

**ANALISIS DAN PREDIKSI CURAH HUJAN KABUPATEN
BREBES DENGAN *RANDOM FOREST REGRESSION*
BERBASIS *APACHE SPARK***

**Yusdi Nugraha¹, Nur Ariesanto Ramdhan², Otong Saeful
Bachri³**

Universitas Muhadi Setiabudi Brebes

E-mail: nugrahayusdi903@gmail.com¹, ariesantoramdhan@gmail.com²,
otongsaifulbahriumus@gmail.com³

Abstrak

Curah hujan merupakan salah satu unsur meteorologi yang memiliki pengaruh terhadap sektor pertanian, pengelolaan sumber daya air, serta mitigasi bencana hidrometeorologi. Kabupaten Brebes memiliki aktivitas pertanian yang cukup tinggi sehingga perubahan pola curah hujan menjadi salah satu faktor yang perlu diperhatikan. Penelitian ini bertujuan untuk menganalisis pola curah hujan Kabupaten Brebes serta membangun model prediksi curah hujan satu hari berikutnya menggunakan algoritma Random Forest Regression berbasis Apache Spark MLlib. Data yang digunakan berupa data meteorologi harian periode 2020–2025 dengan jumlah 2.192 observasi. Variabel penelitian meliputi curah hujan, suhu udara, kelembapan, tekanan udara, dan kecepatan angin. Feature engineering dilakukan dengan membentuk fitur waktu, lag, dan moving average sehingga diperoleh 16 fitur masukan. Pembagian data dilakukan secara kronologis, yaitu data tahun 2020–2023 sebagai training, tahun 2024 sebagai validation, dan tahun 2025 sebagai testing. Optimasi parameter dilakukan terhadap 18 kombinasi Random Forest Regression. Konfigurasi terbaik diperoleh dengan 100 pohon, kedalaman maksimum 5, dan minimum 1 data pada setiap node. Evaluasi model pada data testing menghasilkan MAE sebesar 6,7697 mm, RMSE sebesar 13,9653 mm, dan R^2 sebesar 0,2891. Hasil tersebut menunjukkan bahwa model mampu mengikuti pola perubahan curah hujan secara moderat dan memberikan performa lebih baik dibandingkan metode baseline yang digunakan. Namun, model masih mengalami kesulitan dalam memperkirakan kejadian hujan dengan intensitas tinggi dan ekstrem. Dengan demikian, Random Forest Regression berbasis Apache Spark dapat digunakan sebagai pendekatan untuk analisis dan estimasi curah hujan Kabupaten Brebes, tetapi masih diperlukan pengembangan lebih lanjut untuk meningkatkan kemampuan prediksi terhadap kejadian ekstrem.

Kata Kunci : Curah Hujan, *Random Forest Regression*, *Machine Learning*, *Apache Spark*, Prediksi Cuaca.

1. PENDAHULUAN

Curah hujan merupakan salah satu parameter meteorologi yang memiliki pengaruh besar terhadap berbagai sektor kehidupan, seperti pertanian, pengelolaan sumber daya air, perencanaan wilayah, dan mitigasi bencana hidrometeorologi. Perubahan pola curah hujan yang terjadi secara tidak menentu dapat menyebabkan berbagai permasalahan, seperti banjir akibat intensitas hujan tinggi dan kekeringan akibat rendahnya curah hujan. Informasi mengenai kondisi dan prediksi curah hujan diperlukan untuk membantu proses pengambilan keputusan pada berbagai bidang yang berkaitan dengan kondisi iklim [1].

Indonesia merupakan negara beriklim tropis yang memiliki karakteristik curah hujan dengan tingkat variasi yang tinggi. Pola curah hujan di Indonesia dipengaruhi oleh berbagai faktor atmosfer, seperti suhu udara, kelembapan udara, tekanan udara, kecepatan angin, serta fenomena iklim global seperti *El Niño Southern Oscillation* (ENSO) dan *La Niña*. Kompleksitas hubungan antarvariabel tersebut menyebabkan pola curah hujan memiliki sifat *non-linear* dan sulit diprediksi menggunakan metode statistik sederhana. Kondisi tersebut menjadikan prediksi curah hujan sebagai salah satu permasalahan yang banyak dikaji dalam bidang meteorologi dan pengolahan data [2].

