

Pengembangan Model Prediktif Harga Properti Residensial Berbasis Algoritma Machine Learning

Implementasi Linear Regression untuk Prediksi Harga Rumah

Kelompok 6 - Kelas TI23H

Agenda Presentasi

- 1** Pendahuluan & Latar Belakang
- 2** Rumusan Masalah & Tujuan Tujuan
- 3** Metodologi Penelitian
- 4** Dataset & Preprocessing
- 5** Implementasi Model
- 6** Hasil & Evaluasi
- 7** Kesimpulan & Saran

Latar Belakang

Tantangan Sektor Properti

Sektor properti hadapi tantangan kompleks.

Keterbatasan Metode Konvensional

Metode lama kurang efektif.

Solusi Machine Learning

ML tawarkan solusi inovatif.

Potensi Algoritma ML

Algoritma ML berpotensi besar.



Rumusan Masalah

1

Faktor Harga Properti

Identifikasi faktor-faktor penentu harga.

2

Model Akurat & Efisien

Bangun model prediktif yang tepat.

3

Faktor Paling Berpengaruh

Temukan faktor dominan.

4

Perbandingan Algoritma

Bandingkan performa ML.

5

Akurasi vs Konvensional

Ukur akurasi model.

6

Implementasi Keputusan

Terapkan untuk keputusan.

Tujuan & Manfaat Penelitian

Tujuan

- Identifikasi faktor harga.
- Kembangkan model prediktif.
- Bandingkan kinerja algoritma.
- Evaluasi tingkat akurasi.

Manfaat

- Sistem pendukung keputusan.
- Referensi aplikasi properti.
- Pembelajaran implementasi ML.

Metodologi Penelitian



Jenis Penelitian

Kuantitatif
Prediktif.



Platform & Bahasa

Google
Colaboratory,
Python.



Library

Pandas,
NumPy,
Scikit-learn,
Matplotlib,
Seaborn.

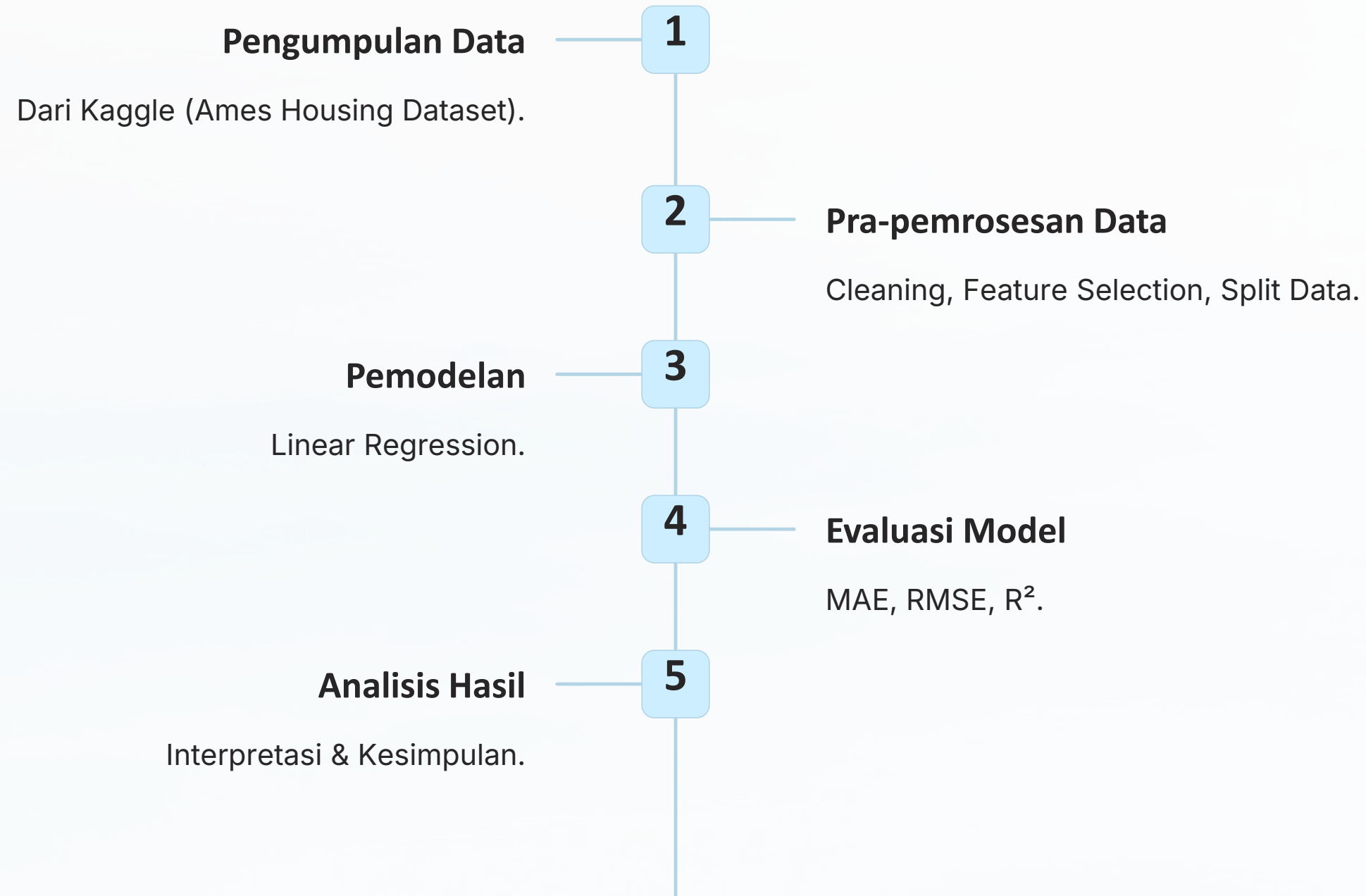


Algoritma

Linear
Regression.



Alur Penelitian



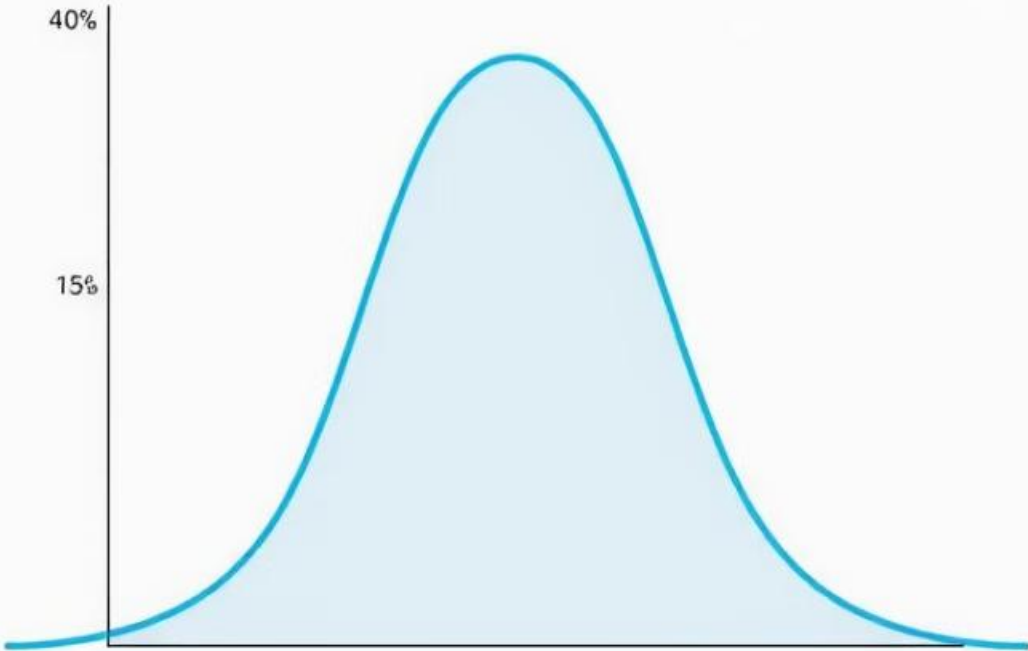
Dataset & Fitur

| | |
|-------------|---------------------------------|
| Dataset | 1.460 data rumah dari Kaggle |
| LotArea | Luas tanah (kaki ²) |
| OverallCond | Kondisi bangunan (1-9) |
| YearBuilt | Tahun pembangunan |
| TotalBsmtSF | Luas basement |
| SalePrice | Harga jual (target) |



