

Analisis dan Prediksi Curah Hujan Bulanan Kota Serang Berbasis Apache Spark Menggunakan Dataset BPS Provinsi Banten

Muhammad Rifqi Maulana^{1,*}, Azwa Fazilatunnisa², M. Yogi Febriansyah³, Alfi Muiz⁴, Irfan Fauzan⁵, Ahmad Tabrani⁶

^{1,2,3,4,5,6} Fakultas Sains dan Teknologi, Program Studi Informatika, UIN Sultan Maulana Hasanuddin Banten, Kota Serang, Indonesia

Email: ^{1,*}maulanarifqi741@gmail.com, ²azwanisaa01@gmail.com, ³yogi534@gmail.com, ⁴alfimuiz5@gmail.com,

⁵irfanfauzan458@gmail.com, ⁶Ahmad.tabrani@uinbanten.ac.id

(* Email Corresponding Author: maulanarifqi741@gmail.com)

Received: | Revision: | Accepted:

Abstrak

Curah hujan merupakan salah satu parameter iklim yang berperan penting dalam sektor pertanian, perencanaan tata kota, serta mitigasi bencana hidrometeorologi seperti banjir dan kekeringan. Namun, pengolahan dan analisis data curah hujan dalam rentang waktu panjang sering menghadapi tantangan akibat besarnya volume data dan kompleksitas pola yang terbentuk. Penelitian ini bertujuan untuk menganalisis pola curah hujan bulanan serta membangun model prediksi curah hujan di Kota Serang dengan memanfaatkan teknologi Big Data menggunakan Apache Spark dan dataset resmi dari Badan Pusat Statistik (BPS) Provinsi Banten periode 2005–2024. Proses penelitian meliputi tahap pembersihan data, analisis eksploratif, perhitungan klimatologi dan anomali, serta pemodelan prediksi menggunakan algoritma Linear Regression, Decision Tree, dan Random Forest. Kinerja model dievaluasi menggunakan Root Mean Square Error (RMSE), koefisien determinasi (R^2), dan Skill Score dengan baseline klimatologi dan baseline lag sebagai pembanding. Hasil penelitian menunjukkan bahwa model Random Forest memberikan performa terbaik dengan nilai RMSE sebesar 84,07 dan nilai R^2 sebesar 0,14, serta Skill Score tertinggi dibandingkan model lainnya. Temuan ini mengindikasikan bahwa Random Forest lebih mampu menangkap pola non-linear pada data curah hujan bulanan. Penerapan Apache Spark terbukti efektif dalam mendukung proses analisis dan pemodelan data iklim berskala besar, sehingga hasil penelitian ini diharapkan dapat menjadi referensi dalam pengembangan sistem prediksi curah hujan serta mendukung pengambilan keputusan di tingkat daerah.

Kata Kunci: Prediksi Curah Hujan, Apache Spark, Machine Learning, Random Forest

Abstract

Rainfall is a crucial climatic parameter that significantly influences agriculture, urban planning, and hydrometeorological disaster mitigation such as floods and droughts. However, analyzing long-term rainfall data often faces challenges due to large data volumes and complex temporal patterns. This study aims to analyze monthly rainfall patterns and develop a rainfall prediction model for Serang City using Big Data technology through Apache Spark and an official dataset from the Central Bureau of Statistics (BPS) of Banten Province covering the period 2005–2024. The research workflow includes data cleaning, exploratory data analysis, climatological and anomaly analysis, and predictive modeling using Linear Regression, Decision Tree, and Random Forest algorithms. Model performance is evaluated using Root Mean Square Error (RMSE), coefficient of determination (R^2), and Skill Score, with climatological and lag-based baselines as references. The results indicate that the Random Forest model outperforms other models, achieving the lowest RMSE value of 84.07 and the highest R^2 value of 0.14, along with the highest Skill Score compared to both baselines. These results demonstrate the capability of Random Forest in capturing non-linear patterns in monthly rainfall data. The implementation of Apache Spark proves effective for large-scale climate data processing and machine learning tasks. This study is expected to contribute as a reference for rainfall prediction research and to support decision-making processes related to climate-sensitive sectors at the regional level.

