

**OPTIMASI DAN PEMILIHAN ALGORITMA MACHINE
LEARNING UNTUK PREDIKSI HARGA RUMAH DI
KABUPATEN TANGERANG**



SKRIPSI

Daffa Kaisha Pratama Chandra

00000055770

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**OPTIMASI DAN PEMILIHAN ALGORITMA MACHINE
LEARNING UNTUK PREDIKSI HARGA RUMAH DI
KABUPATEN TANGERANG**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer

Daffa Kaisha Pratama Chandra

00000055770

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini,

Nama : Daffa Kaisha Pratama Chandra

Nomor Induk Mahasiswa : 00000055770

Program studi : Sistem Informasi

Skripsi dengan judul:

“OPTIMASI DAN PEMILIHAN ALGORITMA MACHINE LEARNING
UNTUK PREDIKSI HARGA RUMAH DI KABUPATEN TANGERANG”

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 24 Maret 2025



(Daffa Kaisha Pratama Chandra)

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

Skripsi dengan judul

OPTIMASI DAN PEMILIHAN ALGORITMA MACHINE LEARNING
UNTUK PREDIKSI HARGA RUMAH DI KABUPATEN TANGERANG

Oleh

Nama : Daffa Kaisha Pratama Chandra
NIM : 00000055770
Program Studi : Sistem Informasi
Fakultas : Fakultas Teknik & Informatika

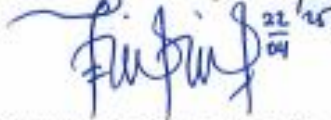
Telah diujikan pada hari Jumat, 11 April 2025

Pukul 13.00 s.d 15.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang



Dr. Imawati, S.Kom., M.M.S.I.
0805097703 / 081431

Penguji



Samuel Ady Sanjaya, S.T., M.T.
0305049402

Pembimbing



Rudi Sutomo, S.Kom., M.Si., M.Kom.
0222057501

Ketua Program Studi Sistem Informasi



Ririn Ikana Desanti, S.Kom., M.Kom

iv

Optimasi dan Pemilihan Algoritma..., Daffa Kaisha Pratama Chandra, Universitas Multimedia Nusantara

M U L T I M E D I A
N U S A N T A R A

iii

Optimasi dan Pemilihan Algoritma..., Daffa Kaisha Pratama Chandra, Universitas Multimedia Nusantara

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Daffa Kaisha Pratama Chandra
NIM : 00000055770
Program Studi : Sistem Informasi
Jenjang : D3/S1/S2* (pilih salah satu)
Judul Karya Ilmiah : OPTIMASI DAN PEMILIHAN ALGORITMA
MACHINE LEARNING UNTUK PREDIKSI
HARGA RUMAH DI KABUPATEN
TANGERANG

Menyatakan dengan sesungguhnya bahwa saya bersedia* (pilih salah satu):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 24 Maret 2025



(Daffa Kaisha Pratama Chandra)

KATA PENGANTAR

Puji dan syukur saya panjatkan kehadiran Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, sehingga tugas akhir dengan judul “Optimasi dan Pemilihan Algoritma *Machine Learning* untuk Prediksi Harga Rumah di Kabupaten Tangerang” dapat diselesaikan dengan tepat waktu. Skripsi ini diajukan sebagai salah satu syarat mencapai gelar Strata 1 Program Studi Sistem Informasi, Fakultas Teknik Informatika, Universitas Multimedia Nusantara.

Laporan skripsi ini dapat diselesaikan dengan bantuan dari berbagai pihak yang memberikan masukan dan dukungan kepada penulis. Maka dari itu, penulis ingin menyampaikan rasa terima kasih kepada:

1. Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Rudi Sutomo, S.Kom., M.Si., M.Kom., sebagai dosen pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat memberikan manfaat berupa informasi untuk ilmu pengetahuan dan sumber inspirasi bagi pembaca.

