



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PERBANDINGAN KINERJA ALGORITMA *ENSEMBLE* LEARNING DALAM KLASIFIKASI TINGKAT KESEHATAN MENTAL PADA MAHASISWA

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer pada
Program Studi Sistem Informasi



Oleh:

RIFSYA AULIA

12250321571



UIN SUSKA RIAU

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2026

LEMBAR PERSETUJUAN

PERBANDINGAN KINERJA ALGORITMA *ENSEMBLE*
***LEARNING* DALAM KLASIFIKASI TINGKAT KESEHATAN**
MENTAL PADA MAHASISWA

TUGAS AKHIR

Oleh:

RIFSYA AULIA
12250321571

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 15 Januari 2026

Ketua Program Studi



Anraini, S.Kom., M.Eng., Ph.D.
NIP. 198408212009012008

Pembimbing



Mustakim, ST., M.Kom.
NIP. 198807022025211003

LEMBAR PENGESAHAN

PERBANDINGAN KINERJA ALGORITMA *ENSEMBLE* *LEARNING* DALAM KLASIFIKASI TINGKAT KESEHATAN MENTAL PADA MAHASISWA

TUGAS AKHIR

Oleh:

RIFSYA AULIA

12250321571

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 09 Januari 2026

Pekanbaru, 09 Januari 2026

Mengesahkan,

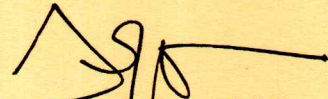
Dekan



Dr. Yuslenita Muda, S.Si., M.Sc.

NIP. 197701032007102001

Ketua Program Studi



Angraini, S.Kom., M.Eng., Ph.D.

NIP. 198408212009012008

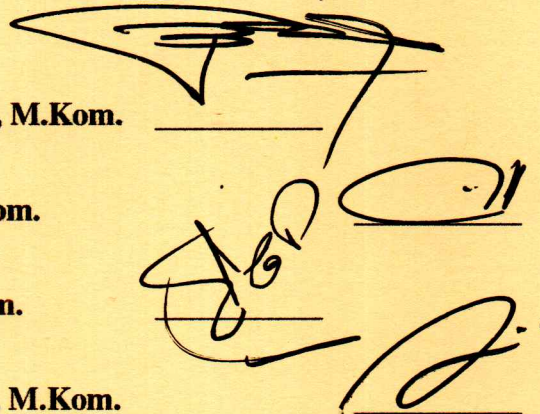
DEWAN PENGUJI:

Ketua : Eki Saputra, S.Kom., M.Kom.

Sekretaris : Mustakim, ST., M.Kom.

Anggota 1 : M. Afdal, ST., M.Kom.

Anggota 2 : Inggih Permana, ST., M.Kom.



Lampiran Surat :

Nomor : Nomor 25/2021

Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Rifsya Aulia

NIM : 12250321571

Tempat/Tgl. Lahir : Emplasment/09 Februari 2004

Fakultas/Pascasarjana : Sains dan Teknologi

Prodi : Sistem Informasi

Judul Skripsi :

PERBANDINGAN KINERJA ALGORITMA *ENSEMBLE LEARNING* DALAM
KLASIFIKASI TINGKAT KESEHATAN MENTAL PADA MAHASISWA

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan Skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas dari plagiat.
4. Apa bila dikemudian hari terbukti terdapat plagiat dalam Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 15 Januari 2026

ig membuat pernyataan



Rifsya Aulia

NIM: 12250321571



LEMBAR PERNYATAAN

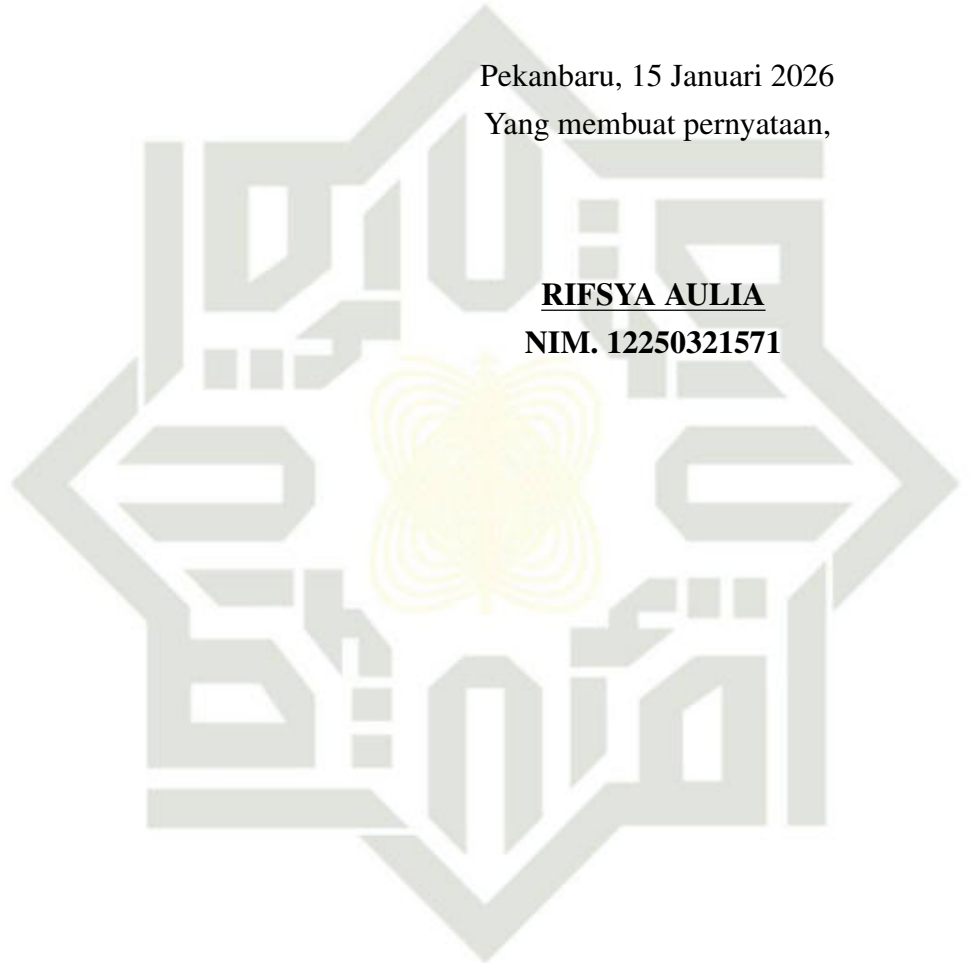
Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 15 Januari 2026

Yang membuat pernyataan,

RIFSYA AULIA

NIM. 12250321571



UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah wa syukurillah, peneliti panjatkan ke hadirat Allah SWT atas segala nikmat, rahmat, dan karunia-Nya yang senantiasa tercurah tanpa kekurangan sedikit pun, sehingga peneliti dapat menyelesaikan Tugas Akhir ini dengan baik. Shalawat serta salam senantiasa peneliti haturkan kepada Nabi Muhammad SAW dengan melafadzkan *Allahumma shalli 'ala Sayyidina Muhammad wa 'ala ali Sayyidina Muhammad*, semoga kita semua senantiasa memperoleh *syafa'at* beliau di dunia maupun di akhirat. *Aamiin ya Rabbal 'Alamiin*.

Tugas Akhir ini peneliti persembahkan secara khusus kepada Ayah tercinta Syarifuddin Lubis dan Ibu tercinta Indah Saedah Munthe, yang senantiasa memberikan doa, kasih sayang, dukungan, serta motivasi tanpa henti kepada peneliti selama menempuh pendidikan hingga akhirnya dapat menyelesaikan studi Strata Satu dengan baik. Keberhasilan penyusunan Tugas Akhir ini peneliti dedikasikan sebagai bentuk tanggung jawab dan bakti kepada kedua orang tua. Peneliti juga mengucapkan terima kasih kepada keluarga tercinta, yaitu Abang Ridho Alfadillah, Kakak Dinda Alawiyah dan Kakak Annisya Putri, yang selalu memberikan semangat, dukungan moral, serta energi positif kepada peneliti selama proses penyusunan Tugas Akhir ini.

Ucapan terima kasih dan penghargaan setinggi-tingginya peneliti sampaikan kepada dosen pembimbing Tugas Akhir, Bapak Mustakim, ST., M.Kom., atas kesabaran, bimbingan, arahan, serta waktu yang telah diberikan kepada peneliti dalam menyelesaikan Tugas Akhir ini. Semoga Allah SWT senantiasa melimpahkan keberkahan, rahmat, dan kesehatan kepada beliau serta kepada kita semua.

Akhir kata, peneliti berharap semoga Tugas Akhir ini dapat memberikan manfaat dan menjadi langkah awal bagi peneliti untuk terus belajar, berkarya, dan berkontribusi secara positif di masa yang akan datang. *Aamiin ya Rabbal 'Alamiin*.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau



KATA PENGANTAR

Puji dan syukur peneliti panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan hidayah-Nya, sehingga Peneliti dapat menyelesaikan Tugas Akhir yang berjudul “Perbandingan Kinerja Algoritma *Ensemble Learning* dalam Klasifikasi Tingkat Kesehatan Mental pada Mahasiswa” dengan baik. Penyusunan Tugas Akhir ini merupakan salah satu syarat untuk menyelesaikan pendidikan Strata Satu (S1) pada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi.

Peneliti menyadari bahwa penyusunan Tugas Akhir ini tidak terlepas dari bantuan, bimbingan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini peneliti menyampaikan ucapan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Ibu Prof. Dr. Hj. Leny Nofianti, M.S., S.E., M.Si., Ak., CA., selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Ibu Dr. Yuslenita Muda, S.Si., M.Sc., selaku Dekan Fakultas Sains dan Teknologi.
3. Ibu Angraini, S.Kom., M.Eng., Ph.D., selaku Ketua Program Studi Sistem Informasi.
4. Ibu Dr. Rice Novita, S.Kom., M.Kom., selaku Sekretaris Program Studi Sistem Informasi.
5. Bapak Saide, S.Kom., M.Kom., ICTPM., M.IM., Ph.D., selaku Dosen Pembimbing Akademik, yang telah memberikan bimbingan, arahan, serta nasihat akademik kepada Peneliti selama masa perkuliahan, sehingga membantu Peneliti dalam menjalani proses studi dengan terarah dan optimal.
6. Bapak Mustakim, ST., M.Kom., selaku Dosen Pembimbing, yang dengan penuh kesabaran, ketulusan, dan dedikasi tinggi telah meluangkan waktu, tenaga, serta pemikiran dalam memberikan bimbingan, arahan, motivasi, dan masukan yang sangat berarti kepada Peneliti.
7. Bapak Eki Saputra, S.Kom., M.Kom., selaku Ketua Sidang, yang telah memimpin jalannya sidang Tugas Akhir dengan penuh kebijaksanaan dan profesionalisme, serta memberikan arahan, penilaian, dan masukan yang objektif dan membangun.
8. Bapak M. Afdal, ST., M.Kom., selaku Penguji I, yang telah memberikan saran, kritik, dan masukan ilmiah yang sangat berharga guna penyempurnaan Tugas Akhir ini.
9. Bapak Inggih Permana, ST., M.Kom., selaku Penguji II, yang telah mem-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

berikan evaluasi, saran, serta masukan yang membangun demi kesempurnaan Tugas Akhir ini.

10. Seluruh Bapak dan Ibu Dosen Program Studi Sistem Informasi, yang telah dengan tulus memberikan ilmu pengetahuan, bimbingan, serta pengalaman akademik yang sangat berharga kepada Peneliti selama proses perkuliahan, sehingga menjadi bekal penting dalam penyelesaian studi dan penyusunan Tugas Akhir ini.
11. Ibu Fara Ulfa, M.Psi., Psikolog., selaku Dosen Fakultas Psikologi sekaligus pakar, yang telah memberikan masukan, arahan, serta pandangan keilmuan yang sangat berarti dan membantu Peneliti dalam memperkuat landasan teoritis dan analisis pada penelitian ini.
12. *Puzzle Research Data Technology* (PREDATECH) sebagai wadah pengembangan riset dan keilmuan telah memberikan lingkungan yang kondusif bagi peneliti dalam memperluas wawasan, meningkatkan keterampilan analisis data, serta memperdalam pemahaman terhadap metodologi penelitian.
13. Kedua orang tua tercinta, Ayah Syarifuddin Lubis dan Ibu Indah Saedah Munthe, yang dengan ketulusan dan kesabaran tanpa batas senantiasa mengiringi setiap langkah Peneliti melalui doa, kasih sayang, pengorbanan, serta dukungan moral yang tak pernah putus. Segala capaian yang diraih Peneliti hingga tahap ini tidak terlepas dari peran dan keikhlasan Ayah dan Ibu dalam mendidik, membimbing, serta menguatkan Peneliti dalam menghadapi setiap tantangan akademik maupun kehidupan.
14. Saudara-saudara Peneliti, Abang Ridho Alfadillah, Kakak Dinda Alawiyah, dan Kakak Annisya Putri, yang telah menjadi bagian penting dalam perjalanan akademik Peneliti. Atas kebersamaan yang terjalin, perhatian yang tulus, serta semangat dan dukungan yang senantiasa diberikan, Peneliti memperoleh kekuatan, motivasi, dan keteguhan hati dalam menghadapi setiap proses dan tantangan, hingga akhirnya mampu menyelesaikan studi dan penyusunan Tugas Akhir ini.
15. Teman-teman seperjuangan Peneliti, yaitu Ciwi Pikmi yang terdiri atas Anggy, Dina, Ara, Emil, Zhara, Mutiara, Alawiyah, dan Aliya, serta *Girl Squad* yang terdiri atas Devi, Irma, dan Farah. Kebersamaan, dukungan, semangat, serta berbagai cerita dan canda yang dibagikan selama masa perkuliahan telah memberikan motivasi, kekuatan, dan pengalaman berharga bagi Peneliti dalam menyelesaikan studi dan penyusunan Tugas Akhir ini.
16. Teman-teman *Classic* yang telah bersama Peneliti dalam menjalani proses



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

perkuliahan, berbagi pengalaman akademik, saling memberikan dukungan, serta bekerja sama dalam menghadapi berbagai dinamika dan tantangan selama masa studi.

7. Seluruh pihak lainnya yang tidak dapat Peneliti sebutkan satu per satu, yang secara langsung maupun tidak langsung telah memberikan bantuan, doa, perhatian, serta dukungan selama proses penyusunan dan penyelesaian Tugas Akhir ini.

Peneliti menyadari bahwa Tugas Akhir ini masih memiliki keterbatasan dan kekurangan. Peneliti mengharapkan saran dan kritik yang bersifat membangun demi penyempurnaan di masa yang akan datang. Kritik dan saran tersebut dapat disampaikan melalui alamat *e-mail* 12250321571@students.uin-suska.ac.id. Semoga Tugas Akhir ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang sistem informasi dan kesehatan mental. Peneliti mengucapkan terima kasih yang sebesar-besarnya kepada semua orang yang telah membantu, mendukung, dan mendoakan Peneliti selama proses penyusunan Tugas Akhir. Semoga Allah SWT membalas segala kebaikan dengan pahala dan keberkahan yang berlipat ganda untuk kita semua.

Pekanbaru, 15 Januari 2026

Peneliti,

RIFSya Aulia
NIM. 12250321571

UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PERBANDINGAN KINERJA ALGORITMA *ENSEMBLE* *LEARNING* DALAM KLASIFIKASI TINGKAT KESEHATAN MENTAL PADA MAHASISWA

RIFSYA AULIA
NIM: 12250321571

Tanggal Sidang: 09 Januari 2026
Periode Wisuda:

Program Studi Sistem Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. Soebrantas, No. 155, Pekanbaru

ABSTRAK

Kesehatan mental mahasiswa merupakan isu penting di lingkungan akademik akibat tingginya tekanan akademik dan sosial yang berpotensi memicu depresi, kecemasan, dan stres. Penelitian ini bertujuan untuk mengklasifikasikan tingkat depresi, kecemasan, dan stres mahasiswa berdasarkan kuesioner *Depression Anxiety Stress Scale-21* (DASS-21) menggunakan pendekatan *machine learning* berbasis *ensemble learning*. Enam algoritma *ensemble learning* digunakan, yaitu *Random Forest*, *Gradient Boosting*, *AdaBoost*, *XGBoost*, *LightGBM*, dan *CatBoost*. Untuk menganalisis pengaruh strategi pengolahan data terhadap kinerja model, diterapkan tiga skema data, yaitu data asli, *Synthetic Minority Oversampling Technique* (SMOTE), dan augmentasi data. *Dataset* dibagi menjadi data latih dan data uji dengan rasio 80:20. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*, serta didukung oleh *confusion matrix* dan kurva ROC-AUC. Hasil penelitian menunjukkan bahwa skema augmentasi data secara konsisten meningkatkan performa model pada seluruh target klasifikasi. *Random Forest* dengan skema augmentasi menghasilkan kinerja terbaik, dengan akurasi sebesar 99.71% pada klasifikasi depresi, 99.42% pada klasifikasi kecemasan, dan 100% pada klasifikasi stres. Temuan ini menunjukkan bahwa kombinasi pendekatan *ensemble learning* dan augmentasi data efektif dalam meningkatkan akurasi klasifikasi kondisi kesehatan mental mahasiswa berbasis DASS-21.

Kata Kunci: augmentasi, DASS-21, *ensemble learning*, kesehatan mental, klasifikasi



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

COMPARATIVE PERFORMANCE ANALYSIS OF ENSEMBLE LEARNING ALGORITHMS IN CLASSIFYING MENTAL HEALTH LEVELS AMONG UNIVERSITY STUDENTS

RIFSYA AULIA
NIM: 12250321571

Date of Final Exam: January 09th 2026
Graduation Period:

Department of Information System
Faculty of Science and Technology
State Islamic University of Sultan Syarif Kasim Riau
Soebrantas Street, No. 155, Pekanbaru

ABSTRACT

Student mental health is an important issue in academic settings due to high academic and social pressures that can potentially trigger depression, anxiety, and stress. This study aims to classify the levels of depression, anxiety, and stress among students based on the Depression Anxiety Stress Scale-21 (DASS-21) questionnaire using an ensemble learning-based machine learning approach. Six ensemble learning algorithms were used, namely Random Forest, Gradient Boosting, AdaBoost, XGBoost, LightGBM, and CatBoost. To analyze the effect of data processing strategies on model performance, three data schemes were applied, namely original data, Synthetic Minority Oversampling Technique (SMOTE), and data augmentation. The dataset was split into training and test sets at 80:20. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics, supported by a confusion matrix and an ROC-AUC curve. The results showed that the data augmentation scheme consistently improved model performance across all classification targets. Random Forest with the augmentation scheme achieved the best performance, with accuracies of 99.71% in depression classification, 99.42% in anxiety classification, and 100% in stress classification. These findings indicate that combining ensemble learning and data augmentation improves the accuracy of classifying students' mental health conditions using the DASS-21.

Keywords: *augmentation, classification, DASS-21, ensemble learning, mental health*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR	vii
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR GAMBAR	xvi
DAFTAR TABEL	xvii
DAFTAR SINGKATAN	xix
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan	5
1.5 Manfaat	6
1.6 Sistematika Penulisan	6
2 LANDASAN TEORI	8
2.1 Depresi	8
2.2 Kecemasan	8
2.3 Stres	9
2.4 <i>Depression Anxiety Stress Scale-21 (DASS-21)</i>	10
2.5 <i>Data Mining</i>	12



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.6	<i>Machine Learning</i>	13
2.7	Klasifikasi	13
2.8	<i>Ensemble Learning</i>	14
2.8.1	<i>Random Forest</i>	15
2.8.2	<i>Gradient Boosting</i>	16
2.8.3	<i>Adaptive Boosting (AdaBoost)</i>	17
2.8.4	<i>Extreme Gradient Boosting (XGBoost)</i>	19
2.8.5	<i>Light Gradient Boosting Machine (LightGBM)</i>	22
2.8.6	<i>Categorical Boosting (CatBoost)</i>	22
2.9	<i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	24
2.10	Augmentasi Data	24
2.11	<i>Confusion Matrix</i>	25
2.12	<i>Receiver Operating Characteristic (ROC)</i>	25
2.13	<i>Area Under the Curve (AUC)</i>	26
2.14	Fakultas Sains dan Teknologi	26
2.15	Penelitian Terdahulu	27
3	METODOLOGI PENELITIAN	31
3.1	Pendahuluan	32
3.1.1	Objek Penelitian	32
3.1.2	Identifikasi Permasalahan	32
3.1.3	Studi Literatur	32
3.2	Pengumpulan Data	33
3.2.1	Kuesioner	33
3.2.2	Wawancara	33
3.3	Augmentasi Data	34
3.4	<i>Preprocessing Data</i>	35
3.4.1	Pembersihan Data	35
3.4.2	Pembentukan Label	35
3.4.3	Transformasi Data	36
3.4.4	Normalisasi Data	36
3.5	Pembagian Data	36
3.6	<i>Balancing Data</i>	37
3.7	Pemodelan	39
3.7.1	<i>Random Forest</i>	39
3.7.2	<i>Gradient Boosting</i>	39

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.7.3	AdaBoost	40
3.7.4	XGBoost	40
3.7.5	LightGBM	40
3.7.6	CatBoost	40
3.8	Evaluasi	42
4	ANALISIS DAN HASIL	44
4.1	Pendahuluan	44
4.1.1	Objek Penelitian	44
4.1.2	Identifikasi Permasalahan	45
4.2	Pengumpulan Data	45
4.2.1	Kuesioner	45
4.2.2	Wawancara	46
4.2.3	Distribusi Data	47
4.2.4	Analisis Korelasi	49
4.3	Augmentasi Data	51
4.4	<i>Preprocessing Data</i>	52
4.4.1	Pembersihan Data	52
4.4.2	Pembentukan Label	53
4.4.3	Transformasi Data	54
4.4.4	Normalisasi Data	55
4.5	Pembagian Data	56
4.6	Pemodelan	58
4.6.1	<i>Random Forest</i>	58
4.6.2	<i>Gradient Boosting</i>	61
4.6.3	AdaBoost	63
4.6.4	XGBoost	66
4.6.5	LightGBM	69
4.6.6	CatBoost	71
4.7	Evaluasi	74
4.7.1	Evaluasi Kinerja Model Klasifikasi	74
4.7.2	Analisis <i>Confusion Matrix</i> dan ROC-AUC	78
4.8	Diskusi	82
5	PENUTUP	85
5.1	Kesimpulan	85
5.2	Saran	86



DAFTAR PUSTAKA

LAMPIRAN A SURAT DAN TRANSKIP WAWANCARA	A - 2
LAMPIRAN B DOKUMENTASI	B - 1
LAMPIRAN C DATA PENELITIAN	C - 1
LAMPIRAN D KUESIONER PENELITIAN	D - 1



