

# Klasifikasi Kesehatan Ibu Hamil Menggunakan Metode *Random Forest* Dengan Optimasi Algoritma Genetika

Wildan Muhtar Alfian Mas<sup>1</sup>, Rissa Nurfitriana Handayani<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Universitas Adhirajasa Reswara Sanjaya

e-mail: <sup>1</sup>wildan.muhtaralfian@gmail.com, <sup>2</sup>rissa@ars.ac.id

## Abstrak

Kesehatan merupakan hak dasar (Undang-Undang Dasar 1945, Pasal 28H ayat 1 dan Undang-Undang No. 23 Tahun 1992) dan juga merupakan investasi. Oleh karena itu, setiap individu dan seluruh bangsa harus berupaya, berjuang, dan meningkatkan kesehatan mereka agar semua orang dapat menikmati kehidupan yang sehat. Pada akhirnya, kesehatan masyarakat harus mencapai tingkat optimal. Meskipun kehamilan adalah proses fisiologis, perawatan kebidanan diperlukan bagi ibu hamil untuk memantau kondisi kesehatan dan perkembangan kehamilan. Pemeriksaan kesehatan secara teratur pada ibu hamil dan janin membantu mereka menjalani kehamilan yang optimal. Wanita mungkin mengalami beberapa ketidaknyamanan fisik selama hamil, yang kadang-kadang mengganggu aktivitas mereka. Dengan kemajuan teknologi yang pesat saat ini, teknologi informasi telah berkembang di berbagai bidang, termasuk bidang kesehatan. Dalam bidang kesehatan, data mining merupakan salah satu bidang yang dapat dimanfaatkan. Data mining adalah serangkaian tindakan untuk menemukan hubungan signifikan melalui pola dan tren dalam dataset yang sangat besar dengan menggunakan metode klasifikasi. Klasifikasi banyak digunakan untuk memprediksi kelas pada label tertentu, yaitu dengan mengelompokkan. Karena tingkat akurasi algoritma *random forest* yang tinggi, penulis akan menggunakan algoritma *random forest* dalam penelitian ini.

**Kata kunci**—Algoritma Genetika, Ibu Hamil, Klasifikasi, *Random Forest*

## Abstract

*The right to health is a foundational entitlement (stipulated in the 1945 Constitution, Article 28H paragraph 1, and Law No. 23 of 1992), and it is also regarded as an investment. Consequently, both individuals and the entire nation should exert efforts, advocate, and enhance their well-being to ensure that everyone can lead a healthy life. The ultimate goal is for public health to attain an optimum level. While pregnancy is a natural physiological process, expectant mothers necessitate obstetric care for monitoring their health status and the progress of the pregnancy. Regular health check-ups for pregnant women and their fetuses contribute to the realization of an ideal pregnancy. Women may encounter certain physical discomforts during pregnancy, potentially disrupting their daily activities. In the context of today's rapid technological advancements, information technology has progressed across various fields, including healthcare. Data mining stands out as a valuable tool in health-related domains. It involves a series of actions aimed at uncovering meaningful relationships by analyzing patterns and trends within extensive datasets using classification methods. Classification is commonly employed to forecast categories associated with specific labels by grouping them. Given the commendable accuracy of the random forest algorithm, the researcher opts to utilize it in this study.*

**Keywords**—Genetic Algorithm, Pregnant Womes, Classification, *Random Forest*

---

### Corresponding Author:

Rissa Nurfitriana Handayani,

Email: rissa@ars.ac.id

---

## 1. PENDAHULUAN

Kesehatan merupakan hak mendasar (berdasarkan Undang-Undang Dasar 1945, Pasal 28H ayat 1, dan Undang-Undang No 23 Tahun 1992) dan juga dianggap sebagai bentuk investasi. Oleh karena itu, setiap individu dan seluruh bangsa harus berusaha, mengadvokasi, dan meningkatkan kesehatan mereka agar semua orang dapat menikmati kehidupan yang sehat. Pada akhirnya, kesehatan masyarakat seharusnya mencapai tingkat optimal. Ini merupakan tanggung jawab bersama antara pemerintah dan masyarakat, termasuk sektor swasta.[1]. Meskipun kehamilan adalah suatu proses fisiologis, perlu adanya perawatan kebidanan untuk memantau kondisi kesehatan ibu hamil dan perkembangan kehamilannya. Pemeriksaan kesehatan rutin pada ibu hamil dan janin membantu ibu menjalani kehamilan dengan kondisi yang optimal. Wanita dapat mengalami beberapa ketidaknyamanan fisik selama kehamilan, yang terkadang dapat mengganggu aktivitas sehari-hari mereka [2].