Statistik Deskriptif Fitur Utama

| Fitur | Rata-rata | Min-Max | Keterangan |
|-------------|--------------------------|------------------|--------------------------|
| LotArea | 10.516 kaki ² | 1.300-215.245 | Ada outlier lahan luas |
| OverallCond | 5,57 | 1-9 | Kondisi sedang-baik |
| YearBuilt | 1971 | 1872-2010 | Banyak tahun 1950-2000 |
| TotalBsmtSF | 1.057 kaki ² | 0-6.110 | Ada rumah tanpa basement |
| SalePrice | \$180.921 | \$34.900-755.000 | Harga sangat bervariasi |



Analisis Korelasi Terhadap Harga Jual

TotalBsmtSF: +0.61 (Kuat)

Luas basement sangat berkorelasi positif dengan harga.

YearBuilt: +0.52 (Sedang-Kuat)

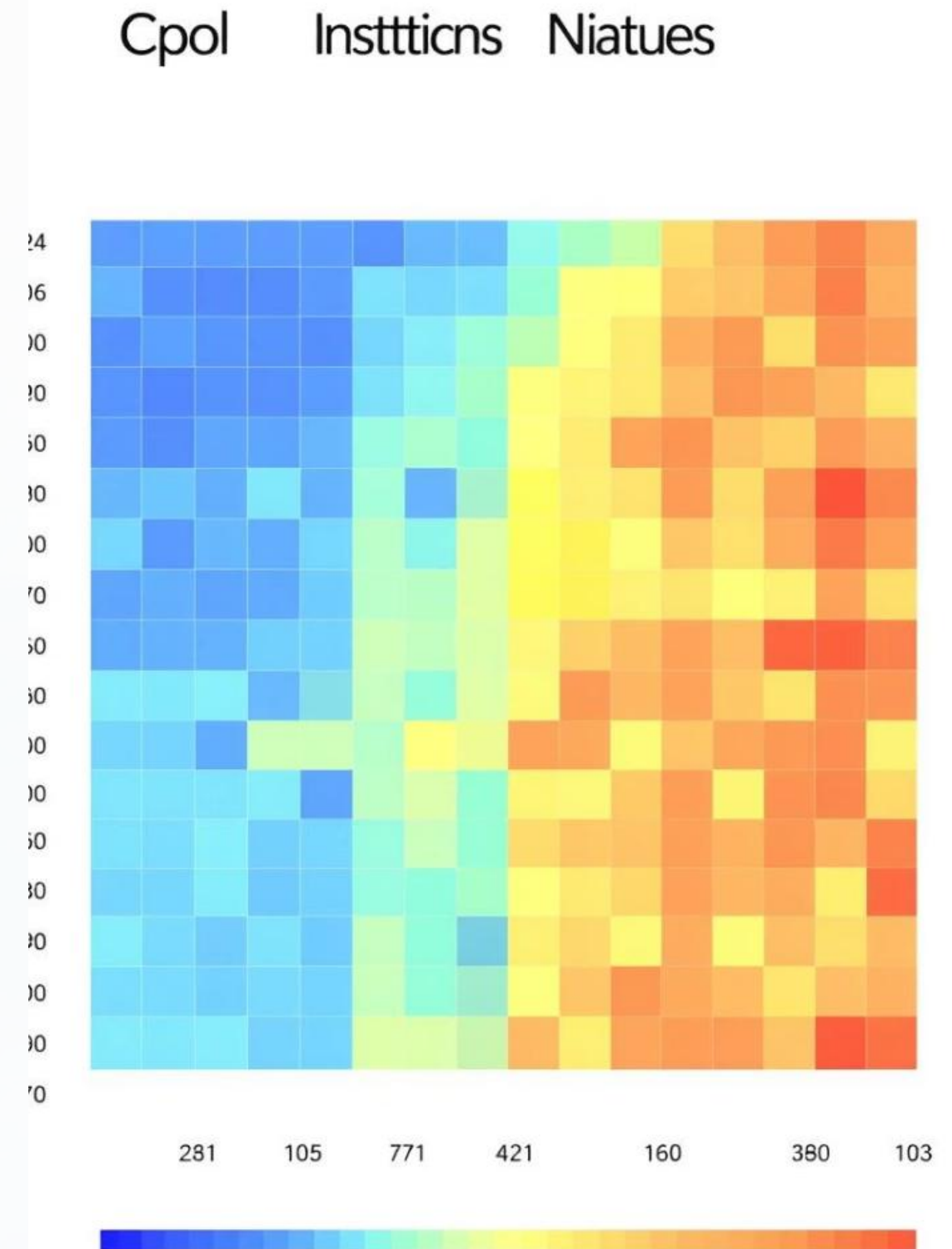
Tahun pembangunan memiliki korelasi sedang-kuat.

