



Analisis Algoritma Klasifikasi Untuk Mengidentifikasi Potensi Risiko Kesehatan Ibu Hamil

Jajang Jaya Purnama¹, Nina Kurnia Hikmawati², Sri Rahayu³

¹Sistem Informasi, Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Jakarta, Indonesia

²Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia, Jakarta, Indonesia

³Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Jakarta, Indonesia

¹jajang.jjp@nusamandiri.ac.id*, ²ninakaha@yahoo.com, ³sriarahayu.ry@nusamandiri.ac.id

Abstract

The health of pregnant women has an important aspect in efforts to achieve the birth of a healthy baby. So early detection of the health of pregnant women has important. In this study the author identified potential maternal health risks for pregnant women by classifying them used machine learning which aims to analyze maternal health datasets with several algorithms including Random Forest, Extra Trees, Extreme Gradient Boosting, Decision Tree, and Light Gradient Boosting Machine. From several classification results carried out analysis and evaluation shown that the Random Forest classification algorithm provided optimal performance with an accuracy of 82,15%. These findings confirmed that the model created could identify complex patterns and relationships between features relevant to the classification of potential health risks for pregnant women at high, medium and low levels. These results have important implications in maternal care, because they can help doctors and medical personnel make more appropriate and effective decisions in dealing with maternal health risks and provide insight into pregnant women from an early age regarding their health conditions.

Keywords: Health of Pregnant Women, Machine Learning, Random Forest.

Abstrak

Kesehatan ibu hamil merupakan aspek penting dalam upaya mewujudkan kelahiran bayi yang sehat. Sehingga untuk mendeteksi dini kesehatan ibu hamil merupakan hal yang penting. Dalam penelitian ini penulis mengidentifikasi potensi risiko kesehatan ibu hamil dengan cara mengklasifikasinya menggunakan *machine learning* yang bertujuan untuk menganalisis *dataset maternal health risk* dengan beberapa algoritma diantaranya *Random Forest, Extra Trees, Extreme Gradient Boosting, Decision Tree, dan Light Gradient Boosting Machine*. Dari beberapa hasil klasifikasi dilakukan analisis dan evaluasi menunjukkan bahwa algoritma klasifikasi *Random Forest* memberikan kinerja optimal dengan akurasi 82,15%. Temuan ini mengkonfirmasi bahwa model yang dibuat dapat mengidentifikasi pola kompleks dan hubungan antara fitur-fitur yang relevan dengan klasifikasi potensi risiko kesehatan ibu hamil dengan tingkatan tinggi, sedang, dan rendah. Hasil ini memiliki implikasi penting dalam perawatan kesehatan ibu hamil, karena dapat membantu dokter dan tenaga medis dalam mengambil keputusan yang lebih tepat dan efektif dalam menangani risiko kesehatan ibu hamil serta memberikan wawasan pada ibu hamil sejak dini mengenai kondisi kesehatannya.

Kata kunci: Kesehatan Ibu Hamil, *Machine Learning, Random Forest*.

1. Pendahuluan

Kesehatan ibu hamil merupakan aspek penting dalam upaya mewujudkan kelahiran bayi yang sehat dan mengurangi angka kematian bayi maupun ibu serta *morbiditas maternal* dan *neonatal* [1]. Menurut UNICEF, setiap jam seorang perempuan meninggal karena melahirkan atau sebab-sebab lain yang berkaitan dengan kehamilan [2]. Sebagian ibu, terutama ibu muda yang baru pertama kali mengalami kehamilan, mungkin masih merasa kurang memahami hal ini secara mendalam. Upaya peningkatan kesehatan ibu hamil masih menghadapi tantangan dalam mengelola faktor risiko untuk ibu dan janin.

Dalam konteks yang lebih luas, edukasi dan pemahaman yang lebih mendalam tentang peran nutrisi yang tepat, gaya hidup sehat, serta deteksi dini terhadap penyakit atau kondisi medis yang dapat memengaruhi kehamilan sangatlah penting. Upaya kolaboratif antara para profesional kesehatan, lembaga pemerintah terkait, dan masyarakat dalam menyediakan akses yang lebih baik terhadap informasi dan layanan kesehatan ibu hamil juga menjadi bagian penting dari strategi pencegahan dan peningkatan kesehatan selama masa kehamilan.

Untuk meningkatkan pemahaman tentang faktor-faktor risiko yang terkait dengan kondisi kesehatan ibu hamil,



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

perlu adanya data yang mengidentifikasi variabel-variabel penting yang berkaitan dengan kesehatan ibu hamil. Data tersebut kemudian digunakan untuk membuat pola pembelajaran klasifikasi tingkat risiko kesehatan pada ibu hamil kemudian dilakukan analisis menggunakan pendekatan *machine learning* [3] [4] [5].

Machine Learning merupakan salah satu model untuk melakukan klasifikasi untuk mengetahui suatu hal tanpa intervensi dari manusia untuk menentukan pilihan, dalam dunia medis sangatlah dibutuhkan untuk pendeteksian dini terhadap kesehatan ibu agar tidak terjadi hal yang fatal yang masih ditangani dari jauh-jauh hari [3].