Perawatan kehamilan, sebagai bagian dari sistem kesehatan, sangat terkait dengan fungsi reproduksi wanita, sehingga tindakan yang dilakukan ibu sangat berpengaruh. Perilaku dipengaruhi oleh pengetahuan atau kognitif. Ibu hamil yang mengetahui tentang gangguan dan penyulit kehamilan dapat bertindak dengan cara yang akan menjaga, mencegah, dan menghindari komplikasi kehamilan.[3]. Dengan kemajuan teknologi yang pesat saat ini, teknologi informasi telah berkembang di berbagai sektor, termasuk di bidang kesehatan. Data mining merupakan proses ekstraksi pengetahuan dari sejumlah besar data yang tersimpan dalam komputer, di mana proses ini sangat bergantung pada teknologi informasi [4].

Dalam ranah kesehatan, data mining menjadi salah satu bidang yang dapat dimanfaatkan. *Data mining* merupakan serangkaian tindakan atau langkah untuk menemukan hubungan yang signifikan melalui pola dan kecenderungan dalam himpunan data yang besar yang disimpan dengan menggunakan metode tertentu. Terdapat berbagai teknik, metode, atau algoritma yang digunakan dalam proses *data mining*. Dalam literatur ilmiah, data mining dikenal sebagai *Knowledge Discovery in Database (KDD)*. Hasilnya dapat diterapkan untuk meningkatkan pengambilan keputusan di masa mendatang. Proses *data mining* umumnya mencakup deskripsi, prediksi, estimasi, klasifikasi, pengelompokan, dan asosiasi. Prediksi dan klasifikasi seringkali digunakan dalam analisis data untuk menggambarkan kelas data atau membuat prediksi untuk data di masa depan.[5]. Metode klasifikasi merupakan salah satu dari berbagai metode yang digunakan dalam *data mining*. Klasifikasi adalah suatu metode pembelajaran data yang digunakan untuk meramalkan nilai dari sekumpulan atribut. Pendekatan ini sering digunakan untuk meramalkan kelas pada label tertentu, yaitu dengan mengelompokkan data (membangun model) berdasarkan serangkaian petunjuk dan nilai (label kelas) dalam mengkategorikan atribut khusus [6].

Dalam proses *data mining*, ini dapat digunakan untuk menilai kecocokan algoritma lainnya. [7]. Keunggulan dari teknik klasifikasi terletak pada metode pembelajaran terawasi, yang berupaya menemukan hubungan antara atribut masukan dan atribut target. Tujuan dari metode klasifikasi adalah meningkatkan kepercayaan hasil yang diperoleh dari data[8]. Klasifikasi adalah proses melatih dan mempelajari fungsi target dengan menetapkan setiap himpunan atribut (fitur) ke salah satu dari label kelas yang tersedia. Ini dilakukan dengan membuat model berdasarkan data latihan yang ada dan selanjutnya menggunakan model tersebut untuk mengklasifikasikan data yang baru [9]

Prinsip kerja teknik klasifikasi melibatkan penempatan suatu catatan data baru ke dalam salah satu dari beberapa kategori (kelas) yang telah ditentukan sebelumnya [10]. Untuk metode klasifikasi ini, Algoritma *Random Forest* digunakan. Studi sebelumnya tentang penggunaan Algoritma *Random Forest* untuk Klasifikasi Kualitas Air Sumur[11] Menyiratkan bahwa algoritma *Random Forest* mampu melakukan klasifikasi kualitas air dengan efektif.

*Random forest* adalah salah satu metode yang digunakan untuk klasifikasi dan regresi. Metode ini merupakan teknik pembelajaran *ensemble* yang menggunakan pohon keputusan sebagai klasifikasi dasar yang dibangun dan digabungkan. Terdapat tiga komponen kunci dalam

metode *Random Forest*, yakni pengambilan sampel *bootstrap* untuk membuat pohon prediksi, setiap pohon keputusan memprediksi dengan variabel prediktor acak, dan *Random Forest* memprediksi dengan menggabungkan hasil dari setiap pohon keputusan melalui mayoritas suara untuk klasifikasi atau rata-rata untuk regresi. [12]. Salah satu *assembly learning* yang dibangun dari pohon keputusan adalah *Random Forest*. Salah satu keuntungan menggunakan pendekatan pembelajaran *ensemble* adalah dapat digunakan untuk kasus klasifikasi dan regresi, mampu memperoleh tingkat akurasi yang tinggi, dan cocok untuk analisis ukuran dataset yang besar dengan banyak dimensi. Selain itu, *ensemble learning* seperti *Random Forest* juga cocok untuk menangani data yang tidak seimbang [13].

