depresi pada mahasiswa merupakan permasalahan kesehatan mental yang semakin meningkat, sementara sebagian besar model machine learning yang memiliki performa tinggi sering kali sulit diinterpretasikan. Oleh karena itu, penelitian ini memanfaatkan Decision Tree sebagai pendekatan Explainable Artificial Intelligence (XAI) yang mampu menghasilkan aturan keputusan yang transparan dan mudah dipahami untuk mengklasifikasikan tingkat depresi mahasiswa. Metode penelitian menggunakan pendekatan supervised learning dengan data Patient Health Questionnaire-9 (PHQ-9) yang direpresentasikan dalam bentuk teks dan dikonversi menjadi numerik menggunakan Label Encoding. Variabel target dibentuk berdasarkan kategori tingkat depresi dari skor total PHQ-9. Untuk meningkatkan performa klasifikasi, dilakukan optimasi parameter Decision Tree dan balancing data menggunakan Synthetic Minority Oversampling Technique (SMOTE). Model dievaluasi menggunakan 10-Fold Cross Validation dengan metrik Accuracy, Precision, Recall, F1-Score, Area Under Curve (AUC), dan Matthews Correlation Coefficient (MCC). Hasil penelitian menunjukkan bahwa model terbaik memperoleh Accuracy sebesar 73,2%, Precision sebesar 73,5%, Recall rata-rata sebesar 73,2%, F1-Score sebesar 73,3%, AUC sebesar 86,8%, dan MCC sebesar 66,5%. Berdasarkan analisis confusion matrix, kategori depresi Severe memiliki tingkat pengenalan tertinggi dengan recall sebesar 89,8%. Hasil penelitian menunjukkan bahwa Decision Tree mampu menghasilkan performa klasifikasi yang baik sekaligus mempertahankan interpretabilitas model melalui aturan keputusan yang mudah dipahami untuk mendukung skrining awal depresi pada mahasiswa.*
Kata kunci: *Depresi, PHQ-9, Decision Tree, Explainable AI, Mahasiswa.*

Abstract: *Depression in college students is a growing mental health problem, while most high-performance machine learning models are often difficult to interpret. Therefore, this study utilizes Decision Tree as an Explainable Artificial Intelligence (XAI) approach capable of generating transparent and easy-to-understand decision rules to classify college students' depression levels. The research method uses a supervised learning approach with Patient Health Questionnaire-9 (PHQ-9) data represented in text form and converted into numeric using Label Encoding. The target variable is formed based on the depression level category from the total PHQ-9 score. To improve classification performance, Decision Tree parameter optimization and data balancing were performed using the Synthetic Minority Oversampling Technique (SMOTE). The model was evaluated using 10-Fold Cross Validation with Accuracy, Precision, Recall, F1-Score, Area Under Curve (AUC), and Matthews Correlation Coefficient (MCC) metrics. The results showed that the best model achieved Accuracy of 73.2%, Precision of 73.5%, Average Recall of 73.2%, F1-Score of 73.3%, AUC of 86.8%, and MCC of 66.5%. Based on the confusion matrix analysis, the Severe depression category had the highest recognition rate with a recall of 89.8%. The results showed that Decision Tree was able to produce good classification performance while maintaining model interpretability through easy-to-understand decision rules to support early screening of depression in college students.*
Keywords: *Depression, PHQ-9, Decision Tree, Explainable AI, Students.*

[*] Corresponding author : Munirah, munirah.mt@umpo.ac.id

Diterima : *Mei, 2026* ***Disetujui :*** *Juni, 2026* ***Dipublikasikan :*** *September, 2026*

1. PENDAHULUAN

Depresi merupakan salah satu gangguan kesehatan mental yang paling banyak dialami masyarakat dunia dan menunjukkan peningkatan yang signifikan pada kalangan mahasiswa akibat tekanan akademik, sosial, dan proses adaptasi lingkungan kampus [1–3]. Untuk mengukur tingkat depresi secara sistematis, berbagai instrumen telah dikembangkan, salah satunya *Patient Health Questionnaire*-9 (PHQ-9) yang terbukti valid dan reliabel untuk mendeteksi serta mengklasifikasikan tingkat keparahan depresi berdasarkan sembilan indikator gejala utama [4–7].

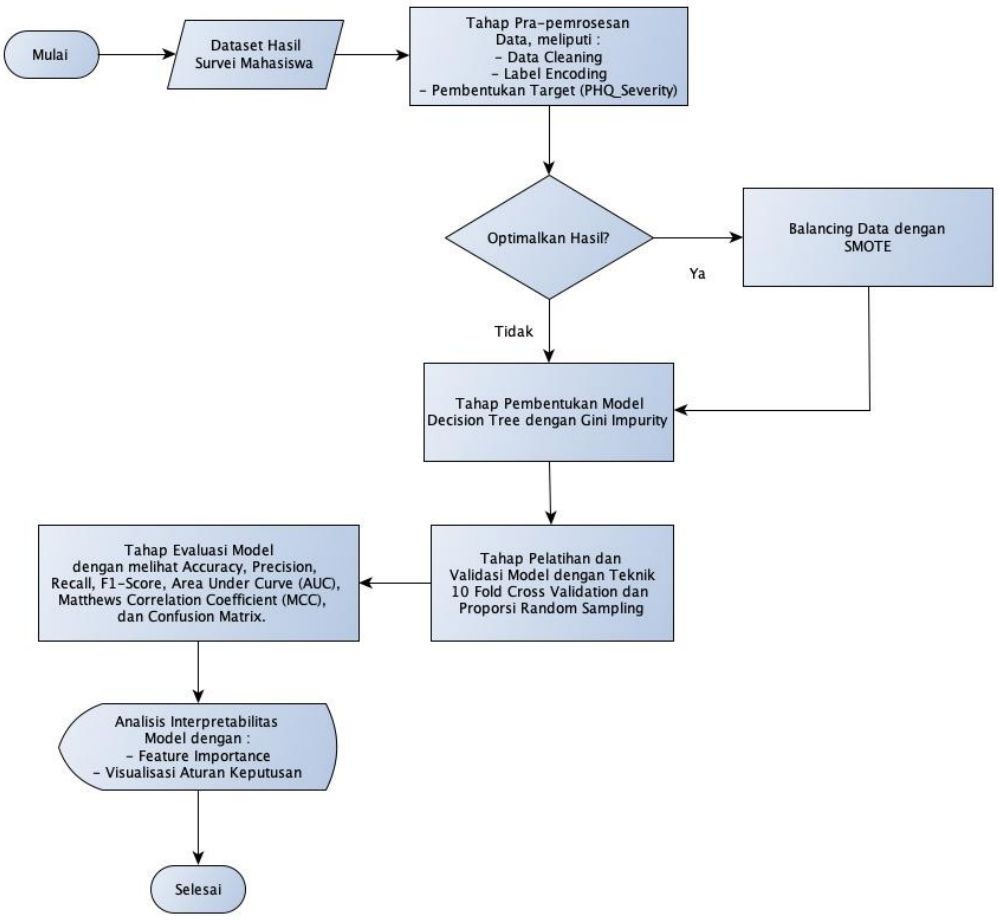
Perkembangan *machine learning* telah mendorong pemanfaatan data kesehatan mental untuk mendukung proses klasifikasi dan prediksi depresi secara lebih efektif [8–10]. Namun, banyak model dengan performa tinggi masih memiliki keterbatasan dalam aspek interpretabilitas sehingga sulit dipahami oleh praktisi maupun pengguna [11,12]. Salah satu metode yang memiliki keunggulan dalam transparansi adalah *Decision Tree* karena mampu menghasilkan aturan keputusan yang mudah dijelaskan dan di visualisasikan [13].

