

**EVALUASI MODEL MACHINE LEARNING
BERBASIS ENSEMBLE LEARNING
DALAM MEMPREDIKSI HARGA RUMAH
SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

DAVID YUSUF DE WASKITO

21.12.1974

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**EVALUASI MODEL MACHINE LEARNING
BERBASIS ENSEMBLE LEARNING
DALAM MEMPREDIKSI HARGA RUMAH**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

DAVID YUSUF DE WASKITO

21.12.1974

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

SKRIPSI

EVALUASI MODEL MACHINE LEARNING BERBASIS ENSEMBLE LEARNING DALAM MEMPREDIKSI HARGA RUMAH

yang disusun dan diajukan oleh

David Yusuf De Waskito

21.12.1974

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 11 Februari 2025

Dosen Pembimbing,


Yoga Pristyanto, S.Kom., M.Eng

NIK. 190302412

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

SKRIPSI

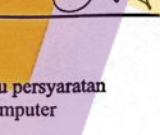
EVALUASI MODEL MACHINE LEARNING
BERBASIS ENSEMBLE LEARNING
DALAM MEMPREDIKSI HARGA RUMAH

yang disusun dan diajukan oleh

David Yusuf De Waskito
21.12.1974

Telah dipertahankan di depan Dewan Penguji
pada tanggal 11 Februari 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
<u>Ainul Yaqin, M.Kom.</u> NIK. 190302255	
<u>Donni Prabowo, S.Kom., M.Kom.</u> NIK. 190302253	
<u>Yoga Pristyanto, S.Kom., M.Eng.</u> NIK. 190302412	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 11 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : David Yusuf De Waskito
NIM : 21.12.1974

Menyatakan bahwa Skripsi dengan judul berikut:

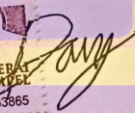
EVALUASI MODEL MACHINE LEARNING BERBASIS ENSEMBLE LEARNING DALAM MEMPREDIKSI HARGA RUMAH

Dosen Pembimbing : Yoga Pristyanto, S.Kom., M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 11 Februari 2025

Yang Menyatakan,


David Yusuf De Waskito

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas segala rahmat dan hidayah-Nya, skripsi ini saya persembahkan kepada:

1. Kedua orang tua tercinta

Terima kasih atas setiap doa yang tak pernah terputus, setiap nasihat yang menjadi cahaya, serta setiap pengorbanan yang tidak akan pernah bisa terbalaskan. Tanpa didikan, dukungan, dan motivasi kalian, saya tidak akan sampai pada titik ini. Skripsi ini adalah wujud kecil dari perjuangan besar yang telah kalian berikan sepanjang hidup saya.

2. Mbah tersayang

Terima kasih atas kasih sayang, doa, dan perhatian yang selalu diberikan. Kehangatan serta petuah-petuahmu menjadi pengingat berharga dalam setiap langkah perjalanan ini.

3. Asri Rizka Pratiwi

Terima kasih telah menjadi pendamping yang selalu ada di sisi saya, memberikan semangat, serta menemani dalam setiap langkah perjuangan ini. Dukungan dan pengertianmu menjadi kekuatan tersendiri dalam menyelesaikan skripsi ini.

4. Bapak Yoga Prisyanto, S.Kom., M.Eng.

Terima kasih atas bimbingan, ilmu, dan arahannya selama proses penyusunan skripsi ini. Setiap masukan dan dukungan yang diberikan sangat berarti dalam perjalanan akademik saya.

5. Teman-teman student staff pengajaran

Terima kasih atas dukungan dan kebersamaan yang telah diberikan selama ini. Khususnya kepada Muh. Naufal, Ibrahim Aji, Hafidz, Welly, Evan Sugeha, Salsabila Balqiis, terima kasih telah menjadi bagian dari perjalanan ini. Setiap pengalaman yang kita lalui bersama akan selalu saya kenang.