Algoritma genetika (GA) adalah suatu algoritma pencarian yang bersifat stokastik, paralel, dan *heuristik*, terinspirasi oleh prinsip dasar seleksi alam yang dikenalkan oleh Charles Darwin [7]. John Holland (1975) dari Universitas Michigan memperkenalkan algoritma genetika (AG). Holland mengatakan bahwa terminologi genetika dapat digunakan untuk setiap masalah adaptasi, baik buatan maupun alami [14].

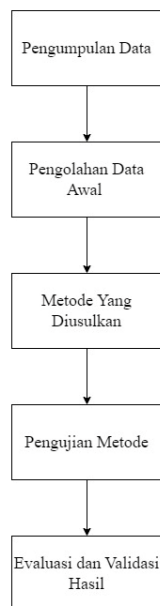
Untuk mengatasi masalah optimalisasi, baik yang terbatas maupun tidak terbatas, algoritma genetika menggunakan seleksi alam, suatu mekanisme yang terinspirasi dari evolusi biologis. Algoritma genetika secara berulang mengubah populasi solusi individu. Populasi saat ini digunakan oleh algoritma genetika untuk secara acak memilih orang tua pada setiap tahap guna menghasilkan keturunan generasi berikutnya. Dari satu generasi ke generasi berikutnya, populasi "berkembang" menuju solusi yang ideal [15].

## 2. METODE PENELITIAN

Penelitian klasifikasi ini memerlukan metode eksperimen, yang sangat sesuai digunakan. karena tujuan penelitian adalah untuk memahami hasil klasifikasi yang dioptimasi dengan algoritma *random forest*.

### 2.1 Perancangan Penelitian

Menurut Dawson Dalam [16] Empat metode penelitian yang sering digunakan melibatkan tindakan penelitian, eksperimen, studi kasus, dan survei. Dalam penelitian ini, akan digunakan metode eksperimen. digunakan untuk mengoptimalkan fitur dengan algoritma genetika fitur dan algoritma *Random Forest* untuk klasifikasi kesehatan ibu hamil. Penelitian ini melakukan beberapa langkah. Hasil penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

## 2.2 Identifikasi Masalah

Dalam upaya melakukan klasifikasi kesehatan ibu hamil dibutuhkan suatu model yang tepat untuk melakukan klasifikasi kesehatan ibu hamil.

## 2.3 PreProcessing Data

Pada tahap preprocessing data dilakukan pengecekan dataset berupa *missing value*, *filter example* dan normalisasi. Operator filter contoh digunakan untuk menyaring atau menghilangkan data yang tidak memiliki nilai atau kosong. Dengan menggunakan metode transformasi Z, dilakukan normalisasi agar variabel atribut memiliki rentang nilai yang seragam, yakni antara 0 dan 1.

## 2.4 Validasi Data

Pada fase validasi data, data penelitian akan dipisahkan menjadi data latih dan data uji. Pembagian data akan dilakukan menggunakan metode *cross-validation* dan *split-validation*. *Cross-validation* mengevaluasi performa terbaik dari model yang akan diuji, sedangkan *split-validation* mengevaluasi model tertentu.

## 2.5 Komparasi Algoritma

Pada tahap perbandingan algoritma, pengujian dilakukan dengan menggunakan berbagai algoritma. Dalam penelitian ini, digunakan empat algoritma, yaitu: *K-Nearest Neighbor (K-NN)*, *Naïve Bayes*, *Decision Tree*, dan *Random Forest*.

## 2.6 Random Forest

Pada tahap ini, algoritma *Random Forest* adalah model terbaik untuk klasifikasi kesehatan ibu hamil. Model ini dipilih berdasarkan tingkat akurasi terbaik dari empat algoritma yang digunakan untuk klasifikasi ibu hamil. Uji *split validation* akan digunakan untuk menguji algoritma tersebut. Ini akan dilakukan dengan menggunakan perbandingan split 0,5 hingga 0,9 untuk mendapatkan nilai rata-rata..

