

Deteksi Keberadaan Manusia dalam Ruangan Berbasis Data Sensor Lingkungan Menggunakan Algoritma K-Nearest Neighbor (KNN)

Nur Shakinah, Adlian Jefiza

Politeknik Negeri Batam, Batam, Indonesia

INFORMASI ARTIKEL	ABSTRAK
Sejarah Artikel: Diterima: Juni 2025 Revisi: Juni 2025 Diterima: Juli 2025 Dipublikasi: Juli 2025 Kata Kunci: K-Nearest Neighbor₁; Deteksi Okupasi; Sensor Lingkungan, Klasifikasi, Machine Learning. *Penulis Korespondensi: nurshakinahhos@gmail.com	Penelitian ini bertujuan untuk menganalisis kemampuan algoritma <i>K-Nearest Neighbor</i> (KNN) dalam mendeteksi keberadaan manusia (okupansi) di dalam suatu ruangan berdasarkan data sensor lingkungan melalui pendekatan klasifikasi biner. Data dikumpulkan dari sensor ruangan yang merekam variabel <i>Temperature</i>, <i>Humidity</i>, <i>Light</i>, <i>CO₂</i>, dan <i>Humidity Ratio</i>, kemudian diolah menggunakan teknik standardisasi fitur (<i>StandardScaler</i>) dan validasi silang (<i>cross-validation</i>) guna memperoleh gambaran mengenai performa model klasifikasi okupansi ruangan. Hasil penelitian menunjukkan bahwa model KNN dengan nilai $K=1$ memberikan akurasi validasi silang tertinggi sebesar 99,06% dan akurasi pada data uji yang sangat tinggi, yang mengindikasikan bahwa variabel <i>Light</i> dan <i>CO₂</i> memiliki korelasi paling kuat terhadap status okupansi. Temuan ini diharapkan dapat memberikan kontribusi terhadap bidang <i>smart building</i> dan efisiensi energi serta menjadi rujukan untuk penelitian selanjutnya. ABSTRACT <i>This study aims to analyze the capability of the K-Nearest Neighbor (KNN) algorithm in detecting human presence (occupancy) in a room based on environmental sensor data through a binary classification approach. Data were collected from room sensors recording Temperature, Humidity, Light, CO₂, and Humidity Ratio, and were processed using feature standardization (StandardScaler) and cross-validation to evaluate the performance of the room occupancy classification model. The findings indicate that the KNN model with $K=1$ achieved the highest cross-validation accuracy of 99.06% and a very high accuracy on the test data, suggesting that Light and CO₂ have the strongest correlation with occupancy status. These results are expected to contribute to the field of smart buildings and energy efficiency and serve as a reference for future research.</i>

PENDAHULUAN

Perkembangan dalam bidang bangunan pintar (*smart building*) dan *Internet of Things* (IoT) menunjukkan bahwa efisiensi energi berbasis deteksi okupansi ruangan semakin menjadi perhatian dalam beberapa tahun terakhir. Kondisi ini dipengaruhi oleh meningkatnya kebutuhan akan sistem otomatisasi pencahayaan, pengaturan suhu, dan ventilasi yang responsif terhadap keberadaan manusia secara *real-time*, sehingga diperlukan kajian lebih mendalam untuk memahami pola hubungan antara data sensor lingkungan dan status okupansi ruangan [1], [2]. Meskipun berbagai penelitian sebelumnya telah membahas deteksi okupansi menggunakan sensor tunggal seperti *Passive Infrared* (PIR) atau kamera, masih terdapat keterbatasan berupa akurasi yang rendah pada kondisi tertentu serta biaya implementasi kamera yang relatif tinggi dan berpotensi melanggar privasi pengguna [3], [4]. Keterbatasan tersebut memunculkan kebutuhan untuk meneliti lebih lanjut mengenai pemanfaatan kombinasi sensor lingkungan non-

intrusif, seperti suhu, kelembapan, intensitas cahaya, dan konsentrasi CO₂, untuk memprediksi okupansi ruangan secara lebih akurat [2], [5].