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR GAMBAR

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.1	Alur Metodologi Penelitian	31
4.1	Distribusi Jenis Kelamin Responden	47
4.2	Distribusi Usia Responden	47
4.3	Distribusi Jurusan Responden	48
4.4	Distribusi Semester Responden	48
4.5	Distribusi IPK Responden	48
4.6	Distribusi Tingkat Kesehatan Mental Mahasiswa	49
4.7	Matriks Korelasi Item DASS-21	50
4.8	Matriks Korelasi Fitur Numerik	51
4.9	Perbandingan Akurasi Target Depresi	77
4.10	Perbandingan Akurasi Target Kecemasan	77
4.11	Perbandingan Akurasi Target Stres	77
4.12	<i>Confusion Matrix</i> Model Terbaik pada Target Depresi	79
4.13	Kurva ROC-AUC Model Terbaik pada Target Depresi	79
4.14	<i>Confusion Matrix</i> Model Terbaik pada Target Kecemasan	80
4.15	Kurva ROC-AUC Model Terbaik pada Target Kecemasan	80
4.16	<i>Confusion Matrix</i> Model Terbaik pada Target Stres	81
4.17	Kurva ROC-AUC Model Terbaik pada Target Stres	81
A.1	Surat Keterangan Selesai Wawancara	A - 2
A.2	Transkrip Hasil Wawancara	A - 3
A.3	Transkrip Hasil Wawancara	A - 4
A.4	Transkrip Hasil Wawancara	A - 5
B.1	Dokumentasi Wawancara	B - 1
B.2	Dokumentasi Pengumpulan Data	B - 1
D.1	Kuesioner Penelitian Tugas Akhir	D - 1

DAFTAR TABEL

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.1	Pernyataan Kuesioner DASS-21	10
2.2	Kategori Tingkat Keparahan Skor DASS-21	11
2.3	Populasi Mahasiswa Setiap Program Studi	27
3.1	Teknik dan Parameter Augmentasi Data	35
3.2	Skema Pengolahan Data	38
3.3	Konfigurasi Parameter Algoritma	41
4.1	Karakteristik Objek Penelitian	44
4.2	Atribut Data Penelitian	46
4.3	Komposisi Data Hasil Augmentasi	52
4.4	Hasil Pembersihan Data	52
4.5	Hasil Pembentukan Label	54
4.6	Hasil Pembentukan Label dengan Augmentasi Data	54
4.7	Hasil Transformasi Data	55
4.8	Hasil Transformasi dengan Augmentasi Data	55
4.9	Hasil Normalisasi Data	56
4.10	Hasil Normalisasi dengan Augmentasi Data	56
4.11	Pembagian Data pada Data Asli	57
4.12	Pembagian Data dengan SMOTE	57
4.13	Pembagian Data dengan Augmentasi Data	57
4.14	Hasil Pemodelan <i>Random Forest</i> pada Data Asli	59
4.15	Hasil Pemodelan <i>Random Forest</i> dengan SMOTE	59
4.16	Hasil Pemodelan <i>Random Forest</i> dengan Augmentasi Data	60
4.17	Hasil Pemodelan <i>Gradient Boosting</i> pada Data Asli	61
4.18	Hasil Pemodelan <i>Gradient Boosting</i> dengan SMOTE	62
4.19	Hasil Pemodelan <i>Gradient Boosting</i> dengan Augmentasi Data	62
4.20	Hasil Pemodelan AdaBoost pada Data Asli	64
4.21	Hasil Pemodelan AdaBoost dengan SMOTE	64
4.22	Hasil Pemodelan AdaBoost dengan Augmentasi Data	65
4.23	Hasil Pemodelan XGBoost pada Data Asli	67
4.24	Hasil Pemodelan XGBoost dengan SMOTE	67
4.25	Hasil Pemodelan XGBoost dengan Augmentasi Data	68
4.26	Hasil Pemodelan LightGBM pada Data Asli	69
4.27	Hasil Pemodelan LightGBM dengan SMOTE	70



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.28 Hasil Pemodelan LightGBM dengan Augmentasi Data	70
4.29 Hasil Pemodelan CatBoost pada Data Asli	72
4.30 Hasil Pemodelan CatBoost dengan SMOTE	72
4.31 Hasil Pemodelan CatBoost dengan Augmentasi Data	73
4.32 Perbandingan Model Terbaik	74
C.1 Data Demografis Responden	C - 1



UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR SINGKATAN

AdaBoost	: <i>Adaptive Boosting</i>
AUC	: <i>Area Under the Curve</i>
BAI	: <i>Beck Anxiety Inventory</i>
CatBoost	: <i>Categorical Boosting</i>
DASS	: <i>Depression, Anxiety, and Stress Scale</i>
DASS-21	: <i>Depression, Anxiety, and Stress Scale-21</i>
DASS-42	: <i>Depression, Anxiety, and Stress Scale-42</i>
FN	: <i>False Negative</i>
FP	: <i>False Positive</i>
GB	: <i>Gradient Boosting</i>
GAD-7	: <i>Generalized Anxiety Disorder-7</i>
HGS	: <i>Hand Grip Strength</i>
IPK	: <i>Indeks Prestasi Kumulatif</i>
ISI-7	: <i>Insomnia Severity Index-7</i>
Kemdikbudristek	: <i>Kementerian Pendidikan, Tinggi, Sains, dan Teknologi</i>
LightGBM	: <i>Light Gradient Boosting Machine</i>
ML	: <i>Machine Learning</i>
PHQ-9	: <i>Patient Health Questionnaire-9</i>
PSS	: <i>Perceived Stress Scale</i>
RF	: <i>Random Forest</i>
Riskesdas	: <i>Riset Kesehatan Dasar</i>
ROC	: <i>Receiver Operating Characteristic</i>
SMOTE	: <i>Synthetic Minority Over-sampling Technique</i>
TN	: <i>True Negative</i>
TP	: <i>True Positive</i>
WHO	: <i>World Health Organization</i>
XGBoost	: <i>Extreme Gradient Boosting</i>

UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

ode ini dapat memengaruhi proses berpikir dan perkembangan psikologis individu (Aleem dkk., 2022), serta berdampak langsung terhadap performa akademik dan keberlangsungan studi mahasiswa (Nurhidayarnis, Mustakim, Novita, Afdal, dan Wahyuni, 2024).

Rendahnya tingkat kesadaran terhadap kesehatan mental menyebabkan proses identifikasi dan penilaian tingkat keparahan gangguan mental masih sering dilakukan secara subjektif serta memerlukan waktu yang relatif lama. Untuk mengatasi keterbatasan tersebut, pemanfaatan teknologi kecerdasan buatan, khususnya *Machine Learning* (ML), semakin banyak diterapkan dalam bidang kesehatan mental. ML bertujuan untuk memprediksi luaran klinis berdasarkan sejumlah *predictor* tertentu (Mateen, Liley, Denniston, Holmes, dan Vollmer, 2020; Roberts dkk., 2021), dan telah terbukti efektif dalam mengelompokkan individu berdasarkan kondisi kesehatan mental yang dialami (Grabowska, Zabielski, dan Senderecka, 2024). Kemampuan ML dalam mengenali pola data yang kompleks serta meminimalkan kesalahan spesifikasi model menjadikannya unggul dalam meningkatkan akurasi prediksi (Tennenhouse, Marrie, Bernstein, dan Lix, 2020). Pendekatan ini sejalan dengan perkembangan riset medis terkini yang secara luas memanfaatkan algoritma ML dalam memprediksi berbagai gangguan mental pada beragam konteks populasi (Deberneh dan Kim, 2021; Saberi-Karimian dkk., 2021; Song dkk., 2021).

Penilaian gangguan mental secara klinis umumnya dilakukan oleh psikiater dengan menggunakan instrumen terstandar, salah satunya adalah kuesioner *Depression Anxiety Stress Scale* (DASS-21) yang terdiri atas 21 butir pernyataan (Vitória dkk., 2024). Instrumen ini dikembangkan berdasarkan model tripartit yang secara konseptual membedakan gejala depresi, kecemasan, dan stres (Ali dan Green, 2019; Ali, Alkhamees, Hori, Kim, dan Kunugi, 2021). Selain memiliki waktu pengisian yang relatif singkat, DASS-21 juga mampu mengidentifikasi gejala stres yang sering muncul bersamaan dengan depresi dan kecemasan, sehingga menjadikannya lebih unggul dibandingkan alat ukur lain yang hanya berfokus pada satu jenis gangguan (Ali dkk., 2021).

Sejumlah penelitian terdahulu telah mengkaji penerapan algoritma *machine learning* dalam mendiagnosis gangguan psikologis, seperti depresi, kecemasan, dan stres pada mahasiswa. M. J. Hasan, Das, Matubber, Shifat, dan Morol (2024) memanfaatkan kuesioner DASS-21 dengan pendekatan *data augmentation* serta menerapkan berbagai algoritma, antara lain *Logistic Regression*, *Random Forest*, *SVM*, dan *Gradient Boosting*. Hasil penelitian tersebut menunjukkan bahwa *Logistic Regression* memberikan kinerja terbaik dengan akurasi masing-masing sebe-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

sar 0.97 untuk kecemasan, 0.975 untuk depresi, dan 0.93 untuk stres. Sementara itu, Duangchaemkarn, Khammarew, dan Aramvith (2024) menerapkan model *deep learning* dan melaporkan akurasi di atas 90% dalam mengklasifikasikan kondisi emosional berbasis DASS-21, yang semakin memperkuat potensi pendekatan *machine learning* dalam diagnosis psikologis.

Penelitian lain menunjukkan bahwa kinerja algoritma dapat bervariasi bergantung pada konteks dan metode yang digunakan. Herman dkk. (2025) mengklasifikasikan gangguan kecemasan pada remaja menggunakan algoritma *Naive Bayes*, *C4.5*, dan *K-Nearest Neighbors*, dengan *C4.5* mencapai akurasi 100% pada skema pengujian *holdout*. Selanjutnya, Srinath dkk. (2022) membandingkan SVM dan *Logistic Regression* dalam memprediksi tingkat keparahan gangguan psikologis dan menemukan bahwa *Logistic Regression* menunjukkan performa yang lebih konsisten. Penelitian oleh S. Usman, Rusli, Jalil, dan Bani (2022) yang menggabungkan pengukuran *Hand Grip Strength* dengan DASS juga menunjukkan bahwa algoritma *Random Forest* mampu mencapai akurasi sebesar 93.75%, lebih tinggi dibandingkan SVM. Temuan serupa dilaporkan oleh Bhatnagar, Agarwal, dan Sharma (2023), Kumar dkk. (2024), Farooq, Konda, Kunwar, dan Rajeev (2023), dan Singh dan Kumar (2021), yang secara konsisten menegaskan efektivitas algoritma *machine learning*, khususnya *Random Forest*, dalam menganalisis gangguan kesehatan mental berbasis kuesioner.

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada penggunaan satu atau dua algoritma utama dan belum menyajikan perbandingan kinerja yang komprehensif antar berbagai algoritma klasifikasi populer dengan skema pengolahan data yang beragam. Selain itu, kajian mengenai pengaruh teknik penyeimbangan dan penambahan data, seperti *Synthetic Minority Over-sampling Technique* (SMOTE) dan *data augmentation*, masih terbatas dan belum dianalisis secara sistematis terhadap seluruh target klasifikasi, yaitu depresi, kecemasan, dan stres secara simultan.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membandingkan kinerja lima algoritma *ensemble learning* yang umum digunakan, yaitu *Random Forest*, *Gradient Boosting*, *Adaptive Boosting* (AdaBoost), *Extreme Gradient Boosting* (XGBoost), *Light Gradient Boosting Machine* (LightGBM), dan *Categorical Boosting* (CatBoost), dalam mengklasifikasikan tingkat depresi, kecemasan, dan stres pada mahasiswa berdasarkan data kuesioner DASS-21. Selain itu, penelitian ini juga mengevaluasi pengaruh tiga skema pengolahan data, yaitu data asli, SMOTE, dan *data augmentation*, terhadap kinerja masing-masing algoritma.



Hasil penelitian ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai kombinasi algoritma dan skema pengolahan data yang paling efektif, serta menjadi landasan dalam pengembangan sistem pemantauan kesehatan mental mahasiswa di lingkungan akademik.

1.2 Rumusan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang telah diuraikan sebelumnya, rumusan masalah dalam penelitian ini disusun untuk mengarahkan fokus kajian secara sistematis sebagai berikut:

1. Bagaimana proses pengklasifikasian tingkat depresi, kecemasan, dan stres pada mahasiswa berbasis data kuesioner DASS-21 menggunakan algoritma *ensemble learning*?
2. Bagaimana perbandingan kinerja masing-masing algoritma *ensemble learning*, yaitu *Random Forest*, *Gradient Boosting*, *XGBoost*, *LightGBM*, dan *CatBoost*, dalam mengklasifikasikan tingkat depresi, kecemasan, dan stres pada mahasiswa berdasarkan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*?
3. Bagaimana pengaruh perbedaan skema pengolahan data, yang meliputi penggunaan data asli, penyeimbangan kelas menggunakan SMOTE, serta penerapan *data augmentation*, terhadap kinerja masing-masing algoritma *ensemble learning* dalam meningkatkan kemampuan klasifikasi kondisi kesehatan mental mahasiswa?

1.3 Batasan Masalah

Untuk menjaga fokus penelitian, memastikan konsistensi metodologis, serta menjamin keterulangan (*reproducibility*) eksperimen, penelitian ini menetapkan sejumlah batasan masalah yang dirumuskan secara terstruktur sebagai berikut:

1. Responden penelitian merupakan mahasiswa aktif Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Data penelitian diperoleh melalui pengisian kuesioner DASS-21 yang digunakan sebagai instrumen pengukuran tingkat depresi, kecemasan, dan stres.
3. Variabel target yang dianalisis dalam penelitian ini mencakup tiga dimensi kesehatan mental, yaitu depresi, kecemasan, dan stres, yang masing-masing diklasifikasikan ke dalam lima tingkat keparahan, yakni Normal, Ringan, Sedang, Berat, dan Sangat Berat.
4. Algoritma *ensemble learning* yang digunakan dalam penelitian ini terdiri atas enam metode, yaitu *Random Forest*, *Gradient Boosting*, *AdaBoost*,

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

XGBoost, LightGBM, dan CatBoost.

5. Pembagian data dilakukan menggunakan metode *hold-out* dengan proporsi 80% sebagai data latih dan 20% sebagai data uji.
6. Penelitian ini menerapkan tiga skema pengolahan data, yang meliputi penggunaan data asli, penyeimbangan kelas menggunakan SMOTE, serta augmentasi data melalui penambahan *Gaussian noise* pada item-item kuesioner DASS-21.
7. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, serta *Receiver Operating Characteristic* (ROC) dan *Area Under the Curve* (AUC) dengan pendekatan *macro average*.
8. Ruang lingkup penelitian ini tidak mencakup pengembangan sistem, implementasi lanjutan, maupun intervensi psikologis atau terapi klinis. Penelitian dibatasi pada tahap pemodelan, evaluasi performa, dan interpretasi hasil klasifikasi.
9. Penelitian ini bersifat analitis dan komputasional, sehingga tidak membahas aspek diagnosis klinis, intervensi medis, maupun rekomendasi penanganan kesehatan mental secara langsung.

1.4 Tujuan

Adapun tujuan yang dapat dicapai melalui penelitian ini adalah sebagai berikut:

1. Mengklasifikasikan tingkat depresi, kecemasan, dan stres pada mahasiswa berbasis data kuesioner DASS-21 dengan memanfaatkan pendekatan *ensemble learning*.
2. Menganalisis serta membandingkan kinerja enam algoritma *ensemble learning*, yaitu *Random Forest*, *Gradient Boosting*, *AdaBoost*, *XGBoost*, *LightGBM*, dan *CatBoost*, dalam mengklasifikasikan tingkat depresi, kecemasan, dan stres pada mahasiswa, serta mengidentifikasi algoritma dengan performa paling optimal berdasarkan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*.
3. Mengevaluasi pengaruh penerapan berbagai skema pengolahan data, yang meliputi penggunaan data asli, penyeimbangan kelas menggunakan SMOTE, serta *data augmentation*, terhadap kinerja masing-masing algoritma *ensemble learning* dalam meningkatkan kemampuan klasifikasi kondisi kesehatan mental mahasiswa.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1.5 Manfaat

Manfaat yang diharapkan dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Memberikan hasil klasifikasi tingkat depresi, kecemasan, dan stres pada mahasiswa berbasis data kuesioner DASS-21 melalui penerapan berbagai algoritma *ensemble learning*, yaitu *Random Forest*, *Gradient Boosting*, *AdaBoost*, *XGBoost*, *LightGBM*, dan *CatBoost*.
2. Menyediakan referensi empiris mengenai pengaruh penerapan berbagai skema pengolahan data, meliputi penggunaan data asli, penyeimbangan kelas menggunakan SMOTE, serta augmentasi data, terhadap kinerja algoritma klasifikasi, sehingga dapat menjadi dasar pertimbangan dalam pemilihan algoritma dan strategi pengolahan data yang tepat untuk klasifikasi tingkat kesehatan mental mahasiswa.
3. Menyajikan analisis perbandingan kinerja antaralgoritma dan skema pengolahan data berdasarkan metrik evaluasi *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan ROC-AUC, sehingga dapat diidentifikasi kombinasi algoritma dan skema data yang menghasilkan performa klasifikasi paling optimal.
4. Memberikan dasar pendukung bagi pihak perguruan tinggi dalam upaya identifikasi dini mahasiswa yang berpotensi mengalami gangguan kesehatan mental, sehingga hasil penelitian ini dapat dimanfaatkan sebagai bahan pertimbangan dalam perencanaan program pendampingan, layanan konseling, dan intervensi psikologis yang bersifat preventif.
5. Menjadi rujukan bagi penelitian selanjutnya yang mengkaji klasifikasi kesehatan mental berbasis DASS-21 dengan pendekatan *ensemble learning*, khususnya dalam aspek pemilihan algoritma, strategi penanganan ketidakseimbangan data, serta metode evaluasi kinerja model.

1.6 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini disusun ke dalam lima bab, dengan uraian sebagai berikut:

BAB I. PENDAHULUAN

Bab ini membahas dasar penelitian, termasuk latar belakang masalah, rumusan masalah, batasan, tujuan, dan manfaat penelitian. Bab ini juga membahas sistematika penulisan yang digunakan dalam penyusunan laporan Tugas Akhir.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II. LANDASAN TEORI

Bab ini menyajikan tinjauan pustaka dan landasan teoretis yang mendukung penelitian. Pembahasan mencakup konsep kesehatan mental yang meliputi depresi, kecemasan, dan stres, instrumen pengukuran DASS-21, konsep dasar *data mining*, *machine learning*, dan *ensemble learning*, serta uraian mengenai algoritma klasifikasi yang digunakan, yaitu *Random Forest*, *Gradient Boosting*, *AdaBoost*, *XGBoost*, *LightGBM*, dan *CatBoost*.

BAB III. METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan penelitian yang dilakukan secara sistematis dan terstruktur, dimulai dari pendahuluan, pengumpulan data, pra-pemrosesan data (*data preprocessing*), pembagian data ke dalam data latih dan data uji, penerapan skema pengolahan data yang meliputi data asli, SMOTE, dan *data augmentation*, proses pemodelan menggunakan algoritma *Random Forest*, *Gradient Boosting*, *AdaBoost*, *XGBoost*, *LightGBM*, dan *CatBoost*, hingga tahap evaluasi kinerja model menggunakan metrik evaluasi yang telah ditetapkan.

BAB IV. ANALISIS DAN HASIL

Bab ini menyajikan hasil eksperimen serta analisis kinerja model klasifikasi untuk setiap target penelitian, yaitu depresi, kecemasan, dan stres. Pembahasan mencakup perbandingan kinerja antar algoritma dan skema pengolahan data berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*, analisis *confusion matrix* dan ROC-AUC pada model dengan performa terbaik, visualisasi grafik perbandingan kinerja, serta pembahasan komparatif antara hasil penelitian ini dan temuan penelitian terdahulu.

BAB V. PENUTUP

Bab ini memuat kesimpulan yang diperoleh berdasarkan hasil penelitian dan analisis yang telah dilakukan, serta saran untuk penelitian selanjutnya yang berkaitan dengan pengembangan metode, pemilihan algoritma, maupun pemanfaatan data dalam klasifikasi kesehatan mental berbasis DASS-21.

UIN SUSKA RIAU



BAB 2

LANDASAN TEORI

2.1 Depresi

Depression (depresi) adalah gangguan mental yang ditandai oleh perasaan sedih yang berlangsung lama, hilangnya minat atau kesenangan terhadap aktivitas sehari-hari, kesulitan berkonsentrasi, masalah daya ingat, dan penurunan tingkat energi (Paykel, 2008). Menurut *Diagnostic and Statistical Manual of Mental Disorders* edisi kelima (DSM-5), depresi diklasifikasikan ke dalam beberapa bentuk, seperti disregulasi *mood* yang tidak teratur, *major depressive disorder*, *persistent depressive disorder* atau distimia, *premenstrual dysphoric disorder*, serta depresi yang muncul sebagai konsekuensi dari kondisi medis tertentu. Meskipun memiliki klasifikasi yang beragam, seluruh bentuk depresi menunjukkan karakteristik umum berupa perasaan hampa, sedih, atau mudah tersinggung yang disertai perubahan fisik maupun kognitif, sehingga menghambat kemampuan individu dalam menjalankan fungsi kehidupan sehari-hari secara optimal (Chand dan Arif, 2025).

Sebagai gangguan yang bersifat multidimensional, depresi menimbulkan beban yang signifikan bagi individu maupun masyarakat luas. Depresi tidak hanya berdampak pada kesehatan psikologis, tetapi juga memengaruhi fungsi sosial, interpersonal, serta kinerja pekerjaan (Asif, Mudassar, Shahzad, Raouf, dan Pervaiz, 2020; Hysenbegasi, Hass, dan Rowland, 2005). Individu dengan depresi sering mengalami penurunan produktivitas, gangguan pola tidur, keterbatasan dalam menjalankan aktivitas harian, serta penurunan kualitas hidup secara keseluruhan (Adler, 2006; Gilmour dan Patten, 2007; Ivanova dkk., 2010; Patel, 2001). Depresi diketahui dapat memperburuk luaran pengobatan dan mempercepat progresivitas berbagai penyakit kronis, seperti diabetes, penyakit kardiovaskular, kanker, dan obesitas, sehingga menambah kompleksitas penanganan kondisi kesehatan secara menyeluruh (Demyttenaere, Bruffaerts, Posada-Villa, dan Al., 2004; Mirza dan Jenkins, 2004).

2.2 Kecemasan

Anxiety (kecemasan) merupakan respons adaptif tubuh terhadap ancaman yang dipersepsikan, yang dipengaruhi oleh keyakinan, emosi, dan proses kognitif individu. Respons ini umumnya ditandai oleh munculnya pikiran yang dipenuhi kekhawatiran, perasaan tegang, serta aktivasi sistem fisiologis, seperti peningkatan tekanan darah, laju pernapasan, dan denyut nadi, disertai gejala somatik berupa berkeringat, kesulitan menelan, pusing, dan nyeri dada (Association, 2013; Yeshaw



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dan Mossie, 2017). Secara fisiologis, kecemasan berkaitan erat dengan respons stres, di mana tubuh menempatkan dirinya dalam kondisi siaga untuk menghadapi ancaman melalui mekanisme *fight or flight*. Pada kondisi ini, detak jantung meningkat, aliran darah dialihkan ke otot-otot utama, pernapasan menjadi lebih cepat dan dangkal, serta terjadi perubahan pada respons pupil (Heck, 2008).