## 2.7 Algoritma Genetika

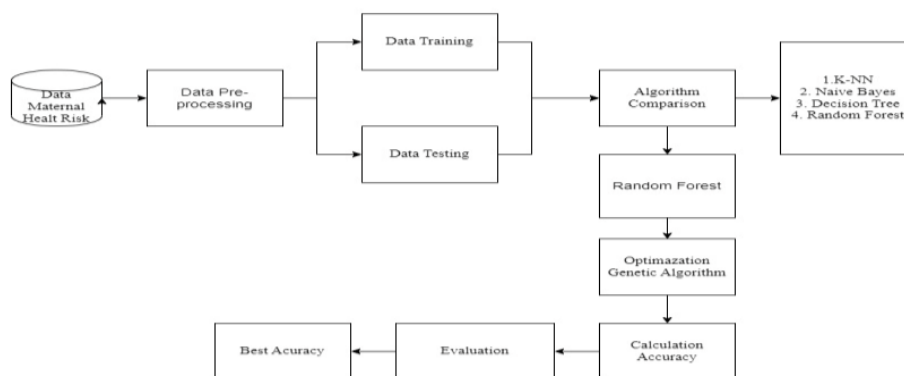
Algoritma genetika merupakan suatu algoritma pencarian optimasi yang bertujuan untuk memaksimalkan atau meminimalkan suatu fungsi yang diberikan. Pada awalnya, algoritma ini memanfaatkan prosedur heuristik. tetapi pengalaman telah menunjukkan bahwa algoritma ini dapat menyelesaikan banyak masalah [17].

## 2.8 Evaluasi

Pada tahap evaluasi ini, nilai akurasi terbaik untuk klasifikasi kesehatan ibu hamil akan diketahui. Untuk menilai apakah ada perbedaan antara sebelum dan setelah optimalisasi klasifikasi kesehatan ibu hamil, peneliti mengevaluasi perbandingan akurasi dengan menggunakan rasio split dari 0,5 hingga 0,9 pada algoritma *Random Forest* dan algoritma *Random Forest* yang dioptimalkan menggunakan Algoritma Genetika (GA).

## 2.9 Metode Yang Diusulkan

Hasil evaluasi klasifikasi bahaya kesehatan ibu yang dilakukan dengan model yang diusulkan menunjukkan bahwa hasil nilai terbaik dicapai melalui optimasi fitur menggunakan algoritma genetik (GA). Ini dapat mempengaruhi hasil klasifikasi terbaik yang dihasilkan oleh algoritma *Random Forest* dalam mengklasifikasikan *maternal health risk*.



Gambar 2. Metode Yang Diusulkan

### 3. HASIL DAN PEMBAHASAN

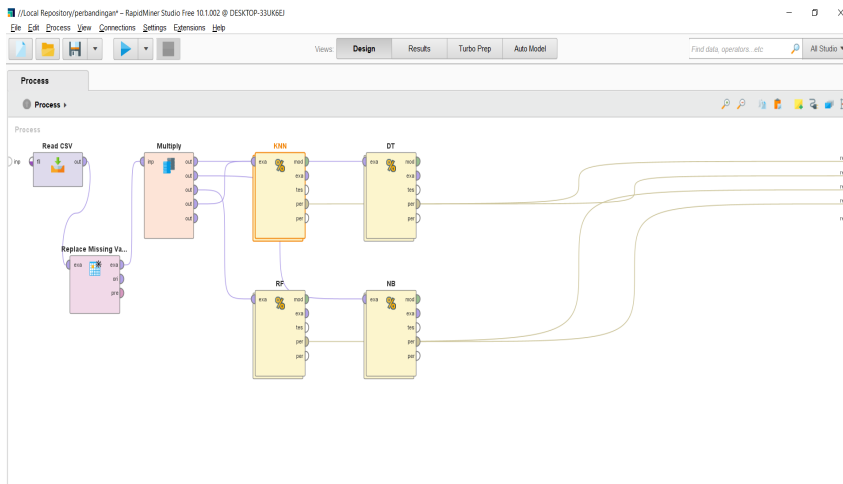
#### 3.1 Pembahasan

Dalam fase ini, *output* dari percobaan untuk mengklasifikasikan dataset Kesehatan Ibu Hamil ditampilkan. Identifikasi masalah adalah langkah pertama yang dilakukan. Diketahui bahwa klasifikasi kesehatan ibu hamil membutuhkan metode atau algoritma yang memiliki model terbaik. Berdasarkan pengetahuan ini, penelitian ini membahas klasifikasi risiko kesehatan ibu hamil. Penelitian ini menggunakan dataset penelitian *Maternal Health Risk*, yang dapat diakses melalui situs Kaggle.com. Dataset penelitian ini terdiri dari 1014 daftar data dengan 7 atribut. Ini ditampilkan pada tabel 1.

