

ANALISIS PREDIKTIF RISIKO POSTPARTUM PADA IBU PRIMIPARA DENGAN ALGORITMA RANDOM FOREST DENGAN RANDOM SEARCH BERDASARKAN FAKTOR PSIKOLOGIS

Angel Fourtuna Dengen*¹⁾, Ema Utami²⁾

1. Magister Informatika, Sekolah Pascasarjana, Universitas Amikom Yogyakarta
2. Magister Informatika, Sekolah Pascasarjana, Universitas Amikom Yogyakarta

Article Info

Kata Kunci: depresi postpartum, ibu primipara, Random Forest, machine learning, prediksi risiko

Article history:

Received 5 January 2025

Revised 15 January 2025

Accepted 14 February 2025

Available online 1 December 2025

DOI :

<https://doi.org/10.29100/jipi.v10i4.7255>

* Corresponding author.

Corresponding Author

E-mail address:

angelf.dengen@students.amikom.ac.id

ABSTRAK

Penelitian ini membahas penerapan algoritma Random Forest untuk memprediksi risiko depresi postpartum pada ibu primipara berdasarkan faktor psikologis yang relevan. Dataset diperoleh dari Kaggle dengan total 1.503 observasi yang mencakup berbagai atribut psikologis seperti tingkat stres, dukungan sosial, dan riwayat kesehatan mental. Data diproses melalui tahapan pra-pemrosesan seperti pembersihan data, normalisasi, dan encoding variabel kategorikal untuk memastikan kualitas data yang optimal. Model Random Forest dikembangkan menggunakan hyperparameter yang dioptimalkan melalui Random Search, seperti jumlah pohon dalam hutan dan kedalaman maksimum, guna meningkatkan akurasi prediksi. Evaluasi performa model menggunakan metrik seperti akurasi, precision, recall, dan F1-score menunjukkan akurasi keseluruhan sebesar 90%. Hasil menunjukkan bahwa model mampu mengenali kelas mayoritas dengan baik (recall 95%), namun performa pada kelas minoritas masih perlu ditingkatkan (recall 79%). Penelitian ini memberikan kontribusi penting dalam pengembangan metode prediksi risiko postpartum yang lebih akurat serta memberikan landasan untuk strategi intervensi dini berbasis data. Selain itu, penelitian ini juga mengidentifikasi atribut psikologis yang paling berpengaruh terhadap risiko postpartum, seperti tingkat stres dan dukungan sosial, sehingga dapat menjadi acuan dalam perancangan kebijakan kesehatan masyarakat.

ABSTRACT

This study discusses the application of the Random Forest algorithm to predict the risk of postpartum depression in primiparous mothers based on relevant psychological factors. The dataset was obtained from Kaggle, comprising a total of 1,503 observations that include various psychological attributes such as stress levels, social support, and mental health history. The data underwent preprocessing steps, including data cleaning, normalization, and encoding of categorical variables to ensure optimal data quality. The Random Forest model was developed using hyperparameters optimized through Random Search, such as the number of trees in the forest and maximum depth, to improve prediction accuracy. Model performance evaluation using metrics like accuracy, precision, recall, and F1-score showed an overall accuracy of 90%. The results indicate that the model effectively identified the majority class (recall 95%), although its performance on the minority class requires improvement (recall 79%). This study makes a significant contribution to developing more accurate methods for predicting postpartum risk and provides a foundation for data-driven early intervention strategies. Furthermore, the study identifies the most influential psychological attributes related to postpartum risk, such as stress levels and social support, offering valuable insights for designing public health policies.

I. PENDAHULUAN

Masa postpartum merupakan periode krusial dalam kehidupan seorang ibu, terutama bagi ibu primipara, yaitu ibu yang baru pertama kali melahirkan [1]. Secara global, sekitar 10% wanita hamil dan 13% wanita yang baru saja melahirkan mengalami gangguan mental, terutama depresi. Di negara berkembang, angkanya bahkan lebih tinggi, yaitu 15,6% selama kehamilan dan 19,8% setelah melahirkan. Dalam kasus yang parah, penderitaan ibu bisa begitu berat hingga mereka mungkin melakukan bunuh diri. Selain itu, ibu yang terdampak tidak dapat berfungsi dengan baik. Akibatnya, pertumbuhan dan perkembangan anak juga dapat terpengaruh secara negatif. [2]. Selama periode ini, ibu sering menghadapi berbagai tantangan fisik dan emosional yang dapat meningkatkan risiko gangguan psikologis, seperti depresi postpartum. Depresi postpartum adalah kondisi yang tidak hanya berdampak pada kesehatan ibu, tetapi juga pada perkembangan anak dan dinamika keluarga. [3], [4]. Oleh karena itu, deteksi dini terhadap risiko postpartum sangat penting untuk dilakukan.

Faktor psikologis, seperti tingkat stres, dukungan sosial, kecemasan, dan sejarah gangguan mental sebelumnya, merupakan variabel utama yang berkontribusi terhadap risiko depresi postpartum. Identifikasi hubungan antara faktor-faktor ini dengan risiko postpartum membutuhkan pendekatan analitik yang cermat. Dalam beberapa tahun terakhir, pendekatan berbasis machine learning telah menunjukkan potensi besar dalam analisis prediktif di bidang kesehatan. [5], [6]. Salah satu algoritma yang banyak digunakan dalam prediksi berbasis data adalah Random Forest.

