

SKRIPSI

**PERBANDINGAN METODE RANDOM FOREST DAN
EXTREME GRADIENT BOOSTING (XGBOOST) PADA
KLASIFIKASI TINGKAT PENGANGGURAN TERBUKA (TPT)
DI INDONESIA**



ROSA AULIA NISA

1519621011

**PROGRAM STUDI
SISTEM DAN TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS NEGERI JAKARTA**

2025

LEMBAR PENGESAHAN SKRIPSI

Judul : Perbandingan Metode Random Forest dan Extreme Gradient Boosting (XGBoost) pada Klasifikasi Tingkat Pengangguran Terbuka (TPT) di Indonesia
Penyusun : Rosa Aulia Nisa
NIM : 1519621011

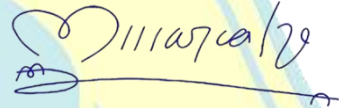
Disetujui oleh:

Pembimbing I,



Fuad Mumtas, M.T.I.
NIP. 198810312019032005

Pembimbing II,



Murien Nugraheni, M.Cs.
NIP. 198710112019032012

Pengesahan Panitia Ujian Skripsi:

Ketua Penguji,



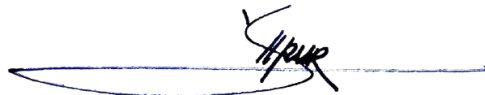
Ali Idrus, M.Kom.
NIP. 198802262019031010

Sekretaris,



Faroh Ladayya, M.Si.
NIP. 199401282020122018

Mengetahui,
Koordinator Program Studi Sistem dan Teknologi Informasi



Lipur Sugiyanta, Ph.D.
NIP. 197612292003121002

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Skripsi ini merupakan Karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 10 Juli 2025
Yang membuat pernyataan



Rosa Aulia Nisa
No. Reg. 1519621011



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Rosa Aulia Nisa
NIM : 1519621011
Fakultas/Prodi : Teknik / Sistem dan Teknologi Informasi
Alamat email : rosaaulia4@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain (.....)

yang berjudul :

Perbandingan Metode Random Forest dan Extreme Gradient Boosting (XGBoost) Pada Klasifikasi Tingkat Pengangguran Terbuka (TPT) di Indonesia.

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 4 Agustus 2025

Penulis

(Rosa Aulia Nisa)
nama dan tanda tangan

**PERBANDINGAN METODE RANDOM FOREST DAN EXTREME
GRADIENT BOOSTING (XGBOOST) PADA KLASIFIKASI TINGKAT
PENGANGGURAN TERBUKA (TPT) DI INDONESIA**

Rosa Aulia Nisa

Dosen Pembimbing: Fuad Mumtas, M.T.I. & Murien Nugraheni, M.Cs.

ABSTRAK

Tingkat Pengangguran Terbuka (TPT) merupakan indikator dalam mengukur pengangguran di Indonesia. Pengangguran sendiri masih menjadi masalah kompleks karena faktor-faktor yang saling berinteraksi dan sulit dipahami. Penelitian ini menyelidiki penerapan algoritma Random Forest dan XGBoost dalam mengklasifikasi status TPT di Indonesia yang dibagi menjadi dua kategori, yaitu tinggi dan rendah, serta membandingkan performa keduanya. Data yang digunakan terdiri atas 514 kabupaten/kota dari 38 Provinsi pada Tahun 2023 dengan variabel-variabel yang digunakan merupakan faktor sosial-ekonomi. Hasil penelitian ini menunjukkan bahwa Random Forest lebih unggul dibandingkan XGBoost. Random Forest memperoleh nilai *accuracy* sebesar 0,7484, *precision* 0,7959, *recall* 0,5735, *F1-score* 0,6667, dan ROC-AUC 0,8745. Sementara itu, XGBoost menghasilkan *accuracy* sebesar 0,7097, *precision* 0,7018, *recall* 0,5882, *F1-score* 0,6400, dan ROC-AUC 0,7847. Hasil *feature importance* menunjukkan bahwa Random Forest menyoroti tiga variabel utama, yakni Indeks Pembangunan Manusia, Kepadatan Penduduk, dan Tingkat Partisipasi Angkatan Kerja, sementara XGBoost memberikan kontribusi yang lebih proposional pada seluruh fitur.

