

Prediksi Cuaca Kota Jakarta menggunakan Metode Random Forest: Studi Optimalitas

ZIAN ASTI DWIYANTI, CAHYO PRIANTO

Universitas Logistik dan Bisnis Internasional, Indonesia
Email : ziandwiasti23@gmail.com

ABSTRAK

Prediksi cuaca berperan penting dalam berbagai bidang kehidupan, seperti pertanian, transportasi, pariwisata, dan mitigasi bencana. Kemampuan memprediksi cuaca secara akurat dan tepat waktu sangat berdampak dalam pengambilan keputusan yang cerdas. Kota Jakarta, sebagai ibu kota Indonesia yang padat penduduk dan memiliki aktivitas ekonomi tinggi, membutuhkan sistem prediksi cuaca yang handal untuk mendukung pengelolaan sektor-sektor tersebut. Studi ini bertujuan memprediksi cuaca di Kota Jakarta dengan menggunakan metode Random Forest dan data cuaca historis yang terpercaya dari website OpenData Jakarta. Evaluasi menunjukkan bahwa model Random Forest memberikan prediksi cuaca yang baik, dengan akurasi, presisi dan recall sebesar 0.71, F1-score sebesar 0.70, serta ROC-AUC sebesar 0.92. Metrik evaluasi ini menggambarkan kinerja model dalam mengklasifikasikan cuaca dengan baik, mempertimbangkan keakuratan, ketepatan, dan keseimbangan antara presisi dan recall. Hasil prediksi cuaca tersebut mencakup kemampuan model untuk mengidentifikasi dengan benar berbagai kelas cuaca, dan memberikan informasi berharga dalam pengambilan keputusan terkait kondisi cuaca di Kota Jakarta.

Kata kunci: akurasi , prediksi cuaca, random forest, ROC-AUC

ABSTRACT

Weather prediction plays a crucial role across various life domains, including agriculture, transportation, tourism, and disaster mitigation. The ability to predict weather accurately and in a timely manner significantly impacts informed decision-making. Jakarta, as Indonesia's populous capital with high economic activity, necessitates a reliable weather forecasting system to support sector management. This study aims to predict Jakarta's weather using the Random Forest method and dependable historical weather data from the OpenData Jakarta website. Evaluation reveals that the Random Forest model offers favorable weather predictions, boasting an accuracy, precision, and recall of 0.71, an F1-score of 0.70, and an ROC-AUC of 0.92. These evaluation metrics epitomize the model's adeptness in effectively classifying weather, striking a balance between precision and recall. The weather prediction outcomes encompass the model's capacity to accurately identify diverse weather categories, thereby furnishing valuable insights for decision-making concerning Jakarta's weather conditions.

Keywords: accuracy, random forest, ROC-AUC, weather prediction

1. PENDAHULUAN

Cuaca adalah fenomena alam yang terjadi di atmosfer Bumi dan berhubungan dengan kondisi udara seperti suhu, kelembaban, waktu, dan wilayah (**Zittis et al., 2022**). Kondisi cuaca yang berubah-ubah dapat memiliki dampak signifikan pada berbagai aspek kehidupan manusia, termasuk pertanian, transportasi, pariwisata, dan mitigasi bencana. Oleh karena itu, kemampuan untuk memprediksi cuaca dengan akurat dan tepat waktu menjadi sangat penting dalam pengambilan keputusan yang berkaitan dengan kegiatan sehari-hari.

Dalam era digital saat ini, teknologi dan metode komputasi telah memungkinkan pengembangan model prediksi cuaca yang lebih canggih dan presisi. Salah satu pendekatan yang populer dalam memprediksi cuaca adalah dengan menggunakan metode *Machine Learning*. *Machine Learning* adalah cabang dari kecerdasan buatan yang memungkinkan komputer belajar dan beradaptasi dari data tanpa harus secara eksplisit diprogram (**Fossaceca & Young, 2018**).

Dalam konteks prediksi cuaca, metode *Machine Learning* dapat digunakan untuk mempelajari pola dan hubungan kompleks antara variabel cuaca yang beragam, seperti suhu, kelembaban, waktu, dan wilayah. Metode ini memanfaatkan teknik statistik dan algoritma yang dapat menghasilkan model prediksi yang dapat memperkirakan kondisi cuaca di masa depan berdasarkan data historis (**Zulkifri et al., 2023**).

Salah satu algoritma *Machine Learning* yang sering digunakan dalam prediksi cuaca adalah *Random Forest*. *Random Forest* adalah sebuah algoritma *ensemble* yang terdiri dari kumpulan pohon keputusan (decision trees) (**Indahyanti et al., 2022**). Setiap pohon keputusan dalam *Random Forest* melakukan prediksi secara independen, dan hasil prediksi dari setiap pohon digabungkan untuk menghasilkan prediksi akhir (**Mishina et al., 2015**). Kelebihan dari *Random Forest* adalah kemampuannya dalam mengatasi masalah prediksi cuaca yang kompleks, termasuk non-linearitas, interaksi antar variabel, dan keberadaan *outliers* dalam data.