Berbagai penelitian menunjukkan bahwa depresi mahasiswa dipengaruhi oleh faktor psikologis, kualitas tidur, kondisi sosial ekonomi, serta karakteristik individu [14,15]. Selain itu, *machine learning* telah banyak digunakan dalam bidang kesehatan mental untuk mendukung pengambilan keputusan berbasis data [16,17]. Sejalan dengan berkembangnya konsep *Explainable Artificial Intelligence* (XAI) [18,19], penelitian ini mengembangkan model *Decision Tree* yang interpretatif untuk mengklasifikasikan tingkat depresi mahasiswa berdasarkan data PHQ-9. Pendekatan ini diharapkan tidak hanya menghasilkan performa klasifikasi yang baik, tetapi juga mampu memberikan penjelasan yang transparan mengenai faktor-faktor yang memengaruhi tingkat depresi mahasiswa [20].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengevaluasi kemampuan algoritma *Decision Tree* dalam mengklasifikasikan tingkat depresi mahasiswa berdasarkan data PHQ-9 serta mengidentifikasi faktor-faktor yang paling berpengaruh terhadap hasil klasifikasi. Selain itu, penelitian ini juga menganalisis interpretabilitas model melalui *feature importance* dan aturan keputusan (*decision rules*) yang dihasilkan. Dengan demikian, penelitian ini tidak hanya berfokus pada performa klasifikasi, tetapi juga pada kemampuan model dalam memberikan penjelasan yang transparan untuk mendukung proses *skrining* awal depresi pada mahasiswa.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *machine learning* terawasi (*supervised learning*) untuk mengklasifikasikan tingkat depresi mahasiswa berdasarkan data PHQ-9. Penelitian terdiri dari beberapa tahapan utama, yaitu pemrosesan data, pembentukan variabel target, pelatihan model, dan evaluasi kinerja model. Seluruh alur dalam penelitian ini dapat terlihat pada diagram alir di Gambar 1.



Gambar 1. Diagram alir penelitian

2.1. *Dataset*

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari PHQ-9 *Enhanced Student Depression Dataset* yang tersedia pada *Mendeley Data* [22]. Dataset ini berisi 682 responden mahasiswa berusia 17–26 tahun yang mengisi kuesioner *Patient Health Questionnaire-9* (PHQ-9) untuk mengukur tingkat depresi. Selain sembilan indikator utama PHQ-9, dataset juga dilengkapi dengan variabel psikososial tambahan berupa kualitas tidur (*Sleep Quality*), tekanan akademik (*Study Pressure*), dan tekanan finansial (*Financial Pressure*). Seluruh proses pengumpulan data dilakukan di bawah supervisi profesional kesehatan mental dan peneliti klinis dengan memperhatikan aspek etika penelitian serta anonimitas responden.

Dataset terdiri atas 682 responden mahasiswa dengan rentang usia 17–26 tahun dan memiliki 14 atribut yang digunakan sebagai variabel prediktor. Atribut Q1–Q9 merupakan sembilan indikator gejala depresi pada instrumen PHQ-9 yang meliputi kehilangan minat, perasaan sedih atau putus asa, gangguan tidur, kelelahan, perubahan nafsu makan, perasaan gagal, kesulitan konsentrasi, perubahan aktivitas psikomotorik, dan pikiran untuk menyakiti diri sendiri.

Sebelum digunakan dalam proses pemodelan, data terlebih dahulu melalui tahap *preprocessing* yang mencakup pembersihan data, transformasi kategori menjadi numerik

menggunakan *Label Encoding*, dan pembentukan variabel target berupa tingkat depresi (*PHQ_Severity*). Variabel target dibagi menjadi lima kategori, yaitu *Minimal, Mild, Moderate, Moderately Severe,* dan *Severe* berdasarkan skor total PHQ-9.

Tabel 1. Cuplikan Dataset

	Age	Gender	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Sleep Quality	Study Pressure	Financial Pressure	PHQ_ Total	PHQ_ Severity
1	22	Male	2	0	0	0	0	0	0	2	0	Good	Good	Average	4	Minimal
2	25	Male	0	0	3	3	3	0	2	2	2	Worst	Bad	Average	15	Moderately severe
3	22	Female	0	0	0	0	0	0	1	0	0	Average	Bad	Average	1	Minimal
4	18	Female	3	3	0	3	2	0	0	0	0	Average	Bad	Worst	11	Moderate
5	24	Male	0	0	0	0	0	0	0	2	0	Good	Average	Good	2	Minimal
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
680	23	Male	0	0	2	2	0	1	1	1	2	Bad	Good	Good	9	Mild
681	19	Male	0	1	0	0	0	1	0	0	0	Average	Average	Bad	2	Minimal
682	21	Male	3	1	3	3	1	2	0	3	2	Average	Bad	Bad	18	Moderately severe

Tabel 1 menampilkan cuplikan 5 data pertama dan 3 data terakhir yang digunakan dalam penelitian. Cuplikan tersebut menunjukkan struktur *dataset* yang terdiri atas atribut demografis, faktor eksternal, dan indikator gejala depresi yang menjadi dasar dalam proses klasifikasi tingkat depresi mahasiswa.

Distribusi data berdasarkan kategori tingkat depresi ditunjukkan pada Tabel 2 yang terlihat bahwa jumlah data pada masing-masing kategori depresi tidak sama. Ketidakseimbangan distribusi kelas ini berpotensi menyebabkan model lebih dominan mempelajari pola pada kelas mayoritas dibandingkan kelas minoritas. Oleh karena itu, penelitian ini menerapkan *Synthetic Minority Oversampling Technique* (SMOTE) untuk meningkatkan representasi kelas minoritas sehingga proses pelatihan model dapat berlangsung lebih seimbang.

