

**PENERAPAN *RECURSIVE FEATURE ELIMINATION WITH CROSS-
VALIDATION* (RFECV) PADA ALGORITMA *RANDOM FOREST* UNTUK
KLASIFIKASI CURAH HUJAN**

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Sistem Informasi



OLEH :

ANIS FAIZUL LAILA

NPM: 2213030066

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI

2026

Skripsi oleh :

ANIS FAIZUL LAILA

NPM : 2213030066

Judul

**PENERAPAN *RECURSIVE FEATURE ELIMINATION WITH CROSS-
VALIDATION* PADA ALGORITMA *RANDOM FOREST* UNTUK
KLASIFIKASI CURAH HUJAN**

Telah disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Program Studi Sistem Informasi
FTIK UN PGRI Kediri

Tanggal : 12 Juni 2026

Pembimbing I



Arie Nugroho, S. Kom, M.M, M.Kom

NIDN. 0712108103

Pembimbing II



Aidina Ristyawan, M. Kom

NIDN. 0721018801

Skripsi oleh :

ANIS FAIZUL LAILA

NPM : 2213030066

Judul

**PENERAPAN *RECURSIVE FEATURE ELIMINATION WITH CROSS-VALIDATION* PADA ALGORITMA *RANDOM FOREST* UNTUK
KLASIFIKASI CURAH HUJAN**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 25 Juni 2026

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua : Arie Nugoroho, M.Kom

[.....]

2. Penguji 1 : Anita Sari Wardani, M.Kom

[.....]

3. Penguji 2 : Aidina Ristyawan, M.Kom

[.....]



Mengetahui,

Dekan Fakultas Teknik dan Ilmu Komputer

Dr. Sunstiono, M.Si

NIDN. 0007076801

PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Anis Faizul Laila
Jenis Kelamin : Perempuan
Tempat/tgl. lahir : Nganjuk, 22 September 2003
NPM : 2213030066
Fak/Prodi. : FTIK/ S1-Sistem Informasi

menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 12 Juni 2026

Yang Menyatakan



ANIS FAIZUL LAILA

NPM: 2213030066

MOTTO

“Banyak hal yang bisa menjatuhkanmu. Tapi satu-satunya hal yang benar-benar dapat menjatuhkanmu adalah sikapmu sendiri”

(Raden Ajeng Kartini)

“Semua jatuh bangunmu hal yang biasa, angan dan pertanyaan waktu yang menjawabnya, berikan tenggat waktu bersedihlah secukupnya, rayakan perasaanmu sebagai manusia”

(Hindia – Mata Air)

“Bagaimana mau sukses, kalau terlalu banyak AH TAPI”

(Sivia Azizah – The Catch Up Club)

ABSTRAK

Anis Faizul Laila: Penerapan RFECV Pada Algoritma *Random Forest* Untuk Klasifikasi Intensitas Curah Hujan.

Kata Kunci: Klasifikasi, Curah Hujan, Seleksi Fitur, *Recursive Feature Elimination with Cross-Validation*

Informasi mengenai intensitas curah hujan memiliki peran penting dalam mendukung mitigasi bencana hidrometeorologi, khususnya banjir. Namun, pada beberapa penelitian sebelumnya, klasifikasi curah hujan menggunakan algoritma *Random Forest* masih menggunakan fitur yang telah ditentukan tanpa proses seleksi fitur yang sistematis, sehingga kemungkinan masih terdapat fitur yang kurang relevan dan memengaruhi performa model. Oleh karena itu, diperlukan metode seleksi fitur yang mampu menentukan fitur optimal untuk meningkatkan kinerja klasifikasi curah hujan.

Penelitian ini bertujuan menerapkan metode *Recursive Feature Elimination with Cross-Validation* (RFECV) pada algoritma *Random Forest* untuk mengetahui fitur-fitur yang berpengaruh dalam klasifikasi intensitas curah hujan serta meningkatkan performa model. Dataset yang digunakan berasal dari Kaggle dengan klasifikasi curah hujan menjadi empat kategori, yaitu ringan, sedang, lebat, dan sangat lebat. Penelitian dilakukan menggunakan pendekatan CRISP-DM yang meliputi tahap *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Penyeimbangan data dilakukan menggunakan metode SMOTE, sedangkan evaluasi model menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-score*.

