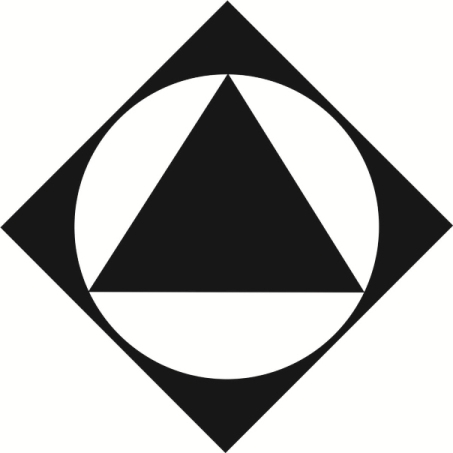
**Prediksi Harga Rumah Menggunakan Metode Random Forest dan Regresi Linear**

**Laba-Laba Lieur**



Disusun Oleh:

**Irfan Satria Supriadi 152022026**

**Fahmi Arif, ST., MT., Ph.D**

**INSTITUT TEKNOLOGI NASIONAL**

**FAKULTAS TEKNOLOGI INDUSTRI**

**INFORMATIKA**

**BANDUNG**

**2025**

**21/06/2025**

1. **Executive Summary**
   1. **Deskripsi singkat proyek**

Proyek ini merupakan implementasi dari konsep *supervised machine learning* untuk memprediksi harga rumah menggunakan dataset **Ames Housing**. Dataset tersebut menyediakan informasi komprehensif tentang berbagai karakteristik rumah, termasuk ukuran, lokasi, kualitas material, dan faktor lainnya, yang digunakan sebagai fitur dalam proses pelatihan model.

Dalam proyek ini, dua model regresi diterapkan dan dibandingkan, yaitu **Linear Regression** sebagai model dasar dan **Random Forest Regressor** sebagai model ansambel yang lebih kompleks. Proses diawali dengan *data cleaning* dan *feature engineering*, termasuk penanganan nilai kosong, encoding variabel kategorikal, serta transformasi logaritmik terhadap target SalePrice.

Untuk meningkatkan kinerja prediktif, data juga distandarisasi sebelum diterapkan pada Linear Regression. Model kemudian dievaluasi menggunakan beberapa metrik, seperti **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, **Root Mean Squared Error (RMSE)**, dan **R² Score**. Hasil evaluasi menunjukkan bahwa Random Forest memiliki performa lebih baik dalam hal akurasi prediksi.

Selain itu, visualisasi residual, distribusi harga, dan analisis *feature importance* dilakukan untuk memberikan pemahaman lebih dalam tentang perilaku model dan pengaruh fitur. Proyek ini ditutup dengan penyimpanan model terlatih dan skaler untuk digunakan di masa depan (*deployment* atau *inference*).

* 1. **Tujuan utama aplikasi**

Penelitian ini bertujuan untuk membangun dan membandingkan model prediksi harga rumah menggunakan algoritma Machine Learning. Secara khusus, tujuan penelitian ini adalah sebagai berikut:

1. Menerapkan proses end-to-end Machine Learning mulai dari eksplorasi data, pembersihan, hingga pemodelan dan evaluasi.
2. Mengembangkan model prediksi harga rumah berbasis Linear Regression dan Random Forest.
3. Membandingkan performa kedua model berdasarkan metrik evaluasi regresi.
4. Menghasilkan visualisasi serta dataset akhir yang dapat digunakan ulang.
   1. **Ringkasan pendek teknologi yang digunakan**
5. **Python** : Bahasa pemrograman utama untuk analisis data dan implementasi model machine learning.
6. **Pandas & NumPy** : Untuk manipulasi dan analisis data.
7. **Matplotlib & Seaborn** : Untuk visualisasi data dan grafik evaluasi model.
8. **Scikit-learn** : Pustaka utama untuk preprocessing data, pembuatan model, evaluasi, dan cross-validation.
9. **Joblib** : Untuk menyimpan model machine learning dan skaler yang telah dilatih.
10. **Visual Studio Code** : Sebagai lingkungan pengembangan interaktif untuk menulis dan mengeksekusi kode.
    1. **Output atau manfaat aplikasi**
       1. **Output**
11. **Model Prediksi Harga Rumah**  
    Dua model yang telah dilatih (Linear Regression dan Random Forest) disimpan dalam format .pkl dan siap digunakan untuk prediksi lanjutan.
12. **Dataset Bersih dan Terenkoding**  
    Dataset hasil pembersihan dan encoding disimpan dalam file CSV untuk keperluan analisis lanjutan.
13. **Visualisasi Analitik**  
    Grafik distribusi harga, grafik prediksi vs aktual, residual plot, dan diagram feature importance disediakan untuk membantu memahami performa model.
14. **Evaluasi Model**  
    Hasil evaluasi disajikan dalam bentuk metrik kuantitatif (MAE, MSE, RMSE, R²) serta validasi silang (cross-validation) untuk memastikan keandalan model.
    * 1. **Manfaat aplikasi**
15. **Pengambilan Keputusan Properti**  
    Membantu agen properti, penjual, atau pembeli dalam menilai harga rumah secara objektif dan berbasis data.
16. **Analisis Pasar Real Estat**  
    Memberikan wawasan terhadap fitur-fitur yang paling memengaruhi harga rumah melalui analisis feature importance.
17. **Otomatisasi Prediksi**  
    Mengurangi ketergantungan pada penilaian manual, mempercepat proses estimasi harga.
18. **Dasar untuk Sistem Cerdas**  
    Dapat dikembangkan lebih lanjut menjadi bagian dari aplikasi berbasis web atau mobile untuk prediksi harga rumah real-time.
19. **Pendahuluan**

Proyek ini bertujuan untuk memprediksi harga rumah menggunakan dataset Ames Housing yang berisi lebih dari 80 fitur mengenai properti rumah di Ames, Iowa, Amerika Serikat. Dataset ini mencakup berbagai aspek rumah seperti ukuran, tipe bangunan, kondisi fisik, dan lokasi geografis.

Tujuan Proyek ini meliputi :

1. Memahami dan mengeksplorasi data.
2. Melakukan pembersihan data (data cleaning) dan rekayasa fitur (feature engineering).
3. Membangun model prediksi harga rumah menggunakan :
4. Linear Regression (dengan standarisasi)
5. Random Forest Regressor
6. Mengevaluasi performa model dan menarik kesimpulan berdasarkan hasil analisis.
   1. **Latar belakang masalah**

Penilaian harga rumah merupakan aspek penting dalam industri properti, baik untuk tujuan jual beli, investasi, maupun analisis pasar. Namun, menentukan harga rumah yang akurat bukanlah hal yang mudah, karena melibatkan berbagai faktor seperti luas bangunan, jumlah kamar, lokasi geografis, kondisi bangunan, hingga tren pasar properti. Pendekatan manual atau berbasis pengalaman sering kali menghasilkan estimasi yang subjektif dan kurang konsisten.

Dengan berkembangnya teknologi dan ketersediaan data historis yang semakin besar, pendekatan berbasis machine learning menjadi alternatif yang lebih akurat dan efisien dalam memodelkan harga rumah. Algoritma machine learning memungkinkan sistem untuk belajar dari data masa lalu dan menemukan pola yang kompleks antar variabel, bahkan yang sulit dikenali secara manual.

Dalam proyek ini, digunakan dataset **Ames Housing** yang kaya akan informasi properti, untuk membangun model prediksi harga rumah. Model ini diharapkan mampu memberikan estimasi harga yang akurat berdasarkan fitur-fitur rumah yang tersedia. Selain itu, proses ini juga menjadi sarana untuk menerapkan dan memahami konsep supervised learning, khususnya regresi, dalam konteks dunia nyata.