Prediksi curah hujan merupakan proses memperkirakan nilai curah hujan pada periode mendatang berdasarkan data historis yang tersedia. Informasi hasil prediksi dapat dimanfaatkan untuk berbagai kebutuhan, seperti menentukan waktu tanam yang sesuai, mengoptimalkan penggunaan sumber daya air, meningkatkan kesiapsiagaan terhadap bencana banjir, serta membantu pengambilan keputusan berbasis kondisi iklim. Namun, karakteristik curah hujan yang memiliki pola musiman, fluktuasi tinggi, dan hubungan kompleks antarvariabel menyebabkan metode prediksi membutuhkan pendekatan yang mampu mempelajari pola data secara lebih mendalam [3].

Kristiana & Sholeha [4] menyebutkan bahwa Kabupaten Brebes merupakan salah satu wilayah di Provinsi Jawa Tengah yang memiliki sektor pertanian sebagai salah satu penggerak utama perekonomian daerah. Kabupaten ini dikenal sebagai salah satu sentra produksi bawang merah terbesar di Indonesia. Berdasarkan data Badan Pusat Statistik Kabupaten Brebes, aktivitas pertanian menjadi sektor penting bagi masyarakat karena sebagian besar wilayahnya memiliki lahan pertanian produktif. Kondisi tersebut menyebabkan sektor pertanian Kabupaten Brebes memiliki ketergantungan yang tinggi terhadap faktor iklim, terutama ketersediaan air dan pola curah hujan.

Salah satu algoritma *Machine Learning* yang banyak digunakan untuk prediksi adalah *Random Forest*. *Random Forest* merupakan algoritma *ensemble learning* yang menggabungkan beberapa *decision tree* untuk menghasilkan model prediksi yang lebih stabil. Algoritma ini memiliki kemampuan dalam menangani data dengan banyak variabel, mampu memodelkan hubungan *non-linear*, serta memiliki risiko *overfitting* yang lebih rendah dibandingkan penggunaan satu *decision tree* [5].

Pada penelitian prediksi curah hujan, metode yang sesuai adalah *Random Forest Regression* karena variabel yang diprediksi berupa nilai numerik kontinu dalam satuan milimeter (mm). Berbeda dengan *Random Forest Classification* yang menghasilkan keluaran berupa kategori tertentu, *Random Forest Regression* mampu menghasilkan nilai prediksi berupa jumlah curah hujan berdasarkan hubungan antara variabel masukan dan target. Oleh karena itu, metode ini dipilih untuk membangun model prediksi curah hujan Kabupaten Brebes.

Penelitian mengenai prediksi curah hujan menggunakan *Machine Learning* telah banyak dilakukan, tetapi sebagian besar penelitian masih berfokus pada penerapan algoritma tanpa mengkaji karakteristik lokal suatu wilayah secara spesifik. Selain itu, penelitian yang melakukan analisis pola dan prediksi nilai curah hujan Kabupaten Brebes menggunakan *Random Forest Regression* berbasis *Apache Spark* masih terbatas. Oleh karena itu, diperlukan penelitian yang mampu mengembangkan model prediksi curah hujan berdasarkan karakteristik data meteorologi lokal Kabupaten Brebes.

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk menganalisis pola curah hujan dan membangun model prediksi curah hujan Kabupaten Brebes menggunakan algoritma *Random Forest Regression* berbasis *Apache Spark*. Model yang dikembangkan

akan dievaluasi menggunakan *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan Koefisien Determinasi (R^2) untuk mengetahui kemampuan model dalam menghasilkan prediksi curah hujan. Berdasarkan uraian tersebut, penelitian ini mengambil judul: " Analisis Dan Prediksi Curah Hujan Kabupaten Brebes Dengan *Random Forest Regression* Berbasis *Apache Spark* ".