Selain respons fisiologis, kecemasan juga memunculkan berbagai gejala psikologis dan perilaku, seperti gemetar pada tangan dan bibir, mulut kering, peningkatan frekuensi buang air kecil, serta gangguan kualitas tidur (Asif dkk., 2020; Ericson dan Gardner, 1992). Meskipun sering dirasakan sebagai kondisi yang tidak nyaman karena mengganggu aktivitas sehari-hari dan menguras energi, kecemasan pada tingkat tertentu memiliki fungsi adaptif dalam membantu individu mengantisipasi dan menghadapi ancaman. Permasalahan justru muncul ketika mekanisme kecemasan berada dalam kondisi tidak seimbang, baik dalam bentuk respons yang berlebihan maupun ketiadaan respons sama sekali. Kecemasan yang berlebihan dapat memicu stres yang tidak produktif dan mengganggu fungsi individu, sementara ketiadaannya dapat mencerminkan lemahnya kemampuan tubuh dalam mengenali dan merespons potensi ancaman (Heck, 2008).

2.3 Stres

Stres merupakan kondisi psikologis yang muncul ketika seseorang merasa bahwa kebutuhan mereka melebihi kemampuan yang dimilikinya untuk mengatasi tuntutan tersebut (A. Usman, Ahmed, Ahmed, dan Akbar, 2011). Dalam konteks ini, stres dapat dipahami sebagai bentuk ancaman atau tantangan terhadap kesejahteraan individu. Apabila kapasitas adaptif organisme tidak mampu memenuhi tuntutan lingkungan, maka kondisi tersebut berpotensi menimbulkan gangguan, baik secara biologis maupun psikologis (Asif dkk., 2020; Cohen, Gianaros, dan Manuck, 2016). Selye mendefinisikan stres sebagai respons nonspecific tubuh terhadap setiap tuntutan adaptasi yang timbul akibat perubahan, baik yang bersifat menyenangkan maupun tidak menyenangkan (Fink, 2017).

Dalam kehidupan sehari-hari, stres sering kali muncul akibat paparan terhadap situasi yang dipersepsikan sebagai tidak terkendali dan merugikan, seperti terjebak dalam kemacetan parah, berhadapan dengan atasan yang bersikap bermusuhan, menghadapi tekanan finansial akibat tagihan yang belum terbayar, maupun ancaman dari lingkungan sekitar. Kondisi tersebut umumnya memicu respons emosional berupa rasa takut dan kecemasan. Bergantung pada konteks dan persepsi individu, respons terhadap ancaman tersebut dapat memunculkan dua



reaksi fisiologis utama, yaitu *fight* (melawan) atau *flight* (melarikan diri). Tingkat intensitas stres serta dampak fisiologis yang ditimbulkannya sangat dipengaruhi oleh bagaimana individu menilai kemampuan dirinya dalam mengelola dan mengatasi stresor yang dihadapi (Fink, 2017).

2.4 Depression Anxiety Stress Scale-21 (DASS-21)

Depression Anxiety Stress Scale-21 (DASS-21) merupakan instrumen pengukuran mandiri yang terdiri atas 21 butir pernyataan yang dirancang untuk menilai gejala depresi, kecemasan, dan stres (DAS). DASS-21 merupakan versi ringkas dari DASS-42 yang tetap mempertahankan karakteristik psikometrik utama dalam mengukur kondisi *stress, anxiety, depression* (SAD) (Antony, Bieling, Cox, Enns, dan Swinson, 1998). Instrumen ini memuat 21 item yang terbagi secara proporsional ke dalam tiga subskala, yaitu masing-masing tujuh item untuk subskala depresi, kecemasan, dan stres (Lovibond dan Lovibond, 1995).

Subskala stres berfokus pada pengukuran gejala ketegangan kronis dan tingkat rangsangan yang berlangsung secara persisten. Subskala kecemasan mengukur respons ketakutan dan aktivasi psikologis yang berkaitan dengan kondisi ancaman, sedangkan subskala depresi menilai suasana hati yang rendah, penurunan harga diri, serta pandangan negatif terhadap masa depan (Fauzi dkk., 2021; Shea, Tennant, dan Pallant, 2009). Tabel 2.1 menampilkan pernyataan instrumen DASS-21. Penilaian pada instrumen DASS-21 dilakukan menggunakan skala Likert empat poin (Fauzi dkk., 2021), yaitu:

0 = Sama sekali tidak berlaku pada saya

1 = Berlaku pada saya dalam tingkat tertentu atau sesekali

2 = Berlaku pada saya dalam tingkat cukup sering atau sebagian besar waktu

3 = Sangat sering berlaku pada saya atau hampir sepanjang waktu

Tabel 2.1. Pernyataan Kuesioner DASS-21

Sumber: Priya dkk. (2020)

No	Depresi	Kecemasan	Stres
1	Tidak dapat merasakan perasaan positif	Mulut terasa kering	Sulit untuk rileks
2	Sulit memiliki inisiatif untuk melakukan sesuatu	Kesulitan bernapas	Bereaksi berlebihan terhadap situasi

Tabel 2.1. Pernyataan Kuesioner DASS-21

No	Depresi	Kecemasan	Stres
1	Tidak memiliki sesuatu yang dinantikan	Mengalami gemetar	Merasa memiliki energi gugup yang berlebihan
2	Merasa sedih dan murung	Khawatir akan panik dan mempermalukan diri sendiri	Mudah gelisah
3	Tidak dapat merasa antusias	Merasa hampir panik	Sulit untuk bersantai
4	Merasa diri tidak berharga	Menyadari detak jantung meskipun tidak melakukan aktivitas fisik	Tidak toleran saat melakukan sesuatu
5	Merasa hidup tidak bermakna	Merasa takut tanpa alasan yang jelas	Mudah tersinggung

Setelah data dikumpulkan, setiap respons partisipan diberikan nilai numerik dengan rentang 0 hingga 3 (M. J. Hasan dkk., 2024). Selanjutnya, skor total dihitung dengan menjumlahkan nilai dari setiap butir pertanyaan sesuai dengan rumus berikut (Henry dan Crawford, 2005):

$$\text{Skor} = (\sum \text{skor pada setiap kategori}) \times 2 \quad (2.1)$$

Tingkat keparahan depresi, kecemasan, dan stres dikategorikan ke dalam lima tingkatan sebagaimana ditunjukkan pada Tabel 2.2 (Asif dkk., 2020; Priya dkk., 2020).

Tabel 2.2. Kategori Tingkat Keparahannya Skor DASS-21

Sumber: Lovibond dan Lovibond (1995)

Kategori	Depresi	Kecemasan	Stres
Normal	0–9	0–7	0–14
Ringan	10–13	8–9	15–18
Sedang	14–20	10–14	19–25
Berat	21–27	15–19	26–33

Tabel 2.2. Kategori Tingkat Keparahan Skor DASS-21

Kategori	Depresi	Kecemasan	Stres
Sangat Berat	28+	20+	34+

Kategori Normal menunjukkan kondisi di mana tidak ditemukan indikasi yang bermakna terkait gangguan kesehatan mental pada dimensi yang diukur. Sementara itu, kategori Ringan hingga Sangat Berat menggambarkan peningkatan tingkat keparahan gejala secara bertahap, yang tercermin dari perolehan skor yang semakin tinggi dan menunjukkan intensitas gangguan yang lebih serius. Pengelompokan tingkat keparahan ini selanjutnya dimanfaatkan sebagai label kelas (*class labels*) dalam pemodelan *machine learning*, yang berperan sebagai variabel target untuk mengklasifikasikan tingkat depresi, kecemasan, dan stres mahasiswa berdasarkan hasil pengisian kuesioner DASS-21.

2.5 Data Mining

Data mining merupakan suatu proses analitis yang bertujuan untuk menemukan pola, keteraturan, serta hubungan yang bermakna di dalam sekumpulan data yang berasal dari domain tertentu, dengan memanfaatkan teknik ilmu komputer dan statistika. Proses ini menjadi salah satu tahapan utama dalam kerangka *Knowledge Discovery in Database* (KDD), yang mencakup seleksi data, pra-pemrosesan (*preprocessing*), transformasi data, penerapan teknik data mining, evaluasi atau interpretasi model, hingga pemanfaatan pengetahuan yang dihasilkan (Fayyad, Piatetsky-Shapiro, dan Smyth, 1996). Dalam praktiknya, data mining banyak diterapkan untuk berbagai tujuan analisis, seperti *classification*, *clustering*, prediksi, serta analisis asosiasi antar data (Birjandi dan Khasteh, 2021).

Dalam data mining, terdapat dua kategori utama model, yaitu model prediktif dan model deskriptif. Model prediktif berfokus pada identifikasi pola tersembunyi dalam data untuk memprediksi kecenderungan atau nilai di masa mendatang, sedangkan model deskriptif bertujuan untuk menginterpretasikan struktur dan pola data agar mudah dipahami oleh ahli pada domain terkait, yang umumnya memanfaatkan pendekatan *unsupervised learning* (Kantardzic, 2011). Pendekatan *supervised learning* banyak digunakan pada model prediktif untuk tugas-tugas seperti *classification*, regresi, analisis deret waktu, dan prediksi. Sebaliknya, pada *unsupervised learning*, label kelas dari setiap sampel tidak diketahui sebelumnya, bahkan jumlah kelas yang terbentuk pun sering kali belum dapat ditentukan di awal proses analisis (Ha, Kambe, dan Pe, 2012).



2.6 Machine Learning

Machine Learning (ML) merupakan cabang ilmu komputer yang berkembang pesat dan memiliki peran penting dalam mendukung analisis data berskala besar, termasuk dalam bidang epidemiologi dan *big data*. Pendekatan ini menawarkan alternatif penyelesaian masalah yang dalam banyak kasus sulit ditangani secara optimal oleh metode statistika klasik (Bi, Goodman, Kaminsky, dan Lessler, 2019). Sebagai bagian dari kecerdasan buatan, *machine learning* memungkinkan sistem komputer untuk belajar dari data dan meningkatkan kinerjanya berdasarkan pengalaman, tanpa memerlukan pemrograman eksplisit untuk setiap aturan pengambilan keputusan (Islam, Hassan, Akter, Jibon, dan Sahidullah, 2024).

Secara konseptual, *machine learning* menggambarkan proses otomatis dalam mengenali pola-pola dari data sebagai dasar untuk menjalankan tugas tertentu, seperti klasifikasi dan prediksi (Mitchell, 1997). ML merupakan subbidang dari kecerdasan buatan (*artificial intelligence*) yang berfokus pada bagaimana sistem komputasi dapat memproses informasi secara adaptif dan cerdas. Meskipun bersifat otomatis, penerapan ML tetap memerlukan sejumlah keputusan penting dari pengguna, antara lain penentuan variabel keluaran (*outcome*) yang akan diprediksi atau diklasifikasikan, pemilihan data yang digunakan dalam proses pembelajaran, serta penetapan atribut yang relevan untuk membangun model (Black, Kueper, dan Williamson, 2023). Dalam konteks klasifikasi dan prediksi, pendekatan yang paling umum digunakan adalah *supervised learning*, di mana model dilatih menggunakan data berlabel untuk menghasilkan prediksi yang akurat (Müller dan Guido, 2016).

2.7 Klasifikasi

Klasifikasi merupakan salah satu tugas utama dalam data mining yang bertujuan untuk memprediksi kelas target dengan tingkat ketepatan yang tinggi. Pada proses ini, algoritma klasifikasi mempelajari hubungan antara atribut masukan dan atribut keluaran dalam proses pembuatan model prediksi melalui tahapan pelatihan (*training*) (N. S. Ahmed dan Hikmat Sadiq, 2018; Anuradha dan Gupta, 2014; S. Patil dan Kulkarni, 2019). Seiring dengan pesatnya pertumbuhan data dalam lingkungan data mining, volume data yang harus dianalisis menjadi semakin besar dan kompleks (Salih Hassan, Abdulazeez, dan Tiryaki, 2018; Zebari, Abdulazeez, Zeebaree, Zebari, dan Saeed, 2020; Zeebaree, Haron, dan Abdulazeez, 2018).

Dalam kondisi tertentu, apabila struktur data tersusun dengan baik dan jumlah *node* yang terbentuk relatif minimal, metode berbasis *decision tree* sering dianggap sebagai pilihan yang efektif dan efisien (O. Ahmed dan Brifcani, 2019;

D. V. Patil dan Bichkar, 2006; Sulaiman, 2020). Secara umum, algoritma klasifikasi memiliki kemampuan untuk menangani volume data yang besar, melakukan prediksi terhadap kelas bersifat kategorikal, memanfaatkan *training set* beserta label kelas untuk membangun model, serta mengklasifikasikan data baru yang belum pernah diamati sebelumnya (Nikam, 2015).

2.8 Ensemble Learning

Ensemble learning merupakan pendekatan dalam *machine learning* yang melibatkan pelatihan beberapa model dasar (*base learners*) dan menggabungkan hasil prediksi dari masing-masing model untuk memperoleh kinerja yang lebih baik serta kemampuan generalisasi yang lebih tinggi dibandingkan penggunaan model tunggal (Zhou, Jiang, Chung, dan Wang, 2021). Pendekatan ini dirancang untuk meningkatkan performa klasifikasi dengan memanfaatkan keunggulan dari berbagai algoritma dasar yang digunakan secara bersamaan.

Salah satu tantangan utama pada algoritma *machine learning* adalah adanya bias, varians yang tinggi, serta keterbatasan akurasi pada model tunggal (Mishra dkk., 2022; Sun, Li, Li, dan Zhang, 2021). Berbagai penelitian menunjukkan bahwa model berbasis *ensemble* cenderung menghasilkan akurasi yang lebih tinggi dan performa yang lebih stabil dibandingkan model individual (Brown, 2011). Secara khusus, metode *ensemble* mampu mengurangi kesalahan akibat varians maupun bias; misalnya, pendekatan *bagging* efektif dalam menurunkan varians tanpa meningkatkan bias, sedangkan *boosting* berfokus pada pengurangan bias melalui pembelajaran bertahap (Alelyani, 2021; Breiman, 1998; Doroudi, 2020).

Berdasarkan mekanisme penggabungan modelnya, metode *ensemble learning* umumnya diklasifikasikan ke dalam tiga kelompok utama, yaitu *boosting*, *bagging*, dan *stacking* (Nguyen, Chen, Lin, dan Seeboonruang, 2021). Contoh algoritma yang termasuk dalam kategori *boosting* antara lain *gradient boosting*, XGBoost, dan AdaBoost, sedangkan algoritma *bagging* yang paling dikenal adalah *Random Forest* dan *Extra Trees Classifier*. Pada pendekatan *stacking*, dikenal pula teknik seperti *super ensemble* dan *blending* (Fawagreh, Gaber, dan Elyan, 2014; Nti, Adekoya, dan Weyori, 2020; Ribeiro dan dos Santos Coelho, 2020). Berbagai algoritma *ensemble* tersebut telah terbukti memberikan kinerja yang unggul dalam beragam aplikasi pembelajaran mesin (Esenogho, Mienye, Swart, Aruleba, dan Obaido, 2022; Q.-F. Li dan Song, 2022; Mohammadi, Narimani, Ashouri, Firouzi, dan Karimi-Jafari, 2022), serta menunjukkan efektivitas yang tinggi dalam penerapan pada permasalahan dunia nyata (Oza dan Tumer, 2008).



Hak Cipta Ditanggung Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.8.1 Random Forest

Random Forest merupakan algoritma *ensemble* nonparametrik yang bersifat *supervised*, yang dibangun melalui penggabungan sejumlah besar *decision tree* (Razavi dkk., 2023; Singh dkk., 2024), menggunakan subset dari sampel pelatihan dan variabel yang dipilih secara acak (Xia, 2020; Dhani, Lestari, Ningrum, Fakhrizal, dan Gandini, 2025). Algoritma ini dapat diterapkan pada permasalahan klasifikasi maupun regresi (Razavi dkk., 2023). Pada tugas klasifikasi, hasil prediksi ditentukan berdasarkan kelas dengan jumlah suara terbanyak (*majority voting*), sedangkan pada regresi, keluaran model diperoleh dari nilai rata-rata prediksi seluruh pohon dalam *ensemble* (Breiman, 2001).

Penggunaan *Random Forest* bertujuan untuk meningkatkan performa dan stabilitas model *decision tree* dengan cara mengurangi risiko *overfitting* serta menurunkan variansi model (Arora dan Dahiya, 2025; Singh dkk., 2024). Algoritma ini diketahui memiliki kinerja yang baik pada *dataset* berukuran besar dan mampu menangani fitur numerik maupun kategorikal secara efektif. Selain itu, *Random Forest* menyediakan mekanisme pengukuran tingkat kepentingan fitur (*feature importance*) yang memungkinkan identifikasi atribut yang memberikan kontribusi terbesar terhadap hasil prediksi (Arora dan Dahiya, 2025). Secara teoritis, performa *Random Forest* dapat dianalisis melalui konsep fungsi margin dan kesalahan generalisasi, sebagaimana dirumuskan pada Persamaan (2.2)–(2.5) (Kullarni dan Sinha, 2013).

1. Fungsi Margin

Kesalahan generalisasi (PE^*) dari *Random Forest* dapat dinyatakan menggunakan fungsi margin, yang merepresentasikan selisih tingkat dukungan terhadap kelas yang benar dibandingkan kelas lainnya:

$$PE^* = P(x, y) (mg(X, Y)) < 0 \quad (2.2)$$

Fungsi margin $mg(X, Y)$ mengukur sejauh mana rata-rata suara untuk kelas target pada titik (X, Y) lebih besar dibandingkan suara rata-rata untuk kelas lain. Dalam hal ini, X merepresentasikan vektor prediktor dan Y adalah kelas target.

Secara matematis, fungsi margin dirumuskan sebagai berikut:

$$mg(X, Y) = \sum_k I(h_k(X) = Y) - \max_{j \neq Y} \sum_k I(h_k(X) = j) \quad (2.3)$$

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dengan $I(\cdot)$ sebagai fungsi indikator yang bernilai 1 jika kondisi terpenuhi dan 0 jika tidak.

2. Kekuatan *Random Forest*

Kekuatan (*strength*) dari *Random Forest* didefinisikan sebagai nilai ekspektasi dari fungsi margin, yang menunjukkan kemampuan rata-rata *ensemble* dalam melakukan klasifikasi yang benar:

$$S = E_{X,Y}[\text{mg}(X,Y)] \quad (2.4)$$

3. Batasan Kesalahan Generalisasi

Kesalahan generalisasi dari sebuah *ensemble classifier* dibatasi oleh korelasi rata-rata antar *base learner* dan kekuatan rata-rata model, yang dirumuskan sebagai berikut:

$$PE^* \leq \rho(1 - s^2)s^2 \quad (2.5)$$

di mana ρ menyatakan rata-rata korelasi antar pohon keputusan, sedangkan s merepresentasikan kekuatan rata-rata pohon-pohon dalam *Random Forest*.

2.8.2 Gradient Boosting

Gradient Boosting merupakan algoritma *machine learning* berbasis *boosting* yang membangun sebuah *ensemble* secara bertahap untuk menghasilkan model yang kuat. Dalam implementasinya, algoritma ini umumnya menggunakan *decision tree* sebagai *base learner*, sehingga sering dikenal sebagai *Gradient Boosted Decision Tree* (GBDT). Konsep awal *boosting* sebagai pendekatan optimasi terhadap *loss function* pertama kali diperkenalkan oleh (Breiman, 1998), kemudian dikembangkan lebih lanjut secara formal oleh (Friedman, 2001).

Proses pembelajaran pada *Gradient Boosting* dilakukan secara iteratif, di mana setiap model baru dilatih untuk memperbaiki kesalahan yang dihasilkan oleh model sebelumnya. Berbeda dengan pendekatan *boosting* konvensional, inti dari algoritma ini terletak pada pemanfaatan gradien negatif dari *loss function* sebagai target pembelajaran bagi *base learner* (Natekin dan Knoll, 2013). Pendekatan ini memungkinkan model untuk secara sistematis meminimalkan kesalahan prediksi. Selain itu, implementasi modern dari *Gradient Boosting*, seperti XGBoost, mampu menangani data dengan nilai hilang (*missing values*), menyediakan informasi tingkat kepentingan fitur (*feature importance*), serta mendukung proses *feature selection* (Mienye dan Sun, 2022). Secara umum, tahapan pembelajaran *Gradient Boosting* dijelaskan pada Persamaan (2.6)–(2.9).

1. Inisialisasi Model

Model diinisialisasi dengan suatu fungsi konstan $F_0(x)$ yang meminimalkan nilai fungsi kerugian pada data pelatihan:

$$F_0(x) = \arg \min_{\alpha} \sum_{i=1}^N L(y_i, \alpha) \quad (2.6)$$

2. Perhitungan Residual Palsu

Residual palsu dihitung sebagai gradien negatif dari fungsi kerugian terhadap prediksi model sebelumnya:

$$r_{mi} = - \frac{\partial L(y_i, F(x))}{\partial F(x)} \Big|_{F(x)=F_{m-1}(x)} \quad (2.7)$$

3. Pelatihan *Base Learner*

Model dasar $h_m(x)$ dilatih untuk memprediksi residual palsu guna memperbaiki kesalahan yang tersisa dari model sebelumnya.

4. Penentuan Bobot Model

Bobot optimal ρ_m ditentukan melalui pencarian baris (*line search*) untuk meminimalkan fungsi kerugian:

$$\rho_m = \arg \min_{\rho} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \rho h_m(x_i)) \quad (2.8)$$

5. Pembaruan Model

Model diperbarui dengan menambahkan kontribusi model baru ke model sebelumnya:

$$F_m(x) = F_{m-1}(x) + \rho_m h_m(x) \quad (2.9)$$

6. Penghentian Iterasi

Iterasi dihentikan ketika peningkatan performa model tidak lagi signifikan atau residual yang dihasilkan mendekati nol, yang mengindikasikan tercapainya konvergensi.

2.3.3 Adaptive Boosting (AdaBoost)

Adaptive Boosting (AdaBoost) merupakan salah satu metode dalam teknik *boosting* yang memanfaatkan sejumlah *weak learners* untuk membentuk model klasifikasi yang lebih kuat. Algoritma ini pertama kali diperkenalkan oleh (Freund dan Schapire, 1999) dan hingga saat ini masih banyak digunakan karena kesederhanaan serta efektivitasnya dalam meningkatkan kinerja model klasifikasi. Ad-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

AdaBoost dikenal sebagai algoritma *boosting* pertama yang berhasil, dengan *base learner* umumnya berupa *decision tree* sederhana dengan satu kali pembelahan, yang sering disebut sebagai *decision stump* karena kedalaman pohonnya yang sangat terbatas (Kuhn dan Johnson, 2013; Mienye dan Sun, 2022).

