



UIN SUSKA RIAU

PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING DALAM KLASIFIKASI DEPRESI MAHASISWA

TUGAS AKHIR

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

Oleh:

FATIMAH AZZAHRA

12250324568



UIN SUSKA RIAU

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2026

© Hak cipta milik 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.

LEMBAR PERSETUJUAN
PERBANDINGAN KINERJA ALGORITMA MACHINE
LEARNING DALAM KLASIFIKASI DEPRESI MAHASISWA

TUGAS AKHIR

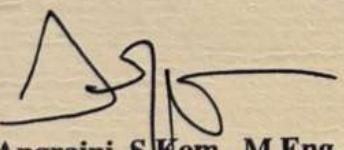
Oleh:

FATIMAH AZZAHRA

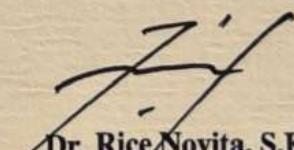
12250324568

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 20 Januari 2026

Ketua Program Studi


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

Pembimbing


Dr. Rice Novita, S.Kom., M.Kom.
NIP. 198511272023212032

LEMBAR PENGESAHAN

PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING DALAM KLASIFIKASI DEPRESI MAHASISWA

TUGAS AKHIR

Oleh:

FATIMAH AZZAHRA

12250324568

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 12 Januari 2026

Pekanbaru, 12 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 : Nesdi Evrilyan Rozanda, S.Kom., M.Sc.

Sekretaris : Dr. Rice Novita, S.Kom., M.Kom.

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

Anggota 2 : Medyantiwi Rahmawita M, ST., M.Kom.

Lampiran Surat :

Nomor : Nomor 25/2021

Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Fatimah Azzahra

NIM : 12250324568

Tempat/Tgl. Lahir : Singkarak / 09 Oktober 2004

Fakultas/Pascasarjana : Sains dan Teknologi

Prodi : Sistem Informasi

Judul Skripsi :

PERBANDINGAN KINERJA ALGORITMA *MACHINE LEARNING* DALAM
KLASIFIKASI DEPRESI 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 penulisan 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, 20 Januari 2026

Yang membuat pernyataan



Fatimah Azzahra

NIM: 12250324568



UIN SUSKA RIAU

LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum, dengan ketentuan bahwa hak cipta ada pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan atas izin penulis dan harus dilakukan mengikuti kaedah dan kebiasaan ilmiah serta menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin tertulis dari Dekan fakultas universitas. Perpustakaan dapat meminjamkan Tugas Akhir ini untuk anggotanya dengan mengisi nama, tanda peminjaman dan tanggal pinjam pada *form* peminjaman.

- 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 meugikan 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.



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 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 dicantumkan dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 20 Januari 2026

Yang membuat pernyataan,

Fatimah Azzahra

NIM. 12250324568



UIN SUSKA RIAU

LEMBAR PERSEMPAHAN



Assalamu'alaikum Warahmatullahi Wabarakatuh

Adhamdulillahi Rabbil 'Alamiin, segala puji bagi Allah Subhanahu Wa Ta'ala sebagai bentuk rasa syukur atas segala nikmat yang telah diberikan tanpa ada kekuatan sedikitpun. Shalawat beserta salam tak lupa pula kita ucapkan kepada Nabi Muhammad Shallallahu 'Alaihi Wa Sallam dengan mengucapkan "Allahumma Sholli 'ala Sayyidina Muhammad Wa 'ala Ali Sayyidina Muhammad". Semoga kita semua selalu senantiasa mendapat syafaat-Nya di dunia maupun di akhirat, Aamiin Ya Rabbal'aalamiin.

Terima kasih kepada ayah, ibu, kakak, abang dan adik tersayang atas setiap doa, dukungan, semangat, dan bimbingan yang selalu diberikan kepada peneliti sampai saat ini. Berkat doa dan kasih sayangmu, anakmu telah berhasil memperoleh gelar sarjana seperti yang engkau harapkan. Tiada apapun di dunia ini yang dapat membala semua jasa-jasa dan pengorbananmu. Peneliti sebagai anakmu ini selalu mendoakan yang terbaik untuk ayah dan ibu agar bahagia dunia akhirat, serta diberikan tempat istimewa di sisi-Nya kelak. Peneliti juga berterima kasih yang tak terhingga kepada saudara kandung peneliti yaitu kakak, abang dan adik yang selalu memberikan dukungan, semangat, pelajaran, serta pemahaman mengenai pengalaman kehidupan yang menjadi ilmu bagi saya.

Kepada Bapak dan Ibu Dosen Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau yang telah memberikan ilmu pengetahuan bermanfaat, pengalaman berharga, dan kebaikan yang tulus selama perkuliahan, peneliti ucapkan terima kasih banyak dan semoga menjadi amal jariyah. Aamiin.

Sahabat-sahabat terdekat yang tidak bisa peneliti sebutkan satu-persatu dan pastinya juga teman-teman seperjuangan, terima kasih berkat kalian masa perkuliahan menjadi lebih bermakna dan menyenangkan semoga di masa mendatang kita bisa bertemu lagi dalam keadaan yang lebih baik.

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 mengujikan 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.

**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 meugikan 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.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillahi Rabbil 'Alamin, puji syukur peneliti panjatkan ke hadirat Allah Subhanahu Wa Ta'ala yang telah melimpahkan rahmat dan hidayah-Nya, sehingga peneliti dapat menyelesaikan Tugas Akhir ini dengan baik. Shalawat serta salam senantiasa tercurah kepada Nabi Muhammad Shallallahu 'Alaihi Wa Sallam dengan mengucapkan *Allahumma Sholli 'ala Sayyidina Muhammad Wa 'ala Ali Sayyidina Muhammad*.

Laporan Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana, sekaligus sebagai bentuk pembelajaran akademis dan spiritual. Oleh karena itu, peneliti mengucapkan terima kasih kepada seluruh pihak yang telah membantu dalam setiap proses penelitian, baik dalam bentuk dukungan material, motivasi, maupun doa. Pada kesempatan ini, peneliti menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Prof. Dr. Hj. Leny Nofianti, MS., SE., M.Si., Ak., CA., Sebagai Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Ibu Dr. Yuslenita Muda,S.Si,M.Sc, Sebagai Dekan Fakultas Sains dan Teknologi.
3. Ibu Angraini, S.Kom., M.Eng., Ph.D., Sebagai Ketua Program Studi Sistem Informasi.
4. Ibu Dr. Rice Novita, S.Kom., M.Kom., Sebagai Sekretaris Program Studi Sistem Informasi sekaligus Dosen Pembimbing I yang telah banyak memberikan arahan, masukan, serta motivasi selama perkuliahan dan penyusunan Tugas Akhir ini.
5. Bapak M. Afdal, ST., M.Kom.,Sebagai Dosen Pengaji I Tugas Akhir yang dengan penuh dedikasi dan ketelitian telah memberikan arahan, masukan, nasihat, serta motivasi dalam penyelesaian Tugas Akhir ini.
6. Ibu Medyantiwi Rahmawita M, ST., M.Kom.,Sebagai Dosen Pengaji II yang dengan penuh dedikasi dan ketelitian telah memberikan arahan, masukan, nasihat, serta motivasi dalam penyelesaian Tugas Akhir ini.
7. Bapak Nesdi Evrilyan Rozanda, S.Kom., M.Sc., Sebagai Ketua Sidang yang telah memimpin sidang Tugas Akhir dengan bijaksana serta memberikan arahan dan masukan yang sangat berarti demi perbaikan Tugas Akhir ini.
8. Ibu Fitriani Muttakin, S.Kom., M.Cs., Sebagai Dosen Pembimbing Akademik peneliti yang telah banyak memberikan arahan, masukan, dan motivasi selama masa perkuliahan.



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 mengujikan 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.

9. Kedua orang tua tercinta, Ayah Iwan Setiawan dan Ibu Lenny, yang senantiasa memberikan kasih sayang, dukungan moral maupun material, perhatian, semangat, serta doa yang tiada henti sejak peneliti kecil hingga berhasil menyelesaikan Laporan Tugas Akhir ini. Segala pengorbanan, kesabaran, dan ketulusan yang telah diberikan menjadi sumber kekuatan dan motivasi utama bagi peneliti dalam menempuh pendidikan dan menghadapi setiap proses serta tantangan selama masa perkuliahan.
10. Abang terbaik sepanjang masa 'aa iki' dan adik tercinta aisyah dan abdullah yang selalu menjadi sumber semangat dalam menyelesaikan Laporan Tugas Akhir ini.
11. KLC PSIKOLOG yang telah bersedia membantu peneliti dalam proses penyusunan Laporan Tugas Akhir ini.
12. Kepada Fitri Nadia Rahima terimakasih banyak telah menemani dari maba hingga selesaiya perkuliahan ini, terimakasih telah memberikan banyak pelajaran dan selalu bersama walaupun banyak rintangan yang harus dilewati, Mari tetap selalu menjaga pertemanan ini.
13. Teman-teman Girls Time (Upiw dan Eka) serta Pemain Inti (Intan, Umi, Nabilah, Habiba, Ade, Satria, dan Koko) yang hadir di penghujung masa perkuliahan dan menjadi sumber keceriaan serta kebersamaan. Kehadiran, tawa, dan dukungan yang diberikan telah membantu peneliti melewati berbagai proses dan tekanan selama penyusunan skripsi, sehingga memberikan semangat dan suasana positif hingga tahap akhir perkuliahan.
14. Peneliti juga mengucapkan terima kasih kepada seseorang yang senantiasa berusaha memberikan yang terbaik, memberikan dukungan secara tulis, serta selalu menemani peneliti dalam setiap proses selama penyusunan skripsi ini. Kehadiran dan perhatian yang diberikan menjadi salah satu sumber kekuatan bagi peneliti untuk tetap bertahan, fokus, dan menyelesaikan Tugas Akhir ini dengan baik.
15. Teman-teman, khususnya Saumi Sintia, meli, Rifda, Zahra dan Intan, yang senantiasa menjadi tempat berbagi cerita, memberikan dukungan, serta semangat selama proses perkuliahan. Ucapan terima kasih juga peneliti sampaiakan kepada seluruh Kelas Apatis dan angkatan 2022 yang tidak dapat disebutkan satu per satu atas bantuan dan motivasi yang telah diberikan.
16. Keluarga besar Puzzle Research Data Technology (Predatech) yang telah menjadi wadah pengembangan minat dan bakat, serta memberikan kepercayaan sebagai Sekretaris Kegiatan. Terima kasih atas kebersamaan, kerja sama, dan pengalaman berharga untuk penulis.



UIN SUSKA RIAU

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 mengujikan 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.

17. Terakhir, peneliti mengucapkan terima kasih kepada diri sendiri yang telah bertahan, tetap kuat, dan tegar dalam menghadapi berbagai proses, tantangan, serta dinamika yang dilalui selama masa perkuliahan. Peneliti mampu menyelesaikan Tugas Akhir ini dengan penuh tanggung jawab dan ketekunan.
18. Seluruh pihak yang tidak dapat disebutkan satu per satu yang telah banyak membantu peneliti dalam menyelesaikan Tugas Akhir ini.

Peneliti mengucapkan terima kasih yang sebesar-besarnya. Semoga segala kebaikan yang telah diberikan menjadi ladang pahala dan mendapatkan balasan dari Allah Subhanahu Wa Ta'ala. Semoga kita semua senantiasa diberikan kebahagiaan dan kesehatan. Aamiin Ya Rabbal 'Aalamiin. Peneliti menyadari bahwa penulisan Tugas Akhir ini masih memiliki keterbatasan dan jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Pertanyaan maupun masukan dapat disampaikan melalui surat elektronik ke alamat 12250324568@students.uin-suska.ac.id. Semoga laporan ini dapat memberikan manfaat bagi semua pihak.

Akhir kata, peneliti mengucapkan terima kasih.

Pekanbaru, 20 Januari 2026

Penulis,

Fatimah Azzahra

NIM. 12250324568

UIN SUSKA RIAU



UIN SUSKA RIAU

PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING DALAM KLASIFIKASI DEPRESI MAHASISWA

FATIMAH AZZAHRA

NIM: 12250324568

Tanggal Sidang: 12 Januari 2026

Program Studi Sistem Informasi

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl. Soebrantas, No. 155, Pekanbaru

ABSTRAK

Depresi merupakan salah satu permasalahan kesehatan mental yang banyak dialami oleh mahasiswa dan memerlukan pendekatan klasifikasi yang akurat serta berbasis data. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Random Forest* dan *Multilayer Perceptron with Stochastic Gradient Descent* (MLP–SGD) dalam klasifikasi depresi mahasiswa menggunakan data sekunder *Student Depression Dataset* dari Kaggle sebanyak 27.901 baris serta data primer berbasis kuesioner *Patient Health Questionnaire-9* (PHQ-9) dari 120 responden. Kedua model diuji pada kondisi *baseline* dan dioptimasi menggunakan *GridSearchCV* dengan tahapan pra-pemrosesan meliputi pembersihan data, transformasi fitur kategorikal, normalisasi fitur numerik, serta pembagian data secara *stratified* dengan rasio 80:20. Hasil evaluasi menunjukkan bahwa MLP–SGD dengan konfigurasi *hyperparameter* optimal memberikan kinerja yang konsisten lebih unggul dibandingkan *Random Forest*, dengan nilai *accuracy* sebesar 0,8366, *F1-score* sebesar 0,8345, dan *AUC-ROC* sebesar 0,9137 pada data sekunder, serta *accuracy* sebesar 83%, *F1-score* sebesar 0,83, dan *AUC-ROC* sebesar 0,9 pada data primer. Hasil ini menunjukkan bahwa MLP–SGD memiliki kemampuan generalisasi dan diskriminasi kelas yang lebih baik dalam klasifikasi depresi mahasiswa.

Kata Kunci: Depresi Mahasiswa, *GridSearchCV*, Klasifikasi, *Machine Learning*



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 meugikan 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.

PERFORMANCE COMPARISON OF MACHINE LEARNING ALGORITHMS FOR STUDENT DEPRESSION CLASSIFICATION

FATIMAH AZZAHRA
NIM: 12250324568

Date of Final Exam: Januari 12th 2026

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

ABSTRACT

Depression is one of the most prevalent mental health problems experienced by university students and requires an accurate and data-driven classification approach. This study aims to compares the performance of the Random Forest and Multilayer Perceptron with Stochastic Gradient Descent (MLP-SGD) algorithms in classifying student depression using secondary data from the Student Depression Dataset obtained from Kaggle, consisting of 27,901 records, as well as primary data collected through the Patient Health Questionnaire-9 (PHQ-9) from 120 respondents. Both models were evaluated under baseline conditions and optimized using GridSearchCV. The preprocessing stages included data cleaning, categorical feature transformation, numerical feature normalization, and stratified data splitting with an 80:20 ratio. The evaluation results indicate that the optimized MLP-SGD model consistently outperformed the Random Forest algorithm. On the secondary dataset, the MLP-SGD achieved an accuracy of 0.8366, an F1-score of 0.8345, and an AUC-ROC value of 0.9137. Similarly, on the primary dataset, the model attained an accuracy of 83%, an F1-score of 0.83, and an AUC-ROC value of 0.9. These results demonstrate that the MLP-SGD model exhibits superior generalization capability and class discrimination performance in student depression classification.

Keywords: Classification, GridSearchCV, Machine Learning, Student Depression



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 ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBERAHAN	vi
KATA PENGANTAR	vii
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvi
DAFTAR SINGKATAN	xvii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
LANDASAN TEORI	6
2.1 Depresi	6
2.2 Depresi Pada Mahasiswa	7
2.3 <i>Machine Learning</i>	8
2.4 <i>Multilayer Perceptron (MLP)</i>	9
2.5 <i>Random Forest</i>	10
2.6 <i>Stochastic Gradient Descent(SGD)</i>	12



UIN SUSKA RIAU

2.7	<i>GridSearchCV</i>	13
2.8	Data Sekunder	15
2.9	Data Primer	15
2.10	<i>Patient Health Questionnaire-9 (PHQ-9)</i>	16
2.11	<i>Confusion Matrix</i>	17
2.12	<i>Flask</i>	18
2.13	Penelitian Terdahulu	20
METODOLOGI PENELITIAN		22
3.1	Tahap Perencanaan Penelitian	23
3.1.1	Identifikasi Masalah	23
3.1.2	Menentukan Tujuan	23
3.1.3	Menentukan Batasan Masalah	23
3.1.4	Studi Pustaka	24
3.2	Tahap Pengumpulan Data	24
3.2.1	Data Sekunder	24
3.2.2	Data Primer	26
3.3	Pre-processing Data	27
3.3.1	Pembersihan Data dan Seleksi Fitur (<i>Data Cleaning & Feature Selection</i>)	27
3.3.2	Transformasi Fitur Kategorikal (<i>Encoding</i>)	28
3.3.3	Pembagian Data (<i>Data Splitting</i>)	29
3.3.4	Normalisasi Data (<i>Feature Scaling</i>)	29
3.4	Penerapan Model	29
3.4.1	<i>Random Forest</i>	29
3.4.2	<i>Multilayer Perceptron</i> dengan <i>Stochastic Gradient Descent</i> (MLP-SGD)	30
3.5	Optimasi <i>Hyperparameter</i>	30
3.6	Evaluasi Model	30
3.7	Tahap <i>Deployment</i>	31
ANALISA DAN HASIL		32
4.1	Tahap Pengumpulan Data	32
4.1.1	Data Sekunder Kaggle	32
4.1.2	Data Primer PHQ-9	33
4.2	Preprocessing Data	34
4.2.1	Data Cleaning & Feature Selection	34
4.2.2	Tahap Transformasi Fitur(<i>Encoding</i>)	38



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 meugikan 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.

