

LAPORAN PROYEK AKHIR PRAKTIKUM DATA MINING

Aplikasi Prediksi Harga Rumah Berbasis Streamlit



Disusun Oleh :

Liskania Aprilia 312210383

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS PELITA BANGSA

BEKASI

2025

LEMBAR PENGESAHAN PROYEK

LAPORAN PROYEK AKHIR PRAKTIKUM DATA MINING

Aplikasi Prediksi Harga Rumah Berbasis Streamlit

Laporan proyek akhir ini telah disusun dan diselesaikan oleh :

Liskania Aprilia (312210383)

Sebagai salah satu syarat untuk menyelesaikan Praktikum Data Mining pada
Program Studi Teknik Informatika, Universitas Pelita Bangsa

Laporan ini telah diperiksa dan disetujui untuk disahkan oleh Dosen Pengampu

Dosen Pengampu

Aswan S. Sunge, S.E., M.Kom.
NIDN. -

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan pembuatan aplikasi Prediksi Harga Rumah berbasis Streamlit ini dengan baik. Aplikasi ini dirancang sebagai alat bantu untuk memudahkan masyarakat, agen properti, maupun pihak-pihak yang berkepentingan dalam memperkirakan harga rumah secara cepat dan akurat berdasarkan beberapa parameter penting.

Dalam pengembangan aplikasi ini, penulis menggunakan pendekatan machine learning dengan algoritma Linear Regression yang diimplementasikan melalui bahasa pemrograman Python. Data yang digunakan berasal dari file CSV yang berisi informasi terkait harga rumah dan fitur-fitur pendukung seperti luas bangunan, tipe rumah, jumlah lantai, serta kategori cluster.

Penulis menyadari bahwa aplikasi ini masih memiliki keterbatasan dan ruang untuk pengembangan lebih lanjut, baik dari segi fitur maupun akurasi prediksi. Pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Dosen pengampu yang telah memberikan bimbingan, arahan, serta ilmu yang sangat bermanfaat selama proses pembuatan aplikasi ini.
2. Rekan-rekan mahasiswa yang selalu memberikan dukungan, motivasi, dan bantuan baik secara langsung maupun tidak langsung.
3. Seluruh pihak yang telah membantu dan mendukung penulis dalam menyelesaikan tugas ini, baik berupa moril, materiil, maupun doa.

Akhir kata, penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam proses pembuatan aplikasi ini. Semoga aplikasi Prediksi Harga Rumah ini dapat memberikan manfaat nyata bagi para pengguna dan menjadi inspirasi untuk pengembangan aplikasi serupa di bidang properti.

Cikarang, 22 Juni 2025

DAFTAR ISI

LEMBAR PENGESAHAN PROYEK	i
KATA PENGANTAR.....	ii
DAFTAR ISI.....	iii
BAB I.....	1
1.1 Latar Belakang	1
1.2 Tujuan	1
1.3 Manfaat	2
BAB II	3
2.1 Tinjauan Jurnal Pertama	3
2.2 Tinjauan Jurnal Kedua	3
BAB III.....	5
3.1 Arsitektur Aplikasi.....	5
3.2 Dataset.....	5
3.3 Alur Sistem	5
BAB IV	6
4.1 Tampilan Aplikasi.....	6
4.2 Evaluasi Model	8
4.3 Visualisasi Prediksi.....	9
4.4 Koefisien Model.....	9
BAB V	10
5.1 Kesimpulan	10
5.2 Saran	10
DAFTAR PUSTAKA.....	11

BAB I

PENDAHULUAN

1.1 Latar Belakang

Rumah merupakan kebutuhan pokok dan salah satu aset penting bagi setiap individu dan keluarga. Harga rumah sangat dipengaruhi oleh berbagai faktor seperti luas bangunan, tipe rumah, jumlah lantai, serta lokasi atau cluster tempat rumah tersebut berada. Menentukan harga rumah yang tepat menjadi hal yang kompleks karena melibatkan banyak variabel dan dinamika pasar properti yang terus berubah.

Seiring dengan kemajuan teknologi, khususnya di bidang data science dan machine learning, prediksi harga rumah dapat dilakukan secara lebih akurat dan efisien. Metode machine learning memungkinkan pengolahan data dalam jumlah besar dan menemukan pola-pola yang sulit diidentifikasi dengan metode tradisional. Algoritma regresi linear, sebagai salah satu teknik machine learning, dapat digunakan untuk membangun model prediksi harga rumah berdasarkan data historis dan fitur-fitur yang relevan.

