

Perbandingan Model *Decision Tree* untuk Prediksi Intensitas Hujan di Juanda, Sidoarjo, Jawa Timur

Rafli Iqbal Taufiqi^{1*}, Veimas Mahardika Pramuji¹⁾, M. Arifudin¹⁾, Agung Hari Saputra¹⁾

¹⁾Program Studi Meteorologi, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Kota Tangerang, Indonesia

*Korespondensi: rafliqbal060204@gmail.com

Abstrak

Intensitas hujan merupakan salah satu parameter penting dalam dinamika iklim dan sangat berkaitan dengan peningkatan kejadian bencana hidrometeorologi. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa tiga model pembelajaran mesin berbasis *decision tree* dalam memprediksi intensitas hujan di wilayah Juanda, Sidoarjo, Jawa Timur. Data yang digunakan berupa data observasi cuaca harian dari Stasiun Meteorologi Kelas I Juanda selama periode 2018–2022, yang mencakup 13 variabel meteorologis. Model yang dibandingkan meliputi *Extra Trees Classifier*, *Random Forest Classifier*, dan *XGBoost Classifier*, dengan evaluasi kinerja berdasarkan metrik akurasi, presisi, *recall*, F1 Score, dan Area Under the Curve (AUC). Hasil penelitian menunjukkan bahwa *Extra Trees Classifier* memiliki performa terbaik dengan nilai akurasi 0,8123; presisi 0,8151; *recall* 0,8123; AUC 0,9158; dan F1 Score 0,8126. Kelembaban relatif (Rh) teridentifikasi sebagai variabel yang paling berpengaruh dalam proses prediksi intensitas hujan pada ketiga model. Temuan ini memberikan pemahaman lebih lanjut mengenai hubungan antara parameter cuaca harian dan intensitas hujan, serta berkontribusi dalam pengembangan sistem prediksi yang lebih akurat sebagai dasar mitigasi bencana hidrometeorologi di wilayah studi.

Kata Kunci: Decision Tree, Intensitas Hujan, Machine Learning, Prediksi, Juanda

Abstract

Rainfall intensity is one of the key parameters in climate dynamics and is strongly associated with the increasing occurrence of hydrometeorological disasters. This study aims to evaluate and compare the performance of three decision tree-based machine learning models in predicting rainfall intensity in the Juanda region, Sidoarjo, East Java. The data used consist of daily weather observations from the Juanda Class I Meteorological Station during the period 2018–2022, covering 13 meteorological variables. The models compared include the Extra Trees Classifier, Random Forest Classifier, and XGBoost Classifier, with performance evaluation based on accuracy, precision, recall, F1 score, and Area Under the Curve (AUC). The results indicate that the Extra Trees Classifier demonstrates the best performance, with an accuracy of 0.8123; precision of 0.8151; recall of 0.8123; AUC of 0.9158; and F1 score of 0.8126. Relative humidity (Rh) was identified as the most influential variable in predicting rainfall intensity across all three models. These findings provide further insights into the relationship between daily weather parameters and rainfall intensity and contribute to the development of more accurate predictive systems as a basis for hydrometeorological disaster mitigation in the study area.

Keywords: Decision Tree, Rainfall Intensity, Machine Learning, Prediction, Juanda

1. PENDAHULUAN

Prediksi merupakan suatu proses untuk memproyeksikan atau memperkirakan suatu peristiwa yang akan datang (Laksmana dkk., 2019). Untuk memprediksi suatu peristiwa diperlukan suatu pendekatan yang memungkinkan kita untuk dapat membuat sebuah perkiraan. Pendekatan yang sering digunakan untuk memprediksi di era modern ini adalah dengan menggunakan *Artificial Intelligence* (AI). *Artificial Intelligence* (AI) berfokus pada pengembangan sistem cerdas yang mampu menjalankan tugas layaknya manusia. Salah satu

pendekatan dalam AI yang berkembang pesat adalah *Machine Learning*, yang memungkinkan sistem untuk belajar secara otomatis dari data tanpa perlu diprogram ulang secara eksplisit. Cara kerja *Machine Learning* dalam memprediksi yaitu dengan menganalisis sifat atau ciri dari objek yang tidak dikenal kemudian mengidentifikasi pola (*pattern*) dalam suatu kelompok data (dataset) (Puteri dkk., 2020). Proses identifikasi ini bertujuan untuk membuat gambaran paling akurat tentang kejadian yang mungkin terjadi di masa depan. Saat ini, upaya prediksi dari metode *Machine Learning* dilakukan dengan pengembangan berbagai pemodelan.

Menurut James (2023), model merupakan representasi atau penyederhanaan dari suatu objek, sistem, atau ide yang bertujuan untuk memudahkan pemahaman dan analisis terhadap sistem nyata dengan menampilkan informasi penting secara optimal. Menurut Bishop (2016), pemodelan dalam *machine learning* bertujuan untuk menciptakan representasi yang efisien dan akurat dari realitas dengan menggunakan kumpulan data (dataset) sebagai dasar. Sebagai pemodelan prediktif yang efisien, *decision tree* melakukan pendekatan sistematis dan interpretable untuk menganalisis serta memprediksi suatu peristiwa.

Decision Tree adalah sebuah model yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan - himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. model *Decision Tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan (Muzakir dan Wulandari, 2016). *Decision Tree* terbuat dari tiga simpul yaitu *leaf*, lalu terdiri juga dari simpul *root* yang merupakan titik awal dari suatu *decision tree*, dan yang terakhir adalah simpul perantara yang berhubungan dengan suatu pengujian.

Decision Tree memiliki beberapa model yang termasuk seperti *Extra Trees Classifier*, *Random Forest Classifier*, dan *XGBoost Classifier*, model tersebut sangat populer dalam melakukan klasifikasi (Solahuddin dkk., 2023). *Extra Trees Classifier* memodifikasi pendekatan *Random Forest* dengan memilih split fitur secara lebih acak dengan meningkatkan kecepatan pelatihan dan potensial dalam menangani banyak data. Sementara itu, *Random Forest* adalah sebuah model *ensemble learning*, mengurangi risiko *overfitting* melalui pembuatan beragam pohon keputusan dan menggabungkan prediksi mereka. Berbeda dengan keduanya, *XGBoost Classifier* menerapkan teknik '*boosting*', di mana pohon dibangun secara bertahap untuk mengoreksi kesalahan dari iterasi sebelumnya, menjadikannya efisien dan efektif dalam berbagai situasi klasifikasi. Masing-masing model ini memiliki keunggulan tersendiri tergantung pada karakteristik data dan spesifikasi masalah yang dihadapi. Beberapa penelitian serupa juga telah dilakukan oleh Mursianto dkk (2021) serta Rachmansyah dan

Astriratma (2023) mengenai model-model tersebut dalam memprediksi pola cuaca dan intensitas hujan.