Hasil penelitian menunjukkan bahwa metode RFECV berhasil memilih tujuh fitur optimal, yaitu T_n , T_x , T_{avg} , RH_{avg} , ss , ddd_x , dan bulan. Model *Random Forest* dengan penerapan RFECV menghasilkan nilai *Accuracy* sebesar 93,53%, *Precision* sebesar 93,50%, *Recall* sebesar 93,27%, dan *F1-score* sebesar 93,38%. Hasil tersebut menunjukkan peningkatan performa dibandingkan penelitian sebelumnya yang memperoleh *Accuracy* sebesar 86,55%.

Berdasarkan hasil penelitian, penerapan RFECV dapat meningkatkan kinerja model *Random Forest* dalam klasifikasi curah hujan melalui pemilihan fitur yang lebih optimal. Penelitian ini masih memiliki keterbatasan pada penggunaan data dari satu sumber dan belum menggunakan data *real-time*. Oleh karena itu, penelitian selanjutnya dapat mengembangkan model menggunakan dataset yang lebih beragam, melakukan penyesuaian parameter model, serta membandingkan *Random Forest* dengan algoritma lain untuk memperoleh performa yang lebih optimal.

KATA PENGANTAR

Puji Syukur Kami panjatkan Kehadirat Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan-Nya penyusunan skripsi ini dapat diselesaikan. Skripsi dengan judul "*Penerapan Recursive Feature Elimination With Cross-Validation Pada Algoritma Random Forest Untuk Klasifikasi Curah Hujan*" ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi Sistem Informasi Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri. Pada kesempatan ini diucapkan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Dr. Zainal Afandi, M.Pd., selaku Rektor UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
2. Dr. Sulistiono, M.Si., selaku Dekan Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
3. Dr. Sucipto, M. Kom., selaku Ka Prodi Sistem Informasi UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
4. Arie Nugroho, M.Kom dan Aidina Ristyawan M.Kom, selaku Pembimbing skripsi yang dengan sabar membimbing dan memotivasi dalam menyelesaikan skripsi ini.
5. Kedua orang tua saya dan Keluarga yang telah mendukung dalam do'a, dana dan motivasi dalam penyelesaian skripsi ini.
6. Semua pihak yang telah membantu, baik secara langsung maupun tidak langsung, dalam proses penyusunan skripsi ini.

Disadari bahwa skripsi ini masih banyak kekurangan, maka kritik dan saran yang membangun sangat diharapkan. Akhirnya, disertai harapan semoga skripsi ini ada manfaatnya bagi kita semua, khususnya bagi dunia pendidikan, meskipun hanya ibarat setitik air bagi samudra yang luas.

Kediri, 12 Juni 2026



ANIS FAIZUL LAILA

NPM: 2213030066

DAFTAR ISI

Halaman Sampul	i
Halaman Persetujuan.....	ii
Halaman Pengesahan	iii
Halaman Pernyataan.....	iv
Halaman Motto.....	v
ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Batasan Masalah.....	2
C. Rumusan Masalah	3
D. Tujuan Penelitian	3
E. Manfaat Penelitian	3
BAB II KAJIAN PUSTAKA	4
A. Kajian Teori	4
1. Curah Hujan	4
2. Klasifikasi	4
3. Data Mining	5
4. Cross Industry Standard process for Data Mining	5
5. Random Forest	6
6. Seleksi Fitur	6
7. Wrapper Methode.....	7
8. Recursive Feature Elimination with Cross-Validation (RFECV).....	7
9. Class Imbalance	8
10. Evaluasi Model.....	9
B. Kajian Hasil Penelitian Terdahulu	11

C. Kerangka Berpikir	13
BAB III METODE PENELITIAN	15
A. Business Understanding (Pemahaman Bisnis).....	15
B. Data Understanding (Pemahaman Data).....	16
C. <i>Data Preparation</i> (Persiapan Data)	18
D. Modeling	20
E. <i>Evaluation</i> (Evaluasi).....	22
F. <i>Deployment</i> (Penerapan)	23
BAB IV HASIL DAN PEMBAHASAN	24
A. Data Hasil Penelitian.....	24
1. Hasil eksplorasi data	24
2. Data Preparation.....	27
3. Hasil Modeling.....	30
4. Perbandingan Model	34
B. Pembahasan.....	36
1. Model Optimal	36
2. Evaluasi Hasil Modeling Terhadap Tujuan Bisnis.....	37
3. Kesesuaian Hasil Penelitian	38
4. Validitas Hasil	39
5. Rekomendasi Penerapan Model	39
BAB V PENUTUP.....	41
A. Kesimpulan	41
B. Saran.....	41
DAFTAR PUSTAKA	42
LAMPIRAN.....	47

