



Perbandingan Algoritma *Machine Learning* (*Logistic Regression*, *SVM*, *KNN*, *Decision Tree*, *Random Forest*, dan *Gradient Boosting*) dalam Prediksi Hujan Harian di Provinsi Lampung

Ayu Aprilia^{1*}, Alka Budi Wahidin², Syafriadi³, Pulung Karo Karo⁴

^{1,3,4} Fisika, Universitas Lampung, Indonesia

² Sains Atmosfer dan Keplanetan, Institut Teknologi Sumatera, Indonesia

*Penulis Korespondensi: ayu.aprilia@fmipa.unila.ac.id

Abstract. In this era, rainfall prediction has become important for meteorological analysis, natural resource management, and hydrometeorological disaster mitigation. The development of rainfall prediction is aligned with modern predictive models, such as popular *Machine Learning* (ML) algorithms. Given these two phenomena, this study aims to evaluate six ML algorithms: *Logistic Regression* (LR), *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), *Decision Tree* (DT), *Random Forest* (RF), and *Gradient Boosting Machine* (GBM). Rainfall prediction combines data from three sources, namely ERA5, NASA POWER, and BMKG Lampung. The data was taken from three BMKG Lampung observation stations (Bandar Lampung, Pesawaran, and Kotabumi) for the period 2020-2025. Performance evaluation was conducted by assessing accuracy, precision, recall, *F1-Score*, confusion matrix, ROC curve, and AUC. The results of the study show that *Random Forest* achieves the best performance across all aspects, with a UAC value of 0.715. AUC value indicates that RF has stable predictions for very complex rainfall cases. The *Decision Tree* model showed low accuracy (0.66) and AUC values of 0.579. Ensemble-based multiple trees are more effective for predictive performance. These results highlight the importance of integrating various meteorological datasets from multiple sources using advanced ensemble learning methods to improve rainfall prediction accuracy and support climate-related decision-making at the regional level.

Keywords: *Ensamble Learning; Machine Learning; Meteorological Data; Rainfall Prediction; Satellite Validation*

Abstrak. Pada era saat ini prediksi hujan menjadi hal penting untuk analisis meteorologi, pengelolaan SDA hingga mitigasi bencana hidrometeorologi. Hal ini sejalan dengan model prediksi modern berupa algoritma *Machine Learning* (ML) yang populer. Melihat dua fenomena tersebut, studi ini bertujuan mengevaluasi 6 algoritma ML yaitu *Logistic Regression* (LR), *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), *Decision Tree* (DT), *Random Forest* (RF), dan *Gradient Boosting Machine* (GBM). Prediksi hujan mengkolaborasikan data dari 3 sumber yaitu ERA5, NASA POWER dan BMKG Lampung. Data diambil dari 3 stasiun observasi BMKG Lampung (Bandar Lampung, Pesawaran dan Kotabumi) pada periode tahun 2020-2025. Evaluasi performa dilakukan dengan melihat nilai akurasi, presisi, *recall*, *F1-Score*, *confusion matrix* serta kurva ROC dan AUC. Hasil penelitian memperlihatkan bahwa *Random Forest* memberikan performa terbaik dari seluruh aspek dengan nilai UAC sebesar 0.715. Hal ini menjadi indikasi bahwa RF memiliki prediksi yang stabil untuk kasus hujan yang sangat kompleks. Pada model *Decision Tree* menunjukkan akurasi dan AUC rendah yaitu 0,66 dan 0,579. Dapat disimpulkan bahwa model yang menggunakan ansambel berbasis banyak pohon lebih efektif dalam performa prediksi. Hasil ini menyampaikan pentingnya integrasi berbagai dataset meteorologi dari berbagai sumber dengan metode pembelajaran ansambel canggih untuk meningkatkan ketepatan prediksi curah hujan dan mendukung proses pengambilan keputusan terkait iklim di tingkat regional.