Intensitas hujan adalah salah satu parameter cuaca yang sangat penting. Menurut Sudjarwadi (1987), intensitas hujan bergantung pada durasi dan volume hujan. Jika hujan berlangsung lebih lama, intensitasnya cenderung meningkat dan sebaliknya, intensitas akan lebih rendah jika hujan berlangsung singkat (Hendri, 2015). Pada penelitian ini studi kasus yang digunakan adalah wilayah Juanda, Sidoarjo, Jawa timur.

Intensitas hujan di sekitar Juanda Sidoarjo Jawa Timur memiliki peran yang signifikan dalam dinamika iklim regional. Indikator perubahan iklim tampak pada peningkatan terjadinya bencana dari waktu ke waktu terutama kejadian bencana yang terkait dengan hidrometeorologi. Salah satu daerah yang dengan kejadian bencana akibat curah hujan ekstrim adalah kabupaten Sidoarjo. Penelitian yang dilakukan oleh (Mondiana dkk., 2021) menunjukkan bahwa Indikator perubahan iklim tampak pada peningkatan terjadinya bencana dari waktu ke waktu, terutama kejadian bencana yang terkait dengan hidrometeorologi di kabupaten Sidoarjo.

Hingga saat ini belum banyak penelitian yang secara spesifik membandingkan performa berbagai model machine learning seperti Extra Trees Classifier, Random Forest Classifier, dan XGBoost Classifier secara bersamaan dalam memprediksi intensitas hujan di Juanda. Perbedaan karakteristik dan mekanisme kerja dari masing-masing model tersebut dapat menghasilkan performa yang bervariasi tergantung pada pola dan kompleksitas data lokal yang digunakan. Dengan memahami model mana yang paling optimal, dapat dihasilkan sistem prediksi yang lebih akurat dan dapat diandalkan untuk mendukung upaya mitigasi bencana. Sehingga, penelitian ini bertujuan untuk membandingkan model Extra Trees Classifier, Random Forest Classifier, dan XGBoost Classifier dalam memprediksi intensitas hujan di Juanda, Sidoarjo. Perbandingan tersebut didasarkan pada nilai Accuracy, AUC, Recall, Presisi, dan skor F1 pada masing-masing model.

2. METODOLOGI

Data yang digunakan dalam penelitian ini berasal dari Stasiun Meteorologi Kelas I Juanda, mencakup periode 5 tahun (2018-2022) berupa data harian pengamatan cuaca yang meliputi 13 variabel meteorologis pada Tabel 1. (Juaeni, 2017).

Tabel 1. Variabel Dataset

No	Variabel	Deskripsi	Keterangan
1	T_{max}	Suhu maksimum	Prediktor
2	T_{min}	Suhu minimum	Prediktor
3	T_{avg}	Suhu rata-rata	Prediktor
4	T_d	Titik embun	Prediktor
5	Rh	Kelembaban relatif	Prediktor
6	$WindDir$	Arah angin	Prediktor
7	$WindSpd$	Kecepatan angin	Prediktor
8	$Pres$	Tekanan	Prediktor
9	$Prec$	Intensitas hujan	Target
10	$TotCLOct$	Jumlah awan	Prediktor
11	$LowCLOct$	Awan rendah	Prediktor
12	$SunD1$	Lama Penyinaran	Prediktor
13	Vis	Visibilitas	Prediktor

Pengolahan dan analisis data dalam penelitian ini dilakukan dengan langkah-langkah berikut:

2. 1. Pengolahan Data Awal: Dataset diolah dengan menghapus data yang tidak lengkap menggunakan dengan menghilangkan data noise, data incomplete dan data inconsistent (Yani dkk., 2022). Hal ini bertujuan untuk mengurangi bias dalam estimasi parameter (Rubin, 1975). Selain itu, variabel target intensitas hujan diklasifikasikan menjadi tiga kategori:

Tabel 2. Klasifikasi Target

Kode biner	Keterangan
0	Tidak hujan
1	Hujan
2	Hujan lebat

2. 2. Penggunaan Pycaret: Dataset yang telah diproses kemudian dimasukkan ke *script* Python menggunakan *library* Pycaret dalam platform Google Colaboratory dengan penyimpanan Google Drive.

2. 3. Pembagian Dataset: Dataset dibagi menjadi set pelatihan (80%) dan pengujian (20%). Pembagian ini dilakukan untuk memastikan model memiliki cukup data untuk pelatihan sementara masih mempertahankan set data yang signifikan untuk pengujian (Gholam ddk., 2018). Salah satu kendala dalam dataset adalah perbedaan *range* nilai antar parameter, sehingga dataset diperlukan normalisasi/*balancing z-score* karena tidak diketahui nilai minimum dan maksimum dari data (Nasution dkk., 2019). Selain itu, dataset menunjukkan adanya ketidakseimbangan kelas. Misalnya, kejadian cuaca ekstrem seperti

badai lebih jarang terjadi dibandingkan dengan hari cuaca normal. Untuk mengatasi masalah ini, diunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*) bertujuan memastikan bahwa tidak ada satu variabel pun yang mendominasi proses klasifikasi (Hastie, 2009).

2. 4.Pembangunan dan Pengujian Model: Tahapan ini menggunakan berbagai hyperparameter untuk membangun dan menguji model prediktif. Hyperparameter adalah parameter yang ditetapkan sebelum proses pelatihan model dimulai dan berfungsi mengatur perilaku serta struktur algoritma pembelajaran mesin, berbeda dengan parameter yang dipelajari dari data selama pelatihan. Pemilihan hyperparameter yang tepat sangat penting karena dapat mempengaruhi kinerja dan akurasi model secara signifikan (Fauzi dkk., 2025). Hyperparameter yang digunakan ditunjukkan dalam Tabel 3.

