

HALAMAN MOTTO

"Those who cannot remember the past are condemned to repeat it."

– George Santayana

"We don't read and write poetry because it's cute. we read and write poetry because we are members of the human race. and the human race is filled with passion. Medicine, law, business, engineering, these are all noble pursuits, and necessary to sustain life. but poetry, beauty, romance, love, these are what we stay alive for".

– John Keating, Dead Poets Society

"Misère et grandeur de ce monde : il ne propose pas de vérités, mais des objets d'amour." –Albert Camus

"Tujuan pendidikan itu untuk mempertajam kecerdasan, memperkuat kemauan serta memperhalus perasaan." – Tan Malaka

"Rasa penasaran itu bisa melukaimu, tapi kalau tidak seperti itu maka aku tak pantas menjadi ilmuwan, lagi pula menjadi ilmuwan adalah hal yang paling romantis"

– Kurisu Makise-

"Mengingat sesuatu yang tidak bisa dilakukan orang lain adalah hal yang menyakitkan. Kamu tidak dapat berbicara dengan siapa pun tentang hal itu. Tidak ada yang akan mengerti Kamu. Kamu akan sendirian." – Rintaro Okabe

HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan untuk:

1. Allah Yang Maha Kuasa, atas segala ciptaan-Mu yang indah dalam wujud atmosfer dan petunjuk-Nya dalam setiap baris kode dan analisis yang penulis susun.
2. Kedua Orang Tua Tercinta, yang cintanya tidak terbatas oleh ruang dan waktu, serta selalu menjadi pemberi semangat utama dalam penyelesaian studi ini.
3. Almamater Universitas Putra Bangsa, tempat penulis menempa diri menjadi seorang peneliti Sains Data yang lebih baik.
4. Teman-teman magang bersama di BPS Kebumen dari sesama almamater Universitas Putra Bangsa, STIS, dan Amikom Yogyakarta.
5. Sahabat-sahabat di Discord, dari Politeknik Negeri Subang Adi Ivani Yusuf, Adrian Ramadhan, Tegar Kusuma, lalu adik tingkat saya satu almamater Bagus Alfian Yusuf dan Indah Ruwahna Anugraheni. Terima kasih telah menjadi saksi bisu malam-malam penuh kopi dan error model, serta selalu memberikan motivasi meskipun kita terhalang jarak antar ruang dan waktu.
6. Engkau yang pernah ku cintai sedalam palung Mariana dan setinggi lapisan eksosfer, namun cintanya habis seperti meteor yang mengarungi dalamnya atmosfer, ku sampaikan rasa terima kasihku yang sebesar-besarnya...

ABSTRAK

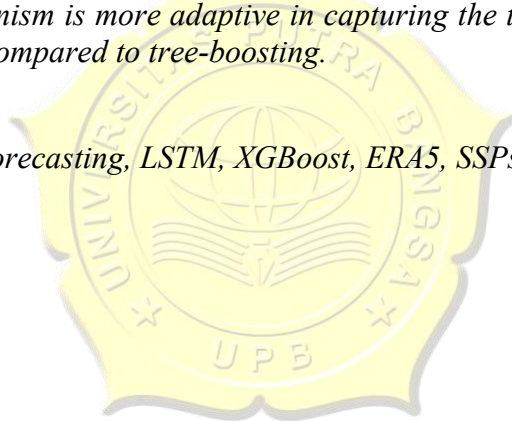
Prediksi curah hujan jangka pendek skala lokal di wilayah Desa Jerukagung, Klirong, Kebumen memiliki kompleksitas tinggi akibat sifat atmosfer yang non-linear dan fluktuatif. Penelitian ini membandingkan performa algoritma *machine learning* (XGBoost) dan *deep learning* (LSTM) menggunakan pendekatan *Two-Stage Modeling* yang memadukan *pre-training* data reanalisis ERA5 dan satelit dengan *fine-tuning* data observasi stasiun *Automatic Weather Station* (AWS) berbasis IoT. Evaluasi dari *satellite-based precipitation products* (SPPs) menunjukkan GSMaP memiliki tingkat kesalahan volume paling minimal terhadap observasi AWS dengan nilai MAE 0,525 mm/jam dan RMSE 1,5 mm/jam, sehingga dipilih sebagai substitusi prediktor curah hujan. Pada pengujian pemodelan lokal, arsitektur LSTM *Fine-Tuned* terbukti sebagai model paling optimal dan presisi, mendominasi klasifikasi kejadian hujan (*F1-Score*: 0,618; *ROC-AUC*: 0,867) serta menghasilkan deviasi error regresi volume paling minimal (MAE: 1,530 mm/jam; RMSE: 3,186 mm/jam). Sementara itu, XGBoost *Fine-Tuned* cenderung lebih selektif dengan *Precision* tertinggi (0,695), namun sensitivitas deteksinya menurun (*Recall* 0,452). Keunggulan LSTM mengonfirmasi bahwa mekanisme gerbang memori sekuensial waktu lebih adaptif dalam menangkap kontinuitas temporal presipitasi lokal dibandingkan *tree-boosting*.