Penelitian ini bertujuan untuk melakukan prediksi cuaca di Kota Jakarta menggunakan metode *Random Forest*. Data cuaca historis yang terpercaya, yang diperoleh dari website resmi OpenData Jakarta, digunakan sebagai dataset untuk melatih dan menguji model prediksi. Metode *Random Forest* dipilih karena kemampuannya dalam mengatasi masalah *overfitting*, serta keandalannya dalam menghasilkan prediksi yang akurat dan stabil (**Sirikulviriya & Sinthupinyo, 2011**).

Selain itu, dalam evaluasi kinerja model prediksi cuaca, salah satu metrik yang penting untuk diperhatikan adalah *Area Under the Receiver Operating Characteristic Curve* (ROC-AUC). ROC-AUC adalah ukuran yang digunakan untuk mengukur sejauh mana model mampu membedakan antara kelas positif dan negatif dalam prediksi (**Kubben et al., 2018**). Grafik ROC-AUC menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai ambang batas (*threshold*) prediksi.

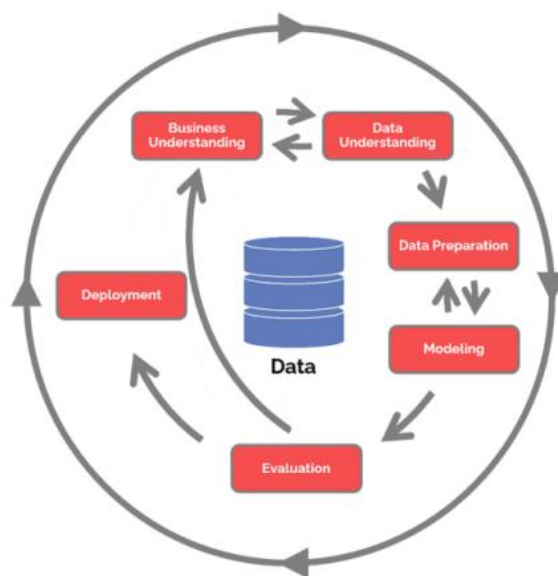
Secara intuitif, ROC-AUC menggambarkan kemampuan model dalam membedakan antara prediksi yang benar positif dengan prediksi yang salah positif. Nilai ROC-AUC berkisar antara 0 hingga 1, di mana nilai 0 menunjukkan kinerja yang buruk (model tidak dapat membedakan antara kelas positif dan negatif) dan nilai 1 menunjukkan kinerja yang sempurna (model mampu membedakan secara sempurna antara kelas positif dan negatif) (**Adawiyah, 2023**).

Dalam penelitian ini, kami akan menggunakan nilai ROC-AUC sebagai salah satu metrik evaluasi untuk mengukur kinerja model prediksi cuaca menggunakan metode *Random Forest*. Nilai ROC-AUC yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan antara prediksi cuaca yang benar dan prediksi yang salah. Dengan menggunakan ROC-AUC, kami dapat mengevaluasi sejauh mana model *Random Forest* mampu menghasilkan prediksi cuaca yang akurat dan dapat diandalkan.

Pada akhir penelitian akan dilakukan analisis hasil evaluasi kinerja model *Random Forest* dalam prediksi cuaca di Kota Jakarta. Kami akan melaporkan nilai akurasi, *precision*, *recall*, dan F1-score sebagai metrik evaluasi lainnya untuk memberikan gambaran yang komprehensif tentang performa model. Selain itu, nilai ROC-AUC akan dijelaskan untuk menunjukkan sejauh mana model mampu membedakan antara prediksi

2. METODE

Dalam penelitian ini, digunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari enam tahapan utama. Metodologi ini memberikan kerangka kerja yang terstruktur dan sistematis untuk melakukan analisis data dan membangun model prediksi cuaca yang akurat (Darono, 2021). Berikut adalah tahapan dalam metodologi CRISP-DM :



Gambar 1. CRISP-DM

2.1 Business Understanding

Pada tahap ini, penelitian akan secara mendalam memahami tujuan bisnis yang mendasari pelaksanaan prediksi cuaca di Kota Jakarta menggunakan berbagai metode klasifikasi. Tujuan bisnis ini memiliki relevansi yang sangat penting dalam upaya memahami dinamika cuaca kota dan mengantisipasi dampaknya terhadap berbagai sektor masyarakat.

