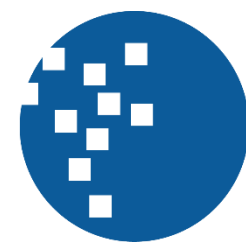


**PENERAPAN ALGORITMA MULTIPLE LINEAR
REGRESSION, GRADIENT BOOSTING, DAN RANDOM
FOREST DALAM MEMPREDIKSI HARGA PROPERTI DI
TANGERANG SELATAN**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

ADITYA MUHAMMAD SAPUTRO

00000047967

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

2025

**PENERAPAN ALGORITMA MULTIPLE LINEAR
REGRESSION, GRADIENT BOOSTING, DAN RANDOM
FOREST DALAM MEMPREDIKSI HARGA PROPERTI DI
TANGERANG SELATAN**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S. Kom)

ADITYA MUHAMMAD SAPUTRO

00000047967

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Aditya Muhammad Saputro

Nomor Induk Mahasiswa : 00000047967

Program studi : Sistem Informasi

Skripsi dengan judul:

PENERAPAN ALGORITMA MULTIPLE LINEAR REGRESSION, GRADIENT BOOSTING, DAN RANDOM FOREST DALAM MEMPREDIKSI HARGA PROPERTI DI TANGERANG SELATAN merupakan hasil karya saya sendiri bukan plagiat dari karya 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 skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 12 Juni 2025



(Aditya Muhammad Saputro)

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

Skripsi dengan judul

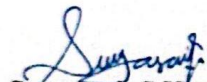
**PENERAPAN ALGORITMA MULTIPLE LINEAR REGRESSION,
GRADIENT BOOSTING, DAN RANDOM FOREST DALAM MEMPREDIKSI
HARGA PROPERTI DI TANGERANG SELATAN**

Oleh

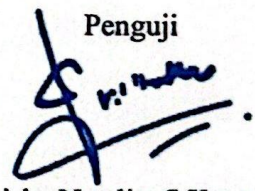
Nama : Aditya Muhammad Saputro
NIM : 00000047967
Program Studi : Sistem Informasi
Fakultas : Fakultas Teknik & Informatika

Telah diujikan pada hari Kamis, 12 Juni 2025
Pukul 15.00 s.d 17.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut.

Ketua Sidang


Suryasari, S.Kom., M.T.
0323088301/067345
25/6-2025.

Penguji


Dr. Friska Natalia, S.Kom., M.T.
306128307

Pembimbing


Ririn Ikana Desanti, S.Kom., M.Kom.
313058001/051314
25/6'25

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.
313058001/051314
25/6'25

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Aditya Muhammad Saputro
NIM : 00000047967
Program Studi : Sistem Informasi
Fakultas : S1
Judul Karya Ilmiah : PENERAPAN ALGORITMA MULTIPLE
LINEAR REGRESSION, GRADIENT BOOSTING, DAN RANDOM FOREST
DALAM MEMPREDIKSI HARGA PROPERTI DI TANGERANG SELATAN

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, 12 Juni 2025

Yang menyatakan,



(Aditya Muhammad Saputro)

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji Syukur kepada Tuhan Yang Maha Esa atas segala Rahmat, karunia, dan petunjuknya seiring dengan berjalannya penulisan dalam menyelesaikan laporan skripsi dengan judul “PENERAPAN ALGORITMA MULTIPLE LINEAR REGRESSION, GRADIENT BOOSTING, DAN RANDOM FOREST DALAM MEMPREDIKSI HARGA PROPERTI DI TANGERANG SELATAN” secara tepat waktu. Saya menyadari tanpa adanya bimbingan saya sangatlah kesulitan untuk menyelesaikan laporan skripsi ini. Oleh karena itu, saya ingin mengucapkan 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. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ririn Ikana Desanti, S.Kom., M.Kom., sebagai 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.
6. Teman-teman saya yang memberikan semangat dan dukungan selama proses penyusunan tugas akhir ini.

Semoga karya ilmiah ini dapat bermanfaat sebagai informasi untuk pengetahuan, sumber, dan referensi bagi pembaca.