Penelitian ini menghadirkan pendekatan baru dalam prediksi risiko postpartum dengan memanfaatkan algoritma Random Forest yang dioptimalkan menggunakan Random Search. Random Forest adalah algoritma ensemble learning yang kuat, mampu menangani dataset dengan dimensi tinggi, dan memberikan hasil prediksi yang akurat. [7], [8]. Algoritma ini bekerja dengan membangun sejumlah besar pohon keputusan dan menggabungkan hasil prediksinya untuk meningkatkan keandalan dan akurasi [9], [10]. Hyperparameter algoritma Random Forest, seperti jumlah pohon keputusan ($n_estimators$), kedalaman maksimum (max_depth), dan lainnya, dioptimalkan menggunakan Random Search. Pendekatan ini dipilih karena efisiensinya dalam menemukan kombinasi parameter terbaik dibandingkan dengan metode tradisional seperti Grid Search [11], [12]. Selain itu, penelitian ini secara khusus fokus pada analisis faktor psikologis sebagai variabel utama, yang jarang dieksplorasi secara mendalam dalam penelitian sebelumnya. Dalam konteks penelitian ini, Random Forest digunakan untuk menganalisis data psikologis ibu primipara dan memprediksi risiko postpartum berdasarkan kombinasi berbagai faktor. Random Forest dipilih karena sejumlah keunggulan yang ditawarkannya. Sebagai algoritma ensemble, Random Forest mampu mengurangi risiko overfitting yang sering terjadi pada pohon keputusan tunggal dengan cara menggabungkan hasil dari banyak pohon Keputusan [13], [14]. Metode ini juga dapat menangani data yang memiliki dimensi tinggi serta variabel input yang saling berinteraksi. Selain itu, Random Forest memiliki keunggulan dalam memberikan informasi tentang pentingnya setiap variabel dalam proses prediksi, yang sangat relevan dalam penelitian ini untuk mengidentifikasi faktor psikologis utama yang memengaruhi risiko postpartum [15]. Di sisi lain, Random Search dipilih sebagai metode optimasi karena sifatnya yang lebih efisien dibandingkan Grid Search dalam menjelajahi ruang hyperparameter yang luas [16]. Alih-alih mencoba semua kombinasi parameter secara sistematis seperti Grid Search, Random Search melakukan sampling acak, yang memungkinkan untuk menemukan kombinasi optimal dengan waktu komputasi yang lebih singkat. Hal ini membuat Random Search lebih praktis dan hemat sumber daya, terutama ketika berhadapan dengan dataset yang besar dan kompleks [17], [18]. Sebagai perbandingan, algoritma seperti Support Vector Machine (SVM) atau Logistic Regression memerlukan preprocessing tambahan dan tidak selalu seefisien Random Forest dalam menangani data yang multivariabel dan tidak seimbang.

Implementasi model prediksi ini dalam praktik klinis memiliki potensi besar untuk membantu tenaga kesehatan mengidentifikasi ibu dengan risiko tinggi depresi postpartum secara lebih dini. Hal ini memungkinkan intervensi seperti konseling psikologis, pemberian dukungan sosial, atau perawatan medis dapat dilakukan dengan lebih cepat dan efisien. Selain itu, penggunaan model ini juga dapat mengurangi beban kerja tenaga medis melalui otomatisasi proses skrining risiko, sehingga alokasi sumber daya kesehatan dapat lebih optimal, terutama di wilayah dengan keterbatasan akses layanan kesehatan. Dengan demikian, hasil penelitian ini diharapkan memberikan wawasan baru bagi tenaga medis dan ahli kesehatan dalam merancang strategi intervensi dini yang lebih efektif. Deteksi dini memungkinkan pemberian layanan kesehatan yang lebih tepat sasaran, seperti konseling psikologis atau penguatan dukungan sosial, sehingga sumber daya dapat difokuskan pada kelompok yang memerlukan perhatian lebih. Langkah ini tidak hanya meningkatkan efisiensi layanan kesehatan tetapi juga efektivitas intervensi yang dilakukan.

Selain itu, penelitian ini memberikan pemahaman lebih mendalam mengenai faktor-faktor psikologis utama yang memengaruhi risiko postpartum. Informasi ini dapat dimanfaatkan untuk merancang program intervensi yang lebih terarah, seperti pelatihan pengelolaan stres atau penguatan jejaring dukungan sosial. Dengan menurunkan

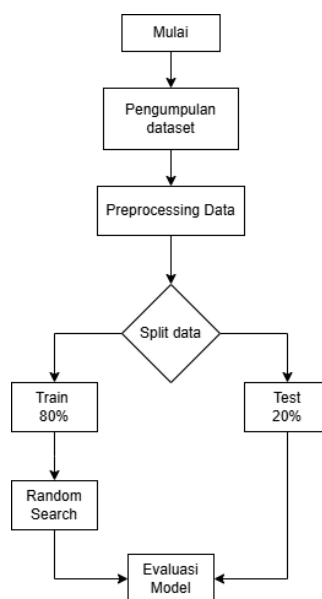
risiko depresi postpartum, kualitas hidup ibu dapat meningkat, yang pada akhirnya juga berdampak positif pada perkembangan anak dan stabilitas keluarga. Pendekatan berbasis machine learning yang digunakan dalam penelitian ini membuka peluang untuk penelitian lebih lanjut, khususnya dalam mengeksplorasi aplikasi teknologi serupa di bidang kesehatan lainnya. Penelitian ini tidak hanya berkontribusi pada pengembangan ilmu pengetahuan di bidang kesehatan, tetapi juga memberikan manfaat praktis yang nyata bagi komunitas medis, tenaga kesehatan, dan masyarakat luas. Dengan mendukung upaya pencegahan dan pengelolaan risiko postpartum yang lebih baik, penelitian ini turut membantu menciptakan sistem kesehatan yang lebih responsif dan berkelanjutan.

Struktur naskah ini akan dibagi ke dalam beberapa bagian. Bagian berikutnya akan mengulas metode yang digunakan dalam penelitian ini, meliputi pengumpulan data, preprocessing data, dan pengembangan model. Selanjutnya, hasil analisis dan interpretasi model akan dijelaskan, diikuti oleh diskusi yang menghubungkan hasil penelitian dengan temuan-temuan sebelumnya. Akhirnya, kesimpulan dan implikasi penelitian akan dirangkum, disertai dengan rekomendasi untuk penelitian lanjutan.

Dengan pendekatan ini, diharapkan penelitian ini dapat berkontribusi pada pengembangan metode prediksi risiko postpartum yang lebih akurat, serta memberikan landasan untuk pengembangan intervensi yang lebih efektif bagi ibu primipara di masa depan.