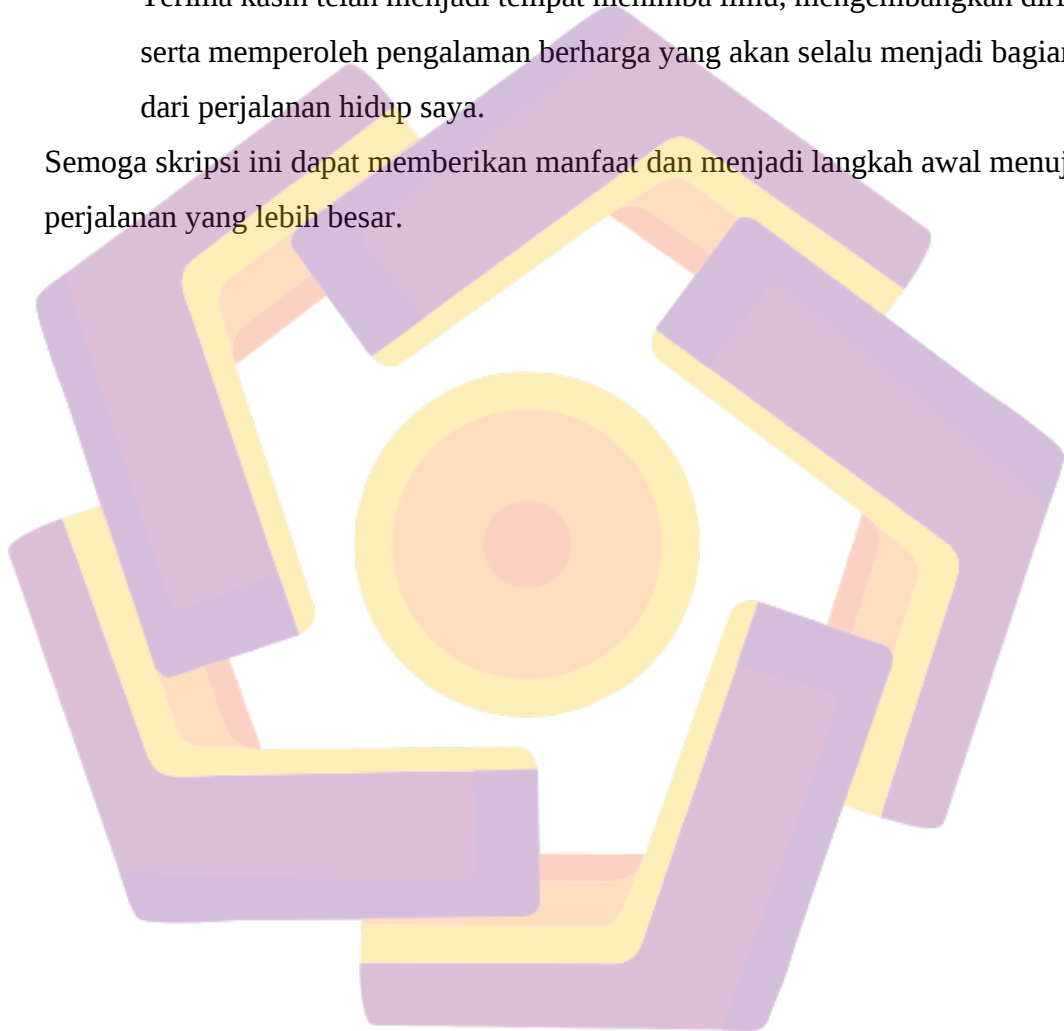
6. Teman-teman baik saya

Terima kasih atas dukungan, semangat, dan kebersamaan yang telah diberikan. Perjalanan ini menjadi lebih ringan dan bermakna karena kehadiran kalian.

7. Universitas Amikom Yogyakarta

Terima kasih telah menjadi tempat menimba ilmu, mengembangkan diri, serta memperoleh pengalaman berharga yang akan selalu menjadi bagian dari perjalanan hidup saya.

Semoga skripsi ini dapat memberikan manfaat dan menjadi langkah awal menuju perjalanan yang lebih besar.



KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir ini dengan lancar. Tugas akhir ini disusun untuk memenuhi salah satu syarat kelulusan Program Sarjana pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selain itu penulis dengan segala kerendahan hati ingin menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa memberikan dukungan dan bantuan untuk menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih kepada :

1. Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta
2. Hanif Al Fatta, S.Kom., M.Kom., Ph.D selaku Dekan Program Fakultas Ilmu Komputer
3. Anggit Dwi Hartanto, M.Kom, selaku Ketua Program Studi Sistem Informasi
4. Yoga Prisyanto, S.Kom., M.Eng., selaku dosen pembimbing yang memberikan arahan, saran, dan motivasi terhadap penulis
5. Kedua orang tua, keluarga besar, dan teman-teman tercinta yang memberikan semangat dan doa kepada penulis.

