



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.

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

PREDIKSI RISIKO KESEHATAN PADA IBU HAMIL MENGGUNAKAN ALGORITMA MACHINE LEARNING

TUGAS AKHIR

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

Oleh:

RAHMA DEVI
12250323758



UIN SUSKA RIAU

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2026

LEMBAR PERSETUJUAN
PREDIKSI RISIKO KESEHATAN PADA IBU HAMIL
MENGGUNAKAN ALGORITMA MACHINE LEARNING

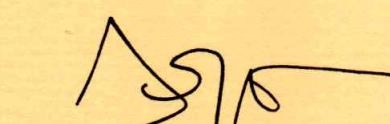
TUGAS AKHIR

Oleh:

RAHMA DEVI
12250323758

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

Ketua Program Studi


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

Pembimbing


Inggih Permana, ST., M.Kom.
NIP. 198812102015031006

LEMBAR PENGESAHAN
PREDIKSI RISIKO KESEHATAN PADA IBU HAMIL
MENGGUNAKAN ALGORITMA MACHINE LEARNING

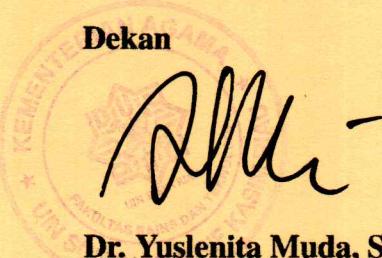
TUGAS AKHIR

Oleh:

RAHMA DEVI
12250323758

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 08 Januari 2026

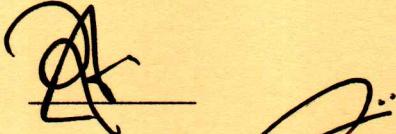
Pekanbaru, 08 Januari 2026
Mengesahkan,

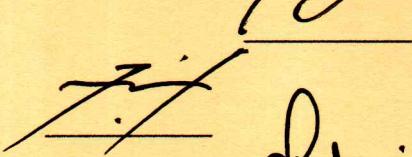
Dekan

Dr. Yuslenita Muda, S.Si., M.Sc.
NIP. 197701032007102001

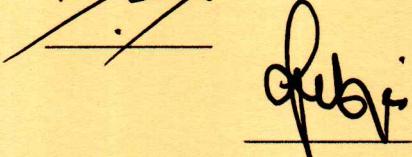
Ketua Program Studi

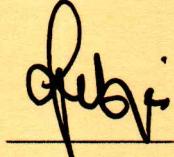

Angraini, S.Kom., M.Eng., Ph.D.
NIP. 198408212009012008

DEWAN PENGUJI:

Ketua : Muhammad Jazman, S.Kom., M.Infosys. 

Sekretaris : Inggih Permana, ST., M.Kom. 

Anggota 1 : Dr. Rice Novita, S.Kom., M.Kom. 

Anggota 2 : Febi Nur Salisah, S.Kom., M.Kom. 

Lampiran Surat :

Nomor : Nomor 25/2021
Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Rahma Devi

NIM : 12250323758

Tempat/Tgl. Lahir : Sibiruang/04 Januari 2003

Fakultas/Pascasarjana : Sains dan Teknologi

Prodi : Sistem Informasi

Judul Skripsi :

**PREDIKSI RISIKO KESEHATAN PADA IBU HAMIL MENGGUNAKAN
ALGORITMA MACHINE LEARNING**

Menyatakan dengan sebenar-benarnya bahwa :

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

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

Pekanbaru, 15 Januari 2026



Rahma Devi

NIM: 12250323758



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.

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.

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 diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 15 Januari 2026

Yang membuat pernyataan,

RAHMA DEVI

NIM. 12250323758

UIN SUSKA RIAU



LEMBAR PERSEMPAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Dengan menyebut nama Allah yang maha pengasih lagi maha penyayang

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillahi Rabbil 'Alamin, 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 syafa'at-Nya di dunia maupun di akhirat, Aamiin Ya Rabbal'Alaamiin.

Tugas Akhir ini peneliti persembahkan kepada orang tua tercinta, sebagai salah satu hadiah istimewa bentuk bakti, rasa terima kasih, dan hormat peneliti kepada orang tua tercinta. Ayahanda M.Hidayat dan Ibunda Zalinar yang telah mengusahakan sekuat tenaga untuk peneliti menempuh pendidikan tinggi. Ayah, ibu, abang, kakak, dan adikku yang tersayang, terima kasih atas segala sumber kekuatan, setiap do'a, motivasi, semangat, kasih sayang serta dukungan yang kalian berikan kepada saya. Terima kasih atas segala sesuatu yang telah Ayah dan ibu berikan yang tak terhitung jumlahnya, Nyawaku nyala karenamu. Semoga Allah Subhanahu Wa Ta'ala membalas segala kebaikan dengan keberkahan yang melimpah

Peneliti juga ingin menyampaikan ucapan terima kasih juga kepada seluruh Dosen Program Studi Sistem Informasi yang telah memberikan ilmu dan pengalaman berharga selama masa studi. Semoga Allah Subhanahu Wa Ta'ala membalas segala kebaikan dengan keberkahan yang melimpah. Wassalamu'alaikum Warahmatullahi Wabarakatuh.

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

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

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

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

**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 menghargai 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, bersyukur kehadirat Allah Subhanahu Wa Ta'ala atas segala rahmat dan karunia-Nya sehingga peneliti dapat menyelesaikan Tugas Akhir ini. Shalawat serta salam kitaucapkan kepada Nabi Muhammad Shallallahu 'Alaihi Wa Sallam dengan mengucapkan Allahumma Sholli'Ala Sayyidina Muhammad Wa'Ala Ali Sayyidina Muhammad. Tugas Akhir ini dibuat sebagai salah satu syarat untuk mendapatkan gelar Sarjana Komputer di Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Pada penulisan Tugas Akhir ini, banyak pihak yang sudah berkontribusi dan mendukung peneliti. Maka dari itu, ungkapan terimakasih Peneliti ucapan 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 Penguji I yang telah memberikan arahan, nasihat, masukan, serta motivasi dalam penyelesaian Tugas Akhir ini.
5. Bapak M. Afdal, ST., M.Kom sebagai Koordinator Tugas Akhir Program Studi Sistem Informasi yang telah banyak memberikan arahan, masukan, dan motivasi selama perkuliahan serta dalam penulisan Laporan Tugas Akhir ini.
6. Bapak Saide, S.Kom., M.Kom., ICTPM., M.IM., Ph.D. Dosen pembimbing Akademik, yang telah memberikan bimbingan, nasihan, dan arahan kepada peneliti dalam menjalani proses studi di Sistem Informasi dengan terarah.
7. Bapak Inggh Permana, ST., M.Kom. sebagai Dosen Pembimbing sejak Kerja Praktek hingga Tugas Akhir yang telah meluangkan waktu, dan membimbing peneliti hingga peneliti dapat menyelesaikan Laporan Tugas Akhir ini.
8. Ibu Febi Nur Salisah, S.Kom., M.Kom. sebagai Dosen Penguji II yang telah memberikan arahan, nasihat, masukan, dan motivasi dalam penyelesaian Tugas Akhir ini.

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. Bapak Muhammad Jazman S.Kom., M.infosys sebagai Ketua Sidang yang telah memberikan arahan, nasihat, masukan, dalam penyelesaian Tugas Akhir ini.
10. Bapak dan Ibu Dosen Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau yang senantiasa memberikan ilmu yang bermanfaat serta memberikan semangat kepada peneliti.
11. Teristimewa untuk keluarga peneliti yang tercinta yakni ayahanda, ibunda, Abang, kakak, dan adik tersayang yang selalu memberikan doa, nasihat, motivasi, support, dan semangat sehingga peneliti dapat menyelesaikan laporan Tugas Akhir ini.
12. Sahabat peneliti sejak awal perkuliahan di grup Girl Squad dan Terserah yaitu Rifsya, irma, farah, ara, mutiara, emil dan endang yang selalu mendukung dan menyemangati peneliti.
13. Partner peneliti Haikal, Faisal, Ilna, Amelia, Gema, yang membantu peneliti dan memberikan saran dalam menyelesaikan Laporan Tugas Akhir.
14. Teman-teman seperjuangan Angkatan 2022, dan SIF D terima kasih telah banyak memberikan pengalaman selama perkuliahan.
15. Semua pihak yang telah mendukung peneliti yang tidak dapat disebutkan satu persatu dalam menyelesaikan tugas akhir ini. Semoga segala doa dan dukungan yang telah diberikan selama ini menjadi amal kebajikan dan mendapat balasan setimpal dari Allah Subhanahu Wa Ta'ala.

Peneliti menyadari bahwa penulisan Tugas Akhir ini masih banyak terdapat kekurangan dan jauh dari kata sempurna. Untuk itu, kritik dan saran atau pertanyaan dapat diajukan melalui e-Mail 12250323758@students.uin-suska.ac.id. Semoga laporan ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya. *Aamiin.*

Wassalamu 'alaikum Warahmatullahi Wabarakatuh

Pekanbaru, 15 Januari 2026

Penulis,

RAHMA DEVI

NIM. 12250323758

PREDIKSI RISIKO KESEHATAN PADA IBU HAMIL MENGGUNAKAN ALGORITMA MACHINE LEARNING

RAHMA DEVI
NIM: 12250323758

Tanggal Sidang: 08 Januari 2026
Periode Wisuda:

Program Studi Sistem Informasi

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. Soebrantas, No. 155, Pekanbaru

ABSTRAK

Kehamilan merupakan periode yang membutuhkan pemantauan kesehatan secara intensif guna mencegah berbagai risiko yang dapat membahayakan ibu maupun janin. Tingginya angka kematian ibu di Indonesia menunjukkan pentingnya upaya deteksi dini terhadap risiko kehamilan untuk meningkatkan kualitas pelayanan kesehatan. Teknologi *machine learning* menawarkan pendekatan prediktif yang efektif dalam mendukung proses identifikasi risiko secara lebih akurat dan cepat. Penelitian ini bertujuan untuk membandingkan performa lima algoritma *machine learning*, yaitu Logistic Regression, Decision Tree c4.5, Random Forest, Support Vector Machine (SVM), dan Naive Bayes dalam memprediksi tingkat risiko kehamilan. Metode yang digunakan adalah hold-out validation dengan proporsi data 80% untuk pelatihan dan 20% untuk pengujian. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dengan akurasi sebesar 93.42%, serta nilai presisi 93.37, recall 93.42%, dan F1-score yang juga mencapai 93.30%. Decision Tree c4.5 berada pada peringkat kedua dengan akurasi 93.15%, sementara SVM memperoleh akurasi 81.92%, Logistic Regression 75.89%, dan Naive Bayes 72.33%. Temuan ini menegaskan bahwa Random Forest merupakan algoritma yang paling optimal untuk prediksi risiko kehamilan. Penelitian ini diharapkan dapat menjadi dasar bagi pengembangan sistem pendukung keputusan berbasis *machine learning* guna meningkatkan efektivitas pelayanan kesehatan bagi ibu hamil.

Kata Kunci: Klasifikasi, Kesehatan ibu, *Machine learning*, Risiko kehamilan, *Random Forest*,



PREDIKSI RISIKO KESEHATAN PADA IBU HAMIL MENGGUNAKAN ALGORITMA MACHINE LEARNING

**RAHMA DEVI
NIM: 12250323758**

*Date of Final Exam: January 08th 2026
Graduation Period:*

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

ABSTRACT

Pregnancy is a period that requires intensive health monitoring to prevent various risks that may endanger both the mother and the fetus. The high maternal mortality rate in Indonesia underscores the importance of early detection of pregnancy risks to improve the quality of healthcare services. Machine learning technology offers an effective predictive approach to support accurate and rapid risk identification. This study aims to analyze and compare the performance of five machine learning algorithms, namely Logistic Regression, Decision Tree C4.5, Random Forest, Support Vector Machine (SVM), and Naive Bayes in predicting pregnancy risk levels using the Maternal Health Risk Dataset. The hold-out validation method was applied with an 80% training and 20% testing data split. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. The results indicate that the Random Forest algorithm achieved the best performance with an accuracy of 93.42%, along with precision 93.37%, recall 93.42%, and F1-score 93.30%. The Decision Tree C4.5 algorithm ranked second with an accuracy of 93.15%, while SVM achieved 81.92%, Logistic Regression 75.89%, and Naive Bayes 72.33%. These findings confirm that Random Forest is the most optimal algorithm for predicting pregnancy risks. This research is expected to serve as a foundation for the development of decision support systems based on machine learning to enhance the effectiveness of maternal healthcare services.

Keywords: Classification, Maternal Health, Machine Learning, Pregnancy Risk, Random Forest.

Hak Cipta Dilindungi Undang-Undang

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

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

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

DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xvi
DAFTAR SINGKATAN	xvii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Batasan Penelitian	4
1.4 Tujuan	5
1.5 Manfaat	5
1.6 Sistematika Penulisan	5
LANDASAN TEORI	7
2.1 Kehamilan dan Resiko Kehamilan	7
2.2 Angka Kematian Ibu	8
2.3 Deteksi Dini Risiko Kehamilan	11
2.4 Kecerdasan Buatan (<i>Artificial Intelligence</i>) dan <i>Machine Learning</i>	13
2.5 <i>Machine Learning</i> dalam Prediksi Risiko Kehamilan	14

2.6	Data Mining	15
2.7	Algoritma <i>Machine Learning</i> Untuk Klasifikasi	16
2.7.1	Logistic Regression	17
2.7.2	Decision Tree C4.5	18
2.7.3	Random Forest (RF)	19
2.7.4	Support Vector Machine (SVM)	20
2.7.5	Naive Bayes (NB)	21
2.8	Preprocessing Data	22
2.9	Matrix Evaluasi	23
2.10	Google Collab	25
2.11	Penelitian Tedahulu	26
3	METODOLOGI PENELITIAN	29
3.1	Metodologi Penelitian	29
3.2	Tahap Perencanaan	29
3.2.1	Identifikasi Masalah	30
3.2.2	Perumusan Tujuan Penelitian	31
3.2.3	Penetapan Batasan Penelitian	31
3.2.4	Studi Pustaka	32
3.3	Pengumpulan Data	32
3.3.1	Identifikasi Sumber Data	33
3.3.2	Deskripsi Dataset	33
3.3.3	Teori Atribut Untuk Prediksi Risiko Kesehatan Kehamilan .	34
3.4	Preprocessing Data	39
3.5	Pemodelan	39
3.5.1	Dataset	40
3.5.2	Data Understanding (Eksplorasi Data)	40
3.5.3	Pemeriksaan Data (<i>Missing Value</i>)	41
3.5.4	Preprocessing Data	41
3.5.5	Data Spliting	42
3.5.6	Pemodelan Algortima	42
3.5.7	Hasil Klasifikasi Model	52
3.5.8	Evaluasi Model	52
3.6	Seleksi Model Terbaik	53

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, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak menghargai 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.

ANALISIS DAN HASIL	54
4.1 Proses Pengumpulan Data	54
4.2 Tabel dataset	54
4.3 Data Understanding	55
4.3.1 Analisis Struktur dan Tipe Data	55
4.3.2 Statistik Deskriptif Variabel	56
4.3.3 Pemeriksaan Nilai Hilang (<i>Missing Values</i>)	58
4.3.4 Identifikasi Kelas Variabel Target (<i>RiskLevel</i>)	58
4.4 <i>Preprocessing</i> Data	59
4.4.1 Encoding Variabel Target (<i>RiskLevel</i>)	59
4.4.2 Pemisahan Data Fitur dan Variabel Target	59
4.4.3 Normalisasi Z-Score (<i>StandardScaler</i>)	60
4.4.4 Verifikasi Statistik Hasil Normalisasi	61
4.5 Pembagian Data	61
4.6 Pemodelan Machine Learning	62
4.6.1 Hasil Klasifikasi Pengukuran Performa dari Setiap Algoritma	62
4.7 Evaluasi Model	73
4.7.1 Evaluasi Performa Model Machine Learning (Per Kelas dan Weighted Average)	77
4.7.2 Hasil Evaluasi Model Algoritma <i>Machine Learning</i>	79
4.8 Hasil Seleksi Model Terbaik	82
PENUTUP	84
5.1 Kesimpulan	84
5.2 Saran	85

DAFTAR PUSTAKA

LAMPIRAN A	SURAT IZIN PENELITIAN	A - 2
LAMPIRAN B	BERITA ACARA DAN PERTANYAAN WAWANCARAB - 1	
LAMPIRAN C	SAMPLE DATA MENTAH	C - 1
LAMPIRAN D	PENELITIAN TERDAHULU	D - 1

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh

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

DAFTAR GAMBAR

2.1	Tren Angka kematian ibu Global 2000-2023	10
2.2	Tren Angka Kematian Ibu di Indonesia 2000–2020.	11
2.3	Perawatan antenatal	12
2.4	kontak ibu hamil dengan tenaga kesehatan	13
2.5	Diagram alur proses KDD	16
2.6	Contoh Visualisasi <i>Logistic Regresion</i> Sumber: (Scipython, 2019).	18
2.7	Contoh Visualisasi <i>Random Forest</i>	20
2.8	Contoh Visualisasi SVM	21
2.9	Contoh Visualisasi <i>Naive Bayes</i>	22
2.10	Tampilan Antarmuka Google Collab	26
2.11	Hasil akurasi tertinggi dari penelitian terdahulu	27
3.1	Alur metodologi penelitian	29
3.2	Alur pemodelan prosedur data	40
3.3	Alur flowchart Algoritma <i>Logistic Regresion</i>	44
3.4	Alur Flowchart Algoritma <i>Decision Tree C4.5</i>	46
3.5	Alur flowchart Algoritma Random Forest	48
3.6	Alur flowchart Algoritma Support Vector Machine	49
3.7	Alur flowchart Algoritma <i>Naive Bayes</i>	51
4.1	Confusion Matrix <i>Random forest</i>	73
4.2	Confusion Matrix <i>Support Vector Machine</i>	74
4.3	Confusion Matrix <i>Logistic Regression</i>	75
4.4	Confusion Matrix <i>Naive Bayes</i>	76
4.5	Confusion <i>Decision Tree C4.5</i>	77
4.6	Hasil Evaluasi Akurasi dari 5 Algoritma <i>Machine Learning</i>	79
4.7	Hasil Evaluasi Presisi dari 5 Algoritma <i>Machine Learning</i>	80
4.8	Hasil Evaluasi F1-Score dari 5 Algoritma <i>Machine Learning</i>	80
4.9	Hasil Evaluasi Recall dari 5 Algoritma <i>Machine Learning</i>	81
4.10	Hasil Seleksi Model Terbaik berdasarkan <i>Accuracy, precision, Recall</i> , dan <i>F1-Score</i> dari 5 Algoritma <i>Machine Learning</i>	82
A.1	Surat Permohonan Izin Peneelitian dari Fakultas	A - 2
A.2	Balasan Surat izin Penelitian DINKES rekomendasi wawancara ke puskesmas simpang baru	A - 3

Hak Cipta Dilindungi Undang-Undang

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

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh kary

ra tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

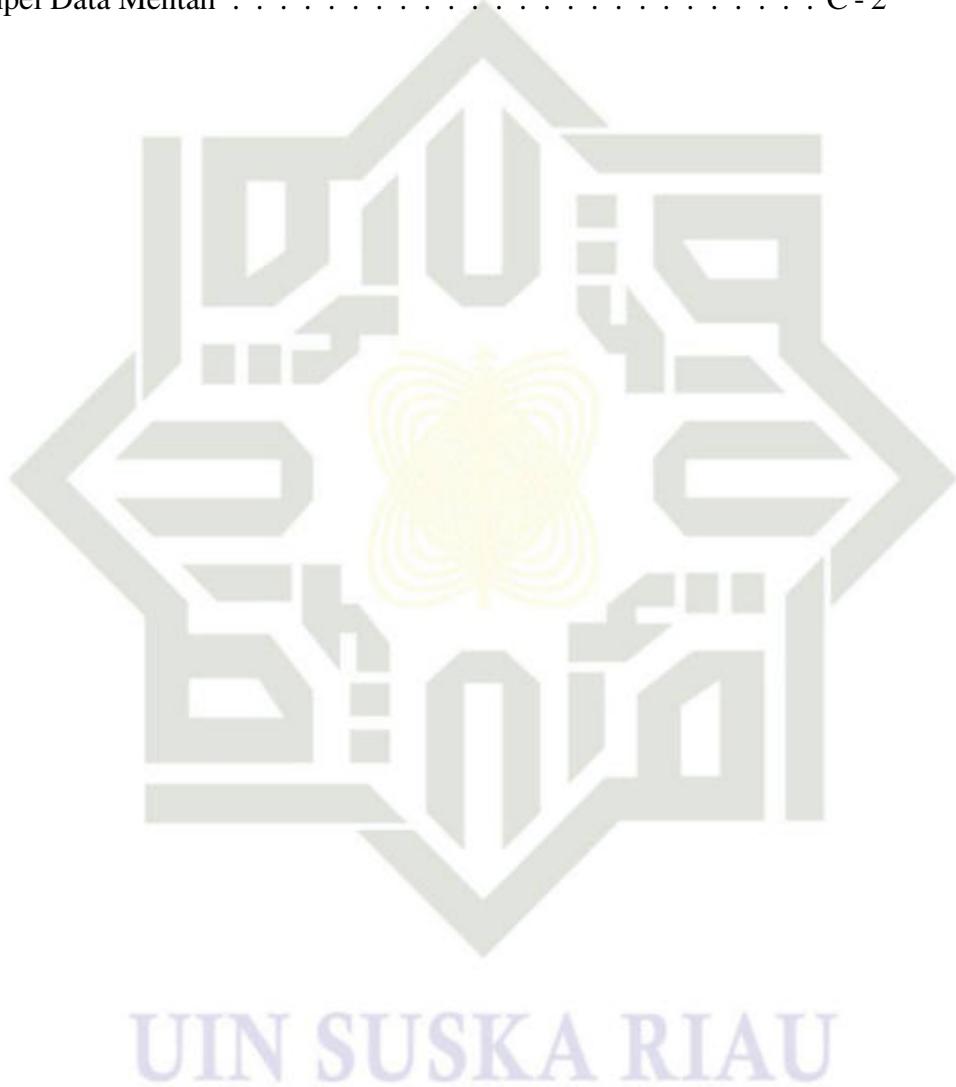


UIN SUSKA RIAU

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.

B.1	Berita wawancara	B - 1
B.2	Pertanyaan Wawancara	B - 2
B.3	Pertanyaan Wawancara (2)	B - 3
B.4	Pertanyaan Wawancara (3)	B - 4
B.5	Pertanyaan Wawancara (4)	B - 5
C.1	Sampel Data Mentah	C - 1
C.2	Sampel Data Mentah (2)	C - 1
C.3	Sampel Data Mentah	C - 2



DAFTAR TABEL

3.1	Penjelasan Atribut Penelitian	34
3.2	Teori Atribut Penelitian	35
3.3	Parameter Tiap Algoritma	43
4.1	Jumlah Data Risk Level	54
4.2	Dataset Kehamilan	54
4.3	Informasi Dataset Kehamilan	55
4.4	Statistik Deskriptif Variabel	56
4.5	Pemeriksaan Missing Values	58
4.6	Distribusi Kategori Risiko	59
4.7	Data Fitur (X) Sebelum Normalisasi	60
4.8	Data Fitur (X) Setelah Normalisasi Z-Score	60
4.9	Verifikasi Statistik Hasil Normalisasi	61
4.10	Pembagian Data	62
4.11	Hasil Evaluasi Algoritma <i>Random Forest</i>	63
4.12	Hasil Pengukuran Performa Random Forest Parameter Terbaik	65
4.13	SVM dengan Pengujian Beberapa Parameter	66
4.14	Hasil Pengukuran Performa SVM Parameter Terbaik	68
4.15	Logistic Regression dengan Pengujian Beberapa Parameter	69
4.16	Hasil Pengukuran Performa <i>Logistic Regression</i> Parameter Terbaik	70
4.17	Naive Bayes dengan Pengujian Beberapa Parameter	70
4.18	Hasil Pengukuran Performa <i>Naive Bayes</i> Parameter Terbaik	71
4.19	<i>Decision Tree C4.5</i> dengan Pengujian Beberapa Parameter	71
4.20	Hasil Pengukuran performa Decision tree C4.5 parameter terbaik	72
4.21	Evaluasi Performa Model per Kelas dan <i>Weighted Avg</i>	78
4.22	Hasil Pemilihan Model Terbaik dari 5 Algoritma	82
D.1	Penelitian Terdahulu	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 SINGKATAN

Age	: Usia ibu
AKI	: Angka Kematian Ibu
ANC	: <i>Antenatal Care</i>
BodyTemp	: Suhu Tubuh
BS	: Kadar Gula Darah
DiastolicBP	: Tekanan darah diastolik
DT	: <i>Decision Tree C4.5</i>
F1-Score	: <i>F1 Measure Score</i>
FN	: <i>False Negative</i>
FP	: <i>False Positive</i>
HeartRate	: Detak Jantung
LR	: <i>Logistic Regression</i>
MMR	: <i>Maternal Mortality Ratio</i>
NB	: <i>Naive Bayes</i>
RF	: <i>Random Forest</i>
RiskLevel	: Kelas Target
SDGs	: <i>Sustainable Development Goals</i>
SVM	: <i>Support Vector Machine</i>
SystolicBP	: Tekanan darah sistolik
TN	: <i>True Negative</i>
TP	: <i>True Positive</i>
UNICEF	: <i>United Nations Children's Fund</i>
UNFPA	: <i>United Nations Population Fund</i>
WHO	: <i>World Health Organization</i>

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, pendekatan *konvensional* yang hanya menganalisa pemeriksaan manual dan penilaian subjektif dinilai kurang optimal dalam memprediksi potensi komplikasi, terutama ketika pola risiko bersifat kompleks dan
 - b. Pengutipan tidak mengutip sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin 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

Kehamilan merupakan fase krusial dalam hidup seorang perempuan yang menuntut perhatian serius terhadap kesehatan dirinya dan janin yang dikandung. Namun demikian, gangguan hipertensi pada kehamilan, seperti preeclampsia, dan tekanan darah tinggi (*hipertensi gestasional*), merupakan penyebab utama kematian ibu terutama khususnya di negara-negara berkembang. Dengan sekitar 46.000 kematian per tahun. Secara global, Menurut data WHO tahun 2023, isu terkait kehamilan dan persalinan bertanggung jawab atas sekitar 260.000 kasus kematian wanita setiap tahun, dan sebagian besar tragedi ini terjadi di negara-negara tersebut (WHO, 2023a). Di Indonesia sendiri, Angka Kematian Ibu (AKI) masih tergolong tinggi, mencapai 189 per 100.000 kelahiran hidup menurut kementerian kesehatan RI berdasarkan Long Form Sensus Penduduk 2023 (Sadikin, 2023). Angka ini menunjukkan tantangan besar karena masih jauh melampaui target *Sustainable Development Goals* (SDGs), yang menargetkan maksimal 70 kematian ibu per 100.000 kelahiran hidup pada tahun 2030.(Sadikin, 2023)

Risiko-risiko ini tidak hanya mengancam nyawa ibu, tetapi juga dapat menyebabkan komplikasi serius seperti kelahiran prematur, gangguan perkembangan bayi, cacat lahir, atau bahkan kematian perinatal (WHO, 2023b). Seringkali, komplikasi kehamilan tidak terdeteksi sejak dini (WHO, UNICEF, UNFPA dan Division, 2025). Menimbulkan trauma psikologis jangka panjang bagi ibu dan keluarga (An, Rahman, Zhou, dan Kang, 2023) Kehamilan yang tidak terpantau dengan baik ini membawa risiko besar, sehingga pendekatan dini menjadi langkah yang sangat penting guna memastikan ibu hamil mendapatkan penanganan yang cepat dan akurat. (Hasan, Shih, Devnath, Eti, dan Rezaul, 2024) (Khoirunnisa dan Lestari, 2023).

