

# Analisis Klasifikasi Tingkat Hidrasi Manusia Menggunakan Algoritma Machine Learning Berbasis Dataset Asupan Air Harian

<sup>1</sup>Arisman, <sup>2</sup>Nurhayati, <sup>3</sup>Tanti

<sup>1,2,3</sup>Universitas mikroskil, Indonesia

<sup>1</sup>arisman@mikroskil.ac.id, <sup>2</sup>nurhayati@mikroskil.ac.id, <sup>3</sup>tanti@mikroskil.ac.id

Submit : 25 Des 2025 | Diterima : 18 Jan 2026 | Terbit : 20 Jan 2026

## ABSTRAK

Kecukupan asupan air harian merupakan faktor penting dalam menjaga keseimbangan fisiologis tubuh manusia. Kekurangan cairan dapat menyebabkan gangguan kesehatan yang berdampak pada penurunan kinerja fisik maupun kognitif. Seiring berkembangnya bidang data science, pendekatan machine learning dapat dimanfaatkan untuk menganalisis dan memprediksi tingkat hidrasi seseorang berdasarkan karakteristik individu dan lingkungan. Penelitian ini bertujuan untuk mengklasifikasikan tingkat hidrasi manusia ke dalam kategori Good dan Poor menggunakan beberapa algoritma machine learning, yaitu Naive Bayes, K-Nearest Neighbor (KNN), dan Random Forest. Dataset yang digunakan merupakan dataset publik Daily Water Intake yang mencakup atribut usia, jenis kelamin, berat badan, tingkat aktivitas fisik, kondisi cuaca, serta jumlah asupan air harian. Tahapan penelitian meliputi proses preprocessing data, pelatihan model, dan evaluasi kinerja menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dalam mengklasifikasikan tingkat hidrasi dibandingkan algoritma lainnya. Penelitian ini diharapkan dapat menjadi referensi dalam penerapan machine learning untuk analisis kesehatan berbasis data.

**Kata Kunci:** Data Science, Machine Learning, Klasifikasi, Tingkat Hidrasi, Asupan Air

## PENDAHULUAN

Air merupakan komponen esensial bagi tubuh manusia dan berperan penting dalam menjaga keseimbangan fisiologis, termasuk proses metabolisme, pengaturan suhu tubuh, serta fungsi kognitif dan fisik. Kekurangan asupan air atau kondisi dehidrasi dapat berdampak negatif terhadap kesehatan, seperti menurunnya konsentrasi, kelelahan, hingga gangguan fungsi organ. Oleh karena itu, pemantauan dan evaluasi tingkat hidrasi menjadi aspek penting dalam upaya menjaga kesehatan individu.

Pada praktiknya, penentuan tingkat hidrasi sering kali didasarkan pada rekomendasi umum atau penilaian subjektif, tanpa mempertimbangkan perbedaan karakteristik individu dan kondisi lingkungan. Faktor-faktor seperti usia, berat badan, tingkat aktivitas fisik, dan kondisi cuaca memiliki pengaruh signifikan terhadap kebutuhan asupan air harian seseorang. Pendekatan konvensional yang tidak mempertimbangkan variasi faktor tersebut berpotensi menghasilkan penilaian tingkat hidrasi yang kurang akurat.

Perkembangan teknologi informasi dan meningkatnya ketersediaan data telah mendorong pemanfaatan data science dalam berbagai bidang, termasuk kesehatan. Data science merupakan bidang multidisipliner yang mengombinasikan statistika, matematika, dan ilmu komputer untuk mengekstraksi pengetahuan dari data dalam jumlah besar (Han et al., 2012). Salah satu pendekatan utama dalam data science adalah machine learning, yang memungkinkan sistem untuk mempelajari pola dari data historis dan menghasilkan prediksi atau klasifikasi secara otomatis (Mitchell, 1997).

Machine learning, khususnya metode *supervised learning*, telah banyak diterapkan dalam permasalahan klasifikasi di bidang kesehatan, seperti prediksi kondisi medis dan analisis perilaku kesehatan (Kotsiantis et al., 2007). Algoritma klasifikasi mampu mengelompokkan data ke dalam kelas tertentu berdasarkan atribut yang dimiliki, sehingga dapat digunakan untuk memprediksi kondisi atau status kesehatan seseorang secara lebih objektif.