Yogyakarta, 7 Februari 2025

David Yusuf De Waskito

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	23
2.2.1 <i>Python</i>	23
2.2.2 <i>Machine learning</i>	23
2.2.3 <i>One-hot encoding</i>	24
2.2.4 <i>Standard scaling</i>	24
2.2.5 <i>Random forest</i>	24
2.2.6 <i>Support Vector Regression</i>	27
2.2.7 <i>XGBoost</i>	29
2.2.8 <i>Train-Test Split</i>	30
2.2.9 <i>R-Squared</i>	30

2.2.10	<i>Mean Absolute Error (MAE)</i>	31
2.2.11	<i>Mean Squared Error (MSE)</i>	32
BAB III METODE PENELITIAN		33
3.1	Objek Penelitian.....	33
3.2	Alur Penelitian	35
3.2.1	Pengumpulan data.....	37
3.2.2	Analisis data.....	37
3.2.3	<i>Data pre-processing</i>	38
3.2.4	<i>Split data</i>	39
3.2.5	Evaluasi model.....	40
3.2.6	Penerapan model.....	40
3.3	Alat dan Bahan.....	41
BAB IV HASIL DAN PEMBAHASAN		42
4.1	Pengumpulan Data.....	42
4.2	Analisis Dan Pra-pemrosesan Data.....	43
4.2.1	Analisis	44
4.2.2	Pembersihan dan pengambilan fitur yang diperlukan.....	46
4.2.3	Analisis outlier	50
4.2.4	Distribusi data	58
4.2.5	<i>One-hot encoding</i>	68
4.2.6	Analisis korelasi.....	71
4.3	Scaling data.....	73
4.4	Training Model	76
4.4.1	<i>Train test split</i>	76
4.4.2	<i>Support Vector Regression</i>	77
4.4.3	<i>Random forest</i>	81
4.4.4	<i>XGBoost</i>	83
4.5	Hasil Perbandingan Evaluasi Model.....	86
4.6	Penerapan Model	88
4.6.1	Menyimpan model <i>random forest</i> dan <i>standard scaler</i>	88
4.6.2	Skema input	88
4.6.3	Membuat API dengan <i>FastAPI</i>	89
4.6.4	Hosting API dengan <i>Ngrok</i> di <i>colab</i>	90
BAB V PENUTUP		93
5.1	Kesimpulan	93
5.2	Saran	94
REFERENSI		95
LAMPIRAN.....		99