Untuk mengatasi permasalahan ini, berbagai solusi telah diusulkan, mulai dari peningkatan mutu layanan antenatal dan pemantauan kondisi ibu hamil secara berkala untuk memastikan deteksi dini faktor risiko (WHO, 2016). Upaya ini sangat penting karena ketepatan waktu dalam mengidentifikasi tanda bahaya berperan besar dalam menurunkan angka komplikasi kehamilan (Lestari, Maulana, Persada, dan Adi, 2023). Namun demikian, pendekatan *konvensional* yang hanya menganalisa pemeriksaan manual dan penilaian subjektif dinilai kurang optimal dalam memprediksi potensi komplikasi, terutama ketika pola risiko bersifat kompleks dan

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.

tidak linear. Oleh karena itu, pendekatan yang semakin populer dan dianggap menjanjikan adalah pemanfaatan metode machine learning sebagai alat bantu deteksi dan klasifikasi risiko kehamilan, mengingat kemampuannya dalam meningkatkan akurasi prediksi berdasarkan data klinis. (Modi dan Kumar, 2025)

Dalam penerapan metode *machine learning* untuk mendeteksi risiko kehamilan, diperlukan pemilihan algoritma yang mampu memberikan hasil klasifikasi dengan tingkat akurasi terbaik. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi dan membandingkan kinerja beberapa algoritma *machine learning* guna menentukan model yang paling optimal dalam memprediksi tingkat risiko kehamilan. Berbagai algoritma *Machine Learning*, seperti *logistic regression*, *Decision tree C4.5*, *Random forest*, *Support Vector Machine* (SVM), dan *Naive Bayes*, dapat diterapkan untuk mengklasifikasikan risiko kehamilan menjadi kategori rendah, sedang, dan tinggi, dengan menjadikan parameter kesehatan ibu hamil sebagai dasarnya. (Hennessy, Tran, Sasikumar, dan Al-Falahi, 2024), (S. A. Rahman, Ibtilum, Bazgir, dan Barai, 2023). Pemanfaatan teknologi Machine Learning dalam proses pengelompokan klasifikasi risiko kehamilan menawarkan solusi yang sangat menjanjikan. Berbagai data vital ibu, seperti usia, suhu tubuh, denyut jantung, kadar hemoglobin, kadar gula darah, serta tekanan darah (sistolik dan diastolik), dapat diolah dan dimodelkan oleh algoritma ini. Salah satu sumber data publik yang ideal untuk penelitian ini adalah Dataset Risiko Kesehatan Ibu Hamil *Maternal Health Risk Dataset* dari Kaggle, yang berisi 1.822 data yang sudah dilabeli sebagai risiko rendah, sedang, dan tinggi. Pemeriksaan antenatal dan terbukti mampu menghasilkan model klasifikasi risiko yang akurat. (Tomar, Sharma, dan Prasad, 2024).

Penelitian-penelitian terdahulu telah membuktikan bahwa model *machine learning* sangat bermanfaat dalam mengidentifikasi risiko kehamilan secara efektif. Studi yang dilakukan oleh R. A. F. Rahman, Harahap, dan Kurnia (2023) telah menunjukkan potensi tinggi dari algoritma *Decision Tree* dalam klasifikasi risiko. Dalam sebuah penelitian yang berfokus pada klasifikasi tingkat risiko ibu hamil, algoritma tersebut berhasil mencapai tingkat akurasi tertinggi sebesar 90%. Sementara itu, metode *Support Vector Machine*(SVM) juga mencatatkan akurasi yang *impresif*, yaitu sebesar 90.9%. (Raja, Mukherjee, dan Sarkar, 2021). Penelitian lain oleh Assaduzzaman, Salim, Tafakori, dan Abdollahian (2024) algoritma Random Forest terbukti menunjukkan akurasi terbaik, mencapai 75,2% dalam klasifikasi risiko *maternal*. Penelitian tersebut juga menyimpulkan bahwa kadar gula darah menjadi faktor yang paling signifikan memengaruhi hasil klasifikasi, diikuti

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.

oleh usia ibu. Penelitian lain oleh Setiawan, Kurniawan, dan Prasetyo (2024) algoritma *Random forest* terbukti menunjukkan akurasi terbaik, yaitu 82%. Lebih lanjut, dalam jurnal *Artificial Intelligence in Medicine* Hennessy dkk. (2024), Model klasifikasi yang menggunakan algoritma gabungan, termasuk *Random Forest*, menunjukkan kemampuan untuk memprediksi komplikasi kehamilan dengan tingkat keakuratan mencapai 76%. Hasil prediksi ini semakin diperkuat oleh temuan serupa dari penelitian lain oleh Zhou dkk. (2024a), Du dkk. (2023) yang menyimpulkan bahwa *Random Forest* dan model kelompok pembelajaran ensemble learning merupakan pilihan ideal karena menawarkan keseimbangan optimal antara akurasi dan efisiensi waktu komputasi. Sementara itu penelitian dari Farida, Tiasti, dan Sari (2023) metode *Logistic Regresion* mencatat akurasi 79,6. Selain itu, algoritma *Naive Bayes* dengan akurasi 85,62% juga menunjukkan performa yang memuaskan saat digunakan untuk memprediksi kehamilan, terutama karena waktu eksekusinya yang cepat, (Khoirunnisa dan Lestari, 2023).

Berdasarkan tinjauan literatur, dapat disimpulkan bahwa berbagai algoritma *machine learning* memiliki keunggulan masing-masing dalam mendeteksi risiko kehamilan. Beberapa studi menunjukkan bahwa *Decision Tree* dan *SVM* memberikan akurasi tinggi, yaitu sekitar 90% dan 90,9%. *Random Forest* juga menunjukkan performa yang konsisten dengan akurasi 75,2%–82%, terutama ketika faktor seperti kadar gula darah dan usia ibu menjadi faktor paling berpengaruh. Oleh sebab itu, variabel-variabel kunci ini mutlak harus dimasukkan dan dipertimbangkan dalam pengembangan model prediksi. Selain itu, penelitian lain melaporkan bahwa model berbasis ensemble, termasuk *Random Forest*, mencapai akurasi sekitar 76% (Hennessy dkk., 2024), sedangkan *Naive Bayes* memperoleh akurasi sebesar 85,62%. (Khoirunnisa dan Lestari, 2023).

Penelitian terdahulu menunjukkan bahwa *machine learning* memiliki potensi besar dalam membantu prediksi risiko kehamilan secara lebih awal. Meskipun berbagai algoritma telah terbukti efektif, masih terdapat beberapa kesenjangan yang perlu dijembatani. Salah satu penelitian terbaru tahun 2024 oleh Setiawan dkk. (2024) juga menyarankan agar penelitian selanjutnya dapat menguji algoritma pembelajaran mesin non-parametrik lainnya seperti *Support Vector Machine*, maupun algoritma parametrik seperti *Logistic Regression*. Pertama, sebagian besar penelitian masih menitikberatkan pada akurasi sebagai ukuran utama performa model, tanpa mempertimbangkan metrik lain seperti *precision*, *recall*, dan *F1 score*, yang penting dalam konteks data medis dengan kemungkinan ketidakseimbangan kelas. Kedua, tidak banyak penelitian yang membandingkan secara

menyeluruh lima algoritma utama *Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *Support Vector Machine*, dan *Naive Bayes* dengan metrik evaluasi yang lengkap. Ketiga, prediksi belum cukup mewakili populasi yang lebih luas karena beberapa penelitian masih menggunakan dataset berukuran kecil. Jadi, penelitian ini dilakukan untuk menyelesaikan masalah ini.

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma *machine learning* dalam memprediksi risiko kehamilan. Lima algoritma yang akan dianalisis adalah *Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *Support Vector Machine* (SVM), dan *Naive Bayes*, dalam memprediksi risiko menggunakan *Maternal Health Risk Dataset* dari *kaggle*. Evaluasi dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk menentukan algoritma yang memberikan prediksi paling optimal. Diharapkan, penelitian ini dapat memberikan gambaran yang jelas mengenai algoritma *machine learning* yang paling efektif untuk memprediksi risiko kehamilan, serta menjadi acuan atau referensi bagi penelitian atau studi lanjutan di kesehatan ibu hamil.

1.2 Perumusan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang telah dipaparkan, rumusan masalah dari penelitian ini adalah Algoritma *Machine Learning* manakah yang memberikan hasil akurasi paling optimal dalam memprediksi tingkat risiko kesehatan pada ibu hamil?

1.3 Batasan Penelitian

Pembahasan dalam penelitian Tugas Akhir ini akan dibatasi pada aspek-aspek berikut:

1. Penelitian ini menggunakan data sekunder dari dataset *Maternal Health Risk Dataset* data yang digunakan 1.822 yang tersedia di platform Kaggle milik Joakim Arvidsson, Dr. Mb Sharma dan juga sudah pernah digunakan sebelumnya oleh Setiawan dkk. (2024).
2. Atribut yang dianalisis adalah usia, suhu tubuh, denyut jantung, kadar gula darah, serta tekanan darah sistolik dan diastolik. Kategori terbatas pada tiga kelas, yaitu *high risk*, *low risk*, dan *mid risk*.
3. Penelitian ini berfokus pada perbandingan lima algoritma *machine learning* yaitu *Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *Support Vector Machine* (SVM), dan *Naive Bayes*, tanpa melakukan pengembangan atau implementasi sistem.
4. Penelitian ini menggunakan metode *hold-out validation* pada rasio 80% data

latih (*training data*) dan 20% data uji (*testing data*).

5. Penelitian ini menggunakan metode *hold-out validation* pada rasio 80% data latih (*training data*) dan 20% data uji (*testing data*).
6. Evaluasi performa model dilakukan menggunakan metrik akurasi dan *confusion matrix* yang terdiri dari *accuracy*, *precision*, *recall*, dan *F1-score*.
7. Hasil penelitian ini hanya ditujukan untuk keperluan analisis akademik dan prediksi risiko kehamilan secara umum, bukan sebagai alat diagnosis medis.
8. Proses analisis dan pelatihan model dilakukan menggunakan bahasa pemrograman *Python* melalui platform *Google Colab*.

1.4 Tujuan

Tujuan tugas akhir ini adalah:

1. Menganalisis dan membandingkan lima algoritma machine learning yaitu *logistic Regression*, *Decision C4.5*, *Random Forest*, *Support Vector Machine* (SVM), dan *Naive Bayes*.
2. Menentukan algoritma yang memberikan hasil akurasi paling optimal dalam memprediksi tingkat risiko kehamilan berdasarkan metrik evaluasi akurasi, *precision*, *recall*, dan *F1-score*.

1.5 Manfaat

Manfaat tugas akhir ini adalah:

1. Penelitian ini diharapkan dapat menambah wawasan dalam memprediksi risiko kehamilan dengan penerapan model *Machine Learning*.
2. Dengan membandingkan kinerja lima algoritma *Logistic Regression*, *Decision C4.5*, *Random Forest*, *SVM*, dan *Naive Bayes*, penelitian ini dapat mengidentifikasi model yang paling optimal. Hasil ini dapat membuka penelitian terbaru kedepannya guna menentukan akurasi model yang lebih baik dari sebelumnya.
3. Menyediakan referensi atau rekomendasi algoritma *Machine Learning* untuk memprediksi risiko kehamilan.
4. Menjadi bahan pertimbangan bagi praktisi kesehatan atau peneliti data dalam melakukan prediksi dini risiko kehamilan berbasis data.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

BAB 1. PENDAHULUAN

Pada bab ini memberikan gambaran mengenai latar belakang penelitian

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.

yang menguraikan kondisi dan permasalahan terait Risiko kesehatan ibu hamil. Bab ini juga memuat perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan sebagai arah pelaksanaan penelitian.

BAB 2. LANDASAN TEORI

Pada bab ini menjelaskan mengenai teori dan konsep yang mendukung penelitian, meliputi pengertian kehamilan dan risiko kehamilan, angka kematian ibu, deteksi dini, *machine learning*, data mining, serta penjelasan algoritma. Selain itu, bab ini juga menyajikan penelitian terdahulu yang relevan sebagai dasar pembanding dan penguat penelitian.

BAB 3. METODOLOGI PENELITIAN

Pada bab ini menjelaskan tahap penelitian yang dilakukan, mulai dari tahap perencanaan, teknik pengumpulan data, tahap pra-pemrosesan data, menguraikan parameter yang digunakan, hingga penerapan Algoritma dalam memprediksi risiko kesehatan ibu hamil.

BAB 4. HASIL ANALISA

Pada bab ini menyajikan hasil yang telah dilakukan dari proses pengumpulan data, pengolahan data dan analisis data. Menyajikan hasil klasifikasi dari lima algoritma yang digunakan yaitu *Random Forest*, *Decisionree C4.5*, *SupportVector Machine*, *Logistic Regresion* dan *Naive Bayes*. Serta analisis hasil seleksi model terbaik berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-Score*.

BAB 5. PENUTUP

Pada bab ini berisikan tentang kesimpulan yang merangkum hasil penelitian berdasarkan tujuan yang telah ditetapkan, dan saran yang dapat dijadikan sebagai acuan untuk pengembangan penelitian selanjutnya.

BAB 2

LANDASAN TEORI

2.1 Kehamilan dan Resiko Kehamilan

Kehamilan adalah proses *fisiologis* yang wajar dialami oleh wanita usia subur, ditandai dengan tumbuh kembangnya janin di dalam rahim. Meskipun ini merupakan peristiwa alami, kehamilan berpotensi menimbulkan risiko kesehatan yang serius jika tidak mendapatkan pemantauan yang memadai. Menurut Bisson dkk. (2023), Kehamilan memicu perubahan *fisiologis* yang signifikan pada sistem kardiovaskular dan metabolisme ibu. Adaptasi ini sangat penting untuk memastikan janin mendapatkan oksigen dan nutrisi yang memadai sekaligus menjaga keseimbangan tubuh (homeostasis). Penyesuaian ini melibatkan peningkatan *volume plasma*, modifikasi tekanan darah, serta perubahan respons hormonal dan vaskular. Sayangnya, kegagalan dalam proses adaptasi ini dapat memicu berbagai komplikasi serius, seperti preeklamsia, hipertensi gestasional, dan gangguan metabolismik, yang berpotensi mengancam keselamatan ibu dan janin. Oleh karena itu, kehamilan harus dipandang bukan sekadar proses biologis, melainkan sebagai peristiwa medis kompleks yang memerlukan pemantauan yang menyeluruh guna mencegah kesakitan (morbidity) dan kematian (*mortalitas*) maternal. (Bisson et al. 2023). Menurut catatan WHO (2023b), lebih dari 40% ibu hamil menghadapi komplikasi, dan 15% dari kasus tersebut berpotensi mengancam jiwa (Nurjanah, Nurfiti, dan Magasida, 2025).

Kegagalan untuk mendeteksi risiko kehamilan sejak awal dapat menimbulkan konsekuensi serius, seperti meningkatnya Angka Kematian Ibu (AKI). Selain itu, kondisi ini juga berpotensi menyebabkan dampak jangka panjang pada bayi, termasuk kelahiran prematur dan gangguan tumbuh kembang.(WHO, UNICEF, UNFPA dan Division, 2025). faktor-faktor yang meningkatkan risiko kehamilan yaitu dari Faktor medis meliputi usia ibu yang berada pada batas ekstrem (di bawah 20 atau di atas 35 tahun), riwayat kehamilan dan persalinan buruk (obstetri buruk), adanya penyakit kronis penyerta seperti diabetes dan hipertensi, serta berbagai kondisi *fisiologis* lain yang memengaruhi jalannya kehamilan (Zhou dkk., 2024b).

Di Indonesia, berdasarkan laporan Kementerian Kesehatan RI tahun 2023, risiko kehamilan tetap menjadi tantangan besar dalam upaya menurunkan Angka Kematian Ibu (AKI). Salah satu kendala utamanya adalah keterlambatan deteksi dini komplikasi. Keterlambatan ini disebabkan oleh berbagai faktor, termasuk rendahnya kesadaran ibu hamil untuk menjalani pemeriksaan antenatal serta keter-

- 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.

batasan tenaga kesehatan di berbagai daerah. Oleh karena itu, jelas bahwa pemanfaatan dan intervensi terhadap faktor risiko kehamilan harus dilakukan secara komprehensif, dengan melibatkan aspek medis, sosial, dan dukungan dari *Machine learning* (Sadikin, 2023). Studi oleh Amalia dkk. (2023) lebih lanjut menegaskan bahwa berbagai atribut kesehatan ibu seperti usia, tekanan darah, suhu tubuh, kadar gula darah, dan denyut jantung dapat menjadi indikator krusial dalam pengelompokan klasifikasi risiko kehamilan. Terbukti, pemanfaatan teknologi seperti *data mining* dan *Machine Learning* di bidang kesehatan reproduksi telah membantu tenaga medis memprediksi risiko kehamilan lebih awal. Hal ini memungkinkan intervensi tepat waktu untuk menekan komplikasi dan angka kematian ibu. (Amalia dkk., 2023)

Dampak dari risiko kehamilan yang tidak terdeteksi sejak dini dapat berakibat serius, baik bagi ibu maupun janin. (WHO, 2023b) melaporkan bahwa sekitar 15% kehamilan di seluruh dunia menghadapi komplikasi serius yang berpotensi mengancam jiwa. Bagi ibu, risiko ini berkorelasi dengan peningkatan angka kematian maternal, sedangkan pada janin, dampaknya dapat mencakup kelahiran prematur, berat badan lahir rendah, hingga kematian neonatal. Di Indonesia sendiri, hipertensi dalam kehamilan dan diabetes gestasional masih menjadi faktor utama tingginya angka kematian ibu (Sadikin, 2023). Hal ini menekankan perlunya deteksi dini yang dilakukan melalui pemeriksaan medis yang teratur.

Dalam konteks pelayanan antenatal (*Antenatal Care/ANC*), indikator medis seperti tekanan darah, kadar gula darah, suhu tubuh, denyut jantung, kadar hemoglobin, serta usia ibu hamil merupakan parameter klinis yang lazim digunakan untuk mendeteksi risiko maternal. Sebagai contoh, tekanan darah sistolik $\geq 140 \text{ mmHg}$ atau diastolik $\geq 90 \text{ mmHg}$ setelah usia kehamilan 20 minggu dapat mengindikasikan hipertensi gestasional (*gestational hypertension*). Pemeriksaan skrining glukosa, seperti pemeriksaan satu jam setelah konsumsi glukosa, dengan kadar gula darah $\geq 140 \text{ mg/dL}$ sering menjadi tanda awal adanya risiko diabetes gestasional. Pemeriksaan rutin terhadap indikator-indikator tersebut memungkinkan tenaga kesehatan melakukan intervensi dini sebelum risiko berkembang menjadi komplikasi kehamilan yang lebih parah.

2.2 Angka Kematian Ibu

Angka Kematian Ibu (AKI), atau dikenal juga sebagai *Maternal Mortality Ratio* (MMR), adalah indikator krusial untuk menilai kualitas pelayanan maternal dan status kesehatan masyarakat di suatu negara. Menurut WHO (2023b), MMR

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.

berfungsi mengukur risiko kematian ibu dibandingkan dengan jumlah kelahiran hidup dalam periode tertentu. Dengan demikian, AKI secara efektif mencerminkan tingkat keberhasilan sistem kesehatan dalam menyediakan layanan persalinan yang aman dan berkualitas. WHO (2023b) mendefinisikan AKI sebagai jumlah kematian ibu selama kehamilan, persalinan, atau dalam 42 hari setelah kehamilan berakhir per 100.000 kelahiran hidup. Indikator ini sangat penting karena sebagian besar kematian maternal dapat dicegah dengan pelayanan kesehatan yang tepat. Oleh sebab itu, penurunan Angka Kematian Ibu (AKI) dijadikan salah satu target utama dalam *Sustainable Development Goals* (SDG), yaitu menurunkan *MMR* global menjadi kurang dari 70 per 100.000 kelahiran hidup pada tahun 2030.(WHO, UNICEF, UNFPA dan Division, 2025).

Secara global, Angka Kematian Ibu (AKI) telah menunjukkan penurunan yang cukup signifikan dalam dua dekade terakhir. WHO, UNICEF, UNFPA dan Division (2025) melaporkan bahwa *Maternal Mortality Ratio (MMR)* dunia turun dari 339 per 100.000 kelahiran hidup pada tahun 2000 menjadi 223 pada 2020, dan diproyeksikan mencapai 197 pada tahun 2023. Meskipun ada kemajuan ini, penurunannya tidak merata. Faktanya, lebih dari 70% kematian maternal masih terpusat di kawasan Sub-Sahara Afrika dan Asia Selatan. Kesenjangan ini jelas menegaskan bahwa kematian ibu bukan semata-mata masalah klinis, melainkan sangat terkait erat dengan ketidaksetaraan akses layanan kesehatan, serta kondisi sosial-ekonomi di berbagai wilayah(WHO, 2023b).

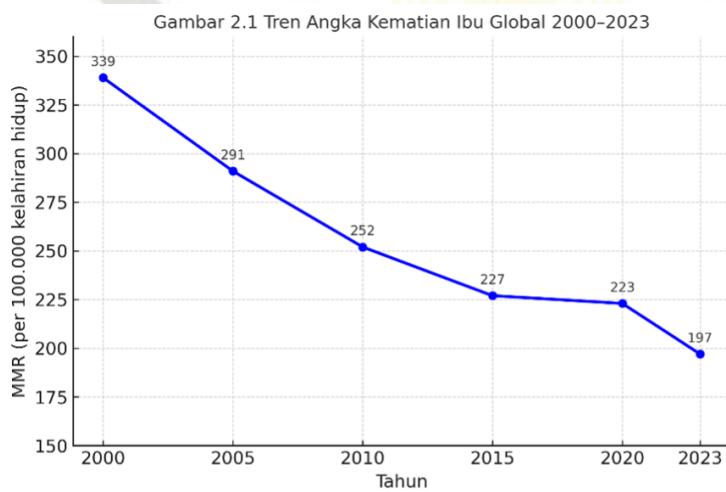
Meskipun Angka Kematian Ibu (AKI) di Indonesia menunjukkan perbaikan, angkanya masih tergolong tinggi jika dibandingkan dengan rata-rata global. Berdasarkan perkiraan gabungan dari WHO/UNICEF (*Inter-Agency Estimates*, 2023), AKI Indonesia berhasil diturunkan dari sekitar 299 per 100.000 kelahiran hidup pada tahun 2000 menjadi 189 per 100.000 pada tahun 2020 (WHO, 2023b). Penurunan ini menggambarkan adanya perbaikan akses layanan kesehatan ibu dan anak, peningkatan cakupan persalinan oleh tenaga kesehatan terlatih, serta berbagai program intervensi gizi ibu hamil. Namun, angka tersebut masih jauh dari target *SDGs* 2030. Penyebab utama kematian ibu di Indonesia hingga saat ini masih relatif konsisten, yakni perdarahan, hipertensi dalam kehamilan (preklamsia/eklamsia), serta komplikasi obstetri lainnya. Sebagian besar kasus kematian ibu di Indonesia terjadi akibat komplikasi langsung yang sebenarnya dapat dicegah melalui peningkatan kualitas layanan deteksi dini risiko kehamilan.(Syairaji dkk., 2024).