Tabel 2. Distribusi Kelas Sebelum an Sesudah SMOTE

Kategori Depresi	Jumlah Data Sebelum SMOTE	Jumlah Data Setelah SMOTE
Minimal	206	206
Mild	155	206
Moderate	128	206
Moderately Severe	125	206
Severe	68	206
Total	**682**	**1030**

2.2. Pra-pemrosesan Data

Beberapa atribut dalam *dataset* masih berbentuk kategori teks, seperti *Gender*, *Sleep Quality, Study Pressure,* dan *Financial Pressure*. Oleh karena itu, dilakukan transformasi menggunakan *Label Encoding* agar seluruh atribut dapat diproses oleh algoritma *Decision Tree*. Sementara itu, atribut PHQ-9 (Q1–Q9) telah tersedia dalam bentuk skor numerik sehingga tidak memerlukan proses *encoding* tambahan.

2.3. Pembentukan Variabel Target

Variabel target dalam penelitian ini dibentuk berdasarkan nilai PHQ_Total yang telah tersedia dalam dataset. Skor tersebut kemudian dikategorikan menjadi lima tingkat depresi sesuai standar klinis sebagai berikut:

1. *Minimal*: 0–4
2. *Mild* (Ringan): 5–9
3. *Moderate* (Sedang): 10–14
4. *Moderately Severe* (Sedang-Berat): 15–19
5. *Severe* (Berat): 20–27

Kategori ini digunakan sebagai label dalam proses klasifikasi. Variabel PHQ_Total hanya digunakan untuk membentuk kategori target (PHQ_*Severity*) dan tidak digunakan sebagai fitur pada proses pelatihan model. Dengan demikian, fitur yang digunakan dalam klasifikasi terdiri atas indikator PHQ-9 (Q1–Q9) dan variabel eksternal yang telah ditentukan sebelumnya.

2.4. Model *Decision Tree*

Decision Tree merupakan salah satu algoritma klasifikasi yang membentuk struktur pohon berdasarkan serangkaian aturan keputusan. Setiap *node* pada pohon merepresentasikan atribut yang digunakan untuk membagi data, sedangkan setiap cabang menunjukkan hasil dari proses pemisahan tersebut. Keunggulan utama *Decision Tree* adalah kemampuannya menghasilkan model yang mudah dipahami dan diinterpretasikan.

Dalam penelitian ini, pemilihan atribut pada setiap *node* dilakukan menggunakan kriteria *Gini Impurity* untuk mengukur tingkat ketidakmurnian suatu *node* berdasarkan distribusi kategori tingkat depresi yang terdapat pada himpunan data D. Semakin kecil nilai *Gini Impurity* maka semakin homogen distribusi kelas pada *node* tersebut. Oleh karena itu, algoritma *Decision Tree* akan memilih atribut yang menghasilkan penurunan nilai *Gini Impurity* terbesar sehingga proses pemisahan data menjadi lebih optimal. Nilai *Gini Impurity* dapat dihitung menggunakan Persamaan (1) [21].

$$Gini\ (\mathrm{D}) = 1 - \sum_{i=1}^{c} [P(i|D)]^2 \qquad (1)$$

dengan:

- Gini (D) : nilai *impurity* pada himpunan data atau node D
- P(i|D) : proporsi data yang termasuk ke dalam kelas ke-i pada node D
- i : indeks kelas
- c : jumlah kelas

Nilai *Gini Impurity* berkisar antara 0 hingga 1. Nilai 0 menunjukkan bahwa seluruh data pada *node* berasal dari satu kategori depresi yang sama (*pure node*), sedangkan nilai yang lebih besar menunjukkan bahwa data pada *node* masih terdiri atas beberapa kategori depresi yang berbeda. Dalam penelitian ini, atribut yang menghasilkan nilai *Gini Impurity* paling rendah dipilih sebagai atribut pemisah (*split attribute*) karena mampu menghasilkan *node* yang lebih homogen terhadap kategori tingkat depresi mahasiswa.

Proses klasifikasi menggunakan algoritma *Decision Tree* diawali dengan memasukkan *dataset* hasil *preprocessing* ke dalam model sebagai data pelatihan. Selanjutnya, nilai *Gini Impurity* dihitung pada setiap atribut untuk menentukan atribut yang mampu memberikan pemisahan data terbaik. Atribut dengan penurunan nilai *impurity* terbesar dipilih sebagai *root node*. Setelah itu, data dibagi ke dalam beberapa subset berdasarkan atribut yang terpilih.

Pada setiap subset, proses perhitungan *Gini Impurity* dilakukan kembali secara rekursif untuk memilih atribut terbaik pada *node* berikutnya. Proses pembentukan cabang akan terus berlanjut hingga memenuhi kriteria penghentian (*stopping criteria*), seperti tercapainya nilai *maximum depth, minimum leaf, minimum split*, atau seluruh data dalam *node* berasal dari kelas yang sama. *Node* akhir (*leaf node*) yang terbentuk kemudian digunakan sebagai keputusan klasifikasi yang merepresentasikan kategori tingkat depresi mahasiswa, yaitu *Minimal*, *Mild*, *Moderate*, *Moderately Severe*, dan *Severe*.

2.5. Pelatihan dan Validasi Model

Model *Decision Tree* dievaluasi menggunakan dua pendekatan, yaitu 10-*Fold Cross Validation* dan *Random Sampling* dengan beberapa variasi proporsi data *training* dan *testing* sebesar 40%:60%, 50%:50%, 60%:40%, 70%:30%, dan 80%:20%. Pendekatan ini digunakan untuk menganalisis pengaruh pembagian data terhadap performa model.

2.6. Evaluasi Kinerja Model

Kinerja model dievaluasi menggunakan beberapa metrik klasifikasi, yaitu *Accuracy*, *Precision, Recall, F1-Score, Area Under Curve* (AUC), *Matthews Correlation Coefficient* (MCC), dan *Confusion Matrix*. *Accuracy* digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data pengujian. *Precision* menunjukkan tingkat ketepatan model dalam mengklasifikasikan suatu kategori, sedangkan *Recall* digunakan untuk mengukur kemampuan model dalam mengenali seluruh data yang termasuk ke dalam kategori tertentu.

F1-Score digunakan sebagai ukuran keseimbangan antara *Precision* dan *Recall*, terutama pada kasus klasifikasi multikelas. Selain itu, AUC digunakan untuk mengevaluasi kemampuan model dalam membedakan kategori depresi yang berbeda, sedangkan MCC digunakan untuk mengukur kualitas klasifikasi secara keseluruhan dengan mempertimbangkan seluruh elemen *confusion matrix*.