Tangerang, 5 Juni 2025



Aditya Muhammad Saputro

PENERAPAN ALGORITMA MULTIPLE LINEAR REGRESSION, GRADIENT BOOSTING, DAN RANDOM FOREST DALAM MEMPREDIKSI HARGA PROPERTI DI TANGERANG SELATAN

(Aditya Muhammad Saputro)

ABSTRAK

Hunian merupakan kebutuhan dasar yang sangat penting bagi setiap individu, khususnya di Indonesia di mana rumah tapak masih menjadi pilihan utama masyarakat dibandingkan jenis hunian lain seperti apartemen. Salah satu wilayah dengan pertumbuhan permintaan dan harga properti yang tinggi adalah Kota Tangerang Selatan. Wilayah ini mengalami peningkatan signifikan dalam pembangunan perumahan, didukung oleh infrastruktur yang berkembang dan migrasi penduduk dari Jakarta. Namun, hingga saat ini belum tersedia standar yang jelas untuk memprediksi harga rumah berdasarkan spesifikasi dan lokasi. Oleh karena itu, tujuan dari penelitian ini adalah untuk menerapkan algoritma *machine learning* guna membangun model prediksi harga rumah yang akurat dan dapat dijadikan acuan oleh masyarakat, khususnya calon pembeli.

Penelitian menggunakan pendekatan CRISP-DM dengan data yang dikumpulkan melalui *web scraping* dari situs Lamudi.co.id. Fitur yang digunakan meliputi lokasi, luas tanah, luas bangunan, jumlah kamar, dan harga. Tiga algoritma diterapkan, yaitu *Multiple Linear Regression*, *Gradient Boosting*, dan *Random Forest*. Evaluasi dilakukan menggunakan metrik R^2 dan RMSE untuk menentukan model terbaik dalam memprediksi harga properti.

Hasil dari penelitian menunjukkan bahwa model *Random Forest* memberikan performa terbaik dengan nilai R^2 sebesar 0.8011 dan nilai RMSE terendah sebesar 536,831,054, menandakan tingkat akurasi yang tinggi dalam memprediksi harga rumah. Faktor-faktor yang paling berpengaruh dalam prediksi adalah luas bangunan dan lokasi properti. Model yang dikembangkan telah diimplementasikan dalam bentuk *prototype* berbasis *web* menggunakan *framework Flask*, yang memungkinkan pengguna untuk memprediksi harga rumah berdasarkan input spesifikasi tertentu.

Kata kunci: CRISP-DM, *Machine Learning*, Prediksi Harga Rumah, *Random Forest*, Tangerang Selatan.

Application of Multiple Linear Regression, Gradient Boosting, and Random Forest Algorithms in Predicting Property Prices in South Tangerang

(Aditya Muhammad Saputro)

ABSTRACT (English)

Housing is a fundamental need for every individual, especially in Indonesia, where landed houses remain the preferred choice over other types of residences such as apartments. One of the regions experiencing rapid growth in property demand and prices is South Tangerang City. This area has seen significant development in housing projects, supported by improving infrastructure and population migration from Jakarta. However, there is still no clear standard for predicting house prices based on specifications and location. Therefore, the purpose of this study is to apply machine learning algorithms to build an accurate house price prediction model that can serve as a reference for the public, particularly prospective buyers.

This research adopts the CRISP-DM approach, using data collected through web scraping from the Lamudi.co.id website. The features used include location, land area, building size, number of rooms, and price. Three algorithms are implemented: Multiple Linear Regression, Gradient Boosting, and Random Forest. Evaluation is conducted using the R^2 and RMSE metrics to determine the best-performing model for predicting property prices.

The results show that the Random Forest model delivers the best performance, with an R^2 value of 0.8011 and the lowest RMSE of 536,831,054, indicating a high level of accuracy in predicting house prices. The most influential factors in the prediction are building size and property location. The best-performing model has been implemented into a web-based prototype using the Flask framework, allowing users to predict house prices based on specific input criteria.