Kata kunci: 5 Data Meteorologi; Pembelajaran Ansambel; Pembelajaran Mesin; Prediksi Hujan; Validasi Satelit

1. LATAR BELAKANG

Indonesia sebagai negara berkembang dan memiliki cangkupan spasial luas seringkali menemui kesulitan terkait ketersediaan data observasi meteorologi. Data observasi meteorologi sangat penting saat ini untuk memahami dinamika atmosfer dan membangun sistem peringatan dini terhadap cuaca ekstrem (Holton, 1992; Wallace & Hobbs, 2006). Selain itu, data meteorologi dibutuhkan untuk memprediksi kondisi cuaca harian. Prediksi cuaca harian dalam era teknologi saat ini memiliki banyak manfaat. Selain menjadi aspek fundamental dalam ilmu

Perbandingan Algoritma Machine Learning (Logistic Regression, SVM, KNN, Decision Tree, Random Forest, dan Gradient Boosting) dalam Prediksi Hujan Harian di Provinsi Lampung atmosfer, prediksi cuaca harian bermanfaat untuk manajemen risiko iklim, perencanaan pertanian hingga pengembangan energi terbarukan. Meninjau pentingnya prediksi data harian tetapi data observasi di Indonesia yang masih terbatas, dibutuhkan peran dataset berbasis satelit. Dataset berbasis satelit dan reanalisis seperti NASA POWER dan ERA5 telah menjadi sumber data penting dalam penelitian lingkungan dan meteorologi. Dataset satelit ini memberikan cakupan data global dengan konsistensi temporal yang baik (Aprilia, Wahidin, Abdurrahman, et al., 2025; Hosmer Jr et al., 2013; Vapnik, 2013).

Ketersediaan data meteorologi yang berkembang seiring dengan berkembangnya Teknik *Machine Learning* (ML). *Machine Learning* menjadi salah satu pendekatan yang baik dalam memodelkan fenomena lingkungan khususnya atmosfer. Algoritma ML memberikan berbagai keunggulan untuk memodelkan hubungan yang tidak hanya hubungan linier. ML dinyatakan unggul dalam memodelkan hubungan kompleks, nonlinier dan multivariat yang sulit ditanggapi oleh metode konvensional. Berbagai penelitian sebelumnya seperti untuk memprediksi curah hujan, radiasi surya, anomali suhu dan variabel atmosfer lainnya telah memanfaatkan algoritma ML (Aprilia, Wahidin, & Abdurrahman, 2025; Aprilia, Wahidin, & Syafriadi, 2025; Arnah Ritonga et al., 2025). Algoritma ML yang memberikan hasil memuaskan antara lain seperti *random forest*, *gradient boosting*, *support vector machine* dan *k-nearest neighbors*. Hasil yang ditunjukkan oleh algoritma lebih baik akurasi dibandingkan metode statistik konvensional (Bell et al., 2021; Breiman, 2001; Friedman, 2001; Tan et al., 2023).

Pada penelitian akan berfokus pada wilayah Lampung. Lampung memiliki variasi spasial curah hujan yang kuat akibat pengaruh topografi, gradien pesisir-darat dan sistem monsun. Namun, jumlah stasiun pengamatan meteorologi dari BMKG masih terbatas. Penelitian yang ada saat ini Sebagian besar hanya membahas tentang performance algoritma ML atau analisis statistik saja. Penelitian yang membahas komprehensif membandingkan performa ML dengan data observasi lokal (BMKG) masih jarang ditemukan. Gap penelitian ini memberikan peluang pentingnya studi komparatif untuk menilai kelebihan dan kekurangan masing-masing algoritma ML terhadap kondisi iklim lokal di Lampung.

Penelitian ini bertujuan untuk mengevaluasi serta membandingkan performa beberapa model *Machine Learning* yaitu *Logistic Regression*, *Support Vector Machine*, *K-Nearest Neighbors*, *Decision Tree*, *Random Forest*, dan *Gradient Boosting Machine* terhadap kejadian hujan harian. Dengan pendekatan model yang sistematis menggabungkan variabel suhu, kelembapan, dan angin dari data observasi BMKG diharapkan bermanfaat dalam sistem prediksi cuaca lokal. Hasil studi ini diharapkan berkontribusi pada aplikasi ML dalam bidang

meteorologi untuk mendukung layanan iklim, perencanaan pertanian dan sistem peringatan dini di Indonesia.

