

Perbandingan Algoritma Klasifikasi Random Forest dengan Naïve Bayes Classifier pada Studi Penyakit Berdasarkan Pola Nutrisi

Akhmad Pandhu Wijaya
Universitas Wahid Hasyim
Semarang, Indonesia

pandhuds@unwahas.ac.id

*Penulis Korespondensi

Diajukan : 17/01/2025

Diterima : 29/01/2025

Dipublikasi : 30/01/2025

ABSTRAK

Penelitian ini bertujuan untuk membandingkan efektivitas algoritma klasifikasi Random Forest dan Naïve Bayes dalam mengklasifikasikan penyakit berdasarkan pola nutrisi. Dataset yang digunakan mencakup atribut seperti usia, jenis kelamin, tinggi badan, berat badan, tingkat aktivitas, preferensi makanan, dan sejumlah informasi terkait nutrisi, termasuk kalori, protein, gula, sodium, karbohidrat, dan serat. Dalam penelitian ini, dua metode klasifikasi yang umum digunakan dalam data mining, yaitu Random Forest dan Naïve Bayes Classifier (NBC), diterapkan untuk menganalisis data dan memprediksi penyakit yang mungkin timbul berdasarkan pola konsumsi makanan. Proses preprocessing data melibatkan pembersihan data, penanganan nilai hilang, dan normalisasi variabel numerik. Untuk evaluasi model, digunakan matriks kinerja seperti akurasi, presisi, recall, dan F1-score untuk menilai kemampuan model dalam memprediksi penyakit secara tepat. Hasil penelitian menunjukkan bahwa model Random Forest memiliki performa terbaik dibandingkan NBC, dengan tingkat akurasi yang lebih tinggi serta presisi dan recall yang lebih baik. Temuan ini mengindikasikan bahwa algoritma Random Forest lebih efektif dalam mengklasifikasikan penyakit berdasarkan pola nutrisi dibandingkan dengan Naïve Bayes Classifier. Kesimpulannya, penelitian ini menunjukkan bahwa algoritma Random Forest lebih unggul dibandingkan Naïve Bayes Classifier dalam mengklasifikasikan penyakit berdasarkan pola nutrisi. Dengan hasil yang menunjukkan tingkat akurasi, presisi, dan recall yang lebih tinggi, Random Forest terbukti lebih efektif dalam menangani kompleksitas data yang melibatkan berbagai atribut nutrisi.

Kata Kunci: Akurasi Model, Data Mining, *Naïve Bayes Classifier*, *Random Forest*.

I. PENDAHULUAN

Dalam era modern ini, kesehatan masyarakat semakin menjadi perhatian utama, terutama dengan meningkatnya prevalensi penyakit yang berkaitan dengan pola makan dan gaya hidup. Pola nutrisi yang tidak seimbang dapat berkontribusi pada berbagai masalah kesehatan, termasuk obesitas, diabetes, dan penyakit jantung. Penting untuk memahami hubungan antara pola konsumsi makanan dan risiko penyakit. Dengan kemajuan teknologi informasi dan data mining, analisis data kesehatan dapat dilakukan dengan lebih efektif, memungkinkan identifikasi pola dan prediksi penyakit berdasarkan data nutrisi (Nomira et al. 2024).

Algoritma klasifikasi merupakan salah satu metode yang umum digunakan dalam analisis data untuk memprediksi kategori atau kelas dari suatu data berdasarkan atribut yang ada. Dua algoritma yang sering digunakan dalam konteks ini adalah Random Forest dan Naïve Bayes Classifier (Arfyanti and Bustomi 2024). Random Forest, yang merupakan metode *ensemble*, dikenal karena kemampuannya dalam menangani data yang kompleks dan memiliki banyak atribut, serta memberikan hasil yang lebih akurat. Di sisi lain, Naïve Bayes Classifier, yang berbasis pada

teorema Bayes, menawarkan kecepatan dan kesederhanaan dalam implementasi, meskipun mungkin kurang efektif dalam menangani interaksi antar atribut.