2. 5.Hyperparameter

Tabel 3. Hyperparameter

No	Description	Value
1	<i>Original data shape</i>	(1826, 13)
2	<i>Transformed data shape</i>	(2859, 13)
3	<i>Transformed train set shape</i>	(2493, 13)
4	<i>Transformed test set shape</i>	(366, 13)
5	<i>Numeric features</i>	12
6	<i>Preprocess</i>	<i>True</i>
7	<i>Numeric imputation</i>	<i>mean</i>
8	<i>Categorical imputation</i>	<i>mode</i>
9	<i>Fix imbalance method</i>	SMOTE
10	<i>Normalize method</i>	<i>zscore</i>
11	<i>Fold Generator</i>	StratifiedKFold
12	<i>Fold Number</i>	10

2. 6.Evaluasi Model: Dalam proses evaluasi, model dinilai berdasarkan sejumlah metrik seperti *Accuracy*, *AUC*, *Recall*, *Precision*, dan *F1 Score*. Perhitungan metrik ini merujuk pada Persamaan (1)– (4), yang didasarkan pada struktur *confusion matrix* sebagaimana disajikan pada Tabel 4 (Han dkk., 2012).

Tabel 4. Confusion Matrix

Confusion Matrix		Observed	
		TRUE	FALSE
Predicted Class	TRUE	True Positive (TP)	False Positive (FP)
	FALSE	False Negative (FN)	True Negative (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1.1)$$

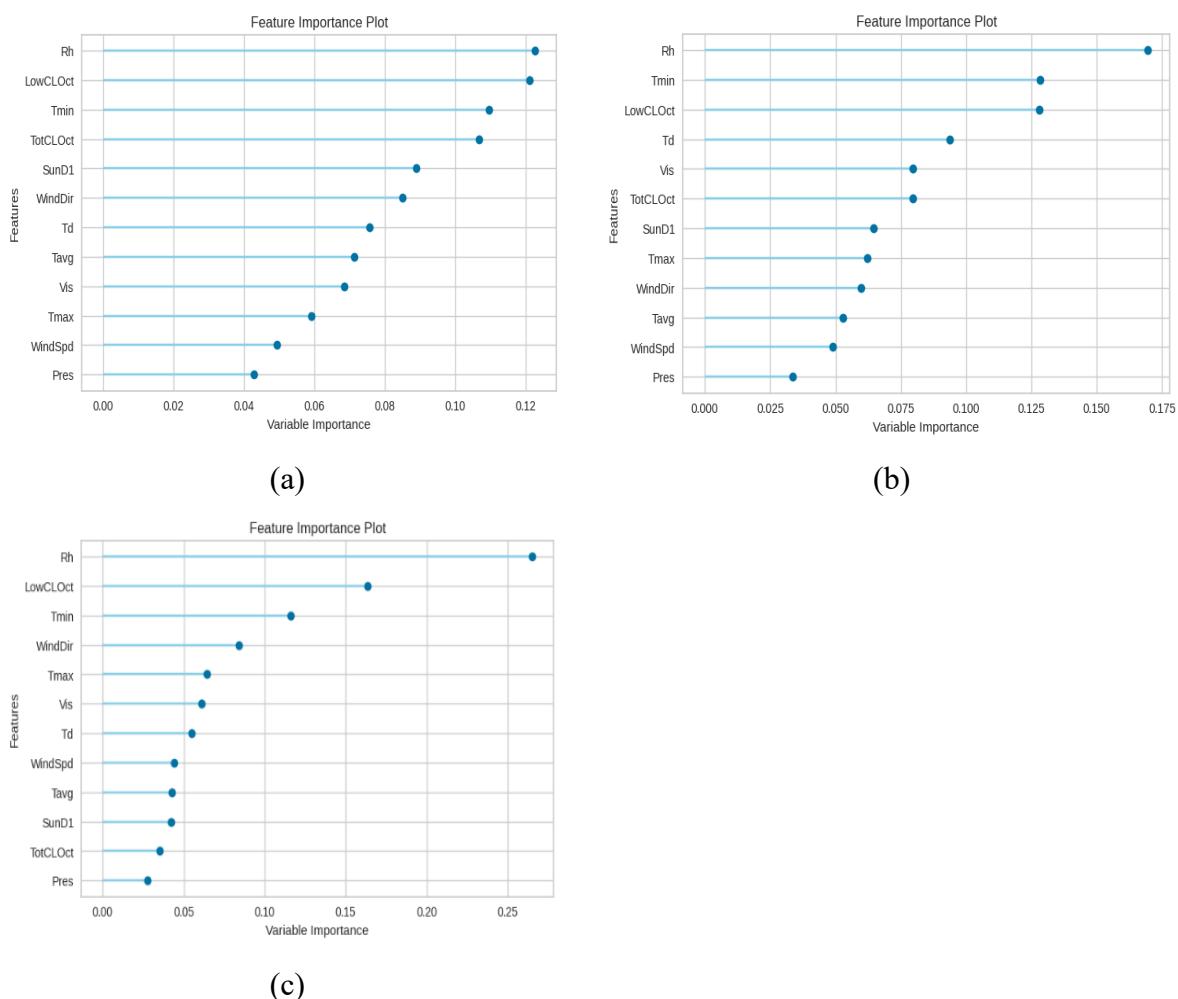
$$Recall \text{ (Sensitivitas)} = \frac{TP}{TP + FN} \quad (1.2)$$

$$Precision \text{ (Presisi)} = \frac{TP}{TP + FP} \quad (1.3)$$

$$F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (1.4)$$

Dengan mengikuti langkah-langkah ini, penelitian dapat menghasilkan analisis yang lebih akurat dan terstruktur, memastikan keandalan dan keefektifan model yang dikembangkan.