Keywords: Rainfall Prediction, Apache Spark, Machine Learning, Random Forest

1. PENDAHULUAN

Curah hujan merupakan salah satu parameter iklim yang memiliki peran sangat penting dalam berbagai sektor, khususnya pertanian, tata ruang perkotaan, pengelolaan sumber daya air[1], serta mitigasi bencana hidrometeorologi seperti banjir dan tanah longsor. Kota Serang sebagai ibu kota Provinsi Banten memiliki karakteristik iklim tropis dengan variasi curah hujan yang cukup signifikan sepanjang tahun. Perubahan pola curah hujan yang semakin sulit diprediksi akibat variabilitas iklim dan perubahan iklim global menuntut adanya metode analisis dan prediksi yang lebih akurat dan adaptif. Ketidakakuratan prediksi curah hujan dapat berdampak pada kesalahan perencanaan pembangunan, gangguan aktivitas pertanian, serta meningkatnya risiko bencana di wilayah perkotaan.

Permasalahan utama dalam analisis curah hujan saat ini tidak hanya terletak pada kompleksitas pola data iklim, tetapi juga pada volume data yang besar dan bersifat time series jangka panjang. Dataset curah hujan dari Badan Pusat Statistik (BPS) Provinsi Banten yang mencakup periode panjang memerlukan pendekatan komputasi yang mampu menangani data dalam skala besar secara efisien. Pengolahan data secara konvensional sering kali mengalami keterbatasan dalam hal performa, waktu komputasi, dan skalabilitas. Oleh karena itu, dibutuhkan pemanfaatan teknologi Big Data yang mampu melakukan pemrosesan data secara terdistribusi, salah satunya adalah Apache Spark.

Apache Spark merupakan framework komputasi Big Data yang dirancang untuk pemrosesan data berskala besar[2] dengan kecepatan tinggi melalui mekanisme in-memory processing. Spark banyak digunakan dalam analisis data iklim dan lingkungan karena kemampuannya dalam mengelola data time series, melakukan transformasi data, serta mengimplementasikan algoritma machine learning secara efisien. Integrasi Apache Spark dengan pendekatan machine learning memberikan peluang besar[3] untuk meningkatkan akurasi analisis pola historis dan prediksi curah hujan bulanan.

Sejumlah penelitian sebelumnya telah mengkaji pemodelan dan prediksi curah hujan menggunakan berbagai metode. Satriyogi Putramulyo *et al.* menggunakan pendekatan regresi dengan prediktor suhu[4] untuk memprediksi curah hujan bulanan di Kota Samarinda dan menunjukkan bahwa metode regresi mampu menangkap pola umum curah hujan, meskipun masih memiliki keterbatasan dalam menangani variabilitas ekstrem [5]. Jesi Pebralia *et al.* menerapkan regresi linier berganda berbasis Python untuk analisis curah hujan dan memperoleh hasil yang cukup baik pada data historis, namun belum membahas aspek skalabilitas pengolahan data [6].

Penelitian lain berfokus pada penggunaan algoritma machine learning. Aji dan Betha menerapkan algoritma Random Forest untuk prediksi curah hujan dan menunjukkan bahwa Random Forest memiliki performa yang lebih baik dibandingkan metode regresi konvensional dalam menangani data nonlinier [7]. Studi serupa juga dilakukan oleh Nicolaus *et al.* yang membandingkan Random Forest dengan Naïve Bayes, di mana Random Forest menunjukkan tingkat akurasi yang lebih tinggi dalam klasifikasi curah hujan [8]. Sementara itu, Mulyani *et al.* menggunakan pendekatan regresi untuk prediksi curah hujan di Kabupaten Majalengka dan menekankan pentingnya pemilihan model yang sesuai dengan karakteristik data lokal [9].

Dalam konteks yang lebih luas, Rahman *et al.* mengembangkan sistem prediksi curah hujan berbasis machine learning untuk mendukung konsep *smart city* dan menunjukkan bahwa kombinasi beberapa model machine learning dapat meningkatkan akurasi prediksi [10]. Selain itu, Rahutomo dan Harjito mengkaji penerapan machine learning untuk prediksi iklim di Indonesia dan menegaskan bahwa pendekatan berbasis data mampu memberikan hasil yang lebih adaptif dibandingkan metode statistik tradisional [11].

