

MACHINE LEARNING-BASED NOWCASTING OF EXTREME RAINFALL OVER ACEH FROM ERA5 REANALYSIS

Muhammad Raza Adzani
2208107010066

Muhammad Siddiq
250820701100001

Kategori Proyek: Application Project - Physical Sciences /
Machine Learning

CONTENTS

MOTIVASI & MASALAH

**TUJUAN & MANFAAT
PENELITIAN**

DEMO SISTEM

DATA & METODE

EKSPERIMEN & HASIL

MOTIVASI: TANTANGAN PREDIKSI HUJAN EKSTREM DI ACEH



Kerentanan Wilayah

- Hujan Monsun & Topografi Kompleks
- Perubahan Tutupan Lahan
- Risiko Tinggi Banjir Bandang

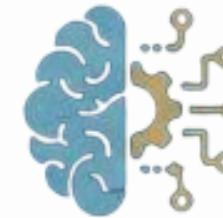
Keterbatasan Sistem Saat Ini

- Radar Terbatas & Gauge Jarang
- Model NWP Sering Bias
- Peringatan Dini Masih Sulit

Kebutuhan Mendesak

- Prediksi Lokal 3-Jam Akurat
- Indikator Risiko yang Handal
- Mengurangi Dampak & Korban Jiwa

ERA5 & MACHINE LEARNING: PENDEKATAN BERBASIS DATA UNTUK NOWCASTING



ERA5: Data Global untuk Aksi Lokal

- Sumber Data: ECMWF Reanalysis v5 (ERA5)
- Resolusi Temporal: Hourly (Per Jam)
- Resolusi Spasial: 0.25° ($\sim 25\text{km}$)
- Periode Data: 2020 - Sekarang (Near Real-Time)

Machine Learning sebagai Ekstraktor Indikator

- Ekstraksi Pola Kompleks: Mengolah multi-variabel atmosfer.
- Tanpa Infrastruktur Fisik: Alternatif radar & gauge yang mahal.
- Indikator Risiko Lokal: Menghasilkan sinyal peringatan dini.
- Model Prediktif: Melatih data historis untuk prediksi masa depan.

Target Nowcasting & Peringatan Dini

- Target: Klasifikasi Biner Hujan Ekstrem.
- Horizon Waktu: 3 Jam ke Depan.
- Domain: Setiap Sel Grid 0.25° di Aceh.
- Tujuan: Early Warning Cepat & Terukur.

FORMULIR MASALAH: NOWCASTING HUJAN EKSTREM DI ACEH

Input



Data ERA5 Single-Level
(2020-2024)



Domain Aceh:
5.0-6.0°N, 95.0-96.0°E
(Grid 5x5, Resolusi 0.25°)



Langkah Waktu: 3 Jam
(3-hourly time step)

Output



Kelas Biner:
Hujan Ekstrem 3 Jam ke Depan vs. Tidak



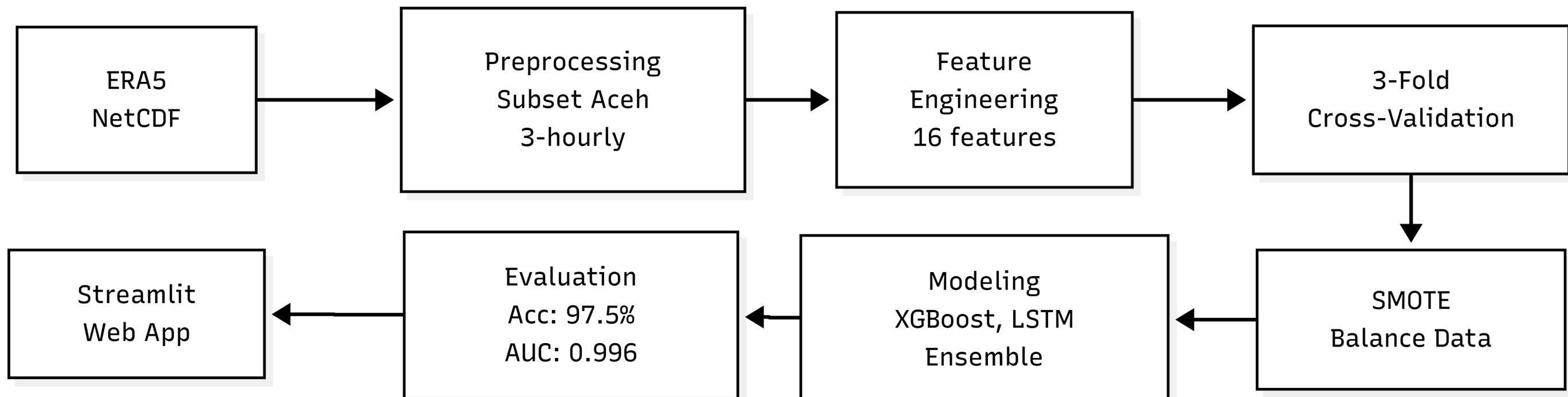
Definisi Ekstrem:
Top 5% Akumulasi Curah Hujan 3 Jam
(tp_next)



