

Analisis Perbandingan Algoritma Machine Learning untuk Klasifikasi Tingkat Risiko Ibu Hamil

Rafiqi Aidil Fitra

Universitas Negeri Medan

E-mail: rfqaidl.4211250001@mhs.unimed.ac.id

Wahyu Abadi Harahap

Universitas Negeri Medan

E-mail: wahyuabadiharahap@mhs.unimed.ac.id

Wahyu Kurnia Rahman

Universitas Negeri Medan

E-mail: wahyukurniarahman59@gmail.com

Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara 20221

Abstract. *This research aims to conduct a comparative analysis of machine learning algorithms for classifying the risk levels of maternal health. With a focus on the significance of identifying and classifying health risks for pregnant women, this study applies supervised learning methods employing Naïve Bayes, Decision Tree, and K-Nearest Neighbors algorithms. Utilizing the "Maternal Health Risk" dataset from UCI Machine Learning, the research is conducted on Google Colaboratory using Python. The results indicate that the Decision Tree algorithm achieves the highest accuracy rate at 90%, surpassing K-Nearest Neighbors (86%) and Naïve Bayes (65%). Consequently, Decision Tree emerges as the preferred choice for predicting maternal health risks, offering the potential for enhanced care and monitoring.*

Keywords: *Classification, Maternal Health Risk, Decision Tree, Naïve Bayes, K-Nearest Neighbors (KNN).*

Abstrak. Penelitian ini bertujuan melakukan analisis perbandingan algoritma machine learning untuk klasifikasi tingkat risiko kesehatan ibu hamil. Dengan latar belakang fokus pada pentingnya identifikasi dan klasifikasi risiko kesehatan ibu hamil, penelitian ini menerapkan metode supervised learning menggunakan algoritma Naïve Bayes, Decision Tree, dan K-Nearest Neighbors. Menggunakan dataset "Maternal Health Risk" dari UCI Machine Learning, penelitian ini dilaksanakan di Google Colaboratory dengan Bahasa Python. Hasil penelitian menunjukkan bahwa Decision Tree mencapai tingkat akurasi tertinggi sebesar 90%, melebihi K-Nearest Neighbors (86%) dan Naïve Bayes (65%). Implikasinya, Decision Tree dapat dijadikan pilihan utama dalam memprediksi risiko kesehatan ibu hamil, memberikan potensi peningkatan dalam perawatan dan pengawasan.

Kata kunci: Klasifikasi, Tingkat Risiko Kesehatan Ibu Hamil, Decision Tree, Naïve Bayes, K-Nearest Neighbors (KNN).

LATAR BELAKANG

Kesehatan ibu hamil adalah salah satu aspek penting dalam perawatan prenatal. Tingkat risiko kesehatan ibu hamil mengacu pada kemungkinan terjadinya komplikasi atau masalah kesehatan selama kehamilan (Susiana, 2019). Identifikasi dan klasifikasi tingkat risiko ini dapat membantu tenaga medis dalam memberikan perawatan yang tepat dan pengawasan yang diperlukan bagi ibu hamil. Perawatan yang tepat dan tepat waktu sangat penting bagi ibu hamil untuk mencegah atau mengurangi risiko komplikasi kesehatan selama kehamilan. Dalam konteks ini, analisis menggunakan algoritma machine learning dapat memberikan pendekatan yang efektif untuk mengklasifikasikan tingkat risiko kesehatan ibu hamil (Amalia et al., 2023) . Dengan memanfaatkan data yang relevan, algoritma machine learning dapat digunakan untuk mengidentifikasi pola dan faktor risiko yang berkontribusi pada risiko kesehatan ibu hamil.

Penelitian ini dapat memberikan manfaat bagi tim medis, dokter kandungan, dan perawat yang terlibat dalam perawatan kesehatan ibu hamil. Dengan menggunakan algoritma machine learning, mereka dapat mendapatkan bantuan dalam mengklasifikasikan tingkat risiko kesehatan ibu hamil dan mengambil tindakan yang sesuai berdasarkan hasil klasifikasi tersebut. Penelitian ini dapat dilakukan di berbagai fasilitas kesehatan yang menyediakan perawatan prenatal. Waktu pelaksanaan penelitian tergantung pada ketersediaan data dan sumber daya yang diperlukan. Penelitian ini dapat dilakukan sepanjang tahun dengan durasi yang ditentukan sesuai dengan metodologi penelitian yang diadopsi.

Kesehatan ibu hamil adalah hal yang penting dalam perawatan prenatal. Untuk mencegah atau mengurangi risiko komplikasi kesehatan selama kehamilan, penting untuk mengidentifikasi dan mengklasifikasikan tingkat risiko kesehatan ibu hamil dengan akurat (Utami et al., n.d.). Dalam hal ini, analisis menggunakan algoritma machine learning dapat menjadi pendekatan yang efektif. Algoritma machine learning dapat membantu dalam mengklasifikasikan tingkat risiko kesehatan ibu hamil berdasarkan data yang relevan, seperti riwayat kesehatan ibu hamil, faktor risiko, dan hasil tes medis.

