



## Analisis Perbandingan Algoritma *Random Forest* dan Algoritma *Naive Bayes* untuk Memprediksi Penyakit Paru-Paru di Indonesia

Eka Wulansari Fidayanthie<sup>1\*</sup>, Asep Sayfullah<sup>2</sup>, Mardiana Rafa Alzena<sup>3</sup>, Nilam Kurnia Sari<sup>4</sup>

<sup>1-4</sup> Universitas Bina Sarana Informatika, Indonesia

[eka.ewf@bsi.ac.id](mailto:eka.ewf@bsi.ac.id)<sup>1</sup>, [asep.alo@bsi.ac.id](mailto:asep.alo@bsi.ac.id)<sup>2</sup>, [alzenar29@gmail.com](mailto:alzenar29@gmail.com)<sup>3</sup>, [nilamkurniasari44@gmail.com](mailto:nilamkurniasari44@gmail.com)<sup>4</sup>

Alamat : Jl. Gatot Subroto No.8, Cimone, Kec. Karawaci, Kota Tangerang, Banten 15114

Korespondensi penulis : [eka.ewf@bsi.ac.id](mailto:eka.ewf@bsi.ac.id)\*

**Abstract-** Lungs are vital organs in the human respiratory system, responsible for fulfilling the body's oxygen needs. If the lungs experience health problems, it can have adverse effects on the human respiratory system. Common causes of lung diseases are usually due to inhaling air contaminated by dust, smoke, viruses, and bacteria. This study aims to compare the performance of two classification algorithms, namely *Random Forest* and *Naive Bayes*, in predicting lung diseases. The data used was obtained from the Kaggle website and processed using *RapidMiner* software. The attributes involved include smoking habits, pre-existing conditions, staying up late, exercise activities, age, and outcomes. Based on the test results, the *Random Forest* algorithm demonstrated the best performance with an accuracy of 93%, while the *Naive Bayes* algorithm achieved an accuracy of 87%. These findings indicate that the *Random Forest* algorithm outperforms the *Naive Bayes* algorithm in terms of lung disease prediction accuracy.

**Keywords:** Data Mining, *Random Forest*, *Naive Bayes*, Lung Disease

**Abstrak-** Paru-paru adalah organ penting dalam sistem pernapasan manusia. Yang memiliki fungsi memenuhi kebutuhan oksigen terhadap tubuh. Jika paru-paru mengalami gangguan kesehatan dapat berakibat buruk terhadap sistem pernapasan manusia. Penyebab penyakit paru-paru yang sering terjadi biasanya dikarenakan menghirup udara tercemar oleh debu, asap, virus dan bakteri. Penelitian ini bertujuan untuk membandingkan performa dua algoritma klasifikasi, yaitu *Random Forest* dan *Naive Bayes*, dalam memprediksi penyakit paru-paru. Data yang digunakan didapat dari situs Kaggle dan diolah dengan *software RapidMiner*. Meliputi atribut merokok, penyakit bawaan, aktivitas begadang, aktivitas olahraga, usia dan hasil. Berdasarkan hasil pengujian, *algoritma Random Forest* menunjukkan performa terbaik dengan akurasi sebesar 93%, sementara itu *algoritma Naive Bayes* memperoleh akurasi 87%. Temuan ini menunjukkan bahwa *algoritma Random Forest* lebih unggul dibandingkan *algoritma Naive Bayes* dalam hal akurasi prediksi penyakit paru-paru.

**Kata Kunci:** Data Mining, *Random Forest*, *Naive Bayes*, Penyakit Paru-paru

### 1. PENDAHULUAN

Paru-paru bertanggung jawab atas sistem pernapasan manusia. Membantu proses pertukaran gas antara udara luar dan darah supaya oksigen dalam tubuh tetap seimbang. Paru-paru memiliki dua peran untuk membuang karbon dioksida dari pemecah metabolik dari aliran darah dan mengambil oksigen melalui udara. Proses kerja paru-paru dapat dikaitkan dengan kondisi fisik yang mempengaruhi seluruh sistem pernapasan. Penyebab penyakit paru-paru yang sering terjadi biasanya dikarenakan menghirup udara tercemar oleh debu, asap, virus dan bakteri. Penyakit paru-paru dapat menyerang siapa saja tanpa melihat umur manusia mulai dari bayi hingga orang dewasa yang tentunya penyakit paru-paru ini sulit untuk disembuhkan. (Haffandi et al., 2022; Ritonga et al., 2025).

Pengambilan keputusan yang mengacu berdasarkan data dan informasi dapat meningkatkan ketepatan dalam membuat keputusan atau memprediksi kondisi kesehatan seseorang. Oleh sebab itu, salah satu cara terbaik untuk menerapkan sistem pendukung keputusan adalah dengan memanfaatkan teknologi dengan baik. (Khoirudin, 2024)

Sesuai dengan latar belakang diatas, untuk memprediksi perbedaan antara algoritma Random Forest dan algoritma Naive Bayes dikarenakan kedua algoritma tersebut memiliki kelebihan dan kekurangan masing-masing. Oleh sebab itu, peneliti perlu melakukan analisis perbandingan antara Random Forest dan Naive Bayes dalam memprediksi penyakit paru-paru untuk mengetahui algoritma mana yang memberikan akurasi terbaik dalam data kesehatan di indonesia.