2. TINJAUAN PUSTAKA

Analisis

Analisis merupakan proses pengolahan dan pemeriksaan suatu data atau informasi untuk memperoleh pola, hubungan, serta karakteristik tertentu yang dapat digunakan sebagai dasar dalam pengambilan keputusan. Dalam bidang teknologi informasi, analisis data dilakukan dengan menerapkan berbagai metode komputasi untuk menemukan informasi yang tersembunyi dari sekumpulan data dalam jumlah besar. Menurut Sarker [6], analisis berbasis *Machine Learning* memungkinkan proses identifikasi pola pada data melalui pembelajaran dari data historis sehingga dapat menghasilkan informasi yang lebih bermakna. Proses analisis tidak hanya berfokus pada pengumpulan data, tetapi juga mencakup tahapan pengolahan, eksplorasi, interpretasi, dan evaluasi hasil untuk memperoleh kesimpulan yang sesuai dengan tujuan penelitian.

Curah Hujan

Curah hujan merupakan salah satu unsur meteorologi yang menunjukkan jumlah air yang jatuh dari atmosfer menuju permukaan bumi dalam bentuk presipitasi pada periode waktu tertentu [7]. Curah hujan menjadi salah satu parameter penting dalam kajian atmosfer karena berhubungan langsung dengan kondisi iklim, siklus hidrologi, ketersediaan air, pertanian, serta potensi terjadinya bencana hidrometeorologi seperti banjir dan kekeringan. Besarnya curah hujan umumnya dinyatakan dalam satuan milimeter (mm), yaitu tinggi air hujan yang terkumpul pada permukaan datar apabila tidak terjadi proses penguapan, infiltrasi, maupun aliran air [8].

Prediksi

Prediksi merupakan suatu proses yang dilakukan untuk memperkirakan kejadian yang kemungkinan akan terjadi di masa yang akan datang berdasarkan data dan informasi yang tersedia saat ini maupun di masa lalu. Kegiatan prediksi biasanya dilakukan dengan menggunakan pendekatan ilmiah, khususnya melalui metode statistik dan matematis, guna menghasilkan estimasi yang lebih akurat.

Tujuan utama dari prediksi adalah untuk meminimalkan tingkat kesalahan dalam peramalan (*forecast error*). Untuk mengukur tingkat akurasi tersebut, digunakan berbagai indikator evaluasi seperti *Standard Error Estimate* (SEE), *Mean Absolute Percentage Error* (MAPE), *Absolute Error*, *Relative Error*, serta *Root Mean Squared Error* (RMSE). Indikator-indikator ini berfungsi sebagai alat untuk menilai sejauh mana hasil prediksi mendekati nilai sebenarnya [9].

Machine Learning

Machine Learning merupakan salah satu cabang dari kecerdasan buatan (*Artificial Intelligence*) yang memungkinkan sistem komputer untuk belajar dari data tanpa harus diprogram secara eksplisit [10]. Dengan memanfaatkan algoritma tertentu, komputer dapat mengenali pola dalam data dan menggunakannya untuk membuat keputusan atau prediksi.

Random Forest

Random Forest merupakan salah satu algoritma dalam metode *ensemble learning* yang digunakan untuk meningkatkan akurasi prediksi dengan menggabungkan banyak model

decision tree. Algoritma ini bekerja dengan cara membangun sejumlah pohon keputusan dari subset data yang berbeda melalui teknik *bootstrap sampling*, serta memilih subset fitur secara acak pada setiap proses pembentukan pohon [11].

Apache Spark

Apache Spark merupakan salah satu kerangka kerja (*framework*) komputasi terdistribusi yang dirancang untuk memproses data dalam skala besar secara cepat dan efisien. *Apache Spark* dikembangkan oleh AMPLab di University of California, Berkeley, dan saat ini dikelola oleh *Apache Software Foundation*. *Spark* menjadi populer karena kemampuannya dalam melakukan pemrosesan data secara *in-memory computing*, yang membuatnya jauh lebih cepat dibandingkan dengan pendekatan tradisional berbasis disk seperti *Hadoop MapReduce* [12].

3. METODE PENELITIAN

Penelitian ini menggunakan jenis penelitian kuantitatif dengan pendekatan eksperimen (*experimental research*). Penelitian kuantitatif digunakan karena data yang digunakan berupa data numerik, yaitu data historis curah hujan Kabupaten Brebes yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Meteorologi Maritim Tegal. Data tersebut dianalisis menggunakan metode komputasi untuk mengetahui pola perubahan curah hujan serta membangun model prediksi. Ada beberapa skema penelitian tersebut dapat dilihat pada Gambar 1.