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis performa *algoritma K-Nearest Neighbor* (KNN) dalam mengklasifikasikan status okupansi ruangan (terisi/tidak terisi) menggunakan pendekatan *supervised learning* berbasis data sensor lingkungan. Algoritma KNN dipilih karena memiliki kompleksitas implementasi yang sederhana, tidak memerlukan proses pelatihan model yang kompleks, serta mampu memberikan performa klasifikasi yang baik pada dataset berdimensi rendah hingga menengah [6]. Hasil penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu pada bidang *machine learning* terapan serta memberikan manfaat praktis bagi pengembang sistem manajemen energi bangunan.

LANDASAN TEORI

Kajian teori ini membahas konsep-konsep utama dan hasil penelitian terdahulu yang menjadi dasar dalam menganalisis deteksi okupansi ruangan berbasis sensor. Secara teoretis, algoritma *K-Nearest Neighbor* (KNN) menjelaskan bahwa klasifikasi suatu data baru dilakukan berdasarkan mayoritas kelas dari K data tetangga terdekat menggunakan perhitungan jarak, yang umumnya adalah *Euclidean Distance*, sehingga algoritma ini mampu mengenali pola hubungan non-linear antara variabel sensor dan status okupansi ruangan [6].

Selain itu, konsep standardisasi fitur (*feature scaling*) memberikan landasan mengenai pentingnya menyamakan skala antarvariabel sebelum proses penghitungan jarak dilakukan. Hal ini menjadi penting karena variabel seperti *Temperature* (°C), *Humidity* (%RH), *Light* (lux), dan *CO₂* (ppm) memiliki satuan dan rentang nilai yang berbeda sehingga dapat memengaruhi hasil klasifikasi apabila tidak dinormalisasi terlebih dahulu [6].

Penelitian-penelitian sebelumnya pada bidang deteksi okupansi menunjukkan bahwa kombinasi sensor lingkungan memiliki performa yang lebih baik dibandingkan penggunaan sensor tunggal. Variabel CO₂, cahaya, dan temperatur merupakan indikator yang paling berpengaruh terhadap keberadaan manusia di dalam ruangan, meskipun kemampuan generalisasi model masih dipengaruhi oleh karakteristik masing-masing ruangan [2], [5]. Selain itu, penelitian lain menunjukkan bahwa penggunaan algoritma KNN pada sistem deteksi okupansi berbasis IoT mampu memberikan akurasi yang tinggi dengan kebutuhan komputasi yang relatif rendah sehingga sesuai diterapkan pada perangkat dengan sumber daya terbatas (*edge device*) [3].

Beberapa studi juga menegaskan bahwa pemilihan nilai K yang optimal melalui proses *cross-validation* berperan penting dalam menyeimbangkan bias dan varians model sehingga mampu meningkatkan performa klasifikasi secara keseluruhan [6]. Berdasarkan keseluruhan teori dan penelitian terdahulu tersebut, penelitian ini mengadopsi kerangka berpikir bahwa kombinasi lima variabel sensor lingkungan dapat digunakan sebagai fitur prediktor yang memadai untuk mengklasifikasikan status okupansi ruangan. Kerangka tersebut selanjutnya digunakan sebagai dasar dalam penyusunan metode penelitian dan analisis hasil pengujian.

METODE PENELITIAN

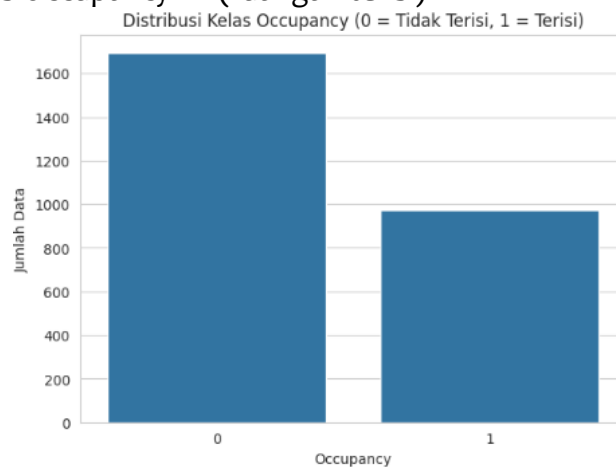
Metodologi penelitian ini dirancang untuk menjelaskan pendekatan, teknik pengumpulan data, serta prosedur analisis yang digunakan dalam mengkaji deteksi okupansi ruangan. Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi *supervised learning* karena dianggap paling sesuai untuk menjawab rumusan masalah mengenai prediksi status okupansi berdasarkan data sensor. Populasi dalam penelitian ini adalah seluruh rekaman data sensor ruangan pada rentang waktu pengamatan, sedangkan sampelnya diperoleh dari berkas *datatest.csv* yang berjumlah 2.665 baris data dengan 8 kolom (id, date, Temperature, Humidity, Light, CO₂, HumidityRatio, dan Occupancy). Data dikumpulkan melalui instrumen sensor lingkungan (suhu, kelembapan, intensitas cahaya, konsentrasi CO₂, dan rasio kelembapan) yang

