

Penerapan Algoritma Random Forest dalam Prediksi Curah Hujan untuk Mendukung Analisis Cuaca

Rizal Torhino*, Pulung Nurtantio Andono

Ilmu Komputer, Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ^{1,*}111202113260@mhs.dinus.ac.id, ²pulung@dsn.dinus.ac.id

Email Penulis Korespondensi: 111202113260@mhs.dinus.ac.id

Submitted: 05/12/2024; Accepted: 17/12/2024; Published: 18/12/2024

Abstrak—Keragaman iklim di Indonesia menyebabkan pola curah hujan yang berbeda-beda di setiap daerah. Kondisi ini menghadirkan tantangan besar dalam upaya menghasilkan prediksi curah hujan yang akurat, yang penting untuk mendukung perencanaan infrastruktur yang efektif serta mitigasi bencana. Tujuan dari penelitian ini adalah untuk menganalisis potensi curah hujan di Kecamatan Purwodadi dengan menggunakan algoritma Random Forest. Dalam analisis ini, beberapa parameter cuaca seperti tekanan udara, suhu, kelembapan, dan kecepatan angin digunakan, sementara curah hujan menjadi variabel target dalam proses prediksi. Dataset yang digunakan dalam penelitian ini diperoleh dari *NASA Prediction Of Worldwide Energy Resources (POWER)* dengan periode waktu antara tahun 2000 hingga 2022. Data tersebut kemudian dibagi menjadi 70% untuk data latih dan 30% untuk data uji. Dalam penelitian ini, algoritma Random Forest digunakan untuk mengklasifikasikan kemungkinan terjadinya hujan berdasarkan kondisi cuaca yang ada. Hasil implementasi menunjukkan bahwa model *Random Forest* mencapai akurasi 100% pada data latih dan 92% pada data uji, yang mengindikasikan performa prediksi yang sangat baik. Hasil dari *confusion matrix* mengonfirmasi bahwa mayoritas prediksi model sesuai dengan data aktual. Temuan ini menunjukkan parameter cuaca yang digunakan efektif dalam memprediksi curah hujan di Kecamatan Purwodadi. Penelitian ini memberikan kontribusi dalam meningkatkan akurasi prediksi curah hujan dan membuka peluang untuk pengembangan model prediksi cuaca yang lebih baik, dengan melibatkan lebih banyak parameter atau menggunakan algoritma lain untuk evaluasi performa yang lebih mendalam.

Kata Kunci: Random Forest; Prediksi Curah Hujan; Tekanan Udara; Suhu; Kelembapan; Kecepatan Angin; Kecamatan Purwodadi

Abstract—Indonesia's climate diversity leads to different rainfall patterns in each region. This condition presents a major challenge in the effort to produce accurate rainfall predictions, which are important to support effective infrastructure planning and disaster mitigation. The purpose of this research is to analyze the rainfall potential in Purwodadi Sub-district using Random Forest algorithm. In this analysis, several weather parameters such as air pressure, temperature, humidity, and wind speed are used, while rainfall becomes the target variable in the prediction process. The dataset used in this study was obtained from *NASA Prediction Of Worldwide Energy Resources (POWER)* with a time period between 2000 and 2022. The data is then divided into 70% for training data and 30% for test data. In this study, the Random Forest algorithm was used to classify the likelihood of rain based on existing weather conditions. The implementation results showed that the Random Forest model achieved 100% accuracy on the training data and 92% on the test data, indicating excellent prediction performance. Results from the confusion matrix confirmed that the majority of the model predictions matched the actual data. This finding shows that the weather parameters used are effective in predicting rainfall in Purwodadi sub-district. This research contributes to improving the accuracy of rainfall prediction and opens up opportunities for the development of better weather prediction models, involving more parameters or using other algorithms for more in-depth performance evaluation.

Keywords: Random Forest; Rainfall Prediction; Surface Pressure; Temperature; Humidity; Wind Speed; Purwodadi sub-district

1. PENDAHULUAN

Curah hujan adalah ukuran volume air yang jatuh ke permukaan bumi selama periode waktu tertentu, biasanya dinyatakan dalam satuan milimeter (mm) [1]. Pengukuran ini menggambarkan tinggi air yang akan menutupi permukaan horizontal jika tidak ada penguapan, aliran, atau penyerapan ke dalam tanah [2]. Periode pengukuran dapat bervariasi, mulai dari harian hingga tahunan, tergantung pada kebutuhan analisis. Data curah hujan yang dikumpulkan dari berbagai lokasi memberikan informasi penting tentang pola hujan yang dapat digunakan untuk berbagai tujuan penelitian dan aplikasi praktis [3]. Pengamatan curah hujan melibatkan berbagai metode dan menghasilkan kumpulan data yang kompleks, mencakup variabilitas temporal dan spasial [4]. Data ini dianalisis untuk mendukung beragam penelitian, termasuk pengembangan model prediksi cuaca. Model semacam itu memanfaatkan data curah hujan bersama parameter cuaca lainnya untuk menghasilkan prediksi yang lebih akurat [5]. Prediksi curah hujan [6] yang andal sangat penting untuk perencanaan mitigasi bencana, pengelolaan sumber daya air, dan perencanaan infrastruktur.

Kota Purwodadi baru-baru ini mengalami banjir besar akibat hujan deras yang mengguyur wilayah tersebut. Bencana ini berdampak pada beberapa desa, seperti Gebangan, Glugu, Jagalan, Jengglong, Jetis, Kemas, Kuripan, Palembang, Plendungan, Sawahan, dan Banaran, sebagaimana dilaporkan oleh BPBD Kabupaten Grobogan. Tidak hanya permukiman, banjir juga merendam berbagai fasilitas umum, termasuk Alun-Alun Purwodadi, Kantor Setda Grobogan, dan sejumlah kantor pemerintahan seperti DPRD, Perhutani, hingga Kejaksaan Negeri Grobogan. Beberapa lokasi strategis lainnya, seperti rumah dinas Bupati dan Wakil Bupati, Jalan Siswomiharjo, Kantor KPU, serta area Segitiga Emas, juga terdampak parah. Bahkan, fasilitas kesehatan seperti RS Panti Rahayu Yakkum dan RSUD Purwodadi tidak luput dari genangan air, yang menghambat mobilitas dan pelayanan masyarakat [7]. Untuk mengantisipasi dampak hujan deras yang menyebabkan banjir, analisis curah hujan menjadi langkah penting. Data mining dapat digunakan untuk menganalisis pola curah hujan di suatu wilayah dan membantu menentukan

kemungkinan terjadinya hujan [8]. Dengan memanfaatkan teknologi ini, pihak berwenang dapat membuat perencanaan yang lebih baik dalam mengelola risiko bencana banjir, terutama di daerah rawan seperti Purwodadi [9]. Selain itu, analisis ini dapat membantu dalam pengambilan keputusan strategis untuk mitigasi bencana dan peningkatan infrastruktur penanggulangan banjir [10].

Data mining adalah proses untuk mengekstraksi informasi penting dan bernilai dari dataset besar [11]. Proses ini melibatkan metode yang kompleks, seperti matematika, teknik statistik, *machine learning*, dan kecerdasan buatan. Dengan memanfaatkan metode tersebut, data mining bertujuan untuk mengidentifikasi pola dan hubungan dalam data yang dapat digunakan untuk mendukung pengambilan keputusan. Dalam beberapa dekade terakhir, data mining telah berkembang menjadi disiplin ilmu yang mencakup berbagai bidang, seperti biologi, lingkungan, finansial, dan teknologi jaringan, menjadikannya alat yang sangat penting dalam analisis data yang kompleks.

