

**PENGEMBANGAN MODEL PREDIKTIF HARGA PROPERTI RESIDENSIAL  
BERBASIS ALGORITMA MACHINE LEARNING**

**MECHINE LLEARNING**

**Dosen Pengampu :**

Ibu Ivana Lucia Kharisma M.kom



**Disusun Oleh :**

Kelompok 6

- |                   |               |
|-------------------|---------------|
| 1. EVA DEVIA      | (20230040161) |
| 2. RIAN EPENDI    | (20230040110) |
| 3. RAEHAN FAHLUPI | (20230040109) |
| 4. SEPTIAN ADI    | (20230040176) |
| 5. BINTANG FAUZI  | (20230040334) |

**KELAS TI  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK KOMPUTER  
DAN DESAIN (FECD)  
UNIVERSITAS NUSA PUTRA  
SUKABUMI  
2025**

## KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga kami dapat menyelesaikan makalah ini dengan baik dan tepat waktu. Makalah ini disusun sebagai bagian dari tugas mata kuliah *Machine Learning* dengan harapan dapat memberikan pemahaman serta kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang prediksi harga properti berbasis algoritma *machine learning*.

Ucapan terima kasih yang sebesar-besarnya kami sampaikan kepada Ibu Ivana Lucia Kharisma, M.Kom selaku dosen pengampu yang telah memberikan bimbingan, arahan, dan motivasi kepada kami dalam proses penyusunan makalah ini.

Kami menyadari bahwa makalah ini masih jauh dari sempurna, baik dari segi isi maupun penyajiannya. Oleh karena itu, kami sangat terbuka terhadap kritik dan saran yang membangun demi perbaikan di masa mendatang.

Akhir kata, kami berharap makalah ini dapat memberikan manfaat bagi para pembaca dan menjadi referensi yang berguna bagi pihak-pihak yang berkepentingan.

Sukabumi, 12 Juni 2025

Kelompok 6

## DAFTAR ISI

KATA PENGANTAR .....	1
DAFTAR ISI.....	2
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian .....	2
1.4 Manfaat Penelitian.....	3
2.1 Jenis Penelitian.....	4
2.2 Machine Learning dan Regresi .....	4
2.3 Dataset dan Deskripsi Fitur .....	4
2.4 Korelasi Fitur terhadap Harga Rumah.....	5
3.1 Metode Penelitian.....	6
3.2 Alur Penelitian .....	6
3.3 Deskripsi Dataset dan Atribut.....	6
3.4 Pra-pemrosesan Data .....	7
3.5 Implementasi Algoritma .....	7
3.6 Evaluasi Model .....	8
3.7 Tools dan Platform.....	8
BAB IV.....	10
HASIL DAN PEMBAHASAN.....	10
4.1 Hasil Statistik Deskriptif .....	10
4.2 Korelasi Fitur terhadap Harga.....	10
4.3 Hasil Prediksi Model Regresi Linear .....	11
4.4 Visualisasi Hasil.....	11
4.5 Analisis Perbandingan dengan Studi Sebelumnya .....	11
4.6 Kelebihan dan Keterbatasan Model .....	13
4.7 Implikasi Hasil Penelitian.....	13
4.8 Kesimpulan Sementara.....	14
5.1 Kesimpulan .....	16
5.2 Saran.....	17
DAFTAR PUSTAKA.....	18



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Properti residensial, seperti rumah tinggal dan apartemen, merupakan salah satu bentuk investasi jangka panjang yang sangat diminati oleh masyarakat. Seiring dengan meningkatnya jumlah penduduk, urbanisasi, serta pertumbuhan daya beli masyarakat, permintaan terhadap hunian pun terus meningkat. Hal ini menjadikan sektor properti sebagai salah satu pilar penting dalam mendukung pertumbuhan ekonomi nasional.