Meskipun berbagai penelitian tersebut telah membahas prediksi curah hujan dengan berbagai algoritma, masih terdapat *research gap* yang perlu dikaji lebih lanjut. Sebagian besar penelitian sebelumnya berfokus pada penerapan algoritma machine learning tanpa mengintegrasikannya dengan teknologi Big Data untuk pengolahan data berskala besar. Selain itu, kajian yang secara khusus menganalisis pola curah hujan bulanan Kota Serang dengan memanfaatkan Apache Spark serta membandingkan beberapa model prediksi seperti Regresi Linier, Random Forest, dan Decision Tree masih terbatas.

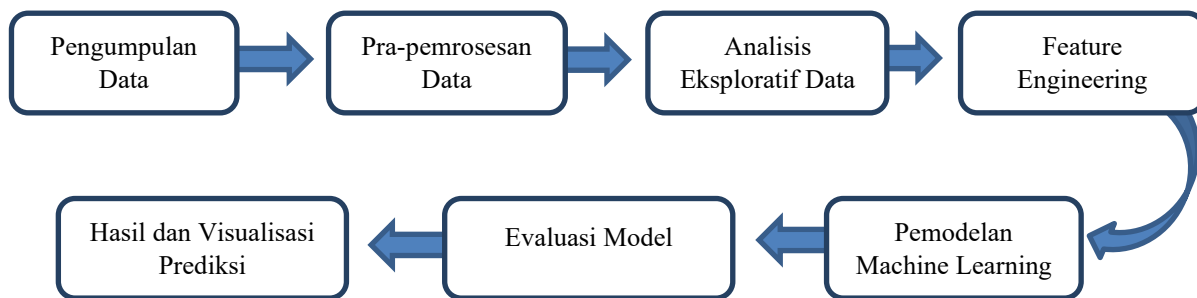
Berdasarkan latar belakang dan *gap analysis* tersebut, penelitian ini bertujuan untuk menganalisis pola historis curah hujan bulanan di Kota Serang menggunakan dataset BPS Provinsi Banten serta membangun model prediksi curah hujan berbasis machine learning dengan memanfaatkan Apache Spark. Penelitian ini juga bertujuan untuk mengevaluasi performa beberapa model prediksi menggunakan metrik evaluasi seperti Root Mean Square Error (RMSE), koefisien determinasi (R^2), dan *skill score*. Diharapkan hasil penelitian ini dapat memberikan kontribusi akademis dalam penerapan teknologi Big Data pada bidang klimatologi serta memberikan manfaat praktis bagi pemerintah daerah dalam mendukung pengambilan keputusan terkait pengelolaan sumber daya air dan mitigasi bencana.

2. METODOLOGI PENELITIAN

2.1 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini bersumber dari Badan Pusat Statistik (BPS) Provinsi Banten, khususnya data curah hujan bulanan Kota Serang. Data tersebut diperoleh melalui publikasi statistik resmi BPS yang dapat diakses secara terbuka, sehingga menjamin validitas dan keterlacakan data [12]. Rentang waktu data mencakup periode 2005 hingga 2024, dengan resolusi temporal bulanan. Variabel utama yang digunakan adalah curah hujan bulanan (mm), sedangkan variabel waktu direpresentasikan secara implisit melalui urutan bulan dalam deret waktu. Pemilihan data BPS dilakukan karena data telah melalui proses verifikasi statistik serta sering digunakan dalam penelitian hidrometeorologi di Indonesia.

2.2 Tahap Penelitian



Gambar 1. Diagram Alir Penelitian

Tahapan penelitian dimulai dari pengumpulan data curah hujan dari situs resmi BPS Provinsi Banten. Data yang diperoleh kemudian melalui proses pembersihan dan pra-pemrosesan, yang meliputi pengecekan nilai kosong, koreksi tipe data, serta penghapusan data tidak valid. Proses pra-pemrosesan ini bertujuan untuk meningkatkan kualitas data dan mengurangi potensi kesalahan dalam proses pemodelan, sebagaimana disarankan dalam penelitian prediksi curah hujan berbasis machine learning [13]. Setelah data bersih, dilakukan analisis eksploratif menggunakan Spark SQL untuk mengidentifikasi pola distribusi, tren musiman, dan fluktuasi curah hujan bulanan. Pada tahap ini juga dihitung nilai klimatologi dan anomali curah hujan untuk mengidentifikasi penyimpangan curah hujan terhadap kondisi normal. Nilai anomali curah hujan dihitung menggunakan persamaan:

$$\text{Anomali} = \text{Curah Hujan Aktual} - \text{Curah Hujan Klimatologi} \quad (1)$$

Analisis ini penting untuk memahami karakteristik data sebelum dilakukan pemodelan prediksi [14]. Setelah tahap eksplorasi, data dibagi menjadi data latih (*training*) dan data uji (*testing*) berdasarkan pendekatan deret waktu (*time-based split*). Data periode 2005–2022 digunakan sebagai data latih, sedangkan periode 2023–2024 digunakan sebagai data uji.