Penelitian oleh Ahmed Marzia dan rekan di Bangladesh mengungkapkan bahwa penggunaan IoT di daerah pedesaan di negara berkembang efektif dalam memantau kesehatan wanita hamil. Studi ini menggunakan algoritma *Logistic Model Tree* (LMT) yang terbukti memiliki akurasi tinggi dalam menilai risiko kehamilan [6].

Penelitian terkait dengan judul *Unintended pregnancy and maternal health complications: cross-sectional analysis of data from rural uttar Pradesh*, India bertujuan untuk mengeksplorasi potensi hubungan antara kehamilan yang tidak diinginkan dan ibu komplikasi kesehatan [7].

Penelitian terkait dengan judul *IoT based risk level prediction model for Maternal health care in the context of Bangladesh*, sebuah sistem telah dikembangkan agar efektif memantau dan memprediksi tingkat risiko ibu hamil di negara bangladesh, penelitian ini dimaksudkan untuk menggunakan alat dan mesin analisis masing-masing algoritma untuk mempelajari menemukan tingkat risiko kesehatan ibu hamil [8].

Penelitian ini dirancang dengan tujuan untuk menganalisis model klasifikasi yang dapat secara efektif mengidentifikasi potensi risiko kesehatan pada ibu hamil [9]. Isu mengenai kesehatan ibu hamil ini menarik untuk dibahas seperti yang diungkap para peneliti terdahulu. Penelitian ini memberikan kontribusi berupa pola klasifikasi tingkat potensi risiko kesehatan pada ibu hamil dengan menggunakan *dataset* publik berjumlah banyak sehingga menghasilkan wawasan yang lebih luas.

Metodologi yang akan digunakan dalam penelitian ini mencakup langkah-langkah analisis data yang komprehensif, penggunaan teknik-teknik *machine learning* atau kecerdasan buatan yang canggih, serta evaluasi terhadap berbagai parameter yang relevan dengan kesehatan ibu hamil. Selain itu, akan dilakukan juga validasi terhadap model klasifikasi yang dikembangkan untuk memastikan akurasi dan keandalannya dalam mengidentifikasi risiko kesehatan dengan tingkat kepercayaan yang tinggi [10].

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam upaya pencegahan dan pengelolaan risiko kesehatan pada ibu hamil. Informasi yang diperoleh dari model klasifikasi yang dikembangkan dapat dimanfaatkan oleh para tenaga kesehatan untuk melakukan tindakan yang tepat waktu dan tepat sasaran guna meningkatkan kesehatan ibu hamil serta mengurangi potensi komplikasi yang dapat terjadi selama kehamilan.

Selain itu, hasil penelitian ini juga diharapkan dapat menjadi landasan bagi pengembangan lebih lanjut terkait sistem pemantauan kesehatan ibu hamil secara *real-time* atau berkelanjutan, sehingga upaya-upaya pencegahan dan intervensi dapat dilakukan secara proaktif dan efisien. Dengan demikian, penelitian ini diharapkan dapat memberikan manfaat yang nyata dalam meningkatkan kesehatan ibu hamil serta meningkatkan kualitas hidup bagi ibu dan bayi yang dikandungnya [11] [12].

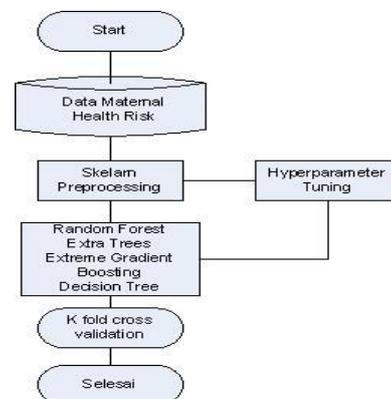
Dengan memanfaatkan *dataset* yang luas dan informatif yang terpublikasi ini, penulis berupaya untuk menganalisa penerapan model klasifikasi dan *pre-processing hyper-parameter tuning* untuk dianalisa kemudian menghasilkan pola untuk mengidentifikasi potensi risiko kesehatan ibu hamil sebagai kontribusi.

Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan baru dalam pemahaman tentang risiko kesehatan ibu hamil dan memperkuat upaya-upaya untuk meningkatkan kesehatan ibu hamil, sehingga dapat menghasilkan hasil yang lebih baik dalam perawatan dan manajemen kehamilan serta kelahiran yang sehat bagi ibu dan bayi [13][14].

Dengan kontribusi yang sudah dijelaskan diatas, penelitian ini diharapkan mampu memberikan manfaat untuk tenaga medis dan ibu hamil dalam mengklasifikasi urgensi kondisi kesehatan ibu hamil.

2. Metode Penelitian

Berikut ini merupakan langkah-langkah metode penelitian yang dilakukan disajikan dalam bentuk diagram pada Gambar 1:



Gambar 1. Metodologi Penelitian

1. Dataset

Dalam studi ini, *dataset* yang digunakan adalah *maternal health risk* [15] yang diperoleh dari UCI *repository* sebanyak 1014 sebagai data utama yang akan diproses menggunakan algoritma. Dari *dataset* tersebut yang digunakan sebagai data latih sebanyak 913 dan data uji sebanyak 101. *Dataset* ini mencakup variabel seperti usia, tekanan darah sistolik, tekanan darah diastolik, kadar gula darah, suhu tubuh, detak jantung, dan tingkat risiko. Informasi detail mengenai atribut-atribut ini dapat ditemukan dalam Tabel 1.