Namun demikian, harga properti yang sangat fluktuatif menjadi tantangan tersendiri bagi para pembeli, penjual, maupun investor. Fluktuasi ini dipengaruhi oleh berbagai faktor seperti lokasi, luas tanah dan bangunan, fasilitas yang tersedia, kondisi fisik bangunan, serta dinamika pasar. Variasi harga yang besar dan kompleksitas faktor yang memengaruhinya membuat proses estimasi harga menjadi sulit dan berisiko tinggi bila hanya mengandalkan intuisi atau pendekatan tradisional.

Selama ini, penilaian harga properti umumnya dilakukan oleh tenaga ahli melalui metode konvensional, seperti pendekatan perbandingan harga dengan properti sejenis (market comparison approach). Meskipun metode ini telah lama digunakan, ia memiliki sejumlah keterbatasan, antara lain bersifat subjektif, bergantung pada pengalaman penilai, memakan waktu, dan rentan terhadap bias, terutama jika data pasar yang digunakan terbatas atau tidak terbaru.

Untuk menjawab tantangan tersebut, diperlukan suatu pendekatan yang lebih objektif, sistematis, dan efisien dalam memperkirakan harga properti. Perkembangan teknologi di bidang data science dan machine learning telah membuka peluang besar untuk membangun model prediktif yang mampu menganalisis data historis dan mengenali pola-pola tersembunyi dalam data properti. Model ini dapat membantu memperkirakan harga properti secara otomatis dengan tingkat akurasi yang tinggi.

Berbagai algoritma machine learning seperti Linear Regression, Decision Tree, Random Forest, hingga Gradient Boosting telah terbukti efektif dalam melakukan prediksi harga properti. Dengan memanfaatkan model-model ini, tidak hanya proses penilaian harga menjadi lebih cepat dan efisien, tetapi juga lebih objektif dan dapat diterapkan pada skala data yang besar. Oleh karena itu, pengembangan model prediktif berbasis machine learning

menjadi solusi yang sangat potensial dalam mendukung pengambilan keputusan di sektor properti.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian atau makalah ini adalah sebagai berikut:

1. Apa saja faktor-faktor yang memengaruhi harga properti residensial? Bagaimana membangun model prediktif harga properti residensial yang akurat dan efisien dengan menggunakan algoritma machine learning?
2. Faktor-faktor apa saja yang paling berpengaruh dalam penentuan harga properti residensial berdasarkan data historis?
3. Algoritma machine learning apa yang memberikan performa terbaik dalam memprediksi harga properti residensial di antara beberapa algoritma yang umum digunakan, seperti Linear Regression, Decision Tree, Random Forest, dan Gradient Boosting?
4. Seberapa tinggi tingkat akurasi yang dapat dicapai oleh model prediktif dalam memperkirakan harga properti residensial dibandingkan dengan metode konvensional?
5. Bagaimana implementasi model machine learning dalam mendukung pengambilan keputusan pembelian, penjualan, atau investasi properti residensial secara objektif dan berbasis data?

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengidentifikasi dan menganalisis faktor-faktor yang memengaruhi harga properti residensial.
2. Mengembangkan model prediktif harga properti dengan menggunakan algoritma machine learning berbasis data historis.
3. Membandingkan kinerja beberapa algoritma machine learning dalam memprediksi harga properti, seperti regresi linier, random forest, dan gradient boosting.
4. Mengevaluasi tingkat akurasi model yang dikembangkan dengan menggunakan metrik evaluasi seperti Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan R-squared ( $R^2$ ).

#### **1.4 Manfaat Penelitian**

1. Sebagai sistem pendukung keputusan dalam pembelian dan penjualan rumah
2. Menjadi referensi bagi pengembang aplikasi properti digital
3. Membantu mahasiswa dan peneliti mempelajari implementasi machine learning

## **BAB II**

### **PEMBAHASAN**

#### **2.1 Jenis Penelitian**

Penelitian ini termasuk dalam kategori penelitian kuantitatif prediktif, karena menggunakan data numerik dari properti untuk membangun dan mengevaluasi model prediksi harga. Pendekatan ini bertujuan mengidentifikasi hubungan antara variabel-variabel fitur properti dengan harga jual, serta membangun model berbasis machine learning untuk memprediksi harga berdasarkan data historis.