telah melalui proses pemeriksaan nilai kosong (*missing value*) dan data duplikat untuk memastikan kualitas data, di mana hasil pemeriksaan menunjukkan tidak ditemukan nilai kosong maupun data duplikat pada dataset. Kolom identitas (*id*) dan waktu (*date*) yang tidak relevan secara prediktif kemudian dihapus, sehingga tersisa lima variabel fitur (*Temperature*, *Humidity*, *Light*, *CO₂*, *HumidityRatio*) dan satu variabel target (*Occupancy*). Data selanjutnya dibagi menjadi data latih (80%, 2.132 baris) dan data uji (20%, 533 baris) menggunakan teknik *stratified random sampling* agar proporsi kelas pada data latih dan data uji tetap seimbang. Sebelum pemodelan, seluruh fitur distandardisasi menggunakan *StandardScaler* agar memiliki skala yang setara. Analisis dilakukan menggunakan algoritma K-Nearest Neighbor (KNN) dengan pencarian nilai K optimal pada rentang 1 hingga 20 melalui 5-fold cross-validation berbasis metrik akurasi, guna memperoleh temuan yang objektif dan sesuai dengan tujuan penelitian. Evaluasi akhir model dilakukan menggunakan *confusion matrix* dan metrik akurasi pada data uji. Seluruh prosedur penelitian dilaksanakan secara sistematis menggunakan bahasa pemrograman Python beserta pustaka *scikit-learn*, *pandas*, dan *seaborn* untuk memastikan bahwa hasil yang diperoleh akurat, konsisten, dan dapat dipertanggungjawabkan secara ilmiah.

HASIL DAN PEMBAHASAN

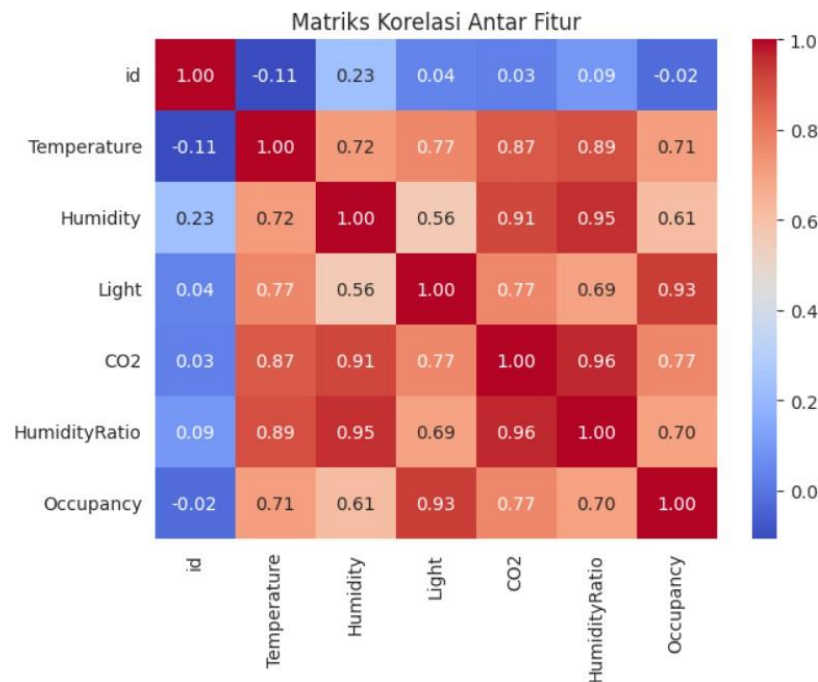
Bab ini menyajikan hasil penelitian yang diperoleh dari proses eksplorasi data, *preprocessing*, pemilihan parameter model, hingga evaluasi performa algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan status okupansi ruangan berdasarkan data sensor lingkungan.