Salah satu algoritma data mining yang efektif untuk menganalisis data cuaca, termasuk curah hujan, adalah *Random Forest*. Algoritma ini bekerja dengan membangun kumpulan pohon keputusan (*decision trees*) untuk membuat prediksi yang akurat dan stabil. *Random Forest* sangat cocok untuk dataset yang kompleks karena mampu menangani hubungan non-linear antar variabel serta meminimalkan *overfitting*. Dalam analisis curah hujan [12], *Random Forest* dapat digunakan untuk memprediksi kondisi cuaca berdasarkan parameter seperti tekanan udara, suhu, kelembapan, dan kecepatan angin. Hasil dari algoritma ini dapat membantu mempermudah pengambilan keputusan, seperti mitigasi bencana banjir atau perencanaan infrastruktur yang lebih baik di wilayah rawan banjir.

Penelitian ini menggunakan data curah hujan dari *NASA Prediction Of Worldwide Energy Resources (POWER)* [13] untuk menganalisis curah hujan di wilayah Kota Purwodadi. Data yang dimanfaatkan mencakup catatan bulanan tentang curah hujan selama periode 2000 hingga 2022. Tujuan utama dari penelitian ini adalah untuk memprediksi pola curah hujan di Kecamatan Purwodadi berdasarkan data historis tersebut.

Dalam penelitian ini, prediksi curah hujan di Kecamatan Purwodadi akan dikelompokkan menjadi dua kategori: tidak hujan dan hujan. Pengelompokan ini dilakukan untuk mempermudah proses klasifikasi dan analisis. Berdasarkan penelitian sebelumnya, metode *Random Forest* sering digunakan sebagai salah satu teknik yang handal dan efisien dalam mengklasifikasikan data untuk mendukung pengambilan keputusan. *Random Forest* bekerja dengan membangun kumpulan pohon keputusan (*decision trees*) secara acak dan menggabungkan hasilnya, sehingga menjadikannya metode yang kuat dan akurat dalam berbagai aplikasi, termasuk analisis curah hujan [14].

Metode algoritma *Random Forest* yang digunakan pada penelitian Akbar Hidayatullah Harahap, Ihsan Muttaqin Bin Abdul Malik, Muhammad Irfan Nur Imam, Muhammad Thariq Sabiq Bilhaq, Aisyah Amini Nur, Siti Lutfia Dwi Agustini di tahun 2021 dengan judul “Klasifikasi Diagnosa Penyakit Jantung menggunakan Algoritma *Random Forest*” dapat menghasilkan tingkat akurasi sebesar 85,3%. Jantung adalah organ vital dalam tubuh manusia, namun tidak jarang mengalami gangguan yang dapat memicu gejala serius hingga penyakit jantung yang berpotensi mematikan. Untuk mendukung diagnosis penyakit jantung yang cepat, tepat, dan akurat, berbagai penelitian telah dilakukan, salah satunya dengan menggunakan algoritma *Random Forest*. Algoritma ini menggabungkan hasil dari beberapa pohon keputusan untuk menghasilkan prediksi yang lebih akurat, sehingga menjadi metode yang efektif dalam mengklasifikasikan data terkait diagnosis penyakit jantung [15].

Penelitian lain yang menggunakan algoritma *Random Forest* adalah penelitian dengan judul “Analisa Perubahan Tutupan Lahan Akibat Letusan Gunung Semeru Tahun 2021 dengan Algoritma *Random Forest* (Studi Kasus: Kecamatan Pronojiwo, Kabupaten Lumajang)”. Penelitian ini dilakukan oleh Husnul Hidayat, Erika Azzahra, Filsa Bioresita. Akurasi yang dihasilkan sebesar 85,02% dengan identifikasi perubahan tutupan lahan akibat letusan Gunung Semeru pada tahun 2021, dengan peningkatan luas pada kelas tutupan sungai, lahan terbuka, dan hutan lahan kering masing-masing sebesar 0,054 km², 6,391 km², dan 6,963 km². Di sisi lain, kelas tutupan lahan area terbangun, sawah, dan perkebunan menunjukkan penurunan luas sebesar 0,055 km², 8,816 km², dan 4,536 km². Perubahan tutupan lahan ini lebih terkonsentrasi di sekitar sungai dan mengarah ke tenggara dari puncak Gunung Semeru [16].

Penelitian berikutnya dengan judul “Analisis Hasil Prediksi Magnitudo Gempa Di Wilayah Kota Padang Menggunakan Teknik *Random Forest*” yang dilakukan oleh Ade Fauzan, Defri Ahmad. Topik prediksi gempa menjadi perhatian penting dalam penelitian karena gempa bumi merupakan bencana alam yang dapat menyebabkan kerusakan besar, baik dari sisi ekonomi maupun kehidupan sosial. Salah satu metode yang digunakan dalam penelitian ini adalah *machine learning* dengan pendekatan *Random Forest* untuk menganalisis prediksi magnitudo gempa. Variabel yang digunakan dalam penelitian ini meliputi waktu, *latitude*, *longitude*, dan magnitudo, dengan data gempa yang diperoleh dari *United States Geological Survey (USGS)*. Hasil penelitian ini menunjukkan nilai RMSE sebesar 0.31758 dan MSE sebesar 0.10085 [17].

“Klasifikasi Wilayah Potensi Risiko Kerusakan Lahan Akibat Bencana Tsunami Menggunakan *Machine Learning*” merupakan salah satu penelitian yang menggunakan algoritma *Random Forest* pada penelitiannya di tahun 2022. Algoritma *Artificial Neural Network (ANN)* memberikan hasil prediksi klasifikasi yang lebih akurat dibandingkan dengan metode *Random Forest* dan *Support Vector Machine (SVM)*. Metode ANN menghasilkan akurasi sebesar 95,45% dan nilai kappa 86,08%. Dengan menggunakan metode ANN, klasifikasi wilayah yang memiliki potensi risiko tinggi kerusakan lahan akibat tsunami menunjukkan tujuh desa, yaitu Wijiharjo, Imogiri, Banguntapan, Pendowoharjo, Bangunharjo, Ngestiharjo, dan Tirtonirmolo. Selain itu, prediksi spasial yang dilakukan dengan metode IDW menghasilkan peta yang menunjukkan persebaran wilayah dengan potensi risiko rendah dan tinggi terhadap kerusakan lahan akibat tsunami [18].

Selanjutnya penelitian yang menggunakan algoritma *random forest* adalah pada penelitian Ilham Kurniawan, Duwi Cahya Putri Buani, Abdussomad, Widya Apriliah, Rizal Amegia Saputra di tahun 2023 dengan judul “Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin” Kemiskinan merupakan isu utama yang menjadi perhatian setiap pemerintah, dengan program Beras Keluarga Miskin (Raskin) sebagai salah satu upaya untuk mengurangi beban rumah tangga miskin melalui bantuan untuk meningkatkan ketahanan pangan. Penelitian ini bertujuan untuk menemukan algoritma klasifikasi prediktif dengan akurasi tertinggi dalam menentukan penerima bantuan Raskin, menggunakan *tools machine learning* Python yang diimplementasikan dalam sebuah *website*. Beberapa algoritma klasifikasi seperti SVM, Naive Bayes, dan Random Forest (RF) diuji dengan *dataset* Raskin dari Kelurahan Gunungparang, Kota Sukabumi. Hasil penelitian menunjukkan bahwa algoritma RF memberikan kinerja terbaik dengan nilai precision, recall, dan f-measure 97%, akurasi 97,26%, serta ROC 0,970, mengungguli SVM dan *Naive Bayes* dengan perbedaan akurasi masing-masing sebesar 5,11% dan 8,87% [19].

