

**ANALISIS PERBANDINGAN *ARTIFICIAL NEURAL NETWORK* (ANN)
DAN *RANDOM FOREST* DALAM PREDIKSI HARGA RUMAH**

SKRIPSI

diajukan untuk memenuhi sebagian syarat penulisan Skripsi Akhir Studi
S-1 Program Studi Rekayasa Perangkat Lunak



Oleh:

Abid Mafahim

2000649

PROGRAM STUDI REKAYASA PERANGKAT LUNAK

KAMPUS UPI DI CIBIRU

UNIVERSITAS PENDIDIKAN INDONESIA

2024

ANALISIS PERBANDINGAN ARTIFICIAL NEURAL NETWORK (ANN)
DAN RANDOM FOREST DALAM PREDIKSI HARGA RUMAH

Oleh
Abid Mafahim
2000649

Diajukan untuk memenuhi syarat mendapatkan gelar Sarjana Komputer Program
Studi Rekayasa Perangkat Lunak

©Abid Mafahim
Universitas Pendidikan Indonesia
2024

Hak cipta dilindungi Undang-Undang
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difotokopi, atau cara lainnya tanpa ijin dari penulis

LEMBAR PENGESAHAN

ABID MAFAHIM

**ANALISIS PERBANDINGAN ARTIFICIAL NEURAL NETWORK (ANN)
DAN RANDOM FOREST DALAM PREDIKSI HARGA RUMAH**

Disetujui dan disahkan oleh pembimbing:

Pembimbing I



Indira Syawanodya, M.Kom.

NIPT 920190219920423201

Pembimbing II



Yulia Retnowati, S.Pd., M.T.

NIPT 920230219960729201

Mengetahui

Ketua Program Studi Rekayasa Perangkat Lunak



M. Iqbal Ardimansyah, S.T., M.Kom.

NIPT 920190219910328101

PERNYATAAN KEASLIAN SKRIPSI DAN BEBAS PLAGIARISME

Dengan ini saya menyatakan bahwa skripsi dengan judul ANALISIS PERBANDINGAN ARTIFICIAL NEURAL NETWORK (ANN) DAN RANDOM FOREST DALAM PREDIKSI HARGA RUMAH ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, 9 Agustus 2024

Yang membuat pernyataan



Handwritten signature of Abid Mafahim.

Abid Mafahim

2000649

UCAPAN TERIMA KASIH

Segala puji dan syukur ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga penelitian ini berhasil diselesaikan sebagai salah satu syarat untuk menyelesaikan Program Studi Rekayasa Perangkat Lunak, Universitas Pendidikan Indonesia. Shalawat dan salam juga senantiasa terlimpah kepada Nabi Muhammad SAW, yang telah membimbing umat menuju jalan yang lebih baik.

Pada kesempatan yang berharga ini, ungkapan terima kasih yang tulus disampaikan kepada berbagai pihak yang telah berperan dalam memberikan dorongan dan dukungan, sehingga penelitian ini dapat diselesaikan dengan baik. Dukungan yang diberikan tidak hanya berupa saran atau rekomendasi yang sangat berharga dalam setiap tahap penelitian, tetapi juga dalam bentuk dukungan emosional yang memberikan kekuatan dan semangat. Adapun pihak-pihak yang dimaksud antara lain:

1. Bapak M. Iqbal Ardiansyah, S.T., M.Kom. selaku kepala Program Studi Rekayasa Perangkat Lunak yang sudah memberikan kontribusinya dalam memberikan dukungan kepada mahasiswa RPL.
2. Ibu Asyifa I. Septiana, S.Pd., M.Eng. selaku dosen pembimbing akademik yang telah membantu dalam memberikan bimbingan secara non-akademik dan akademik.
3. Ibu Indira Syawanodya, M.Kom., selaku dosen pembimbing skripsi pertama yang sudah memberikan banyak saran dan rekomendasi serta dukungan lainnya yang sangat membantu untuk selesainya penyusunan penelitian ini.
4. Ibu Yulia Retnowati, S.Pd., M.T., sebagai dosen pembimbing skripsi kedua, yang telah memberikan berbagai rekomendasi dan saran berharga, sehingga karya ilmiah ini dapat tersusun dengan baik dan berkualitas.
5. Seluruh dosen Rekayasa Perangkat Lunak yang sudah memberikan ilmu-ilmu yang bermanfaat untuk penelitian ini melalui proses perkuliahan.
6. Keluarga penulis yang telah memberikan bantuan moral dan materi kepada penulis selama berlangsungnya pengerjaan skripsi ini.