Dataset yang digunakan terdiri atas 2.665 data dengan enam variabel prediktor, yaitu *Temperature*, *Humidity*, *Light*, *CO₂*, *HumidityRatio*, dan *id*, serta satu variabel target yaitu *Occupancy*. Berdasarkan hasil eksplorasi data diketahui bahwa dataset tidak memiliki nilai kosong (*missing value*) maupun data duplikat sehingga seluruh data dapat digunakan pada tahap analisis berikutnya. Selain itu, distribusi kelas menunjukkan adanya ketidakseimbangan (*class imbalance*), yaitu sebanyak 1.693 data termasuk kelas *Occupancy* = 0 (ruangan tidak terisi) dan 972 data termasuk kelas *Occupancy* = 1 (ruangan terisi).



Gambar 1. Distribusi Kelas Occupancy

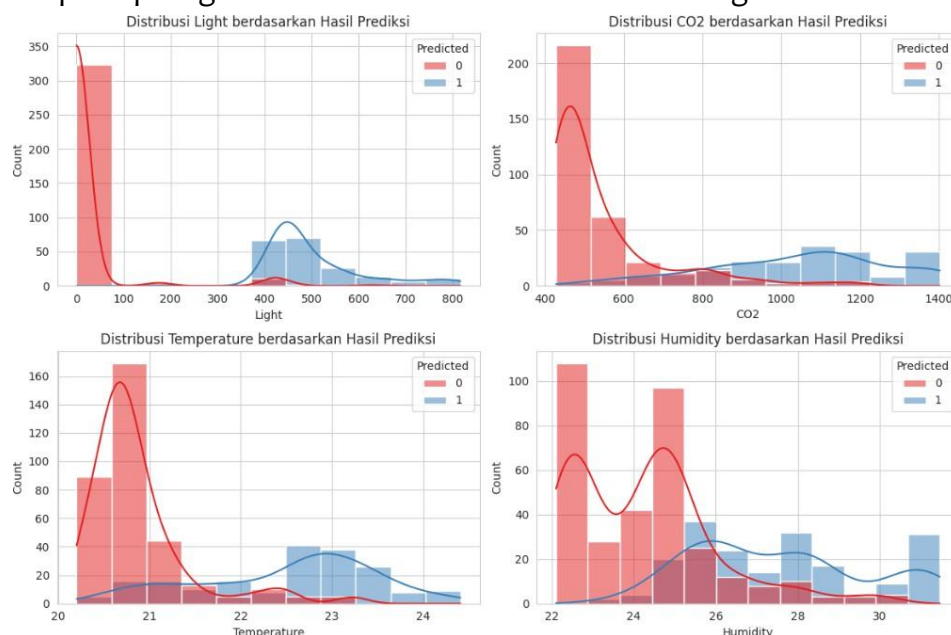
Distribusi kelas tersebut menunjukkan bahwa jumlah data pada kelas ruangan tidak terisi lebih banyak dibandingkan kelas ruangan terisi. Kondisi ini perlu diperhatikan karena dapat memengaruhi proses pelatihan dan evaluasi model klasifikasi apabila tidak dianalisis dengan baik. Selanjutnya dilakukan analisis hubungan antarvariabel menggunakan matriks korelasi untuk mengetahui tingkat pengaruh masing-masing fitur terhadap status okupansi ruangan.