Untuk memperoleh evaluasi yang lebih komprehensif, penelitian ini juga menggunakan *Confusion Matrix* guna menganalisis pola kesalahan klasifikasi pada masing-masing kategori depresi. Analisis ini memungkinkan identifikasi kategori yang paling mudah maupun paling sulit dikenali oleh model.

Karena distribusi data antar kategori depresi tidak seimbang, penelitian ini menerapkan *Synthetic Minority Oversampling Technique* (SMOTE) sebelum proses pelatihan model. Teknik ini digunakan untuk meningkatkan representasi kelas minoritas sehingga model dapat mempelajari pola klasifikasi secara lebih seimbang dan menghasilkan performa yang lebih baik.

Selain *Decision Tree* sebagai model utama, penelitian ini juga menggunakan *Random Forest* sebagai model pembanding (*benchmark*) untuk mengevaluasi *trade-off* antara akurasi dan interpretabilitas model.

2.7. Analisis Interpretabilitas

Analisis interpretabilitas dilakukan untuk memahami kontribusi masing-masing fitur dalam proses klasifikasi. Hal ini dilakukan melalui:

- Analisis *feature importance* untuk mengidentifikasi variabel yang paling berpengaruh
- Visualisasi struktur pohon keputusan untuk mengekstrak aturan keputusan (*decision rules*)

Pendekatan ini sejalan dengan konsep XAI yang menekankan transparansi dalam model *machine learning*.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Evaluasi Model

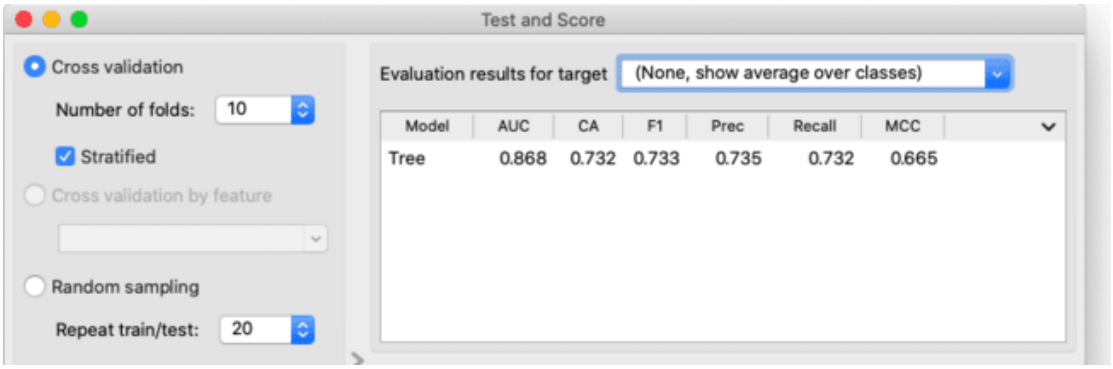
Model *Decision Tree* dievaluasi menggunakan beberapa skenario pengujian untuk memperoleh konfigurasi parameter yang optimal. Evaluasi dilakukan menggunakan *10-Fold Cross Validation* dan *Random Sampling* dengan berbagai proporsi data *training* dan *testing*. Selain itu, dilakukan pula proses *balancing* data menggunakan SMOTE untuk mengatasi ketidakseimbangan distribusi kelas pada *dataset*. Hasil pengujian menunjukkan bahwa konfigurasi terbaik diperoleh pada parameter *maximum depth* 10, *minimum leaf* 2, dan *minimum split* 5 setelah penerapan SMOTE. Pada konfigurasi tersebut, model mencapai akurasi sebesar 73,2% menggunakan metode *10-Fold Cross Validation*. Hasil ini menunjukkan peningkatan dibandingkan model sebelum *balancing data* yang hanya memperoleh akurasi sebesar 66,6%.

Tabel 3. Hasil Eksperimen pada *Decision Tree*

Eksp	Depth	Min Leaf	Min Split	Cross Validation 10 Fold (%)	Proporsi Random Sampling (%)				
					40 : 60	50 : 50	60 : 40	70 : 30	80 : 20
DT-1	5	2	5	60,7	59,8	59,1	60,2	60,1	58,0
DT-2	7	2	5	65,0	62,0	62,5	62,0	63,4	63,1
DT-3	10	2	5	65,2	62,1	62,8	62,3	64,1	63,8
DT-4	10	5	10	66,6	60,4	62,0	62,1	62,9	61,5
DT-5	15	5	10	66,6	60,4	62,0	62,1	62,9	61,5
Setelah dilakukan optimalisasi dan *balancing data* menggunakan SMOTE									
DT-1	5	2	5	65,1	63,1	63,1	63,1	63,8	64,0
DT-2	7	2	5	72,1	66,8	67,6	68,6	69,9	71,6
DT-3	10	2	5	73,2	67,0	68,1	70,3	71,5	72,8
DT-4	10	5	10	70,3	66,2	66,1	68,8	70,2	70,1
DT-5	15	5	10	69,8	66,2	66,1	68,8	70,3	70,1

Tabel 3 menunjukkan hasil eksperimen beberapa konfigurasi parameter *Decision Tree* menggunakan berbagai metode pembagian data. Secara umum, penggunaan data *balancing*

melalui SMOTE memberikan peningkatan performa pada seluruh skenario pengujian. Selain itu, metode *Cross Validation* menghasilkan nilai akurasi yang lebih stabil dibandingkan *Random Sampling* karena seluruh data memperoleh kesempatan menjadi data pelatihan dan data pengujian secara bergantian.

