

LAPORAN PROYEK AKHIR PRATIKUM DATA MINING 2025

ANALISIS HARGA RUMAH DI INDONESIA MENGUNAKAN STREAMLIT DAN ALGORITMA DATA MINING



Disusun oleh :

1. Nurul Asma As-shidiq (312210325)
2. Alfina Damayanti (312210379)
3. Henfi Putri Intan (312210354)

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS PELITA BANGSA
BEKASI**

2025

LEMBAR PENGESAHAN
LAPORAN PROYEK AKHIR PRATIKUM DATA MINING 2025

**ANALISIS HARGA RUMAH DI INDONESIA MENGGUNAKAN
STREAMLIT DAN ALGORITMA DATA MINING**

Diajukan untuk memenuhi salah satu syarat menyelesaikan praktikum pada
Program Studi Teknik Informatika Universitas Pelita Bangsa

Disusun oleh :

1. Nurul Asma As-shidiq (312210325)
2. Alfina Damayanti (312210379)
3. Henfi Putri Intan (312210354)

Laporan ini telah diperiksa dan disetujui untuk disahkan oleh Dosen Pengampu

Dosen Pengampu:

Asep Muhidin, S.Kom., M.Kom.

NIDN

BAB I

PENDAHULUAN

1.1 Latar belakang

Analisis harga rumah di Indonesia menjadi penting seiring dengan pertumbuhan populasi dan urbanisasi yang pesat. Penggunaan Streamlit dan algoritma data mining memungkinkan visualisasi data yang interaktif dan pengolahan informasi yang efisien untuk memahami faktor-faktor yang mempengaruhi harga rumah.

Kemajuan teknologi informasi telah mengubah cara manusia dalam mengelola dan memanfaatkan data. Salah satu bidang ilmu komputer yang berkembang pesat adalah data mining, yaitu teknik menemukan pola dan informasi penting dari kumpulan data besar. Dalam berbagai sektor, termasuk properti, data mining dimanfaatkan untuk menganalisis tren harga rumah, mengenali faktor-faktor yang memengaruhi harga, hingga memperkirakan nilai properti di masa depan.

Streamlit sebagai alat untuk membangun aplikasi web interaktif memudahkan pengguna dalam mengakses dan memahami data, sementara algoritma data mining dapat mengidentifikasi pola dan tren yang mungkin tidak terlihat secara langsung. Melalui pendekatan ini, diharapkan dapat ditemukan faktor-faktor yang mempengaruhi harga rumah, seperti lokasi, fasilitas, dan kondisi pasar. Penelitian ini bertujuan untuk memberikan gambaran yang lebih jelas mengenai dinamika harga rumah di Indonesia dan membantu stakeholder dalam membuat keputusan yang lebih informasional.

1.2 Tujuan

1. Memberikan pengalaman langsung dalam menerapkan metode data mining pada analisis harga rumah di Indonesia.
2. Mahasiswa diharapkan dapat mengimplementasikan algoritma supervised dan unsupervised learning, seperti Regresi Linier dan K-Means Clustering, dalam aplikasi web menggunakan Streamlit.

3. Mahasiswa tidak hanya memahami teori, tetapi juga mampu mengaplikasikannya dalam sistem nyata yang dapat diuji secara langsung.

1.3 Manfaat

1. Mendapatkan pengalaman langsung dalam menerapkan metode data mining, yang membantu memahami bagaimana teori yang dipelajari di kelas diterapkan dalam situasi nyata.
2. Dengan menggunakan algoritma seperti Regresi Linier dan K-Means Clustering, mahasiswa belajar cara menganalisis data dan menemukan pola yang berguna dalam menentukan harga rumah.
3. Menggunakan Streamlit untuk membuat aplikasi web interaktif memberikan mahasiswa keterampilan teknis yang berharga, yang dapat meningkatkan daya saing mereka di dunia kerja.

BAB II

TINJAUAN PUSTAKA

2.1 Data Mining

Data mining merupakan proses untuk menemukan pola atau informasi penting dari kumpulan data yang sangat besar. Proses ini memanfaatkan teknik statistik, machine learning, dan basis data guna mengidentifikasi pola tersembunyi serta melakukan prediksi. Data mining sendiri adalah bagian dari tahapan knowledge discovery in databases (KDD), yang meliputi seleksi data, praproses, transformasi, penambangan data, hingga interpretasi hasil. Dalam penelitian ini, data mining digunakan untuk menganalisis harga rumah dengan mempertimbangkan atribut seperti luas bangunan, jumlah kamar, dan lokasi.