Dengan demikian, fokus utama tahap ini adalah untuk merumuskan dan mengklarifikasi aspek-aspek fundamental yang melandasi pelaksanaan prediksi cuaca. Lebih dari sekadar menyajikan prediksi yang akurat, penelitian ini memiliki misi untuk memberikan pandangan yang lebih dalam terkait bagaimana metode-metode klasifikasi dapat saling dibandingkan dan diukur kinerjanya dalam menghadapi tantangan kompleksitas prediksi cuaca.

Lebih jauh, tujuan bisnis yang tercantum adalah melampaui sekadar menghadirkan hasil prediksi cuaca yang lebih baik. Penelitian ini juga bertujuan untuk memberikan kontribusi signifikan dalam pengembangan metode analisis cuaca di lingkungan perkotaan, yang pada gilirannya dapat memberikan panduan berharga bagi berbagai sektor, mulai dari transportasi hingga pertanian, dalam menghadapi variasi cuaca yang berpotensi mengganggu.

2.2 Data Understanding

Dalam tahap *Data Understanding*, dilaksanakan serangkaian langkah yang terarah untuk memastikan bahwa data yang digunakan sebagai dasar penelitian memiliki integritas dan relevansi yang diperlukan. Penyelidikan ini melibatkan beberapa tahap, mulai dari proses pencarian dan pengumpulan data hingga analisis mendalam terhadap atribut-atribut yang ada serta evaluasi keseluruhan kualitas dataset (**Chen et al., 2014**).

Studi ini memfokuskan pada objek penelitian yang melibatkan pemahaman mendalam mengenai kondisi cuaca di Provinsi DKI Jakarta. Data yang diambil berasal dari sumber open data Jakarta, memberikan wawasan yang berharga tentang variabilitas cuaca di kawasan tersebut. Rentang waktu pengumpulan data ini meliputi periode dari bulan Januari hingga Desember 2018, memungkinkan pemodelan berdasarkan data historis yang cukup representatif. Dataset ini terdiri dari enam atribut utama yang mencakup informasi penting, seperti tanggal, wilayah yang diobservasi, waktu pengamatan, kondisi cuaca aktual, persentase kelembaban udara, dan suhu udara dalam skala Celsius.

Total data yang tersedia dalam dataset ini mencapai 8535 entri. Atribut-atribut ini bukan hanya sekadar sekumpulan data, tetapi juga kunci untuk memahami pola dan dinamika cuaca di wilayah tersebut. Tabel 1, yang disajikan dalam penelitian ini, mendetailkan informasi mengenai setiap atribut dataset, memberikan pandangan menyeluruh mengenai kontribusi masing-masing variabel terhadap pemodelan prediksi cuaca.

Dalam upaya menjaga kualitas data, tahap analisis juga mencakup evaluasi yang ketat terhadap keandalan dataset. Ini melibatkan identifikasi dan penanganan nilai yang hilang, perbaikan duplikasi data, serta verifikasi konsistensi dan keteraturan data. Dengan menjalankan tahap ini secara komprehensif, dataset yang dihasilkan menjadi sumber daya yang dapat diandalkan untuk mendukung pengembangan model prediksi cuaca yang akurat dan efektif dalam menciptakan pemahaman tentang perubahan cuaca di Kota Jakarta.

Tabel 1. Detail Atribut Dataset

Nama Atribut	Keterangan
Tanggal	Tanggal data diambil
Wilayah	Wilayah data
Waktu	Waktu data dari kelembaban dan suhu diambil
Cuaca	Jenis cuaca
Kelembaban	Nilai kelembaban dalam presentase
Suhu	Nilai suhu dalam derajat celcius

Sedangkan Potongan dataset cuaca di DKI Jakarta terdapat dalam Tabel 2.

Tabel 2. Preview Dataset Cuaca DKI Jakarta

Tanggal	Wilayah	Waktu	Cuaca	Kelembaban	Suhu
2018-01-02	Jakarta Timur	Pagi	Cerah Berawan	65-95 %	23-32°
2018-01-02	Jakarta Barat	Pagi	Cerah	65-95 %	23-32°
2018-01-02	Jakarta Selatan	Pagi	Cerah Berawan	65-95 %	23-32°
2018-01-02	Kepulauan Seribu	Pagi	Cerah	65-95 %	23-32°
2018-01-02	Jakarta Pusat	Pagi	Cerah	65-95 %	23-32°

2.3 Data Preparation

Data Preparation atau adalah proses yang dilakukan untuk mengubah data mentah menjadi data yang berkualitas dan cocok sebagai input untuk alat-alat analisis data (**Lenz et al., 2018**). Data yang digunakan terdistribusi ke dalam 12 dataset berdasarkan bulan. Untuk menyimpan data dari bulan Januari hingga Desember, dataset per bulan perlu diintegrasikan menjadi satu dataset utuh. Metode yang digunakan untuk menggabungkan semua file adalah dengan menempatkannya dalam satu folder. Proses iterasi dilakukan untuk membaca setiap dataset dalam folder tersebut dan menggabungkannya dengan menggunakan operasi penggabungan (*concatenation*).