Gambar 1. Skema Penelitian

Identifikasi Masalah ↓	1. Tingginya variasi curah hujan di Kabupaten Brebes. 2. Kesulitan dalam memprediksi curah hujan secara akurat menggunakan metode konvensional. 3. Dibutuhkan metode prediksi yang cepat, akurat, dan mampu menangani data dalam skala besar.
Pengumpulan Data ↓	1. Data curah hujan (BMKG) 2. Data suhu udara 3. Data kelembapan udara 4. Data tekanan udara 5. Data kecepatan angin 6. Data periode tahun 2020–2025
Preprocessing Data ↓	1. Data Cleaning 2. Penanganan Missing Value 3. Transformasi/Normalisasi Data 4. Feature Selection 5. Pembagian Data Training dan Testing
Metode ↓	Random Forest Berbasis Apache Spark
Proses Random Forest ↓	1. Input Data Training 2. Bootstrap Sampling 3. Pembentukan Decision Tree 4. Pemilihan Fitur Secara Acak 5. Pembentukan Random Forest (Ensemble) 6. Prediksi Curah Hujan
Evaluasi Model ↓	1. Mean Absolute Error (MAE) 2. Root Mean Square Error (RMSE) 3. R² Score (Coefficient of Determination)
Hasil	1. Prediksi Curah Hujan 2. Perbandingan Data Aktual dan Hasil Prediksi 3. Visualisasi Hasil Prediksi 4. Analisis Kinerja Model Random Forest Berbasis Apache Spark

Diagram tersebut menampilkan sejumlah langkah atau tahapan dalam proses analisis dan prediksi curah hujan kabupaten brebes dengan *Random Forest Regression* berbasis *Apache Spark*, yang dapat dijelaskan sebagai berikut:

1. Pengumpulan Data (Dataset)

Pengumpulan data merupakan salah satu tahapan penting dalam penelitian karena kualitas data yang digunakan akan mempengaruhi hasil analisis dan prediksi yang dihasilkan oleh model. Pada penelitian ini, data yang digunakan merupakan data sekunder yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data tersebut digunakan sebagai dasar dalam membangun model prediksi curah hujan menggunakan metode *Random Forest* yang diimplementasikan pada platform *Apache Spark*.

Data yang dikumpulkan meliputi beberapa parameter meteorologi yang memiliki keterkaitan terhadap terjadinya curah hujan. Parameter tersebut terdiri atas data curah hujan, suhu udara, kelembapan udara, tekanan udara, dan kecepatan angin. Seluruh data dikumpulkan dalam periode tahun 2020 sampai dengan tahun 2025 sehingga mampu menggambarkan kondisi cuaca secara historis dan dapat digunakan sebagai data pelatihan (training) maupun data pengujian (testing) pada proses pemodelan.

2. Preprocessing

Preprocessing data merupakan tahap awal dalam pengolahan Dataset yang bertujuan untuk memastikan data memiliki kualitas yang baik sebelum digunakan dalam proses analisis dan pembangunan model prediksi. Tahapan *preprocessing* diperlukan karena data meteorologi yang diperoleh dari sumber eksternal dapat memiliki permasalahan seperti data kosong (*missing value*), data tidak sesuai format, data ganda (*duplicate data*), maupun nilai yang tidak konsisten.

Pada penelitian ini, *preprocessing* dilakukan menggunakan *Apache Spark* melalui library *PySpark*. Tahapan *preprocessing* dilakukan untuk menyiapkan data meteorologi Kabupaten Brebes agar dapat digunakan sebagai input dalam algoritma *Random Forest Regression*.

Berdasarkan pemeriksaan awal terhadap dataset, diperoleh sebanyak 2.192 pengamatan harian dengan rentang waktu 1 Januari 2020 sampai 31 Desember 2025. Seluruh tanggal pengamatan bersifat unik, tidak ditemukan nilai kosong, dan tidak ditemukan tanggal duplikat. Dengan demikian, tidak diperlukan penghapusan data akibat *missing value* maupun duplikasi pada tahap pemeriksaan awal.