Tangerang, 24 Maret 2025



(Daffa Kaisha Pratama Chandra)

OPTIMASI DAN PEMILIHAN ALGORITMA MACHINE LEARNING UNTUK PREDIKSI HARGA RUMAH DI KABUPATEN TANGERANG

(Daffa Kaisha Pratama Chandra)

ABSTRAK

Tempat tinggal merupakan kebutuhan primer bagi setiap individu, dengan rumah sebagai salah satu jenis hunian utama yang paling diminati. Di Indonesia, mayoritas masyarakat masih lebih memilih untuk membeli rumah dibandingkan jenis hunian lainnya seperti apartemen. Di wilayah Tangerang, harga rumah terus mengalami peningkatan dari tahun ke tahun, namun hingga saat ini belum terdapat acuan standar mengenai harga rumah berdasarkan spesifikasinya. Oleh karena itu, dibutuhkan sebuah model yang mampu memprediksi harga rumah secara akurat dan dapat dijadikan sebagai acuan bagi calon pembeli dalam merencanakan pembelian properti. Penelitian ini bertujuan untuk membandingkan kinerja model regresi *machine learning* dalam memprediksi harga rumah berdasarkan data dari rumah123.com dengan menggunakan tiga algoritma, yaitu *Multiple Linear Regression*, *Random Forest* (dengan dan tanpa *tuning hyperparameter Optuna*), serta *XGBoost* (dengan dan tanpa *tuning hyperparameter Optuna*). Evaluasi dilakukan menggunakan koefisien determinasi (R^2), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Hasil analisis menunjukkan bahwa seluruh model XGBoost yang telah dioptimalkan menggunakan Optuna menunjukkan hasil terbaik dengan nilai R^2 sebesar 0,9098, RMSE sebesar 0,2450, dan MAPE sebesar 0,0084 pada data uji. Model ini terbukti efektif dalam menangkap hubungan non-linier yang kompleks pada data properti, sehingga menjadi model paling andal untuk estimasi harga rumah di Kabupaten Tangerang.

Kata kunci: *Hyperparameter Tuning*, *Machine Learning*, Prediksi Harga Rumah

UNIVERSITAS
MULTIMEDIA
NUSANTARA

OPTIMIZING AND MACHINE LEARNING ALGORITHM

SELECTION FOR HOUSE PRICE PREDICTION IN

TANGERANG REGENCY

(Daffa Kaisha Pratama Chandra)

ABSTRACT (English)

Housing is a primary need for every individual, with houses being one of the most preferred types of dwellings. In Indonesia, the majority of people still prefer buying a house over other types of residences such as apartments. In the Tangerang region, house prices continue to rise year by year; however, there is currently no standardized reference for house prices based on their specifications. Therefore, a model capable of accurately predicting house prices is needed as a reference for prospective buyers in planning property purchases. This study aims to compare the performance of regression-based machine learning models in predicting house prices using data from rumah123.com by applying three algorithms: Multiple Linear Regression, Random Forest (with and without hyperparameter tuning using Optuna), and XGBoost (with and without hyperparameter tuning using Optuna). Evaluation was conducted using three key metrics: coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The analysis results show that among all models, the XGBoost model optimized with Optuna achieved the best performance, with an R^2 value of 0.9098, an RMSE of 0.2450, and a MAPE of 0.0084 on the test data. This model has proven effective in capturing complex non-linear relationships in real estate data, making it the most reliable choice for estimating house prices in Kabupaten Tangerang.

Keywords: House Price Prediction, Hyperparameter Tuning, Machine Learning

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
ABSTRACT (English).....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah.....	6
1.4 Tujuan dan Manfaat Penelitian	6
1.4.1 Tujuan Penelitian	6
1.4.2 Manfaat Penelitian	7
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Penelitian Terdahulu	9
2.2 Objek Penelitian	13
2.2.1 Rumah	13
2.2.2 Sumber Data.....	14
2.3 Algoritma dan <i>Framework</i>	15
2.3.1 CRISP-DM.....	15
2.3.2 <i>Multiple Linear Regression</i>	16
2.3.3 <i>Random Forest</i>	17
2.3.4 <i>Extreme Gradient Boosting Tree (XGBoost)</i>	18
2.3.5 <i>Hyperparameter Tuning</i> dengan Optuna	20