2.2 Regresi Linier

Regresi linier adalah teknik statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen dengan variabel dependen yang bersifat kontinu. Metode ini mencari garis lurus terbaik yang meminimalkan selisih kuadrat antara nilai prediksi dan nilai aktual. Dalam konteks prediksi harga rumah, regresi linier membantu memperkirakan harga berdasarkan atribut seperti luas tanah dan jumlah kamar, serta menghasilkan koefisien yang menunjukkan pengaruh masing-masing variabel terhadap harga.

2.3 K-Means Clustering

K-Means adalah algoritma unsupervised learning yang berfungsi untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan fitur. Prosesnya diawali dengan menentukan sejumlah centroid secara acak, lalu data dikelompokkan ke centroid terdekat. Centroid diperbarui secara iteratif hingga hasilnya stabil. Dalam penelitian ini, K-Means digunakan untuk mengelompokkan rumah berdasarkan fitur seperti luas dan lokasi, sehingga dapat diketahui segmen pasar atau kategori rumah berdasarkan harga.

2.4 Algoritma Klasifikasi

Beberapa algoritma klasifikasi yang digunakan antara lain:

- Naive Bayes: Algoritma probabilistik yang didasarkan pada Teorema Bayes dan mengasumsikan independensi antar fitur. Cocok untuk data kategorikal dan sering digunakan untuk klasifikasi teks atau deteksi spam.
- Support Vector Machine (SVM): Algoritma yang memaksimalkan margin antar kelas untuk mendapatkan pemisahan terbaik.
- K-Nearest Neighbor (K-NN): Mengklasifikasikan data baru berdasarkan kedekatan jarak dengan data latih terdekat.
- Decision Tree: Menggunakan struktur pohon untuk membuat keputusan klasifikasi secara hierarkis.
- Regresi Logistik: Digunakan untuk klasifikasi biner dengan memodelkan probabilitas sebagai fungsi logistik dari variabel input.

2.5 Visualisasi dan Evaluasi Model

Hasil model divisualisasikan menggunakan confusion matrix (untuk klasifikasi), scatter plot (untuk regresi), dan visualisasi cluster (untuk K-Means). Evaluasi performa model dilakukan menggunakan metrik seperti akurasi, mean squared error (MSE), dan R-squared (R^2) guna mengukur tingkat ketepatan prediksi.

2.6 Streamlit

Streamlit adalah framework Python open-source yang memudahkan pembuatan aplikasi web interaktif, terutama untuk kebutuhan data science dan machine learning. Dengan sintaks sederhana, Streamlit memungkinkan pembuatan dashboard visualisasi dan aplikasi prediktif yang dapat langsung diakses melalui browser tanpa perlu keahlian HTML atau JavaScript. Kelebihan utama Streamlit terletak pada kemudahan integrasi dengan pustaka Python lain seperti Pandas, scikit-learn, dan Matplotlib, serta proses pengembangan yang cepat. Dalam penelitian ini, Streamlit digunakan untuk membangun aplikasi analisis harga rumah berbasis web

BAB III

IMPLEMENTASI

3.1 Deskripsi Umum

Aplikasi berbasis web ini dirancang untuk mengembangkan aplikasi interaktif yang memungkinkan pengguna menganalisis data harga rumah secara langsung. Tujuan utamanya adalah untuk mengeksplorasi faktor-faktor yang mempengaruhi harga rumah di Indonesia, menerapkan berbagai algoritma data mining untuk menemukan pola dan hubungan dalam data, serta menyediakan fitur prediksi harga berdasarkan input pengguna. Dengan demikian, praktikum ini bertujuan untuk membantu pembeli, penjual, dan investor dalam membuat keputusan yang lebih informasi terkait transaksi properti, serta meningkatkan pemahaman tentang bagaimana teknologi modern dapat digunakan untuk analisis data.