Beberapa algoritma klasifikasi yang umum digunakan dalam penelitian data science antara lain Naive Bayes, K-Nearest Neighbor (KNN), dan Random Forest. Naive Bayes bekerja berdasarkan prinsip

probabilitas dengan asumsi independensi antar atribut (Mitchell, 1997). KNN melakukan klasifikasi berdasarkan kedekatan jarak antar data dalam ruang fitur (Cover & Hart, 1967). Sementara itu, Random Forest merupakan metode ensemble yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko overfitting (Breiman, 2001). Metode ensemble seperti Random Forest dilaporkan memiliki kinerja yang lebih stabil pada permasalahan klasifikasi dengan data yang kompleks (Witten et al., 2016).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan dan mengevaluasi algoritma machine learning dalam mengklasifikasikan tingkat hidrasi manusia menggunakan dataset publik *Daily Water Intake*. Penelitian ini membandingkan kinerja algoritma Naive Bayes, K-Nearest Neighbor, dan Random Forest berdasarkan metrik evaluasi accuracy, precision, recall, F1-score, serta confusion matrix. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan pendekatan data science berbasis machine learning di bidang kesehatan, khususnya sebagai dasar dalam pembuatan sistem pendukung keputusan untuk pemantauan tingkat hidrasi manusia.

## TINJAUAN PUSTAKA

### Data Science dalam Bidang Kesehatan

Data science merupakan bidang multidisipliner yang mengintegrasikan statistika, matematika, dan ilmu komputer untuk mengekstraksi pengetahuan bermakna dari data dalam jumlah besar (Han et al., 2012). Pendekatan data science mencakup tahapan pengumpulan data, pembersihan data, eksplorasi data, pemodelan, serta evaluasi hasil guna mendukung pengambilan keputusan berbasis data (Witten et al., 2016). Dalam bidang kesehatan, data science telah banyak dimanfaatkan untuk menganalisis pola perilaku kesehatan, memprediksi kondisi medis, serta mengembangkan sistem pendukung keputusan. Pendekatan ini memungkinkan pemanfaatan data kesehatan secara lebih objektif dan personal dibandingkan metode konvensional yang bersifat umum dan subjektif (Kotsiantis et al., 2007).

### Machine Learning dan Supervised Learning

Machine learning merupakan cabang dari kecerdasan buatan yang memungkinkan sistem komputer untuk belajar dari data dan meningkatkan kinerjanya tanpa pemrograman eksplisit (Mitchell, 1997). Berdasarkan cara pembelajarannya, machine learning diklasifikasikan menjadi supervised learning, unsupervised learning, dan reinforcement learning (Shalev-Shwartz & Ben-David, 2014). Supervised learning merupakan pendekatan yang paling umum digunakan dalam permasalahan klasifikasi, di mana model dilatih menggunakan data berlabel untuk mempelajari hubungan antara atribut input dan kelas target. Pendekatan ini banyak diterapkan pada studi kesehatan karena mampu menghasilkan prediksi yang terukur dan dapat dievaluasi secara kuantitatif (Kotsiantis et al., 2007).

### Pra-pemrosesan Data dan Rekayasa Fitur

Keberhasilan model machine learning sangat dipengaruhi oleh kualitas data dan proses pra-pemrosesan. Pra-pemrosesan data bertujuan untuk mengatasi permasalahan seperti nilai hilang, perbedaan skala data, serta representasi variabel kategorikal agar dapat diproses oleh algoritma machine learning (Han et al., 2012). Teknik seperti encoding variabel kategorikal, normalisasi atau standardisasi data numerik, serta pemilihan fitur yang relevan terbukti mampu meningkatkan kinerja dan stabilitas model klasifikasi. Rekayasa fitur juga berperan penting dalam membantu model menangkap pola data yang lebih representatif, khususnya pada dataset kesehatan yang bersifat heterogen (Witten et al., 2016).

### Algoritma Naive Bayes

Naive Bayes merupakan algoritma klasifikasi probabilistik yang didasarkan pada teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen satu sama lain (Mitchell, 1997). Keunggulan utama Naive Bayes terletak pada kesederhanaan, efisiensi komputasi, serta kemampuannya dalam menangani dataset dengan dimensi tinggi. Namun, pada data dunia nyata, khususnya data kesehatan, asumsi independensi antar atribut sering kali tidak terpenuhi. Ketergantungan antar fitur dapat menyebabkan penurunan kinerja Naive Bayes dalam memodelkan hubungan kompleks pada data (Kotsiantis et al., 2007).