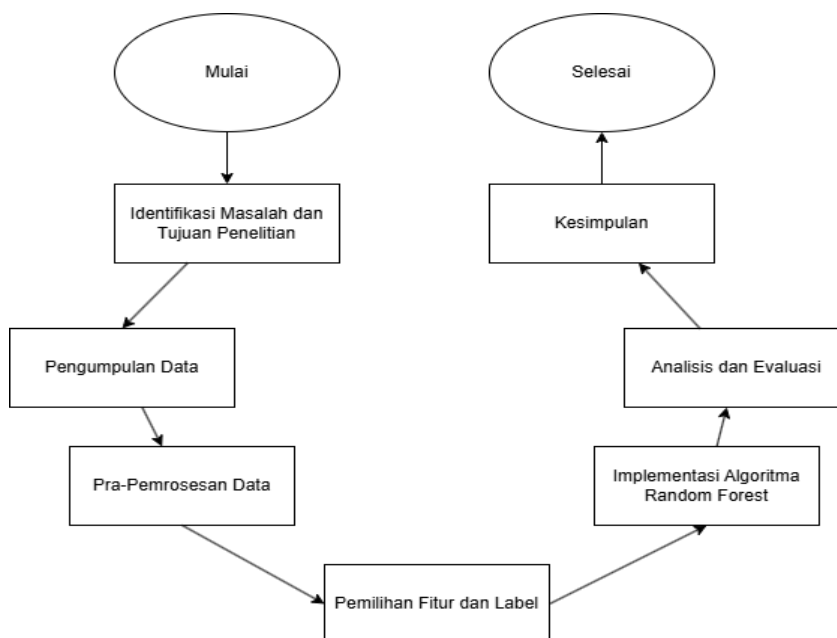
Curah hujan memiliki dampak yang dapat dilihat dari kedua sisi, baik positif maupun negatif. Di sisi positif, curah hujan menyediakan sumber air yang cukup, memberikan udara yang lebih segar, dan menjadikan tanah lebih subur, yang mendukung pertumbuhan tanaman. Namun, sisi negatifnya, curah hujan yang berlebihan dapat menyebabkan banjir dan mengganggu aktivitas masyarakat, terutama ketika hujan berlangsung dalam durasi panjang. Algoritma *Random Forest* merupakan salah satu metode yang efektif digunakan untuk analisis prediksi curah hujan. Dengan menggunakan algoritma ini, kita dapat memprediksi kemungkinan terjadinya hujan atau tidak, yang dapat membantu dalam perencanaan dan mitigasi terhadap dampak buruk curah hujan yang berlebihan [20].

Penelitian ini bertujuan untuk menganalisis dan memprediksi curah hujan di Kecamatan Purwodadi dengan penerapan algoritma *Random Forest*. Melalui penelitian ini, diharapkan dapat diperoleh prediksi yang lebih tepat mengenai curah hujan di wilayah tersebut, yang nantinya dapat digunakan untuk mendukung perencanaan yang lebih baik. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang lebih akurat dalam menghadapi potensi risiko bencana banjir. Dengan demikian, dapat membantu dalam pengambilan keputusan yang lebih efisien dalam upaya mitigasi bencana dan perencanaan yang terkait dengan manajemen risiko di wilayah Purwodadi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Berdasarkan uraian masalah pada bagian pendahuluan, penelitian ini akan melakukan analisis prediksi curah hujan di Kecamatan Purwodadi menggunakan algoritma *Random Forest* [21]. Hasil klasifikasi curah hujan akan dikelompokkan menjadi dua kategori, yaitu tidak hujan dan hujan. Implementasi algoritma *Random Forest* akan dilakukan di Google Colab dengan menggunakan bahasa pemrograman Python [22]. Alur penelitian untuk analisis prediksi curah hujan dimulai dengan identifikasi masalah dan penentuan tujuan penelitian, yang dijelaskan dalam bagian pendahuluan. Tahap berikutnya adalah pengumpulan data, diikuti oleh pra-pemrosesan data untuk memastikan data siap digunakan. Selanjutnya, dilakukan pemilihan fitur dan label yang relevan sebagai input bagi model. Setelah itu, masuk ke tahap implementasi algoritma untuk memprediksi curah hujan. Hasil prediksi kemudian dianalisis dan dievaluasi pada tahap analisis dan evaluasi, sebelum akhirnya disimpulkan pada bagian kesimpulan. Alur penelitian ini dapat dilihat secara visual pada Gambar 1.



Gambar 1. Alur Penelitian

2.2 Pengumpulan Data

Pada Pada tahap pengumpulan data, penelitian ini memanfaatkan data dari *NASA Prediction Of Worldwide Energy Resources (POWER)* [13], yang menyajikan berbagai parameter cuaca seperti tekanan udara, suhu, kecepatan angin, dan kelembapan. Data ini mencakup informasi yang penting untuk analisis cuaca, namun fokus utama penelitian ini adalah pada data parameter curah hujan yang terakumulasi secara kontinu dari tahun 2000 hingga 2022. Data curah hujan yang diperoleh disajikan dalam format *Comma Separated Values (CSV)* dan diorganisir berdasarkan bulan, mulai dari Januari hingga Desember. Data yang lengkap dan terstruktur dengan baik ini akan digunakan sebagai dasar untuk analisis prediksi, dengan penerapan algoritma Random Forest, untuk memprediksi curah hujan di wilayah yang diteliti.

2.3 Pra-Pemrosesan Data

Tahap pra-pemrosesan memiliki tujuan utama untuk memastikan data yang digunakan dalam analisis bersih dan siap pakai, sehingga hasil yang diperoleh lebih akurat. Proses ini mencakup serangkaian langkah penting yang bertujuan untuk mengubah data mentah menjadi format yang terorganisir dan bebas dari ketidaksesuaian atau anomali yang dapat mengganggu analisis. Dengan begitu, tahap ini mempermudah proses selanjutnya. Beberapa langkah yang dilakukan meliputi:

- a. Pembersihan Data: Mengevaluasi data untuk mengatasi nilai yang hilang, outlier, atau inkonsistensi dalam data [11].
- b. Transformasi Data: Mengonversi data ke format yang sesuai agar bisa digunakan dalam algoritma analisis (misalnya normalisasi atau pengkategorian).
- c. Pembagian Data: Memisahkan data menjadi set pelatihan dan pengujian (misalnya 70% untuk pelatihan dan 30% untuk pengujian).

Dengan menjalani tahap pra-pemrosesan ini, kualitas data meningkat signifikan, memungkinkan algoritma dan model prediktif bekerja lebih optimal dan menghasilkan output yang lebih dapat dipercaya. Tahap ini sangat penting dalam proses analisis data karena data yang terorganisir dan terstruktur dengan baik menjadi dasar utama untuk keberhasilan analisis [23].

2.4 Pemilihan Fitur dan Label

Pada ada tahap ini, fitur dan label yang akan digunakan dalam penelitian prediksi curah hujan dengan algoritma *Random Forest* ditentukan. Fitur yang digunakan mencakup data parameter cuaca, seperti curah hujan, yang diambil dari bulan Januari hingga Februari antara tahun 2000 hingga 2022 di Kecamatan Purwodadi. Parameter cuaca yang dimasukkan adalah tekanan udara, kecepatan angin, suhu, dan kelembapan. Fitur-fitur ini dipilih karena dapat memberikan gambaran yang relevan tentang kondisi cuaca yang mempengaruhi curah hujan di wilayah tersebut. Sedangkan label yang digunakan dalam penelitian ini adalah dua kategori curah hujan, yaitu tidak hujan dan hujan. Penentuan fitur dan label ini sangat penting dalam membangun model, karena hal tersebut akan memastikan bahwa model yang dikembangkan dapat menghasilkan prediksi yang akurat dan dapat diandalkan. Dengan memilih fitur yang tepat dan label yang jelas, model akan lebih efektif dalam memprediksi kemungkinan terjadinya hujan berdasarkan data yang tersedia [24].

2.5 Implementasi Algoritma Random Forest

Tahap selanjutnya adalah implementasi algoritma Random Forest. Implementasi algoritma Random Forest akan menggunakan *google colab* dengan bahasa pemrograman *python*. *Google Colab*, atau *Colaboratory*, adalah platform berbasis *cloud* yang salah satu nya digunakan untuk melakukan implementasi *data mining* [25].