Tabel 1. Daftar Data Atribut

No.	Data Input	Rincian
1	Age	Usia
2	SystolicBP	Tekanan darah tinggi
3	DiastolicBP	Tekanan darah rendah
4	BS	Kadar gula darah
5	BodyTemp	Temperature suhu badan
6	HeartRate	Denyut jantung
7	Risk Level	Memberikan prediksi tingkat resiko selama kehamilan

Adapun alur proses penelitian yang dilakukan di Rapidminer yang terdapat di Gambar 3.



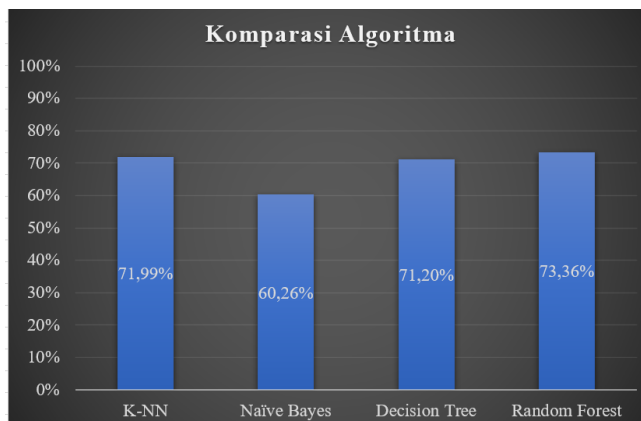
Gambar 3. Komparasi Algoritma

Setelah menyelesaikan proses preprocessing data, langkah selanjutnya adalah melakukan perbandingan algoritma. Untuk membandingkan keempat algoritma yang diuji dalam penelitian ini, dilakukan perbandingan algoritma. Algoritma yang digunakan meliputi Random Forest, Decision Tree, K-Nearest Neighbor (K-NN), dan Naive Bayes. Untuk menilai kinerja keempat algoritma tersebut, digunakan metode Cross Validation, yang menghasilkan nilai presisi, akurasi, dan recall.. Nilai ketepatan yang dihasilkan oleh masing-masing algoritma dapat dilihat di sini tabel 2.

Tabel 2. Komparasi Algoritma

<i>Algorithm</i>	<i>Validation</i>	<i>Accuracy</i>
<i>Naïve Bayes (NB)</i>	<i>Cross Validation</i>	60.26%
<i>K-Nearest Neighbor (K-NN)</i>	<i>Cross Validation</i>	71.99%
<i>Decision Tree (DT)</i>	<i>Cross Validation</i>	71.20%
<i>Random Forest (RF)</i>	<i>Cross Validation</i>	73.36%

Berdasarkan perbandingan algoritma, Dapat diperhatikan bahwa algoritma Random Forest menunjukkan nilai akurasi tertinggi sebesar 73,36%, mengungguli algoritma lainnya. Informasi lebih lanjut dapat ditemukan dalam Gambar 4.



Gambar 4 Komparasi Algoritma

Selanjutnya, *confusion matrix* adalah metode yang digunakan untuk mengukur akurasi Data Mining dalam bentuk tabel. Tabel *confusion matrix* yang dihasilkan dari klasifikasi algoritma *Random Forest (RF)* dapat ditemukan pada Tabel 3.

Tabel 3. *Confusion Matrix Random Forest*

	<i>True High Risk</i>	<i>True Low Risk</i>	<i>True Mid Risk</i>	<i>Class Precession</i>
<i>Pred.High Risk</i>	233	5	21	89,96%
<i>Pred.Low Risk</i>	16	370	174	66,07%
<i>Pred. Mid Risk</i>	23	31	141	72,31%
<i>Class Recall</i>	85,56%	91,13%	41,96%	

Berikut adalah hitungan manual dari hasil *Accuracy*, *Precision*, dan *Recall*:

### 1. Akurasi

Dari hasil pengujian, diperoleh nilai akurasi sebesar 73,36%. Berikut adalah perhitungan akurasi secara manual.

$$\begin{aligned}
 \text{Akurasi} &= \frac{233+370+141}{370+233+174+141+31+23+21+16+5} * 100\% \\
 &= \frac{744}{1014} * 100\% \\
 &= 0,7336 * 100\% \\
 &= 73,36\%
 \end{aligned}$$

### 2. Precision

Dari hasil pengujian, nilai presisi yang didapatkan adalah sebesar 89,96% untuk kelas High Risk. 66,07% Low Risk, dan 72,31% Mid Risk Berikut perhitungan presisinya.