DAFTAR GAMBAR

Gambar 2.1 Tahapan Mendapatkan Model Terbaik	9
Gambar 2.2 Kerangka Berpikir	14
Gambar 4.1 Kode Konversi Datetime	24
Gambar 4.2 Kode Untuk Ekstraksi Fitur Bulan	24
Gambar 4.3 Histogram Fitur Numerik	26
Gambar 4.4 Mendeteksi Outlier	27
Gambar 4.5 Mapping Arah Angin	27
Gambar 4.6 Distribusi Kelas Sebelum Dan Sesudah Diseimbangkan	30
Gambar 4.7 <i>Confusion matrix</i> Model RFECV	32
Gambar 4.8 <i>Confusion matrix</i> Metode RFE	33
Gambar 4.9 <i>Confusion matrix</i> Metode SFS	34
Gambar 4.10 Perbandingan Kinerja Model Menggunakan RFECV dengan Penelitian Sebelumnya	38
Gambar 4.11 Grafik Jumlah Fitur vs Skor Performa RFECV	39
Gambar 4.12 Tampilan Input Early Warning System Curah Hujan	40
Gambar 4.13 Tampilan Hasil Analisis Sistem	40

DAFTAR TABEL

Tabel 4.1 Deskripsi Atribut Dataset.....	25
Tabel 4.2 Jumlah Missing Value Pada Atribut	25
Tabel 4.3 Kategori Intensitas Curah Hujan.....	28
Tabel 4.4 Skema Eksperimen Seleksi Fitur	29
Tabel 4.5 Hasil Klaisfikasi RFECV	31
Tabel 4.6 Hasil Klasifikasi RFE.....	33
Tabel 4.7 Hasil Evaluasi SFS.....	34
Tabel 4.8 Perbandingan Performa Model dengan Penelitian Sebelumnya	35

DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan Skripsi	47
Lampiran 2. Surat Keterangan Bebas Similarity dari PPI.....	48
Lampiran 3. Bukti Skor Cek Similarity	49
Lampiran 4. Bukti Screenshot Submit Artikel.....	50
Lampiran 5. Lembar Berita Acara Ujian.....	51
Lampiran 6. Lembar Revisi Ujian.....	52

BAB I

PENDAHULUAN

A. Latar Belakang

Tingginya intensitas curah hujan di Indonesia menjadi salah satu pemicu terjadinya bencana hidrometeorologi, seperti banjir. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB, 2025), bencana banjir menempati urutan tertinggi di Indonesia dengan frekuensi mencapai lebih dari 1.400 kejadian dalam satu tahun. Guna meminimalkan dampak risiko bencana tersebut, Badan Meteorologi, Klimatologi, dan Geofisika (BMKG, 2020) telah menetapkan standar klasifikasi curah hujan harian ke dalam empat kategori, yaitu hujan ringan (0–20 mm/hari), sedang (20–50 mm/hari), lebat (50–100 mm/hari), dan sangat lebat (>100 mm/hari). Informasi klasifikasi yang presisi dari standar ini sangat dibutuhkan sebagai indikator awal dalam sistem deteksi dini (Faradila, 2026). Penentuan kategori tersebut juga menjadi landasan penting dalam perancangan manajemen risiko bencana oleh instansi pemerintah terkait (Safriani et al., 2025).

Seiring perkembangan teknologi komputasi, ekstraksi pola dari data historis parameter cuaca dapat dilakukan melalui pendekatan *data mining* (Asiah et al., 2025). Salah satu algoritma *ensemble learning* yang sering digunakan untuk klasifikasi adalah *Random Forest*, karena mampu menangani hubungan non-linier serta tidak sensitif terhadap penciran (*outliers*) pada data berdimensi tinggi (Sangaji et al., 2024). Meskipun algoritma ini dikenal memiliki performa yang kuat, akurasi akhir model *Random Forest* sangat bergantung pada kualitas dan relevansi fitur yang dimasukkan ke dalam proses pelatihan model (Elsa Damayanti et al., 2025).

Kelemahan metodologis berupa pengabaian seleksi fitur statistik pada pemodelan iklim multikelas ditemukan dalam beberapa penelitian terdahulu. Penelitian acuan pertama oleh (Indaryono et al., 2024) yang mengklasifikasikan empat tingkat intensitas curah hujan di Indonesia dengan algoritma *Random Forest* berhasil mencapai akurasi 86%. Namun, nilai tersebut diperoleh melalui pembatasan empat variabel saja (Tavg, RH_avg, ss, dan Ddd_x) yang dipilih secara manual berdasarkan intuisi peneliti tanpa pengujian statistik formal. Masalah ini diperkuat oleh penelitian kedua dari (Maheswara & Al'aziz, 2025) pada data curah

hujan multikelas di Bogor, di mana model *Random Forest* langsung memproses seluruh parameter cuaca tanpa seleksi fitur dan hanya menghasilkan nilai *F1-score* sebesar 65,5%, yang mengindikasikan akurasi model kurang optimal akibat dimensi data yang redundan.