## **2. LANDASAN TEORI**

### ***Data Mining***

*Data mining* didefinisikan sebagai proses otomatis atau semi-otomatis yang menganalisis sejumlah data besar untuk mengidentifikasi pola, anomali dan korelasi dalam sejumlah data besar yang sebelumnya tidak diketahui. Ini melibatkan penggunaan kecerdasan buatan, matematika, statistik, dan machine learning untuk menemukan hubungan dan tren yang dapat digunakan untuk pengambilan keputusan. Data mining juga merupakan proses menemukan informasi penting dari kumpulan data dan kemudian mengubahnya menjadi pengetahuan yang dapat digunakan. Proses ini dikenal dengan *Knowledge Discovery in Database* (KDD). Gregory Piatetsky Shapiro adalah *Founder* dan presiden sebuah situs *website* belajar dan diskusi *KDNuggets*, yang berfokus pada *Business, Analytics, Data Mining*, dan *Data Science*. Pada *Workshop* pertama kalinya, dia menggunakan istilah *KDD* (*Knowledge Discovery in Database*). Karena sumber pengetahuan tidak lagi mengacu pada basis data. (Fajri et al., 2024)

### ***Random Forest***

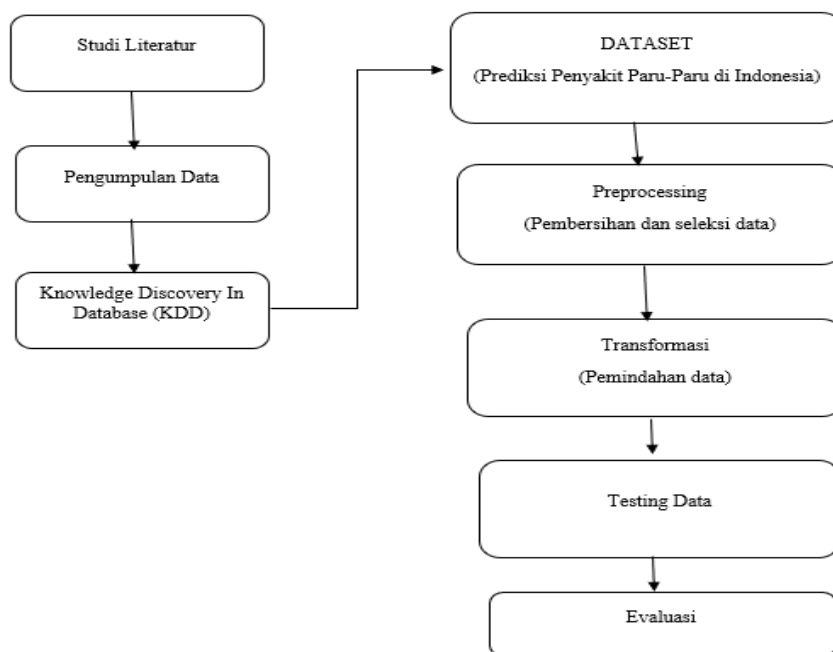
*Random forest* adalah teknik pembelajaran yang digunakan pada pengklasifikasian dan prediksi. *Random forest* termasuk bagian dari algoritma machine learning, dalam penerapan metode ini dibutuhkan ketersediaan data yang berkualitas tinggi untuk mendukung proses pengumpulan dan pengolahan data. Informasi mengenai kinerja sistem berperan penting untuk meningkatkan performa perangkat lunak dan perangkat keras, menganalisis serta mengidentifikasi pola yang saling berhubungan. (Salman et al., 2024).

## Naive Bayes

Algoritma *Naive Bayes* merupakan salah satu metode yang umum digunakan dalam pengklasifikasian statistik yang didasarkan pada *Teorema Bayes*. Metode ini memberikan prediksi *probabilitas* untuk kategori atau kelas tertentu berdasarkan fitur-fitur yang berkaitan dengan data. Proses pengklasifikasian menggunakan *Naive Bayes* dilakukan melalui langkah-langkah berikut: (Junaidi et al., 2024)

## Kerangka Pemikiran

Kerangka pemikiran disusun oleh peneliti sebagai landasan dalam pelaksanaan penelitian, yang dapat dilihat pada gambar berikut:



**Gambar 1.** Kerangka Pemikiran

## Evaluasi Model

*Confusion Matrix* merupakan metode standar dalam bidang data mining untuk mengevaluasi akurasi model. *Matrix* ini memungkinkan penilaian menyeluruh melalui empat parameter utama: akurasi, *recall*, presisi, dan tingkat kesalahan (*error rate*). Tujuan utama dari *Confusion Matrix* adalah untuk menghitung jumlah prediksi yang benar (*true positives* dan *true negatives*) dan prediksi yang salah (*false positives* dan *false negatives*) yang dihasilkan oleh model. *True Positives (TP)*, *True Negatives (TN)*, *False Positives (FP)*, *False Negatives (FN)* adalah *matrix* yang digunakan untuk menilai kinerja model secara khusus. Ilustrasi dan

formulasi yang relevan menawarkan rincian perhitungan dan representasi visual dari *matrix* ini. Seperti pada tabel dibawah ini : (Argarini Pratama & Hellyana, 2022)