$$\begin{aligned}
 \text{High Risk} &= \frac{233}{233+5+21} * 100\% \\
 &= \frac{233}{259} * 100\% \\
 &= 0,8996 * 100\% \\
 &= 89,96\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Low Risk} &= \frac{370}{370+16+174} * 100\% \\
 &= \frac{370}{560} * 100\% \\
 &= 0,6607 * 100\% \\
 &= 66,07\%
 \end{aligned}$$

$$\begin{aligned} \text{Mid Risk} &= \frac{141}{141+23+31} * 100\% \\ &= \frac{141}{195} * 100\% \\ &= 0,72,31 * 100\% \\ &= 72,31\% \end{aligned}$$

### 3. Recall

Dari hasil pengujian, nilai recall yang didapatkan adalah sebesar 85,66% untuk kelas *High Risk*., 91,13% *Low Risk*, dan 41,56% *Mid Risk* Berikut perhitungan recallnya.

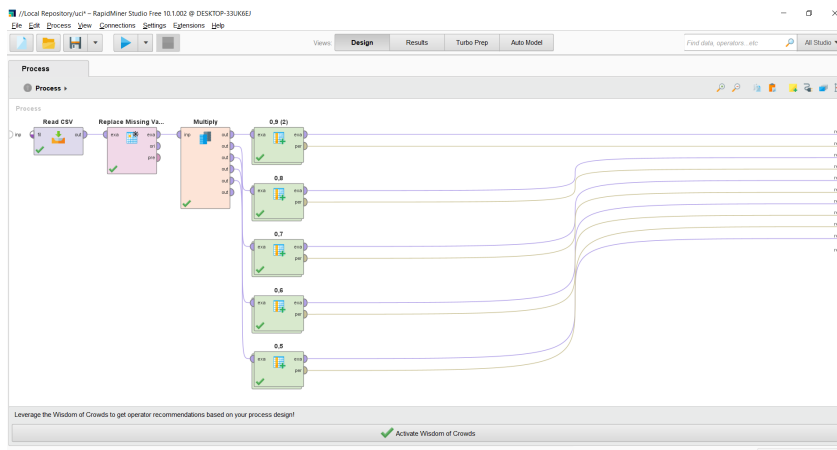
$$\begin{aligned} \text{High Risk} &= \frac{233}{233+16+23} * 100\% \\ &= \frac{233}{272} * 100\% \\ &= 0,8566 * 100\% \\ &= 85,66\% \end{aligned}$$

$$\begin{aligned} \text{Low Risk} &= \frac{370}{370+5+31} * 100\% \\ &= \frac{370}{406} * 100\% \\ &= 0,9113 * 100\% \\ &= 91,13\% \end{aligned}$$

$$\begin{aligned} \text{Mid Risk} &= \frac{141}{141+21+174} * 100\% \\ &= \frac{141}{336} * 100\% \\ &= 0,419 * 100\% \\ &= 41,96\% \end{aligned}$$

Berikut adalah urutan langkah dalam validasi data menggunakan *Split Ratio* mulai dari 0,5 hingga 0,9 pada algoritma *Random Forest* dengan memanfaatkan fitur optimasi algoritma genetika, yang dapat dilihat pada gambar 4..





Gambar 5. Proses Penelitian Optimasi

Berdasarkan penelitian yang dilakukan, Berikut ini merupakan nilai akurasi dari fitur optimasi, yang terdokumentasi dalam Tabel 4.

Tabel 4. Split Ratio 0,5 - 0.9 RF Setelah Di Optimasi

Algoritma	Validasi	Ratio	Akurasi
Random Forest	Split Validation	0,5	77.71%
Random Forest	Split Validation	0.6	78.02%
Random Forest	Split Validation	0,7	79.34%
Random Forest	Split Validation	0,7	81.68%
Random Forest	Split Validation	0,9	85.29%
<b>Rata rata</b>			<b>80,40%</b>

Dari Tabel 4, dapat diketahui bahwa algoritma *Random Forest* berbasis Optimasi Algoritma Genetika dengan split ratio 0,5 hingga 0,9 memiliki nilai rata-rata sebesar 80,40%. Validasi dengan split ratio 0,9 menghasilkan hasil akhir yang tinggi, yakni 85,29% dibandingkan dengan split ratio lainnya. Berikutnya, tabel Matriks Keberbingungan setelah dioptimalkan menggunakan fitur algoritma genetika dapat dilihat pada Tabel 5.