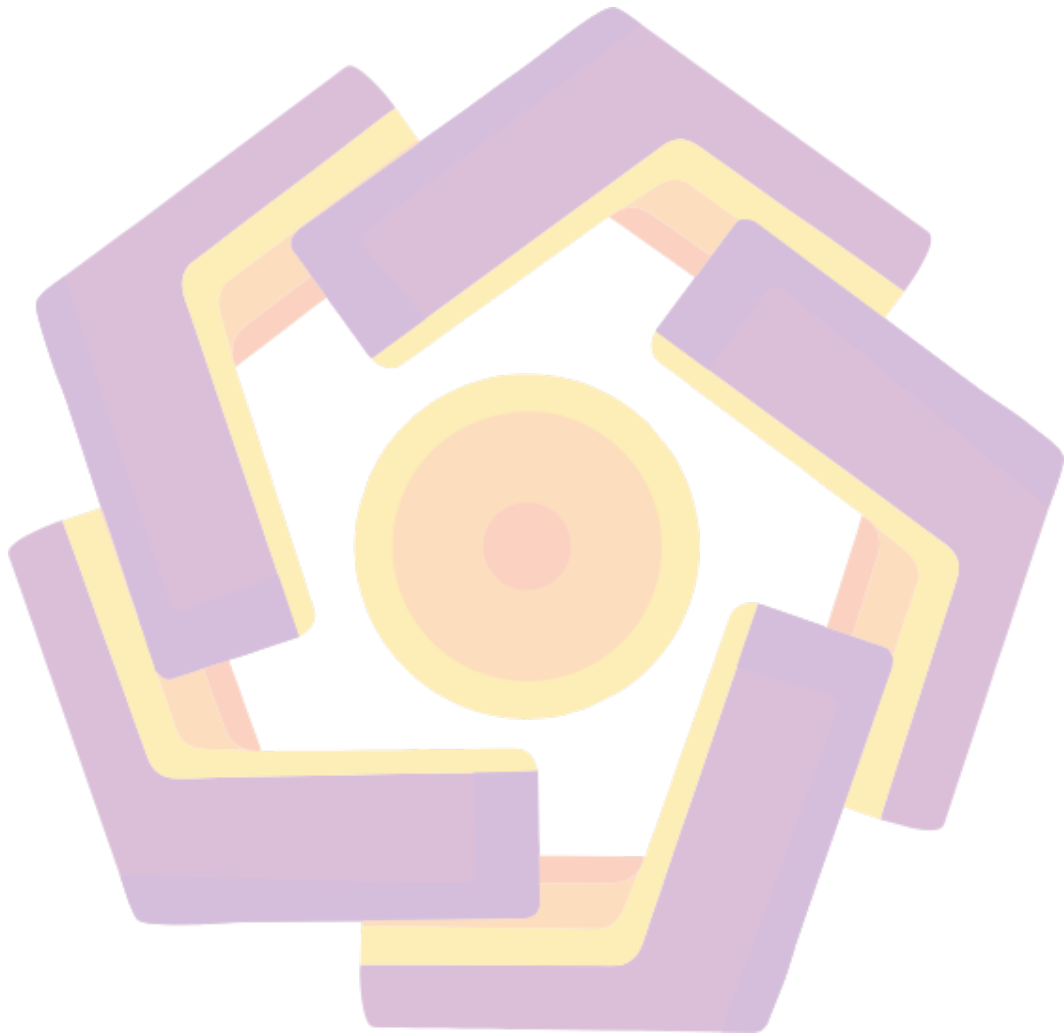
DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	12
Tabel 4.1 Tabel Dataset	42
Tabel 4.2 Data <i>surface</i> dan <i>building area</i>	48
Tabel 4.3 Data <i>surface</i> dan <i>building area</i> dalam m ²	48
Tabel 4.4 Hasil konversi <i>surface</i> dan <i>building area</i> ke numerik	49
Tabel 4.5 Tabel <i>Outlier building_area</i>	52
Tabel 4.6 Hasil <i>one-hot encoding</i>	70
Tabel 4.7 Tabel proses <i>scaling</i> fitur <i>price</i>	74
Tabel 4.8 Tabel proses <i>scaling</i> fitur <i>bed</i>	75
Tabel 4.9 Tabel proses <i>scaling</i> fitur <i>bath</i>	75
Tabel 4.10 Hasil optimisasi dan evaluasi SVR	79
Tabel 4.11 Hasil evaluasi akhir model SVR	80
Tabel 4.12 Hasil optimisasi dan evaluasi <i>Random Forest</i>	82
Tabel 4.13 Hasil optimisasi dan evaluasi <i>XGBoost</i>	85
Tabel 4.14 Hasil evaluasi akhir <i>XGBoost</i>	86
Tabel 4.15 Hasil evaluasi model	86

