

**PENERAPAN METODE *RANDOM FOREST*
REGRESSION UNTUK MEMPREDIKSI HARGA
RUMAH DI JAKARTA**

Skripsi

**Disusun untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Statistika**



*Mencerdaskan dan
Memartabatkan Bangsa*

Disusun Oleh:

Muhammad Sandi Kuncoro

1314619029

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA**

2024

ABSTRAK

MUHAMMAD SANDI KUNCORO. Penerapan Metode *Random Forest Regression* Untuk Memprediksi Harga Rumah di Jakarta. Skripsi, Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. Juli 2024.

Jumlah penduduk di kota Jakarta selalu mengalami peningkatan dalam beberapa tahun terakhir. Pada tahun 2021, jumlah penduduk yang tinggal di Kota Jakarta adalah 10.605.437 jiwa. Angka ini kemudian meningkat menjadi 10.640.007 jiwa di tahun 2022, menunjukkan kenaikan sebesar 0,33%. Tren peningkatan ini berlanjut di tahun 2023, di mana jumlah penduduk Kota Jakarta mencapai 10.672.100 jiwa, atau naik 0,30% dibandingkan tahun 2022. Jumlah penduduk yang terus meningkat membawa dampak besar pada kebutuhan dasar manusia. Salah satu kebutuhan primer yang paling terpengaruh adalah rumah. Sehingga prediksi terhadap harga rumah akan sangat berguna bagi produsen maupun konsumen. Teknologi yang semakin canggih menjadi sarana yang tepat dan efisien dalam mengiklankan rumah yang dijual sehingga teknologi dimanfaatkan oleh para investor atau warga yang menjual rumah sebagai tempat iklan salah satunya situs *online OLX.co.id*. Tujuan penelitian ini adalah untuk memprediksi harga rumah di Jakarta dengan metode *Random Forest Regression* dan membandingkan hasilnya dengan metode Regresi Linier Berganda. Variabel yang digunakan yaitu luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, jumlah lantai, dan lokasi. Hasil dari penelitian ini, *Random Forest Regression* mendapat skor akurasi 82%, lebih tinggi dibandingkan Regresi Linier Berganda yang hanya mendapat skor akurasi 59% untuk memprediksi harga rumah di Jakarta. Berdasarkan *Random Forest Regression*, faktor yang dianggap paling penting adalah luas tanah dan luas bangunan. Sementara faktor lainnya mendapat tingkat kepentingan fitur dibawah 5%.

Kata Kunci: Harga Rumah, Kepentingan Fitur, Prediksi, *Random Forest Regression*.

ABSTRACT

MUHAMMAD SANDI KUNCORO. Application of the Random Forest Regression Method to Predict House Prices in Jakarta. Thesis, Study Program Statistics, Faculty of Mathematics and Natural Sciences, State University Jakarta. July 2024.






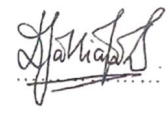
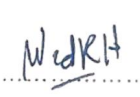
The population of Jakarta has always increased in recent years. In 2021, the population living in Jakarta was 10,605,437 people. This figure then increased to 10,640,007 people in 2022, showing an increase of 0.33%. This increasing trend continues in 2023, where the population of Jakarta will reach 10,672,100 people, or an increase of 0.30% compared to 2022. The increasing population has a major impact on basic human needs. One of the most affected primary needs is housing. So that predictions of house prices will be very useful for both producers and consumers. Increasingly sophisticated technology is an appropriate and efficient means of advertising houses for sale so that technology is utilized by investors or residents who sell houses as a place to advertise, one of which is the online site OLX.co.id. The purpose of this study is to predict house prices in Jakarta using the Random Forest Regression method and compare the results with the Multiple Linear Regression method. The variables used are land area, building area, number of bedrooms, number of bathrooms, number of floors, and location. The results of this study, Random Forest Regression got an accuracy score of 82%, higher than Multiple Linear Regression which only got an accuracy score of 59% to predict house prices in Jakarta. Based on Random Forest Regression, the factors considered to be the most important are land area and building area. Meanwhile, other factors received a feature importance level below 5%.

Keywords: Feature Importance, House Prices, Prediction, Random Forest Regression.

**LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI
PENERAPAN METODE *RANDOM FOREST REGRESSION*
UNTUK MEMPREDIKSI HARGA RUMAH DI JAKARTA**

Nama : Muhammad Sandi Kuncoro

NIM : 1314619029

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab Dekan	: <u>Prof. Dr. Muktiningsih N., M.St.</u> NIP. 196405111989032001		
Wakil Penanggung Jawab Wakil Dekan I	: <u>Dr. Esmar Budi, S.Si., MT</u> NIP. 197207281999031002		
Ketua Penguji	: <u>Dr. Dian Handayani, M.Si</u> NIP. 197404151998032001		23/24 /07
Sekretaris	: <u>Faroh Ladayya, M.Si.</u> NIP. 199401282020122018		24/24 /07
Anggota Pembimbing I	: <u>Prof. Dr. Suyono, M.Si.</u> NIP. 195712181993031005		23/24 /07
Pembimbing II	: <u>Dania Siregar, S.Stat., M.Si.</u> NIDN. 0027029106		22/24 /07
Penguji Ahli	: <u>Dra. Widianti Rahayu, M.Si</u> NIP. 196611032001122001		24/24 /07



Dinyatakan lulus ujian skripsi tanggal: 23 Juli 2024

LEMBAR PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul **“Penerapan Metode *Random Forest Regression* untuk Memprediksi Harga Rumah di Jakarta”**. yang disusun sebagai salah satu syarat memperoleh gelar Sarjana Statistika dari Program Studi Statistika Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari dosen pembimbing.

Sumber informasi yang diperoleh dari penulis lain yang telah dipublikasikan yang disebutkan dalam teks skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidah dan etika penulisan ilmiah pada umumnya dan ketentuan yang berlaku di Universitas Negeri Jakarta.

Jika dikemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sanding dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Jakarta, 19 Juli 2024



Muhammad Sandi Kuncoro



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 : Muhammad Sandi Kuncoro

NIM : 1314619029

Fakultas/Prodi : Fakultas Matematika dan Ilmu Pengetahuan Alam

Alamat email : muhammadsandikuncoro_1314619029@mhs.unj.ac.id

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 :

Penerapan Metode Random Forest Regression Untuk Memprediksi Harga Rumah Di Jakarta

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, 29 Juli 2024

Penulis

(Muhammad Sandi Kuncoro)

KATA PENGANTAR

Alhamdulillah rabbil'aalaamiin, segala puji dan syukur penulis panjatkan kepada Allah SWT karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi ini dengan judul “Penerapan Metode *Random Forest Regression* untuk Memprediksi Harga Rumah di Jakarta”. yang disusun sebagai salah satu syarat untuk memperoleh gelar sarjana statistika di Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta.

Penulis menyadari bahwa selama penyelesaian skripsi ini mendapatkan banyak bantuan, doa, dan dukungan, dari berbagai pihak. Oleh karena itu, izinkan penulis menyampaikan ucapan rasa terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua penulis, Bapak Andi Alfian dan Ibu Susanti yang selalu memotivasi, memberi dukungan, doa, dan memfasilitasi segala kebutuhan selama kuliah.
2. Ibu Dr. Dian Handayani, M.Si. selaku Koordinator Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Jakarta.
3. Bapak Prof. Dr. Suyono, M.Si. dan Ibu Dania Siregar, S.Stat, M.Si. selaku Dosen Pembimbing I dan II yang telah meluangkan waktunya untuk membimbing, mengarahkan, serta memberi dorongan sehingga skripsi ini dapat menjadi lebih baik dan terarah.
4. Seluruh Bapak/Ibu dosen pengajar dan Staf Administrasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Jakarta yang turut membantu proses penyusunan skripsi ini.
5. Teman-teman mahasiswa/i Program Studi Statistika UNJ angkatan 2019 yang selalu memotivasi, memberikan dukungan, bantuan, dan memberikan semangat kepada penulis selama penyelesaian skripsi ini.
6. Semua pihak yang tidak bisa disebutkan satu persatu karena telah membantu baik secara langsung maupun tidak langsung dalam penyusunan skripsi ini.