Dalam dataset ini, kelas yang akan diprediksi terdapat pada kolom Cuaca. Setelah melakukan eksplorasi data, ditemukan bahwa terdapat 27 kelas yang berbeda, antara lain Cerah Berawan, Berawan, Hujan Lokal, Hujan Ringan, Cerah, Hujan Petir, Berawan Tebal, Hujan Sedang, Cerah Berawan, Cerah, Berawan, Hujan Lokal, Berawan, Cerah Berawa, Hujan, Hujan Ringan, Cerah berawan, Beawan, Hujan Ringan, Cerang Berawan, Cerah, Hujan Loka, Hujan Sedang, Hujan Petir, Berawa, Hujan Sedang dan Cerah Berawan. Namun, terdapat perbedaan penulisan dan variasi nama untuk jenis cuaca yang sebenarnya sama. Sebagai contoh, "Hujan Loka" dan "Hujan Lokal" sebenarnya mengacu pada jenis cuaca yang sama. Hal yang serupa juga terjadi pada "Beawan" dan "Berawa" yang memiliki arti yang sama dengan "Berawan". Oleh karena itu, diperlukan proses normalisasi kelas untuk menghindari adanya duplikasi jenis cuaca. Dari 27 kelas awal, akan dilakukan normalisasi menjadi 8 kategori cuaca yang lebih umum, seperti Cerah, Cerah Berawan, Berawan, Berawan Tebal, Hujan Lokal, Hujan Ringan, Hujan Sedang, dan Hujan Petir.

Atribut waktu terdiri dari tahun, bulan, dan tanggal. Untuk atribut tanggal, digunakan ekspresi lambda untuk memisahkan nilai tanggal berdasarkan separator "-" dalam format "yyyy-mm-dd". Hasilnya akan diubah menjadi format "yyyy", "mm", dan "dd". Selanjutnya, DataFrame 'df' akan diperbarui dengan penambahan dua kolom baru, yaitu kolom "hari" dan kolom "bulan", yang berisi informasi hari dan bulan yang diekstraksi dari atribut tanggal.

Atribut suhu dan kelembaban awalnya memiliki tipe data *string* dengan *postfix* tambahan "%" dan "°". Agar nilai suhu dan kelembaban dapat diolah secara numerik, *postfix* tersebut akan dihilangkan. Selain itu, jika nilai suhu dan kelembaban berada dalam interval tertentu, maka nilai tersebut akan dipisah menjadi dua kolom baru, yaitu kolom "suhu_min" dan kolom "suhu_max" untuk suhu, serta kolom "kelembaban_min" dan kolom "kelembaban_max" untuk kelembaban.

Atribut waktu memiliki 8 jenis waktu, yaitu Dini Hari, Pagi, Siang, Malam, dini hari, pagi, siang, dan malam. Terdapat beberapa entitas yang merujuk pada objek yang sama, misalnya "Pagi" dan "pagi". Untuk mengatasi hal ini, dilakukan transformasi bentuk waktu menjadi *lowercase* sehingga terdapat 4 jenis waktu yang dihasilkan, yaitu dini hari, pagi, siang, dan malam.

Sebelum data dapat digunakan sebagai input untuk algoritma pembelajaran mesin, perlu dilakukan konversi atribut kategorikal yang awalnya berbentuk *string* menjadi bentuk numerik. Pada dataset ini, terdapat 3 atribut yang memerlukan konversi, yaitu wilayah, waktu, dan cuaca. Untuk melakukan konversi, digunakan pendekatan *mapping* atau pemetaan nilai. Setiap nilai unik akan dihubungkan dengan sebuah nilai numerik yang mewakili kategori tersebut. Dengan menggunakan pemetaan ini, atribut wilayah, cuaca dan waktu akan diubah menjadi representasi numerik yang dapat digunakan dalam algoritma pembelajaran mesin.

2.4 Modelling

Pada tahap *modelling* dalam metodologi CRISP-DM, dilakukan pengembangan model prediksi cuaca menggunakan pendekatan yang cermat dan terstruktur. Dalam konteks ini, model prediksi cuaca menjadi inti dari upaya penelitian, yang ditujukan untuk menghadirkan solusi yang efektif dalam mengatasi tantangan prediksi cuaca yang kompleks di Kota Jakarta.