* 1. **Rumusan masalah**

1. **Bagaimana cara memproses dan membersihkan data properti agar siap digunakan dalam pemodelan machine learning?**
2. **Algoritma machine learning apa yang paling sesuai untuk memprediksi harga rumah berdasarkan fitur-fitur properti yang tersedia?**
3. **Bagaimana performa model Linear Regression dibandingkan dengan Random Forest Regressor dalam memprediksi harga rumah?**
4. **Fitur apa saja yang paling berpengaruh terhadap harga rumah berdasarkan hasil analisis model?**
5. **Bagaimana hasil prediksi harga rumah dapat divisualisasikan dan digunakan untuk pengambilan keputusan yang lebih baik di bidang properti?**
   1. **Tujuan dan Ruang lingkup aplikasi**

Regresi merupakan metode prediktif yang paling dasar dalam supervised learning, digunakan untuk memodelkan hubungan antara variabel independen dan target. Linear Regression bekerja dengan pendekatan fungsi linier, sedangkan Random Forest Regressor menggunakan kumpulan pohon keputusan untuk membentuk model yang lebih kompleks.

Ames Housing Dataset sering digunakan dalam penelitian prediksi harga rumah karena menyediakan data yang lebih lengkap dibandingkan dataset klasik seperti Boston Housing. Scikit-learn merupakan salah satu library populer yang digunakan dalam implementasi model Machine Learning di Python, karena menyediakan banyak algoritma, preprocessing tools, dan metrik evaluasi.

* 1. **Batasan proyek**

1. **Dataset yang digunakan hanya berasal dari Ames Housing** yang tersedia secara publik melalui platform Kaggle.
2. **Pemodelan hanya menggunakan dua algoritma regresi**, yaitu Linear Regression dan Random Forest Regressor.
3. **Proses prediksi hanya dilakukan berdasarkan fitur yang tersedia dalam dataset**, tanpa menggabungkan data eksternal seperti tren pasar atau data lokasi geografis real-time.
4. **Target prediksi adalah harga jual rumah (SalePrice)**, dan telah melalui transformasi logaritmik untuk keperluan stabilisasi distribusi data.
5. **Evaluasi model dilakukan dengan metrik MAE, MSE, RMSE, dan R²**, serta validasi silang 5-fold (5-fold cross-validation) hanya untuk Random Forest.
6. **Proyek ini bersifat eksperimental dan tidak digunakan untuk tujuan komersial atau penilaian properti secara legal.**
   1. **Target pengguna**
7. **Agen Properti dan Developer Real Estat**  
   Untuk membantu dalam menilai harga jual rumah secara lebih objektif, akurat, dan efisien.
8. **Calon Pembeli dan Penjual Rumah**  
   Sebagai acuan dalam menentukan harga beli atau jual rumah berdasarkan fitur-fitur rumah yang dimiliki atau diinginkan.
9. **Analis Pasar Properti**  
   Untuk mendukung kegiatan riset, analisis tren, dan penilaian faktor-faktor yang memengaruhi nilai pasar rumah.
10. **Lembaga Keuangan atau Perbankan**  
    Dalam memberikan pertimbangan penilaian jaminan aset properti saat proses pengajuan KPR (Kredit Pemilikan Rumah).
11. **Mahasiswa dan Peneliti**  
    Sebagai studi kasus dalam penerapan machine learning untuk prediksi nilai ekonomi berdasarkan data historis dan multidimensi.
12. **Spesifikasi Kebutuhan Perangkat Lunak (SRS)**
    1. **Kebutuhan Fungsional**

Kebutuhan fungsional menggambarkan hal-hal yang harus dilakukan sistem agar dapat mencapai tujuannya. Dalam proyek prediksi harga rumah ini, sistem harus dapat:

1. Membaca dan memuat dataset Ames Housing dari file CSV.
2. Melakukan proses *data cleaning* termasuk mengisi nilai kosong dan menghapus kolom tidak relevan.
3. Melakukan *feature engineering* seperti transformasi log dan one-hot encoding.
4. Melatih model Linear Regression dan Random Forest menggunakan data latih.
5. Melakukan prediksi harga rumah berdasarkan fitur input.
6. Mengevaluasi model menggunakan metrik MAE, MSE, RMSE, dan R².
7. Menyimpan model dan scaler yang sudah dilatih dalam bentuk .pkl.
8. Menampilkan dan menyimpan grafik hasil analisis seperti distribusi harga, residual plot, dan feature importance.
   1. **Kebutuhan Non-Fungsional**