Perlu dicatat bahwa terdapat variasi data mengenai Angka Kematian Ibu (AKI) dari berbagai sumber. Sebagai contoh, *World Bank Gender Data Portal*

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.

memperkirakan MMR Indonesia pada tahun 2023 adalah 140 per 100.000 kelahiran hidup. Perbedaan angka ini timbul akibat perbedaan metode estimasi yang digunakan *WHO/UNICEF* mengandalkan model *inter-agency* berbasis data global, sedangkan *World Bank* menggunakan proyeksi statistik dengan dataset yang lebih baru. Dalam konteks penulisan akademik, sangat penting untuk mengakui adanya variasi data ini. Namun, demi menjaga konsistensi dan kesesuaian dengan standar global, penelitian ini akan tetap merujuk pada angka yang dikeluarkan oleh *WHO/UNICEF inter-agency estimates*. (Key, 2025).

Meskipun Indonesia telah mencatat kemajuan dalam upaya penurunan angka kematian ibu, hasil yang dicapai masih belum optimal. Kondisi ini menunjukkan adanya tantangan signifikan dalam peningkatan sistem pelayanan kesehatan, terutama terkait deteksi dini risiko kehamilan, pemerataan kualitas tenaga kesehatan, dan akses layanan di wilayah terpencil. Dalam kerangka penelitian ini, urgensi tersebut menjadi sangat relevan: penerapan analisis data medis menggunakan algoritma *Machine Learning* diharapkan dapat membantu mendeteksi risiko kehamilan lebih awal. Dengan demikian, hasil penelitian ini berpotensi memberikan kontribusi strategis dalam upaya nasional untuk menekan angka kematian ibu di Indonesia. Berikut Trend Angka Kematian Ibu Global tahun 2000-2023 dapat dilihat pada Gambar 2.1



Gambar 2.1. Trend Angka kematian ibu Global 2000-2023

Grafik menunjukkan penurunan global MMR dari 339 pada tahun 2000 menjadi 223 pada 2020, dan diperkirakan turun lagi menjadi 197 pada 2023. Meski menurun, angka tersebut masih cukup tinggi di kawasan tertentu, terutama Afrika Sub-Sahara dan Asia Selatan. Selanjutnya pada Gambar 2.2 Trend Angka Kematian

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.

ibu Indonesia tahun 2000-2020.

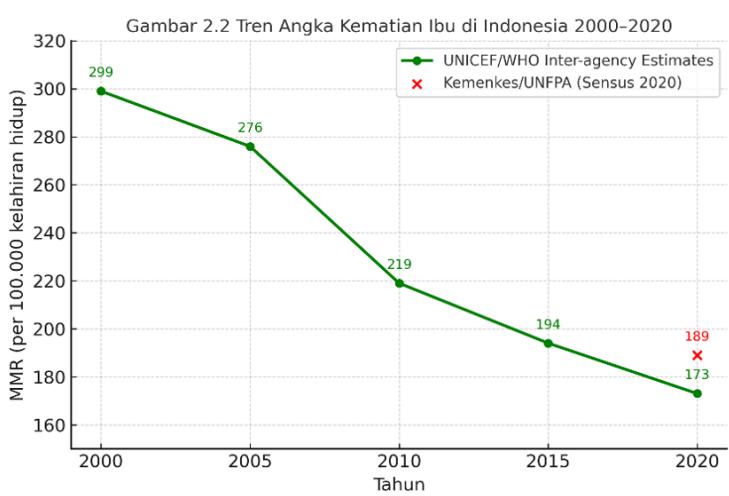
Hak Cipta Dilindungi Undang-Undang

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 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.



Gambar 2.2. Tren Angka Kematian Ibu di Indonesia 2000–2020.

Grafik memperlihatkan penurunan AKI Indonesia dari 299 (2000) menjadi 173 (estimasi WHO/UNICEF 2020). Namun, data sensus nasional yang dicatat UNFPA/Kemenkes menunjukkan angka 189 pada 2020 dan sampai 2023 juga masih memakai di angka 189. Perbedaan ini terjadi karena perbedaan metode estimasi, tetapi keduanya menegaskan bahwa AKI Indonesia masih berada jauh di atas target SDGs.

2.3 Deteksi Dini Risiko Kehamilan

Sebenarnya, risiko saat hamil dapat diketahui sejak dini melalui pemeriksaan rutin ke bidan atau puskesmas. Namun, karena keterbatasan tenaga kesehatan, jarak layanan yang jauh, dan rendahnya kesadaran masyarakat akan pentingnya pemeriksaan kehamilan, seringkali masalah kesehatan baru terdeteksi setelah terlambat. Oleh karena itu, deteksi dini risiko kehamilan menjadi salah satu strategi krusial untuk menurunkan angka kematian ibu dan bayi. Proses deteksi dini ini dilakukan melalui Pemeriksaan Kehamilan Berkala (*Antenatal Care/ANC*) yang bertujuan mengidentifikasi baik kondisi medis maupun faktor sosial yang berpotensi memicu komplikasi selama masa kehamilan dan persalinan. *Menurut World Health Organization* (WHO, 2022), pemeriksaan ANC minimal harus dilakukan delapan kali selama kehamilan agar komplikasi dapat terdeteksi lebih awal. Pemeriksaan ini mencakup evaluasi tekanan darah, kadar hemoglobin, status gizi, serta penelusuran riwayat penyakit yang dapat memengaruhi kehamilan. (WHO, 2016)

Di Indonesia, Kementerian Kesehatan RI (2023) menekankan bahwa ANC merupakan layanan esensial yang wajib diakses semua ibu hamil, terutama dalam

Hak Cipta Dilindungi Undang-Undang

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

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

- Pengutipan tidak menghargai kepentingan yang wajar UIN Suska Riau.

mendukung target penurunan Angka Kematian Ibu (AKI). Pemeriksaan ini juga mencakup deteksi tanda bahaya kehamilan seperti perdarahan, preeklamsia, anemia, dan infeksi. Melalui deteksi dini, risiko komplikasi dapat diantisipasi lebih cepat, sehingga penanganan medis dapat dilakukan tepat waktu sebelum kondisi memburuk (Sadikin, 2023). Perkembangan teknologi juga telah mendukung upaya deteksi dini risiko kehamilan. Penelitian Amalia dkk. (2023) menunjukkan bahwa penggunaan data mining dan algoritma klasifikasi, dapat membantu tenaga medis dalam memprediksi tingkat risiko ibu hamil berdasarkan atribut klinis seperti usia, tekanan darah, kadar gula darah, suhu tubuh, dan denyut jantung. Integrasi teknologi ini dapat meningkatkan akurasi deteksi dan membantu pengambilan keputusan medis secara lebih cepat (Amalia dkk., 2023).

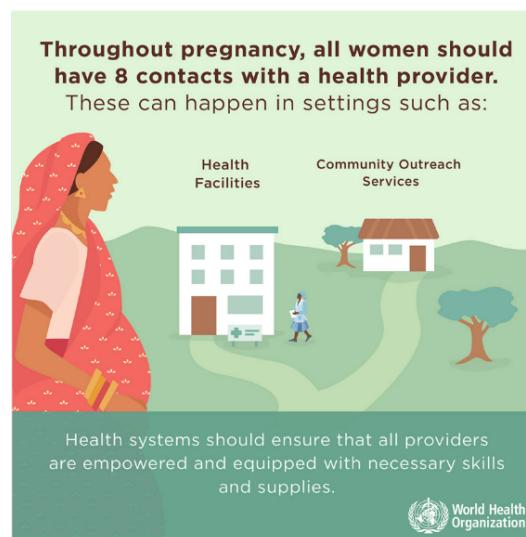
Selain pendekatan medis dan teknologi, edukasi ibu hamil dan dukungan keluarga memiliki peran yang sangat penting dalam keberhasilan deteksi dini risiko kehamilan. UNICEF (2023) secara khusus menyoroti bahwa peningkatan literasi kesehatan reproduksi terutama bagi masyarakat yang memiliki akses layanan terbatas sangatlah krusial. Tujuannya adalah memastikan ibu hamil mampu mengenali tanda-tanda bahaya dan segera mendapatkan pertolongan. Dengan kata lain, deteksi dini risiko kehamilan hanya akan berhasil melalui kombinasi dari layanan medis yang berkualitas, dukungan teknologi, dan pemberdayaan masyarakat.(WHO, UNICEF, UNFPA dan Division, 2025). Perawatan antenatal dapat dilihat pada Gambar 2.3, Gambar kontak ibu hamil dengan tenaga kesehatan pada Gambar 2.4.



Gambar 2.3. Perawatan antenatal

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.



Gambar 2.4. kontak ibu hamil dengan tenaga kesehatan

Pada Gambar 2.3 terdapat Pesan utamanya: Segera setelah Anda tahu sedang hamil, lakukan perawatan *antenatal*. Tujuan perawatan *antenatal*:

1. Dukungan emosional dan nasihat, agar ibu hamil merasa tenang, mendapat bimbingan, dan tidak stres selama kehamilan.
2. Perawatan medis, pemeriksaan kesehatan ibu dan janin secara rutin untuk mendeteksi risiko sejak dini.
3. Informasi relevan dan tepat waktu tentang kehamilan, memberi pengetahuan yang benar agar ibu bisa menjaga kesehatan dirinya dan bayinya.

Kemudian pada Gambar 2.4 Pesan utamanya: Selama kehamilan, semua ibu hamil sebaiknya melakukan 8 kali kontak dengan tenaga kesehatan. Tempat kontak atau perawatan bisa dilakukan di:

1. fasilitas kesehatan (puskesmas, rumah sakit, klinik).
2. Layanan masyarakat/*outreach services* (misalnya posyandu, kunjungan rumah, layanan lapangan).

2.4 Kecerdasan Buatan (*Artificial Intelligence*) dan *Machine Learning*

Perkembangan teknologi informasi dan komputasi modern telah mendorong munculnya kecerdasan buatan (*Artificial Intelligence/AI*) sebagai pendekatan untuk menyelesaikan masalah yang kompleks dan memerlukan analisis data secara otomatis. AI didefinisikan sebagai kemampuan sistem komputer untuk melakukan tugas yang biasanya membutuhkan kecerdasan manusia, seperti pengambilan keputusan, pemecahan masalah, pengenalan pola, dan pembelajaran dari pengalaman. (Hawale, Chavan, Timalsina, dan Thatere, 2024) Dalam konteks kesehatan, AI dapat diap-

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.

likasikan untuk membantu tenaga medis menganalisis data pasien, memprediksi risiko penyakit, dan memberikan rekomendasi berbasis bukti yang lebih cepat dan akurat dibanding metode *konvensional*.

Machine Learning (ML), yaitu metode yang memungkinkan komputer belajar dari data tanpa harus diprogram secara eksplisit (Rajagopal dkk., 2024). ML menggunakan algoritma untuk menemukan pola atau hubungan dalam dataset, kemudian membuat model prediktif yang dapat digunakan untuk klasifikasi atau prediksi. *Machine Learning* sendiri terbagi menjadi beberapa jenis utama, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. Penelitian ini menggunakan pendekatan supervised learning karena dataset maternal health memiliki label kategori risiko yang jelas, sehingga model dapat dilatih untuk klasifikasi risiko kehamilan (Morales dan Escalante, 2021).

Keunggulan ML dalam prediksi risiko kesehatan ibu antara lain kemampuan untuk menangani data besar dan *kompleks*, mengurangi kesalahan prediksi berbasis intuisi manusia, serta mendukung pengambilan keputusan yang cepat dan akurat (Tzimourta, Tsipouras, Angelidis, Tsalikakis, dan Orovou, 2025). Misalnya, algoritma *Decision Tree C4.5* atau *Random Forest* dapat digunakan untuk menentukan variabel paling berpengaruh terhadap risiko komplikasi kehamilan, sedangkan Logistic Regression digunakan untuk menghitung probabilitas risiko berdasarkan kondisi ibu. Penerapan ini membantu tenaga medis dalam mengambil tindakan *preventif* yang lebih tepat sasaran. (Oktarina, Alamsyah, Nurhalissa, dan Satria, 2025). Dengan demikian, pemahaman AI dan ML menjadi pondasi penting dalam penelitian ini untuk menghasilkan prediksi risiko kehamilan yang lebih awal.

2.5 *Machine Learning* dalam Prediksi Risiko Kehamilan

Machine Learning (ML) adalah cabang dari kecerdasan buatan yang memungkinkan komputer belajar dari data dan membuat prediksi atau keputusan tanpa pemrograman eksplisit (Shailaja, Seetharamulu, and Jabbar 2018). Dalam konteks kesehatan, *Machine Learning* digunakan untuk menganalisis data medis dan memprediksi risiko kesehatan, termasuk risiko kehamilan. (A. Rahman dan Alam, 2023) (S. A. Rahman dkk., 2023). Adapun Jenis-Jenis *Machine Learning*

1. *Supervised learning* (pembelajaran terawasi)

Model dilatih menggunakan dataset yang sudah dilabeli, sehingga setiap input memiliki output yang diketahui. (Ghassemi dkk., 2020) Model belajar memetakan input ke output yang benar. Dalam prediksi risiko kehamilan, data ibu hamil seperti usia, tekanan darah, kadar gula darah, dan label

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.

risiko (rendah, sedang, tinggi) digunakan untuk melatih model. Algoritma yang umum digunakan termasuk *Logistic Regression*, *Decision Tree*, *Support Vector Machine (SVM)*, dan *Naive Bayes*.

2. *Unsupervised Learning* (Pembelajaran tak terawasi)

Digunakan ketika dataset tidak memiliki label. Model berusaha menemukan pola atau struktur tersembunyi contohnya *clustering* untuk mengelompokkan ibu hamil dengan karakteristik serupa tanpa mengetahui kategori risiko terlebih dahulu.

3. *Reinforcement Learning* (Pembelajaran Penguatan)

Agen belajar membuat keputusan melalui aksi dan menerima umpan balik berupa reward atau punishment. Pendekatan ini lebih umum untuk sistem yang *interaktif*, seperti *robotika* atau *game*, dan jarang digunakan untuk prediksi risiko kehamilan.

Penerapan Machine Learning dalam Prediksi Risiko Kehamilan. Pendekatan yang digunakan dalam penelitian ini adalah *supervised learning*, karena dataset yang tersedia memiliki label risiko yang jelas. Algoritma seperti *Logistic Regression*, *Decision Tree C4.5*, *SVM*, dan *Naive Bayes* dapat diterapkan untuk menganalisis data kesehatan ibu hamil dan memprediksi tingkat risiko kehamilan. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk menentukan algoritma yang paling efektif dalam prediksi risiko kehamilan.

2.6 Data Mining

Data mining adalah proses eksplorasi dan analisis data dalam jumlah besar untuk menemukan pola, tren, atau informasi yang sebelumnya tidak diketahui, yang berguna untuk pengambilan keputusan (Han et al., 2022). Menurut (Fayyad et al., 1996), data mining merupakan bagian inti dari *Knowledge Discovery in Database* (KDD), yang mencakup serangkaian tahapan dari pengumpulan data hingga penarikan pengetahuan yang dapat digunakan. Dalam konteks penelitian kesehatan, data mining memungkinkan identifikasi faktor risiko dan prediksi kondisi pasien secara lebih akurat.

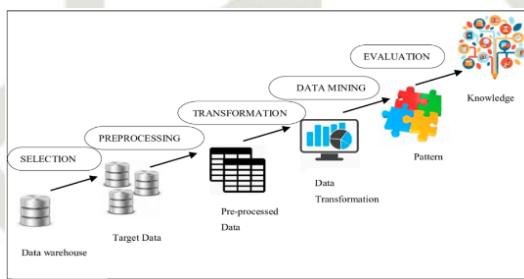
Tujuan utama data mining adalah menemukan pola dan pengetahuan baru yang dapat digunakan untuk pengambilan keputusan yang lebih baik (Han et al., 2022). Manfaat praktis data mining antara lain: Prediksi Risiko membantu memprediksi kemungkinan komplikasi kehamilan, Klasifikasi mengelompokkan ibu hamil ke dalam kategori risiko rendah, sedang, atau tinggi. Asosiasi, menemukan

korelasi antara faktor kesehatan dan hasil kehamilan, dan Deteksi Anomali, mengidentifikasi kasus yang menyimpang dari pola normal (Fayyad et al., 1996). Dalam penelitian ini, data mining digunakan untuk menganalisis data maternal health guna membangun model klasifikasi risiko kehamilan.

Proses data mining biasanya terdiri dari beberapa tahapan penting yang saling berkaitan:

1. Pengumpulan Data, mengumpulkan data dari berbagai sumber
2. *Pra-pemrosesan* Data, membersihkan dan menyiapkan data untuk analisis;
3. Transformasi Data, mengubah data ke format yang sesuai
4. Pemodelan Data, menggunakan algoritma untuk menemukan pola
5. Evaluasi Model, mengukur akurasi dan efektivitas model
6. Penerapan Pengetahuan, mengimplementasikan hasil ke dalam keputusan praktis.(Cheng, Chen, Sun, Zhang, dan Tao, 2017)

Diagram alur proses KDD dapat dilihat pada Gambar 2.5 berikut:



Gambar 2.5. Diagram alur proses KDD

2.7 Algoritma *Machine Learning* Untuk Klasifikasi

Machine Learning (ML) untuk klasifikasi merupakan metode untuk memprediksi kategori atau kelas suatu data berdasarkan fitur yang dimilikinya. Klasifikasi adalah salah satu cabang utama dari supervised learning, di mana model dilatih dengan dataset berlabel untuk mempelajari pola hubungan antara input dan output Nurhalizah, Ardianto, dan Purwono (2024). Dalam penelitian prediksi risiko kehamilan, algoritma klasifikasi digunakan untuk mengelompokkan ibu hamil ke dalam kategori risiko: rendah, sedang, atau tinggi, berdasarkan variabel klinis seperti usia, tekanan darah, kadar hemoglobin, dan riwayat medis sebelumnya. Ada 5 algoritma *Machine learning* yang umum digunakan untuk klasifikasi data medis antara lain:

2.7.1 Logistic Regression

Logistic regression, salah satu metode klasifikasi dalam statistik dan pembelajaran mesin, digunakan untuk memprediksi probabilitas kejadian dari variabel dependen kategorikal biasanya biner: 0/1. Model *logistic regression* tidak seperti regresi linear, yang memprediksi nilai kontinu. Sebaliknya, itu menghasilkan nilai probabilitas yang kemudian diterjemahkan menjadi kelas. Menurut *Applied Logistic Regression*, referensi penting untuk aplikasi kesehatan oleh Scott, Hosmer, dan Lemeshow (1991), logistic regression sangat cocok untuk penelitian bidang medis karena dapat menginterpretasikan koefisien langsung (odds ratio) dan dapat menangani variabel prediktor kategorikal dan kontinu.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (2.1)$$

$P(Y = 1 | X)$ menyatakan *probabilitas* bahwa suatu observasi termasuk ke dalam kelas “1”, misalnya ibu hamil dengan risiko tinggi, berdasarkan sekumpulan variabel prediktor X . Nilai e merupakan basis bilangan eksponensial (sekitar 2,71828) yang digunakan dalam fungsi sigmoid. Parameter β_0 adalah intercept atau konstanta model, sedangkan β_i merupakan koefisien regresi untuk variabel prediktor X_i . Koefisien ini menunjukkan perubahan *log-odds* dari kejadian $Y = 1$ akibat peningkatan satu satuan pada X_i , dengan asumsi variabel prediktor lainnya tetap konstan.

$P(Y = 1 | X)$ probabilitas bahwa kejadian kelas “1” (misalnya ibu hamil berisiko tinggi) terjadi berdasarkan prediktor X .

e basis eksponensial ($\approx 2,71828$) digunakan dalam fungsi sigmoid.

β₀ intercept (konstanta) model.

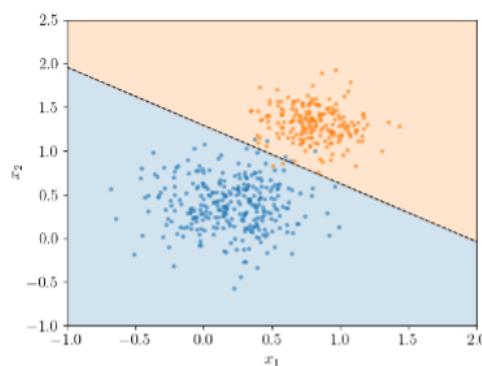
β_i koefisien regresi untuk variabel X_i . Koefisien ini menunjukkan perubahan *log-odds* dari hasil jika X_i naik satu unit, dengan prediktor lain dianggap tetap.

X_i variabel prediktor dalam konteks penelitian ini bisa berupa usia ibu, tekanan darah, kadar gula darah, dan denyut jantung. Fungsi sigmoid

$$\frac{1}{1+e^{-z}}$$

mengubah output linear menjadi nilai antara 0 dan 1, yang kemudian dikonversi menjadi kelas. Menurut panduan (Hilbe, 2017) yang juga menekankan bahwa *interpretasi koefisien* bisa dilakukan melalui *odds ratio* e^{β_i} . Gambar contoh visualisasi *Logistic Regression* dapat dilihat pada Gambar 2.6 berikut.

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



Gambar 2.6. Contoh Visualisasi *Logistic Regresion* Sumber: (Scipython, 2019).

Visualisasi ini menunjukkan decision boundary (garis batas keputusan) pada model *Logistic Regresion*. Model ini memisahkan dua kelas data merah dan biru menggunakan fungsi sigmoid untuk menghitung *probabilitas*. *Logistic regression* bisa diperluas menjadi multi-class lebih dari dua kelas.

2.7.2 Decision Tree C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membangun pohon keputusan. Pohon keputusan adalah metode yang sangat kuat dan terkenal untuk klasifikasi dan prediksi. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Aturan mudah dipahami dalam bahasa alami. Itu juga dapat diekspresikan dalam bahasa database seperti *Structured Query Language* untuk menemukan catatan dari kategori tertentu Setio, Saputro, dan Winarno (2020). Berikut adalah rumus mencari *Entropy* dan *Gain* pada Algoritma C4.5: Untuk perhitungan nilai *Entropy* adalah sebagai berikut:

$$\text{Entropy} = - \sum_{i=1}^n P_i \log_2 P_i \quad (2.2)$$

Keterangan: S: himpunan kasus

A: fitur

n: jumlah partisi S

P_i : proporsi dari S, terhadap S

Kemudian hitung nilai Gain dengan metode information gain:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Value}(A)} \frac{S_v}{S} \text{Entropy}(S_v) \quad (2.3)$$

S = Himpunan Kasus.

A = Atribut.

n = Jumlah Partisi Atribut.

$A_{|Si|}$ = Jumlah Kasus pada partisi ke- i .

$|S|$ = Jumlah Kasus dalam S .

Pohon keputusan mirip sebuah struktur pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Pohon keputusan bekerja mulai dari akar paling atas, jika diberikan sejumlah data uji, misalnya X dimana kelas dari data X belum diketahui, maka pohon keputusan akan menelusuri mulai dari akar sampai node dan setiap nilai dari atribut sesuai data X diuji apakah sesuai dengan aturan pohon keputusan, kemudian pohon keputusan akan memprediksi kelas dari tupel X (Prajoko, Sembiring, dan S, 2021).

2.7.3 Random Forest (RF)

Menurut Sandri dan Zuccolotto (2008), tingkat kepentingan suatu variabel penjelas X_h dalam algoritma Random Forest dapat dihitung menggunakan *Mean Decrease in Gini* (MDG) sebagai berikut:

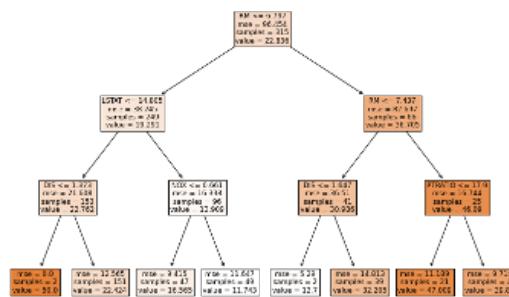
$$MDG_h = \frac{1}{k} \sum_t [d(h,t) I(h,t)] \quad (2.4)$$

Dengan: K = Banyaknya pohon dalam *random forest* $D(h,t)$ = besar penuaan *indeks Gini* untuk Variabel penjelas X_h pada simpul(node) t $I(h,t) \in \{0, 1\}$ Jika X_h memilih simpul t , 0, lainnya Nilai MDG menunjukkan seberapa besar suatu variabel berkontribusi dalam mengurangi impurity (ketidakteraturan) pada pohon-pohon di dalam model *Random Forest* Semakin tinggi nilai MDG, semakin penting variabel tersebut dalam menentukan hasil prediksi. Gambar Contoh visualisasi Random Forest dapat dilihat pada Gambar 2.7 berikut.

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, 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.



Gambar 2.7. Contoh Visualisasi *Random Forest*

Random Forest adalah ensemble dari banyak decision tree yang digunakan untuk meningkatkan akurasi. Gambar ini memperlihatkan salah satu pohon keputusan di dalam *Random Forest*. Setiap pohon melakukan klasifikasi, lalu hasil akhirnya ditentukan berdasarkan voting mayoritas dari seluruh pohon. *Random Forest* mendukung multi-class classification, sehingga cocok digunakan untuk dataset dengan lebih dari dua kelas.