Penelitian sebelumnya menunjukkan bahwa Random Forest sering kali menghasilkan akurasi yang lebih tinggi dibandingkan dengan Naïve Bayes dalam berbagai aplikasi klasifikasi, termasuk dalam bidang kesehatan. Misalnya, Khan dan Khan (2022) menemukan bahwa Random Forest lebih unggul dalam memprediksi penyakit berdasarkan data nutrisi (Khan, M. A. 2022), sementara Sari dan Prabowo (2023) menekankan pentingnya pemilihan algoritma yang tepat untuk analisis data kesehatan (Sari, D. P. 2023). Temuan ini menunjukkan bahwa pemilihan algoritma yang tepat dapat mempengaruhi hasil analisis dan prediksi yang dihasilkan.

Mengingat perbedaan karakteristik dan keunggulan masing-masing algoritma, penelitian ini bertujuan untuk membandingkan efektivitas Random Forest dan Naïve Bayes dalam mengklasifikasikan penyakit berdasarkan pola nutrisi. Penelitian ini diharapkan dapat memberikan wawasan yang lebih dalam tentang algoritma mana yang lebih baik dalam memprediksi penyakit dengan menggunakan dataset yang mencakup berbagai atribut terkait nutrisi dan kesehatan. Hasil dari penelitian ini tidak hanya akan berkontribusi pada pengembangan metode analisis data kesehatan, tetapi juga dapat memberikan rekomendasi yang berguna bagi praktisi kesehatan dalam upaya pencegahan dan penanganan penyakit.

Selain itu, penelitian ini juga akan mengeksplorasi proses preprocessing data yang meliputi pembersihan data, penanganan nilai hilang, dan normalisasi variabel numerik. Proses ini sangat penting untuk memastikan kualitas data yang digunakan dalam analisis, sehingga hasil yang diperoleh lebih akurat dan dapat diandalkan. Dengan latar belakang tersebut, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam bidang kesehatan dan teknologi informasi, serta membuka peluang untuk penelitian lebih lanjut dalam analisis pola nutrisi dan klasifikasi penyakit.

Referensi terkini yang mendukung penelitian ini mencakup studi-studi yang telah dilakukan sebelumnya, seperti yang dipublikasikan oleh Khan dan Khan (2022) serta Sari dan Prabowo (2023), yang menunjukkan relevansi dan pentingnya penggunaan algoritma klasifikasi dalam konteks kesehatan. Penelitian ini tidak hanya berfokus pada perbandingan algoritma, tetapi juga pada implikasi praktis dari hasil analisis dalam upaya meningkatkan kesehatan masyarakat. Penelitian ini memiliki urgensi yang tinggi mengingat meningkatnya angka penyakit yang berkaitan dengan pola makan di seluruh dunia. Dengan semakin banyaknya data yang tersedia mengenai pola nutrisi dan kesehatan, ada kebutuhan mendesak untuk menerapkan metode analisis yang efektif untuk mengidentifikasi faktor-faktor risiko. Pemahaman yang lebih baik tentang bagaimana pola nutrisi mempengaruhi kesehatan dapat membantu dalam pengembangan strategi pencegahan yang lebih efektif. Selain itu, perbandingan antara algoritma klasifikasi yang berbeda dapat memberikan wawasan tentang metode mana yang paling sesuai untuk aplikasi di bidang kesehatan, sehingga dapat meningkatkan akurasi dalam diagnosis dan pengobatan.

Manfaat penelitian ini sangat signifikan dalam berbagai aspek. Pertama, hasil dari perbandingan algoritma Random Forest dan Naïve Bayes Classifier dapat memberikan wawasan yang lebih dalam tentang efektivitas masing-masing metode dalam menganalisis data kesehatan, khususnya dalam konteks pola nutrisi dan penyakit. Kedua, penelitian ini dapat membantu praktisi kesehatan dan ahli gizi dalam merumuskan strategi pencegahan dan intervensi yang lebih tepat berdasarkan data yang akurat. Selain itu, temuan dari penelitian ini dapat menjadi dasar bagi pengembangan sistem pendukung keputusan yang lebih canggih, yang dapat digunakan untuk diagnosis dini dan pengelolaan penyakit. Penelitian ini tidak hanya berkontribusi pada ilmu pengetahuan, tetapi juga memiliki implikasi praktis yang dapat meningkatkan kualitas hidup masyarakat melalui pemahaman yang lebih baik tentang hubungan antara pola nutrisi dan kesehatan.