---

### Algoritma K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan metode klasifikasi non-parametrik yang menentukan kelas suatu data berdasarkan mayoritas kelas dari sejumlah tetangga terdekatnya dalam ruang fitur (Cover & Hart, 1967). Algoritma ini memiliki konsep yang sederhana dan mudah diimplementasikan. Meskipun demikian, kinerja KNN sangat dipengaruhi oleh pemilihan parameter  $k$ , metrik jarak, serta skala data. Oleh karena itu, proses normalisasi data menjadi langkah penting untuk memastikan hasil klasifikasi yang optimal (Shalev-Shwartz & Ben-David, 2014).

### Algoritma Random Forest

Random Forest merupakan algoritma ensemble yang menggabungkan banyak pohon keputusan untuk menghasilkan prediksi akhir melalui mekanisme pemungutan suara (*majority voting*) (Breiman, 2001). Pendekatan ensemble ini bertujuan untuk meningkatkan akurasi dan mengurangi risiko *overfitting* yang sering terjadi pada model pohon keputusan tunggal. Random Forest dikenal memiliki kinerja yang stabil pada dataset dengan karakteristik kompleks dan kombinasi fitur numerik serta kategorikal. Selain itu, algoritma ini juga menyediakan informasi mengenai tingkat kepentingan fitur (*feature importance*), yang bermanfaat dalam interpretasi model, khususnya pada konteks kesehatan (Witten et al., 2016).

### Evaluasi Model Klasifikasi

Evaluasi kinerja model klasifikasi merupakan tahap penting untuk menilai efektivitas algoritma yang digunakan. Metrik evaluasi yang umum digunakan antara lain accuracy, precision, recall, dan F1-score. Akurasi mengukur ketepatan prediksi secara keseluruhan, sedangkan precision dan recall memberikan gambaran yang lebih rinci terkait kesalahan prediksi pada masing-masing kelas (Shalev-Shwartz & Ben-David, 2014). F1-score digunakan untuk menyeimbangkan nilai precision dan recall, terutama pada kondisi data yang tidak seimbang. Selain itu, confusion matrix digunakan untuk menganalisis distribusi kesalahan klasifikasi, seperti *false positive* dan *false negative*, yang memiliki implikasi penting dalam aplikasi kesehatan (Pedregosa et al., 2011).

### Machine Learning untuk Analisis dan Prediksi Hidrasi

Penelitian terkini menunjukkan bahwa pendekatan machine learning dapat digunakan untuk menganalisis dan memprediksi status hidrasi manusia dengan memanfaatkan data fisiologis, aktivitas fisik, serta kondisi lingkungan. Model berbasis machine learning mampu memberikan rekomendasi hidrasi yang lebih personal dibandingkan pendekatan konvensional yang bersifat umum (Wang et al., 2022). Selain itu, studi mengenai rekomendasi asupan air harian berbasis data menunjukkan bahwa model prediktif yang mempertimbangkan karakteristik individu dan lingkungan dapat meningkatkan akurasi penilaian hidrasi (Dolci et al., 2022). Temuan ini memperkuat relevansi penggunaan algoritma machine learning dalam penelitian klasifikasi tingkat hidrasi.

## METODE PENELITIAN

Metodologi penelitian ini disusun untuk menjelaskan tahapan-tahapan yang dilakukan dalam proses klasifikasi tingkat hidrasi manusia menggunakan algoritma machine learning. Alur penelitian dimulai dari pengumpulan dataset, preprocessing data, pemodelan, hingga evaluasi kinerja model.

### Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan dataset publik *Daily Water Intake* yang diperoleh dari repositori data terbuka. Dataset ini berisi data individu terkait karakteristik fisik, aktivitas, kondisi lingkungan, serta tingkat hidrasi harian. Atribut yang terdapat dalam dataset terdiri dari usia (*Age*), jenis kelamin (*Gender*), berat badan (*Weight*), tingkat aktivitas fisik (*Physical Activity Level*), kondisi cuaca (*Weather*), jumlah asupan air harian (*Daily Water Intake*), dan tingkat hidrasi (*Hydration Level*).