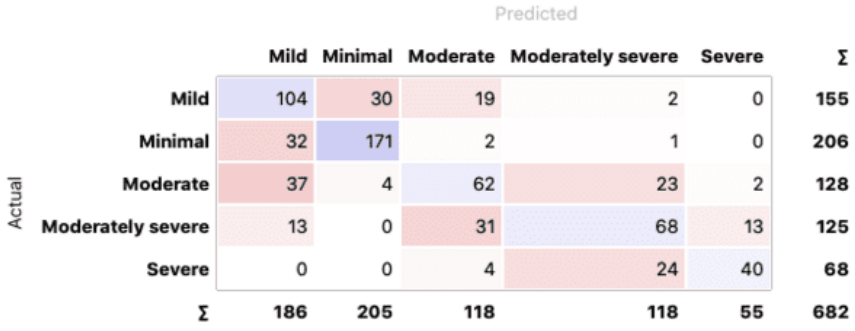


Gambar 2. Hasil Evaluasi Model

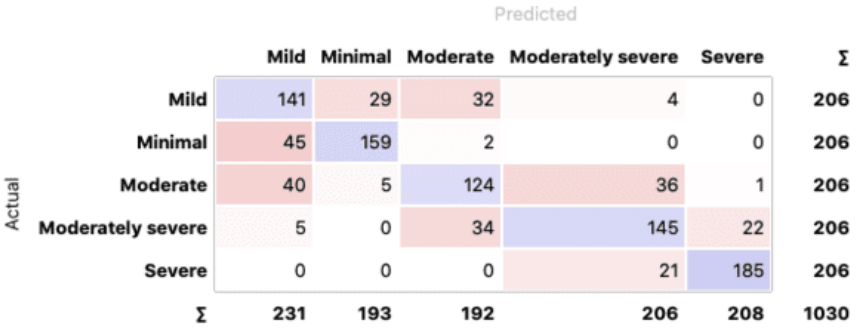
Berdasarkan Gambar 2, model *Decision Tree* yang dievaluasi menggunakan 10-*Fold Cross Validation* dan *dataset* yang telah diseimbangkan menggunakan SMOTE menghasilkan akurasi sebesar 73,2%. Nilai *precision* sebesar 73,5%, *recall* sebesar 73,2%, dan *F1-score* sebesar 73,3% menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengklasifikasikan tingkat depresi mahasiswa secara konsisten pada seluruh kategori. Selain itu, nilai AUC sebesar 86,8% mengindikasikan bahwa model memiliki kemampuan diskriminasi yang sangat baik dalam membedakan kategori depresi yang berbeda. Nilai MCC sebesar 66,5% juga menunjukkan adanya hubungan yang cukup kuat antara hasil prediksi model dengan kelas sebenarnya. Secara keseluruhan, hasil evaluasi menunjukkan bahwa model *Decision Tree* mampu memberikan performa klasifikasi yang baik sekaligus mempertahankan sifat interpretatif yang menjadi keunggulan utama model.

3.2. Analisis *Confusion Matrix*

Sebelum proses *balancing* data menggunakan SMOTE, *dataset* yang digunakan dalam penelitian ini terdiri atas 682 data dengan distribusi kelas yang tidak seimbang. Setelah penerapan SMOTE, jumlah data meningkat menjadi 1030 data karena dilakukan proses *oversampling* pada kelas-kelas minoritas hingga distribusi setiap kategori depresi menjadi lebih seimbang. Oleh karena itu, penelitian ini menampilkan *confusion matrix* sebelum dan sesudah penerapan SMOTE untuk menunjukkan pengaruh *balancing data* terhadap performa klasifikasi model *Decision Tree*.



(a) Sebelum SMOTE



	Mild	Minimal	Moderate	Moderately severe	Severe	Σ
Mild	141	29	32	4	0	**206**
Minimal	45	159	2	0	0	**206**
Moderate	40	5	124	36	1	**206**
Moderately severe	5	0	34	145	22	**206**
Severe	0	0	0	21	185	**206**
Σ	**231**	**193**	**192**	**206**	**208**	**1030**

(b) Sesudah SMOTE

Gambar 3. Hasil *confusion matrix* sebelum dan sesudah SMOTE

Gambar 3 menunjukkan *confusion matrix* hasil klasifikasi tingkat depresi mahasiswa menggunakan model *Decision Tree* setelah dilakukan optimasi parameter dan *balancing data* menggunakan SMOTE. Pada *confusion matrix* sebelum SMOTE di Gambar 3(a), total data yang dievaluasi adalah 682 sampel sesuai jumlah data asli. Sedangkan pada *confusion matrix* setelah SMOTE di Gambar 3(b), total data menjadi 1030 sampel sebagai hasil proses *oversampling* pada kelas minoritas. Berdasarkan hasil tersebut, model menunjukkan performa yang cukup baik dalam mengklasifikasikan lima kategori tingkat depresi yang terdiri atas *Minimal, Mild, Moderate, Moderately Severe,* dan *Severe*.

Kategori *Severe* memiliki performa terbaik dengan nilai *recall* sebesar 89,8%, yang menunjukkan bahwa sebagian besar mahasiswa dengan tingkat depresi berat berhasil diidentifikasi dengan benar oleh model. Selain itu, kategori *Minimal* juga menunjukkan performa yang tinggi dengan *recall* sebesar 82,4%. Hasil ini mengindikasikan bahwa model mampu membedakan kategori depresi ekstrem dengan cukup baik.

Sebaliknya, kategori *Mild* memiliki nilai *recall* terendah sebesar 61,0%, sedangkan kategori *Moderate* dan *Moderately Severe* masing-masing sebesar 64,6% dan 70,4%. Berdasarkan *confusion matrix*, sebagian besar kesalahan klasifikasi terjadi pada kategori-kategori tersebut. Sebagai contoh, sebanyak 17,3% data pada kategori *Moderate* diklasifikasikan sebagai *Mild* dan 17,5% lainnya diklasifikasikan sebagai *Moderately Severe*. Selain itu, 17,7% data pada kategori *Moderately Severe* diprediksi sebagai *Moderate*.

Fenomena tersebut menunjukkan adanya kemiripan karakteristik antar kategori depresi yang berdekatan. Kondisi ini dapat dijelaskan oleh mekanisme penentuan tingkat depresi pada PHQ-9 yang menggunakan rentang skor yang relatif berdekatan. Sebagai contoh, skor 9 dikategorikan sebagai *Mild*, sedangkan skor 10 sudah termasuk kategori *Moderate*. Demikian pula skor 14 termasuk kategori *Moderate*, sedangkan skor 15 masuk kategori *Moderately Severe*. Perbedaan skor yang relatif kecil menyebabkan batas antar kategori menjadi kurang tegas sehingga model mengalami kesulitan dalam memisahkan kategori-kategori menengah secara sempurna.

Meskipun masih terdapat kesalahan klasifikasi pada kategori *Mild, Moderate,* dan *Moderately Severe*, sebagian besar kesalahan tersebut terjadi pada kategori yang berdekatan secara klinis. Dengan demikian, hasil prediksi yang dihasilkan model masih dapat memberikan informasi yang bermanfaat untuk mendukung proses deteksi awal tingkat depresi mahasiswa.

Secara keseluruhan, hasil *confusion matrix* menunjukkan bahwa model *Decision Tree* yang dikombinasikan dengan teknik SMOTE mampu menghasilkan performa klasifikasi yang cukup baik dengan kemampuan yang tinggi dalam mengenali kategori depresi berat dan depresi minimal.