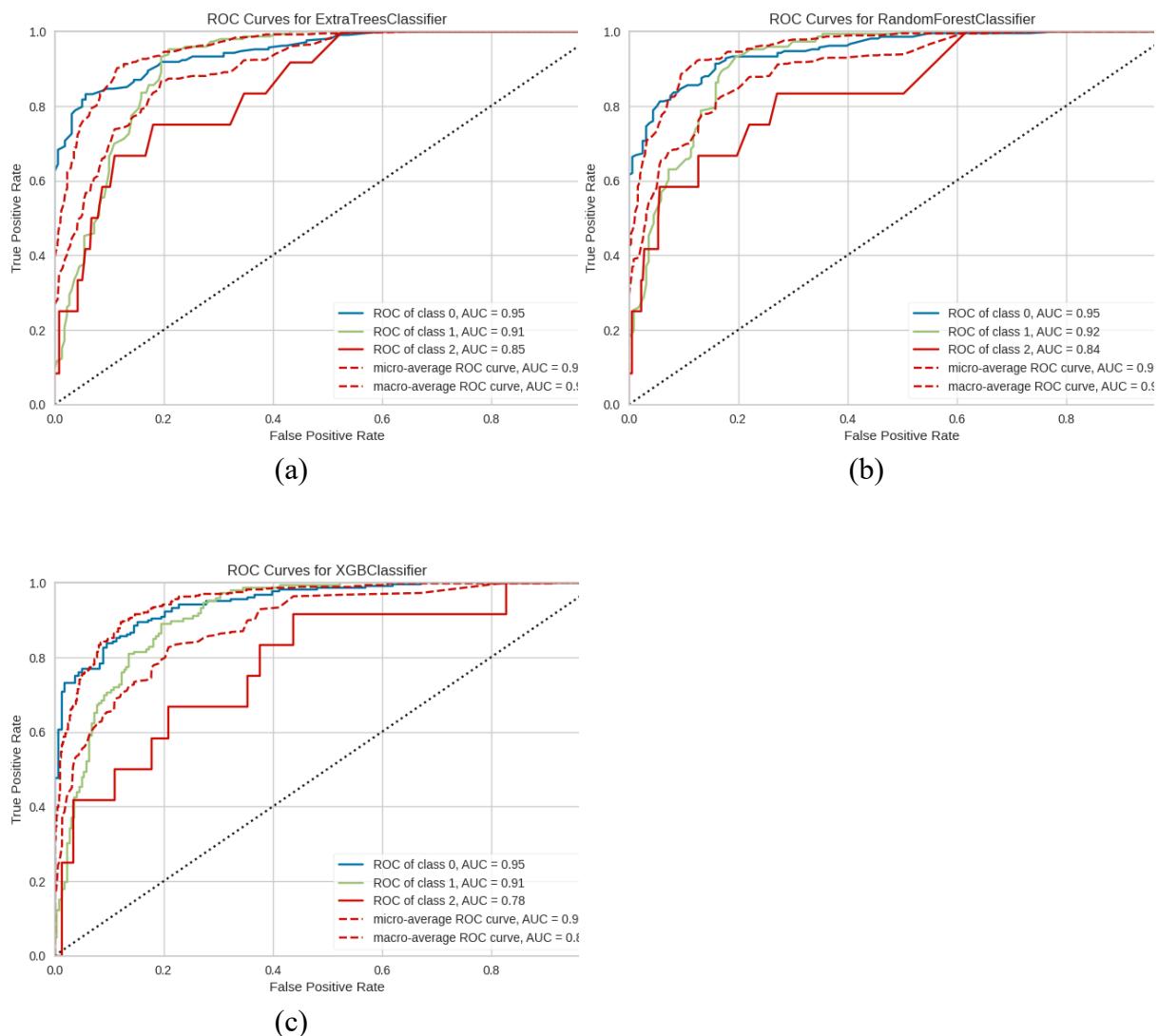
3. HASIL DAN PEMBAHASAN



Gambar 1. Feature Importance Plot (a) Extra Trees Classifier (b) Random Forest Classifier (c) XGBoost Classifier

Gambar 1 merupakan *feature importance plot* atau plot yang digunakan untuk menampilkan pentingnya setiap fitur dalam suatu model. Melalui plot ini, kontribusi relatif dari masing-masing variabel terhadap prediksi model dapat diidentifikasi berdasarkan nilai

