

**Prediksi Curah Hujan Pemicu Banjir Menggunakan
Model XGBoost dan Analisis SHAP di Kota Bandung**

SKRIPSI SARJANA KOMPUTER

Oleh

Alfi Savio Rasyad

227006516083



PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL

2026

**PREDIKSI CURAH HUJAN PEMICU BANJIR
MENGUNAKAN MODEL XGBOOST DAN
ANALISIS SHAP DI KOTA BANDUNG**

Karya ilmiah sebagai salah satu syarat untuk memperoleh
gelar Sarjana Sistem Informasi dari Fakultas Teknologi

Komunikasi dan Informatika

Oleh

Alfi Savio Rasyad

227006516083



PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI
DAN INFORMATIKA
UNIVERSITAS NASIONAL

2026

LEMBAR PENGESAHAN

HALAMAN PENGESAHAN
TUGAS AKHIR

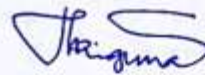


Dosen Pembimbing 1



Dr. Ir. Andrianingsih, S.Kom., MMSI.

Dosen Pembimbing 2



Dr. Arie Gunawan, S.Kom., MMSI.

PERNYATAAN KEASLIAN TUGAS AKHIR

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

Prediksi Curah Hujan Pemicu Banjir Menggunakan Model XGBoost dan Analisis SHAP di Kota Bandung

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 27 Februari 2026

(TT)
10000
METRA
TEMPEL
788ESANX271229132

Alfi Savio Rasyad

227006516083

LEMBAR PERSETUJUAN REVIEW AKHIR

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

Prediksi Curah Hujan Pemicu Banjir Menggunakan Model XGBoost dan Analisis SHAP di Kota Bandung

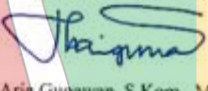
Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2025-2026 pada tanggal 24 Februari Tahun 2026

Dosen Pembimbing 1


Dr. Ir. Andrianaesih, S.Kom., MMSI


MIDN: 0303097902

Dosen Pembimbing 2


Dr. Arie Gunawan, S.Kom., MMSI

NIDN: 0410047808

Ketua Program Studi


Dr. Ir. Andrianaesih, S.Kom., MMSI

MIDN: 0303097902

UNIVERSITAS NASIONAL

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Alfi Savio Rasyad
 NPM : 227006516083
 Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
 Program Studi : Sistem Informasi
 Tanggal Sidang : 24 Februari 2026

JUDUL DALAM BAHASA INDONESIA :

Prediksi Curah Hujan Pemicu Banjir Menggunakan Model XGBoost dan Analisis SHAP di Kota Bandung

JUDUL DALAM BAHASA INGGRIS :

Flood-Triggering Rainfall Prediction Using the XGBoost Model and SHAP Analysis in Bandung City

TANDA TANGAN DAN TANGGAL		
Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 27/2/2026	TGL : 27/2/2026	TGL : 27 / 02 / 2026

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

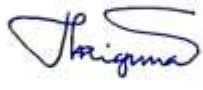
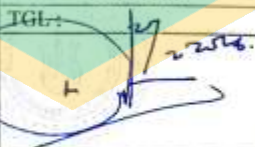

Nama : Alfi Savio Rasyad
 NPM : 227006516083
 Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
 Program Studi : Sistem Informasi
 Tanggal Sidang : 24 Februari 2026

JUDUL DALAM BAHASA INDONESIA :

Prediksi Curah Hujan Pemicu Banjir Menggunakan Model XGBoost dan Analisis SHAP di Kota Bandung

JUDUL DALAM BAHASA INGGRIS :

Flood-Triggering Rainfall Prediction Using the XGBoost Model and SHAP Analysis in Bandung City

TANDA TANGAN DAN TANGGAL		
Pembimbing 2	Ka. Prodi	Mahasiswa
TGL :	TGL :	TGL : 27/02/2026
		

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia sehingga penulis dapat menyelesaikan skripsi dengan judul **“Prediksi Curah Hujan Pemicu Banjir Menggunakan Model XGBoost dan Analisis SHAP di Kota Bandung”** sebagai salah satu syarat kelulusan Program Studi Sarjana Fakultas Teknologi Komunikasi dan Informatika. Penelitian dan penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak terima kasih terutama kepada dosen pembimbing Tugas Akhir, Dr.Ir. Andrianingsih, S.Kom.,MMSI. dan Dr. Arie Gunawan, S.Kom., MMSI. yang telah meluangkan banyak waktu, tenaga, pikiran, arahan, motivasi serta memaklumi segala kekurangan penulis selama penelitian tugas akhir dan penyusunan skripsi. Penulis juga mengucapkan banyak terima kasih kepada:

1. Orang tua tercinta, atas doa, kasih sayang, serta dukungan material dan spiritual yang menjadi motivasi utama penulis.
2. Seluruh dosen pengajar di Program Studi Sistem Informasi FTKI maupun dosen di Program Studi lain yang memberikan banyak ilmu.
3. Teman-teman seangkatan dan sehimpunan berbagai angkatan yang telah membantu dan mendukung.
4. Sashika Vinora Tasliano, atas kasih sayang, kesabaran, dan dukungan semangat yang tak henti diberikan kepada penulis.