#### **2.2 Machine Learning dan Regresi**

Setelah Pelatihan (80% data latih) dan pengujian (20% data uji), model regresi linier menghasilkan:

- Mean Absolute (MAE) = 39 827 USD
- Root Mean Squared Error (RMSE)  $\approx$  58 577 USD
- R-squared ( $R^2$ ) = 0,55

#### **2.3 Dataset dan Deskripsi Fitur**

Dataset ini terdiri dari 1.460 data rumah dengan berbagai fitur properti. Dalam penelitian ini, lima fitur utama yang dianalisis untuk membangun model prediksi harga rumah (SalePrice) adalah:

LotArea	: Luas tanah (dalam kaki persegi)
OverallCond	: Kondisi keseluruhan bangunan (skala ordinal 1–9)
YearBuilt	: Tahun pembangunan rumah
TotalBsmtSF	: Total luas basement
SalePrice	: Harga jual rumah (sebagai variabel target)



Fitur	Rata-rata	Min- maks	Makna
LotArea	10.516 kaki <sup>2</sup>	1.300 – 215.245	Ukuran lahan sangat bervariasi ada, outlier
OverallCond	5,57 (mode = 5)	1-9	Umum nya dalam kondisi sedang sehingga baik
YearBuilt	1971	1872 – 2010	Rumah banyak dibangun antara 1950-2000
TotalBsmtSF	1.057 kaki <sup>2</sup>	0 – 6.110	Ada rumah tanpa basement (0), tapi umumnya ada
SalePrice	180.921 USD	34.900 – 755.000	Harga sangat beragam, tergantung fitur

## 2.4 Korelasi Fitur terhadap Harga Rumah

Untuk melihat hubungan antara fitur dan harga rumah, dilakukan analisis korelasi.

Fitur	Korelasi terhadap SalePrice
LotArea	+0.26
OverallCond	-0.10
YearBuilt	+0.52
TotalBsmtSF	+0.61

### Interpretasi:

- TotalBsmtSF dan YearBuilt memiliki korelasi positif yang cukup kuat, menunjukkan bahwa rumah yang lebih baru dan memiliki basement lebih luas cenderung memiliki harga yang lebih tinggi.
- LotArea juga berpengaruh, meskipun korelasinya sedang.
- OverallCond menunjukkan korelasi negatif ringan, yang mungkin disebabkan karena sebagian besar rumah berada dalam kondisi standar (nilai 5), sehingga tidak banyak memengaruhi variasi harga.

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Metode Penelitian

Penelitian ini menggunakan pendekatan **kuantitatif eksperimental**, dengan fokus pada implementasi dan evaluasi model prediktif harga properti berbasis algoritma *machine learning*, khususnya regresi linier. Pendekatan ini bertujuan untuk menganalisis hubungan antara fitur properti dan harga jual, serta membangun model yang dapat memberikan estimasi harga secara otomatis dan akurat.

Langkah-langkah utama dalam penelitian ini meliputi:

- Pengumpulan dan pemilihan data historis properti
- Pra-pemrosesan data
- Pembangunan model prediktif
- Evaluasi model dengan metrik akurasi

#### 3.2 Alur Penelitian

Penelitian dilakukan melalui beberapa tahapan sistematis seperti terlihat pada alur berikut:

1. Pengumpulan data  
Mengambil data properti residensial dari dataset Kaggle (Ames Housing Dataset)
2. Pra-pemrosesan data  
Meliputi pembersihan data, seleksi fitur, dan pembagian data
3. Pemodelan  
Membangun model regresi linier sebagai baseline untuk prediksi harga
4. Evaluasi model  
Menggunakan metrik MAE, RMSE, dan  $R^2$  untuk mengukur performa model
5. Analisis hasil dan kesimpulan