Random Forest adalah salah satu algoritma machine learning berbasis *ensemble* yang menggunakan kombinasi dari beberapa pohon keputusan untuk meningkatkan akurasi prediksi. Algoritma ini bekerja dengan cara membangun banyak pohon keputusan selama pelatihan dan memberikan hasil prediksi berdasarkan voting mayoritas untuk klasifikasi atau rata-rata untuk regresi. Dalam konteks prediksi curah hujan, *Random Forest* sangat cocok karena mampu menangani data yang kompleks dan variabel yang saling berkorelasi. Dengan menggunakan *dataset* curah hujan dari tahun 2000 hingga 2022, algoritma ini dapat digunakan untuk mengklasifikasikan apakah akan terjadi hujan atau tidak berdasarkan parameter cuaca seperti tekanan udara, suhu, kecepatan angin, dan kelembapan. Proses utama dalam algoritma *Random Forest* dimulai dengan pembentukan *bootstrap samples* dari data latih. Setiap pohon keputusan dibangun dari subset data ini dengan memilih sejumlah fitur secara acak untuk setiap split. Proses pemilihan ini mengurangi risiko *overfitting* [14].

Selama implementasi, algoritma *Random Forest* menghasilkan prediksi dengan dua langkah utama: *bagging (Bootstrap Aggregating)* dan voting. Untuk masalah klasifikasi curah hujan, setiap pohon memberikan prediksi apakah akan terjadi hujan atau tidak. Keputusan akhir ditentukan oleh mayoritas suara dari semua pohon. Selain itu, evaluasi model dilakukan dengan metrik seperti akurasi, *precision*, *recall*, *F1-score*, dan *ROC-AUC*. Hal ini penting untuk menilai seberapa baik model mampu memprediksi curah hujan dengan data uji [14].

Keunggulan *Random Forest* dalam analisis prediksi curah hujan adalah kemampuannya menangani *dataset* dengan dimensi tinggi, resistansi terhadap *overfitting*, serta fleksibilitas dalam menangani data kontinu maupun kategorikal. Implementasi algoritma ini di *Python* biasanya menggunakan pustaka *scikit-learn*, yang menyediakan



fungsi seperti *RandomForestClassifier* untuk klasifikasi. Dengan pengoptimalan parameter seperti jumlah pohon (*n_estimators*) dan kedalaman pohon maksimum (*max_depth*), *Random Forest* dapat memberikan prediksi yang sangat akurat dan dapat diandalkan[14].

Dataset yang digunakan dalam penelitian ini terdiri dari 193 data curah hujan, yang merupakan 70% dari total data. *Dataset* ini dialokasikan untuk pelatihan algoritma *Random Forest*. Data curah hujan yang dianalisis mencakup rentang waktu dari tahun 2000 hingga 2022 di Kecamatan Purwodadi, dengan empat parameter utama: tekanan udara, suhu, kelembapan, dan kecepatan angin. Pembagian *dataset* ini bertujuan untuk melatih model agar dapat mengidentifikasi pola curah hujan berdasarkan parameter cuaca tersebut.

Struktur *dataset* terbagi menjadi dua bagian utama. Bagian pertama adalah kolom fitur yang memuat data parameter cuaca sebagai variabel prediktor. Sementara itu, bagian kedua adalah kolom label yang berfungsi untuk klasifikasi hasil prediksi. Label untuk kelompok curah hujan dibagi menjadi dua kategori, yaitu "tidak hujan" yang diwakili dengan angka 0 dan "hujan" yang diwakili dengan angka 1. Pembagian label ini dirancang untuk mempermudah proses klasifikasi dalam model. Untuk evaluasi kinerja model, 30% dari *dataset*, sebanyak 83 data curah hujan, digunakan sebagai data pengujian. Data ini diterapkan untuk menilai akurasi model dalam memprediksi kelompok hujan dan tidak hujan. Pembagian *dataset* yang terstruktur ini memungkinkan algoritma *Random Forest* untuk melakukan klasifikasi dengan lebih efektif, sehingga hasil prediksi dapat diandalkan dalam analisis curah hujan.

2.6 Analisis dan Evaluasi

Evaluasi data adalah tahap akhir dalam proses klasifikasi dengan algoritma *Random Forest*. Pada tahap ini, *dataset* yang telah dikelompokkan berdasarkan kategori tertentu akan dianalisis untuk mengukur kualitas hasil klasifikasi. Evaluasi ini menggunakan berbagai metrik, seperti *recall*, presisi, *F1-score*, dan akurasi, yang berfungsi sebagai indikator kinerja algoritma. Langkah ini penting untuk memahami sejauh mana algoritma dapat memberikan prediksi yang andal terkait curah hujan berdasarkan data yang telah diproses[26].

Hasil evaluasi dari pengujian curah hujan di Kecamatan Purwodadi dengan algoritma *Random Forest* akan memberikan gambaran tentang tingkat akurasi, presisi, *recall*, dan *F1-score*. Selain itu, evaluasi ini juga akan menunjukkan hasil prediksi untuk kategori tidak hujan dan hujan, serta mengidentifikasi prediksi mana yang kurang akurat. Dengan informasi tersebut, penelitian diharapkan dapat memberikan kontribusi yang signifikan dalam memprediksi kemungkinan hujan di wilayah tersebut, sehingga mendukung perencanaan dan pengambilan keputusan yang lebih baik[14].

3. HASIL DAN PEMBAHASAN

Penelitian ini berfokus pada prediksi kemungkinan hujan di Kecamatan Purwodadi dengan menganalisis berbagai parameter cuaca yang memiliki pengaruh signifikan terhadap curah hujan. Parameter yang digunakan dalam analisis meliputi tekanan udara, suhu, kelembapan, dan kecepatan angin, yang masing-masing berkontribusi terhadap dinamika atmosfer yang memicu terjadinya hujan.

Tekanan udara berperan dalam memantau pergerakan sistem cuaca, sementara suhu memengaruhi proses penguapan dan tingkat kelembapan udara. Kelembapan mencerminkan jumlah uap air di atmosfer, dan kecepatan angin dapat memengaruhi pembentukan awan serta proses presipitasi. Dengan mengevaluasi keterkaitan antarparameter ini, penelitian bertujuan untuk membangun model prediksi yang dapat memperkirakan peluang terjadinya hujan dengan akurasi tinggi.

Hasil penelitian diharapkan dapat memberikan wawasan yang lebih mendalam mengenai faktor-faktor yang memengaruhi pola cuaca di Kecamatan Purwodadi. Selain itu, temuan ini juga diharapkan dapat membantu masyarakat dan pihak terkait dalam merencanakan aktivitas atau mengambil langkah mitigasi yang lebih efektif terkait potensi hujan di wilayah tersebut.

3.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan sumber data dari *NASA Prediction Of Worldwide Energy Resources (POWER)*[13]. *Dataset* yang digunakan mencakup rentang waktu dari tahun 2000 hingga 2022, sehingga memungkinkan analisis tren jangka panjang dan variasi cuaca di wilayah yang diteliti. Data ini dipilih karena kualitasnya yang terpercaya dan cakupannya yang luas, yang menyediakan berbagai parameter meteorologi yang relevan untuk penelitian terkait prediksi cuaca, khususnya prediksi curah hujan. Sampel dari data yang dikumpulkan dapat dilihat pada Tabel 1.