Akhir kata, semoga Tuhan Yang Maha Esa membalas kebaikan dan bantuan yang telah diberikan dengan hal yang lebih baik. Penulis mengharapkan kritik dan saran yang bersifat membangun dan semoga skripsi ini dapat memberikan manfaat di bidang Teknologi Informatika.

Jakarta, Februari 2026

Alfi Savio Rasyad

ABSTRAK

Banjir di Kota Bandung merupakan tantangan hidrometeorologis persisten yang sering kali dipicu oleh curah hujan ekstrem dengan intensitas tinggi dan durasi singkat. Metode prediksi konvensional sering kali gagal menangkap pola non-linier dan karakteristik distribusi data hujan yang bersifat heavy-tailed dan zero-inflated. Penelitian ini bertujuan untuk mengembangkan model prediksi curah hujan harian yang akurat menggunakan algoritma Extreme Gradient Boosting (XGBoost) serta menerapkan metode Shapley Additive exPlanations (SHAP) untuk mengatasi masalah "kotak hitam" (black-box) pada model machine learning demi kepentingan mitigasi bencana. Penelitian ini memanfaatkan data meteorologis harian dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) periode 2020–2024.

Metodologi penelitian melibatkan pra-pemrosesan data yang komprehensif, termasuk penanganan outlier dan interpolasi berbasis waktu, serta rekayasa fitur (feature engineering) yang membentuk fitur lag dan statistik bergerak (rolling statistics) untuk menangkap persistensi cuaca. Kinerja model dievaluasi menggunakan metrik RMSE, MAE, dan R^2 , serta dibandingkan dengan model baseline seperti Regresi Linier, Random Forest, dan LSTM. Hasil eksperimen menunjukkan bahwa model XGBoost dengan optimasi hiperparameter Bayesian menghasilkan kinerja paling unggul dan stabil dibandingkan model pembanding. Analisis SHAP berhasil mengungkap bahwa fitur historis curah hujan (akumulasi hujan 3 hari sebelumnya) dan kelembapan rata-rata merupakan variabel paling dominan yang memengaruhi prediksi hujan ekstrem. Selain itu, penelitian ini mengidentifikasi ambang batas fisik (physical threshold) pemicu kenaikan probabilitas hujan ekstrem secara interpretatif. Temuan ini diimplementasikan dalam sebuah dashboard sistem informasi yang dapat berfungsi sebagai pendukung keputusan dalam sistem peringatan dini banjir yang transparan dan berbasis data bagi pemangku kepentingan di Kota Bandung.

Kata Kunci: Prediksi Curah Hujan, Banjir, XGBoost, SHAP, Feature Engineering, Kota Bandung.

ABSTRACT

Flooding in Bandung City represents a persistent hydrometeorological challenge, predominantly triggered by extreme rainfall events characterized by high intensity and short duration. Conventional prediction methods often fail to capture non-linear patterns and the heavy-tailed, zero-inflated distribution characteristics of rainfall data. This study aims to develop an accurate daily rainfall prediction model using the Extreme Gradient Boosting (XGBoost) algorithm and applies the Shapley Additive exPlanations (SHAP) method to address the "black-box" issue in machine learning models for disaster mitigation purposes. The research utilizes daily meteorological data from the Meteorology, Climatology, and Geophysics Agency (BMKG) for the period 2020–2024.

The methodology involves comprehensive data preprocessing, including outlier handling and time-based interpolation, as well as feature engineering that constructs lag features and rolling statistics to capture weather persistence. Model performance was evaluated using RMSE, MAE, and R^2 metrics, and benchmarked against baseline models such as Linear Regression, Random Forest, and LSTM. Experimental results demonstrate that the XGBoost model with Bayesian hyperparameter optimization yielded the most superior and robust performance compared to the comparative models. SHAP analysis successfully revealed that historical rainfall features (3-day rainfall accumulation) and average humidity are the most dominant variables influencing extreme rainfall predictions. Furthermore, this study identified interpretative physical thresholds that trigger a spike in extreme rainfall probability. These findings are implemented in an information system dashboard that serves as a decision support tool within a transparent, data-driven flood early warning system for stakeholders in Bandung City.