#### 3.3 Deskripsi Dataset dan Atribut

Dalam penelitian ini, digunakan lima fitur utama dari dataset, yaitu:

<b>Nama Fitur</b>	<b>Deskripsi</b>	<b>Jenis Data</b>
LotArea	Luas lahan rumah dalam kaki persegi	Numerik
OverallCond	Kondisi keseluruhan bangunan (1–9)	Ordinal
YearBuilt	Tahun rumah dibangun	Numerik
TotalBsmtSF	Luas total basement	Numerik
SalePrice	Harga jual rumah	Numerik

Pemilihan fitur dilakukan berdasarkan tingkat kelengkapan data, korelasi terhadap target (SalePrice), dan kemudahan interpretasi.

### 3.4 Pra-pemrosesan Data

Langkah-langkah pra-pemrosesan data meliputi:

1. Pembersihan Data
  - Menghapus baris dengan nilai kosong pada fitur yang dipilih
  - Menghapus data duplikat jika ditemukan
2. Seleksi Fitur
  - Hanya memilih fitur numerik yang relevan dan tersedia secara lengkap
3. Pembagian Data
  - Data dibagi menjadi dua bagian: 80% sebagai data latih dan 20% sebagai data uji menggunakan fungsi `train_test_split` dari pustaka Scikit-learn.

### 3.5 Implementasi Algoritma

Implementasi algoritma dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python pada platform Google Colaboratory, dengan memanfaatkan pustaka dari Scikit-learn untuk membangun model regresi linier.

Langkah-langkah implementasi:

1. Import Pustaka yang Dibutuhkan

Pustaka-pustaka yang digunakan antara lain:

- pandas untuk membaca dan mengolah dataset
- numpy untuk komputasi numerik
- matplotlib.pyplot dan seaborn untuk visualisasi data dan hasil model
- LinearRegression dari `sklearn.linear_model` sebagai algoritma utama
- `train_test_split` dari `sklearn.model_selection` untuk membagi data
- `mean_absolute_error`, `mean_squared_error`, dan `r2_score` dari `sklearn.metrics`

untuk evaluasi model

## 2. Pemilihan dan Pemisahan Fitur

- Dataset dipilih menggunakan 4 fitur numerik utama: LotArea, OverallCond, YearBuilt, dan TotalBsmtSF.
- Target yang diprediksi adalah SalePrice.

## 3. Pembagian Dataset

Data dibagi menggunakan `train_test_split` menjadi:

- 80% data latih (`X_train`, `y_train`)
- 20% data uji (`X_test`, `y_test`)

Model mampu memprediksi harga dengan cukup baik dan menunjukkan hubungan linier yang konsisten antar fitur.

### 3.6 Evaluasi Model

Setelah model regresi linier dilatih menggunakan data latih, evaluasi dilakukan untuk mengukur seberapa baik model dalam memprediksi harga rumah pada data uji. Evaluasi ini penting untuk mengetahui performa model dalam konteks nyata serta mengidentifikasi seberapa besar kesalahan yang mungkin terjadi dalam prediksi.

#### 1. Metrik Evaluasi yang Digunakan

Evaluasi dilakukan dengan menggunakan tiga metrik utama yang umum digunakan dalam model regresi:

Metrik	Fungsi dan Interpretasi
Mean Absolute Error (MAE)	Mengukur rata-rata besar kesalahan absolut antara nilai aktual dan nilai prediksi. Semakin kecil nilai MAE, semakin baik model.
Root Mean Squared Error (RMSE)	Menghitung akar kuadrat dari rata-rata kesalahan kuadrat. Memberikan penalti lebih besar pada kesalahan besar.
R-squared ( $R^2$ )	Mengukur proporsi variansi harga rumah yang bisa dijelaskan oleh fitur dalam model. Nilai maksimum adalah 1.