Penulis menyadari bahwa penulisan skripsi ini masih jauh dari kesempurnaan, dikarenakan adanya keterbatasan ilmu. Oleh karena itu, penulis menerima segala masukan, saran, serta kritik yang membangun untuk pengembangan pengetahuan

penulis di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat kepada penulis dan semua pihak yang membaca.

Jakarta, 19 Juli 2024



Muhammad Sandi Kuncoro



DAFTAR ISI

ABSTRAK	iii
DAFTAR ISI	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
BAB II KAJIAN PUSTAKA	6
2.1 Web Scraping.....	6
2.2 Preprocessing Data	8
2.3 Regresi Linier Berganda.....	8
2.4 Penduga Parameter Regresi Linier Berganda.....	9
2.5 Variabel Dummy Pada Model Regresi	10
2.6 Uji Asumsi.....	11
2.6.1 Uji Normalitas.....	11
2.6.2 Uji Homoskedastisitas.....	11
2.6.3 Uji Multikolinieritas.....	12
2.7 Uji-t Parsial.....	13
2.8 Machine Learning.....	13
2.9 Classification And Regression Tree (CART)	15
2.10 Ensemble Learning.....	18
2.11 Bootstrap Aggregation (Bagging)	19

2.12	Random Forrest Regression.....	20
2.13	Hyperparameter Tuning	22
2.14	K-Fold Cross-Validation	23
2.15	Feature Importance	24
2.16	Evaluasi Model	25
2.16.1	R-squared (R^2).....	25
2.16.2	Mean Square Error (MSE)	26
2.16.3	Mean Absolute Error (MAE)	26
2.16.4	Mean Absolute Percentage Error (MAPE).....	26
2.17	Rumah.....	27
2.17.1	Fungsi Rumah	27
2.17.2	Spesifikasi Rumah.....	27
2.18	Situs Jual Beli Rumah.....	28
BAB III		30
Metodologi Penelitian		30
3.1	Data Penelitian	30
3.2	Variabel Penelitian.....	30
3.3	Analisis Data	31
BAB IV Hasil dan Pembahasan		35
4.1	Exploratory Data Analysis (EDA).....	35
4.2	Penduga Parameter Regresi Linier Berganda.....	39
4.3	Uji Asumsi Klasik Regresi Linier Berganda	40
4.3.1	Uji Normalitas	40
4.3.2	Uji Homoskedastisitas.....	41
4.3.3	Uji Multikolinieritas.....	41
4.4	Hasil Fitting Model Regresi Linier Berganda	41

4.5	Hasil Prediksi RFR.....	42
4.6	Feature Importance.....	43
BAB V PENUTUP.....		44
5.1	Kesimpulan.....	44
5.2	Saran.....	44
DAFTAR PUSTAKA.....		45
LAMPIRAN.....		49



DAFTAR TABEL

Tabel 3.1 Variabel Penelitian	31
Tabel 4.1 Statistik Deskriptif Untuk Setiap Variabel Numerik	36
Tabel 4.2 Harga Rata-rata, Minimum, dan Maksimum Berdasarkan Lokasi	36
Tabel 4.3 Hasil Estimasi Parameter Regresi Linier Berganda.....	39
Tabel 4.4 Uji Multikolinieritas	41
Tabel 4.5 Evaluasi Model Regresi Linier Berganda.....	42
Tabel 4.6 Evaluasi Metrik RFR	43



DAFTAR GAMBAR

Gambar 2.1 Decision Tree	15
Gambar 2.2 Bootstrap Agreggating (Bagging).....	19
Gambar 2.3 Random Forest.....	21
Gambar 2.4 Logo OLX.....	29
Gambar 3.1 <i>Flowchart</i> Penelitian.....	34
Gambar 4.1 Jumlah Rumah Dijual Berdasarkan Lokasi	35
Gambar 4.2 Distribusi Harga Rumah Berdasarkan Lokasi	37
Gambar 4.3 Scatter Plot Variabel Bebas Dengan Variabel Terikat.....	38
Gambar 4.4 Metrik Korelasi Antar Variabel Numerik.....	39
Gambar 4.5 Plot Distribusi Dan Residu	40
Gambar 4.6 Feature Importance	43



Lampiran

Lampiran 1. Syntax analisis	49
-----------------------------------	----