Perawatan yang tepat dan tepat waktu bagi ibu hamil sangat penting untuk meminimalkan risiko komplikasi kesehatan dan memastikan kelahiran yang sehat. Namun, mengklasifikasikan tingkat risiko kesehatan ibu hamil secara akurat dapat menjadi tugas yang kompleks dan sulit bagi tenaga medis. Metode tradisional dalam

mengklasifikasikan risiko kesehatan ibu hamil sering kali didasarkan pada penilaian subjektif, yang dapat menghasilkan hasil yang bervariasi dan tidak konsisten (Amalia et al., 2023). Dengan menggunakan teknik-teknik yang lebih objektif dan berbasis data, algoritma machine learning dapat membantu mengidentifikasi pola dan faktor risiko yang berkontribusi pada risiko kesehatan ibu hamil secara lebih akurat dan konsisten. Penelitian ini dapat memberikan metode yang lebih andal dan efisien dalam mengklasifikasikan tingkat risiko kesehatan ibu hamil, yang pada gilirannya dapat meningkatkan perawatan dan pengawasan yang diberikan kepada mereka.

Dengan menerapkan berbagai algoritma machine learning untuk klasifikasi seperti Naïve Bayes, Decision Tree, K-Nearest Neighbors kita dapat mengidentifikasi pola dan faktor risiko yang berkontribusi pada risiko kesehatan ibu hamil. Secara umum, algoritma klasifikasi adalah metode atau teknik yang digunakan untuk mengklasifikasikan data ke dalam kategori atau kelas yang telah ditentukan berdasarkan fitur-fitur yang ada (Nurahmadan et al., 2021). Tujuan utama algoritma klasifikasi adalah untuk mempelajari pola atau hubungan antara fitur-fitur data yang diberikan, sehingga dapat melakukan prediksi atau pengklasifikasian pada data baru yang belum pernah dilihat sebelumnya. Implementasi algoritma machine learning ini akan memberikan bantuan kepada tim medis, dokter kandungan, dan perawat dalam mengklasifikasikan tingkat risiko kesehatan ibu hamil secara objektif. Dengan demikian, mereka dapat memberikan perawatan yang tepat dan pengawasan yang diperlukan bagi ibu hamil, serta mengambil tindakan yang sesuai berdasarkan hasil klasifikasi yang diperoleh.

Dalam konteks ini, penelitian ini dilakukan untuk mengeksplorasi penggunaan algoritma machine learning dalam mengklasifikasikan tingkat risiko kesehatan ibu hamil. Dengan menerapkan pendekatan berbasis data yang lebih objektif, algoritma machine learning dapat mengidentifikasi pola dan faktor risiko yang berkontribusi pada risiko kesehatan ibu hamil secara akurat dan konsisten. Penelitian ini bertujuan untuk menghasilkan metode yang lebih andal dan efisien dalam mengklasifikasikan tingkat risiko kesehatan ibu hamil, dengan potensi meningkatkan perawatan dan pengawasan yang diberikan kepada mereka.

KAJIAN TEORITIS

Decision Tree

Decision Tree merupakan algoritma klasifikasi yang mengadopsi struktur pohon, menyerupai diagram alir pada *flowchart*. Tiap node internal dalam pohon mengindikasikan pengujian pada suatu atribut, cabangnya mencerminkan hasil pengujian, dan node daun mewakili kelas atau distribusi kelas (Nasrullah, 2021). Penggunaan konsep entropi menjadi kunci dalam menentukan bagaimana pohon akan terbagi (*split*) berdasarkan atribut. Semakin tinggi entropi suatu sampel, semakin tidak murni sampel tersebut. Dalam proses ini, Rumus yang di gunakan untuk menghitung entropy sample s adalah:

$$Entropy(S) = \sum_i^c -p_i \log_2 p_i$$

Dimana c merujuk pada jumlah nilai yang ada dalam atribut target (yaitu, jumlah kelas) sementara itu, symbol pi menggambarkan proporsi atau rasio antara jumlah sampel dalam kelas tertentu (kelas I) dengan jumlah sampel dalam dataset.