2.3.6	Validasi Model	21
2.3.7	Metrik Evaluasi	21
2.4	<i>Tools</i> dan Teknik	23
2.4.1	Python	23
2.4.2	Jupyter	24
BAB III METODOLOGI PENELITIAN		26
3.1	Objek Penelitian	26
3.2	Metode Penelitian	26
3.2.1	Alur Penelitian	27
3.3	Variabel Penelitian	33
3.3.1	Variabel Dependen	33
3.3.2	Variabel Independen	33
3.4	Teknik Pengumpulan Data	33
3.5	Teknik Analisis Data	34
BAB IV ANALISIS DAN HASIL PENELITIAN		36
4.1	<i>Business Understanding</i>	36
4.2	<i>Data Understanding</i>	36
4.2.1	Pengumpulan Data	36
4.2.2	Analisis Fitur dan Pemilihan Fitur	39
4.3	<i>Data Preparation</i>	41
4.3.1	Konversi Tipe Data	41
4.3.2	<i>Feature Extraction</i>	43
4.3.3	<i>Filter Data</i>	44
4.3.4	<i>Outliers Handling</i>	46
4.3.5	<i>Encoding</i>	47
4.3.6	<i>Feature Selection & Data Splitting</i>	49
4.4	<i>Modeling</i>	55
4.4.1	<i>Modeling Base Model Multiple Linear Regression</i>	55
4.4.2	<i>Modeling Base Model Random Forest</i>	57
4.4.3	<i>Modeling Base Model XGBoost</i>	60
4.4.4	<i>Modeling Optuna Random Forest</i>	62
4.4.5	<i>Modeling Optuna XGBoost</i>	64

4.5	<i>Evaluation</i>	67
4.6	<i>Deployment</i>	68
4.7	Pembahasan Hasil dan Diskusi	79
BAB V		81
5.1	Simpulan	81
5.2	Saran.....	82
DAFTAR PUSTAKA		84
LAMPIRAN		90



DAFTAR TABEL

Tabel 1.1 Komparasi Performa Algoritma <i>Machine Learning</i>	5
Tabel 2.1 Tabel Penelitian Terdahulu Mengenai Algoritma <i>Machine Learning</i>	10
Tabel 2.2 Tabel Penelitian Terdahulu Mengenai Optuna	12
Tabel 2.2.3 Interpretasi Nilai R^2	22
Tabel 3.1 Perbandingan Metode <i>Data Mining</i>	26
Tabel 3.2 Variabel Data yang akan digunakan	29
Tabel 3.3 Perbandingan Kelebihan & Kekurangan Algoritma	35
Tabel 4.1 Selisih Harga Prediksi & Sebenarnya <i>Multiple Linear Regression</i>	56
Tabel 4.2 Performa Model <i>Multiple Linear Regression</i>	57
Tabel 4.3 Hyperparameter <i>Default Random Forest</i>	57
Tabel 4.4 Selisih Harga Prediksi & Sebenarnya <i>Base Model Random Forest</i>	59
Tabel 4.5 Performa <i>Base Model Random Forest</i>	59
Tabel 4.6 Hyperparameter <i>Default Base Model XGBoost</i>	60
Tabel 4.7 Selisih Harga Prediksi & Sebenarnya <i>Base Model XGBoost</i>	61
Tabel 4.8 Performa <i>Base Model XGBoost</i>	61
Tabel 4.9 <i>Optuna Random Forest Hyperparameter Value</i>	62
Tabel 4.10 Selisih Harga Prediksi & Sebenarnya <i>Optuna Random Forest</i>	63
Tabel 4.11 Performa Model <i>Optuna Random Forest</i>	64
Tabel 4.12 <i>Optuna XGBoost Hyperparameter Value</i>	64
Tabel 4.13 Selisih Harga Sebenarnya dan Prediksi <i>Optuna XGBoost</i>	66
Tabel 4.14 Performa <i>Optuna XGBoost</i>	66
Tabel 4.15 Perbandingan Performa dari Tiap Model	67
Tabel 4.16 Komparasi Hasil Penelitian dengan Penelitian Terdahulu	80