3.2 Alat dan Bahan

- Bahasa Pemrograman: Python 3.11
- Library: pandas, scikit-learn, matplotlib, seaborn, streamlit
- Platform Hosting: Streamlit Cloud
- Dataset: Data harga rumah Indonesia (.csv)

3.3 Desain Sistem

Aplikasi terdiri dari 4 fitur utama:

- Upload dataset
- Pemilihan algoritma
- Visualisasi serta hasil analisis
- Desain antarmuka dibuat dengan CSS sederhana agar tampilan tetap profesional dan modern.

3.4 Alur Pengguna

Pengguna dapat mengakses aplikasi melalui tautan yang tersedia, mengunggah dataset, memilih kolom target, lalu menentukan algoritma yang akan digunakan untuk analisis. Hasil analisis langsung ditampilkan dalam bentuk teks dan grafik.

3.5 Link Aplikasi

<https://harga-rumah-2025.streamlit.app>

QR CODE



SCAN ME

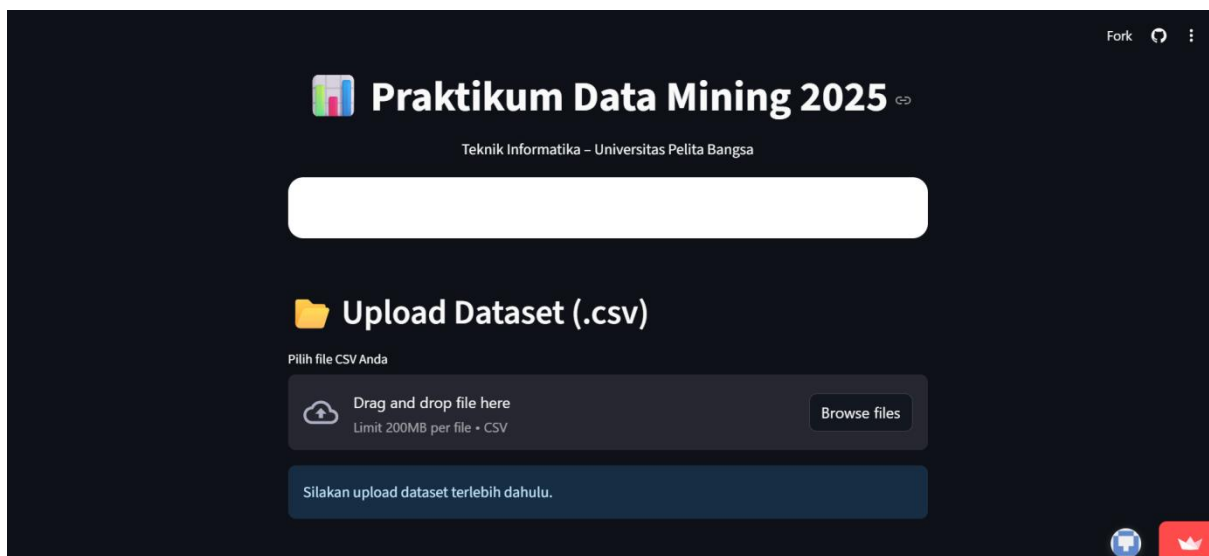
BAB IV

HASIL DAN PEMBAHASAN

4.1 Tampilan Aplikasi

Merupakan halaman awal atau beranda (homepage) dari aplikasi praktikum Data Mining 2025, yang bertujuan untuk:

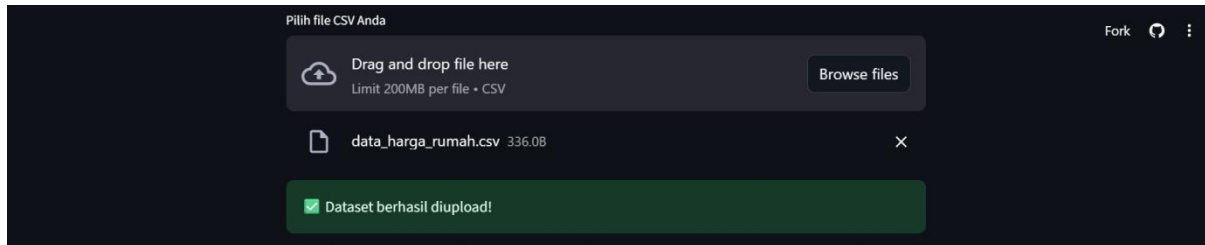
- Memberikan identitas aplikasi (judul, nama program studi, institusi).
- Menyediakan fitur untuk mengunggah file dataset berformat .csv.
- Menampilkan instruksi dan batasan ukuran file (maks 200MB).