Kata Kunci: Prediksi Curah Hujan, LSTM, XGBoost, ERA5, SSPs.

ABSTRACT

Short-term, local-scale rainfall forecasting in the Jerukagung Village area, Klirong, Kebumen, is highly complex due to the nonlinear and fluctuating nature of the atmosphere. This study compares the performance of a machine learning algorithm (XGBoost) and a deep learning algorithm (LSTM) using a Two-Stage Modeling approach that combines pre-training with ERA5 reanalysis and satellite data with fine-tuning using IoT-based Automatic Weather Station (AWS) observation data. An evaluation of satellite-based precipitation products (SPPs) showed that GSMap had the lowest volume error relative to AWS observations, with an MAE of 0.525 mm/hour and an RMSE of 1.5 mm/hour; therefore, it was selected as the substitute precipitation predictor. In local modeling tests, the Fine-Tuned LSTM architecture proved to be the most optimal and precise model, dominating rainfall event classification (F1-Score: 0.618; ROC-AUC: 0.867) and producing the smallest volume regression error (MAE: 1.530 mm/hour; RMSE: 3.186 mm/hour). Meanwhile, the fine-tuned XGBoost model tended to be more selective with the highest precision (0.695), but its detection sensitivity was lower (recall 0.452). The superiority of LSTM confirms that the sequential temporal memory gate mechanism is more adaptive in capturing the temporal continuity of local precipitation compared to tree-boosting.

Keyword: Rainfall forecasting, LSTM, XGBoost, ERA5, SSPs.



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah Subhanahu wa ta'ala Tuhan semesta alam atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penelitian ini. Skripsi ini disusun untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Strata-1 pada Program Studi Sains Data Universitas Putra Bangsa. Dalam penelitian ini, penulis melakukan penelitian perbandingan performa algoritma *machine learning* berbasis XGBoost dengan algoritma *deep learning* LSTM menggunakan data historis reanalisis ERA5, data satelit, dan data observasi lapangan dari sensor AWS lokal berbasis IoT untuk studi kasus wilayah Desa Jerukagung, Klirong, Kebumen.

Dalam proses penyusunan skripsi ini, penulis mendapatkan banyak bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang tulus ini kepada:

1. Kedua orang tua tercinta dan keluarga atas dukungan doa dan moral.
2. Bapak Dr. Gunarso Wiwoho, S.E., M.M., selaku Rektor Universitas Putra Bangsa.
3. Bapak Yulianto, S.Kom., M.Kom., selaku Dekan Fakultas Sains dan Teknologi
4. Bapak Sarjimin, S.Kom., M.Kom, selaku Ketua Program Studi S-1 Sains Data.
5. Bapak Ir. Rokhmatullah Muhammad Ikhsanudin, S.Kom., M.Cs, selaku Dosen Pembimbing yang telah memberikan arahan berharga bagi penulis dalam penulisan skripsi ini.

6. Ibu Lolanda Hamim Annisa, S.Kom, M.Kom selaku Dosen wali yang telah membimbing penulis selama menjadi mahasiswa dalam hal akademik maupun non-akademik selama masa kuliah
7. Segenap Dosen, Laboran dan Pegawai Fakultas Sains dan Teknologi khususnya Program Studi Sains Data Universitas Putra Bangsa yang senantiasa memberikan ilmu pengetahuan dan pengalaman kepada penulis.

Semoga dengan ditulisnya dan menyebarluaskan laporan penelitian ini dapat menjadi salah satu usaha dalam meningkatkan pemahaman manusia akan ilmu pengetahuan dan alam semesta.

Penulis menyadari bahwa laporan penelitian ini masih jauh dari kata sempurna. Oleh karena itu, penulis berharap kritik dan saran dari pembaca dan sesama penulis yang bersifat membangun dan dapat menjadikan laporan ini sebagai referensi untuk penyusunan laporan kegiatan penelitian yang sejenis.

Kebumen, 29 Juli 2026

Penulis

Evan Alif Widhyatma

DAFTAR ISI

HALAMAN SAMPUL DEPAN.....	i
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN UJIAN	iv
HALAMAN BEBAS PLAGIARISME	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
ABSTRAK	viii
<i>ABSTRACT</i>	ix
KATA PENGANTAR	x
DAFTAR ISI.....	xii
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Penelitian	8
1.3 Tujuan Penelitian	8
1.4 Manfaat Penelitian	9
1.5 Batasan Penelitian.....	10
BAB II KAJIAN PUSTAKA	12
2.1 Tinjauan Teori	12
2.1.1 Meteorologi	12
2.1.2 Suhu udara	13
2.1.3 Titik Embun.....	15
2.1.4 Tekanan Udara.....	16
2.1.5 Angin	17
2.1.6 Curah Hujan.....	18
2.1.7 Pemodelan Prediksi Cuaca	19

2.1.8 Machine Learning.....	22
2.1.9 Long Short-Term Memory	24
2.1.10 Extreme Gradient Boosting	29
2.1.11 Metrik Evaluasi.....	35
2.2 Penelitian Terdahulu.....	38
2.3 Kerangka Pemikiran	48
2.4 Keaslian Penelitian	49
BAB III METODOLOGI PENELITIAN.....	51
3.1 Tahapan Penelitian.....	51
3.2 Metode Penelitian	56
3.3 Alat dan Bahan.....	57
3.3.1 Alat	57
3.3.2 Bahan	60
3.4 Objek dan Subjek Penelitian.....	61
3.4.1 Objek Penelitian	61
3.4.2 Subjek Penelitian	61
3.5 Variabel Penelitian.....	63
3.6 Metode Pengumpulan Data.....	65
3.6.1 Pengumpulan Data Primer.....	65
3.6.2 Pengumpulan Data Sekunder.....	67
3.7 Metode Analisis Data.....	73
3.7.1 Pembagian Dataset	73
3.7.2 Praproses Data	74
3.7.3 Perancangan Model	77
3.7.4 Optimalisasi Model.....	79
3.7.5 Validasi Model	86
3.7.6 Pengujian Model.....	86
BAB IV HASIL DAN PEMBAHASAN	89
4.1 Analisis dan Eksplorasi Data	89
4.2 Pemodelan Model Prediksi LSTM.....	98
4.2.1 Pre-training Model LSTM	99
4.2.2 Fine-tuning Model LSTM	105

4.3 Pemodelan Model Prediksi XGBoost	109
4.3.1 Pre-training Model XGBoost.....	110
4.3.2 Fine-tuning Model XGBoost.....	116
4.4 Perbandingan Performa dan Metrik antar Model.....	120
BAB V PENUTUP.....	127
5.1 Kesimpulan	127
5.2 Saran	128
DAFTAR PUSTAKA	130
LAMPIRAN.....	139