II. METODE PENELITIAN

Tahapan berikutnya adalah pengembangan model prediktif menggunakan algoritma Random Forest. Pengolahan dilakukan menggunakan perangkat lunak Python dengan pustaka *scikit-learn*. Dataset dibagi menjadi data latih (80%) dan data uji (20%) untuk mengoptimalkan pelatihan dan evaluasi model. Hyperparameter algoritma Random Forest, seperti jumlah pohon keputusan ($n_estimators$) dan kedalaman maksimum (max_depth), dioptimalkan menggunakan metode Random Search, yang memberikan efisiensi lebih tinggi dalam menemukan kombinasi parameter terbaik dibandingkan Grid Search. Setelah itu, model dilatih menggunakan data latih dengan parameter yang telah dioptimalkan untuk meningkatkan performanya. Model ini kemudian digunakan untuk menganalisis pola-pola yang relevan dari data yang telah diproses dan membangun model prediktif yang dapat mengidentifikasi individu dengan risiko tinggi depresi postpartum.



Gambar 1 Alur Penelitian

A. Pengumpulan Dataset

Pengumpulan data dilakukan untuk mendapatkan berbagai macam data yang diperlukan dalam penelitian. Dataset yang digunakan pada penelitian ini diperoleh website Kaggle yang dikelola oleh Parvez Al Muqtadir, mencakup dari 11 atribut dan memilih 9 atribut. Link sebagai dataset sebagai berikut <https://www.kaggle.com/datasets/parvezalmuqtadir2348/postpartum-depression>. Dataset "Postpartum Depression" termasuk dalam kategori medis dan kesehatan mental, yang berfokus pada analisis depresi setelah melahirkan. Dataset ini menggabungkan data kategorikal dan numerik yang diperoleh dari survei kuesioner

berbasis Google Form yang diisi oleh pasien rumah sakit, dengan total 1.503 observasi dan berbagai variabel. Variabel-variabel tersebut meliputi aspek demografi, kesehatan, sosial, ekonomi, dan psikologi. Variabel demografi mencakup usia ibu, tingkat pendidikan, status pernikahan, dan jumlah kelahiran sebelumnya. Variabel kesehatan mencakup riwayat depresi sebelum kehamilan, komplikasi selama kehamilan, status menyusui, dan kualitas tidur pasca melahirkan. Dari sisi sosial dan ekonomi, dataset ini mencatat kondisi ekonomi keluarga, dukungan sosial, dan status pekerjaan ibu. Sedangkan variabel psikologis meliputi tingkat stres selama kehamilan, kesehatan mental selama kehamilan, dan kepuasan terhadap proses persalinan. Variabel target dalam dataset ini adalah depresi pascapersalinan, yang diukur dengan skala biner (ada/tidak ada) atau skala tertentu. Dataset ini tersedia di Kaggle dan bertujuan untuk analisis statistik, pengembangan model prediktif menggunakan pembelajaran mesin, serta penelitian kesehatan untuk memahami dan mencegah depresi pascapersalinan. Data ini penting untuk mengidentifikasi faktor risiko dan merancang strategi intervensi bagi ibu yang berisiko

B. Pra-pemrosesan Data

Tahapan *preprocessing* data dilakukan untuk memastikan kualitas data yang digunakan dalam penelitian ini. Langkah pertama adalah pembersihan data, yaitu menghapus entri yang tidak lengkap atau tidak valid untuk menjaga integritas dataset. Selanjutnya, dilakukan normalisasi data guna menyesuaikan skala variabel agar cocok untuk analisis menggunakan algoritma Random Forest. Setelah itu, variabel kategorikal dalam dataset diubah menjadi format numerik melalui teknik *encoding*, sehingga dapat digunakan dalam proses analisis lebih lanjut. Setelah transformasi variabel kategorikal selesai, fitur-fitur yang paling relevan dipilih menggunakan metode *Sequential Forward Selection* (SFS). Metode ini memastikan bahwa hanya faktor-faktor psikologis yang memiliki kontribusi signifikan terhadap model prediktif yang digunakan dalam pengembangan model, sehingga meningkatkan akurasi dan efisiensi analisis.

C. Pengembangan Model

Pengembangan model prediktif dilakukan dengan menggunakan algoritma Random Forest yang diimplementasikan melalui perangkat lunak Python menggunakan pustaka *scikit-learn*. Tahapan awal pengembangan model adalah pembagian dataset menjadi dua subset, yaitu data latih (*training data*) dan data uji (*testing data*), dengan proporsi 80:20. Pembagian ini dilakukan untuk memastikan bahwa model dapat dilatih dengan data yang memadai sekaligus dievaluasi pada data yang belum pernah dilihat sebelumnya, sehingga menghindari risiko *overfitting*. Proses ini bertujuan untuk memastikan bahwa model tidak hanya bekerja baik pada data latih, tetapi juga memiliki kemampuan generalisasi yang baik pada data baru.

Penelitian ini mengusulkan pendekatan baru dalam prediksi risiko depresi postpartum pada ibu primipara dengan menerapkan algoritma *Random Forest* yang dioptimalkan menggunakan metode *Random Search*. Pendekatan ini memiliki keunggulan efisiensi dalam menentukan kombinasi hyperparameter terbaik dibandingkan metode tradisional seperti *Grid Search*. Selanjutnya, untuk meningkatkan performa model, dilakukan optimasi *hyperparameter* algoritma Random Forest. Beberapa *hyperparameter* utama yang dioptimalkan meliputi jumlah pohon keputusan (*n_estimators*), kedalaman maksimum pohon (*max_depth*), serta parameter lain yang memengaruhi struktur dan fungsi model. Optimasi ini dilakukan menggunakan metode *Random Search*, yang merupakan pendekatan lebih efisien dibandingkan *Grid Search*, karena metode ini secara acak mencari kombinasi parameter terbaik dalam ruang pencarian yang telah ditentukan sebelumnya. Keunggulan *Random Search* terletak pada kemampuannya untuk mengurangi waktu komputasi sambil tetap menemukan kombinasi *hyperparameter* yang optimal untuk model.