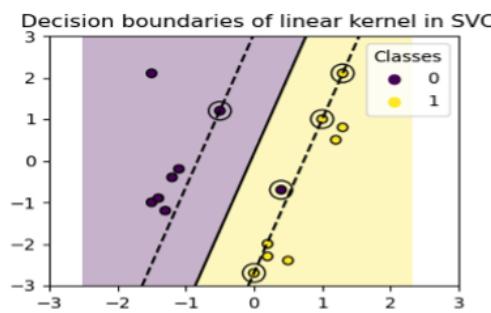
2.7.4 Support Vector Machine (SVM)

Support Vector Machine adalah salah satu algoritma yang paling umum digunakan untuk menganalisis dan mengurutkan data. SVM dibedakan menjadi dua jenis, yaitu SVM linear dan SVM non-linear. Algoritma SVM ini termasuk dalam kategori pengajaran yang diawasi, di mana yang dimaksud dengan pengajaran yang diawasi adalah ketika data yang didapat sudah memiliki label dan tinggal mengolahnya. Kernel *trick* terdapat dalam algoritma SVM. Kernel trick adalah teknik untuk mengubah data pada dimensi tertentu, seperti mengubah dua dimensi menjadi tiga dimensi, dengan tujuan menghasilkan hyperline yang lebih baik. Kernel dapat melakukan berbagai fungsi, seperti linear, RBF, polinomial, dan sigmoid (Saitta, 2023). Dalam kasus ini, kernel polinomial digunakan. Formulasi untuk menghitung algoritma SVM dapat menggunakan rumus berikut:

$$\frac{1}{2} \|\mathbf{w}\|^2 = \frac{1}{2} (w_1^2 + w_2^2 + w_3^2 + w_4^2 + w_5^2) \quad (2.5)$$

Keterangan \mathbf{W} = Fitur yang dimiliki y_i = Label yang dimiliki b = Bias i = Nilai ke- i x_i = Variable support vector Contoh gambar visualisasi SVM dapat dilihat pada Gambar 2.8 berikut.

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



Gambar 2.8. Contoh Visualisasi SVM

Gambar ini menunjukkan *hyperplane* (garis pemisah) dari algoritma *Support Vector Machine*(SVM). SVM mencari garis terbaik yang memisahkan kelas-kelas data dengan jarak margin terbesar. Titik di tepi margin disebut *support vectors*, yang menjadi penentu posisi *hyperplane*.

2.7.5 Naive Bayes (NB)

Naive Bayes Classifier adalah metode sederhana untuk statistik bayesian. Karena asumsi bahwa semua variabel berkontribusi terhadap klasifikasi dan berkorelasi satu sama lain, hal ini disebut *class conditional independence*. Hal ini didasarkan pada teorema *probabilitas bayes*. (S. A. Rahman dkk., 2023)

$$p(c_i | X) = \frac{P(X | C_i) p(C_i)}{P(X)} \quad (2.6)$$

$$P(X) = \sum_{i=1}^n P(X | C_i) P(C_i) \quad (2.7)$$

Keterangan:

$P(C_i | X)$ = Probabilitas posterior kelas (c, target) yang diberikan predictor (x, atribut).

$P(C_i)$ = Probabilitas ke*as* sebelumnya.

$P(X | C_i)$ = Kemungkinan yang merupakan probabilitas dan prediktor kelas tertentu.

$P(X)$ = Sebagai probabilitas prior dari prediktor.

Klasifikasi *Naïve Bayes* memiliki langkah-langkah sebagai berikut:

1. Pertimbangkan dataset D sebagai data pelatihan yang terkait dengan label kelas, dengan setiap pasangan data diwakili oleh vektor elemen n-dimensi $X = (X_1, X_2, X_3, \dots, X_n)$.
2. Anggap ada kelas M $C_1, C_2, C_3, \dots, C_m$. Dalam klasifikasi Naive Bayes,

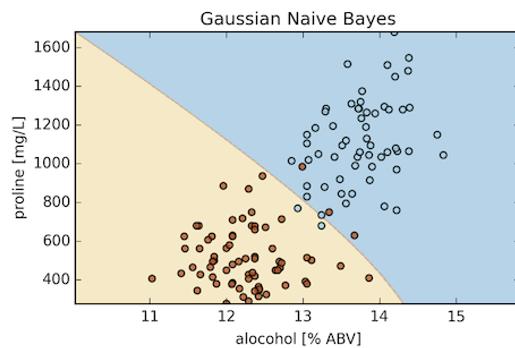
Hak Cipta Dilindungi Undang-Undang

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

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

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

pasangan data X yang tidak diketahui diklasifikasikan ke dalam kelas C_i jika dan hanya jika $P(C_i - X) \geq P(C_j - X)$ untuk $1 \leq j \leq m$, dan $i \neq j$. Klasifier akan memprediksi bahwa X termasuk ke dalam kelas dengan probabilitas posterior tertinggi, yang dikondisikan pada X . Teorema Bayes digunakan untuk menghitung probabilitas ini. Contoh gambar visualisasi *Naive Bayes* dapat dilihat pada Gambar 2.9 berikut.



Gambar 2.9. Contoh Visualisasi *Naive Bayes*

Visualisasi ini menunjukkan *decision boundary* dari algoritma *Gaussian Naive Bayes*, di mana setiap warna mewakili area keputusan untuk kelas yang berbeda. *Naive Bayes* bekerja dengan menghitung *probabilitas* berdasarkan *teorema Bayes* dan mengasumsikan bahwa fitur saling *independen*.

2.8 Preprocessing Data

Preprocessing data adalah tahap awal yang sangat penting dalam *machine learning* karena kualitas data mentah menentukan performa model klasifikasi. Data kesehatan ibu hamil sering kali mengandung *missing values*, *outlier*, duplikasi, dan format yang tidak konsisten. Oleh karena itu, tahap ini meliputi pembersihan, transformasi, dan pengaturan skala variabel agar data siap digunakan oleh algoritma seperti *Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *SVM*, dan *Naive Bayes*. Preprocessing memungkinkan model untuk mengenali pola yang relevan dalam data kesehatan, sehingga prediksi risiko kehamilan menjadi lebih akurat (Han, Pei, dan Tong, 2022).

Salah satu masalah utama dalam dataset kesehatan adalah *missing values* atau data yang hilang. *Missing values* dapat muncul akibat kesalahan pencatatan atau pemeriksaan medis yang tidak lengkap. Sebagai contoh, jika kadar hemoglobin seorang ibu hamil tidak tercatat, nilai median populasi dapat digunakan agar algoritma *Logistic Regression* tetap dapat menghitung probabilitas risiko preeklamsia

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.

tanpa error. Selain itu, beberapa algoritma klasifikasi sensitif terhadap skala fitur. Normalisasi dan standardisasi data bertujuan untuk menyamakan skala variabel sehingga setiap fitur memiliki kontribusi proporsional terhadap model (Butwall, 2021). Standardisasi dapat dilakukan menggunakan Z-Score: $Z = \frac{X-\mu}{\sigma}$, dengan X sebagai nilai asli, μ rata-rata, dan σ standar deviasi. Contoh aplikatifnya, tekanan darah dan indeks massa tubuh distandarisasi agar tidak mendominasi hasil prediksi risiko kehamilan, sehingga model *SVM* atau *Logistic Regression* dapat bekerja optimal.

Fitur kategorikal, juga perlu diubah menjadi format numerik agar algoritma dapat memprosesnya. Metode yang umum digunakan adalah *One-Hot Encoding* atau *Label Encoding* Géron (2022). Selain itu, outlier atau nilai ekstrem dapat memengaruhi akurasi model, terutama algoritma berbasis jarak seperti *SVM*. *Outlier* dapat diidentifikasi melalui *Interquartile Range (IQR)* atau *threshold Z-Score*, kemudian dikoreksi, dihapus, atau dijadikan kategori khusus risiko tinggi. Misalnya, tekanan darah 180 mmHg dapat diberi label khusus sebagai indikator risiko tinggi preeklamsia. Preprocessing data menjadi fondasi yang menghubungkan data mentah dengan algoritma klasifikasi. Dengan data yang telah bersih, distandarisasi, dan dikodekan dengan benar, model seperti *Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *SVM*, dan *Naive Bayes* dapat memprediksi risiko kehamilan dengan lebih akurat dan konsisten. Contoh aplikatifnya, seorang ibu hamil dengan tekanan darah tinggi dan kadar hemoglobin rendah dapat diklasifikasikan sebagai risiko tinggi *preeklamsia* sehingga intervensi medis dapat diberikan lebih cepat dan tepat.

29 Matrix Evaluasi

Matriks evaluasi adalah alat fundamental dalam menilai kinerja model klasifikasi dalam machine learning. Salah satu pendekatan evaluasi paling umum adalah dengan menggunakan *Confusion Matrix*, yang menyajikan perbandingan antara hasil prediksi model dengan nilai aktual. Matriks ini memungkinkan peneliti untuk mengidentifikasi kesalahan prediksi dan menilai performa model secara kuantitatif Chicco dan Jurman (2020). Dalam konteks prediksi risiko kehamilan, *Confusion Matrix* membantu mengevaluasi kemampuan model dalam mengklasifikasikan ibu hamil ke dalam kategori risiko rendah, sedang, dan tinggi.

Confusion Matrix terdiri dari empat elemen utama: *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. TP adalah jumlah kasus positif yang berhasil diprediksi dengan benar, TN adalah jumlah kasus negatif

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 tepat, FP adalah jumlah kasus negatif yang salah diklasifikasikan sebagai positif, dan FN adalah jumlah kasus positif yang gagal diidentifikasi. Elemen-elemen ini menjadi dasar untuk menghitung metrik evaluasi seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

Perhitungan matematika untuk *confusion matrix* berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.8)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.9)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.10)$$

$$F1 - Score = \frac{precision \times recall}{precision + recall} \quad (2.11)$$

Keterangan:

TP (True Positive): Prediksi positif yang benar.

TN (True Negative): Prediksi negatif yang benar.

FP (False Positive): Prediksi positif yang salah.

FN (False Negative): Prediksi negatif yang salah.

Accuracy memberikan gambaran umum kinerja model, namun kurang sensitif jika terdapat ketidakseimbangan kelas (misal jumlah ibu hamil risiko tinggi lebih sedikit daripada risiko rendah).

Accuracy adalah metrik yang digunakan untuk mengukur persentase prediksi yang benar dari seluruh prediksi yang dilakukan. *Accuracy* memberikan gambaran umum tentang kinerja model, namun kurang sensitif terhadap ketidakseimbangan kelas, yang umum terjadi pada dataset risiko kehamilan, di mana kasus risiko tinggi mungkin lebih sedikit dibanding risiko rendah (Ghanem dkk., 2023).

Precision dan *Recall* adalah metrik yang lebih spesifik untuk mengevaluasi kemampuan model dalam mengidentifikasi kelas positif. *Precision* mengukur proporsi prediksi positif yang benar, dihitung. *Precision* penting untuk mengetahui tingkat keakuratan prediksi kasus risiko tinggi, sedangkan *Recall* penting untuk mengevaluasi kemampuan model menangkap semua kasus risiko tinggi yang ada. *F1-Score* adalah rata-rata harmonis antara *Precision* dan *Recall*, memberikan ukuran kinerja yang seimbang. Sedangkan *Recall* mengukur proporsi kasus positif yang berhasil diidentifikasi. *F1-Score* sangat berguna ketika dataset tidak seimbang, seperti pada prediksi risiko kehamilan, karena mempertimbangkan baik kesalahan positif maupun kesalahan negatif.

Secara aplikatif, dalam penelitian prediksi risiko kehamilan, keempat metrik ini digunakan untuk menilai performa model klasifikasi seperti *Logistic Regression*, *Decision Tree*, *Random Forest*, *SVM*, dan *Naive Bayes*. Misalnya, *Recall* tinggi menunjukkan model berhasil menangkap sebagian besar ibu hamil yang berisiko tinggi, sehingga intervensi medis dapat diberikan lebih cepat, sedangkan *Precision* tinggi menunjukkan sebagian besar prediksi risiko tinggi benar-benar akurat.

2.10 Google Collab

Google Colab *Google Colaboratory*, atau *Google Colab*, adalah platform berbasis cloud yang menyediakan lingkungan notebook Jupyter secara gratis untuk pengembangan dan eksekusi kode *Python*, termasuk proyek *machine learning* dan data science. *Google Colab* memungkinkan peneliti menjalankan kode tanpa memerlukan konfigurasi perangkat keras lokal yang kompleks, karena seluruh proses komputasi dilakukan di server cloud Google Bisong (2019). Keunggulan ini sangat relevan untuk penelitian prediksi risiko kehamilan, yang membutuhkan pengolahan dataset besar dan algoritma klasifikasi kompleks.

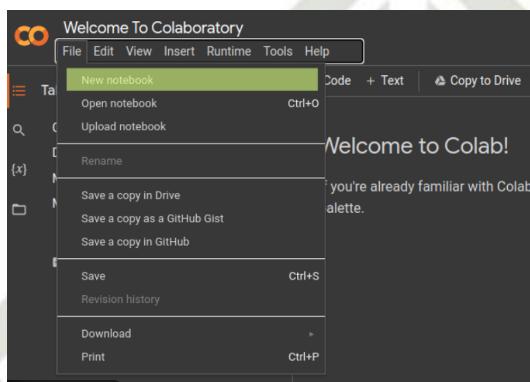
Salah satu fitur penting *Google Colab* adalah kemampuannya untuk mengakses GPU dan TPU secara gratis atau dengan biaya minimal, sehingga mempercepat proses pelatihan model *machine learning*. Hal ini memungkinkan implementasi algoritma seperti *Random Forest*, *SVM*, dan *Neural Network* lebih efisien, terutama pada dataset yang besar atau ketika eksperimen *hyperparameter* dilakukan secara iteratif Bisong (2019). Dengan demikian, *Colab* menyediakan lingkungan eksperimental yang fleksibel bagi peneliti tanpa harus menginvestasikan perangkat keras khusus.

Google Colab juga mendukung integrasi langsung dengan *Google Drive*, *GitHub*, dan pustaka *Python* populer seperti *Pandas*, *NumPy*, *Scikit-learn*, *TensorFlow*, dan *PyTorch*. Integrasi ini mempermudah manajemen dataset, penyimpanan model, dan kolaborasi tim, sehingga peneliti dapat melakukan preprocessing, pelatihan, evaluasi, dan visualisasi data secara terpadu. Dalam penelitian prediksi risiko kehamilan, hal ini memungkinkan implementasi pipeline *machine learning* yang efisien dari *preprocessing* hingga evaluasi model.

Selain itu, *Colab* mendukung dokumentasi interaktif melalui Markdown dan visualisasi grafik, yang memudahkan peneliti menjelaskan alur kerja dan hasil analisis. Kemampuan ini sangat membantu dalam penelitian ilmiah dan publikasi, karena setiap langkah eksperimen dapat dicatat, direproduksi, dan dibagikan dengan mudah (Bisong, 2020). Contohnya, proses pelatihan model klasifikasi untuk

prediksi risiko kehamilan dapat disajikan lengkap dengan kode, *output*, dan grafik performa model dalam satu notebook yang dapat diakses oleh pembimbing atau rekan penelitian.

Secara konseptual, penggunaan *Google Colab* dalam penelitian *machine learning* menyediakan lingkungan yang efisien, *kolaboratif*, dan terintegrasi untuk eksperimen model klasifikasi. Dengan memanfaatkan *Colab*, peneliti dapat fokus pada pengembangan model dan interpretasi hasil, tanpa terbebani oleh keterbatasan perangkat keras atau masalah kompatibilitas perangkat lunak. Hal ini meningkatkan produktivitas penelitian, khususnya dalam prediksi risiko kehamilan berbasis data kesehatan ibu. Berikut gambar tampilan *Google colab* pada Gambar 2.10



Gambar 2.10. Tampilan Antarmuka Google Collab

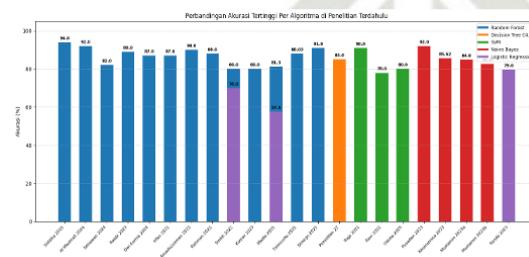
2.11 Penelitian Tedahulu

Berdasarkan analisis terhadap 30 penelitian terdahulu mengenai prediksi risiko kesehatan ibu hamil, terlihat bahwa *Random Forest* menjadi algoritma yang paling sering digunakan dan menunjukkan performa paling konsisten. *Random Forest* muncul dalam 14 penelitian dan tercatat sebagai algoritma dengan akurasi tertinggi pada 13 penelitian, dengan akurasi tertinggi mencapai 94% (Siddika dan SultanaM, 2025). Algoritma *Decision Tree C4.5* digunakan dalam 10 penelitian, namun hanya tercatat satu kali sebagai algoritma dengan akurasi tertinggi, yaitu pada penelitian nomor 27, dengan akurasi sekitar 85%. Sementara itu, *Support Vector Machine (SVM)* muncul dalam 10 penelitian dan menjadi algoritma dengan akurasi tertinggi pada 3 penelitian, dengan akurasi bervariasi antara 59,6% hingga 90,9%. Algoritma *Naïve Bayes* digunakan dalam 9 penelitian dan berhasil menjadi yang tertinggi pada 4 penelitian, dengan akurasi tertinggi sekitar 85%. Sedangkan *Logistic Regression* digunakan dalam 6 penelitian, namun hanya 3 penelitian yang menampilkan hasil akurasi, yang menunjukkan bahwa algoritma ini lebih jarang unggul dibanding algoritma lain. Secara keseluruhan, *Random Forest* 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 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.

bukti memberikan performa prediksi yang tinggi dan konsisten dibandingkan algoritma lainnya, sedangkan algoritma lain seperti *Decision Tree C4.5*, *SVM*, dan *Naïve Bayes* hanya unggul pada beberapa kasus tertentu. Temuan ini menunjukkan bahwa *Random Forest* merupakan pilihan algoritma yang efektif untuk prediksi risiko kesehatan ibu hamil, khususnya dalam konteks dataset yang bervariasi dan kompleks. Berikut gambar diagram hasil akurasi tiap algoritma dari penelitian terdahulu yang menyatakan hasil penelitiannya sebagai algoritma dengan akurasi tertinggi.

1. Diagram hasil akurasi algoritma yang menyatakan akurasi tertinggi dari penelitian terdahulu pada Gambar 2.11



Gambar 2.11. Hasil akurasi tertinggi dari penelitian terdahulu

Berdasarkan analisis dari 30 penelitian terdahulu mengenai prediksi risiko kesehatan ibu hamil, dapat dilihat perbandingan akurasi dari lima algoritma utama: *Random Forest*, *Decision Tree C4.5*, *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *Logistic Regression*. Dari grafik batang, terlihat bahwa *Random Forest* secara konsisten menghasilkan akurasi tertinggi di sebagian besar penelitian, dengan nilai akurasi berkisar antara 75.2% hingga 92%, dan muncul sebagai algoritma dengan performa terbaik sebanyak 13 kali. Akurasi tersebut tercatat secara detail pada penelitian: Irfan, Basuki, dan Azhar (2021) (87%), S. A. Rahman dkk. (2023) (88%), A. Rahman dan Alam (2023) (89%), Mutlu, Yü journal = Unknown Journal cel, Durmaz, Cengİl, dan Yildirim (2023) (89.16%), Nwokoro, Duke, dan Nwokoro (2024) (92%), Assaduzzaman dkk. (2024) (75.2%), Malde dkk. (2025) (81.3%), Simegn dan Degu (2025) (91%), Tzimourta dkk. (2025) (88,03%), S. A. Rahman dkk. (2023) (88%), Dwi Kurnia Putra, 2024 (87%).

Untuk algoritma lainnya, *Decision Tree C4.5* tercatat satu kali sebagai algoritma dengan akurasi tertinggi pada penelitian Kurniawan (2025) (85,25%). *Support Vector Machine (SVM)* unggul pada tiga penelitian dengan rincian akurasi: Raja dkk. (2021) (90,9%), Udon et al., 2025 (80%), dan Malde dkk. (2025) (81.3%).

Naïve Bayes berhasil menjadi yang tertinggi pada empat penelitian dengan nilai akurasi: Pusadan, Ghifari, dan Anshori (2023) (92%), Khoirunnisa dan Lestari (2023)(85,62%), “Classification Of Maternal Health Risk Using Three Models Naïve Bayes Method” (2023)(84,8%), dan Mustamin, Aziz, Firmansyah, dan Ishak (2023) 82,6%). Terakhir, *Logistic Regression* menunjukkan performa yang relatif lebih rendah dan kurang stabil dibandingkan algoritma lainnya, dengan hasil yang tercatat pada Farida et al., 2023 (79,6%), Smith et al., 2022 (70%), dan Malde dkk. (2025) (57,6%).

Meskipun model *Machine Learning* telah terbukti efektif, penelitian terdahulu mengidentifikasi gap penting, terutama terkait kurangnya konsistensi dalam penggunaan metrik evaluasi model. Banyak studi cenderung hanya berfokus pada akurasi tanpa mempertimbangkan metrik *Precision*, *Recall*, dan *F1-Score* yang sangat krusial dalam domain medis untuk menghindari kesalahan diagnosis. Selain itu, kurangnya perbandingan *komprehensif* antar-algoritma pada dataset yang berbeda memberikan peluang bagi penelitian ini untuk memberikan hasil yang lebih mendalam dan reliabel. Hasil penelitian terdahulu dapat dilihat pada (Lampiran D).

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.

Hak Cipta Dilindungi Undang-Undang

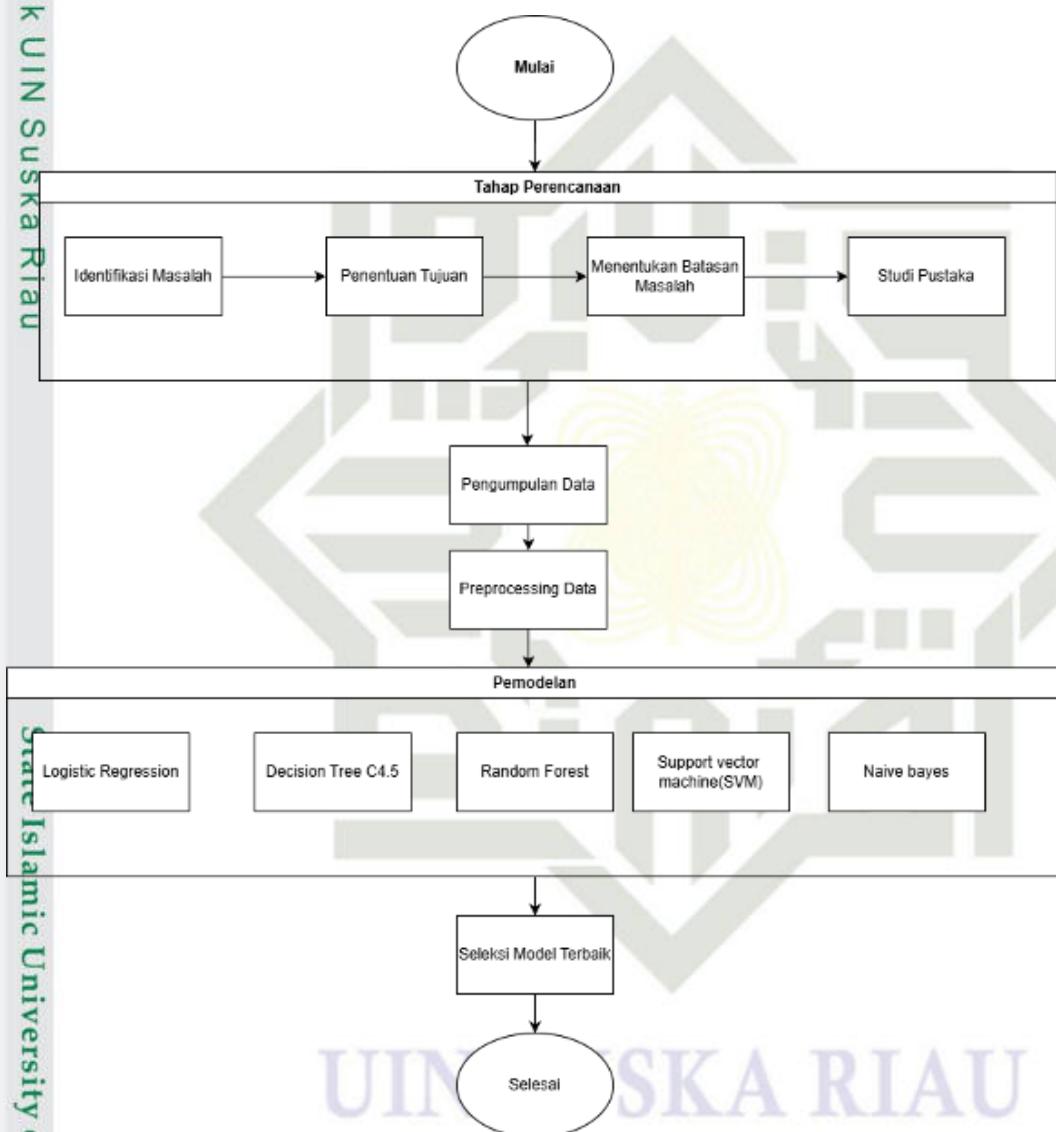
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan 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

3.1 Metodologi Penelitian

Adapun proses penelitian adalah Tahap Perencanaan, Pengumpulan Data, Pemodelan Machine Learning, serta Evaluasi dan Seleksi Model. Alur metodologi untuk penelitian kali ini dapat dilihat pada Gambar 3.1.