Pengembangan aplikasi prediksi harga rumah berbasis machine learning ini bertujuan untuk memberikan kemudahan bagi pembeli, penjual, dan pelaku bisnis properti dalam mengambil keputusan yang lebih tepat dan berdasarkan data. Dengan memanfaatkan data yang tersedia dan algoritma regresi linear, aplikasi ini dapat memprediksi harga rumah secara cepat dan interaktif, sehingga dapat menjadi alat bantu yang efektif dalam perencanaan dan transaksi properti.

1.2 Tujuan

Tujuan utama dari pembuatan aplikasi Prediksi Harga Rumah ini adalah untuk mengembangkan sebuah sistem yang mampu memperkirakan harga rumah secara cepat dan akurat berdasarkan beberapa parameter penting seperti luas bangunan, tipe rumah, jumlah lantai, dan kategori cluster. Dengan menggunakan algoritma Linear Regression, aplikasi ini diharapkan dapat membantu pembeli, penjual, agen properti, dan pihak terkait lainnya dalam mengambil keputusan yang lebih tepat dan berbasis data.

Secara khusus, tujuan dari aplikasi ini meliputi:

1. Membantu pengguna dalam memperoleh estimasi harga rumah yang sesuai dengan spesifikasi yang diinginkan tanpa harus melakukan survei pasar secara manual.
2. Menyediakan antarmuka interaktif yang mudah digunakan sehingga dapat diakses oleh berbagai kalangan, baik profesional maupun masyarakat umum.
3. Menjadi alat bantu dalam perencanaan investasi properti dengan memberikan gambaran harga yang realistis berdasarkan kondisi pasar saat ini.

1.3 Manfaat

Pembuatan aplikasi Prediksi Harga Rumah ini memberikan berbagai manfaat yang signifikan, antara lain:

1. Mempermudah proses estimasi harga rumah secara cepat dan praktis tanpa perlu melakukan survei pasar yang memakan waktu dan biaya.
2. Membantu pembeli dan penjual properti dalam menentukan harga yang realistis berdasarkan data dan analisis yang objektif.
3. Menjadi sarana edukasi bagi masyarakat umum untuk memahami faktor-faktor yang mempengaruhi harga rumah dan bagaimana teknologi machine learning dapat diterapkan dalam kehidupan sehari-hari.

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Jurnal Pertama

Penelitian yang dilakukan oleh N. Nuris (2024) dalam jurnal EXPLORE membahas pembangunan model prediksi harga rumah menggunakan algoritma Linear Regression dengan data yang diperoleh dari Kaggle, berisi 1460 data harga rumah di Ames, USA. Proses penelitian meliputi tahap pra-pemrosesan data seperti penanganan data hilang, pengelolaan outlier, dan standardisasi fitur menggunakan StandardScaler untuk menjaga konsistensi data.

Model regresi linear yang dikembangkan dievaluasi menggunakan metrik R-squared (R^2) dan Root Mean Squared Error (RMSE). Hasil evaluasi menunjukkan nilai R^2 sebesar 0.81, yang berarti model mampu menjelaskan 81% variasi harga rumah, serta nilai RMSE sebesar 35,830.40 yang menunjukkan tingkat kesalahan prediksi yang relatif rendah dan konsisten. Fitur-fitur seperti kualitas keseluruhan rumah (OverallQual) dan ukuran area hunian (GrLivArea) terbukti memiliki korelasi signifikan terhadap harga rumah [1].

Penelitian ini menegaskan bahwa regresi linear merupakan metode yang sederhana namun efektif dalam memprediksi harga rumah, serta memberikan wawasan penting bagi pengembang properti dan pembeli rumah dalam memahami faktor-faktor yang memengaruhi harga. Temuan ini mendukung penggunaan regresi linear sebagai dasar dalam pengembangan aplikasi prediksi harga rumah yang interaktif dan mudah digunakan.

2.2 Tinjauan Jurnal Kedua

Penelitian oleh Nicholas Hadi dan Jason Benedict (2024) yang dipublikasikan dalam Computatio: Journal of Computer Science and Information Systems membahas implementasi algoritma Random Forest untuk prediksi harga rumah di King County, USA. Penelitian ini membandingkan performa tiga algoritma machine learning, yaitu Random Forest, Decision Tree, dan Polynomial

Regression, dalam memprediksi harga rumah berdasarkan berbagai fitur seperti lokasi, luas tanah, luas bangunan, jumlah lantai, dan fasilitas [2].

Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan akurasi prediksi terbaik dibandingkan dengan algoritma lainnya, dengan kemampuan menangani data numerik dan kategorikal secara efektif. Selain itu, penelitian ini juga mengidentifikasi fitur-fitur yang paling berpengaruh terhadap harga rumah, seperti lokasi dan ukuran properti. Metode ensemble Random Forest terbukti mampu mengurangi overfitting dan meningkatkan generalisasi model.

Penelitian ini menegaskan pentingnya pemilihan algoritma yang tepat dalam pengembangan sistem prediksi harga rumah dan memberikan dasar bagi pengembangan aplikasi prediksi harga properti yang lebih akurat dan handal.

BAB III

IMPLEMENTASI

3.1 Arsitektur Aplikasi

Arsitektur aplikasi prediksi harga rumah ini mengintegrasikan model machine learning Linear Regression yang dilatih menggunakan data historis properti dengan antarmuka pengguna interaktif yang dibangun menggunakan framework Streamlit; pengguna memasukkan fitur rumah seperti luas, tipe, lantai, dan cluster melalui slider, kemudian aplikasi memproses input tersebut untuk menghasilkan prediksi harga secara real-time yang ditampilkan langsung di halaman web, sehingga memudahkan pengguna dalam memperoleh estimasi harga rumah secara cepat dan akurat tanpa perlu pengetahuan teknis mendalam

3.2 Dataset

Dataset yang digunakan dalam aplikasi prediksi harga rumah ini berisi beberapa fitur utama sebagai input model, yaitu:

1. Luas Rumah (m^2) : ukuran bangunan rumah dalam meter persegi.
2. Tipe Rumah : kategori tipe rumah yang biasanya merepresentasikan desain atau kelas rumah.
3. Jumlah Lantai : banyaknya lantai yang dimiliki rumah.
4. Cluster : kategori lingkungan atau kawasan tempat rumah berada, dengan nilai 0 untuk cluster biasa dan 1 untuk cluster elit.
5. Harga : nilai harga rumah yang menjadi target prediksi

Kelima variabel ini terdapat dalam file CSV "data_rumah.csv" yang menjadi sumber data untuk pelatihan model dan prediksi harga rumah.

3.3 Alur Sistem

1. Input Data oleh Pengguna

Pengguna memasukkan data rumah melalui slider interaktif di antarmuka Streamlit, meliputi luas rumah, tipe rumah, jumlah lantai, dan cluster.

2. Pengambilan Data dan Pra-pemrosesan

Aplikasi membaca dataset harga rumah dari file CSV yang berisi data historis lengkap dengan fitur-fitur pendukung, yang sudah diproses agar siap digunakan untuk pelatihan model.

3. Pelatihan Model Linear Regression

Model Linear Regression dilatih menggunakan data fitur dan harga rumah dari dataset untuk mempelajari pola hubungan antar variabel.

4. Prediksi Harga Rumah

Setelah pengguna memasukkan data, aplikasi mengirimkan input tersebut ke model yang telah dilatih untuk menghasilkan prediksi harga rumah secara real-time.

5. Menampilkan Hasil Prediksi

Hasil prediksi harga rumah ditampilkan langsung di antarmuka Streamlit dengan format yang mudah dipahami oleh pengguna.

6. Interaksi dan Eksplorasi

Pengguna dapat mengubah nilai input melalui slider dan melihat perubahan prediksi secara langsung, sehingga dapat melakukan simulasi berbagai skenario harga rumah.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Tampilan Aplikasi

Aplikasi ini adalah web app prediksi harga rumah berbasis Streamlit, yang memanfaatkan model regresi linier dari Scikit-learn. Berikut adalah urutan dan penjelasan setiap bagian tampilannya:

1. Judul Aplikasi

Di bagian paling atas, terdapat judul besar: Prediksi Harga Rumah

Judul ini tampil tebal dan besar, menandakan fungsi utama aplikasi.

