

Prediksi Curah Hujan Harian di India Menggunakan Long Short-Term Memory (LSTM): Pendekatan Multi-Domain untuk Mitigasi Bencana, Perencanaan Pertanian, dan Manajemen Sumber Daya Air

Progress Report Penelitian Deep Learning

Peneliti: Fikri Armia Fahmi **Program Studi:** Informatika **Pembimbing:** Dr. Ida Nurhaida, M.T

SLIDE 1: LATAR BELAKANG

Mengapa Prediksi Curah Hujan Penting?

- **Sektor Pertanian:** Menyumbang 18% PDB India dengan populasi 1.4 miliar jiwa
- **Variabilitas Tinggi:** Dipengaruhi sistem monsun, ENSO, dan perubahan iklim
- **Tantangan:** Perencanaan sumber daya air, manajemen pertanian, mitigasi bencana

Solusi: Deep Learning dengan LSTM

- Menangkap dependensi temporal jangka panjang
 - Performa superior dibanding metode tradisional (ARIMA, SVR)
 - Peningkatan akurasi 15-25% dalam RMSE dan MAE
-

SLIDE 2: RUMUSAN MASALAH

Pertanyaan Penelitian Utama:

1. Bagaimana merancang arsitektur LSTM optimal untuk prediksi curah hujan harian?
 2. Bagaimana pengaruh hiperparameter terhadap performa model?
 3. Bagaimana performa LSTM dibandingkan konfigurasi alternatif?
 4. Bagaimana mengintegrasikan faktor eksternal (ENSO, temperatur)?
 5. Bagaimana aplikasi praktis untuk pertanian, peringatan dini, dan manajemen air?
-

SLIDE 3: TUJUAN PENELITIAN

Target Utama:

1. **Implementasi Model:** LSTM untuk prediksi curah hujan harian (dataset 1901-2015, 36 subdivisi)
2. **Eksperimen Sistematis:** 13 konfigurasi model berbeda
3. **Target Performa:** $R^2 \geq 0.97$, $RMSE \leq 0.55$ mm, $MAE \leq 0.15$ mm
4. **Integrasi Fitur:** Faktor eksternal untuk meningkatkan akurasi
5. **Aplikasi Praktis:** Sistem peringatan dini, perencanaan pertanian, manajemen air

6. **Dokumentasi:** Kode reproducible untuk peneliti lain

SLIDE 4: DATASET

Karakteristik Dataset

Sumber: Rainfall in India (Rajanand) via Kaggle

Periode: 1901-2015 (115 tahun)

Cakupan: 36 subdivisi meteorologi di seluruh India

Statistik Dataset:

- **Data Original:** 4,116 observasi bulanan
- **Setelah Transformasi:** 1,503,342 observasi harian
- **Training Sequences:** 1,201,809 (80%)
- **Test Sequences:** 300,453 (20%)
- **Sequence Length:** 30 hari (sliding window)

Preprocessing:

- Transformasi bulanan → harian (distribusi uniform)
- Normalisasi per-subdivisi (MinMaxScaler)
- Train-test split kronologis

[Gambar: Peta India dengan 36 subdivisi meteorologi]

SLIDE 5: ARSITEKTUR MODEL LSTM

Arsitektur Baseline:

```
Input Layer (30 timesteps, 1 feature)
  ↓
LSTM Layer 1 (64 units, return_sequences=True)
  ↓
Dropout (0.2)
  ↓
LSTM Layer 2 (32 units)
  ↓
Dense Layer (16 units, ReLU)
  ↓
Output Layer (1 unit, Linear)
```

Total Parameters: 29,857

Variasi Arsitektur:

- **Simple:** 32-16-8 units

- **Baseline:** 64-32-16 units
- **Deep:** 128-64-32 units

[Gambar: Diagram arsitektur LSTM]

SLIDE 6: DESAIN EKSPERIMEN

13 Konfigurasi Model yang Diuji:

Kategori	Variasi
Learning Rate	0.0001, 0.001, 0.01
Batch Size	16, 32, 64, 128
Dropout Rate	0.0, 0.2, 0.5
Optimizer	Adam, RMSprop, SGD
Arsitektur	Simple, Baseline, Deep
Sequence Length	15, 30, 60 hari

Metrik Evaluasi:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- R² Score (Coefficient of Determination)

SLIDE 7: HASIL EKSPERIMEN - TOP 5 MODEL

Perbandingan Performa:

Rank	Model	RMSE (mm)	MAE (mm)	R ² Score
1	No Dropout	0.5022	0.1070	0.9746
2	Low LR (0.0001)	0.5292	0.1126	0.9718
3	High LR (0.01)	0.6306	0.2638	0.9599
4	Large Batch (64)	0.6330	0.2512	0.9596
5	Simple Arch	0.6480	0.2695	0.9577

