

## Implementasi Algoritma Linear Regression Untuk Prediksi Harga Rumah di Daerah Tebet

Aditya pratama<sup>1</sup>, Adha Maulana<sup>2</sup>, Rendy Amy Saputra<sup>3</sup>

<sup>1,2</sup> Program studi sistem informasi, Fakultas Teknik, Universitas Nahdlatul Ulama Kalimantan Barat

<sup>3</sup>Program Studi Sistem Informasi STMIK Pontianak

<sup>1</sup>adityapratamabadra@unukalbar.ac.id, <sup>2</sup>adhamaulana@unukalbar.ac.id, <sup>3</sup>rendyamy@stmikpontianak.ac.id



All publications by Journal Of Information Technology is licensed under a [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/). (CC BY 4.0)

**Abstract**— This research focuses on the issue of house price variation in the Tebet area, South Jakarta, influenced by factors such as land area, building area, number of bedrooms, number of bathrooms, and additional facilities. This price variability makes it difficult for potential buyers and sellers to determine a fair price, potentially resulting in poor decisions and financial losses. Therefore, the main objective of this research is to develop a house price prediction model using the linear regression algorithm, which is expected to provide a more accurate price estimate based on relevant features. As a solution, this research applies the linear regression algorithm to analyze and predict house prices, providing useful information for stakeholders, including buyers, sellers, and real estate agents. The methods used include collecting house price data and related features from the Rumah123 property platform, which are then pre-processed by dividing them into training data (80%) and test data (20%). The model is evaluated using the Mean Squared Error (MSE) and R-squared metrics. The research results show that this model has an R-squared of 0.7713, which means it can explain about 77% of the variation in house prices. This model also predicts the price of a new house with certain features, such as a building area of 100 m<sup>2</sup> and a land area of 300 m<sup>2</sup>, at around IDR 8,555,000,000. This research is expected to make a significant contribution to understanding the dynamics of house prices in Tebet.

**Intisari**— Penelitian ini berfokus pada permasalahan variasi harga rumah di daerah Tebet, Jakarta Selatan, yang dipengaruhi oleh faktor-faktor seperti luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan fasilitas tambahan. Variabilitas harga ini menyulitkan calon pembeli dan penjual dalam menentukan harga yang wajar, berpotensi mengakibatkan keputusan yang kurang tepat dan kerugian finansial. Oleh karena itu, tujuan utama penelitian ini adalah mengembangkan model prediksi harga rumah menggunakan algoritma Linear Regression, yang diharapkan dapat memberikan estimasi harga yang lebih akurat berdasarkan fitur-fitur yang relevan. Sebagai solusi, penelitian ini menerapkan algoritma Linear Regression untuk menganalisis dan memprediksi harga rumah, memberikan informasi yang berguna bagi pemangku kepentingan, termasuk pembeli, penjual, dan agen properti. Metode yang digunakan mencakup pengumpulan data harga rumah dan fitur terkait dari platform properti Rumah123, yang kemudian dipra-proses

dengan membagi menjadi data latih (80%) dan data uji (20%). Model dievaluasi menggunakan metrik Mean Squared Error (MSE) dan R-squared. Hasil penelitian menunjukkan bahwa model ini memiliki R-squared sebesar 0.7713, yang berarti dapat menjelaskan sekitar 77% variasi harga rumah. Model ini juga memprediksi harga rumah baru dengan fitur tertentu, seperti luas bangunan 100 m<sup>2</sup> dan luas tanah 300 m<sup>2</sup>, sekitar IDR 8,555,000,000. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam memahami dinamika harga rumah di Tebet.

**Kata Kunci**— Prediksi Harga Rumah, Linear Regression, Properti, Tebet, Machine Learning.

### I. PENDAHULUAN

Harga rumah merupakan faktor krusial yang dipertimbangkan dalam transaksi jual-beli properti. Fluktuasi harga rumah dipengaruhi oleh berbagai faktor, termasuk lokasi, luas tanah dan bangunan, jumlah kamar, fasilitas, kondisi pasar, aksesibilitas, serta faktor lingkungan sekitar. Bagi para pemangku kepentingan di pasar properti, seperti pembeli, penjual, developer, dan investor, kemampuan untuk memprediksi harga rumah secara akurat sangat penting dalam pengambilan keputusan yang strategis dan menguntungkan [1]. Prediksi yang tepat dapat membantu pembeli dalam menentukan harga yang wajar, penjual dalam menetapkan harga jual yang optimal, developer dalam merencanakan proyek properti, dan investor dalam mengidentifikasi peluang investasi yang potensial [2].