Goal

- Maksimalkan Recall Kelas Ekstrem
- Pertahankan False Alarm pada Tingkat Wajar untuk Peringatan Dini

DIAGRAM ALIR



TINJAUAN DATASET

Ringkasan Dataset



Periode: 2020-2024 (5 tahun)

Training: [██████████] 75%

Testing: [███] 25%

Samples: ~273,500 train | ~91,900 test



Resolusi Spasial: $0.25^\circ \times 0.25^\circ$
(5x5 sel grid)

Training: [██████████] 75%

Testing: [███] 25%

Samples: ~273,500 train | ~91,900 test



Resolusi Temporal: 3-jam (3-hourly)

Training: [██████████] 75%

Testing: [███] 25%

Samples: ~273,500 train | ~91,900 test



Total Sampel: 365.400

- ✓ Chronological splits
- ✓ SMOTE on training data
- ✓ No temporal leakage
- ✓ Average performance



Proposi Ekstrem: ~5% (Kelas Tidak Seimbang)

REKAYASA FITUR



Koordinat Spasial

- Latitude (lat)
- Longitude (lon)
- Resolusi grid: 0.25°



Meteorologi & Kondisi Saat Ini

- Total Precipitation (tp)
- Runoff (ro)
- Komponen Angin (u10, v10)
- Suhu 2m (t2m)
- Kelembapan Tanah (swvl1)
- Kecepatan Angin (wind_speed)



Temporal (Siklus Waktu)

- Jam Sin/Cos (hour_sin, hour_cos)
- Hari dalam Tahun Sin/Cos (doy_sin, doy_cos)
- Menangkap siklus harian & musiman