4.2.3	Split Data	39
4.2.4	Tahapan Normalisasi	40
4.3	Pelatihan Model	41
4.3.1	Pelatihan Model Baseline	41
4.3.1.1	Random Forest	42
4.3.1.2	MLP-SGD	44
4.3.2	Proses Pelatihan Model dengan <i>Hyperparameter Tuning</i>	49
4.3.2.1	Random Forest	49
4.3.2.2	MLP-SGD	53
4.3.3	Perbandingan Hasil Model Optimal	56
4.3.4	Feature Importance	58
4.4	Evaluasi Model	59
4.5	Evaluasi dengan Data Primer	61
4.6	Implementasi Model Berbasis Web	62
4.7	Diskusi	64
5	PENUTUP	66
5.1	Kesimpulan	66
5.2	Saran	66
DAFTAR PUSTAKA		
LAMPIRAN A IZIN PENELITIAN DAN VALIDASI DATA		A - 1
LAMPIRAN B KUESIONER PENELITIAN		B - 1
LAMPIRAN C UJI COBA		C - 1
LAMPIRAN D DOKUMENTASI		D - 1

UIN SUSKA RIAU



DAFTAR GAMBAR

2.1	Lapisan Utama MLP	9
2.2	Diagram <i>Confusion Matrix</i>	17
2.3	Logo <i>Flask</i>	19
3.1	<i>Flowchart</i> Tahapan Penelitian	22
4.1	Dataset Klasifikasi derpresi	32
4.2	Hasil <i>filtering</i> mahasiswa	35
4.3	Hasil Enconding Gender	39
4.4	Visualisasi hasil <i>Min-Max Scaling</i>	41
4.5	<i>Confusion Matrix Random Forest Baseline</i>	43
4.6	<i>ROC Curve Random Forest Baseline</i>	44
4.7	<i>Confusion Matrix MLP-SGD Baseline</i>	46
4.8	<i>ROC Curve MLP-SGD Baseline</i>	47
4.9	Perbandingan Hasil Pelatihan Model <i>Baseline</i>	48
4.10	<i>ROC Curve</i> Perbandingan Hasil Pelatihan Model <i>Baseline</i>	49
4.11	<i>Confusion Matrix</i> Model <i>Random Forest Optimasi</i>	51
4.12	<i>Learning Curve</i> Model <i>Random Forest Optimasi</i>	52
4.13	<i>Confusion Matrix</i> Model <i>MLP-SGD Optimasi</i>	55
4.14	<i>ROC Curve</i> Model <i>MLP-SGD Optimasi</i>	56
4.15	Perbandingan Hasil Pelatihan Model Optimasi	57
4.16	Visualisasi <i>Feature Importance</i> Depresi	58
4.17	Hasil Perbandingan Model <i>Baseline</i> dan Optimasi	60
4.18	Arsitektur Sistem Berbasis <i>Framework Flask</i>	62
4.19	Antarmuka <i>Input</i> Variabel Psikososial dan Akademik	63
4.20	Tampilan Hasil Prediksi dan <i>Confidence Score</i>	63
A.1	Surat Izin Pemngambilan data	A - 1
A.2	Surat Izin Validasi data	A - 2
B.1	Tampilan Kuesioner PHQ-9	B - 1
B.2	Pengisian Data Pribadi	B - 1
B.3	Faktor Terkait Depresi	B - 1
B.4	Pernyataan Instrumen PHQ-9	B - 1
D.1	Dokumentasi bersama pihak KLC psikolog	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 TABEL

3.1	Karakteristik Data	25
3.2	Karakteristik Responden Data Primer	26
3.3	Instrumen Pernyataan PHQ-9	26
4.1	Dataset Depresi pada Mahasiswa	32
4.2	Distribusi Kelas Data Depresi	33
4.3	Distribusi Status Depresi Responden (Data Primer PHQ-9)	33
4.4	Dataset Responden PHQ-9	33
4.5	Tampilan Data pada Variabel Profession	34
4.6	Hasil Penghapusan Kolom yang Tidak Relevan	36
4.7	Hasil Akhir Fitur Dataset Setelah Pembersihan	36
4.8	Hasil Cleaning Data pada Variabel Financial Stress	37
4.9	Variabel Degree	37
4.10	Hasil eliminasi Variabel Degree	37
4.11	Variabel Degree	38
4.12	Hasil Pengelompokan pada Variabel Sleep Duration	38
4.13	Pembagian Subset Data (Train-Test Split)	39
4.14	Dataset Setelah Tahap Normalisasi dan Encoding	40
4.15	Parameter <i>Random Forest</i>	42
4.16	Hasil Performa Model <i>Random Forest Baseline</i>	42
4.17	Parameter MLP <i>Baseline</i>	44
4.18	Hasil MLP-SGD <i>Baseline</i>	45
4.19	Perbandingan Hasil Pelatihan Model <i>Baseline</i>	47
4.20	Hasil <i>Best Hyperparameter Random Forest</i>	50
4.21	Hasil Performa Model <i>Random Forest</i> Setelah Optimasi	50
4.22	Hasil <i>Best Hyperparameter MLP-SGD</i>	53
4.23	Hasil Performa Model MLP-SGD Setelah Optimasi	54
4.24	Perbandingan Hasil Evaluasi Model Optimal	56
4.25	Perbandingan Kinerja Model <i>Baseline</i> dan Optimal	59
4.26	Perbandingan <i>Confusion Matrix</i> Model <i>Baseline</i> dan Optimal	61
4.27	Hasil Evaluasi Menggunakan Data Primer	61

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 mengujikan 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.



UIN SUSKA RIAU

© 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.

DAFTAR SINGKATAN

ML	:	Machine Learning
RF	:	Random Forest
MLP	:	Multilayer Perceptron
SGD	:	Stochastic Gradient Descent
PHQ-9	:	Patient Health Questionnaire-9
CV	:	Cross-Validation
TP	:	True Positive
TN	:	True Negative
FP	:	False Positive
FN	:	False Negative

**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 mengikuti 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 1

PENDAHULUAN

1.1 Latar Belakang

Depresi pada mahasiswa merupakan masalah kesehatan mental yang signifikan dengan prevalensi tinggi dan dampak luas terhadap kehidupan akademik dan sosial. *Meta-analisis* global (Kiekens dkk., 2023); (Santomauro dkk., 2021) mencatat bahwa prevalensi depresi pada mahasiswa secara global rata-rata mencapai 30,6%. Studi *Healthy Minds* di Amerika Serikat melaporkan bahwa 44% mahasiswa mengalami gejala depresi sedang hingga berat, dengan 15% di antaranya pernah mempertimbangkan bunuh diri (Bantjes, 2024). Di Asia Tenggara, prevalensi rata-rata depresi mahasiswa mencapai 29,4% (Dessauvagie, Dang, Nguyen, dan Groen, 2022), menegaskan bahwa fenomena ini tidak terbatas pada wilayah tertentu, melainkan menjadi isu lintas negara.

Kondisi serupa juga terjadi di Indonesia, berdasarkan laporan Survei Kesehatan Indonesia(SKI) 2023 yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia, prevalensi depresi secara nasional mencapai 1,4%, dengan kelompok usia 15–24 tahun menunjukkan angka yang lebih tinggi yaitu 2,0% (Wilaksono, Damry, Aminuddin, dan Alfiandari, 2025). Pada tingkat pendidikan tinggi, penelitian (Ramadianto, Kusumadewi, Agiananda, dan Raharjanti, 2022) menemukan bahwa 22,2% mahasiswa kedokteran di Indonesia mengalami depresi. Temuan serupa pada penelitian (Laoli, Nauli, dan Karim, 2022) melaporkan bahwa mahasiswa Universitas Riau menunjukkan 35% mahasiswa tingkat akhir berada pada kategori depresi sedang.

Faktor penyebab utama meliputi tekanan akademik, *stres* finansial, dan dukungan sosial yang lemah ((Han dkk., 2025); (Ramadianto dkk., 2022)). Penelitian yang dilakukan oleh (Kristensen, Larsen, Urke, dan Danielsen, 2023) menunjukkan bahwa tingginya tekanan akademik, rendahnya *self-esteem*, serta lemahnya resiliensi dalam menghadapi tuntutan perkuliahan memperburuk kondisi psikologis mahasiswa.

Faktor-faktor tersebut berpotensi meningkatkan risiko depresi yang selanjutnya dapat berdampak pada penurunan prestasi akademik, risiko putus kuliah, hingga munculnya keinginan untuk bunuh diri (Steare, Muñoz, Sullivan, dan Lewis, 2023). Kondisi ini menunjukkan bahwa depresi mahasiswa merupakan isu serius yang memerlukan pendekatan sistematis berbasis analisis data untuk memahami pola dan faktor penyebabnya. Berbagai metode telah diterapkan dalam mengidentifikasi depresi. Pada penelitian ini, klasifikasi depresi digunakan untuk menentukan

secara objektif apakah seorang mahasiswa termasuk kategori depresi atau tidak, sehingga proses penilaian menjadi lebih terstruktur.

Namun, metode identifikasi depresi secara manual seperti wawancara klinis sering kali bersifat subjektif, memakan waktu, dan kurang efektif jika diterapkan dalam skala besar (Al Masud, Shanto, Sakin, dan Kabir, 2025). Oleh karena itu, metode *machine learning*(ML) semakin berkembang sebagai solusi karena mampu mengolah data kompleks secara efisien dan objektif. ML juga memungkinkan analisis komparatif antar-algoritma untuk menentukan metode klasifikasi yang paling tepat dan akurat dalam mendeteksi depresi (Mumenin dkk., 2024).

Salah satu algoritma yang banyak digunakan dalam klasifikasi kesehatan mental adalah *Random Forest*(RF). Algoritma ini dikenal efektif dalam mengolah data dengan variabel campuran dan menghasilkan performa stabil karena merupakan metode *ensemble* dari pohon keputusan (Elovanio dkk., 2020). Penelitian pada data kesehatan mental menunjukkan bahwa RF mampu mencapai akurasi tinggi, misalnya pada klasifikasi tingkat kecemasan mahasiswa dengan akurasi 98,14% (Mohamed dkk., 2023).

Sebaliknya, *Multilayer Perceptron*(MLP) adalah jaringan saraf tiruan yang mampu memodelkan hubungan *non-linear* yang lebih kompleks antar fitur (Mumenin dkk., 2024). Optimasi MLP sangat bergantung pada algoritma pembelajaran seperti *Stochastic Gradient Descent*(SGD), yang secara iteratif memperbarui bobot jaringan untuk meminimalkan fungsi *loss* (Kabir, Hossain, Rahman, dan Mishu, 2023). Penggunaan SGD memungkinkan MLP beradaptasi dengan pola data yang rumit dan meningkatkan kemampuan generalisasi model (Li, Zou, dan Li, 2024). Penelitian (Li dkk., 2024) melaporkan bahwa MLP dengan optimisasi SGD menghasilkan akurasi 87,1%, melampaui RF yang mencapai 72,3% pada klasifikasi depresi.

Kinerja optimal dari MLP dan RF sangat bergantung pada pemilihan *hyperparameter* yang tepat (Rasheed, Kumar, Rani, Kantipudi, dan Anila, 2024). Pada MLP, *hyperparameter* seperti *learning rate*, jumlah neuron, dan jumlah lapisan tersembunyi harus dioptimalkan agar SGD dapat bekerja efektif dan model dapat berkonvergensi dengan baik (Li dkk., 2024). Sedangkan pada RF, parameter seperti jumlah pohon (*n_estimators*) dan kedalaman pohon (*max_depth*) perlu disesuaikan untuk mencapai keseimbangan antara *bias* dan *varians*, sehingga menghindari *underfitting* atau *overfitting* ((Saha dkk., 2024); (Cukic, Lopez, dan Pavon, 2020)).

Teknik optimasi *hyperparameter* yang umum digunakan adalah *Grid-SearchCV*, yang melakukan pencarian sistematis pada kombinasi parameter yang telah ditentukan untuk menemukan konfigurasi terbaik. Studi oleh (Saha dkk.,

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 meugikan 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.

2024) melaporkan bahwa penerapan *GridSearchCV* pada MLP meningkatkan akurasi secara signifikan. Hasil serupa ditemukan pada RF, di mana *tuning* jumlah pohon dan kedalaman pohon dengan *GridSearchCV* menghasilkan akurasi sebesar 98,3%, meningkat dari 95,7% pada konfigurasi *default* (Rasheed dkk., 2024).

Meskipun banyak penelitian telah mengaplikasikan MLP dan RF untuk klasifikasi depresi, studi yang secara sistematis membandingkan kedua algoritma ini dengan penerapan optimasi *hyperparameter* masih sangat terbatas (Cukic dkk., 2020). Urgensi penelitian ini muncul dari keterbatasan studi terdahulu yang belum melakukan perbandingan terstruktur antara algoritma *Random Forest* dan *Multi-layer Perceptron-SGD*, terutama dengan penerapan teknik *GridSearchCV*. Dengan melakukan perbandingan tersebut, penelitian ini diharapkan dapat memberikan rekomendasi model yang paling efektif serta meningkatkan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah bagaimana perbandingan kinerja algoritma *Random Forest* dan *Multi-layer Perceptron-SGD* dalam klasifikasi depresi mahasiswa dengan optimasi *hyperparameter GridSearchCV*?

1.3 Batasan Masalah

Batasan masalah tugas akhir ini adalah:

1. Penelitian ini menggunakan *dataset* publik dari *Kaggle* sebagai data sekunder yang berisi data mahasiswa dengan status depresi yang telah dianonimkan untuk menjaga kerahasiaan identitas responden, serta data kuesioner PHQ-9 sebagai data primer.
2. Penelitian dibatasi pada implementasi hasil model ke dalam sistem nyata secara sederhana, tanpa pengembangan sistem berskala besar.
3. Penelitian berfokus pada pencarian algoritma yang optimal untuk klasifikasi depresi mahasiswa menggunakan algoritma MLP-SGD dan RF.
4. Pengujian menerapkan *optimizer*, yaitu *Stochastic Gradient Descent(SGD)* untuk algoritma MLP dan menggunakan *hyperparameter GridSearchCV* untuk kedua algoritma.
5. Evaluasi performa model dilakukan menggunakan metrik dari *confusion matrix*, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-score*.

1.4 Tujuan Penelitian

Tujuan tugas akhir ini adalah:



- Hak Cipta Dilindungi Undang-Undang**
1. Menganalisis perbandingan kinerja algoritma Random Forest dan Multilayer Perceptron-SGD dalam klasifikasi depresi mahasiswa.
 2. Mengidentifikasi pengaruh optimasi hyperparameter terhadap performa kedua algoritma dengan menggunakan GridSearchCV.
 3. Menghasilkan model klasifikasi depresi mahasiswa yang memiliki performa terbaik berdasarkan hasil evaluasi komparatif.

1.5 Manfaat Penelitian

Manfaat tugas akhir ini adalah:

1. Menjadi referensi bagi penelitian sejenis yang membahas efektivitas perbandingan algoritma *Random Forest* dan *Multilayer Perceptron with Stochastic Gradient Descent*(MLP-SGD) yang dioptimasi menggunakan *GridSearchCV*, khususnya dalam bidang *machine learning* untuk kesehatan mental.
2. Menghasilkan model klasifikasi risiko depresi pada mahasiswa melalui penerapan algoritma *Random Forest* dan *MLP-SGD* dengan optimasi *GridSearchCV*.
3. Membantu praktisi kesehatan mental(layanan psikologi) serta institusi perguruan tinggi dalam mengidentifikasi mahasiswa yang berisiko mengalami depresi secara lebih cepat, tepat, dan berbasis data.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

BAB 1. PENDAHULUAN

Bab ini membahas deskripsi umum dari tugas akhir, meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan tugas akhir.

BAB 2. LANDASAN TEORI

Bab ini membahas teori yang mendasari penelitian, seperti depresi pada mahasiswa, algoritma *Multilayer Perceptron*, optimisasi *Stochastic Gradient Descent* (SGD), *Random Forest*, teknik optimasi *hyperparameter* dengan *GridSearchCV*, serta studi terdahulu yang relevan.

BAB 3. METODOLOGI PENELITIAN

Bab ini menguraikan metode yang digunakan dalam penelitian, meliputi deskripsi *dataset* depresi pada mahasiswa, tahapan *preprocessing*, penerapan algoritma *Multilayer Perceptron-SGD* dan *Random Forest*, proses *tuning hyperparameter* menggunakan *GridSearchCV*, serta metode evaluasi model untuk membandingkan performa klasifikasi.



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 meugikan 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 4. ANALISA DAN HASIL

Bab ini menyajikan hasil klasifikasi depresi pada mahasiswa menggunakan algoritma *Multilayer Perceptron-SGD* dan *Random Forest*. Pembahasan difokuskan pada evaluasi kinerja model berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta analisis efektivitas optimasi *hyperparameter* melalui *GridSearchCV* dalam meningkatkan performa klasifikasi.

BAB 5. PENUTUP

Bab ini menyajikan kesimpulan dari penelitian tugas akhir yang dilakukan, serta saran untuk penelitian selanjutnya.



**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 mengurangi 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 2

LANDASAN TEORI

2.1 Depresi

Depresi merupakan suatu gangguan suasana hati yang kompleks, ditandai oleh perasaan sedih yang menetap, kehilangan minat atau kesenangan terhadap aktivitas sehari-hari, gangguan tidur maupun pola makan, serta penurunan fungsi sosial dan kognitif individu (Kamran, Bibi, ur. Rehman, dan Morris, 2022). Definisi ini menekankan bahwa depresi bukan sekadar merasa sedih, melainkan suatu kondisi klinis yang memengaruhi berbagai aspek kehidupan sehari-hari. Interaksi antara faktor biologis, psikologis, dan sosial memainkan peran penting dalam munculnya dan berlanjutnya kondisi ini (Kamran dkk., 2022).

WHO melaporkan bahwa lebih dari 264 juta orang di seluruh dunia menderita depresi (Luo dkk., 2025). Depresi merupakan kontributor utama terhadap beban penyakit global dan menempati peringkat teratas sebagai penyebab kecacatan (Luo dkk., 2025). Pada tahun 2022, depresi menjadi masalah kesehatan global yang signifikan, menempati peringkat keempat di antara penyakit-penyakit dunia, dengan sekitar 55% penderita mengalami pikiran untuk bunuh diri (Agustin, Prastika, Kendrasti, Fajriyah, dan Le-Quy, 2025) Prevalensi depresi yang luas ini memengaruhi individu dari berbagai demografi, dengan insiden yang lebih tinggi ditemukan pada perempuan dan remaja. Depresi memiliki kaitan yang erat dengan bunuh diri, yang menyebabkan lebih dari 700.000 kematian setiap tahun (Agustin dkk., 2025)

Berdasarkan hasil *Global Burden of Disease Study* tahun 2019, jumlah kasus gangguan depresif diperkirakan mencapai sekitar 297,6 juta jiwa, dengan tren peningkatan beban *disability-adjusted life years*(DALY) akibat depresi mayor sejak tahun 1990 hingga 2019 (Wu dkk., 2024). Peningkatan ini cenderung lebih tinggi pada kelompok usia muda dan wilayah dengan indeks pembangunan sosial-ekonomi yang tinggi. Hal ini menegaskan pentingnya penanganan depresi sebagai masalah kesehatan global, mengingat potensi kondisi ini untuk berkembang menjadi lebih parah apabila tidak ditangani dengan baik.