Gambar 3.1. Alur metodologi penelitian

3.2 Tahap Perencanaan

Tahap perencanaan merupakan langkah awal dalam proses penelitian yang berfungsi sebagai pondasi untuk memastikan arah, fokus, dan keberhasilan peneli-

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.

tian. Menurut Kothari (2014), tahap ini melibatkan penetapan masalah, tujuan, variabel, dan pendekatan *metodologis* agar penelitian dapat dilaksanakan secara sistematis dan menghasilkan temuan yang valid. Dalam penelitian ini, tahap perencanaan difokuskan untuk merancang analisis perbandingan algoritma machine learning dalam memprediksi risiko kehamilan, yang bertujuan menghasilkan temuan analisis dan evaluatif berdasarkan data medis. Untuk mencapai hasil yang maksimal, tahap perencanaan dibagi menjadi beberapa langkah, yaitu: identifikasi masalah, perumusan tujuan, penetapan batasan penelitian, studi pustaka, dan perancangan metodologi penelitian.

3.2.1 Identifikasi Masalah

Langkah pertama dalam perencanaan adalah identifikasi masalah, yang bertujuan memahami isu utama yang akan diteliti. Berdasarkan laporan World Health Organization (WHO, 2023b), lebih dari 295.000 wanita meninggal setiap tahun akibat komplikasi kehamilan dan persalinan, dengan sebagian besar kasus terjadi di negara berkembang. Di Indonesia, data Kementerian Kesehatan Republik Indonesia (2023) menunjukkan bahwa angka kematian ibu (AKI) masih berada pada 189 per 100.000 kelahiran hidup, jauh di atas target *Sustainable Development Goals (SDGs)* 2030, yaitu ≤ 70 per 100.000 kelahiran hidup.

Salah satu penyebab tingginya AKI adalah keterlambatan dalam mendeteksi risiko kehamilan sejak dini akibat keterbatasan sumber daya manusia di bidang kesehatan dan kurangnya analisis berbasis data. Oleh karena itu, dibutuhkan pendekatan analitis berbasis *machine learning* yang dapat membandingkan algoritma untuk mengidentifikasi model terbaik dalam memprediksi tingkat risiko kehamilan. Permasalahan inilah yang menjadi dasar penelitian, yaitu bagaimana membandingkan algoritma *machine learning* untuk menentukan yang memiliki akurasi optimal dalam mengklasifikasikan ibu hamil ke dalam kategori *low risk*, *mid risk*, dan *high risk*, akurasi tinggi.

Masalah yang diangkat dalam penelitian ini adalah tingginya angka kematian ibu (AKI) di Indonesia yang masih menjadi tantangan besar di bidang kesehatan. Salah satu faktor penyebab adalah keterlambatan dalam mengenali risiko kehamilan sejak dini. Oleh karena itu, diperlukan machine learning yang dapat membantu mendeteksi risiko kehamilan lebih cepat dengan memanfaatkan data medis yang tersedia.

3.2.2 Perumusan Tujuan Penelitian

Setelah permasalahan diidentifikasi, langkah berikutnya adalah merumuskan tujuan penelitian. Tujuan utama penelitian ini adalah menganalisis dan membandingkan performa algoritma machine learning dalam memprediksi tingkat risiko kehamilan berdasarkan parameter kesehatan ibu hamil seperti usia, tekanan darah sistolik, tekanan darah diastolik, kadar gula darah, suhu tubuh, kadar hemoglobin, dan denyut jantung. Selain itu, penelitian ini bertujuan untuk mengidentifikasi algoritma yang memberikan akurasi paling optimal dalam mengklasifikasikan ibu hamil ke dalam tiga kategori risiko, yaitu *Low Risk*, *Mid Risk* dan *High Risk*.

Secara spesifik, penelitian ini bertujuan untuk:

1. Membandingkan lima algoritma machine learning, yaitu *Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *Support Vector Machine (SVM)*, dan *Naive Bayes*, berdasarkan metrik evaluasi akurasi, presisi, *recall*, dan *F1-score* untuk memperoleh hasil prediksi yang paling optimal.
2. Mengidentifikasi algoritma yang memberikan akurasi paling optimal dalam memprediksi tingkat risiko kehamilan.

3.2.3 Penetapan Batasan Penelitian

Tahap berikutnya adalah penetapan batasan penelitian (scoping), yang bertujuan membatasi ruang lingkup penelitian agar tetap fokus dan efisien batasan dalam penelitian ini meliputi:

1. Sumber Data Penelitian ini hanya menggunakan data sekunder, yaitu Maternal Health Risk Dataset yang terdiri dari 1.822 baris dan 7 kolom, tersedia di platform *Kaggle*.
2. Atribut yang digunakan terbatas pada enam parameter medis utama, yaitu usia, tekanan darah sistolik, tekanan darah diastolik, kadar gula darah, suhu tubuh, hemoglobin, dan detak jantung.
3. Algoritma Penelitian ini berfokus pada penerapan dan perbandingan lima algoritma *machine learning*: *Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *Support Vector Machine (SVM)*, dan *Naive Bayes*.
4. Evaluasi Model: Kinerja model dinilai menggunakan metrik klasifikasi, yaitu akurasi, presisi, *recall*, dan *F1-score*.

Batasan ini ditetapkan agar penelitian dapat dilaksanakan dengan cakupan yang proporsional terhadap waktu, sumber daya, serta data yang tersedia.

3.2.4 Studi Pustaka

Studi pustaka dilakukan untuk memperkuat dasar teoretis penelitian. Menurut Sugiyono (2019), studi pustaka berfungsi untuk mengidentifikasi teori-teori dan hasil penelitian terdahulu yang relevan, serta menentukan posisi penelitian yang sedang dilakukan. Dalam penelitian ini, studi pustaka dilakukan terhadap beberapa sumber ilmiah seperti jurnal, dan laporan lembaga resmi (WHO, Kemenkes, dan UNICEF). Beberapa penelitian terdahulu menunjukkan efektivitas algoritma machine learning dalam prediksi risiko kehamilan. Misalnya, penelitian oleh S. A. Rahman dkk. (2023) menemukan bahwa algoritma *Decision Tree* mampu mencapai akurasi 90% dalam mengklasifikasikan tingkat risiko kehamilan. Semenitara itu, penelitian oleh Assaduzzaman dkk. (2024) menunjukkan bahwa *Random Forest* mencapai akurasi 82% dengan variabel usia dan kadar gula darah sebagai faktor dominan. Selain itu, penelitian oleh Khoirunnisa dan Lestari (2023) menunjukkan bahwa algoritma *Naive Bayes* mampu memberikan akurasi sebesar 85,62%, dengan keunggulan pada efisiensi waktu komputasi.

Berdasarkan hasil kajian terhadap penelitian di atas, dapat disimpulkan bahwa penelitian terdahulu telah berhasil menunjukkan efektivitas berbagai algoritma machine learning dalam mendeteksi risiko kehamilan. Namun demikian, masih terdapat beberapa kesenjangan yang perlu dijembatani. Pertama, sebagian besar penelitian masih menitikberatkan pada akurasi sebagai ukuran utama performa model, tanpa mempertimbangkan metrik lain seperti *precision*, *recall*, dan *F1-score*, yang penting dalam konteks data medis dengan kemungkinan ketidakseimbangan kelas. Kedua, belum banyak penelitian yang membandingkan lima algoritma utama secara komprehensif, yaitu *Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *Support Vector Machine (SVM)*, dan *Naive Bayes*, dengan metrik evaluasi yang lengkap. Ketiga, beberapa penelitian masih menggunakan dataset berukuran kecil atau bersifat lokal, sehingga hasil prediksi belum cukup representatif terhadap populasi yang lebih luas. Maka Penelitian ini dilakukan untuk menjawab kesenjangan tersebut dengan cara membandingkan kinerja lima algoritma *machine learning Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *SVM*, dan *Naive Bayes* dalam memprediksi risiko kehamilan.

3.3 Pengumpulan Data

Data yang digunakan merupakan data yang diperoleh dari sumber dataset Kaggle. Dataset berisi informasi Risiko kesehatan ibu hamil dengan tiga label risiko kehamilan rendah, sedang, dan tinggi. Variabelnya adalah usia, suhu tubuh, denyut

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.

jantung, kadar hemoglobin, kadar gula darah, serta tekanan darah (sistolik dan diastolik) Dataset yang digunakan dalam penelitian ini berjumlah 1.822 *record*. Secara umum, tahap pengumpulan data dalam penelitian ini meliputi empat langkah utama yaitu identifikasi sumber data, deskripsi dataset, penentuan atribut data penelitian. Bukti pengumpulan data yang telah dilakukan ditunjukkan pada (Lampiran A), (Lampiran B), dan (Lampiran C).

3.3.1 Identifikasi Sumber Data

Sumber data yang digunakan dalam penelitian ini terdiri dari data kaggle, *Maternal Health Risk Dataset* yang tersedia secara publik di platform *Kaggle*. Dataset *Maternal Health Risk Dataset* (diunggah oleh Joakim Arvidsson tahun 2023) berasal dari penelitian yang dilakukan di *Bangladesh*. Bangladesh dan Indonesia punya karakteristik kesehatan ibu yang mirip, karena dua-duanya termasuk negara berkembang di Asia Selatan dan Asia Tenggara. Dataset ini digunakan Karena atribut datanya (usia, tekanan darah, kadar gula, suhu tubuh, denyut jantung) adalah parameter fisiologis *universal* yang digunakan oleh *WHO* untuk menilai risiko kehamilan di seluruh dunia. Dataset ini dikembangkan untuk mendukung penelitian di bidang kesehatan ibu dan mengandung data medis yang berkaitan dengan kondisi ibu hamil, seperti tekanan darah, kadar gula darah, suhu tubuh, hemoglobin, dan detak jantung.(World Health Organization journal = Unknown Journal , 2013). Selain itu, baik Indonesia maupun Bangladesh memiliki profil kesehatan ibu yang mirip sama-sama negara berkembang dengan angka kematian ibu yang tinggi dan penyebab utamanya adalah preeklampsia. Menurut penelitian di Bangladesh,Khan dkk. (2023) komplikasi *Pre eclampsia/Eclampsia* menyumbang sekitar 20–24% dari kematian ibu.

3.3.2 Deskripsi Dataset

Dataset yang digunakan, yaitu *Maternal Health Risk Dataset Kaggle*, 2021, berisi 1.822 baris data dengan 7 atribut (fitur) dan 1 label kelas. Label tersebut menunjukkan tingkat risiko kehamilan yang dikategorikan menjadi tiga kelas:

1. *Low Risk* (Risiko Rendah)
2. *Mid Risk* (Risiko Sedang)
3. *High Risk* (Risiko Tinggi)

Tabel 3.1 menjelaskan atribut yang digunakan dalam penelitian ini:

Tabel 3.1. Penjelasan Atribut Penelitian

No	Atribut	Tipe Data	Deskripsi
1	<i>Age</i>	Numerik	Usia ibu hamil (tahun). Usia ibu hamil dalam satuan tahun. Menunjukkan usia seorang wanita selama masa kehamilan.
2	<i>SystolicBP</i>	Numerik	Tekanan darah sistolik (mmHg). Nilai tekanan darah bagian atas (mmHg). Merupakan salah satu atribut penting dalam memantau kesehatan ibu selama kehamilan.
3	<i>DiastolicBP</i>	Numerik	Tekanan darah diastolik (mmHg). Nilai tekanan darah bagian bawah (mmHg). Juga merupakan atribut penting untuk menentukan kondisi kesehatan ibu hamil.
4	<i>BS</i>	Numerik	Kadar gula darah (Blood Sugar, mmol/L). Menunjukkan kadar glukosa darah dalam satuan mmol/L. Nilai ini berperan penting untuk mendeteksi risiko diabetes gestasional selama kehamilan.
5	<i>BodyTemp</i>	Numerik	Suhu tubuh (°C). Suhu tubuh ibu hamil dalam satuan Fahrenheit (°F). Suhu tubuh yang tidak normal dapat mengindikasikan kondisi kesehatan yang tidak stabil.
6	<i>HeartRate</i>	Numerik	Denyut jantung (BPM). Detak jantung normal saat istirahat, diukur dalam satuan bpm (beats per minute). Dapat digunakan untuk memantau kondisi jantung ibu selama kehamilan.
7	<i>RiskLevel</i>	Kategorikal	Label atau kelas target yang menunjukkan tingkat risiko kehamilan berdasarkan atribut-atribut sebelumnya (misalnya: rendah, sedang, tinggi). Tingkat risiko: Low Risk, Mid Risk, High Risk.

3.3.3 Teori Atribut Untuk Prediksi Risiko Kesehatan Kehamilan

Tabel 3.2 berisi teori atribut yang digunakan.

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

Tabel 3.2. Teori Atribut Penelitian

Atribut	Ringkasan Temuan dari Relevansi untuk Model Literatur
Age (Usia ibu)	<p>Usia ibu berhubungan dengan risiko komplikasi; usia sangat muda dan usia lanjut (sering didefinisikan ≥ 35 tahun) dikaitkan dengan peningkatan risiko <i>pre-eclampsia</i> dan komplikasi persalinan. (Vital dan Reports, 2025) Usia ≤ 20 tahun atau ≥ 35 tahun meningkatkan risiko komplikasi seperti <i>preeklampsia</i>, persalinan prematur, dan kematian ibu. Usia ibu yang muda (≤ 19 tahun) atau usia lanjut (≥ 35 tahun) dikaitkan dengan peningkatan risiko hasil persalinan perinatal yang merugikan (Cavazos-rehg dkk., 2016). Banyak penelitian menunjukkan usia ibu hamil merupakan faktor risiko komplikasi kehamilan, termasuk <i>pre-eclampsia</i> atau hipertensi gestasional (Syahfirda, Hamid, Santi, dan Mulawardhana, 2023).</p>

- 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.

Tabel 3.2. Teori Atribut Penelitian (lanjutan)

Atribut	Ringkasan Temuan dari Literatur	Relevansi untuk Model
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.	SystolicBP / DiastolicBP Hipertensi kehamilan (termasuk pre-eclampsia) adalah salah satu penyebab utama <i>morbiditas dan mortalitas maternal</i> . Nilai SBP/DBP tinggi (SBP ≥ 160 atau DBP ≥ 110 mmHg) menandakan kondisi berat. Kedua komponen (sistolik & diastolik) relevan, dan perubahan BP selama kehamilan memprediksi risiko (Magee dkk., 2022). Tekanan darah ≥ 140 mmHg menjadi indikasi hipertensi gestasional atau <i>preeklampsia</i> , dengan diastolik ≥ 90 mmHg sebagai kriteria diagnostik preeklampsia (Magee dkk., 2022).	Sangat kuat sebagai fitur; BP langsung terkait dengan diagnosis preeclampsia/hipertensi gestasional. Nilai SBP/DBP $\geq 140/90$ atau $\geq 160/110$.

Tabel 3.2. Teori Atribut Penelitian (lanjutan)

Atribut	Ringkasan Temuan dari Literatur	Relevansi untuk Model
BS (Blood Sugar / Glukosa darah)	<p>Kadar glukosa meningkatkan risiko makrosomia, <i>pre-eclampsia</i>, serta komplikasi neonatal. Pengukuran glukosa adalah indikator kuat keadaan metabolisme ibu selama kehamilan. Screening GDM biasanya di trimester kedua (24–28 minggu), namun penelitian merekomendasikan perhatian lebih dini pada kelompok berisiko (World Health Organization journal = Unknown Journal, 2013). Gula darah tinggi (≥ 140 mg/dL) adalah indikator diabetes gestasional, meningkatkan risiko komplikasi, hipertensi, termasuk preeklamsia (WHO, 2016).</p>	<p>Indikator metabolisme untuk prediksi risiko komplikasi.</p>
BodyTemp	<p>Demam (hipertermia) pada kehamilan; suhu tinggi bisa menandakan infeksi atau stres sistemik yang berkontribusi pada komplikasi (More, 2017). Pedoman klinis mendefinisikan demam puerperal sebagai kenaikan suhu tubuh $\geq 38^\circ\text{C}$ sebagai ambang batas (journal = Unknown Journal, 2008).</p>	<p>BodyTemp dapat berperan sebagai indikator kondisi akut (infeksi) yang menaikkan risiko.</p>

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.

Tabel 3.2. Teori Atribut Penelitian (lanjutan)

Atribut	Ringkasan Temuan dari Literatur	Relevansi untuk Model
HeartRate	<p>Selama kehamilan terjadi peningkatan fisiologis HR (rata-rata +10–20 bpm). Tachycardia patologis dengan HR istirahat \geq 100–120 bpm dapat menunjukkan kondisi kardiovaskular, anemia, infeksi, atau stres yang berpotensi menaikkan risiko maternal. Dokumen menyebut: “Commonly a pregnant woman’s resting heart rate rises by 10–20 beats per minute, but there is no accepted upper limit value heart rate in pregnancy.”</p>	HeartRate dapat dimasukkan sebagai fitur tambahan.
3-kelas Risiko (Low, Mid, High)	<p><i>Low Risk:</i> kondisi normal, komplikasi minimal, cukup ditangani melalui pelayanan antenatal standar. <i>Mid Risk:</i> satu atau beberapa parameter kesehatan mulai menyimpang sehingga memerlukan pemantauan lebih intensif. <i>High Risk:</i> kondisi kehamilan dengan faktor risiko berat atau kombinasi faktor signifikan, membutuhkan pengawasan dan intervensi medis khusus (Afolabi, Esienumoh, dan Afolabi, 2023).</p>	<p>Target klasifikasi model; evaluasi menggunakan Accuracy, Precision, Recall, dan F1 Score.</p>

Semua atribut yang ada di dataset *Kaggle Age, BP, BS, BodyTemp*,

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

1. 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.

HeartRate telah diakui secara medis oleh WHO dan Kemenkes sebagai indikator utama dalam skrining risiko kehamilan, sehingga secara teori dan empiris valid untuk digunakan dalam model prediksi risiko maternal. Adapun Penelitian oleh (Simegn and Degu 2025) ini menemukan korelasi yang kuat antara risiko kehamilan yang tinggi, tekanan darah sistolik dan diastolik yang tinggi, kadar gula darah yang lebih tinggi, dan usia yang lebih tua dibandingkan dengan parameter klinis lainnya. Dan kinerja yang unggul yaitu random forest dengan akurasi klasifikasi 91% dan dipilih untuk klasifikasi tingkat risiko. Selanjutnya penelitian oleh Siddika dan SultanaM (2025) Pasien maternal, misalnya ketika seorang ibu berpikir tekanan darahnya mungkin naik, glukosanya mungkin naik, atau turun, dan terdapat banyak data yang dapat membahayakan bayi. Penelitian ini menganalisis data tersebut dan melatihnya dengan berbagai algoritma *Machine Learning*.

3.4 Preprocessing Data

Tahap ini bertujuan untuk menyiapkan data agar siap digunakan dalam proses pelatihan model. Langkah yang dilakukan meliputi:

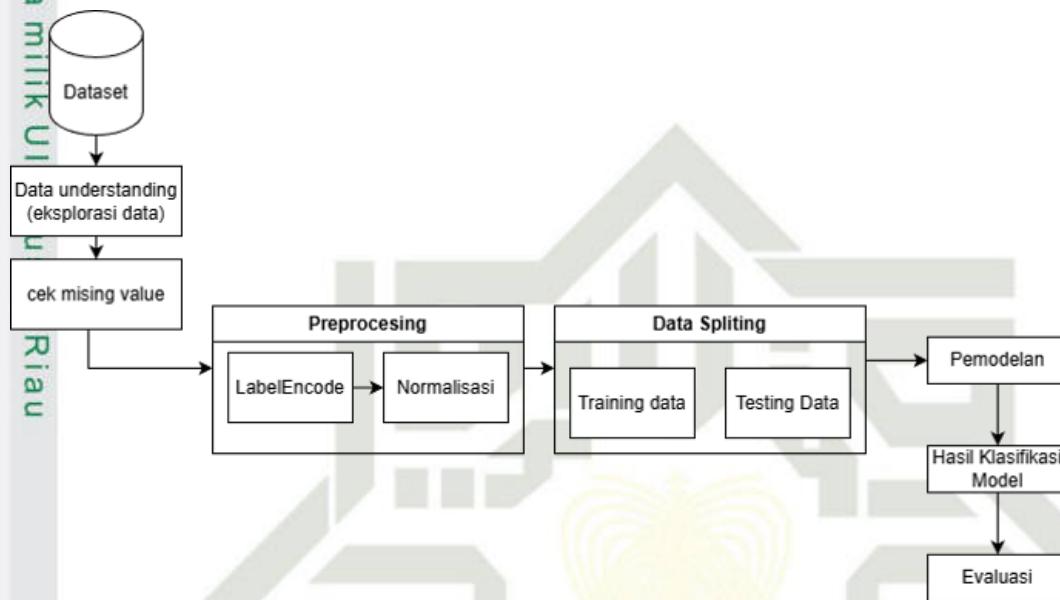
1. Pembersihan data (*Data cleaning*) untuk memastikan tidak ada data kosong (missing values) atau nilai ekstrem yang dapat mempengaruhi hasil model.
2. Encoding variabel kategori menggunakan Label Encoder agar seluruh atribut dapat diolah oleh algoritma berbasis numerik.
3. Normalisasi data dengan *z score* sehingga setiap fitur memiliki rentang nilai yang seragam antara 0-1, guna meningkatkan stabilitas perhitungan algoritma.
4. Pembagian data *train-test split* menjadi data latih dan data uji dengan rasio 80:20 agar performa model dapat dievaluasi secara objektif.

3.5 Pemodelan

Tahap pemodelan merupakan inti dari penelitian ini karena berfungsi untuk membangun model prediksi risiko kehamilan menggunakan berbagai algoritma *machine learning*. Menurut Han, Kamber, dan Pei (2022), pemodelan adalah proses mengubah data yang telah diproses menjadi representasi matematis untuk menge-nali pola dan menghasilkan prediksi. Dalam penelitian ini, tahap pemodelan dilakukan dengan menggunakan lima algoritma klasifikasi yaitu *Logistic Regression*, *Decision Tree C4.5*, *Random Forest*, *Support Vector Machine (SVM)*, dan *Naïve Bayes*, dengan tujuan untuk menentukan algoritma yang paling efektif dalam memprediksi tingkat risiko kehamilan. Agar proses ini lebih sistematis, tahap pemodelan dibagi menjadi lima sub-tahap, yaitu: penentuan pendekatan pemodelan, pema-

gian data, penerapan algoritma, pelatihan dan pengujian model, serta evaluasi awal hasil model.

Alur pemodelan klasifikasi digambarkan pada Gambar 3.2 Tahapan ini lebih teknis dibanding metodologi umum, karena menjelaskan bagaimana dataset diproses hingga menghasilkan model terbaik.



Gambar 3.2. Alur pemodelan prosedur data

3.5.1 Dataset

yang digunakan dalam penelitian ini adalah *Maternal Health Risk Dataset* yang bersumber dari Kaggle. Dataset ini dipilih karena memuat informasi medis ibu hamil yang relevan dengan tujuan yang memiliki parameter seperti usia, tekanan darah, kadar gula darah, hemoglobin, suhu tubuh, dan detak jantung penelitian. Jumlah data sebanyak 1.822 record dianggap cukup representatif untuk membangun dan melatih model klasifikasi. Dataset ini kemudian diubah ke dalam format CSV agar dapat diolah menggunakan bahasa pemrograman Python dengan bantuan pustaka *scikit-learn, pandas, dan numpy*.

3.5.2 Data Understanding (Eksplorasi Data)

Tahap data understanding dilakukan untuk memperoleh pemahaman awal terhadap karakteristik dataset. Proses ini meliputi peninjauan struktur dan tipe data, statistik deskriptif, serta identifikasi kelas variabel target *RiskLevel*.

3.5.3 Pemeriksaan Data (*Missing Value*)

Pemeriksaan data dilakukan untuk mendeteksi keberadaan nilai hilang pada setiap atribut dalam dataset. Jika Hasil pemeriksaan menunjukkan bahwa seluruh data bersifat lengkap maka tidak diperlukan proses penanganan missing value pada tahap selanjutnya.

3.5.4 Preprocessing Data

Pada tahap *preprocessing*, variabel target *RiskLevel* yang masih berbentuk data kategorikal ditransformasikan ke dalam bentuk numerik menggunakan metode *Label Encoding*. Proses ini dilakukan agar data dapat diproses oleh algoritma *machine learning*.

1. Label Encoding

Tahap ini merupakan dari tahap *preprocessing* data yang bertujuan untuk menyiapkan data agar siap digunakan dalam proses pelatihan model. Langkah yang dilakukan yaitu Encoding variabel kategori menggunakan *LabelEncoder* agar seluruh atribut dapat diolah oleh algoritma berbasis numerik.

2. Normalisasi

Sebelum data digunakan untuk pelatihan model pembelajaran mesin, *preprocessing* data adalah tahap penting yang bertujuan untuk meningkatkan kualitas data serta menyesuaikan data agar dapat di proses secara optimal oleh algoritma *machine learning*. Salah satu tahap *preprocessing* yang dilakukan adalah normalisasi data. Normalisasi bertujuan untuk menyamakan skala dari fitur-fitur numerik sehingga tidak ada fitur yang mendominasi proses pelatihan model hanya karena memiliki rentang nilai yang lebih besar. Dengan data yang telah dinormalisasi, proses pembelajaran model menjadi lebih stabil dan efektif. Metode normalisasi yang digunakan adalah z-score, metode ini menstandarkan data dengan mengubah nilai setiap fitur sehingga memiliki rata-rata mendekati nol dan standar deviasi mendekati satu.(Han dkk., 2022)

Normalisasi dilakukan dengan rumus:

$$z = \frac{x - \mu}{\sigma} \quad (3.1)$$

Penjelasan Rumus Z-Score

x : nilai asli dari suatu data atau fitur

μ adalah nilai rata-rata (*mean*) dari fitur.