Naïve Bayes

Naïve Bayes Classifier adalah metode statistik bayesian sederhana. Hal ini disebut Naïve karena mengasumsi bahwa semua variabel berkontribusi terhadap klasifikasi dan berkorelasi satu sama lain, asumsi ini disebut class conditional independence (Hasanah et al., n.d.). Hal ini didasarkan pada teorema probabilitas bayes.

$$P(C_i|X) = \frac{P(X|C_i) P(C_i)}{P(X)}$$

$$P(X) = (P(X_1|C_1)P(C_1)) + (P(X_2|C_2)P(C_2)) + \dots + (P(X_n|C_n)P(C_n))$$

Keterangan:

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

$P(C_i)$ = Probabilitas kelas 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. Ambil dataset D sebagai data pelatihan yang terkait dengan label kelas, dimana setiap pasangan data direpresentasikan oleh vektor elemen n -dimensi $X = (X_1, X_2, X_3, \dots, X_n)$.
2. Anggap ada M kelas $C_1, C_2, C_3, \dots, C_m$. Jika kita ingin mengklasifikasikan suatu pasangan data X yang tidak diketahui, Maka clasifier akan memprediksi bahwa X termasuk kedalam kelas dengan propabilitas posterior tertinggi, yang dikondisikan pada X . Dalam klasifikasi Naïve Bayes, pasangan data X yang tidak diketahui diklasifikasikan ke dalam kelas C_i jika dan hanya jika $P(C_i | X) > P(C_j | X)$ untuk $1 \leq j \leq m$, dan $i \neq j$, propabilitas tersebut dihitung menggunakan teorema bayes.

K-Nearest Neighbor

K-nearest neighbor (K-NN) merupakan algoritma supervised yang melakukan klasifikasi berdasarkan mayoritas. Algoritma ini umumnya memanfaatkan metric jarak, seperti Eucilidean atau Manhattan distance (Brawijaya et al., 2019). Selain itu, jarak lain seperti norma Chebyshev atau Mahalanobis juga dapat digunakan. Dalam KNN, fungsi Euclidean Distance diaplikasikan, sebagaimana dinyatakan dalam persamaan berikut.

$$euc = \sqrt{\sum_i^n ((X_2)_i - (X_1)_i)^2}$$

Keterangan:

X_2 = Data Latih

X_1 = Data Uji

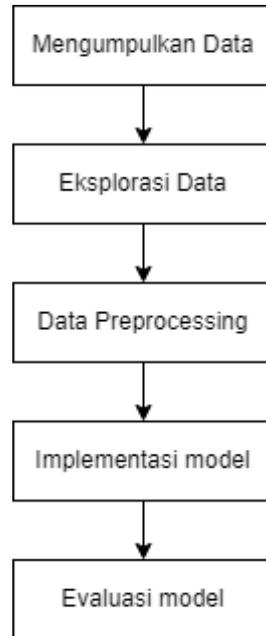
i = Variable data

n = Dimensi data

METODE PENELITIAN

Metode penelitian yang digunakan dalam komparasi antara tiga perbandingan model atau metode klasifikasi dengan menggunakan metode algoritma *data mining* yaitu metode Naïve bayes, K-Nearest Neighbor (KNN) dan Decision Tree yang mana merupakan bagian dari metode *supervised learning*. Penelitian ini dimulai dengan pengambilan dataset dari UCI Machine Learning yang berjudul *Maternal Health Risk*. Selanjutnya, data tersebut diproses dalam Google Colaboratory menggunakan Bahasa python, yang mana dilakukan eksplorasi dataset tersebut dan dilanjutkan dengan

preprocessing. Setelah itu, akan dilakukan implementasi model menggunakan ketiga algoritma tersebut yang mana akan didapatkan hasil setelah melakukan evaluasi model. Adapun alur penelitian ini dapat dilihat pada gambar dibawah ini.



Gambar 1. Alur penelitian

HASIL DAN PEMBAHASAN

Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah dataset *Maternal Health Risk* yang diambil dari UCI Machine Learning Repository. Pada website tersebut menyediakan kumpulan dataset-dataset yang sudah siap gunakan untuk implementasi machine learning. Dataset yang digunakan dalam penelitian ini terdiri dari 7 Variabel, yaitu: *Age*, *SystolicBP*, *DiastolicBP*, *BS*, *BodyTemp*, *HeartRate*, dan variabel target yaitu *Risk Level*. Dataset yang digunakan pada penelitian ini tidak memiliki *Missing Value* sama sekali. Untuk penjelasan dataset dapat dilihat pada tabel dibawah ini.

Tabel 1. Dataset

No	Nama Variabel	Deskripsi
1.	Age	Usia dari seorang ibu yang sedang mengandung.
2.	SystolicBP	Nilai Tekanan Darah tertinggi dalam mmHg.
3.	DiastolicBP	Nilai Tekanan Darah terendah dalam mmHg.
4.	BS	Kadar glukosa darah dinyatakan dalam konsentrasi molar.
5.	BodyTemp	Suhu badan dalam fahrenheit.
6.	HeartRate	Denyut jantung istirahat normal dalam <i>beats per minute</i> .
7.	RiskLevel	Prediksi Tingkat Intensitas Risiko selama kehamilan.