Atribut *Hydration Level* digunakan sebagai variabel target atau kelas, yang terdiri dari dua kategori, yaitu *Good* dan *Poor*. Sementara itu, atribut lainnya digunakan sebagai variabel independen yang memengaruhi kondisi hidrasi individu.

## Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pemodelan. Beberapa langkah preprocessing yang dilakukan meliputi:

1. Pemeriksaan Data  
Data diperiksa untuk memastikan tidak terdapat nilai kosong (*missing values*) dan inkonsistensi data. Apabila ditemukan data yang tidak valid, maka dilakukan penyesuaian atau pembersihan data.
2. Transformasi Data Kategorikal  
Atribut kategorikal seperti *Gender*, *Physical Activity Level*, dan *Weather* diubah ke dalam bentuk numerik menggunakan teknik *encoding* agar dapat diproses oleh algoritma machine learning.
3. Normalisasi Data  
Atribut numerik seperti usia, berat badan, dan jumlah asupan air harian dinormalisasi untuk menyamakan skala data sehingga tidak terjadi dominasi atribut tertentu terhadap proses pembelajaran model.
4. Pembagian Data  
Dataset dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*), dengan perbandingan 80% untuk data latih dan 20% untuk data uji. Pembagian ini bertujuan untuk menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

## Pemodelan Machine Learning

Pada tahap pemodelan, dilakukan penerapan beberapa algoritma klasifikasi untuk memprediksi tingkat hidrasi manusia. Algoritma yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Naive Bayes  
Algoritma Naive Bayes digunakan karena kesederhanaannya serta kemampuannya dalam menangani data dengan dimensi yang relatif kecil. Algoritma ini bekerja berdasarkan teorema Bayes dengan asumsi independensi antar atribut.
2. K-Nearest Neighbor (KNN)  
KNN melakukan klasifikasi berdasarkan kedekatan jarak antara data uji dan data latih. Nilai  $k$  ditentukan melalui pengujian untuk memperoleh hasil klasifikasi terbaik.
3. Random Forest  
Random Forest merupakan algoritma ensemble yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko *overfitting*. Algoritma ini dipilih karena kemampuannya dalam menangani data dengan karakteristik kompleks dan hubungan non-linear. Setiap model dilatih menggunakan data latih yang sama untuk memastikan perbandingan kinerja yang adil antar algoritma.

## Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan beberapa metrik evaluasi yang umum digunakan dalam permasalahan klasifikasi, yaitu:

1. Akurasi, untuk mengukur tingkat ketepatan model dalam mengklasifikasikan data
2. Precision, untuk mengukur tingkat ketepatan prediksi kelas positif
3. Recall, untuk mengukur kemampuan model dalam mengenali kelas positif
4. F1-Score, sebagai rata-rata harmonis antara precision dan recall
5. Confusion Matrix, untuk memberikan gambaran detail mengenai hasil klasifikasi model

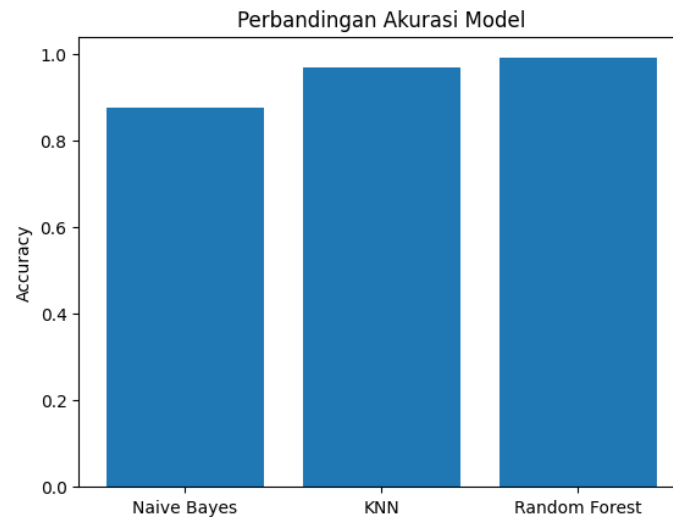
Hasil evaluasi dari masing-masing algoritma kemudian dibandingkan untuk menentukan algoritma dengan performa terbaik dalam mengklasifikasikan tingkat hidrasi manusia.

## HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengujian model klasifikasi tingkat hidrasi manusia yang diperoleh melalui eksperimen menggunakan dataset *Daily Water Intake*. Seluruh proses pengujian dilakukan sesuai dengan tahapan metodologi penelitian yang telah dijelaskan sebelumnya, mulai dari preprocessing data, pelatihan model, hingga evaluasi kinerja. Eksperimen dilaksanakan dengan membagi dataset ke dalam data latih dan data uji, sehingga kinerja model dapat diuji terhadap data yang belum pernah digunakan dalam proses pelatihan.

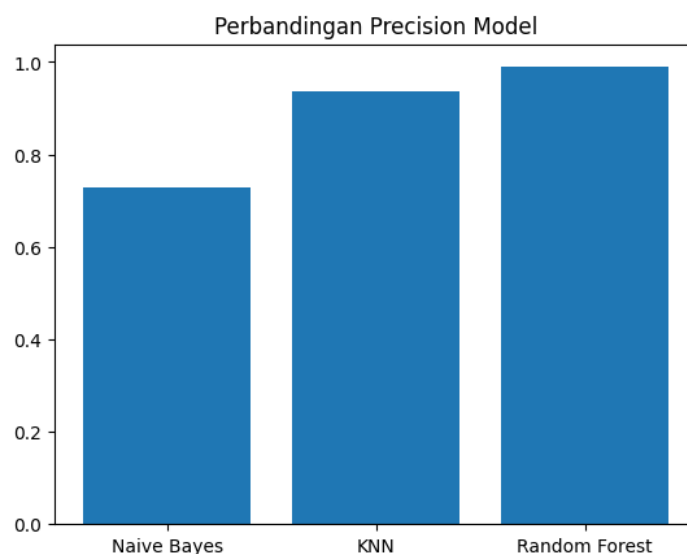
Pengujian dilakukan terhadap tiga algoritma klasifikasi, yaitu Naive Bayes, K-Nearest Neighbor (KNN), dan Random Forest. Setiap algoritma dilatih menggunakan konfigurasi parameter yang sama serta dataset yang identik, sehingga perbandingan kinerja dapat dilakukan secara objektif dan adil. Evaluasi kinerja model dilakukan menggunakan beberapa metrik evaluasi yang umum digunakan dalam permasalahan klasifikasi, yaitu accuracy, precision, recall, F1-score, serta confusion matrix.

Hasil pengujian yang disajikan pada bagian ini tidak hanya bertujuan untuk menunjukkan tingkat akurasi masing-masing algoritma, tetapi juga untuk memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam mengklasifikasikan tingkat hidrasi secara tepat dan konsisten. Dengan demikian, hasil eksperimen ini dapat digunakan sebagai dasar dalam menentukan algoritma terbaik serta mengevaluasi efektivitas pendekatan machine learning yang diterapkan dalam penelitian ini.



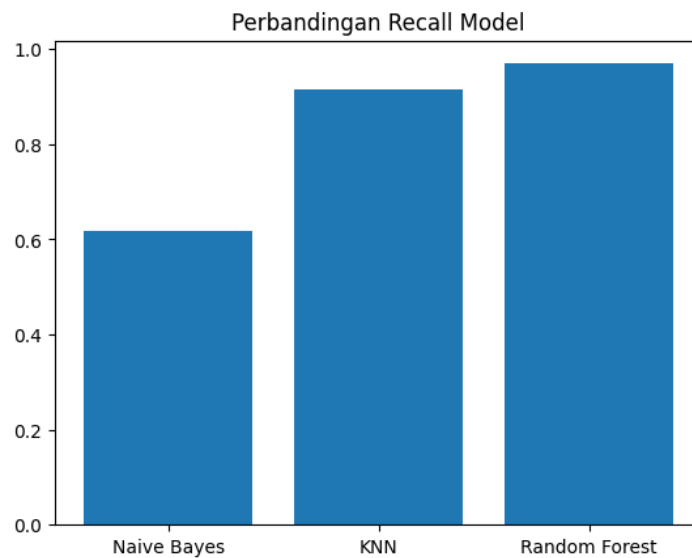
Gambar 1 Perbandingan Nilai Akurasi Dari Algoritma Naive Bayes, K-Nearest Neighbor (KNN), Dan Random Forest

Gambar 1 menunjukkan perbandingan nilai akurasi dari algoritma Naive Bayes, K-Nearest Neighbor (KNN), dan Random Forest. Hasil pengujian menunjukkan bahwa algoritma Random Forest menghasilkan akurasi tertinggi dibandingkan algoritma lainnya. Hal ini mengindikasikan bahwa Random Forest memiliki kemampuan generalisasi yang lebih baik dalam mengklasifikasikan tingkat hidrasi berdasarkan dataset asupan air harian.