2.3 Pemodelan Machine Learning

Pemodelan prediksi dilakukan menggunakan Apache Spark MLlib dengan tiga algoritma utama, yaitu *Linier Regression*, *Decision Tree Regression*, dan *Random Forest Regression*. *Linier Regression* digunakan sebagai model dasar karena kesederhanaannya dalam menangkap hubungan linier antar variabel. *Decision Tree* digunakan untuk menangkap pola nonlinier dan interaksi antar fitur, sedangkan *Random Forest* diterapkan untuk meningkatkan akurasi dan stabilitas prediksi melalui pendekatan ensemble. Selain model *machine learning*, penelitian ini juga menggunakan model baseline berupa prediksi berbasis nilai lag atau rata-rata historis sebagai pembandingan kinerja. Penggunaan baseline penting untuk memastikan bahwa model prediktif yang dibangun benar-benar memberikan peningkatan performa dibandingkan pendekatan sederhana, sebagaimana disarankan dalam penelitian prediksi hidrologi dan iklim sebelumnya [15].

2.4 Model Evaluasi

Evaluasi kinerja model dilakukan menggunakan *Root Mean Square Error (RMSE)*, *koefisien determinasi (R²)*, dan *Skill Score*. RMSE digunakan untuk mengukur besar kesalahan prediksi secara absolut dan didefinisikan sebagai:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

Nilai R² digunakan untuk menilai kemampuan model dalam menjelaskan variasi data aktual, sedangkan Skill Score digunakan untuk membandingkan performa model prediksi terhadap model baseline. Skill Score dihitung berdasarkan selisih kesalahan model terhadap kesalahan baseline, sebagaimana dikemukakan oleh Onyutha [15]. Penggunaan ketiga metrik ini memberikan evaluasi yang lebih komprehensif terhadap akurasi dan keandalan model prediksi curah hujan.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Kinerja Model Berdasarkan RMSE

Tabel 1. Evaluasi Model Berdasarkan RMSE

Model	RMSE (mm)
Baseline Klimatologi	180.30
Baseline Lag-1	112.69
Linear Regression	96.04
Decision Tree	84.32
Random Forest	84.07

Hasil evaluasi menunjukkan bahwa model baseline klimatologi menghasilkan nilai RMSE tertinggi sebesar 180,30 mm, yang mengindikasikan bahwa pendekatan rata-rata klimatologis kurang mampu merepresentasikan variasi curah hujan bulanan secara akurat. Baseline lag-1 memberikan perbaikan dengan RMSE sebesar 112,69 mm, namun masih menunjukkan tingkat kesalahan yang relatif besar.

Model *machine learning* menunjukkan performa yang lebih baik. Linear Regression menghasilkan RMSE sebesar 96,04 mm, sementara Decision Tree dan Random Forest memberikan hasil paling optimal dengan RMSE masing-masing 84,32 mm dan 84,07 mm. Hasil ini menunjukkan bahwa model berbasis pohon lebih efektif dalam menangkap pola nonlinier pada data curah hujan dibandingkan model linier.

3.2 Analisis Koefisien Determinasi (R^2)

Tabel 2. Evaluasi Model Berdasarkan R^2

Model	R^2
Baseline Klimatologi	- 2.970
Baseline Lag-1	- 0.551
Linear Regression	- 0.126
Decision Tree	0.132
Random Forest	0.137

Nilai R^2 pada baseline klimatologi dan baseline lag-1 bernilai negatif, masing-masing sebesar -2,97 dan -0,55, yang menunjukkan bahwa kedua pendekatan baseline tidak mampu menjelaskan variabilitas data curah hujan dengan baik. Model Linear Regression menghasilkan nilai R^2 sebesar -0,13, yang masih menunjukkan keterbatasan dalam menjelaskan pola curah hujan. Sebaliknya, Decision Tree dan Random Forest menghasilkan nilai R^2 positif masing-masing sebesar 0,13 dan 0,14, yang menandakan bahwa kedua model tersebut mampu menangkap sebagian variasi data aktual. Meskipun nilai R^2 yang diperoleh belum tergolong tinggi, hasil ini dapat dimaklumi mengingat karakteristik data curah hujan yang memiliki fluktuasi tinggi dan dipengaruhi oleh banyak faktor eksternal.