Gambar 1. Tampilan Awal Streamlit web app interface

Tampilan ini muncul setelah pengguna berhasil mengunggah file data_harga_rumah.csv ke sistem. Tampak:

- Nama file yang berhasil diunggah (data_harga_rumah.csv) beserta ukurannya (336.0B).
- Tanda centang hijau dan pesan “Dataset berhasil diupload!” sebagai notifikasi bahwa proses upload telah sukses



Gambar 2. Tampilan Konfirmasi Upload Dataset (.csv)

Menampilkan kolom-kolom dari dataset:

- Luas_Rumah(m2), Jumlah_Kamar, Jenis_Lokasi, Harga_Jual, dan Label
- Setiap baris mewakili 1 data rumah.
- Tampilan seperti ini sering digunakan di aplikasi data mining berbasis web, terutama yang dibuat dengan Streamlit, Flask, atau Jupyter Notebook (dengan pandas).

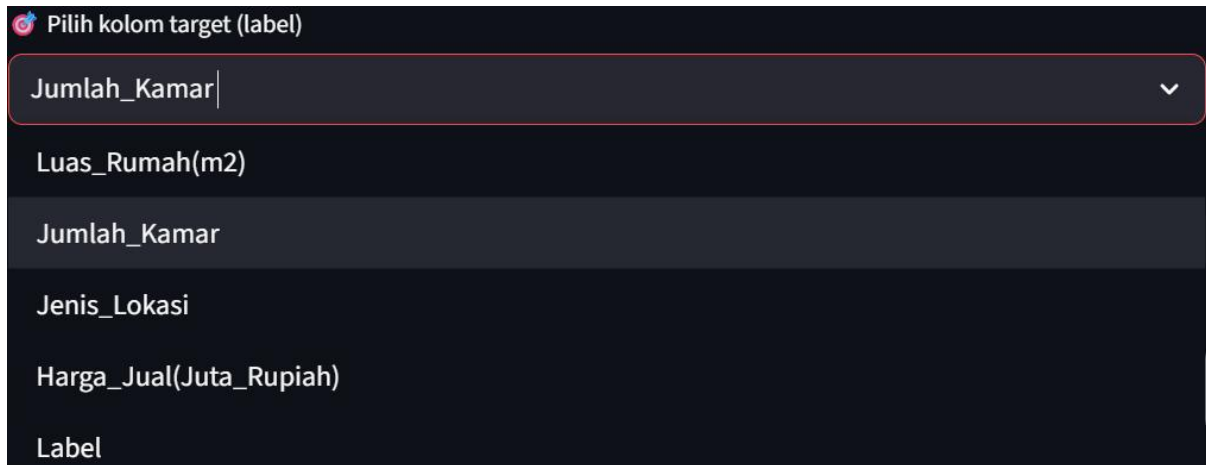
	Luas_Rumah(m2)	Jumlah_Kamar	Jenis_Lokasi	Harga_Jual(Juta_Rupiah)	Label
0	60	2	Perkotaan	750	Mahal
1	45	1	Pinggiran	480	Murah
2	85	3	Perkotaan	1200	Mahal
3	30	1	Perkotaan	350	Murah
4	100	4	Perkotaan	1450	Mahal
5	40	2	Pinggiran	500	Murah
6	70	2	Pinggiran	850	Mahal
7	55	3	Perkotaan	700	Murah
8	90	3	Perkotaan	1100	Mahal
9	35	1	Pinggiran	400	Murah

Gambar 3. Tampilan Preview atau Visualisasi Dataset

Tampilan ini digunakan untuk memilih kolom target dari dataset, yaitu kolom yang akan diprediksi dalam proses data mining, seperti klasifikasi atau regresi. Pada gambar:

- Dropdown berisi semua nama kolom dari dataset: Luas_Rumah(m2), Jumlah_Kamar, Jenis_Lokasi, Harga_Jual(Juta_Rupiah), dan Label.