Tabel 1. Keterangan Dataset

No	Atribut	Keterangan
1	<i>Age</i>	Usia pada saat seorang wanita mengalami kehamilan
2	<i>systolicBP</i>	Angka tekanan darah maksimum dalam satuan mmHg dan atribut penting lainnya yang relevan selama masa kehamilan
3	<i>diastolicBP</i>	Angka tekanan darah minimal dalam satuan mmHg, serta atribut penting lainnya yang relevan selama masa kehamilan
4	<i>BS</i>	Konsentrasi glukosa darah diukur dalam nilai molar
5	<i>BodyTemp</i>	Suhu tubuh ibu hamil selama periode tertentu dalam kehamilan
6	<i>HeartRate</i>	Denyut jantung istirahat normal
7	<i>Riskleve</i>	Memprediksi tingkat risiko yang intens selama kehamilan dengan mempertimbangkan atribut sebelumnya

Sumber: [15]

2. Preprocessing

Sebelum menjalankan proses klasifikasi, langkah awal yang dilakukan adalah *preprocessing* untuk

mempermudah interpretasi. *Preprocessing* menggunakan *sklearn* yang disediakan *library* oleh *python* yaitu `from sklearn import preprocessing, sklearn.preaprocessing` sebuah modul dalam pustaka *Scikit-learn* (atau *sklearn*) yang digunakan untuk *processing* dari *dataset maternal health risk* sebelum memasukannya kedalam model pembelajaran mesin. pra - pemrosesan data ini penting untuk memastikan bahwa data yang digunakan untuk melatih model memiliki format dan skala yang tepat sehingga model dapat belajar dengan baik dan menghasilkan hasil yang terbaik [16].

Sklearn.preprocessing menyediakan beberapa fungsi untuk normalisasi dan pengolahan data. *StandardScaler* adalah teknik yang menormalkan fitur data sehingga memiliki rata-rata nol dan varians satu, membantu menghindari dominasi fitur skala besar dalam pembelajaran. *MinMaxScaler* mengubah fitur ke rentang tertentu, biasanya antara 0 dan 1. *OneHotEncoder* mengkonversi variabel kategori menjadi format biner, yang cocok untuk data kategorikal tanpa urutan. *LabelEncoder* mengkodekan label kelas menjadi bilangan bulat, sesuai untuk masalah klasifikasi yang membutuhkan target numerik. Terakhir, *Imputer* mengisi nilai yang hilang dalam data, menggunakan statistik seperti *mean* atau *median* dari fitur terkait.

Hyperparameter tuning adalah proses mengoptimalkan parameter yang digunakan dalam model pembelajaran mesin untuk meningkatkan kinerjanya. Ini melibatkan pengujian berbagai kombinasi parameter untuk menemukan set yang memberikan hasil terbaik dalam hal akurasi dan efisiensi model. Metode yang sering digunakan untuk *hyperparameter tuning* termasuk *grid search*, *random search*, dan *Bayesian optimization*, yang semuanya bertujuan untuk mengeksplorasi ruang parameter secara sistematis dan efektif.

Dataset maternal health risk ini sebanyak 1014 dengan label *high risk*, *mid risk*, dan *low risk*. Sampel dari *dataset* ini disajikan dalam Tabel 2 berikut:

Tabel 2. Sampel Dataset *maternal health risk*[15]

No	<i>Age</i>	<i>SystolicBP</i>	<i>DiastolicBP</i>	<i>BS</i>	<i>BodyTemp</i>	<i>HeartRate</i>	<i>RiskLevel</i>
1	25	130	80	15.0	98.0	86	high risk
2	35	140	90	13.0	98.0	70	high risk
3	29	90	70	8.0	100.0	80	high risk
4	30	140	85	7.0	98.0	70	high risk
5	35	120	60	6.1	98.0	76	low risk
...
145	14	90	65	7.0	101.0	70	high risk
146	15	80	60	6.7	98.0	80	low risk
147	15	100	65	6.7	98.0	76	low risk
148	12	95	60	6.7	98.0	77	low risk
149	37	120	90	11.0	98.0	88	high risk

3. Klasifikasi

Klasifikasi adalah proses dimana model belajar untuk memetakan himpunan atribut atau fitur dari objek data ke dalam kelas yang tersedia, berdasarkan pada fungsi target yang telah dilatih. Dengan kata lain, klasifikasi melibatkan penilaian terhadap objek data untuk menentukan kelas yang paling sesuai. Pada penelitian ini menggunakan *library Pycaret* untuk klasifikasi untuk langkah-langkahnya sendiri *import* terlebih dahulu modulnya *from pycaret.classification*, selanjutnya *clf = setup (data=data, target='RiskLevel', session_id=123, preprocess=False)* digunakan untuk mempersiapkan pemodelan. *Best = compare_models()* digunakan untuk membandingkan kinerja beberapa model pembelajaran mesin yang berbeda secara otomatis menggunakan metrik evaluasi yang telah ditentukan. Setelah pemodelan selesai, ini akan mengembalikan model terbaik berdasarkan metrik evaluasi yang dipilih secara otomatis [17].