2. KAJIAN TEORITIS

Secara fisika, curah hujan terbentuk akibat proses kondensasi di atmosfer akibat perubahan suhu, kelembapan, tekanan dan dinamika angin. Sehingga variabel meteorologi seperti suhu udara, kelembapan relatif dan kecepatan angin memiliki korelasi yang kuat terhadap pembentukan awan. Selain pembentukan awan, sistem konveksi juga dipengaruhi oleh variabel meteorologi tersebut. Model fisis atmosfer dijelaskan dalam persamaan Navier-Stokes, Persamaan Termodinamika dan Persamaan Kontinuitas (Wallace & Hobbs, 2006). Persamaan ini berhasil menjelaskan perubahan kecil pada kondisi awal mampu menghasilkan respon atmosfer yang sangat berbeda. Sehingga prediksi hujan menjadi fenomena yang dipengaruhi oleh dinamika fisik yang cukup kompleks dan bersifat nonlinier sehingga metode deterministik tradisional sulit menghadapi keterbatasan ini (Holton, 1992).

Pendekatan ML dinilai baik untuk mencari Solusi mengatasi kompleksitas tersebut. Sehingga saat ini pendekatan ML sangat populer dalam meteorologi modern. Algoritma ML seperti *Logistic Regression*, *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Tree*, *Random Forest*, dan *Gradient Boosting Machine* (GBM) menjadi model yang mengatasi pola nonlinier (Sutanto et al., 2024). Selain itu, algoritma tersebut mampu menerjemahkan hubungan multivariabel yang tidak terdefinisi secara eksplisit dalam model empiris (Bukhari & Mohamed, 2015).

Masing-masing algoritma memiliki Langkah-langkah prediksi yang berbeda. Hal ini menjadi menarik untuk menilik lebih lanjut bagaimana performa masing-masing algoritma. LR beroperasi berdasarkan fungsi logit, fungsi logit Adalah fungsi dalam statistika dan ML yang mengubah probabilitas (0 dan 1) menjadi skala logarimik dari $-\infty$ hingga $+\infty$. SVM menjadi model yang dapat mengoptimalkan margin pemisah antara kelas data. Selain itu, SVM terbukti efektif untuk data dengan pola nonlinier menggunakan kernel Gaussian (LaValley, 2008). Algoritma KKN memiliki keunggulan dalam kedekatan antar titik data dan cocok untuk pola local (Peterson, 2009). Selanjutnya algoritma *Decision Tree* dan algoritma ensemble methods seperti *Random Forest*, *Gradient Boosting Machine* terbukti baik dalam meningkatkan akurasi prediksi dalam berbagai masalah cuaca dan iklim (Aprilia, Zakiya, Pauzi, et al., 2025). Sifat-sifat algoritma tersebut yang mampu meminimalkan *overfitting* dan membaca interaksi kompleks antar variabel menjadi point unggulan (Shaharudin, 2020; Sutanto et al., 2024; Wangwongchai et al., 2023).

Penelitian sebelumnya berhasil membuktikan bahwa metode ML mampu meningkatkan performa prediksi curah hujan dibandingkan metode konvensional. Penelitian sebelumnya membandingkan model regresi linier dengan model ensemble (Tan et al., 2023). Hasil ditunjukkan bahwa model ensemble lebih akurat. Penelitian lain mengungkapkan bahwa algoritma pohon Keputusan seperti *Random Forest* cenderung lebih stabil (Usman et al., 2023). Penelitian ini mengambil konteks pembahasan dalam wilayah tropis dengan variabilitas tinggi khususnya ada musim monsun. Penelitian lain menunjukkan bahwa *Gradient Boosting* mampu memprediksi fase hujan dan tidak hujan dengan baik. Hasil penelitian mendapati akurasi diatas 80% untuk data harian BMKG (Hikmayanti Handayani et al., 2025). Hasil-hasil penelitian ini menguatkan bahwa kondisi atmosfer Indonesia sangat kompleks dipengaruhi oleh dinamika monsoon, topografi dan fenomena ENSO dan IOD tetapi masih dapat diprediksi baik oleh algoritma ML.