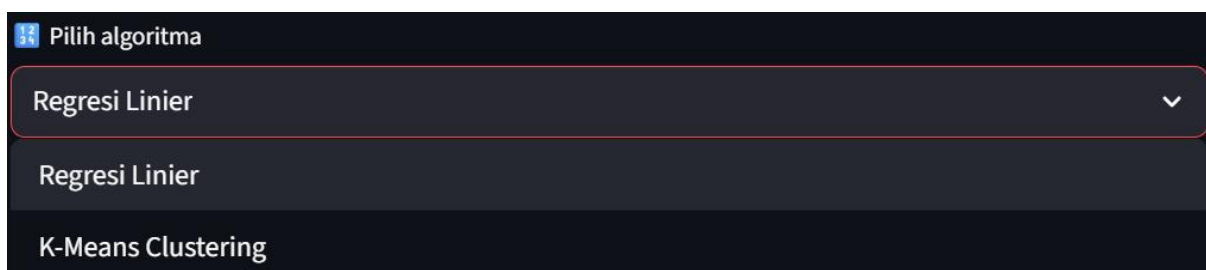
- Pengguna saat ini memilih "Jumlah_Kamar", meskipun secara umum yang paling tepat untuk target klasifikasi adalah kolom Label.



Gambar 4. Tampilan Pemilihan Kolom Target

Tampilan ini digunakan untuk memilih algoritma data mining yang akan digunakan dalam proses analisis dataset. Pengguna diberikan opsi melalui dropdown menu yang berisi jenis-jenis algoritma.

- Regresi Linier – digunakan untuk prediksi nilai numerik (misalnya: harga rumah).
- K-Means Clustering – digunakan untuk mengelompokkan data tanpa label (unsupervised learning).