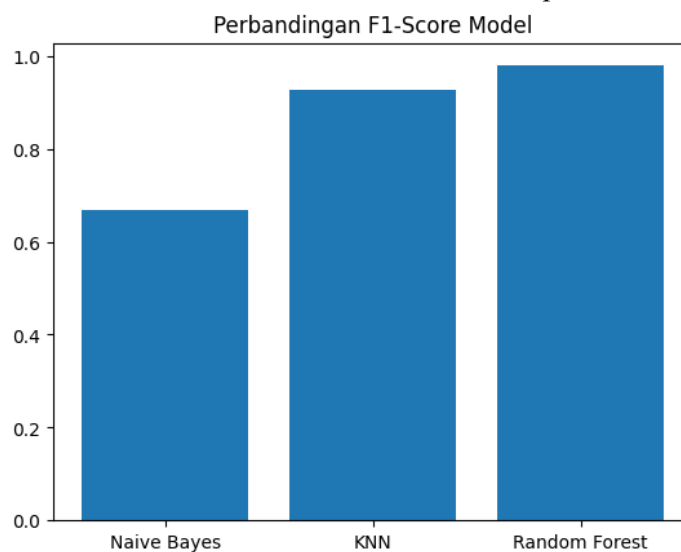
Gambar 2 Perbandingan Precision Model

Berdasarkan Gambar 2, algoritma Random Forest menunjukkan nilai precision yang paling tinggi. Nilai precision yang tinggi menunjukkan bahwa model mampu meminimalkan kesalahan prediksi positif, sehingga lebih andal dalam mengklasifikasikan tingkat hidrasi secara tepat.



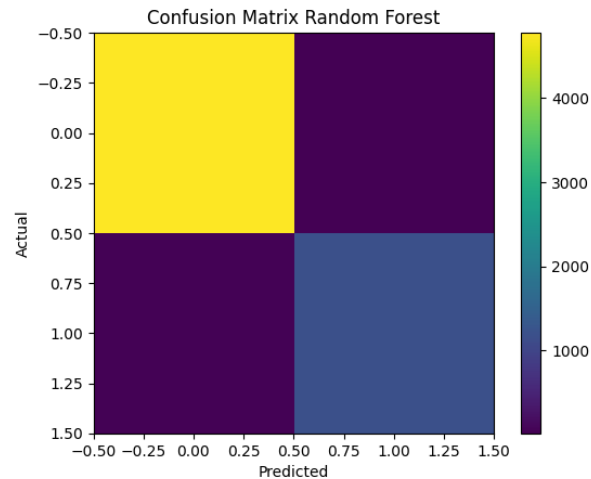
Gambar 3 Perbandingan Nilai Recall

Gambar 3 memperlihatkan perbandingan nilai recall dari masing-masing algoritma. Random Forest memiliki nilai recall tertinggi, yang menunjukkan kemampuannya dalam mengenali sebagian besar data yang benar-benar termasuk ke dalam kelas tertentu, baik *Good* maupun *Poor*.



Gambar 4 Perbandingan Nilai F1-Score

Nilai F1-score yang ditampilkan pada Gambar 4 menunjukkan bahwa algoritma Random Forest memiliki keseimbangan terbaik antara precision dan recall. Hal ini menandakan bahwa algoritma tersebut mampu menghasilkan kinerja klasifikasi yang optimal secara keseluruhan.



Gambar 5 Confusion Matrix

Gambar 5 menunjukkan confusion matrix dari algoritma Random Forest sebagai model terbaik. Terlihat bahwa jumlah prediksi benar lebih dominan dibandingkan kesalahan klasifikasi, baik *false positive* maupun *false negative*. Hal ini menguatkan hasil evaluasi bahwa Random Forest merupakan algoritma paling efektif dalam penelitian ini.