**Tabel 1.** Komposisi Tabel Hasil Confusion Matrix

Kelas	Terklarifikasi Positif	Terklarifikasi Negatif
<b>Positif</b>	<b>True Positif (TP)</b>	<b>False Negative (FN)</b>
<b>Negatif</b>	<b>False Positif (FP)</b>	<b>True Negative (TN)</b>

Sumber : (Argarini Pratama & Hellyana, 2022)

*True Positives (TP)* merepresentasikan jumlah data secara inheren bernilai benar dan nilai kebenaran datanya juga diidentifikasi benar oleh model. Sebaliknya, *False Negatives (FN)* adalah jumlah data yang sebenarnya bernilai benar, tetapi dianggap salah oleh sistem. *True Negatives (TN)* mencatat jumlah data yang intrinsik dan nilai kebenarannya dikenali sebagai benar oleh sistem. Sebaliknya *False Positives (FP)* menunjukkan jumlah data yang sejatinya salah, tetapi sistem menganggapnya benar.:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

### 3. METODOLOGI PENELITIAN

#### Proses Pengumpulan Data

Dataset pada penelitian ini diperoleh dari website <https://www.kaggle.com> . Dataset yang peneliti gunakan yang telah diunduh dalam format CSV dengan judul Prediksi Terkena Penyakit Paru-Paru, dataset ini diambil dari sumber aslinya <https://www.kaggle.com/datasets/andot03bsrc/dataset-predic-terkena-penyakit-paruparu> data sampel berjumlah 1000 data dan mempunyai 10 atribut di dalamnya seperti usia, jenis kelamin, merokok, bekerja, rumah tangga, aktivitas begadang, aktivitas olahraga, aktivitas lapangan, penyakit bawaan dan hasil.

Berikut adalah data sampel yang telah diperoleh :

No	Usia	Jenis_Kelamin	Merokok	Bekerja	Rumah_Tangga	Aktivitas_Begadang	Aktivitas_Olahraga	Aktivitas_Lapangan	Penyakit_Bawaan	Hasil
1	Tua	Pria	Pasif	Tidak	Ya	Ya	Sering	Ada	Tidak	Ya
2	Tua	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Ada	Tidak
3	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
4	Tua	Pria	Aktif	Ya	Tidak	Tidak	Jarang	Ada	Ada	Tidak
5	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Ya
6	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Tidak
7	Tua	Wanita	Pasif	Tidak	Ya	Tidak	Sering	Tidak	Tidak	Ya
8	Muda	Pria	Aktif	Tidak	Ya	Ya	Sering	Tidak	Tidak	Tidak
9	Tua	Wanita	Aktif	Ya	Ya	Ya	Jarang	Ada	Ya	Ya
10	Muda	Wanita	Pasif	Ya	Tidak	Ya	Jarang	Ada	Ada	Ya
11	Tua	Wanita	Pasif	Ya	Ya	Tidak	Sering	Ada	Ada	Ya
12	Tua	Wanita	Aktif	Tidak	Ya	Tidak	Jarang	Ada	Tidak	Tidak
13	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
14	Tua	Wanita	Aktif	Ya	Tidak	Ya	Jarang	Ada	Ada	Tidak
15	Muda	Wanita	Pasif	Ya	Tidak	Ya	Sering	Tidak	Ada	Ya
16	Muda	Wanita	Pasif	Ya	Tidak	Ya	Jarang	Ada	Ada	Ya
17	Tua	Wanita	Aktif	Ya	Ya	Tidak	Sering	Ada	Ada	Ya
18	Tua	Wanita	Aktif	Tidak	Ya	Tidak	Jarang	Ada	Tidak	Tidak
19	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
20	Tua	Wanita	Aktif	Ya	Tidak	Ya	Jarang	Ada	Ada	Tidak
21	Muda	Wanita	Pasif	Ya	Tidak	Ya	Sering	Tidak	Ada	Ya
22	Tua	Pria	Pasif	Tidak	Ya	Ya	Sering	Ada	Tidak	Ya
23	Tua	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Ada	Tidak
24	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
25	Tua	Pria	Aktif	Ya	Tidak	Tidak	Jarang	Ada	Ada	Tidak
26	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Ya
27	Muda	Wanita	Pasif	Ya	Tidak	Ya	Jarang	Ada	Ada	Ya
28	Tua	Wanita	Pasif	Ya	Ya	Tidak	Sering	Ada	Ada	Ya
29	Tua	Wanita	Aktif	Tidak	Ya	Tidak	Jarang	Ada	Tidak	Tidak
30	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak

Sumber: Penelitian, 2025

**Gambar 2.** Sampel dataset yang belum diolah

### 1. Data Selection

Beberapa atribut yang dianggap kurang relevan dalam penelitian ini digantikan dengan atribut aktivitas lapangan yang lebih relevan. Tujuan utama tahapan ini dilakukan untuk mengidentifikasi atribut yang sesuai dan berkontribusi terhadap munculnya gejala penyakit paru-paru. Tabel berikut menunjukkan atribut yang akan diolah oleh *Rapidminer*.