Proses pembelajaran pada AdaBoost dilakukan secara iteratif dengan melatih *base classifier* secara berurutan, di mana bobot setiap sampel dalam data pelatihan disesuaikan berdasarkan hasil prediksi pada iterasi sebelumnya (Mienye dan Sun, 2022). Sampel yang salah diklasifikasikan akan diberikan bobot yang lebih besar, sehingga memperoleh perhatian lebih pada proses pelatihan selanjutnya. Salah satu keunggulan AdaBoost terletak pada kemudahan implementasinya serta kebutuhan penyetelan *hyperparameter* yang relatif minimal (Wu dan Zhao, 2011). AdaBoost bekerja sama dengan berbagai algoritma pembelajaran untuk meningkatkan hasil kinerjanya (Shakir, Aziz, Al, dan Alkhazraji, 2025). Dalam setiap iterasi, *base learner* ditambahkan ke dalam ensemble dan diberi bobot tertentu untuk membentuk sebuah *strong classifier* yang memiliki kinerja lebih baik dibandingkan model tunggal (F. Wang dkk., 2019). Secara umum, mekanisme pembelajaran AdaBoost mengikuti tahapan matematis sebagaimana dirumuskan pada Persamaan (2.10)–(2.14) (Mienye dan Sun, 2022).

1. Inisialisasi Bobot Sampel

Pada tahap awal, setiap sampel dalam data pelatihan diberikan bobot yang sama. Untuk *dataset* yang terdiri atas m sampel, bobot awal masing-masing sampel i didefinisikan sebagai berikut:

$$D_1(i) = \frac{1}{n}, \quad i = 1, 2, \dots, m \quad (2.10)$$

dengan n menyatakan jumlah total sampel dalam data pelatihan. Bobot ini menentukan kontribusi setiap sampel dalam proses pelatihan *weak learner* pada iterasi awal.

2. Pembaruan Bobot pada Setiap Iterasi

Pada iterasi ke- t , bobot sampel diperbarui berdasarkan kesalahan prediksi yang dilakukan oleh *weak learner* $h_t(x)$. Proses pembaruan bobot dirumuskan sebagai berikut:

$$D_{t+1}(i) = Z_t \cdot \exp(-\alpha_t \cdot y_i \cdot h_t(x_i)) \quad (2.11)$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dengan:

- α_t sebagai bobot model pada iterasi ke- t ,
- y_i sebagai label target untuk sampel ke- i ,
- $h_t(x_i)$ sebagai hasil prediksi *weak learner* ke- t ,
- Z_t sebagai faktor normalisasi untuk memastikan distribusi bobot tetap terjaga.

3. Perhitungan Bobot Model

Bobot model α_t ditentukan berdasarkan tingkat kesalahan (ϵ_t) dari *weak learner* pada iterasi ke- t , yang dihitung menggunakan persamaan:

$$\alpha_t = \frac{1}{2} \cdot \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (2.12)$$

dengan tingkat kesalahan ϵ_t dirumuskan sebagai:

$$\epsilon_t = \sum_{i=1}^m D_t(i) \cdot I[h_t(x_i) \neq y_i] \quad (2.13)$$

di mana $I[\cdot]$ merupakan fungsi indikator yang bernilai 1 jika prediksi salah dan 0 jika prediksi benar.

4. Pembentukan Model Akhir

Setelah seluruh iterasi sebanyak T selesai, model akhir $H(x)$ dibentuk dengan menggabungkan seluruh *weak learner* secara berbobot, yang dirumuskan sebagai berikut:

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t \cdot h_t(x) \right) \quad (2.14)$$

dengan $\text{sign}(\cdot)$ sebagai fungsi tanda yang menghasilkan nilai 1 untuk kelas positif dan -1 untuk kelas negatif.

2.3.4 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) merupakan algoritma *ensemble* berbasis *decision tree* yang dikembangkan dalam kerangka kerja *gradient boosting*. Algoritma ini dikenal memiliki tingkat skalabilitas dan akurasi yang tinggi, sehingga banyak digunakan pada permasalahan klasifikasi maupun regresi. Dalam beberapa tahun terakhir, XGBoost telah menjadi salah satu algoritma yang paling dominan dalam penerapan *machine learning* praktis dan banyak digunakan pada



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

berbagai kompetisi analitik, termasuk kompetisi yang diselenggarakan di platform Kaggle. Algoritma ini diperkenalkan oleh T. Chen dan Guestrin (2016) dengan sejumlah pengembangan signifikan dibandingkan metode *gradient boosting* konvensional. Salah satu perbedaan utama XGBoost terletak pada formulasi fungsi kerugiannya, yang secara eksplisit menambahkan istilah *regularization* untuk mengendalikan kompleksitas model dan mencegah terjadinya *overfitting* (Y. Li dan Chen, 2020).

Proses optimasi dalam XGBoost dilakukan dengan memodifikasi fungsi objektif menjadi pendekatan kuadratik, sehingga memungkinkan pemanfaatan informasi gradien dan hessian secara lebih efisien. Keberadaan istilah *regularization* dalam fungsi objektif menjadikan model lebih stabil dan tahan terhadap *overfitting* (Liang, Luo, Zhao, dan Wu, 2020). Sejalan dengan pendekatan *gradient boosting*, XGBoost juga menerapkan berbagai mekanisme pengendalian kompleksitas model, seperti pengaturan *max depth*, *learning rate*, dan *subsampling*. Kombinasi strategi tersebut memungkinkan XGBoost mencapai performa yang lebih baik dengan risiko *overfitting* yang lebih rendah. Secara matematis, tahapan optimasi dan pembentukan model pada XGBoost dapat dijelaskan melalui Persamaan (2.15)–(2.18) (Mienye dan Sun, 2022).

1. Fungsi Objektif XGBoost

Pada iterasi ke- m , fungsi objektif XGBoost dirumuskan sebagai kombinasi antara fungsi kerugian dan term regulasi yang berfungsi untuk mengontrol kompleksitas pohon keputusan. Fungsi tersebut dinyatakan sebagai:

$$L_M(F(x_i)) = \sum_{i=1}^n L(y_i, F(x_i)) + \sum_{m=1}^M \Omega(h_m) \quad (2.15)$$

dengan:

- $F(x_i)$ sebagai nilai prediksi untuk sampel x_i ,
- $L(y_i, F(x_i))$ sebagai fungsi kerugian yang mengukur selisih antara prediksi dan label target y_i ,
- $\Omega(h_m)$ sebagai term regulasi yang memberikan penalti terhadap kompleksitas model pada iterasi ke- m .

2. Term Regulasi

Term regulasi digunakan untuk membatasi kompleksitas pohon keputusan yang dibangun pada setiap iterasi. Formulasi term regulasi $\Omega(h_m)$ dituliskan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

sebagai berikut:

$$\Omega(h) = \gamma T + \lambda \|\omega\|_2^2 \quad (2.16)$$

di mana:

- γ merupakan parameter kompleksitas yang mengontrol pengurangan minimum fungsi kerugian agar suatu simpul dapat dibagi,
- T menyatakan jumlah daun pada pohon keputusan,
- λ merupakan parameter penalti terhadap kompleksitas pohon,
- ω adalah nilai keluaran pada simpul daun pohon keputusan.

3. Aproksimasi Taylor Orde Kedua

Untuk menyederhanakan proses optimasi, XGBoost menggunakan aproksimasi Taylor orde kedua terhadap fungsi kerugian, yang melibatkan gradien dan hessian. Bentuk aproksimasi fungsi objektif tersebut dinyatakan sebagai:

$$L_M \approx \sum_{i=1}^n \left(g_i f_m(x_i) + \frac{1}{2} h_i f_m^2(x_i) \right) + \Omega(h_m) \quad (2.17)$$

dengan:

- g_i sebagai gradien atau turunan pertama fungsi kerugian untuk sampel ke- i ,
- h_i sebagai hessian atau turunan kedua fungsi kerugian untuk sampel ke- i ,
- $f_m(x_i)$ sebagai nilai prediksi pada iterasi ke- m .

4. Perhitungan Kerugian pada Daun

Setelah gradien dan hessian diperoleh, nilai kerugian dihitung untuk setiap daun pohon keputusan. Jika I_j menyatakan himpunan sampel yang berada pada daun ke- j , maka fungsi objektif untuk pohon keputusan dapat dirumuskan sebagai:

$$L_M = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \quad (2.18)$$

dengan ω_j sebagai nilai keluaran pada daun ke- j .

5. Optimasi Bobot Daun

Bobot pada setiap daun pohon keputusan, yaitu ω_j , ditentukan melalui proses optimasi dengan tujuan meminimalkan fungsi objektif yang telah

dirumuskan sebelumnya.

6. Pengendalian *Overfitting*

Untuk mengendalikan risiko *overfitting*, XGBoost menerapkan sejumlah mekanisme, antara lain pembatasan kedalaman maksimum pohon, penggunaan *learning rate* yang lebih kecil, serta penerapan *subsampling* guna menurunkan variansi model.

2.8.5 *Light Gradient Boosting Machine (LightGBM)*

Light Gradient Boosting Machine (LightGBM) merupakan implementasi efisien dari algoritma *gradient boosting* yang dikembangkan pada tahun 2017 oleh peneliti di Microsoft (Ke dkk., 2017). Algoritma ini dirancang untuk menangani berbagai permasalahan pembelajaran mesin, seperti klasifikasi, perankingan, dan regresi, dengan fokus utama pada efisiensi komputasi dan skalabilitas. LightGBM memperkenalkan dua teknik utama, yaitu *Gradient-based One-Sided Sampling (GOSS)* dan *Exclusive Feature Bundling (EFB)*, yang memungkinkan proses pelatihan model berlangsung lebih cepat tanpa mengorbankan akurasi.

Metode GOSS merupakan pengembangan dari pendekatan *gradient boosting* yang menitikberatkan proses pembelajaran pada sampel-sampel dengan nilai gradien yang lebih besar, karena sampel tersebut dianggap memiliki kontribusi yang lebih signifikan terhadap pembaruan model (R. Wang dkk., 2019). Sementara itu, metode EFB bertujuan untuk mengurangi dimensi fitur dengan menggabungkan atribut-atribut yang saling eksklusif dalam ruang fitur yang jarang (*sparse feature space*), sehingga kompleksitas komputasi dapat ditekan secara efektif (Neshatian dan Varn, 2019).

Keunggulan utama LightGBM terletak pada kecepatan pelatihan dan efisiensi penggunaan memori. Algoritma ini mengonversi nilai kontinu menjadi sejumlah *bins* diskret, yang memungkinkan proses pembentukan pohon keputusan dilakukan dengan lebih cepat dan hemat memori. Selain itu, kombinasi teknik GOSS dan EFB menjadikan LightGBM mampu mencapai tingkat akurasi yang kompetitif, bahkan lebih tinggi dibandingkan beberapa metode *boosting* lainnya, khususnya pada *dataset* berukuran besar (Mienye dan Sun, 2022).

2.8.6 *Categorical Boosting (CatBoost)*

Categorical Boosting (CatBoost) merupakan algoritma *machine learning* berbasis *gradient boosting* yang diperkenalkan oleh Prokhorenkova *et al.* pada tahun 2017 (Prokhorenkova, Gusev, Vorobev, Dorogush, dan Gulin, 2018). Algoritma ini dikembangkan untuk mengatasi dua keterbatasan utama pada metode *gra-*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

diend boosting konvensional, yaitu kesulitan dalam menangani fitur kategorikal secara langsung serta tingginya risiko terjadinya *overfitting*. Salah satu inovasi utama CatBoost adalah penggunaan estimasi gradien yang tidak bias, di mana suatu sampel tidak dilibatkan dalam proses pelatihan pada saat gradien untuk sampel tersebut dihitung. Pendekatan ini bertujuan untuk mengurangi fenomena *prediction shift* yang kerap muncul pada algoritma *boosting* tradisional (Tanha, Abdi, Samadi, Razzaghi, dan Asadpour, 2020).

Selain itu, CatBoost memiliki kemampuan bawaan untuk memproses fitur kategorikal tanpa memerlukan tahap pra-pemrosesan tambahan. Fitur kategorikal merupakan atribut dengan nilai diskret yang tidak memiliki hubungan numerik atau ordinal secara langsung, sehingga tidak dapat digunakan secara langsung dalam pembentukan pohon keputusan (Zhang, Zhao, dan Zheng, 2020). Meskipun metode *target encoding* konvensional sering digunakan untuk menangani fitur semacam ini, pendekatan tersebut memiliki kelemahan berupa kecenderungan *overfitting*, terutama ketika suatu kategori hanya memiliki jumlah sampel yang terbatas, karena nilai hasil transformasi sangat dipengaruhi oleh label individual (Niu dkk., 2021).

Untuk mengatasi permasalahan tersebut, CatBoost mengusulkan pendekatan *ordered target encoding* yang lebih stabil dan robust. Algoritma ini terlebih dahulu melakukan permutasi acak terhadap data pelatihan, kemudian menghitung nilai fitur kategorikal berdasarkan rata-rata label dari sampel-sampel sebelumnya dalam urutan permutasi yang memiliki kategori yang sama. Secara matematis, transformasi fitur kategorikal pada CatBoost dapat dirumuskan pada Persamaan (2.19) (Mienye dan Sun, 2022):

$$\text{CatBoost}(x_{\sigma_p,k}) = \frac{\sum_{j=1}^{p-1} \mathbb{I}(x_{\sigma_j,k} = x_{\sigma_p,k}) \cdot y_{\sigma_j} + a \cdot P}{\sum_{j=1}^{p-1} \mathbb{I}(x_{\sigma_j,k} = x_{\sigma_p,k}) + a} \quad (2.19)$$

Rumus 2.19 menggambarkan proses transformasi fitur kategorikal pada CatBoost, di mana P merepresentasikan nilai prior dan $a > 0$ merupakan bobot prior. Penambahan nilai prior bertujuan untuk menstabilkan estimasi pada kategori dengan frekuensi kemunculan yang rendah serta mengurangi pengaruh *noise* pada data, sehingga risiko terjadinya *overfitting* dapat diminimalkan (Lu, Zhang, Xue, Xiao, dan Liu, 2022).



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.9 Synthetic Minority Over-sampling Technique (SMOTE)

Synthetic Minority Over-sampling Technique (SMOTE) merupakan metode *oversampling* yang digunakan untuk menangani permasalahan ketidakseimbangan kelas, khususnya pada kondisi ketika jumlah sampel pada kelas minoritas jauh lebih sedikit dibandingkan dengan kelas mayoritas. Metode ini bekerja dengan membangkitkan sampel sintetis baru pada kelas minoritas melalui proses interpolasi linier antar sampel minoritas yang berdekatan, sehingga distribusi data antar kelas menjadi lebih seimbang (Huang, 2025). Keunggulan utama SMOTE terletak pada kemampuannya dalam memperluas representasi kelas minoritas di ruang fitur serta mengurangi ketergantungan model terhadap pengulangan data yang sama, sehingga risiko *overfitting* dapat ditekan hingga tingkat tertentu (Malhotra, Leslie, Ludwig, dan Bogacz, 2017). Meskipun demikian, SMOTE juga memiliki keterbatasan, terutama pada wilayah ruang fitur yang mengalami tumpang tindih antara kelas minoritas dan mayoritas, di mana sampel sintetis yang dihasilkan berpotensi mengandung *noise* atau memperparah tumpang tindih kelas, sehingga dapat meningkatkan kemungkinan terjadinya kesalahan klasifikasi (Karthikeyan dan Vasuki, 2016; Huang, 2025).

2.10 Augmentasi Data

Data augmentation dalam konteks *machine learning* merujuk pada sekumpulan teknik yang digunakan untuk menghasilkan data buatan dari data yang telah tersedia dengan tujuan meningkatkan ukuran dan keragaman *dataset* pelatihan, terutama pada kondisi data yang terbatas atau tidak seimbang (Shorten dan Khoshgoftaar, 2019). Pendekatan ini bertujuan untuk memperkaya variasi data latih agar model dapat mempelajari pola yang lebih beragam dan membangun representasi yang lebih mendekati karakteristik data sebenarnya (Z. Wang dkk., 2026). Proses augmentasi umumnya dilakukan dengan memodifikasi data untuk menghasilkan titik data baru tanpa memerlukan proses pengumpulan data tambahan.

Selain meningkatkan keberagaman data, augmentasi data juga berperan penting dalam mengurangi risiko *overfitting* serta meningkatkan kemampuan generalisasi model terhadap data baru yang belum pernah diamati sebelumnya (J. Chen, Tam, Raffel, Bansal, dan Yang, 2023). Dengan bertambahnya variasi data latih, model diharapkan mampu menangkap pola yang lebih kompleks dan stabil, sehingga performa model *machine learning* dapat ditingkatkan secara lebih optimal, khususnya pada penelitian yang menghadapi keterbatasan jumlah data dan variasi sampel (Omoniyi dkk., 2025).



2.11 Confusion Matrix

Confusion matrix merupakan representasi visual yang digunakan untuk menunjukkan perbandingan antara kelas aktual dan kelas hasil prediksi yang dihasilkan oleh suatu algoritma *classification* (Yang dan Berdine, 2024). Melalui matriks ini, kinerja model dapat diamati secara menyeluruh dengan menampilkan jumlah data yang berhasil diklasifikasikan secara benar maupun yang mengalami kesalahan prediksi. Informasi yang disajikan dalam *confusion matrix* mencakup empat komponen utama, yaitu *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN) (Rois, Ray, Rahman, dan Roy, 2021).

Berdasarkan *confusion matrix*, berbagai metrik evaluasi kinerja model dapat dihitung, seperti *accuracy*, *precision*, *recall*, dan *F1-Score* (Amin, 2023). Metrik-metrik tersebut digunakan untuk menilai sejauh mana efektivitas algoritma *classification* dalam melakukan prediksi secara tepat (Rois dkk., 2021). Konsep *confusion matrix* yang awalnya diterapkan pada kasus *binary classification* juga dapat diperluas penggunaannya pada permasalahan *multiclass classification* dengan prinsip perhitungan yang sama (Amin, 2023). Oleh karena itu, parameter kinerja yang digunakan dalam penelitian ini dirumuskan sebagaimana ditunjukkan pada Persamaan 1–4 (Priya dkk., 2020; Sathyanarayanan, 2024; Singh dkk., 2024; Yin dkk., 2023).

Dalam *confusion matrix*, terdapat empat komponen utama yang mendeskripsikan hasil klasifikasi model (Priya dkk., 2020; Sathyanarayanan, 2024; Yin dkk., 2023).

1. *True Positive* (TP): Data yang berasal dari suatu kelas dan berhasil diprediksi sesuai dengan kelas tersebut.
2. *False Negative* (FN): Data yang seharusnya termasuk ke dalam suatu kelas, tetapi diprediksi sebagai kelas lain.
3. *False Positive* (FP): Data yang tidak berasal dari suatu kelas, namun keliru diprediksi sebagai kelas tersebut.
4. *True Negative* (TN): Data yang tidak berasal dari suatu kelas dan berhasil diprediksi dengan benar.

2.12 Receiver Operating Characteristic (ROC)

Receiver Operating Characteristic (ROC) *Curve* merupakan metode evaluasi visual yang digunakan untuk menilai kinerja model klasifikasi secara intuitif (Vercellis, 2009). Kurva ini memetakan nilai *false positive rate* pada sumbu horizontal dan *true positive rate* pada sumbu vertikal. Pembentukan kurva ROC di-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

lakukan dengan memvariasikan nilai ambang batas pada skor prediksi, sehingga setiap nilai ambang menghasilkan titik yang berbeda pada kurva (Fawcett, 2006). Pada kurva ROC, *true positive rate* yang juga dikenal sebagai *sensitivity* atau *recall* ditempatkan pada sumbu y, sedangkan *false positive rate* berada pada sumbu x. Luas area di bawah kurva ROC, yang dikenal sebagai *Area under the Curve* (ROC-AUC) atau *c statistic*, merupakan salah satu ukuran yang paling umum digunakan dalam penelitian ilmiah untuk mengevaluasi kinerja model *binary classification*, dengan nilai berkisar antara 0 sebagai hasil terburuk hingga 1 sebagai hasil sempurna (Chicco dan Jurman, 2023).

2.13 Area Under the Curve (AUC)

Area Under the Curve (AUC) merupakan ukuran yang menyatakan luas area di bawah kurva ROC dan digunakan untuk merepresentasikan kemampuan model dalam membedakan kelas secara keseluruhan. Nilai AUC diperoleh dengan menghitung luas trapesium yang terbentuk di bawah kurva ROC. Secara teoritis, nilai AUC berada pada rentang 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan bahwa model memiliki kemampuan klasifikasi yang semakin baik (Bradley, 1997). Untuk memudahkan interpretasi, nilai AUC umumnya diklasifikasikan ke dalam beberapa kategori sebagai berikut (Gorunescu, 2011).

1. 0.90 – 1.00 = Klasifikasi Sangat Baik
2. 0.80 – 0.90 = Klasifikasi Baik
3. 0.70 – 0.80 = Kualitas Cukup
4. 0.60 – 0.70 = Kualitas Buruk
5. 0.50 – 0.60 = Kualitas Sangat Buruk

2.14 Fakultas Sains dan Teknologi

Sebagai bagian dari persiapan untuk mengubah status Institut Agama Islam Negeri Sultan Syarif Qasim (IAIN SUSQA) Pekanbaru menjadi Universitas Islam Negeri Sultan Syarif Kasim Riau (UIN Suska) Riau, Fakultas Sains dan Teknologi didirikan pada akhir tahun 2001. Fakultas ini awalnya didirikan dengan Program Studi Teknik Informatika yang dibuka pada tahun 1999. Pada tahun 2001, Program Studi Teknik Industri juga dibuka. Lima program studi ditawarkan di Fakultas Sains dan Teknologi saat ini seiring dengan perkembangan institusi, yaitu Teknik Informatika, Teknik Industri, Sistem Informasi, Matematika, dan Teknik Elektro.

Berdasarkan data pelaporan resmi yang bersumber dari Pangkalan Data Pendidikan Tinggi (PDDikti) Kemdiktisaintek Semester Genap Tahun 2024, jumlah mahasiswa di Fakultas Sains dan Teknologi dapat dilihat pada Tabel 2.3.

Tabel 2.3. Populasi Mahasiswa Setiap Program Studi

No	Kode	Program Studi	Akreditasi	Jumlah Mahasiswa
1	55201	Teknik Informatika	Unggul	629
2	26201	Teknik Industri	Baik Sekali	470
3	57201	Sistem Informasi	Baik Sekali	454
4	44201	Matematika	Unggul	169
5	20201	Teknik Elektro	Baik Sekali	482
Total				2.204

2.15 Penelitian Terdahulu

Duangchaemkarn dkk. (2024) mengkaji efektivitas berbagai model *machine learning* dalam mengklasifikasikan kondisi emosional mahasiswa, khususnya depresi, kecemasan, dan stres, berdasarkan profil DASS-21. Penelitian ini menguji enam model ML menggunakan 2.602 data yang diperoleh dari Pusat Konseling Universitas Phayao. Hasil analisis menunjukkan bahwa model *Deep Learning* menghasilkan tingkat akurasi dan presisi tertinggi, dengan nilai di atas 90% untuk seluruh kategori emosional serta nilai AUC yang melebihi 99%. Temuan ini menunjukkan potensi pendekatan berbasis pembelajaran mendalam dalam klasifikasi kondisi emosional menggunakan DASS-21 sebagai instrumen pengukuran.