DAFTAR GAMBAR

Gambar 2.1 Model SVR	27
Gambar 3.1 Alur Penelitian	35
Gambar 4.1 Menghubungkan Google Drive dan Memasukkan Dataset ke Google Colab	42
Gambar 4.2 Fitur dataset dan jumlah data	44
Gambar 4.3 Data kategorikal lokasi	45
Gambar 4.4 Menghapus fitur nav-link dan description	46
Gambar 4.5 Konversi fitur price ke numerik	47
Gambar 4.6 Membersihkan data surface dan building area	48
Gambar 4.7 Konversi surface dan building area ke numerik	49
Gambar 4.8 Generalisasi Lokasi	49
Gambar 4.9 Analisis outlier menggunakan boxplot	51
Gambar 4.10 Identifikasi outlier	52
Gambar 4.11 Penghapusan Outlier pada building_area	57
Gambar 4.12 Hasil analisis pembersihan outlier	57
Gambar 4.13 Distribusi data lokasi rumah	59
Gambar 4.14 Distribusi fitur price	60
Gambar 4.15 Distribusi fitur bed	61
Gambar 4.16 Distribusi fitur bath	61
Gambar 4.17 Distribusi fitur carport	62
Gambar 4.18 Distribusi fitur surface area	63
Gambar 4.19 Distribusi fitur building area	64
Gambar 4.20 Distribusi fitur price di Sleman	65
Gambar 4.21 Distribusi fitur price di Bantul	65
Gambar 4.22 Distribusi fitur price di Yogyakarta	66
Gambar 4.23 Distribusi fitur price di Kulon Progo	66
Gambar 4.24 Distribusi fitur price di Gunung Kidul	67
Gambar 4.25 Distribusi lokasi setelah encoding	68
Gambar 4.26 One-hot encoding	69
Gambar 4.27 Analisis heatmap korelasi	71
Gambar 4.28 Matriks korelasi	72
Gambar 4.29 Standard Scaling	74
Gambar 4.30 Train-test split	76
Gambar 4.31 Optimisasi dan evaluasi SVR	78
Gambar 4.32 Grafik performa model SVR terhadap C	79
Gambar 4.33 Evaluasi model SVR dengan parameter terbaik	80
Gambar 4.34 Optimisasi dan evaluasi random forest	81
Gambar 4.35 Grafik performa model Random forest terhadap estimator	82
Gambar 4.36 Optimisasi dan evaluasi XGBoost	84
Gambar 4.37 Grafik performa model XGBoost terhadap estimator	85
Gambar 4.38 Evaluasi model XGBoost dengan parameter terbaik	86
Gambar 4.39 Simpan model dan scaler ke file	88
Gambar 4.40 Pembuatan skema input	89

Gambar 4.41 Implementasi FastAPI	90
Gambar 4.42 Hosting API	91
Gambar 4.43 Tampilan utama API	91
Gambar 4.44 Tampilan docs API	91
Gambar 4.45 Aplikasi web prediksi harga rumah di DIY	92