**Tabel 2.** Atribut Terpilih

No	Atribut Terpilih	Keterangan
1	Merokok	Aktif/Pasif
2	Penyakit_Bawaan	Ada/Tidak
3	Aktivitas_Begadang	Tidak/Ya
4	Aktivitas_Olahraga	Tidak/Ya
5	Usia	Muda/Tua
6	Hasil	Ya/Tidak

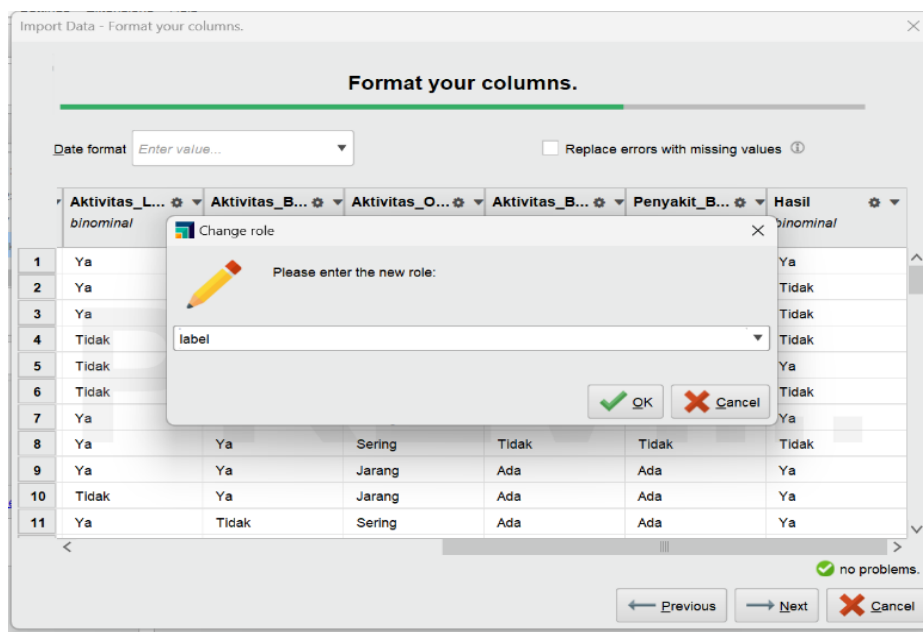
Sumber : Penelitian 2025

Atribut yang terpilih adalah merokok, penyakit\_bawaan, aktivitas\_begadang, aktivitas\_olahraga, usia dan hasil. Selanjutnya, atribut-atribut tersebut akan diproses menggunakan algoritma *Random Forest* dan perangkat lunak *RapidMiner*. Proses pengolahan ini akan menghasilkan pohon keputusan dan nilai akurasi yang digunakan untuk prediksi penyakit paru-paru.

### Data Transformation

Tahapan ini melibatkan transformasi data ke dalam format yang mendukung proses pemodelan dan analisis data mining. Proses transformasi dataset ini dilakukan menggunakan perangkat lunak *Rapidminer*, yang meliputi transformasi data dari kategori polynominal menjadi binominal guna mendukung kebutuhan analisis pada atribut-atribut yang relevan.

Selain itu, fase ini menetapkan hasil karakteristik sebagai label untuk penentuan keputusan dalam prediksi penyakit paru-paru.



Gambar 3. Perubahan tipe data dan set role

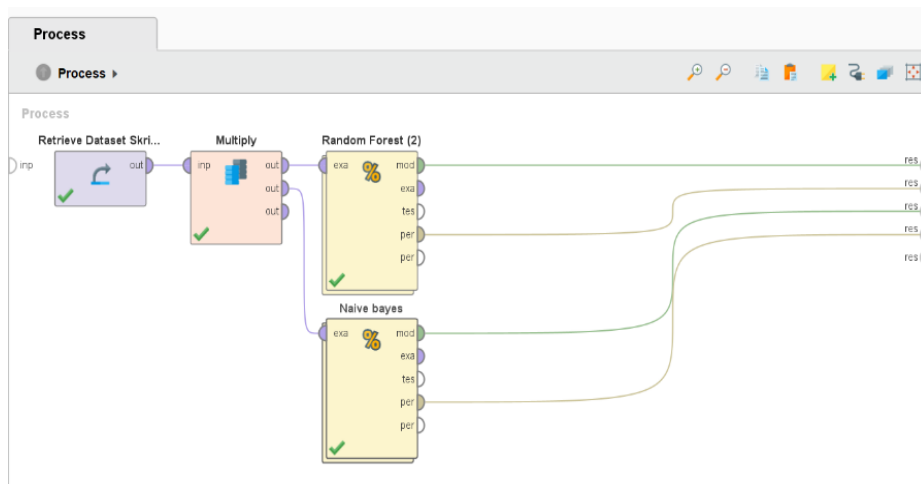
#### 4. HASIL DAN PEMBAHASAN

##### Pengujian pada Tools *RapidMiner*

Data yang telah diklasifikasikan, selanjutnya diproses menggunakan kapabilitas *RapidMiner*. Proses ini dilakukan untuk memastikan *validitas* dan *reliabilitas* data, sehingga menjamin akurasi. Kemudian, data yang telah *terverifikasi* akurat tersebut diolah guna menghasilkan luaran komparatif untuk prediksi penyakit paru-paru. Tahapan ini bertujuan untuk mendukung pengambilan keputusan yang didasarkan pada hasil olah data.