II. STUDI LITERATUR

Penelitian Terdahulu

Algoritma klasifikasi merupakan metode yang digunakan untuk memprediksi kategori dari data berdasarkan atribut yang ada. Dua algoritma yang sering digunakan dalam analisis data kesehatan adalah Random Forest dan Naïve Bayes Classifier. Random Forest adalah metode

ensemble yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi overfitting. Menurut Khan dan Khan (2022), Random Forest menunjukkan kinerja yang lebih baik dalam menangani data yang kompleks dan memiliki banyak atribut dibandingkan dengan algoritma lain. Perbandingan dua metode menjadi bahan evaluasi dalam meningkatkan kualitas proses klasifikasi pada dataset sejenis (Rofiani et al. 2024). Di sisi lain, Naïve Bayes Classifier, yang berbasis pada teorema Bayes, menawarkan kecepatan dan kesederhanaan dalam implementasi, meskipun mungkin kurang efektif dalam menangani interaksi antar atribut (Sari, D. P. 2023).

Dalam konteks kesehatan, penelitian sebelumnya menunjukkan bahwa pemilihan algoritma yang tepat dapat mempengaruhi hasil prediksi penyakit. Sebuah studi oleh Ahmad et al. (2021) menemukan bahwa Random Forest lebih unggul dalam memprediksi diabetes dibandingkan Naïve Bayes, dengan tingkat akurasi yang lebih tinggi. Hal ini sejalan dengan temuan oleh Junaidi dan Rahman (2020), yang menunjukkan bahwa Random Forest lebih efektif dalam menganalisis data nutrisi dan kesehatan.

Selain itu, penelitian oleh Setiawan et al. (2022) menekankan pentingnya preprocessing data dalam meningkatkan kinerja model klasifikasi. Proses ini mencakup pembersihan data, penanganan nilai hilang, dan normalisasi variabel numerik, yang semuanya berkontribusi pada akurasi model. Dengan demikian, pemahaman yang lebih baik tentang algoritma dan teknik preprocessing dapat membantu dalam pengembangan sistem pendukung keputusan yang lebih efektif dalam bidang kesehatan. Secara keseluruhan, studi literatur ini menunjukkan bahwa perbandingan antara Random Forest dan Naïve Bayes Classifier dalam konteks pola nutrisi dan penyakit sangat relevan. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan metode analisis data kesehatan yang lebih baik dan lebih akurat.

Random Forest adalah algoritma berbasis ensemble yang dapat menangani data kompleks dengan performa tinggi, terutama dalam memproses dataset dengan banyak fitur (Breiman 2001). Naïve Bayes Classifier (NBC), di sisi lain, sangat efektif dalam memisahkan kelas pada dataset yang tidak linear melalui penggunaan kernel (Cortes, C., & Vapnik 1995). Penelitian oleh (Ahmad, T. 2020) menunjukkan bahwa algoritma Random Forest memiliki keunggulan dalam akurasi prediksi dibandingkan metode lain dalam analisis data kesehatan, sementara NBC unggul dalam menangani data dengan dimensi tinggi. Aplikasi Data Mining dalam Nutrisi Penggunaan data mining untuk menganalisis pola nutrisi dan risiko penyakit telah berkembang pesat. Misalnya, penelitian oleh (Johnson, R. K. 2018) menggunakan algoritma K- Nearest Neighbors (KNN) untuk mengklasifikasikan pola makan berdasarkan tingkat risiko diabetes. Selain itu, studi oleh (Kim, J. 2020) menggunakan kombinasi metode klasifikasi dan regresi untuk memprediksi asupan kalori harian yang optimal bagi pasien dengan gangguan metabolik.

Meskipun asumsi independensi antar fitur mungkin tidak selalu terpenuhi, NBC tetap menunjukkan kinerja yang baik dalam banyak aplikasi, termasuk dalam klasifikasi penyakit berdasarkan pola nutrisi (Rish, 2001). Penelitian sebelumnya menunjukkan bahwa kombinasi kedua algoritma ini dapat memberikan wawasan yang lebih dalam tentang faktor-faktor yang mempengaruhi kesehatan, serta membantu dalam pengembangan strategi pencegahan yang lebih efektif (Khan, M. A. 2022; Sari, D. P. 2023).