Model Terbaik: **No Dropout**

- **Peningkatan 24%** dibanding baseline
- R² = 0.9746 (97.46% varians dijelaskan)
- RMSE = 0.5022 mm

[Gambar: Bar chart perbandingan RMSE 13 model]

SLIDE 8: ANALISIS HIPERPARAMETER

Temuan Kunci:

1. Dropout Regularization

- ☒ **No Dropout (0.0)**: Performa terbaik
- ☐ **Baseline (0.2)**: Performa moderat
- ☒ **High Dropout (0.5)**: Degradasi parah (RMSE 1.5867 mm)

2. Learning Rate

- **Low (0.0001)**: Konvergensi stabil, performa terbaik
- **Baseline (0.001)**: Keseimbangan baik
- **High (0.01)**: Konvergensi cepat, lebih volatil

3. Batch Size

- **Large (64)**: Optimal untuk efisiensi dan performa
- **Small (16)**: Gradient berisik, performa menurun

4. Optimizer

- **Adam & RMSprop**: Performa superior
- **SGD**: Performa tertinggal (RMSE 0.7337 mm)

[Gambar: Line plot training curves untuk berbagai learning rate]

SLIDE 9: ANALISIS PREDIKSI

Visualisasi Performa Model Terbaik:

Scatter Plot: Prediksi vs Aktual

- Clustering ketat di sekitar garis diagonal
- $R^2 = 0.9746$
- Sedikit underprediksi pada nilai ekstrem

Time Series: 500 Hari Sampel

- Model menangkap pola musiman dengan baik
- Tracking akurat pada periode transisi
- Deviasi pada kejadian ekstrem

Distribusi Residual

- Mean: -0.003 mm (mendekati nol)

- Std Dev: 0.502 mm
- Distribusi sekitar normal

[Gambar: Scatter plot prediksi vs aktual] [Gambar: Time series plot 500 hari]

SLIDE 10: PERFORMA GEOGRAFIS

Analisis Per-Subdivisi:

- **Konsistensi Tinggi:** $R^2 > 0.95$ di mayoritas subdivisi
- **Performa Terbaik:** Wilayah dengan variabilitas moderat
- **Tantangan:** Wilayah pesisir dan topografi kompleks

Pola Geografis:

- Wilayah timur laut: Curah hujan tinggi, variabilitas tinggi
- Wilayah barat: Curah hujan rendah, lebih stabil
- Coastal regions: Pengaruh angin laut, performa sedikit lebih rendah

[Gambar: Choropleth map India dengan R^2 score per subdivisi]

SLIDE 11: APLIKASI MULTI-DOMAIN

1. Transformasi ke Klasifikasi (5 Kategori)

- Very Low (0-1 mm), Low (1-5 mm), Medium (5-10 mm)
- High (10-20 mm), Very High (>20 mm)
- **Akurasi:** Random Forest classifier untuk sistem peringatan diskrit

2. Sistem Peringatan Dini

- **Banjir:** Alert jika curah hujan 7 hari > 100-150 mm
- **Kekeringan:** Alert jika curah hujan 30 hari < 20-30% normal
- **Lead Time:** 3-7 hari (banjir), 2-4 minggu (kekeringan)

3. Manajemen Sumber Daya Air

- Optimasi operasi waduk
- Perencanaan alokasi air
- **Potensi:** Peningkatan 10-15% efisiensi penggunaan air

[Gambar: Distribusi kategori curah hujan - 5 kelas] [Gambar: Confusion matrix klasifikasi]

SLIDE 12: ANALISIS MUSIMAN & CLUSTERING

Pola Musiman:

- **Monsun (Jun-Sep):** Curah hujan tertinggi
- **Musim Kering (Dec-Feb):** Curah hujan terendah
- **Variasi Regional:** Timur laut memiliki pola bimodal

Clustering 641 Distrik (K-Means):

Cluster	Karakteristik	Curah Hujan Tahunan	Wilayah
0	Kering	< 800 mm	Rajasthan, Gujarat barat
1	Moderat	800-1500 mm	India tengah & utara
2	Basah	1500-2500 mm	Pantai barat, India timur
3	Sangat Basah	> 2500 mm	Timur laut, Western Ghats

[Gambar: Heatmap curah hujan per wilayah dan bulan] [Gambar: PCA scatter plot 4 cluster distrik]

SLIDE 13: DETEKSI ANOMALI & TREND JANGKA PANJANG

Deteksi Anomali Curah Hujan:

- **45 distrik (7%)** dengan curah hujan anomali terdeteksi
- **Metode:** Z-score dan IQR analysis
- **Lokasi Anomali:**
 - Curah hujan sangat tinggi: Meghalaya, Assam
 - Curah hujan sangat rendah: Rajasthan barat

Trend Historis 1901-2015:

- **Penurunan gradual:** -0.12 mm/tahun ($p < 0.05$)
- **Variabilitas meningkat** dalam dekade terakhir
- **Indikasi:** Potensi dampak perubahan iklim

Proyeksi Masa Depan:

- Penurunan berkelanjutan dengan ketidakpastian meningkat
- Penting untuk adaptasi dan mitigasi

[Gambar: Box plot deteksi anomali per state] [Gambar: Time series trend 1901-2015 dengan proyeksi]

SLIDE 14: KETERBATASAN PENELITIAN

Keterbatasan Utama:

1. **Transformasi Data:** Distribusi uniform tidak menangkap variabilitas harian aktual
2. **Underprediksi Ekstrem:** Model cenderung underprediksi kejadian curah hujan ekstrem
3. **Kendala Komputasi:** Training pada CPU, eksperimen terbatas

4. **Fitur Terbatas:** Hanya menggunakan data curah hujan historis
5. **Interpretabilitas:** Sifat black-box LSTM
6. **Validasi Stakeholder:** Belum divalidasi dengan pengguna akhir

Solusi yang Direncanakan:

- Ensemble methods
 - Weighted loss functions untuk kejadian ekstrem
 - Integrasi fitur eksternal (temperatur, kelembapan, ENSO)
 - Probabilistic forecasting untuk kuantifikasi ketidakpastian
-





SLIDE 15: RENCANA PENGEMBANGAN

Roadmap Menuju UAS:

Technical Improvements:

- ☒ Ensemble methods (top-5 models)
- ☒ Attention mechanisms
- ☒ Bidirectional LSTM
- ☒ Probabilistic predictions (Monte Carlo Dropout)
- ☒ Multi-task learning (regresi + klasifikasi)

Application Development:

-  Web application (Streamlit/Flask)
-  Mobile app untuk petani
-  Dashboard untuk pembuat kebijakan
-  API untuk integrasi sistem

Target Capaian:

- **Performa:** $R^2 \geq 0.98$, $RMSE \leq 0.45$ mm
 - **Aplikasi:** 100+ pengguna uji, 2+ kemitraan
 - **Publikasi:** 1+ konferensi/jurnal peer-reviewed
-

SLIDE 16: KONTRIBUSI PENELITIAN

Kontribusi Akademis:

1. **Metodologis:** Studi hiperparameter komprehensif untuk prediksi curah hujan
2. **Teknis:** Mencapai $R^2 > 0.97$ dengan arsitektur efisien
3. **Ilmiah:** Analisis 115 tahun pola curah hujan India

Kontribusi Praktis:

1. **Pertanian:** Framework dukungan keputusan untuk petani
2. **Bencana:** Sistem peringatan dini banjir dan kekeringan

3. **Air**: Optimasi manajemen sumber daya air

Impact Potensial:

- 🍽️ Ketahanan pangan
 - 🚒 Kesiapsiagaan bencana
 - 💧 Penggunaan air berkelanjutan
 - 🌍 Adaptasi perubahan iklim
-

SLIDE 17: KESIMPULAN

Pencapaian Utama:

- ✓ **Model LSTM berhasil dikembangkan** dengan performa terdepan ($R^2 = 0.9746$)
- ✓ **13 eksperimen komprehensif** mengidentifikasi konfigurasi optimal
- ✓ **Temuan kunci**: Dataset besar tidak memerlukan dropout agresif
- ✓ **Framework aplikasi** untuk pertanian, peringatan dini, dan manajemen air
- ✓ **Analisis mendalam** 115 tahun data curah hujan India

Wawasan Penting:

- Learning rate rendah → optimasi stabil
- Batch size besar → efisiensi tinggi
- Optimizer adaptif (Adam) → performa superior
- Kapasitas moderat → keseimbangan optimal

Next Steps:

Pengembangan ensemble methods, integrasi fitur eksternal, deployment aplikasi, dan validasi dengan stakeholder.

SLIDE 18: TERIMA KASIH

Kontak & Informasi:

Email: [email mahasiswa]

GitHub: [repository link]

LinkedIn: [profile link]

Pertanyaan & Diskusi

"Deep learning bukan hanya tentang akurasi model, tetapi bagaimana teknologi dapat memberikan dampak nyata untuk masyarakat dalam menghadapi tantangan iklim."