Tabel 1. Sampel *Dataset* Awal

PARAMETER	YEA R	JAN	FEB	MA R	AP R	MA Y	JUN	JUL	AU G	SEP	OC T	NO V	DE C
PS	2000	99.2	99.2	99.2	99.2	99.3	99.4	99.3	99.4	99.4	99.2	99.1	99.2
		1		3	2	6	1	9	2		9	6	1
PS	2001	99.1	99.1	99.2	99.3	99.3	99.3	99.4	99.4	99.3	99.3	99.3	99.4
		9	5	7		8	9	4	8	9	4	2	2



PS	2002	99.3	99.4	99.3	99.2	99.3	99.4	99.5	99.5	99.5	99.4	99.3	99.3
		3	6	3	9	9	3	3	4	6		6	9
T2M	2000	25.4	25.3	25.5	26.2	26.0	25.6	26.7	27.7	28.4	27.9	27.0	26.7
		8	9		8	9	3		4		2	8	7
T2M	2001	26.1	25.9	25.8	26.3	26.4	25.6	25.9	27.3	29.2	27.6	26.8	26.6
		4		9	5	1	6	1	4		9	9	4
T2M	2002	26.2	25.5	26.4	26.5	26.6	26.9	27.8	28.1	28.8	30.2	27.7	26.9
		2	3	1	8	7	2		2	9		2	7
QV2M	2000	18.1	18.0	18.1	18.3	18.0	16.5	15.1	14.2	15.3	16.4	17.7	17.3
		3	7	3	1	7	4	4	2	8	2		3
QV2M	2001	18.1	18.0	18.3	18.4	18.0	17.3	15.9	14.2	14.8	17.1	17.7	17.0
		9	7	7	3	1	3	9	2	3	5		3
QV2M	2002	18.3	18.2	18.4	18.5	17.7	15.8	14.1	12.5	12.8	13.7	17.7	18.4
		1	5	9	5		1		7	2	9	6	3
WS10M	2000	3.28	3.63	2.95	1.8	2.43	2.62	2.8	2.93	3.36	2.72	2.59	2.36
WS10M	2001	2.77	3.87	2.39	2.06	2.55	2.45	2.8	2.73	3.32	2.38	2.45	2.49
WS10M	2002	2.82	4.34	2.11	2.34	2.62	2.93	3.04	3.13	3.21	3.08	2.38	2.05
PRECTOTCORR	2000	5.27	5.27	5.27	0	5.27	0	0	0	0	0	5.27	5.27
PRECTOTCORR	2001	10.5	10.5	10.5	5.27	5.27	0	0	0	0	5.27	5.27	5.27
		5	5	5									
PRECTOTCORR	2002	5.27	10.5	5.27	0	0	0	0	0	0	0	10.5	0
			5									5	

Tabel 1 menampilkan sampel data dari dataset yang mencakup periode waktu antara tahun 2000 hingga 2022, diperoleh dari *NASA Prediction Of Worldwide Energy Resources (POWER)*. Dataset ini mengandung berbagai parameter cuaca yang relevan untuk analisis prediksi curah hujan di Kecamatan Purwodadi. Rentang waktu yang luas dalam *dataset* ini memberikan basis data yang kaya untuk mempelajari pola cuaca jangka panjang secara lebih mendalam.

Analisis *dataset* ini berfokus pada parameter cuaca utama, yaitu tekanan udara di permukaan (*PS*), suhu udara pada ketinggian dua meter (*T2M*), kelembapan spesifik pada ketinggian dua meter (*QV2M*), kecepatan angin pada ketinggian sepuluh meter (*WS10M*), dan curah hujan terkoreksi (*PRECTOTCORR*). Parameter-parameter ini dipilih karena pengaruhnya yang signifikan terhadap pola cuaca dan potensi presipitasi di wilayah tersebut. Dengan menggunakan parameter cuaca tersebut sebagai fitur analisis dan *PRECTOTCORR* sebagai label utama.

3.2 Pra-Pemrosesan Data

Data parameter curah hujan yang diperoleh akan melalui tahap pra-pemrosesan untuk mempermudah proses analisis dalam penelitian ini. Proses pra-pemrosesan bertujuan untuk menyusun ulang data agar lebih terstruktur dan siap untuk dianalisis lebih lanjut. Salah satu langkah utamanya adalah mengubah orientasi data dari format horizontal menjadi vertikal sehingga lebih mudah dipahami dan diolah.

Selain itu, beberapa elemen dalam *dataset*, seperti tahun dan nama bulan, dihapus karena tidak relevan dengan analisis prediksi. Nama atribut dalam dataset juga mengalami penyesuaian agar lebih deskriptif dan sesuai dengan tujuan penelitian. Sebagai contoh, parameter *PS* (*Surface Pressure*) diubah menjadi "tekanan udara," *T2M* (*Temperature at 2 Meters*) menjadi "suhu," *QV2M* (*Specific Humidity at 2 Meters*) menjadi "kelembapan," *WS10M* (*Wind Speed at 10 Meters*) menjadi "kecepatan angin," dan *PRECTOTCORR* (*Precipitation Corrected*) menjadi "curah hujan."

Hasil dari tahap pra-pemrosesan ini disusun ulang dalam format yang lebih bersih dan dapat digunakan untuk analisis berikutnya. Tabel 2 memuat data yang telah diproses, menunjukkan parameter cuaca yang telah disederhanakan dan diberi nama yang lebih intuitif. Proses ini memastikan data siap digunakan dalam model prediksi untuk menghasilkan hasil yang akurat dan dapat diandalkan.

Tabel 2. Sampel Dataset Setelah Pra-Pemrosesan Data

Tekanan Udara	Suhu	Kelembapan	Kecepatan Angin	Curah Hujan
99,26	25,87	18,19	2,07	5,27
99,28	26,11	18,07	2,16	0
99,45	25,74	16,91	2,8	0
99,47	26,37	16,05	2,75	0
99,47	27,02	13,67	3,27	0
99,44	27,94	14,53	3,27	0
99,46	28,37	14,77	3,06	0
99,41	27,73	17,4	2,59	5,27
99,25	26,61	18,43	2,05	10,55



Tekanan Udara	Suhu	Kelembapan	Kecepatan Angin	Curah Hujan
99,24	26,17	18,01	2,12	10,55
99,34	25,71	18,37	3,69	15,82
99,24	25,51	18,43	3,61	15,82
99,35	26,4	18,01	1,95	5,27
99,34	26,61	18,62	1,88	5,27
99,31	26,3	18,49	2,16	5,27
99,47	25,69	17,46	2,65	5,27
99,52	25,23	15,87	3,01	0
99,48	26,71	15,01	2,73	0
99,46	28,42	14,71	3,23	0
99,45	28,76	15,93	2,73	5,27
99,29	27,98	17,21	2,23	5,27
99,38	27,13	17,64	1,91	10,55
99,46	26,02	18,25	3,16	10,55
99,44	26,71	18,86	2,11	10,55
99,42	26,92	18,92	1,84	10,55
99,39	27,08	19,1	1,94	5,27
99,26	26,75	19,41	2,27	5,27
99,45	26,05	18,55	2,3	5,27
99,42	25,87	17,94	2,72	5,27
99,43	26,12	17,64	2,7	5,27
99,37	26,45	18,07	2,18	5,27
99,32	26,23	17,94	2,27	5,27
99,25	26,53	18,25	1,88	5,27
99,11	25,74	18,19	3,07	10,55
99,19	25,37	18,19	4,25	10,55
99,26	25,64	18,07	3,51	5,27
99,24	25,95	17,88	2,16	5,27

3.3 Pemilihan Fitur dan Label

Setelah tahap pra-pemrosesan selesai, data yang telah diproses akan diberi label untuk mengelompokkan kemungkinan terjadinya hujan. Proses pemberian label ini bertujuan untuk membedakan antara dua kategori, yaitu hujan dan tidak hujan, berdasarkan data yang telah disiapkan. Hasilnya, data yang sudah terstruktur dan diberi label ini dapat dilihat pada Tabel 3, yang menunjukkan pengelompokan berdasarkan prediksi curah hujan di wilayah yang diteliti.

