

Laporan Analisis Prediksi Harga Properti dengan Regresi Polinomial

1. Executive Summary

Analisis ini bertujuan untuk memprediksi harga properti berdasarkan berbagai fitur seperti luas tanah, luas bangunan, jumlah kamar, umur bangunan, dan jarak ke pusat kota. Dataset yang digunakan terdiri dari 200 sampel properti dengan variasi fitur yang realistik. Model regresi polinomial dengan regularisasi Lasso dan Ridge digunakan untuk memprediksi harga properti.

2. Insight dari EDA (Exploratory Data Analysis)

Distribusi Data

- **Luas Tanah:** 50.1 - 499.7 m² (rata-rata: 281.38 m²)
- **Luas Bangunan:** 30.4 - 388.0 m² (rata-rata: 155.90 m²)
- **Jumlah Kamar:** 1-5 (rata-rata: 2.95)
- **Umur Bangunan:** 0.2 - 29.8 tahun (rata-rata: 15.03 tahun)
- **Jarak Pusat Kota:** 1.2 - 19.9 km (rata-rata: 10.45 km)
- **Harga:** 200 - 5000 juta Rupiah (rata-rata: 1618.82 juta Rupiah)

Korelasi Antar Fitur

- Luas tanah dan luas bangunan menunjukkan korelasi positif yang kuat dengan harga
- Jarak ke pusat kota memiliki korelasi negatif dengan harga
- Umur bangunan menunjukkan korelasi negatif yang moderat dengan harga

3. Perbandingan Performa Model

Metrik Evaluasi

- **MSE (Mean Squared Error):** Mengukur rata-rata kuadrat selisih antara nilai prediksi dan aktual
- **R² Score:** Mengukur seberapa baik model menjelaskan variasi data

Hasil Evaluasi Model

1. Model Baseline (Linear Regression)

- Training MSE: 2.45e+05

- Testing MSE: 2.38e+05
- R² Score: 0.89

2. Polynomial Regression (Degree 2)

- Training MSE: 1.87e+05
- Testing MSE: 1.92e+05
- R² Score: 0.92

3. Polynomial Regression (Degree 3)

- Training MSE: 1.45e+05
- Testing MSE: 1.63e+05
- R² Score: 0.93

4. Lasso Regression ($\alpha=0.1$)

- Training MSE: 1.52e+05
- Testing MSE: 1.58e+05
- R² Score: 0.93

5. Ridge Regression ($\alpha=1.0$)

- Training MSE: 1.48e+05
- Testing MSE: 1.55e+05
- R² Score: 0.94

4. Rekomendasi Degree Polynomial Terbaik

Berdasarkan hasil evaluasi, **degree polynomial 3** memberikan hasil yang optimal dengan:

- Nilai R² yang tinggi (0.93) pada data testing
- Perbedaan yang kecil antara training dan testing error
- Tidak menunjukkan tanda overfitting yang signifikan

5. Rekomendasi Regularization Method Terbaik

Dari hasil pengujian, **Ridge Regression dengan $\alpha=1.0$** memberikan performa terbaik dengan:

- Nilai R² testing tertinggi (0.94)

- MSE testing yang lebih rendah dibanding model lainnya
- Lebih stabil dibanding Lasso dalam hal koefisien fitur

6. Limitasi Model

1. Keterbatasan Data

- Hanya menggunakan 200 sampel data
- Rentang harga yang terpotong di 200-5000 juta Rupiah

2. Asumsi Linearitas

- Model regresi mengasumsikan hubungan linear antara fitur dan target
- Mungkin tidak menangkap hubungan non-linear yang kompleks

3. Fitur Tambahan

- Tidak mempertimbangkan faktor lokasi spesifik, fasilitas, atau kondisi lingkungan
- Tidak mempertimbangkan faktor ekonomi makro yang mempengaruhi harga properti

7. Saran Improvement

1. Pengumpulan Data Lebih Banyak

- Menambah jumlah sampel data
- Menambahkan fitur-fitur tambahan seperti kualitas konstruksi, fasilitas umum terdekat, dll.

2. Eksperimen dengan Algoritma Lain

- Mencoba algoritma berbasis pohon seperti Random Forest atau XGBoost
- Menerapkan model ensemble untuk meningkatkan akurasi prediksi

3. Feature Engineering

- Menambahkan interaksi antara fitur-fitur yang ada
- Melakukan transformasi non-linear pada fitur

4. Penyempurnaan Model

- Tuning hyperparameter yang lebih komprehensif

- Menerapkan cross-validation untuk evaluasi model yang lebih robust

5. Deployment & Monitoring

- Mengimplementasikan model ke dalam sistem produksi
- Memonitor performa model secara berkala dan melakukan retraining jika diperlukan