3.3. Analisis Faktor Yang Mempengaruhi Model

Nilai akurasi yang diperoleh oleh model *Decision Tree* tidak hanya dipengaruhi oleh karakteristik algoritma yang digunakan, tetapi juga oleh beberapa faktor lain, seperti metode pembagian data, distribusi kelas pada dataset, serta konfigurasi parameter model. Untuk memahami faktor-faktor yang mempengaruhi performa klasifikasi, dilakukan serangkaian eksperimen menggunakan berbagai skenario pengujian yang meliputi variasi proporsi data *training* dan *testing*, *balancing data* menggunakan SMOTE, serta optimasi parameter *Decision Tree*.

3.3.1. Pengaruh Pembagian Data (*Train-Test Split*)

Berdasarkan hasil pengujian pada Tabel 2 sebelumnya, proporsi pembagian data *training* dan *testing* memberikan pengaruh terhadap performa model. Pada seluruh skenario pengujian, peningkatan jumlah data *training* cenderung meningkatkan nilai akurasi yang diperoleh. Pada model *Decision Tree* setelah penerapan SMOTE, akurasi meningkat dari 64,0% pada skenario *random sampling* 80:20 menjadi 73,2% pada skenario 10-*Fold Cross Validation*. Hasil ini menunjukkan bahwa semakin banyak data yang digunakan untuk proses pelatihan, semakin baik kemampuan model dalam mempelajari pola klasifikasi tingkat depresi mahasiswa.

Selain itu, metode 10-*Fold Cross Validation* menghasilkan performa yang lebih stabil dibandingkan *random sampling* karena seluruh data memperoleh kesempatan menjadi data pelatihan dan data pengujian secara bergantian. Oleh karena itu, penelitian ini menggunakan hasil 10-*Fold Cross Validation* sebagai hasil evaluasi utama model.

3.3.2. Pengaruh Data *Imbalance* dan SMOTE

Distribusi kelas pada dataset menunjukkan adanya ketidakseimbangan jumlah data antar kategori depresi. Sebelum proses *balancing*, kategori *Severe* memiliki jumlah data yang lebih sedikit dibandingkan kategori lainnya sehingga berpotensi menyebabkan model lebih sulit mempelajari pola pada kelas minoritas. Untuk mengatasi masalah tersebut, penelitian ini menerapkan SMOTE.

Hasil pengujian menunjukkan bahwa penerapan SMOTE meningkatkan akurasi *Decision Tree* dari 66,6% menjadi 73,2%. Peningkatan ini menunjukkan bahwa ketidakseimbangan data memberikan pengaruh terhadap performa model. Dengan menambah representasi data pada kelas minoritas, model menjadi lebih mampu membentuk aturan keputusan yang lebih baik dan meningkatkan kemampuan klasifikasi pada seluruh kategori depresi.

3.3.3. Pengaruh Parameter *Decision Tree*

Optimasi parameter dilakukan dengan memvariasikan nilai *maximum depth, minimum leaf,* dan *minimum split* pada algoritma *Decision Tree*. Hasil pengujian menunjukkan bahwa peningkatan kedalaman pohon dari 5 menjadi 10 mampu meningkatkan akurasi model secara

signifikan. Namun, peningkatan kedalaman hingga 15 tidak menghasilkan peningkatan performa yang berarti.

Hasil pengujian menunjukkan bahwa konfigurasi terbaik diperoleh pada parameter *maximum depth* 10, *minimum leaf* 2, dan *minimum split* 5 setelah penerapan SMOTE dengan akurasi sebesar 73,2%. Temuan ini menunjukkan bahwa peningkatan kompleksitas pohon hingga tingkat tertentu mampu meningkatkan kemampuan model dalam membentuk aturan klasifikasi yang lebih akurat. Namun, peningkatan kedalaman pohon yang lebih tinggi tidak selalu menghasilkan peningkatan performa yang signifikan.

3.3.4. Perbandingan *Decision Tree* dan *Random Forest*

Sebagai model pembanding, penelitian ini juga menguji algoritma *Random Forest* menggunakan skenario evaluasi yang sama. Hasil pengujian menunjukkan bahwa *Random Forest* menghasilkan performa yang lebih tinggi dibandingkan *Decision Tree* pada seluruh skenario pengujian. Setelah penerapan SMOTE, *Random Forest* mencapai akurasi sebesar 82,6%, sedangkan *Decision Tree* mencapai 73,2% seperti yang di tampilkan pada Tabel 4.

Tabel 4. Hasil Eksperimen untuk *Random Forest*

Eksp	Depth	Min Leaf	Min Split	Cross Validation 10 Fold (%)	Proporsi Random Sampling				
					40 : 60	50 : 50	60 : 40	70 : 30	80 : 20
RF-1	5	2	5	75,8	70,1	71,6	71,8	72,3	73,0
RF-2	7	2	5	76,7	71,0	71,2	73,6	73,8	74,1
RF-3	10	2	5	77,0	70,3	70,1	72,4	72,6	73,5
RF-4	10	5	10	74,5	69,9	71,7	73,3	74,0	73,7
RF-5	15	5	10	75,5	69,6	71,4	73,3	73,5	74,0
Setelah dilakukan optimalisasi dan *balancing data* menggunakan SMOTE									
RF-1	5	2	5	80,3	75,5	77,4	78,4	80,7	80,5
RF-2	7	2	5	79,9	75,2	76,9	78,6	79,4	81,4
RF-3	10	2	5	82,6	75,9	77,1	78,4	80,6	81,1
RF-4	10	5	10	82,0	75,7	77,3	78,9	80,6	80,1
RF-5	15	5	10	80,5	75,2	77,5	78,8	80,2	80,9

Keunggulan *Random Forest* disebabkan oleh mekanisme *ensemble* yang menggabungkan banyak pohon keputusan sehingga mampu mengurangi variansi model dan meningkatkan kemampuan generalisasi. Meskipun demikian, *Decision Tree* tetap memiliki keunggulan dalam aspek interpretabilitas karena menghasilkan aturan keputusan yang lebih mudah dipahami dan dijelaskan. Oleh karena itu, *Decision Tree* tetap dipilih sebagai model utama dalam penelitian ini karena tujuan penelitian tidak hanya berfokus pada akurasi, tetapi juga pada kemampuan menjelaskan faktor-faktor yang mempengaruhi tingkat depresi mahasiswa.