Meskipun kesadaran terhadap pentingnya kesehatan mental semakin meningkat, tingkat cakupan perawatan dan efektivitas intervensi terhadap depresi masih tergolong rendah. Studi terkini menunjukkan bahwa sebagian besar individu dengan *Major Depressive Disorder*(MDD) tidak memperoleh layanan pengobatan yang memadai, bahkan di negara dengan sumber daya kesehatan yang relatif baik (Santomauro dkk., 2021). Hambatan utama meliputi stigma sosial, keterbatasan

tenaga profesional di bidang kesehatan mental, serta akses layanan yang belum merata (Roberts dkk., 2022). Oleh karena itu, diperlukan pendekatan ilmiah berbasis data dan teknologi guna memperkuat upaya klasifikasi, diagnosis akurat, dan intervensi yang lebih efektif terhadap depresi pada skala global.

2.2 Depresi Pada Mahasiswa

Mahasiswa merupakan kelompok populasi yang berada pada masa transisi kritis antara remaja dan dewasa muda, di mana individu dituntut untuk beradaptasi dengan berbagai tantangan akademik, sosial, dan emosional (Wilks dkk., 2020). Pada fase ini, tekanan psikologis sering kali meningkat akibat tuntutan akademik, ketidakpastian karier, dan perubahan gaya hidup, yang apabila tidak dikelola dengan baik dapat berkembang menjadi depresi (Pascual dkk., 2020). (Alsubaie, Stain, Webster, dan Wadman, 2019) menjelaskan bahwa prevalensi depresi pada mahasiswa menunjukkan peningkatan signifikan dalam dekade terakhir, seiring meningkatnya beban akademik dan kompleksitas lingkungan sosial. Tekanan akademik, kompetisi antarindividu, serta ekspektasi terhadap prestasi yang tinggi sering kali menjadi pemicu gangguan emosional.

Studi global menunjukkan bahwa prevalensi depresi di kalangan mahasiswa cukup tinggi, sekitar 30% melaporkan gejala depresi sedang hingga berat (Wilks dkk., 2020). Di Tiongkok, *National Report on Mental Health Development*(2021–2022) mengidentifikasi mahasiswa sebagai kelompok berisiko tinggi terhadap depresi. Sebuah survei terhadap hampir 80.000 mahasiswa melaporkan tingkat prevalensi depresi sebesar 21,48% (Luo dkk., 2025)). Tekanan akademik, masalah keuangan, dan kurangnya dukungan sosial menjadi faktor risiko utama yang memicu depresi (Pascual dkk., 2020). Kondisi ini berdampak negatif pada interaksi sosial, prestasi akademik, dan kesejahteraan mental secara keseluruhan (Dong, Li, Chen, He, dan Li, 2024).

Jika tidak ditangani dengan baik, depresi dapat menyebabkan konsekuensi serius seperti *putus sekolah*, penurunan prestasi akademik, dan risiko bunuh diri(Mumenin dkk., 2024). Penelitian (Meda, Pardini, Rigobello, Visioli, dan Novara, 2023) yang melibatkan 1.388 mahasiswa menemukan bahwa hampir 20% responden mengalami gejala depresi berat hingga memiliki ide untuk bunuh diri (*suicidal ideation*), dengan faktor kekhawatiran ekonomi sebagai prediktor signifikan. Dalam konteks ini, pendekatan berbasis *machine learning* mulai dimanfaatkan untuk menganalisis faktor-faktor penyebab depresi secara objektif. Melalui pemrosesan data seperti tekanan akademik, durasi tidur, kebiasaan makan, *stres* finansial, dan tingkat kepuasan belajar, model klasifikasi seperti *Random Forest* dan *Multilay-*

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 mengujikan 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. 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 mengurangi 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.

er Perceptron-SGD mampu memetakan risiko depresi dengan tingkat akurasi bagus (Teng, Yang, dan Liu, 2024). Dengan demikian, pemahaman teoritis dan empiris tentang depresi pada mahasiswa menjadi dasar penting dalam pengembangan sistem deteksi dini berbasis data (*data-driven*) untuk mendukung kesejahteraan mental generasi muda.

2.3 Machine Learning

Machine Learning(ML) merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang memungkinkan sistem komputer untuk belajar dari data dan membuat prediksi tanpa harus diprogram secara eksplisit. Secara umum, ML bekerja dengan menganalisis pola dari data historis untuk menghasilkan model yang dapat menggeneralisasi pada data baru (Chen, Mangalathu, dan Jeon, 2022).

Dalam bidang medis dan psikologi, ML digunakan untuk memproses data kompleks seperti hasil survei, ekspresi wajah, aktivitas media sosial, atau rekam medis guna mengidentifikasi kondisi emosional dan mental seseorang (El Massari, Sabouri, Mhammedi, dan Gherabi, 2022). Keunggulan utama ML terletak pada kemampuannya menangani data besar, *multivariat*, dan *non-linear* yang sulit diinterpretasikan oleh metode statistik konvensional.

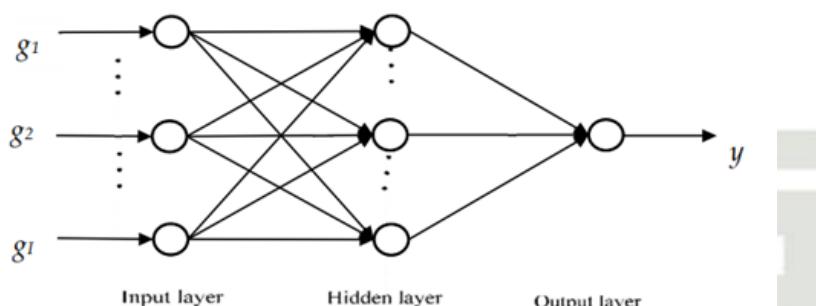
Dalam klasifikasi, ML memiliki berbagai algoritma yang digunakan untuk membedakan antara dua atau lebih kategori, seperti individu dengan dan tanpa depresi. Algoritma seperti *Random Forest*, *Support Vector Machine*(SVM), *Naïve Bayes*, dan *Multilayer Perceptron*(MLP) memiliki karakteristik berbeda dalam cara mereka memproses data dan menangkap pola. Menurut (Trofimov dkk., 2025), pemilihan algoritma yang tepat bergantung pada kompleksitas data dan tujuan analisis. Misalnya, *Random Forest* unggul dalam stabilitas dan *interpretabilitas*, sedangkan MLP lebih kuat dalam mengenali pola *non-linear* dan interaksi kompleks *antarfitur*.

Relevansi ML dalam penelitian ini terletak pada kemampuannya untuk mengidentifikasi kombinasi faktor yang berkontribusi terhadap depresi mahasiswa berdasarkan variabel seperti tekanan akademik, durasi tidur, stres keuangan, serta kepuasan belajar. Dengan membandingkan algoritma yang berbeda, peneliti dapat menentukan model yang memiliki performa terbaik dalam mengklasifikasikan kondisi depresi secara akurat. Selain itu, penggunaan metode optimasi seperti *GridSearchCV* dan algoritma pelatihan seperti *Stochastic Gradient Descent*(SGD) memastikan bahwa model yang dihasilkan bukan hanya akurat tetapi juga efisien dan *generalizable* terhadap data baru (Saha dkk., 2024).

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 meugikan 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.

2.4 Multilayer Perceptron (MLP)

Multilayer Perceptron(MLP) merupakan salah satu bentuk arsitektur jaringan saraf tiruan *Artificial Neural Network*(ANN) yang paling dasar dan umum digunakan dalam permasalahan klasifikasi maupun regresi (Al Masud dkk., 2025). Secara umum, MLP terdiri dari tiga jenis lapisan utama, yaitu lapisan input (*input layer*), lapisan tersembunyi (*hidden layer*), dan lapisan output (*output layer*) yang ditunjukkan pada Gambar Gambar 2.1.



Gambar 2.1. Lapisan Utama MLP

Sumber: (Agustin dkk., 2025)

Setiap neuron dalam lapisan-lapisan tersebut saling terhubung melalui bobot(*weights*) yang akan disesuaikan selama proses pelatihan (Abdelbasset dkk., 2022). Mekanisme kerja MLP terinspirasi dari cara kerja otak manusia dalam memproses informasi di mana neuron menerima sinyal, mengolahnya, lalu meneruskan hasilnya ke neuron berikutnya melalui fungsi aktivasi (Nosratpour, Tafakori, dan Abdollahian, 2025). Pada tahap pelatihan, MLP menggunakan algoritma *backpropagation* untuk memperbarui bobot setiap neuron berdasarkan kesalahan prediksi yang dihasilkan. Proses ini dimulai dengan perhitungan *error* antara *output* model dengan target yang sebenarnya, kemudian *error* tersebut dipropagasi mundur untuk memperbaiki bobot agar model semakin akurat (Ecer, Ardabili, Band, dan Mosavi, 2020). Penyesuaian bobot dilakukan menggunakan *optimizers*, salah satunya *Stochastic Gradient Descent*(SGD) yang bekerja dengan cara memperbarui bobot secara bertahap berdasarkan subset data (*mini-batch*). Pendekatan ini membuat proses pelatihan menjadi lebih efisien dan mencegah model terjebak pada *local minima* (Sopelsa Neto dkk., 2021).

Performa MLP sangat dipengaruhi oleh fungsi aktivasi, yang berperan dalam menentukan aliran sinyal antar neuron serta memungkinkan jaringan mempelajari hubungan *non-linear* dalam data. Fungsi aktivasi populer seperti *Sigmoid*, *Tanh*, dan *ReLU* membantu MLP menangani permasalahan kompleks yang tidak

dapat diselesaikan oleh model linear sederhana (Al Masud dkk., 2025). Semakin banyak lapisan tersembunyi yang digunakan, semakin besar pula kemampuan MLP mengenali pola data yang rumit, meskipun berisiko menimbulkan *overfitting* jika tidak dikontrol dengan baik (Ecer dkk., 2020)

Dalam konteks klasifikasi depresi pada mahasiswa, MLP digunakan karena mampu mengenali pola kompleks antara faktor psikologis, akademik, dan sosial. Persamaan *output* dari model MLP tersebut dapat dituliskan sebagai berikut 2.1 (Sumayli, 2023):

$$y = \delta_2 \left(\sum_{i=1}^m (w_i^{(2)} \delta_1 \text{hgh}(x)) + b^{(2)} \right), \quad X = \sum_{i=1}^n (X_j W_{xj}^{(1)}) + b^{(1)} \quad (2.1)$$

Dalam persamaan tersebut, y merepresentasikan vektor prediksi yang dihasilkan oleh model MLP. Nilai ini dihitung berdasarkan vektor *input* X , yang merupakan fitur dari data. Bobot yang menghubungkan *input* ke lapisan tersembunyi dilambangkan dengan $W^{(1)}$, sedangkan bobot dari lapisan tersembunyi ke lapisan *output* dilambangkan dengan $w^{(2)}$. Fungsi aktivasi pada lapisan tersembunyi adalah δ_1 , sementara fungsi aktivasi pada lapisan *output* adalah δ_2 .

Bias pada masing-masing lapisan dilambangkan dengan $b^{(1)}$ untuk lapisan tersembunyi dan $b^{(2)}$ untuk lapisan *output*. Parameter m dan n masing-masing menunjukkan jumlah sampel dan jumlah fitur dalam *dataset*. Dengan arsitektur ini, MLP dapat mempelajari hubungan *non-linear* antara *input* dan *output*. MLP mampu mempelajari representasi kompleks dari data melalui proses pelatihan, di mana bobot dan bias disesuaikan untuk meminimalkan kesalahan prediksi. Berkat kemampuannya dalam menangani data *non-linear* dan kompleks, MLP banyak digunakan dalam berbagai bidang, termasuk pengolahan data tabular, citra, dan teks (Sumayli, 2023).

2.5 Random Forest

Random Forest(RF) merupakan salah satu algoritma *ensemble learning* yang banyak digunakan dalam klasifikasi dan regresi karena kemampuannya menghasilkan model yang akurat dan stabil. Algoritma ini bekerja dengan membangun sejumlah pohon keputusan (*decision trees*) secara acak dan menggabungkan hasil prediksi masing-masing pohon untuk menentukan keputusan akhir berdasarkan prinsip *majority voting* (Mendoza, Lee, Huang, dan Sun, 2021). Pendekatan ini meningkatkan kemampuan generalisasi model serta mengurangi risiko *overfitting* yang umumnya terjadi pada pohon keputusan tunggal (Asadi, Roshan, dan Kat-

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 mengujikan 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.

tan, 2021). Dengan menerapkan proses *bagging (bootstrap aggregating)*, RF memanfaatkan *subset* data dan fitur yang berbeda untuk setiap pohon, sehingga memperkaya variasi model tanpa mengorbankan konsistensi hasil (Yuan, Liu, Feng, dan Dauphin, 2023).

Keunggulan utama RF terletak pada fleksibilitasnya dalam menangani berbagai jenis data, baik numerik maupun kategorikal, serta kemampuannya untuk bekerja dengan *dataset* berukuran besar dan kompleks (Obaido dkk., 2024). Selain itu, RF memiliki ketahanan terhadap *noise* dan data yang hilang (*missing values*), menjadikannya salah satu algoritma yang efisien dalam analisis data multidimensi (Mendoza dkk., 2021). Dalam konteks penelitian berbasis data psikologis seperti klasifikasi depresi mahasiswa, karakteristik ini sangat penting karena data yang dikumpulkan sering kali bersifat heterogen dan tidak sepenuhnya bersih. Dengan sifatnya yang tidak terlalu sensitif terhadap skala dan distribusi data, RF mampu mempertahankan performa prediksi yang tinggi tanpa memerlukan prapemrosesan (*preprocessing*) yang kompleks (Yuan dkk., 2023). Untuk menjelaskan proses pengambilan keputusan pada algoritma *Random Forest*, dapat digunakan persamaan berikut 2.2 (Aruleba dkk., 2022):

$$H(N(X)) = \arg \max_y - \sum_{k=1}^K \mathbb{1}(H_k(x) = j), \quad \text{untuk } j = 1, \dots, c \quad (2.2)$$

Persamaan tersebut menunjukkan bahwa hasil prediksi akhir \hat{C} ditentukan berdasarkan mayoritas suara (*majority voting*) dari seluruh pohon keputusan yang ada dalam hutan (*forest*). Setiap pohon h_i memberikan prediksi terhadap kelas tertentu, dan fungsi indikator I akan bernilai 1 jika pohon ke- i memprediksi kelas j serta 0 jika tidak. Kemudian, nilai-nilai ini dijumlahkan untuk semua pohon, dan kelas dengan jumlah suara terbanyak dipilih sebagai hasil akhir oleh operator $\arg \max$. Dengan demikian, penggunaan rumus ini memungkinkan *Random Forest* menggabungkan hasil prediksi dari banyak model sederhana menjadi satu keputusan yang lebih akurat dan stabil, sekaligus mengurangi risiko *overfitting* yang sering terjadi pada model pohon tunggal ((Aruleba dkk., 2022); (Yuan dkk., 2023)).

Selain itu, RF memiliki kemampuan untuk memberikan estimasi *feature importance*, yaitu nilai yang menunjukkan sejauh mana suatu fitur berkontribusi terhadap hasil klasifikasi. Mekanisme ini memungkinkan peneliti untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap prediksi, sehingga memberikan interpretasi yang lebih baik terhadap hubungan antar variabel (Yuan dkk., 2023). Dalam studi kesehatan mental, seperti klasifikasi depresi, *feature impor-*

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 mengujikan 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.

tance dapat membantu mengungkap variabel psikososial atau akademik yang paling berperan dalam menentukan tingkat kerentanan seseorang terhadap depresi (Aruleba dkk., 2022). Hal ini menjadikan RF tidak hanya sebagai alat prediksi, tetapi juga sebagai sarana eksplorasi pengetahuan yang bermakna secara praktis.

Secara metodologis, RF terbukti efektif dalam menghasilkan hasil yang konsisten melalui validasi silang (*cross-validation*) dan optimasi *hyperparameter*. Parameter penting yang perlu diatur mencakup jumlah pohon (*n_estimators*), kedalam maksimum pohon (*max_depth*), dan jumlah fitur yang digunakan pada setiap pembagian cabang (*max_features*). Optimasi parameter ini, seperti melalui teknik *GridSearchCV*, mampu meningkatkan performa model secara signifikan dengan menemukan kombinasi terbaik yang memberikan keseimbangan antara akurasi dan generalisasi (Wallace dkk., 2023).

Secara keseluruhan, RF menawarkan keseimbangan yang baik antara akurasi, efisiensi, dan interpretabilitas. Berbeda dengan model *black box* seperti *deep neural networks*, RF masih memungkinkan analisis mendalam terhadap struktur keputusan dan kontribusi tiap fitur tanpa kehilangan ketepatan prediksi. Oleh karena itu, pemilihan RF dalam penelitian ini tidak hanya didasarkan pada performa empirisnya, tetapi juga pada kemampuannya untuk memberikan pemahaman konseptual yang lebih jelas terhadap fenomena depresi mahasiswa melalui pendekatan berbasis data ((Mendoza dkk., 2021);(Yuan dkk., 2023)).

2.6 Stochastic Gradient Descent(SGD)

Stochastic Gradient Descent(SGD) merupakan salah satu algoritma optimasi paling fundamental dan banyak digunakan dalam proses pelatihan model, khususnya pada jaringan saraf tiruan seperti *Multilayer Perceptron* (MLP) (Alharbi dan Khan, 2025). Secara umum, tujuan utama algoritma optimasi adalah untuk meminimalkan fungsi kehilangan (*loss function*) dengan cara menyesuaikan parameter model agar kesalahan prediksi menjadi sekecil mungkin. Dalam metode *gradient descent* konvensional, proses pembaruan bobot dilakukan berdasarkan gradien dari keseluruhan *dataset*, yang meskipun akurat, membutuhkan waktu komputasi yang besar dan memori yang tinggi. Untuk mengatasi keterbatasan tersebut, dikembangkanlah pendekatan *stochastic*, di mana pembaruan bobot dilakukan berdasarkan satu atau beberapa sampel secara acak pada setiap iterasi (Jentzen dan Rieker, 2022).