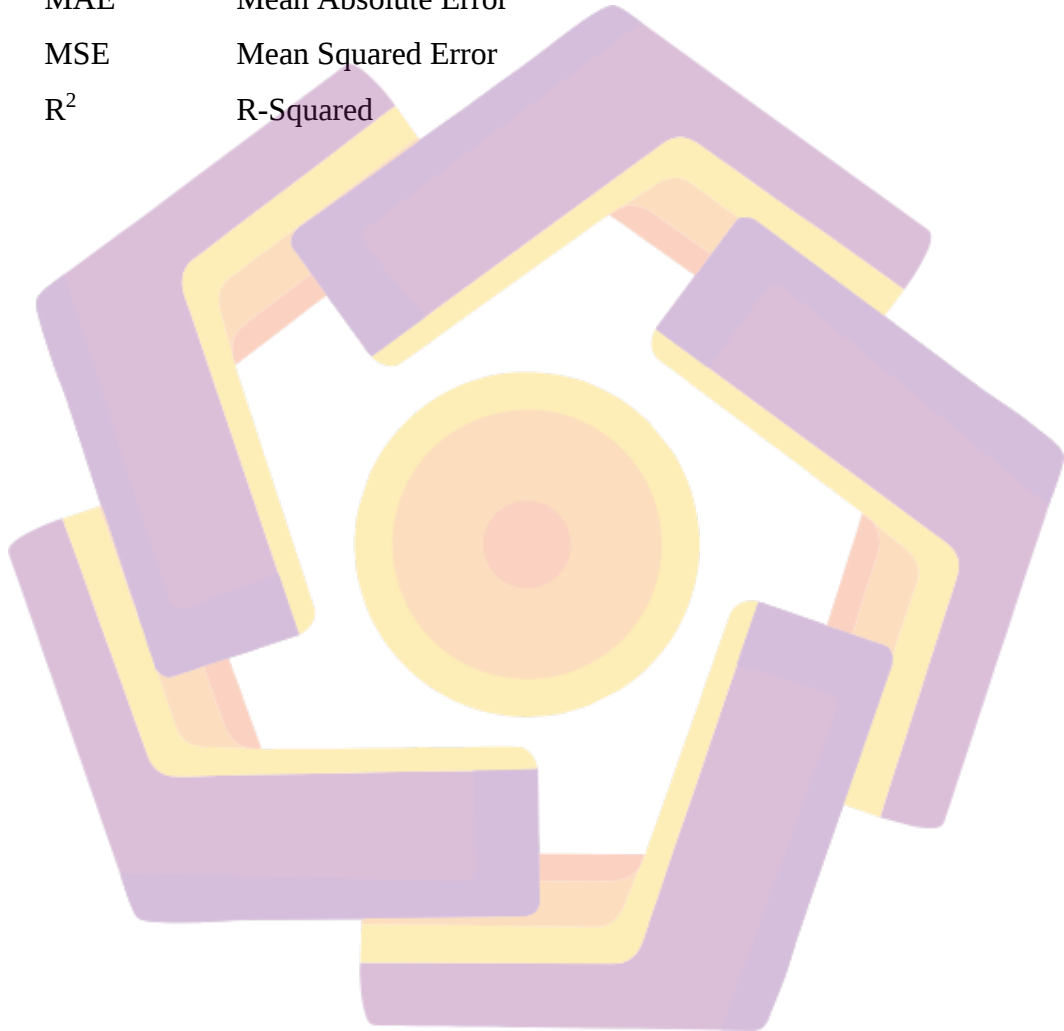
DAFTAR LAMPIRAN

Lampiran 1 Colab Skripsi	99
Lampiran 2 Colab FastAPI	99
Lampiran 3 Front-end Aplikasi Prediksi Harga Rumah	99



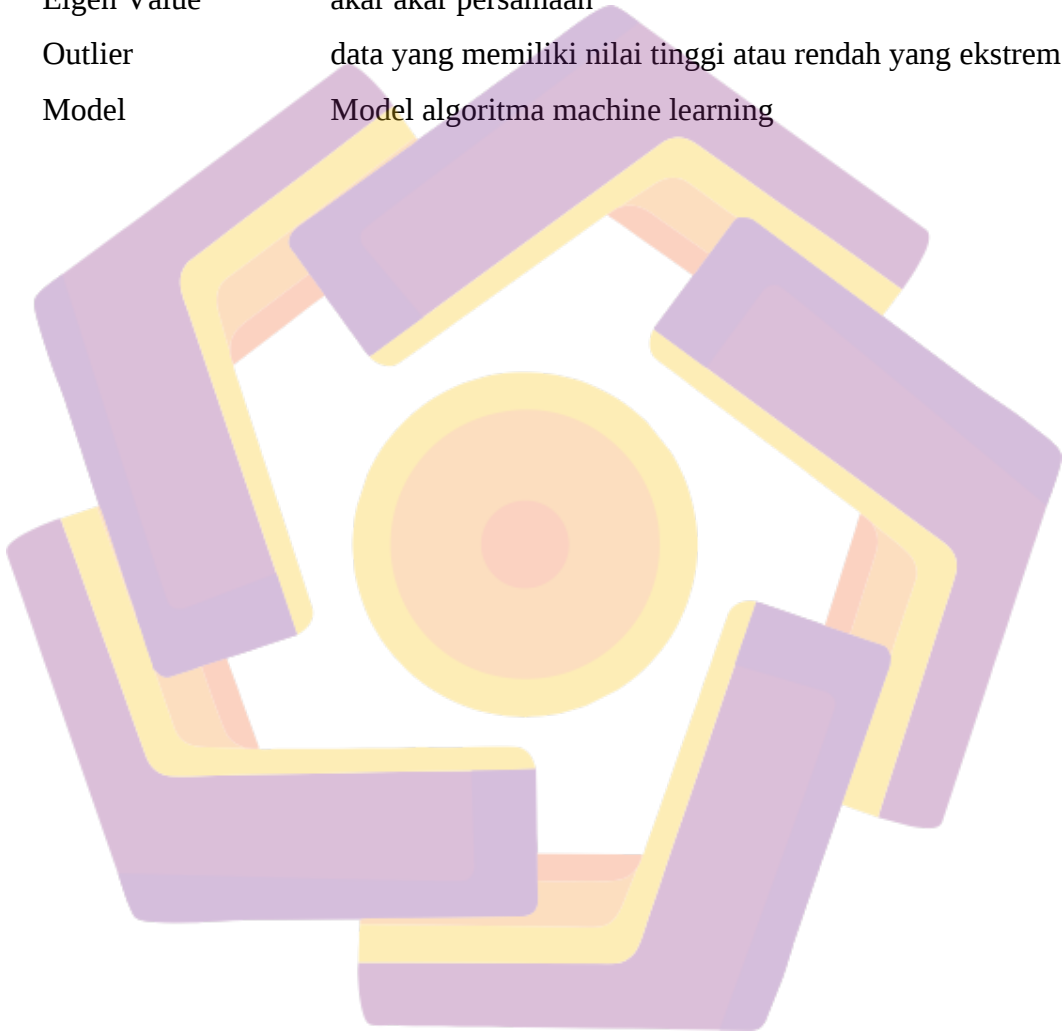
DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
SVR	Support Vector Regression
RBF	Radial Basis Function
MAE	Mean Absolute Error
MSE	Mean Squared Error
R^2	R-Squared