Eksplorasi Data

Dalam langkah eksplorasi data ini, tujuannya adalah untuk mengidentifikasi jenis dan keadaan variabel-variabel dalam dataset. Berdasarkan visualisasi pada Tabel 2 berikut, dapat diidentifikasi bahwa terdapat 4 variabel dengan tipe data *integer*, 2 variabel dengan tipe data *float*, dan 1 variabel dengan tipe data *object*. Selain itu, observasi menunjukkan bahwa seluruh variabel dalam dataset memiliki nilai non-null, atau dengan kata lain, tidak terdapat nilai yang hilang.

Tabel 2. Info dataset

Kolom	Jumlah non-null	Dtype
Age	1014	Int64
SystolicBP	1014	Int64
DiastolicBP	1014	Int64
BS	1014	Float64
BodyTemp	1014	Float64
HeartRate	1014	Int64
RiskLevel	1014	Object

Pada Tabel 3 berikut, terdapat informasi mengenai deskripsi setiap variabel. Informasi yang disajikan mencakup data tentang jumlah data, rata-rata (mean), standar deviasi, nilai minimum, nilai maksimum, dan quartile-quartile dari masing-masing variabel. Rinciannya memberikan gambaran mendalam mengenai sebaran data dan karakteristik statistik pada dataset *Maternal Health Risk*.

Tabel 3. Deskripsi untuk setiap variabel

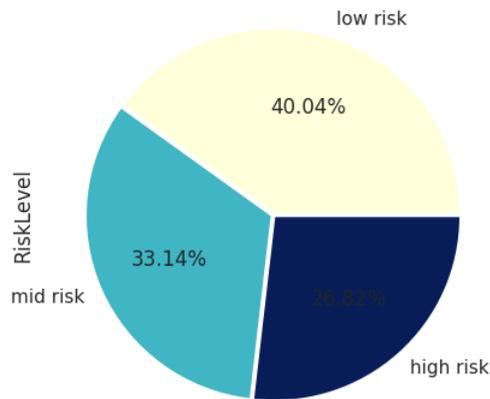
	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate
count	1014	1014	1014	1014	1014	1014
mean	29.87	113.1	76.46	8.725	98.66	74.30
std	13.47	18.40	13.88	3.293	1.371	8.088
min	10.00	70.00	49.00	6.000	98.00	7.000
25%	19.00	100.0	65.00	6.900	98.00	70.00
50%	26.00	120.0	80.00	7.500	98.00	76.00
75%	39.00	120.0	90.00	8.000	98.00	80.00
max	70.00	160.0	100.0	19.00	103.0	90.00

Dalam Tabel 4 yang terlampir, disajikan informasi mengenai korelasi antar variabel bebas.

Tabel 4. Korelasi antar variabel bebas

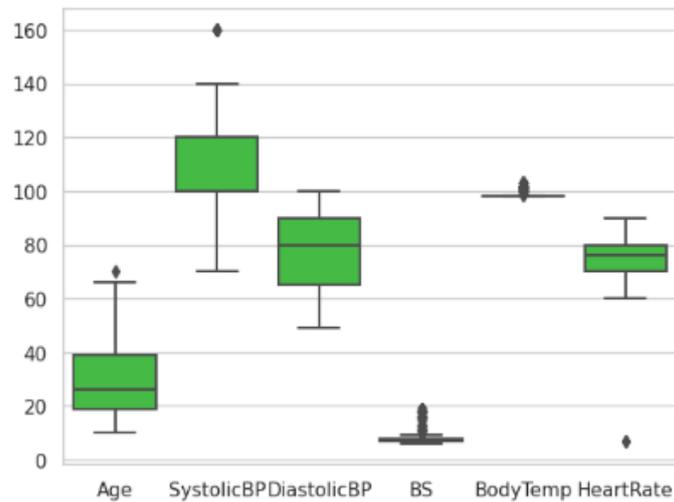
	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate
Age	1.000000	0.416045	0.398026	0.473284	-0.255323	0.079798
SystolicBP	0.416045	1.000000	0.787006	0.425172	-0.286616	-0.02310
DiastolicBP	0.398026	0.787006	1.000000	0.423824	-0.257538	-0.04615
BS	0.473284	0.425172	0.423824	1.000000	-0.103493	0.142867
BodyTemp	-0.25532	-0.28661	-0.257538	-0.10349	1.000000	0.098771
HeartRate	0.079798	-0.02310	-0.046151	0.142867	0.098771	1.000000

Selanjutnya, peneliti melakukan visualisasi jumlah data pada dataset berdasarkan tiga kelas dengan menggunakan diagram lingkaran, sebagaimana terlihat pada gambar berikut. Diagram lingkaran ini memberikan informasi bahwa selisih jumlah data antar tiap kelas cukup seimbang, dengan catatan bahwa jumlah data yang terkecil yaitu pada kelas 'high risk', mencapai 26,82%, yang hampir setara dengan 1/3 dari keseluruhan dataset. Oleh karena itu, dapat disimpulkan bahwa tidak diperlukan proses *resampling*, mengingat distribusi data pada setiap kelas cenderung seimbang. Keputusan ini didasarkan pada pertimbangan bahwa *resampling* mungkin tidak diperlukan untuk memitigasi ketidakseimbangan kelas, mengingat proporsi data antar kelas relatif mendekati proporsi yang diinginkan.