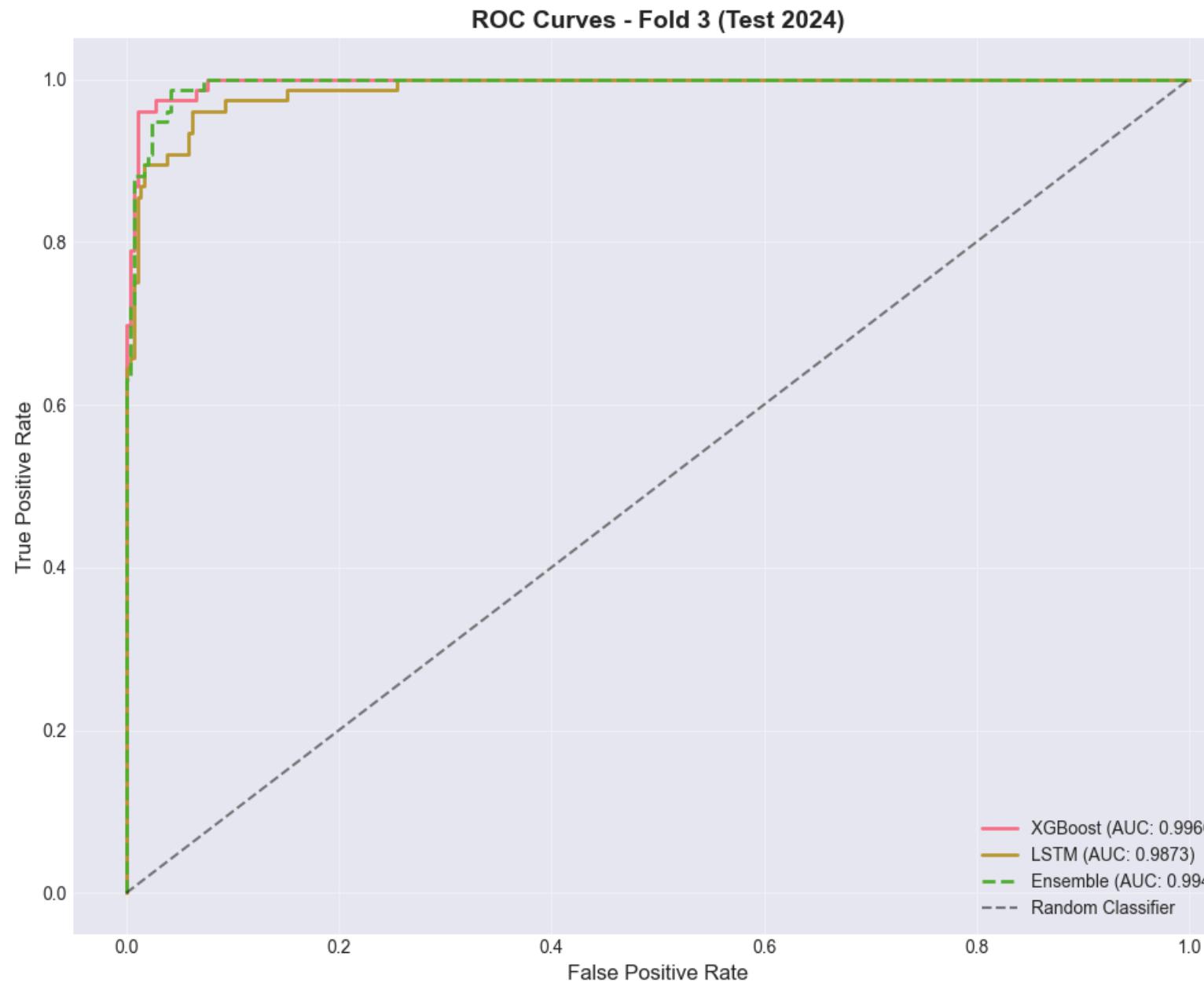


Histori Hujan

- tp_lag1 (3 jam lalu)
- tp_lag2 (6 jam lalu)
- tp_roll3_mean (rata-rata 0-6 jam)
- Indikator Intensitas & tren



ANALISIS KURVA ROC & KINERJA MODEL (VALIDASI)