Tabel 3. Pemilihan Fitur dan Label

Tekanan Udara	Suhu	Kelembapan	Kecepatan Angin	Curah Hujan	Hasil
99,26	25,87	18,19	2,07	5,27	1
99,28	26,11	18,07	2,16	0	0
99,45	25,74	16,91	2,8	0	0
99,47	26,37	16,05	2,75	0	0
99,47	27,02	13,67	3,27	0	0
99,44	27,94	14,53	3,27	0	0
99,46	28,37	14,77	3,06	0	0
99,41	27,73	17,4	2,59	5,27	1
99,25	26,61	18,43	2,05	10,55	1
99,24	26,17	18,01	2,12	10,55	1
99,34	25,71	18,37	3,69	15,82	1
99,24	25,51	18,43	3,61	15,82	1
99,35	26,4	18,01	1,95	5,27	1
99,34	26,61	18,62	1,88	5,27	1
99,31	26,3	18,49	2,16	5,27	1
99,47	25,69	17,46	2,65	5,27	1
99,52	25,23	15,87	3,01	0	0
99,48	26,71	15,01	2,73	0	0
99,46	28,42	14,71	3,23	0	0
99,45	28,76	15,93	2,73	5,27	1
99,29	27,98	17,21	2,23	5,27	1
99,38	27,13	17,64	1,91	10,55	1
99,46	26,02	18,25	3,16	10,55	1

Tekanan Udara	Suhu	Kelembapan	Kecepatan Angin	Curah Hujan	Hasil
99,44	26,71	18,86	2,11	10,55	1
99,42	26,92	18,92	1,84	10,55	1
99,39	27,08	19,1	1,94	5,27	1
99,26	26,75	19,41	2,27	5,27	1
99,45	26,05	18,55	2,3	5,27	1
99,42	25,87	17,94	2,72	5,27	1
99,43	26,12	17,64	2,7	5,27	1
99,37	26,45	18,07	2,18	5,27	1
99,32	26,23	17,94	2,27	5,27	1
99,25	26,53	18,25	1,88	5,27	1
99,11	25,74	18,19	3,07	10,55	1
99,19	25,37	18,19	4,25	10,55	1
99,26	25,64	18,07	3,51	5,27	1
99,24	25,95	17,88	2,16	5,27	1

Tabel 3 menyajikan parameter yang digunakan dalam penelitian, yang meliputi tekanan udara, suhu, kelembapan, dan kecepatan angin sebagai fitur, serta curah hujan sebagai penentu label. Fitur-fitur ini akan dianalisis untuk memprediksi kemungkinan terjadinya hujan di Kecamatan Purwodadi. Sedangkan curah hujan akan digunakan untuk menentukan label kelompok, dengan kelompok "tidak hujan" diberi angka 0 dan kelompok "hujan" diberi angka 1. Proses pemberian label dilakukan berdasarkan nilai curah hujan. Jika nilai curah hujan adalah 0, maka data tersebut dikelompokkan ke dalam kategori "tidak hujan", sedangkan jika nilai curah hujan lebih besar dari 0, data tersebut akan dikelompokkan sebagai "hujan". Kolom yang berisi hasil klasifikasi ini akan digunakan sebagai label dalam model prediksi yang dikembangkan.

3.4 Implementasi Algoritma Random Forest

Pada tahap ini, algoritma *Random Forest* akan diimplementasikan menggunakan *Google Colab* dengan *dataset* yang mencakup parameter curah hujan dari tahun 2000 hingga 2022. Proses ini dimulai dengan memuat *dataset* ke dalam *Google Colab*, di mana sampel dari 10 data pertama akan ditampilkan untuk memberikan gambaran awal mengenai struktur *dataset* yang digunakan.

Proses ini bertujuan untuk memastikan bahwa data telah berhasil dimasukkan dengan benar dan siap untuk diproses lebih lanjut. Menampilkan 10 data awal seperti pada gambar 2 akan membantu peneliti memverifikasi format data dan memastikan bahwa semua kolom dan nilai telah terorganisir dengan baik sebelum melanjutkan ke langkah analisis lebih lanjut.

	tekanan_udara	suhu	kelembapan	kecepatan_angin	curah_hujan	hasil
0	99.21	25.48	18.13	3.28	5.27	1
1	99.20	25.39	18.07	3.63	5.27	1
2	99.23	25.50	18.13	2.95	5.27	1
3	99.22	26.28	18.31	1.80	0.00	0
4	99.36	26.09	18.07	2.43	5.27	1
5	99.41	25.63	16.54	2.62	0.00	0
6	99.39	26.70	15.14	2.80	0.00	0
7	99.42	27.74	14.22	2.93	0.00	0
8	99.40	28.40	15.38	3.36	0.00	0
9	99.29	27.92	16.42	2.72	0.00	0

Gambar 2. Menampilkan 10 Data Awal

Setelah *dataset* berhasil dimasukkan, langkah selanjutnya adalah membagi data menjadi fitur dan label. Fitur yang digunakan meliputi tekanan udara, suhu, kelembapan, dan kecepatan angin, yang akan digunakan untuk memprediksi curah hujan. Sedangkan untuk label, angka 0 digunakan untuk menunjukkan kondisi "tidak hujan," sementara angka 1 menunjukkan kondisi "hujan" pada kolom hasil.

Pembagian data ini bertujuan untuk memisahkan informasi yang digunakan dalam prediksi (fitur) dan kategori yang ingin diprediksi (label). Proses ini akan memastikan bahwa model *Random Forest* dapat dilatih dengan baik untuk memprediksi kemungkinan hujan berdasarkan parameter cuaca yang ada, dan hasil pembagian tersebut dapat dilihat pada gambar 3.

```
x = rain[['tekanan_udara', 'suhu', 'kelembapan', 'kecepatan_angin']] # Fitur
y = rain['hasil'] # Label
```

Gambar 3. Penentuan Fitur dan Label

Dalam penerapan algoritma *Random Forest* untuk memprediksi curah hujan, langkah pertama yang harus dilakukan adalah membagi dataset yang masih terintegrasi menjadi dua bagian terpisah. Hal ini penting karena data perlu dibagi menjadi data latih dan data uji agar model dapat dilatih dengan baik dan diuji kinerjanya. Pembagian ini dilakukan dengan proporsi 70% untuk data latih dan 30% untuk data uji. Proses pembagian data ini memungkinkan model untuk belajar dari data latih dan kemudian dievaluasi menggunakan data uji yang terpisah, sehingga dapat memberikan gambaran yang lebih objektif mengenai performa model. Pembagian data ini akan terlihat pada gambar 4, yang menunjukkan alokasi data latih dan data uji secara jelas.

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=42)
```

Gambar 4. Pemisahan Data Latih dan Uji

Kode pada gambar 5 adalah inisialisasi sebuah model klasifikasi menggunakan algoritma *Random Forest*, yang diterapkan melalui objek *RandomForestClassifier* dari pustaka *sklearn.ensemble* di Python. Pada kode ini, *n_estimators=100* menunjukkan bahwa model *Random Forest* akan menggunakan 100 pohon keputusan (*decision trees*) untuk melakukan klasifikasi. Setiap pohon keputusan akan dilatih dengan data yang berbeda untuk meningkatkan ketahanan dan akurasi model secara keseluruhan. Sementara itu, parameter *random_state=42* digunakan untuk memastikan bahwa hasil yang diperoleh dapat direproduksi. Dengan kata lain, pengaturan ini memberikan kontrol terhadap hasil acak yang dihasilkan oleh algoritma, memastikan bahwa setiap kali model dijalankan dengan parameter yang sama, hasil yang konsisten akan didapat.