Pendekatan ini menunjukkan potensi data mining dalam mengidentifikasi pola dan hubungan yang kompleks antara pola nutrisi dan risiko penyakit. Dengan menerapkan algoritma klasifikasi seperti Random Forest dan Naïve Bayes, data mining dapat memberikan wawasan yang mendalam mengenai faktor-faktor yang berkontribusi terhadap kesehatan individu. Metode ini memungkinkan prediksi penyakit dengan akurasi yang lebih tinggi, sehingga memfasilitasi intervensi dini dan strategi pencegahan yang lebih efektif. Selain itu, kemampuan untuk menganalisis dataset besar dan beragam memungkinkan peneliti untuk mengeksplorasi interaksi antar variabel yang mungkin tidak terlihat melalui analisis tradisional. Data mining tidak hanya berfungsi sebagai alat analisis, tetapi juga sebagai dasar untuk pengembangan kebijakan kesehatan yang berbasis bukti dan peningkatan kualitas hidup masyarakat. Menurut (Zhang, Y. 2021), preprocessing data seperti normalisasi dan seleksi fitur dapat meningkatkan akurasi model prediksi. Selain itu, penelitian oleh (Gupta, A. 2019) menunjukkan pentingnya penanganan ketidakseimbangan kelas dalam dataset kesehatan untuk memastikan hasil prediksi yang valid.

Penelitian ini memberikan kontribusi signifikan dalam beberapa aspek. Pertama, penelitian ini memperkaya literatur yang ada mengenai penggunaan algoritma klasifikasi dalam analisis data kesehatan, khususnya dalam konteks pola nutrisi dan penyakit. Dengan membandingkan efektivitas Random Forest dan Naïve Bayes Classifier, penelitian ini memberikan wawasan baru tentang algoritma mana yang lebih sesuai untuk aplikasi di bidang kesehatan. Kedua, hasil penelitian ini dapat digunakan sebagai dasar untuk pengembangan sistem pendukung keputusan yang lebih baik dalam diagnosis dan pencegahan penyakit. Dengan pemahaman yang lebih baik tentang hubungan antara pola nutrisi dan risiko penyakit, praktisi kesehatan dan ahli gizi dapat merumuskan strategi intervensi yang lebih tepat dan berbasis data. Ketiga, penelitian ini juga membuka peluang untuk penelitian lebih lanjut dalam penerapan algoritma lain dan teknik analisis yang lebih canggih, serta mendorong eksplorasi lebih dalam mengenai faktor-faktor yang mempengaruhi kesehatan masyarakat.

III. METODE

1. Desain Penelitian

Penelitian ini merupakan studi kuantitatif berbasis analisis data yang bertujuan untuk membandingkan efektivitas algoritma klasifikasi Random Forest dan Naïve Bayes Classifier dalam menganalisis hubungan antara pola nutrisi dan risiko penyakit. Proses penelitian dimulai dengan pengumpulan data nutrisi harian, di mana data dikumpulkan melalui survei atau kuesioner yang dirancang untuk mengumpulkan informasi tentang kebiasaan makan responden, termasuk jenis makanan, frekuensi konsumsi, dan informasi nutrisi lainnya. Setelah data terkumpul, langkah selanjutnya adalah preprocessing data, yang mencakup pembersihan data untuk menghapus nilai yang hilang atau tidak valid, normalisasi variabel numerik, dan pengkodean variabel kategorikal agar dapat digunakan dalam model klasifikasi.

Setelah data diproses, tahap berikutnya adalah pembangunan model klasifikasi. Di sini, algoritma Random Forest dan Naïve Bayes Classifier diterapkan pada dataset untuk membangun model yang dapat memprediksi risiko penyakit berdasarkan pola nutrisi. Model yang dibangun kemudian dievaluasi menggunakan metrik kinerja seperti akurasi, presisi, recall, dan F1-score untuk menilai efektivitas masing-masing algoritma dalam mengklasifikasikan data. Terakhir, hasil dari evaluasi model akan diinterpretasikan untuk memberikan wawasan tentang algoritma mana yang lebih efektif dalam memprediksi penyakit berdasarkan pola nutrisi, serta implikasi dari temuan tersebut untuk praktik kesehatan dan penelitian lebih lanjut.