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

σ adalah nilai standar deviasi dari fitur.

z : nilai hasil normalisasi *Z-Score*

Nilai *Z-Score* menunjukkan seberapa jauh suatu nilai dari rata-rata fitur dalam satuan standar deviasi:

$Z = 0$ berarti nilai berada tepat di rata-rata.

$Z > 0$: di atas rata-rata

$Z < 0$: di bawah rata-rata

$|Z|$ besar: semakin jauh dari rata-rata

3.5.5 Data Splitting

Data Splitting, langkah berikutnya adalah membagi dataset menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*testing set*). Data latih digunakan untuk membangun dan melatih model, sedangkan data uji digunakan untuk menguji kinerja model. Pada penelitian ini digunakan perbandingan 80:20, artinya 80% data digunakan untuk melatih model, sementara 20% sisanya digunakan untuk menguji model. Teknik ini penting agar model yang dihasilkan tidak hanya menghafal data yang ada (*overfitting*), tetapi juga mampu melakukan prediksi pada data baru. Tahap berikutnya adalah normalisasi, yaitu proses praproses data untuk menyamakan skala antar atribut. Hal ini dilakukan agar perbedaan rentang nilai antar variabel, seperti usia (tahun) dan tekanan darah (mmHg), tidak menimbulkan bias dalam pembentukan model. Normalisasi dilakukan dengan mengubah nilai setiap variabel ke dalam rentang tertentu, misalnya 0–1, sehingga semua fitur memiliki kontribusi yang seimbang.

3.5.6 Pemodelan Algoritma

Tahap pemodelan merupakan inti dari penelitian, di mana data latih digunakan untuk membangun dan menguji kemampuan prediksi lima algoritma klasifikasi dalam mengklasifikasikan risiko kehamilan. Setiap model diimplementasikan dengan parameter yang telah ditentukan untuk menilai kinerja dasar. Proses umum pemodelan meliputi inisialisasi model, pelatihan (fit) pada data latih, prediksi (predict) pada data uji, dan evaluasi hasilnya. Berikut Parameter yang digunakan bisa dilihat pada Tabel 3.3 berikut.

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. Parameter Tiap Algoritma

Algoritma	Nama & Nilai Parameter Utama
Logistic Regression (beberapa parameter yang digunakan)	max_iter = 1000, C = [0.01, 0.1, 1, 10], penalty = [l1, l2], solver = liblinear
Decision Tree (beberapa parameter yang digunakan)	max_depth = [3, 5, 7, None], min_samples_split = [2, 5, 10], criterion = entropy
Random Forest (beberapa parameter yang digunakan)	n_estimators = [50, 100, 150, 200], max_depth = [None, 20], min_samples_split = [2, 5]
SVM (beberapa parameter yang digunakan)	C = [0.1, 1, 10, 100], kernel = [linear, rbf, poly], gamma = [scale, auto], degree = [2, 3]
Naive Bayes (beberapa parameter yang digunakan)	var_smoothing = [1e-9, 1e-8, 1e-7, 1e-6]

Berikut ini Pemodelan dari tiap Algoritma:

1. Logistic Regression

Logistic Regression adalah metode klasifikasi yang memodelkan hubungan antara variabel independen dengan variabel dependen kategorikal menggunakan fungsi *logistik*. Algoritma ini tidak menghasilkan nilai kontinu seperti regresi linier, melainkan probabilitas antara 0–1 yang kemudian dikonversi ke kelas target. Keunggulan utamanya adalah kesederhanaan dan interpretasi yang jelas, menjadikannya pilihan yang baik untuk memberikan gambaran probabilitas risiko kehamilan. (Bertini et al., 2022). Algoritma *Logistic Regression* bekerja dengan memodelkan peluang suatu kelas menggunakan fungsi logit. Model ini sangat sensitif terhadap pengaturan parameter regularisasi untuk mengontrol kompleksitas model. Pada proses ini, *Logistic Regression* diuji menggunakan beberapa nilai C yang mengatur kekuatan regularisasi, yaitu 0.01, 0.1, 1, dan 10. Selain itu, dua metode penalti digunakan: l1 (Lasso) dan l2 (Ridge). Solver liblinear dipilih karena merupakan solver yang kompatibel untuk kedua *penalty* tersebut. Setiap kombinasi parameter diuji dengan cara membangun model *Logistic Regression* menggunakan parameter yang sedang diuji, kemudian melatih model pada data training. Hasil prediksi pada data testing dievaluasi menggunakan akurasi

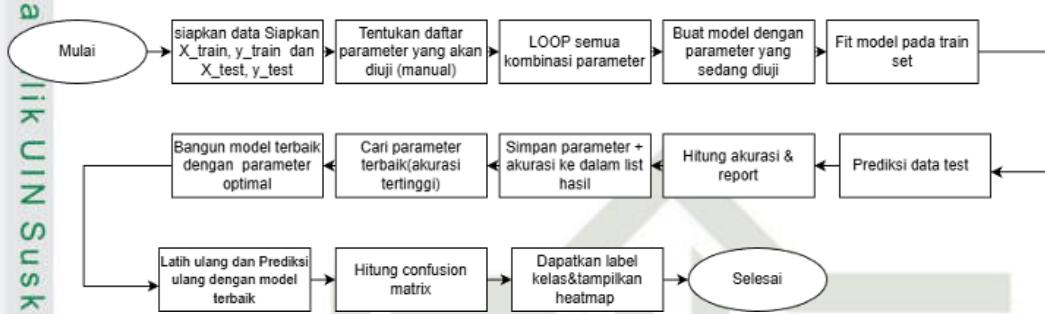
Hak Cipta Dilindungi Undang-Undang

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

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

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

dan *classification report*. Semua hasil disimpan untuk memilih parameter terbaik. Setelah parameter terbaik diperoleh, model dilatih ulang menggunakan parameter tersebut dan dievaluasi menggunakan confusion matrix. Gambar Alur *flowchart Algoritma Logistic Regresion* pada Gambar 3.3 berikut.



Gambar 3.3. Alur flowchart Algoritma Logistic Regresion

Flowchart dimulai dari proses inisialisasi di mana beberapa parameter *Logistic Regression* didefinisikan secara manual. Parameter pertama adalah nilai C, yaitu 0.01, 0.1, 1, dan 10. Nilai C berfungsi sebagai kontrol regularisasi: semakin kecil nilainya, semakin kuat regularisasi diterapkan. Parameter kedua adalah *penalty*, yaitu 'l1' dan 'l2'. *Penalty l1* melakukan seleksi fitur dengan mendorong beberapa koefisien menjadi nol, sedangkan l2 menekan nilai koefisien tetapi tidak menghilangkannya. Solver yang digunakan adalah *liblinear* karena hanya *solver* inilah yang mendukung kedua *penalty* tersebut secara bersamaan. Proses kemudian memasuki tahap looping di mana setiap nilai C dikombinasikan dengan setiap jenis *penalty*. Untuk setiap kombinasi, program mencetak parameter yang sedang diuji, kemudian membangun model *Logistic Regression* dengan parameter tersebut. Nilai *max_iter* ditetapkan 1000 untuk memastikan proses konvergensi stabil. Setelah model dibuat, langkah selanjutnya adalah melatih model menggunakan data *training* *X_train* dan *y_train*. Pelatihan ini menghasilkan koefisien regresi logistik yang digunakan untuk memprediksi kelas risiko kehamilan. Model yang telah terlatih kemudian digunakan untuk memprediksi data uji *X_test*. Hasil prediksi disimpan dalam *y_pred_lr*. Prediksi ini dibandingkan dengan data asli *y_test* untuk menghitung akurasi menggunakan rumus *accuracy_score*. Nilai akurasi kemudian dikalikan dengan 100 sehingga dipresentasikan dalam bentuk persentase. *Classification report* juga ditampilkan, berisi nilai *precision*, *recall*, dan *f1-score* untuk setiap kelas.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak 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.

las yaitu *High Risk* (0), *Low Risk* (1), dan *Mid Risk* (2). Setiap hasil evaluasi, termasuk nilai C, *penalty*, dan akurasi, disimpan dalam list *results_lr* sebagai dictionary. Proses ini berjalan sampai seluruh 8 kombinasi parameter selesai diuji. Setelah itu, program mencari parameter terbaik dengan memilih dictionary yang memiliki nilai akurasi tertinggi. Parameter terbaik ini dicetak, lengkap dengan nilai C, *penalty*, dan akurasinya. Mengikuti identifikasi parameter terbaik, model Logistic Regression dibuat ulang menggunakan nilai C terbaik dan *penalty* terbaik tersebut. Model terbaik ini kemudian dilatih kembali dengan seluruh data training untuk memastikan performanya optimal. Setelah pelatihan, prediksi dilakukan kembali pada *X_test* dan disimpan sebagai *y_pred_best*. Tahap terakhir adalah evaluasi menggunakan confusion matrix yang memperlihatkan jumlah prediksi benar dan salah untuk tiap kelas: *High Risk*, *Low Risk*, dan *Mid Risk*. Label kelas diatur secara manual sesuai mapping dataset. Confusion matrix kemudian divisualisasikan dalam bentuk heatmap menggunakan Seaborn sehingga lebih mudah diamati. Proses berakhir setelah visualisasi selesai, yang menampilkan performa final *Logistic Regression* dengan parameter terbaik.

2. Decision Tree C4.5.

Algoritma Decision Tree C4.5 adalah pengembangan dari ID3 yang menggunakan entropy sebagai dasar pemilihan atribut terbaik dengan menghitung information gain ratio. Pada penelitian ini, pendekatan tersebut diterapkan menggunakan *DecisionTreeClassifier* dengan *criterion* = 'entropy', sehingga mekanisme pemisahan node mengikuti prinsip C4.5. Proses dimulai dengan mengimpor seluruh library yang dibutuhkan, termasuk *DecisionTreeClassifier*, *accuracy_score*, *classification_report*, serta library visualisasi seperti *matplotlib* dan *seaborn*. Selanjutnya, dua parameter utama diuji secara manual, yaitu *max_depth* (kedalaman pohon) dan *min_samples_split* (jumlah minimum sampel untuk membagi sebuah node). Parameter-parameter tersebut diuji dalam beberapa kombinasi, yaitu *max_depth* = [3, 5, 7, None] dan *min_samples_split* = [2, 5, 10]. Setiap kombinasi diuji melalui looping sehingga seluruh parameter dapat dibandingkan. Untuk setiap kombinasi parameter, model Decision Tree dibangun menggunakan nilai *criterion*='entropy' dan *random_state*=42 untuk memastikan reproduktibilitas. Model kemudian dilatih menggunakan data latih (*x_train*, *y_train*), dan hasilnya digunakan untuk melakukan prediksi terhadap data uji (*x_test*). Setelah prediksi dilakukan, nilai akurasi dihitung dan laporan klasifikasi dic-

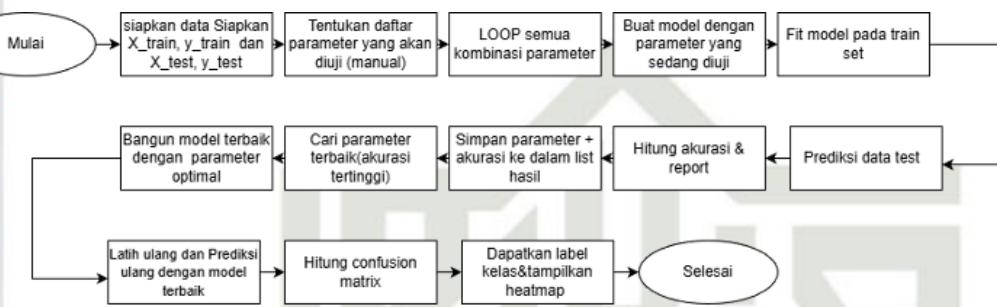
Hak Cipta Dilindungi Undang-Undang

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

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

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

etak agar performa tiap kombinasi parameter dapat terlihat. Seluruh hasil disimpan dalam list agar dapat diseleksi nanti. Di akhir proses, model dengan akurasi tertinggi dipilih sebagai model terbaik, serta confusion matrix dari model terbaik tersebut divisualisasikan agar terlihat pola klasifikasi yang dihasilkan *Decision Tree C4.5*. Berikut Gambar Alur flowchart Algoritma *Decision Tree C4.5*. Gambar alur flowchart Algoritma *Decision Tree C4.5* pada Gambar 3.4 berikut.



Gambar 3.4. Alur Flowchart Algoritma *Decision Tree C4.5*

Proses pengujian algoritma *Decision Tree C4.5* dimulai dengan melakukan inisialisasi library yang diperlukan untuk pemodelan, evaluasi, dan visualisasi. Selanjutnya, penelitian ini menguji dua parameter utama, yaitu *max_depth* dan *min_samples_split*, yang sangat mempengaruhi bentuk dan kompleksitas pohon keputusan. Nilai *max_depth* diuji pada 3, 5, 7, dan None, sedangkan nilai *min_samples_split* diuji pada 2, 5, dan 10. Seluruh kombinasi parameter tersebut diuji menggunakan proses looping sehingga total 12 model diuji. Pada setiap iterasi, model Decision Tree dibuat menggunakan *criterion='entropy'*, yang merepresentasikan metode C4.5 dalam menentukan pemilihan atribut terbaik. Model dilatih menggunakan data training, dan hasil pelatihan digunakan untuk memprediksi kelas pada data uji. Nilai akurasi dihitung dengan mengalikan skor dengan 100 untuk mendapatkan persentase, kemudian classification report dicetak untuk menilai performa tiap kelas. Semua hasil pengujian disimpan untuk dibandingkan, setelah itu model dengan akurasi tertinggi dipilih dan ditetapkan sebagai model terbaik. Parameter terbaik yang diperoleh meliputi nilai *max_depth* dan *min_samples_split* yang menghasilkan performa optimal. Untuk model terbaik tersebut, confusion matrix dihitung untuk melihat distribusi kesalahan prediksi antar kelas. Matriks ini divisualisasikan menggunakan heatmap untuk memberikan gambaran yang lebih jelas mengenai

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.

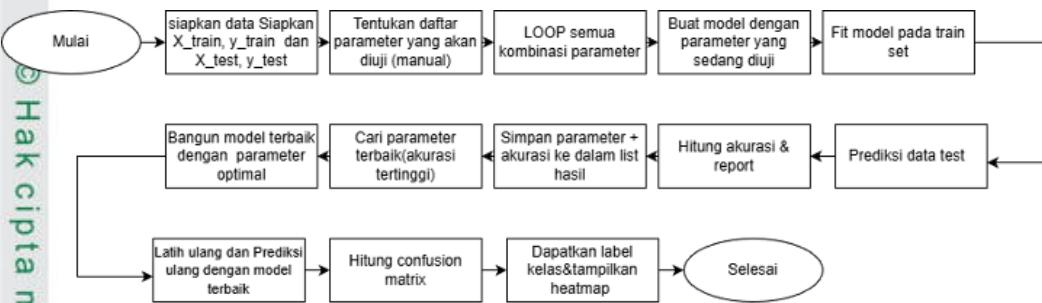
kinerja model dalam membedakan kelas *high risk, low risk, dan mid risk*. Dengan demikian, seluruh proses mulai dari pengujian beberapa parameter, pemilihan parameter terbaik, hingga visualisasi confusion matrix memberikan gambaran yang komprehensif tentang bagaimana algoritma *Decision Tree C4.5* bekerja dalam memprediksi risiko kehamilan berdasarkan dataset yang digunakan.

3. Random Forest

Random Forest merupakan pengembangan dari Decision Tree yang bekerja dengan cara membangun banyak pohon keputusan secara acak dan menggabungkan hasilnya untuk memperoleh prediksi akhir. Setiap pohon dilatih menggunakan subset data dan subset fitur yang dipilih secara acak (*bootstrap sampling*). Prediksi akhir ditentukan melalui voting mayoritas dari seluruh pohon. Sifatnya yang robust dan kemampuannya menangani data medis yang kompleks membuatnya menjadi pilihan yang kuat. Pada penelitian ini, proses pengujian dimulai dengan menentukan beberapa kombinasi parameter yang ingin diuji, yaitu jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), jumlah minimum sampel untuk pemisahan node (*min_samples_split*), dan minimum sampel pada daun (*min_samples_leaf*). Setiap kombinasi parameter diuji melalui proses iteratif (looping), di mana model RandomForestClassifier dibangun, dilatih menggunakan data latih, dan menghasilkan prediksi pada data uji untuk dihitung akurasinya. Laporan klasifikasi dicetak untuk mengetahui performa tiap kelas risiko. Seluruh hasil disimpan dalam list untuk menentukan parameter terbaik berdasarkan akurasi tertinggi. Setelah parameter optimal diperoleh, model Random Forest terbaik dibangun ulang dan digunakan untuk prediksi final. Prediksi ini dihitung *confusion matrix*-nya dan divisualisasikan menggunakan heatmap untuk mengetahui pola kesalahan model dalam memprediksi kelas *high risk, low risk, dan mid risk*.

Gambar Alur flowchart Algoritma Random Forest pada Gambar 3.5 berikut.

UIN SUSKA RIAU



Gambar 3.5. Alur flowchart Algoritma Random Forest

Proses dimulai dari tahap import library, yaitu `RandomForestClassifier` untuk membangun model dan `accuracy_score` serta `classification_report` untuk evaluasi performa. Setelah itu, ditentukan empat kelompok parameter yang akan diuji: `n_estimators` (50 hingga 200), `max_depth` (None dan 20), `min_samples_split` (2 dan 5), serta `min_samples_leaf` (1). Setiap kombinasi parameter diuji menggunakan struktur looping bertingkat, di mana pada setiap iterasi model Random Forest baru diinisialisasi menggunakan parameter tersebut. Model kemudian dilatih pada `X_train` dan `y_train`, sehingga terbentuk kumpulan pohon keputusan dengan struktur yang berbeda-beda tergantung kombinasi parameter. Setelah pelatihan selesai, model digunakan untuk memprediksi `X_test`, menghasilkan `output y_pred` yang kemudian dibandingkan dengan `y_test` untuk menghitung akurasi. `Classification report` juga dicetak untuk mengevaluasi presisi, `recall`, dan `F1-score` pada masing-masing kelas risiko (*high*, *low*, dan *mid*). Seluruh hasil, termasuk nilai parameter dan akurasi, disimpan ke dalam list `results` untuk kepentingan pemilihan model terbaik. Setelah seluruh parameter diuji, proses berlanjut pada tahap pemilihan parameter terbaik, yaitu kombinasi yang menghasilkan akurasi tertinggi. Parameter terbaik ini kemudian digunakan untuk membangun model Random Forest baru yang dilatih ulang pada data latih. Prediksi baru (`y_pred_rf`) dihasilkan berdasarkan model terbaik tersebut. Selanjutnya dilakukan perhitungan `confusion matrix` untuk melihat seberapa baik model dalam membedakan tiga kelas risiko kehamilan. `Confusion matrix` divisualisasikan dalam bentuk `heatmap` menggunakan `seaborn` dengan label kelas *high risk*, *low risk*, dan *mid risk*. Visualisasi ini membantu mengidentifikasi pola kesalahan model, apakah model lebih sering keliru pada kelas tertentu atau sudah mampu mengenali kelas dengan baik. Alur berakhir setelah seluruh evaluasi selesai dilakukan, menghasilkan model `Random Forest` terbaik yang dapat dianalisis lebih lanjut.

Hak Cipta Dilindungi Undang-Undang

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

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

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

Hak Cipta Dilindungi Undang-Undang

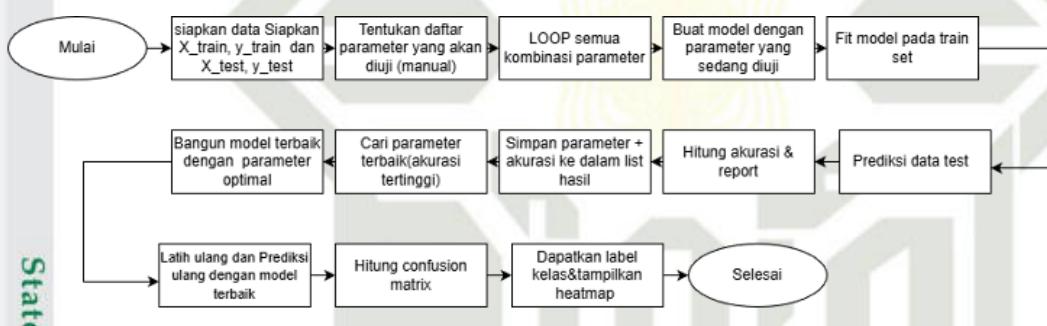
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

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

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

4. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma supervised learning yang bertujuan menemukan hyperplane terbaik untuk memisahkan kelas dengan margin maksimum. SVM efektif digunakan untuk klasifikasi linier maupun non-linier melalui pemilihan fungsi kernel yang sesuai. Dalam penelitian ini, *SVM* diterapkan untuk mengklasifikasikan tingkat risiko kehamilan menjadi tiga kategori: rendah, sedang, dan tinggi. Model diinisialisasi menggunakan kelas SVC dengan konfigurasi parameter *default*, termasuk *kernel* = *rbf* (*radial basis function*), *C* = 1.0, dan *gamma* = *scale*. Selanjutnya, model dilatih pada data latih (*X_train*, *y_train*) untuk menemukan hyperplane pemisah yang optimal. Hasil prediksi (*y_pred_svm*) pada data uji digunakan untuk mengevaluasi kinerja model melalui metrik akurasi dan laporan klasifikasi. Pemilihan kernel RBF membantu menangani non-linearitas pada data kesehatan ibu hamil, sehingga SVM diharapkan mampu memberikan prediksi yang akurat terhadap risiko komplikasi kehamilan. Gambar Alur flowchart Algoritma Support Vector Machine pada Gambar 3.6 berikut.



Gambar 3.6. Alur flowchart Algoritma Support Vector Machine

Alur proses algoritma SVM dimulai dengan mengimpor *library* yang dibutuhkan, termasuk SVC dari *sklearn.svm*, *accuracy_score* dan *classification_report* untuk evaluasi, serta *matplotlib* dan *seaborn* untuk visualisasi. Selanjutnya, beberapa parameter *SVM* didefinisikan, yaitu *C* (regulasi kesalahan), *kernel* (*linear*, *rbf*, *poly*), *gamma* (*scale*, *auto*), dan *degree* (hanya untuk *kernel poly*). Proses pelatihan dilakukan menggunakan loop untuk setiap kombinasi parameter. Jika *kernel* yang dipilih adalah *poly*, maka parameter *degree* juga dicoba, sedangkan untuk *kernel linear* atau *rbf*, *degree* tidak digunakan. Setiap kombinasi parameter membentuk model *SVM* yang diinisialisasi dengan nilai yang sesuai. Model kemudian dilatih pada data

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.

pelatihan (X_train , y_train) dan melakukan prediksi pada data uji (X_test). Setelah prediksi, akurasi dihitung menggunakan *accuracy_score* dan laporan klasifikasi dicetak melalui *classification_report* untuk melihat performa pada masing-masing kelas. Semua hasil parameter dan akurasinya disimpan ke dalam *list results_svm*. Setelah seluruh kombinasi diuji, parameter terbaik ditentukan berdasarkan nilai akurasi tertinggi dari list tersebut. Terakhir, model dengan parameter terbaik digunakan untuk membuat prediksi ulang, kemudian dihitung confusion matrix dan divisualisasikan dalam bentuk heatmap menggunakan seaborn untuk melihat performa klasifikasi tiap kelas secara jelas. Proses ini memastikan bahwa model SVM yang dihasilkan memiliki performa optimal berdasarkan pengujian manual parameter, sekaligus memberikan interpretasi yang jelas mengenai klasifikasi dan kesalahan prediksi.

5. Naive Bayes

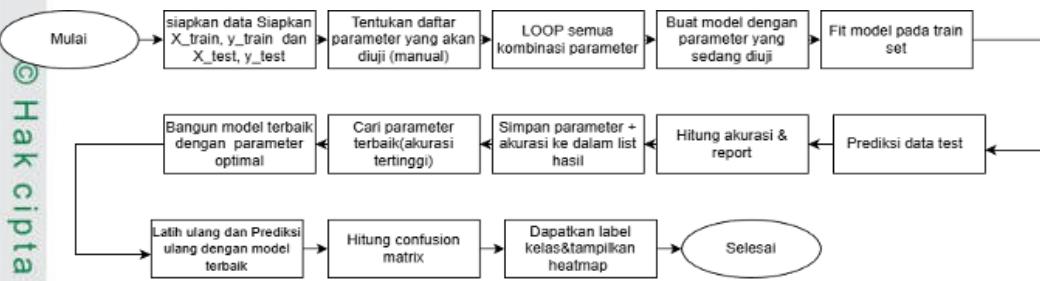
Algoritma *Naive Bayes* merupakan metode klasifikasi berbasis probabilistik yang menggunakan Teorema *Bayes* dengan asumsi bahwa setiap fitur saling independen. Pada Gaussian Naive Bayes, data diasumsikan mengikuti distribusi normal (*Gaussian*). Salah satu parameter penting dalam GaussianNB adalah *var_smoothing*, yaitu nilai kecil yang ditambahkan ke varians agar model tetap stabil saat menemukan data dengan varians yang sangat kecil. Dalam implementasi ini, beberapa nilai *var_smoothing* diuji secara manual untuk mencari kombinasi terbaik. Proses diawali dengan menyiapkan daftar nilai *var_smoothing* ($1e-9$, $1e-8$, $1e-7$, $1e-6$), kemudian model dilatih dan dievaluasi satu per satu menggunakan data latih dan data uji. Setiap hasil yang diperoleh disimpan, lalu dipilih nilai *var_smoothing* dengan akurasi tertinggi sebagai parameter terbaik. Model kemudian dibangun ulang menggunakan parameter tersebut dan dievaluasi kembali menggunakan confusion matrix untuk melihat performa secara menyeluruh antar kelas. Berikut alur *flowchart* dari proses pengujian Gaussian Naive Baye. (Khoirunnisa dan Lestari, 2023). Gambar alur Flowchart Algoritma Naive Bayes dilihat pada Gambar 3.7 berikut.