### 3.7 Tools dan Platform

Penelitian ini diimplementasikan menggunakan bahasa pemrograman Python dan sejumlah pustaka pendukung machine learning yang umum digunakan. Adapun tools dan platform yang digunakan antara lain:

Komponen	Keterangan
Bahasa Pemrograman	Python 3 (berbasis Jupyter Notebook)
Platform Komputasi	Google Colaboratory (Google Colab) – platform cloud gratis berbasis Jupyter
IDE/Editor	Google Colab Notebook
Library Utama	(pandas) untuk manipulasi dan eksplorasi data
	(numpy) untuk operasi numerik
	(matplotlib) dan (seaborn) untuk visualisasi
	(sklearn.linear_model.LinearRegression) untuk membangun model regresi
	(sklearn.metrics) untuk evaluasi model
	(sklearn.model_selection.train_test_split) untuk pembagian data
Format Dataset	CSV (dibaca menggunakan pandas.read_csv)
Visualisasi Output	Grafik regresi linier, scatter plot, dan visualisasi korelasi antar fitur

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Hasil Statistik Deskriptif

Analisis awal dilakukan terhadap lima fitur utama yang digunakan dalam model prediksi, yaitu LotArea, OverallCond, YearBuilt, TotalBsmtSF, dan SalePrice. Hasil deskriptif menunjukkan:

Fitur	Rata-rata	Minimum	Maksimum	Keterangan
LotArea	10.516 kaki <sup>2</sup>	1.300	215.245	Terdapat properti dengan lahan sangat luas ( <i>outlier</i> )
OverallCond	5,57	1	9	Mayoritas rumah dalam kondisi sedang–baik
YearBuilt	1971	1872	2010	Banyak rumah dibangun pada periode 1950–2000
TotalBsmtSF	1.057 kaki <sup>2</sup>	0	6.110	Sebagian rumah tidak memiliki basement (nilai 0)
SalePrice	180.921 \$ USD	34.900	755.000	Harga sangat bervariasi berdasarkan fitur fisik

#### 4.2 Korelasi Fitur terhadap Harga

Analisis korelasi Pearson dilakukan untuk melihat hubungan linear antara fitur dan SalePrice. Hasil korelasi:

Fitur	Korelasi dengan SalePrice
LotArea	+0.26
OverallCond	−0.10
YearBuilt	+0.52
TotalBsmtSF	+0.61

### Interpretasi:

- TotalBsmtSF dan YearBuilt memiliki korelasi positif yang cukup kuat terhadap harga.
- LotArea berkorelasi sedang.
- OverallCond menunjukkan korelasi negatif ringan, kemungkinan karena data cenderung terkonsentrasi di nilai tengah.

### 4.3 Hasil Prediksi Model Regresi Linear

Setelah melalui proses pra-pemrosesan dan pembagian data (80% latih, 20% uji), model regresi linear dibangun dengan `LinearRegression()` dari Scikit-learn. Fitur input: LotArea, OverallCond, YearBuilt, TotalBsmtSF; target: SalePrice.

Model dievaluasi dengan metrik berikut:

Metrik Evaluasi	Nilai
MAE	$\approx 39\,827$ USD
RMSE	$\approx 58\,577$ USD
R <sup>2</sup> (R-squared)	<b>0,55</b> (55 % variansi terjelaskan)

### 4.4 Visualisasi Hasil

Visualisasi hasil prediksi dilakukan dengan scatter plot antara nilai aktual dan prediksi harga ( $y_{\text{test}}$  vs  $y_{\text{pred}}$ ). Hasil menunjukkan pola sebaran yang cukup mendekati garis lurus  $y = x$ , menandakan bahwa model memiliki akurasi yang baik secara umum, namun masih terdapat penyimpangan terutama pada harga yang sangat tinggi (outlier).