Kata Kunci: Klasifikasi, Random Forest, *Extream Gradient Boosting* (XGBoost), Tingkat Pengangguran Terbuka.

**COMPARISON OF RANDOM FOREST AND EXTREME GRADIENT
BOOSTING (XGBOOST) METHODS IN CLASSIFICATION OF OPEN
UNEMPLOYMENT RATE (TPT) IN INDONESIA**

Rosa Aulia Nisa

Dosen Pembimbing: Fuad Mumtas, M.T.I. & Murien Nugraheni, M.Cs.

ABSTRACT

The Open Unemployment Rate (TPT) is an indicator used to measure unemployment in Indonesia. Unemployment remains a complex issue due to interaction and difficult-to-understand factors. This study investigates the application of the Random Forest and XGBoost algorithms in classifying the TPT status in Indonesia, which is divided into two categories: high and low, and conduct a comparative evaluation of both performances. The data used consists of 514 districts/cities from 38 provinces in 2023, with variables based on selected socio-economic factors. The results of this study indicate that Random Forest is superior to XGBoost. Random Forest obtained an accuracy value of 0,7484, precision 0,7959, recall 0,5735, F1-score 0,6667, and ROC-AUC 0,8745. Meanwhile, XGBoost obtained an accuracy of 0,7097, precision of 0,7018, recall of 0,5882, F1-score of 0,6400, and ROC-AUC of 0,7847. The feature importance results show that Random Forest highlights three main variables, namely the Human Development Index, Population Density, and Labor Force Participation Rate, while XGBoost provides a more proportional contribution to all features.

Key Word: *Classification, Random Forest, Extreme Gradient Boosting (XGBoost), The Open Unemployment Rate.*

KATA PENGANTAR

Alhamdulillahirobbil'alamiin. Puji syukur kehadiran Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat mencapai tahap ini dalam menempuh gelar sarjana. Skripsi yang berjudul “Perbandingan Random Forest dan Extreame Gradient Boosting (XGBoost) Pada Klasifikasi Tingkat Pengangguran Terbuka (TPT) di Indonesia” telah diselesaikan dengan sebaik-baiknya dalam memenuhi salah satu syarat memperoleh gelar Sarjana Komputer strata satu (S1) Program Studi Sistem dan Teknologi Informasi di Fakultas Teknik Universitas Negeri Jakarta.

Selama penyusunan skripsi ini, penulis menyadari banyak kesulitan dan tantangan yang dihadapi. Namun demikian, proses penyusunan skripsi ini dapat terselesaikan atas kerjasama, bantuan, dan dukungan dari berbagai pihak. Untuk itu, ucapan terima kasih dengan tulus penulis sampaikan khusus kepada:

1. Bapak Lipur Sugiyanta, Ph.D. selaku Koordinator Program Studi Sistem dan Teknologi Informasi Universitas Negeri Jakarta
2. Bapak Fuad Mumtas, S.Kom., M.T.I. dan Ibu Murien Nugraheni, M.Cs. selaku dosen pembimbing 1 dan dosen pembimbing 2 yang senantiasa menyediakan waktu untuk memberikan ilmu, saran serta dukungan selama proses penyusunan skripsi ini.
3. Seluruh Bapak dan Ibu Dosen Program Studi Sistem dan Teknologi Informasi yang telah memberikan ilmu-ilmu bermanfaat kepada penulis selama perkuliahan baik di dalam maupun di luar kelas.
4. Kedua orang tua penulis Bapak Kohar dan Ibu Tintin yang selalu memberikan dukungan dan doa semaksimal mungkin serta Kakak dan Adik penulis Teteh Qori, Ica, dan Zidane yang memberikan semangat dan doa kepada penulis hingga dapat melanjutkan perkuliahan sampai akhir dan menyelesaikan skripsi ini.
5. Nanda, Nia, dan Nurul yang telah menemani dan mewarnai kehidupan penulis dengan canda, tawa, dan tangis yang telah dilalui bersama-sama selama masa perkuliahan.