Sebelum proses klasifikasi pengujian pembelajaran menggunakan algoritma pengembangan model menggunakan *hyperparameter tuning* yang akan dibandingkan dengan tanpa penyetelan *hyperparameter tuning*, beberapa *hyperparameter* umum yang disetel dalam berbagai algoritma *machine learning* diantaranya : *learning rate*, *layer*, ukuran *batch*, jumlah pohon (untuk model berbasis pohon) dan tipe parameter kernel biasanya untuk SVM [18].

Algoritma yang akan diuji berkaitan dengan *decision tree learning* yang biasanya menggunakan model klasifikasi dalam *data mining* [19] diantaranya *Random Forest* [20], *Extra Trees* [21], *Extreme Gradient Boosting* [22], *Decision Tree* [23], dan *Light Gradient Boosting Machine* [24].

4. Uji Cross Validation

Cross validation adalah teknik standar yang digunakan untuk menguji dan mengevaluasi kinerja model atau algoritma dengan memprediksi tingkat kesalahan dan melakukan evaluasi pada data. Salah satu metode *cross validation* yang umum adalah *K-Fold Cross Validation*, di mana data dibagi menjadi K kelompok yang saling tumpang tindih secara acak. Dalam contoh ini, data dibagi menjadi 10 bagian, di mana 90% digunakan untuk pelatihan model dan 10% sisanya digunakan untuk pengujian. Proses ini diulang sebanyak 10 kali sehingga setiap rekaman data memiliki kesempatan menjadi data pengujian.

K-fold Cross Validation adalah salah satu metode yang sangat berguna dalam mengevaluasi kinerja model atau algoritma secara obyektif. Dengan membagi data menjadi K kelompok yang saling tumpang tindih secara acak dan mengulang proses pelatihan dan pengujian sebanyak K kali, kita dapat memastikan bahwa setiap rekaman data memiliki kesempatan untuk menjadi bagian dari data pengujian. Proses ini tidak hanya

memberikan perkiraan yang lebih andal tentang kinerja model, tetapi juga membantu dalam mengidentifikasi apakah model tersebut *overfitting* atau *underfitting* terhadap data. Dengan demikian, *K-Fold Cross Validation* menjadi salah satu alat penting dalam pengembangan dan penyetelan model *machine learning* yang efektif [25].

3. Hasil dan Pembahasan

Setelah *dataset* didapatkan kemudian dilakukan *preprocessing* dengan *library sklearn preprocessing*. Selain itu dilakukan penambahan *hyperparameter tuning* agar meningkatkan performa model *machine learning* sehingga mampu memberikan prediksi yang lebih akurat dan generalisasi yang lebih baik terhadap data baru. Proses ini dilakukan setelah mendapatkan hasil awal dari model. Hasil *preprocessing* ini kemudian diterapkan pada lima model klasifikasi diantaranya *Random Forest*, *Extra Trees*, *Extreme Gradient Boosting*, *Decision Tree*, dan *Light Gradient Boosting Machine*. Dalam penelitian ini, terdapat tiga kelas yang diidentifikasi, yakni *high risk*, *mid risk*, dan *low risk*. Hasil dari pengujian masing-masing algoritma dijelaskan pada bagian ini hingga menemukan akurasi paling tinggi.

Penentuan akurasi merupakan proses menghitung seberapa baik model *machine learning* dalam memprediksi data dengan benar, dengan rumus akurasi (1).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

TP = *True Positive*, Kasus di mana model memprediksi positif (kelas yang diprediksi benar) dan hasil aktualnya juga positif.

FP = *False Positive*, Kasus di mana model memprediksi positif, tetapi hasil aktualnya negatif (kelas yang diprediksi salah).

TN = *True Negative*, Kasus di mana model memprediksi negatif (kelas yang diprediksi benar) dan hasil aktualnya juga *negative*.

FN = *False Negative*, Kasus di mana model memprediksi negatif, tetapi hasil aktualnya positif (kelas yang diprediksi salah).

Untuk mengukur model lebih akurat menggunakan *K-fold cross validation* dari jumlah data dataset 1014 dengan rumus pada persamaan (2).

$$\text{Ukuran Fold} = \frac{\text{jumlah data}}{k} \quad (2)$$

Ukuran *fold* adalah jumlah data yang akan digunakan dalam satu *fold*, dengan *k* adalah jumlah fold yang digunakan, jumlah data akan dibagi menjadi k bagian yang sama besar, setiap bagian tersebut akan digunakan sebagai *dataset* validasi satu persatu. Untuk *K-fold* nya sendiri sebanyak 10 kali dengan alasan evaluasi lebih

stabil dan lebih tahan terhadap *overfitting* dibanding 5 dan lebih 10 memerlukan proses lebih lama. *K-fold cross-validation* dengan $k=10$ cenderung memberikan perkiraan yang cukup stabil tentang kinerja model tanpa terlalu meningkatkan komputasi yang diperlukan [$K=10$] [26].