Pendekatan ini memberikan keunggulan utama dalam hal efisiensi dan kemampuan keluar dari perangkap *local minima*. Karena setiap pembaruan bobot dilakukan menggunakan *subset* kecil data, arah gradien yang dihasilkan bersifat

lebih bervariasi dan dinamis. Variasi ini membantu model untuk tidak mudah terjebak pada solusi lokal dan mendorong pencarian solusi global yang lebih optimal. Selain itu, pembaruan yang dilakukan secara acak juga mempercepat konvergensi, terutama ketika berhadapan dengan *dataset* berskala besar (Tian, Zhang, dan Zhang, 2023). Dengan demikian, SGD menjadi salah satu teknik yang sangat efektif dalam mengoptimalkan model dengan jumlah parameter yang kompleks dan *non-linear*, seperti yang dimiliki oleh MLP (Fatima dkk., 2020).

Dalam praktiknya, efektivitas SGD sangat dipengaruhi oleh nilai *learning rate*, yaitu besarnya langkah pembaruan bobot pada setiap iterasi. Nilai yang terlalu besar dapat menyebabkan model melompat-lompat dan gagal mencapai titik optimal, sedangkan nilai yang terlalu kecil akan membuat proses pelatihan lambat dan berpotensi berhenti sebelum mencapai konvergensi. Untuk mengatasi hal ini, dikembangkan berbagai varian SGD seperti *SGD with Momentum*, *Nesterov Accelerated Gradient*, dan *Adaptive Gradient Methods* yang menyesuaikan langkah pembaruan bobot secara adaptif berdasarkan sejarah gradien. Pemilihan varian yang tepat sangat bergantung pada karakteristik data dan kompleksitas model yang digunakan (Smith, Dherin, Barrett, dan De, 2021).

Dalam konteks penelitian klasifikasi depresi pada mahasiswa, penggunaan SGD sebagai algoritma optimasi pada MLP bertujuan untuk meningkatkan efisiensi proses pelatihan dan memastikan bahwa model mampu mempelajari pola kompleks antar variabel secara efektif. Dengan data yang mencakup berbagai aspek seperti tekanan akademik, stres finansial, dan kepuasan belajar, SGD membantu model untuk menemukan kombinasi bobot yang paling representatif terhadap pola depresi yang muncul dalam populasi mahasiswa. Selain itu, optimasi berbasis SGD juga memungkinkan model untuk mencapai keseimbangan antara akurasi dan kemampuan generalisasi, sehingga hasil klasifikasi tidak hanya akurat pada data pelatihan, tetapi juga andal saat diuji pada data baru (*unseen data*).

2.7 GridSearchCV

Dalam pembelajaran mesin, optimasi *hyperparameter* merupakan tahap penting yang bertujuan untuk menemukan kombinasi parameter terbaik agar model menghasilkan performa yang optimal ((Saha dkk., 2024); (Bischl dkk., 2023)). *Hyperparameter* sendiri adalah parameter yang tidak dipelajari secara langsung dari data selama proses pelatihan, melainkan ditentukan sebelum pelatihan dimulai. Contohnya termasuk jumlah neuron pada lapisan tersembunyi (*hidden layer*), *learning rate*, jumlah *estimator* pada *Random Forest*, atau fungsi aktivasi pada jaringan saraf. Pemilihan nilai *hyperparameter* yang tepat sangat berpengaruh ter-

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 mengurangi 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. 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 mengutip 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.

hadap kinerja model baik dari segi akurasi, kemampuan generalisasi, maupun waktu komputasi (Candelieri, Ponti, dan Archetti, 2024). Model dengan konfigurasi *hyperparameter* yang tidak sesuai dapat mengalami *overfitting* (terlalu menyesuaikan diri pada data pelatihan) atau *underfitting* (gagal menangkap pola penting dari data) (Tetko, van Deursen, dan Godin, 2024).

Untuk memperoleh konfigurasi *hyperparameter* terbaik, berbagai metode optimasi telah dikembangkan. Dua pendekatan yang paling umum digunakan adalah *Grid Search* dan *Random Search*. *Grid Search* bekerja dengan mengevaluasi model pada seluruh kombinasi parameter yang mungkin dalam ruang pencarian tertentu, sehingga menghasilkan hasil yang akurat namun memerlukan waktu komputasi yang besar. Sebaliknya, *Random Search* memilih kombinasi parameter secara acak dalam ruang pencarian, sehingga lebih efisien dan sering kali mampu menemukan kombinasi optimal dengan jumlah iterasi yang lebih sedikit (Fajri dan Primajaya, 2023). Selain dua metode tersebut, pendekatan modern seperti *Bayesian Optimization* dan *Genetic Algorithm* juga mulai digunakan karena kemampuannya memperkirakan area pencarian yang paling menjanjikan berdasarkan hasil sebelumnya (Zulfiqar, Gamage, Kamran, dan Rasheed, 2022).

Dalam konteks model jaringan saraf seperti *Multilayer Perceptron*(MLP), optimasi *hyperparameter* menjadi sangat krusial karena struktur model yang kompleks dan jumlah parameter yang besar (Zhang, Wang, Wang, Chen, dan Wang, 2023). Nilai *hyperparameter* seperti jumlah lapisan tersembunyi, jumlah neuron per lapisan, *learning rate*, *batch size*, dan fungsi aktivasi sangat memengaruhi stabilitas proses pelatihan serta kualitas hasil klasifikasi. Pengaturan yang tidak tepat dapat menyebabkan konvergensi lambat, performa buruk, atau model yang tidak stabil. Oleh karena itu, kombinasi metode optimasi seperti *Grid Search* yang dikombinasikan dengan *Cross-Validation* sering digunakan untuk menjamin hasil yang tidak hanya optimal terhadap data pelatihan, tetapi juga konsisten terhadap data validasi.

Dalam penelitian klasifikasi depresi pada mahasiswa, optimasi *hyperparameter* berperan penting dalam meningkatkan performa model baik pada algoritma *Random Forest* maupun MLP-SGD. Melalui optimasi ini, model dapat mencapai keseimbangan antara kompleksitas dan kemampuan generalisasi, sehingga hasil prediksi menjadi lebih akurat dan stabil. Misalnya, pada MLP-SGD, *tuning* dilakukan terhadap parameter seperti *learning rate* dan jumlah neuron, sementara pada *Random Forest* parameter yang dioptimasi meliputi jumlah pohon (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*). Dengan pendekatan ini, model tidak hanya menghasilkan nilai akurasi tinggi, tetapi juga mampu mengenali pola depresi mahasiswa dari berbagai faktor akademik, sosial, dan psikologis (Saha dkk., 2024).

2.8 Data Sekunder

Data sekunder dalam penelitian *machine learning* merupakan data yang telah tersedia sebelumnya dan dikumpulkan oleh orang lain untuk tujuan tertentu, kemudian digunakan kembali oleh peneliti sebagai sumber data penelitian (Koch, Denton, Hanna, dan Foster, 2021). Penggunaan data sekunder sangat umum dalam penelitian pengajaran mesin karena memungkinkan pengumpulan *dataset* berskala besar yang sulit dilakukan secara mandiri (Koesten, Vougiouklis, Simperl, dan Groth, 2020). Dalam penelitian akademik, *dataset* publik seperti *Kaggle* sangat sering digunakan karena memberikan data yang terstruktur, terdokumentasi, dan dapat direplikasi oleh peneliti lain.

Data sekunder yang dikumpulkan melalui *Kaggle* digunakan untuk membangun model klasifikasi depresi mahasiswa. Algoritma pembelajaran yang diawasi (*supervised learning*), seperti *Random Forest* dan *Multilayer Perceptron*(MLP), dapat mempelajari pola hubungan yang kompleks dan *non-linear* dengan lebih stabil saat menggunakan *dataset* dengan jumlah sampel yang besar. Selain itu, *dataset* berskala besar membantu mengurangi kemungkinan *overfitting*. *Overfitting* adalah situasi ketika model terlalu menyesuaikan diri terhadap data pelatihan sehingga tidak dapat melakukan generalisasi pada data baru (Bradshaw, Huemann, Hu, dan Rahmim, 2023). Data sekunder divalidasi menggunakan metode *cross-validation k-fold*, khususnya *5-fold cross-validation*. *Cross-validation* adalah metode evaluasi internal yang membagi *dataset* menjadi beberapa *subset*, dengan setiap *subset* berfungsi secara bergantian sebagai data validasi dan sisanya digunakan sebagai data pelatihan. Dibandingkan dengan pembagian data tunggal (*hold-out*), metode ini memberikan estimasi performa model yang lebih objektif dan stabil (Teodorescu dan Obreja Brașoveanu, 2025).

2.9 Data Primer

Data primer adalah data yang dikumpulkan secara langsung dari responden sesuai dengan tujuan dan konteks penelitian. Berbeda dengan data sekunder, data primer memungkinkan peneliti mendapatkan data yang lebih relevan dan kontekstual dengan populasi sasaran penelitian (Ho, Phua, Wong, dan Goh, 2020). Dalam penelitian ini, kuesioner *Patient Health Questionnaire-9* (PHQ-9) diberikan kepada mahasiswa sebagai representasi dari populasi sasaran.

Data primer ini digunakan sebagai data uji eksternal (*external test data*). Metode ini bertujuan untuk menilai kapasitas model pembelajaran mesin yang telah dilatih menggunakan data sekunder ketika diterapkan pada data independen dari berbagai distribusi (Ho dkk., 2020). Proses evaluasi model menggunakan *dataset*

- 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 mengurangi 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. 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 mengurangi 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.

yang tidak terlibat sama sekali dalam tahap pelatihan atau validasi internal dikenal sebagai validasi luar (*external validation*) (Verma, Bach, dan Mork, 2023). Dalam penelitian *machine learning*, khususnya di bidang kesehatan mental, validasi eksternal sangat penting karena dapat menunjukkan seberapa baik model dapat menggeneralisasi pada data dunia nyata (He dkk., 2024). Studi terbaru menekankan bahwa model yang sangat baik pada validasi internal belum tentu berhasil ketika diuji pada data eksternal, sehingga pengujian menggunakan data primer menjadi langkah evaluasi yang sangat penting (He dkk., 2024). Karena itu, peran data primer bersifat evaluatif dan strategis dalam menilai keandalan model, meskipun jumlah data sekunder lebih kecil.

2.10 *Patient Health Questionnaire-9 (PHQ-9)*

PHQ-9 merupakan salah satu instrumen skrining depresi yang paling banyak digunakan dalam praktik klinis dan penelitian kesehatan mental karena mudah digunakan dan mengukur depresi dengan akurat (Kroenke, Spitzer, dan Williams, 2001); (Levis dkk., 2024). Instrumen PHQ-9 dikembangkan oleh Kroenke, Spitzer, dan Williams, serta telah divalidasi secara luas pada berbagai populasi, termasuk mahasiswa dan kelompok usia dewasa muda (Gao dan Liu, 2024). Kriteria yang ditemukan dalam Diagnostic and Statistical Manual of Mental Disorders (DSM) digunakan oleh PHQ-9 untuk mengukur intensitas gejala depresi. Oleh karena itu, hasil pengukurannya relevan secara akademik dan klinis (Levis dkk., 2024).

PHQ-9 terdiri dari sembilan pertanyaan yang masing-masing menunjukkan gejala depresi utama, termasuk suasana hati depresif, kehilangan minat, masalah tidur, kelelahan, gangguan konsentrasi, perubahan nafsu makan, dan pikiran untuk menyakiti diri sendiri (Alves, Marci, Cohen-Stavi, Whelan, dan Boussios, 2025). PHQ-9 ini menangkap kondisi depresi terkini dengan meminta responden menilai frekuensi gejala dalam dua minggu terakhir (Alves dkk., 2025).

Dalam penelitian ini, skor PHQ-9 digunakan sebagai label kelas (*target variable*) untuk proses klasifikasi. Kategori PHQ-9 diubah menjadi kelas biner (*binary classification*), yaitu “tidak depresi” dan “depresi”, untuk menyederhanakan proses pelatihan dan evaluasi model (Rolnick dkk., 2020). Pendekatan biner ini umum digunakan dalam penelitian berbasis ML karena mampu meningkatkan stabilitas model dan mengurangi kompleksitas dibandingkan klasifikasi *multi-class* (Raith, Kamp, Stoiber, Jakl, dan Wagner, 2022). PHQ-9 dipilih karena validitasnya yang tinggi serta kemudahannya untuk diintegrasikan ke dalam pemodelan klasifikasi berbasis data numerik terstruktur (Rolnick dkk., 2020).

2.11 Confusion Matrix

Dalam penelitian berbasis *machine learning*, evaluasi performa model menjadi aspek penting untuk memastikan bahwa model yang dikembangkan tidak hanya menghasilkan prediksi, tetapi juga memiliki tingkat keakuratan yang dapat dianalisa (Moulaei, Shanbehzadeh, Mohammadi-Taghiabad, dan Kazemi-Arpanahi, 2022). Salah satu metode evaluasi yang paling umum digunakan dalam klasifikasi adalah *Confusion Matrix* (Pratama dkk., 2023). *Confusion Matrix* merupakan tabel dua dimensi yang menggambarkan hubungan antara nilai aktual (*true class*) dan nilai prediksi (*predicted class*) dari suatu model klasifikasi (Moulaei dkk., 2022). *Confusion Matrix* terdiri dari empat komponen utama, yaitu *True Positive*(TP), *True Negative*(TN), *False Positive*(FP), dan *False Negative*(FN). Keempat komponen ini menjadi dasar untuk menghitung berbagai metrik evaluasi model seperti akurasi, presisi, *recall*, spesifisitas, dan *F1-score* yang ditunjukkan pada gambar Gambar 2.2 (Pratama dkk., 2023).

		ACTUAL VALUE	
		POSITIVE	NEGATIVE
PREDICTED VALUE	NEGATIVE	TP	FP
	POSITIVE	FN	TN

Gambar 2.2. Diagram *Confusion Matrix*

Sumber: (Dwinnie dkk., 2023)

Masing-masing metrik tersebut memiliki fungsi yang berbeda dalam menilai kualitas model. Akurasi mengukur proporsi prediksi yang benar dari keseluruhan data dan dihitung dengan rumus 2.3:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3)$$

Sementara *Recall* mengukur kemampuan model dalam mendeteksi seluruh data positif 2.4:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.4)$$



Presisi mengukur ketepatan model dalam memprediksi kelas positif 2.5:

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2.5)$$

F1-score kemudian digunakan untuk memberikan gambaran seimbang antara presisi dan *recall* melalui rata-rata harmonis dari keduanya 2.6:

$$F1\text{-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (2.6)$$

Dalam konteks penelitian ini, penggunaan *Confusion Matrix* menjadi sangat penting untuk menilai sejauh mana model *Random Forest* dan *MLP-SGD* mampu mengklasifikasikan mahasiswa yang mengalami depresi dan yang tidak. Misalnya, nilai *recall* yang tinggi pada model *MLP-SGD* akan menunjukkan bahwa model tersebut efektif dalam mengenali mahasiswa dengan gejala depresi(minim FN), semintara presisi tinggi pada *Random Forest* dapat menunjukkan bahwa model mampu meminimalkan kesalahan dalam mengklasifikasikan mahasiswa non-depresi sebagai depresi(minim FP). Dengan membandingkan nilai-nilai ini, dapat disimpulkan model mana yang memiliki kinerja lebih baik secara keseluruhan dalam klasifikasi depresi.

Selain itu, *Confusion Matrix* juga berperan dalam memahami bias model terhadap salah satu kelas, yang sering terjadi pada data tidak seimbang. Dalam penelitian ini, analisis *Confusion Matrix* digunakan sebagai dasar untuk mengevaluasi dan menginterpretasikan hasil model, bukan sekadar melihat nilai akurasi secara global. Dengan demikian, pemanfaatan *Confusion Matrix* memberikan gambaran yang lebih komprehensif dan objektif mengenai performa model klasifikasi depresi, memastikan hasil penelitian memiliki validitas empiris yang kuat(Helmut, Fitriyani, Romadiana, dkk., 2024).

2.12 *Flask*

Flask merupakan *micro web framework* berbasis bahasa pemrograman Python yang dirancang untuk pengembangan aplikasi web secara sederhana, fleksibel, dan ringan, sehingga memungkinkan pengembang membangun aplikasi tanpa ketergantungan konfigurasi yang kompleks (Suherlan, Arti, Nabilah, dan Hazimah, 2025). Dalam konteks pengembangan sistem berbasis *machine learning*, *Flask* banyak digunakan karena kemampuannya untuk mengintegrasikan model prediksi ke dalam aplikasi web atau *Application Programming Interface*(API) secara efisien, sehingga hasil pemodelan dapat diakses oleh pengguna akhir melalui antarmuka berbasis web (Narayanan, Balamurugan, Palas, dkk., 2022) Gambar 2.3 merupakan tampilan dari logo *Flask*.

- 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 mengujikan 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.



Gambar 2.3. Logo *Flask*

Sumber:(Suherlan dkk., 2025)

Flask sering digunakan dalam penelitian *machine learning* pada tahap *deployment* model, yaitu proses menerapkan model yang telah dilatih agar dapat digunakan secara efektif dalam lingkungan nyata (Suherlan dkk., 2025). Tahap *deployment* ini sangat penting karena memungkinkan sistem klasifikasi depresi untuk digunakan sebagai alat bantu deteksi dini berbasis web yang dapat diakses secara langsung oleh pengguna dan tidak hanya berhenti pada tahap pengujian model.

Flask menggunakan konsep *routing* dalam arsitekturnya, yang berarti bahwa fungsi tertentu di sisi server menangani setiap permintaan (*request*) pengguna untuk menghasilkan tanggapan yang sesuai (Suherlan dkk., 2025). Model *machine learning* seperti *Random Forest* dan *Multilayer Perceptron with Stochastic Gradient Descent Optimization*(MLP-SGD) disimpan dalam bentuk berkas menggunakan pustaka seperti *joblib* atau *pickle*. Selanjutnya, model tersebut dimuat kembali ke dalam aplikasi *Flask* untuk melakukan proses prediksi berdasarkan data input dari pengguna (Dani, Bhople, Waghmare, Munginwar, dan Patil, 2022).