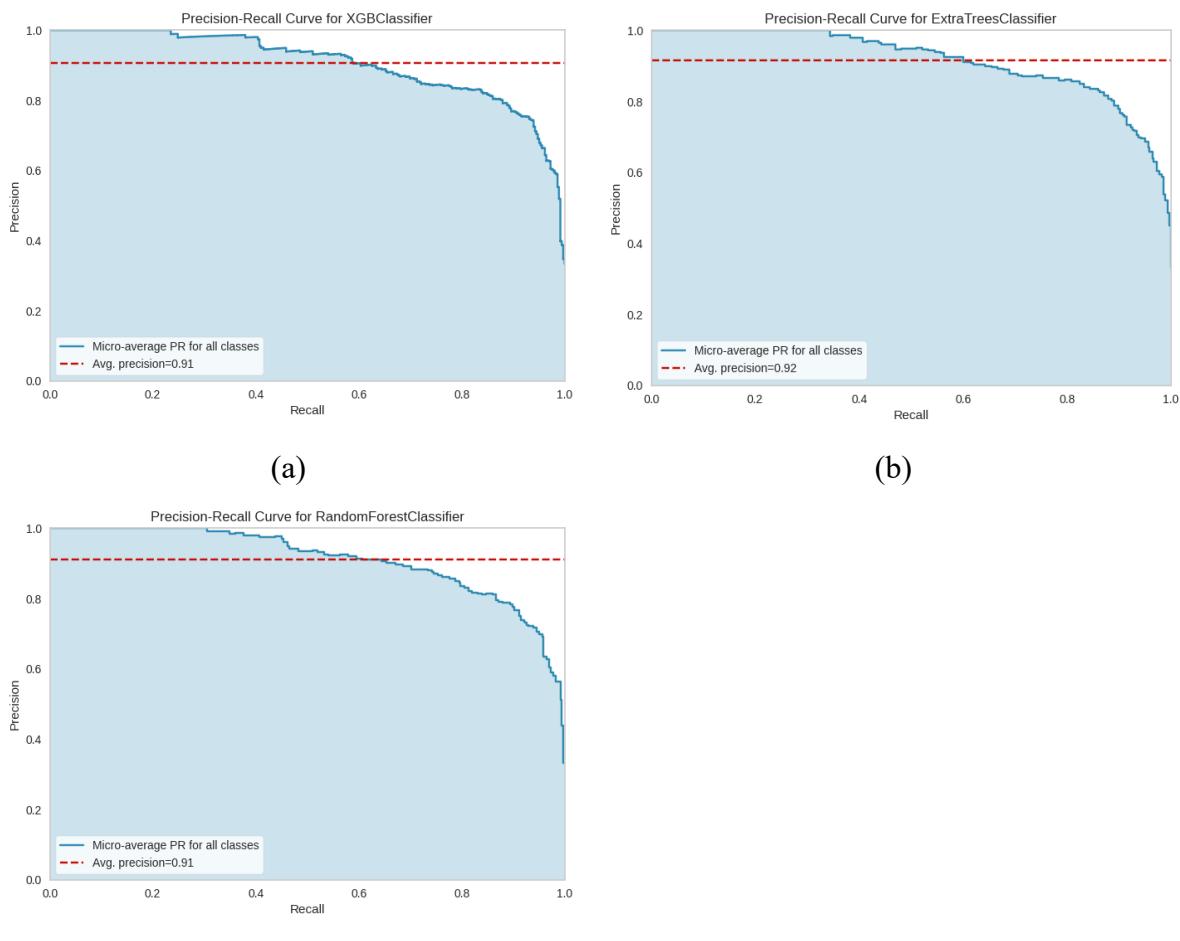
importance score yang dihitung oleh algoritma masing-masing model. Nilai ini merepresentasikan seberapa besar peran suatu fitur dalam pengambilan keputusan model prediktif. Kelembaban relatif (Rh) adalah variabel yang paling berkontribusi pada ketiga model. Sementara itu, variabel tekanan (Pres) merupakan variabel dengan persentase kontribusi terendah di semua model. Secara keseluruhan, terdapat perbedaan dalam peringkat kepentingan variabel antara model, namun variabel kelembaban relatif (Rh), awan rendah (TotCLOct), dan suhu minimum (Tmin) konsisten muncul sebagai variabel yang paling penting untuk ketiga model dalam memprediksi intensitas hujan.



Gambar 2. Kurva Receiver Operating Characteristic (ROC) (a) Extra Trees Classifier (b) Random Forest Classifier (c) XGBoost Classifier

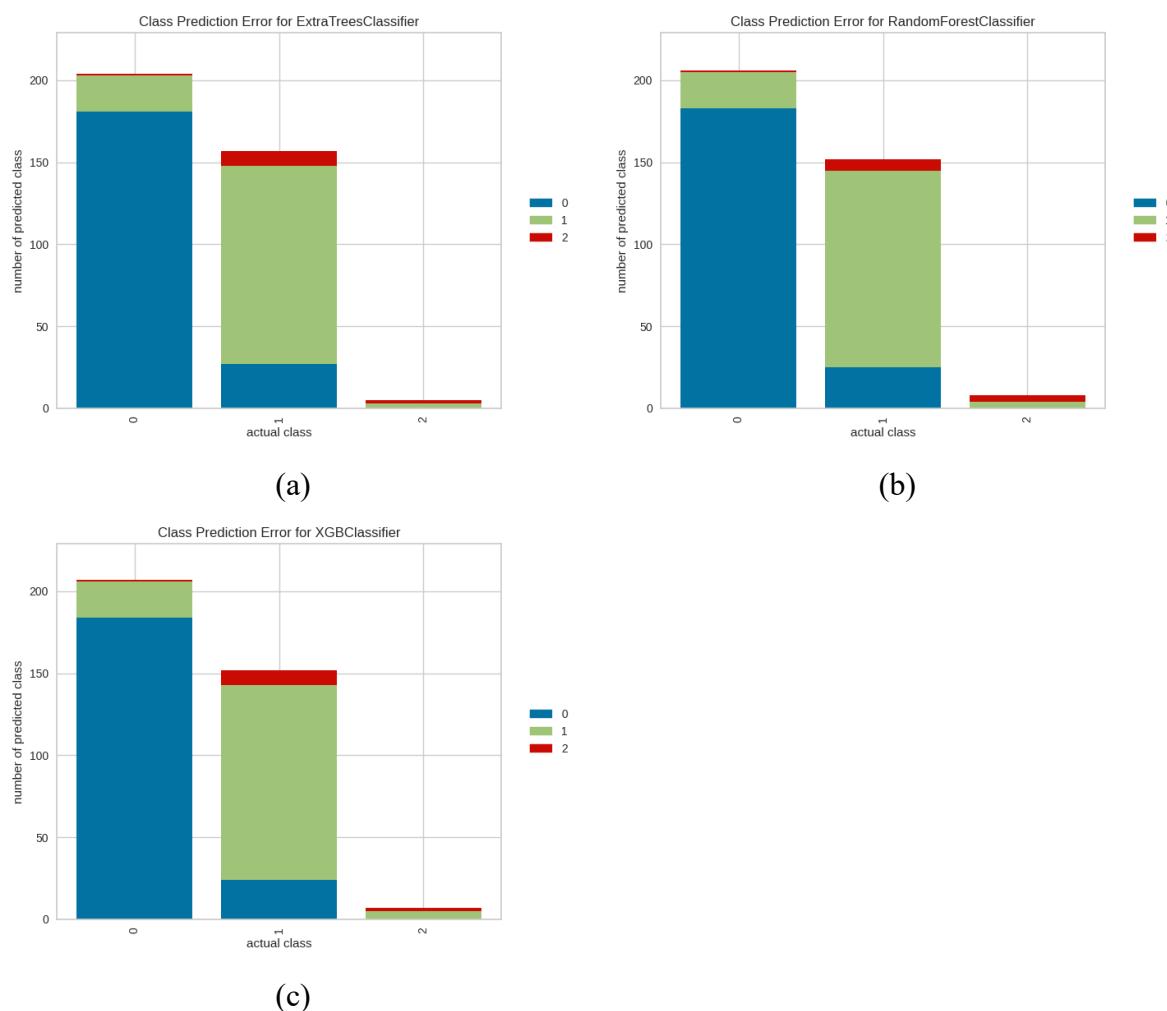
Gambar 2 menggambarkan kurva *Receiver Operating Characteristic* (ROC) pada tiga model klasifikasi yang berbeda: *Extra Trees Classifier*, *Random Forest Classifier*, dan *XGB*