Keywords: *CRISP-DM, House Price Prediction, Machine Learning, Random Forest, South Tangerang.*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR	v
ABSTRAK	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR RUMUS	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	7
1.4 Tujuan dan Manfaat Penelitian	7
1.4.1 Tujuan Penelitian	7
1.4.2 Manfaat Penelitian	7
1.5 Sistematika Penulisan	8
BAB II LANDASAN TEORI	9
2.1 Penelitian Terdahulu	9
2.2 Objek Penelitian	16
2.3 Teori tentang Framework / Algoritma yang digunakan	17
2.4 Teori tentang Tools / Software yang digunakan	24
BAB III METODOLOGI PENELITIAN	26
3.1 Objek Penelitian	26
3.2 Metode Penelitian	26
3.2.1 <i>Business Understanding</i>	27
3.2.2 <i>Data Understanding</i>	27
3.2.3 <i>Data Preperation</i>	29

3.2.4	<i>Modeling</i>	30
3.2.5	<i>Evaluation</i>	31
3.2.6	<i>Deployment</i>	31
3.3	Teknik Pengumpulan Data	32
3.4	Teknik Analisis Data	33
BAB IV ANALISIS DAN HASIL PENELITIAN		35
4.1	<i>Business Understanding</i>	35
4.2	<i>Data Understanding</i>	35
4.3	<i>Data Preparation</i>	49
4.4	<i>Modeling</i>	55
4.4.1	<i>Multiple Linear Regression</i>	56
4.4.2	<i>Gradient Boosting</i>	60
4.4.3	<i>Random Forest</i>	68
4.5	<i>Evaluation</i>	76
4.6	<i>Deployment</i>	78
4.7	Hasil Penelitian	84
4.8	Pembahasan Hasil Penelitian	86
BAB V SIMPULAN DAN SARAN		88
5.1	Simpulan	88
5.2	Saran	89
DAFTAR PUSTAKA		90
LAMPIRAN		95



DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu Mengenai Algoritma Machine Learning ...	9
Tabel 2. 2 Interpretasi Nilai R^2	23
Tabel 3. 1 perbandingan Metode <i>Data Mining</i>	26
Tabel 3. 2 Perbandingan antara Lamudi dan Rumah123	28
Tabel 3. 3 Variabel Data yang akan digunakan	29
Tabel 3. 4 Perbandingan Kelebihan dan Kekurangan Algoritma	34
Tabel 4. 1 Hasil performa model <i>Multiple Linear Regression</i>	56
Tabel 4. 2 Hasil perbandingan prediksi model <i>Multiple Linear Regression</i>	57
Tabel 4. 3 Hasil parameter <i>cross-validation</i> model <i>Gradient Boosting</i>	62
Tabel 4. 4 Hasil performa model <i>Gradient Boosting</i>	62
Tabel 4. 5 Hasil perbandingan prediksi model <i>Gradient Boosting</i>	63
Tabel 4. 6 Hasil parameter <i>cross-validation</i> model <i>Random Forest</i>	70
Tabel 4. 7 Hasil performa model <i>Random Forest</i>	70
Tabel 4. 8 Hasil perbandingan prediksi model <i>Random Forest</i>	71
Tabel 4. 9 Perbandingan performa dari tiap model.....	76
Tabel 4. 10 Validasi prediksi model <i>Multiple Linear Regression</i>	81
Tabel 4. 11 Validasi prediksi model <i>Gradient Boosting</i>	82
Tabel 4. 12 Validasi prediksi model <i>Random Forest</i>	83
Tabel 4. 13 Hasil Penelitian	84
Tabel 4. 14 Hasil Akurasi	85