Gambar 2. Diagram lingkaran variabel terikat

Kemudian, dilakukan eksplorasi lanjutan pada dataset, dan peneliti mengamati keberadaan outlier (pencilan). Pada Gambar 3 yang tertera, hampir seluruh variabel menunjukkan keberadaan outlier. Saat ini, peneliti memutuskan untuk mengabaikan nilai-nilai yang berbeda jauh tersebut, mengingat nilai-nilai tersebut tampaknya menjadi fenomena yang alami dalam konteks kasus ini, kecuali untuk variabel HeartRate. Variabel tersebut menonjol dengan nilai yang sangat jauh dari rentang nilai lainnya. Oleh karena itu, perhatian khusus akan difokuskan pada variabel HeartRate guna mengevaluasi dan menangani nilai-nilai outlier yang signifikan pada variabel tersebut.



Gambar 3. Boxplot

Data Preprocessing

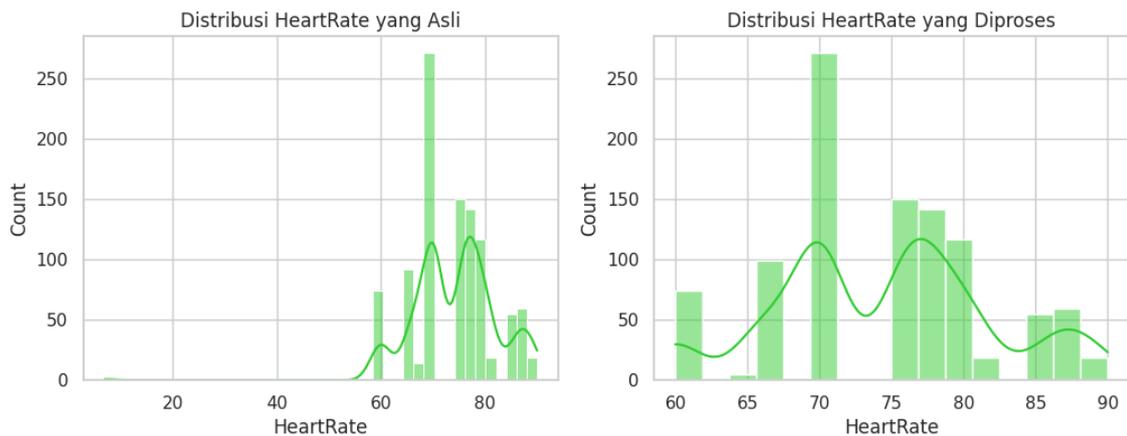
Pada tahap ini, dilakukan proses pembersihan data dengan tujuan mempersiapkan dataset sebelum langkah penerapan algoritma. Dari hasil analisis sebelumnya, terlihat bahwa dataset ini mengandung beberapa variabel dengan nilai pencilan (outlier). Meskipun demikian, mayoritas nilai-nilai tersebut masih dapat dianggap toleran dalam konteks kehidupan nyata. Satu-satunya variabel yang menonjol dengan outlier nilai yang tidak masuk akal adalah HeartRate.

HeartRate	7	60	65	66	67	68	70	75	76	77	78	80	82	86	88	90
RiskLevel																
high risk	0	20	0	25	5	0	43	6	20	25	9	48	0	22	30	19
low risk	2	33	2	44	2	0	126	7	55	56	2	39	10	7	21	0
mid risk	0	21	3	18	5	2	102	6	56	15	35	30	9	26	8	0

Gambar 4. Data pada kolom ‘RiskLevel’ dan ‘HeartRate’

Pada Gambar 4 tersebut, terdapat dua pengamatan dengan nilai denyut jantung sebesar 7 bpm (denyut per menit). Menurut standar medis, denyut jantung normal saat istirahat untuk dewasa berkisar antara 60 hingga 100 denyut per menit, dan denyut jantung terendah yang pernah tercatat dalam sejarah manusia adalah 25 bpm. Oleh karena itu, dengan mempertimbangkan bahwa nilai-nilai ini mungkin merupakan kesalahan input, peneliti memutuskan untuk menghapus dua data ini untuk menjaga kualitas dan validitas analisis data.

Selanjutnya, peneliti membuat dua plot distribusi, yang pertama menampilkan distribusi asli dari variabel HeartRate, dan yang kedua menunjukkan distribusi setelah proses pembersihan data. Dapat terlihat pada grafik berikut bahwa plot kedua menunjukkan distribusi yang lebih terorganisir dibandingkan dengan plot pertama. Hal ini menunjukkan bahwa setelah mengatasi nilai-nilai pencilan pada variabel HeartRate, distribusi data menjadi lebih representatif dan memungkinkan analisis yang lebih akurat. Proses pembersihan data pada variabel HeartRate memberikan kontribusi positif dalam meningkatkan kejelasan pola distribusi, sehingga memperkuat integritas hasil analisis selanjutnya.