Dalam menjalankan tahap ini, dataset cuaca historis yang telah melalui *preprocessing* akan menjadi bahan baku utama. Dataset ini akan dipisahkan menjadi dua subset: data pelatihan yang akan digunakan untuk melatih model dan data pengujian yang akan digunakan untuk mengevaluasi kinerja model. Melalui pendekatan ini, tahap *modelling* tidak hanya merujuk pada proses pembentukan model semata, tetapi juga menegaskan pentingnya pengujian yang akurat dan representatif.

Model prediksi cuaca akan dibangun dengan menerapkan algoritma *Random Forest* pada data pelatihan. *Random Forest* merupakan pendekatan yang kuat dalam klasifikasi dan regresi yang memanfaatkan sekumpulan pohon keputusan untuk menghasilkan prediksi yang lebih kuat dan tahan terhadap *overfitting* (anwar, 2022). Dalam konteks ini, algoritma ini akan diterapkan untuk memahami hubungan kompleks antara berbagai variabel cuaca historis dan kondisi cuaca aktual di Kota Jakarta.

Penting untuk dicatat bahwa tahap *modelling* ini bukan hanya tentang menerapkan algoritma secara teknis, tetapi juga melibatkan eksplorasi parameter-parameter yang optimal untuk memastikan model memiliki kinerja yang baik. Hasil dari tahap ini akan memberikan wawasan berharga mengenai kemampuan model dalam menggeneralisasi dan mengatasi situasi cuaca yang beragam. Dengan memanfaatkan alat komputasi dan analisis yang canggih, tahap ini menjadi landasan untuk mengukur efektivitas metode *Random Forest* dalam memprediksi cuaca di lingkungan yang dinamis seperti Kota Jakarta.

2.5 Evaluation

Dalam evaluasi kinerja model, kami menggunakan *confusion matrix* untuk menganalisis hasil prediksi model. *Confusion matrix* digunakan untuk menghitung jumlah *true positive* (TP), *false positive* (FP), *true negative* (TN), dan *false negative* (FN) (Indrayuni, 2018).

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	<p>TP (True Positive)</p>	<p>FP (False Positive) <i>Type I Error</i></p>
	0 (Negative)	<p>FN (False Negative) <i>Type II Error</i></p>	<p>TN (True Negative)</p>

Gambar 2. Confusion Matrix

Dari *confusion matrix* ini, kita dapat menghitung beberapa metrik evaluasi kinerja model, yaitu akurasi, presisi, *recall*, dan *F1 score*. Akurasi merupakan persentase prediksi yang benar dari seluruh prediksi yang dilakukan oleh model (Arini, 2021). Presisi mengukur persentase prediksi positif yang benar dari seluruh prediksi positif yang dilakukan oleh model (Purwono et al., 2022). *Recall* mengukur persentase prediksi positif yang benar dari seluruh data positif yang ada pada dataset (Ragab et al., 2021). *F1 score* merupakan rata-rata harmonis dari presisi dan *recall*, memberikan indikasi keseluruhan kinerja model dalam memprediksi kelas yang berbeda. ROC-AUC memberikan gambaran tentang seberapa baik model dapat membedakan antara kelas positif dan negatif dalam prediksi cuaca. Nilai ROC-AUC mendekati 1 menunjukkan kinerja yang lebih baik, yang berarti model memiliki kemampuan yang tinggi dalam memprediksi cuaca dengan akurasi yang tinggi. Selain itu, akurasi juga menjadi metrik penting yang mengukur sejauh mana model dapat mengklasifikasikan data cuaca dengan benar. Semakin tinggi nilai akurasi, semakin baik model dalam melakukan prediksi cuaca secara tepat.

2.6 Deployment

Tahap *deployment* dalam metodologi CRISP-DM membawa kita ke langkah implementasi praktis dari model prediksi cuaca yang telah dikembangkan. Setelah model prediksi cuaca melalui proses pelatihan dan evaluasi, langkah selanjutnya adalah memastikan bahwa model ini dapat diakses dan dimanfaatkan secara efisien. Pada tahap ini, model prediksi cuaca akan diintegrasikan ke dalam alur kerja yang dapat memberikan manfaat nyata bagi berbagai pemangku kepentingan.

Konkretnya, untuk mengimplementasikan model ini, kita akan menggunakan modul *pickle* yang terdapat dalam bahasa pemrograman *Python*. Modul ini memungkinkan kita untuk menyimpan model yang telah dilatih dalam bentuk file, sehingga kita dapat dengan mudah mengakses dan memuatnya kembali di masa depan. Dengan pendekatan ini, model prediksi cuaca dapat diintegrasikan ke dalam berbagai platform, seperti *Jupyter Notebook* atau lingkungan pemrograman *Python* lainnya, untuk penggunaan yang lebih lanjut.