M. J. Hasan dkk. (2024) meneliti kondisi kesehatan mental mahasiswa dalam konteks akademik dengan menerapkan teknik augmentasi data dan model *machine learning* untuk memprediksi tingkat keparahan berdasarkan respons kuesioner DASS-21. *Dataset* yang digunakan berasal dari 96 mahasiswa American International University–Bangladesh (AIUB) dan dianalisis menggunakan beberapa algoritma ML, termasuk *Random Forest*, *Logistic Regression*, *Support Vector Machine* (SVM), dan *Gradient Boosting*. Hasil penelitian menunjukkan bahwa *Logistic Regression* memiliki kinerja paling unggul berdasarkan metrik *F1-score*, *recall*, akurasi, dan presisi, dengan tingkat akurasi sebesar 0.97 untuk kecemasan, 0.975 untuk depresi, dan 0.93 untuk stres.

Herman dkk. (2025) melakukan klasifikasi gangguan kecemasan pada remaja dengan memanfaatkan tiga algoritma pembelajaran mesin, yaitu *Naive Bayes*, *C4.5*, dan *K-Nearest Neighbors* (K-NN). Data penelitian diperoleh dari kuesioner DASS-21, yang mencakup 418 sampel dan 11 fitur, termasuk tujuh item yang berkaitan dengan kecemasan. Pengujian dilakukan menggunakan teknik *holdout* dengan beberapa skema pembagian data dan validasi silang *k-fold*. Hasil penelitian menunjukkan bahwa algoritma *C4.5* adalah yang terbaik, mengungguli *Naive*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Bayes dan K-NN, dengan akurasi 100% pada skema *holdout* dan 98% pada validasi silang.

Srinath dkk. (2022) melakukan analisis terhadap kinerja algoritma *machine learning* yang digunakan untuk memprediksi tingkat depresi, kecemasan, dan stres dengan menggunakan instrumen *Depression, Anxiety, and Stress Scale* (DASS42). Studi ini membandingkan dua algoritma utama yaitu *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Fokus penelitian adalah untuk meningkatkan akurasi klasifikasi melalui penyetelan parameter. Nilai akurasi LR masing-masing mencapai 98.15% untuk depresi, 98.05% untuk kecemasan, dan 98.45% untuk stres.

S. Usman dkk. (2022) mengombinasikan pengukuran kekuatan genggaman tangan (HGS) dengan *Depression Anxiety Stress Scale* (DASS) sebagai metode penayangan awal tingkat stres. Setelah data dinormalisasi, algoritma *Random Forest* dan *Support Vector Machine* (SVM) digunakan untuk menganalisis data yang diperoleh. Hasil pengujian menunjukkan bahwa *Random Forest* memiliki akurasi 93.75%, sensitivitas 93.80%, dan spesifisitas 94.90% dibandingkan dengan SVM.

Bhatnagar dkk. (2023) mengembangkan kuesioner berbasis skala Likert yang disusun dengan mengacu pada GAD-7 dan DASS-21 untuk mengukur tingkat kecemasan mahasiswa teknik di India. Penelitian ini melibatkan 127 peserta dan menggunakan algoritma klasifikasi seperti SVM, *Random Forest*, *Decision Tree*, dan *Naive Bayes*. Hasilnya menunjukkan bahwa *Random Forest* memiliki akurasi tertinggi sebesar 78.9%, diikuti oleh SVM dengan akurasi 75.5%.

Anand dkk. (2023) mengusulkan pendekatan pembelajaran *ensemble* untuk mengklasifikasikan tingkat stres mahasiswa dengan mempertimbangkan berbagai faktor perilaku dan akademik, seperti durasi tidur, waktu layar, beban tugas, dan metode belajar. Data survei digunakan untuk mengelompokkan tingkat stres ke dalam tiga kategori, yaitu stres tinggi, stres terkelola, dan tanpa stres. Beberapa algoritma diterapkan, termasuk *Decision Tree*, *Random Forest*, AdaBoost, dan *Gradient Boosting*, dengan hasil akurasi mencapai 93.48% dan *F1-score* sebesar 93.14%.

Daza, Arroyo-Paz, Bobadilla, Apaza, dan Pinto (2023) memprediksi tingkat kecemasan mahasiswa menggunakan pendekatan *Stacking* dengan mengombinasikan beberapa model klasifikasi. *Dataset* yang digunakan terdiri dari 284 mahasiswa dan diproses menggunakan bahasa pemrograman *Python*, termasuk penerapan teknik *oversampling* untuk menyeimbangkan data. Setelah pelatihan menggunakan metode *cross-validation*, model *Stacking 4A* menunjukkan kinerja terbaik dengan akurasi 97.83%, sensitivitas 98.44%, dan *F1-score* 97.88%.

Daza Vergaray, Miranda, Cornelio, López Carranza, dan Ponce Sánchez



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

(2023) menerapkan teknik *Stacking* dalam memprediksi depresi pada mahasiswa Teknik Komputer dan Sistem dari universitas negeri. *Dataset* yang digunakan terdiri dari 284 responden dan melalui tahap pembersihan serta penyeimbangan data menggunakan metode *oversampling*. Dengan menerapkan *cross-validation*, model *Stacking* 1 menunjukkan performa terbaik dengan akurasi 94.69%, sensitivitas 94.22%, dan *F1-score* 94.12%, yang menegaskan keunggulan pendekatan *ensemble* dibandingkan model individual.

M. E. Hasan dkk. (2025) mengevaluasi prevalensi depresi, kecemasan, dan stres pada mahasiswa universitas di Bangladesh sekaligus menilai kinerja prediktif berbagai model pembelajaran mesin. Survei dilakukan terhadap 1.697 mahasiswa dari dua universitas negeri, yang menunjukkan prevalensi depresi sebesar 56.9%, kecemasan 69.5%, dan stres 32.2%. Dalam pemodelan prediktif, SVM memberikan akurasi tertinggi untuk depresi, *Logistic Regression* paling efektif untuk kecemasan, sementara CatBoost menunjukkan performa terbaik dalam memprediksi stres.

El Morr dkk. (2024) mengembangkan model prediktif berbasis *machine learning* untuk menilai tingkat depresi, kecemasan, dan stres mahasiswa di Lebanon selama masa *pandemic*. Penilaian dilakukan menggunakan instrumen PHQ-9, BAI, dan PSS, dengan delapan algoritma klasifikasi yang diuji. Hasil penelitian menunjukkan bahwa *Random Forest* menghasilkan kinerja terbaik untuk depresi dengan nilai AUC sebesar 78.27%, sementara *Naive Bayes* dan AdaBoost lebih unggul dalam memprediksi kecemasan dan stres.

Chowdhury, Rad, dan Rahman (2024) memanfaatkan tiga instrumen psikologis, yaitu GAD-7, PHQ-9, dan ISI-7, untuk memprediksi tingkat depresi, kecemasan, dan insomnia pada mahasiswa di Bangladesh. Beberapa algoritma diterapkan, dengan XGBoost menunjukkan performa terbaik melalui akurasi sebesar 86% dan nilai AUC mencapai 94%. Temuan ini menegaskan keterkaitan antara gangguan tidur dan kondisi psikologis mahasiswa.

Vitória dkk. (2024) mengeksplorasi penggunaan algoritma *machine learning* dalam memprediksi skor kecemasan dan depresi berdasarkan kuesioner PROMIS®. Studi komparatif terhadap beberapa model *supervised learning* menunjukkan bahwa model yang diusulkan mampu mencapai rata-rata MAPE sebesar 6.31%, nilai R^2 sebesar 0.76, serta koefisien *Spearman* sebesar 88.86%, yang melampaui performa model linier tradisional seperti SVM, *Random Forest*, dan *Gradient Boosting*.

Rois dkk. (2021) menganalisis tingkat stres mahasiswa serta faktor-faktor yang memengaruhinya dengan melibatkan 355 mahasiswa di Bangladesh. Bebe-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

rapa algoritma klasifikasi diterapkan, antara lain *Decision Tree*, *Random Forest*, *SVM*, dan *Logistic Regression*. Hasil analisis menunjukkan bahwa *Random Forest* merupakan model dengan performa terbaik, dengan tingkat akurasi mencapai 89.72%. Faktor utama yang berkontribusi terhadap stres meliputi kualitas tidur, tekanan darah, detak jantung, kebiasaan merokok, dan latar belakang akademik.

Kumar dkk. (2024) melakukan deteksi dini potensi gangguan kesehatan mental pada lebih dari 15.000 mahasiswa dengan mempertimbangkan faktor akademik, perilaku, sosial, dan digital. Beberapa algoritma diterapkan, termasuk *Logistic Regression*, *Random Forest*, *XGBoost*, *SVM* dengan kernel RBF, *Feed-Forward Neural Network*, dan *LSTM*. Hasil penelitian menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan *F1-score* sebesar 0.811 dan *AUC* sebesar 0.892, dengan variabel kunci meliputi *IPK*, beban akademik, dan keterlibatan sosial.

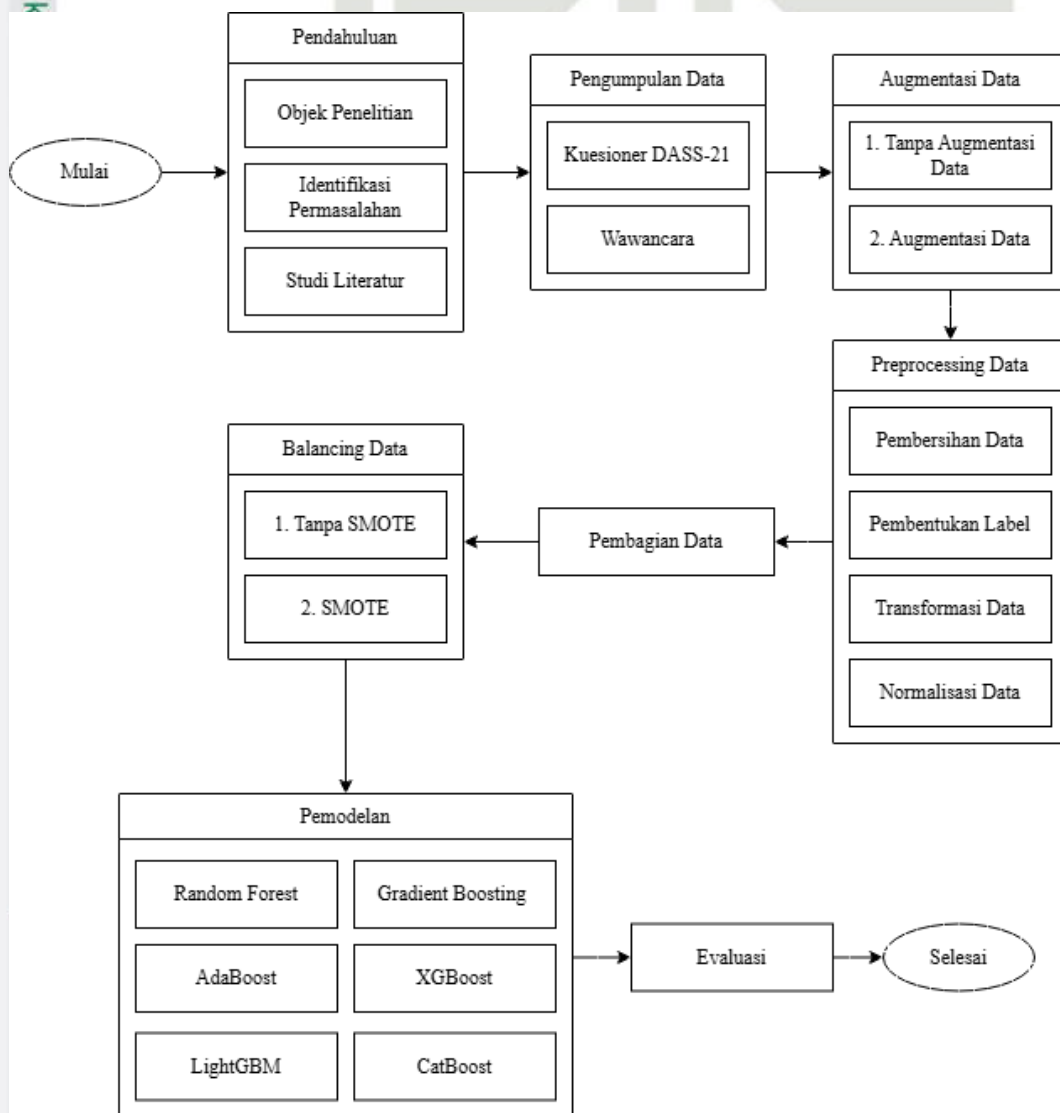
Farooq dkk. (2023) mengembangkan model pembelajaran mesin untuk melakukan diagnosis dini kecemasan menggunakan *dataset real-time* yang dikumpulkan di wilayah Kashmir. Ada sejumlah algoritma yang digunakan, seperti *Naive Bayes*, *Logistic Regression*, *Random Forest*, *AdaBoost*, *Gradient Boosting Classifier*, dan *LightGBM*. Hasil pengujian menunjukkan bahwa model yang diusulkan mampu memprediksi kecemasan dengan tingkat akurasi 95% pada skema pembagian data 70:30.

Singh dan Kumar (2021) mengkaji pemanfaatan teknik pembelajaran mesin dan kecerdasan buatan dalam mengidentifikasi aktivitas emosional yang berkaitan dengan kondisi psikologis manusia. Penelitian ini menunjukkan relevansi pendekatan pembelajaran mesin dalam mengenali berbagai emosi, seperti kesedihan, kemarahan, dan kebahagiaan, menggunakan beragam jenis data. Instrumen *DASS-21* digunakan untuk mengidentifikasi gejala depresi dan kecemasan dalam kuesioner, dengan melakukan evaluasi performa lima algoritma klasifikasi: *SVM*, *Decision Tree*, *Random Forest*, *Naive Bayes*, dan *KNN*.

BAB 3

METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* untuk mengklasifikasikan tingkat depresi, kecemasan, dan stres berdasarkan hasil pengisian kuesioner DASS-21. Tahapan penelitian ini disusun secara sistematis dan mencakup pendahuluan, pengumpulan data, tahap *preprocessing*, pembagian data, proses pemodelan klasifikasi, serta evaluasi kinerja model. Proses klasifikasi dilakukan dengan membandingkan kinerja beberapa algoritma, yaitu *Random Forest*, *Gradient Boosting*, *AdaBoost*, *XGBoost*, *LightGBM*, dan *CatBoost*. Gambar 3.1 menunjukkan alur metodologi penelitian secara keseluruhan.



Gambar 3.1. Alur Metodologi Penelitian

3.1 Pendahuluan

Tahap pendahuluan merupakan landasan awal dalam pelaksanaan penelitian ini. Pada tahap ini, rangkaian kegiatan disusun secara sistematis agar proses penelitian dapat berjalan terarah dan selaras dengan tujuan yang telah ditetapkan. Pendahuluan mencakup sejumlah kegiatan awal yang menjadi pijakan dalam pelaksanaan tahapan penelitian selanjutnya, yang meliputi penentuan objek penelitian, identifikasi permasalahan, serta studi literatur.

3.1.1 Objek Penelitian

Objek penelitian ini adalah mahasiswa aktif Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Pemilihan objek penelitian difokuskan pada mahasiswa yang telah menjalani aktivitas perkuliahan secara berkelanjutan, sehingga diasumsikan memiliki pengalaman akademik yang cukup untuk merefleksikan kondisi psikologis selama masa studi. Mahasiswa Fakultas Sains dan Teknologi dipilih karena karakteristik bidang keilmuannya yang menuntut kemampuan analitis, pemecahan masalah, serta intensitas akademik yang relatif tinggi. Karakteristik tersebut dinilai relevan dalam menggambarkan tingkat depresi, kecemasan, dan stres yang dapat muncul sebagai respons terhadap dinamika dan tuntutan perkuliahan, khususnya dalam menghadapi beban tugas, evaluasi akademik, serta tekanan pencapaian prestasi.

3.1.2 Identifikasi Permasalahan

Identifikasi permasalahan dalam penelitian ini berfokus pada rendahnya kesadaran mahasiswa terhadap kondisi kesehatan mental, seiring dengan meningkatnya indikasi stres, kecemasan, dan depresi di lingkungan akademik. Kondisi tersebut berpotensi memberikan dampak negatif terhadap performa akademik, kesejahteraan psikologis, serta kualitas hidup mahasiswa secara menyeluruh. Proses identifikasi permasalahan dilakukan melalui observasi awal terhadap lingkungan perkuliahan, disertai dengan penelaahan terhadap gejala umum yang kerap dialami mahasiswa, seperti tekanan akademik, kelelahan mental, kesulitan mengelola waktu, serta tuntutan studi yang semakin kompleks dan kompetitif.

3.1.3 Studi Literatur

Studi literatur dilakukan sebagai landasan teoritis dan metodologis dalam pelaksanaan penelitian ini. Kajian literatur mencakup pembahasan mengenai konsep dasar depresi, kecemasan, dan stres, serta instrumen pengukuran kesehatan mental menggunakan DASS-21 yang telah banyak digunakan dan tervalidasi dalam



penelitian di bidang psikologi dan kesehatan mental. Selain itu, studi literatur juga difokuskan pada penerapan algoritma *machine learning*, khususnya pendekatan *ensemble learning* seperti *Random Forest*, *Gradient Boosting*, *AdaBoost*, *XGBoost*, *LightGBM*, dan *CatBoost*, dalam konteks klasifikasi kesehatan mental. Hasil kajian ini menjadi acuan utama dalam perancangan metode penelitian, pemilihan algoritma, serta penentuan strategi evaluasi yang sesuai dengan karakteristik data.

3.2 Pengumpulan Data

Pengumpulan data merupakan tahap krusial dalam penelitian ini karena kualitas data yang diperoleh akan sangat memengaruhi hasil analisis dan kinerja model klasifikasi. Data dikumpulkan untuk merepresentasikan kondisi kesehatan mental mahasiswa secara objektif dan kontekstual, sehingga dapat digunakan secara andal dalam proses pemodelan dan evaluasi. Dua metode utama yang digunakan dalam pengumpulan data adalah kuesioner dan wawancara, yang saling melengkapi antara informasi kuantitatif dan pemahaman kontekstual terhadap kondisi responden.

3.2.1 Kuesioner

Pengumpulan data utama dalam penelitian ini dilakukan menggunakan kuesioner *Depression Anxiety Stress Scales (DASS-21)* yang dibagikan kepada mahasiswa Fakultas Sains dan Teknologi. Kuesioner disebarakan secara *online* melalui *Google Forms* serta secara *offline* untuk menjangkau responden yang lebih luas dan beragam. Pendekatan ini memungkinkan proses pengumpulan data dilakukan secara efisien sekaligus tetap memperhatikan kenyamanan responden. Data yang terkumpul selanjutnya diolah menggunakan pendekatan *data mining* dan *machine learning* untuk memodelkan klasifikasi tingkat depresi, kecemasan, dan stres mahasiswa secara sistematis.

3.2.2 Wawancara

Wawancara dilakukan dengan melibatkan pakar psikologi klinis sebagai narasumber untuk memperoleh validasi konseptual dan kontekstual terhadap instrumen penelitian yang digunakan. Tujuan dari kegiatan ini adalah untuk memastikan bahwa kuesioner *DASS-21* yang digunakan sesuai untuk mengukur depresi, kecemasan, dan stres mahasiswa, baik dari segi substansi pertanyaan maupun interpretasi hasil skor. Selain itu, wawancara juga dimanfaatkan untuk menggali pandangan pakar mengenai faktor-faktor yang berkontribusi terhadap gangguan kesehatan mental pada mahasiswa, seperti tekanan akademik, tuntutan sosial, manajemen waktu, serta dinamika transisi kehidupan perkuliahan. Informasi yang diperoleh

- ### 3.3

3.3

3.3

3.3

3.3

3.3

3.3

3.3

Tabel 3.1. Teknik dan Parameter Augmentasi Data

No	Teknik	Parameter	Nilai	Keterangan
1	<i>Gaussian Noise</i>	<code>noise_std</code>	0.5	Standar deviasi dari Gaussian noise yang ditambahkan ke setiap item DASS-21.
2	<i>Data Replication</i>	<code>augment_count</code>	3	Jumlah data sintetis yang dihasilkan dari setiap data asli melalui proses augmentasi.
3	<i>Value Clipping</i>	<code>value_range</code>	0–3	Pembatasan nilai hasil augmentasi agar tetap berada dalam rentang skala DASS-21.
4	<i>Rounding</i>	<code>rounding_type</code>	<i>integer</i>	Pembulatan nilai ke bilangan bulat untuk menjaga validitas respons kuesioner.

3.4 Preprocessing Data

Selama proses klasifikasi, data yang diperoleh dari hasil observasi dan pengisian kuesioner tidak digunakan secara langsung. Tahap pra-pemrosesan data (*preprocessing*), diperlukan untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik dan layak dianalisis. Tahap ini dilakukan untuk meminimalkan kesalahan yang dapat memengaruhi kinerja model. Secara umum, terdapat beberapa proses utama yang dilakukan dalam tahap pra-pemrosesan data, sebagaimana dijelaskan pada subbagian berikut.

3.4.1 Pembersihan Data

Data dibersihkan untuk menghindari konten yang tidak lengkap, duplikat, atau tidak konsisten. Pada tahap ini, dilakukan pemeriksaan terhadap keberadaan nilai kosong (*missing value*) serta kesesuaian format data pada setiap atribut. Data yang tidak memenuhi kriteria kelayakan analisis, seperti data yang tidak lengkap atau tidak valid, akan diperbaiki apabila memungkinkan atau dihapus dari *dataset* agar tidak memengaruhi proses pemodelan dan hasil klasifikasi.

3.4.2 Pembentukan Label

Data respons kuesioner DASS-21 selanjutnya diolah untuk membentuk skor pada masing-masing dimensi kesehatan mental, yaitu depresi, kecemasan, dan stres. Skor untuk setiap dimensi diperoleh melalui penjumlahan nilai dari tujuh item per-



tanyaan yang relevan, kemudian dikalikan dengan faktor pengali sesuai dengan pedoman DASS-21. Nilai skor yang dihasilkan selanjutnya dikonversi ke dalam kategori tingkat keparahan, yaitu Normal, Ringan, Sedang, Berat, dan Sangat Berat. Proses pembentukan label ini bertujuan untuk mengubah data numerik menjadi kelas kategorikal yang akan digunakan sebagai variabel target dalam proses klasifikasi.

3.4.3 Transformasi Data

Beberapa atribut dalam *dataset*, seperti jenis kelamin dan program studi, bersifat kategorikal sehingga tidak dapat diproses secara langsung oleh algoritma *ensemble learning*. Oleh karena itu, dilakukan proses transformasi data menggunakan teknik *encoding*. Pada penelitian ini, metode `OrdinalEncoding` digunakan untuk mengonversi nilai kategorikal menjadi representasi numerik, sehingga atribut tersebut dapat digunakan sebagai fitur masukan dalam model klasifikasi.