DAFTAR GAMBAR

Gambar 1. 1 Pertumbuhan Kuartal (QoQ) Proyek Perumahan di Kuartal 4 2023..	2
Gambar 2. 1 Siklus CRISP-DM	17
Gambar 3. 1 Proses CRISP-DM	27
Gambar 3. 2 <i>Flowchart</i> Tahap <i>Data Understanding</i>	28
Gambar 3. 3 <i>Flowchart</i> Tahap <i>Data Preparation</i>	30
Gambar 3. 4 <i>Flowchart</i> Tahap <i>Modeling</i>	30
Gambar 3. 5 <i>flowchart</i> Tahap <i>Evaluation</i>	31
Gambar 3. 6 <i>Flowchart</i> Teknik Pengumpulan Data	32
Gambar 4. 1 Halaman iklan rumah data <i>location</i> dan <i>price</i>	36
Gambar 4. 2 Halaman rincian iklan rumah	36
Gambar 4. 3 Halaman fasilitas <i>outdoor</i> iklan rumah	37
Gambar 4. 4 <i>Inspect element</i> halaman iklan rumah	38
Gambar 4. 5 Wilayah di Kota Tangerang Selatan.....	39
Gambar 4. 6 Jangkauan harga iklan rumah	39
Gambar 4. 7 <i>Library</i> dan <i>package web scraping</i>	40
Gambar 4. 8 Kode <i>webdriver-manager</i>	41
Gambar 4. 9 Kode <i>scraping link</i> href halaman hasil iklan rumah	41
Gambar 4. 10 Kode verifikasi data duplikat	42
Gambar 4. 11 Kode jeda waktu antar halaman	42
Gambar 4. 12 Kode <i>scraping</i> halaman iklan.....	43
Gambar 4. 13 Kode penyimpanan data	44
Gambar 4. 14 Diagram distribusi data wilayah.....	45
Gambar 4. 15 Diagram distribusi jumlah kamar tidur	46
Gambar 4. 16 Diagram distribusi jumlah kamar mandi.....	46
Gambar 4. 17 Diagram distribusi jumlah halaman parkir.....	47
Gambar 4. 18 Diagram distribusi luas bangunan	48
Gambar 4. 19 Diagram distribusi luas tanah.....	48
Gambar 4. 20 Kode penggabungan <i>csv</i>	49
Gambar 4. 21 Hasil <i>dataframe</i> penggabungan <i>file csv</i>	50
Gambar 4. 22 Informasi <i>dataset</i>	50
Gambar 4. 23 Total duplikat	51
Gambar 4. 24 Kode transformasi <i>null value</i>	51
Gambar 4. 25 Kode <i>transform</i> data <i>price</i>	52
Gambar 4. 26 Kode <i>OneHotEncoder</i>	53
Gambar 4. 27 Kode <i>upper limits</i>	54
Gambar 4. 28 Hasil akhir setelah <i>cleaning</i>	54
Gambar 4. 29 List kolom pada data x dan y	55
Gambar 4. 30 Perbandingan rata-rata per- <i>location Multiple Linear Regression</i> ..	58
Gambar 4. 31 <i>Scatter Plot</i> harga prediksi dan aktual <i>Multiple Linear Regression</i>	59
Gambar 4. 32 Grafik perbandingan prediksi 100 data pertama <i>Multiple Linear</i> ..	60
Gambar 4. 33 Grafik perbandingan prediksi <i>Multiple Linear Regression</i>	60