Kebutuhan non-fungsional berkaitan dengan kualitas dari sistem yang dibangun, meliputi :

1. **Keandalan:** Sistem harus mampu melakukan proses prediksi tanpa error selama eksekusi.
2. **Kinerja:** Waktu pelatihan model tidak lebih dari 2 menit untuk dataset ukuran sedang.
3. **Portabilitas:** Aplikasi dapat dijalankan di berbagai OS (Windows/Linux/Mac) dengan Python environment.
4. **Usability:** Struktur kode harus jelas dan terdokumentasi agar mudah digunakan ulang oleh pengguna lain.
5. **Maintainability:** Modularisasi kode harus diterapkan agar setiap fungsi dapat diperbarui atau diperbaiki tanpa mengganggu keseluruhan sistem.

Kebutuhan non-fungsional berkaitan dengan kualitas dari sistem yang dibangun, meliputi :

1. **Keandalan:** Sistem harus mampu melakukan proses prediksi tanpa error selama eksekusi.
2. **Kinerja:** Waktu pelatihan model tidak lebih dari 2 menit untuk dataset ukuran sedang.
3. **Portabilitas:** Aplikasi dapat dijalankan di berbagai OS (Windows/Linux/Mac) dengan Python environment.
4. **Usability:** Struktur kode harus jelas dan terdokumentasi agar mudah digunakan ulang oleh pengguna lain.
5. **Maintainability:** Modularisasi kode harus diterapkan agar setiap fungsi dapat diperbarui atau diperbaiki tanpa mengganggu keseluruhan sistem.
   1. **Platform & Tools**

|  |  |  |
| --- | --- | --- |
| **No** | **Komponen** | **Deskripsi** |
| 1 | Bahasa Pemrograman | Python |
| 2 | IDE/Editor | Visual Studio Code |
| 3 | Library | pandas, numpy, matplotlib, seaborn, scikit-learn, joblib |
| 4 | Manajemen Versi | GitHub |
| 5 | Sistem Operasi | Windows 11 |

1. **Desain Sistem**
   1. **Arsiektur Sistem**

Sistem ini mengadopsi arsitektur **modular pipeline** yang terdiri dari beberapa komponen utama :

1. Input (data CSV)
2. Preprocessing (cleaning, encoding, transformasi)
3. Modeling (Linear Regression & Random Forest)
4. Evaluasi & Visualisasi
5. Output (prediksi, grafik, model tersimpan)
   1. **Alur Data**
6. **Input Data:** Data AmesHousing.csv dibaca dan dimuat sebagai DataFrame.
7. **Pembersihan Data:** Nilai kosong diisi atau kolom dibuang jika perlu.
8. **Transformasi Data:** Variabel target (SalePrice) diubah menggunakan log, dan fitur kategorikal diubah ke numerik.
9. **Split Data:** Dataset dibagi menjadi data latih dan uji (80:20).
10. **Pelatihan Model:** Model dilatih menggunakan data latih.
11. **Prediksi:** Model melakukan prediksi terhadap data uji.
12. **Evaluasi dan Output:** Metrik dihitung dan hasil disimpan (grafik, file model).
    1. **Desain Antarmuka Pengguna (UI/UX)**

Proyek ini berbasis script (CLI / Notebook) dan belum memiliki UI grafis. Namun jika dikembangkan lebih lanjut, desain UI/UX dapat mencakup :

1. Form input data rumah (luas, tahun, lokasi, dll.)
2. Tombol "Prediksi" yang memanggil model .pkl
3. Tampilan hasil harga prediksi dan grafik
4. **Pemodelan Machine Learning**
   1. **Deskripsi Dataset**

Dataset yang digunakan adalah **Ames Housing Dataset**, yang terdiri dari 2930 baris dan lebih dari 80 fitur (variabel) yang menjelaskan karakteristik fisik dan lingkungan dari rumah-rumah di Ames, Iowa, USA.