Dengan menggunakan *Flask*, penelitian ini tidak hanya menghasilkan analisis model secara teoretis, tetapi juga memungkinkan penerapan hasil penelitian dalam bentuk sistem yang fungsional. Beberapa keuntungan penggunaan *Flask* antara lain sifatnya yang ringan dan fleksibel, sehingga sesuai untuk pengembangan prototipe sistem klasifikasi depresi. Selain itu, *Flask* mendukung pengembangan sistem *end-to-end*, mulai dari input data, proses prediksi, hingga penyajian hasil kepada pengguna (Narayanan dkk., 2022)). Dengan implementasi berbasis *Flask*, penelitian ini tidak hanya berfokus pada perbandingan kinerja algoritma, tetapi juga membahas bagaimana hasil pemodelan dapat diterapkan dalam sistem klasifikasi depresi mahasiswa yang dapat digunakan secara nyata. Pendekatan ini sejalan dengan tren penelitian terkini yang menekankan pentingnya sistem *machine learning* yang aplikatif dalam bidang kesehatan mental (Mulugeta, Zewotir, Tegegne, Juhar, dan Muleta, 2023).

- 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 mengujikan 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.

2.13 Penelitian Terdahulu

Beberapa penelitian terdahulu telah membahas penerapan algoritma *machine learning* seperti *Random Forest*(RF) dan *Multilayer Perceptron*(MLP) dalam klasifikasi depresi pada mahasiswa.

Penelitian yang dilakukan oleh Mohamed et al. dengan judul “*Hybrid Machine Learning Models for Anxiety Detection in University Students*” mengembangkan model hibrida dengan mengombinasikan algoritma *Multilayer Perceptron*(MLP) dan *Random Forest*(RF) untuk klasifikasi tingkat kecemasan mahasiswa di Kashmir. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* mampu mencapai akurasi sebesar 98,14% dengan nilai *Kappa* yang tinggi, bahkan pada kelas minoritas. Studi ini menekankan bahwa kombinasi algoritma serta pemilihan parameter yang tepat dapat menghasilkan performa klasifikasi yang stabil (Mohamed dkk., 2023).

Penelitian yang dilakukan oleh Mumenin et al. dengan judul “*DDNet: A Deep Learning-based Hybrid Model for Depression Detection*” mengusulkan arsitektur DDNet, yaitu *stacked ensemble* berbasis *Multilayer Perceptron*(MLP), *Stochastic Gradient Descent*(SGD), dan *CatBoost*. Penelitian ini menggunakan dataset depresi mahasiswa dan menekankan pentingnya proses optimisasi parameter. Hasil akhir menunjukkan bahwa model DDNet berhasil mencapai akurasi sebesar 99,16%, yang membuktikan bahwa kombinasi arsitektur berbasis MLP-SGD dengan optimisasi mampu meningkatkan performa klasifikasi depresi secara signifikan (Mumenin dkk., 2024).

Penelitian yang dilakukan oleh Saha et al. dengan judul “*Ensemble of hybrid model based technique for early detecting of depression based on SVM and neural networks*” membandingkan algoritma RF dan MLP pada data berbasis faktor sosiodemografis. Penelitian ini menerapkan teknik optimisasi *hyperparameter* menggunakan *GridSearchCV* untuk memperoleh konfigurasi model terbaik. Hasil penelitian menunjukkan bahwa algoritma RF mencapai akurasi sebesar 94,67%, sedangkan MLP mencapai 97,33%. Temuan ini menunjukkan bahwa proses *hyperparameter tuning* yang sistematis berperan penting dalam meningkatkan performa algoritma klasifikasi (Saha dkk., 2024).

Penelitian yang dilakukan oleh Masud et al. dengan judul “*Effective depression detection and interpretation: Integrating machine learning, deep learning, language models, and explainable AI*” menggunakan berbagai algoritma, termasuk *Random Forest*(RF) dan *Multilayer Perceptron*(MLP), untuk klasifikasi depresi pada mahasiswa di Bangladesh. Penelitian ini juga menerapkan *GridSearchCV* untuk optimasi *hyperparameter* dan terbukti mampu meningkatkan performa model

secara signifikan. Setelah dilakukan proses *tuning*, algoritma *Random Forest* mencapai akurasi sebesar 91,1% dengan nilai *F1-score* sebesar 91,6%, sedangkan MLP memperoleh akurasi sebesar 81,8% dengan *F1-score* sebesar 87,8%. Hasil ini menunjukkan bahwa optimasi *hyperparameter* berperan penting dalam meningkatkan kemampuan model mendeteksi depresi secara akurat pada data sosial dan akademik mahasiswa (Al Masud dkk., 2025).

Penelitian yang dilakukan oleh Elujide et al. dengan judul “*Informatics in Medicine Unlocked: Application of Deep and Machine Learning Techniques for Multi-label Classification Performance on Psychotic Disorder Diseases*” menggunakan algoritma RF dan MLP untuk klasifikasi gangguan psikotik. Namun, hasil yang diperoleh relatif rendah, yaitu sebesar 64,07% untuk RF dan 58,44% untuk MLP. Temuan ini mengindikasikan bahwa tanpa konfigurasi *hyperparameter* yang optimal, performa algoritma dapat menurun secara signifikan, terutama pada dataset dengan kompleksitas *multi-label* yang tinggi (Elujide dkk., 2021)

Penelitian yang dilakukan oleh Zhao et al. dengan judul “*A Machine-learning-derived Online Prediction Model for Depression Risk in COPD Patients: A Retrospective Cohort Study from CHARLS*” menerapkan berbagai algoritma *machine learning*, termasuk RF dan MLP, pada data kesehatan dan ekspresi genetik untuk mendeteksi risiko depresi. Hasil penelitian menunjukkan bahwa akurasi yang diperoleh relatif moderat, yaitu sebesar 69,56% untuk RF dan 69,24% untuk MLP, meskipun telah dilakukan tahap pra-pemrosesan data. Penelitian ini menegaskan bahwa kualitas data serta strategi pemilihan parameter memiliki pengaruh yang signifikan terhadap hasil klasifikasi depresi (Zhao, Fang, Cui, dan Bai, 2020).

Penelitian berjudul “*A Machine Learning Model Using Clinical Notes to Estimate PHQ-9 Symptom Severity Scores in Depressed Patients*” mengembangkan sebuah model *machine learning* untuk mengestimasi tingkat keparahan depresi berdasarkan skor *Patient Health Questionnaire-9*(*PHQ-9*) dengan memanfaatkan catatan klinis sebagai data sekunder. Instrumen *PHQ-9* digunakan sebagai *ground truth*, sementara model dilatih untuk mengekstraksi pola linguistik serta informasi klinis dari *electronic health records* yang berkorelasi dengan gejala depresi. Hasil penelitian menunjukkan bahwa pendekatan ini mampu menghasilkan estimasi skor *PHQ-9* dengan kinerja yang baik, sehingga data sekunder berbasis catatan klinis berpotensi digunakan sebagai pelengkap data primer kuesioner dalam pengembangan sistem pendukung keputusan kesehatan mental (Alves dkk., 2025)

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 mengurangi 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.

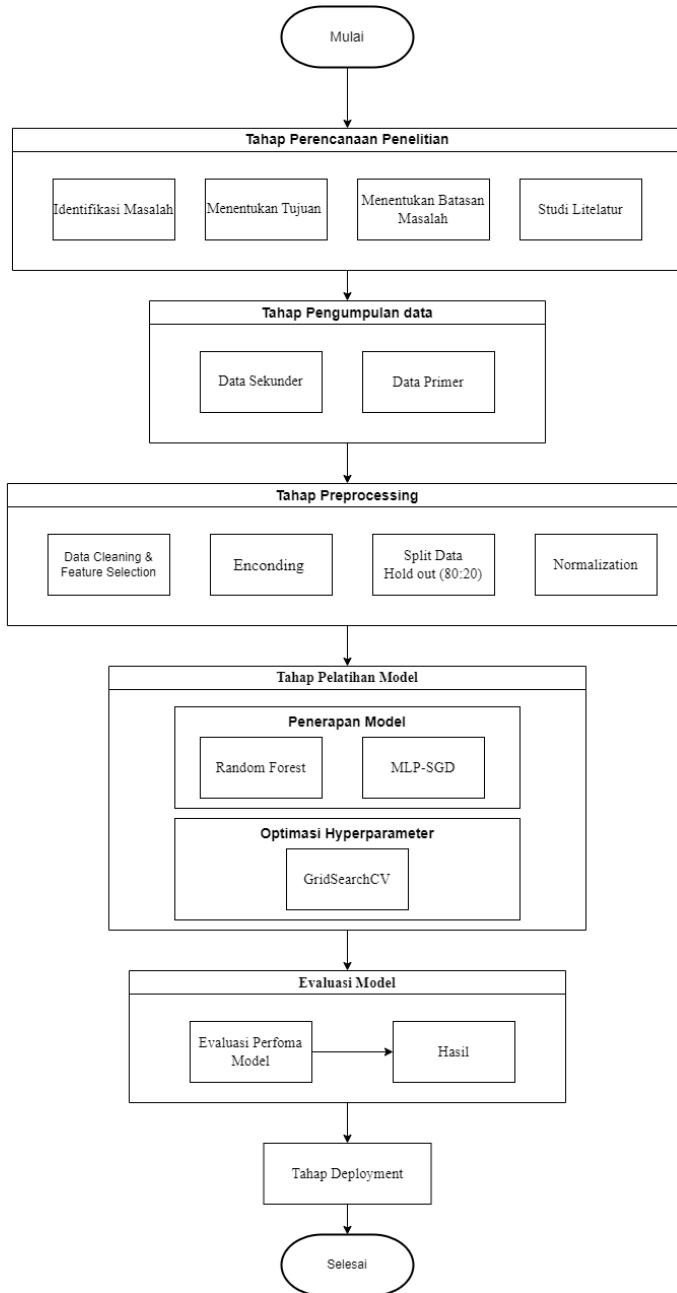
Hak Cipta Dilindungi Undang-Undang

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 3

METODOLOGI PENELITIAN

Langkah-langkah yang dilakukan dalam penelitian Tugas Akhir ini dapat dilihat pada Gambar Gambar 3.1.



Gambar 3.1. Flowchart Tahapan Penelitian

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 mengurangi 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 Tahap Perencanaan Penelitian

Tahap perencanaan penelitian ini disusun secara sistematis untuk membangun dasar yang kuat dalam studi komparatif antara algoritma *Random Forest*(RF) dan *Multilayer Perceptron* dengan optimasi *Stochastic Gradient Descent*(MLP-SGD) pada klasifikasi depresi mahasiswa. Tahapan ini mencakup proses identifikasi masalah, menentukan tujuan, menentukan batasan masalah, serta studi literatur. Seluruh tahapan tersebut digambarkan dalam *flowchart* penelitian yang berfungsi untuk memastikan arah penelitian tetap fokus, relevan, dan terukur. Pendekatan yang digunakan menekankan pada peningkatan akurasi model melalui optimasi *hyperparameter* menggunakan *GridSearchCV*, sehingga hasil klasifikasi yang diperoleh memiliki performa terbaik.

3.1.1 Identifikasi Masalah

Prevalensi depresi di kalangan mahasiswa cukup tinggi, mencapai 30,6% secara global dan 22,4% di Indonesia. Metode tradisional seperti wawancara klinis dinilai subjektif dan kurang efisien untuk populasi besar. Pendekatan *machine learning*, seperti *Random Forest* yang tahan terhadap *overfitting* dan MLP-SGD yang mampu menangkap pola *non-linear* kompleks. Namun, penelitian yang membandingkan kedua algoritma ini dengan optimasi *hyperparameter* masih terbatas.

3.1.2 Menentukan Tujuan

Penelitian ini bertujuan membandingkan performa *Random Forest* dan MLP-SGD dalam klasifikasi depresi mahasiswa menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, setelah optimasi *GridSearchCV*. *Hyperparameter* yang dioptimasi mencakup *learning rate* dan jumlah neuron tersembunyi untuk MLP-SGD, serta *n_estimators* dan *max_depth* untuk *Random Forest*. Diharapkan model yang dihasilkan akurat, efisien, dan memiliki kemampuan generalisasi yang baik.

3.1.3 Menentukan Batasan Masalah

Batasan masalah ditetapkan untuk memastikan bahwa penelitian tetap berada dalam ruang lingkup yang telah ditentukan, sehingga fokus kajian tidak melebar dan tujuan penelitian dapat dicapai secara optimal. Penelitian ini difokuskan pada permasalahan klasifikasi biner, yaitu depresi dan non-depresi, dengan memanfaatkan *Student Depression Dataset* dari Kaggle sebagai data sekunder serta data primer yang diperoleh melalui kuesioner *Patient Health Questionnaire-9*(PHQ-9).

Proses optimasi model dilakukan menggunakan metode *GridSearchCV* dengan pendekatan *5-fold cross-validation* untuk memperoleh konfigurasi *hyperpara-*

parameter yang optimal. Evaluasi kinerja model didasarkan pada metrik yang diturunkan dari *confusion matrix*, meliputi *accuracy*, *precision*, *recall*, dan *F1-score*, guna memberikan gambaran performa model secara komprehensif. Implementasi hasil pemodelan ke dalam sistem nyata secara sederhana, tanpa mencakup pengembangan sistem berskala besar atau implementasi lanjutan yang kompleks.

3.1.4 Studi Pustaka

Tahap ini mencakup penelaahan berbagai penelitian terdahulu terkait penerapan *machine learning* dalam klasifikasi depresi. Sumber studi berasal dari literatur relevan, baik jurnal nasional maupun internasional, yang membahas penerapan *Random Forest* dan *MLP-SGD* dalam klasifikasi depresi pada mahasiswa.

3.2 Tahap Pengumpulan Data

Tahap pengumpulan data dalam penelitian ini dilakukan melalui dua sumber, yaitu data sekunder dan data primer. Data sekunder diperoleh dari platform Kaggle berupa *Student Depression Dataset* yang digunakan sebagai data pelatihan (*training data*) dalam pemodelan dan evaluasi klasifikasi depresi mahasiswa menggunakan algoritma *Random Forest* dan *Multilayer Perceptron-Stochastic Gradient Descent*(*MLP-SGD*) dengan optimasi *hyperparameter GridSearchCV*. Dataset ini dipilih karena memiliki jumlah data yang besar, atribut yang relevan dengan aspek demografis, akademik, dan psikososial, serta telah digunakan dalam penelitian sebelumnya.

Selain itu, data primer dikumpulkan melalui penyebaran kuesioner *Patient Health Questionnaire-9*(PHQ-9) kepada mahasiswa aktif Universitas Islam Negeri Sultan Syarif Kasim Riau dan digunakan sebagai data uji eksternal (*external testing data*) untuk menilai kemampuan generalisasi model. Seluruh proses pengumpulan data dilakukan dengan memperhatikan prinsip etika penelitian, kerahasiaan identitas responden, serta validasi profesional, sehingga data yang diperoleh layak dan dapat dipertanggungjawabkan secara ilmiah dan klinis.

3.2.1 Data Sekunder

Dataset diambil dari platform *Kaggle* melalui *Student Depression Dataset* (www.kaggle.com/datasets/adilshamim8/student-depression-dataset). *Dataset* ini berisi 27.901 entri data mahasiswa dengan 18 atribut yang meliputi usia, tekanan akademik, kepuasan belajar, kebiasaan tidur, dukungan sosial, dan depresi.

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 meugikan 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.

Tabel 3.1. Karakteristik Data

No	Nama Atribut	Keterangan	Nilai / Contoh
1	Gender	Jenis kelamin	Laki-laki, Perempuan
2	Age	Usia responden	21, 23, 25, dll
3	City	Kota tempat tinggal	Delhi, Mumbai, Bangalore, dll
4	Profession	Status responden	Mahasiswa, Pekerja Profesional
5	Academic Pressure	Tekanan akademik	1.0, 2.0, 3.0, 4.0, 5.0
6	Work Pressure	Tekanan kerja	1.0, 2.0, 3.0, 4.0, 5.0
7	CGPA	Indeks prestasi kumulatif	3.32, 3.45, dll
8	Study Satisfaction	Kepuasan belajar	1.0, 2.0, 3.0, 4.0, 5.0
9	Job Satisfaction	Kepuasan kerja	1.0, 2.0, 3.0, 4.0, 5.0
10	Sleep Duration	Durasi tidur	5–6 Jam, ; 5 Jam, 7–8 Jam, ; 8 Jam
11	Dietary Habits	Pola makan	Sehat, Seimbang, Tidak Sehat
12	Degree	Jenjang pendidikan	SMA, S1, S2, S3
13	Suicidal	Pikiran bunuh diri	Tidak, Ya
14	Thoughts		
15	Work/Study Hours	Durasi kerja/belajar	3.0, 9.0, 10.0, dll
16	Financial Stress	Tekanan finansial	1.0, 2.0, 3.0, 4.0, 5.0
17	Family History	Riwayat penyakit mental	Tidak, Ya
18	Depression	Variabel target	1 = Depresi, 0 = Tidak
	Id	Identifier unik	1, 2, 3, dst

Tabel Tabel 3.1 menunjukkan atribut pada *dataset* yang mencakup aspek demografis, akademik, psikososial, dan perilaku kesehatan. Seluruh data telah dianonimkan untuk menjaga kerahasiaan identitas responden. *Dataset* tersebut dinilai representatif karena mencerminkan karakteristik populasi mahasiswa secara umum.

Dataset ini juga telah digunakan oleh beberapa peneliti sebelumnya dalam studi yang berfokus pada klasifikasi depresi mahasiswa menggunakan berbagai algoritma *machine learning* (Verdiana dkk., 2025);(Effendi, Irianto, Fauzi, dan Vitrani, 2025);(Sonjaya dkk., 2025). Hal ini menunjukkan bahwa *dataset* tersebut telah teruji dan diakui secara empiris dalam konteks penelitian kesehatan mental berbasis data.Dalam penelitian ini, *dataset Kaggle* tersebut digunakan sebagai data pelatihan(*training data*) untuk membangun dan membandingkan model klasifikasi depresi mahasiswa menggunakan algoritma *Random Forest(RF)* dan *Multi-layer Perceptron* dengan optimisasi *Stochastic Gradient Descent(MLP-SGD)* guna memperoleh model yang paling optimal dan akurat.