3. Model Random Forest

Metode *Random Forest* adalah salah satu algoritma *Machine Learning* yang termasuk dalam kategori *ensemble learning*. *Ensemble learning* melibatkan penggabungan hasil dari beberapa model untuk meningkatkan kinerja dan ketepatan prediksi dibandingkan dengan penggunaan satu model tunggal. Dalam konteks *Random Forest*, model yang digunakan adalah pohon keputusan (*decision trees*).

4. Pelatihan Model

Proses pelatihan model dilakukan menggunakan algoritma *Random Forest Regression* berbasis *Apache Spark MLlib* dengan memanfaatkan 16 fitur yang telah dibentuk melalui tahap *feature engineering*. Fitur tersebut terdiri atas variabel meteorologi, fitur waktu, fitur lag, dan moving average yang digunakan untuk memprediksi curah hujan satu hari berikutnya. Data dibagi secara kronologis untuk mempertahankan urutan waktu, yaitu data tahun 2020–2023 digunakan sebagai data training, sedangkan tahun 2024 digunakan sebagai data validation dan tahun 2025 sebagai data testing. Pada tahap pelatihan, dilakukan pencarian parameter terhadap 18 kombinasi model yang berasal dari variasi jumlah pohon,

kedalaman maksimum pohon, dan jumlah minimum data pada setiap node. Pemilihan konfigurasi terbaik dilakukan berdasarkan nilai *RMSE* terendah pada data validation. Hasil pencarian memperoleh konfigurasi optimal berupa 100 pohon (*numTrees*), kedalaman maksimum 5 (*maxDepth*), dan minimum 1 data pada setiap node (*minInstancesPerNode*), dengan *maxBins* sebesar 32, *featureSubsetStrategy* *auto*, serta *seed* 42. Setelah parameter terbaik diperoleh, model dilatih kembali menggunakan gabungan data training tahun 2020–2023 dan validation tahun 2024 sebanyak 1.819 pengamatan sebelum digunakan untuk menghasilkan prediksi pada data tahun 2025.

5. Pengujian dan Evaluasi

Tahap pengujian dilakukan menggunakan data tahun 2025 yang terdiri atas 365 target harian dan tidak digunakan dalam proses pelatihan maupun pemilihan parameter model. Penggunaan data testing yang terpisah bertujuan untuk mengetahui kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah digunakan sebelumnya. Model final menghasilkan nilai prediksi curah hujan yang kemudian dibandingkan dengan nilai aktual pada periode pengujian. Selisih antara nilai aktual dan prediksi digunakan sebagai dasar untuk mengetahui tingkat kesalahan model. Selain menghasilkan nilai prediksi, hasil pengujian juga digunakan untuk menganalisis hubungan antara nilai aktual dan prediksi serta mengetahui kemampuan model dalam mengikuti perubahan curah hujan selama tahun 2025. Evaluasi performa model dilakukan menggunakan Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan Koefisien Determinasi (R^2). Berdasarkan pengujian pada data tahun 2025, model menghasilkan nilai MAE sebesar 6,7697 mm, yang menunjukkan rata-rata selisih absolut antara curah hujan aktual dan hasil prediksi sekitar 6,77 mm per hari. Nilai RMSE sebesar 13,9653 mm menunjukkan adanya beberapa kesalahan prediksi dengan nilai yang cukup besar, terutama pada kejadian curah hujan dengan intensitas tinggi. Sementara itu, nilai R^2 sebesar 0,2891 menunjukkan bahwa model mampu menjelaskan sekitar 28,91% variasi curah hujan aktual pada data pengujian. Selain ketiga metrik tersebut, diperoleh nilai bias sebesar -0,9397 mm dan korelasi sebesar 0,6361, yang menunjukkan adanya kecenderungan underprediction serta hubungan positif pada tingkat sedang antara nilai aktual dan prediksi.

4. HASIL DAN PEMBAHASAN

1. Gambaran Umum Dataset

Dataset penelitian diperoleh dari BMKG Stasiun Meteorologi Maritim Tegal dan merepresentasikan kondisi meteorologi harian wilayah Kabupaten Brebes. Periode data mencakup enam tahun penuh, yaitu 2020 sampai 2025. Tahun 2020 dan 2024 masing-masing memiliki 366 observasi karena merupakan tahun kabisat, sedangkan tahun lainnya memiliki 365 observasi. Dengan demikian, jumlah keseluruhan data adalah 2.192 baris dan tidak terdapat tanggal yang terlewat pada rentang penelitian.