Berdasarkan teori dan penelitian sebelumnya menjadi penguat bahwa algoritma ensemble memiliki performa yang cukup baik. Namun, evaluasi empiris masih dibutuhkan untuk melihat bagaimana performa masing algoritma ML saat dibandingkan satu persatu. Oleh karena itu, penelitian ini berfokus pada wilayah lampung dengan karakteristik spasial lampung dan melihat algoritma yang paling cocok diterapkan. Algoritma ML digabungkan dengan data observasi BMKG khusus di Lampung.

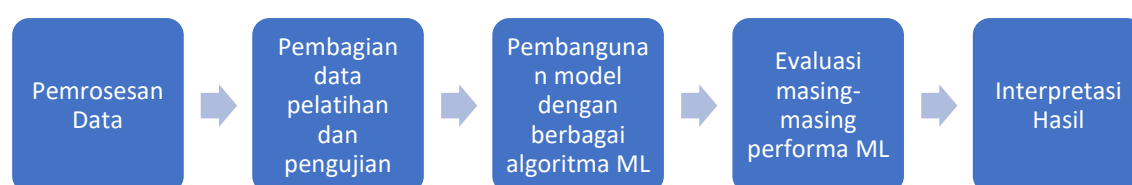
3. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan ML dan membandingkan bagaimana performa masing-masing ML. Dalam kasus ini, performa ML dinilai dari algoritma mana yang paling baik dalam memprediksi hujan di Lampung. Penelitian ini mengambil data dari 3 stasiun observasi BMKG di lampung dengan Lokasi detail ditampilkan pada Tabel 1. Data observasi harian BMKG disbanding dengan data satelit era5 dan NASA POWER selama periode tahun 2020-2024. Perbandingan data satelit seperti yang telah dilakukan oleh (A.Aprilia et al ,2025) dengan hanya menggunakan algoritma *Gradient Boosting Machine* (GMB) (Aprilia, Zakiya, Pauzi, et al., 2025) .

Tabel 1. Lokasi Koordinat Stasiun Pengamatan BMKG.

Nama Stasiun	Latitude	Longitude
Stasiun BMKG Lampung (Stamar Panjang)	-5.45	105.31
Stasiun Klimatologi Pesawaran	-5.17	105.18
BMKG Geophysics stasiun class III Kotabumi	-4.83	104.87

Data meteorologi, data satelit dan data reanalisis diolah melalui beberapa tahap yang ditampilkan pada Gambar 1. Selanjutnya, model akan dianalisis dengan metode perhitungan berdasarkan teori dan penelitian sebelumnya.

**Gambar 1.** Alur Pengolahan Data Penelitian.

Pada tahap pertama seluruh data dibersihkan dari data hilang dan data eror menggunakan metode interpolasi liner. Data juga diberlakukan normalisasi *feature scaling* menggunakan *Min-Max Scalling*. Setelah data dilihat telah bersih dari data missing dan eror maka data dibagi menjadi data latih dan data uji. Data latih dengan proporsi 80% dan data uji dengan proporsi 20% dengan *stratified split*. Pada penelitian ini akan menggunakan 6 model machine learning yaitu *Logistic Regression* (LR), *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), *Decision Tree* (DT), *Random Forest* (RF), dan *Gradient Boosting Machine* (GBM). Pemilihan 6 model berlandaskan tren penelitian yang memberikan performa baik dalam prediksi hujan. Berikut akan dijelaskan lebih detail bagaimana alur perhitungan masing-masing algoritma.