##### Proses Pengolahan data Tools *RapidMiner*

Tahap selanjutnya melibatkan implementasi prosedur perhitungan tingkat akurasi. Proses ini diterapkan pada algoritma Naive Bayes ataupun *Random Forest* menggunakan *RapidMiner*. Gambar III.3 merupakan desain model proses utama perbandingan algoritma untuk klasifikasi data penyakit paru-paru.

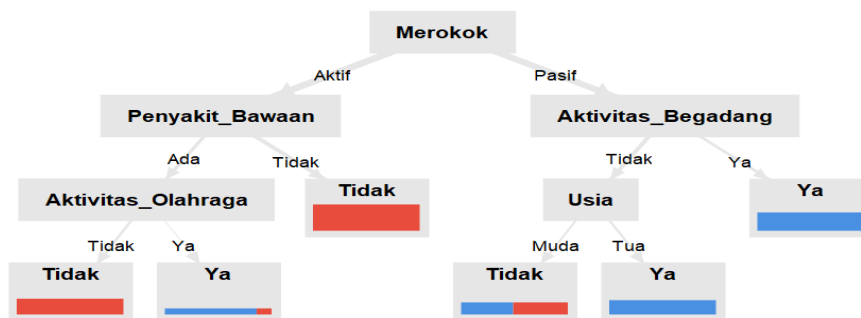


**Gambar 4.** Desain model perbandingan algoritma klasifikasi

Perancangan desain proses utama ini menggunakan satu dataset yang sama untuk penerapan dua algoritma secara bersamaan, yaitu *Random Forest* dan *Naive Bayes* diterapkan untuk memastikan bahwa setiap algoritma menggunakan dataset yang konsisten dan identik dalam proses pengolahan.

### **Random Forest**

Berdasarkan proses yang telah dilakukan, gambar berikut menampilkan salah satu hasil pohon acak dari *Random Forest* yang digunakan pada penelitian ini



**Gambar 5.** *Random Forest* (Pohon Acak)

Gambar berikut ini merupakan representasi aturan model dalam bentuk teks yang dihasilkan *Rapidminer*. Ini memberikan penjelasan mengenai struktur model *Random Forest* yang telah dibentuk.

## Tree

```
Merokok = Aktif
| Penyakit_Bawaan = Ada
| | Aktivitas_Olahraga = Tidak: Tidak {Ya=0, Tidak=170}
| | Aktivitas_Olahraga = Ya: Ya {Ya=49, Tidak=8}
| Penyakit_Bawaan = Tidak: Tidak {Ya=0, Tidak=293}
Merokok = Pasif
| Aktivitas_Begadang = Tidak
| | Usia = Muda: Tidak {Ya=62, Tidak=65}
| | Usia = Tua: Ya {Ya=153, Tidak=0}
| Aktivitas_Begadang = Ya: Ya {Ya=200, Tidak=0}
```

**Gambar 6.** Rule Model Random Forest (Pohon Acak)

Hasil dari pohon acak ini yaitu :

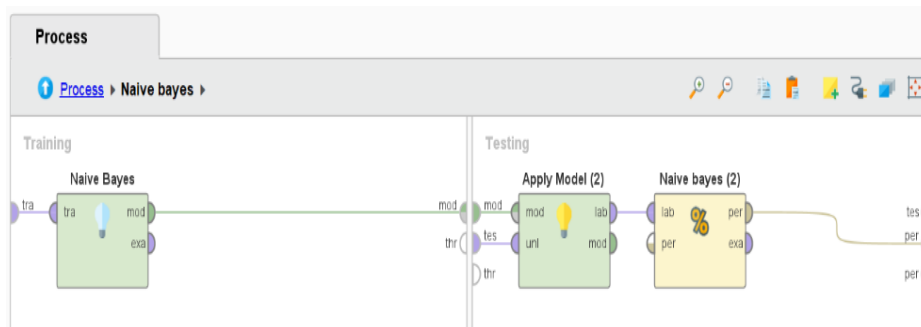
Perokok Aktif

- Jika merokok aktif dan mempunyai penyakit bawaan maka diprediksi terkena penyakit paru-paru.
- Jika merokok aktif dan tidak mempunyai penyakit bawaan maka tidak terkena penyakit paru-paru.
- Jika merokok aktif, mempunyai penyakit bawaan dan mempunyai aktivitas olahraga maka tidak terkena penyakit paru-paru.
- Jika merokok aktif, mempunyai penyakit bawaan dan tidak mempunyai aktivitas olahraga maka diprediksi terkena penyakit paru-paru.

Perokok Pasif

- Jika merokok pasif dan tidak melakukan aktivitas begadang maka tidak terkena penyakit paru-paru.
- Jika merokok pasif dan melakukan aktivitas begadang maka diprediksi terkena penyakit paru-paru.
- Jika merokok pasif, melakukan aktivitas begadang, dan berusia muda maka tidak terkena penyakit paru-paru.
- Jika merokok pasif, melakukan aktivitas begadang, dan berusia tua maka diprediksi terkena penyakit paru-paru.

## Naive Bayes



**Gambar 7.** Pengujian *Naive Bayes*

Algoritma *Naive bayes* digunakan pada gambar diatas sebagai model desain untuk pengumpulan data penyakit paru-paru, dan aturan modelnya sebagai berikut :

### SimpleDistribution

Distribution model for label attribute Hasil

Class Ya (0.478)  
10 distributions

Class Tidak (0.522)  
10 distributions

**Gambar 8.** Hasil Pengujian *Naive Bayes*

Hasil algoritma *Naive bayes* ditunjukkan pada gambar diatas untuk kelas Ya dan Tidak masing-masing dengan nilai  $0,478$  dan  $0,522$  sehingga dapat dikatakan bahwa kelas *No* merupakan kelas tertinggi untuk memprediksi penyakit paru-paru.

## Evaluasi

### *Random Forest*

Tahapan pengujian akurasi dilaksanakan setelah proses pengolahan data, dengan tujuan untuk menilai performa setiap metode yang diterapkan *Random Forest*, tujuan penelitian ini untuk menentukan jenis data yang mampu memberikan hasil prediksi akurat. Dalam upaya memprediksi penyakit paru-paru, sistem ini dimanfaatkan sebagai sumber referensi yang dapat diandalkan. *Confusion matrix* dan kurva *ROC/AUC (Area Under Cover)* digunakan sebagai metode evaluasi. Hasil perhitungan data training menggunakan algoritma *Random Forest*. Diketahui tingkat akurasinya adalah 93.00%. Gambar *confusion matrix* disajikan pada gambar III.8 dan gambar III.9 merupakan grafik *ROC/AUC (Area Under Cover)* dan model *confusion matrix* yaitu 0.992

accuracy: 93.00% +/- 1.83% (micro average: 93.00%)

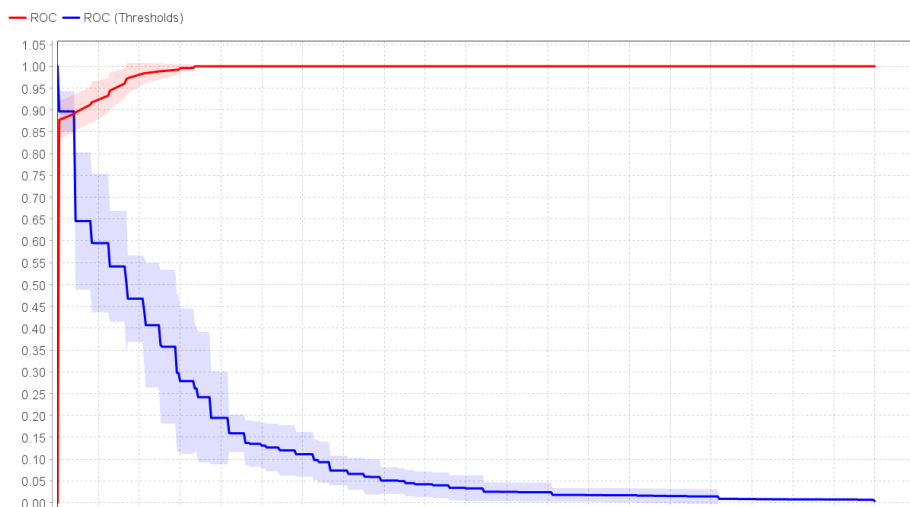
	true Ya	true Tidak	class precision
pred. Ya	451	43	91.30%
pred. Tidak	27	479	94.66%
class recall	94.35%	91.76%	

**Gambar 9.** Akurasi Random Forest

Berdasarkan hasil pengujian pada gambar diatas tingkat akurasi yang didapat dengan melakukan perhitungan berikut, dimana jumlah *True Positive* mencapai 413 data, *True Negative* berjumlah 457 data , *False Positive* sebanyak 65 data dan *False Negative* juga 65 data, tingkat akurasi dari model algoritma bisa dihitung dengan rumus berikut :

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+Fp+FN} \\
 &= \frac{451+479}{451+43+27+479} \\
 &= \frac{930}{1000} \\
 &= 0,93
 \end{aligned}$$

AUC: 0.992 +/- 0.004 (micro average: 0.992) (positive class: Tidak)