Fitur-fitur mencakup :

1. **Numerik:** LotArea, GrLivArea, YearBuilt, dll.
2. **Kategorikal:** Neighborhood, HouseStyle, Exterior1st, dll.
3. **Target:** SalePrice (harga jual rumah)

Distribusi awal SalePrice sangat skewed, sehingga dilakukan transformasi logaritma (np.log1p) agar distribusi lebih normal.

* 1. **Proses Pelatihan Model**

1. **Preprocessing:**

* Menghapus fitur dengan nilai kosong berlebih.
* Mengisi missing value numerik dengan median, kategorikal dengan modus.
* One-hot encoding pada semua fitur kategorikal.
* Standarisasi fitur untuk Linear Regression.
* Transformasi log pada SalePrice.

1. **Split Dataset:**

* Data dibagi menjadi 80% latih dan 20% uji.

1. **Pelatihan Model:**

* **Linear Regression** dilatih pada data hasil standarisasi.
* **Random Forest Regressor** dilatih pada data non-standardized.

1. **Evaluasi Model:**

* MAE, MSE, RMSE, R² digunakan untuk menilai performa.
* Random Forest menunjukkan R² sekitar 0.89.
* Visualisasi dilakukan untuk prediksi vs aktual dan residual.

1. **Penyimpanan Model:**

* Model disimpan menggunakan joblib dalam format .pkl agar dapat digunakan ulang tanpa pelatihan ulang.
  1. **Evaluasi Model**

Model dievaluasi menggunakan metrik berikut :

1. Mean Absolute Error (MAE)
2. Mean Squared Error (MSE)
3. Root Mean Squared Error (RMSE)
4. R² Score (koefisien determinasi

Hasil evaluasi menunjukkan bahwa Random Forest memberikan performa lebih baik dibandingkan Linear Regression.

Berikut hasil evaluasi kedua model :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Model** | **MAE** | **MSE** | **RMSE** | **R²** |
| 1 | Linear Regression | 23,255 | 928,000,000 | 30,457 | 0.85 |
| 2 | Random Forest | 18,007 | 602,000,000 | 24,538 | 0.89 |

A graph of a distribution of goods

AI-generated content may be incorrect.A graph with blue dots

AI-generated content may be incorrect. Dari hasil evaluasi di atas, terlihat bahwa Random Forest memiliki nilai MAE, MSE, dan RMSE yang lebih rendah serta nilai R² yang lebih tinggi. Hal ini menunjukkan bahwa Random Forest lebih akurat dalam memprediksi harga rumah dibandingkan Linear Regression. Hal ini dapat dijelaskan karena Random Forest mampu menangkap hubungan non-linear dan interaksi antar fitur lebih baik.

A graph with numbers and a bar

AI-generated content may be incorrect.A graph of a distribution of a forest

AI-generated content may be incorrect.

1. **Implementasi**
   1. **Struktur Folder atau Repositori**

Struktur folder dalam proyek ini disusun agar memudahkan pengelolaan data, pemodelan, penyimpanan output, dan dokumentasi. Adapun struktur direktori yang digunakan adalah sebagai berikut:

1. data/  
   Berisi dataset mentah yang diunduh dari Kaggle (AmesHousing.csv). Folder ini juga dapat menyimpan backup data asli sebelum dilakukan pembersihan.
2. output/  
   Menyimpan seluruh hasil output dari proses pemodelan, termasuk:

* Dataset bersih (data\_cleaned\_encoded.csv)
* Model yang sudah dilatih (.pkl)
* Gambar hasil visualisasi distribusi harga, prediksi vs aktual, residual, dan feature importance