3.3 Analisis Skill Score terhadap Baseline

Skill score digunakan untuk mengukur peningkatan kinerja model machine learning relatif terhadap model baseline. Hasil perhitungan menunjukkan bahwa seluruh model machine learning memiliki nilai skill score positif, baik terhadap baseline klimatologi maupun baseline lag-1, yang berarti seluruh model memberikan peningkatan kinerja dibandingkan pendekatan konvensional.

Tabel 3. Evaluasi Skill Score Model terhadap Baseline Klimatologi

Model	Skill vs Klim
Linear Regression	0.467
Decision Tree	0.532
Random Forest	0.534

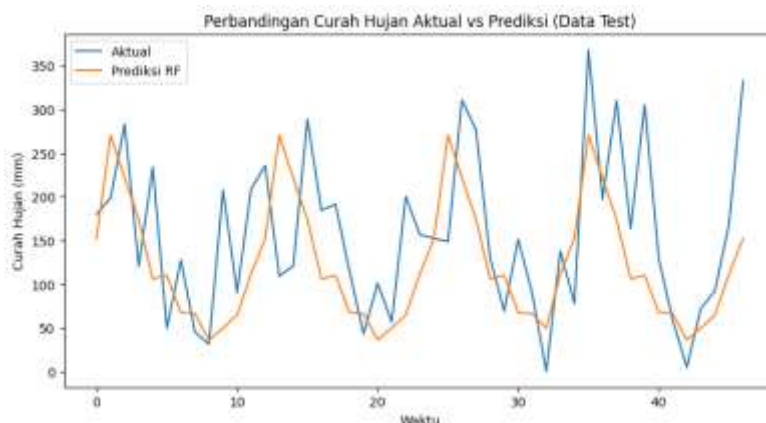
Pada Tabel 3. Terhadap baseline klimatologi, *Linear Regression*, *Decision Tree*, dan *Random Forest* masing-masing menghasilkan skill score sebesar 0,467, 0,532, dan 0,534. Nilai ini menunjukkan bahwa model *Random Forest* dan *Decision Tree* mampu meningkatkan akurasi prediksi hingga lebih dari 50% dibandingkan pendekatan klimatologi.

Tabel 4. Evaluasi Skill Score Model terhadap Baseline Lag-1

Model	Skill vs Lag-1
Linear Regression	0.147
Decision Tree	0.251
Random Forest	0.253

Sementara itu, Tabel 4. terhadap baseline lag-1, Random Forest kembali menunjukkan performa terbaik dengan skill score sebesar 0,253, diikuti oleh Decision Tree sebesar 0,251 dan Linear Regression sebesar 0,147. Hal ini menegaskan bahwa model ensemble berbasis pohon memberikan peningkatan yang konsisten dibandingkan pendekatan berbasis nilai historis sederhana.

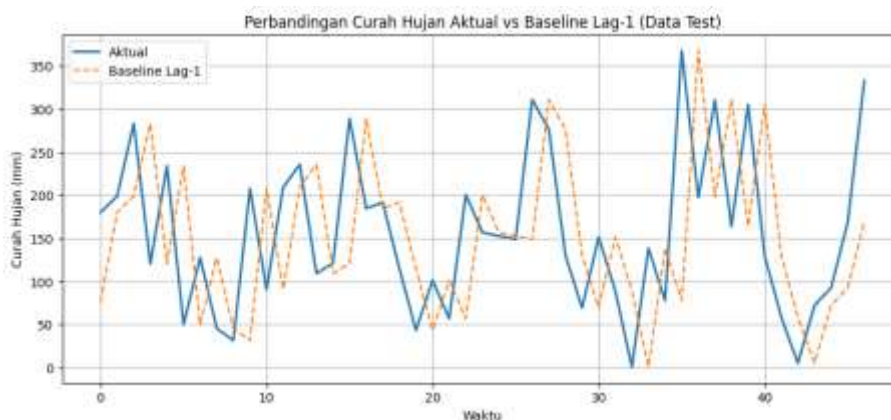
3.4 Analisis Visualisasi Hasil Prediksi Curah Hujan