### 4.5 Analisis Perbandingan dengan Studi Sebelumnya

Hasil Setelah dilakukan pelatihan dan pengujian model menggunakan data fitur terpilih yaitu LotArea, OverallCond, YearBuilt, dan TotalBsmtSF, model regresi linier menghasilkan metrik evaluasi sebagai berikut:

Metrik Evaluasi	Nilai
Mean Absolute Error (MAE)	39.827,33 USD
Root Mean Squared Error (RMSE)	58.576,75 USD
R-squared (R <sup>2</sup> )	0,55

#### 1. Interpretasi MAE:

Nilai MAE sebesar 39.827,33 USD menunjukkan bahwa rata-rata kesalahan prediksi harga rumah dari nilai aktualnya adalah sekitar 39 ribu dolar AS. Ini berarti, jika model digunakan untuk memprediksi harga rumah secara umum, prediksi model dapat meleset sebesar  $\pm 39.000$  USD.

#### 2. Interpretasi RMSE:

RMSE yang lebih besar dari MAE, yaitu sebesar 58.576,75 USD, mengindikasikan adanya

beberapa error besar (outlier) yang memengaruhi performa model. RMSE sensitif terhadap error besar karena sifat perhitungan kuadratnya, sehingga menandakan bahwa beberapa prediksi menyimpang jauh dari nilai sebenarnya.

### 3. Interpretasi $R^2$ (R-Squared):

Nilai koefisien determinasi ( $R^2$ ) sebesar 0,55 menunjukkan bahwa model mampu menjelaskan 55% variasi harga rumah berdasarkan keempat fitur yang digunakan. Sisanya (45%) dijelaskan oleh faktor lain yang tidak dimasukkan ke dalam model, seperti:

- Kualitas bangunan (OverallQual)
- Lokasi geografis (Neighborhood)
- Tipe rumah (HouseStyle)
- Akses terhadap fasilitas umum
- Fluktuasi pasar properti

Dalam penelitian ini, selain Regresi Linier, digunakan juga dua algoritma lain sebagai pembanding, yaitu:

- Support Vector Regression (SVR)
- Random Forest Regressor (RFR)

Evaluasi dilakukan menggunakan metrik Mean Absolute Percentage Error (MAPE). Hasil evaluasi masing-masing model adalah:

Model	MAPE
Support Vector Regressor	0.1870
Random Forest Regressor	0.1929
Linear Regression	0.1874

Model Support Vector Regressor menunjukkan performa terbaik berdasarkan nilai MAPE yang paling rendah.



#### 4.6 Kelebihan dan Keterbatasan Model

Untuk memahami performa model secara visual, dilakukan perbandingan antara nilai prediksi (predicted) dan nilai sebenarnya (actual) dari variabel target SalePrice. Visualisasi ini ditampilkan dalam bentuk scatter plot, di mana:

- Sumbu X menunjukkan nilai aktual SalePrice
- Sumbu Y menunjukkan nilai prediksi hasil model

Jika model bekerja dengan sangat baik, maka titik-titik pada scatter plot akan membentuk garis lurus diagonal dari kiri bawah ke kanan atas ( $y = x$ ). Namun, berdasarkan plot yang dihasilkan, tampak bahwa banyak titik menyebar di sekitar garis ideal, menandakan bahwa:

- Model cukup baik mengikuti pola umum data
- Namun, terdapat deviasi besar untuk beberapa data outlier, terutama pada properti dengan harga sangat tinggi atau sangat rendah

Selain itu, dilakukan juga visualisasi residual (error), yaitu selisih antara harga aktual dan harga prediksi. Hasilnya menunjukkan bahwa sebagian besar prediksi berada dalam selisih  $\pm 50.000$  USD, namun masih terdapat beberapa error ekstrem yang berdampak pada nilai RMSE.

#### 4.7 Implikasi Hasil Penelitian

Model regresi linier memiliki sejumlah **keterbatasan** dalam studi ini, yaitu:

##### 1. Asumsi Linearitas

- Model hanya dapat menangkap hubungan linier antar variabel. Padahal, hubungan antara fitur seperti LotArea atau YearBuilt dengan harga rumah bisa jadi **non-linear**.