Tabel 1 Struktur Dataset Penelitian

No.	Nama Kolom	Tipe Data Setelah Validasi	Satuan/Keterangan
1	Tanggal	Date	Tanggal pengamatan
2	curah_hujan_mm	Double	Curah hujan harian (mm)
3	suhu_udara_c	Double	Suhu udara (°C)
4	kelembapan_pct	Double	Kelembapan udara (%)
5	tekanan_hpa	Double	Tekanan udara (hPa)

No.	Nama Kolom	Tipe Data Setelah Validasi	Satuan/Keterangan
6	kecepatan_angin_ms	Double	Kecepatan angin (m/s)
7	kategori_hujan	String	Kategori intensitas hujan

Sumber: Hasil pengolahan data, 2026

2. Hasil Validasi dan Preprocessing Data

Validasi awal dilakukan menggunakan Apache Spark dengan membaca seluruh kolom sebagai string. Langkah ini diperlukan karena file CSV menggunakan tanda titik koma sebagai pemisah kolom dan tanda koma sebagai desimal. Setelah data terbaca, tanda desimal dikonversi menjadi titik, lima kolom numerik diubah menjadi tipe Double, dan kolom tanggal dikonversi menjadi tipe Date dengan pola yyyy-MM-dd. Spasi pada kategori hujan juga dibersihkan agar penamaan kategori konsisten.

Tabel 2 Hasil Pemeriksaan Kualitas Data

Aspek Pemeriksaan	Hasil	Keterangan
Jumlah data	2.192 baris	Sesuai jumlah hari 2020–2025
Rentang tanggal	01-01-2020 s.d. 31-12-2025	Rentang lengkap dan berurutan
Nilai kosong	0	Tidak ditemukan missing value
Tanggal ganda	0	Tidak ditemukan duplikasi tanggal
Baris duplikat	0	Tidak ditemukan baris identik
Nilai negatif	0	Tidak ada nilai numerik negatif
Konsistensi kategori	5 kategori	Tidak hujan sampai hujan sangat lebat