Classifier. Kurva ROC adalah grafik yang merepresentasikan kinerja suatu model klasifikasi pada semua ambang batas klasifikasi. *Macro-average ROC* menghitung setiap kelas secara terpisah dan merata-ratakannya yang mana memperlakukan semua kelas dengan sama. *Micro-average ROC* dihitung dengan memperhatikan kontribusi kelas (Malakouti dkk., 2023). Ketiga model memiliki nilai AUC yang berbeda untuk setiap kelas target. Sesuai dengan klasifikasi target pada Tabel 2, kelas 0 mengacu pada kondisi *tidak hujan*, kelas 1 menunjukkan *hujan*, dan kelas 2 merepresentasikan *lebat*. *Random Forest Classifier* memiliki AUC tertinggi untuk kelas 1 (0,92), sedangkan *Extra Trees Classifier* menunjukkan kinerja terbaik untuk kelas 2 (0,85). AUC ketiga model berada dalam kategori sangat baik (Peryanto dkk., 2020). *Random Forest Classifier* juga unggul dibandingkan dua model lainnya berdasarkan nilai tertinggi pada *macro-average AUC* (0,96) dan *micro-average AUC* (0,91).



Gambar 3. Kurva Precision-Recall (a) Extra Trees Classifier (b) Random Forest Classifier (c) XGBoost Classifier

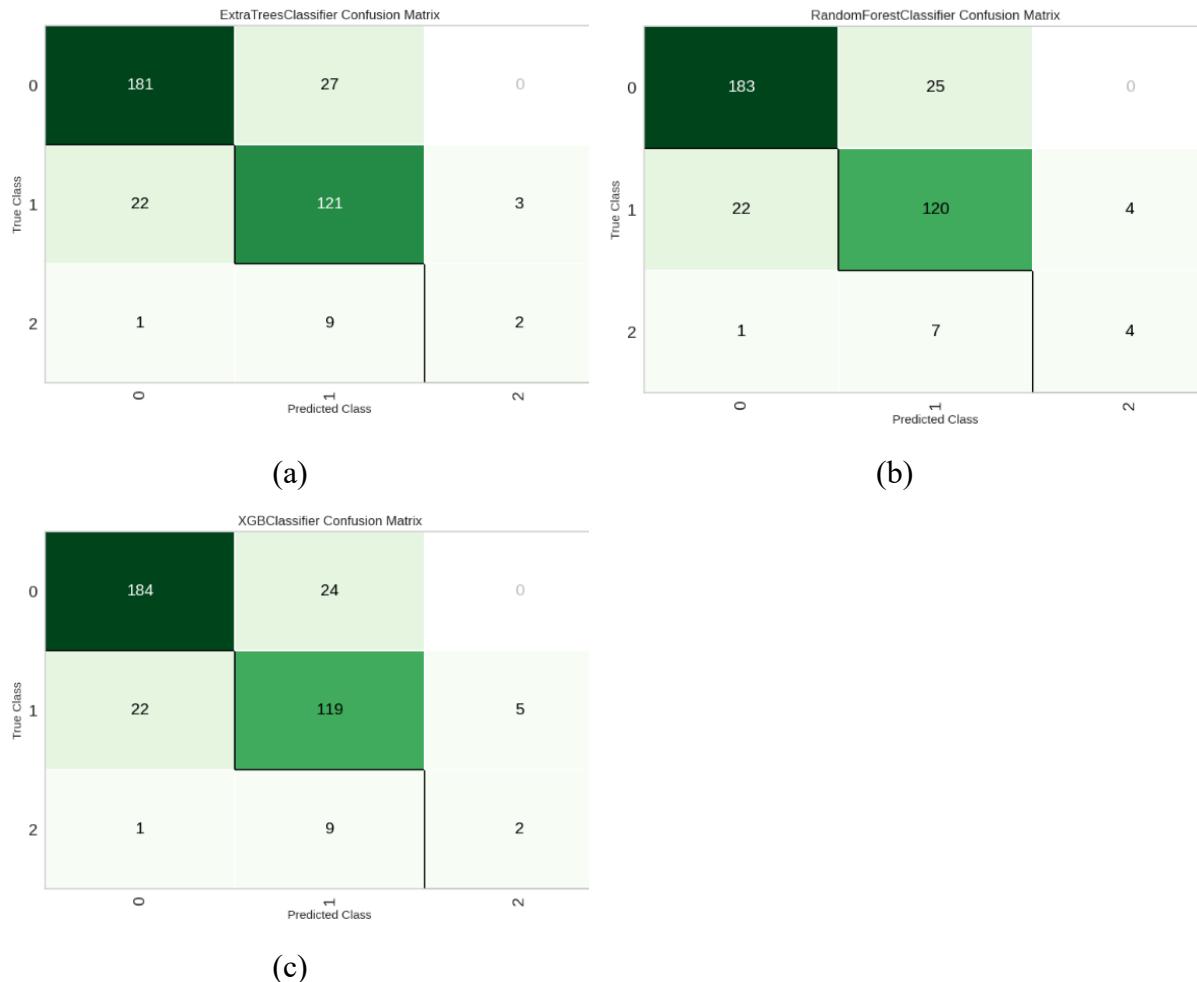
Gambar 3 menggambarkan kurva *Precision-Recall* untuk model klasifikasi *Random Forest Classifier*, *XGB Classifier*, dan *Extra Trees Classifier*. Ketiga model menunjukkan rata-rata yang serupa untuk semua kelas, dengan presisi rata-rata pada nilai 0,91 hingga 0,92, yang menunjukkan tingkat presisi yang konsisten di berbagai tingkat recall. Kurva *Precision-Recall* juga menunjukkan bahwa ketiga model memiliki presisi yang tinggi ketika *recall* rendah, tetapi presisi menurun seiring peningkatan *recall*. Ini merupakan indikasi bahwa ketika model berusaha untuk meningkatkan *recall*, meskipun juga menghasilkan lebih banyak *false positive*, yang menurunkan presisi.