Tabel 1 Hasil Evaluasi Kinerja Algoritma

Algoritma	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0,8765	0,7289	0,6188	0,6693
KNN	0,9707	0,9375	0,9158	0,9265
Random Forest	0,9917	0,9891	0,9695	0,9792

Tabel 1 menunjukkan hasil evaluasi kinerja algoritma Naive Bayes, K-Nearest Neighbor (KNN), dan Random Forest berdasarkan metrik accuracy, precision, recall, dan F1-score. Berdasarkan hasil pengujian, algoritma Random Forest memperoleh nilai tertinggi pada seluruh metrik evaluasi. Hal ini menunjukkan bahwa Random Forest memiliki kemampuan klasifikasi yang paling optimal dalam menentukan tingkat hidrasi manusia berdasarkan dataset asupan air harian. Algoritma KNN menunjukkan performa yang cukup baik, sementara Naive Bayes memiliki kinerja yang relatif lebih rendah dibandingkan algoritma lainnya.

### Pembahasan

Berdasarkan hasil pengujian berbasis eksperimen yang telah dilakukan, diperoleh perbedaan kinerja yang cukup signifikan antara algoritma Naive Bayes, K-Nearest Neighbor (KNN), dan Random Forest dalam mengklasifikasikan tingkat hidrasi manusia. Perbedaan ini menunjukkan bahwa karakteristik algoritma dan kemampuannya dalam memodelkan hubungan antar atribut sangat memengaruhi hasil klasifikasi.

Algoritma Random Forest menunjukkan performa terbaik pada seluruh metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score. Tingginya nilai akurasi dan F1-score mengindikasikan bahwa Random Forest mampu mengklasifikasikan data dengan tingkat kesalahan yang sangat rendah serta menjaga keseimbangan antara ketepatan dan kelengkapan prediksi. Keunggulan ini disebabkan oleh kemampuan Random Forest dalam menangani hubungan non-linear antar atribut serta mekanisme ensemble yang menggabungkan banyak pohon keputusan, sehingga model menjadi lebih stabil dan tidak mudah mengalami overfitting.

Algoritma K-Nearest Neighbor (KNN) menunjukkan performa yang cukup baik dan berada di bawah Random Forest. Nilai precision dan recall yang tinggi pada KNN menunjukkan bahwa algoritma ini mampu mengenali pola data dengan baik berdasarkan kedekatan karakteristik antar sampel. Namun, kinerja

KNN sangat bergantung pada pemilihan parameter  $k$  serta sensitivitas terhadap skala data, sehingga memerlukan proses normalisasi dan penentuan parameter yang tepat agar menghasilkan performa optimal.

Sementara itu, algoritma Naive Bayes menghasilkan performa yang relatif lebih rendah dibandingkan dua algoritma lainnya. Hal ini disebabkan oleh asumsi independensi antar atribut yang digunakan oleh Naive Bayes, yang pada kenyataannya tidak sepenuhnya terpenuhi pada dataset *Daily Water Intake*. Beberapa atribut, seperti tingkat aktivitas fisik, berat badan, dan jumlah asupan air harian, memiliki hubungan yang saling memengaruhi, sehingga pendekatan probabilistik sederhana pada Naive Bayes kurang mampu menangkap kompleksitas pola data secara menyeluruh.

Analisis confusion matrix pada model terbaik, yaitu Random Forest, menunjukkan bahwa jumlah prediksi benar jauh lebih dominan dibandingkan kesalahan klasifikasi. Kesalahan yang masih terjadi umumnya berasal dari data dengan karakteristik yang berada di ambang batas antara kelas *Good* dan *Poor*. Hal ini mengindikasikan bahwa terdapat individu dengan pola asupan air dan aktivitas yang hampir serupa, sehingga sulit dibedakan secara tegas oleh model klasifikasi.

Secara keseluruhan, hasil penelitian ini membuktikan bahwa pendekatan machine learning efektif digunakan untuk mengklasifikasikan tingkat hidrasi manusia berdasarkan data asupan air harian dan karakteristik individu. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa metode ensemble seperti Random Forest cenderung memberikan performa yang lebih unggul pada permasalahan klasifikasi dengan data yang kompleks. Dengan demikian, algoritma Random Forest direkomendasikan sebagai metode yang paling optimal dalam penelitian ini.