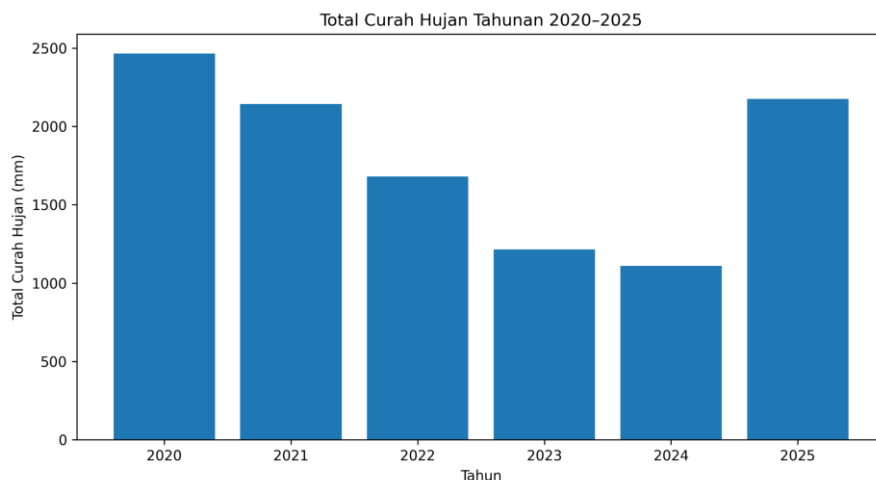
Sumber: Hasil pengolahan data, 2026

3. Pola Curah Hujan Tahunan

Tabel 3 Ringkasan Curah Hujan Tahunan

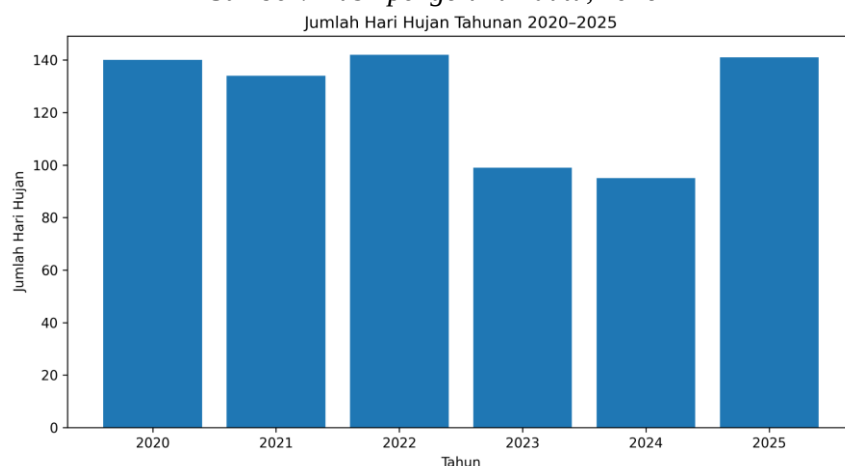
Tahun	Data	Total Hujan (mm)	Rata-rata Harian (mm)	Maksimum (mm)	Hari Hujan	Tanpa Hujan
2020	366	2.465,60	6,74	106,93	140	226
2021	365	2.142,12	5,87	124,97	134	231
2022	365	1.679,43	4,60	100,08	142	223
2023	365	1.214,92	3,33	50,04	99	266
2024	366	1.109,20	3,03	82,04	95	271
2025	365	2.175,36	5,96	124,97	141	224

Sumber: Hasil pengolahan data, 2026



Gambar 2 Total Curah Hujan Tahunan Periode 2020–2025

Sumber: Hasil pengolahan data, 2026



Gambar 3 Jumlah Hari Hujan Tahunan Periode 2020–2025

Sumber: Hasil pengolahan data, 2026

Total curah hujan enam tahun mencapai 10.786,63 mm. Tahun 2020 memiliki total curah hujan tertinggi sebesar 2.465,60 mm, sedangkan nilai terendah terjadi pada 2024 sebesar 1.109,20 mm. Pola tahunan menunjukkan penurunan bertahap dari 2020 sampai 2024, kemudian meningkat kembali secara tajam pada 2025 menjadi 2.175,36 mm. Kenaikan pada 2025 juga diikuti peningkatan jumlah hari hujan menjadi 141 hari, mendekati kondisi 2020 dan 2022.

Jumlah hari hujan tertinggi terjadi pada 2022 sebanyak 142 hari, meskipun total curah hujannya tidak setinggi 2020. Hal ini menunjukkan bahwa banyaknya hari hujan tidak selalu menghasilkan total curah hujan terbesar karena intensitas setiap kejadian berbeda. Tahun 2024 hanya memiliki 95 hari hujan dan 271 hari tanpa hujan, sehingga menjadi periode paling kering dalam dataset. Temuan tersebut memperlihatkan adanya variasi antartahun yang kuat dan menguatkan penggunaan pembagian data berdasarkan waktu.

5. KESIMPULAN

Berdasarkan hasil penelitian, data meteorologi Kabupaten Brebes periode 2020–2025 menunjukkan pola musiman yang cukup kuat. Curah hujan relatif tinggi pada periode November hingga Maret, sedangkan periode Juni hingga September cenderung lebih kering.

Data juga memiliki distribusi yang tidak seimbang karena sebagian besar observasi merupakan hari tanpa hujan dan hanya sebagian kecil yang termasuk kejadian hujan lebat maupun sangat lebat. Algoritma *Random Forest Regression* berbasis *Apache Spark MLlib* berhasil diterapkan untuk memprediksi curah hujan satu hari berikutnya. Proses pemodelan menggunakan 16 fitur yang mencakup kondisi meteorologi, informasi waktu, nilai lag, dan moving average. Setelah dilakukan pengujian terhadap 18 kombinasi parameter, konfigurasi terbaik diperoleh dengan 100 pohon, kedalaman maksimum 5, dan minimum satu data pada setiap node.