Logistic Regression (LR)

Model LR bekerja dengan mengestimasi fungsi logit. Fungsi logit sering digunakan dalam ilmu statistika dan ML dengan mengubah probabilitas menjadi skala logaritmik (Hosmer Jr et al., 2013).

K-Nearest Neighbors (KNN)

Algoritma KNN bekerja dengan cara mengklasifikasi objek baru berdasarkan mayoritas nilai terdekat dalam himpunan data latih. KKN mengklasifikasi data dalam ruang fitur

Perbandingan Algoritma Machine Learning (Logistic Regression, SVM, KNN, Decision Tree, Random Forest, dan Gradient Boosting) dalam Prediksi Hujan Harian di Provinsi Lampung menggunakan jarak Euclidean. Prediksi kelas disesuaikan dengan mayoritas nilai terdekat (Peterson, 2009).

Support Vector Machine

Support Vektor Machine (SVM) adalah algoritma ML yang bekerja dengan cara menemukan hyperplane (garis pemisah) terbaik untuk memaksimalkan margin antara kelas data berbeda. Algoritma ini baik untuk mengklasifikasi dan menemukan data-data penting sehingga menciptakan pemisah optimal (Huang et al., 2018).

Decision Tree

Decision Tree yaitu algoritma pemodelan dengan prinsip kerja menggunakan struktur seperti pohon. Maksudnya disini, pengambilan Keputusan memenuhi serangkaian aturan dan kondisi, apabila divisualisasikan mirip seperti pohon. Algoritma ini cocok digunakan untuk klasifikasi dan regresi. Selanjutnya, algoritma ini membagi data rekursi menggunakan impurity Gini. Gini Impurity yakni peluang melakukan kesalahan dalam mengklasifikasi titik data jika dilabeli terhadap distribusi kelas data (Bukhari & Mohamed, 2015).

Random Forest

Random Forest (RF) Adalah algoritma ML mirip seperti Decision Tree, Dimana RF menggabungkan banyak pohon keputusan untuk membuat Keputusan (prediksi Tunggal). RF dibangun dari ratusan pohon dengan pola acak dan diprediksi dengan voting (Breiman, 2001).

Gradient Boosting Machine (GBM)

GBM Adalah algoritma ML yang membangun model prediktif secara bertahap dengan menggabungkan banyak model lemah. Model lemah yang dimaksud adalah pohon keputusan sederhana yang akan dioptimalkan akurasi. GBM dibangun dengan model aditif secara bertahap. Model GBM dapat dirumuskan pada persamaan 9. Nilai residual dihitung melalui fungsi turunan loss (Friedman, 2001).

Evaluasi Model

Evaluasi model dilakukan dengan menggunakan nilai akurasi, presisi, recall dan F-1. Evaluasi ini mengikuti aturan evaluasi model klasifikasi yang populer digunakan. Akurasi menilai banyak prediksi yang benar dan presisi adalah rasio antara prediksi positif yang benar dengan total prediksi positif. Selanjutnya Adalah Recall, recall mengukur rasio antara prediksi positif yang benar dengan total jumlah data yang sebenarnya positif. F1-Score Adalah metrik yang menggabungkan precision dan recall (Aprilia, Wahidin, Abdurrahman, et al., 2025; Hikmayanti Handayani et al., 2025).

Visualisasi model ditambah dengan *Confusion Matrix* dan *ROC Curve-AUC*. *Confusion Matrix* sebagai tabel evaluasi yang memberikan jumlah prediksi benar dan salah