Pada tahap ini, model yang telah dilatih akan disimpan menggunakan modul *pickle*. Proses ini mengemas model beserta semua parameter dan informasi pentingnya ke dalam suatu file yang dapat diarsipkan dan dibagikan. Ketika model perlu digunakan untuk prediksi cuaca pada data baru, kita cukup memuat model dari file yang telah disimpan. Langkah ini dapat dilakukan dengan mudah menggunakan fungsi *pickle* yang tersedia dalam bahasa pemrograman *Python*. Setelah model dimuat, kita dapat langsung memanfaatkannya untuk melakukan prediksi cuaca

pada data baru. Hal ini cukup dengan mengimpor model yang telah dimuat, memuat data cuaca baru, dan memanggil metode prediksi dari model tersebut.

Dengan demikian, tahap *deployment* bukan hanya sekadar mengakhiri proses penelitian, tetapi juga membuka pintu bagi pemanfaatan praktis dari hasil penelitian. Dengan model prediksi cuaca yang telah diimplementasikan, pemangku kepentingan dari berbagai sektor dapat memanfaatkannya untuk mengambil keputusan yang lebih baik dan lebih terinformasi dalam menghadapi variasi cuaca yang kompleks di Kota Jakarta.

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, evaluasi kinerja model *Random Forest* dalam prediksi cuaca di Kota Jakarta dilakukan dengan menganalisis metrik-metrik evaluasi yang dijelaskan sebelumnya. Metrik-metrik ini memberikan pemahaman yang lebih dalam tentang keakuratan, presisi, *recall*, *F1-score*, dan ROC-AUC dari model dalam melakukan prediksi cuaca.

Dalam tahap evaluasi, model *Random Forest* dievaluasi menggunakan dataset pengujian yang terpisah dari dataset pelatihan dengan perbandingan 80:20 antara data latih dan data tes. Hasil evaluasi ini memberikan gambaran yang obyektif tentang kinerja model dalam menggeneralisasi pola dari data pelatihan ke data yang belum pernah dilihat sebelumnya.

Tabel hasil evaluasi kinerja model *Random Forest* di bawah ini menyajikan rata-rata dari metrik-metrik evaluasi yang dihasilkan dari pengujian model pada dataset pengujian.

Table 3. Hasil Evaluasi

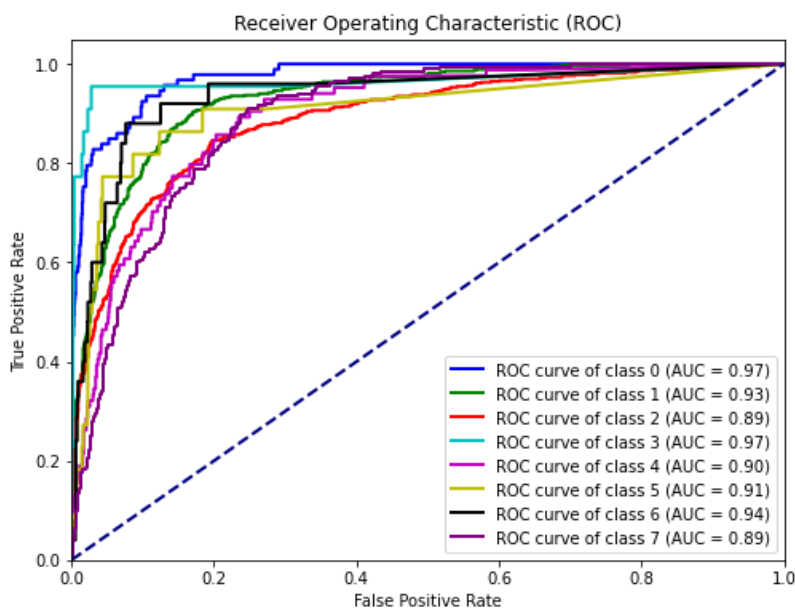
Metrik	Hasil
<i>Accuracy</i>	0.71
<i>Precision</i>	0.71
<i>Recall</i>	0.71
<i>F1-score</i>	0.70
ROC-AUC	0.92

Tabel yang dipaparkan di atas mengungkapkan hasil evaluasi yang komprehensif terhadap kinerja model *Random Forest* dalam konteks prediksi cuaca di Kota Jakarta. Berdasarkan hasil evaluasi ini, model *Random Forest* menunjukkan kualitas yang mengesankan dalam mengklasifikasikan kondisi cuaca dengan akurasi mencapai 0.71. Angka ini mengindikasikan bahwa model mampu dengan benar mengklasifikasikan cuaca dalam 71% dari total prediksi yang dilakukan. Selain itu, aspek presisi, *recall*, dan *F1-score* juga menunjukkan performa yang kuat dengan nilai 0.71. Hal ini menggambarkan kemampuan model dalam memprediksi cuaca dengan akurasi yang tinggi, sambil menjaga keseimbangan antara *false positive* dan *false negative*.