1. Hasil Pengujian Tanpa Menggunakan *Hyperparameter Tuning*

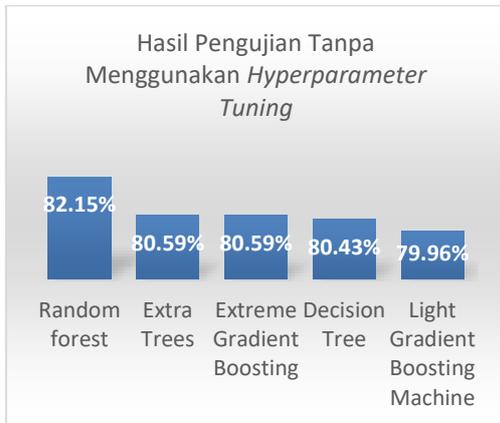
Setelah *dataset* siap untuk diklasifikasi, kelima algoritma yang digunakan menghasilkan nilai akurasi sebagaimana disajikan dalam Tabel 3.

Tabel 3. Data Hasil Pengujian Tanpa Menggunakan *Hyperparameter Tuning*

Model	Accuracy
<i>Random Forest</i>	82,15%
<i>Extra Trees</i>	80,59%
<i>Extreme Gradient Boosting</i>	80,59%
<i>Decision Tree</i>	80,43%
<i>Light Gradient Boosting Machine</i>	79,96%

Sumber: Penelitian (2024)

Dari Tabel 3 diatas, dapat dilihat data hasil pengujian algoritma tanpa menggunakan *hyperparameter tuning* menghasilkan akurasi paling tinggi menggunakan algoritma *Random Forest* dengan nilai akurasi sebesar 82,15% dan yang paling rendah hasil algoritma *Light Gradient Boosting Machine* sebesar 79,96%.



Gambar 2. Diagram Hasil Pengujian Tanpa Menggunakan *Hyperparameter Tuning*

Dari Gambar 2, dapat dilihat diagram hasil pengujian algoritma tanpa menggunakan *hyperparameter tuning* menghasilkan akurasi paling tinggi menggunakan algoritma *Random Forest* dengan nilai akurasi sebesar 82,15%, dilanjutkan dengan hasil algoritma *Extra Trees* sebesar 80,59%, hasil algoritma *Extreme Gradient Boosting* sebesar 80,59%, hasil algoritma *Decision Tree* sebesar 80,43%, dan yang paling rendah hasil algoritma *Light Gradient Boosting Machine* sebesar 79,96%.

2. Hasil Pengujian Algoritma Dengan *Hyperparameter Tuning*

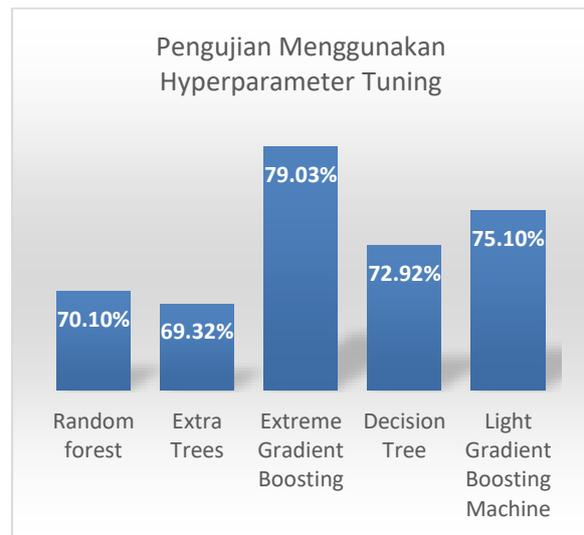
Pengujian algoritma dengan menggunakan *hyperparameter tuning* juga telah dilakukan pada kelima algoritma yang digunakan dengan hasil akurasi masing-masing algoritma dapat dilihat pada Tabel 4.

Tabel 4. Hasil Pengujian Algoritma Dengan *Hyperparameter Tuning*

Algoritma	Hasil Pengujian dengan <i>Hyperparameter Tuning</i>
<i>Random Forest</i>	70,10%
<i>Extra Trees</i>	69,32%
<i>Extreme Gradient Boosting</i>	79,03%
<i>Decision Tree</i>	72,92%
<i>Light Gradient Boosting Machine</i>	75,10%

Sumber: Penelitian (2024)

Dari Tabel 4 diatas, dapat dilihat data hasil pengujian algoritma dengan menggunakan *hyperparameter tuning* menghasilkan akurasi paling tinggi menggunakan algoritma *Extreme Gradient Boosting* dengan nilai akurasi sebesar 79,03% dan yang paling rendah hasil akurasi *Extra Trees* sebesar 69,32%.



Gambar 3. Diagram Hasil Pengujian Algoritma Dengan *Hyperparameter Tuning*

Dari Gambar 3 diatas, dapat dilihat diagram hasil pengujian algoritma dengan menggunakan *hyperparameter tuning* menghasilkan akurasi paling tinggi menggunakan algoritma *Extreme Gradient Boosting* dengan nilai akurasi sebesar 79,03%, dilanjutkan dengan hasil algoritma *Light Gradient Boosting Machine* sebesar 75,10%, hasil algoritma *Decision Tree* sebesar 72,92%, hasil algoritma *Random Forest* sebesar 70,10%, dan yang paling rendah hasil algoritma *Extra Trees* sebesar 69,32%.