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.



Gambar 3.7. Alur flowchart Algoritma *Naive Bayes*

dan dievaluasi kembali menggunakan confusion matrix untuk melihat performa secara menyeluruh antar kelas. Berikut alur *flowchart* dari proses pengujian *Gaussian Naive Bayes*. (Khoirunnisa dan Lestari, 2023).

Flowchart dimulai dari Start, kemudian proses pertama adalah melakukan impor *library* yang dibutuhkan, yaitu *GaussianNB* dari *sklearn.naive_bayes* untuk membangun model *Naive Bayes*, serta *accuracy_score* dan *classification_report* untuk evaluasi model. Kemudian *matplotlib* dan *seaborn* digunakan untuk visualisasi *Confusion Matrix*. Setelah *library* siap, sistem menyiapkan daftar parameter yang akan diuji yaitu *var_smoothing* = $[1e-9, 1e-8, 1e-7, 1e-6]$. Nilai-nilai ini dipilih karena *var_smoothing* mengontrol kestabilan perhitungan probabilitas dan biasanya berada pada skala $1e-n$. Selanjutnya, proses masuk ke tahap looping. Untuk setiap nilai *var_smoothing*, model *GaussianNB* dibuat menggunakan konfigurasi *GaussianNB*(*var_smoothing* = nilai_saat_ini). Setelah model dibuat, proses pelatihan dilakukan menggunakan *X_train* dan *y_train*, sehingga model mempelajari hubungan antar fitur dan label berdasarkan asumsi distribusi Gaussian. Setelah proses pelatihan selesai, model digunakan untuk melakukan prediksi terhadap *X_test*, menghasilkan nilai *y_pred_nb*. Hasil prediksi ini kemudian dibandingkan dengan label sebenarnya (*y_test*) untuk menghitung akurasi menggunakan *accuracy_score* $\times 100$, sehingga nilai akurasi ditampilkan dalam persen. Pada setiap iterasi, juga ditampilkan *Classification Report*, yang berisi nilai precision, recall, f1-score, dan support untuk setiap kelas (*high risk*, *low risk*, *mid risk*). Seluruh hasil, termasuk nilai *var_smoothing* dan akurasi yang diperoleh, disimpan dalam list *results_nb* agar bisa dibandingkan di akhir proses. Setelah seluruh iterasi selesai, sistem mencari parameter terbaik menggunakan fungsi *max()*, yaitu nilai *var_smoothing* dengan akurasi tertinggi. Nilai inilah yang akan digunakan untuk membangun model final. Tahap selanjutnya adalah membuat

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.

ulang model *GaussianNB* menggunakan nilai *var_smoothing* terbaik. Model tersebut kembali dilatih dengan data latih, kemudian melakukan prediksi ulang terhadap data uji. Hasil prediksi terbaik ini digunakan untuk menghitung Confusion Matrix menggunakan fungsi *confusion_matrix*. Selanjutnya, Confusion Matrix divisualisasikan dalam bentuk heatmap dengan label kelas manual yaitu *high risk (0)*, *low risk (1)*, dan *mid risk (2)*. Visualisasi ini memberikan gambaran jumlah prediksi benar dan salah pada setiap kelas, sehingga memudahkan dalam mengevaluasi apakah model *Naive Bayes* lebih cenderung salah pada kelas tertentu. Alur berakhir pada End, menekankan bahwa proses eksplorasi parameter sederhana seperti *var_smoothing* pada *Gaussian Naive Bayes* dapat meningkatkan performa model, namun tetap harus diinterpretasikan hati-hati karena *Naive Bayes* memiliki asumsi independensi fitur yang tidak selalu sesuai dengan karakteristik dataset kesehatan.

3.5.7 Hasil Klasifikasi Model

Hasil Klasifikasi, setelah pemodelan dilakukan dengan setiap algoritma, diperoleh hasil berupa prediksi tingkat risiko kehamilan untuk data uji. Setiap model menghasilkan prediksi terhadap tiga kategori risiko kehamilan, yaitu rendah, sedang, dan tinggi, sesuai dengan label pada dataset *Maternal Health Risk* dari *Kaggle*. Hasil prediksi dibandingkan dengan label aktual untuk menghitung nilai akurasi, presisi, *recall*, dan *f1-score*, yang digunakan sebagai dasar evaluasi performa model.

3.5.8 Evaluasi Model

Tahap terakhir adalah evaluasi performa model. Evaluasi dilakukan dengan menggunakan *Confusion Matrix* yang menunjukkan jumlah prediksi benar dan salah pada tiap kelas (*Low*, *Mid*, *High Risk*). Selain itu, digunakan juga metrik evaluasi lain, yaitu:

1. Akurasi (*Accuracy*)

Mengukur proporsi prediksi yang benar dari data uji keseluruhan. Akurasi tinggi menunjukkan bahwa model mampu mengklasifikasikan data secara umum dengan benar.

2. Presisi (*Precision*)

Menghitung jumlah prediksi positif yang relevan. Untuk menghindari kesalahan dalam mengidentifikasi ibu hamil yang tidak berisiko sebagai berisiko, sangat penting untuk memiliki presisi tinggi.

3. Recall (*Sensitivity*)
Menghitung jumlah kasus risiko kehamilan yang berhasil dideteksi model. Jumlah ulang yang tinggi menunjukkan bahwa model peka terhadap kasus risiko kehamilan.
4. *F1-Score*
F1-Score digunakan ketika terjadi ketidakseimbangan kelas (*imbalance class*), seperti jumlah data risiko ibu hamil yang lebih sedikit dibandingkan tidak berisiko. *F1-Score* adalah rata-rata harmonis dari *precision* dan *recall*.

3.6 Seleksi Model Terbaik

Tahap akhir dalam proses pengujian model pengajaran mesin adalah pemilihan model terbaik. Pada tahap ini, hasil evaluasi sebelumnya digunakan untuk memilih model dengan performa terbaik dengan membandingkan nilai metrik seperti akurasi, *precision*, *recall*, dan skor F1. Untuk mengklasifikasikan status risiko kehamilan pada ibu hamil di kelas mayoritas dan minoritas, model yang dipilih harus memberikan hasil prediksi yang paling akurat dan konsisten. Keputusan ini juga dapat dipengaruhi oleh faktor lain, seperti kompleksitas model, waktu komputasi, dan interpretabilitas. Dengan memilih model terbaik, diharapkan model tersebut dapat digunakan sebagai alat analisis yang kuat untuk mendeteksi risiko kehamilan pada ibu hamil dan dapat digunakan dalam studi lanjutan atau proses pengambilan keputusan oleh pihak terkait.

5.1 Kesimpulan

Hasil penelitian ini mengenai prediksi risiko kesehatan ibu hamil menggunakan lima algoritma machine learning: Random Forest, Decision Tree C4.5, Support Vector Machine (SVM), Logistic Regresion, dan Naive Bayes. Proses evaluasi model dengan metode hold-out 80:20. Berdasarkan hasil pengolahan dan analisa data, peneliti menghasilkan beberapa kesimpulan berikut:

1. Penelitian ini berhasil menganalisis dan membandingkan lima algoritma machine learning, yaitu Logistic Regression, Decision Tree C4.5, Random Forest, Support Vector Machine (SVM), dan Naive Bayes, dalam memprediksi tingkat risiko kehamilan menggunakan data kesehatan ibu hamil. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-Score. Hasil pengujian menunjukkan bahwa setiap algoritma memiliki kemampuan yang berbeda dalam mengenali pola data dan mengklasifikasikan risiko kehamilan ke dalam kategori rendah, sedang, dan tinggi. Algoritma berbasis pohon, khususnya Random Forest dan Decision Tree C4.5, menunjukkan performa yang lebih stabil dibandingkan algoritma lainnya, sedangkan SVM, Logistic Regression, dan Naive Bayes memiliki kinerja yang relatif lebih rendah.
 2. Berdasarkan hasil perbandingan kinerja, Random Forest merupakan algoritma dengan performa terbaik, dengan nilai akurasi 93.42%, presisi 93.37%, recall 93.42%, dan F1-Score 93.30%. Model ini mampu mengklasifikasikan seluruh kelas risiko secara konsisten dan menunjukkan kemampuan yang unggul dalam menangani pola data yang kompleks. Decision Tree C4.5 dengan akurasi 93.15%, presisi 93.02%, recall 93.15%, dan F1-Score 93.01%, berada pada peringkat kedua dengan performa yang mendekati Random Forest dan memiliki keunggulan dalam kemudahan interpretasi model. Kelebihan Decision Tree C4.5 terletak pada struktur model yang lebih sederhana dan mudah dipahami, sehingga dapat menjadi alternatif yang lebih praktis untuk implementasi. Sementara itu, Support Vector Machine menunjukkan performa menengah dengan akurasi 81.92%, presisi 80.56%, recall 81.92%, dan F1-Score 79.61%. Sedangkan Logistic Regression akurasi 75.89%, presisi 77.65%, recall 75.89%, dan F1-Score 70.15%, dan Naive Bayes akurasi 72.33%, presisi 66.73%, recall 72.33%, dan F1-Score 69.61%.

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.

Score 67.77%. memiliki performa paling rendah. Hal ini menunjukkan bahwa kedua algoritma terakhir kurang optimal dalam memodelkan hubungan nonlinier pada data risiko kehamilan.

Saran

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan, Peneliti mengajukan beberapa saran sebagai berikut:

1. Menguji algoritma lain seperti XGBoost, LightGBM, AdaBoost, atau KNN untuk mengetahui apakah terdapat model yang dapat memberikan performa lebih baik dari Random Forest.
2. Mengembangkan sistem prediksi risiko kehamilan berbasis aplikasi web atau mobile agar dapat dimanfaatkan oleh tenaga kesehatan sebagai alat bantu deteksi dini.

DAFTAR PUSTAKA

- 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.
- Afolabi, A. O., Esienumoh, E., dan Afolabi, K. A. j. . U. (2023). Risk Assessment among Pregnant Women in Nigeria : Requisite Step towards Reduced Maternal Morbidity and Mortality. , 8(2).
- Analia, H., Rahmadanti, R., Syaiin, A., Salsabila4, S., Yunita, Y., dan Sriyadi, S. j. . U. (2023). Prediksi Resiko Kesehatan Ibu Hamil Dengan Menggunakan Metode Decision Tree. *Swabumi*, 11(1), 48–53. doi: 10.31294/swabumi.v11i1.15270
- An, Q., Rahman, S., Zhou, J., dan Kang, J. J. j. . U. (2023). A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. *Sensors*, 23(9). doi: 10.3390/s23094178
- Assaduzzaman, A. M., Salim, S., Tafakori, L., dan Abdollahian, M. j. . U. (2024). Predicting maternal risk level using machine learning models. *BMC Pregnancy and Childbirth*, 24(1). Retrieved from <https://doi.org/10.1186/s12884-024-07030-9> doi: 10.1186/s12884-024-07030-9
- Bisong, E. (2019). Google colaboratory. Dalam *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners* (hal. 59–64). Springer.
- Bisson, C., Dautel, S., Patel, E., Suresh, S., Dauer, P., dan Rana, S. j. . U. (2023). Preeclampsia pathophysiology and adverse outcomes during pregnancy and postpartum. *Frontiers in Medicine*, 10(March), 1–10. doi: 10.3389/fmed.2023.1144170
- Butwall, M. j. . U. (2021). Data Normalization and Standardization : Impacting Classification Model Accuracy. , 183(35), 8–11.
- Cavazos-rehg, P. A., Krauss, M. J., Spitznagel, E. L., Bommarito, K., Madden, T., Olsen, M. A., ... Bierut, L. J. j. . U. (2016). HHS Public Access. , 19(6), 1202–1211. doi: 10.1007/s10995-014-1624-7.Maternal
- Cheng, Y., Chen, K., Sun, H., Zhang, Y., dan Tao, F. j. . U. (2017). Data and Knowledge Mining with Big Data towards Smart Production. *Journal of Industrial Information Integration*. Retrieved from <http://dx.doi.org/10.1016/j.jii.2017.08.001> doi: 10.1016/j.jii.2017.08.001
- Chicco, D., dan Jurman, G. j. . U. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1), 6.

- chine Learning Models for Predicting Risky Pregnancies in Early Clinical Interventions. , 2024.
- Hawale, D., Chavan, A., Timalsina, D., dan Thatere, A. B. j. . U. (2024). The Role of Artificial Intelligence in Healthcare: A Review. *AIP Conference Proceedings*, 3188(1), 1–22. doi: 10.1063/5.0240194
- Hennessy, A., Tran, T. H., Sasikumar, S. N., dan Al-Falahi, Z. j. . U. (2024). Machine learning, advanced data analysis, and a role in pregnancy care? How can we help improve preeclampsia outcomes? *Pregnancy Hypertension*, 37(March), 101137. Retrieved from <https://doi.org/10.1016/j.preghy.2024.101137> doi: 10.1016/j.preghy.2024.101137
- idayat, R., dan Astuti, T. j. . U. (2020). Diagnosis Preeklamsia pada Ibu Hamil Berdasarkan Algoritme K- Nearest Neighbour. *Jurnal Ilmu-ilmu Informatika dan Manajemen STMIK*, 14(2), 106–116.
- Hilbe, J. M. (2017). *Testing and Fitting a Logistic Model*. doi: 10.1201/b18678-7
- Irfan, M., Basuki, S., dan Azhar, Y. j. . U. (2021). Giving more insight for automatic risk prediction during pregnancy with interpretable machine learning. *Bulletin of Electrical Engineering and Informatics*, 10(3), 1621–1633. doi: 10.11591/eei.v10i3.2344
- Jefri Jaya journal = Unknown Journal . (2025). Perbandingan Algoritma Machine Learning Untuk Prediksi Tingkat Risiko Ibu Hamil. *Jurnal Ilmu Komputer dan Sistem Informasi*, 13(1). doi: 10.24912/jiksi.v13i1.32856
- journal = Unknown Journal, W. (2008). Education material for teachers of midwifery.
- Key, C. j. . U. (2025). The gender data landscape highlights opportunities for enhancing gender equality outcomes in Indonesia to support productivity and wealth gains, reduce poverty and increase shared prosperity. (1), 2020–2021.
- Khalil, A., Syngelaki, A., Maiz, N., Zinevich, Y., dan Nicolaides, K. H. j. . U. (2013). Maternal age and adverse pregnancy outcome : a cohort study. (March), 634–643. doi: 10.1002/uog.12494
- Khan, S., Bakkar, A., Hossain, A. T., Zohora, F. T., Arifeen, S. E., Rahman, A. E., dan Jamil, K. j. . U. (2023). Preeclampsia and eclampsia-specific maternal mortality in Bangladesh : Levels , trends , timing , and care-seeking practices. , 13, 1–12. doi: 10.7189/jogh.13.07003
- Khoirunnisa, V., dan Lestari, S. j. . U. (2023). Implementasi Klasifikasi Kehamilan Beresiko Dengan Metode Naive Bayes Pada Puskesmas Kelurahan Malaka Jaya. *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, 4(3),

- 1680–1693. doi: 10.35870/jimik.v4i3.396
- Kurniawan, A. j. . U. (2025). Perbandingan Optimasi Algoritma Klasifikasi Decision Tree , Naive Bayes dan KNN Menggunakan Optimize Parameter Grid Pada Tingkat Resiko Ibu Hamil. , 18(2), 111–119.
- Lestari, D., Maulana, F. I., Persada, S. F., dan Adi, P. D. P. (2023). Machine learning for perinatal complication prediction: A systematic review. Dalam *International conference on information, communication and computing technology* (hal. 789–803). Springer.
- Magee, L. A., Brown, M. A., Hall, D. R., Gupte, S., Hennessy, A., Karumanchi, S. A., ... Dadelszen, P. V. j. . U. (2022). Pregnancy Hypertension : An International Journal of Women ' s Cardiovascular Health The 2021 International Society for the Study of Hypertension in Pregnancy classification , diagnosis management recommendations for international practice . *Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health*, 27(September 2021), 148–169. Retrieved from <https://doi.org/10.1016/j.preghy.2021.09.008> doi: 10.1016/j.preghy.2021.09.008
- Malde, A., Prabhu, V. G., Banga, D., Hsieh, M., Renduchintala, C., dan Pirrallo, R. j. . U. (2025). A Machine Learning Approach for Predicting Maternal Health Risks in Lower-Middle-Income Countries Using Sparse Data and Vital Signs.
- Modi, N., dan Kumar, Y. (2025). Automated Machine Learning-Based System for the Prediction of Maternal Health Indicators and High-Risk Pregnancy. Dalam *2025 7th international conference on energy, power and environment (icepe)* (hal. 1–6). IEEE.
- Morales, E. F., dan Escalante, H. J. j. . U. (2021). A brief introduction to supervised, unsupervised, and reinforcement learning. *Biosignal Processing and Classification Using Computational Learning and Intelligence: Principles, Algorithms, and Applications*(January 2022), 111–129. doi: 10.1016/B978-0-12-820125-1.00017-8
- More, V. S. j. . U. (2017). Fever in pregnancy and its maternal and fetal outcomes. , 6(12), 5523–5527.
- Mustamin, N. F., Aziz, F., Firmansyah, F., dan Ishak, P. j. . U. (2023). Classification Of Maternal Health Risk Using Three Models Naive Bayes Method. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 17(4), 395. doi: 10.22146/ijccs.84242
- Mutlu, H. B., Yü journal = Unknown Journal cel, N., Durmaz, F., Cengİl, E., dan

- Yildirim, M. (2023). Prediction of Maternal Health Risk with Traditional Machine Learning Methods. , 4(1), 16–23. doi: 10.46572/naturengs.1293185
- Nurhalizah, R. S., Ardianto, R., dan Purwono, P. j. . U. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer dan Informatika*, 4(1), 61–72.
- Nurjanah, N., Nurfitia, N. R., dan Magasida, D. j. . U. (2025). Asuhan sayang ibu (ASIH) dalam optimalisasi kesehatan ibu hamil melalui interprofesional collaboration dan peningkatan pengetahuan deteksi dini dan komplikasi kehamilan di klinik pratama akbid muhammadiyah cirebon. *Jurnal PKM*, 1(2), 7–13.
- Nwokoro, C. O., Duke, O. A., dan Nwokoro, K. C. j. . U. (2024). Predicting Maternal Outcomes Using Tree-based Methods in Machine Learning. , 0–14.
- Octaviani, D. A., Widiastuti, D., Amelia, R., Salam, A., dan Nuswantoro, U. D. j. . U. (2025). Implementasi Data Mining Untuk Memprediksi Dalam Kehamilan Poltekkes Kemenkes Semarang , Indonesia Implementation Of Data Mining To Predict Pre-Eclampsia In Pregnancy Alamat korespondensi : , 18(1), 29–38. doi: 10.32763/ps5qpj75
- Oktarina, E. S., Alamsyah, G., Nurhalissa, R., dan Satria, R. F. j. . U. (2025). Transformasi Perawatan Kesehatan Ibu Hamil dengan IoT: Solusi Cerdas untuk Pemantauan Real-Time di Daerah Terpencil. *Jurnal Algoritma*, 22(1), 458–467. doi: 10.33364/algoritma/v.22-1.2290
- Pawar, L., Malhotra, J., Sharma, A., Arora, D., dan Vaidya, D. (2022). A robust machine learning predictive model for maternal health risk. Dalam *2022 3rd international conference on electronics and sustainable communication systems (icesc)* (hal. 882–888). IEEE.
- Prajoko, A., Sembiring, R. W., dan S, S. j. . U. (2021). Penerapan Algoritma C4.5 Dalam Klasifikasi Penerima Kartu Keluarga Sejahtera (KKS). *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 6(1), 171. doi: 10.30645/jurasik.v6i1.281
- Purnama, J. J., Hikmawati, N. K., dan Rahayu, S. j. . U. (2024). Hamil, kesehatan ibu. *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Analisis Algoritma Klasifikasi Untuk Mengidentifikasi Potensi Risiko*, 5(1), 120–127.
- Pusadan, M. Y., Ghifari, A., dan Anshori, Y. j. . U. (2023). Implementasi Data Mining untuk Prediksi Status Proses Persalinan pada Ibu Hamil Menggu-

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.

nakan Algoritma Naive Bayes. *Technomedia Journal*, 8(1 Juni), 137–153. doi: 10.33050/tmj.v8i1.1980

Rabbi, F., dan MEM, J. C. (2023). Predicting Maternal Health Risk by Utilizing Different Machine Learning Classifiers. Dalam *Proceedings of the international annual conference of the american society for engineering management*. (hal. 1–10). American Society for Engineering Management (ASEM).

Rahman, A., dan Alam, M. G. R. (2023). Explainable AI based maternal health risk prediction using machine learning and deep learning. Dalam *2023 ieee world ai iot congress (aiiot)* (hal. 13–18). IEEE.

Rahman, R. A. F., Harahap, W. A., dan Kurnia, W. (2023). Analisis Perbandingan Algoritma Machine Learning untuk Klasifikasi Tingkat Risiko Ibu Hamil. *Student Research Journal*, 1(6), 246–261.

Rahman, S. A., Ibtisum, S., Bazgir, E., dan Barai, T. j. . U. (2023). The Significance of Machine Learning in Clinical Disease Diagnosis: A Review. *International Journal of Computer Applications*, 185(36), 10–17. doi: 10.5120/ijca2023923147

Raja, R., Mukherjee, I., dan Sarkar, B. K. j. . U. (2021). A Machine Learning-Based Prediction Model for Preterm Birth in Rural India. *Journal of Healthcare Engineering*, 2021. doi: 10.1155/2021/6665573

Rajagopal, A., Ayanian, S., Ryu, A. J., Qian, R., Legler, S. R., Peeler, E. A., ... Kawamoto, K. j. . U. (2024). Machine Learning Operations in Health Care: A Scoping Review. *Mayo Clinic Proceedings: Digital Health*, 2(3), 421–437. Retrieved from <https://doi.org/10.1016/j.mcpdig.2024.06.009> doi: 10.1016/j.mcpdig.2024.06.009

Sadikin, B. G. j. . U. (2023). Laporan Kinerja Kementerian Kesehatan RI. *Laporan Kinerja Kementerian Kesehatan RI*, 1–23.

Saitta, L. j. . U. (2023). Support-Vector Networks. , 297, 273–297.

Sandri, M., dan Zuccolotto, P. j. . U. (2008). A bias correction algorithm for the gini variable importance measure in classification trees. *Journal of Computational and Graphical Statistics*, 17(3), 611–628. doi: 10.1198/106186008X344522

Scott, A. J., Hosmer, D. W., dan Lemeshow, S. (1991). *Applied Logistic Regression*. (Vol. 47) (No. 4). doi: 10.2307/2532419

Setiawan, K. E., Kurniawan, A., dan Prasetyo, S. Y. j. . U. (2024). Comparative analysis of machine learning decision tree-based models for predicting maternal health risks. *Procedia Computer Science*, 245(C), 57–64. Retrieved from <https://doi.org/10.1016/j.procs.2024.10.229> doi:

- 10.1016/j.procs.2024.10.229
- Setio, P. B. N., Saputro, D. R. S., dan Winarno, B. (2020). Klasifikasi dengan pohon keputusan berbasis algoritme C4. 5. Dalam *Prisma, prosiding seminar nasional matematika* (Vol. 3, hal. 64–71).
- Siddika, A., dan SultanaM, M. j. . U. (2025). JTIE J . of Tech . , 1–35.
- Simegn, G. L., dan Degu, M. Z. j. . U. (2025). Maternal health risk analysis , automated pregnancy risk level identification and status monitoring system based on multivariable clinical data. (November 2024), 0–3. doi: 10.21037/ht-23-10
- Syahfirda, V. A., Hamid, F. S., Santi, A. D., dan Mulawardhana, P. j. . U. (2023). Analysis of risk factor of preeclampsia : A literature review.
- Syairaji, M., Nurdiani, D. S., Wiratama, B. S., Prü journal = Unknown Journal st, Z. D., Bloemenkamp, K. W., dan Verschueren, K. J. (2024). Trends and causes of maternal mortality in Indonesia: a systematic review. *BMC Pregnancy and Childbirth*, 24(1), 1–14. doi: 10.1186/s12884-024-06687-6
- Tomar, K., Sharma, C. M., dan Prasad, T. j. . U. (2024). A Machine Learning-Based Risk Prediction Model During Pregnancy in Low-Resource Settings †. , 1–9.
- Tzimourta, K. D., Tsipouras, M. G., Angelidis, P., Tsalikakis, D. G., dan Orovou, E. j. . U. (2025). Maternal Health Risk Detection: Advancing Midwifery with Artificial Intelligence. *Healthcare (Switzerland)*, 13(7). doi: 10.3390/healthcare13070833
- Vital, N., dan Reports, S. j. . U. (2025). Health E-Stats. (3).
- WHO. (2016). WHO recommendations on antenatal care for a positive pregnancy experience (Executive summary). *World Health Organization*, 10(January), 152. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/77770/9789241501996_eng.pdf%0Ajsessionid=9F903CEF68613DB19A238B5D2D73CCA9?sequence=1
- WHO. (2023a). *Improving maternal and newborn health and survival and reducing stillbirth* (Vol. 11) (No. 1). Retrieved from http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM PEMBETUNGAN TERPUSAT STRATEGI MELESTARI
- WHO. (2023b). *Trends in maternal mortality 2000 to 2020: estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Di-*



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.

vision. Retrieved from <https://www.who.int/publications/i/item/9789240068759>

WHO, UNICEF, UNFPA, W. B. G., dan Division, U. (2025). *Trends in maternal mortality estimates 2000 to 2023: estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division.* Retrieved from <https://www.who.int/reproductivehealth/publications/maternal-mortality-2000-2017/en/>

World Health Organization journal = Unknown Journal . (2013). Diagnostic Criteria and Classification of Hyperglycaemia First Detected in Pregnancy. , 1–62.