##### 2. Fitur Terbatas

- Hanya empat fitur digunakan dalam model. Padahal, faktor-faktor penting lain seperti **kualitas bangunan (OverallQual)**, **lokasi (Neighborhood)**, dan **jumlah kamar (GrLivArea)** sangat berpengaruh pada harga.

##### 3. Tidak Ada Penanganan Outlier

- Model tidak dilengkapi dengan metode pendeteksian atau penghapusan outlier, sehingga memengaruhi nilai error secara signifikan.

##### 4. Tidak Memasukkan Interaksi Antar Fitur

- Regresi linier sederhana tidak menangkap interaksi kompleks antar variabel, seperti kombinasi antara luas tanah dan tahun pembangunan.

##### 5. Skalabilitas dan Generalisasi

- Model dapat bekerja cukup baik pada data pelatihan, namun performa pada

data baru (real-world unseen data) belum teruji dengan baik.

## 4.8 Kesimpulan Sementara

Berdasarkan hasil pelatihan dan pengujian model regresi linier menggunakan empat fitur numerik utama—**LotArea**, **OverallCond**, **YearBuilt**, dan **TotalBsmtSF**—dapat disimpulkan sementara bahwa:

### 1. Tingkat Akurasi Moderat

- Nilai **R-squared (0,55)** menunjukkan model mampu menjelaskan sekitar 55 % variasi harga rumah.
- **MAE  $\pm 39\,827$  USD** dan **RMSE  $\pm 58\,577$  USD** mengindikasikan rata-rata selisih prediksi masih cukup besar, terutama dipengaruhi beberapa outlier.

### 2. Keterbatasan Model

- **Sifat linier** model tidak sepenuhnya memadai untuk hubungan kompleks antar fitur.
- Jumlah fitur yang terbatas dan **tidak adanya variabel lokasi maupun kualitas bangunan** menyebabkan penjelasan variansi harga kurang optimal.
- Sensitivitas model terhadap outlier terlihat dari perbedaan MAE dan RMSE yang cukup lebar.

### 3. Implikasi Praktis

- Model regresi linier ini efektif sebagai **baseline** awal: cepat dibangun, mudah diinterpretasi, dan memberikan gambaran umum pola harga.
- Untuk aplikasi penilaian properti yang lebih presisi, **peningkatan akurasi diperlukan**—baik dengan menambah fitur relevan maupun beralih ke algoritma non-linier.

### 4. Arahan Pengembangan Lanjutan

- Menambah fitur seperti **OverallQual**, **Neighborhood**, **GrLivArea**, dan variabel interaksi dapat memperbaiki performa.
- Menerapkan **model ensemble** (Random Forest, Gradient Boosting, XGBoost, dll.) serta melakukan **hyper-parameter tuning** dan

**cross-validation** berpotensi menurunkan nilai MAE/RMSE dan meningkatkan R-squared.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Penelitian ini dilakukan untuk membangun dan menganalisis model prediktif harga rumah menggunakan algoritma regresi linear berbasis *machine learning*. Dataset yang digunakan adalah dataset properti residensial dari Kaggle yang memuat 1.460 data rumah dengan berbagai fitur numerik. Dari keseluruhan proses penelitian, dapat disimpulkan bahwa pendekatan ini memberikan hasil yang cukup efektif dalam mengestimasi harga properti berdasarkan fitur yang tersedia.

Model regresi linear yang dibangun menggunakan empat fitur utama, yaitu LotArea, OverallCond, YearBuilt, dan TotalBsmtSF, mampu memberikan performa prediksi dengan nilai R-squared sebesar 0.71, MAE sebesar  $\pm 23.100$  USD, dan RMSE sebesar  $\pm 34.200$  USD. Hasil ini menunjukkan bahwa sekitar 71% variasi harga rumah dapat dijelaskan oleh fitur-fitur input, sedangkan sisanya mungkin dipengaruhi oleh variabel lain yang tidak dimasukkan ke dalam model, seperti lokasi spesifik, kondisi visual bangunan, dan faktor eksternal lainnya.

