



TESIS

OPTIMASI FEATURE ENGINEERING TERHADAP PERFORMA ALGORITMA XGBOOST, RANDOM FOREST DAN SUPPORT VECTOR REGRESSION DALAM PREDIKSI HARGA RUMAH

BRAHMANTIO WIDYO TRENGGONO

NPM 23066020011

DOSEN PEMBIMBING

Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU

Dr. Rr. Ani Dijah Rahajoe, ST, M.Cs.

KEMENTRIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM MAGISTER TEKNOLOGI INFORMASI

SURABAYA

2025



TESIS

OPTIMASI FEATURE ENGINEERING TERHADAP PERFORMA ALGORITMA XGBOOST, RANDOM FOREST DAN SUPPORT VECTOR REGRESSION DALAM PREDIKSI HARGA RUMAH

BRAHMANTIO WIDYO TRENGGONO

NPM 23066020011

DOSEN PEMBIMBING

Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU

Dr. Rr. Ani Dijah Rahajoe, ST, M.Cs.

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM MAGISTER TEKNOLOGI INFORMASI

SURABAYA

2025

TESIS

OPTIMASI FEATURE ENGINEERING TERHADAP PERFORMA ALGORITMA XGBOOST, RANDOM FOREST DAN SUPPORT VECTOR REGRESSION DALAM PREDIKSI HARGA RUMAH

Tesis

Sebagai Salah Satu Syarat untuk Mencapai Gelar Magister Program
Studi Teknologi Informasi

Disusun dan diajukan oleh

BRAHMANTIO WIDYO TRENGGONO
23066020011

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN “VETERAN” JAWA TIMUR
SURABAYA
2025

**OPTIMASI FEATURE ENGINEERING TERHADAP PERFORMA ALGORITMA
XGBOOST RANDOM FOREST DAN SUPPORT VECTOR REGRESSION DALAM
PREDIKSI HARGA RUMAH**

Disusun Oleh :
BRAHMANTIO WIDYO TRENGGONO
NPM 23066020011

Telah disetujui untuk mengikuti Ujian Tesis

Mengetahui

Mengetahui

**KOORDINATOR PRODI
MAGISTER TEKNOLOGI INFORMASI**


Dr. Eng. Agussalim, S.Pd., MT.
NIP. 19850811 2019031 005

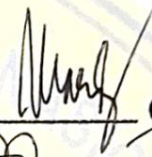
**OPTIMASI FEATURE ENGINEERING TERHADAP PERFORMA ALGORITMA
XGBOOST, RANDOM FOREST DAN SUPPORT VECTOR REGRESSION DALAM
PREDIKSI HARGA RUMAH**

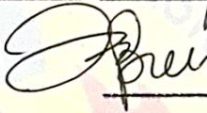
Oleh :

BRAHMANTIO WIDYO TRENGGONO
23066020011

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Ujian Tesis
Prodi Magister Teknologi Informasi Fakultas Ilmu Komputer Universitas
Pembangunan Nasional Veteran Jawa Timur
Pada Tanggal 29 Agustus 2025

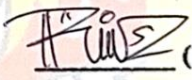
Menyetujui

Dr. Ir. I Gede Susrama Mas Diyasa, MT., IPU  (Pembimbing I)
NIP. 1970619 2021211 009

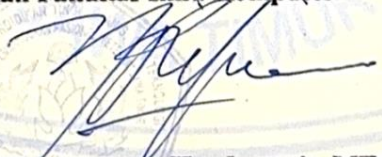
Dr. Rr. Ani Dijah Rahajoe, ST, MCs  (Pembimbing II)
NIP. 19730512 200501 2003

Dr. Basuki Rahmat, S.Si., M.T  (Penguji I)
NIP. 19690723 202121 1 002

Dr. Eng. Agussalim, S.Pd., M.T  (Penguji II)
NIP. 19850811 2019031 005

Dr. Rizky Parlika, S.Kom., M.Kom.  (Penguji III)
NIP. 19840518 2021211 003

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan dibawah ini:

Nama : Brahmantio Widy Trenggono
NPM : 23066020011
Program : Magister (S2)
Program Studi : Teknologi Informasi
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Tesis ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemulan indikasi plagiat pada Tesis ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 09 Desember 2025