3. Perbandingan Hasil Pengujian Dengan dan Tanpa *Hyperparameter Tuning*

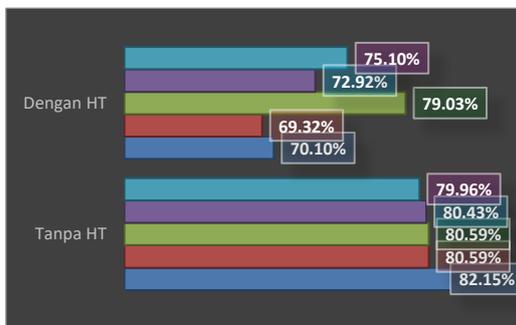
Hasil pengujian kelima algoritma baik dengan maupun tanpa *hyperparameter tuning* masing-masing sudah didapatkan kemudian perlu dibandingkan sebagai evaluasi algoritma yang paling cocok diterapkan terhadap *dataset* yang digunakan. Hasil perbandingan ini disajikan pada Tabel 5.

Tabel 5. Perbandingan Hasil Pengujian Dengan dan Tanpa *Hyperparameter Tuning*

Algoritma	Tanpa HT	Dengan HT
<i>Random Forest</i>	82,15%	70,10%
<i>Extra Trees</i>	80,59%	69,32%
<i>Extreme Gradient Boosting</i>	80,59%	79,03%
<i>Decision Tree</i>	80,43%	72,92%
<i>Light Gradient Boosting Machine</i>	79,96%	75,10%

Sumber: Penelitian (2024)

Dari Tabel 5 diatas, dapat dilihat data perbandingan hasil pengujian algoritma dengan dan tanpa menggunakan *hyperparameter tuning* menghasilkan akurasi paling tinggi menggunakan algoritma *Random Forest* dengan nilai akurasi sebesar 82,15% dan yang paling rendah hasil dari algoritma *Extra Trees* dengan menggunakan *hyperparameter tuning* yaitu sebesar 69,32%.



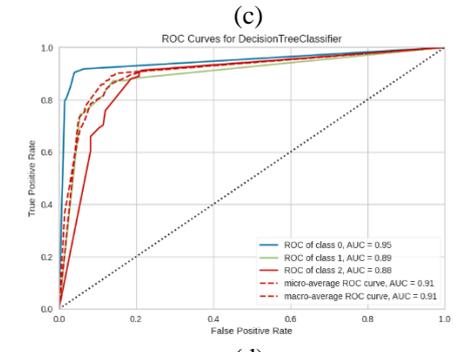
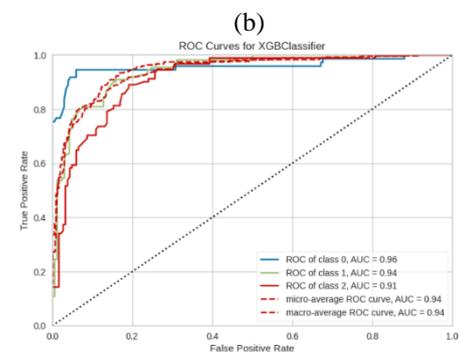
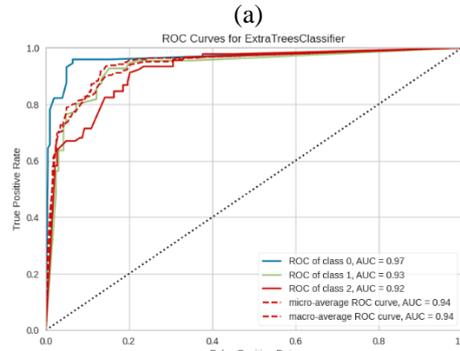
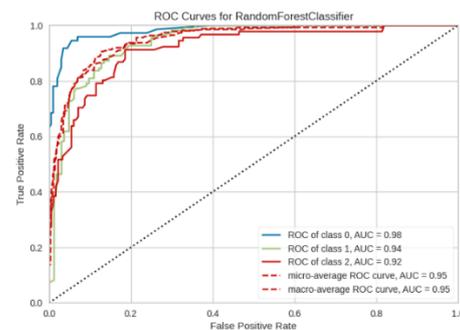
Gambar 4. Diagram Perbandingan Hasil Pengujian Algoritma Dengan dan Tanpa Menggunakan *Hyperparameter Tuning*

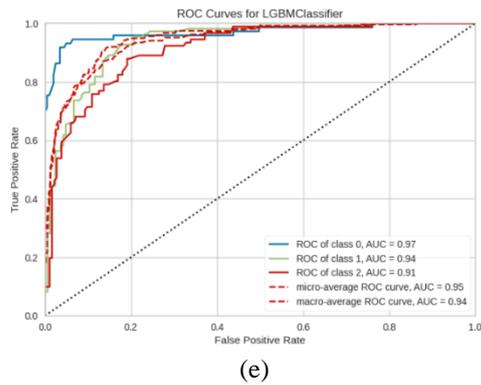
Dari Gambar 4 diatas, dapat dilihat diagram perbandingan hasil pengujian algoritma dengan dan tanpa menggunakan *hyperparameter tuning* menghasilkan akurasi paling tinggi menggunakan algoritma *Random Forest* dengan nilai akurasi sebesar 82,15%.