Gambar 2. Heatmap korelasi antar variabel

Berdasarkan Gambar 2, variabel Light memiliki korelasi tertinggi terhadap Occupancy dengan nilai sebesar 0,93, diikuti oleh CO₂ (0,77), Temperature (0,71), HumidityRatio (0,70), dan Humidity (0,61). Sementara itu, variabel id memiliki korelasi yang sangat rendah sehingga tidak memberikan kontribusi yang signifikan terhadap proses klasifikasi. Hasil ini menunjukkan bahwa perubahan intensitas cahaya dan konsentrasi CO₂ merupakan indikator utama keberadaan manusia di dalam ruangan. Temuan tersebut sejalan dengan penelitian sebelumnya yang menyatakan bahwa kombinasi sensor lingkungan mampu meningkatkan akurasi deteksi okupansi dibandingkan penggunaan sensor tunggal.

Untuk memperoleh gambaran distribusi setiap fitur terhadap hasil klasifikasi, dilakukan visualisasi menggunakan histogram. Histogram digunakan untuk melihat penyebaran nilai masing-masing variabel pada setiap kelas hasil prediksi sehingga dapat diketahui fitur yang memiliki kemampuan paling baik dalam membedakan kondisi ruangan terisi dan tidak terisi.

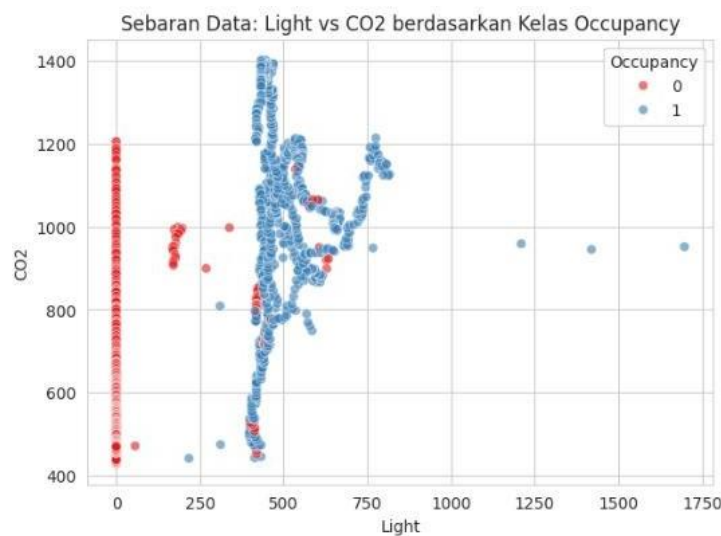


Gambar 3. Histogram Distribusi Fitur Berdasarkan Hasil Prediksi

Berdasarkan Gambar 3, variabel Light menunjukkan pemisahan distribusi yang paling jelas antara kelas Occupancy = 0 dan Occupancy = 1. Data pada kelas Occupancy = 0 didominasi oleh nilai intensitas cahaya yang rendah, sedangkan kelas Occupancy = 1 cenderung memiliki nilai Light yang lebih tinggi. Pola serupa juga terlihat pada variabel CO₂, di mana data pada kelas Occupancy = 1 memiliki konsentrasi CO₂ yang lebih tinggi dibandingkan kelas Occupancy = 0. Hal ini menunjukkan bahwa keberadaan manusia di dalam ruangan berkaitan dengan meningkatnya intensitas pencahayaan dan konsentrasi CO₂ akibat aktivitas manusia.

Sementara itu, variabel Temperature dan Humidity masih menunjukkan adanya tumpang tindih (*overlap*) antara kedua kelas. Meskipun demikian, kedua variabel tersebut tetap memberikan informasi tambahan yang membantu proses klasifikasi ketika dikombinasikan dengan variabel lain. Hasil ini memperkuat temuan pada analisis korelasi bahwa Light dan CO₂ merupakan fitur yang paling berpengaruh dalam mendeteksi status okupansi ruangan.

Selain histogram, dilakukan pula visualisasi sebaran data menggunakan scatter plot antara variabel Light dan CO₂ untuk melihat tingkat pemisahan kedua kelas secara visual.



Gambar 4. Sebaran Data Light dan CO₂ Berdasarkan Kelas Occupancy

Berdasarkan Gambar 4, terlihat bahwa data dengan Occupancy = 0 cenderung terkonsentrasi pada nilai Light yang rendah dengan rentang CO₂ yang relatif lebih kecil. Sebaliknya, data dengan Occupancy = 1 sebagian besar berada pada nilai Light dan CO₂ yang lebih tinggi. Walaupun masih terdapat beberapa titik yang saling berdekatan, secara umum kedua kelas menunjukkan pola pemisahan yang cukup jelas. Kondisi ini menunjukkan bahwa kombinasi variabel Light dan CO₂ mampu membedakan status okupansi dengan baik sehingga mendukung tingginya performa algoritma KNN dalam proses klasifikasi.