DAFTAR GAMBAR

Gambar 1.1 Perbandingan Responden Membeli Rumah/Tanah/Apartemen	2
Gambar 1.2 Jenis Tempat Tinggal yang Dipilih oleh Milenial (Februari 2022)	2
Gambar 1.3 Median Harga Rumah di Tangerang (Mei 2022-Mei 2023)	3
Gambar 2.1 Diagram PRISMA	9
Gambar 2.2 Siklus <i>CRISP-DM</i>	15
Gambar 2.3 <i>K-Fold Cross Validation</i>	21
Gambar 3.1 Alur Metode <i>CRISP-DM</i>	27
Gambar 3.2 <i>Flowchart</i> Tahap Data Understanding	28
Gambar 3.3 <i>Flowchart</i> Tahap Data Preparation	31
Gambar 3.4 <i>Flowchart</i> Tahap Modeling.....	32
Gambar 3.5 <i>Flowchart</i> Tahap <i>Evaluation</i>	32
Gambar 3.6 <i>Flowchart</i> Alur Teknik Pengumpulan Data	34
Gambar 4.1 Halaman Awal Rumah123.com	37
Gambar 4.2 Penyaringan Tipe Properti di Rumah123.com	37
Gambar 4.3 Melacak <i>Tag</i> HTML URL Spesifikasi Rumah	38
Gambar 4.4 Hasil <i>Web Scraping</i> URL Rumah	38
Gambar 4.5 Hasil <i>Web Scraping</i> Spesifikasi Rumah.....	39
Gambar 4.6 Sebaran Kolom Jumlah Lantai	39
Gambar 4.7 Sebaran Kolom <i>Carport</i>	40
Gambar 4.8 Sebaran Kolom Sertifikat Rumah	40
Gambar 4.9 Sebaran Kolom Kondisi Properti	41
Gambar 4.10 Kode Konversi Tipe Data Kolom Harga.....	42
Gambar 4.11 Kode Untuk Konversi Tipe Data Kolom Lainnya	43
Gambar 4.12 Kode untuk Mengekstrak Nama Kecamatan dari Kolom Lokasi ...	43
Gambar 4.13 Kode untuk Melakukan <i>Filter Data</i>	44
Gambar 4.14 Sebaran Jumlah Lantai, <i>Carport</i> , Sertifikat, Kondisi Properti	45
Gambar 4.15 Distribusi Sebelum <i>Log Tranformation</i>	46
Gambar 4.16 Kode untuk Melakukan <i>Log Transformation</i>	46
Gambar 4.17 Distribusi Setelah <i>Log Transformation</i>	47
Gambar 4.18 Kode untuk Melakukan <i>Encoding</i>	48
Gambar 4.19 Bentuk <i>Dataset</i> yang Akan Digunakan.....	48
Gambar 4.20 Kode untuk Membuat <i>Heatmap</i> Korelasi.....	49
Gambar 4.21 <i>Heatmap</i> Korelasi Antar Variabel Numerik.....	49
Gambar 4.22 Visualisasi Variabel Target vs Variabel Prediktor (Numerik)	50
Gambar 4.23 Visualisasi Variabel Prediktor dengan Korelasi Kuat.....	52
Gambar 4.24 Visualisasi Variabel Prediktor dengan Korelasi Lemah	52
Gambar 4.25 Visualisasi Variabel Target vs Variabel Prediktor (Kecamatan)	53
Gambar 4.26 Kecamatan dengan Harga Rumah Tertinggi - Terendah.....	54
Gambar 4.27 Kode untuk Melakukan <i>Data Splitting</i>	55
Gambar 4.28 Hasil Prediksi <i>Multiple Linear Regression</i>	56