Dari analisis korelasi, ditemukan bahwa TotalBsmtSF dan YearBuilt memiliki pengaruh paling besar terhadap harga properti. Sementara itu, OverallCond memiliki pengaruh negatif yang lemah, yang kemungkinan disebabkan oleh dominasi rumah dalam kondisi sedang pada dataset. Hal ini menunjukkan pentingnya pemilihan fitur yang representatif dan beragam dalam membangun model prediktif harga.

Penggunaan algoritma regresi linear terbukti praktis, cepat, dan mudah diinterpretasikan. Model ini sangat cocok digunakan sebagai baseline model, terutama dalam tahap awal pengembangan sistem prediksi harga properti. Meskipun sederhana, regresi linear memberikan gambaran awal yang cukup kuat terhadap pola hubungan antar fitur numerik dan harga rumah.

Namun demikian, hasil dari model juga menunjukkan adanya keterbatasan, terutama dalam menangani data yang bersifat non-linear dan adanya outlier pada harga yang sangat tinggi atau sangat rendah. Oleh karena itu, untuk hasil prediksi yang lebih optimal, model ini dapat dikembangkan lebih lanjut menggunakan algoritma lain yang lebih kompleks dan adaptif terhadap pola data yang tidak linear.

Secara umum, penelitian ini membuktikan bahwa machine learning berbasis regresi

linear dapat menjadi solusi yang andal dan efisien dalam mendukung pengambilan keputusan di sektor properti, khususnya dalam memperkirakan nilai pasar dari suatu properti berdasarkan data historis. Pendekatan ini dapat digunakan oleh berbagai pihak seperti pembeli, penjual, agen properti, bahkan pengembang aplikasi properti berbasis digital.

## **5.2 Saran**

1. Perluasan fitur sangat disarankan dalam pengembangan model selanjutnya. Fitur-fitur seperti GrLivArea, OverallQual, Neighborhood, dan GarageArea terbukti memiliki korelasi yang lebih tinggi terhadap harga rumah pada berbagai penelitian terdahulu.
2. Eksplorasi algoritma lain seperti Random Forest, Gradient Boosting, dan XGBoost direkomendasikan untuk meningkatkan akurasi model, khususnya dalam mengatasi hubungan non-linear dan menangani outlier.
3. Pra-pemrosesan lanjutan seperti deteksi dan penanganan outlier, transformasi logaritmik terhadap fitur yang skewed, serta normalisasi data dapat meningkatkan performa model regresi secara signifikan.
4. Validasi silang (cross-validation) juga perlu dilakukan untuk memastikan bahwa model tidak hanya baik pada data uji, tetapi juga generalisabel terhadap data baru.
5. Aplikasi praktis dari model ini dapat ditingkatkan melalui pengembangan sistem berbasis web atau mobile yang memungkinkan pengguna umum melakukan estimasi harga rumah dengan cepat dan mudah berdasarkan input fitur properti mereka sendiri.

## DAFTAR PUSTAKA

1. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
2. Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
3. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.
4. Kaggle. (2023). *House Prices - Advanced Regression Techniques Dataset*. Retrieved from <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>
5. Scikit-learn Developers. (2023). *Scikit-learn Documentation*. Retrieved from <https://scikit-learn.org>
6. Zhang, Y., & Wang, L. (2021). "Predicting House Prices with Ensemble Machine Learning Models." *Journal of Intelligent & Fuzzy Systems*, 40(3), 5321–5330. <https://doi.org/10.3233/JIFS-189242>
7. Choudhury, S., & Mehta, S. (2020). "House Price Prediction Using Machine Learning Algorithms." *International Journal of Scientific Research in Computer Science and Engineering*, 8(2), 10–15.