6. Teman-teman Sistem dan Teknologi Informasi Angkatan 2021 yang berjuang bersama dan selalu mendukung satu sama lain serta menemani perjalanan kuliah dari awal hingga akhir.
7. Seluruh pihak yang tidak bisa penulis sebutkan satu per satu namun tidak menghilangkan rasa hormat dan rasa terima kasih penulis dalam menyelesaikan studi ini.
8. Terakhir, kepada diri sendiri yang telah melalui proses panjang dalam membentuk diri ini menjadi lebih kuat dan tangguh. Terima kasih telah bertahan sejauh ini dan tidak menyerah meski sering merasa takut dan ragu selama proses perkuliahan hingga penyusunan skripsi ini. Terima kasih telah percaya bahwa setiap usaha akan menemukan hasilnya pada waktunya.

Sebagai penutup, penulis berharap agar tulisan skripsi ini dapat membantu dan memberikan manfaat bagi yang membacanya. Penulis menyadari bahwa penyusunan skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis tidak menutup diri dari segala bentuk kritik dan saran yang membangun dari berbagai pihak agar dapat lebih baik kedepannya.

Jakarta, 10 Juli 2025
Penyusun,

Rosa Aulia Nisa
1519621011

DAFTAR ISI

LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Pembatasan Masalah	4
1.4 Perumusan Masalah	4
1.5 Tujuan Penelitian.....	5
1.6 Manfaat Penelitian	5
BAB II KAJIAN PUSTAKA	6
2.1 Konsep dan Teori.....	6
2.1.1 Pengangguran.....	6
2.1.2 <i>Machine learning</i>	9
2.1.3 Klasifikasi	11
2.1.4 <i>Knowledge Discovery in Database (KDD)</i>	11
2.1.5 <i>Ensemble Learning</i>	15
2.1.6 Random Forest.....	16
2.1.7 Extreme Gradient Boosing (XGBoost).....	18
2.1.8 <i>Hyperparameter Tuning</i>	20
2.1.9 <i>Confusion Matrix</i>	23
2.1.10 Kurva <i>Receiver Operating Characteristic (ROC)</i>	25
2.1.11 <i>Variable Importance</i>	26
2.2 Penelitian yang Relevan	27
2.3 <i>State of The Art</i>	30

2.4 Kerangka Berpikir	31
BAB III METODOLOGI PENELITIAN	34
3.1 Tempat dan Waktu Penelitian	34
3.2 Alat dan Bahan Penelitian	34
3.2.1 Alat.....	34
3.2.2 Bahan Penelitian	35
3.3 Alur Penelitian	36
3.4 Teknik Pengumpulan Data.....	42
3.5 Metode Evaluasi Data.....	43
BAB IV HASIL DAN PEMBAHASAN.....	44
4.1 Hasil Penelitian	44
4.1.1 Data Selection.....	44
4.1.2 Data <i>Preprocessing</i>	50
4.1.3 Data Transformation	51
4.1.4 Data Mining	54
4.1.5 Data Evaluation.....	57
4.2 Pembahasan.....	62
BAB V KESIMPULAN DAN SARAN	66
5.1 Kesimpulan	66
5.2 Saran.....	66
DAFTAR PUSTAKA.....	67
LAMPIRAN.....	73
RIWAYAT HIDUP.....	92