7. Rekan-rekan seperjuangan khususnya kepada Sanjaya Wisnu, Rangga Kalam Sidiq, Dimas Aji Kurniawan, dan Adrian Sugandi Wijaya yang sudah memberikan dukungan dan bantuan yang berarti selama pengerjaan skripsi.
8. Abdul Faliq Habbi Nawa, Miftah Firdaus, dan Hasya sebagai teman yang sudah mendukung secara emosional selama berlangsungnya pengerjaan skripsi ini.

Dengan tersusunnya skripsi ini, besar harapan dapat bermanfaat bagi para pembacanya dan bagi yang sudah turut mendukung dalam berjalannya penelitian ini. Terima kasih.

Bandung, 9 Agustus 2024

Abid Mafahim
2000649

ABSTRAK

ANALISIS PERBANDINGAN *ARTIFICIAL NEURAL NETWORK* (ANN) DAN *RANDOM FOREST* DALAM PREDIKSI HARGA RUMAH

Abid Mafahim

2000649

Seiring perkembangan teknologi informasi, penerapan *machine learning* dalam industri properti rumah, khususnya untuk prediksi harga rumah, menjadi semakin penting dan akurasi pada prediksi yang lebih baik semakin dibutuhkan. Peran teknologi dapat membantu mempercepat dan meningkatkan akurasi dalam proses jual-beli properti. Maka dari itu, peran teknologi *machine learning* dapat dimanfaatkan untuk memenuhi kebutuhan meningkatkan akurasi prediksi harga rumah di kota besar negara berkembang seperti Kota Bandung. Penelitian ini bertujuan untuk menganalisis efektivitas algoritma *Artificial Neural Network* dan *Random Forest* dalam memprediksi harga rumah di Kota Bandung. Data yang digunakan adalah data penjualan rumah di Kota Bandung dengan mencakup luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, jumlah tempat parkir, dan lokasi kecamatannya. Analisis algoritma dilakukan dengan membandingkan hasil performa pengujian antara kedua algoritma dengan menggunakan metrik-metrik performa untuk model regresi seperti *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R Square* (R^2). Selain itu, penelitian ini juga menganalisis rasio data antara data *Training*, data validasi, dan data uji mana yang memiliki hasil terbaik. Hasil penelitian menunjukkan bahwa model dengan rasio data sebesar 60:20:20 menghasilkan performa yang paling baik pada kedua algoritma. Algoritma *Random Forest* menunjukkan performa yang lebih unggul dengan hasil MAE: 0.0470; MSE: 0.0079; RMSE: 0.0888; dan R^2 : 0.7085.

Kata kunci: *machine learning*, prediksi harga rumah, *Neural Network*, *Random Forest*.

ABSTRACT

COMPARATIVE ANALYSIS OF ARTIFICIAL NEURAL NETWORK (ANN) AND RANDOM FOREST IN HOUSE PRICE PREDICTION

Abid Mafahim

2000649

With the advancement of information technology, the application of machine learning in the housing industry, particularly for house price prediction, has become increasingly important, and the need for better accuracy in predictions is growing. Technology plays a crucial role in accelerating and enhancing accuracy in property transactions. Therefore, the role of machine learning technology can be leveraged to meet the demand for improving the accuracy of house price predictions in large cities of developing countries, such as Bandung. This research aims to analyze the effectiveness of the Artificial Neural Network (ANN) and Random Forest algorithms in predicting house prices in Bandung. The data used include house sales data in Bandung, covering land area, building area, number of bedrooms, number of bathrooms, number of parking spaces, and the subdistrict location. The analysis of the algorithms was conducted by comparing the testing performance of both algorithms using regression performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R Square (R^2). Additionally, this study also analyzed which data ratio among the training, validation, and test data provided the best results. The findings show that a data ratio of 60:20:20 yields the best performance for both algorithms, with the Random Forest algorithm outperforming the ANN with results of MAE: 0.0470, MSE: 0.0079, RMSE: 0.0888, and R^2 : 0.7085.