Hal itu juga berlaku pada klasifikasi biner (dua kelas), ketiadaan seleksi fitur tetap membatasi performa model. Penelitian ketiga oleh (Hapsari & Pandya Wisesa, 2025) mengenai prediksi kejadian hujan di Bandung menunjukkan bahwa model *Random Forest* tanpa seleksi fitur menghasilkan nilai *F1-score* yang rendah, yaitu berkisar antara 66% hingga 72%, karena model kesulitan mengenali pola akibat gangguan *noise* data. Berdasarkan penjelasan tersebut, dapat disimpulkan bahwa, ketiadaan seleksi fitur statistik membuat algoritma *Random Forest* rentan terhambat oleh dimensi data yang redundan atau mengandung *noise*.

Oleh karena itu, penelitian ini menerapkan metode *Recursive Feature Elimination with Cross-Validation* (RFECV) yang bekerja secara induktif dengan mengeliminasi fitur berkinerja rendah secara berulang berdasarkan bobot kepentingan fitur (*feature importance*) dari model *Random Forest*. Mekanisme *Cross-Validation* di dalamnya juga menjamin validasi subset fitur yang ketat untuk menghindari *overfitting* (Sumathi & lowast, 2025). Masalah dalam penelitian ini adalah pemilihan fitur pada data curah hujan masih dilakukan secara subjektif dan tanpa seleksi fitur sehingga menurunkan performa model. Penerapan metode ini bertujuan untuk memperoleh subset fitur terbaik guna mengoptimalkan performa model *Random Forest* dalam klasifikasi curah hujan sebagai rujukan komputasi dalam mendukung sistem mitigasi bencana secara dini.

B. Batasan Masalah

1. Menggunakan data curah hujan dari Kaggle, dengan klasifikasi curah hujan menjadi Ringan, Sedang, Lebat, dan Sangat Lebat.
2. Penyeimbangan data dilakukan menggunakan metode *SMOTE* sesuai penelitian sebelumnya.
3. Model yang digunakan adalah *Random Forest*, dan penelitian ini fokus pada penerapan seleksi fitur menggunakan *Recursive Feature Elimination with Cross-Validation* (RFECV) untuk menghilangkan fitur yang tidak relevan dan

memperoleh subset fitur optimal sehingga dapat meningkatkan kinerja klasifikasi.

4. Evaluasi kinerja model dibatasi pada metrik akurasi, presisi, *Recall* dan *F1-score*, dengan fokus utama pada nilai presisi dalam menilai kinerja model.

C. Rumusan Masalah

1. Fitur-fitur apa saja yang terpilih sebagai subset optimal berdasarkan hasil seleksi fitur pada data curah hujan?
2. Seberapa besar pengaruh kinerja model *Random Forest* sebelum dan sesudah penerapan seleksi fitur dalam mengklasifikasikan intensitas curah hujan?

D. Tujuan Penelitian

1. Menentukan subset fitur optimal pada data curah hujan menggunakan *Recursive Feature Elimination with Cross-validation*.
2. Membandingkan kinerja model *Random Forest* sebelum dan sesudah penerapan seleksi fitur *Recursive Feature Elimination with Cross-validation*.

E. Manfaat Penelitian

1. Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi pada bidang data mining, khususnya sebagai referensi empiris terkait efektivitas metode *Recursive Feature Elimination with Cross-Validation* (RFECV) dalam menyeleksi fitur optimal pada model *Random Forest* untuk klasifikasi intensitas curah hujan multikelas.

2. Manfaat Praktis

Penelitian ini diharapkan dapat memberikan informasi bagi Stasiun Klimatologi BMKG mengenai fitur-fitur atmosfer dominan hasil seleksi RFECV. Informasi tersebut diharapkan dapat menjadi acuan untuk menyederhanakan model prediksi curah hujan sehingga beban komputasi lebih efisien tanpa menurunkan akurasi secara signifikan. Selain itu, hasil penelitian ini diharapkan dapat menjadi bahan pertimbangan awal bagi peneliti lain yang mengembangkan sistem deteksi dini bencana hidrometeorologi berbasis data BMKG.