DAFTAR ISTILAH

Vektor	besaran yang mempunyai arah
Kernel	fungsi matematika yang digunakan dalam metode Support
Vector Regression (SVR)	
Eigen Value	akar akar persamaan
Outlier	data yang memiliki nilai tinggi atau rendah yang ekstrem
Model	Model algoritma machine learning



INTISARI

Prediksi harga rumah penting dalam industri properti untuk membantu pengambilan keputusan. Namun, faktor seperti luas bangunan, jumlah kamar, dan lokasi membuat estimasi harga menjadi kompleks. Penelitian ini membandingkan metode *Ensemble Learning* dan *Support Vector Regression* (SVR) untuk memperoleh model prediksi terbaik. Dataset diperoleh dari Kaggle dan diproses melalui penanganan outlier, normalisasi *Standard Scaler*, serta encoding lokasi. Model yang diuji meliputi *Random Forest*, *XGBoost*, dan SVR (RBF kernel). Evaluasi menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *R-squared* (R^2) menunjukkan bahwa *Random Forest* unggul (MAE 0.1971, MSE 0.1451, R^2 0.8437), diikuti *XGBoost* (MAE 0.2146, MSE 0.1475, R^2 0.84107), sementara SVR memiliki performa terendah (MAE 0.267, MSE 0.19, R^2 0.79). Hasil ini menunjukkan bahwa metode *Ensemble Learning* lebih efektif dalam prediksi harga rumah dibandingkan SVR. Penelitian ini bermanfaat bagi pengembang properti, agen real estat, dan calon pembeli dalam memperkirakan harga rumah. Studi selanjutnya disarankan menggunakan dataset lebih besar dan meningkatkan tuning parameter SVR.

Kata kunci: Prediksi harga rumah, *Machine Learning*, *Ensemble Learning*, *Random forest*, *XGBoost*, *Support Vector Regression*.

ABSTRACT

House price prediction is crucial in the real estate industry to support decision-making. However, factors such as building area, number of rooms, and location make price estimation complex. This study compares Ensemble Learning and Support Vector Regression (SVR) to determine the best predictive model. The dataset was obtained from Kaggle and processed through outlier handling, Standard Scaler normalization, and location encoding. The tested models include Random Forest, XGBoost, and SVR (RBF kernel). Evaluation using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) indicates that Random Forest performs best (MAE 0.1971, MSE 0.1451, R^2 0.8437), followed by XGBoost (MAE 0.2146, MSE 0.1475, R^2 0.84107), while SVR shows the lowest performance (MAE 0.267, MSE 0.19, R^2 0.79). These results demonstrate that Ensemble Learning is more effective for house price prediction than SVR. This research benefits property developers, real estate agents, and potential buyers in estimating house prices. Future studies are recommended to use a larger dataset and improve SVR parameter tuning.

Keywords: House price prediction, Machine Learning, Ensemble Learning, Random Forest, XGBoost, Support Vector Regression.