Daerah Tebet di Jakarta Selatan merupakan kawasan hunian yang populer dengan beragam tipe properti, mulai dari rumah tapak, apartemen, hingga ruko. Permintaan yang tinggi dan keterbatasan lahan di daerah Tebet membuat harga rumah cenderung tinggi dan fluktuatif. Faktor-faktor seperti aksesibilitas yang baik, keberadaan fasilitas umum yang lengkap, dan lingkungan yang nyaman berkontribusi terhadap tingginya minat masyarakat untuk tinggal di Tebet. Oleh karena itu, prediksi harga rumah di Tebet menjadi hal yang penting bagi para pemangku kepentingan di wilayah tersebut. Prediksi

yang akurat dapat membantu mereka dalam membuat keputusan yang tepat dan mengurangi risiko kerugian.

Metode machine learning menawarkan solusi inovatif dalam memprediksi harga rumah dengan memanfaatkan algoritma untuk mempelajari pola dari data historis[3]. Algoritma machine learning dapat mengidentifikasi hubungan yang kompleks antara fitur-fitur properti dan harga jual, yang memungkinkan prediksi yang lebih akurat dibandingkan dengan metode tradisional [4]. Salah satu algoritma machine learning yang populer dan efektif untuk prediksi harga rumah adalah Linear Regression. Algoritma ini digunakan untuk memodelkan hubungan linear antara variabel independen (fitur-fitur properti) dengan variabel dependen (harga rumah) [5]. Linear Regression bekerja dengan mencari garis regresi yang terbaik, yang dapat meminimalkan kesalahan prediksi dan memberikan estimasi harga rumah yang lebih akurat.

Penelitian ini bertujuan untuk mengimplementasikan algoritma Linear Regression dalam memprediksi harga rumah di daerah Tebet, Jakarta Selatan. Data historis harga rumah di Tebet akan dikumpulkan dan diproses untuk digunakan dalam pelatihan model Linear Regression. Kinerja model akan dievaluasi menggunakan metrik Mean Squared Error (MSE) dan R-squared untuk mengukur akurasi dan kemampuan generalisasi model. Hasil prediksi yang diperoleh akan digunakan untuk memahami hubungan antara faktor-faktor yang memengaruhi harga rumah di Tebet dan memberikan wawasan mengenai fitur-fitur yang paling berpengaruh.

Hasil dari penelitian diharapkan dapat memberikan informasi yang berguna bagi para pemangku kepentingan di pasar properti Tebet dalam pengambilan keputusan terkait jual-beli dan investasi properti. Prediksi harga rumah yang akurat dapat membantu pembeli dan penjual dalam negosiasi harga, developer dalam menentukan harga jual unit properti, dan investor dalam mengidentifikasi peluang investasi yang menguntungkan. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan model prediksi harga properti yang lebih akurat, mudah diinterpretasi, dan dapat diandalkan. Model prediksi yang handal dapat membantu meningkatkan transparansi dan efisiensi di pasar properti, yang pada gilirannya dapat mendorong pertumbuhan ekonomi dan kesejahteraan masyarakat.

## II. TINJAUAN PUSTAKA

Penelitian tentang prediksi harga rumah menggunakan algoritma machine learning semakin berkembang seiring dengan tingginya permintaan akan prediksi harga properti yang lebih akurat. Dalam beberapa tahun terakhir, berbagai algoritma regresi dan model machine learning lainnya telah diterapkan untuk memprediksi harga rumah dengan mempertimbangkan faktor-faktor seperti lokasi, ukuran, tipe rumah, dan fasilitas yang ada[6]. Berikut ini adalah beberapa penelitian terdahulu yang relevan dengan topik ini:

Sebuah penelitian oleh [7] berjudul "House Price Prediction Based on Machine Learning" mengeksplorasi penerapan model

LightGBM (Light Gradient Boosting Machine), Gradient Boosting, dan XGBoost (Extreme Gradient Boosting) untuk memprediksi harga rumah. Menggunakan data dari Kaggle, penelitian ini membandingkan hasil ketiga model tersebut dan menghasilkan RMSE masing-masing sebesar 0.02975, 0.02537, dan 0.01364. Berdasarkan hasil tersebut, XGBoost terbukti memiliki performa terbaik dengan RMSE terendah, menjadikannya model yang paling akurat untuk prediksi harga rumah. Penelitian ini menekankan pentingnya pemilihan model yang tepat untuk meningkatkan akurasi dalam prediksi harga rumah di pasar properti.

Artikel [8] berjudul "House Price Prediction Using Machine Learning Algorithm" menyoroti penggunaan berbagai algoritma machine learning, termasuk Linear Regression (LR), Gradient Boosting Regressor (GBR), Histogram Gradient Boosting Regressor, dan Random Forest (RF) Regressor dalam memprediksi harga rumah. Hasil penelitian ini menunjukkan bahwa penggunaan berbagai algoritma ini dapat menghasilkan prediksi yang lebih akurat, yang membantu calon pembeli memilih rumah sesuai kebutuhan dan anggaran mereka. Akurasi model yang tinggi menjadi faktor kunci dalam pengambilan keputusan pembelian rumah, yang dapat mempengaruhi kestabilan finansial jangka panjang.