Gambar 5. Distribusi HeartRate Sebelum & Sesudah

Setelah melewati tahap pembersihan data, langkah berikutnya adalah melakukan transformasi pada data bertipe objek, khususnya pada kolom RiskLevel, menjadi data dalam bentuk angka. Sebagaimana terlihat pada Tabel 5, dataset telah mengalami perubahan di mana nilai 'high risk' direpresentasikan sebagai 0, 'mid risk' sebagai 1, dan 'low risk' sebagai 2. Data tersebut telah berhasil digabung ke dalam dataframe, dan perubahan dilakukan untuk mengubah tipe data menjadi integer. Proses ini memiliki tujuan untuk memudahkan analisis dan implementasi algoritma, sekaligus meningkatkan interpretabilitas data. Data yang telah melalui transformasi ini siap digunakan untuk tahap implementasi model.

Tabel 5. Dataset setelah pra-pemrosesan data

	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
0	25	130	80	15	98	86	0
1	35	140	90	13	98	70	0
2	29	90	70	8	100	80	0
3	30	140	85	7	98	70	0
4	35	120	60	61	98	76	2
...
1009	22	120	60	15	98	80	0
1010	55	120	90	18	98	60	0
1011	35	85	60	19	98	86	0
1012	43	120	90	18	98	70	0
1013	32	120	65	6	101	76	1

Implementasi Model

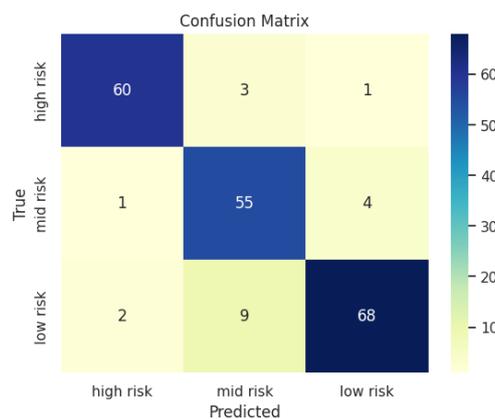
1. Decision Tree

Dalam mengimplementasikan model klasifikasi Decision Tree, peneliti menggunakan metode GridSearchCV untuk mencari parameter terbaik. Sebagai hasil dari penelitian ini, parameter yang diadopsi pada model adalah {'criterion': 'entropy', 'max_depth': 49}. Laporan klasifikasi dari pemodelan ini dapat dilihat pada Gambar 6 yang disertakan. Laporan klasifikasi memberikan gambaran komprehensif mengenai evaluasi model, termasuk akurasi, presisi, recall, dan F1-score.

	precision	recall	f1-score	support
high risk	0.95	0.94	0.94	64
mid risk	0.82	0.92	0.87	60
low risk	0.93	0.86	0.89	79
accuracy			0.90	203
macro avg	0.90	0.90	0.90	203
weighted avg	0.91	0.90	0.90	203

Gambar 6. Laporan klasifikasi Decision Tree

Dari hasil evaluasi model Decision Tree yang telah dibangun, tercatat bahwa akurasi yang berhasil dicapai sebesar 90%. Gambar 7 berikut, yang menyajikan Confusion Matrix, memberikan visualisasi yang lebih rinci terkait performa model Decision Tree



Gambar 7. Confusion matrix model Decision Tree

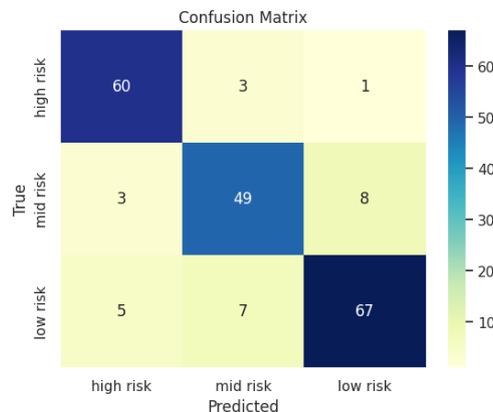
2. KNN (K-Nearest Neighbors)

Dalam implementasi model klasifikasi KNN, peneliti menggunakan metode GridSearchCV untuk mencari parameter terbaik. Hasil penelitian menunjukkan bahwa parameter yang diadopsi pada model KNN adalah {'leaf_size': 41, 'n_neighbors': 1}. Laporan klasifikasi dari pemodelan ini dapat ditemukan pada Gambar 8 yang disertakan.

	precision	recall	f1-score	support
high risk	0.88	0.94	0.91	64
mid risk	0.83	0.82	0.82	60
low risk	0.88	0.85	0.86	79
accuracy			0.87	203
macro avg	0.86	0.87	0.87	203
weighted avg	0.87	0.87	0.87	203

Gambar 8. Laporan klasifikasi KNN

Dari hasil evaluasi model KNN yang telah dibangun, tercatat bahwa akurasi yang berhasil dicapai sebesar 87%. Gambar 9 berikut, yang menyajikan Confusion Matrix, memberikan visualisasi yang lebih rinci terkait performa model KNN.