DAFTAR TABEL

Tabel II-1. Penelitian Terdahulu	44
Tabel III-1. Sensor dan Variabel Output.....	58
Tabel III-2. Daftar Library Python.....	58
Tabel III-3. Identifikasi Variabel Penelitian	63
Tabel III-4. Variabel AWS Lokal IoT.....	66
Tabel III-5. Variabel ERA5 Reanalisis	68
Tabel III-6. Variabel IMERG.....	71
Tabel III-7. Variabel GSMap	72
Tabel III-8. Pembagian Data Pre-training.....	73
Tabel III-9. Pembagian Data Fine-tuning	74
Tabel III-10. Nilai Pencarian Parameter Klasifikasi (XGBoost)	80
Tabel III-11. Nilai Pencarian Parameter Regresi (XGBoost)	82
Tabel III-12. Nilai Pencarian Parameter Klasifikasi (LSTM).....	84
Tabel III-13. Nilai Pencarian Parameter Regresi (LSTM).....	85
Tabel IV-1. Perbandingan Korelasi dan <i>Error</i> variabel terhadap AWS	97
Tabel IV-2. Nilai Optimal Hyperparameter Klasifikasi LSTM.....	100
Tabel IV-3. Nilai Optimal Hyperparameter Regresi LSTM	103
Tabel IV-4. Perbandingan Hasil Prediksi Model LSTM	108
Tabel IV-5. Nilai Optimal Hyperparameter Klasifikasi XGboost	110
Tabel IV-6. Nilai Optimal Hyperparameter Regresi XGBoost.....	113
Tabel IV-7. Perbandingan hasil prediksi model XGBoost.....	119
Tabel IV-8. Perbandingan Metrik Klasifikasi Pre-training.....	121
Tabel IV-9. Perbandingan Metrik Regresi Pre-training	122
Tabel IV-10. Perbandingan Klasifikasi Model Fine-tune	123
Tabel IV-11. Perbandingan Regresi Model Fine-tune	125

DAFTAR GAMBAR

Gambar I-1. Tren dan Pola Musiman di Wilayah Jerukagung.....	5
Gambar II-1. Arsitektur LSTM.....	25
Gambar II-2. Arsitektur XGBoost	30
Gambar II-3. Konsep Confusion Matrix	36
Gambar II-4. Diagram Kerangka Pemikiran.....	48
Gambar III-1. Diagram Tahap Penelitian.....	51
Gambar IV-1. Korelasi Pearson data curah hujan.....	90
Gambar IV-2. Korelasi Spearman data curah hujan	91
Gambar IV-3. Perbandingan MAE AWS dengan dataset lain	92
Gambar IV-4. Perbandingan RMSE AWS dengan dataset lain	93
Gambar IV-5. Perbandingan MBE AWS dengan dataset lain	94
Gambar IV-6. Korelasi Pearson Variabel Prediktor	95
Gambar IV-7. Train and Val Loss Klasifikasi LSTM.....	100
Gambar IV-8. Confusion Matriks LSTM Pretrained	101
Gambar IV-9. Train and Val Loss Regresi LSTM.....	103
Gambar IV-10. Perbandingan Observasi dan Prediksi LSTM.....	104
Gambar IV-11. Perbandingan Kurva ROC LSTM	105
Gambar IV-12. Perbandingan PR Model LSTM	106
Gambar IV-13. Perbandingan hasil prediksi model LSTM	108
Gambar IV-14. Train and Val Loss Klasifikasi XGBoost	111
Gambar IV-15. Confusion Matriks XGBoost Pretrained.....	112
Gambar IV-16. Train and Val Loss Regresi XGBoost	114
Gambar IV-17. Perbandingan Observasi dan Prediksi XGBoost	115
Gambar IV-18. Perbandingan Kurva ROC XGBoost	117
Gambar IV-19. Perbandingan PR Model XGBoost.....	118
Gambar IV-20. Perbandingan hasil prediksi model XGBoost.....	119
Gambar IV-21. Grafik Perbandingan Metrik Klasifikasi Pre-training.....	121
Gambar IV-22. Perbandingan Metrik Regresi Pre-training	122
Gambar IV-23. Grafik Perbandingan Klasifikasi Model Fine-tune.....	123
Gambar IV-24. Grafik Perbandingan Metrik Regresi Model Fine-tune	125

DAFTAR LAMPIRAN

Lampiran 1. Sampel Data AWS IoT	141
Lampiran 2. Sample Data GSMap	142
Lampiran 3. Sampel Data IMERG	143
Lampiran 4. Sampel Data ERA5 (Parameter Permukaan).....	144
Lampiran 5. Sampel Data ERA5 (Parameter Dinamika Atmosfer Atas).....	145
Lampiran 6. Sampel Data ERA5 (Parameter Termodinamika)	146
Lampiran 7. Kode GEE data ERA5	147
Lampiran 8. Kode GEE data IMERG	149
Lampiran 9. Kode GEE data GSMap.....	150
Lampiran 10. Kode Model LSTM	151
Lampiran 11. Kode Model XGBoost.....	154
Lampiran 12. Kartu Bimbingan Skripsi.....	157
Lampiran 13. Kartu Tanda Peserta Seminar Proposal	158