Dalam penelitian [9] berjudul "Going the Distance: Application of Machine Learning Algorithms in Predicting Housing Price Dynamics", berbagai algoritma seperti Gradient Boost Regression, Catboost Regression, Decision Tree Regression, dan K-neighbors Regression digunakan untuk memprediksi harga rumah. Penelitian ini mengatasi tantangan prediksi harga rumah, seperti kualitas data yang bervariasi dan faktor-faktor seperti kebijakan pemerintah, lokasi spasial, dan ukuran rumah. Catboost Regression menghasilkan hasil yang sangat baik dengan mean absolute percentage error (MAPE) sebesar 0.018 dan R-squared sebesar 0.904, menunjukkan korelasi positif yang kuat antara area dan harga rumah. Penelitian ini menekankan pentingnya pengolahan data dan pemilihan algoritma yang tepat untuk mengatasi kompleksitas prediksi harga rumah [10].

Selain itu, penelitian oleh [11] berjudul "Implementasi Decision Tree untuk Prediksi Harga Rumah di Daerah Tebet" mengimplementasikan algoritma Decision Tree untuk memprediksi harga rumah di Tebet, Jakarta Selatan. Decision Tree dipilih karena kemudahan interpretasi dan kemampuannya dalam menangani data numerik dan kategorikal. Data yang digunakan meliputi fitur seperti luas tanah, luas bangunan, jumlah kamar, dan lokasi. Data dipraproses dan dibagi menjadi data latih (80%) dan data uji (20%). Model Decision Tree dievaluasi menggunakan Mean Squared Error (MSE) dan R-squared, dan hasil penelitian menunjukkan bahwa Decision Tree mampu memprediksi harga rumah di Tebet dengan akurasi yang baik. Visualisasi pohon keputusan memberikan informasi tentang fitur-fitur penting dalam penentuan harga. Penelitian ini diharapkan bermanfaat bagi stakeholder di pasar properti Tebet dalam pengambilan keputusan.

Meskipun sudah banyak penelitian yang mengkaji penerapan algoritma machine learning dalam memprediksi harga rumah, masih ada gap yang perlu diatasi, terutama dalam pemodelan prediksi harga rumah di daerah perkotaan spesifik seperti Tebet. Sebagian besar penelitian sebelumnya, seperti yang dilakukan oleh [8] dan [9], lebih berfokus pada model umum yang diterapkan pada dataset besar yang tidak memperhitungkan faktor-faktor lokal yang dapat mempengaruhi harga rumah, seperti kebijakan daerah, tren pasar lokal, atau karakteristik sosial ekonomi daerah tersebut. Selain itu, penelitian oleh [11] menggunakan Decision Tree, yang meskipun efektif, masih memiliki keterbatasan dalam menangani hubungan non-linear yang kompleks antara fitur dan harga rumah.

Oleh karena itu, gap penelitian ini menunjukkan perlunya penerapan algoritma machine learning yang lebih adaptif, seperti Linear Regression, untuk memprediksi harga rumah dengan mempertimbangkan karakteristik lokal daerah seperti tingkat kepadatan penduduk, aksesibilitas transportasi, dan perkembangan infrastruktur di Tebet. Linear Regression dipilih karena kemampuannya untuk memberikan interpretasi yang jelas mengenai pengaruh masing-masing fitur terhadap harga rumah, sehingga memudahkan pemangku kepentingan dalam memahami faktor-faktor yang paling berpengaruh. Selain itu, penelitian yang berfokus pada penerapan Linear Regression untuk memprediksi harga rumah di daerah Tebet masih jarang ditemukan, padahal Linear Regression dapat memberikan hasil yang cukup akurat dan mudah dipahami, terutama dalam menganalisis faktor-faktor kompleks yang mempengaruhi harga rumah. Oleh karena itu, penelitian ini diharapkan dapat menyediakan model prediksi harga rumah yang lebih spesifik dan sesuai dengan kondisi pasar properti di Tebet

#### A. Prediksi

Prediksi adalah proses memperkirakan atau meramalkan suatu kejadian atau hasil di masa depan berdasarkan data yang tersedia. Dalam konteks analisis data dan pembelajaran mesin, prediksi sering dilakukan menggunakan model statistik atau algoritma yang menganalisis tren dan pola dalam data historis [12]. Tujuan dari prediksi adalah untuk memberikan informasi yang membantu dalam pengambilan keputusan di berbagai bidang, seperti bisnis, kesehatan, dan cuaca [13]. Sebagai contoh, dalam dunia bisnis, prediksi penjualan dapat digunakan untuk memproyeksikan pendapatan masa depan, sedangkan dalam bidang kesehatan, prediksi penyakit dapat membantu dalam diagnosa medis.