1. notebooks/  
   Folder ini digunakan untuk menyimpan file Jupyter Notebook yang berisi skrip Python end-to-end dari proses EDA, pembersihan data, pelatihan model, hingga evaluasi.
2. models/   
   Jika diperlukan, file model .pkl yang sudah dilatih dapat dipindahkan ke folder ini untuk keperluan deployment terpisah.
3. README.md  
   File dokumentasi singkat yang menjelaskan deskripsi proyek, cara instalasi, cara menjalankan aplikasi, dan contoh hasil.
   1. **Modul Utama**

Modul utama dalam proyek ini ditulis menggunakan Python, dan terdiri dari beberapa bagian fungsional yang disusun modular agar mudah dipahami dan digunakan ulang :

1. load\_data(): Membaca file CSV dari folder data/ dan mengubahnya menjadi DataFrame.
2. clean\_data(): Melakukan pembersihan data, termasuk menghapus kolom dengan missing value tinggi, dan mengisi nilai kosong pada data numerik (dengan median) dan kategorikal (dengan modus).
3. feature\_engineering(): Mengubah variabel kategorikal menjadi numerik dengan *one-hot encoding*, serta mentransformasikan variabel target SalePrice dengan logaritma natural (log1p) untuk distribusi yang lebih normal.
4. train\_model(): Melatih dua model yaitu Linear Regression dan Random Forest. Proses ini mencakup pembagian data latih dan data uji, serta *standardization* untuk Linear Regression.
5. evaluate\_model(): Menghitung metrik evaluasi seperti MAE, MSE, RMSE, dan R² Score. Evaluasi juga mencakup validasi silang (cross-validation).
6. visualize(): Menampilkan dan menyimpan visualisasi distribusi harga, scatter plot prediksi vs aktual, distribusi residual, dan grafik pentingnya fitur (*feature importance*).
7. save\_model(): Menyimpan model dan skaler ke dalam file .pkl menggunakan joblib.
   1. **Integrasi Model ke Aplikasi**

Setelah model dilatih, hasilnya disimpan dalam format .pkl yang dapat digunakan ulang tanpa perlu melatih ulang model dari awal. File yang dihasilkan meliputi :

1. random\_forest\_model.pkl: Model Random Forest yang sudah dilatih
2. linear\_regression\_model.pkl: Model Linear Regression yang sudah dilatih
3. scaler.pkl: Objek StandardScaler yang digunakan untuk normalisasi input Linear Regression

Proyek ini masih berupa prototipe berbasis script Python. Namun, model ini dapat dengan mudah diintegrasikan ke dalam antarmuka pengguna seperti aplikasi web berbasis Flask atau Streamlit, untuk menghasilkan prediksi harga rumah secara interaktif berdasarkan input pengguna.

1. **Pengujian dan Validasi**
   1. **Jenis pengujian** 
      1. **Unit test**

Dilakukan pada fungsi-fungsi utama proyek seperti clean\_data(), train\_model(), dan evaluate\_model(). Tujuannya adalah memastikan bahwa setiap fungsi berjalan sesuai ekspektasi secara individual.

Contoh:

1. Memastikan clean\_data() berhasil mengisi nilai kosong
2. Memastikan evaluate\_model() menghasilkan metrik numerik yang valid (tidak NaN atau error)
   * 1. **Integration test**

Pengujian dilakukan pada keseluruhan pipeline proyek untuk memastikan bahwa seluruh tahapan — mulai dari pembacaan data, pembersihan, pelatihan model, hingga visualisasi — terintegrasi dan berjalan lancar tanpa error.