Gambar 4. 34 Kode <i>cross-validation</i> model <i>Gradient Boosting</i>	61
Gambar 4. 35 Perbandingan rata-rata per- <i>location</i> <i>Gradient Boosting</i>	64
Gambar 4. 36 <i>Scatter Plot</i> harga prediksi dan aktual <i>Gradient Boosting</i>	65
Gambar 4. 37 Grafik perbandingan prediksi 100 data pertama <i>Gradient Boost</i> ...	66
Gambar 4. 38 Grafik perbandingan prediksi <i>Gradient Boosting</i>	66
Gambar 4. 39 Grafik <i>feature importance</i> model <i>Gradient Boosting</i>	67
Gambar 4. 40 <i>Scatter plot</i> harga prediksi dan luas bangunan <i>Gradient Boosting</i>	68
Gambar 4. 41 Kode <i>cross-validation</i> model <i>Random Forest</i>	69
Gambar 4. 42 Perbandingan rata-rata per- <i>location</i> <i>Random Forest</i>	72
Gambar 4. 43 <i>Scatter plot</i> harga prediksi dan aktual <i>Random Forest</i>	73
Gambar 4. 44 Grafik perbandingan prediksi 100 data pertama <i>Random Forest</i> ..	74
Gambar 4. 45 Grafik perbandingan prediksi <i>Random Forest</i>	74
Gambar 4. 46 Grafik <i>feature importance</i> model <i>Random Forest</i>	75
Gambar 4. 47 <i>Scatter plot</i> harga prediksi dan luas bangunan <i>Random Forest</i>	76
Gambar 4. 48 <i>Import</i> model pada <i>website</i>	78
Gambar 4. 49 Halaman <i>home</i>	79
Gambar 4. 50 Halaman <i>prediction</i>	80
Gambar 4. 51 Hasil percobaan model <i>Multiple Linear Regression</i>	81
Gambar 4. 52 Hasil percobaan model <i>Gradient Boosting</i>	82
Gambar 4. 53 Hasil percobaan model <i>Random Forest</i>	83



DAFTAR RUMUS

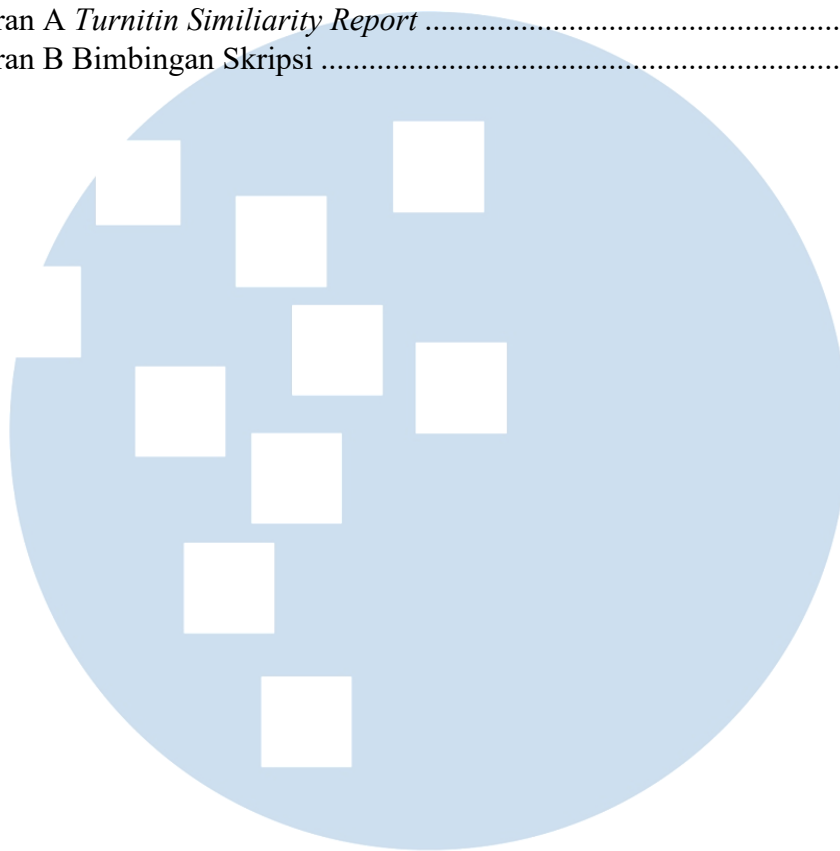
2. 1 Rumus <i>Multiple Linear Regression</i>	19
2. 2 Rumus Inisialisasi <i>Gradient Boosting</i>	20
2. 3 Rumus Kalkulasi <i>Gradient Boosting</i>	20
2. 4 Rumus Akhir <i>Gradient Boosting</i>	20
2. 5 Rumus <i>Random Forest</i>	21
2. 6 Rumus R-Squared	22
2. 7 Rumus RMSE	23

UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A <i>Turnitin Similiarity Report</i>	95
Lampiran B Bimbingan Skripsi	101



UMN

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