Gambar 4. *Class Prediction Error* (a) *Extra Trees Classifier* (b) *Random Forest Classifier* (c) *XGBoost Classifier*

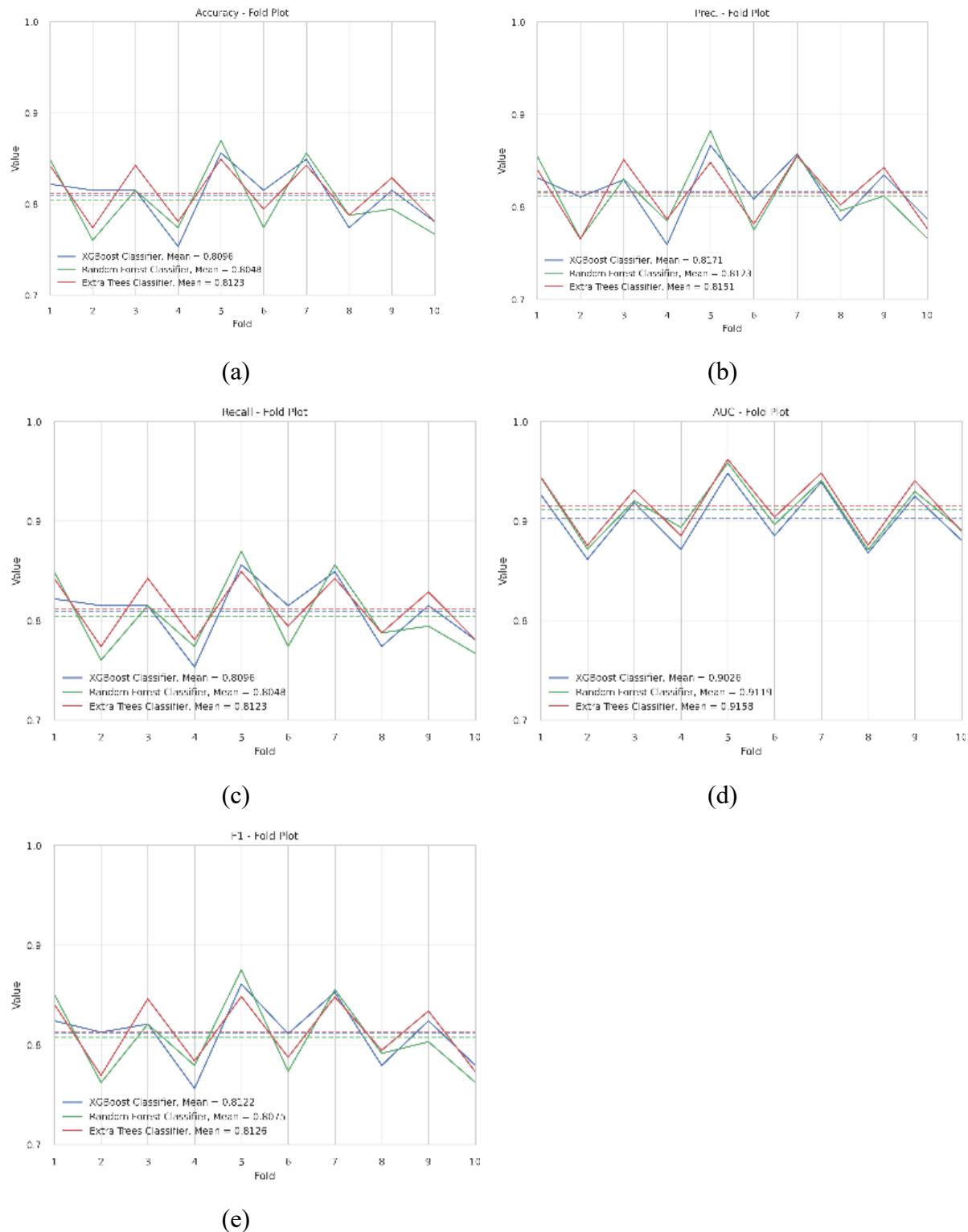
Gambar 4 menunjukkan bahwa ketiga model *Extra Trees*, *Random Forest*, dan *XGBoost* memiliki performa baik dalam memprediksi kelas 0 (tidak hujan), yang ditunjukkan oleh dominasi prediksi benar (warna biru) pada kolom kelas aktual 0. Untuk kelas 1 (hujan), meskipun masih ada kesalahan prediksi ke kelas 0 atau 2, ketiga model cukup akurat dengan

dominasi prediksi benar (warna hijau). Namun, pada kelas 2 (hujan lebat), semua model mengalami kesulitan karena prediksi benar (warna merah) sangat sedikit. Model Random Forest tampak memiliki jumlah prediksi yang benar dan kesalahan yang lebih sedikit dibanding Extra Trees dan XGBoost, terutama pada kelas 1. Hal ini menunjukkan bahwa *Random Forest Classifier* memiliki performa terbaik di antara ketiganya.



Gambar 5. *Confusion Matrix (a) Extra Trees Classifier (b) Random Forest Classifier (c) XGBoost Classifier*

Confusion Matrix dari ketiga model menunjukkan model dapat mengklasifikasikan kelas 0 dan kelas 1 dengan baik, *XGBoost Classifier* paling baik memprediksi kelas 1 dengan 184 prediksi benar. Sedang *Extra Trees Classifier* menunjukkan peforma terbaik pada kelas 2 dengan 121 prediksi benar. *Random Forest Classifier* menjadi yang terbaik dalam memprediksi kelas 2 dengan 4 prediksi benar.