3.2.2 Data Primer

Selain menggunakan data sekunder, penelitian ini juga memanfaatkan data primer yang diperoleh melalui penyebaran kuesioner *Patient Health Questionnaire-9*(PHQ-9) kepada mahasiswa aktif Universitas Islam Negeri Sultan Syarif Kasim Riau (Lampiran B). Jumlah responden yang berhasil dihimpun dalam penelitian ini sebanyak 120 responden, yang berasal dari 8 fakultas yang ada di lingkungan UIN Sultan Syarif Kasim Riau (Tarbiyah dan Keguruan, Syariah dan Hukum, Dakwah dan Komunikasi, Ekonomi dan Ilmu Sosial, Sains dan Teknologi, Pertanian dan Peternakan, Ushuluddin, serta Psikologi). Responden mencakup mahasiswa dari berbagai jenjang pendidikan, mulai dari Sarjana(S1), Magister(S2), hingga Doktor(S3), sehingga data yang diperoleh merepresentasikan variasi tingkat akademik mahasiswa. Berdasarkan karakteristik demografis, responden terdiri dari mahasiswa berjenis kelamin laki-laki dan perempuan yang ditunjukan pada tabel Tabel 3.2

Tabel 3.2. Karakteristik Responden Data Primer

Karakteristik	Keterangan
Jumlah responden	120 mahasiswa aktif
Asal institusi	UIN Sultan Syarif Kasim Riau
Fakultas	8 fakultas
Jenjang pendidikan	S1, S2, S3
Jenis kelamin	Laki-laki dan Perempuan

Data primer digunakan sebagai data uji eksternal (*external testing data*) untuk mengevaluasi kemampuan generalisasi model klasifikasi depresi yang sebelumnya telah dilatih menggunakan dataset sekunder dari Kaggle. *Patient Health Questionnaire-9*(PHQ-9) merupakan instrumen skrining depresi yang terdiri dari sembilan pertanyaan mengenai gejala utama depresi yang dinilai berdasarkan frekuensi kemunculannya selama dua minggu terakhir, sebagaimana ditunjukkan pada Tabel Tabel 3.3.

Tabel 3.3. Instrumen Pernyataan PHQ-9

No	Kode	Pernyataan
1	PHQ1	Kurang berminat atau merasa tidak tertarik melakukan aktivitas sehari-hari
2	PHQ2	Merasa sedih, murung, atau putus asa
3	PHQ3	Mengalami kesulitan tidur atau tidur berlebihan
4	PHQ4	Merasa lelah atau kehilangan energi
5	PHQ5	Mengalami penurunan atau peningkatan nafsu makan
6	PHQ6	Merasa buruk terhadap diri sendiri atau merasa gagal

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 mengujikan 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.

Tabel 3.3 Instrumen Pernyataan PHQ-9 (Lanjutan)

No	Kode	Pernyataan
7	PHQ7	Mengalami kesulitan berkonsentrasi, misalnya saat belajar atau membaca
8	PHQ8	Bergerak atau berbicara lebih lambat dari biasanya, atau sebaliknya menjadi sangat gelisah
9	PHQ9	Memiliki pikiran untuk menyakiti diri sendiri atau merasa lebih baik mati

Setiap item dinilai menggunakan skala *Likert* empat tingkat (0–3), sehingga menghasilkan skor total dalam rentang 0–27, di mana skor yang lebih tinggi menunjukkan tingkat keparahan depresi yang lebih besar. Dalam penelitian ini, skor PHQ-9 digunakan sebagai dasar pelabelan data primer dan ditransformasikan menjadi klasifikasi biner, yaitu Tidak Depresi (skor < 10) dan Depresi (skor ≥ 10). Jumlah responden yang berhasil dihimpun dalam penelitian ini sebanyak 120 responden. Data primer tersebut digunakan sebagai data uji eksternal (*external testing data*) untuk mengevaluasi kemampuan generalisasi model klasifikasi depresi yang sebelumnya telah dilatih menggunakan *dataset* sekunder dari *Kaggle*. Ambang batas tersebut dipilih karena memiliki validitas klinis yang baik serta dinilai lebih sesuai untuk meningkatkan stabilitas dan kemudahan evaluasi model *machine learning*. Seluruh hasil pengisian kuesioner PHQ-9 telah melalui validasi profesional oleh Psikolog Rumah Sakit Jiwa(RSJ) Tampan dan dikonsultasikan dengan *KLC Psychology* yang ditampilkan pada (Lampiran A) dan (Lampiran D), sehingga pelabelan data primer dapat dipertanggungjawabkan secara klinis dan etis.

3.3 Pre-processing Data

Tahap *preprocessing* data merupakan langkah penting dalam penelitian ini untuk memastikan bahwa *dataset* berada dalam kondisi optimal sebelum digunakan pada proses pelatihan model *machine learning*. Oleh karena itu, dilakukan serangkaian tahapan pra-pemrosesan yang meliputi pembersihan dan seleksi fitur, transformasi variabel (*encoding*), pembagian data, dan *normalization*. Proses ini bertujuan agar seluruh fitur memiliki format dan skala yang seragam serta sesuai dengan kebutuhan algoritma yang digunakan, yaitu *Random Forest* dan *Multilayer Perceptron* (MLP).

3.3.1 Pembersihan Data dan Seleksi Fitur (*Data Cleaning & Feature Selection*)

Tahap ini diawali dengan proses penyaringan data (*filtering*) untuk memastikan bahwa populasi penelitian hanya mencakup responden dengan status sebagai mahasiswa. Pada tahap ini, variabel-variabel yang tidak relevan dengan konteks

analisis depresi pada mahasiswa dihapus guna meminimalkan potensi bias informasi. Langkah berikutnya adalah pembersihan data, yang mencakup penanganan terhadap entri yang tidak lengkap (*missing values*), penghapusan data duplikat, serta data yang mengandung nilai ekstrem atau tidak logis. Tahapan ini penting dilakukan untuk menjaga kualitas dan mencegah bias yang dapat mengganggu kinerja model dalam melakukan klasifikasi. Terakhir, dilakukan pemeriksaan terhadap konsistensi pada label target guna memastikan bahwa hanya terdapat dua kelas utama, yaitu Depresi(1) dan Tidak Depresi(0), sesuai dengan pendekatan klasifikasi biner dalam penelitian ini.

3.3.2 Transformasi Fitur Kategorikal (*Encoding*)

Algoritma *machine learning* tidak dapat memproses data *non-numeric* secara langsung, maka fitur yang bersifat kategorikal dikonversi ke dalam bentuk numerik menggunakan metode *Label Encoding*. Fitur kategorikal ditransformasikan menggunakan dua pendekatan berbeda sesuai dengan karakteristik datanya:

1. **Ordinal Encoding (Pemetaan Manual):** Diterapkan pada fitur yang memiliki tingkatan atau urutan logis (*ordinal*).
 - **Sleep Duration:** Dilakukan pemetaan manual untuk memastikan urutan waktu yang logis (misalnya: "5-6 hours" diberi nilai 0, hingga "7-8 hours" diberi nilai 3). Hal ini penting agar model dapat mempelajari pola bahwa durasi tidur yang lebih lama memiliki bobot nilai yang lebih besar.
 - **Degree:** Dilakukan penyederhanaan kategori (*bining*) dari 28 variasi jurusan menjadi 4 tingkatan jenjang pendidikan (0=SMA/*Class 12*, 1=*Undergraduate/S1*, 2=*Master/S2*, 3=*PhD/S3*). Teknik ini, yang disebut sebagai *Feature Engineering*, bertujuan mengurangi dimensi data agar model MLP dapat konvergen lebih optimal.
2. **Label Encoding:** Diterapkan pada fitur nominal atau biner yang tidak memiliki tingkatan hierarkis, seperti *Gender*, *Dietary Habits*, *Have you ever had suicidal thoughts?*, dan *Family History of Mental Illness*. Setiap kategori unik dikonversi menjadi representasi bilangan bulat(0 dan 1).

Dengan penerapan kedua metode tersebut, seluruh fitur kategorikal berhasil dikonversi menjadi bentuk numerik yang sesuai untuk kebutuhan algoritma klasifikasi, tanpa menghilangkan struktur informasi penting yang terkandung dalam masing-masing variabel.

3.3.3 Pembagian Data (*Data Splitting*)

Setelah data diubah ke numerik, *dataset* dibagi menjadi data latih(80%) dan data uji(20%) menggunakan *train-test split*. Proses pembagian dilakukan secara Acak Terstratififikasi(*Stratified Random Sampling*) berdasarkan variabel target(Depresi) untuk menjaga proporsi kelas Depresi dan Tidak Depresi tetap seimbang di data latih dan data uji. Strategi ini penting untuk mengatasi ketidakseimbangan kelas dan memastikan evaluasi model yang objektif. Data latih digunakan untuk pelatihan dan validasi model, sedangkan data uji digunakan untuk mengukur performa model dengan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*.

3.3.4 Normalisasi Data (*Feature Scaling*)

Tahapan terakhir dalam pra-pemrosesan adalah normalisasi data menggunakan metode *Min-Max Scaling*. Teknik ini mengubah rentang nilai fitur numerik menjadi antara 0 dan 1. Normalisasi diperlukan agar semua fitur(seperti *Age*, *CGPA*) berada pada skala yang sama dan tidak ada atribut yang mendominasi proses pembelajaran, terutama pada algoritma MLP yang sensitif terhadap perbedaan skala *input*. Proses ini dilakukan menggunakan fungsi *MinMaxScaler()* dengan rumus transformasi sebagai berikut:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.1)$$

Langkah ini memastikan distribusi data menjadi lebih seragam sehingga mempercepat dan menstabilkan proses konvergensi model pada saat pelatihan.

3.4 Penerapan Model

Tahapan penerapan model merupakan inti dari penelitian ini karena pada tahap ini dilakukan proses pelatihan dan pengujian algoritma *machine learning* yang digunakan untuk melakukan klasifikasi depresi pada mahasiswa. Dua algoritma yang diterapkan adalah *Random Forest(RF)* dan *Multilayer Perceptron* dengan *Stochastic Gradient Descent(MLP-SGD)*. Kedua model ini dipilih karena memiliki kemampuan yang kuat dalam menangani data multivariat dan mampu melakukan proses klasifikasi biner dengan performa yang kompetitif.

3.4.1 *Random Forest*

Dalam penelitian ini, model *Random Forest* menggunakan *scikit-learn* dengan parameter dasar yang kemudian dioptimalkan menggunakan *GridSearchCV* untuk menemukan kombinasi parameter terbaik. Parameter utama yang disesuaikan antara lain jumlah *estimator* (*n_estimators*), kedalaman maksimum po-

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 mengutip 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.

hon (*max_depth*), dan jumlah fitur yang dipertimbangkan pada setiap percabangan (*max_features*). Proses optimasi dilakukan dengan menggunakan teknik *cross-validation* guna memastikan bahwa hasil model tidak hanya baik pada data latih tetapi juga mampu beradaptasi dengan data uji.

3.4.2 Multilayer Perceptron dengan Stochastic Gradient Descent(MLP-SGD)

Penggunaan *MLP-SGD* pada penelitian ini dipilih karena mampu menangkap hubungan non-linear antar fitur yang mungkin tidak dapat ditangani dengan baik oleh model berbasis pohon. Selain itu, penerapan *activation function* seperti *ReLU* pada lapisan tersembunyi membantu model dalam mempelajari pola yang kompleks dalam data, sedangkan fungsi *sigmoid* pada lapisan output digunakan untuk menghasilkan probabilitas klasifikasi biner antara kelas Depresi dan Tidak Depresi. Proses pelatihan *MLP* dilakukan dengan jumlah *epoch* yang cukup untuk mencapai kestabilan *loss*, disertai penerapan *early stopping* untuk mencegah terjadinya *overfitting*.

3.5 Optimasi Hyperparameter

Tahap optimasi *hyperparameter* bertujuan untuk memperoleh konfigurasi model dengan performa terbaik pada data latih. Proses ini dilakukan menggunakan metode *GridSearchCV*, yang secara sistematis menguji berbagai kombinasi parameter berdasarkan hasil *cross-validation* sebanyak *5-fold*. Pendekatan ini memastikan model tidak hanya menyesuaikan diri dengan data latih (*overfitting*), tetapi juga mampu melakukan generalisasi terhadap data baru.

Pada algoritma *Random Forest*, parameter yang dioptimasi meliputi jumlah pohon dalam *ensemble* (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*). Sedangkan pada *MLP-SGD*, parameter yang disesuaikan mencakup jumlah neuron pada lapisan tersembunyi (*hidden_layer_sizes*), laju pembelajaran awal (*learning_rate_init*), serta fungsi aktivasi (*activation: ReLU, tanh*). *Solver* ditegakkan pada ‘sgd’ karena penelitian ini berfokus pada optimasi algoritma *Stochastic Gradient Descent* sebagai metode pembaruan bobot. *GridSearchCV* mengevaluasi seluruh kombinasi parameter menggunakan metrik akurasi, kemudian memilih konfigurasi dengan performa tertinggi ditunjukkan pada (Lampiran C). Melalui proses ini, model yang dihasilkan menjadi lebih stabil, adaptif, dan memiliki kemampuan prediksi yang lebih baik terhadap data yang belum pernah dilihat sebelumnya.

3.6 Evaluasi Model

Evaluasi model dilakukan untuk menilai sejauh mana performa algoritma *Random Forest* dan *Multilayer Perceptron-SGD(MLP-SGD)* dalam mengklasifikasi

fikasikan status depresi mahasiswa secara akurat. Setelah proses pelatihan dan optimasi *hyperparameter* menggunakan *GridSearchCV*, kedua model diuji pada data uji yang terpisah dari data latih guna mengukur kemampuan generalisasi terhadap data baru.

Metrik evaluasi yang digunakan meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. *Accuracy* mengukur persentase prediksi yang benar, *precision* menilai kualitas prediksi positif, *recall* mengukur kemampuan model dalam mendeteksi semua kasus positif yang sebenarnya, dan *F1-score* memberikan gambaran keseimbangan antara *precision* dan *recall*. Selain itu, *confusion matrix* juga digunakan untuk melihat distribusi prediksi benar dan salah pada masing-masing kelas (Depresi dan Tidak Depresi). Tahapan ini bertujuan untuk memberikan gambaran komprehensif mengenai performa kedua algoritma setelah melalui proses optimasi.

3.7 Tahap Deployment

Tahap *deployment* dilakukan dengan mengimplementasikan model *machine learning* terbaik ke dalam sistem berbasis web menggunakan *framework Flask*. Aplikasi ini memungkinkan pengguna untuk memasukkan data karakteristik mahasiswa dan memperoleh hasil prediksi status depresi secara *real-time*. *Deployment* bertujuan untuk menunjukkan bahwa model yang dikembangkan tidak hanya memiliki performa yang baik secara akademik, tetapi juga dapat diterapkan sebagai sistem pendukung keputusan awal dalam skrining depresi mahasiswa.

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan analisis yang telah dilakukan pada klasifikasi depresi mahasiswa menggunakan algoritma *Random Forest* dan *Multilayer Perceptron* dengan *Stochastic Gradient Descent*(MLP-SGD), maka dapat disimpulkan sebagai berikut:

1. Penerapan algoritma *Random Forest* dan *Multilayer Perceptron* dengan *Stochastic Gradient Descent*(MLP-SGD) menunjukkan bahwa kedua model mampu melakukan klasifikasi depresi mahasiswa dengan performa yang baik. Namun, berdasarkan hasil evaluasi komparatif, MLP-SGD secara konsisten menunjukkan kinerja yang sedikit lebih unggul pada metrik *accuracy*, *F1-score*, dan AUC-ROC, baik pada data sekunder maupun pada pengujian data primer berbasis kuesioner PHQ-9. Hal ini terlihat dalam menjaga keseimbangan antara *recall* dan *F1-score* yang krusial untuk klasifikasi depresi, sehingga model ini lebih efektif dalam memodelkan hubungan kompleks dan nonlinier pada data depresi mahasiswa.
2. Optimasi *hyperparameter* menggunakan *GridSearchCV* terbukti memberikan peningkatan performa pada kedua algoritma. *Random Forest* menunjukkan peningkatan yang stabil dan konsisten pada hampir seluruh metrik evaluasi, mencerminkan karakteristiknya sebagai algoritma *ensemble* yang *robust*. Sementara itu, pada MLP-SGD, optimasi *hyperparameter* mampu meningkatkan kemampuan diskriminatif dan generalisasi model, sehingga menghasilkan performa yang lebih optimal dibandingkan konfigurasi *baseline*.
3. Konfigurasi *hyperparameter* terbaik pada MLP-SGD, yaitu penggunaan dua *hidden layer* dengan kombinasi(64, 32) neuron, *learning rate* sebesar 0,01, momentum sebesar 0,9, serta fungsi aktivasi *tanh*, menghasilkan performa tertinggi dengan nilai *accuracy* sebesar 0,8366, *F1-score* sebesar 0,8345, dan AUC-ROC sebesar 0,9137. Berdasarkan hasil tersebut, MLP-SGD dengan optimasi *hyperparameter* ditetapkan sebagai model klasifikasi depresi mahasiswa terbaik dalam penelitian ini.