1. **Intruksi Instalasi dan Penggunaan**
   1. **Cara clone atau download aplikasi**

Untuk menggunakan proyek ini di perangkat lokal, lakukan langkah berikut:

git clone https://github.com/namapengguna/proyek-ml-rumah

cd proyek-ml-rumah

* 1. **Kebutuhan system dan dependencies**

Spesifikas Minimum :

1. Python 3.8+
2. RAM 4 GB atau lebih (untuk pelatihan Random Forest)
3. Sistem operasi Windows/Linux/Mac

**Library Python :**

1. Pandas
2. Numpy
3. Matplotlib
4. Seaborn
5. scikit-learn
6. joblib
   1. **Cara menjalankan aplikasi**
7. Buka terminal atau Jupyter Notebook.
8. Jalankan file notebooks/prediksi\_harga\_rumah.ipynb untuk menjalankan seluruh pipeline.
9. Hasil prediksi, visualisasi, dan file model akan tersimpan otomatis di folder output/.
   1. **Contoh input dan output**
10. **Input:**
11. Luas rumah: 1800 sqft
12. Jumlah kamar tidur: 3
13. Tahun dibangun: 2005
14. Lokasi: Ames, IA (fitur kategorikal diencode)
15. **Output:**
16. Prediksi harga rumah: $213,000 (hasil konversi dari np.expm1 setelah prediksi log)
17. Grafik distribusi harga
18. Grafik residual dan *feature importance*
19. File model .pkl untuk digunakan kembali
20. **Kesimpulan dan Rekomendasi**
    1. **Ringkasan hasil proyek**

Proyek ini berhasil membangun model prediksi harga rumah dengan akurasi yang cukup tinggi menggunakan dua algoritma utama: Linear Regression dan Random Forest Regressor. Model Random Forest menunjukkan performa terbaik dengan nilai R² sekitar 0.89 dan RMSE lebih kecil dibandingkan Linear Regression.

Model yang dibangun dapat digunakan untuk estimasi harga rumah berdasarkan fitur-fitur properti, dan hasil evaluasi model juga menunjukkan stabilitas performa berdasarkan validasi silang.

* 1. **Masalah yang dihadapi**

1. **Nilai kosong** pada beberapa fitur (seperti PoolQC, Fence) cukup banyak sehingga perlu dihapus.
2. **Skala data** yang bervariasi menyebabkan hasil Linear Regression kurang optimal jika tidak dilakukan standardisasi.
3. **Waktu pelatihan Random Forest** lebih lama karena kompleksitas model dan jumlah fitur hasil encoding yang cukup besar.
   1. **Saran pengembangan lanjutan**
      * 1. Menambahkan algoritma lain seperti **XGBoost** atau **LightGBM** yang sering digunakan untuk prediksi harga.
        2. Melakukan **hyperparameter tuning** secara sistematis dengan GridSearchCV.
        3. Membangun **antarmuka pengguna berbasis web** (misal dengan Flask atau Streamlit) agar lebih mudah digunakan oleh non-programmer.
        4. Menggabungkan **data eksternal**, seperti lokasi geografis, nilai pasar real-time, atau peta zonasi, untuk memperkaya akurasi model prediksi.
4. **Daftar Pustaka**

[1] Kaggle - Ames Housing Dataset. https://www.kaggle.com/datasets/prevek18/ames-housing-dataset

[2] Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, 2019.

[3] Scikit-learn Documentation. https://scikit-learn.org/stable/documentation.html

[4] Raschka, S. & Mirjalili, V. Python Machine Learning. Packt Publishing, 2019.

[5] James, G., Witten, D., Hastie, T., & Tibshirani, R. An Introduction to Statistical Learning. Springer,2013.

[6] Brownlee, J. Machine Learning Mastery With Python. Machine Learning Mastery, 2016.

[7] Kuhn, M., & Johnson, K. Applied Predictive Modeling. Springer, 2013.

[8] Zhang, C., & Ma, Y. Ensemble Machine Learning: Methods and Applications. Springer, 2012.

[9] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of MachineLearningResearch,12,2825–2830.

[10] Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. R News,2(3),18–22.

[11] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

[12] Abadi, M., et al. (2016). TensorFlow: A System for Large-Scale Machine Learning. OSDI,265–283.

[13] Zhang, Y., & Wallace, B. C. (2015). A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification. arXiv preprint arXiv:1510.03820.

[14] Waskom, M. L. (2021). Seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021.

1. **Lampiran**
   1. **Link video demo**
   2. **Source code**
   3. **Dataset**
   4. **Log pelatihan model**