Setelah optimasi selesai, model dilatih menggunakan data latih dengan kombinasi parameter yang telah dioptimalkan. Proses pelatihan ini melibatkan pembentukan sejumlah besar pohon keputusan, di mana setiap pohon dilatih pada subset data yang berbeda untuk memaksimalkan keanekaragaman model. Hasil dari setiap pohon keputusan kemudian digabungkan melalui teknik agregasi untuk menghasilkan prediksi akhir yang lebih akurat. Proses ini memungkinkan Random Forest untuk mengatasi tantangan dataset dengan dimensi tinggi, menghadirkan kemampuan yang andal dalam memprediksi risiko postpartum pada ibu primipara berdasarkan faktor-faktor psikologis yang telah ditentukan sebelumnya. Dengan pendekatan ini, model yang dihasilkan diharapkan dapat memberikan prediksi yang akurat dan konsisten, sehingga bermanfaat untuk implementasi dalam konteks klinis.

D. Pelatihan Model

Setelah optimasi *hyperparameter*, model dilatih menggunakan data latih dengan parameter default sebagai baseline, yang kemudian disesuaikan berdasarkan hasil optimasi untuk meningkatkan performa. Algoritma

Random Forest membangun sejumlah besar pohon keputusan menggunakan *bootstrap sampling* untuk meningkatkan keanekaragaman dan mengurangi risiko *overfitting*. Prediksi akhir diperoleh melalui teknik agregasi, seperti *majority voting*, untuk menghasilkan model yang akurat dan andal. Pelatihan ini memastikan model dapat mengenali pola-pola kompleks dalam data, sehingga efektif untuk memprediksi risiko postpartum berdasarkan faktor psikologis.

E. Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa prediktif algoritma Random Forest dengan menggunakan berbagai metrik evaluasi, seperti akurasi, *precision*, *recall*, dan F1-score. Akurasi mengukur proporsi prediksi yang benar dibandingkan dengan keseluruhan data uji, sementara *precision* mengindikasikan proporsi prediksi positif yang benar-benar relevan. *Recall* menunjukkan kemampuan model untuk mendeteksi seluruh data positif sebenarnya, sedangkan F1-score memberikan keseimbangan antara *precision* dan *recall*, terutama ketika terdapat ketidakseimbangan kelas dalam dataset.

$$P = \frac{TP}{TP+FP} \quad (1)$$

$$R = \frac{TP}{TP+FN} \quad (2)$$

$$F1 = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (3)$$

$$A = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Selain itu, evaluasi dilengkapi dengan analisis *confusion matrix* yang memberikan gambaran detail mengenai prediksi model, termasuk jumlah *true positives* (TP), *true negatives* (TN), *false positives* (FP), dan *false negatives* (FN). Hasil evaluasi ini memberikan wawasan mendalam tentang kekuatan dan kelemahan model, sehingga memungkinkan penyempurnaan lebih lanjut jika diperlukan.

F. Kesimpulan

Kesimpulan merupakan tahap untuk merangkum hasil analisis dan evaluasi dari penelitian ini. Pada bagian ini, hasil evaluasi model seperti akurasi, presisi, recall, dan nilai F1-score dianalisis untuk menilai performa algoritma Random Forest dengan Random Search dalam memprediksi risiko postpartum pada ibu primipara berdasarkan faktor psikologis. Selain itu, kesimpulan memberikan wawasan mengenai keunggulan metode yang digunakan serta saran untuk penelitian lanjutan yang melibatkan dataset yang lebih besar dan model yang lebih kompleks guna meningkatkan akurasi prediksi.

III. HASIL DAN PEMBAHASAN

Penelitian ini mengusulkan pendekatan prediksi risiko postpartum menggunakan algoritma Random Forest yang dioptimalkan dengan metode Random Search. Hasil eksperimen dibandingkan dengan temuan dari penelitian sebelumnya untuk menilai keunggulan pendekatan ini. Sebanyak enam rujukan ilmiah digunakan untuk mendukung pembahasan hasil eksperimen, termasuk studi tentang algoritma Random Forest, optimasi hyperparameter, dan tantangan analisis data tidak seimbang.

Sebagai contoh penerapan, data sampel diambil dari dataset yang mencakup berbagai faktor psikologis ibu pascapersalinan. Dataset ini dibagi menjadi data latih dan data uji dengan proporsi 80:20. Algoritma Random Forest diimplementasikan menggunakan pustaka scikit-learn di Python untuk membangun model prediksi risiko depresi postpartum. Sebelum pelatihan model, fitur-fitur yang paling relevan dipilih menggunakan metode Sequential Forward Selection (SFS), yang membantu menyaring faktor-faktor psikologis utama, seperti tingkat stres, dukungan sosial, dan riwayat kesehatan mental, untuk meningkatkan akurasi model. Selama pengembangan model, hyperparameter utama seperti jumlah pohon keputusan (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*) dioptimalkan menggunakan metode Random Search. Model yang dihasilkan kemudian dilatih menggunakan data latih untuk memprediksi label pada data uji. Evaluasi kinerja dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score.

Hasil evaluasi menunjukkan bahwa model memiliki performa yang cukup tinggi, mengindikasikan efektivitas algoritma Random Forest dalam menangani data yang kompleks dan multivariabel. Hal ini menegaskan potensi algoritma tersebut dalam membantu identifikasi risiko depresi postpartum secara lebih akurat dan efisien.