Keywords: *machine learning, house price prediction, Neural Network, Random Forest.*

DAFTAR ISI

LEMBAR PENGESAHAN	iii
PERNYATAAN KEASLIAN SKRIPSI DAN BEBAS PLAGIARISME	iv
UCAPAN TERIMA KASIH	v
ABSTRAK	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR FORMULA	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Penelitian	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Penelitian	4
1.6 Struktur Organisasi Skripsi	5
BAB II KAJIAN PUSTAKA.....	6
2.1 Pasar dan Harga Rumah	6
2.2 <i>Machine Learning</i>	7
2.2.1 <i>Supervised Learning</i>	7
2.2.2 <i>Unsupervised Learning</i>	8
2.2.3 <i>Semi-supervised Learning</i>	8
2.3 <i>Artificial Neural Networks (ANN)</i>	8
2.4 <i>Hyperparameter Tuning</i> Pada Model ANN.....	10
2.5 <i>Random Forest Regression</i>	12
2.6 <i>Overfitting dan Underfitting</i>	14
2.7 Penelitian Terkait	15
BAB III METODE PENELITIAN	21
3.1 Desain Penelitian.....	21
3.1.1 Klarifikasi Penelitian.....	22
3.1.2 Studi Deskriptif 1	22
3.1.3 Studi Perskriptif	22
3.1.4 Studi Deskriptif 2	26
3.2 <i>Dataset</i> Harga Rumah Di Kota Bandung.....	26

3.3	Instrumen Penelitian.....	26
3.3.1	<i>Mean Absolut Error (MAE)</i>	27
3.3.2	<i>Mean Squared Error (MSE)</i>	28
3.3.3	<i>Root Mean Squared Error (RMSE)</i>	28
3.3.4	<i>R Square (R²)</i>	29
3.4	Alat dan Bahan Penelitian.....	29
BAB IV TEMUAN DAN PEMBAHASAN		31
4.1	Pengolahan Data.....	31
4.1.1	Persiapan Data.....	31
4.1.2	<i>Data Cleaning</i>	33
4.1.3	<i>Data Splitting</i>	40
4.2	Pengembangan Model Prediksi Harga Rumah	41
4.2.1	Model ANN.....	41
4.2.2	Model <i>Random Forest</i>	50
4.3	Evaluasi Model Prediksi Harga Rumah	53
BAB V KESIMPULAN, IMPLIKASI, DAN REKOMENDASI		59
5.1	Kesimpulan	59
5.2	Implikasi.....	59
5.3	Rekomendasi.....	60
DAFTAR PUSTAKA.....		61

DAFTAR GAMBAR

Gambar 2.1 Cara kerja algoritma <i>supervised learning</i> (Mahesh, 2018).....	8
Gambar 2.2 Ilustrasi algoritma ANN beserta lapisan-lapisannya (Zhang, 2018).	10
Gambar 2.3 Ilustrasi Cara Kerja Random Forest Regression Dengan 25 Pohon (Harris & Grzes, 2019).....	13
Gambar 2.4 Grafik perbandingan setiap algoritma pada penelitian (Zong, 2023) dalam metrik performa model R^2 , MSE, dan MAE.	18
Gambar 3.1 Alur DRM	21
Gambar 3.2 Alur Kegiatan pada Tahapan Studi Perskriptif	23
Gambar 3.3 Diagram GQM dari instrumen penelitian yang digunakan	27
Gambar 4.1 <i>Dataset</i> Sebelum Ditambahkan Nama Kolom	31
Gambar 4.2 Contoh <i>Dataset</i> Setelah Ditambahkan Nama Kolom.....	33
Gambar 4.3 Diagram Proses <i>Data Cleaning</i> yang Dilakukan.	34
Gambar 4.4 Kolom-Kolom <i>Dataset</i> Setelah Proses Penghapusan Kolom	34
Gambar 4.5 Penghapusan Data <i>Null</i> Sebelum (Atas) dan Sesudah (Bawah)	35
Gambar 4.6 Hasil Pencarian Data Dengan Filter "tanah" Sebelum (a) dan Sesudah (b) Proses Penghapusan Data Hanya Jual Lahan	37
Gambar 4.7 Contoh Penghapusan Data yang Terduplikasi	38
Gambar 4.8 Diagram Perbandingan Antara Sebelum Pengurangan <i>Outlier</i> (a) dan Setelah Pengurangan <i>Outlier</i> (b).....	39
Gambar 4.9 Hasil Transformasi Data Kedalam Bentuk yang Ternormalisasi.....	40
Gambar 4.10 Grafik Hasil <i>Training</i> Pada Model ANN Rasio Data 80:10:10 Berdasarkan Rentang <i>Epochs</i> -nya.....	44
Gambar 4.11 Grafik Hasil <i>Training</i> Pada Model ANN Rasio Data 70:15:15 Berdasarkan Rentang <i>Epochs</i> -nya.....	46
Gambar 4.12 Grafik Hasil <i>Training</i> Pada Model ANN Rasio Data 60:20:20 Berdasarkan Rentang <i>Epochs</i> -nya.....	48
Gambar 4.13 Grafik Hasil <i>Training</i> Pada Model ANN Rasio Data 50:25:25 Berdasarkan Rentang <i>Epochs</i> -nya.....	50
Gambar 4.14 Diagram Tebar pada Model ANN Rasio Data 60:20:20	55
Gambar 4.15 Diagram Tebar pada Model <i>Random Forest</i> Rasio Data 60:20:20.	57