```
rfc = RandomForestClassifier(n_estimators=100, random_state=42)
```

Gambar 5. Fungsi *Random Forest*

Gambar 6 merupakan perintah untuk melatih model *Random Forest* yang telah dibuat sebelumnya dengan menggunakan *dataset* pelatihan. Fungsi *fit()* pada objek *rfc* digunakan untuk melatih model dengan data yang disediakan. Parameter *x_train* berisi fitur atau variabel *input* (seperti tekanan udara, suhu, kelembapan, dan kecepatan angin), sementara *y_train* berisi label atau target *output* yang sesuai (seperti kategori hujan atau tidak hujan). Dengan menjalankan *rfc.fit(x_train, y_train)*, model *Random Forest* akan belajar pola hubungan antara fitur dan label dari data pelatihan untuk dapat melakukan prediksi pada data yang belum terlihat sebelumnya.

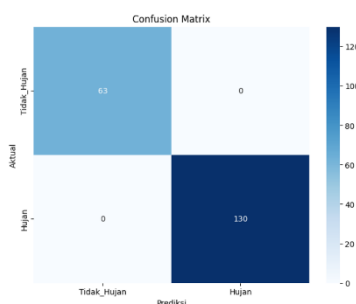
```
rfc.fit(x_train, y_train)
```

Gambar 6. Melatih Model

3.5 Analisis dan Evaluasi

Hasil implementasi algoritma *Random Forest* dalam penelitian ini menunjukkan kinerja yang memuaskan dengan tingkat akurasi yang tinggi pada kedua dataset, baik data latih maupun data uji. Pada data latih, model mencapai akurasi 100%, yang mengindikasikan bahwa model berhasil mempelajari pola dari data latih dengan sangat baik, sehingga dapat mengklasifikasikan data dengan akurat sesuai dengan label yang telah ditentukan.

Sementara itu, saat diuji dengan data uji yang belum pernah dilihat sebelumnya, model memperoleh akurasi sebesar 92%. Hasil ini menunjukkan bahwa *Random Forest* memiliki kemampuan untuk menggeneralisasi dengan baik dan memprediksi kelas data yang tidak terdapat dalam data latih. Ini memberikan indikasi bahwa algoritma *Random Forest* dapat diandalkan untuk memprediksi curah hujan berdasarkan parameter cuaca seperti tekanan udara, suhu, kelembapan, dan kecepatan angin. Untuk mengevaluasi lebih lanjut kinerja model, *confusion matrix* juga digunakan, yang hasilnya dapat dilihat pada gambar 7.



Gambar 7. Hasil *Confusion Matrix*

Kode yang ditunjukkan pada gambar 8 digunakan untuk menampilkan laporan klasifikasi dari model, baik untuk data latih maupun data uji. Pada baris pertama, perintah `print("Laporan klasifikasi training:")` mencetak judul "Laporan klasifikasi training" yang menunjukkan bahwa laporan berikutnya berfokus pada evaluasi model terhadap data latih. Di baris kedua, fungsi `classification_report()` dari pustaka `scikit-learn` digunakan untuk menghasilkan dan menampilkan laporan kinerja model berdasarkan label asli dari data latih (`y_train`) dan hasil prediksi (`y_predict_train`). Argumen `target_names=['Tidak_Hujan', 'Hujan']` digunakan untuk memberikan label yang lebih jelas pada kategori 0 dan 1.

Kemudian, pada baris ketiga, perintah `print("Laporan klasifikasi testing:")` mencetak judul "Laporan klasifikasi testing" untuk menandakan bahwa laporan selanjutnya berfokus pada evaluasi model terhadap data uji. Baris terakhir kembali menggunakan fungsi `classification_report()` untuk menghasilkan laporan klasifikasi untuk data uji, dengan menggunakan label asli (`y_test`) dan hasil prediksi (`y_predict_test`).