DAFTAR PUSTAKA

- Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, & Apriade Voutama. (2022). Analisis Cluster Faktor Penunjang Menggunakan ALgoritma K-Means (Studi Kasus: Kabupaten Karawang). *Jurnal Informatika Dan Rekayasa Elektronik*, 5(2), 244–252. <https://doi.org/10.36595/JIRE.V5I2.701>
- Aini, N., & Nasution, M. I. P. (2025). Akurasi Kualitas Data Informasi Pada Sistem Manajemen. *JURNAL RUMPUN MANAJEMEN DAN EKONOMI*, 2(1), 40–50. <https://doi.org/10.61722/JRME.V2I1.3259>
- Ardan, T. S., Anggara Mujiatno, S., & Akbar, B. A. (2025). Implementasi Data Analytics Menggunakan Metodologi CRISP-DM (Studi Kasus Data Bike Sales In Europe Pada Platform Kaggle). *JSiI | Jurnal Sistem Informasi*, 12(2), 10–15. <https://doi.org/10.30656/jsii.v12i2.10761>
- Asiah, S., Riana, W., Chryston Purba, D., Ilham Azharsum, M., Asido Elyakim, V. P., Tunas Bangsa, S., Artikel, G., Kunci, K., & Fuzzy Prediksi Curah Hujan Ketidakpastian Data, A. (2025). Prediksi Curah Hujan Tahunan di Indonesia Menggunakan Model Hybrid Jaringan Syaraf Tiruan dan Algoritma Fuzzy Annual Rainfall Prediction in Indonesia Using A Hybrid Artificial Neural Network and Fuzzy Algorithm Model. *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, 4(2), 2828–9099. <https://doi.org/10.55123/jomlai.v4i2.5964>
- Aufa, R. A., Altirika, E., Pramudyantoro, A., Pratama, Y. B., & Wahyuzi, Z. (2025). Pengenalan Pola Kasus Potensi Banjir Di Pangkalpinang Dengan Algoritma Random Forest Dan XGBoost Menggunakan Google Earth Engine. *Journal of Information Technology and Society*, 3(1), 1–6. <https://doi.org/10.35438/JITS.V3I1.1415>
- Aulia, S. P., Rahmat, B., & Junaidi, A. (2025). Enhancing Heart Disease Prediction through SMOTE-ENN Balancing and RFECV Feature Selection. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(3), 1968–1973. <https://doi.org/10.59934/JAIEA.V4I3.1057>
- Awad, M., & Fraihat, S. (2023). Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems. *Journal of Sensor and Actuator Networks 2023, Vol. 12, Page 67*, 12(5), 67. <https://doi.org/10.3390/JSAN12050067>
- Ayu Syaharani, M., Arif Dermawan, B., & Ibnu Adam, R. (2025). Prediksi Curah Hujan Kota Bogor Menggunakan Algoritma Random Forest. *Jurnal Sistem Informasi Dan Aplikasi*, 3(1), 52–62. <https://doi.org/10.52958/JSIA.V3I1.12459>
- Aziz, F., Supriyadi, ;, Wungo, L., & Irmawati, ; (2026). Adaptive Ensemble LS-SVM with Stability Feature Selection for Drug-Induced Autoimmunity Prediction. *Innovation Journal of Pharmaceutical Science*, 2(1), 22–30. <https://journal.irmexdigika.com/index.php/injophars/article/view/92>
- Bintoro, P., Ratnasari, Wihardjo, E., Putri, I. P., & Asari, A. (2024). *Pengantar Machine Learning* (A. Asari, Ed.). PT MAFY MEDIA LITERASI INDONESIA. www.penerbitmafy.com
- BMKG. (2020). *Badan Meteorologi Klimatologi Dan Geofisika Stasiun Meteorologi Umu Mehang Kunda Sumba Timur Buletin Stasiun Meteorologi Umu Mehang Sumba Timur*. www.meteowangapu.com
- BNPB. (2025). *Buletin Info Bencana Desember 2025* (A. Rosyida, Ed.). <https://surl.li/dlnqgk>