Gambar 9. Confusion matrix model KNN

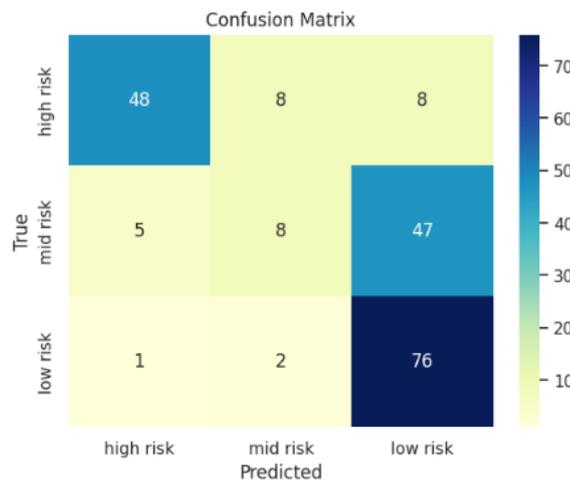
3. Naïve Bayes

Dalam implementasi model klasifikasi Naïve Bayes, peneliti menggunakan metode GridSearchCV untuk mencari parameter terbaik. Hasil penelitian menunjukkan bahwa parameter yang diadopsi pada model Naïve Bayes adalah {'var_smoothing': 0.00200923300256505}. Laporan klasifikasi dari pemodelan ini dapat ditemukan pada Gambar 10 yang disertakan.

	precision	recall	f1-score	support
high risk	0.89	0.75	0.81	64
mid risk	0.44	0.13	0.21	60
low risk	0.58	0.96	0.72	79
accuracy			0.65	203
macro avg	0.64	0.62	0.58	203
weighted avg	0.64	0.65	0.60	203

Gambar 10. Laporan klasifikasi Naïve Bayes

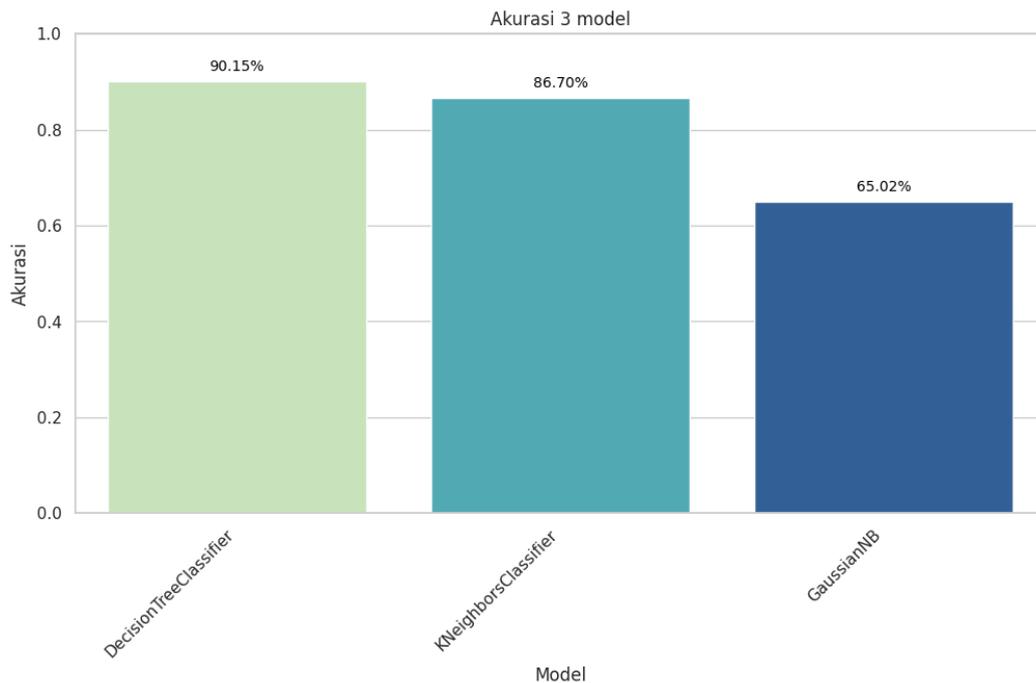
Dari hasil evaluasi model Naïve Bayes yang telah dibangun, dicapai akurasi sebesar 65%. Gambar 11, yang menyajikan Confusion Matrix, memberikan visualisasi yang lebih rinci terkait performa model Naïve Bayes dalam mengklasifikasikan data.