Selain metrik-metrik tersebut, penting untuk melihat metrik ROC-AUC yang memiliki nilai sebesar 0.92. Metrik ROC-AUC memberikan gambaran tentang sejauh mana model mampu membedakan antara prediksi positif dan negatif menggunakan kurva ROC. Nilai ROC-AUC yang mendekati 1 mengindikasikan tingkat kinerja yang sangat baik dalam membedakan berbagai kelas cuaca. Hasil ini menunjukkan bahwa model *Random Forest* memiliki kemampuan yang tangguh dalam mengenali perbedaan antara kondisi cuaca yang berbeda secara efisien.

Lebih lanjut, grafik ROC-AUC yang ditampilkan di bawah ini memberikan pandangan visual mengenai performa model. Kurva ROC ini menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai ambang batas (*threshold*) untuk membedakan kelas cuaca yang berbeda. Grafik ini membantu untuk secara lebih terperinci memahami kemampuan model dalam menyesuaikan pengaturan ambang batas yang berbeda untuk meminimalkan kesalahan klasifikasi.

Dengan kombinasi antara nilai-nilai metrik dan visualisasi ini, hasil evaluasi memberikan dukungan kuat terhadap kinerja model *Random Forest* dalam menjalankan tugas penting dalam prediksi cuaca di lingkungan yang dinamis seperti Kota Jakarta. Hasil ini memberikan dasar yang meyakinkan bagi kesimpulan mengenai efektivitas model dalam memberikan solusi yang handal dan akurat dalam prediksi cuaca.



Gambar 3. Grafik ROC-AUC

Berdasarkan grafik ROC-AUC yang tersaji, dapat kita amati bahwa kurva ROC menjulur ke arah sudut kiri atas, mengindikasikan performa model yang luar biasa dalam membedakan antara kelas cuaca yang berbeda. Semakin mendekati kurva ROC ke garis diagonal, semakin terbatas kemampuan model dalam membedakan kelas cuaca, namun dalam kasus ini, bentuk kurva yang mendekati sudut kiri atas menandakan kemampuan memisahkan kelas cuaca yang tinggi.

Dengan nilai ROC-AUC sebesar 0.92 dan bentuk kurva ROC yang mendekati posisi optimal di sudut kiri atas, dapat diambil kesimpulan bahwa model *Random Forest* memiliki kinerja yang unggul dalam memprediksi cuaca di wilayah Kota Jakarta. Kemampuan model dalam membedakan berbagai variasi cuaca memberikan keyakinan dalam hasil prediksi yang dihasilkan, mengarah pada prediksi yang akurat dan dapat diandalkan.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model *Random Forest* adalah alat prediksi cuaca yang efisien dan andal untuk wilayah Kota Jakarta. Dengan akurasi yang mencolok, serta nilai *precision*, *recall*, *F1-score*, dan ROC-AUC yang mencerminkan kinerja yang solid, model ini memiliki potensi untuk menjadi alat yang sangat berharga dalam menyediakan informasi cuaca yang sangat diperlukan oleh berbagai sektor, mulai dari pertanian yang tergantung pada kondisi cuaca hingga industri transportasi yang memerlukan pemahaman yang akurat tentang cuaca untuk mengatur operasional mereka. Tidak hanya itu,

model ini juga memiliki potensi untuk mendukung upaya mitigasi bencana dan pengambilan keputusan yang lebih baik dalam situasi cuaca yang tidak pasti.

4. KESIMPULAN

Penelitian ini menggunakan metode *Random Forest* untuk melakukan prediksi cuaca di Kota Jakarta. Evaluasi kinerja model menunjukkan hasil yang menggembirakan, dengan akurasi sebesar 0.71, menunjukkan kemampuan model untuk memprediksi cuaca dengan tingkat keberhasilan sebesar 71%. Selain itu, nilai ROC-AUC sebesar 0.92 menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan antara kelas cuaca yang berbeda. Dengan hasil ini, model *Random Forest* memiliki potensi untuk digunakan dalam pengembangan sistem prediksi cuaca yang lebih akurat dan dapat diandalkan di Kota Jakarta. Hasil penelitian ini memberikan kontribusi yang penting dalam bidang prediksi cuaca dan dapat memberikan manfaat bagi berbagai sektor yang terkait dengan cuaca seperti pertanian, transportasi, pariwisata, dan mitigasi bencana.

Dalam kesimpulan, penelitian ini menunjukkan bahwa metode *Random Forest* memiliki kinerja yang baik dalam prediksi cuaca di Kota Jakarta. Meskipun masih ada ruang untuk pengembangan lebih lanjut, hasil evaluasi yang positif menunjukkan potensi penggunaan model ini dalam mendukung pengambilan keputusan yang lebih baik terkait dengan cuaca di Kota Jakarta. Dengan memanfaatkan data historis cuaca yang tersedia dan metode *machine learning*, model *Random Forest* dapat memberikan prediksi cuaca yang lebih akurat dan berguna dalam berbagai sektor kehidupan sehari-hari.