A. Evaluasi Kinerja Model

Berikut adalah hasil evaluasi model yang telah dilakukan menggunakan metrik evaluasi seperti precision, recall, F1-score, confusion matrix, dan akurasi keseluruhan. Evaluasi ini bertujuan untuk mengukur performa model dalam melakukan prediksi pada data uji. Hasil evaluasi disajikan dalam bentuk tabel sebagai berikut:

TABEL I.
EVALUASI MODEL

Kelas	Presisi	Recall	F1-Score	Support
0	0.91	0.95	0.87	205
1	0.87	0.79	0.82	95
Accuracy			0.90	301
Macro Avg	0.89	0.87	0.88	301
Weighted avg	0.90	0.90	0.90	301

Hasil evaluasi model menunjukkan bahwa untuk kelas 0, model memiliki precision sebesar 91%, recall sebesar 95%, dan F1-score sebesar 93%, dengan total 206 data pada kelas ini. Sementara untuk kelas 1, model memiliki precision sebesar 87%, recall sebesar 79%, dan F1-score sebesar 82%, dengan total 95 data. Rata-rata performa berdasarkan macro average menunjukkan precision, recall, dan F1-score masing-masing sebesar 89%, 87%, dan 87%, sedangkan berdasarkan weighted average, nilai precision, recall, dan F1-score masing-masing sebesar 90%, 90%, dan 90%. Secara keseluruhan, model memiliki akurasi sebesar 90%, yang menunjukkan bahwa model mampu melakukan prediksi dengan cukup baik, meskipun performa pada kelas 1 masih lebih rendah dibandingkan kelas 0. Akurasi tinggi sebesar 90% yang dicapai oleh algoritma *Random Forest* selaras dengan penelitian [19], yang menunjukkan kemampuan algoritma ini dalam menangani data berdimensi tinggi secara akurat, sementara penggunaan *Random Search* untuk optimasi hyperparameter terbukti meningkatkan efisiensi, sesuai dengan temuan [11], [12] yang membandingkan metode ini dengan *Grid Search*. Temuan ini juga menegaskan pentingnya penerapan model *machine learning* dalam mendukung analisis kesehatan mental, sebagaimana dijelaskan oleh [20], bahwa model prediktif berbasis *machine learning* memiliki potensi sebagai alat yang efektif untuk mendeteksi risiko dan mendukung intervensi dini pada masalah kesehatan mental, termasuk depresi postpartum.

Hasil penelitian ini menunjukkan bahwa algoritma Random Forest mampu memberikan performa yang sangat baik dalam memprediksi risiko depresi postpartum. Dengan nilai Area Under Curve (AUC) sebesar 0.96, model ini memiliki kemampuan yang sangat tinggi dalam membedakan individu dengan risiko tinggi dan rendah terhadap depresi postpartum. Selain itu, Average Precision (AP) sebesar 0.90 menegaskan keseimbangan yang baik antara precision dan recall, menunjukkan kemampuan model untuk mendeteksi kasus positif tanpa mengorbankan akurasi prediksi secara keseluruhan.

Temuan dalam penelitian ini sejalan dengan berbagai studi sebelumnya yang juga menggarisbawahi keunggulan algoritma Random Forest dalam analisis data kesehatan. Studi sebelumnya, seperti penelitian oleh Olisah et al., yang juga melaporkan bahwa Random Forest memberikan kinerja yang sangat baik dalam prediksi penyakit kronis, termasuk dalam situasi dengan variabel yang saling berinteraksi dan data berdimensi tinggi. Keandalan algoritma ini dalam menghasilkan prediksi yang akurat menjadikannya pilihan yang populer dalam pengembangan model prediktif di bidang Kesehatan [21]. Namun, hasil penelitian ini juga menunjukkan bahwa performa algoritma dapat dipengaruhi oleh distribusi data yang tidak seimbang. Hal ini didukung oleh temuan dalam studi Xu et al., di mana Random Forest mengalami penurunan kinerja ketika digunakan pada dataset dengan ketidakseimbangan kelas yang ekstrem. Tantangan ini menyoroti pentingnya langkah preprocessing data yang tepat, seperti penerapan teknik penyeimbangan data, untuk memastikan model tetap optimal dalam mendeteksi kasus minoritas. [22]. Dari perbandingan ini, dapat disimpulkan bahwa keberhasilan implementasi algoritma Random Forest sangat dipengaruhi oleh kualitas data yang digunakan. Teknik preprocessing, seperti resampling data, normalisasi, atau penerapan metode balancing seperti SMOTE (Synthetic Minority Oversampling Technique), memainkan peran penting dalam meningkatkan performa model. Dengan demikian, penelitian ini tidak hanya menggarisbawahi keunggulan Random Forest dalam konteks analisis kesehatan mental, tetapi juga memberikan panduan penting untuk memastikan model prediktif dapat diaplikasikan secara efektif dalam berbagai kondisi dataset.

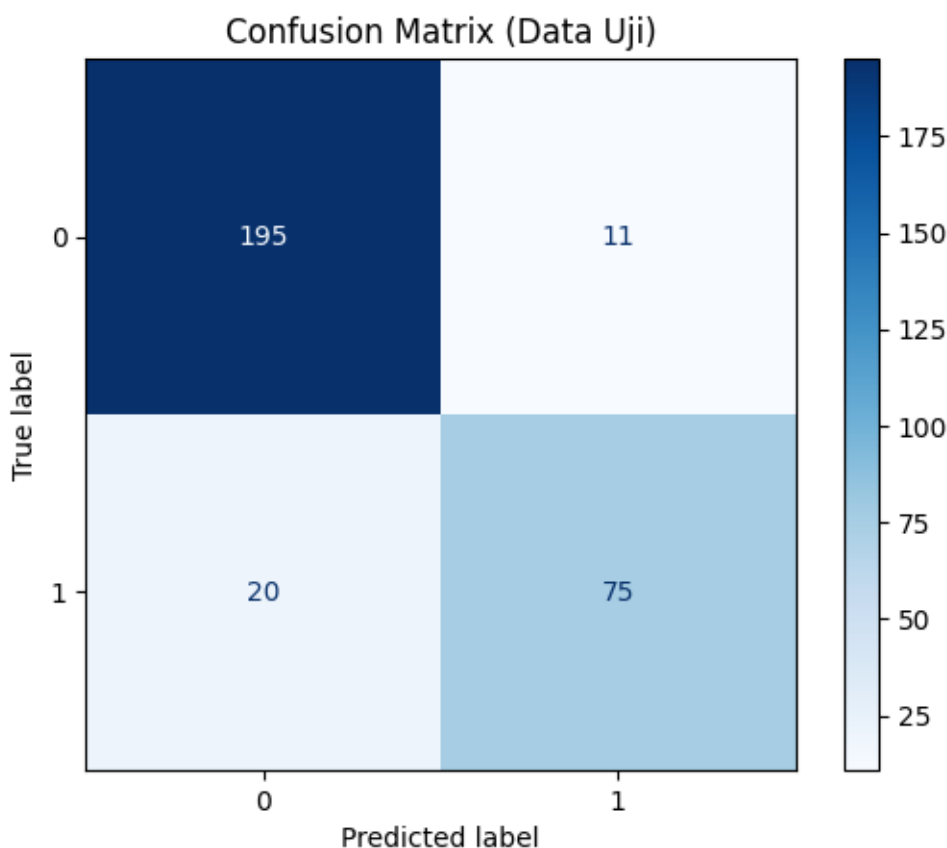
Meskipun hasil penelitian ini cukup memuaskan, terdapat beberapa kelemahan dan tantangan yang dihadapi selama proses penelitian. Salah satu tantangan utama adalah ketidakseimbangan data, di mana jumlah responden dengan risiko tinggi jauh lebih sedikit dibandingkan kelompok lainnya. Ketidakseimbangan ini secara signifikan memengaruhi performa model, terutama dalam mendeteksi kelas minoritas, sebagaimana terlihat dari nilai recall