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- Cahyo, A. D., & Wijayanto, A. W. (2026). Penerapan CRISP-DM Model dengan Algoritma Decision Tree, Naive Bayes, SVM, dan KNN untuk Klasifikasi Curah Hujan di Kabupaten Malang. *EKSPONENSIAL*, 17, 1–8. <https://doi.org/10.30872/eksponensial.v17i1.1649>
- Catursari, S. A. (2024). *Analisis Pengaruh Teknana Udara, Kelembaban Udara, Dan Suhu Udara Terhadap Curah Hujan Di Kota Surabaya* [UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG]. <http://etheses.uin-malang.ac.id/id/eprint/65664>
- Drogkoula, M., Kokkinos, K., & Samaras, N. (2023). A Comprehensive Survey of Machine Learning Methodologies with Emphasis in Water Resources Management. *Applied Sciences* 2023, Vol. 13, Page 12147, 13(22), 12147. <https://doi.org/10.3390/APP132212147>
- Dwinanda, I. G., Cassiophea, L., Saragih, O. K., Thareq, S. I., Permanasuri, N. P. D. A., & Kurniawati, M. (2025). Penerapan Teknologi Sistem Peringatan Dini dalam Upaya Pengurangan Resiko Bencana Hidrometeorologi di Kelurahan Tumbang Tahai. *PengabdianMu: Jurnal Ilmiah Pengabdian Kepada Masyarakat*, 10, 361–370. <https://doi.org/10.33084/pengabdianmu.v10isuppl-1.9266>
- Eina, M. F., Chrisnanto, Y. H., & Melina, M. (2024). Klasifikasi Telemarketing Menggunakan Naïve Bayes Classification Dan Wrapper Sequential Feature Selection. *INTECOMS: Journal of Information Technology and Computer Science*, 7(4), 1189–1198. <https://doi.org/10.31539/INTECOMS.V7I4.10846>
- Elsa Damayanti, A., Andriani, E., & Giyanfranco Zola, M. (2025). Implementasi Machine Learning Untuk Prediksi Curah Hujan di Daerah Rawan Banjir. In *Prosiding Seminar Nasional Ilmu Matematika dan Sains* (Number 1). <https://prosiding.arimsi.or.id/index.php/PROSEMNASIMSI>
- Faisal, M. R., & Nugrahadi, D. T. (2019). *Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R - Google Books*. Scripta Cendekia.
- Faradila, R. N. (2026). Pemetaan Kerentanan Banjir Berbasis Penginderaan Jauh dan Sistem Informasi Geografis untuk Mendukung Mitigasi Bencana di Kecamatan Panti, Kabupaten Jember. *Jurnal Penginderaan Jauh Indonesia*, 5(1), 25–34. <https://doi.org/10.12962/JPJI.V5I1.9513>
- Fatah, Z., & Muhfidha, L. (2026). Penerapan Metode Random Forest Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu di Pondok Pesantren Salafiyah Syafi'iyah. In *JAMASTIKA* (Vol. 5). <https://doi.org/https://doi.org/10.35473/jamastika.v5i1.4638>
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1–3), 389–422. <https://doi.org/10.1023/A:1012487302797/METRICS>
- Han, J., Kamber, M., & Pei, J. (2011). Data Transformation by Normalization. *Data Mining: Concepts and Techniques*, 113–115. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Han, J., Kamber, M., & Pei, Jian. (2012). *Data Mining Concepts and Techniques Third Edition*. In *ICB Research Reports* (Third, Number 9). Elsevier. <https://openlibrary.telkomuniversity.ac.id/home/catalog/id/172825/slug/data-mining-concepts-and-techniques-third-edition.html>