Sebelum proses klasifikasi dilakukan, seluruh variabel numerik melalui tahap *preprocessing* menggunakan metode StandardScaler. Standardisasi bertujuan untuk menyamakan skala setiap variabel sehingga proses perhitungan jarak Euclidean pada algoritma KNN tidak dipengaruhi oleh perbedaan satuan maupun rentang nilai antarfitur.

Contoh data sebelum standardisasi:

```
[3.390000e+02 2.254000e+01 2.516000e+01 4.274000e+02 8.536000e+02
 4.249921e-03]
```

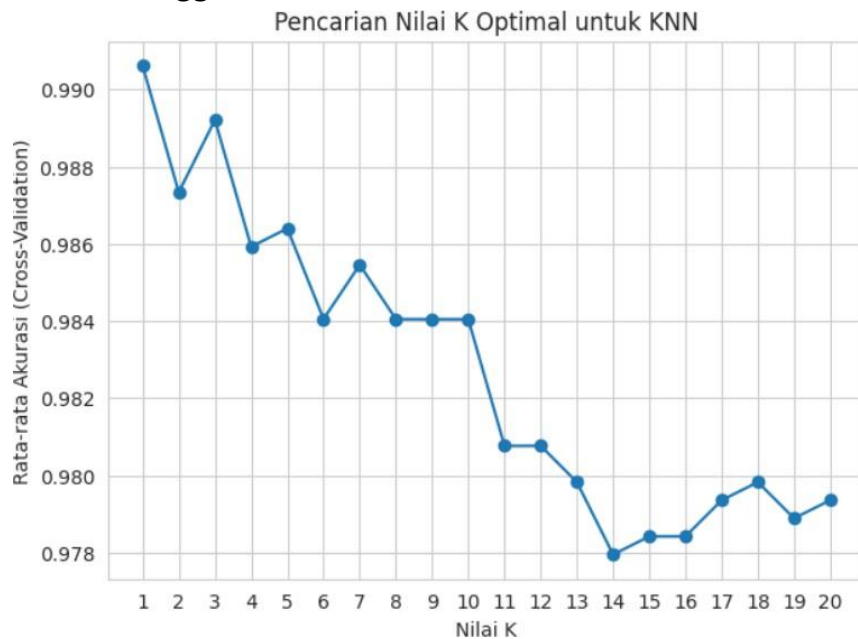
Contoh data setelah standardisasi:

```
[-1.45789905 1.09383554 -0.07246815 0.93928435 0.46989738 0.37406335]
```

Gambar 5. Hasil standardisasi data menggunakan StandardScaler

Berdasarkan hasil standardisasi pada Gambar 5, seluruh fitur telah berada pada skala yang seragam sehingga proses klasifikasi dapat dilakukan secara lebih optimal. Tahapan ini menjadi penting karena algoritma KNN merupakan metode berbasis jarak (*distance-based algorithm*), sehingga performanya sangat dipengaruhi oleh skala data.

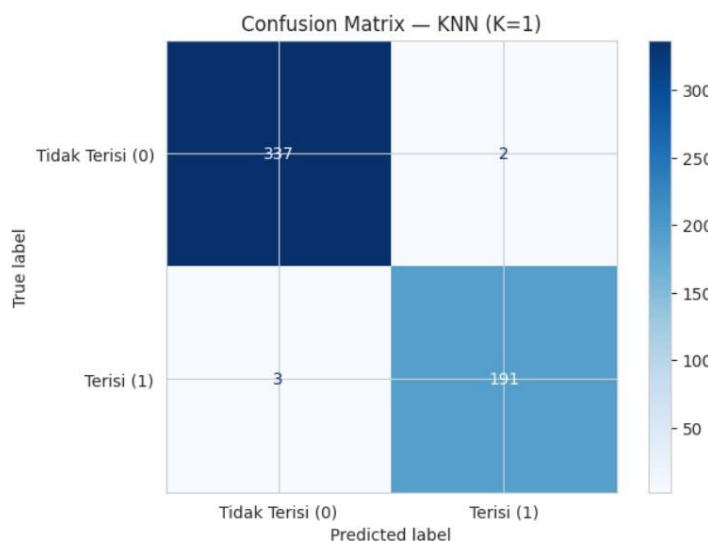
Tahap berikutnya adalah menentukan nilai K terbaik menggunakan metode 5-fold Cross Validation. Pengujian dilakukan pada beberapa nilai K untuk memperoleh parameter yang menghasilkan akurasi tertinggi.