sebesar 79%. Keterbatasan ini menunjukkan bahwa, meskipun hasil penelitian cukup baik, masih diperlukan upaya lebih lanjut untuk meningkatkan performa model, khususnya dalam mengenali kelas minoritas. Penelitian di masa depan dapat mempertimbangkan penggunaan dataset yang lebih seimbang, baik melalui pengumpulan data tambahan maupun penerapan teknik alternatif untuk menangani ketidakseimbangan data. Selain itu, validasi model pada dataset yang lebih beragam dan representatif diharapkan dapat meningkatkan kemampuan generalisasi hasil penelitian ini.

Penelitian lanjutan dapat mengeksplorasi penggunaan metode ensemble lain, seperti Gradient Boosting atau XGBoost, untuk meningkatkan performa model dalam menghadapi data tidak seimbang. Validasi model pada dataset yang lebih besar dan beragam perlu dilakukan untuk meningkatkan generalisasi hasil. Variabel tambahan, seperti faktor sosial-ekonomi atau riwayat kesehatan keluarga, dapat digunakan untuk membuat model prediktif yang lebih komprehensif. Selain itu, pengujian model di lingkungan klinis direkomendasikan untuk memastikan efektivitasnya dalam mendukung pengambilan keputusan tenaga kesehatan. Rekomendasi ini diharapkan dapat meningkatkan akurasi model dan manfaatnya dalam deteksi dini risiko postpartum.

B. Analisis Confussion Matrix

Berdasarkan confusion matrix pada gambar 2, performa model dalam memprediksi data uji dapat dijelaskan sebagai berikut: Dari 206 data aktual kelas 0, sebanyak 195 data berhasil diprediksi dengan benar sebagai kelas 0 (*True Positive*), sementara 11 data salah diprediksi sebagai kelas 1 (*False Negative*). Untuk kelas 1, dari total 95 data, sebanyak 75 data berhasil diprediksi dengan benar sebagai kelas 1 (*True Prositvie*), sedangkan 20 data salah diprediksi sebagai kelas 0 (*False Positive*). Hasil ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali kelas mayoritas (0), namun masih terdapat kesalahan yang cukup signifikan dalam mengenali kelas minoritas (1). Hal ini tercermin dari jumlah *False negative* dan *False positive* yang lebih tinggi pada kelas 1.



Gambar 2 Confussion Matrix

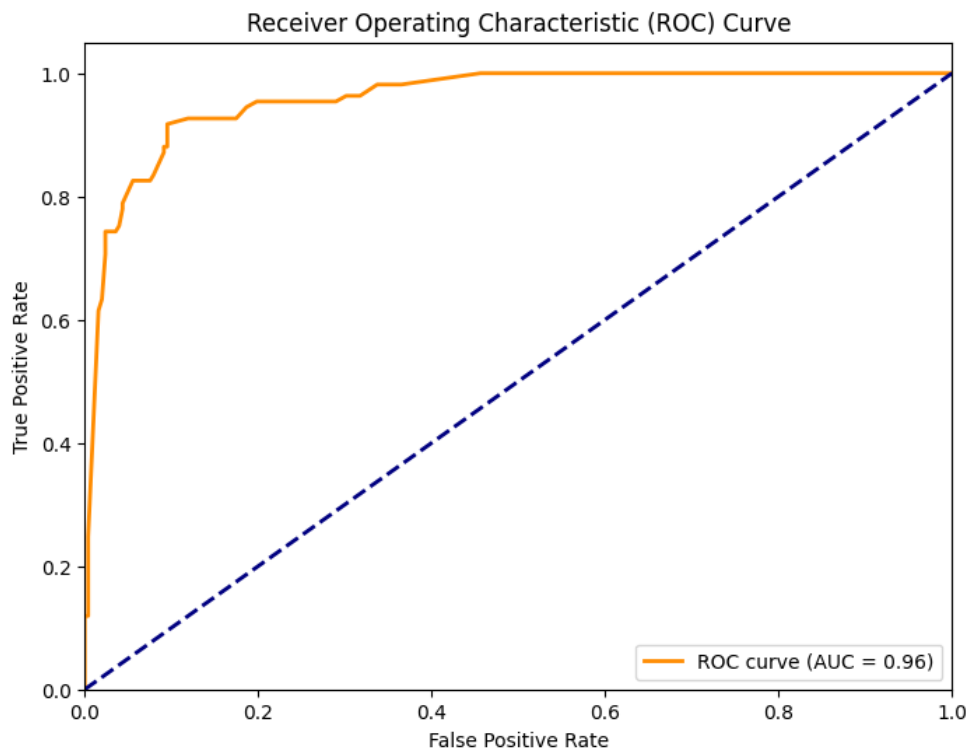
C. Analisis Grafik ROC

Kurva *Receiver Operating Characteristic* (ROC) yang ditampilkan pada gambar 3 menunjukkan kinerja model prediksi menggunakan algoritma Random Forest dalam mendeteksi risiko postpartum pada ibu primipara. ROC adalah representasi grafis dari kemampuan model dalam membedakan antara kelas positif dan negatif berdasarkan

berbagai nilai ambang (threshold). Pada sumbu horizontal ditampilkan False Positive Rate (FPR), sedangkan sumbu vertikal menunjukkan True Positive Rate (TPR).