LotArea: +0.26 (Sedang)

Luas lahan berkorelasi sedang dengan harga.

OverallCond: -0.10 (Lemah-Negatif)

Kondisi keseluruhan berkorelasi lemah negatif.



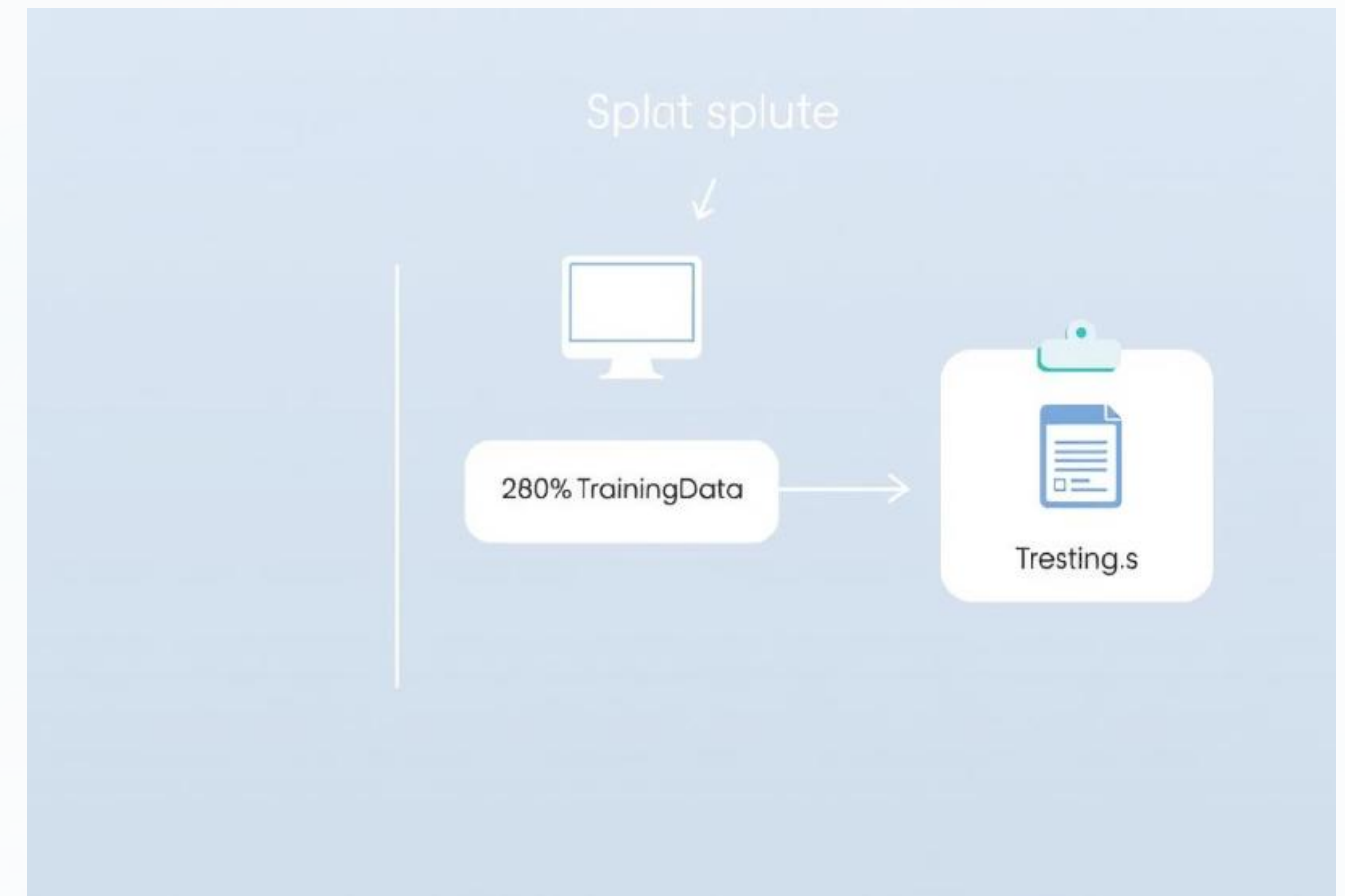
Implementasi Model Regresi Linier

Formula Regresi Linier

$$\text{SalePrice} = \beta_0 + \beta_1(\text{LotArea}) + \beta_2(\text{OverallCond}) + \beta_3(\text{YearBuilt}) + \beta_4(\text{TotalBsmtSF}) + \varepsilon$$

Pembagian Data

Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian.



Hasil Evaluasi Model

39.827,33
USD

MAE

Rata-rata kesalahan absolut
model.

58.576,75
USD

RMSE

Akar kuadrat rata-rata kesalahan
kuadrat.

0.55

R^2 Score

Model menjelaskan 55% variasi
harga.

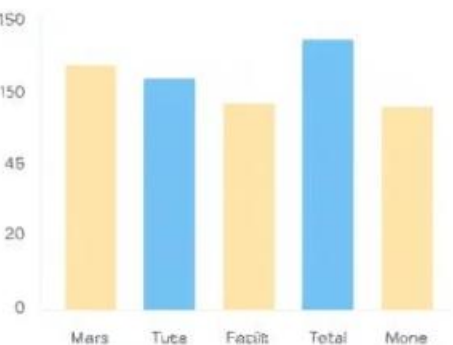
evaluation mest

Process meatly yr inredpless.

Emails

Asif-ings

Conversion Rate TM



| | | |
|------------|-----------|---------------|
| Sales | | |
| Gonace | \$305 kan | \$,547,300K |
| Tasteol | \$205 kan | \$15,300,390K |
| Trand Room | \$500 kan | \$3,566,200K |