Tabel 5 *Confusion Matrix* Setelah Di Optimasi

	<i>True High Risk</i>	<i>True Low Risk</i>	<i>True Mid Risk</i>	<i>Class Precession</i>
<i>Pred. High Risk</i>	26	0	1	96,30%
<i>pred. Low Risk</i>	1	40	12	75,47%
<i>Pred. Mid Risk</i>	0	1	21	95,45%
<i>Class Recall</i>	96,30%	97,56%	61,76%	

Berikut adalah hitungan manual dari hasil *Accuracy*, *Precision*, dan *Recall*:

1. Akurasi

Dari hasil pengujian, diperoleh nilai akurasi sebesar 85,29%. Berikut adalah perhitungan akurasi secara manual.

$$\begin{aligned} Accuracy &= \frac{26+40+21}{26+0+1+1+40+12+0+1+21} * 100\% \\ &= \frac{87}{102} * 100\% \\ &= 0,85,29 * 100\% \\ &= 85,29\% \end{aligned}$$

## 2. *Precesion*

Dari hasil pengujian, nilai presisi yang didapatkan adalah sebesar 96,30% untuk kelas *High Risk*, 75,47% *Low Risk*, dan 95,45% *Mid Risk* Berikut perhitungan presisinya.

$$\begin{aligned} High Risk &= \frac{26}{26+0+1} * 100\% \\ &= \frac{26}{27} * 100\% \\ &= 0,9630 * 100\% \\ &= 96,30\% \end{aligned}$$

$$\begin{aligned} Low Risk &= \frac{40}{40+1+12} * 100\% \\ &= \frac{40}{53} * 100\% \\ &= 0,7547 * 100\% \\ &= 75,47\% \end{aligned}$$

$$\begin{aligned} Mid Risk &= \frac{21}{21+0+1} * 100\% \\ &= \frac{21}{22} * 100\% \\ &= 0,9545 * 100\% \\ &= 95,45\% \end{aligned}$$

## 3. *Recall*

Dari hasil pengujian, nilai recall yang didapatkan adalah sebesar 96,30% untuk kelas *High Risk*, 97,56% *Low Risk*, dan 61,76% *Mid Risk* Berikut perhitungan *recall*nya.

$$\begin{aligned} High Risk &= \frac{26}{26+1+0} * 100\% \\ &= \frac{26}{27} * 100\% \\ &= 0,9630 * 100\% \\ &= 96,30\% \end{aligned}$$

$$\begin{aligned} \text{Low Risk} &= \frac{40}{40+1+0} * 100\% \\ &= \frac{40}{41} * 100\% \\ &= 0,9756 * 100\% \\ &= 97,56\% \end{aligned}$$

$$\begin{aligned} \text{Mid Risk} &= \frac{21}{21+1+12} * 100\% \\ &= \frac{21}{34} * 100\% \\ &= 0,6176 * 100\% \\ &= 61,76\% \end{aligned}$$

### 3.2 Hasil

Berdasarkan pengujian yang telah diperoleh pada dataset *maternal healt risk*, dapat diketahui bahwa algoritma *Random Forest* berbasis optimasi algoritma genetika memiliki akurasi yang baik yaitu 85,29%

## 4. KESIMPULAN

Penelitian ini dilakukan pemodelan algoritma *Random Forest* dan di optimasi dengan *Feature* Algoritma Genetika dengan menggunakan dataset *Maternal Healt Risk* yang didapatkan dari website Kaggle. Validasi data pada Algoritma Random Forest menunjukkan tingkat kinerja terbaik dibandingkan dengan empat algoritma lainnya, seperti Naïve Bayes, K-Nearest Neighbor (K-NN), dan Decision Tree. Algoritma Random Forest menghasilkan tingkat akurasi sebesar 73,36%. Untuk menguji algoritma Random Forest menggunakan operator Split Validation dengan split ratio 0,5 hingga 0,9. Berdasarkan pengujian yang telah dilakukan, nilai rata-rata yang diperoleh adalah 74,40% untuk akurasi. Validasi data dengan split ratio 0,9 mencapai tingkat akurasi sebesar 78,43%.

Untuk meningkatkan kinerja akurasi algoritma Random Forest, dilakukan optimasi menggunakan algoritma genetika dengan validasi data, mencapai akurasi sebesar 85,29%. Pengujian dilakukan pada algoritma Random Forest menggunakan operator Split Validation dengan split ratio 0,5 hingga 0,9. Berdasarkan pengujian yang dilakukan, nilai rata-rata yang dihasilkan adalah 80,40% untuk akurasi, dengan 85,29%. Pengujian algoritma dengan split ratio 0,9 mencapai tingkat akurasi tertinggi, yakni sebesar 85,29% untuk akurasi.

## UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terima kasih kepada Ibu Rissa Nurfitriana Handayani yang telah memberi dukungan bimbingan terhadap penelitian ini

## DAFTAR PUSTAKA

- [1] Nur Hafifah and Zaenal Abidin, "Peran Posyandu dalam Meningkatkan Kualitas Kesehatan Ibu dan Anak di Desa Sukawening, Kabupaten Bogor," *Jurnal Pusat Inovasi Masyarakat*, vol. 2, no. 5, pp. 893–900, Jul. 2020.
- [2] Islami and Titik Ariyanti, "Prenatal Yoga Dan Kondisi Kesehatan Ibu Hamil," *Jurnal Ilmu Keperawatan dan Kebidanan*, vol. 10, no. 1, pp. 49–56, 2019.

- [3] Kadek Dwitya Widi Adnyani, "Prenatal Yoga Untuk Kondisi Kesehatan Ibu Hamil," *JURNAL YOGA DAN KESEHATAN*, vol. 4, no. 1, pp. 35–48, Mar. 2021.
- [4] Rissa Nurfitriana Handayani and Djajasukma Tjahjadi, "Optimasi Support Vector Machine Menggunakan Particle Swarm Optimazion Untuk Klasifikasi Multiclass Dataset," *JURNAL RESPONSIF*, vol. 5, no. 1, pp. 120–126, Feb. 2023.
- [5] Sanni Ucha Putri, Eka Irawan, and Fitri Rizky, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5," *Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 2, no. 1, pp. 39–46, Jan. 2021.
- [6] Aldi Tangkelayuk, "Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 2, pp. 1109–1119, Jun. 2022.
- [7] Rizki Tri Prasetio, "Seleksi Fitur Dan Optimasi Parameter K-Nn Berbasis Algoritma Genetika Pada Dataset Medis," *JURNAL RESPONSIF*, vol. 2, no. 2, pp. 213–221, Aug. 2020.
- [8] Senna Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Faktor Exacta*, vol. 11, no. 3, pp. 266–274, 2018.
- [9] Dito Putro Utomo and Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 2, pp. 437–444, Apr. 2020.
- [10] Dinda Nabila Batubara, Agus Perdana Windarto, and Eka Irawan, "Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi Decision Tree dan Support Vector Machine," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 1, pp. 102–108, Feb. 2022.
- [11] Muhamad Malik Mutoffar, Muchammad Naseer, and Ariansyah Fadillah, "KLASIFIKASI KUALITAS AIR SUMUR MENGGUNAKAN ALGORITMA RANDOM FOREST," *Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika*, vol. 4, no. 2, pp. 138–146, Dec. 2022.
- [12] Aji Primajaya and Betha Nurina Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, vol. 1, no. 1, pp. 27–31, Mar. 2018.
- [13] Faqih Hamami and Iqbal Ahmad Dahlan, "Klasifikasi Cuaca Provinsi Dki Jakarta Menggunakan Algoritma Random Forest Dengan Teknik Oversampling," *Jurnal TEKNOINFO*, vol. 16, no. 1, pp. 87–92, 2022.
- [14] Duwi Cahya Putri Buani, "Prediksi Penyakit Hepatitis Menggunakan Algoritma Naïve Bayes Dengan Seleksi Fitur Algoritma Genetika," *Jurnal Evolusi*, vol. 6, no. 2, pp. 1–5, 2018.
- [15] Raman Kumar, Jasgurpreet Singh Chohan, Sandeep Singh, Shubham Sharma, Yadvinder Singh, and S. Rajkumar, "Implementation of Taguchi and Genetic Algorithm Techniques for Prediction of Optimal Part Dimensions for Polymeric Biocomposites in Fused Deposition Modeling," *Hindawi International Journal of Biomaterials*, 2022.
- [16] Amrin, Irawan Satriadi, and Oki Rosanto, "Algoritma C4. 5 Untuk Diagnosa Penyakit Tuberkulosis," *JURNAL KHATULISTIWA INFORMATIKA*, vol. VII, no. 2, pp. 79–84, Dec. 2019.
- [17] Yuslena Sari, Muhammad Alkaff, Eka Setya Wijaya, Syarifah Soraya, and Dany Primanita Kartikasari, "Optimasi Penjadwalan Mata Kuliah Menggunakan Metode Algoritma Genetika dengan Teknik Tournament Selection," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 6, no. 1, pp. 85–92, Feb. 2019.