Zhou, X., Cai, F., Li, S., Li, G., Zhang, C., Xie, J., dan Yang, Y. j. . U. (2024a). Machine learning techniques for prediction in pregnancy complicated by autoimmune rheumatic diseases: Applications and challenges. *International Immunopharmacology*, 134(March), 112238. Retrieved from <https://doi.org/10.1016/j.intimp.2024.112238> doi: 10.1016/j.intimp.2024.112238

Zhou, X., Cai, F., Li, S., Li, G., Zhang, C., Xie, J., dan Yang, Y. j. . U. (2024b). Machine learning techniques for prediction in pregnancy complicated by autoimmune rheumatic diseases: Applications and challenges. *International Immunopharmacology*, 134(March), 112238. Retrieved from <https://doi.org/10.1016/j.intimp.2024.112238> doi: 10.1016/j.intimp.2024.112238

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 A

SURAT IZIN PENELITIAN



KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
FAKULTAS SAINS DAN TEKNOLOGI

كلية العلوم و التكنولوجيا

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) 589 025 Web: www.uin-suska.ac.id E-mail: faoste@uin-suska.ac.id

Nomor : B- 6795 / F.V/PP.009/09/2025 Pekanbaru, 29 September 2025
Sifat : Biasa
Hal : Mohon Izin Penelitian dan Pengambilan Data
Tugas Akhir/Skripsi

Kepada Yth.
Kepala Dinas Kesehatan Kota Pekanbaru
Komplek Perkantoran Walikota Pekanbaru
Gedung Belah Bubung, Lt. IV
Jl. Abdul Rahman Harnid, Kelurahan Tuah Negeri
Kecamatan Tenayan Raya

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	:	Rahma Devi
NIM	:	12250323758
Fakultas	:	Sains dan Teknologi
Program Studi/Smt	:	Sistem Informasi / 7(Tujuh)
No. Hp/E-mail	:	082285142993

untuk pengambilan data yang sangat dibutuhkan dalam Tugas akhir mahasiswa tersebut
yang berjudul "PREDIKSI RISIKO KEHAMILAN PADA IBU HAMIL MENGGUNAKAN
ALGORITMA MACHINE LEARNING". 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
ucapkan terima kasih.

Wassalam
Dekan


S. Slenita Muda, M.Sc
NIP. 197701032007102001



Dokumen ini telah ditanda tangani secara elektronik.
Token : Qq9FDG0m

Gambar A.1. Surat Permohonan Izin Peneelitian dari Fakultas

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.



**PEMERINTAH KOTA PEKANBARU
DINAS KESEHATAN**

Jl. Abdul Rahman Hamid - Gedung Belah Bubung Lt. 1-2, Kel. Tuah Negeri, Kec. Tenayan Raya Kota Pekanbaru - 28289, Telepon/Fax., Laman www.dinkes pekanbaru.go.id, Email dinkespu@pekanbaru.go.id

Pekanbaru, 27 Oktober 2025

Nomor : B.400.14.5.4/Dinkes-Umum/1777/2025
Sifat : Biasa
Lamp : -
Hal : Riset an. Rahmi Devi

Yth. Kepala Puskesmas Simpang Baru
di
Pekanbaru

Menindaklanjuti surat dari Dekan Fakultas Sains dan Teknologi UIN Suska Riau Nomor : B-6795/F.V/PP.009/09/2025 Tanggal 29 September 2025, tentang rekomendasi Penggambilan Data kepada :

Nama	:	Rahmi Devi
NIM	:	12250323758
Instansi	:	UIN Suska Riau
Fakultas/Jurusan	:	Sistem Informasi
Judul Penelitian	:	Prediksi Resiko Kehamilan Pada Ibu Hamil Menggunakan Algoritma Machine Learning

Sehubungan dengan hal tersebut di atas, kami mengharapkan kepada Saudara untuk dapat membantu kelancaran pengumpulan data kepada yang bersangkutan di atas.

Atas perhatian dan kerjasama Saudara, kami ucapkan terima kasih

Pekanbaru, 27 Oktober 2025



Tembusan :
Yth. Dekan Fakultas Sains dan Teknologi UIN Suska Riau

Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan oleh Balai Besar Sertifikasi Elektronik (BSrE), Badan Siber dan Sandi Negara (BSSN).

Gambar A.2. Balasan Surat izin Penelitian DINKES rekomendasi wawancara ke puskesmas simpang baru

Hak Cipta Dilindungi Undang-Undang

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

LAMPIRAN B

BERITA ACARA DAN PERTANYAAN WAWANCARA

BERITA WAWANCARA ATAU HASIL OBSERVASI

Pada hari ini,

Telah dilaksanakan wawancara yang berkaitan dengan penelitian yang akan dilakukan untuk penyusunan Tugas Akhir

Tempat : Puskesmas Simpang Baru

Pukul : 09:30

Nama Narasumber : Dina Dwi Hastuti, S.Tr.Kel

Jabatan : Bidan Pelaksana Program KIA-KB

Pihak pewawancara melakukan wawancara dengan pihak narasumber yang berkaitan dengan penelitian yang akan dilakukan di Puskesmas Simpang Baru, kemudian narasumber memberikan jawaban terkait pertanyaan yang diajukan oleh pewawancara.

Adapun pertanyaan yang diajukan serta hasil wawancara terlampir.

Mengetahui,

Puskesmas Simpang Baru, 19 November 2025

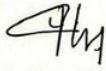
Peneliti

Narasumber



Rahma Devi

NIM. 12250323758



...DINA DWI HASTUTI, S.Tr.Kel

Gambar B.1. Berita wawancara

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak menghargai 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 PERTANYAAN WAWANCARA

A. Terkait Risiko

1. Bagaimana gambaran kasus Risiko ibu Hamil?

Dari ANC kunjungan ibu hamil, dimana dikunjungan itu nanti kita bisa mengkaji atau ketemu apakah nanti beresiko jika beresiko tidak beresiko. Dan untuk penanganannya kalo bisa ditangani kami tangani jika tidak batu nanti itu dirujuk.

2. Seperti apa Tingkat Risiko ibu hamil?

Tingkat Risikonya itu ada kategori Tinggi, rendah, dan sedang, juga itu kadang dari umurnya misalnya terlalu tua diatas 36 tahun kalo terlalu muda juga ga bagus. beresiko dibawa ke tahun depan, kalo perbaikan temui risiko temui kategori kecil dan cukup deangani melalui relazanah standar. Risiko sedang ditandai kalau ada satu atau dua faktor risiko yg mutu meningkat seiring meningkatnya kemajuan lebih tinggi. Risiko Tinggi memiliki faktor yg meningkatkan kalo ada faktor yg memperbaikinya.

3. Apakah ada penyuluhan ibu hamil? Seperti apa?

Ada, pastien harus konseling, persiapan konseling dan mengikuti kelas ibu hamil. dari yang penting itu dia ANC

B. Terkait (Hipertensi Gestasional dan Preklampsia)

1. Apa perbedaan mendasar antara Hipertensi Gestasional dan Pre-eklampsia? Dan apa saja kriteria klinis (tanda dan gejala) yang harus dipenuhi untuk mendiagnosis Pre-eklampsia di Puskesmas?

Prekampsia bisa didekati, misalnya disertai dengan kuku bergejala, dimana ini harus terjadi tiap bulan. Tabel darah kaya HB tidak HB kurang harus ditengahnya dengan. Sedang karen anemia istimewa banyak seperti perdarahan perdarahan, Jantung kurang berkembang karena kurang darah

Gambar B.2. Pertanyaan Wawancara

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak menghargai 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. Apa saja ciri-ciri atau keluhan awal yang paling sering disampaikan oleh ibu hamil yang akhirnya didiagnosis Pre-eklampsia di wilayah ini?
- diagnosa diketahui dan diketahui ada indikatornya seperti sebelum hamil dia sudah punya suhu tubuh tinggi. Pada perumah, MNC sebagian besar, tidak narsis diagnostik berdasarkan indikator. Misal dengan hipertensi juga ke RS untuk perawatan lebih lanjut. Maka yang pentingnya pemeriksaan kesehatan.
3. Apa dampak terburuk dari Pre-eklampsia jika tidak terdeteksi sejak dini, baik bagi kesehatan ibu (misalnya Eklampsia) maupun janin (misalnya kelahiran prematur)?
- Bisa menyebabkan komplikasi ibu juga bisa menyebabkan kewalahan, dan tenses tinggi. Kalau ibu tinggi tetapi tidak atribut fatal, fatal dan harus dikontrol dan dirujuk ke RS
4. Secara sederhana, apa ciri-ciri utama (gejala) dari Pre-eklampsia yang paling sering Ibu/Bapak temui pada ibu hamil di sini?
- Keluhan awal dari ibu hamil itu terdiri preklampsia, kadang ada yang tanpa gejala. Misal sering pusing, dan sebaliknya cek tenses tinggi.

C. Terkait Atribut

1. Untuk atribut seperti tekanan darah sistolik dan diastolik, kadar gula darah, usia, kadar HB, suhu tubuh, denyut jantung, apakah atribut ini valid digunakan untuk risiko ibu hamil?
 2. Menurut Ibu/Bapak, berapa angka tekanan darah yang menjadi batas mutlak seorang ibu harus dikategorikan risiko tinggi (berpotensi Pre-eklampsia)?
- 140/90. Tegangan Darah Sistolik BP 140 diatas 140 sudah mulai tidak halal, kalau 160 itu sudah tinggi. Dan Diastolik BP diatas 90 sudah harus halal-halal

Gambar B.3. Pertanyaan Wawancara (2)

© 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.

3. (kadar gula darah) Seberapa besar pengaruh tingginya kadar gula darah terhadap peningkatan risiko kehamilan, terutama kaitannya dengan Pre-eklampsia atau komplikasi lain di sini pk/bu?
- daerah kpd mg/dl itu sudah tinggi, dimana batas batas patolog tingginya yg dpt 20 mg/dl Jika mmol nya itu diatas 7,8 mmol/L itu sudah tinggi dari batasnya sampai 11,1 mmol/L, dan normalnya dibawah 7,8.
4. (Usia) Dalam pengalaman di Puskesmas, apakah usia terlalu muda (di bawah 20 tahun) dan usia terlalu tua (di atas 35 tahun) sama-sama meningkatkan risiko? Mana yang dampaknya lebih besar pada kasus risiko tinggi?
- dari segi usia jika terlalu tua diatas 35 tahun, dan kalau terlalu muda juga tidak bagus seperti usia di bawah 20 tahun. Bisa menyebabkan bercacik
5. (kadar ~~BB~~) Apakah kadar ~~BB~~ yang rendah (~~anemia~~) dianggap sebagai faktor yang secara signifikan meningkatkan risiko komplikasi kehamilan?
- Untuk mengelati secara dini faktor risiko komplikasi atau gugup terhadap kehamilan yaitu dengan ANC dengan rutin melakukannya pemeriksaan. Dimana standarnya tentu dengan teratur kesehatan Ibu 6-8 kali. Kadang faktor kesehatan itu juga membuat Ibu tersebut stress pada ANC.
6. (Suhu tubuh dan jenut jantung) Kapan peningkatan Suhu Tubuh atau Denyut Jantung seorang ibu hamil (di luar demam biasa) menjadi tanda bahaya yang mendorong Ibu/Bapak melakukan pemeriksaan lanjutan untuk risiko kehamilan?
- Melalui Ibu hamil ini kadang memang meningkat, kadang juga mereka ibu hamil kepanasan terus. Kalau suhu tubuh tsb itu sudah terangkat, ilah 38°C Ibu udah mau demam. Kalau dalam f Ibu 38,6°F keadaan Ibu sudah tinggi. Denyut jantung pada Ibu hamil biasanya meningkat 80-100 Marsh dianggap normal. Tuntas 100 BPM sudah tinggi

Gambar B.4. Pertanyaan Wawancara (3)

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.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau



Gambar B.5. Pertanyaan Wawancara (4)

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 C
SAMPLE DATA MENTAH

1	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
2	25	130	80	15	98	86	high risk
3	35	140	90	13	98	70	high risk
4	29	90	70	8	100	80	high risk
5	30	140	85	7	98	70	high risk
6	35	120	60	06.01	98	76	low risk
7	23	140	80	07.01	98	70	high risk
8	35	85	60	11	102	86	high risk
9	42	130	80	18	98	70	high risk
10	23	90	60	07.01	98	76	low risk
11	35	85	60	11	102	86	high risk
12	25	110	89	07.01	98	77	low risk
13	15	120	80	07.01	98	70	low risk
14	50	140	90	15	98	90	high risk
15	25	140	100	07.01	98	80	high risk
16	10	70	50	06.09	98	70	low risk
17	40	140	100	18	98	90	high risk
18	21	90	65	07.05	98	76	low risk

Gambar C.1. Sampel Data Mentah

A	B	C	D	E	F	G
869	15	76	49	06.04	98	77
870	15	120	80	07.02	98	70
871	15	80	60	7	98	80
872	12	95	60	07.02	98	77
873	29	90	70	06.07	98	80
874	31	120	60	06.01	98	76
875	29	130	70	06.07	98	78
876	17	85	60	9	102	86
877	19	120	80	7	98	70
878	20	110	60	7	100	70
879	32	120	65	6	101	76
880	26	85	60	6	101	86
881	29	130	70	07.07	98	78
882	19	120	80	7	98	70
883	54	130	70	12	98	67
884	44	120	90	16	98	80
885	23	130	70	06.09	98	70
886	22	85	60	06.00	98	76

Gambar C.2. Sampel Data Mentah (2)

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.

	A	B	C	D	E	F	G
1805	32	140	90	18	98	88	high risk
1806	12	90	60	8	102	66	high risk
1807	35	140	100	9	98	66	high risk
1808	54	140	100	15	98	66	high risk
1809	40	120	95	11	98	80	high risk
1810	12	90	60	11	102	60	high risk
1811	60	120	85	15	98	60	high risk
1812	55	140	95	19	98	77	high risk
1813	50	130	100	16	98	76	high risk
1814	13	90	65	9	101	80	high risk
1815	17	90	65	07.07	103	67	high risk
1816	17	85	60	06.03	102	86	high risk
1817	40	120	75	07.07	98	70	high risk
1818	48	120	80	11	98	88	high risk
1819	22	120	60	15	98	80	high risk
1820	55	120	90	18	98	60	high risk
1821	35	85	60	19	98	86	high risk
1822	43	120	90	18	98	70	high risk

Gambar C.3. Sampel Data Mentah

LAMPIRAN D

PENELITIAN TERDAHULU

Tabel D.1. Penelitian Terdahulu

No	Penulis	Judul	Algoritma	Hasil
	Hidayat dan Astuti (2020)	Diagnosis Preeklamsia dengan K-NN	K-Nearest Neighbour	Sensitivity 98.19%, Specificity 100%, Accuracy 98.33%.
	Irfan dkk. (2021)	Giving more insight for automatic risk prediction during pregnancy with interpretable machine learning	XGBoost, Random Forest, Naïve Bayes, KNN	Akurasi RF 87%. Gap: Dataset terbatas (400 data), belum membandingkan precision, recall, dan F1-score secara mendalam.
3	Raja dkk. (2021)	A Machine Learning-Based Prediction Model for Preterm Birth in Rural India	Decision tree, logistic regression, SVM	SVM menghasilkan akurasi tertinggi 90.9%. Menggunakan metrik Accuracy, Sensitivity, Specificity.
4	Pawar, Malhotra, Sharma, Arora, dan Vaidya (2022)	A Robust Machine Learning Predictive Model for Maternal Health Risk	Traditional ML & Robust ML Model	Gap: Model tradisional belum stabil dan akurasi terbatas, perlu pendekatan yang lebih robust.
	Fahira, Dwiyanti, dan Habibi (2023))	Pendekatan Supervised Learning untuk Diagnosis Kehamilan	Decision tree, Random Forest	Gap: Model sangat bergantung pada data seimbang; terbatas pada dua algoritma saja.
	S. A. Rahman dkk. (2023)	Explainable AI based Maternal Health Risk Prediction using ML and DL	XGBoost, ANN, RF	Akurasi RF: 88%. Gap: Belum mengevaluasi dampak ketidakseimbangan kelas (low, high, very high).

Tabel D.1. Penelitian Terdahulu (lanjutan)

No	Penulis	Judul	Algoritma	Hasil
1	Fahira dkk. (2023)	Classification of Hypertension in Pregnant Women Using Multinomial Logistic Regression	Multinomial Logistic Regression	Akurasi 79.6%. Gap: Tidak menyebut precision, recall, atau F1; hanya fokus pada akurasi.
2	Khoirunnisa dan Lestari (2023)	Implementasi Klasifikasi Kehamilan Berisiko dengan Naive Bayes	Naïve Bayes	Akurasi 85.62%. Gap: Belum dibandingkan dengan DT atau RF untuk validitas.
3	Rabbi dan MEM (2023)	Predicting Maternal Health Risk by Utilizing Different ML Classifiers	Logistic Regression, RF, KNN, Naïve Bayes	Akurasi RF \pm 88–90%. Gap: Evaluasi masih dominan akurasi, belum detail pada F1-score tiap kelas risiko.
10	(Amalia dkk. (2023)	Prediksi Resiko Kesehatan Ibu Hamil Dengan Menggunakan Metode Decision Tree	Decision Tree	Akurasi rendah (61.54%).
11	S. A. Rahman dkk. (2023)	Analisis Perbandingan Algoritma ML untuk Klasifikasi Tingkat Risiko	DT, Naïve Bayes, KNN	Decision Tree mencapai akurasi 90%, mengungguli NB (65%).
12	Pusadan dkk. (2023)	Implementasi Data Mining Prediksi Status Persalinan	Naïve Bayes	Accuracy 92%, Precision 82.4%, Recall 94%.
13	(Mustamin dkk., 2023)	Maternal Health Risk Classification (3 Naïve Bayes Models)	Naïve Bayes (Gaussian, Multinomial, Bernoulli)	Akurasi tertinggi pada Multinomial & Bernoulli (84.8%).

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

Tabel D.1. Penelitian Terdahulu (lanjutan)

No	Penulis	Judul	Algoritma	Hasil
14	Naive Bayes Method” (2023)	(“Classification Of Maternal Health Risk Using Three Models Naive Bayes	Naive Bayes	Akurasi Gaussian NB: 82.6%.
15	Muthu dkk. (2023)	Prediction of Maternal Health Risk with Traditional ML Methods	Decision Tree dan KNN	Random Forest terbaik (89.16%), KNN terendah (68.47%). Analisis parameter vital sign.
16	(Nwokoro dkk. (2024)	Predicting Maternal Outcomes Using Tree-based Methods in ML	DT, RF, AdaBoost, Gradient Boost, XGBoost	XGBoost precision tertinggi (0.95). Gap: Tidak melaporkan akurasi, recall, dan F1 secara lengkap.
17	Assaduzzamar dkk. (2024))	Predicting Maternal Risk Level Using Machine Learning Models	Random Forest, XGBoost, SVM	Model ensemble memberikan akurasi tertinggi 92%.
18	Setiawan dkk. (2024))	Comparative Analysis of DT-Based Models for Predicting Maternal Risks	DT, Bagging DT, AdaBoost, RF	RF Akurasi 82%. Disarankan menguji SVM, Logistic Regression, dan Naïve Bayes.
	(Hennessy dkk., 2024)	Machine Learning, Advanced Data Analysis, and a Role in Pregnancy Care?	ML, AI, DT, RF	Artikel tinjauan umum mengenai peningkatan prediksi preeklamsia.

Hak Cipta Dilindungi Undang-Undang

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

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

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

Tabel D.1. Penelitian Terdahulu (lanjutan)

No	Penulis	Judul	Algoritma	Hasil
20	Purnama, Hikmawati, dan Rahayu (2024)	Analisis Klasifikasi Mengidentifikasi Potensi Risiko	RF, Extra Trees, DT, LightGBM	RF Akurasi 82.15%. Gap: Hanya menggunakan akurasi, performa per kelas belum dianalisis.
21	Jefri Jaya journal = Unknown Journal (2025)	Perbandingan Algoritma ML untuk Prediksi Risiko Ibu Hamil	XGBoost, SVM, Decision Tree	XGBoost terbaik (Akurasi 0.827).
22	(Harnal dkk. (2025))	Predicting Maternal Health Risk: Percentage Split vs K-Fold	SVM, KNN, Logistic Regression	Menyoroti pentingnya metode evaluasi yang robust (K-Fold). Data hasil akurasi spesifik tidak disebutkan.
23	Assaduzzamar dkk. (2024)	Early Prediction of Maternal Health Risk Factors Using ML Techniques	Random Forest	Akurasi 75.2%, Presisi 85.7%, F1 73% setelah penerapan PCA.
24	Malde dkk. (2025))	Predicting Maternal Health Risks in LMIC Using Sparse Data	SVM, LR, RF, DT	RF 81.3%, DT 80.3%, SVM 59.6%, LR/GNB 57.6%.
25	Octaviani, Widiastuti, Amelia, Salam, dan Nuswan-toro (2025)	Implementasi Data Mining Prediksi Pre-Eklampsia Menggunakan C4.5	Algoritma C4.5	Riwayat PE berkorelasi erat (0.71) dengan kejadian PE.
26	Tzimourta dkk. (2025)	Maternal Health Risk Detection: Advancing Midwifery with AI	Random Forest	Akurasi 88.03%, Presisi 88.10%. Menunjukkan ketahanan dalam klasifikasi risiko.

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, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel D.1. Penelitian Terdahulu (lanjutan)

No	Penulis	Judul	Algoritma	Hasil
27	Simegn dan Degu (2025)	Maternal health risk analysis, automated identification and monitoring	KNN, SVM, XGBoost, RF	RF unggul dengan akurasi 91%. Korelasi kuat pada tekanan darah dan gula darah.
28	Kurniawan (2025)	Perbandingan Optimasi Algoritma: DT, Naive Bayes, KNN	DT (ID3, C4.5), NB, KNN	C4.5/ID3 mencapai akurasi terbaik 85.25%.
29	Modi dan Kumar (2025)	A Contextual Framework for Predicting Pregnancy Complications	SVM	Akurasi 80%, Presisi/Recall/F1 77%. Menggunakan variabel vital sign.
30	Siddika dan SultanaM (2025)	Maternal Health Risk Analysis using EDA and ML Algorithms	XGBoost, DT, SVM, NB	XGBoost dan DT mencapai akurasi 94%, SVM 72%, NB 64%.

Hak Cipta Dilindungi Undang-Undang

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

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

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

**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.

DAFTAR RIWAYAT HIDUP

Rahma Devi lahir di Sibiruang pada tanggal 04 Januari 2003 dan merupakan anak ketiga dari empat bersaudara dari pasangan M.Hidayat dan Zalinar. Pada tahun 2010 Peneliti memulai pendidikan Sekolah Dasar di SD Negeri 007 Sibiruang, dan diselesaikan pada tahun 2016 . kemudian peneliti menempuh pendidikan menengah pertama di Madrasah Tsana wiyah Nurul Falah Sibiruang dan melanjutkan pendidikan menengah atas di SMA Negeri 2 Koto Kampar Hulu dari tahun 2019 hingga 2022. Setelah menyelesaikan pendidikan menengah atas, peneliti melanjutkan dengan menjadi mahasiswa Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Selama masa perkuliahan, peneliti mencoba aktif dengan mengikuti berbagai kegiatan di lingkungan kampus, Seperti mengikuti Himpunan mahasiswa pada tahun 2023 sebagai *copy writing* divisi mendkominfo, kemudian pada tahun 2024 peneliti di percaya menjadi Bendahara Himpunan mahasiswa Sistem Informasi (HIMASI). Dan juga pada Desember 2024 peneliti diterima sebagai Tim Media Sistem Informasi (MEDIASI) di bagian *MC dan Copy Writing*. Sebagai penutup, peneliti mengucapkan rasa syukur yang mendalam serta terima kasih kepada seluruh pihak yang telah memberikan dukungan dan bantuan hingga terselesaiannya Tugas Akhir ini yang berjudul “ Prediksi risiko kesehatan pada ibu hamil menggunakan Algoritma *Machine Learning*”

UIN SUSKA RIAU