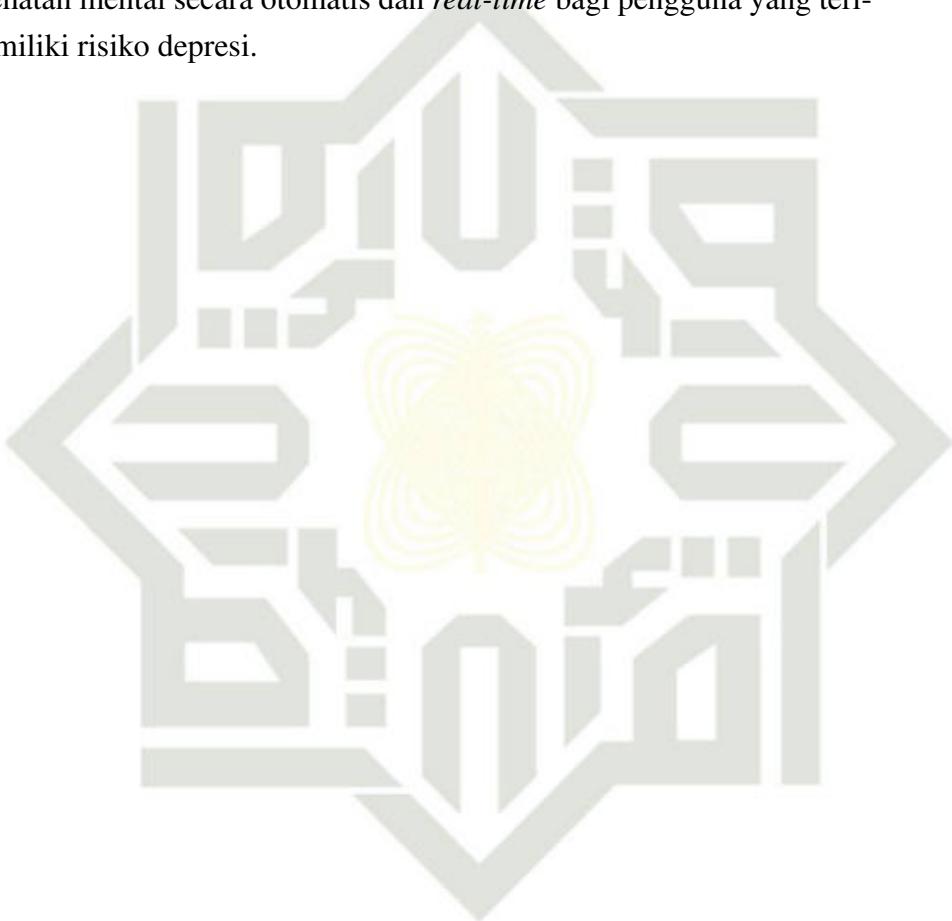
5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, pengembangan penelitian selanjutnya disarankan untuk mengeksplorasi metode optimasi *hyperparameter*

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 mengikuti 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.

yang lebih adaptif dan efisien, seperti *Bayesian Optimization* atau *Random Search*, guna mencakup ruang pencarian parameter yang lebih luas dalam waktu yang lebih singkat. Selain itu, eksperimen lebih lanjut pada variasi arsitektur jaringan saraf yang lebih kompleks serta penerapan strategi pembelajaran tambahan perlu dilakukan untuk meningkatkan stabilitas performa, khususnya dalam menekan kesalahan klasifikasi *false negative* yang sangat krusial dalam konteks kesehatan mental mahasiswa. Integrasi model ke dalam sistem pendukung keputusan yang lebih dinamis pada platform *website* yang diharapkan dapat memberikan rekomendasi intervensi kesehatan mental secara otomatis dan *real-time* bagi pengguna yang teridentifikasi memiliki risiko depresi.





- 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 meugikan 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 PUSTAKA

- Abdelbasset, W. K., Elkholi, S. M., Opulencia, M. J. C., Diana, T., Su, C.-H., Alashwal, M., ... Nguyen, H. C. (2022). Development of multiple machine-learning computational techniques for optimization of heterogenous catalytic biodiesel production from waste vegetable oil. *Arabian Journal of Chemistry*, 15(6), 103843.
- Agustin, W. S., Prastika, H. A., Kendrasti, G. K., Fajriyah, R., dan Le-Quy, V. (2025). Clustering the depression prevalence in indonesia provinces through natural breaks jenks method. *Clinical Practice and Epidemiology in Mental Health*, 21, 1–10.
- Atharbi, Y., dan Khan, S. S. (2025). Classifying multi-lingual reviews sentiment analysis in arabic and english languages using the stochastic gradient descent model. *Computers, Materials & Continua*, 83(1).
- Al Masud, G. n., Shanto, R. I., Sakin, I., dan Kabir, M. R. (2025). Effective depression detection and interpretation: Integrating machine learning, deep learning, language models, and explainable ai. *Array*, 25, 100375.
- Alsubaie, M. M., Stain, H. J., Webster, L. A. D., dan Wadman, R. (2019). The role of sources of social support on depression and quality of life for university students. *International journal of adolescence and youth*, 24(4), 484–496.
- Alves, P., Marci, C. D., Cohen-Stavi, C. J., Whelan, K. M., dan Boussios, C. (2025). A machine learning model using clinical notes to estimate phq-9 symptom severity scores in depressed patients. *Journal of affective disorders*, 376, 216–224.
- Aruleba, R. T., Adekiya, T. A., Ayawei, N., Obaido, G., Aruleba, K., Mienye, I. D., ... Ogbuokiri, B. (2022). Covid-19 diagnosis: a review of rapid antigen, rt-pcr and artificial intelligence methods. *Bioengineering*, 9(4), 153.
- Asadi, S., Roshan, S., dan Kattan, M. W. (2021). Random forest swarm optimization-based for heart diseases diagnosis. *Journal of biomedical informatics*, 115, 103690.
- Bantjes, J. (2024). Global knowledge-power asymmetries and student mental health in sub-saharan africa: A case study from south africa. *Journal of International Students*, 14(2), 136–149.
- Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., ... others (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(2), e1484.

- Bradshaw, T. J., Huemann, Z., Hu, J., dan Rahmim, A. (2023). A guide to cross-validation for artificial intelligence in medical imaging. *Radiology: Artificial Intelligence*, 5(4), e220232.
- Candelieri, A., Ponti, A., dan Archetti, F. (2024). Fair and green hyperparameter optimization via multi-objective and multiple information source bayesian optimization. *Machine Learning*, 113(5), 2701–2731.
- Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., ... Jianzhong (2020). The psychological impact of the covid-19 epidemic on college students in china. *Psychiatry research*, 287, 112934.
- Chen, M., Mangalathu, S., dan Jeon, J.-S. (2022). Machine learning-based seismic reliability assessment of bridge networks. *Journal of Structural Engineering*, 148(7), 06022002.
- Cukic, M., Lopez, V., dan Pavon, J. (2020). Classification of depression through resting-state electroencephalogram as a novel practice in psychiatry. *Journal of medical Internet research*, 22(11).
- Dani, H., Bhople, P., Waghmare, H., Munginwar, K., dan Patil, A. (2022). Review on frameworks used for deployment of machine learning model. *Int J Res Appl Sci Eng Technol*, 10(2), 211–215.
- Dessauvagie, A. S., Dang, H.-M., Nguyen, T. A. T., dan Groen, G. (2022). Mental health of university students in southeastern asia: a systematic review. *Asia Pacific Journal of Public Health*, 34(2-3), 172–181.
- Dong, J., Li, Y., Chen, J., He, L., dan Li, Y. (2024). The relationship between depression heterogeneity and academic performance in chinese university freshmen: a latent profile analysis. *Current Psychology*, 43(23), 20484–20495.
- Dwinnie, Z. C., Khairani, L., Putri, M. A. M., Adhiva, J., dan Tsamarah, M. I. F. (2023). Application of the supervised learning algorithm for classification of pregnancy risk levels. *Public Research Journal of Engineering, Data Technology and Computer Science*, 1(1), 26–33.
- Eeer, F., Ardabili, S., Band, S. S., dan Mosavi, A. (2020). Training multilayer perceptron with genetic algorithms and particle swarm optimization for modeling stock price index prediction. *Entropy*, 22(11), 1239.
- Effendi, M. K., Irianto, S. Y., Fauzi, C., dan Vitriani, Y. (2025). Optimizing student depression prediction using particle swarm optimization and random forest. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 11(1), 44–50.
- El Massari, H., Sabouri, Z., Mhammedi, S., dan Gherabi, N. (2022). Diabetes prediction using machine learning algorithms and ontology. *Journal of ICT*

- Standardization*, 10(2), 319–337.
- Elovanio, M., Hakulinen, C., Pulkki-Råback, L., Aalto, A.-M., Virtanen, M., Partonen, T., dan Suvisaari, J. (2020). General health questionnaire (ghq-12), beck depression inventory (bdi-6), and mental health index (mhi-5): psychometric and predictive properties in a finnish population-based sample. *Psychiatry Research*, 289, 112973.
- Elujide, I., Fashoto, S. G., Fashoto, B., Mbunge, E., Folorunso, S. O., dan Olamijuwon, J. O. (2021). Application of deep and machine learning techniques for multi-label classification performance on psychotic disorder diseases. *Informatics in medicine Unlocked*, 23, 100545.
- Fajri, M., dan Primajaya, A. (2023). Komparasi teknik hyperparameter optimization pada svm untuk permasalahan klasifikasi dengan menggunakan grid search dan random search. *Journal of Applied Informatics and Computing*, 7(1), 10–15.
- Fatima, N., dkk. (2020). Enhancing performance of a deep neural network: A comparative analysis of optimization algorithms. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*.
- Gao, X., dan Liu, Z. (2024). Analyzing the psychometric properties of the phq-9 using item response theory in a chinese adolescent population. *Annals of General Psychiatry*, 23(1), 7.
- Han, S.-S., Zhang, Y.-S., Zhu, W., Ye, Y.-P., Li, Y.-X., Meng, S.-Q., ... others (2025). Status and epidemiological characteristics of depression and anxiety among chinese university students in 2023. *BMC Public Health*, 25(1), 1189.
- He, Y., Sakuma, K., Kishi, T., Li, Y., Matsunaga, M., Tanihara, S., ... Ota, A. (2024). External validation of a machine learning model for schizophrenia classification. *Journal of Clinical Medicine*, 13(10), 2970.
- Helmund, E., Fitriyani, F., Romadiana, P., dkk. (2024). Classification comparison performance of supervised machine learning random forest and decision tree algorithms using confusion matrix. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 13(1), 92–97.
- idayat, T., Ridwan, M., Iqbal, M. F., Sukisno, S., Rizky, R., dan Manongga, W. E. (2025). Determining toddler's nutritional status with machine learning classification analysis approach. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 24(2), 235–246.
- Ho, S. Y., Phua, K., Wong, L., dan Goh, W. W. B. (2020). Extensions of the external validation for checking learned model interpretability and generalizability. *Patterns*, 1(8).

- Jentzen, A., dan Riekert, A. (2022). A proof of convergence for stochastic gradient descent in the training of artificial neural networks with relu activation for constant target functions. *Zeitschrift für angewandte Mathematik und Physik*, 73(5), 188.
- Kabir, F., Hossain, M. A., Rahman, A. M., dan Mishu, S. Z. (2023). Depression detection from social media textual data using natural language processing and machine learning techniques. Dalam *2023 26th international conference on computer and information technology (iccit)* (hal. 1–6).
- Kamran, M., Bibi, F., ur. Rehman, A., dan Morris, D. W. (2022). Major depressive disorder: existing hypotheses about pathophysiological mechanisms and new genetic findings. *Genes*, 13(4), 646.
- Kiekens, G., Hasking, P., Bruffaerts, R., Alonso, J., Auerbach, R. P., Bantjes, J., ... Claes, L. (2023). Non-suicidal self-injury among first-year college students and its association with mental disorders: results from the world mental health international college student (wmh-ics) initiative. *Psychological medicine*, 53(3), 875–886.
- Koch, B., Denton, E., Hanna, A., dan Foster, J. G. (2021). Reduced, reused and recycled: The life of a dataset in machine learning research. *arXiv preprint arXiv:2112.01716*.
- Koesten, L., Vougiouklis, P., Simperl, E., dan Groth, P. (2020). Dataset reuse: Toward translating principles to practice. *Patterns*, 1(8).
- Kristensen, S. M., Larsen, T. M. B., Urke, H. B., dan Danielsen, A. G. (2023). Academic stress, academic self-efficacy, and psychological distress: A moderated mediation of within-person effects. *Journal of youth and adolescence*, 52(7), 1512–1529.
- Kroenke, K., Spitzer, R. L., dan Williams, J. B. (2001). The phq-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9), 606–613.
- Laoli, S. E. W., Nauli, F. A., dan Karim, D. (2022). Gambaran tingkat depresi mahasiswa dan faktor penyebab depresi mahasiswa tingkat akhir di universitas riau. *Coping: Community of Publishing in Nursing*, 10(1), 115.
- Le, H.-L., Le, T.-T., Vu, T., Tran, D., Chau, D., dan Ngo, T. (2023). A survey on the impact of hyperparameters on random forest performance using multiple accelerometer datasets. *Int. J. Comput. Their Appl.*, 30, 351–361.
- Levis, B., Bhandari, P. M., Neupane, D., Fan, S., Sun, Y., He, C., ... others (2024). Data-driven cutoff selection for the patient health questionnaire-9 depression screening tool. *JAMA network open*, 7(11), e2429630–e2429630.

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 meugikan 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.

- Li, X., Zou, L., dan Li, H. (2024). Multilayer perceptron-based wearable exercise-related heart rate variability predicts anxiety and depression in college students. *Sensors*, 24(13), 4203.
- Luo, L., Yuan, J., Wu, C., Wang, Y., Zhu, R., Xu, H., ... Zhang, Z. (2025). Predictors of depression among chinese college students: a machine learning approach. *BMC Public Health*, 25(1), 470.
- Meda, N., Pardini, S., Rigobello, P., Visioli, F., dan Novara, C. (2023). Frequency and machine learning predictors of severe depressive symptoms and suicidal ideation among university students. *Epidemiology and Psychiatric Sciences*, 32, e42.
- Mendoza, T., Lee, C.-H., Huang, C.-H., dan Sun, T.-L. (2021). Random forest for automatic feature importance estimation and selection for explainable postural stability of a multi-factor clinical test. *Sensors*, 21(17), 5930.
- Mohamed, E. S., Naqishbandi, T. A., Bukhari, S. A. C., Rauf, I., Sawrikar, V., dan Hussain, A. (2023). A hybrid mental health prediction model using support vector machine, multilayer perceptron, and random forest algorithms. *Healthcare Analytics*, 3, 100185.
- Moulaei, K., Shanbehzadeh, M., Mohammadi-Taghiabad, Z., dan Kazemi-Arpanahi, H. (2022). Comparing machine learning algorithms for predicting covid-19 mortality. *BMC medical informatics and decision making*, 22(1), 2.
- Mulugeta, G., Zewotir, T., Tegegne, A. S., Juhar, L. H., dan Muleta, M. B. (2023). Classification of imbalanced data using machine learning algorithms to predict the risk of renal graft failures in ethiopia. *BMC Medical Informatics and Decision Making*, 23(1), 98.
- Mumenin, N., Hossain, A. K., Hossain, M. A., Debnath, P. P., Della, M. N., Rashed, M. M. H., ... Hossain, M. S. (2024). Screening depression among university students utilizing ghq-12 and machine learning. *Heliyon*, 10(17).
- Narayanan, S., Balamurugan, N., Palas, P. B., dkk. (2022). Leveraging machine learning methods for multiple disease prediction using python ml libraries and flask api. Dalam 2022 *international conference on applied artificial intelligence and computing (icaaic)* (hal. 694–701).
- Nesratpour, R., Tafakori, L., dan Abdollahian, M. (2025). Satellite-based extreme precipitation estimation using convolution neural networks and ant colony optimized multi-layers perceptron. *Atmospheric Research*, 320, 108037.
- Nugraha, W., dan Sasongko, A. (2022). Hyperparameter tuning on classification algorithm with grid search. *Sistemasi*, 11(2), 391–401.
- Obaido, G., Mienye, I. D., Egbelowo, O. F., Emmanuel, I. D., Ogunleye, A., Og-



UIN SUSKA RIAU

- buokiri, B., ... Aruleba, K. (2024). Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects. *Machine Learning with Applications*, 17, 100576.
- Osmam, S. H., Abu, N., Aziz, H., Chow, Y. P., Wan Mohamad Nazarie, W. F., Ab Mutualib, N.-S., ... Jamal, R. (2020). Deep transcriptome sequencing of pediatric acute myeloid leukemia patients at diagnosis, remission and relapse: Experience in 3 malaysian children in a single center study. *Frontiers in Genetics*, 11, 66.
- Pascual, M. A., Levenson, J. C., Merranko, J., Gill, M. K., Hower, H., Yen, S., ... others (2020). The effect of traumatic events on the longitudinal course and outcomes of youth with bipolar disorder. *Journal of affective disorders*, 274, 126–135.
- Pratama, P. F., Rahmadani, D., Nahampun, R. S., Harmutika, D., Rahmadeyan, A., dan Evizal, M. F. (2023). Random forest optimization using particle swarm optimization for diabetes classification. *Public Research Journal of Engineering, Data Technology and Computer Science*, 1(1), 41–46.
- Raith, A., Kamp, C., Stoiber, C., Jakl, A., dan Wagner, M. (2022). Augmented reality in radiology for education and training—a design study. Dalam *Healthcare* (Vol. 10, hal. 672).
- Ramadianto, A. S., Kusumadewi, I., Agiananda, F., dan Raharjanti, N. W. (2022). Symptoms of depression and anxiety in indonesian medical students: association with coping strategy and resilience. *BMC psychiatry*, 22(1), 92.
- Rasheed, S., Kumar, G. K., Rani, D. M., Kantipudi, M. P., dan Anila, M. (2024). Heart disease prediction using gridsearchcv and random forest. *EAI Endorsed Transactions on Pervasive Health and Technology*, 10.
- Roberts, T., Esponda, G. M., Torre, C., Pillai, P., Cohen, A., dan Burgess, R. A. (2022). Reconceptualising the treatment gap for common mental disorders: a fork in the road for global mental health? *The British Journal of Psychiatry*, 221(3), 553–557.
- Rolnick, J. A., Liao, J. M., Emanuel, E. J., Huang, Q., Ma, X., Shan, E. Z., ... others (2020). Spending and quality after three years of medicare's bundled payments for medical conditions: quasi-experimental difference-in-differences study. *bmj*, 369.
- Saha, D. K., Hossain, T., Safran, M., Alfarhood, S., Mridha, M., dan Che, D. (2024). Ensemble of hybrid model based technique for early detecting of depression based on svm and neural networks. *Scientific Reports*, 14(1), 25470.
- Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott,

- D. M., ... Aravkin, A. Y. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the covid-19 pandemic. *The Lancet*, 398(10312), 1700–1712.
- Smith, S. L., Dherin, B. R. U., Barrett, D. G., dan De, S. (2021). Stochastic gradient descent. *Machine learning with neural networks*.
- Sonjaya, R. P., Gintara, A. R., Riza, L. S., Nursalman, M., Nugraha, E., dan Wahyudin, D. (2025). Predicting student depression using the naive bayes model on the student depression dataset from kaggle. *JENTIK: Jurnal Pendidikan Teknologi Informasi dan Komunikasi*, 4(1), 1–12.
- Sopelsa Neto, N. F., Stefenon, S. F., Meyer, L. H., Bruns, R., Nied, A., Seman, L. O., ... Yow, K.-C. (2021). A study of multilayer perceptron networks applied to classification of ceramic insulators using ultrasound. *Applied Sciences*, 11(4), 1592.
- Steare, T., Muñoz, C. G., Sullivan, A., dan Lewis, G. (2023). The association between academic pressure and adolescent mental health problems: A systematic review. *Journal of affective disorders*, 339, 302–317.
- Suherlan, E., Arti, S., Nabilah, S., dan Hazimah, Z. (2025). Penerapan flask framework untuk deployment model machine learning dalam mendukung analisis adaptasi mahasiswa pada pembelajaran daring. *PINTER: Jurnal Pendidikan Teknik Informatika dan Komputer*, 9(1), 110–117.
- Sumayli, A. (2023). Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (gpr), multilayer perceptron (mlp), and k-nearest neighbor (knn) regression models. *Arabian Journal of Chemistry*, 16(7), 104833.
- Sun, Y., Wang, J., Wang, T., Li, J., Wei, Z., Fan, A., ... Chen, Y. (2024). Post-fracture production prediction with production segmentation and well logging: Harnessing pipelines and hyperparameter tuning with gridsearchcv. *Applied Sciences*, 14(10), 3954.
- Teng, C., Yang, C., dan Liu, Q. (2024). Utilising ai technique to identify depression risk among doctoral students. *Scientific Reports*, 14(1), 31978.
- Teodorescu, V., dan Obreja Brașoveanu, L. (2025). Assessing the validity of k-fold cross-validation for model selection: Evidence from bankruptcy prediction using random forest and xgboost. *Computation*, 13(5), 127.
- Tetko, I. V., van Deursen, R., dan Godin, G. (2024). Be aware of overfitting by hyperparameter optimization! *Journal of Cheminformatics*, 16(1), 139.
- Tian, Y., Zhang, Y., dan Zhang, H. (2023). Recent advances in stochastic gradient descent in deep learning. *Mathematics*, 11(3), 682.