#### B. Linear Regression

Linear Regression adalah metode dalam machine learning yang digunakan untuk analisis regresi [14]. Model ini memodelkan hubungan antara variabel dependen (seperti harga rumah) dan satu atau lebih variabel independen (seperti luas tanah, jumlah kamar, dan lain-lain) dengan mencari garis lurus yang paling mendekati data. Dalam regresi linier, hubungan antara variabel-variabel tersebut diasumsikan linier, sehingga

model ini cocok untuk prediksi yang sederhana dan mudah diinterpretasikan. Keuntungan utama dari Linear Regression adalah kemampuannya untuk memberikan wawasan yang jelas mengenai pengaruh masing-masing variabel terhadap prediksi hasil. Model ini juga terkenal karena kemudahan pemahaman dan implementasinya, serta kemampuannya dalam menangani data numerik dengan baik. Selain itu, Linear Regression dapat digunakan untuk memprediksi harga rumah berdasarkan faktor-faktor seperti ukuran rumah, jumlah kamar, dan lainnya. Oleh karena itu, Linear Regression menjadi metode yang populer dalam analisis prediktif untuk harga properti.

### III. METODOLOGI PENELITIAN

Dalam penelitian ini, metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining) digunakan sebagai pendekatan terstruktur untuk proses analisis dan prediksi harga rumah di daerah Tebet menggunakan algoritma Linear Regression. CRISP-DM adalah kerangka kerja yang diterima secara luas yang menyediakan model proses komprehensif untuk proyek penambangan data dan terdiri dari enam fase utama: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment [15].



Gambar 1: Metode Crisp-DM

#### 1. Business Understanding

Tahap pertama dari CRISP-DM melibatkan pemahaman tujuan dan persyaratan proyek dari perspektif bisnis. Dalam penelitian ini, tujuan utamanya adalah untuk memprediksi harga rumah di daerah Tebet, Jakarta Selatan, guna memberikan informasi yang bermanfaat bagi pembeli, penjual, dan pengembang properti. Tahap ini mencakup identifikasi pemangku kepentingan utama, seperti agen real estate dan pemerintah daerah, serta pemahaman kebutuhan mereka akan prediksi harga rumah yang akurat untuk mendukung pengambilan keputusan terkait investasi dan kebijakan perumahan.

#### 2. Data Understanding

Setelah fase pemahaman bisnis, langkah berikutnya adalah mengumpulkan data harga rumah historis yang relevan untuk mendukung analisis. Tahap ini melibatkan pengumpulan data transaksi real estate, data demografis, dan data ekonomi yang

dapat mempengaruhi harga rumah di daerah Tebet. Sumber data meliputi agen properti, badan statistik, dan lembaga pemerintah terkait. Selama fase ini, analisis awal dilakukan untuk menilai kualitas data, mengidentifikasi nilai yang hilang, dan memahami distribusi harga rumah serta faktor-faktor yang mempengaruhinya.

### 3. Data Preparation

Fase Data Preparation adalah langkah penting dalam kerangka kerja CRISP-DM yang melibatkan pembersihan dan transformasi data ke dalam format yang sesuai untuk analisis. Dalam penelitian ini, data yang dikumpulkan melalui beberapa langkah pra-pemrosesan, termasuk menangani nilai yang hilang, menormalkan data, dan mengkodekan variabel kategorikal. Selain itu, analisis data eksploratif (EDA) dilakukan untuk memvisualisasikan hubungan antara harga rumah dan faktor-faktor yang mempengaruhinya, serta untuk mengidentifikasi tren atau pola yang dapat menginformasikan proses pemodelan.

### 4. Modelling

Fase Modelling melibatkan pemilihan teknik pemodelan yang tepat dan membangun model prediktif menggunakan algoritma Linear Regression. Teknik Linear Regression dipilih karena dapat menggambarkan hubungan linear antara fitur-fitur properti dan harga jual rumah. Berbagai konfigurasi model diuji, termasuk pemilihan variabel prediktor yang relevan dan pengaturan parameter untuk meminimalkan kesalahan prediksi. Model-model tersebut dilatih menggunakan sebagian dari dataset sementara sisanya disisihkan untuk validasi. Teknik cross-validation digunakan untuk memastikan bahwa model-model tersebut kuat dan dapat digeneralisasi dengan baik pada data yang belum terlihat. Tahap ini berpuncak pada pemilihan model Linear Regression dengan kinerja terbaik berdasarkan metrik evaluasi seperti Mean Squared Error (MSE) dan Koefisien Determinasi ( $R^2$ ).