Keywords: Rainfall Prediction, Flood, XGBoost, SHAP, Feature Engineering, Bandung City.

DAFTAR ISI

PREDIKSI CURAH HUJAN PEMICU BANJIR MENGGUNAKAN MODEL XGBOOST DAN ANALISIS SHAP DI KOTA BANDUNG	II
LEMBAR PENGESAHAN	ERROR! BOOKMARK NOT DEFINED.
PERNYATAAN KEASLIAN TUGAS AKHIR.....	IV
LEMBAR PERSETUJUAN REVIEW AKHIR	V
LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI	VI
KATA PENGANTAR.....	VIII
ABSTRAK	IX
ABSTRACT	X
DAFTAR ISI.....	XI
DAFTAR TABEL	XVI
DAFTAR GAMBAR	XVII
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH	2
1.3 TUJUAN PENELITIAN	3
1.4 MANFAAT PENELITIAN	3
1.5 BATASAN MASALAH.....	4
BAB II TINJAUAN PUSTAKA	6
2.1 LITERATUR REVIEW	6
2.1.1 <i>Prediksi Curah Hujan Menggunakan Pendekatan Statistik dan Machine Learning.....</i>	<i>7</i>
2.1.2 <i>Penerapan XGBoost dalam Konteks Hidrometeorologi</i>	<i>13</i>
2.1.3 <i>Urgensi Interpretabilitas Model dan SHAP dalam Studi Lingkungan</i>	

2.1.4	<i>Analisis Kesenjangan Penelitian</i>	16
2.2	LANDASAN TEORI	17
2.2.1	<i>Sistem Informasi</i>	18
2.2.2	<i>Data Mining dan Machine Learning</i>	18
2.2.3	<i>Gradient Boosting dan XGBoost</i>	20
2.2.4	<i>Interpretabilitas Model dan SHAP</i>	24
2.2.5	<i>Ambang Curah Hujan Pemicu Banjir</i>	25
2.2.6	<i>Metrik Evaluasi Model Prediksi</i>	26
2.2.7	<i>Algoritma Pembandingan (Baseline Models)</i>	29
2.3	VOSVIEWER	34
2.4	ANALISIS DIAGRAM <i>FISHBONE</i>	35
2.5	HIPOTESIS PENELITIAN.....	37
BAB III	METODOLOGI PENELITIAN	39
3.1	KERANGKA KERJA PENELITIAN.....	39
3.1.1	<i>Prosedur Kerangka Penelitian</i>	41
3.2	LOKASI PENELITIAN.....	43
3.2.1	<i>Gambaran Umum Kota Bandung</i>	43
3.2.2	<i>Peta Lokasi Penelitian</i>	43
3.2.3	<i>Waktu Penelitian</i>	45
3.3	RANCANGAN PENELITIAN	46
3.3.1	<i>Identifikasi dan Pengumpulan Data</i>	46
3.3.2	<i>Pra-Pemrosesan dan Rekayasa Fitur</i>	48
3.3.3	<i>Pengembangan dan Evaluasi Model</i>	51
3.4	DATA DAN SUMBER DATA.....	53
3.4.1	<i>Jenis Data</i>	53
3.4.2	<i>Sumber Data</i>	53
3.4.3	<i>Struktur dan Contoh Dataset</i>	54
3.4.4	<i>Kesesuaian Data dengan Tujuan Penelitian</i>	55
3.5	ANALISIS DATA EKSPLORATIF (EXPLORATORY DATA ANALYSIS / EDA)..	55
3.5.1	<i>Analisis Statistik Deskriptif</i>	56