3.4.4 Normalisasi Data

Fitur numerik dalam *dataset*, seperti usia, semester, IPK, serta skor item DASS-21, memiliki rentang nilai yang berbeda-beda. Perbedaan skala ini berpotensi menyebabkan fitur tertentu mendominasi proses pembelajaran model. Untuk mengatasi hal tersebut, dilakukan normalisasi data menggunakan teknik `StandardScaler`. Proses normalisasi ini bertujuan untuk menyamakan skala fitur sehingga algoritma *ensemble learning* dapat bekerja secara lebih stabil dan menghasilkan performa yang lebih optimal.

3.5 Pembagian Data

Setelah tahapan pra-pemrosesan data selesai, *dataset* dibagi menjadi data latih dan data uji menggunakan metode *hold-out*. Teknik *sampling stratified* digunakan untuk membagi 80% *dataset* sebagai data latih dan 20% sebagai data uji, sehingga proporsi *sampling* kedua subset data tetap seimbang. Tujuan dari metode ini adalah untuk mencegah bias yang dapat disebabkan oleh distribusi kelas yang tidak merata serta memastikan bahwa semua tingkat depresi, kecemasan, dan stres diwakili baik dalam data latih maupun uji. Model klasifikasi dibangun dan dilatih dengan data latih. Sementara itu, data uji digunakan untuk mengevaluasi kinerja model secara objektif dan mengukur seberapa baik model dapat digeneralisasi terhadap data yang belum pernah digunakan sebelumnya.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.6 Balancing Data

Balancing data diterapkan dalam penelitian ini sebagai salah satu strategi pengolahan data untuk mengatasi permasalahan ketidakseimbangan distribusi kelas pada *dataset* DASS-21. Ketidakseimbangan kelas dapat menyebabkan model klasifikasi mempelajari pola dari kelas mayoritas dengan lebih dominan. Akibatnya, kemampuan model untuk memahami pola dari kelas minoritas menjadi kurang optimal. Kondisi ini dapat mengurangi kinerja model, terutama untuk metrik evaluasi seperti *recall* dan *F1-score*. Oleh karena itu, penelitian ini membandingkan dua kondisi pengolahan data, yaitu tanpa penerapan teknik *balancing* dan dengan penerapan *balancing* menggunakan SMOTE, untuk mengevaluasi pengaruh penyeimbangan kelas terhadap kinerja algoritma *ensemble learning*.

1. Tanpa *Balancing* (Tanpa SMOTE)

Pada skema ini, data digunakan sesuai dengan distribusi kelas aslinya tanpa dilakukan proses penyeimbangan. Pendekatan ini dimaksudkan untuk menggambarkan kinerja dasar (*baseline*) dari masing-masing algoritma *ensemble learning* ketika dihadapkan pada *dataset* DASS-21 yang secara alami memiliki perbedaan jumlah sampel antar kelas depresi, kecemasan, dan stres. Hasil yang diperoleh dari skema tanpa *balancing* selanjutnya digunakan sebagai pembanding dalam menilai efektivitas penerapan teknik penyeimbangan kelas pada skema pengolahan data lainnya.

2. *Balancing* dengan SMOTE

Pada skema ini, *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan sebagai metode penyeimbangan kelas untuk meningkatkan representasi kelas minoritas pada data latih. SMOTE bekerja dengan membangkitkan sampel sintetis melalui proses interpolasi antara satu sampel kelas minoritas dan beberapa tetangga terdekatnya (*k-nearest neighbors*) di ruang fitur. Pendekatan ini memungkinkan terbentuknya data sintetis yang lebih bervariasi serta tetap mencerminkan pola data asli, dibandingkan dengan metode *oversampling* sederhana yang hanya melakukan penggandaan data. Penerapan SMOTE dilakukan secara eksklusif pada data latih setelah proses pembagian data (*train-test split*) untuk menghindari terjadinya kebocoran informasi (*data leakage*).

Adapun skema pengolahan data yang digunakan dalam penelitian ini disajikan pada Tabel 3.2. Tabel tersebut merangkum tiga skenario pengolahan data yang diterapkan, yaitu data asli, penyeimbangan kelas menggunakan SMOTE, dan augmentasi data, yang masing-masing digunakan untuk mengevaluasi pengaruh



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

perlakuan data terhadap kinerja model klasifikasi.

Tabel 3.2. Skema Pengolahan Data

No	Skema Data	Data Asli	SMOTE	Augmentasi	Keterangan
1	Data Asli	✓	–	–	<i>Dataset</i> digunakan dalam kondisi awal tanpa perlakuan tambahan, sehingga distribusi kelas dan jumlah data tetap sesuai dengan data hasil pengumpulan kuesioner DASS-21.
2	SMOTE	✓	✓	–	<i>Dataset</i> diproses dengan teknik penyeimbangan kelas menggunakan SMOTE pada data latih untuk mengatasi ketidakseimbangan distribusi kelas tanpa menambah variasi pola data.
3	Augmentasi	✓	–	✓	<i>Dataset</i> diperluas dengan penambahan data sintetis melalui teknik augmentasi berbasis <i>Gaussian noise</i> untuk memperkaya variasi data tanpa menerapkan penyeimbangan kelas.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.7 Pemodelan

Tahap pemodelan klasifikasi merupakan tahapan inti dalam penelitian ini yang berfokus pada penerapan algoritma *ensemble learning* untuk membuat model klasifikasi berdasarkan data yang telah dipersiapkan pada tahap sebelumnya. Pada tahap ini, setiap algoritma diuji menggunakan tiga skenario pengolahan data, yaitu data asli, data yang telah melalui proses penyeimbangan menggunakan SMOTE, serta data hasil augmentasi. Pengujian pada berbagai skenario tersebut bertujuan untuk menganalisis pengaruh perlakuan data terhadap performa model klasifikasi yang dihasilkan.

3.7.1 Random Forest

Random Forest adalah algoritma klasifikasi berbasis kelompok pembelajaran yang menggunakan pendekatan *bagging*. Untuk menghasilkan keputusan akhir, *Random Forest* membangun sejumlah pohon keputusan secara independen dan menggabungkan hasil prediksinya. Dalam penelitian ini, *Random Forest* diuji dengan mengubah jumlah pohon (`n_estimator`) sebagai parameter utama. Parameter lain yang digunakan termasuk pengaturan kedalaman maksimum pohon (`max_depth`), jumlah sampel minimum pada setiap daun (`min_samples_leaf`), kedalaman maksimum pohon (`max_depth`), jumlah minimum sampel pada setiap daun (`min_samples_leaf`), pembobotan kelas (`class_weight`) untuk mengakomodasi ketidakseimbangan data, pengaturan `random state`, serta pemanfaatan kemampuan komputasi melalui `parallel processing`. Algoritma ini digunakan untuk mengevaluasi kestabilan model serta kemampuan generalisasi dalam mengklasifikasikan tingkat kesehatan mental mahasiswa.

3.7.2 Gradient Boosting

Gradient Boosting merupakan algoritma *ensemble learning* berbasis *boosting* yang membangun model secara bertahap dengan tujuan meminimalkan kesalahan prediksi dari model yang dibangun sebelumnya. Pada penelitian ini, *Gradient Boosting* diuji dengan memvariasikan jumlah estimator (`n_estimators`) sebagai parameter utama. Parameter lain yang digunakan mencakup `learning rate`, kedalaman maksimum pohon (`max_depth`), serta pengaturan `random state`. Algoritma ini dimanfaatkan untuk menganalisis kemampuan model dalam menangkap hubungan *non-linear* yang terdapat pada data DASS-21.



3.7.3 AdaBoost

AdaBoost merupakan algoritma *boosting* yang bekerja dengan menyesuaikan bobot data secara adaptif, di mana sampel yang salah diklasifikasikan pada iterasi sebelumnya akan diberikan bobot yang lebih besar pada iterasi berikutnya. Dalam penelitian ini, AdaBoost diuji dengan memvariasikan jumlah estimator (`n_estimators`). Parameter lain yang digunakan meliputi `learning rate` dan pengaturan `random state`. Penerapan AdaBoost bertujuan untuk mengevaluasi efektivitas mekanisme pembelajaran adaptif dalam meningkatkan performa klasifikasi pada data dengan tingkat kompleksitas yang beragam.

3.7.4 XGBoost

XGBoost adalah versi baru dari algoritma *Gradient Boosting* yang dimaksudkan untuk meningkatkan produktivitas komputasi dan performa model melalui mekanisme optimasi dan *regularization*. Pada penelitian ini, XGBoost diuji dengan memvariasikan jumlah estimator (`n_estimators`). Parameter lain yang digunakan mencakup kedalaman maksimum pohon (`max_depth`), `learning rate`, rasio subsampling, rasio pemilihan fitur (`feature sampling`), fungsi objektif untuk klasifikasi multikelas (`objective`), jumlah kelas target (`num_class`), metrik evaluasi (`evaluation metric`), pengaturan `random state`, serta pemanfaatan `multi-threading`. Algoritma ini digunakan untuk mengevaluasi kinerja pendekatan *boosting* tingkat lanjut pada permasalahan klasifikasi multikelas.

3.7.5 LightGBM

LightGBM merupakan algoritma *gradient boosting* yang mengadopsi pendekatan berbasis *histogram* serta strategi pertumbuhan pohon secara *leaf-wise* untuk meningkatkan efisiensi proses pelatihan. Pada penelitian ini, LightGBM diuji dengan memvariasikan jumlah estimator (`n_estimators`). Parameter lain yang digunakan meliputi `learning rate`, kedalaman maksimum pohon (`max_depth`), subsampling, pemilihan fitur (`feature sampling`), fungsi objektif multikelas (`objective`), jumlah kelas target (`num_class`), pengaturan `random state`, serta pemanfaatan `parallel processing`. Algoritma ini digunakan untuk menilai efisiensi waktu pelatihan dan performa model pada data yang telah melalui proses penyeimbangan dan augmentasi.

3.7.6 CatBoost

CatBoost merupakan algoritma *gradient boosting* yang dirancang untuk menangani fitur kategorikal secara lebih efektif serta mengurangi risiko *overfitting*.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dalam penelitian ini, CatBoost diuji dengan memvariasikan jumlah iterasi pelatihan (iterations) sebagai parameter utama. Parameter lain yang digunakan mencakup kedalaman pohon (depth), learning rate, fungsi kerugian untuk klasifikasi multikelas (loss function), metrik evaluasi (evaluation metric), pengaturan random seed, serta penonaktifan tampilan proses pelatihan (verbose). Algoritma ini digunakan untuk mengevaluasi kestabilan model serta konsistensi performa pada berbagai skenario pengolahan data.

Untuk memperoleh model dengan performa optimal, setiap algoritma diuji dengan beberapa variasi parameter utama yang dianggap paling berpengaruh terhadap kinerja model. Seluruh eksperimen dilakukan secara manual menggunakan pendekatan *hold-out*. Rincian parameter yang digunakan pada masing-masing algoritma disajikan pada Tabel 3.3.

Tabel 3.3. Konfigurasi Parameter Algoritma

No	Algoritma	Parameter	Nilai
1	<i>Random Forest</i>	n_estimators	10, 50, 100, 200, 300, 500
		max_depth	10
		min_samples_leaf	2
		criterion	gini
		class_weight	balanced
		random_state	42
2	<i>Gradient Boosting</i>	n_estimators	10, 50, 100, 200, 300, 500
		learning_rate	0.05
		max_depth	3
		subsample	1.0
		loss	log_loss
		random_state	42
3	AdaBoost	n_estimators	10, 50, 100, 200, 300, 500
		learning_rate	0.05
		base_estimator	DecisionTreeClassifier
		algorithm	SAMME
4	XGBoost	random_state	42
		n_estimators	10, 50, 100, 200, 300, 500
		learning_rate	0.05
		max_depth	6
		subsample	0.8

Tabel 3.3. Konfigurasi Parameter Algoritma

No	Algoritma	Parameter	Nilai
4		colsample_bytree	0.8
		objective	multi:softprob
		eval_metric	mlogloss
		num_class	5
		random_state	42
5	LightGBM	n_estimators	10, 50, 100, 200, 300, 500
		learning_rate	0.05
		max_depth	-1
		num_leaves	31
		subsample	0.8
		colsample_bytree	0.8
		objective	multiclass
		num_class	5
		random_state	42
6	CatBoost	iterations	10, 50, 100, 200, 300, 500
		learning_rate	0.05
		depth	6
		loss_function	MultiClass
		eval_metric	TotalF1
		random_seed	42
		verbose	False

3.3 Evaluasi

Tahap evaluasi merupakan proses lanjutan yang bertujuan untuk menilai performa model klasifikasi yang telah dikembangkan pada tahap pemodelan. Evaluasi ini dilakukan untuk mengetahui kemampuan model dalam mengklasifikasikan tingkat depresi, kecemasan, dan stres secara tepat berdasarkan data uji. Penilaian kinerja model dilakukan dengan menggunakan sejumlah metrik evaluasi yang umum diterapkan pada permasalahan klasifikasi multi-kelas, sebagaimana dijelaskan berikut ini.

1. *Accuracy* digunakan untuk menunjukkan tingkat ketepatan prediksi model secara keseluruhan, yaitu perbandingan antara jumlah prediksi yang benar dengan total data uji. Metrik ini memberikan gambaran umum mengenai performa model, namun kurang mampu merefleksikan kinerja secara

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

menyeluruh ketika distribusi kelas tidak seimbang.

2. *Precision* mengukur ketepatan model dalam memprediksi suatu kelas tertentu, yang dihitung berdasarkan perbandingan antara jumlah prediksi benar pada kelas tersebut dengan total prediksi yang dihasilkan untuk kelas yang sama. Metrik ini berperan penting dalam mengevaluasi sejauh mana model mampu mengurangi kesalahan prediksi positif.
3. *Recall* digunakan untuk menilai kemampuan model dalam mengidentifikasi seluruh data yang seharusnya termasuk ke dalam suatu kelas. Nilai *recall* mencerminkan tingkat keberhasilan model dalam menangkap seluruh anggota kelas target tanpa mengabaikan data yang relevan.
4. *F1-score* merupakan metrik gabungan yang mengintegrasikan *precision* dan *recall* melalui perhitungan rata-rata harmonis. Nilai *F1-score* digunakan sebagai salah satu indikator utama dalam menentukan model terbaik, karena metrik ini mencerminkan keseimbangan kemampuan model dalam mengidentifikasi kelas secara tepat dan lengkap.
5. *Confusion Matrix* digunakan untuk menyajikan hasil prediksi model secara lebih detail pada setiap kelas. Melalui matriks ini, dapat diketahui jumlah prediksi yang benar maupun salah pada masing-masing tingkat keparahan, sehingga membantu dalam mengidentifikasi pola kesalahan klasifikasi yang dihasilkan oleh model.
6. ROC-AUC digunakan untuk menilai kemampuan model dalam membedakan antar kelas berdasarkan nilai skor prediksi yang dihasilkan. Pada klasifikasi multi-kelas, perhitungan ROC-AUC dilakukan menggunakan pendekatan *One-vs-Rest*. Nilai AUC yang mendekati angka 1 mengindikasikan bahwa model memiliki kemampuan diskriminasi kelas yang semakin baik.

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan secara menyeluruh, diperoleh beberapa kesimpulan utama sebagai berikut:

1. Penelitian ini membuktikan bahwa pendekatan *ensemble learning* mampu mengklasifikasikan tingkat depresi, kecemasan, dan stres mahasiswa berdasarkan data kuesioner DASS-21 secara komputasional dan efektif. Seluruh model yang dikembangkan berhasil mengenali pola klasifikasi multikelas pada data kesehatan mental mahasiswa, meskipun dengan tingkat kinerja yang bervariasi. Temuan ini menegaskan bahwa pendekatan *ensemble learning* memiliki potensi yang kuat sebagai alat bantu awal dalam pemetaan kondisi kesehatan mental mahasiswa berbasis instrumen psikologis terstandar.
2. Hasil perbandingan kinerja enam algoritma *ensemble learning* menunjukkan bahwa *Random Forest* merupakan algoritma dengan performa paling optimal dan konsisten pada seluruh target klasifikasi. Pada target depresi, *Random Forest* dengan skema augmentasi data mencapai akurasi sebesar 99.71%, *precision* 99.26%, *recall* 99.09%, dan *F1-score* 99.16%. Pada target kecemasan, model yang sama menghasilkan akurasi sebesar 99.42%, dengan *precision* 99.59%, *recall* 98.79%, dan *F1-score* 99.17%. Sementara itu, pada target stres, *Random Forest* dengan skema augmentasi data mencapai performa klasifikasi sempurna dengan nilai akurasi, *precision*, *recall*, dan *F1-score* masing-masing sebesar 100%. Selain *Random Forest*, algoritma CatBoost dan *Gradient Boosting* juga menunjukkan performa yang kompetitif, khususnya pada skema augmentasi data dengan tingkat akurasi di atas 85%. LightGBM dan XGBoost memberikan kinerja menengah yang relatif stabil, sedangkan AdaBoost secara konsisten menunjukkan kinerja terendah pada seluruh target klasifikasi, dengan tingkat akurasi berada pada kisaran 50–65%.
3. Evaluasi terhadap perbedaan skema pengolahan data menunjukkan bahwa skema augmentasi data memberikan peningkatan kinerja yang paling signifikan dibandingkan data asli maupun data dengan penyeimbangan kelas menggunakan SMOTE. Augmentasi data terbukti mampu meningkatkan nilai akurasi, serta memperbaiki keseimbangan *precision*, *recall*, dan *F1-score*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

pada hampir seluruh algoritma dan target klasifikasi. Sebaliknya, penerapan SMOTE mampu meningkatkan kinerja pada beberapa algoritma tertentu, namun tidak selalu menghasilkan performa yang optimal dibandingkan skema augmentasi data, terutama pada algoritma dengan karakteristik *weak learner* seperti AdaBoost.

5.2 Saran

Penelitian di masa mendatang dianjurkan untuk menggunakan *dataset* dengan jumlah yang lebih banyak dan karakteristik yang lebih beragam agar model yang dikembangkan mampu merepresentasikan kondisi populasi secara lebih akurat. Penambahan data diharapkan dapat meningkatkan stabilitas dan keandalan model serta mengurangi potensi bias akibat keterbatasan jumlah sampel. Selain itu, penerapan metode evaluasi seperti *K-Fold Cross Validation* dapat dipertimbangkan untuk meningkatkan hasil pengujian yang lebih objektif dan konsisten. Eksplorasi teknik pengolahan data, khususnya metode augmentasi dan penyeimbangan kelas, dapat dikembangkan lebih lanjut guna meningkatkan performa model pada kelas dengan distribusi data yang tidak seimbang. Penelitian lanjutan juga disarankan untuk melakukan eksplorasi *hyperparameter tuning* secara lebih sistematis, misalnya melalui *Grid Search* atau *Random Search*, guna memperoleh kombinasi parameter yang optimal pada setiap algoritma. Di sisi lain, analisis interpretabilitas model, seperti *feature importance* atau pendekatan *Explainable Artificial Intelligence* (XAI), perlu dikaji agar hasil klasifikasi dapat lebih mudah dipahami dan dimanfaatkan oleh praktisi di bidang kesehatan mental. Temuan penelitian ini diharapkan dapat berfungsi sebagai dasar untuk pembangunan sistem pendukung keputusan deteksi dini kondisi kesehatan mental mahasiswa berbasis kuesioner DASS-21.

DAFTAR PUSTAKA

- Adler, D. (2006). Job Performance Deficits Due to Depression. *American Journal of Psychiatry*, 163(9), 1569. Retrieved from <http://psychiatryonline.org/article.aspx?doi=10.1176/appi.ajp.163.9.1569> doi: 10.1176/appi.ajp.163.9.1569
- Ahmed, N. S., dan Hikmat Sadiq, M. (2018). Clarify of the Random Forest Algorithm in an Educational Field. Dalam *2018 international conference on advanced science and engineering (icoase)* (hal. 179–184). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/8548804/> doi: 10.1109/ICOASE.2018.8548804
- Ahmed, O., dan Brifcani, A. (2019). Gene Expression Classification Based on Deep Learning. Dalam *2019 4th scientific international conference najaf (sicc)* (hal. 145–149). Al-Najaf, Iraq: IEEE. Retrieved from <https://ieeexplore.ieee.org/document/9019357/> doi: 10.1109/SICN47020.2019.9019357
- Aleem, S., ul Huda, N., Amin, R., Khalid, S., Alshamrani, S. S., dan Alshehri, A. (2022). Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions. *Electronics*, 11(7), 1111. Retrieved from <https://www.mdpi.com/2079-9292/11/7/1111> doi: 10.3390/electronics11071111
- Alelyani, S. (2021). Stable bagging feature selection on medical data. *Journal of Big Data*, 8(1), 11. Retrieved from <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00385-8> doi: 10.1186/s40537-020-00385-8
- Ali, A. M., Alkamees, A. A., Hori, H., Kim, Y., dan Kunugi, H. (2021). The Depression Anxiety Stress Scale 21: Development and Validation of the Depression Anxiety Stress Scale 8-Item in Psychiatric Patients and the General Public for Easier Mental Health Measurement in a Post COVID-19 World. *International Journal of Environmental Research and Public Health*, 18(19), 10142. Retrieved from <https://www.mdpi.com/1660-4601/18/19/10142> doi: 10.3390/ijerph181910142
- Ali, A. M., dan Green, J. (2019). Factor structure of the depression anxiety stress Scale-21 (DASS-21): Unidimensionality of the Arabic version among Egyptian drug users. *Substance Abuse Treatment, Prevention, and Policy*, 14(1), 40. Retrieved from <https://doi.org/10.1186/s13011-019-0226>



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

-1<https://substanceabusepolicy.biomedcentral.com/articles/10.1186/s13011-019-0226-1> doi: 10.1186/s13011-019-0226-1

Amin, M. F. (2023). Confusion matrix in three-class classification problems: A step-by-step tutorial. *Journal of Engineering Research*, 7(1). Retrieved from <https://digitalcommons.aaru.edu.jo/erjeng/vol7/iss1/26>

Anand, R. V., Md, A. Q., Urooj, S., Mohan, S., Alawad, M. A., dan C., A. (2023). Enhancing Diagnostic Decision-Making: Ensemble Learning Techniques for Reliable Stress Level Classification. *Diagnostics*, 13(22), 3455. Retrieved from <https://www.mdpi.com/2075-4418/13/22/3455> doi: 10.3390/diagnostics13223455

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., dan Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment*, 10(2), 176–181. Retrieved from <https://doi.apa.org/doi/10.1037/1040-3590.10.2.176> doi: 10.1037/1040-3590.10.2.176

Anuradha, dan Gupta, G. (2014). A self explanatory review of decision tree classifiers. Dalam *International conference on recent advances and innovations in engineering (icraie-2014)* (hal. 1–7). Jaipur, India: IEEE. Retrieved from <https://ieeexplore.ieee.org/document/6909245> doi: 10.1109/ICRAIE.2014.6909245

Arora, P., dan Dahiya, S. (2025). Enhancing Mental Health Diagnosis with DA-TabSVM: A Multi-Class Hybrid Approach for Detecting Depression and Anxiety. *Procedia Computer Science*, 258, 1348–1364. Retrieved from <https://doi.org/10.1016/j.procs.2025.04.368> doi: 10.1016/j.procs.2025.04.368