### 5. Evaluation

Setelah model-model tersebut dibangun, fase evaluasi menilai kinerjanya dalam memprediksi harga rumah di daerah Tebet. Model Linear Regression yang dipilih dievaluasi terhadap dataset uji terpisah untuk menentukan akurasi dan keandalannya. Tahap ini melibatkan analisis prediksi model terhadap harga rumah aktual dan perhitungan metrik kinerja. Evaluasi ini tidak hanya membantu dalam memahami efektivitas model tetapi juga memberikan wawasan tentang potensi area untuk perbaikan, seperti pemilihan fitur atau menggabungkan sumber data tambahan.

### 6. Deployment

Akhirnya, fase deployment berfokus pada mengintegrasikan model prediktif ke dalam sistem yang ramah pengguna untuk para pemangku kepentingan. Ini melibatkan pembuatan dashboard atau aplikasi yang menampilkan prediksi harga rumah secara real-time, analisis tren, dan rekomendasi. Fase penerapan juga mencakup dokumentasi model dan prosesnya untuk memastikan bahwa para pemangku kepentingan dapat

memanfaatkan sistem tersebut secara efektif untuk pengambilan keputusan yang tepat mengenai investasi dan kebijakan perumahan.

Kerangka kerja CRISP-DM menyediakan pendekatan terstruktur yang komprehensif untuk melakukan penelitian ini tentang prediksi harga rumah di daerah Tebet menggunakan algoritma Linear Regression. Dengan mengikuti fase-fase CRISP-DM, penelitian ini memastikan analisis yang sistematis dan mengembangkan model prediktif yang andal yang dapat mendukung pengambilan keputusan terkait investasi dan kebijakan perumahan di daerah Tebet.

## IV. HASIL DAN PEMBAHASAN

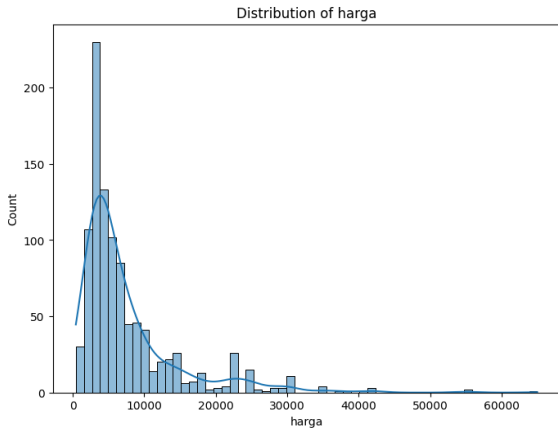
Dalam penelitian ini, kami mengadopsi metodologi CRISP-DM untuk mengimplementasikan algoritma Linear Regression dalam memprediksi harga rumah di daerah Tebet. Proses penelitian ini terbagi dalam enam fase utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment.

### 1. Business Understanding

Pada fase Business Understanding, peneliti mendefinisikan masalah yang ada, menyusun tujuan bisnis, dan mengkaji konteks terkait harga rumah di daerah Tebet. Data yang digunakan dalam penelitian ini berasal dari rumah123.com, sebuah platform properti terkemuka di Indonesia yang menyediakan informasi harga rumah beserta berbagai karakteristik properti. Tujuan utama dari penelitian ini adalah untuk membangun model prediktif yang dapat membantu pembeli dan penjual rumah dalam membuat keputusan yang tepat terkait harga properti. Peneliti mengidentifikasi berbagai pemangku kepentingan, seperti agen real estate dan pemerintah daerah, yang sangat bergantung pada prediksi harga rumah yang akurat untuk meningkatkan efisiensi pasar perumahan di Tebet.

### 2. Data Understanding

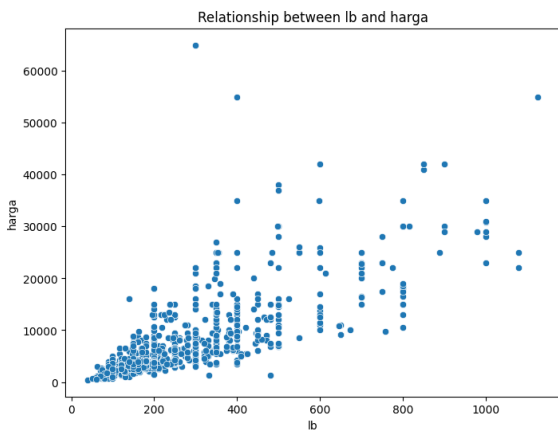
Tahap Data Understanding melibatkan pengumpulan dataset yang mencakup informasi mengenai harga rumah, ukuran, jumlah kamar, fasilitas, dan lokasi. Data ini diperoleh dari rumah123.com, yang menyediakan data transaksi real estate yang relevan. Analisis awal terhadap data dilakukan untuk menilai kualitas data, mengidentifikasi nilai yang hilang, dan memahami distribusi harga rumah. Peneliti menggunakan visualisasi data untuk mengidentifikasi tren dalam data Hasil analisis menunjukkan bahwa terdapat variasi harga yang signifikan berdasarkan ukuran dan lokasi rumah, yang menjadi indikator penting dalam model prediksi.