3.5.2	<i>Analisis Distribusi Data</i>	57
3.5.3	<i>Analisis Korelasi Antarvariabel</i>	58
3.5.4	<i>Identifikasi Missing Values dan Outlier</i>	59
3.5.5	<i>Analisis Korelasi Antarvariabel Meteorologis</i>	60
3.5.6	<i>Ringkasan Temuan EDA dan Implikasinya</i>	62
3.6	PRA-PEMROSESAN DATA.....	62
3.6.1	<i>Data Cleaning</i>	62
3.6.2	<i>Penanganan Missing Values dengan Interpolasi Berbasis Waktu</i>	63
3.6.3	<i>Penanganan Outlier</i>	63
3.6.4	<i>Normalisasi Data dengan Min–Max Scaler</i>	64
3.6.5	<i>Transformasi Data Kategorikal dengan One-Hot Encoding</i>	64
3.6.6	<i>Ringkasan Hasil Pra-Pemrosesan Data</i>	65
3.7	REKAYASA FITUR (FEATURE ENGINEERING).....	65
3.7.1	<i>Pembentukan Fitur Lag (Time-Lagged Features)</i>	66
3.7.2	<i>Pembentukan Fitur Keterlambatan Waktu (Lagged Features)</i>	66
3.7.3	<i>Pembentukan Fitur Curah Hujan Kumulatif</i>	67
3.7.4	<i>Pembentukan Fitur Statistik Temporal (Rolling Statistics)</i>	67
3.7.5	<i>Integrasi Fitur Meteorologis dan Kesiapan Data untuk Pemodelan</i>	68
3.8	TEKNIK PEMODELAN.....	68
3.8.1	<i>Pembagian Data Latih dan Data Uji</i>	68
3.8.2	<i>Pembangunan Model Utama Menggunakan XGBoost</i>	69
3.8.3	<i>Proses Pelatihan dan Penyelelan Parameter</i>	69
3.8.4	<i>Model Pembandingan (Baseline Models)</i>	69
3.8.5	<i>Kesiapan Model untuk Tahap Evaluasi dan Interpretasi</i>	71
3.9	VALIDASI DAN REPLIKASI.....	71
3.9.1	<i>Validasi Model</i>	71
3.9.2	<i>Replikasi Eksperimen</i>	72
3.9.3	<i>Stabilitas Hasil</i>	73
3.9.4	<i>Generalisasi Model terhadap Data Baru</i>	73
3.10	IMPLEMENTASI ALUR PENELITIAN DALAM DASHBOARD PREDIKSI CURAH HUJAN	74

3.10.1	<i>Tahap Input Data (Raw Data Upload)</i>	75
3.10.2	<i>Tahap Pembersihan Data (Data Cleaning)</i>	75
3.10.3	<i>Tahap Eksplorasi Data (Exploratory Data Analysis / EDA)</i>	75
3.10.4	<i>Tahap Persiapan Data dan Feature Engineering</i>	76
3.10.5	<i>Tahap Pembagian Data dan Pelatihan Model</i>	76
3.10.6	<i>Tahap Evaluasi Model</i>	76
3.10.7	<i>Tahap Prediksi Curah Hujan</i>	76
3.11	ALAT DAN PERANGKAT LUNAK	77
3.11.1	<i>Bahasa Pemrograman Python</i>	77
3.11.2	<i>Google Colab sebagai Lingkungan Komputasi</i>	77
3.11.3	<i>Library dan Framework Pendukung</i>	78
3.11.4	<i>Tools Visualisasi</i>	78
BAB IV HASIL DAN PEMBAHASAN		79
4.1	HASIL EKSPLORASI DATA DAN PRA-PEMROSESAN	79
4.1.1	<i>Karakteristik Dataset Awal</i>	79
4.1.2	<i>Pola Curah Hujan dan Variabel Meteorologis</i>	80
4.1.3	<i>Hasil Pra-Pemrosesan Data</i>	82
4.2	HASIL REKAYASA FITUR DAN PERSIAPAN DATA	84
4.2.1	<i>Fitur Input yang Digunakan dalam Model</i>	84
4.2.2	<i>Dampak Feature Engineering terhadap Data</i>	87
4.3	HASIL PEMBANGUNAN DAN PELATIHAN MODEL XGBOOST.....	90
4.3.1	<i>Konfigurasi Model dan Hasil Tuning Hiperparameter</i>	90
4.3.2	<i>Output Pelatihan Model</i>	91
4.4	EVALUASI KINERJA MODEL PREDIKSI	92
4.4.1	<i>Evaluasi Statistik Model</i>	93
4.4.2	<i>Evaluasi Visual Prediksi terhadap Data Aktual</i>	96
4.4.3	<i>Stabilitas dan Generalisasi Model</i>	97
4.5	ANALISIS INTERPRETABILITAS MODEL MENGGUNAKAN SHAP	98
4.5.1	<i>Identifikasi Fitur Dominan Secara Global</i>	98
4.5.2	<i>Analisis Kontribusi Lokal Variabel Meteorologis</i>	99