```
print("Laporan klasifikasi training:")
print(classification_report(y_train, y_predict_train, target_names=['Tidak_Hujan', 'Hujan']))
print("Laporan klasifikasi testing:")
print(classification_report(y_test, y_predict_test, target_names=['Tidak_Hujan', 'Hujan']))
```

Gambar 8. Laporan Klasifikasi Data Latih dan Uji

Hasil laporan klasifikasi dapat dilihat pada gambar 9, untuk data latih menunjukkan bahwa model memiliki performa yang sangat baik dengan nilai *precision*, *recall*, dan *f1-score* sebesar 1.00 untuk kedua kategori "Tidak_Hujan" dan "Hujan". Hal ini menunjukkan bahwa model berhasil memprediksi semua data dengan benar tanpa kesalahan, menghasilkan akurasi 100%. Secara keseluruhan, baik untuk kategori "Tidak_Hujan" maupun "Hujan", model memiliki kemampuan klasifikasi yang sempurna pada data latih, yang tercermin pada nilai *accuracy*, *macro avg*, dan *weighted avg* yang semuanya bernilai 1.00.

Sementara itu, untuk data uji, meskipun akurasi menurun menjadi 92%, model masih menunjukkan kinerja yang cukup baik. Untuk kategori "Tidak_Hujan", *precision* sebesar 0.81 dan *recall* 0.96, yang berarti model memiliki ketepatan yang lebih rendah dalam memprediksi "Tidak_Hujan", tetapi sangat baik dalam mengidentifikasinya. Sementara untuk kategori "Hujan", model memperoleh *precision* sebesar 0.98 dan *recall* 0.89, yang menandakan bahwa model lebih baik dalam mengidentifikasi hujan dan memiliki akurasi yang sangat tinggi dalam memprediksi "Hujan". Nilai *f1-score* yang dihasilkan untuk kedua kategori juga menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Secara keseluruhan, meskipun ada penurunan sedikit dalam performa dibandingkan data latih, model tetap dapat diandalkan dengan hasil yang baik pada data uji.

Laporan klasifikasi training:				
	precision	recall	f1-score	support
Tidak_Hujan	1.00	1.00	1.00	63
Hujan	1.00	1.00	1.00	130
accuracy			1.00	193
macro avg	1.00	1.00	1.00	193
weighted avg	1.00	1.00	1.00	193
Laporan klasifikasi testing:				
	precision	recall	f1-score	support
Tidak_Hujan	0.81	0.96	0.88	27
Hujan	0.98	0.89	0.93	56
accuracy			0.92	83
macro avg	0.90	0.93	0.91	83
weighted avg	0.93	0.92	0.92	83

Gambar 9. Hasil Laporan Klasifikasi Data Latih dan Uji

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa implementasi algoritma *Random Forest* berhasil diterapkan untuk analisis prediksi curah hujan di Kecamatan Purwodadi. Penelitian ini menggunakan beberapa parameter cuaca sebagai fitur, yaitu tekanan udara, suhu, kelembapan, dan kecepatan angin, sementara curah hujan digunakan sebagai label untuk menentukan kemungkinan hujan, yang diklasifikasikan menjadi dua kelompok: "hujan" dan "tidak hujan." Proses implementasi algoritma *Random Forest* dimulai dengan pembagian dataset menjadi data latih dan data uji, di mana 70% data digunakan untuk pelatihan dan 30% untuk pengujian. Model dilatih menggunakan 193 data latih dan diuji dengan 83 data uji. Hasil pengujian model menunjukkan akurasi yang sangat baik, dengan



akurasi sebesar 100% pada data latih, yang mengindikasikan bahwa model mampu mengklasifikasikan semua data latih dengan tepat. Sementara itu, pada data uji, model menghasilkan akurasi sebesar 92%, yang menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dalam memprediksi curah hujan pada data yang tidak terlihat sebelumnya. Selain itu, laporan klasifikasi untuk data latih dan data uji menunjukkan performa yang baik dengan nilai *precision*, *recall*, dan *f1-score* yang tinggi, baik untuk kategori "hujan" maupun "tidak hujan." Hasil penelitian ini membuktikan bahwa algoritma *Random Forest* dapat diterapkan dengan efektif untuk memprediksi curah hujan di Kecamatan Purwodadi berdasarkan parameter cuaca yang relevan. Dengan tingkat akurasi yang tinggi, model ini dapat menjadi alat yang bermanfaat untuk memprediksi kondisi cuaca dan mendukung perencanaan serta upaya mitigasi terkait curah hujan di wilayah tersebut. Meskipun demikian, penelitian ini memiliki keterbatasan dalam hal cakupan data dan wilayah yang terbatas. Oleh karena itu, disarankan untuk memperluas penelitian dengan menambahkan data dari wilayah lain dan mencoba algoritma klasifikasi lain untuk membandingkan kinerja serta akurasi prediksi lebih lanjut.

REFERENCES

- [1] W. W. G. Ameto *et al.*, "Perhitungan Debit Limpasan Pada Cv. Klasigi Kabupaten Sorong Provinsi Papua Barat," *INTAN J. Penelit. Tambang*, vol. 7, no. 1, pp. 1–6, 2024.
- [2] Marthin Luter Laia and Yudi Setyawan, "Perbandingan Hasil Klasifikasi Curah Hujan Menggunakan Metode SVM dan NBC," *J. Stat. Ind. dan Komputasi*, vol. 5, no. 2, pp. 51–61, 2020.
- [3] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [4] R. F. Firdaus, "Prediksi Curah Hujan Menggunakan Metode Long Short Term Memory (Studi Kasus : Kota Bandung) Prediksi Curah Hujan Menggunakan Metode Long Short Term Memory (Studi Kasus : Kota Bandung)," 2022.
- [5] A. Luthfiarta, A. Febriyanto, H. Lestiawan, and W. Wicaksono, "Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 10–17, 2020, doi: 10.33633/joins.v5i1.2760.
- [6] S. S. Sanusi W., *Statistika Untuk Pemodelan Data Curah Hujan*, vol. 53, no. 9, 2016.
- [7] B. L. Al Zakina, Z. Zainuddin, T. B. Santoso, and L. Safitri, "Penyuluhan Mitigasi Bencana Banjir Sebagai Upaya Kesiapsiagaan Bencana Di Kota Purwodadi," *J. Hum. Educ.*, vol. 4, no. 4, pp. 80–84, 2024, doi: 10.31004/jh.v4i4.1147.
- [8] M. I. Ananta, L. M. Limantara, J. S. Fidari, and H. Nurdin, "Analisa Curah Hujan Rancangan Di Daerah Aliran Sungai Bendungan Manik Kabupaten Kupang," *J. Tek. Sipil*, vol. 13, no. 1, pp. 67–78, 2024.
- [9] A. S. Nugroho and A. G. Daniamiseno, "Pengembangan E-Book Mitigasi Bencana Gunung Api Berbasis Prinsip-Prinsip Desain Pesan Pembelajaran untuk Siswa Sekolah Menengah Pertama," *J. Inov. Teknol. Pendidik.*, vol. 9, no. 1, pp. 114–122, 2022, doi: 10.21831/jitp.v9i1.21690.
- [10] Y. Setiawan, E. P. Purwandari, A. Wijanarko, and E. Sunandi, "Pemetaan Zonasi Rawan Banjir Dengan Analisis Indeks Rawan Banjir Menggunakan Metode Fuzzy Simple Adaptive Weighting," *Pseudocode*, vol. 7, no. 1, pp. 78–87, 2020, doi: 10.33369/pseudocode.7.1.78-87.
- [11] P. Rahayu *et al.*, *Buku Ajar Data Mining*, vol. 1, no. January 2024. PT. Sonpedia Publishing Indonesia, 2024.
- [12] A. Fahraini, D. Achmad Rusdiansyah, J. A. Jenderal Yani Km, and K. Selatan, "Analisis Keandalan Metode Analisa Frekuensi Dan Intensitas Hujan Berdasarkan Data Curah Hujan Klimatologi Banjarbaru," *Sustain. Technol. Journal*, vol. 9, no. 1, pp. 11–23, 2020, [Online]. Available: <http://jtb.ulm.ac.id/index.php/JTB>
- [13] NASA, "NASA Prediction Of Worldwide Energy Resources (POWER) | Data Access Viewer (DAV)," <https://power.larc.nasa.gov/>. Accessed: Oct. 16, 2024. [Online]. Available: <https://power.larc.nasa.gov/data-access-viewer/>
- [14] M. Haditama Rifan, "Analisis Dan Pembuatan Dashboard Prediksi Kelulusan Mahasiswa Menggunakan Metode Random Forest, Naïve Bayes Dan Support Vector Machine," *Anal. Dan Pembuatan Dashboard Prediksi Kelulusan Mhs. Menggunakan Metod. Random For. Naïve Bayes Dan Support Vector Mach.*, p. 1159, 2023.
- [15] A. H. Harahap, I. M. B. A. Malik, M. I. N. Imam, M. T. S. Bilhaq, A. A. Nur, and S. L. D. Agustini, "Klasifikasi Diagnosa Penyakit Jantung menggunakan Algoritma Random Forest," *Tek. Inform. UIN Sunan Gunung Djati Bandung*, vol. 3, pp. 1–51, 2021.
- [16] H. Hidayat, E. Azzahra, F. Bioresita, and K. Pronojiwo, "Analisa Perubahan Tutupan Lahan Akibat Letusan Gunung Semeru Tahun 2021 dengan Algoritma Random Forest (Studi Kasus : Kecamatan Pronojiwo , Kabupaten Lumajang) Pendahuluan Indonesia merupakan negara dengan luas wilayah sekitar 1, 905 juta km² serta ter," vol. 19, no. 2, pp. 236–245, 2024.
- [17] A. Fauzan and D. Ahmad, "Analisis Hasil Prediksi Magnitudo Gempa Di Wilayah Kota Padang Menggunakan Teknik Random Forest," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 4, no. 3, pp. 1569–1576, 2023, doi: 10.46306/lb.v4i3.450.
- [18] A. Y. Isnaeni and S. Y. J. Prasetyo, "Klasifikasi Wilayah Potensi Risiko Kerusakan Lahan Akibat Bencana Tsunami Menggunakan Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 1, pp. 33–42, 2022, doi: 10.28932/jutisi.v8i1.4056.
- [19] I. Kurniawan, D. C. P. Buani, A. Abdussomad, W. Apriliah, and R. A. Saputra, "Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 421–428, 2023, doi: 10.25126/jtiik.20231026225.
- [20] N. H. Putri, R. H. Dananjaya, and N. S. Surjandari, "Aplikasi Machine Learning Method pada Pemetaan Kerawanan Tanah Longsor di Kabupaten Karangayar," *Media Komun. Dunia Ilmu Sipil*, vol. 6, no. 1, pp. 23–34, 2024, [Online]. Available: <https://journal.univetbantara.ac.id/index.php/modulus/article/view/4490>
- [21] N. A. Prakoso Indaryono, "Analisa Perbandingan Algoritma Random Forest Dan Naïve Bayes Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Di Indonesia," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 1, pp. 158–167, 2024, doi: 10.29100/jupi.v9i1.4421.
- [22] A. S. Syuhada, A. M. Simanullang, D. S. Lewa, and S. Jefry Marthin, "Makalah Pembelajaran Mesin (Machine Learning),"



Makal. Pembelajaran Mesin (machine Learn.), pp. 1–11, 2021.

- [23] Y. N. R. Putro, A. Afriansyah, and R. Bagaskara, “Penggunaan Algoritma Gaussian Naïve Bayes & Decision Tree Untuk Klasifikasi Tingkat Kemenangan Pada Game Mobile Legends,” *JUKI J. Komput. dan Inform.*, vol. 6, no. 1, pp. 10–26, 2024, doi: 10.53842/juki.v6i1.472.
- [24] D. Prasetyawan and R. Gatra, “Algoritma K-Nearest Neighbor untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan dan Ekonomi,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 56–67, 2022, doi: 10.14421/jiska.2022.7.1.56-67.
- [25] Google, “Google Colab,” <https://colab.google/>. Accessed: Oct. 21, 2024. [Online]. Available: <https://colab.google/>
- [26] P. Kurniawan and H. W. Nugroho, “Implementasi Data Mining Dalam Klasifikasi Tingkat Kesenjangan Kompetensi PNS Menggunakan Metode Naive Bayes,” vol. 6, no. 2, pp. 835–846, 2024, doi: 10.47065/bits.v6i2.5641.