Hasil pengujian pada data tahun 2025 menghasilkan *MAE* sebesar 6,7697 mm, *RMSE* sebesar 13,9653 mm, dan R^2 sebesar 0,2891. Model menunjukkan kinerja yang lebih baik dibandingkan baseline rata-rata dan persistence. Meskipun demikian, model masih memiliki keterbatasan dalam memprediksi kejadian curah hujan ekstrem karena distribusi data didominasi oleh kondisi tanpa hujan dan intensitas rendah. Oleh karena itu, model lebih sesuai digunakan sebagai alat estimasi dan analisis pola daripada sebagai pengganti prakiraan cuaca resmi.

6. REFERENCES

- E. Surmaini and T. W. Hadi, “Verifikasi prediksi curah hujan *ensemble* menggunakan metode roc,” *J. Meteorol. dan Geofis.*, vol. 21, no. 1 SE-Article, pp. 37–44, Nov. 2020, doi: 10.31172/jmg.v21i1.618.
- A. S. Wahyuni, M. T. Edkayasa, S. Johan, S. H. Norfahmi, and L. Lizalidiawati, “Analisis Pengaruh Tekanan Udara, Kelembaban, dan Suhu Udara terhadap Curah Hujan di Kota Bengkulu selama El Niño 2023,” *J. Kelaut. Indones. J. Mar. Sci. Technol.*, vol. 18, no. 2, pp. 109–117, 2025.
- A. S. Siddiq, “Optimasi peramalan harga pangan di sulawesi selatan menggunakan *machine learning* (studi perbandingan *Random Forest* dan *extreme gradient boosting*)= Optimizing Food Price Prediction in South Sulawesi with *Machine Learning* (A Comparative Study of Random Fore.” Universitas Hasanuddin, 2026.
- A. Kristiana and A. Sholeha, “Analisis Faktor-Faktor Yang Mempengaruhi Produksi Bawang Merah di Desa Sengon, Kecamatan Tanjung, Kabupaten Brebes,” *JECMER J. Econ. Manag. Entrepreneursh. Res.*, vol. 1, no. 2, pp. 11–20, 2023.
- L. M. Harahap, C. B. Sitanggang, D. M. Tambunan, D. A. Pinem, and A. B. Hasugian, “Pengaruh Perubahan Iklim Terhadap Strategi Manajemen Agribisnis: Studi Kasus Di Wilayah Pertanian Indonesia (the Effect of Climate Change on Agribusiness Management Strategies: a Case Study in Indonesian Agriculture),” *J. Ilm. Ekon. Dan Manaj.*, vol. 3, no. 3, pp. 451–462, 2025.
- M. A. Syaharani, B. A. Dermawan, and R. I. Adam, “Prediksi Curah Hujan Kota Bogor Menggunakan Algoritma *Random Forest*,” *J. Sist. Inf. dan Apl.*, vol. 3, no. 1, pp. 52–62, 2025.
- H. Harisdianto, A. K. Affandi, M. Ariani, and S. Suhadi, “Downscaling spasial data curah hujan TRMM di wilayah Sumatera Selatan dengan algoritma artificial neural network dan *Random Forest Regression*,” *J. Penelit. Sains*, vol. 25, no. 3, pp. 283–290, 2023.
- G. Fibarkah, M. A. Tondang, N. Wafaa’Yulistyaningrum, and M. Afrad, “Prediksi Curah Hujan di Kabupaten Rembang dengan Model *Random Forest*,” in *Proceedings of the National Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative Media*, 2024, pp. 863–871.
- R. Torhino and P. N. Andono, “Penerapan Algoritma *Random Forest* dalam Prediksi Curah Hujan untuk Mendukung Analisis Cuaca,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 3, pp. 1688–1699, 2024.
- G. A. RAKHMAT and W. MUTOHAR, “Prakiraan Hujan menggunakan Metode *Random Forest* dan *Cross Validation*,” *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 8, no. 2, pp. 173–187, 2023.
- I. Hapsari and S. Pandya Wisesa, “Evaluasi Model Prediksi Curah Hujan Berbasis *Machine Learning*

- di Kota Bandung,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 11, no. 2 SE-, pp. 136–143, Sep. 2025, doi: 10.25077/TEKNOSI.v11i2.2025.136-143.
- A. Syahreza, N. K. Ningrum, and M. A. Syahrazy, “Perbandingan Kinerja Model Prediksi Cuaca: *Random Forest*, *Support Vector Regression*, dan *XGBoost*,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 2, pp. 526–534, 2024.