- Hapsari, I., & Pandya Wisesa, S. (2025). Evaluasi Model Prediksi Curah Hujan Berbasis Machine Learning di Kota Bandung. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 11(2), 136–143. <https://doi.org/10.25077/teknosi.v11i2.2025.136-143>
- Hidayat, W., Yoga Siswa, T. A., & Hasudungan, R. (2025). Penerapan Algoritma Naive Bayes, Recursive Feature Elimination, dan Adaptive Synthetic Sampling Pada Klasifikasi Penyakit Dermatitis. *Jurnal INSTEK (Informatika Sains Dan Teknologi)*, 10, 424–436. <https://doi.org/10.24252/instek.v10i2.59737>
- Huizen, L. M., Ardima, M. B., & Idris, M. (2025). Meningkatkan kinerja SVM: Dampak berbagai teknik seleksi fitur pada akurasi prediksi. *AITI*, 22(1), 1–14. <https://doi.org/10.24246/aiti.v22i1.1-14>
- Indaryono, N. A. P., Saedudin, Rd. R., & Hamami, F. (2024). Analisa Perbandingan Algoritma Random Forest Dan Naive Bayes Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Di Indonesia. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(1), 158–167. <https://doi.org/10.29100/jupi.v9i1.4421>
- J., S., & U., Dr. K. (2022). Comparison of Sentiment Analysis on Online Product Reviews Using Optimised RNN-LSTM with Support Vector Machine. *Webology*, 19(1), 3883–3898. <https://doi.org/10.14704/web/v19i1/web19256>
- Kiwasi Wororomi, J., Felix Reba, Ms., Aleksander Mandowen, S., Alvian Sroyer, M. M., Halomoan Edy Manurung, Ms., Mayko Edison Koibur, Mc., Erna Hudianti Pujiarini, Me., Marwan Ramdhany Edy, Ms., Hajjar Yuliana, Mk., Mukarramah Yusuf, M., & Marta Dinata, R. (2024). *Data Mining (Memahami Pola Di Balik Angka)* (N. R. Jayanti, Ed.). EUREKA MEDIA AKSARA.
- Lukman, S., & Priyadi. (2024). Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI. *Jurnal Elektronika Dan Komputer*, 17(2). <https://doi.org/10.51903/elkom.v17i2.2311>
- Maheswara, I. D. G. L., & Al'aziz, A. H. (2025). Perbandingan Model Machine Learning Pada Klasifikasi Curah Hujan Di Bogor. *INTI Nusa Mandiri*, 19(2), 202–210. <https://doi.org/10.33480/INTI.V19I2.6296>
- Novelan, M. S., & Aryza, S. (2025). *Belajar Mengenal Dasar Machine Learning* (R. M. Sari, Ed.). Serasi Media Teknologi.
- Novita, L., & Aziza, D. K. (2025). *Komunikasi Dan Mitigasi Bencana: Analisis Peran Media Sosial dan Website BMKG serta Perannya dalam Pengembangan Ekonomi Hijau* (M. H. Fuadi, Ed.; 1st ed.). K-Media.
- Nugroho, A., Fanani, A. Z., & Shidik, G. F. (2021). Evaluation of Feature Selection Using Wrapper for Numeric Dataset with Random Forest Algorithm. *Proceedings - 2021 International Seminar on Application for Technology of Information and Communication: IT Opportunities and Creativities for Digital Innovation and Communication within Global Pandemic, ISemantic 2021*, 179–183. <https://doi.org/10.1109/ISEMANTIC52711.2021.9573249>
- Nugroho, A., & Harini, D. (2024). Teknik Random Forest untuk Meningkatkan Akurasi Data Tidak Seimbang. *JSITIK: Jurnal Sistem Informasi Dan Teknologi Informasi Komputer*, 2(2), 128–140. <https://doi.org/10.53624/JSITIK.V2I2.379>
- Pirnando Sitepu, J., Mongi, C. E., Widiyana, S. A., Ketaren, E., Program,), Sistem, S., Fakultas, I., Dan, M., & Pengetahuan, I. (2025). Prediksi Banjir Di Kota Manado Menggunakan Metode Random Forest. *Jurnal TIMES*, 14(1), 70–75. <https://doi.org/10.51351/JTM.14.1.2025826>