Gambar 2. Perbandingan Curah Hujan Aktual dan Hasil Prediksi Menggunakan Model Random Forest (Data Uji)

Visualisasi hasil prediksi digunakan untuk mengevaluasi kemampuan model dalam menangkap pola curah hujan bulanan Kota Serang. Gambar 2. menampilkan perbandingan antara curah hujan aktual dan hasil prediksi menggunakan model Random Forest pada data uji. Sumbu horizontal merepresentasikan urutan waktu (bulan), sedangkan sumbu vertikal menunjukkan besaran curah hujan dalam satuan milimeter (mm). Secara umum, hasil prediksi Random Forest menunjukkan pola yang mengikuti tren curah hujan aktual, khususnya dalam menangkap gambaran musiman.

Meskipun demikian, pada beberapa periode tertentu terlihat adanya selisih antara nilai aktual dan nilai prediksi, terutama pada saat terjadi lonjakan curah hujan yang tinggi. Model Random Forest cenderung menghasilkan prediksi yang lebih halus sehingga belum sepenuhnya mampu merepresentasikan nilai ekstrem. Namun, secara keseluruhan model ini tetap menunjukkan performa yang stabil dan konsisten dalam memprediksi pola curah hujan bulanan.



Gambar 3. Perbandingan Curah Hujan Aktual dan Model Baseline Lag-1 (Data Uji)

Gambar 3. memperlihatkan perbandingan antara curah hujan aktual dan model baseline Lag-1, yang menggunakan nilai curah hujan bulan sebelumnya sebagai prediksi bulan berikutnya. Visualisasi menunjukkan bahwa model baseline ini sangat bergantung pada nilai sebelumnya sehingga sering mengalami keterlambatan dalam merespons perubahan curah hujan yang signifikan. Hal ini menyebabkan kesalahan prediksi meningkat pada periode dengan fluktuasi curah hujan yang tinggi.

Berdasarkan perbandingan kedua visualisasi tersebut, dapat disimpulkan bahwa model Random Forest memiliki kemampuan prediksi yang lebih baik dibandingkan model baseline Lag-1. Random Forest mampu menangkap pola historis dan musiman dengan lebih efektif, sedangkan model baseline hanya cocok sebagai pembanding dasar. Dengan dukungan Apache Spark dalam pengolahan data, pendekatan berbasis machine learning terbukti lebih unggul untuk analisis dan prediksi curah hujan bulanan di Kota Serang.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis pola curah hujan bulanan serta membangun model prediksi curah hujan di Kota Serang dengan memanfaatkan teknologi Big Data melalui Apache Spark dan dataset curah hujan dari BPS Provinsi Banten. Berdasarkan hasil analisis data historis, diperoleh bahwa pola curah hujan bulanan di Kota Serang menunjukkan karakteristik musiman yang jelas, dengan intensitas curah hujan tinggi pada awal dan akhir tahun serta penurunan signifikan pada pertengahan tahun. Pola ini menjadi dasar dalam proses pemodelan dan evaluasi prediksi curah hujan bulanan. Penggunaan Apache Spark terbukti efektif dalam menangani proses pembersihan data, transformasi, analisis eksploratif, hingga pemodelan machine learning secara terintegrasi dan efisien, sehingga sesuai untuk pengolahan data iklim yang berskala besar dan berulang. Dalam tahap pemodelan, tiga algoritma machine learning yaitu Linear Regression, Decision Tree, dan Random Forest dibandingkan dengan dua baseline, yaitu baseline klimatologi dan baseline lag. Hasil evaluasi menunjukkan bahwa seluruh model machine learning mampu memberikan kinerja yang lebih baik dibandingkan baseline, terutama Random Forest dan Decision Tree yang menghasilkan nilai RMSE terendah serta nilai koefisien determinasi (R^2) positif. Random Forest menjadi model dengan performa terbaik, ditunjukkan oleh nilai RMSE paling kecil, nilai R^2 tertinggi, serta nilai Skill Score yang paling besar baik terhadap baseline klimatologi maupun baseline lag. Hal ini mengindikasikan bahwa Random Forest lebih mampu menangkap pola non-linear dan variabilitas curah hujan dibandingkan model lainnya. Dengan demikian, penelitian ini menyimpulkan bahwa penerapan Apache Spark dan algoritma Random Forest efektif digunakan untuk analisis dan prediksi curah hujan bulanan di Kota Serang, serta berpotensi mendukung pengambilan keputusan di bidang perencanaan wilayah, pertanian, dan mitigasi risiko hidrometeorologi.