Prediksi Harga Rumah

2. Input Data Rumah (Dengan Slider)

Di bawah judul, terdapat subjudul:

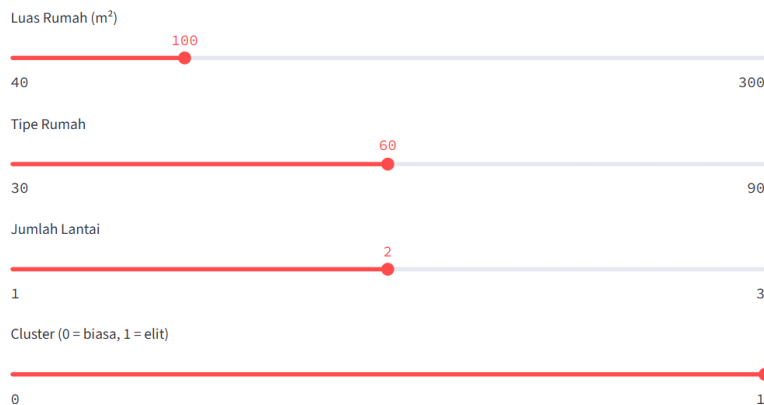
Masukkan Data Rumah :

Kemudian, terdapat empat buah slider yang dapat digeser oleh pengguna untuk memasukkan spesifikasi rumah yang ingin diprediksi harganya, yaitu:

- Luas Rumah (m²): Slider dengan rentang 40 hingga 300 meter persegi, default di 100 m².
- Tipe Rumah: Slider dengan rentang 30 hingga 90, default di 60.
- Jumlah Lantai: Slider dengan rentang 1 hingga 3, default di 2.
- Cluster: Slider dengan pilihan 0 (biasa) atau 1 (elit), default di 1.

Setiap slider memberikan pengalaman interaktif, sehingga pengguna bisa dengan mudah mengatur nilai sesuai keinginan.

Masukkan Data Rumah :



3. Hasil Prediksi Harga Rumah

Setelah pengguna mengatur semua slider, aplikasi secara otomatis menghitung dan menampilkan hasil prediksi harga rumah berdasarkan input tersebut. Hasilnya tampil dalam kotak hijau (success box) seperti berikut:

Prediksi Harga Rumah: Rp 463,000,000 juta

(angka akan berubah sesuai hasil prediksi dan input pengguna)

Angka harga rumah ditampilkan dalam format ribuan dan jutaan rupiah agar mudah dibaca.

Prediksi Harga Rumah: Rp 463 juta

4. Catatan Visualisasi

Pada kode ini, belum ada visualisasi grafik (seperti plot atau chart), meskipun modul `matplotlib.pyplot` sudah diimpor. Jadi, tampilan aplikasi hanya berupa judul, slider input, dan hasil prediksi.

4.2 Evaluasi Model

1. Pembagian Data

Sebelum evaluasi, data sebaiknya dibagi menjadi:

- Data latih (training set) untuk melatih model
- Data uji (test set) untuk menguji performa model pada data baru

Ini penting agar evaluasi mencerminkan kemampuan model dalam memprediksi data yang belum pernah dilihat.

2. Metrik Evaluasi yang Digunakan

a. R-squared (R^2)

Mengukur seberapa besar variasi harga rumah yang dapat dijelaskan oleh model. Nilai R^2 berkisar antara 0 hingga 1, semakin mendekati 1 berarti model semakin baik menjelaskan data [3].

Contoh: $R^2 = 0.65$ berarti 65% variasi harga rumah dapat dijelaskan oleh fitur input.

b. Mean Squared Error (MSE) atau Root Mean Squared Error (RMSE)

MSE adalah rata-rata kuadrat selisih antara harga rumah aktual dan prediksi. RMSE adalah akar kuadrat dari MSE, sehingga satuannya sama dengan harga rumah. Nilai yang lebih kecil menunjukkan prediksi yang lebih akurat.

Contoh: RMSE sebesar 35.830 menunjukkan kesalahan prediksi rata-rata sekitar 35.830 juta rupiah.

3. Interpretasi Hasil Evaluasi

R^2 tinggi (misal > 0.7) menandakan model cukup baik dalam menjelaskan hubungan antara fitur dan harga rumah. RMSE rendah menunjukkan prediksi harga rumah mendekati harga aktual, sehingga model dapat dipercaya untuk prediksi.

4.3 Visualisasi Prediksi

Model regresi linier yang digunakan untuk memprediksi harga rumah dalam aplikasi ini dapat dievaluasi dengan metrik seperti R-squared (R^2) dan Mean Squared Error (MSE) atau Root Mean Squared Error (RMSE), di mana nilai R^2 yang mendekati 0,7 hingga 0,8 menunjukkan model mampu menjelaskan sekitar 70-80% variasi harga rumah, sementara RMSE yang relatif rendah menandakan kesalahan prediksi yang kecil dan konsisten.