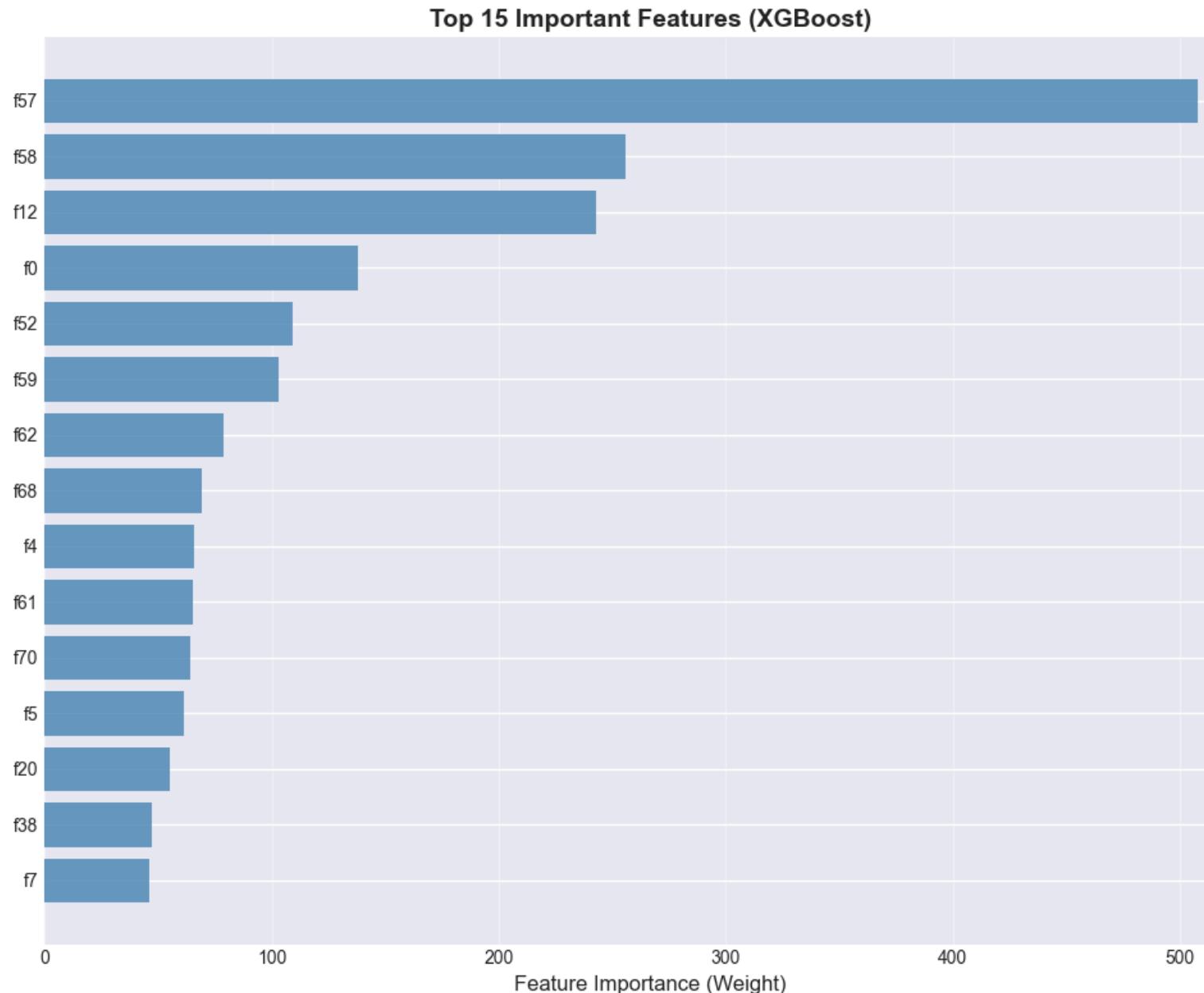
OC Curve Analysis (Fold 3):

- XGBoost: AUC = 0.996 (Exceptional)
- LSTM: AUC = 0.987
- Ensemble: AUC = 0.995

XGBoost menunjukkan diskriminasi hampir sempurna dengan recall 96.1% dan precision 92.4%

Performa superior menjadikan XGBoost pilihan terbaik untuk early warning system

ANALISIS FEATURE IMPORTANCE MODEL XGBOOST



Fitur Paling Berpengaruh:

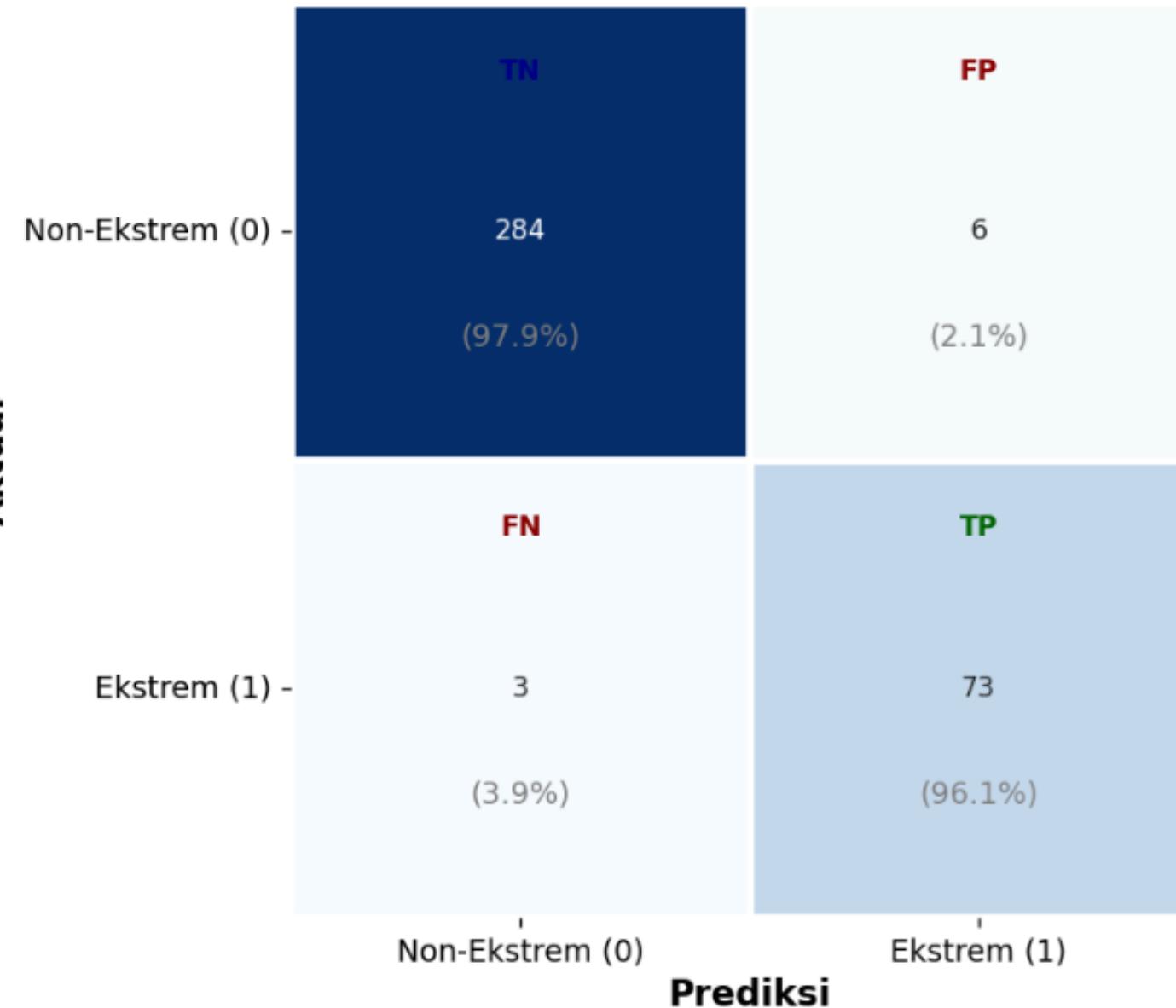
1. Total Precipitation (tp) Intensitas hujan saat ini
2. Rolling Mean (tp_roll3) Akumulasi 3 jam terakhir
3. Lag Features (tp_lag1/2) Persistensi hujan

Variabel Pendukung:

- Soil moisture (swvl1)
- Temperature (t2m)
- Wind speed dan direction

Kesimpulan: Model fokus pada riwayat hujan recent dan kondisi atmosfer untuk prediksi akurat.

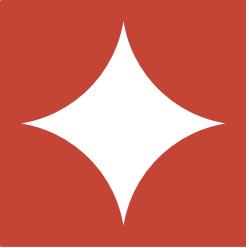
CONFUSION MATRIX



Deteksi 96.05%: Hanya 3 dari 76 kejadian ekstrem terlewat (FN).
Model sangat sensitif mendeteksi hampir semua event berbahaya untuk early warning.

Precision 92.41%: False alarm hanya 6 dari 79 alarm (FP). Tingkat kepercayaan tinggi menjaga kredibilitas sistem peringatan.

Akurasi 97.54%: Overall performance sangat baik dengan balanced detection antara ekstrem dan non-ekstrem, cocok untuk deployment operasional.



KESIMPULAN



1. Model Performance Excellent

Ensemble XGBoost-LSTM mencapai akurasi 97.5% dengan recall 96.1% untuk deteksi ekstrem dan precision 92.4% untuk minimalisasi false alarm.

2. Prediksi Real-Time Reliable

Demo menunjukkan confidence 99.92% dengan konsensus kuat kedua model, membuktikan sistem robust untuk prediksi operasional 3 jam ke depan.

3. Early Warning Effective

Lead time 3 jam cukup untuk preparedness. Threshold optimal (0.495) menghasilkan balance ideal antara sensitivitas dan kredibilitas warning.

4. Deployment-Ready

Model teruji dengan data ERA5 2020-2024 spesifik Aceh, siap diintegrasikan dengan BMKG untuk pilot project mitigasi banjir.



THANK YOU!