### KESIMPULAN

Berdasarkan hasil penelitian dan pengujian berbasis eksperimen yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma machine learning efektif digunakan untuk mengklasifikasikan tingkat hidrasi manusia berdasarkan dataset asupan air harian. Penelitian ini memanfaatkan atribut karakteristik individu dan lingkungan, seperti usia, jenis kelamin, berat badan, tingkat aktivitas fisik, kondisi cuaca, serta jumlah asupan air harian, untuk memprediksi tingkat hidrasi ke dalam kelas *Good* dan *Poor*. Hasil evaluasi menunjukkan bahwa algoritma Random Forest menghasilkan kinerja terbaik dibandingkan algoritma Naive Bayes dan K-Nearest Neighbor (KNN). Random Forest memperoleh nilai tertinggi pada seluruh metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score, serta menghasilkan jumlah kesalahan klasifikasi yang paling rendah berdasarkan analisis confusion matrix. Hal ini menunjukkan bahwa Random Forest mampu memodelkan hubungan antar atribut yang kompleks dan memberikan hasil klasifikasi yang lebih stabil serta akurat. Algoritma KNN menunjukkan performa yang cukup baik, namun kinerjanya sangat dipengaruhi oleh pemilihan parameter dan skala data. Sementara itu, algoritma Naive Bayes memiliki keterbatasan dalam menangani ketergantungan antar atribut, sehingga menghasilkan performa yang lebih rendah pada dataset yang memiliki hubungan antar fitur. Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan data science berbasis machine learning dapat dimanfaatkan sebagai dasar dalam pengembangan sistem pendukung keputusan di bidang kesehatan, khususnya dalam pemantauan dan evaluasi tingkat hidrasi manusia. Untuk penelitian selanjutnya, disarankan untuk menggunakan dataset dengan jumlah dan variasi atribut yang lebih besar, menerapkan teknik optimasi parameter, serta mengeksplorasi algoritma lain guna meningkatkan kinerja model yang dihasilkan.

### REFERENSI

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Dhany, H. W., & Izhari, F. (2025). *Multicategory CNN for prediction sleep quality from feature physiological and behavioral daily*. *Proceedings of International Conference on Islamic Community Studies*, 697–706.
- Dolci, A., et al. (2022). Personalized prediction of optimal water intake in adults: A data-driven approach. *Scientific Reports*, 12, Article 21869. <https://doi.org/10.1038/s41598-022-21869-y>
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

- 
- Ignatia, G. (2024). *Klasifikasi kualitas air yang layak pakai dengan machine learning (KNN & ANN). Jurnal Komputer dan Informatika, 19(2)*, 148–155.
- Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2007). Supervised machine learning: A review of classification techniques. *Informatica, 31(3)*, 249–268.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
- Rohana, T., Novita, H. Y., & Nurlaelasari, E. (2025). *Komparasi algoritma machine learning dalam memprediksi kapasitas produksi air bersih di Indonesia. Jurnal Teknologi dan Terapan, (tahun 2025)*.
- Salmi, M. (2024). Handling imbalanced medical datasets: Review of a decade of methods and applications. *Artificial Intelligence Review*.
- Sangaji, D., & Sutabri, T. (2024). *Optimalisasi prediksi indeks kualitas air di Indonesia dengan menggunakan machine learning melalui pendekatan metode Prophet. Switch: Jurnal Sains dan Teknologi Informasi, 2(6)*, 01–14. <https://doi.org/10.62951/switch.v2i6.277>
- Saputra, A. W., Purnamasari, A. I., & Ali, I. (2024). *Implementasi algoritma Naïve Bayes untuk memprediksi kualitas air yang dapat dikonsumsi. Jurnal Teknologi dan Sistem Komputer, 12(1)*, 45–52.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107298019>
- Wang, S., Lafaye, C., Saubade, M., Besson, C., Margarit-Taule, J. M., Gremaux, V., & Liu, S.-C. (2022). Predicting hydration status using machine learning models from physiological and sweat biomarkers during endurance exercise: A single case study. *IEEE Journal of Biomedical and Health Informatics, 26(9)*, 4725–4732. <https://doi.org/10.1109/JBHI.2022.3186150>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques (4th ed.)*. Morgan Kaufmann
- Zalti, U. R. (2025). *Analisis dan prediksi kelayakan air minum menggunakan algoritma klasifikasi machine learning. Jurnal Ilmiah Kualitas Air Minum (ejurnal UMRI, materi 2025)*.