



SKRIPSI

***STACKING ENSEMBLE* BERBASIS OPTIMASI *HYPERPARAMETER* PADA PERAMALAN INDEKS STANDAR PENCEMAR UDARA KOTA JAKARTA**

FAHMI HAMMAM TAQIYUDDIN
NPM 21083010048

DOSEN PEMBIMBING

Aviolla Terza Damaliana, S.Si., M.Stat
Andri Fauzan Adziima, S.Si., M.Si.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

STACKING ENSEMBLE BERBASIS OPTIMASI HYPERPARAMETER PADA PERAMALAN INDEKS STANDAR PENCEMAR UDARA KOTA JAKARTA

Oleh:
FAHMI HAMMAM TAQIYUDDIN
NPM. 22083010026

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 24 Juni 2026.

Menyetujui,

Aviolla Terza Damaliana, S.Si., M.Stat
NIP. 19940802 202203 2 015


.....

(Pembimbing I)

Andri Fauzan Adziima, S.Si., M.Si.
NIP. 19950512 202406 1 001


.....

(Pembimbing II)

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 19830310 202121 1 006


.....

(Ketua Penguji)

Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 19950723 202406 1 002


.....

(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

***STACKING ENSEMBLE* BERBASIS OPTIMASI *HYPERPARAMETER*
PADA PERAMALAN INDEKS STANDAR PENCEMAR UDARA KOTA**

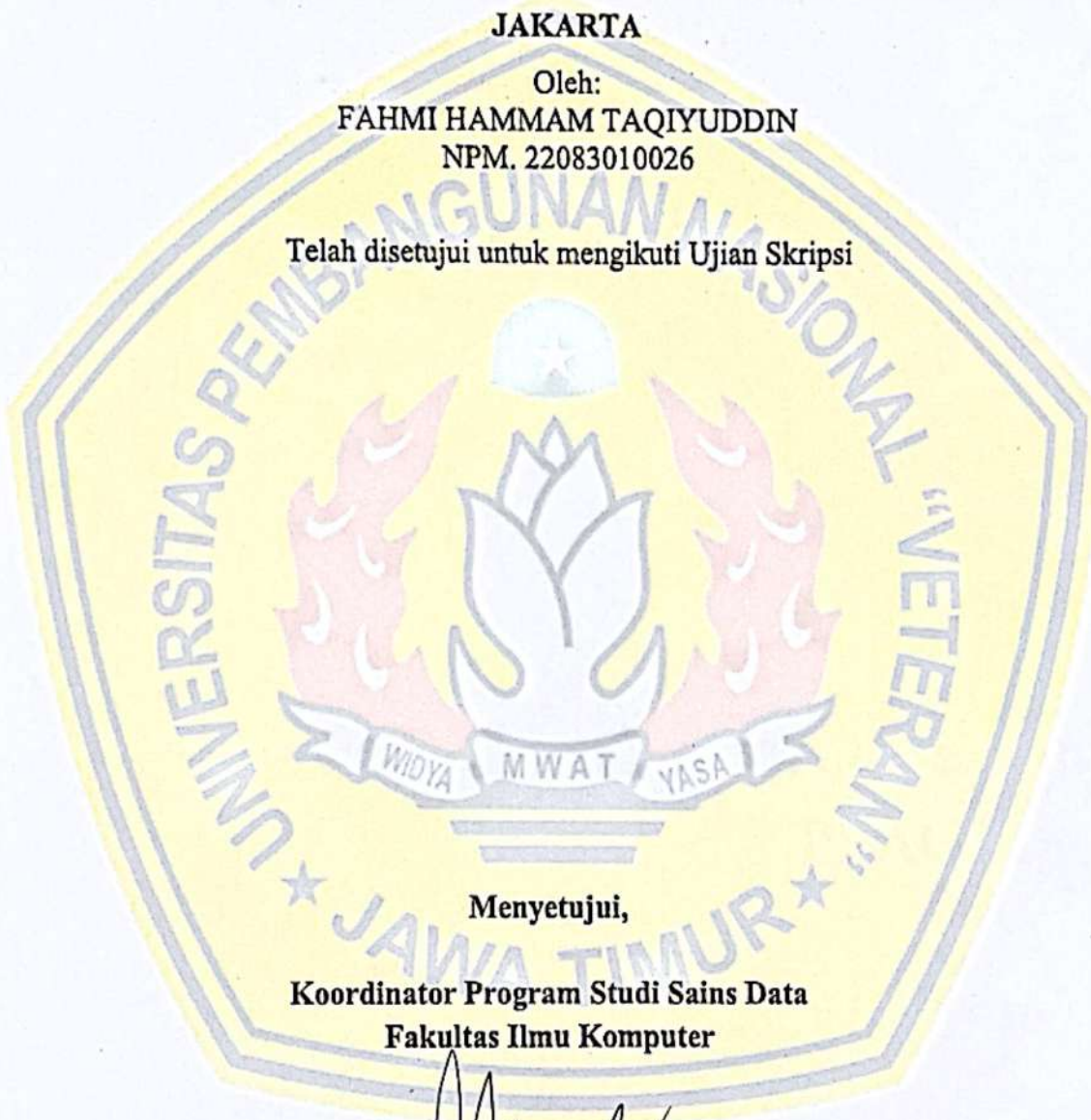
JAKARTA

Oleh:

FAHMI HAMMAM TAQIYUDDIN

NPM. 22083010026

Telah disetujui untuk mengikuti Ujian Skripsi



Menyetujui,

**Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**

Dr. Ir. I Gede Susrama Mas Divasa, S.T., M.T.

NIP. 19700619 202121 1 009

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Fahmi Hammam Taqiyuddin
NPM : 22083010026
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 19 Juni 2026
Yang Membuat Pernyataan,



FAHMI HAMMAM TAQIYUDDIN
NPM. 22083010026

ABSTRAK

Nama Mahasiswa / NPM : Fahmi Hammam Taqiyuddin / 22083010026
Judul Skripsi : *Stacking Ensemble* Berbasis Optimasi *Hyperparameter* pada Peramalan Indeks Standar Pencemar Udara Kota Jakarta
Dosen Pembimbing : 1. Aviolla Terza Damaliana, S.Si., M.Stat
2. Andri Fauzan Adziima, S.Si., M.Si.

Pencemaran udara masih menjadi permasalahan lingkungan yang berdampak terhadap kesehatan masyarakat dan kualitas hidup perkotaan. Laporan *State of Global Air 2024* menunjukkan bahwa pencemaran udara berkontribusi terhadap sekitar 8,1 juta kematian global pada tahun 2021. Di Jakarta, pencemaran udara diperkirakan menyebabkan lebih dari 10.000 kematian serta kerugian kesehatan yang mencapai miliaran dolar setiap tahun. Penelitian ini bertujuan membangun model prediksi Indeks Standar Pencemar Udara (ISPU) Kota Jakarta menggunakan arsitektur *Stacking Ensemble* yang mengombinasikan *Random Forest* dan *XGBoost* sebagai *base learner* serta *Ridge Regression* sebagai *meta learner*. Untuk meningkatkan performa model, diterapkan optimasi *hyperparameter* menggunakan *Bayesian Optimization* dan *Genetic Algorithm*. Data penelitian menggunakan data kualitas udara harian Kota Jakarta periode 2023–2025 yang diperoleh dari portal Satu Data Jakarta. Hasil penelitian menunjukkan bahwa kedua metode optimasi mampu meningkatkan performa model dibandingkan *Baseline*, yang ditunjukkan oleh penurunan nilai RMSE dari 17,7586 menjadi 16,7993 pada *Bayesian Optimization* dan 16,7869 pada *Genetic Algorithm*, serta peningkatan nilai R^2 dari 0,2546 menjadi 0,3330 dan 0,3362. Model terbaik kemudian diimplementasikan ke dalam *dashboard* berbasis *Streamlit* untuk menyajikan hasil *forecasting* ISPU secara interaktif.

Kata Kunci: *ISPU, Kualitas Udara, Stacking Ensemble, Random Forest, XGBoost, Genetic Algorithm, Forecasting.*

ABSTRACT

Student Name / NPM : Fahmi Hammam Taqiyuddin / 22083010026
Undergraduate thesis title : *Stacking Ensemble Based on Hyperparameter Optimization for Forecasting the Air Pollution Standard Index in Jakarta City*
Advisors : 1. Aviolla Terza Damaliana, S.Si., M.Stat
2. Andri Fauzan Adziima, S.Si., M.Si.

Air pollution is still an environmental problem that has an impact on public health and urban quality of life. The State of Global Air 2024 report shows that air pollution contributed to about 8.1 million global deaths in 2021. In Jakarta, air pollution is estimated to cause more than 10,000 deaths and health losses that amount to billions of dollars each year. This study aims to build a prediction model of the Jakarta City Air Pollution Standard Index (ISPU) using the Stacking Ensemble architecture that combines Random Forest and XGBoost as the base learner and Ridge Regression as the meta learner. To improve model performance, hyperparameter optimization is applied using Bayesian Optimization and Genetic Algorithm. The research data uses daily air quality data for the City of Jakarta for the 2023–2025 period obtained from the One Data Jakarta portal. The results showed that both optimization methods were able to improve the performance of the model compared to Baseline, which was shown by the decrease in RMSE values from 17.7586 to 16.7993 in Bayesian Optimization and 16.7869 in the Genetic Algorithm, as well as the increase in R^2 values from 0.2546 to 0.3330 and 0.3362. The best model is then implemented into a Streamlit-based dashboard to present ISPU forecasting results interactively.

Keywords: *Air Pollution Standard Index (ISPU), Air Quality, Stacking Ensemble, Random Forest, XGBoost, Genetic Algorithm, Forecasting.*

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga proposal skripsi dengan judul “***Stacking Ensemble* Berbasis Optimasi *Hyperparameter* pada Peramalan Indeks Standar Pencemar Udara Kota Jakarta**” dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Ibu Aviolla Terza Damaliana, S.Si., M.Stat. selaku Dosen Pembimbing I dan Bapak Andri Fauzan Adziima, S.Si., M.Si. selaku Dosen Pembimbing II yang telah meluangkan waktunya untuk memberikan bimbingan, nasihat, arahan, serta motivasi kepada penulis. Selain itu, selama penyusunan skripsi ini penulis juga banyak menerima bantuan, doa, dan dukungan dari berbagai pihak. Untuk itu, penulis mengucapkan terima kasih kepada:

1. Allah SWT, Tuhan Semesta Alam, atas izin, rahmat, dan kemudahan yang telah diberikan kepada penulis sehingga dapat menempuh pendidikan serta menyelesaikan penyusunan skripsi ini dengan baik.
2. Kedua orang tua penulis yang telah menjadi support system utama dalam kehidupan penulis, baik dalam bentuk dukungan finansial, doa, motivasi, kasih sayang, maupun segala hal yang menunjang kehidupan akademik dan pribadi penulis selama menempuh perkuliahan.
3. Keluarga “Masa Depan Cerah”, yaitu Regia, Jefran, dan Alip, yang telah berada di sisi penulis selama menempuh perkuliahan. Terima kasih telah menjadi teman berbagi cerita, sobat healing, teman ngopskuy, serta bagian dari perjalanan penulis selama menjalani kehidupan perkuliahan.
4. Seluruh teman-teman perkuliahan yang telah kebersamai penulis selama menempuh pendidikan di perguruan tinggi. Terima kasih atas bantuan, dukungan, kebersamaan, serta pengalaman berharga yang telah diberikan selama masa perkuliahan.
5. Pihak-pihak lain yang tidak dapat disebutkan satu per satu, yang telah membantu dan memberikan dukungan kepada penulis dalam proses penyusunan skripsi ini.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 19 Juni 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xvi
DAFTAR NOTASI.....	xvii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	7
1.3. Batasan Masalah	7
1.4. Tujuan Penelitian	8
1.5. Manfaat Penelitian	8
BAB II TINJAUAN PUSTAKA	9
2.1. Penelitian Terdahulu	9
2.2. Kerangka Teori	15
2.2.1. Indeks Standar Pencemar Udara (ISPU).	16
2.2.2. Parameter Pencemar Udara	17
2.2.2.1. <i>Particulate Matter</i> (PM ₁₀ dan PM _{2.5})	17

2.2.2.2.	Karbon Monoksida (CO)	18
2.2.2.3.	Sulfur Dioksida (SO ₂)	19
2.2.2.4.	Nitrogen Dioksida (NO ₂)	19
2.2.2.5.	Ozon (O ₃)	20
2.2.3.	Prediksi	21
2.2.4.	<i>Autocorrelation Function</i> (ACF)	21
2.2.5.	<i>Partial Autocorrelation Function</i> (PACF)	22
2.2.6.	<i>Time Series Split</i> (TSS)	23
2.2.7.	<i>Random Forest</i> (RF)	24
2.2.8.	<i>Extreme Gradient Boosting</i> (XGBoost)	26
2.2.9.	<i>Ridge Regression</i>	29
2.2.10.	<i>Stacking Ensemble Learning</i>	33
2.2.11.	Optimasi Hyperparameter	36
2.2.11.1.	<i>Bayesian Optimization</i> (BO)	37
2.2.11.2.	<i>Genetic Algorithm</i> (GA)	41
2.2.12.	Matriks Evaluasi	45
2.2.12.1.	Mean Absolute Error (MAE)	46
2.2.12.2.	Root Mean Square Error (RMSE)	46
2.2.12.3.	Mean Absolute Percentage Error (MAPE)	47
2.2.12.4.	R-Squared (R ²)	48
BAB III METODOLOGI PENELITIAN		51
3.1.	Variabel Penelitian dan Sumber Data	51
3.2.	Langkah Analisis	52
3.3.	Desain Sistem	58
BAB IV HASIL DAN PEMBAHASAN		61
4.1.	Pengumpulan Data	61
4.2.	<i>Preprocessing</i> Data	62
4.2.1.	Agregasi Data	62
4.2.2.	Penanganan <i>Missing value</i>	63

4.2.3.	<i>Exploratory Data Analysis (EDA)</i>	64
4.2.4.	Uji Stasioneritas	68
4.2.5.	<i>Differencing Data</i>	70
4.2.6.	Analisis ACF dan PACF	71
4.2.7.	Penentuan <i>Lag</i>	72
4.2.8.	<i>Feature engineering</i>	74
4.2.9.	Pembagian Data	76
4.2.10.	Normalisasi Data.....	77
4.3.	Pemodelan.....	79
4.3.1.	Model <i>Baseline Random Forest</i>	79
4.3.2.	Model <i>Baseline XGBoost</i>	80
4.3.3.	Model <i>Baseline Stacking</i>	81
4.4.	Optimasi <i>Hyperparameter</i>	85
4.4.1.	<i>Random Forest Bayesian Optimization</i>	85
4.4.2.	<i>XGBoost Bayesian Optimization</i>	93
4.4.3.	<i>Stacking Bayesian Optimization</i>	95
4.4.4.	<i>Random Forest Genetic Algorithm</i>	99
4.4.5.	<i>XGBoost Genetic Algorithm</i>	108
4.4.6.	<i>Stacking Genetic Algorithm</i>	110
4.5.	Hasil Pengujian dan Analisis Model.....	113
4.5.1.	Evaluasi dan Perbandingan Kinerja Model.....	113
4.5.2.	Hasil <i>Forecasting</i> dan Analisis Prediksi ISPU	117
4.6.	<i>Deployment</i> dan Implementasi Sistem Prediksi ISPU	120
BAB V PENUTUP.....		125
5.1.	Kesimpulan	125
5.2.	Saran Pengembangan	126
DAFTAR PUSTAKA.....		129
LAMPIRAN.....		133

DAFTAR GAMBAR

Gambar 2. 1 <i>Workflow Random Forest</i>	24
Gambar 2. 2 <i>Workflow XGBoost</i>	27
Gambar 2. 3 <i>Workflow Ridge Regression</i>	30
Gambar 2. 4 <i>Workflow Stacking</i>	34
Gambar 3. 1 Diagram Alir Penelitian	52
Gambar 3. 2 Dashboard ISPU	58
Gambar 4. 1 Grafik Time Series Nilai Ymax	66
Gambar 4. 2 Grafik Plot ACF dan PACF	71
Gambar 4. 3 Perbandingan Nilai R_2 pada data latih dan data uji	116
Gambar 4. 4 Visualisasi Hasil <i>Forecasting</i> ISPU 7 Hari ke Depan	120
Gambar 4. 5 Dashboard Utama Sistem Prediksi ISPU	121
Gambar 4. 6 Halaman Prediksi Sistem	122
Gambar 4. 7 Halaman Analisis Pola Prediksi	123

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	9
Tabel 2. 2 Batas Indeks Standar Pencemar Udara	16
Tabel 3. 1 Struktur Data Penelitian	51
Tabel 3. 2 Parameter <i>Random Forest</i>	56
Tabel 3. 3 Parameter XGBoost	56
Tabel 3. 4 Kategori Kualitas Udara Berdasarkan ISPU	58
Tabel 3. 5 Kategori Warna Indeks Standar Pencemar Udara	59
Tabel 4. 1 Data Mentah Kualitas Udara.....	61
Tabel 4. 2 Variabel Penelitian.....	62
Tabel 4. 3 Hasil Agregasi Data Harian	63
Tabel 4. 4 Statistik Deskriptif Data.....	64
Tabel 4. 5 Statistik Nilai Ymax Berdasarkan Hari.....	66
Tabel 4. 6 Hasil Uji ADF	69
Tabel 4. 7 Hasil <i>Differencing</i> Data	70
Tabel 4. 8 Ringkasan Analisis Numerik ACF dan PACF	72
Tabel 4. 9 Hasil Signifikansi ACF dan PACF	73
Tabel 4. 10 Hasil <i>Feature engineering</i>	75
Tabel 4. 11 Hasil Pembagian Data.....	76
Tabel 4. 12 Ringkasan Normalisasi Data pada Setiap Skenario Lag.....	78
Tabel 4. 13 Sampel Data Model RF.....	80
Tabel 4. 14 Hasil Prediksi XGB	81
Tabel 4. 15 Hasil Eksperimen Lag pada Model <i>Baseline Stacking</i>	83
Tabel 4. 16 Hasil Prediksi Model Dasar	83
Tabel 4. 17 Parameter Model <i>Meta Ridge</i>	84
Tabel 4. 18 Parameter Model <i>Meta Ridge</i>	85
Tabel 4. 19 Ruang Pencarian <i>Hyperparameter</i> RF BO	86
Tabel 4. 20 Kombinasi <i>Hyperparameter Trial</i>	87
Tabel 4. 21 Pembagian <i>Fold TSS</i>	87
Tabel 4. 22 Hasil Prediksi Empat <i>Trial</i> pada <i>Fold 1</i>	88

Tabel 4. 23 Perhitungan <i>Error Trial</i>	88
Tabel 4. 24 <i>Total Squared Error Fold 1</i>	89
Tabel 4. 25 Hasil RMSE setiap <i>fold</i>	90
Tabel 4. 26 Hasil Rata-rata RMSE <i>Trial</i>	92
Tabel 4. 27 Parameter Terbaik <i>Random Forest Bayesian Optimization</i>	92
Tabel 4. 28 Hasil Prediksi <i>Random Forest Bayesian Optimization</i>	93
Tabel 4. 29 Parameter Terbaik XGBoost BO	94
Tabel 4. 30 Hasil Prediksi XGBoost <i>Bayesian Optimization</i>	95
Tabel 4. 31 Hasil Eksperimen Lag pada Model <i>Stacking Bayesian Optimization</i>	96
Tabel 4. 32 Hasil Prediksi Model Dasar.....	97
Tabel 4. 33 Parameter Model Meta Ridge.....	98
Tabel 4. 34 Hasil Prediksi <i>Stacking Bayesian Optimization</i>	99
Tabel 4. 35 Ruang Pencarian <i>Hyperparameter RF GA</i>	100
Tabel 4. 36 Pembentukan Individu RF GA	101
Tabel 4. 37 Pembagian <i>Fold TSS</i>	101
Tabel 4. 38 Hasil Prediksi Empat Individu pada <i>Fold 1</i>	102
Tabel 4. 39 Perhitungan <i>Error</i> Prediksi Empat Individu.....	103
Tabel 4. 40 <i>Total Squared Error Fold 1</i>	104
Tabel 4. 41 Hasil RMSE Setiap <i>Fold</i>	105
Tabel 4. 42 Hasil Rata-rata RMSE Individu.....	106
Tabel 4. 43 Nilai <i>Fitness</i> Setiap Individu	107
Tabel 4. 44 Parameter Terbaik <i>Random Forest Genetic Algorithm</i>	107
Tabel 4. 45 Hasil Prediksi <i>Random Forest Genetic Algorithm</i>	108
Tabel 4. 46 Parameter Terbaik XGBoost <i>Genetic Algorithm</i>	109
Tabel 4. 47 Hasil Prediksi XGBoost <i>Genetic Algorithm</i>	110
Tabel 4. 48 Hasil Eksperimen Lag pada Model <i>Stacking Genetic Algorithm</i>	111
Tabel 4. 49 Hasil Prediksi Model Dasar.....	111
Tabel 4. 50 Parameter Model Meta Ridge.....	112
Tabel 4. 51 Hasil Prediksi <i>Stacking Genetic Algorithm</i>	113
Tabel 4. 52 Matriks Evaluasi Data Latih	114
Tabel 4. 53 Matriks Evaluasi Data Uji	115

Tabel 4. 54 Hasil <i>Forecasting</i> ISPU 7 Hari Ke Depan.....	119
--	-----

DAFTAR LAMPIRAN

Lampiran 1. Dataset.....	133
Lampiran 2. Kode Program	134
Lampiran 3. GUI.....	135
Lampiran 4. LOA Jurnal.....	136

DAFTAR NOTASI

I	:	Intensitas
W_R	:	<i>Weight factor</i>
H	:	<i>Hue</i>
S	:	<i>Saturation</i>
V	:	<i>Value</i>
dst	:	Gambar akumulator
scr	:	Gambar input
F	:	<i>Foreground</i>
B	:	<i>Background</i>
f	:	<i>Frame</i>
SE	:	<i>Structuring element</i>
ECD	:	<i>Equivalent circular diameter</i>
b	:	Bias
$f(x)m$	:	Fungsi vektor masukan
$d(x', x)$	:	Jarak di antara data uji z ke setiap vector data latih
$K(x, y)$	:	Fungsi kernel linear
$G(x)$	:	Fungsi gaussian satu dimensi
σ	:	Standard deviasi dari distribusi
$G(x, y)$	:	Fungsi gaussian dua dimensi