Dalam konteks ini, penerapan model tanpa penyetulan *hyperparameter tuning* menghasilkan akurasi tertinggi pada algoritma *Random Forest* sebesar 82,15%, *Extra Trees* sebesar 80,59%, *Extreme Gradient Boosting* sebesar 80,59%, *Decision Tree* sebesar 80,43%, dan *Light Gradient Boosting Machine* sebesar 79,96%.

Sementara itu, penerapan model dengan optimasi *Hyperparameter Tuning* menghasilkan akurasi tertinggi pada algoritma *Extreme Gradient Boosting* sebesar 79,03%, *Light Gradient Boosting Machine* sebesar 75,10%, *Decision Tree* sebesar 72,92%, *Random Forest* sebesar 70,10%, dan *Extra Trees* sebesar 69,32%.

Metrik evaluasi pembelajaran mesin memberikan gambaran yang bisa mengukur seberapa baik model dapat membedakan antara dua kelas dengan menghitung area dibawah kurva seperti Gambar 5.





Gambar 5. Hasil AUC

AUC dapat berkisar dari 0 hingga 1 dan memberikan ukuran kinerja yang agregat di semua ambang batas klasifikasi yang mungkin. AUC 0,5 menunjukkan tidak ada diskriminasi (yaitu, model tidak memiliki kemampuan untuk membedakan antara kelas positif dan negatif), sedangkan AUC 1,0 menunjukkan diskriminasi yang sempurna (yaitu, model menggunakan dengan sempurna antara kelas positif dan negatif tanpa kesalahan). (a) untuk algoritma *Random Forest*, (b) untuk algoritma *Extra Trees*, (c) untuk algoritma *Extreme Gradient Boosting*, (d) untuk algoritma *Decision Tree* dan (e) untuk algoritma *Light Gradient Boosting Machine*.

4. Kesimpulan

Evaluasi dari pembelajaran mesin yang diolah dari *dataset maternal health risk* disajikan dalam penelitian ini dengan komprehensif dari awal *dataset* diolah menggunakan pemroses *processing* data *sklearn.preprocessing* yang meliputi *StandardScaler*, *MinMaxScaler*, *OneHotEncoder* dan *LabelEncoder* yang kemudian dilanjut dengan proses klasifikasi penyetelan *hyperparameter tuning* dan tanpa penyetelan *hyperparameter tuning* terhadap algoritma, diantaranya *Random Forest*, *Extra Trees*, *Extreme Gradient Boosting*, *Decision Tree* dan *Light Gradient Boosting Machine*. Berdasarkan analisis dan pembahasan, kesimpulan dapat diambil bahwa algoritma *Random Forest* (RF) dalam model tanpa penyetelan *hyperparameter* menunjukkan kinerja terbaik dalam mendeteksi kondisi kesehatan ibu hamil, dengan akurasi 82.15% dengan demikian, dapat disimpulkan bahwa meskipun algoritma *Random Forest* (RF) telah menunjukkan kinerja yang sangat baik dalam mendeteksi kondisi kesehatan ibu hamil dengan akurasi sebesar 82,15% tanpa penyetelan *hyperparameter*, masih terdapat ruang untuk peningkatan kinerja model. Langkah selanjutnya yang dapat diambil adalah melakukan penyetelan *hyperparameter* yang lebih cermat untuk memperbaiki performa model lebih lanjut. Selain itu, eksplorasi terhadap teknik *feature engineering* dan pemilihan fitur yang lebih baik juga dapat membantu meningkatkan

keakuratan prediksi. Selanjutnya, evaluasi lebih lanjut terhadap generalisasi model diharapkan bahwa model akan lebih handal dan dapat digunakan secara efektif dalam mendukung keputusan klinis yang berkaitan dengan kesehatan ibu hamil.