REFERENCES

- [1] Riski Efriandana, "Pengendalian Wireless Terpusat Dengan Capsman Dalam Mewujudkan Quality Of Service Jaringan Wireless," *J. Komput. Teknol. Inf. dan Sist. Inf.*, vol. 2, no. 2, pp. 400–408, 2023, doi: 10.62712/juktisi.v2i2.114.
- [2] Salbiah Salbiah and Muhammad Irwan Padli Nasution, "Analysis of Data Processing in Management Information Systems," *J. Ilm. Ekon. Dan Manaj.*, vol. 2, no. 12, pp. 01–12, 2024, doi: 10.61722/jiem.v2i11.3017.
- [3] M. F. Hajri, "Pengembangan kurikulum di Era Digital : Tantangan dan Peluang pada Abad 21," *J. Stud. Islam dan Hum.*, vol. 4, no. 1, pp. 33–41, 2023.
- [4] A. Khumaidi, R. Raafi'udin, and I. P. Solihin, "Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung," *J. Telemat.*, vol. 15, no. 1, pp. 13–18, 2020, doi: 10.61769/telematika.v15i1.340.
- [5] S. Putramulyo and S. Alaa, "Prediksi Curah Hujan Bulanan Di Kota Samarinda Menggunakan Persamaan Regresi Dengan Prediktor Data Suhu dan Kelembapan Udara," *Eig. Math. J.*, vol. 1, no. 2, pp. 13–16, 2018, doi: 10.29303/emj.v2i2.20.
- [6] J. Pebralia, "Analisis Curah Hujan Menggunakan Machine Learning Metode Regresi Linier Berganda Berbasis Python dan Jupyter Notebook Rainfall Analysis using Machine Learning-Multiple Linear Regression Method Based on Python and Jupyter Notebook," *J. Ilmu Fis. dan Pembelajarannya*, vol. 6, no. 2, pp. 23–30, 2022.
- [7] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.
- [8] N. A. Prakoso Indaryono, R. R. Saedudin, and F. Hamami, "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.
- [9] E. Dewi, S. Mulyani, I. Septianingrum, N. Nurjanah, and R. Rahmawati, “Prediksi Curah Hujan Di Kabupaten Majalengka Dengan Menggunakan Algoritma Regresi,” *e-Jurnal JUSITI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 8–1, no. 1, pp. 67–77, 2019, doi: 10.36774/jusiti.v8i1.602.
- [10] A. Rahman *et al.*, “Rainfall Prediction System Using Machine Learning Fusion for Smart Cities,” *MDPI*, pp. 1–51, 2022.
- [11] F. Rahutomo and B. Harjito, “Machine Learning-Based Climate Prediction in Indonesia: A Baseline Experiment,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 8, pp. 797–810, 2025, doi: 10.14569/IJACSA.2025.0160877.
- [12] Badan Pusat Statistik Provinsi Banten, “Curah Hujan menurut Bulan dan Stasiun Pengamatan di Provinsi Banten.” 2025.
- [13] C. D. Usman, A. P. Widodo, K. Adi, and R. Gernowo, “Rainfall prediction model in Semarang City using machine learning,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 2, pp. 1224–1231, 2023, doi: 10.11591/ijeecs.v30.i2.pp1224-1231.
- [14] F. Schiavio, D. Bonetta, and W. Binder, “Dynamic speculative optimizations for SQL compilation in apache spark,” *Proc. VLDB Endow.*, vol. 13, no. 5, pp. 754–767, 2020, doi: 10.14778/3377369.3377382.
- [15] C. Onyutha, “A hydrological model skill score and revised R-squared,” *Hydrol. Res.*, vol. 53, no. 1, pp. 51–64, 2022, doi: 10.2166/NH.2021.071.