DAFTAR TABEL

Tabel 2.1. Pratinjau Penelitian Terkait.....	15
Tabel 2.2 Hasil penelitian yang dilakukan oleh Mansa dkk.	19
Tabel 3.1 Spesifikasi Perangkat yang Digunakan.....	29
Tabel 3.2 Perangkat lunak dan <i>library</i> yang digunakan	30
Tabel 4.1 Daftar Kolom yang Ditambahkan Pada <i>Dataset</i> Mentah Harga Rumah di Bandung	32
Tabel 4.2 Beberapa Contoh Nilai <i>building_area</i> (m2), <i>land_area</i> , dan <i>price</i> Pada <i>Dataset</i> Sebelum Dilakukan Mengubah Tipe Data.....	35
Tabel 4.3 Contoh Nilai <i>building_area</i> (m2), <i>land_area</i> , dan <i>price</i> Pada <i>Dataset</i> Setelah Dilakukan Mengubah Tipe Data.	36
Tabel 4.4 Contoh Data Dengan Lokasi yang Sudah Diubah Menjadi Bentuk <i>Number</i>	39
Tabel 4.5 Jumlah Setiap Jenis Data Berdasarkan Besaran Data <i>Training</i>	40
Tabel 4.6 Arsitektur Model ANN yang Dihasilkan Dari Proses <i>Hyperparameter Tuning</i>	42
Tabel 4.7 Hasil <i>Training</i> dan Validasi Pada Model ANN Rasio Data 80:10:10 .	43
Tabel 4.8 Hasil <i>Training</i> dan Validasi Pada Model ANN Rasio Data 70:15:15 .	45
Tabel 4.9 Hasil <i>Training</i> dan Validasi Pada Model ANN Rasio Data 60:20:20 .	47
Tabel 4.10 Hasil <i>Training</i> dan Validasi Pada Model ANN Rasio Data 50:25:25	49
Tabel 4.11 Hasil <i>Training</i> dan Validasi Pada Model <i>Random Forest</i> Rasio Data 80:10:10	51
Tabel 4.12 Hasil <i>Training</i> dan Validasi Pada Model <i>Random Forest</i> Rasio Data 70:15:15	51
Tabel 4.13 Hasil <i>Training</i> dan Validasi Pada Model <i>Random Forest</i> Rasio Data 60:20:20	52
Tabel 4.14 Hasil <i>Training</i> dan Validasi Pada Model <i>Random Forest</i> Rasio Data 50:25:25	53
Tabel 4.15 Hasil Pengujian Model ANN	54
Tabel 4.16 Hasil Pengujian Model <i>Random Forest</i>	56

DAFTAR FORMULA

Formula 3.1 Rumus <i>Mean Absolut Error</i> (MAE) (Rahayuningtyas dkk., 2021) .	27
Formula 3.2 Rumus <i>Mean Squared Error</i> (MSE) (Rahayuningtyas dkk., 2021) .	28
Formula 3.3 Rumus <i>Root Mean Squared Error</i> (RMSE) (Chai dan Draxler, 2014)	28
Formula 3.4 Rumus R square yang Digunakan Untuk Model Regresi (Plevris dkk., 2022)	29