2. Pengumpulan Data

Dataset yang digunakan mencakup berbagai atribut, termasuk usia, jenis kelamin, tinggi badan, berat badan, tingkat aktivitas, preferensi makanan, serta informasi nutrisi seperti kalori, protein, gula, sodium, karbohidrat, dan serat. Setelah pengumpulan data, langkah selanjutnya adalah preprocessing data, yang meliputi pembersihan data untuk menghapus entri yang tidak valid, penanganan nilai hilang dengan metode imputasi, dan normalisasi variabel numerik agar berada dalam skala yang sama. Imputasi digunakan untuk menangani nilai hilang dalam dataset yang dapat mempengaruhi akurasi model klasifikasi. Metode yang umum digunakan dalam penelitian ini adalah imputasi rata-rata, di mana nilai hilang pada variabel numerik diisi dengan rata-rata dari nilai yang ada. Untuk variabel kategorikal, metode imputasi yang digunakan adalah modus, yaitu nilai yang paling sering muncul dalam dataset. Dengan cara ini, data yang hilang dapat diatasi tanpa mengurangi ukuran dataset secara signifikan, sehingga model dapat dilatih dengan data yang lebih lengkap. Setelah proses imputasi, langkah selanjutnya adalah normalisasi variabel numerik agar berada dalam skala yang sama. Normalisasi penting untuk memastikan bahwa semua variabel memiliki kontribusi yang seimbang dalam proses pelatihan model.

Setelah data siap, kedua algoritma klasifikasi diterapkan. Random Forest digunakan sebagai metode ensemble yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko overfitting. Di sisi lain, Naïve Bayes Classifier diterapkan untuk memanfaatkan asumsi independensi antar atribut, yang membuatnya cepat dan efisien dalam memproses data. Untuk mengevaluasi kinerja model, digunakan matriks kinerja seperti akurasi, presisi, recall, dan F1-score. Hasil dari kedua algoritma dibandingkan untuk

menentukan mana yang lebih efektif dalam memprediksi penyakit berdasarkan pola nutrisi. Dengan pendekatan ini, penelitian diharapkan dapat memberikan wawasan yang lebih dalam tentang hubungan antara pola nutrisi dan risiko penyakit serta efektivitas metode klasifikasi yang digunakan.

3. Preprocessing Data

Proses preprocessing data meliputi:

a. Pembersihan Data

Pada tahap ini, data yang tidak relevan atau tidak konsisten dihapus. Proses ini mencakup penghapusan data duplikat yang dapat menyebabkan bias dalam analisis. Selain itu, nilai yang hilang (missing values) ditangani dengan menggunakan metode imputasi yang sesuai, seperti imputasi rata-rata untuk variabel numerik. Outlier juga diidentifikasi dan ditangani, baik dengan menghapusnya atau dengan menggunakan teknik transformasi untuk mengurangi dampaknya terhadap model.

b. Normalisasi Data

Atribut numerik, seperti usia, tinggi badan, berat badan, kalori, dan komponen nutrisi lainnya, dinormalisasi menggunakan metode Min-Max Scaling. Metode ini mengubah skala variabel sehingga semua atribut berada dalam rentang yang sama, biasanya antara 0 dan 1. Normalisasi ini penting untuk meningkatkan kinerja algoritma klasifikasi, karena banyak algoritma sensitif terhadap skala data.

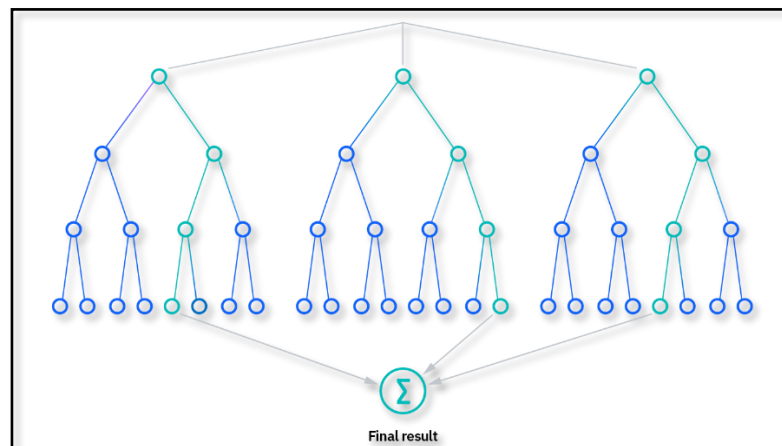
c. Pengkodean Kategorikal

Data kategorikal, seperti jenis kelamin dan preferensi makanan, dikodekan menggunakan metode One-Hot Encoding. Metode ini mengubah setiap kategori menjadi variabel biner terpisah, sehingga memungkinkan algoritma klasifikasi untuk memproses data kategorikal dengan lebih efektif. Dengan cara ini, setiap kategori diwakili oleh kolom baru yang berisi nilai 1 (jika kategori tersebut ada) atau 0 (jika tidak ada), sehingga menghindari masalah yang mungkin timbul dari pengkodean ordinal.

4. Metode Klasifikasi

Penelitian ini menggunakan dua algoritma klasifikasi:

- Trees random forest*, adalah algoritma machine learning yang menggabungkan beberapa decision tree untuk mencapai hasil akhir. Algoritma ini digunakan untuk mengklasifikasikan data set dalam jumlah besar.



Gambar 1. Model *Random Forest*

Sumber gambar : <https://www.ibm.com/id>

- Naive bayes*, adalah algoritma pembelajaran mesin yang digunakan untuk mengklasifikasikan objek berdasarkan probabilitasnya untuk termasuk dalam kelompok tertentu.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Dengan :

P(A|B) = Probabilitas bersyarat A yang diberikan oleh B

P(B|A) = Probabilitas bersyarat B yang diberikan oleh A

P(A) = Probabilitas kejadian A

P(B) = Probabilitas kejadian B

5. Evaluasi Model

Model dievaluasi berdasarkan beberapa metrik berikut:

- a. Akurasi persentase prediksi yang benar.
- b. Precision dan recall kemampuan model dalam memprediksi kelas target dengan benar.
- c. F1-Score rata-rata harmonis dari precision dan recall, untuk mengatasi ketidakseimbangan kelas.

6. Implementasi dan Validasi

- a. Implementasi menggunakan algoritma Trees Random Forest dan Naive Bayes Updateable untuk klasifikasi dataset.
- b. Validasi model menggunakan metode Cross-Validation (k-fold) untuk memastikan generalisasi model terhadap dataset baru. Penelitian ini menggunakan 10-fold cross-validation.

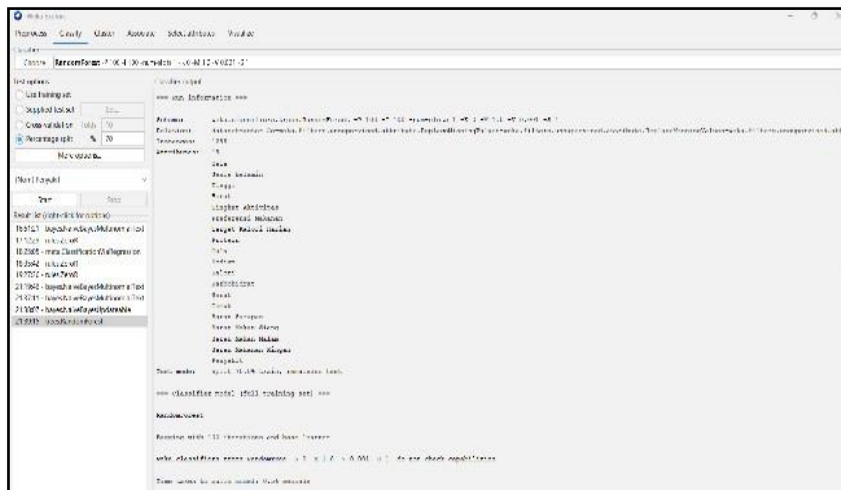
7. Alur Penelitian

- a. Pengumpulan dataset.
- b. Preprocessing data.
- c. Pembagian dataset menjadi data latih (80%) dan data uji (20%).
- d. Pembangunan model menggunakan algoritma Trees Random Forest dan Naive Bayes Updateable.
- e. Evaluasi performa model berdasarkan data uji.
- f. Interpretasi hasil dan analisis perbandingan algoritma.

8. Alat dan Perangkat Lunak

Software yang digunakan adalah menggunakan Weka untuk klasifikasi dataset yang akan diuji.

IV. HASIL DAN PEMBAHASAN



Gambar 2. Uji Algoritma *Random Forest*

Grafik yang menunjukkan performa kedua algoritma dalam menangkap data kelas positif dan menghindari kesalahan.

```

    === Evaluation on test split ===

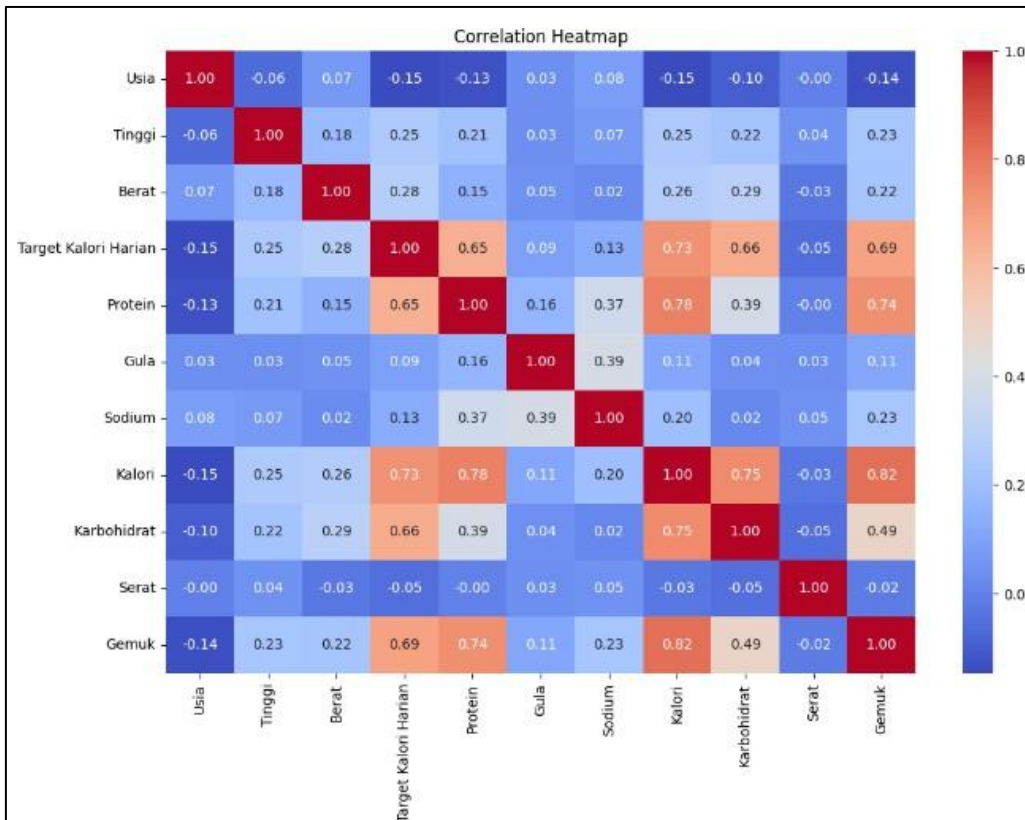
    Time taken to test model on test split: 0.01 seconds

    === Summary ===

    Correctly Classified Instances      465      91.3556 %
    Incorrectly Classified Instances    44       8.6444 %
    Kappa statistic                    0.7705
    Mean absolute error                 0.0117
    Root mean squared error            0.0942
    Relative absolute error            23.9941 %
    Root relative squared error        62.9831 %
    Total Number of Instances          509
    
```

Gambar 9. Perbandingan Waktu Eksekusi Algoritma

Gambar 9. menunjukkan waktu yang dibutuhkan oleh masing-masing model untuk melatih dan menguji dataset.