UCAPAN TERIMA KASIH

Terima kasih kepada Bapak Cahyo Prianto selaku pembimbing pada penelitian ini dan kepada teman-teman D4 Teknik Informatika B Angkatan 2020 yang telah memberikan *support*. Tak lupa untuk mengucapkan terima kasih kepada kedua orang tua yang telah mendukung baik itu materil maupun non materil.

DAFTAR RUJUKAN

- Adawiyah, R. (2023). *Eksplorasi Kapasitas Pengkodean Amplitudo Untuk Model Quantum Machine Learning*. 3(1), 38–58. <http://journal.politeknik-pratama.ac.id/index.php/JTIM>• page38
- anwar, S. R. (2022). Perbandingan Prediksi Harga Saham menggunakan Metode SVR, RFR, dan DTR. *Prosiding Seminar Pendidikan Matematika Dan Matematika*, 6(2). <http://prosiding.himatikauny.org/index.php/prosidinglsm/article/view/250>
- Arini, I. N. (2021). Analisis Akurasi Model-Model Prediksi Financial Distress. *Jurnal Ilmu Manajemen*, 9(3), 1196–1204. <https://doi.org/10.26740/jim.v9n3.p1196-1204>
- Chen, H., Hailey, D., Wang, N., & Yu, P. (2014). A review of data quality assessment methods for public health information systems. *International Journal of Environmental Research and Public Health*, 11(5), 5170–5207. <https://doi.org/10.3390/ijerph110505170>

- Darono, A. (2021). Sains Data Perpajakan di Indonesia: Implementasi Praktis, Pembelajaran, dan Agenda Kajiannya. *Simposium Nasional Perpajakan*, 1(1), 9–17.
- Fossaceca, J. M., & Young, S. H. (2018). *Artificial intelligence and machine learning for future army applications*. 1063507(May 2018), 6. <https://doi.org/10.1117/12.2307753>
- Indahyanti, U., Azizah, N. L., & Setiawan, H. (2022). Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa. *Jurnal Sains Dan Informatika*, 8(2), 160–169. <https://doi.org/10.34128/jsi.v8i2.459>
- Indrayuni, E. (2018). Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film. *Jurnal Pilar Nusa Mandiri*, 14(2), 175. <https://doi.org/10.33480/pilar.v14i2.918>
- Kubben, P., Dumontier, M., & Dekker, A. (2018). Fundamentals of clinical data science. In *Fundamentals of Clinical Data Science*. <https://doi.org/10.1007/978-3-319-99713-1>
- Lenz, J., Wuest, T., & Westkämper, E. (2018). Holistic approach to machine tool data analytics. *Journal of Manufacturing Systems*, 48(July 2021), 180–191. <https://doi.org/10.1016/j.jmsy.2018.03.003>
- Mishina, Y., Murata, R., Yamauchi, Y., Yamashita, T., & Fujiyoshi, H. (2015). Boosted random forest. *IEICE Transactions on Information and Systems*, E98D(9), 1630–1636. <https://doi.org/10.1587/transinf.2014OPP0004>
- Purwono, P., Dewi, P., Wibisono, S. K., & Dewa, B. P. (2022). Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network. *Insect (Informatics and Security): Jurnal Teknik Informatika*, 7(2), 82–90. <https://doi.org/10.33506/insect.v7i2.1828>
- Ragab, M., Abdel Aal, A. M. K., Jifri, A. O., & Omran, N. F. (2021). Enhancement of Predicting Students Performance Model Using Ensemble Approaches and Educational Data Mining Techniques. *Wireless Communications and Mobile Computing*, 2021. <https://doi.org/10.1155/2021/6241676>
- Zittis, G., Almazroui, M., Alpert, P., Ciais, P., Cramer, W., Dahdal, Y., Fnais, M., Francis, D., Hadjinicolaou, P., Howari, F., Jrrar, A., Kaskaoutis, D. G., Kulmala, M., Lazoglou, G., Mihalopoulos, N., Lin, X., Rudich, Y., Sciare, J., Stenchikov, G., ... Lelieveld, J. (2022). Climate Change and Weather Extremes in the Eastern Mediterranean and Middle East. *Reviews of Geophysics*, 60(3). <https://doi.org/10.1029/2021RG000762>
- Zulfikri, A., Gunawan, & Andriani, W. . (2023). Tinjauan Pustaka Sistematis: Penerapan Metode Naives Bayes untuk Klasifikasi dalam Dataset Cuaca. *Journal of Practical Computer Science*, 2(2), 90-95.