DAFTAR TABEL

Nomor	Judul Tabel	Halaman
2.1	Parameter-parameter pada Random Forest	21
2.2	Parameter-parameter pada XGBoost	21
2.3	Ilustrai <i>Confusion Matrix</i> pada Pengukuran Klasifikasi	23
2.4	Keakuratan Hasil Klasifikasi Berdasarkan Nilai AUC	25
2.5	Penelitian Relevan	27
3.1	Alat Perangkat Keras Penelitian	34
3.2	Alat Perangkat Lunak Penelitian	34
3.3	Detail Fitur-Fitur yang Digunakan pada Penelitian	35
3.4	Data IPM Menurut Kabupaten/Kota di Indonesia	37
3.5	Data TPT dan TPAK menurut Kabupaten/Kota di Indonesia	38
3.6	Data Keseluruhan Setelah Penggabungan Berdasarkan Kabupaten/Kota	38
3.7	Hasil Evaluasi menggunakan <i>Confusion Matrix</i>	43
4.1	Dataset TPT beserta Variabel Lainnya pada Tahun 2023	44
4.2	Statistika Deskriptif	45
4.3	Hasil Pelabelan pada Dataset	52
4.4	Hasil Setelah Melakukan Label Encoder	52
4.5	Rasio Jumlah Data pada Data Latih dan Data Uji	53
4.6	Daftar Parameter Random Forest Beserta Nilai Inputnya	55
4.7	Daftar Parameter XGBoost Beserta Nilai Inputnya	55
4.8	Hasil Kombinasi Hyperparameter Terbaik pada Model Random Forest	56
4.9	Hasil Kombinasi Hyperparameter Terbaik pada Model XGBoost	56
4.10	Hasil Klasifikasi Random Forest dan XGBoost pada Nilai TPT	61
4.11	Feature Importance pada Random Forest dan XGBoost	62

DAFTAR GAMBAR

Nomor	Judul Gambar	Halaman
1.1	Tingkat Pengangguran Terbuka (TPT) di Indonesia Berdasarkan Provinsi Tahun 2023	1
2.1	Skema <i>Artificial Intelligence</i> dan <i>Machine learning</i>	9
2.2	Proses klasifikasi dengan <i>Input X</i> dan <i>Ouput Y</i>	11
2.3	Gambaran Langkah-Langkah Proses KDD	12
2.4	Ilustrasi Algoritma Random Forest	16
2.5	Ilustrasi Algoritma XGBoost	18
2.6	Ilustrasi <i>Grid Search</i> dan <i>Cross Validation</i>	21
2.7	Kurva ROC	25
2.8	Kerangka Berpikir Penelitian	33
3.1	Alur Penelitian	36
3.2	Menemukan Duplikasi Data	39
3.3	Pemberian Label Pada Variabel Target	40
3.4	Penerapan <i>Label Encoder</i> Pada Variabel Kategorik	40
3.5	Pembagian Data Menjadi Data Latih dan Data Uji	41
3.6	Pemodelan Random Forest dengan <i>RandomSearchCV</i>	41
3.7	Pemodelan XGBoost dengan <i>RandomSearchCV</i>	41
4.1	<i>Scatterplot</i> antara TPT dan TPAK	47
4.2	<i>Scatterplot</i> antara TPT dan IPM	48
4.3	<i>Scatterplot</i> antara TPT dan KP	49
4.4	<i>Scatterplot</i> TPT dan PPM	49
4.5	<i>Scatterplot</i> antara TPT dan PDRB	50
4.6	Tampilan Hasil Pemeriksaan Duplikasi Data	50
4.7	<i>Box Plot</i> untuk Deteksi <i>Outliers</i>	51
4.9	Jumlah Masing-Masing Kelas Pada Data Latih	53
4.10	Hasil Normalisasi Data menggunakan <i>Robust Scaler</i>	54
4.11	<i>Confusion Matrix</i> pada Random Forest dan XGBoost	58
4.12	Hasil ROC-AUC pada Random Forest dan XGBoost	60
4.13	Heatmap Performa Random Forest dan XGBoost	63
4.14	Visulisasi <i>Feature Importance</i> pada Random Forest dan XGBoost	64

DAFTAR LAMPIRAN

Nomor	Judul Lampiran	Halaman
1	Data Tingkat Pengangguran Terbuka (TPT) Kabupaten/Kota Tahun 2023	73
2	<i>Script</i> Penelitian Menggunakan Bahasa Pemograman Python	85