Kurva ROC yang hampir mendekati sudut kiri atas grafik menunjukkan bahwa model memiliki kinerja yang sangat baik dalam klasifikasi. Hal ini didukung oleh nilai *Area Under Curve* (AUC) sebesar 0,96, yang mendekati nilai maksimum 1, menandakan kemampuan prediksi yang sangat tinggi. Dengan AUC sebesar ini, model dapat dikatakan memiliki akurasi yang sangat baik dalam membedakan antara individu dengan risiko tinggi dan rendah untuk mengalami depresi postpartum.

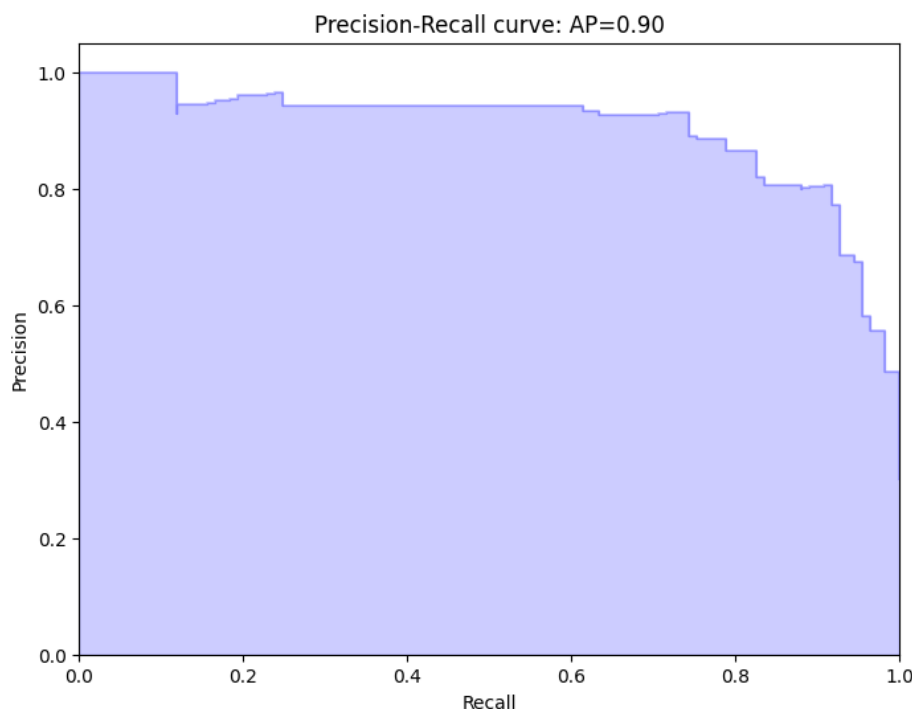
Hasil ini menunjukkan bahwa model Random Forest yang digunakan tidak hanya akurat, tetapi juga mampu meminimalkan kesalahan prediksi positif palsu dan negatif palsu. Performanya sangat menjanjikan untuk digunakan dalam deteksi dini dan intervensi risiko postpartum pada ibu primipara.



Gambar 3 Curva ROC

D. Analisis PR Curve

Kurva Precision-Recall yang ditampilkan pada gambar 4 menunjukkan kinerja model dalam mempertahankan keseimbangan antara precision dan recall pada berbagai ambang batas klasifikasi. Dengan area di bawah kurva (AP) sebesar 0,90, model ini menunjukkan kemampuan yang sangat baik dalam mendeteksi kasus positif postpartum depression, sambil meminimalkan prediksi positif palsu. Nilai AP yang tinggi mengindikasikan bahwa model ini tidak hanya mampu mengenali individu yang berisiko tinggi dengan tingkat akurasi yang baik, tetapi juga efektif dalam menjaga tingkat recall untuk memastikan tidak ada kasus yang terlewat. Kurva ini juga mencerminkan keandalan model dalam menangani potensi ketidakseimbangan jumlah data antara kelas positif dan negatif. Dengan demikian, model yang digunakan memberikan kontribusi signifikan dalam mendukung analisis prediktif postpartum depression, yang sangat penting untuk perencanaan intervensi yang tepat di bidang klinis.



Gambar 4 PR Curve

IV. KESIMPULAN

Kesimpulan dari penelitian diatas, model memiliki performa yang cukup baik dengan akurasi keseluruhan sebesar 90%. Model dapat mengenali data dari kelas mayoritas (0) dengan baik, terlihat dari nilai recall sebesar 95% dan jumlah prediksi benar yang tinggi. Namun, pada kelas minoritas (1), performa model masih perlu ditingkatkan, terutama dalam mengurangi jumlah kesalahan prediksi (false positive dan false negative), sebagaimana terlihat dari nilai recall yang lebih rendah sebesar 79%. Hal ini menunjukkan bahwa model memiliki potensi, tetapi perlu dilakukan optimasi lebih lanjut, seperti penyeimbangan data atau penyempurnaan parameter, untuk meningkatkan kemampuan dalam mengenali kelas minoritas secara lebih akurat. Penelitian ini memberikan kontribusi baru dalam pengembangan metode prediksi risiko postpartum dengan memanfaatkan algoritma *Random Forest* yang dioptimalkan menggunakan *Random Search*. Fokus pada faktor psikologis sebagai variabel utama memperkuat relevansi hasil dalam mendukung strategi intervensi dini untuk mencegah depresi postpartum pada ibu primipara.