Gambar 10. Matriks Perbandingan Waktu Eksekusi

Hasil penelitian menunjukkan bahwa Random Forest unggul dalam memprediksi penyakit berdasarkan pola nutrisi harian. Hal ini disebabkan oleh kemampuan algoritma ini dalam menangkap hubungan kompleks antar fitur tanpa overfitting. Sebaliknya, Naive Bayes yang mengasumsikan independensi antar fitur kurang efektif untuk dataset nutrisi yang memiliki hubungan antar variabel yang kompleks.

Faktor-faktor seperti asupan kalori, keseimbangan makronutrien, dan preferensi makanan ditemukan memiliki pengaruh signifikan terhadap risiko penyakit tertentu. Misalnya, asupan kalori yang melebihi target harian cenderung meningkatkan risiko obesitas, sedangkan rendahnya konsumsi serat berkaitan dengan risiko gangguan pencernaan. Analisis lebih lanjut mengungkapkan bahwa preprocessing data, terutama normalisasi dan seleksi fitur, memainkan peran penting dalam meningkatkan performa model. Metode Feature Importance pada Random Forest menunjukkan bahwa atribut kalori, protein, dan gula memiliki kontribusi terbesar dalam prediksi penyakit.

V. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa algoritma Random Forest dan Naive Bayes dapat digunakan untuk memprediksi penyakit berdasarkan pola nutrisi harian. Dari hasil evaluasi, Random Forest terbukti lebih unggul dalam hal akurasi, precision, recall, dan F1-score dibandingkan Naive Bayes. Hal ini menunjukkan potensi besar penggunaan algoritma Random bahwa asupan nutrisi seperti kalori, protein, dan gula memiliki pengaruh signifikan terhadap risiko penyakit. Preprocessing data, terutama normalisasi dan seleksi fitur, penting untuk meningkatkan kinerja model. Random Forest lebih cocok untuk analisis data nutrisi yang kompleks dibandingkan Naive Bayes. Penelitian ini memberikan kontribusi dalam pengembangan sistem prediksi penyakit berbasis data nutrisi. Untuk penelitian selanjutnya, disarankan menggunakan dataset yang lebih besar dan algoritma yang lebih canggih, seperti Deep Learning, untuk meningkatkan akurasi prediksi dan generalisasi model.

VII. REFERENSI

- Ahmad, T., et al. 2020. "Comparative Analysis of Random Forest and Other Classification Algorithms in Healthcare Data." *Journal of Medical Informatics*.
- Arfyanti, Ita, and Tommy Bustomi. 2024. "Perbandingan Kinerja Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Darah Tinggi." 6(3): 1987–94. doi:10.47065/bits.v6i3.6477.
- Breiman, L. 2001. *Random Forests*.
- Cortes, C., & Vapnik, V. 1995. "Support- Vector Networks." *Machine Learning* 3(20): 273–97.
- Gupta, A., et al. 2019. "Handling Class Imbalance in Healthcare Data Using Machine Learning Algorithms." *Journal of Biomedical Informatics* (98): 103–12.
- Johnson, R. K., et al. 2018. "Dietary Sugars Intake and Cardiovascular Health: A Scientific Statement from the American Heart Association." *Circulation* 11(120): 1011–20.
- Khan, M. A., & Khan. 2022. "A Comparative Study of Machine Learning Algorithms for Disease Prediction." *Journal of Healthcare Engineering*.
- Kim, J., et al. 2020. "Predicting Daily Caloric Intake Using Machine Learning Algorithms." *Journal of Nutrition and Metabolism*.
- Nomira, Syarifah Rozati, Studi Kesehatan Masyarakat, Fakultas Ilmu-ilmu Kesehatan, Universitas Abulyatama, and Aceh Besar. 2024. "Relevansi Gizi Dan Kesehatan; Relevance of Nutrition And Health." *Public Health Science Journal*: 1–16.
- Rofiani, Riska, Luluk Oktaviani, Dwi Vernanda, and Tri Hendriawan. 2024. "Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5." *Jurnal Tekno Kompak* 18(1): 126. doi:10.33365/jtk.v18i1.3525.
- Sari, D. P., & Prabowo. 2023. "Application of Machine Learning in Nutritional Studies." *Journal of Health Sciences*.
- Zhang, Y., et al. 2021. "Data Preprocessing Techniques for Improving the Performance of Machine Learning Models in Healthcare." *Journal of Healthcare Informatics Research* 2(5): 123–45.