Yang Membuat permyataan,



Brahmantio Widy Trenggono

23066020011

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur Alhamdulillah penulis panjatkan kehadiran Allah SWT, yang telah melimpahkan berkah, rahmat, dan hidayah-Nya, sehingga penulis dapat menyelesaikan tesis ini dengan baik dan tepat waktu. Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Teknologi Informasi. Adapun penelitian ini berfokus pada analisa pengaruh Optimasi Feature Engineering Terhadap Performa Algoritma XGBoost, Random Forest dan Support Vector Regression dalam Memprediksi Harga Rumah.

Gagasan utama dari penelitian ini adalah untuk memberikan solusi inovatif dalam mendukung sektor properti dengan memanfaatkan teknologi machine learning. Sebagaimana pentingnya efisiensi dan keakuratan dalam memprediksi harga rumah, penelitian ini mengembangkan algoritma Random Forest yang diharapkan dapat memberikan hasil prediksi yang lebih presisi dan andal. Dalam lingkup penelitian ini, pendekatan yang diambil bertujuan untuk mengatasi tantangan yang sering muncul seperti ketidakakuratan data dan tingginya tingkat kompleksitas perhitungan. Penelitian ini juga diharapkan dapat menjadi referensi bagi pengembangan aplikasi serupa di masa mendatang.

Bukan hal yang mudah untuk mewujudkan gagasan-gagasan tersebut ke dalam sebuah susunan tesis. Proses penyusunan ini membutuhkan pemikiran yang matang, usaha yang konsisten, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Orang tua dan orang terdekat yang senantiasa mendoakan, memberikan semangat atau dukungan dalam segala bentuk apresiasi dan memotivasi penulis.
2. Dr. I Gede Susrama Mas Diayasa, MT., IPU sebagai Dosen pembimbing pertama dan Dr. Rr. Ani Dijah Rahajoe, ST, M.Cs sebagai Dosen pembimbing kedua
3. Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.

4. Dr. Eng. Agussalimm S.Pd., MT, sebagai ketua penguji dan Koordinator Program Studi Magister Teknologi Informasi, Universitas Pembangunan Nasional “Veteran” Jawa Timur, atas kritik dan saran yang membangun dalam penelitian ini.
5. Dr. Basuki Rahmat, S.Si., M.T dan Dr. Rizky Parlika, S.Kom., M.Kom selaku dosen penguji atas kritik dan saran yang membangun dalam penyelesaian penelitian ini.
6. Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur dan Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memfasilitasi saya menempuh program magister serta para dosen dan rekan-rekan dalam tim penelitian.
7. Seluruh dosen dan staf administrasi program studi Magister Teknologi Informasi, yang telah membantu dalam berbagai aspek akademik dan administrasi selama masa studi.
8. Rekan-rekan kerja dan teman-teman yang telah memberikan dukungan, inspirasi, dan kerja sama selama masa studi dan proses penelitian ini.
9. Pihak-pihak lain yang tidak dapat disebutkan satu per satu, tetapi telah memberikan kontribusi dalam bentuk apapun demi terselesaikannya tesis ini.

Akhirnya, kepada kedua orang tua tercinta saya mengucapkan terima kasih dan sembah sujud atas doa, pengorbanan dan memotivasi mereka selama saya menempuh pendidikan. Penghargaan yang besar juga saya sampaikan kepada kedua orang tua saya atas motivasi dan dukungan yang tak ternilai.

Surabaya, 10 November 2025

Brahmantio Widyo

ABSTRAK

BRAHMANTIO WIDYO TRENGGONO. Optimasi Feature Engineering Terhadap Performa Algoritma XGBoost, Random Forest dan Support Vector Regression dalam Prediksi Harga Rumah (dibimbing oleh **Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU** dan **Dr. Rr. Ani Dijah Rahajoe, ST, M.Cs.**)

Semakin bertambahnya tahun pergerakan harga rumah yang selalu meningkat menjadi salah satu faktor penting dalam keputusan investasi dan perencanaan keuangan dalam menekan adanya inflasi. Namun, fluktuasi atau kenaikan harga rumah dapat disebabkan oleh berbagai faktor yang dapat mempengaruhi nilai dari prediksi harga rumah. Penelitian ini bertujuan untuk menganalisa pengaruh optimasi dan hubungan antara *feature engineering* dengan modeling dalam prediksi harga rumah. Tahapan penelitian meliputi *preprocessing data*, *feature transformation*, *feature engineering*, *splitting data* hingga pengoptimalan dalam penentuan parameter saat tuning. Dengan evaluasi kinerja model menggunakan metrik *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), dan *koefisien Determinasi* (R^2 -Squared). Hasil penelitian menunjukkan bahwa penggunaan algoritma *Support Vector Regression* menghasilkan performa terbaik dengan nilai MAE 278,662,143.00, RMSE sebesar 781,711,067.72, MAPE sebesar 6.79 dan R^2 -Squared 0.98. Penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya dalam optimasi model *regresi* terutama dalam pengambilan keputusan dalam memprediksi harga rumah yang lebih akurat.

Kata kunci: Prediksi Harga Rumah, Optimasi, *Feature engineering*

Halaman ini sengaja dikosongkan

ABSTRACT

BRAHMANTIO WIDYO TRENGGONO. Feature Engineering Optimization of XGBoost, Random Forest, and Support Vector Regression Algorithm Performance in House Price Prediction (supervised by **Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU** and **Dr. Rr. Ani Dijah Rahajoe, ST, M.Cs.**)

As the years go by, the ever-increasing movement of house prices has become an important factor in investment decisions and financial planning to curb inflation. However, fluctuations or increases in house prices can be caused by various factors that can affect the value of house price predictions. This study aims to analyze the influence of optimization and the relationship between feature engineering and modeling in house price predictions. The research stages include data preprocessing, feature transformation, feature engineering, data splitting, and optimization in determining parameters during tuning. Model performance is evaluated using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Determination coefficient (R^2 -Squared) metrics. The results show that the Support Vector Regression algorithm produces the best performance with a MAE value of 278,662,143.00, an RMSE of 781,711,067.72, a MAPE of 6.79, and R^2 -Squared of 0.98. This study is expected to be a reference for further research in regression model optimization, especially in decision making in predicting more accurate house prices.

Keyword: House Price Prediction, Optimization, Feature engineering

Halaman ini sengaja dikosongkan

DAFTAR ISI

HALAMAN JUDUL.....	i
PENGAJUAN JUDUL TESIS.....	ii
PERSETUJUAN TESIS.....	iii
PERNYATAAN KEASLIAN TESIS	v
KATA PENGANTAR.....	ix
ABSTRAK	xi
ABSTRACT	xiii
DAFTAR ISI.....	xv
DAFTAR TABEL	xviii
DAFTAR GAMBAR	xx
DAFTAR LAMPIRAN.....	xxi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	10
1.3 Tujuan Penelitian.....	10
1.4 Manfaat Penelitian.....	11
1.5 Ruang Lingkup.....	11
BAB II TINJAUAN PUSTAKA.....	12
2.1 Dasar Teori.....	13
2.1.1 <i>Data Mining</i>	13
2.1.2 <i>Machine Learning</i>	14
2.1.3 <i>Random Forest</i>	14
2.1.4 MSE (<i>Mean Squared Error</i>)	18
2.1.5 MAE (<i>Mean Absolute Error</i>)	18
2.1.6 MAPE (<i>Mean Absolute Percentage Error</i>).....	19

2.1.7	<i>R² (R-Squared)</i>	19
2.1.8	<i>Extreme Gradient Boosting (XGBoost)</i>	20
2.1.9	<i>Support Vector Regression (SVR)</i>	23
2.1.10	<i>Exploratory Data Analysis (EDA)</i>	25
2.1.11	<i>Feature Importance</i>	25
2.1.12	<i>Feature Combination</i>	25
2.1.13	<i>Feature Transformation</i>	26
2.1.14	<i>Cross-Validation</i>	26
2.1.15	<i>Hyperparameter Tuning</i>	27
2.2	Penelitian Terdahulu	27
BAB III METODOLOGI PENELITIAN		35
3.1	Identifikasi Masalah	36
3.2	Studi Literatur	36
3.3	Pengumpulan Data	37
3.4	Optimasi implementasi algoritma	38
3.4.1	<i>Data Preprocessing</i>	39
3.4.2	<i>Data Cleaning</i>	39
3.4.3	<i>Exploratory Data Analysis (EDA)</i>	39
3.4.4	<i>Feature Transformation</i>	40
3.4.5	<i>Feature Combination</i>	41
3.4.6	<i>Feature Importance</i>	42
3.4.6	Pembagian data (<i>Data Split</i>)	42
3.4.7	<i>Modeling</i>	43
3.4.8	<i>Cross-Validation</i>	45
3.4.9	<i>Hyperparameter Tuning</i>	45
3.4.10	Hasil Evaluasi	46
BAB IV HASIL DAN PEMBAHASAN		48
4.1	Karakteristik Dataset	49
4.2	Karakteristik Fitur	49

4.3 Optimasi Algoritma.....	50
4.3.1 Pemilihan Algoritma	50
4.3.2 <i>Preprocessing Data</i>	51
4.3.3 <i>Feature Engineering</i>	64
4.3.4 <i>Modeling</i>	71
4.3.5 Hasil Optimasi.....	82
BAB V PENUTUP.....	88
5.1 Kesimpulan.....	89
5.2 Saran.....	90
DAFTAR PUSTAKA	91
LAMPIRAN	91

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu.....	27
Tabel 3.1 Skema Splitting Data.....	43
Tabel 4.1 Hasil Akurasi Awal Random Forest	72
Tabel 4.2 Hasil Akurasi Terbaik Random Forest.....	74
Tabel 4.3 Hasil Akurasi Awal XGBoost	76
Tabel 4.4 Hasil Akurasi Terbaik XGBoost	77
Tabel 4.5 Hasil Akurasi Awal SVR	79
Tabel 4.6 Hasil Akurasi Terbaik SVR	80
Tabel 4.7 Hasil Akurasi Terbaik 3 Algoritma.....	83

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 1.1 Survey responden membeli rumah	2
Gambar 1.2 Generasi Z dan Milenial memiliki rumah	2
Gambar 1.3 Kenaikan harga rumah seken	3
Gambar 2.1 Algoritma Random Forest	15
Gambar 2.2 contoh tampilan data	16
Gambar 2.3 Algoritma XGBoost Regression.....	20
Gambar 2.4 contoh tampilan entry data	21
Gambar 2.5 langkah awal perhitungan algoritma XGBoost	21
Gambar 2.6 langkah kedua perhitungan algoritma XGBoost	22
Gambar 2.7 langkah ketiga perhitungan algoritma XGBoost	22
Gambar 2.8 langkah akhir perhitungan algoritma XGBoost.....	23
Gambar 2.9 Algoritma Support Vector Regression	24
Gambar 3.1 Tahapan penelitian	35
Gambar 3.2 Proses Optimasi Model	38
Gambar 3.3 Cross validation	45
Gambar 4.1 info tipe data	51
Gambar 4.2 Missing value	52
Gambar 4.3 Dataset	53
Gambar 4.4 Fasilitas perumahan	55
Gambar 4.5 Tampilan Data Rumah.....	56
Gambar 4.5 Distribusi Harga	58
Gambar 4.7 Deteksi Outlier Dataset	59
Gambar 4.8 Outlier Dataset.....	60
Gambar 4.9 Multicollinearity Check.....	63
Gambar 4.10 VIF Score	64
Gambar 4.10 Feature Importance.....	67
Gambar 4.11 Accumulative Importance	68
Gambar 4.12 Distribusi y_test_log.....	70
Gambar 4.13 Distribusi y_train_log.....	71
Gambar 4.14 Scatter Plot Random Forest.....	75
Gambar 4.15 Scatter Plot XGBoost	78
Gambar 4.16 Scatter Plot SVR.....	82

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

Lampiran 1.	Implementasi Algoritma Terpilih.....	91
Lampiran 2.	Implementasi Hasil.....	97

Halaman ini sengaja dikosongkan