Gambar 4.29 Hasil Prediksi Base Model <i>Random Forest</i>	58
Gambar 4.30 Hasil Prediksi Base Model <i>XGBoost</i>	60
Gambar 4.31 Hasil Prediksi Model <i>Optuna Random Forest</i>	63
Gambar 4.32 Hasil Prediksi Model <i>Optuna XGBoost</i>	65
Gambar 4.33 Kode Menyimpan Model	69
Gambar 4.34 Library Pembuatan <i>Website</i>	69
Gambar 4.35 Kode Konfigurasi Halaman <i>Website</i> dan <i>Tab</i> Navigasi.....	70
Gambar 4.36 Kode Pembuatan Halaman <i>Home</i>	70
Gambar 4.37 Halaman <i>Home</i>	71
Gambar 4.38 Kode Pembuatan Halaman <i>Disclaimer</i>	71
Gambar 4.39 Halaman <i>Disclaimer</i>	72
Gambar 4.40 Kode Pembuatan User Input Halaman <i>House Price Prediction</i>	73
Gambar 4.41 Kode Prediksi Halaman <i>House Price Prediction</i>	74
Gambar 4.42 Halaman <i>Prediction</i> (Sebelum Melakukan Prediksi).....	75
Gambar 4.43 Halaman <i>Prediction</i> (Setelah Melakukan Prediksi)	76
Gambar 4.44 Halaman <i>Prediction</i> (Kecamatan Tidak Terdaftar).....	76
Gambar 4.45 Kode Pembuatan Halaman <i>House Specs Prediction</i>	78
Gambar 4.46 Halaman <i>House Specs Prediction</i>	79



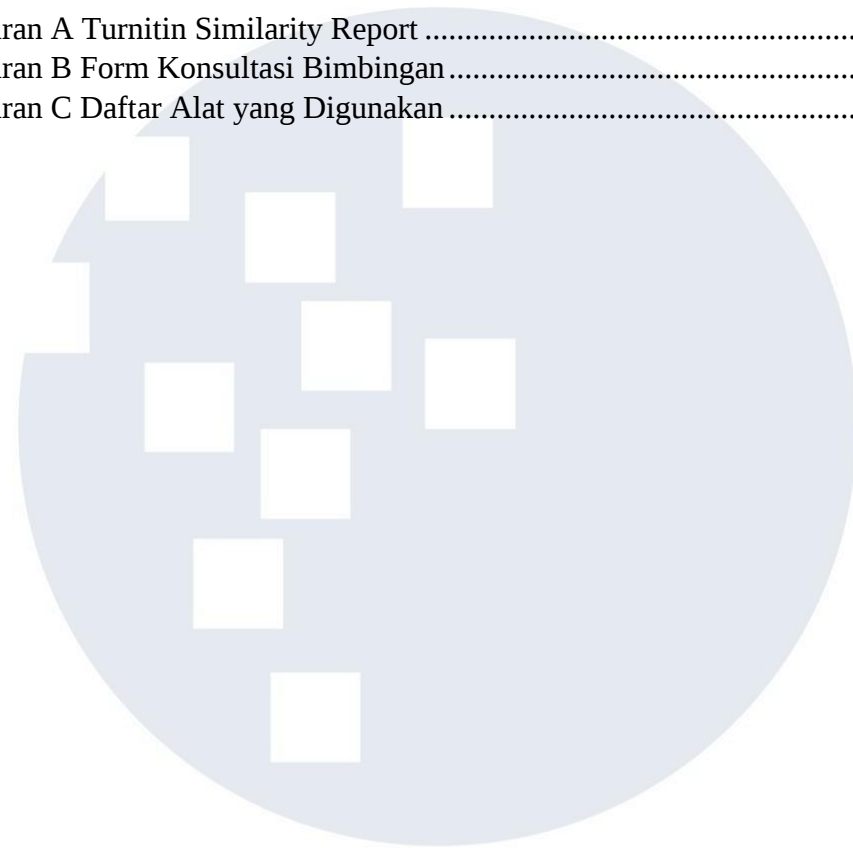
DAFTAR RUMUS

(2.1) Rumus <i>Multiple Linear Regression</i>	16
(2.2) Rumus Prediksi <i>Random Forest</i>	18
(2.3) Rumus Prediksi <i>XGBoost</i>	19
(2.4) Rumus Meminimalisir <i>Loss Function XGBoost</i>	19
(2.13) Rumus R^2	22
(2.14) Rumus MAPE.....	22
(2.15) Rumus RMSE.....	23



DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report	90
Lampiran B Form Konsultasi Bimbingan	97
Lampiran C Daftar Alat yang Digunakan	98



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA