



SKRIPSI

IMPLEMENTASI HYBRID TABNET-XGBOOST BERBASIS DATA RADIOSONDE UNTUK PREDIKSI INTENSITAS HUJAN DI KOTA SURABAYA

**ANNABEL GRACIA PURYANI
NPM 22083010048**

DOSEN PEMBIMBING

Aviolla Terza Damaliana, S.Si., M.Stat.

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.

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

LEMBAR PENGESAHAN

IMPLEMENTASI HYBRID TABNET-XGBOOST BERBASIS DATA RADIOSONDE UNTUK PREDIKSI INTENSITAS HUJAN DI KOTA SURABAYA

Oleh:
ANNABEL GRACIA PURYANI
NPM. 22083010048

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 31 Agustus 2026:

Menyetujui,

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



(Pembimbing I)

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.
NIP. 19990606 202406 1 001



(Pembimbing II)

Trimono, S.Si., M.Si
NIP. 19950908 202203 1 003



(Ketua Penguji)

Wahvu Syaifullah J. S., S.Kom., M.Kom
NIP. 19860825 202121 1 003



(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

IMPLEMENTASI HYBRID TABNET-XGBOOST BERBASIS DATA RADIOSONDE UNTUK PREDIKSI INTENSITAS HUJAN DI KOTA SURABAYA

Oleh:
ANNABEL GRACIA PURYANI
NPM. 22083010048

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer



Amri Muhaimin, S.Stat., M.Stat., M.S.

NIP, 19950723 202406 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Annabel Gracia Puryani
NPM : 22083010048
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, 3 September 2026
Yang Membuat Pernyataan,



ANNABEL GRACIA PURYANI
NPM. 22083010048

ABSTRAK

Nama Mahasiswa / NPM : Annabel Gracia Puryani / 22083010048
Judul Skripsi : Implementasi Hybrid TabNet-XGBoost Berbasis Data Radiosonde untuk Prediksi Intensitas Hujan di Kota Surabaya
Dosen Pembimbing : 1. Aviolla Terza Damaliana, S.Si., M.Stat.
2. Alfian Rizaldy Pratama, S.Tr.T., M.Tr.Kom.

Curah hujan merupakan salah satu unsur cuaca yang berperan penting dalam aktivitas masyarakat dan berkaitan erat dengan risiko bencana hidrometeorologi, terutama banjir di wilayah perkotaan seperti Kota Surabaya. Variabilitas intensitas curah hujan yang tinggi dan keberadaan data ekstrem menjadi tantangan dalam membangun model prediksi yang mampu menghasilkan kesalahan prediksi rendah, sementara pemanfaatan data radiosonde dalam pendekatan *machine learning* dengan kombinasi *TabNet* dan model regresi masih perlu dikembangkan. Penelitian ini bertujuan mengembangkan model prediksi intensitas curah hujan harian di Kota Surabaya berbasis data radiosonde periode 2019–2025 melalui seleksi variabel menggunakan *Pearson Correlation* dan *TabNet*, penentuan *lag* menggunakan *Cross-Correlation Function (CCF)*, serta penerapan tiga skenario, yaitu *Hybrid TabNet-XGBoost*, *Hybrid TabNet-XGBoost* dengan penanganan data ekstrem, dan *Hybrid TabNet-SVR*. Penanganan data ekstrem dilakukan dengan tidak menyertakan data curah hujan lebih dari 80 mm dalam proses pelatihan, sedangkan optimasi parameter model dilakukan menggunakan *Optuna* dan evaluasi menggunakan RMSE, MAE, dan MAPE. Hasil penelitian menunjukkan bahwa *Hybrid TabNet-XGBoost* menghasilkan RMSE 19,1404 mm, MAE 11,8515 mm, dan MAPE 18,53%, sedangkan *Hybrid TabNet-XGBoost* dengan penanganan data ekstrem menghasilkan RMSE 14,4482 mm, MAE 10,1127 mm, dan MAPE 19,03%; sementara *Hybrid TabNet-SVR* menghasilkan RMSE 20,8319 mm, MAE 10,6938 mm, dan MAPE 12,36%. Dengan demikian, *Hybrid TabNet-XGBoost* dengan penanganan data ekstrem memberikan kinerja terbaik berdasarkan RMSE dan MAE, sedangkan *Hybrid TabNet-SVR* menghasilkan MAPE terendah. Inovasi penelitian ini terletak pada penggabungan *TabNet* sebagai metode seleksi variabel dengan *XGBoost* dan *SVR*, pemanfaatan *lag* berbasis CCF, serta pengujian pengaruh penanganan data curah hujan ekstrem dalam tiga skenario model. Ketiga model selanjutnya digunakan untuk menghasilkan prediksi intensitas curah hujan tahun 2026 yang disajikan melalui aplikasi berbasis *Streamlit* dalam bentuk tabel dan visualisasi.