3.4. Analisis *Feature Importance*

Tabel 5. Hasil *Feature Importance*

	Feature	Importance
3	Q2	0,126277
10	Q9	0,117910

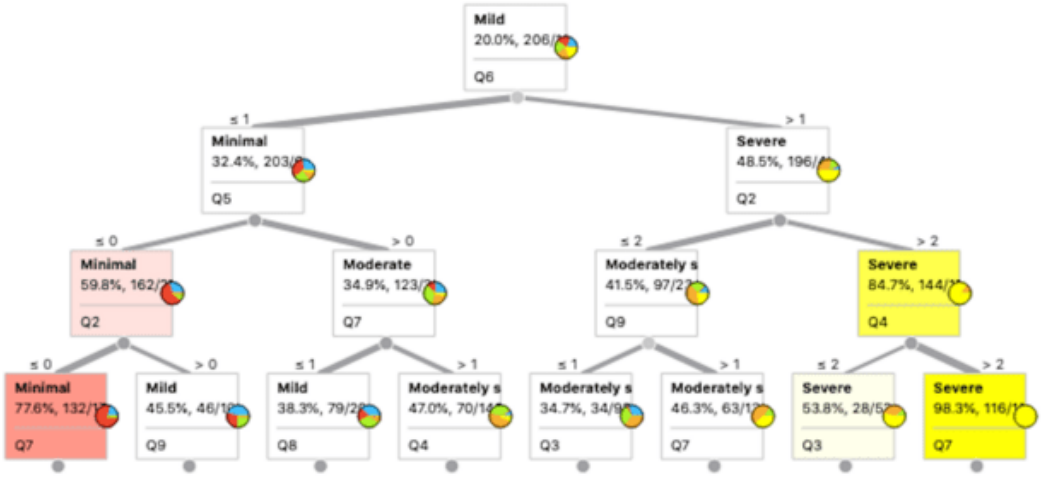
6	Q5	0,103344
2	Q1	0,100351
9	Q8	0,095757
4	Q3	0,080953
8	Q7	0,071342
5	Q4	0,071080
7	Q6	0,067451
0	Age	0,049139
13	Financial Pressure	0,047855
12	Study Pressure	0,043249
11	Sleep Quality	0,018723
1	Gender	0,006570

Hasil analisis *feature importance* pada Tabel 5 menunjukkan bahwa variabel Q2 memiliki kontribusi terbesar dalam proses klasifikasi tingkat depresi mahasiswa dengan nilai *importance* sebesar 0,1263. Variabel tersebut diikuti oleh Q9 (0,1179), Q5 (0,1033), Q1 (0,1004), dan Q8 (0,0958). Hasil ini menunjukkan bahwa indikator-indikator utama dalam instrumen PHQ-9 memiliki pengaruh dominan dalam menentukan tingkat depresi mahasiswa.

Q2 yang merepresentasikan perasaan sedih, murung, atau putus asa menjadi fitur yang paling berpengaruh dalam model. Selain itu, Q9 yang berkaitan dengan pikiran untuk menyakiti diri sendiri juga memiliki kontribusi yang tinggi. Temuan ini menunjukkan bahwa gejala emosional dan psikologis yang diukur oleh PHQ-9 memiliki kemampuan diskriminatif yang kuat dalam membedakan tingkat depresi mahasiswa.

Sebaliknya, variabel eksternal seperti usia, tekanan finansial, tekanan akademik, kualitas tidur, dan jenis kelamin memiliki nilai *importance* yang relatif lebih rendah. Hal ini mengindikasikan bahwa proses klasifikasi lebih banyak dipengaruhi oleh indikator depresi yang terdapat pada PHQ-9 dibandingkan faktor-faktor demografis maupun lingkungan yang digunakan dalam penelitian ini.

3.5. Analisis Aturan Keputusan



Gambar 4. Visualisasi Aturan Keputusan

Gambar 4 menunjukkan struktur pohon keputusan yang dihasilkan oleh model *Decision Tree*. Meskipun Q2 memiliki nilai *feature importance* tertinggi, visualisasi pohon menunjukkan bahwa Q6 dipilih sebagai *root node* atau atribut pertama yang digunakan dalam proses klasifikasi. Hal ini menunjukkan bahwa Q6 merupakan atribut yang paling efektif untuk melakukan pemisahan awal data, sedangkan Q2 memberikan kontribusi yang lebih besar secara keseluruhan terhadap proses klasifikasi pada seluruh cabang pohon.

Pada cabang kiri pohon, mahasiswa dengan nilai Q6 yang rendah cenderung diklasifikasikan ke dalam kategori depresi *Minimal*. Sebaliknya, pada cabang kanan pohon, mahasiswa dengan nilai Q6 yang lebih tinggi diproses lebih lanjut menggunakan atribut Q2 dan Q7. Hasil visualisasi menunjukkan bahwa kombinasi skor tinggi pada Q6, Q2, dan Q7 mengarahkan model pada kategori *Severe* dengan tingkat kemurnian klasifikasi mencapai 98,3%.

Pemilihan atribut Q6 sebagai *root node* menunjukkan bahwa atribut tersebut menghasilkan penurunan nilai *Gini Impurity* terbesar pada tahap awal pembentukan pohon keputusan. Dengan kata lain, Q6 merupakan atribut yang paling efektif dalam memisahkan data berdasarkan kategori tingkat depresi. Selanjutnya, atribut Q2, Q5, Q7, dan Q9 dipilih pada cabang-cabang berikutnya untuk meningkatkan homogenitas data pada setiap *node* hingga diperoleh aturan klasifikasi yang optimal.

Keunggulan utama *Decision Tree* terletak pada kemampuannya menghasilkan aturan keputusan yang mudah dipahami. Sebagai contoh, mahasiswa dengan skor tinggi pada Q6, Q2, dan Q7 cenderung diklasifikasikan ke dalam kategori depresi berat (*Severe*), sedangkan mahasiswa dengan skor rendah pada Q6 dan Q5 cenderung berada pada kategori depresi minimal (*Minimal*). Dengan demikian, model tidak hanya menghasilkan prediksi, tetapi juga mampu memberikan penjelasan yang transparan mengenai faktor-faktor yang mempengaruhi hasil klasifikasi.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model *Decision Tree* yang interpretatif untuk mengklasifikasikan tingkat depresi mahasiswa berdasarkan data PHQ-9. Hasil pengujian menunjukkan bahwa optimasi parameter dan penerapan teknik balancing data menggunakan SMOTE mampu meningkatkan performa model secara signifikan. Model terbaik diperoleh pada konfigurasi *maximum depth* 10, *minimum leaf* 2, dan *minimum split* 5, dengan akurasi sebesar 73,2% menggunakan 10-*Fold Cross Validation*.

Selain memperoleh *Accuracy* sebesar 73,2%, model juga menghasilkan *Precision* sebesar 73,5%, *Recall* sebesar 73,2%, *F1-Score* sebesar 73,3%, AUC sebesar 86,8%, dan MCC sebesar 66,5%. Hasil *confusion matrix* menunjukkan bahwa kategori depresi Severe memiliki performa klasifikasi terbaik dengan *recall* sebesar 89,8%, sedangkan kesalahan klasifikasi sebagian besar terjadi pada kategori *Moderate* dan *Moderately Severe* yang memiliki karakteristik gejala yang berdekatan.