**Gambar 10.** Kurva Confusion Matrix ROC Random Forest

### ***Naive Bayes***

Pada tahap selanjutnya, pengujian dilakukan dengan menerapkan algoritma Naive Bayes, yang menghasilkan tingkat akurasi sebesar 87.00% . Hasil evaluasi disajikan dalam gambar III.10 dalam bentuk *Confusion Matrix*, sedangkan gambar III.11 menampilkan kurva ROC/AUC (*Area Under Cover*) dari model nilai sebesar 0.852.

accuracy: 87.00% +/- 1.70% (micro average: 87.00%)

	true Ya	true Tidak	class precision
pred. Ya	413	65	86.40%
pred. Tidak	65	457	87.55%
class recall	86.40%	87.55%	

**Gambar 11.** Akurasi *Naive Bayes*

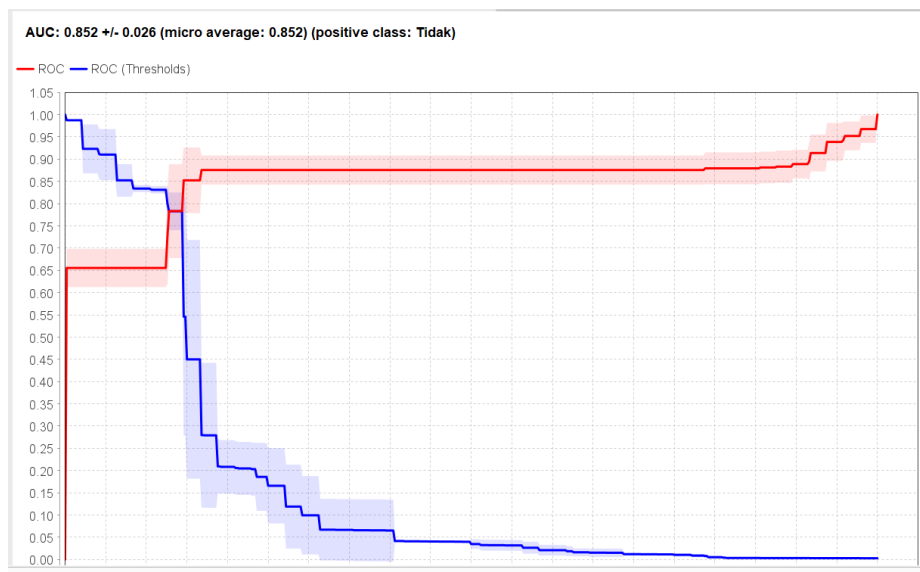
Berdasarkan hasil pengujian pada gambar diatas tingkat akurasi yang didapat dengan melakukan perhitungan berikut, dimana jumlah *True Positive* mencapai 413 data, *True Negative* berjumlah 457 data, *False Positive* sebanyak 65 data dan *False Negative* juga 65 data, tingkat akurasi dari model algoritma bisa dihitung dengan rumus berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$= \frac{413+457}{413+65+65+457}$$

$$= \frac{870}{1000}$$

$$= 0,87$$



**Gambar 12.** Kurva *ROC Matrix Naive Bayes*

**Tabel 3.** Komparasi Akurasi dan AUC

	<i>Random Forest</i>	<i>Naive Bayes</i>
<i>Accuracy</i>	<b>93.00%</b>	<b>87.00%</b>
<i>AUC</i>	<b>0.992</b>	<b>0.852</b>

## 5. KESIMPULAN

### Kesimpulan

Sebagai penutup, penelitian ini menunjukkan bahwa penerapan algoritma dalam proses prediksi penyakit paru-paru memiliki peran yang signifikan, khususnya dengan membandingkan metode *Random Forest* dan *Naive Bayes*. Paru-paru yang merupakan organ vital dalam pernapasan manusia sangat rentan terhadap gangguan kesehatan, terutama akibat dari paparan polusi dan faktor risiko lainnya. Dalam penelitian ini, peneliti memanfaatkan aplikasi *RapidMiner* untuk mengolah data yang diperoleh dari situs *Kaggle* yang mencakup atribut penting seperti merokok, penyakit bawaan, aktivitas begadang, aktivitas olahraga, usia dan hasil. Berdasarkan hasil pengujian, *algoritma Random Forest* menunjukkan performa terbaik dengan akurasi sebesar 93%, sementara itu *algoritma Naive Bayes* memperoleh akurasi 87%. Temuan ini memperkuat bahwa *algoritma Random Forest* lebih unggul dalam memprediksi penyakit paru-paru. Secara keseluruhan, penelitian menegaskan bahwa pemanfaatan teknologi dan pengolahan data yang tepat dapat mendukung upaya deteksi dini penyakit paru-paru, serta berkontribusi pada peningkatan kualitas layanan kesehatan masyarakat.

### Saran

Adapun saran yang dapat peneliti sampaikan berdasarkan kesimpulan yaitu:

Untuk penelitian selanjutnya, diperlukan studi lanjutan untuk memprediksi model penyakit paru-paru di Indonesia dengan menggunakan algoritma lain seperti decision tree, KNN, ID3, SVM atau algoritma lainnya dengan menggunakan aplikasi *rapidminer* agar dapat dibandingkan dengan hasil prediksi penyakit paru-parunya.