4.6	ANALISIS AMBANG CURAH HUJAN PEMICU BANJIR	100
4.6.1	<i>Identifikasi Ambang Curah Hujan Berdasarkan Model Prediksi</i>	<i>100</i>
4.6.2	<i>Keterkaitan Ambang Curah Hujan dengan Variabel Dominan</i>	<i>101</i>
4.7	ANALISIS PERBANDINGAN DENGAN MODEL PEMBANDING	102
4.7.1	<i>Perbandingan Metrik Evaluasi Antar Model</i>	<i>102</i>
4.7.2	<i>Perbandingan Pola Prediksi Model</i>	<i>103</i>
4.7.3	<i>Kesimpulan Analisis Perbandingan</i>	<i>103</i>
4.8	IMPLEMENTASI SISTEM DASHBOARD PREDIKSI CURAH HUJAN	104
4.8.1	<i>Tampilan dan Fitur Utama Dashboard</i>	<i>104</i>
4.8.2	<i>Alur Kerja Sistem Prediksi Curah Hujan</i>	<i>106</i>
4.9	PEMBAHASAN HASIL PENELITIAN	107
4.9.1	<i>Implikasi Kinerja Model XGBoost</i>	<i>107</i>
4.9.2	<i>Relevansi Analisis SHAP terhadap Fenomena Meteorologis</i>	<i>108</i>
4.9.3	<i>Keterkaitan Temuan dengan Literatur Terdahulu</i>	<i>108</i>
4.9.4	<i>Keterbatasan Penelitian</i>	<i>109</i>
BAB V PENUTUP		110
5.1	KESIMPULAN	110
5.2	SARAN	111
DAFTAR PUSTAKA		113
LAMPIRAN		120

DAFTAR TABEL

Tabel II.1 Literature Penelitian Terdahulu	8
Tabel II.2 Perbandingan Studi Interpretabilitas dan Penentuan Ambang	15
Tabel III.1 Sampel Data Meteorologis BMKG	54
Tabel IV.1 Tampilan awal dataset meteorologis harian Kota Bandung.	79
Tabel IV.2 Kondisi data sesudah pra-pemrosesan.....	83



DAFTAR GAMBAR

Gambar II.1 Arsitektur XGBoost	22
Gambar II.2 Arsitektur Regresi Linier	30
Gambar II.3 Arsitektur Random Forest.....	31
Gambar II.4 Arsitektur LSTM.....	33
Gambar II.5 VOSviewer	35
Gambar II.6 Diagram Fishbone.....	36
Gambar III.1 Diagram Flowchart Kerangka Kerja Penelitian	40
Gambar III.2 Peta Lokasi Penelitian Kota Bandung.....	44
Gambar III.3 Diagram Tahap Identifikasi dan Pengumpulan Data.....	48
Gambar III.4 Diagram Tahap Pra-Pemrosesan dan Rekayasa Fitur.....	50
Gambar III.5 Diagram Tahap Pengembangan dan Evaluasi Model.....	52
Gambar III.6 Histogram Distribusi Curah Hujan Harian dengan Estimasi Densitas Kernel (KDE).....	57
Gambar III.7 Matriks Korelasi Spearman Lengkap untuk Deteksi Multikolinearitas	58
Gambar III.8 Visualisasi Boxplot untuk Identifikasi Outlier Alami pada Data Curah Hujan.....	59
Gambar III.9 Plot ACF dan PACF untuk Identifikasi <i>Lag</i> Optimal.....	61
Gambar III.10 Alur Workflow Dashboard Prediksi Curah Hujan.....	74
Gambar IV.1 Deret waktu curah hujan harian Kota Bandung.	81
Gambar IV.2 Distribusi curah hujan harian.....	81
Gambar IV.3 Korelasi awal antara curah hujan dan variabel meteorologis.	82
Gambar IV.4 Hasil penanganan nilai hilang pada data meteorologis.	83
Gambar IV.5 Visualisasi Feature Importance / SHAP Feature Importance Model.....	86
Gambar IV.6 Distribusi Curah Hujan Harian (RR).....	87
Gambar IV.7 Visualisasi SHAP Summary Plot Model XGBoost	89
Gambar IV.8 Visualisasi Pembagian Data Latih dan Data Uji (Train–Test Split) ..	91
Gambar IV.9 Perbandingan Curah Hujan Aktual dan Prediksi Model XGBoost ..	92

Gambar IV.10 Perbandingan Nilai R^2 Antar Model.....	94
Gambar IV.11 Perbandingan Nilai RMSE Antar Model.....	95
Gambar IV.12 Perbandingan Curah Hujan Aktual dan Prediksi Model.....	96
Gambar IV.13 Visualisasi Batch Prediction Model.....	97
Gambar IV.14 Tampilan Utama Dashboard Prediksi Curah Hujan Berbasis Streamlit.....	105
Gambar IV.15 Alur Kerja Sistem Dashboard Prediksi Curah Hujan	106