DAFTAR PUSTAKA

- [1] L. C. Shuffrey, M. E. Thomason, and N. H. Brito, "Improving Perinatal Maternal Mental Health Starts with Addressing Structural Inequities," *JAMA Psychiatry*, vol. 79, no. 5, pp. 387–388, May 2022, doi: 10.1001/jamapsychiatry.2022.0097.
- [2] WHO, "Mental Health, Brain Health and Substance Use."
- [3] S. Molgora, E. Saita, M. Barbieri Carones, E. Ferrazzi, and F. Facchin, "Predictors of Postpartum Depression among Italian Women: A Longitudinal Study," *Int J Environ Res Public Health*, vol. 19, no. 3, Feb. 2022, doi: 10.3390/ijerph19031553.
- [4] R. Sun, M. Zhao, L. Ma, Y. Duan, and J. Wei, "High psychological stress levels related to delivery can increase the occurrence of postpartum mental disorders," *Front Psychiatry*, vol. 14, 2023, doi: 10.3389/fpsy.2023.1273647.
- [5] A. B. Zheutlin *et al.*, "Improving postpartum hemorrhage risk prediction using longitudinal electronic medical records," *Journal of the American Medical Informatics Association*, vol. 29, no. 2, pp. 296–305, Feb. 2022, doi: 10.1093/jamia/ocab161.
- [6] S. Holcroft, I. Karangwa, F. Little, J. Behoor, and O. Bazirete, "Predictive Modelling of Postpartum Haemorrhage Using Early Risk Factors: A Comparative Analysis of Statistical and Machine Learning Models," *Int J Environ Res Public Health*, vol. 21, no. 5, May 2024, doi: 10.3390/ijerph21050600.
- [7] V. Ignatenko, A. Surkov, and S. Koltcov, "Random forests with parametric entropybased information gains for classification and regression problems," *PeerJ Comput Sci*, vol. 10, 2024, doi: 10.7717/peerj-cs.1775.
- [8] S. Wang, Z. Zhang, S. Geng, and C. Pang, "Research on optimization of random forest algorithm based on spark," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 3721–3731, 2022, doi: 10.32604/cmc.2022.015378.
- [9] C. Xu, J. Wang, T. Zheng, Y. Cao, and F. Ye, "Prediction of prognosis and survival of patients with gastric cancer by a weighted improved random forest model: an application of machine learning in medicine," *Archives of Medical Science*, vol. 18, no. 5, pp. 1208–1220, 2022, doi: 10.5114/aoms/135594.
- [10] V. Ignatenko, A. Surkov, and S. Koltcov, "Random forests with parametric entropybased information gains for classification and regression problems," *PeerJ Comput Sci*, vol. 10, 2024, doi: 10.7717/peerj-cs.1775.
- [11] A. Stuke, P. Rinke, and M. Todorovic, "Efficient hyperparameter tuning for kernel ridge regression with Bayesian optimization," *Mach Learn Sci Technol*, vol. 2, no. 3, Sep. 2021, doi: 10.1088/2632-2153/abee59.
- [12] Y. A. Ali, E. M. Awwad, M. Al-Razgan, and A. Maarouf, "Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity," *Processes*, vol. 11, no. 2, Feb. 2023, doi: 10.3390/pr11020349.

- [13] K. Gao, Y. Wang, and L. Ma, "Belief Entropy Tree and Random Forest: Learning from Data with Continuous Attributes and Evidential Labels," *Entropy*, vol. 24, no. 5, May 2022, doi: 10.3390/e24050605.
- [14] A. Dmitry Devyatkin and G. Oleg Grigoriev, "Random Kernel Forests," *IEEE Access*, vol. 10, pp. 77962–77979, 2022, doi: 10.1109/ACCESS.2022.3193385.
- [15] B. F. Hutchens and J. Kearney, "Risk Factors for Postpartum Depression: An Umbrella Review," Jan. 01, 2020, *John Wiley and Sons Inc.* doi: 10.1111/jmwh.13067.
- [16] Y. A. Ali, E. M. Awwad, M. Al-Razgan, and A. Maarouf, "Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity," *Processes*, vol. 11, no. 2, Feb. 2023, doi: 10.3390/pr11020349.
- [17] C. Valsecchi, V. Consonni, R. Todeschini, M. E. Orlandi, F. Gosetti, and D. Ballabio, "Parsimonious optimization of multitask neural network hyperparameters," *Molecules*, vol. 26, no. 23, Dec. 2021, doi: 10.3390/molecules26237254.
- [18] Z. Lin, A. Matta, S. Du, and E. Sahin, "A Partition-Based Random Search Method for Multimodal Optimization," *Mathematics*, vol. 11, no. 1, Jan. 2023, doi: 10.3390/math11010017.
- [19] C. Xu, J. Wang, T. Zheng, Y. Cao, and F. Ye, "Prediction of prognosis and survival of patients with gastric cancer by a weighted improved random forest model: an application of machine learning in medicine," *Archives of Medical Science*, vol. 18, no. 5, pp. 1208–1220, 2022, doi: 10.5114/aoms/135594.
- [20] S. Holcroft, I. Karangwa, F. Little, J. Behoor, and O. Bazirete, "Predictive Modelling of Postpartum Haemorrhage Using Early Risk Factors: A Comparative Analysis of Statistical and Machine Learning Models," *Int J Environ Res Public Health*, vol. 21, no. 5, May 2024, doi: 10.3390/ijerph21050600.
- [21] H. Hairani, A. Anggrawan, and D. Priyanto, "INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION journal homepage : www.joiv.org/index.php/joiv INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link." [Online]. Available: www.joiv.org/index.php/joiv
- [22] Z. Xu, D. Shen, T. Nie, and Y. Kou, "A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data," *J Biomed Inform*, vol. 107, Jul. 2020, doi: 10.1016/j.jbi.2020.103465.