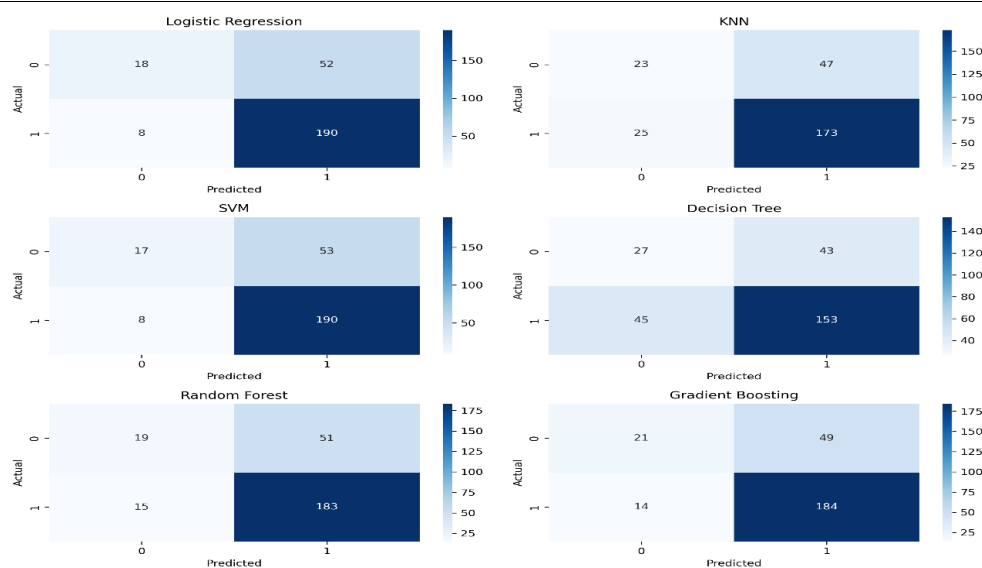
pada model klasifikasi. Tabel ini memuat 4 elemen yaitu *True Positif* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). *ROC Curve* adalah grafik yang menunjukkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai ambang keputusan (*threshold*). Sedangkan, AUC adalah nilai numerik yang menunjukkan luas area di bawah kurva ROC (Hoo et al., 2017).

4. HASIL DAN PEMBAHASAN

Hasil performa dari enam model ditampilkan pada Tabel 2. Berdasarkan Tabel 2 nampak variasi signifikan dalam kemampuan memprediksi hujan harian. Berdasarkan hasil performa algoritma RF tampak unggul dalam akurasi dengan nilai F-1 score tertinggi. Hal ini menunjukkan bahwa RF mampu mendeteksi hampir seluruh kejadian hujan secara benar. Algoritma SVM memiliki nilai recall tinggi tetapi nilai presisi lebih rendah. Hal ini mengindikasikan bahwa SVM menghasilkan prediksi berlebih pada kondisi hujan. Pada hasil performa tampak algoritma yang memiliki performa paling rendah yaitu Decision Tree. Decision Tree tampak overfitting dan kurang mampu untuk generalisasi data pengujian.

Tabel 2. Hasil Performa 6 Model ML.

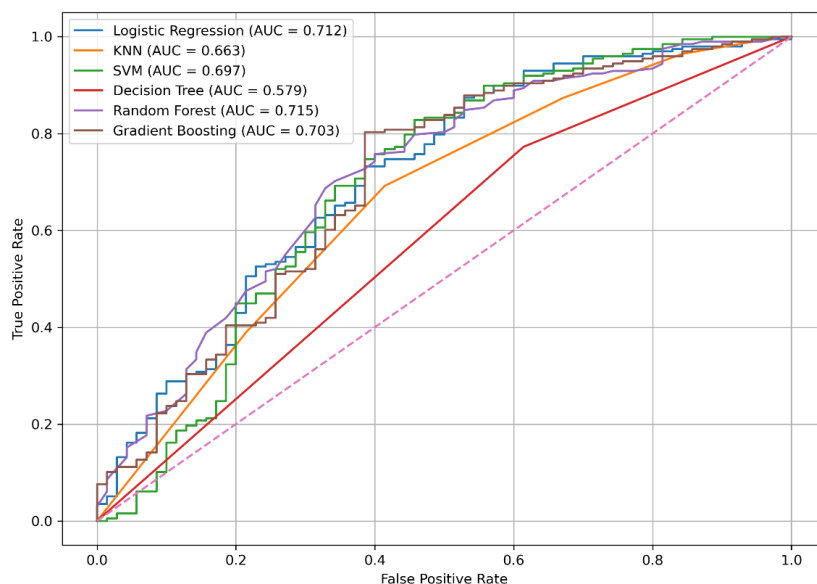
Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0,78972	0,80102	0,96319	0,874652
KNN	0,785047	0,816216	0,92638	0,867816
SVM	0,78972	0,789216	0,98773	0,877384
Decision Tree	0,663551	0,789809	0,760736	0,775
Random Forest	0,813084	0,809045	0,98773	0,889503
Gradient Boosting	0,775701	0,801047	0,93865	0,864407