Gambar 11. Confusion matrix model Naïve Bayes

Evaluasi Model

Pada tahap evaluasi model ini, peneliti membentuk sebuah diagram batang untuk membandingkan hasil akurasi dari ketiga algoritma yang digunakan. Diperoleh bahwa algoritma Decision Tree menghasilkan akurasi sebesar 90,15%, algoritma KNN mencapai akurasi sebesar 86,70%, sementara algoritma Naïve Bayes hanya mampu mencapai akurasi sebesar 65,02%. Diagram ini memberikan gambaran visual yang jelas mengenai performa relatif dari setiap algoritma dalam mengklasifikasikan data, memungkinkan peneliti untuk mengevaluasi dan memilih algoritma yang paling sesuai dengan karakteristik dataset yang bersangkutan.



Gambar 12. Diagram batang dari ketiga algoritma

KESIMPULAN DAN SARAN

Berdasarkan hasil uji coba yang dilakukan menggunakan algoritma Decision Tree, K-Nearest Neighbors (KNN), dan Naïve Bayes dalam bahasa pemrograman Python di Google Colab, dapat disimpulkan bahwa Decision Tree memberikan hasil terbaik dalam hal akurasi dan kinerja untuk memprediksi Risiko kesehatan ibu hamil. Algoritma Decision Tree mencapai tingkat akurasi sebesar 90%, sementara KNN mencapai 86%, dan Naïve Bayes hanya mencapai 65%. Hasil tersebut menunjukkan bahwa Decision Tree lebih unggul dalam memodelkan dan memprediksi Risiko kesehatan ibu hamil berdasarkan dataset yang digunakan. Oleh karena itu, disarankan untuk mengadopsi penggunaan Decision Tree sebagai pilihan utama dalam konteks ini, mengingat performa yang lebih baik dibandingkan dengan KNN dan Naïve Bayes.

Sejalan dengan hasil uji coba yang menunjukkan keunggulan Decision Tree dalam memprediksi Risiko Kesehatan Ibu Hamil, disarankan untuk tetap mengambil langkah-langkah pencegahan dengan menguji algoritma lain sebagai langkah validasi tambahan. Meskipun Decision Tree memberikan hasil akurasi yang tinggi, pengujian dengan algoritma lain seperti Random Forest, Support Vector Machine (SVM), atau model deep learning seperti neural networks, dapat memberikan pemahaman yang lebih

komprehensif terkait performa model. Dengan menggabungkan hasil dari beberapa algoritma, dapat memperkuat kehandalan prediksi dan memberikan perspektif yang lebih luas terhadap kompleksitas data. Oleh karena itu, disarankan untuk melibatkan uji coba dengan algoritma-algoritma tambahan guna memastikan hasil prediksi yang lebih akurat dan dapat diandalkan.

DAFTAR REFERENSI

- Amalia, H., Rahmadanti, R., Syaiin, A., Salsabila, S., & Bina Sarana Informatika, U. (2012). Prediksi Risiko Kesehatan Ibu Hamil Dengan Menggunakan Metode Decision Tree. *JURNAL SWABUMI*, 11(1), 2023.
- Brawijaya, H., Samudi, & Widodo, S. (2019). Komparasi Algoritma K-Nearest Neighbor dan Naïve Bayes pada Pengobatan Penyakit Kutil Menggunakan Cryotherapy. *JUITA: Jurnal Informatika*, VII(2), 93–100.
- Hasanah, Q., Andrianto, A., & Hidayat, M. A. (n.d.). *Sistem Informasi Posyandu Ibu Hamil dengan Penerapan Klasifikasi Risiko Kehamilan Menggunakan Metode Naïve Bayes (Implementing Classification Risk in Posyandu System Information for Pregnant Using Naïve Bayes Method)*.
- Nasrullah, A. H. (2021). *IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS*. 7(2). <http://ejournal.fikom-unasman.ac.id>
- Nurahmadan, I. F., Agusta, A., Winarno, P. A., Sazali, B. H., Thurfah, Y., & Rosaliah, A. (2021). Perbandingan Algoritma Machine Learning Untuk Klasifikasi Denyut Jantung Janin. In *Nurul Chamidah, S.Kom., M.Kom.*
- Susiana, S. (2019). Angka Kematian Ibu: Faktor Penyebab Dan Upaya Penanganannya. *Bidang Kesejahteraan Sosial Info Singkat*, XI(24), 13–18.
- Utami, N., Dewi Puspitasari, R., Kurniawati, I., Graharti, R., & Yudho, A. (n.d.). *Tingkat Pengetahuan Ibu Hamil Mengenai Kesehatan Ibu dalam Masa Kehamilan dan Nifas di RSUD Dr. H. Abdul Moeloek Bandar Lampung (Vol. 3)*.