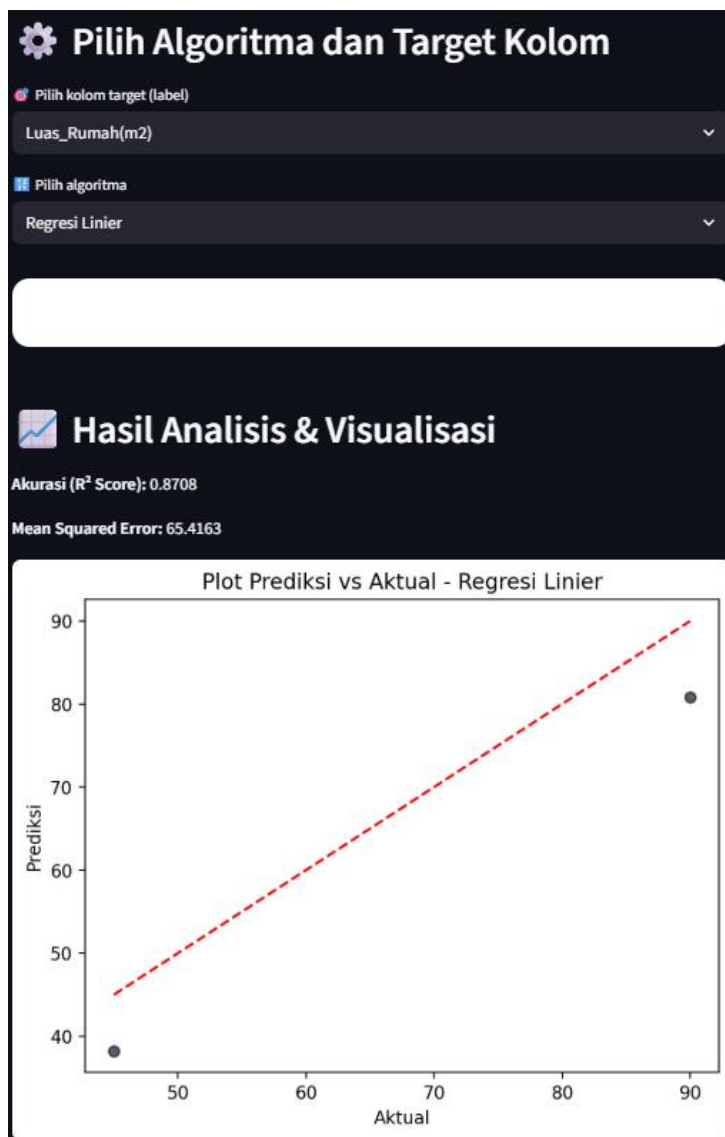


Gambar 5. Tampilan Pemilihan Algoritma

Tampilan ini muncul setelah pengguna memilih **kolom target** (dalam hal ini Luas_Rumah(m2)) dan algoritma **Regresi Linier** dari menu yang tersedia. Setelah model dijalankan, sistem menampilkan hasil evaluasi berupa metrik

akurasi dengan nilai **R^2 (R-Squared)** sebesar **0.8708**, yang menunjukkan bahwa sekitar 87% variasi data aktual dapat dijelaskan oleh model.

Selain itu, ditampilkan juga **Mean Squared Error (MSE)** sebesar 65.4163 yang mengindikasikan rata-rata kesalahan prediksi. Di bagian bawah terdapat grafik visualisasi berupa **plot Prediksi vs Aktual**, dengan garis merah putus-putus sebagai garis ideal (perfect fit) dan titik-titik hitam sebagai hasil prediksi sebenarnya. Grafik ini membantu pengguna memahami sejauh mana prediksi model mendekati nilai aktual pada data. Tampilan seperti ini sangat penting untuk memberikan umpan balik visual dan statistik terhadap performa model yang telah dipilih.



Gambar 6. Tampilan Output Analisis dan Visualisasi Model Regresi

BAB V

PENUTUP

Kesimpulan

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa pemanfaatan Streamlit sebagai antarmuka visual memudahkan proses eksplorasi dan analisis data harga rumah di Indonesia secara interaktif. Dengan mengunggah dataset harga rumah dan menerapkan algoritma data mining seperti Regresi Linier dan K-Means Clustering, sistem mampu melakukan prediksi dan pengelompokan data secara otomatis. Hasil yang diperoleh menunjukkan bahwa model regresi linier mampu memprediksi harga atau luas rumah dengan tingkat akurasi yang baik, sedangkan algoritma clustering dapat membantu mengidentifikasi pola kelompok harga rumah berdasarkan lokasi, jumlah kamar, dan luas bangunan. Secara keseluruhan, penggunaan teknologi Streamlit yang dipadukan dengan algoritma data mining terbukti efektif dalam membantu analisis harga rumah secara praktis, cepat, dan mudah dipahami oleh pengguna.

Saran

Pengembangan aplikasi dapat diarahkan pada:

- Penambahan algoritma lain seperti Random Forest atau XGBoost
- Fitur ekspor hasil analisis
- Validasi data sebelum analisis
- Evaluasi model dengan lebih banyak metrik (MAE, RMSE, dll.)

DAFTAR PUSTAKA

Berikut adalah beberapa referensi lokal yang dapat digunakan untuk tinjauan pustaka mengenai analisis harga rumah di Indonesia dengan menggunakan Streamlit dan algoritma data mining:

1. Prabowo, H. (2020). Analisis Faktor-Faktor yang Mempengaruhi Harga Rumah di Kota Yogyakarta Menggunakan Metode Regresi Linier Berganda. *Jurnal Ilmiah Mahasiswa Ekonomi Akuntansi*, 5(1), 1-10.
2. Sari, D. P., & Rahardjo, B. (2021). Penerapan Algoritma K-Means dalam Pengelompokan Harga Rumah di Jakarta. *Jurnal Teknologi dan Sistem Komputer*, 9(2), 123-130.
3. Wibowo, A. (2019). Pemanfaatan Data Mining untuk Analisis Harga Properti di Indonesia. *Jurnal Informatika*, 15(3), 45-52.
4. Nugroho, A. (2022). Visualisasi Data Harga Rumah Menggunakan Streamlit: Studi Kasus di Surabaya. *Jurnal Sistem Informasi*, 18(1), 67-75.
5. Setiawan, B. (2020). Analisis Prediksi Harga Rumah Menggunakan Metode Machine Learning di Indonesia. *Jurnal Ilmiah Komputer dan Informatika*, 8(2), 89-97.