Gambar 2. Confusion Matrix Setiap Model.

Evaluasi performa masing-masing model ditambahkan visualisasi berupa *Confusion Matrix*. Secara keseluruhan seluruh model menunjukkan kecenderungan sama yakni hasil prediksi cenderung lebih baik untuk kelas positif dibandingkan untuk kelas negatif. Hal ini tampak dari nilai *True Positive* (TP) yang tinggi untuk semua model. Pada hasil keseluruhan Kembali model *Random Forest* dan *SVM* lebih stabil. Sebaliknya *Decision Tree* tampak memiliki performa terendah. Hasil *Confusion Matrix* lebih detail tampak pada Gambar 2.

Visualisasi performa ditambahkan dengan kurva ROC -AUC yang disajikan pada Gambar 3. Metrik evaluasi ini dapat menilai performa model untuk membedakan kelas positif dan negative untuk berbagai ambang batas. Hasil ROC-AUC menunjukkan Kembali performa *Random Forest* memiliki kemampuan diskriminasi terbaik untuk membedakan kelas. Selanjutnya performa model logistic regression tampak sebaik RF. Untuk model *SVM*, *GBM* dan *KNN* menunjukkan performa menengah. *Decision Tree* memiliki performa terendah yang cenderung tidak stabil sehingga prediksi data kurang akurat.



Gambar 3. Kurva ROC dan AUC Tiap Model.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi performa 6 model secara umum model RF memberikan performa terbaik dalam kemampuan memprediksi hujan. Hal ini tampak dari akurasi, presisi, recall, F1-Score, confusion matrix dan AUC tertinggi. Model Logistic Regression dan SVM juga menunjukkan evaluasi baik untuk aspek recall dan AUC. Untuk model decision tree tampak memiliki performa yang kurang stabil. Dengan melihat seluruh metrics evaluasi dapat dinilai untuk kasus curah hujan Lampung pada periode 2020-2024 model RF Adalah model paling direkomendasikan untuk memprediksi hujan.