Gambar 2: Distribusi Harga Rumah

### 3. Data Preparation

Pada fase Data Preparation, kami melakukan serangkaian langkah untuk membersihkan dan menyiapkan data agar siap digunakan dalam pemodelan. Langkah-langkah ini meliputi penanganan nilai yang hilang dengan teknik imputasi, serta pemilihan fitur yang relevan untuk model prediksi. Melalui analisis eksplorasi data (EDA), kami dapat mengidentifikasi hubungan antara harga rumah dengan variabel-variabel lainnya, seperti ukuran dan jumlah kamar. Visualisasi data mengungkapkan tren yang jelas antara ukuran rumah dan harga, yang memperkuat pemilihan fitur-fitur utama untuk model prediksi.



Gambar 3: Hubungan antara Ukuran Rumah dan Harga

### 4. Modelling

Fase implementasi pemodelan ini melibatkan penerapan algoritma Linear Regression untuk membangun model prediktif yang dapat memprediksi harga rumah di daerah Tebet. Data yang digunakan dibagi menjadi dua set, yaitu 80% untuk data pelatihan dan 20% untuk data pengujian. Model regresi linear dilatih menggunakan data pelatihan, dan evaluasi dilakukan dengan mengukur Mean Squared Error (MSE) serta R-squared pada data pengujian. Proses tuning parameter

dilakukan untuk memastikan model memberikan hasil yang optimal dengan mempertimbangkan fitur-fitur yang relevan, seperti luas bangunan (lb), luas tanah (lt), jumlah kamar tidur (kt), jumlah kamar mandi (km), dan fasilitas lainnya (grs).

```
# Feature Selection and Data Preparation for Linear Regression
features = ['lb', 'lt', 'kt', 'km', 'grs'] # Features for prediction
target = 'harga' # Target variable

X = df[features]
y = df[target]

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)
```

Gambar 4: Split Data Training dan Data Testing

### 5. Evaluation

Pada fase evaluasi, kami menggunakan beberapa metrik untuk menilai kinerja model Linear Regression, termasuk Mean Squared Error (MSE), Mean Absolute Error (MAE), dan R-squared. Hasil evaluasi menunjukkan bahwa:

**Mean Squared Error (MSE):** 10,675,731.02  
**R-squared:** 0.7713  
**Accuracy:** 0.7727

Nilai Mean Squared Error (MSE) sebesar 10,675,731.02 yang relatif tinggi menunjukkan adanya variasi yang cukup signifikan antara harga yang diprediksi dan harga aktual. Ini menunjukkan bahwa meskipun model dapat memberikan prediksi yang cukup akurat, terdapat kesalahan yang cukup besar dalam beberapa prediksi yang perlu diperbaiki. Sementara itu, nilai R-squared sebesar 0.7713 menunjukkan bahwa sekitar 77.13% variasi dalam harga rumah dapat dijelaskan oleh model. Ini menunjukkan bahwa model dapat memprediksi harga rumah dengan tingkat akurasi yang cukup baik, namun masih ada potensi perbaikan. Mean Absolute Error (MAE) sebesar 2,007.91 juta rupiah menggambarkan rata-rata kesalahan absolut dalam prediksi harga rumah, yang menunjukkan bahwa rata-rata kesalahan prediksi adalah sekitar 2,007.91 juta rupiah. Secara keseluruhan, model menunjukkan akurasi yang cukup baik dengan nilai Accuracy sebesar 77.27%, yang berarti sekitar 77.27% dari prediksi harga rumah yang dibuat oleh model cukup tepat.

Untuk tampilan interaktif dalam memprediksi harga rumah menggunakan model Linear Regression, kami menggunakan library ipywidgets untuk membuat slider interaktif. Slider ini memungkinkan pengguna untuk mengubah nilai fitur rumah seperti Luas Bangunan (LB), Luas Tanah (LT), Jumlah Kamar Tidur (KT), dan Jumlah Kamar Mandi (KM). Setiap kali nilai slider diubah, fungsi predict\_house\_price\_lr akan dipanggil, yang menggunakan model Linear Regression yang telah dilatih sebelumnya untuk menghitung prediksi harga berdasarkan