Kata Kunci: Curah Hujan, Radiosonde, TabNet, XGBoost, SVR, Prediksi

ABSTRACT

Nama Mahasiswa / NPM : Annabel Gracia Puryani / 22083010048
Judul Skripsi : Implementasi Hybrid TabNet-XGBoost Berbasis
rainosonde data for Rainfall Intensity Preinction in
Kota Surabaya
Supervisors : 3. Aviolla Terza Damaliana, S.Si., M.Stat.
4. Alfian Rizaldy Pratama, S.Tr.T., M.Tr.Kom.

rainfall is salah satu unsur cuaca that berperan important in aktivitas masyarakat and berkaitan erat with risiko bencana hidrometeorologi, especially banjir in wilayah perkotaan seperti Kota Surabaya. Variabilitas intensity rainfall that high and toberadaan extreme data becomes tantangan in membangun preinction model that mampu generate preinction error low, sementara pebenefitan rainosonde data in pendekatan *machine learning* with kombinasi *TabNet* and model regresi still perlu intombangkan. this study berobjective mengembangkan preinction model intensity rainfall dayan in Kota Surabaya berbasis rainosonde data periode 2019–2025 through seleksi variable using Pearson *Correlation* and *TabNet*, penentuan *lag* using *Cross-Correlation Function* (CCF), as well as implementation tiga scenario, namely *Hybrid TabNet–XGBoost*, *Hybrid TabNet–XGBoost* with extreme-data handling, and *Hybrid TabNet–SVR*. extreme-data handling was performed using not menyertwill data rainfall more from 80 mm in trathisng process, whereas model optimization parameters conducted using *Optuna* and evaluation using RMSE, MAE, and MAPE. research results shows that *Hybrid TabNet–XGBoost* generate RMSE 19,1404 mm, MAE 11,8515 mm, and MAPE 18,53%, whereas *Hybrid TabNet–XGBoost* with extreme-data handling generate RMSE 14,4482 mm, MAE 10,1127 mm, and MAPE 19,03%; sementara *Hybrid TabNet–SVR* generate RMSE 20,8319 mm, MAE 10,6938 mm, and MAPE 12,36%. therefore, *Hybrid TabNet–XGBoost* with extreme-data handling provides performance best based on RMSE and MAE, whereas *Hybrid TabNet–SVR* generate MAPE telow. Inovasi this study terletak on penggabungan *TabNet* as method seleksi variable with *XGBoost* and SVR, pebenefitan *lag* berbasis CCF, as well as testing pengaruh handling extreme rainfall data in tiga scenario model. third model subsequently is used to generate preinction intensity rainfall year 2026 that insajikan through aplikasi berbasis *Streamlit* in the form of table and visualisasi.

Keyword: Rainfall, Rainosonde, TabNet, XGBoost, SVR, Prediction

KATA PENGANTAR

Puji syukur kehadirat Tuhan Yang Maha Esa atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga proposal skripsi dengan judul **“Implementasi *Hybrid TabNet-XGBoost* Berbasis Data Radiosonde untuk Prediksi Intensitas Hujan di Kota Surabaya”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Ibu Terza Damaliana, S.Si., M.Stat selaku Dosen Pembimbing I dan Bapak Alfian Rizaldy Pratama, S.Tr.T., M.Tr.Kom. selaku Dosen Pembimbing 2 yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Selain itu, selama penyusunan proposal skripsi penulis juga banyak menerima bantuan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Bapak Amri Muhaimin, S.Stat., M.Stat., M.S. selaku Koordinator Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
3. Dosen-dosen Program Studi Sains Data yang telah memberikan ilmu dan bimbingan kepada penulis.
4. Mama, Bapak, Adek, dan seluruh keluarga besar penulis yang selalu mendukung dan mendoakan penulis sampai titik ini.
5. Teman baik, teman dekat, dan seluruh teman sains data gen 3 yang telah menemani dan membantu penulis selama perkuliahan.

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, September 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
ABSTRACT.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xiii
DAFTAR NOTASI.....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	11
1.3. Batasan Masalah	11
1.4. Tujuan Penelitian	12
1.5. Manfaat Penelitian	13
BAB II TINJAUAN PUSTAKA	15
2.1. Penelitian Terdahulu	15
2.2. Dasar Teori.....	19
BAB III METODOLOGI PENELITIAN	53

3.1.	Variabel Penelitian dan Sumber Data	53
3.2.	Langkah Analisis	55
3.3.	Desain Sistem.....	68
BAB IV HASIL DAN PEMBAHASAN		74
4.1.	Pengumpulan Data	75
4.2.	Preprocessing Data.....	78
4.3.	Exploratory Data Analysis (EDA)	86
4.4.	Uji Korelasi dengan Korelasi Pearson	91
4.5.	Seleksi Variabel menggunakan TabNet.....	96
4.6.	Penentuan Lag menggunakan Cross-Correlation Function (CCF)	115
4.7.	Pembangunan Model Prediksi Curah Hujan.....	117
4.8.	Perbandingan Kinerja Model	160
4.9.	Prediksi 2026	164
4.10,	Tampilan Antar Muka (Graphical User Interface).....	166
BAB V PENUTUP.....		176
5.1	Kesimpulan	177
5.2	Saran Pengembangan	179
DAFTAR PUSTAKA.....		181
LAMPIRAN.....		184

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Model TabNet	34
Gambar 2.2 Arsitektur Model XGBoost	40
Gambar 2.3 <i>Flowchart Hybrid TabNet-XGBoost</i>	43
Gambar 2.4 Arsitektur Model SVR.....	45
Gambar 2. 5 <i>Flowchart Hybrid TabNet-SVR</i>	48
Gambar 3.1 Langkah Analisis	56
Gambar 3.2 Desain sistem prediksi	69
Gambar 4.1 Visualisasi Variabel Radiosonde	89
Gambar 4.2 Visualisasi Distribusi CH	91
Gambar 4.3 Plot <i>Validation Loss TabNet</i>	105
Gambar 4.4 <i>Validation Loss TabNet</i> penanganan ekstrem	142
Gambar 4.5 Aktual vs Prediksi XGBoost	162
Gambar 4.6 Aktual vs Prediksi <i>XGBoost</i> Penanganan Ekstrem	163
Gambar 4.7 Aktual vs Prediksi SVR.....	163
Gambar 4.8 Halaman Utama Sistem Prediksi Curah Hujan	167
Gambar 4.9 Halaman Prediction Sistem Prediksi Curah Hujan.....	168
Gambar 4. 10 Pemilihan Model Pipeline	170
Gambar 4.11 Halaman <i>Analysis & Results Tab Preprocessing</i>	171
Gambar 4. 12 Halaman <i>Analysis & Results Tab Feature Engineering</i>	172
Gambar 4. 13 Halaman <i>Analysis & Results Tab</i> Evaluasi Model	173
Gambar 4. 14 Halaman Analysis & Results Tab Grafik	174
Gambar 4. 15 Halaman <i>Analysis & Results Tab</i> Prediksi	174

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	15
Tabel 2. 2 Ambang Batas SWEAT	21
Tabel 2. 3 Ambang Batas LI	22
Tabel 2. 4 Ambang Batas KI.....	22
Tabel 2. 5 Ambang Batas CAPE.....	23
Tabel 2. 6 Ambang Batas SI	24
Tabel 2. 7 Ambang Batas TT	24
Tabel 2. 8 Ambang Batas TPW	25
Tabel 2. 9 Ambang Batas CIN.....	25
Tabel 2. 10 Ambang Batas MVV	26
Tabel 2. 11 Ambang Batas Hujan	30
Tabel 2. 12 Kriteria Akurasi MAPE	51
Tabel 3.1 Variabel Penelitian.....	53
Tabel 3.2 Struktur Data.....	54
Tabel 3. 3 Hyperparameter TabNet	60
Tabel 3. 4 Hyperparameter Optuna.....	63
Tabel 3. 5 Hyperparameter XGBoost	64
Tabel 3. 6 <i>Hyperparameter</i> SVR.....	67
Tabel 4.1 <i>Dataset</i> Radiosonde Tahun 2019-2025.....	76
Tabel 4.2 Jumlah <i>Missing Value</i>	79
Tabel 4.3 <i>Dataset</i> Radiosonde setelah <i>preprocessing</i>	83
Tabel 4.4 Statistika Deskriptif Variabel Radiosonde.....	86
Tabel 4.5 Haisl Korelasi Pearson.....	93
Tabel 4. 6 Evaluasi Regresi Linear	97
Tabel 4. 7 Hasil Uji Multikolinearitas	99
Tabel 4. 8 Hasil multikolinearitas setelah penanganan.....	100
Tabel 4.9 Variabel <i>Importance TabNet</i>	107
Tabel 4.10 Data Perhitungan <i>TabNet</i>	108

Tabel 4.11 Rata-Rata dan Standar Deviasi Variabel	109
Tabel 4.12 Hasil standarisasi Manual	110
Tabel 4.13 Hasil Sparsemax Manual	113
Tabel 4.14 Hasil <i>Prior Scales</i> Manual.....	114
Tabel 4.15 Hasil Feature Importance Manual.....	115
Tabel 4.16 Hasil <i>lag</i> CCF	116
Tabel 4.17 Dataset skenario pertama	119
Tabel 4.18 Hyperparameter optuna skenario 1	123
Tabel 4.19 <i>Dataset</i> perhitungan manual <i>XGBoost</i>	128
Tabel 4.20 Nilai <i>Gradient</i> Manual.....	129
Tabel 4.21 Kandidat <i>Split XGBoost</i> Manual.....	131
Tabel 4.22 Hasil Gain Manual <i>XGBoost</i>	136
Tabel 4.23 Hasil Nilai <i>Leaf</i> Manual.....	137
Tabel 4.24 Evaluasi Model <i>XGBoost</i>	138
Tabel 4.25 Korelasi Pearson <i>XGBoost</i> Penanganan Ekstrem	141
Tabel 4.26 Variabel <i>Importance</i> <i>XGBoost</i> Penanganan Ekstrem	143
Tabel 4.27 Hasil <i>lag</i> CCF Penanganan Ekstrem.....	144
Tabel 4.28 <i>Hyperparameter Optuna</i> Penanganan Ekstrem	145
Tabel 4.29 Evaluasi Model <i>XGBoost</i> Penanganan Ekstrem	146
Tabel 4.30 <i>Dataset</i> Perhitungan Manual SVR.....	153
Tabel 4.31 Rata-Rata dan Standar Deviasi Data SVR.....	153
Tabel 4.32 Hasil Standarisasi Manual SVR.....	154
Tabel 4.33 Hasil Prediksi Manual SVR.....	158
Tabel 4.34 Evaluasi Model SVR	160
Tabel 4.35 Perbandingan Evaluasi Model	161
Tabel 4.36 Hasil prediksi skenario 1.....	164
Tabel 4.37 Hasil prediksi skenario 2.....	165
Tabel 4.38 Hasil prediksi skenario 3.....	166

DAFTAR LAMPIRAN

Lampiran 1. LoA	184
Lampiran 2 Dokumentasi Kode Program Analisis	185
Lampiran 3 Dokumentasi Kode Program GUI	186
Lampiran 4 Data Penelitian.....	187

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