## REFERENSI

- Abdul Khalim, K., Hayati, U., & Bahtiar, A. (2023). Perbandingan prediksi penyakit hipertensi menggunakan metode random forest dan Naïve Bayes. *Jurnal Mahasiswa Teknik Informatika*, 7(1). <https://garuda.kemdikbud.go.id/documents/detail/3374524>
- Buana, I., & Harahap, D. A. (2022). Asbestos, radon dan polusi udara sebagai faktor risiko kanker paru pada perempuan bukan perokok. *AVERROUS: Jurnal Kedokteran dan Kesehatan Malikussaleh*, 8(1). <https://ojs.unimal.ac.id/averrous/article/view/7088/0>
- Ernawati, A., Sitorus, Z., Iqbal, M., & Nasution, D. (2025). Penerapan data mining untuk klasifikasi penduduk miskin di Kabupaten Labuhanbatu menggunakan Random Forest dan K-Nearest Neighbors. *Bulletin of Information Technology (BIT)*, 6, 23–35. <https://doi.org/10.47065/bit.v5i2.1783>
- Fajri, T. I., Sari, H. L., Dinata, R. K., Hasdyna, N., Hasugian, B. S., Retno, S., Wahyuni, S., Fadhilah, C., & Pohan, N. (2024). *Data Mining (J. Prayoga, Ed.; Pertama)*. PT Serasi Media Teknologi. <https://books.google.co.id/books?id=YykdEQAAQBAJ&printsec=copyright&hl=id>

- Haffandi, M. Y., Haerani, E., Syafria, F., & Oktavia, L. (2022). Klasifikasi penyakit paru-paru dengan menggunakan metode Naïve Bayes classifier. *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, 5(2), 176. <https://doi.org/10.37600/tekinkom.v5i2.649>
- Hidayat, M. T., Suarna, N., & Rahaningsih, N. (2023). Implementasi algoritma Naïve Bayes untuk prediksi persediaan barang PT Dilmoni Citra Mebel Indonesia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 693–699.
- Junaidi, S., Beno, I. S., Farkhan, M., Supartha, I. K. D. G., Pasaribu, A. A., Kmurawak, R. M., Supriyanto, S., Sroyer, A. M., Reba, F., Fitriyanto, R., Syafiqoh, U., & Rizal, A. A. (2024). *Buku Ajar Machine Learning (Efitra, Ed.; Pertama)*. PT Sonpedia Publishing Indonesia.  
[https://www.google.co.id/books/edition/Buku\\_Ajar\\_Machine\\_Learning/ACT2EAAAQBAJ](https://www.google.co.id/books/edition/Buku_Ajar_Machine_Learning/ACT2EAAAQBAJ)
- Karima, R. A., & Fatah, Z. (2024). Implementasi metode K-Nearest Neighbor untuk klasifikasi penyakit paru-paru pada anak. *Jurnal Ilmiah Multidisiplin Ilmu*, 1(6), 10–17.
- Khoirudin. (2024). Klasifikasi penyakit jantung menggunakan perbandingan algoritma Decision Tree dan Naïve Bayes. *Kopertip: Jurnal Ilmiah Manajemen Informatika dan Komputer*, 8(1), 19–25. <https://doi.org/10.32485/kopertip.v8i1.351>
- Pambudi, R., Rahman Harahap, A., Saputra, F. D., & Jusub, M. (2024). Klasifikasi penyakit paru-paru menggunakan metode Decision Tree. *Oktal: Jurnal Ilmu Komputer dan Science*, 3, 2397–2398. <https://journal.mediapublikasi.id/index.php/oktal>
- Pratama, E. A., Hellyana, C. M., & Fadlilah, N. I. (2022). Perbandingan 3 algoritma klasifikasi data mining dalam pro-kontra bahaya rokok elektrik. *Jurnal Teknoinfo*, 16(1), 93.
- Rasyid, N., Nawaf, T. B., Nuraminah, A., & Pulungan, M. P. (2022). Sistem pakar untuk diagnosa penyakit paru-paru menggunakan metode Certainty Factor. *Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer*, 2(1), 18–21.
- Ritonga, M. J., Khoirudin, & Albahi, M. (2025). Akad dalam transaksi keuangan syariah. *Al-Kharaj: Jurnal Ekonomi Keuangan & Bisnis Syariah*, 7, 2282. <https://doi.org/10.47467/alkharaj.v7i6.8065>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest algorithm overview. *Babylonian Journal of Machine Learning*, 69–79. <https://doi.org/10.58496/BJML/2024/007>
- Sirojul Munir, A., Saputra, A. B., Aziz, A., Agung Barata, M., Yani, A., 10, N., & Bojonegoro, K. (2024). Perbandingan akurasi algoritma Naïve Bayes dan algoritma Decision Tree dalam pengklasifikasian penyakit kanker payudara.
- Sugiharto, S., Putri, R. A., Simanjuntak, S., & Larissa, O. (2021). *Seminar Nasional Hasil Penelitian dan Pengabdian Kepada Masyarakat 2021: Pengembangan Ekonomi Bangsa melalui Inovasi Digital*. Jakarta.