Daftar Rujukan

- [1] T. R. P. Lestari, "Achievement of Mother and Baby Health Status As One of the Successes of Mother and Child Health Programs," *Kajian*, vol. 25, no. 1, pp. 75–89, 2020.
- [2] P. Carr, "Add to my library Features :," *Implement. Cult. Chang.*, vol. 2003, no. 0404, p. 114, 2006.
- [3] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, "Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest," *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, 2022.
- [4] S. R. Chowdhury, S. Mishra, A. O. Miranda, and P. K. Mallick, "Energy Consumption Prediction Using Light Gradient Boosting Machine Model," *Lect. Notes Electr. Eng.*, vol. 690, pp. 413–422, 2021.
- [5] L. Ni *et al.*, "Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model," *J. Hydrol.*, vol. 586, no. March, p. 124901, 2020.
- [6] M. Ahmed, M. A. Kashem, M. Rahman, and S. Khatun, "Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT)," *Lect. Notes Electr. Eng.*, vol. 632, pp. 357–365, 2020.
- [7] N. Dehingia *et al.*, "Unintended pregnancy and maternal health complications: Cross-sectional analysis of data from rural Uttar Pradesh, India," *BMC Pregnancy Childbirth*, vol. 20, no. 1, pp. 1–11, 2020.
- [8] M. Ahmed and M. A. Kashem, "2022 4th International Conference on Sustainable Technologies for Industry 4.0, STI 2022," *2022 4th Int. Conf. Sustain. Technol. Ind. 4.0, STI 2022*, vol. 0, pp. 19–20, 2022.
- [9] A. Pinandito, S. A. Wicaksono, and S. H. Wijoyo, "Implementasi Machine Learning dalam Deteksi Risiko Tinggi Diabetes Melitus pada Kehamilan," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 4, pp. 739–746, 2023.
- [10] C. Oganis, S. Musdalifah, and D. Lusiyanti, "Klasifikasi Status Gizi Ibu Hamil Untuk Mengidentifikasi Bayi Berat Lahir Rendah (Bblr) Menggunakan Metode Support Vector Machine (Svm) (Studi Kasus Di Puskesmas Labuan)," *J. Ilm. Mat. Dan Terap.*, vol. 14, no. 2, pp. 144–151, 2017.
- [11] J. Khatib Sulaiman, M. Dzakwan Ar Rosyid, and I. Artikel Abstrak, "Klasifikasi Tingkat Risiko Kesehatan Ibu Hamil Menggunakan Algoritma Support Vectore Machine Universitas AMIKOM Yogyakarta," *Indones. J. Comput. Sci. Attrib.*, vol. 12, no. 5, pp. 2023–2798, 2023.
- [12] R. A. F. W. A. H. W. K. Rahman, "Analisis Perbandingan Algoritma Machine Learning untuk Klasifikasi Tingkat Risiko Ibu Hamil," *Student Res. J.*, vol. 1, no. 6, pp. 246–261, 2023.
- [13] O. Sagi and L. Rokach, "Approximating XGBoost with an interpretable decision tree," *Inf. Sci. (Ny.)*, vol. 572, pp. 522–542, 2021.
- [14] T. Rahman, "Perbandingan Produkt. Kerja Karyawan Sebelum Dan Seteleah Pelatih. Pada PT Kuwera Jaya Makassar", pp. 12–26, 2018.
- [15] I. Budi and R. R. Suryono, "Application of named entity recognition method for Indonesian datasets: a review," *Bull. Electr. Eng. Informatics*, vol. 12, no. 2, pp. 969–978, 2023.
- [16] F. T. Admojo and S. R. Jabir, "Analisis performa metode Naïve Bayesh Classifier pada Electronic Nose dalam identifikasi formalin pada tahu," *Indones. J. Data Sci.*, vol. 4, no. 1, pp. 1–16, 2023.
- [17] E. Eka Citra, D. Hatta Fudholi, and C. Kusuma Dewa, "Implementasi Arsitektur EfficientNetV2 Untuk Klasifikasi Gambar Makanan Tradisional Indonesia," *J. Media Inform. Budidarma*, vol. 7, no. 2, pp. 766–777, 2023.

- [18] I. Carolina and T. Haryanto, "JITE (Journal of Informatics and Telecommunication Engineering) Modeling Of Hyperparameter Tuned RNN-LSTM and Deep Learning," vol. 7, no. January, pp. 502–513, 2024.
- [19] S. S. Gavankar and S. D. Sawarkar, "Eager decision tree," *2017 2nd Int. Conf. Conver. Technol. I2CT 2017*, vol. 2017-Janua, pp. 837–840, 2017.
- [20] S. M. Ahmed *et al.*, "Derivation and external validation of clinical prediction rules identifying children at risk of linear growth faltering," *Elife*, vol. 12, pp. 1–18, 2023.
- [21] Z. Arif Ali, Z. H. Abduljabbar, H. A. Tahir, A. Bibo Sallow, and S. M. Almufti, "eXtreme Gradient Boosting Algorithm with Machine Learning: a Review," *Acad. J. Nawroz Univ.*, vol. 12, no. 2, pp. 320–334, 2023.
- [22] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021.
- [23] E. Pekel, "Estimation of soil moisture using decision tree regression," *Theor. Appl. Climatol.*, vol. 139, no. 3–4, pp. 1111–1119, 2020.
- [24] S. Islam and S. H. Amin, "Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques," *J. Big Data*, vol. 7, no. 1, 2020.
- [25] A. N. Safira, B. Warsito, and A. Rusgiyono, "Analisis Support Vector Regression (Svr) Dengan Algoritma Grid Search Time Series Cross Validation Untuk Prediksi Jumlah Kasus Terkonfirmasi Covid-19 Di Indonesia," *J. Gaussian*, vol. 11, no. 4, pp. 512–521, 2023.
- [26] K. K. Metode, D., & Neighbors, "Jurnal Indonesia : Manajemen Informatika dan Komunikasi Aspect-Based Sentiment Analysis Terhadap Ulasan Aplikasi Flip Menggunakan Pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) Abstrak Jurnal Indonesia : Manajemen Informatika dan Komunika," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 3, pp. 1858–1873, 2023.