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 meugikan kepentingan yang wajar UIN Suska Riau.

- Trofimov, I., Korotaev, P., Ivanov, I., Malginov, A., Tokhtamyshev, A., Yanilkin, A., dan Kruglov, I. (2025). Machine learning assisted design of reactor steels with high long-term strength and toughness. *Materials & Design*, 114014.
- Verdiana, M., Nugroho, E. D., Anggraini, L., Bagaskara, R., Yulita, W., Afriansyah, A., dan Algifari, M. H. (2025). Analisis hubungan dan prediksi depresi mahasiswa berdasarkan faktor akademik dan gender. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 33–42.
- Verma, D., Bach, K., dan Mork, P. J. (2023). External validation of prediction models for patient-reported outcome measurements collected using the selfback mobile app. *International Journal of Medical Informatics*, 170, 104936.
- Wallace, M. L., Mentch, L., Wheeler, B. J., Tapia, A. L., Richards, M., Zhou, S., ... Buysse, D. J. (2023). Use and misuse of random forest variable importance metrics in medicine: demonstrations through incident stroke prediction. *BMC medical research methodology*, 23(1), 144.
- Wilaksono, B., Damry, N. M., Aminuddin, N. M., dan Alfiandari, D. (2025). Mental health and depression in indonesia: Risk factors and service utilization for healthy aging using ski 2023 data. *Jurnal Penelitian dan Pengabdian Masyarakat*, 4(10), 3430–3436.
- Wilks, C. R., Auerbach, R. P., Alonso, J., Benjet, C., Bruffaerts, R., Cuijpers, P., ... others (2020). The importance of physical and mental health in explaining health-related academic role impairment among college students. *Journal of psychiatric research*, 123, 54–61.
- Wu, Y., Fan, L., Xia, F., Zhou, Y., Wang, H., Feng, L., ... others (2024). Global, regional, and national time trends in incidence for depressive disorders, from 1990 to 2019: an age-period-cohort analysis for the gbd 2019. *Annals of General Psychiatry*, 23(1), 28.
- Yang, Y., dan Wang, H. (2025). Random forest-based machine failure prediction: A performance comparison. *Applied Sciences*, 15(16), 8841.
- Yuan, X., Liu, S., Feng, W., dan Dauphin, G. (2023). Feature importance ranking of random forest-based end-to-end learning algorithm. *Remote Sensing*, 15(21), 5203.
- Zhang, J., Wang, T., Wang, B., Chen, C., dan Wang, G. (2023). Hyperparameter optimization method based on dynamic bayesian with sliding balance mechanism in neural network for cloud computing. *Journal of Cloud Computing*, 12(1), 109.
- Zhao, Y., Fang, L., Cui, L., dan Bai, S. (2020). Application of data mining for predicting hemodynamics instability during pheochromocytoma surgery. *BMC*



UIN SUSKA RIAU

- medical informatics and decision making*, 20(1), 165.
- Zulfiqar, M., Gamage, K. A., Kamran, M., dan Rasheed, M. B. (2022). Hyper-parameter optimization of bayesian neural network using bayesian optimization and intelligent feature engineering for load forecasting. *Sensors*, 22(12), 4446.

©Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

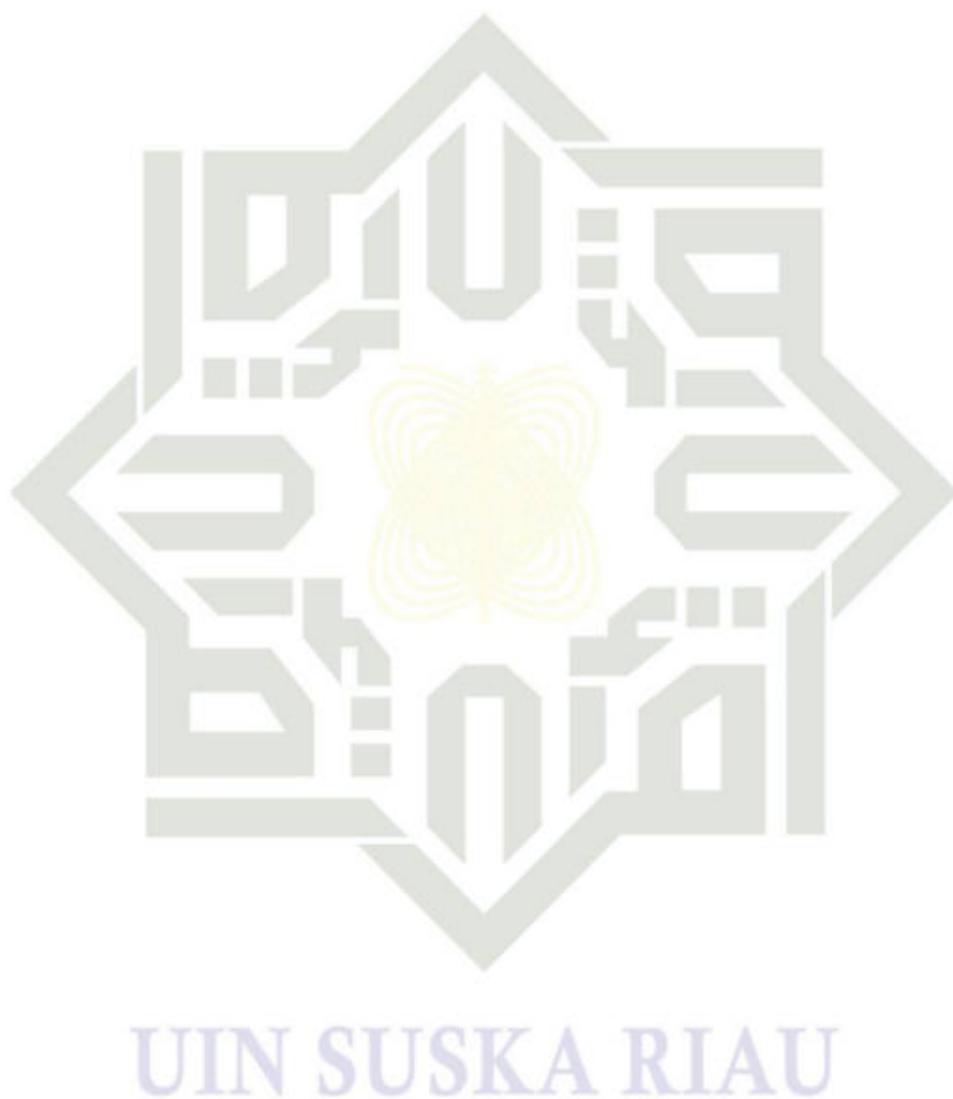
Hak Cipta Dilindungi Undang-Undang

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.





UIN SUSKA RIAU

© 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 mengujikan 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

IZIN PENELITIAN DAN VALIDASI DATA



KEMENTERIAN AGAMA REPUBLIK INDONESIA

UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

FAKULTAS SAINS DAN TEKNOLOGI

كلية العلوم والتكنولوجيا

FACULTY OF SCIENCES AND TECHNOLOGY

Jl. H.R Soebrantas KM.15 No. 155 Tuah Madani Kec. Tuah Madani- Pekanbaru 28298 PO Box. 1004
Fax. (0761) 589 0259 Web: www.uin-suska.ac.id E-mail: layanan.faste@uin-suska.ac.id

Nomor : B-8798 / F/PP.009/12/2025
Sifat : Biasa
Hal : Mohon Izin Penelitian dan Pengambilan Data
Tugas Akhir/Skripsi

Pekanbaru, 22 Desember 2025

Kepada Yth.
Kepala KLC PSYCHOLOGY
Jl. Sekuntum Raya No.33, Delima
Kec. Tampan, Kota Pekanbaru, Riau

Assalamualaikum Wr. Wb.
Dengan hormat, Sehubungan telah dimulainya mata kuliah Tugas Akhir pada Fakultas Sains dan Teknologi UIN Suska Riau, kami bermaksud mengirimkan mahasiswa:

Nama : Fatimah Azzahra
NIM : 12250324568
Fakultas : Sains dan Teknologi
Program Studi/Smt : Sistem Informasi / 7
No. Hp/E-mail : 08989317916

untuk pengambilan data yang sangat dibutuhkan dalam Tugas akhir mahasiswa tersebut yang berjudul "Perbandingan Kinerja algoritma Random Forest dan Multilayer Perceptron-SGD dalam klasifikasi depresi mahasiswa". Kami mohon kiranya saudara berkenan memberikan izin dan fasilitas demi kelancaran Tugas Akhir mahasiswa yang bersangkutan.

Demikian surat ini kami sampaikan, atas bantuan dan kerjasama Saudara kami ucapan terima kasih.

Wassalam

Dekan



Dr. Yuslenita Muda, M.Sc
NIP. 197701032007102001



Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan Balai Besar Sertifikasi Elektronik (BBSrE).
Token : xC2yyvK1

Gambar A.1. Surat Izin Pemngambilan data

UIN SUSKA RIAU



UIN SUSKA RIAU

State Islamic University of Sultan Syarif Kasim Riau

© Hak cipta milik 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.

1 dari 1

KLC.

Education & Psychology
Psychology

KLC PSYCHOLOGY

Jl. Sekuntum Raya No. 33 C, Kelurahan Delima, Pekanbaru, Riau, 28290 Tel: 0821-
6045-9759 | Web: <https://klc-edp.com> | Email: info@klc-edp.com

SURAT KETERANGAN VALIDASI DATA

No.09.113/KLC-Psy/l/2026

Yang bertanda tangan dibawah ini :

Nama : Masni Kamallia S.Psi.,M.Psi
Jabatan : Direktur KLC.Psycholg

Menerangkan dengan sesungguhnya bahwa yang bersangkutan dibawah ini :

Nama : Fatimah Azzahra
NIM : 12250324568
Program Studi : Sistem Informasi
Instansi : UIN Suska Riau

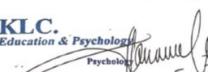
Telah melakukan Validasi Data di KLC Psychology yang berada dibawah naungan KLC Educational & Psychologi pada taggal: 24 Desember 2026. Validasi data yang dilakukan oleh Fatimah Azzahra bertujuan untuk memastikan keakuratan dan keandalan data yang digunakan dalam penelitiannya yang berjudul "Perbandingan Kinerja Algoritma Random Forest dan Multilayer Perceptron-SGD dalam Klasifikasi Depresi Mahasiswa" dan telah mengikuti prosedur serta peraturan yang berlaku di KLC Psychology.

Demikian surat keterangan ini dibuat untuk digunakan sebagaimana mestinya.

Pekanbaru, 10 Januari 2026

Pemberi keterangan

Direktur KLC Psychology


KLC.
Education & Psychology
Psychology

Masni Kamallia S.Psi., M.Psi

Gambar A.2. Surat Izin Validasi data



UIN SUSKA RIAU

State Islamic University of Sultan Syarif Kasim Riau

© Hak cipta milik 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 laporan, pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 - b. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN B KUESIONER PENELITIAN

KLASIFIKASI DEPRESI PADA MAHASISWA MENGGUNAKAN INSTRUMEN PHQ-9

Anda tidak akan diwajibkan memberitahukan:

Catatan Pendahuluan

- Seluruh jawaban yang diberikan berbantuan rohani dan bantuan untuk kepentingan akademik dan penelitian.
- Tidak solusi jawaban benar atau salah. Cukup karena itu, sebaik mengisi jawaban ini secara jujur sesuai dengan kondisinya.
- Perintah pada kuesioner ini merupakan hasil pengalaman Anda selama 2 minggu terakhir.
- Waktu perintah kuesioner ini sekitar 5-10 menit.

Atas kebaikan dan partisipasi bapak/ibu, saya siapkan terima hasil yang sebaik-baiknya. Terima kasih SMT memberikan kebaikan dan dengan telitiannya dan kerjakan.

Wawancara dilakukan Wawancara oleh Wawancara.

*Menunjukkan pertanyaan yang negatif.

Gambar B.1. Tampilan Kuesioner PHQ-9

1. Nama Lengkap *

2. Umur *
berikan jawaban dalam bentuk angka, contoh : 21

3. Jenis Kelamin *
Tanda catu oval saja

Perempuan
Laki-laki

4. Jenjang Pendidikan *
Tanda catu oval saja

Diploma
Belajar ISS
Magister ISS
Doktor ISS

Gambar B.2. Pengisian Data Pribadi

5. Seberapa besar tekanan akademik yang Anda rasakan saat ini? *

1 = Sangat Rendah - 5 = Sangat Tinggi

Tanda catu oval saja

Sangat Rendah
Rendah
Sedang
Tinggi
Sangat Tinggi

Gambar B.3. Faktor Terkait Depresi

16. Kehilangan minat atau keinginan dalam melakukan aktivitas *
Tanda catu oval saja

0 1 2 3
○ ○ ○ ○

Gambar B.4. Pernyataan Instrumen PHQ-9

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 meugikan 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 C

UJI COBA

Lampiran 1: Hasil Eksperimen 10 Terbaik MLP-SGD (GridSearchCV)

No	Activation	Alpha	Hidden Layer	LR Init	Accuracy	Precision	Recall	F1-Macro
1	tanh	0,0001	(64, 32)	0,01	0,8457	0,8451	0,8427	0,8436
2	tanh	0,0010	(64, 32)	0,010	0,8457	0,8451	0,8427	0,8436
3	relu	0,0001	(64, 32)	0,001	0,8441	0,8435	0,8408	0,8419
4	relu	0,0010	(64, 32)	0,001	0,8441	0,8435	0,8408	0,8419
5	tanh	0,0001	(128, 64)	0,01	0,8442	0,8436	0,8408	0,8419
6	tanh	0,0010	(128, 64)	0,01	0,8442	0,8436	0,8408	0,8419
7	relu	0,0010	(128, 64)	0,01	0,8442	0,8439	0,8405	0,8418
8	relu	0,0001	(128, 64)	0,01	0,8439	0,8438	0,8400	0,8415
9	relu	0,0001	(64, 32)	0,01	0,8430	0,8415	0,8421	0,8415
10	relu	0,0010	(64, 32)	0,01	0,8431	0,8420	0,8420	0,8416

Lampiran 2: Hasil Eksperimen 10 Terbaik Random Forest (GridSearchCV)

No	N estimators	Max depth	Min split	Min leaf	Akurasi	Precision	Recall	F1-score
1	100	10	5	2	0,8413	0,8404	0,8383	0,8392
2	300	15	5	4	0,8412	0,8403	0,8382	0,8390
3	300	None	10	2	0,8411	0,8401	0,8381	0,8389
4	300	25	10	2	0,8411	0,8401	0,8381	0,8389
5	300	10	5	2	0,8411	0,8404	0,8377	0,8388
6	100	15	10	2	0,8409	0,8401	0,8379	0,8388
7	300	15	10	2	0,8409	0,8399	0,8378	0,8387
8	200	25	10	4	0,8408	0,8400	0,8378	0,8387
9	200	None	10	4	0,8408	0,8400	0,8378	0,8387
10	200	15	5	4	0,8407	0,8398	0,8377	0,8386

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.

LAMPIRAN D

DOKUMENTASI



Gambar D.1. Dokumentasi bersama pihak KLC psikolog



DAFTAR RIWAYAT HIDUP

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.

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

Fatimah Azzahra lahir di Singkarak pada tanggal 09 Oktober 2004 dan merupakan anak ke-2 dari pasangan Bapak Syahrial dan Ibu Salbetri. Peneliti menempuh pendidikan dasar di SD Muhammadiyah Perawang pada tahun 2010 dan menyelesaikannya pada tahun 2016, kemudian melanjutkan pendidikan menengah pertama di SMP Negeri 1 Tualang dan lulus pada tahun 2019, serta pendidikan menengah atas di SMA Negeri 1 Tualang yang diselesaikan pada tahun 2022. Pada tahun yang sama, peneliti melanjutkan pendidikan tinggi di Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Selama masa perkuliahan, peneliti aktif dalam organisasi Himpunan Mahasiswa Sistem Informasi (HIMASI) dan *Puzzle Research Data Technology* (Predatech). Peneliti telah melaksanakan Kerja Praktek (KP) di PT Indah Kiat Pulp and Paper Perawang pada Divisi AES (*Automatic Electrical Specialist*), dengan fokus pada pembelajaran di bidang kelistrikan serta analisis penggunaan sistem PI-Vision sebagai sistem monitoring data proses, yang turut memberikan pengalaman dalam pengembangan dan analisis sistem berbasis website. Selain itu, peneliti beberapa kali mengikuti lomba desain poster dan memperoleh beasiswa prestasi dari Pemerintah Provinsi (PEMPROV). Peneliti menyelesaikan pendidikan Strata Satu (S1) dalam waktu 7 semester setelah berhasil menyelesaikan Tugas Akhir (TA) dengan judul “Perbandingan Kinerja Algoritma Machine Learning dalam Klasifikasi Depresi Mahasiswa”, dengan bidang minat meliputi *machine learning*, *data mining*, *design grafis* dan pengembangan sistem berbasis web.

UIN SUSKA RIAU