Arif, S., Mudassar, A., Shahzad, T. Z., Raouf, M., dan Pervaiz, T. (2020). Frequency of depression, anxiety and stress among university students. *Pakistan Journal of Medical Sciences*, 36(5), 971–976. doi: 10.12669/pjms.36.5.1873

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5®)*. Arlington, VA: American Psychiatric Publishing. Retrieved from <https://www.appi.org/products/dsm-5>

Bhatnagar, S., Agarwal, J., dan Sharma, O. R. (2023). Detection and classification of anxiety in university students through the application of machine learning. *Procedia Computer Science*, 218(2022), 1542–1550. Retrieved from <https://doi.org/10.1016/j.procs.2023.01.132><https://doi.org/10.1016/j.procs.2023.01.132>



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

linkinghub.elsevier.com/retrieve/pii/S1877050923001321 doi:
10.1016/j.procs.2023.01.132

- Bo Q., Goodman, K. E., Kaminsky, J., dan Lessler, J. (2019). What is machine learning? a primer for the epidemiologist. *American Journal of Epidemiology*, 188(12), 2222-2239. Retrieved from <https://doi.org/10.1093/aje/kwz189> doi: 10.1093/aje/kwz189
- Birjandi, S. M., dan Khasteh, S. H. (2021). A survey on data mining techniques used in medicine. *Journal of Diabetes & Metabolic Disorders*, 20(2), 2055–2071. Retrieved from <https://doi.org/10.1007/s40200-021-00884-2> doi: 10.1007/s40200-021-00884-2
- Black, J. E., Kueper, J. K., dan Williamson, T. S. (2023). An introduction to machine learning for classification and prediction. *Family practice*, 40(1), 200–204.
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145–1159. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0031320396001422> doi: 10.1016/S0031-3203(96)00142-2
- Breiman, L. (1998). Arcing Classifiers. *The Annals of Statistics*, 26(3), 801–824. Retrieved from <https://www.jstor.org/stable/120055>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. Retrieved from <https://link.springer.com/10.1023/A:1010933404324> doi: 10.1023/A:1010933404324
- Brown, G. (2011). Ensemble Learning. Dalam *Encyclopedia of machine learning* (hal. 312–320). Boston, MA: Springer US. Retrieved from https://doi.org/10.1007/978-0-387-30164-8_252 doi: 10.1007/978-0-387-30164-8_252
- Center for Reproductive Health, University of Queensland, dan Johns Bloomberg Hopkins School of Health Public. (2022). *Indonesia – National Adolescent Mental Health Survey (I-NAMHS): Laporan Penelitian*. Pusat Kesehatan Reproduksi. Retrieved from <https://qcmhr.org/outputs/reports/12-i-namhs-report-bahasa-indonesia>
- Chand, S. P., dan Arif, H. (2025). *Depression*. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/0http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6396718>



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Chen, J., Tam, D., Raffel, C., Bansal, M., dan Yang, D. (2023). An Empirical Survey of Data Augmentation for Limited Data Learning in NLP. *Transactions of the Association for Computational Linguistics*, 11, 191–211. Retrieved from https://doi.org/10.1162/tacl_a-00542 doi: 10.1162/tacl_a-00542
- Chen, T., dan Guestrin, C. (2016). XGBoost. Dalam *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (hal. 785–794). New York, NY, USA: ACM. Retrieved from <https://dl.acm.org/doi/10.1145/2939672.2939785> doi: 10.1145/2939672.2939785
- Chicco, D., dan Jurman, G. (2023). The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Mining*, 16(1), 4. Retrieved from <https://doi.org/10.1186/s13040-023-00322-4https://biodatamining.biomedcentral.com/articles/10.1186/s13040-023-00322-4> doi: 10.1186/s13040-023-00322-4
- Chowdhury, A. H., Rad, D., dan Rahman, M. S. (2024). Predicting anxiety, depression, and insomnia among Bangladeshi university students using tree-based machine learning models. *Health Science Reports*, 7(4). Retrieved from <https://onlinelibrary.wiley.com/doi/10.1002/hsr2.2037> doi: 10.1002/hsr2.2037
- Cohen, S., Gianaros, P. J., dan Manuck, S. B. (2016). A Stage Model of Stress and Disease. *Perspectives on Psychological Science*, 11(4), 456–463. Retrieved from <https://journals.sagepub.com/doi/10.1177/1745691616646305> doi: 10.1177/1745691616646305
- Daza, A., Arroyo-Paz, Bobadilla, J., Apaza, O., dan Pinto, J. (2023). Stacking ensemble learning model for predict anxiety level in university students using balancing methods. *Informatics in Medicine Unlocked*, 42(July), 101340. Retrieved from <https://doi.org/10.1016/j.imu.2023.101340https://linkinghub.elsevier.com/retrieve/pii/S2352914823001867> doi: 10.1016/j.imu.2023.101340
- Daza Vergaray, A., Miranda, J. C. H., Cornelio, J. B., López Carranza, A. R., dan Ponce Sánchez, C. F. (2023). Predicting the depression in university students using stacking ensemble techniques over oversampling method. *Informatics in Medicine Unlocked*, 41(June), 101295. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S2352914823001417> doi: 10.1016/j.imu.2023.101295



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Deberneh, H. M., dan Kim, I. (2021). Prediction of Type 2 Diabetes Based on Machine Learning Algorithm. *International Journal of Environmental Research and Public Health*, 18(6), 3317. Retrieved from <https://www.mdpi.com/1660-4601/18/6/3317> doi: 10.3390/ijerph18063317
- Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., dan Al., E. (2004). Prevalence, Severity, and Unmet Need for Treatment of Mental Disorders in the World Health Organization World Mental Health Surveys. *JAMA*, 291(21), 2581. Retrieved from <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.291.21.2581> doi: 10.1001/jama.291.21.2581
- Dhani, A., Lestari, D., Ningrum, M. P., Fakhri, M. A., dan Gandini, G. L. (2025). Applying A Supervised Model for Diabetes Type 2 Risk Level Classification. *Public Research Journal of Engineering, Data Technology and Computer Science*, 2(2), 60–67. doi: 10.57152/predatecs.v2i2.1105
- Doroudi, S. (2020). The Bias-Variance Tradeoff: How Data Science Can Inform Educational Debates. *AERA Open*, 6(4). Retrieved from <https://journals.sagepub.com/doi/10.1177/2332858420977208> doi: 10.1177/2332858420977208
- Duangchaemkarn, K., Khammarew, P., dan Aramvith, S. (2024). Machine Learning-Based Classification of Mental Health State Using the DASS-21 Profile. Dalam *2024 16th biomedical engineering international conference (bmeicon)* (hal. 1–5). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/10896316/> doi: 10.1109/BMEiCON64021.2024.10896316
- El Morr, C., Jammal, M., Bou-Hamad, I., Hijazi, S., Ayna, D., Romani, M., dan Hoteit, R. (2024). Predictive Machine Learning Models for Assessing Lebanese University Students' Depression, Anxiety, and Stress During COVID-19. *Journal of Primary Care & Community Health*, 15. Retrieved from <https://journals.sagepub.com/doi/10.1177/21501319241235588> doi: 10.1177/21501319241235588
- Ericson, P. M., dan Gardner, J. W. (1992). Two longitudinal studies of communication apprehension and its effects on college students' success. *Communication Quarterly*, 40(2), 127–137. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/01463379209369828> doi: 10.1080/01463379209369828
- Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., dan Obaido, G. (2022). A Neural Network Ensemble With Feature Engineering for Improved Credit



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Card Fraud Detection. *IEEE Access*, 10, 16400–16407. Retrieved from <https://ieeexplore.ieee.org/document/9698195/> doi: 10.1109/ACCESS.2022.3148298

Farooq, S. A., Konda, O., Kunwar, A., dan Rajeev, N. (2023). Anxiety Prediction and Analysis- A Machine Learning Based Approach. Dalam *2023 4th international conference for emerging technology (incet)* (hal. 1–7). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/10170115/> doi: 10.1109/INCET57972.2023.10170115

Fauzi, M. F., Anuar, T. S., Teh, L. K., Lim, W. F., James, R. J., Ahmad, R., ... Salleh, M. Z. (2021). Stress, anxiety and depression among a cohort of health sciences undergraduate students: The prevalence and risk factors. *International Journal of Environmental Research and Public Health*, 18(6), 1–14. doi: 10.3390/ijerph18063269

Fawagreh, K., Gaber, M. M., dan Elyan, E. (2014). Random forests: from early developments to recent advancements. *Systems Science & Control Engineering*, 2(1), 602–609. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/21642583.2014.956265> doi: 10.1080/21642583.2014.956265

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X> doi: 10.1016/j.patrec.2005.10.010

Fayyad, U., Piatetsky-Shapiro, G., dan Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Proceedings - 2nd International Conference on Knowledge Discovery and Data Mining, KDD 1996*, 82–88. doi: 10.5555/3001460.3001477

Fink, G. (2017). Stress: Concepts, definition and history. Dalam *Reference module in neuroscience and biobehavioral psychology*. Elsevier. Retrieved from <https://www.sciencedirect.com/science/article/pii/B9780128093245022082> doi: <https://doi.org/10.1016/B978-0-12-809324-5.02208-2>

Freund, Y., dan Schapire, R. E. (1999). A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5), 771–780.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5). Retrieved from <https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full> doi:



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

10.1214/aos/1013203451

- Gilmour, H., dan Patten, S. B. (2007). Depression and work impairment. *Health reports*, 18(1), 9–22. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17441440>
- Gorunescu, F. (2011). *Data Mining* (Vol. 12). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from <http://link.springer.com/10.1007/978-3-642-19721-5> doi: 10.1007/978-3-642-19721-5
- Grabowska, A., Zabielski, J., dan Senderecka, M. (2024). Machine learning reveals differential effects of depression and anxiety on reward and punishment processing. *Scientific Reports*, 14(1), 8422. Retrieved from <https://doi.org/10.1038/s41598-024-58031-9> <https://www.nature.com/articles/s41598-024-58031-9> doi: 10.1038/s41598-024-58031-9
- Ha, J., Kambe, M., dan Pe, J. (2012). *Data Mining*. Elsevier. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/C20090618195> doi: 10.1016/C2009-0-61819-5
- Hasan, M. E., Arif, M., Rakibul Hasan, S. M., Muwanguzi, M., Abaatyo, J., Kag-gwa, M. M., ... Mamun, M. A. (2025). Prevalence, associated factors, and machine learning-based prediction of depression, anxiety, and stress among university students: a cross-sectional study from Bangladesh. *Journal of Health, Population and Nutrition*, 44(1), 361. Retrieved from <https://jhpn.biomedcentral.com/articles/10.1186/s41043-025-01095-8> doi: 10.1186/s41043-025-01095-8
- Hasan, M. J., Das, A., Matubber, J., Shifat, S. H., dan Morol, M. K. (2024). Enhanced Classification of Anxiety, Depression, and Stress Levels: A Comparative Analysis of DASS21 Questionnaire Data Augmentation and Classification Algorithms. Dalam *Proceedings of the 3rd international conference on computing advancements* (hal. 435–442). New York, NY, USA: ACM. Retrieved from <https://dl.acm.org/doi/10.1145/3723178.3723236> doi: 10.1145/3723178.3723236
- Heck, J. (2008). *Anxiety Has Healing*. Rio de Janeiro: Imprimatur.
- Henry, J. D., dan Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British journal of clinical psychology*, 44(2), 227–239. doi: 10.1348/014466505X29657
- Herman, Darwis, H., Nurfauziyah, Puspitasari, R., Widyawati, D., dan Faradibah,



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- A. (2025). Comparative Analysis of Anxiety Disorder Classification Using Algorithm Naïve Bayes, Decision Tree and K-NN. Dalam *2025 19th international conference on ubiquitous information management and communication (imcom)* (hal. 1–6). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/10857485/> doi: 10.1109/IMCOM64595.2025.10857485
- Huang, Y. (2025). Fs-smote: An improved smote method based on feature space scoring mechanism for solving class-imbalanced problems. *IEEE Access*, 13, 148074-148082. doi: 10.1109/ACCESS.2025.3597794
- Hyssenbegasi, A., Hass, S. L., dan Rowland, C. R. (2005). The impact of depression on the academic productivity of university students. *The journal of mental health policy and economics*, 8(3), 145–151. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16278502>
- Islam, M. M., Hassan, S., Akter, S., Jibon, F. A., dan Sahidullah, M. (2024). A comprehensive review of predictive analytics models for mental illness using machine learning algorithms. *Healthcare Analytics*, 6, 100350. Retrieved from <https://www.sciencedirect.com/science/article/pii/S2772442524000522> doi: <https://doi.org/10.1016/j.health.2024.100350>
- Ivanova, J. I., Birnbaum, H. G., Kidolezi, Y., Subramanian, G., Khan, S. A., dan Stensland, M. D. (2010). Direct and indirect costs of employees with treatment-resistant and non-treatment-resistant major depressive disorder. *Current Medical Research and Opinion*, 26(10), 2475–2484. Retrieved from <http://www.tandfonline.com/doi/full/10.1185/03007995.2010.517716> doi: 10.1185/03007995.2010.517716
- Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms* (2nd ed.). Wiley-IEEE Press.
- Karthikeyan, P., dan Vasuki, S. (2016). Hybrid approach of efficient decision-based algorithm and fuzzy logic for the removal of high density salt and pepper noise in images. *Journal of Circuits, Systems and Computers*, 25(10), 1650130. Retrieved from <https://doi.org/10.1142/S0218126616501309> doi: 10.1142/S0218126616501309
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 3147–3155. doi: 10.5555/3294996.3295074
- Kohn, M., dan Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY: Springer New York. Retrieved from <http://link.springer.com/10>



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

.1007/978-1-4614-6849-3 doi: 10.1007/978-1-4614-6849-3

- Kullarni, V. Y., dan Sinha, P. K. (2013). Random Forest Classifier: A Survey and Future Research Directions. *International Journal of Advanced Computing*, 36(1), 1144–1156.
- Kumar, A., Saxena, J. M., Joshi, J. P., Aljapurkar, A., Paranjpye, R., Talla, T., ... Gupta, D. (2024). Early Detection of College Student Mental Health Issues Using Machine Learning. Dalam *2024 international conference on integrated intelligence and communication systems (iciics)* (hal. 1–5). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/10859910/> doi: 10.1109/ICIICS63763.2024.10859910
- Li, Q.-F., dan Song, Z.-M. (2022). High-performance concrete strength prediction based on ensemble learning. *Construction and Building Materials*, 324, 126694. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0950061822003841> doi: 10.1016/j.conbuildmat.2022.126694
- Li, Y., dan Chen, W. (2020). A Comparative Performance Assessment of Ensemble Learning for Credit Scoring. *Mathematics*, 8(10), 1756. Retrieved from <https://www.mdpi.com/2227-7390/8/10/1756> doi: 10.3390/math8101756
- Liang, W., Luo, S., Zhao, G., dan Wu, H. (2020). Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms. *Mathematics*, 8(5), 765. Retrieved from <https://www.mdpi.com/2227-7390/8/5/765> doi: 10.3390/math8050765
- Liu, J., Ning, W., Zhang, N., Zhu, B., dan Mao, Y. (2024). Estimation of the Global Disease Burden of Depression and Anxiety between 1990 and 2044: An Analysis of the Global Burden of Disease Study 2019. *Healthcare*, 12(17), 1721. Retrieved from <https://www.mdpi.com/2227-9032/12/17/1721> doi: 10.3390/healthcare12171721
- Lovibond, P., dan Lovibond, S. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33(3), 335–343. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/7726811/> <https://linkinghub.elsevier.com/retrieve/pii/000579679400075U> doi: 10.1016/0005-7967(94)00075-U
- Li, C., Zhang, S., Xue, D., Xiao, F., dan Liu, C. (2022). Improved estimation of coalbed methane content using the revised estimate of depth and CatBoost



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

algorithm: A case study from southern Sichuan Basin, China. *Computers & Geosciences*, 158, 104973. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0098300421002570> doi: <https://doi.org/10.1016/j.cageo.2021.104973>

Malhotra, G., Leslie, D. S., Ludwig, C. J., dan Bogacz, R. (2017). Overcoming indecision by changing the decision boundary. *Journal of Experimental Psychology: General*, 146(6), 776–805. doi: 10.1037/xge0000286

Mateen, B. A., Liley, J., Denniston, A. K., Holmes, C. C., dan Vollmer, S. J. (2020). Improving the quality of machine learning in health applications and clinical research. *Nature Machine Intelligence*, 2(10), 554–556. Retrieved from <https://doi.org/10.1038/s42256-020-00239-1> <https://www.nature.com/articles/s42256-020-00239-1> doi: 10.1038/s42256-020-00239-1

Mienye, I. D., dan Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, 10, 99129–99149. Retrieved from <https://ieeexplore.ieee.org/document/9893798/> doi: 10.1109/ACCESS.2022.3207287

Mirza, I., dan Jenkins, R. (2004). Risk factors, prevalence, and treatment of anxiety and depressive disorders in Pakistan: systematic review. *BMJ*, 328(7443), 794. Retrieved from <https://www.bmj.com/lookup/doi/10.1136/bmj.328.7443.794> doi: 10.1136/bmj.328.7443.794

Mishra, S., Shaw, K., Mishra, D., Patil, S., Kotecha, K., Kumar, S., dan Bajaj, S. (2022). Improving the Accuracy of Ensemble Machine Learning Classification Models Using a Novel Bit-Fusion Algorithm for Healthcare AI Systems. *Frontiers in Public Health*, 10. Retrieved from <https://www.frontiersin.org/articles/10.3389/fpubh.2022.858282/full> doi: 10.3389/fpubh.2022.858282

Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.

Mohammadi, S., Narimani, Z., Ashouri, M., Firouzi, R., dan Karimi-Jafari, M. H. (2022). Ensemble learning from ensemble docking: revisiting the optimum ensemble size problem. *Scientific Reports*, 12(1), 410. Retrieved from <https://www.nature.com/articles/s41598-021-04448-5> doi: 10.1038/s41598-021-04448-5

Müller, A. C., dan Guido, S. (2016). *Introduction to Machine Learning with Python*. O'Reilly Media, Inc. Retrieved from <https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/>



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
- Natekin, A., dan Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7. Retrieved from <http://journal.frontiersin.org/article/10.3389/fnbot.2013.00021/abstract> doi: 10.3389/fnbot.2013.00021
- Neshatian, K., dan Varn, L. (2019). Feature Bundles and their Effect on the Performance of Tree-based Evolutionary Classification and Feature Selection Algorithms. Dalam *2019 IEEE Congress on Evolutionary Computation (CEC)* (hal. 1612–1619). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/8789951/> doi: 10.1109/CEC.2019.8789951
- Nguyen, K. A., Chen, W., Lin, B.-S., dan Seeboonruang, U. (2021). Comparison of Ensemble Machine Learning Methods for Soil Erosion Pin Measurements. *ISPRS International Journal of Geo-Information*, 10(1), 42. Retrieved from <https://www.mdpi.com/2220-9964/10/1/42> doi: 10.3390/ijgi10010042
- Nikam, S. S. (2015). A comparative study of classification techniques in data mining algorithms. *Oriental Journal of Computer Science & Technology*, 8(1), 13–19. Retrieved from <http://www.computerscijournal.org/?p=1592>
- Niu, D., Diao, L., Zang, Z., Che, H., Zhang, T., dan Chen, X. (2021). A Machine-Learning Approach Combining Wavelet Packet Denoising with Catboost for Weather Forecasting. *Atmosphere*, 12(12). Retrieved from <https://www.mdpi.com/2073-4433/12/12/1618> doi: 10.3390/atmos12121618
- Nti, I. K., Adekoya, A. F., dan Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, 7(1), 20. Retrieved from <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00299-5> doi: 10.1186/s40537-020-00299-5
- Nurhidayarnis, S., Mustakim, Novita, R., Afdal, M., dan Wahyuni, S. (2024). Backpropagation Neural Network-Based Classification in Prioritizing Stress Attributes in Students Using the Analytic Hierarchy Process. Dalam *2024 IEEE International Conference on Control & Automation, Electronics, Robotics, Internet of Things, and Artificial Intelligence (CERIA)* (hal. 1–6). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/10915004/> doi: 10.1109/CERIA64726.2024.10915004
- Omoniyi, T. M., Abel, B., Omoebamiye, O., Onimisi, Z. M., Matos, J. C., Tinoco, J., dan Minh, T. Q. (2025). The Effect of Data Augmentation on Perfor-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
- mance of Custom and Pre-Trained CNN Models for Crack Detection. *Applied Sciences*, 15(22). Retrieved from <https://www.mdpi.com/2076-3417/15/22/12321> doi: 10.3390/app152212321
- Oza, N. C., dan Tumer, K. (2008). Classifier ensembles: Select real-world applications. *Information Fusion*, 9(1), 4–20. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S1566253507000620> doi: 10.1016/j.inffus.2007.07.002
- Patel, V. (2001). Cultural factors and international epidemiology. *British Medical Bulletin*, 57(1), 33–45. Retrieved from <https://academic.oup.com/bmb/article/57/1/33/301595> doi: 10.1093/bmb/57.1.33
- Patil, D. V., dan Bichkar, R. S. (2006). A Hybrid Evolutionary Approach To Construct Optimal Decision Trees With Large Data Sets. Dalam *2006 ieee international conference on industrial technology* (hal. 429–433). Chengdu, China: IEEE. Retrieved from <http://ieeexplore.ieee.org/document/4237572/> doi: 10.1109/ICIT.2006.372250
- Patil, S., dan Kulkarni, U. (2019). Accuracy Prediction for Distributed Decision Tree using Machine Learning approach. Dalam *2019 3rd international conference on trends in electronics and informatics (icoei)* (hal. 1365–1371). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/8862580/> doi: 10.1109/ICOEI.2019.8862580
- Paykel, E. S. (2008). Basic concepts of depression. *Dialogues in Clinical Neuroscience*, 10(3), 279–289. Retrieved from <https://www.tandfonline.com/doi/full/10.31887/DCNS.2008.10.3/espaykel> doi: 10.31887/DCNS.2008.10.3/espaykel
- Priya, A., Garg, S., dan Tigga, N. P. (2020). Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. *Procedia Computer Science*, 167(2019), 1258–1267. Retrieved from <https://doi.org/10.1016/j.procs.2020.03.442><https://linkinghub.elsevier.com/retrieve/pii/S1877050920309091> doi: 10.1016/j.procs.2020.03.442
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., dan Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Dalam *Proceedings of the 32nd international conference on neural information processing systems* (hal. 6639–6649). Red Hook, NY, USA: Curran Associates Inc.
- Qasrawi, R., Polo, S. P. V., Al-Halawa, D. A., Hallaq, S., dan Abdeen, Z. (2022). Assessment and Prediction of Depression and Anxiety Risk Factors in



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Schoolchildren: Machine Learning Techniques Performance Analysis. *JMIR Formative Research*, 6(8), 1–15. doi: 10.2196/32736

Rahzavi, M., Ziyadidegan, S., Jahromi, R., Kazeminasab, S., Baharlouei, E., Janfaza, V., ... Sasangohar, F. (2023). Machine Learning, Deep Learning and Data Preprocessing Techniques for Detection, Prediction, and Monitoring of Stress and Stress-related Mental Disorders: A Scoping Review. *JMIR Mental Health*, 11, 1–28. doi: 10.2196/53714