Gambar 6. Grafik akurasi terhadap variasi nilai K

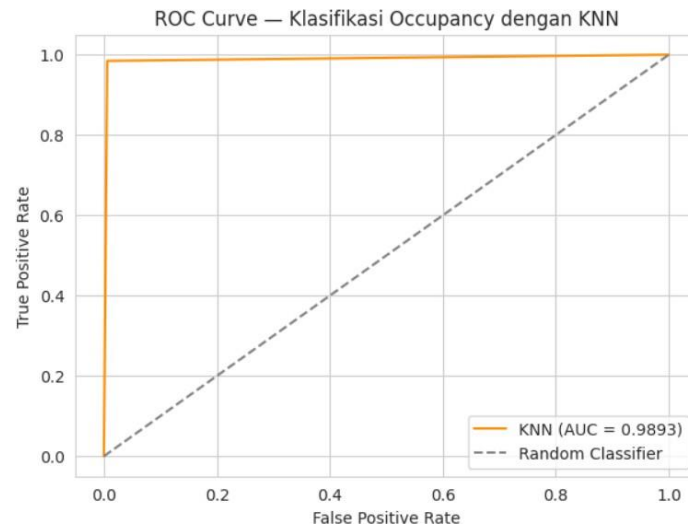
Hasil pengujian menunjukkan bahwa $K = 1$ memberikan rata-rata akurasi tertinggi sebesar 99,06%, sehingga nilai tersebut dipilih sebagai parameter akhir dalam proses klasifikasi. Tingginya akurasi menunjukkan bahwa karakteristik data memiliki pemisahan kelas yang baik sehingga satu tetangga terdekat sudah mampu memberikan keputusan klasifikasi yang optimal.

Setelah parameter terbaik diperoleh, dilakukan proses klasifikasi menggunakan algoritma KNN dan dievaluasi menggunakan Confusion Matrix, ROC Curve, serta metrik evaluasi berupa Accuracy, Precision, Recall, dan F1-score.



Gambar 7. Confusion Matrix hasil klasifikasi KNN

Berdasarkan Confusion Matrix pada Gambar 7, sebagian besar data berhasil diklasifikasikan dengan benar. Jumlah kesalahan klasifikasi (*False Positive* dan *False Negative*) relatif kecil dibandingkan jumlah prediksi yang benar (*True Positive* dan *True Negative*), sehingga menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan kondisi ruangan terisi dan tidak terisi.



Gambar 8. Kurva ROC algoritma KNN

Kurva ROC pada Gambar 8 menunjukkan kemampuan model dalam membedakan kedua kelas secara efektif. Nilai Area Under Curve (AUC) yang mendekati 1 menunjukkan bahwa model memiliki kemampuan diskriminasi yang sangat baik dalam mengidentifikasi status okupansi ruangan.

	Metrik	Nilai
0	Akurasi	0.9906
1	Presisi	0.9896
2	Recall	0.9845
3	F1-Score	0.9871
4	AUC	0.9893

Gambar 9. Hasil Akurasi, Presisi, Recall, dan F-1 score

Hasil ini didukung oleh nilai Accuracy, Precision, Recall, dan F1-score yang tinggi, sehingga menunjukkan bahwa model tidak hanya memiliki tingkat akurasi yang baik, tetapi juga mampu memberikan keseimbangan antara kemampuan mendeteksi kelas positif dan meminimalkan kesalahan klasifikasi.

Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) mampu mengklasifikasikan status okupansi ruangan dengan performa yang sangat baik. Tingginya akurasi model dipengaruhi oleh penggunaan fitur-fitur yang memiliki hubungan kuat terhadap status okupansi, khususnya variabel Light dan CO₂, serta didukung oleh proses *preprocessing* dan pemilihan nilai K yang optimal melalui 5-fold Cross Validation. Temuan ini sejalan dengan penelitian terdahulu yang menyatakan bahwa kombinasi sensor lingkungan non-intrusif dapat dimanfaatkan secara efektif untuk mendeteksi keberadaan manusia di dalam ruangan. Meskipun demikian, penelitian ini masih memiliki keterbatasan berupa distribusi kelas yang tidak seimbang dan pengujian yang hanya dilakukan pada satu dataset. Oleh karena itu, penelitian selanjutnya dapat mengkaji penerapan teknik penanganan *class imbalance* serta melakukan pengujian menggunakan dataset dari lingkungan bangunan yang berbeda agar kemampuan generalisasi model dapat ditingkatkan.

KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa algoritma K-Nearest Neighbor mampu mengklasifikasikan status okupansi ruangan dengan sangat baik berdasarkan data sensor lingkungan, memberikan gambaran yang jelas mengenai hubungan antara variabel Light, CO₂, Temperature, HumidityRatio, dan Humidity terhadap keberadaan manusia dalam ruangan. Berdasarkan hasil analisis, dapat disimpulkan bahwa nilai K=1 merupakan parameter optimal untuk model KNN pada dataset ini dengan akurasi cross-validation sebesar 99,06%, yang menguatkan bahwa fitur pencahayaan dan konsentrasi CO₂ menjadi indikator dominan dalam deteksi okupansi. Penelitian ini juga mengungkap bahwa dataset yang digunakan bersih dari missing value dan duplikat namun memiliki ketidakseimbangan kelas yang perlu diperhatikan pada penelitian lanjutan, sehingga memberikan kontribusi terhadap pengembangan pengetahuan dalam bidang machine learning terapan untuk sistem bangunan pintar. Selain itu, hasil penelitian ini diharapkan dapat menjadi dasar bagi penelitian selanjutnya, khususnya dalam mengkaji penerapan teknik penanganan class imbalance serta pengujian model pada dataset okupansi dari ruangan dan kondisi lingkungan yang berbeda. Dengan demikian, studi ini tidak hanya memberikan pemahaman baru mengenai efektivitas KNN untuk deteksi okupansi, tetapi juga membuka peluang untuk pengembangan kajian lebih lanjut yang dapat memperkaya literatur terkait sistem deteksi okupansi berbasis sensor non-intrusif.

Daftar Pustaka

- [1] A. R. Mena, H. G. Ceballos, and J. Alvarado-Urbe, "Measuring Indoor Occupancy through Environmental Sensors: A Systematic Review on Sensor Deployment," *Sensors*, vol. 22, no. 10, pp. 3770, 2022.
- [2] A. Tsanousa, C. Moschou, E. Bektsis, S. Vrochidis, and I. Kompatsiaris, "Fusion of Environmental Sensors for Occupancy Detection in a Real Construction Site," *Sensors*, vol. 23, no. 23, pp. 9596, 2023.
- [3] M. Emad-Ud-Din and Y. Wang, "Promoting Occupancy Detection Accuracy Using On-Device Lifelong Learning," *IEEE Sensors Journal*, vol. 23, no. 9, pp. 9595–9608, 2023.
- [4] A. N. Sayed, Y. Himeur, and F. Bensaali, "Deep and Transfer Learning for Building Occupancy Detection: A Review and Comparative Analysis," *Engineering Applications of Artificial Intelligence*, vol. 115, 2022.
- [5] C. Wang, J. Jiang, T. Roth, C. Nguyen, Y. Liu, and H. Lee, "Integrated Sensor Data Processing for Occupancy Detection in Residential Buildings," *Energy and Buildings*, vol. 237, 2021.
- [6] M. Emad-Ud-Din and Y. Wang, "Promoting Occupancy Detection Accuracy Using On-Device Lifelong Learning," *IEEE Sensors Journal*, 2023.