Gambar 4. Plot Metrik Terhadap Fold (a) *Accuracy* (b) *Precision* (c) *Recall* (d) *AUC* (e) *F1 Score*

Tabel 5. Mean Mariks Model

Model	Accuracy	Precision	Recall	AUC	F1
<i>Extra Trees Classifier</i>	0.8123	0.8151	0.8123	0.9158	0.8126
<i>Random Forest Classifier</i>	0.8048	0.8123	0.8048	0.9119	0.8075
<i>XGB Classifier</i>	0.8096	0.8171	0.8096	0.9026	0.8122

Gambar 4 dan Tabel 5 menunjukkan matriks evaluasi performa dari tiga model klasifikasi, yaitu *Extra Trees Classifier*, *Random Forest Classifier*, dan *XGBoost Classifier*. Secara umum, perbedaan performa di antara ketiga model tersebut tidak menunjukkan signifikansi yang besar. Hal ini ditunjukkan oleh nilai metrik evaluasi yang relatif memiliki nilai yang sama. Sebagai contoh, nilai akurasi berkisar antara 0,8048 hingga 0,8123; presisi antara 0,8123 hingga 0,8171; recall antara 0,8048 hingga 0,8123; serta F1-score antara 0,8075 hingga 0,8126. Meskipun demikian, *Extra Trees Classifier* menunjukkan performa paling baik dibandingkan dua model lainnya, ditandai dengan pencapaian nilai akurasi (0,8123), presisi (0,8151), recall (0,8123), AUC (0,9158), dan F1-score (0,8126) yang lebih tinggi. Berdasarkan hasil tersebut, dapat disimpulkan bahwa *Extra Trees Classifier* merupakan model dengan performa terbaik pada eksperimen ini, meskipun keunggulannya relatif tipis.

4. KESIMPULAN

Penelitian ini berhasil mengevaluasi performa tiga model klasifikasi berbasis *machine learning*, yaitu *Extra Trees Classifier*, *Random Forest Classifier*, dan *XGBoost Classifier* dalam memprediksi intensitas hujan di wilayah Juanda, Sidoarjo, Jawa Timur. Hasil evaluasi menunjukkan bahwa ketiga model memiliki performa yang relatif setara dengan selisih nilai metrik yang tidak terlalu signifikan. Namun, *Extra Trees Classifier* secara konsisten mencatatkan nilai tertinggi pada metrik akurasi (0,8123), presisi (0,8151), recall (0,8123), AUC (0,9158), dan *F1 Score* (0,8126) yang mengindikasikan performa terbaik dalam mengklasifikasikan intensitas hujan. Selain itu, kelembaban relatif (*relative humidity*) teridentifikasi sebagai variabel cuaca yang paling berkontribusi dalam proses prediksi intensitas hujan. Hasil ini memberikan pemahaman lebih lanjut tentang prediksi intensitas hujan di wilayah Juanda, Sidoarjo, Jawa Timur yang dapat digunakan untuk meningkatkan kesiapsiagaan dan penanganan dampak hidrometeorologi, seperti banjir dan bencana terkait cuaca lainnya.

REFERENCES

- Bishop, C. M. (2016). *Pattern recognition and machine learning*. Springer. ISBN: 978-1-4939-3843-8.
- Fauzi, N. P. N., Khomsah, S., & Wicaksono, A. D. P. (2025). Penerapan feature engineering dan hyperparameter tuning untuk meningkatkan akurasi model Random Forest pada klasifikasi risiko kredit. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(2).
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. *Departmental Technical Reports (CS)*, 1209.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Hendri, A. (2015). Analisis metode intensitas hujan pada Stasiun Hujan Pasar Kampar Kabupaten Kampar. *Annual Civil Engineering Seminar*. ISBN: 978-979-792-636-6.
- James, A. O. (2020). *Pengantar sistem informasi*. Salemba Infotek. ISBN: 978-979-061-004-0.
- Juaeni, I. (2017). Penerapan metode statistik untuk perbandingan parameter atmosfer permukaan antara Padang dan Selaparang. *Jurnal Meteorologi dan Geofisika*, 18(2), 73–81.
- Laksmana, R. D., Santoso, E., & Rahayudi, B. (2019). Prediksi penjualan roti menggunakan metode exponential smoothing (Studi Kasus: Harum Bakery). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(5), 4933–4941.
- Malakouti, S. M., Menhaj, M. B., & Suratgar, A. A. (2023). Machine learning techniques for classifying dangerous asteroids. *MethodsX*, 11, 102337.
- Mondiana, Y. Q., Zairina, A., & Sari, R. K. (2021). Prediksi peluang kejadian curah hujan ekstrem dan implikasi pengelolaan sumberdaya air. *Journal of Forest Science Avicenna*, 4(2), 96–101.
- Mursianto, G. A., Falih, I. M., Irfan, M., Sakinah, T., & Prasvita, D. S. (2021). Perbandingan metode klasifikasi Random Forest dan XGBoost serta implementasi teknik SMOTE pada kasus prediksi hujan. *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, Jakarta, 15 September 2021. e-ISBN: 978-623-93343-4-5.

- Muzakir, A., & Wulandari, R. A. (2016). Model data mining sebagai prediksi penyakit hipertensi kehamilan dengan teknik decision tree. *Scientific Journal of Informatics*, 3(1), 19–26.
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan normalisasi data untuk klasifikasi wine menggunakan algoritma K-NN. *CESS (Journal of Computer Engineering, System and Science)*, 4(1), 78–82.
- Peryanto, A., Yudhana, A., & Umar, R. (2020). Klasifikasi citra menggunakan convolutional neural network dan K-Fold cross validation. *Journal of Applied Informatics and Computing*, 4(1), 45–51.
- Puteri, K., & Silvanie, A. (2020). Machine learning untuk model prediksi harga sembako dengan metode regresi linear berganda. *Jurnal Nasional Informatika (JUNIF)*, 1(2), 82–94.
- Rachmansyah, R. K., & Astriratma, R. (2023). Implementasi algoritma Extra Trees untuk klasifikasi cuaca Provinsi DKI Jakarta dengan oversampling SMOTE. *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, 4(2).
- Solahuddin, M., Purnamasari, A. I., & Dikananda, A. R. (2023). Klasifikasi kualitas berita pada majalah menggunakan metode decision tree. *Jurnal Teknologi Ilmu Komputer*, 1(2), 48–54.
- Sudjarwadi. (1987). *Teknik sumber daya air*. Yogyakarta: PAU Ilmu Teknik, Universitas Gadjah Mada.
- Yani, V. I., Aradea, & Mubarak, H. (2022). Optimasi prakiraan cuaca menggunakan metode ensemble pada Naïve Bayes dan C4.5. *Jurnal Teknik Informatika dan Sistem Informasi*, 8(3), 607–619.