DAFTAR REFERENSI

- Aprilia, A., Wahidin, A. B., & Abdurrahman, A. F. (2025). Integration of machine learning and NASA POWER dataset for predicting coffee production in Lampung. *Jurnal Fisika Flux: Jurnal Ilmiah Fisika FMIPA Universitas Lambung Mangkurat*, 22(1), 44. <https://doi.org/10.20527/flux.v22i1.20980>
- Aprilia, A., Wahidin, A. B., & Syafriadi. (2025). Seismic activity analysis in Indonesia: Integrating machine learning, geospatial data, and environmental factors for risk assessment. *JPF (Jurnal Pendidikan Fisika) Universitas Islam Negeri Alauddin Makassar*, 13(1), 12–26. <https://doi.org/10.24252/jpf.v13i1.54081>
- Aprilia, A., Wahidin, A. B., Abdurrahman, A. F., Prihanto, S., & Al-Y. (2025). Comparative validation of NASA POWER and ERA5 satellite-based meteorological data using BMKG observations in Bandar Lampung, Indonesia. *InfoSains*, 15(2), 360–369. <https://doi.org/10.54209/infosains.v15i02>
- Aprilia, A., Zakiya, H., Pauzi, G. A., & Supriyanto, A. (2025). Integration of Magnus thermodynamic parameters and machine learning algorithms in rainfall prediction. *Jurnal Ilmiah*, 11(2), 82–91.
- Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Soci, C., Villaume, S., Bidlot, J. R., Haimberger, L., Woollen, J., Buontempo, C., & Thépaut, J. N. (2021). The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, 147(741), 4186–4227. <https://doi.org/10.1002/qj.4174>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bukhari, S., & Mohamed, S. (2015). Applications of applied scientific research. *Journal of Advanced Applied Scientific Research*, 5(1), 68–83.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Handayani, H. H., Lestari, S. A. P., Cahyana, Y., & Karawang, P. (2025). Evaluasi kinerja algoritma random forest dan gradient boosting untuk klasifikasi penyakit jantung. *KOMTIKA*, 9(1). <https://doi.org/10.31603/komtika.v9i1.13450>
- Holton, J. R. (1992). *An introduction to dynamic meteorology* (Vol. 48). Academic Press. <https://doi.org/10.1016/B978-0-12-354355-4.50005-X>
- Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve? *Emergency Medicine Journal*, 34(6), 357–359. <https://doi.org/10.1136/emered-2017-206735>
- Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons. <https://doi.org/10.1002/9781118548387>
- Huang, S., Cai, N., Pacheco, P. P., Narandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics and Proteomics*, 15(1), 41–51. <https://doi.org/10.21873/cgp.20063>
- LaValley, M. P. (2008). Logistic regression. *Circulation*, 117(18), 2395–2399. <https://doi.org/10.1161/CIRCULATIONAHA.106.682658>
- Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, 4(2), 1883. <https://doi.org/10.4249/scholarpedia.1883>

- Perbandingan Algoritma Machine Learning (Logistic Regression, SVM, KNN, Decision Tree, Random Forest, dan Gradient Boosting) dalam Prediksi Hujan Harian di Provinsi Lampung*
- Ritonga, A., Al Amini, A., Mutianda, L., Singarimbun, R., Baeha, A. H., Pasaribu, G. R., & Damanik, J. A. D. (2025). Analisis probabilitas hujan menggunakan data historis dari BMKG Wilayah I tahun 2013–2015. *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam*, 4(1), 1–20. <https://doi.org/10.55606/jurrimipa.v4i1.4367>
- Shaharudin, S. M. (2020). Imputation methods for addressing missing data of monthly rainfall in Yogyakarta, Indonesia. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(1.4), 646–651. <https://doi.org/10.30534/ijatcse/2020/9091.42020>
- Sutanto, T., Aditya, M. R., Budiman, H., Noor Ridha, M. R., Syapotro, U., & Azijah, N. (2024). Comparison of logistic regression, random forest, SVM, and KNN algorithm for water quality classification based on contaminant parameters. *INTI Journal*, 2022(1). <https://doi.org/10.61453/jods.v2023no48>
- Tan, M. L., Armanuos, A. M., Ahmadianfar, I., Demir, V., Heddam, S., Al-Areeq, A. M., Abba, S. I., Halder, B., Kilinc, H. C., & Yaseen, Z. M. (2023). Evaluation of NASA POWER and ERA5-Land for estimating tropical precipitation and temperature extremes. *Journal of Hydrology*, 624, 129940. <https://doi.org/10.1016/j.jhydrol.2023.129940>
- Usman, C. D., Widodo, A. P., Adi, K., & Gernowo, R. (2023). Rainfall prediction model in Semarang City using machine learning. *Indonesian Journal of Electrical Engineering and Computer Science*, 30(2), 1224–1231. <https://doi.org/10.11591/ijeecs.v30.i2.pp1224-1231>
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer.
- Wallace, J. M., & Hobbs, P. V. (2006). *Atmospheric science: An introductory survey* (2nd ed., Vol. 92). Elsevier.
- Wangwongchai, A., Waqas, M., Dechpichai, P., Hlaing, P. T., Ahmad, S., & Humphries, U. W. (2023). Imputation of missing daily rainfall data: A comparison between artificial intelligence and statistical techniques. *MethodsX*, 11, 102459. <https://doi.org/10.1016/j.mex.2023.102459>