Ribeiro, M. H. D. M., dan dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing*, 86, 105837. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S1568494619306180> doi: 10.1016/j.asoc.2019.105837

Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., ... Schönlieb, C.-B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*, 3(3), 199–217. Retrieved from <https://doi.org/10.1038/s42256-021-00307-0> <https://www.nature.com/articles/s42256-021-00307-0> doi: 10.1038/s42256-021-00307-0

Rois, R., Ray, M., Rahman, A., dan Roy, S. K. (2021). Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms. *Journal of Health, Population and Nutrition*, 40(1), 50. Retrieved from <https://doi.org/10.1186/s41043-021-00276-5> <https://jhpn.biomedcentral.com/articles/10.1186/s41043-021-00276-5> doi: 10.1186/s41043-021-00276-5

Saberi-Karimian, M., Khorasanchi, Z., Ghazizadeh, H., Tayefi, M., Saffar, S., Ferns, G. A., dan Ghayour-Mobarhan, M. (2021). Potential value and impact of data mining and machine learning in clinical diagnostics. *Critical Reviews in Clinical Laboratory Sciences*, 58(4), 275–296. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/10408363.2020.1857681> doi: 10.1080/10408363.2020.1857681

Saleem, S., Yousif, M., Abubakar, M., Rehman, F., dan Saima Yousaf. (2025). Ethical Considerations in Utilizing Machine Learning for Depression and Anxiety Detection in College Students. *Journal of Computing & Biomedical Informatics*, 8(02). doi: 10.56979/802/2025

Salih Hassan, O. M., Abdulazeez, A. M., dan Tiryaki, V. M. (2018). Gait-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Based Human Gender Classification Using Lifting 5/3 Wavelet and Principal Component Analysis. Dalam *2018 international conference on advanced science and engineering (icoase)* (hal. 173–178). Duhok. doi: 10.1109/ICOASE.2018.8548909

Sathyanarayanan, S. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, 27(4), 4023–4031. Retrieved from <https://africanjournalofbiomedicalresearch.com/index.php/AJBR/article/view/4345/3327> doi: 10.53555/AJBR.v27i4S.4345

Shakir, Y. H., Aziz, E., Al, A., dan Alkhazraji, A. (2025). Leveraging Machine Learning for Early Risk Prediction in Cirrhosis Outcome Patients. *Public Research Journal of Engineering, Data Technology and Computer Science*, 3(1), 22–30. doi: <https://doi.org/10.57152/predatecs.v3i1.2015>

Shea, T. L., Tennant, A., dan Pallant, J. F. (2009). Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). *BMC Psychiatry*, 9(1), 21. Retrieved from <https://bmcpsy psychiatry.biomedcentral.com/articles/10.1186/1471-244X-9-21> doi: 10.1186/1471-244X-9-21

Shorten, C., dan Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*. Retrieved from <https://doi.org/10.1186/s40537-019-0197-0> doi: 10.1186/s40537-019-0197-0

Singh, A., dan Kumar, D. (2021). Identification of Anxiety and Depression Using DASS-21 Questionnaire and Machine Learning. Dalam *2021 first international conference on advances in computing and future communication technologies (icacfct)* (hal. 69–74). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/9837365/> doi: 10.1109/ICACFCT53978.2021.9837365

Singh, A., Singh, K., Kumar, A., Shrivastava, A., dan Kumar, S. (2024). Machine Learning Algorithms for Detecting Mental Stress in College Students. *2024 IEEE 9th International Conference for Convergence in Technology, I2CT 2024*. doi: 10.1109/I2CT61223.2024.10544243

Song, X., Zheng, Q., Zhang, R., Wang, M., Deng, W., Wang, Q., ... Ma, X. (2021). Potential Biomarkers for Predicting Depression in Diabetes Mellitus. *Frontiers in Psychiatry*, 12(November), 1–8. Retrieved from <https://www.frontiersin.org/articles/10.3389/fpsyt.2021.731220/full> doi: 10.3389/fpsyt.2021.731220

Srinath, K. S., Kiran, K., Pranavi, S., Amrutha, M., Shenoy, P. D., dan Venugopal,



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- K. R. (2022). Prediction of Depression, Anxiety and Stress Levels Using Dass-42. Dalam *2022 IEEE 7th international conference for convergence in technology (i2ct)* (hal. 1–6). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/9824087/> doi: 10.1109/I2CT54291.2022.9824087
- Salaiman, M. A. (2020). Evaluating Data Mining Classification Methods Performance in Internet of Things Applications. *Journal of Soft Computing and Data Mining*, 1(2), 11–25. Retrieved from <https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/7127>
- Sim, Y., Li, Z., Li, X., dan Zhang, J. (2021). Classifier Selection and Ensemble Model for Multi-class Imbalance Learning in Education Grants Prediction. *Applied Artificial Intelligence*, 35(4), 290–303. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/08839514.2021.1877481> doi: 10.1080/08839514.2021.1877481
- Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., dan Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: an experimental review. *Journal of Big Data*, 7(1), 70. Retrieved from <https://doi.org/10.1186/s40537-020-00349-y> doi: 10.1186/s40537-020-00349-y
- Tennenhouse, L. G., Marrie, R. A., Bernstein, C. N., dan Lix, L. M. (2020). Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease. *Journal of Psychosomatic Research*, 134(September 2019), 110126. Retrieved from <https://doi.org/10.1016/j.jpsychores.2020.110126> <https://linkinghub.elsevier.com/retrieve/pii/S0022399919308785> doi: 10.1016/j.jpsychores.2020.110126
- Usman, A., Ahmed, Z., Ahmed, I., dan Akbar, Z. (2011). Work stress experienced by the teaching staff of University of the Punjab, Pakistan: antecedents and consequences. *International Journal of Business and Social Science*, 2(8), 202–210.
- Usman, S., Rusli, F. L., Jalil, S. Z. A., dan Bani, N. A. (2022). Depression Anxiety Stress Scale and Handgrip using Machine Learning Analysis. Dalam *2022 4th international conference on smart sensors and application (icssa)* (hal. 76–80). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/9870948/> doi: 10.1109/ICSSA54161.2022.9870948
- Vercellis, C. (2009). *Business Intelligence*. Wiley. Retrieved from <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470753866> doi: 10.1002/9780470753866



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Vitória, A. R. S., Guimarães, M. O., Fazzioni, D., Díaz-Salazar, A. A., Zara, A. L. S. A., Junior, I. G. S., dan Bulcão-Neto, R. F. (2024). A Machine Learning Approach for Anxiety and Depression Prediction Using PROMIS Questionnaires. Dalam *Annals of computer science and intelligence systems* (Vol. 39, hal. 187–194). Retrieved from https://annals-csis.org/Volume_39/drp/3586.html doi: 10.15439/2024F3586
- Wang, F., Li, Z., He, F., Wang, R., Yu, W., dan Nie, F. (2019). Feature learning viewpoint of adaboost and a new algorithm. *IEEE Access*, 7, 149890-149899. doi: 10.1109/ACCESS.2019.2947359
- Wang, R., Liu, Y., Ye, X., Tang, Q., Gou, J., Huang, M., dan Wen, Y. (2019). Power System Transient Stability Assessment Based on Bayesian Optimized LightGBM. Dalam *2019 IEEE 3rd conference on energy internet and energy system integration (ei2)* (hal. 263–268). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/9062027/> doi: 10.1109/EI247390.2019.9062027
- Wang, Z., Wang, P., Liu, K., Wang, P., Fu, Y., Lu, C.-T., ... Zhou, Y. (2026). A comprehensive survey on data augmentation. *IEEE Transactions on Knowledge and Data Engineering*, 38(1), 47-66. doi: 10.1109/TKDE.2025.3622600
- Wu, P., dan Zhao, H. (2011). Some Analysis and Research of the AdaBoost Algorithm. Dalam (hal. 1–5). Retrieved from http://link.springer.com/10.1007/978-3-642-18129-0_1 doi: 10.1007/978-3-642-18129-0_1
- Xia, Y. (2020). Chapter Eleven - Correlation and association analyses in microbiome study integrating multiomics in health and disease. Dalam J. Sun (Ed.), *The microbiome in health and disease* (Vol. 171, hal. 309–491). Academic Press. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1877117320300478> doi: 10.1016/bs.pmbts.2020.04.003
- Yang, S., dan Berdine, G. (2024). Confusion matrix. *The Southwest Respiratory and Critical Care Chronicles*, 12(53), 75–79. Retrieved from <https://pulmonarychronicles.com/index.php/pulmonarychronicles/article/view/1391> doi: 10.12746/swrccc.v12i53.1391
- Yeshaw, Y., dan Mossie, A. (2017). Depression, anxiety, stress, and their associated factors among Jimma university staff, Jimma, southwest Ethiopia, 2016: A cross-sectional study. *Neuropsychiatric Disease and Treatment*, 13, 2803–2812. doi: 10.2147/NDT.S150444
- Yin, Y., Jang-Jaccard, J., Xu, W., Singh, A., Zhu, J., Sabrina, F., dan Kwak,



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

J. (2023). IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset. *Journal of Big Data*, 10(1), 15. Retrieved from <https://doi.org/10.1186/s40537-023-00694-8><https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00694-8> doi: 10.1186/s40537-023-00694-8

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., dan Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, 1(1), 56–70. Retrieved from <https://jastt.org/index.php/jastt/path/article/view/24> doi: 10.38094/jastt1224

Zeebaree, D. Q., Haron, H., dan Abdulazeez, A. M. (2018). Gene Selection and Classification of Microarray Data Using Convolutional Neural Network. Dalam *2018 international conference on advanced science and engineering (icoase)* (hal. 145–150). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/8548836/> doi: 10.1109/ICOASE.2018.8548836

Zhang, Y., Zhao, Z., dan Zheng, J. (2020). CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. *Journal of Hydrology*, 588, 125087. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0022169420305473> doi: <https://doi.org/10.1016/j.jhydrol.2020.125087>

Zhou, J., Jiang, Z., Chung, F.-L., dan Wang, S. (2021). Formulating Ensemble Learning of SVMs Into a Single SVM Formulation by Negative Agreement Learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(10), 6015–6028. Retrieved from <https://ieeexplore.ieee.org/document/8950142/> doi: 10.1109/TSMC.2019.2958647

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN A

SURAT DAN TRANSKIP WAWANCARA

SURAT KETERANGAN

Saya yang bertanda tangan dibawah ini:

Nama : Fara Ulfa, M.Psi., Psikolog
 NIP : 199301032020122020
 Jabatan : Lektor
 Fakultas : Psikologi
 Instansi : Universitas Islam Negeri Sultan Syarif Kasim Riau

Dengan ini menerangkan bahwa:

Nama : Rifsya Aulia
 NIM : 12250321571
 Program Studi : Sistem Informasi
 Fakultas : Sains dan Teknologi
 Universitas : Universitas Islam Negeri Sultan Syarif Kasim Riau
 Judul Tugas Akhir : PENERAPAN ALGORITMA ENSEMBLE LEARNING UNTUK
 KLASIFIKASI TINGKAT KESEHATAN MENTAL MAHASISWA
 MENGGUNAKAN DEPRESSION ANXIETY STRESS SCALE-21

Nama tersebut di atas telah melakukan wawancara dalam rangka penyusunan Tugas Akhir dengan judul tersebut di atas.

Demikian Surat Keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya.

Pekanbaru, 2 Desember 2025

Fara Ulfa, M.Psi., Psikolog
 NIP. 199301032020122020

Gambar A.1. Surat Keterangan Selesai Wawancara

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

TRANSKRIP HASIL WAWANCARA

Narasumber : Ibu Fara Ulfa, M.Psi, Psikolog

Hari/Tanggal : Selasa/02 Desember 2025

Lokasi : Fakultas Psikologi UIN Sultan Syarif Kasim Riau

1. Berdasarkan pengalaman Ibu, apakah saat ini gangguan psikologis seperti depresi, kecemasan, dan stres menjadi masalah yang signifikan di kalangan mahasiswa?

Berdasarkan pengalaman saya dalam praktik klinis serta pengamatan selama mendampingi mahasiswa, gangguan psikologis seperti depresi, kecemasan, dan stres memang menjadi permasalahan yang cukup signifikan di kalangan mahasiswa saat ini. Mahasiswa berada pada fase dewasa awal, yang secara psikologis tergolong rentan terhadap berbagai tekanan. Tekanan tersebut dapat berasal dari tuntutan akademik, proses penyesuaian sosial, maupun ekspektasi terhadap diri sendiri. Dalam praktiknya, cukup banyak mahasiswa yang menunjukkan gejala stres dan kecemasan, dan pada beberapa kasus, kondisi tersebut dapat berkembang menjadi depresi apabila tidak ditangani sejak tahap awal.

2. Menurut Ibu, faktor-faktor apa yang paling sering memengaruhi kondisi kesehatan mental mahasiswa, khususnya dalam konteks akademik?

Faktor yang paling sering memengaruhi kondisi kesehatan mental mahasiswa umumnya berasal dari tekanan akademik yang berlangsung secara terus-menerus. Beban tugas yang tinggi, tuntutan untuk mempertahankan prestasi akademik, jadwal perkuliahan yang padat, serta keterbatasan waktu istirahat yang berdampak pada kualitas tidur, sering kali menjadi pemicu utama munculnya stres dan kecemasan. Kondisi ini dapat semakin diperberat oleh perbedaan cara mahasiswa dalam menghadapi tekanan psikologis. Misalnya, berdasarkan pengalaman saya, mahasiswa laki-laki cenderung memiliki tingkat kepedulian terhadap kesehatan mental yang lebih rendah dan kurang terbuka dalam mengekspresikan permasalahan yang dialami. Akibatnya, dukungan sosial yang terbentuk sering kali kurang optimal. Kombinasi antara tekanan akademik yang tinggi, rendahnya kesadaran terhadap kesehatan mental, serta strategi koping yang kurang adaptif dapat meningkatkan risiko munculnya gangguan psikologis pada mahasiswa.

Gambar A.2. Transkrip Hasil Wawancara

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Apakah atribut seperti jenis kelamin, usia, jurusan, semester, dan IPK relevan digunakan sebagai variabel demografis dalam kajian kesehatan mental mahasiswa?

Atribut demografis seperti jenis kelamin, usia, jurusan, semester, dan IPK cukup relevan untuk digunakan dalam kajian kesehatan mental mahasiswa. Variabel-variabel tersebut berkaitan dengan tahapan perkembangan psikologis, perbedaan tuntutan akademik, serta tingkat tekanan yang dihadapi oleh mahasiswa. Perbedaan usia dan jenis kelamin dapat memengaruhi cara individu merespons stres, sementara jurusan dan semester biasanya berkaitan dengan variasi beban akademik. IPK juga dapat merefleksikan tekanan prestasi maupun kemampuan adaptasi mahasiswa dalam menjalani proses perkuliahan.

4. Apakah penggunaan instrumen DASS-21 sudah tepat untuk mengukur depresi, kecemasan, dan stres pada mahasiswa?

Penggunaan instrumen DASS-21 menurut saya sudah tepat untuk mengukur tingkat depresi, kecemasan, dan stres pada mahasiswa. DASS-21 merupakan alat ukur psikologis yang telah terstandarisasi serta memiliki validitas dan reliabilitas yang baik, khususnya pada populasi mahasiswa. Item-item dalam DASS-21 dirancang untuk merepresentasikan aspek emosional utama yang berkaitan langsung dengan kondisi psikologis, serta mencerminkan dampak tekanan yang umum dialami mahasiswa, seperti tuntutan akademik, kelelahan emosional, dan kesulitan untuk relaksasi. Dengan cakupan tersebut, DASS-21 mampu memberikan gambaran kondisi kesehatan mental secara cukup komprehensif dan konsisten, sehingga layak digunakan sebagai instrumen utama dalam penelitian.

5. Apakah kategorisasi tingkat gejala (Normal, Ringan, Sedang, Berat, dan Sangat Berat) sudah sesuai dengan standar psikologi?

Kategorisasi tingkat gejala tersebut sudah sesuai dengan standar psikologi dan memang umum digunakan, baik dalam praktik klinis maupun penelitian. Pembagian tingkat keparahan ini memudahkan proses interpretasi hasil pengukuran dan membantu peneliti maupun praktisi dalam mengidentifikasi tingkat gangguan psikologis yang dialami individu secara lebih jelas.

Gambar A.3. Transkrip Hasil Wawancara



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

6. Apakah penggunaan model machine learning untuk mengklasifikasikan depresi, kecemasan, dan stres pada mahasiswa merupakan pendekatan yang tepat?

Penggunaan model machine learning merupakan pendekatan yang tepat dan relevan, terutama ketika berhadapan dengan data dalam jumlah besar. Pendekatan ini dapat membantu proses klasifikasi dan deteksi dini kondisi kesehatan mental secara lebih objektif dan konsisten. Namun demikian, hasil yang diperoleh dari model machine learning tetap harus diposisikan sebagai alat bantu pendukung pengambilan keputusan, bukan sebagai pengganti diagnosis klinis. Peran tenaga profesional di bidang kesehatan mental tetap menjadi hal yang utama dalam penentuan diagnosis dan penanganan lanjutan.

Gambar A.4. Transkrip Hasil Wawancara

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN B DOKUMENTASI



Gambar B.1. Dokumentasi Wawancara



Gambar B.2. Dokumentasi Pengumpulan Data

LAMPIRAN C

DATA PENELITIAN

Tabel C.1. Data Demografis Responden

No	Nama	NIM	Jenis Kelamin	Jurusan	Usia	Semester	IPK
1	Muhammad Irfan	12450314060	Laki-Laki	Sistem Informasi	19	3	3.60
2	Danial						
3	Diva	12550125251	Laki-Laki	Teknik Informatika	18	1	0
4	Judiya Oktofiyana	12250421112	Perempuan	Matematika	21	7	3.68
5	Muhammad Ilham	12550410320	Laki-Laki	Matematika	18	1	0
6	Munzir						
7	Sarah Aulia Rianda	12450324612	Perempuan	Sistem Informasi	21	3	3.50
8	Rafa Essenza Quraniq	12450213066	Laki-Laki	Teknik Industri	19	3	3.30
9	Afif Alfarisi Hernas	12450312874	Laki-Laki	Sistem Informasi	18	3	3.97
10	Anugrah Ananda Adi Guna	12050113299	Laki-Laki	Teknik Informatika	23	11	3.23
11	Yoga Natadisastro	12150113125	Laki-Laki	Teknik Informatika	22	9	3.45
12	Akses Fadhillah	12550320084	Perempuan	Sistem Informasi	19	1	0
13	Alfarhozy Arya	12250314166	Laki-Laki	Sistem Informasi	21	7	3.40
14	Lyra Zulyanda Daulay	12250120375	Perempuan	Teknik Informatika	20	7	3.62
15	Fairuz Habibi	12550515542	Laki-Laki	Teknik Elektro	20	1	0
16	Nia Nurmahani	12450324499	Perempuan	Sistem Informasi	20	3	3.67
17	Diah Afsari	12450321251	Perempuan	Sistem Informasi	19	3	3.70
18	Muhammad Ravy Rahman Widjaya	12450214248	Laki-Laki	Teknik Industri	19	3	4.00
19	Muhammad Kemal Reza	12450211186	Laki-Laki	Teknik Industri	20	3	3.40
20	Zaky Dhiva Zulfachri	12450211448	Laki-Laki	Teknik Industri	19	3	3.55
21	Sarah Aulia Rianda	12450324612	Perempuan	Sistem Informasi	21	3	3.50
22	Rafa Essenza Quraniq	12450213066	Laki-Laki	Teknik Industri	19	3	3.30
23	Ahmad Latif	12450213096	Laki-Laki	Teknik Industri	20	3	3.56
24	Afifa	12450320814	Perempuan	Sistem Informasi	19	3	3.43
25	Nabila Armelia Sardi	12550320101	Perempuan	Sistem Informasi	18	1	0
26	Akses Fadhillah	12550320084	Perempuan	Sistem Informasi	19	1	0
27	Randy Rifaldi	12050516199	Laki-Laki	Teknik Elektro	24	11	3.34
28	...	...	...	...	...	...	...
29	Muhammad Rizki Kurniawan	12050512459	Laki-Laki	Teknik Elektro	23	11	3.02

LAMPIRAN D

KUESIONER PENELITIAN

Kuesioner Penelitian Tugas Akhir

Pertanyaan Jawaban 432 Setelan

Bagian 1 dari 2

Kuesioner Penelitian Tugas Akhir

Assalamualaikum Warahmatullahi Wabarakatuh

Saya Rifsyia Aulia, mahasiswa Program Studi Sistem Informasi Semester 7, sedang melakukan penelitian Tugas Akhir. Mohon ketersediannya untuk mengisi kuesioner ini.

Catatan Penting:

- Jawaban Anda akan dirahasiakan dan hanya digunakan untuk kepentingan penelitian
- Mohon isi kuesioner ini sesuai kondisi dalam 7 hari terakhir
- Tidak ada jawaban benar atau salah, jadi jawablah dengan jujur agar hasil penelitian akurat
- Waktu pengisian ± 5–10 menit

Terima kasih atas partisipasi Anda!
Partisipasi Anda sangat berarti untuk membantu menciptakan lingkungan akademik yang lebih sehat.

Gambar D.1. Kuesioner Penelitian Tugas Akhir



DAFTAR RIWAYAT HIDUP



Peneliti bernama Rifsya Aulia, lahir di Kabupaten Pelalawan pada tanggal 9 Februari 2004. Peneliti berasal dari Desa Sialang Godang, Kecamatan Bandar Petalangan, Kabupaten Pelalawan, Provinsi Riau. Peneliti merupakan anak keempat dari empat bersaudara, dari pasangan Syarifuddin Lubis dan Indah Saedah Munthe. Pendidikan formal Peneliti dimulai dari TK Putra Mandiri 2 yang diselesaikan pada tahun 2009, kemudian melanjutkan pendidikan di SD Negeri 007 Sialang Godang dan lulus pada tahun 2016. Selanjutnya Peneliti menempuh pendidikan di SMP Negeri 1 Pangkalan Kuras dan lulus pada tahun 2019, serta melanjutkan ke SMA Negeri 1 Pangkalan Kuras hingga lulus pada tahun 2022. Pada tahun 2022, Peneliti melanjutkan pendidikan ke jenjang perguruan tinggi di Universitas Islam Negeri Sultan Syarif Kasim Riau, Fakultas Sains dan Teknologi, Program Studi Sistem Informasi. Pada tahun 2024, Peneliti melaksanakan Kerja Praktek di PT. Adei Plantation & Industry. Selanjutnya, pada tahun 2025, Peneliti melaksanakan Kuliah Kerja Nyata (KKN) di Desa Pulau Gadang, Kecamatan XIII Koto Kampar, Kabupaten Kampar, Provinsi Riau. Selain kegiatan akademik, Peneliti juga aktif dalam sebuah organisasi/komunitas riset di bawah Fakultas Sains dan Teknologi (FST) UIN Sultan Syarif Kasim Riau bernama *Puzzle Research Data Technology* (Predatech) dan menjabat sebagai Koordinator Divisi Riset dan Publikasi.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.