Analisis lebih lanjut menunjukkan bahwa ketidakseimbangan distribusi kelas mempengaruhi performa model. Penerapan SMOTE berhasil meningkatkan akurasi dari

66,6% menjadi 73,2%, sehingga model menjadi lebih mampu mengenali kategori depresi yang sebelumnya kurang terwakili. Selain itu, pengujian terhadap berbagai proporsi pembagian data menunjukkan bahwa metode 10-*Fold Cross Validation* menghasilkan performa yang lebih stabil dibandingkan *random sampling*.

Meskipun *Random Forest* menghasilkan akurasi yang lebih tinggi, yaitu sebesar 82,6%, *Decision Tree* tetap dipilih sebagai model utama karena memiliki keunggulan dalam aspek interpretabilitas. Struktur pohon keputusan yang dihasilkan memungkinkan proses klasifikasi dipahami secara transparan sehingga lebih sesuai untuk mendukung pendekatan XAI dalam skrining awal depresi mahasiswa.

Penelitian ini memiliki keterbatasan karena kategori PHQ_*Severity* dibentuk berdasarkan skor total PHQ-9. Oleh karena itu, model yang dihasilkan lebih ditujukan untuk mengeksplorasi pola klasifikasi dan interpretabilitas faktor-faktor yang berkontribusi terhadap tingkat depresi mahasiswa, serta bukan dimaksudkan sebagai alat diagnosis klinis yang berdiri sendiri.

Untuk penelitian selanjutnya, disarankan menambahkan variabel eksternal lain seperti dukungan sosial, aktivitas organisasi, penggunaan media sosial, prestasi akademik, riwayat kesehatan mental, serta faktor lingkungan keluarga guna meningkatkan kemampuan prediksi model. Selain itu, penerapan metode seleksi fitur dan pendekatan XAI yang lebih lanjut dapat menjadi arah pengembangan penelitian berikutnya.

DAFTAR PUSTAKA

[1] World Health Organization, Depression, WHO Fact Sheets (2021). URLhttps://www.who.int/news-room/fact-sheets/detail/ depression

[2] R. Beiter, et al., The prevalence and correlates of depression, anxiety, and stress in a sample of college students, Journal of Affective Disorders (2015). doi:10.1016/j.jad.2014.10.054

[3] T. M. Evans, et al., Evidence for a mental health crisis in graduate education, Nature Biotechnology (2018). doi:10.1038/nbt.4089. URL https://doi.org/10.1038/nbt.4089

[4] K. Kroenke, R. L. Spitzer, J. B. Williams, The phq-9: validity of a brief depression severity measure, Journal of General Internal Medicine (2001). doi:10.1046/j.1525-1497.2001.016009606.x.

[5] R. L. Spitzer, K. Kroenke, J. B. W. Williams, Validation and utility of a self-report version of prime-md, JAMA (1999). doi:10.1001/jama. 282.18.1737.

[6] B. Levis, et al., Accuracy of patient health questionnaire-9 (phq-9) for screening to detect major depression, BMJ (2019). doi:10.1136/bmj.l146. URL https://doi.org/10.1136/bmj.l1476

[7] American Psychiatric Association, Diagnostic and statistical manual of mental disorders (dsm-5)Official manual, often via APA or library (2013). doi:10.1176/appi.books.9780890425596.

[8] A. Rajkomar, J. Dean, I. Kohane, Machine learning in medicine, New England Journal of Medicine (2019). doi:10.1056/NEJMra1814259. URL https://doi.org/10.1056/NEJMra1814259

[9] A. B. R. Shatte, et al., Understanding the effect of adding auto-mated and human coaching to a mobile health physical activity app for afghanistan and iraq veterans: Protocol for a randomized con-trolled trial of the stay strong intervention, JMIR Mental Health (2019). doi:10.2196/12526. URL https://doi.org/10.2196/12526

[10] E. J. Topol, High-performance medicine: the convergence of human and artificial intelligence, Nature Medicine (2019). doi:10.1038/ s41591-018-0300-7. URL https://doi.org/10.1038/s41591-018-0300-7

[11] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable ml, arXiv (2017). doi:10.48550/arXiv.1702.08608. URL https://arxiv.org/abs/1702.08608

[12] C. Molnar, Interpretable machine learningOnline book (free) (2020). URL https://christophm.github.io/interpretable-ml-book/
[13] L. Breiman, J. Friedman, R. A. Olshen, C. J. Stone, Classification and Regression Trees, 1984, classic book, often available via library or archive. doi:10.1201/9781315139470. URL https://doi.org/10.1201/9781315139470
[14] R. H. Salk, J. S. Hyde, L. Y. Abramson, Gender differences in de-pression in representative national samples: Meta-analyses of diag-noses and symptoms, Psychological Bulletin 143 (8) (2017) 783–822. doi:10.1037/bul0000102. URL https://doi.org/10.1037/bul0000102
[15] J. Dinis, M. Bragança, Quality of sleep and depression in college students: A systematic review, Sleep Science (2018). doi:10.5935/ 1984-0063.20180045. URL https://doi.org/10.5935/1984-0063.20180045
[16] D. Bzdok, A. Meyer-Lindenberg, Machine learning for precision psychia-try, Biological Psychiatry (2018). doi:10.1016/j.bpsc.2017.11.007. URL https://doi.org/10.1016/j.bpsc.2017.11.007
[17] I. Goodfellow, Y. Bengio, A. Courville, Deep learningMIT Press book (2016). URL https://www.deeplearningbook.org/
[18] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems (NeurIPS) (2017). URL https://papers.nips.cc/paper/ 7062-a-unified-approach-to-interpreting-model-predictions
[19] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair,R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable ai for trees, Nature Machine Intelligence 2 (1) (2020) 56–67. doi:10.1038/s42256-019-0138-9.URL https://doi.org/10.1038/s42256-019-0138-9
[20] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learn-ingSecond edition 2009 (most cited) (2001). URL https://hastie.su.domains/ElemStatLearn/
[21] Rokach, L., & Maimon, O. (2005). *Top-Down Induction of Decision Trees Classifiers—A Survey.* IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews, 35(4), 476–487.
[22] M.A.I.A. Miraz, S. Mondal, and S. Ahmed, PHQ-9 Enhanced Student Depression Dataset, Mendeley Data, Version 6, 2025. Doi: 10.17632/kkzjk253cy.6