- Putra, A. F. D., Putra, A. F. D., Azmi, M. N., Wijayanto, H., Utama, S., & Wirawan, I. G. P. W. W. (2024). Optimizing Rain Prediction Model Using Random Forest and Grid Search Cross-Validation for Agriculture Sector. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 23(3), 519–530. <https://doi.org/10.30812/matrik.v23i3.3891>
- Rahayu, P. W., Sudipa, I. G. I., Suryani, Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, Muh. N., Slamet, I., Harlina, S., & Maysanjaya, I. M. D. (2024). *Buku Ajar Data Mining* (Efitra, Ed.). PT. Sonpedia Publishing Indonesia. <https://www.researchgate.net/publication/377415198>
- Rahmi, R., Antoni, D., Syaputra, H., Fatoni, F., & Kurniawan, T. B. (2023). Metode Klasifikasi Gejala Penyakit Coronavirus Disease 19 (COVID-19) Menggunakan Algoritma Neural Network. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 12, 16–23. <https://doi.org/10.32736/sisfokom.v12i1.1406>
- Ramadhan, N. C., H. H. H., Rohana, T., & Siregar, A. M. (2024). Optimasi Algoritma Machine Learning Menggunakan Seleksi Fitur Xgboost Untuk Klasifikasi Kanker Payudara. *TIN: Terapan Informatika Nusantara*, 5(2), 162–171. <https://doi.org/10.47065/tin.v5i2.5408>
- Ristyawan, A., Nugroho, A., & Amarya, T. K. (2025). Optimasi Preprocessing Model Random Forest untuk Prediksi Stroke. *JATISI*, 12(1). <https://doi.org/10.35957/JATISI.V12I1.9587>
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27. <https://doi.org/10.1002/WIDM.1075;PAGE:STRING:ARTICLE/CHAPTER>
- Rosidah, N., Prathivi, R., & Susanto, S. (2025). Optimasi Metode Random Forest Untuk Klasifikasi Resiko Obesitas Berdasarkan Pola Makan. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, 7, 56–62. <https://doi.org/10.51401/jinteks.v7i1.5065>
- Ruswanti, D., Ruswanti, D., Susilo, D., & Riani, R. (2024). Implementasi CRISP-DM pada Data Mining untuk Melakukan Prediksi Pendapatan dengan Algoritma C.45. *Go Infotech: Jurnal Ilmiah STMIK AUB*, 30(1), HAL. 111–121. <https://doi.org/10.36309/goi.v30i1.266>
- Ruswanti, D., Susilo, D., & Riani, R. (2024). Implementasi CRISP-DM pada Data Mining untuk Melakukan Prediksi Pendapatan dengan Algoritma C.45. *Go Infotech: Jurnal Ilmiah STMIK AUB*, 30, 111–121. <https://doi.org/10.36309/goi.v30i1.266>
- Safriani, M., Arsyad, Agusalm, M., Apriyanto, Zulharnah, Hasdanita, F., Suprpti, Mansida, A., Rohani, I., & Fachruddin. (2025). *Hidrologi* (W. Yuliani, Ed.; Cetakan pertama). Lingkar Edukasi Indonesia. <https://share.google/31eutQRGKAICvWwGa>
- Sangaji, D., Sutabri, T., Jend, J. A., Yani, N., 0711-515582 Palembang, T., & Selatan, S. (2024). *Optimalisasi Prediksi Indeks Kualitas Air di Indonesia dengan Menggunakan Machine Learning Melalui Pendekatan Metode Prophet*. (2), 1–14. <https://doi.org/10.62951/switch.v2i6.277>
- Setiawan, I., Fatah Yasin, I., & Tri Desianti, Y. (2025). Komparasi Kinerja Algoritma Random Forest, Decision Tree, Naïve Bayes, dan KNN dalam Prediksi Tingkat Depresi Mahasiswa menggunakan Student Depression Dataset. *Jurnal Ilmu Komputer Dan Teknologi*, 6, 47–58. <https://doi.org/10.35960/ikomti.v6i1.1756>
- Singgalen, Y. A. (2023). Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT). *JURNAL MEDIA INFORMATIKA BUDIDARMA*. <https://doi.org/10.30865/mib.v7i3.6461>

- Siregar, A. M., & Puspabhuana, D. A. N. A. (2017). *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group. https://books.google.com/books/about/DATA_MINING.html?hl=id&id=rTlmDwAAQBAJ
- Sukman, S., Rustan, F. R., Yusaman, Y., Tanje, H. W., Sukri, A. S., Amir, Muh. K., Syajruddin, S., Sriwati, M., & Rachman, R. M. (2024). *Hidrologi tohar media* (Number June). TOHAR MEDIA.
- Sulung, U., & Muspawi, M. (2024). Memahami sumber data penelitian: Primer, sekunder, dan tersier. *Edu Reserch*, 5, 110–116.
- Sumathi, S., & lowast, R. R. (2025). Feature-Optimized Random Forest Model for Wildfire Prediction using Weather Information. *Indian Journal of Science and Technology*, 18(19), 1530–1537. <https://doi.org/10.17485/IJST/V18I19.442>
- Suparnoh, Sumarno, Maksun, T., Hartono, I. D., & Masnun. (2025). *Python: Python untuk Para Penjelajah Data Seni dan Ilmu Memproses Informasi Panduan Komprehensif dengan Studi Kasus Praktis* (A. F. Qohar & W. H. Y. Afandi, Eds.). PT Penerbit Qriset Indonesia. www.qrisetindonesia.com
- Tallar, R. Y. (2023). *Dasar-Dasar Hidrologi Terapan* (R. T. Manurung, Ed.). www.ideaspublishing.co.id
- Wallace, M. L., Mentch, L., Wheeler, B. J., Tapia, A. L., Richards, M., Zhou, S., Yi, L., Redline, S., & Buysse, D. J. (2023). Use and misuse of random forest variable importance metrics in medicine: demonstrations through incident stroke prediction. *BMC Medical Research Methodology* 2023 23:1, 23(1), 1–12. <https://doi.org/10.1186/S12874-023-01965-X>
- Wirth, R., & Hipp, J. (n.d.). *CRISP-DM: Towards a Standard Process Model for Data Mining*.