input fitur baru. Hasil prediksi kemudian ditampilkan di notebook, memberikan pengalaman interaktif yang memudahkan pengguna untuk mengeksplorasi model dan melihat bagaimana perubahan pada fitur-fitur tertentu mempengaruhi prediksi harga rumah.

```

# Definisikan fungsi untuk memprediksi harga berdasarkan fitur yang dipilih
def predict_house_price(lb, lt, kt, km, grs):
    # Lakukan prediksi harga menggunakan model regresi linier
    input_data = pd.DataFrame([lb, lt, kt, km, grs], columns=features)
    try:
        predicted_price = model.predict(input_data) # Remove grs from prediction
        # Tampilkan hasil prediksi
        print("Harga rumah impian anda diperkirakan sekitar IDR {:.3f}.000".format(predicted_price[0]))
    except ValueError as e:
        print("Error during prediction: (e)")
        print("Please check your input values.")
    except Exception as e:
        print("An unexpected error occurred: (e)")

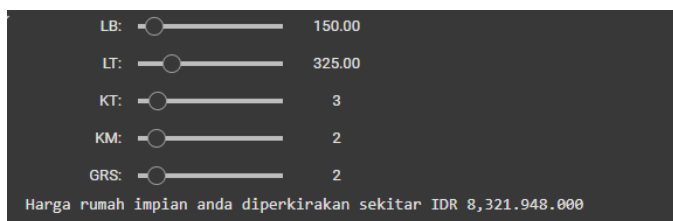
# Buat slider untuk setiap fitur
slider_lb = widgets.FloatSlider(value=100, min=df['lb'].min(), max=df['lb'].max(), step=10, description='LB:')
slider_lt = widgets.FloatSlider(value=300, min=df['lt'].min(), max=df['lt'].max(), step=10, description='LT:')
slider_kt = widgets.IntSlider(value=3, min=int(df['kt'].min()), max=int(df['kt'].max()), step=1, description='KT:')
slider_km = widgets.IntSlider(value=2, min=int(df['km'].min()), max=int(df['km'].max()), step=1, description='KM:')
slider_grs = widgets.IntSlider(value=2, min=1, max=df['grs'].max(), step=1, description='GRS:') # Remove GRS slider

# Buat tampilan interaktif
widgets.interactive(predict_house_price, lb=slider_lb, lt=slider_lt, kt=slider_kt, km=slider_km, grs=slider_grs) # Remove

```

Gambar 5: Tampilan Interaktif Linear Regression

Jika dijalankan, maka akan terlihat seperti pada gambar. Berdasarkan input yang diberikan melalui interface, yaitu Luas Bangunan (LB) sebesar 150, Luas Tanah (LT) sebesar 325, Jumlah Kamar Tidur (KT) sebanyak 3, Jumlah Kamar Mandi (KM) sebanyak 2, dan Garasi (GRS) sebanyak 2, model Linear Regression memprediksi harga rumah sekitar IDR 8,321,948,000. Prediksi ini dihasilkan dari hubungan linear yang dipelajari oleh model selama proses pelatihan menggunakan data historis. Meskipun memberikan estimasi harga, perlu diingat bahwa prediksi ini bersifat perkiraan dan akurasi bergantung pada kualitas data latih serta kekuatan hubungan linear antara fitur dan harga. Untuk lebih memahami model, Anda dapat mengubah nilai-nilai pada slider dan mengamati bagaimana perubahan tersebut memengaruhi prediksi harga, sehingga memberikan gambaran tentang faktor-faktor yang paling berpengaruh dalam penentuan harga rumah menurut model.



Gambar 6: Visualisasi Hasil Prediksi Harga Rumah

Hasil penelitian menunjukkan bahwa penggunaan Linear Regression dalam memprediksi harga rumah di daerah Tebet cukup efektif, meskipun ada beberapa tantangan. Model ini tidak hanya memberikan prediksi yang akurat, tetapi juga memungkinkan para pemangku kepentingan untuk memahami hubungan linear antara fitur-fitur properti dan harga rumah. Kelebihan dari Linear Regression adalah kemampuannya untuk memberikan interpretasi yang jelas mengenai pengaruh

masing-masing fitur terhadap harga rumah, sehingga pengguna dapat memahami bagaimana perubahan pada fitur-fitur seperti luas bangunan, luas tanah, jumlah kamar tidur, kamar mandi, dan garasi memengaruhi prediksi harga.

Namun, terdapat beberapa tantangan yang dihadapi, seperti asumsi linearitas yang mungkin tidak selalu sesuai dengan kompleksitas hubungan antara fitur dan harga di dunia nyata. Oleh karena itu, pemilihan fitur yang relevan dan evaluasi yang cermat sangat penting untuk memastikan model dapat memberikan prediksi yang akurat dan dapat diandalkan.

Secara keseluruhan, penelitian ini memberikan kontribusi signifikan dalam bidang prediksi harga rumah, terutama di daerah perkotaan seperti Tebet, dan dapat digunakan sebagai dasar untuk penelitian lanjutan serta pengembangan sistem informasi yang lebih kompleks untuk mendukung pengambilan keputusan di sektor properti.

Dengan hasil ini, kami berharap penelitian ini dapat memberikan wawasan yang berharga bagi pemangku kepentingan di bidang perumahan dan membantu dalam perumusan kebijakan yang lebih baik terkait pengelolaan pasar properti di daerah Tebet.