Autistier Met: TM



Sale Quarter



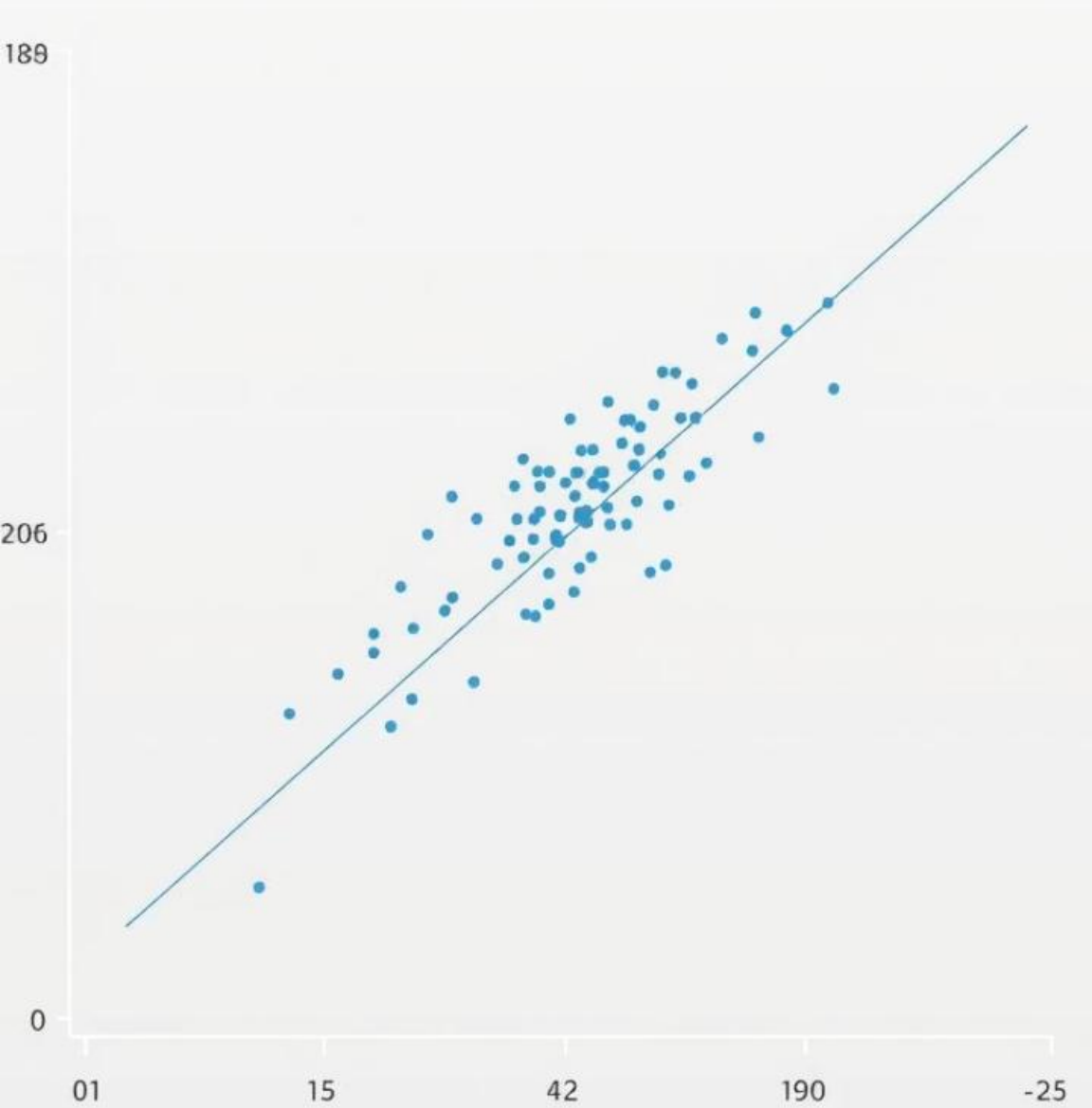
+ \$155 Acquloristion Cost TM

✓ Monthly Recurring Revenue

\$ \$113,0000

\$ Ary Moard

\$ \$112,00,000



Visualisasi Hasil Prediksi

Scatter plot menunjukkan harga aktual vs. prediksi.

Pola sebaran mendekati garis $y=x$, menunjukkan akurasi baik.

Namun, masih ada penyimpangan pada harga tinggi (outlier).

Kelebihan dan Keterbatasan Model

Kelebihan

- Mudah dipahami dan diimplementasikan.
- Cocok untuk hubungan linier.
- Baseline model yang baik.

Keterbatasan

- Tidak menangkap hubungan non-linear.
- Sensitif terhadap outlier.
- Kurang optimal untuk fitur kompleks.

Perbandingan dengan Studi Sebelumnya



Vaibhav Badne et al. (2022)

$R^2 > 0.70$ untuk prediksi harga di India.



Jaykumar Parekh Parekh (2023)

$R^2 = 0.73$ dengan fitur serupa.



Ayushi Bhagat et al. (2023)

Regresi Linier sebagai model baseline.

Model kami memberikan hasil yang kompetitif dan konsisten dengan studi sebelumnya.

Kesimpulan dan Saran Pengembangan

Kesimpulan

- Model Regresi Linier berhasil dibangun dengan $R^2 = 0.55$.
- TotalBsmtSF dan YearBuilt paling berpengaruh.
- Model praktis, cepat, dan mudah diinterpretasikan.
- Cocok sebagai baseline model untuk properti.

Saran Pengembangan

- Perluasan fitur: GrLivArea, OverallQual, Neighborhood.
- Algoritma lanjutan: Random Forest, XGBoost.
- Preprocessing: Deteksi outlier, transformasi logaritmik.
- Validasi: Cross-validation.
- Aplikasi: Sistem berbasis web/mobile.