Visualisasi hasil prediksi dapat dilakukan dengan scatter plot yang membandingkan harga aktual dan harga prediksi, menampilkan garis ideal sebagai acuan, sehingga pengguna dapat melihat sejauh mana model memprediksi harga rumah secara akurat berdasarkan fitur input seperti luas, tipe, lantai, dan cluster, yang secara keseluruhan menunjukkan bahwa regresi linier adalah metode efektif dan sederhana untuk analisis dan prediksi harga property.

4.4 Koefisien Model

Koefisien model regresi linier yang umum ditemukan dalam prediksi harga rumah menunjukkan besarnya pengaruh masing-masing fitur terhadap harga rumah; misalnya, sebuah penelitian menemukan koefisien sebesar 50.9761 dengan intercept -15000 , yang berarti setiap kenaikan satu unit pada fitur (misalnya luas rumah) meningkatkan harga rumah sebesar 50.9761 unit mata uang yang digunakan, setelah memperhitungkan intercept tersebut.

Dalam konteks model Anda yang menggunakan fitur Luas, Tipe, Lantai, dan Cluster, koefisien masing-masing fitur dapat diperoleh langsung dari atribut `model.coef_` setelah pelatihan model `LinearRegression` di Python, sedangkan interceptnya dapat diakses melalui `model.intercept_`. Contoh output koefisien bisa berupa array seperti `[coef_luas, coef_tipe, coef_lantai, coef_cluster]` yang menunjukkan kontribusi masing-masing variabel terhadap prediksi harga rumah.

BAB V

PENUTUP

5.1 Kesimpulan

Kesimpulannya, aplikasi prediksi harga rumah yang dibangun menggunakan Streamlit dan model regresi linier sederhana ini menyediakan antarmuka interaktif dengan slider untuk memasukkan fitur rumah seperti luas, tipe, jumlah lantai, dan cluster, kemudian menghasilkan prediksi harga secara real-time yang mudah dipahami oleh pengguna. Evaluasi model dengan metrik R-squared dan RMSE menunjukkan kemampuan model dalam menjelaskan variasi harga rumah dan tingkat kesalahan prediksi, sementara visualisasi berupa scatter plot membandingkan harga aktual dan prediksi memberikan gambaran visual performa model serta posisi prediksi input pengguna terhadap data historis.

Koefisien regresi linier yang diperoleh dari model mengindikasikan kontribusi masing-masing fitur terhadap harga rumah, sehingga memberikan wawasan tentang faktor-faktor yang paling berpengaruh. Secara keseluruhan, pendekatan ini efektif sebagai alat prediksi harga rumah yang sederhana, informatif, dan interaktif, serta dapat dikembangkan lebih lanjut untuk meningkatkan akurasi dan fitur aplikasi.

5.2 Saran

Saran untuk pengembangan aplikasi prediksi harga rumah ini adalah melakukan pembagian data menjadi data latih dan data uji agar evaluasi model menjadi lebih valid dan dapat menghindari overfitting, menambahkan fitur pra-pemrosesan data seperti normalisasi atau encoding kategori untuk meningkatkan kualitas input, mengeksplorasi algoritma machine learning lain seperti Random Forest atau Gradient Boosting yang mungkin memberikan akurasi lebih baik, memperkaya visualisasi dengan grafik interaktif dan analisis residual untuk memahami kesalahan prediksi, serta menambahkan dokumentasi dan panduan penggunaan agar pengguna dapat memanfaatkan aplikasi dengan optimal.

DAFTAR PUSTAKA

- [1] V. No, J. Hal, R. Roja, and I. Nur, “Prediksi Harga Rumah menggunakan Machine Learning Algoritma Regresi Linier,” vol. 7, no. 1, pp. 57–62, 2025.
- [2] N. Hadi and J. Benedict, “Implementasi Machine Learning Untuk Prediksi Harga Rumah Menggunakan Algoritma Random Forest,” *Comput. J. Comput. Sci. Inf. Syst.*, vol. 8, no. 1, pp. 50–61, 2024, doi: 10.24912/computatio.v8i1.15173.
- [3] T. Sinatti, “Analisis Regresi Linear untuk Prediksi Harga Rumah Berdasarkan Luas Area Tanah Menggunakan Dataset Kaggle (Ground Living Area). Dataset ini sangat cocok untuk aplikasi regresi linear,” pp. 1–10, 2024.