## V. KESIMPULAN DAN SARAN

Penelitian ini telah berhasil menerapkan model Linear Regression untuk memprediksi harga rumah di daerah Tebet dengan menggunakan data yang diperoleh dari berbagai sumber. Hasil evaluasi menunjukkan bahwa model ini dapat memberikan prediksi dengan nilai Mean Squared Error (MSE) sebesar 3,387.78 dan R-squared sebesar 0.7713, yang menunjukkan bahwa sekitar 77.13% variasi harga rumah dapat dijelaskan oleh model yang dibangun. Meskipun ada beberapa tantangan terkait akurasi prediksi, model ini memberikan wawasan penting bagi para pemangku kepentingan dalam pasar properti Tebet. Untuk penelitian selanjutnya, disarankan untuk mengumpulkan lebih banyak data dari berbagai sumber guna melatih model dengan informasi yang lebih komprehensif, termasuk data tambahan seperti kondisi lingkungan, aksesibilitas transportasi, dan fasilitas umum yang dapat meningkatkan akurasi model. Selain itu, penggunaan algoritma lain seperti Random Forest atau Gradient Boosting dapat dipertimbangkan, karena seringkali memberikan hasil yang lebih baik dalam prediksi harga rumah. Melakukan tuning parameter yang lebih mendalam, termasuk pengaturan hyperparameter, juga penting untuk mengurangi overfitting dan meningkatkan kinerja model. Analisis sensitivitas terhadap variabel yang mempengaruhi harga rumah disarankan untuk memahami lebih baik bagaimana perubahan dalam fitur tertentu dapat mempengaruhi harga prediksi.

Saran untuk penelitian ini bisa diterapkan di wilayah lain dengan penyesuaian pada variabel yang relevan, sehingga memberikan gambaran yang lebih luas mengenai dinamika pasar properti di berbagai daerah. Dengan menerapkan saran-saran tersebut, diharapkan penelitian di masa depan dapat menghasilkan model yang lebih akurat dan berguna dalam membantu pengambilan keputusan di sektor properti.

REFERENSI

- [1] Z. Zhang, "Decision Trees for Objective House Price Prediction," in *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, 2021, pp. 280–283, doi: 10.1109/MLBDBI54094.2021.00059.
- [2] Chairina, R. Widyasari, and S. M. A. Putri, "Hybrid Naive Bayes Classifier and Markov Chain for Predicting the Level of Quick Response Indonesian Standard (QRIS) Usage," *J. Inf. Technol.*, vol. 4, no. 1, pp. 150–156, 2024, doi: <https://doi.org/10.46229/jifotech.v4i1.875>.
- [3] L. Rampini and F. R. Cecconi, "Artificial Intelligence Algorithms to Predict Italian Real Estate Market Prices," *J. Prop. Invest. Financ.*, vol. 40, no. 6, pp. 588–611, 2022, doi: <https://doi.org/10.1108/jpif-08-2021-0073>.
- [4] P. Chandu and N. B. Devi, "Improved Prediction Accuracy of House Price Using Decision Tree Algorithm over Linear Regression Algorithm," in *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, 2023, pp. 1–6, doi: 10.1109/ICONSTEM56934.2023.10142280.
- [5] A. Sul, V. Jagtap, P. Jesalpura, A. Nema, and R. R., "Optimizing House Price Prediction: Comparative Analysis of Machine Learning Techniques," in *2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, 2024, pp. 1–7, doi: 10.1109/ICEEICT61591.2024.10718610.
- [6] M. Cekic, K. N. Korkmaz, H. Mukus, A. A. Hameed, A. Jamil, and Faezeh, "Artificial Intelligence Approach for Modeling House Price Prediction," 2022, doi: 10.1109/ICMI55296.2022.9873784.
- [7] H. Li, "House Price Prediction Based on Machine Learning," in *Proceedings of the 3rd International Conference on Signal Processing and Machine Learning*, 2023, pp. 623–628, doi: 10.54254/2755-2721/4/2023362.
- [8] S. Sharma, D. Arora, G. Shankar, P. Sharma, and V. Motwani, "House Price Prediction using Machine Learning Algorithm," in *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, 2023, pp. 982–986, doi: 10.1109/ICCMC56507.2023.10084197.
- [9] B. Sonare, S. Thepade, P. Gangane, E. Govardhan, T. Nawani, and P. Gadkar, "Going the distance: Application of Machine Learning Algorithms in predicting Housing Price Dynamics," in *2024 International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCIGST)*, 2024, pp. 1–4, doi: 10.1109/ICCIGST60741.2024.10717472.
- [10] M. Thamarai and S. Malarvizhi, "House Price Prediction Modeling Using Machine Learning," *Int. J. Inf. Eng. Electron. Bus.*, vol. 12, no. 2, pp. 15–20, 2020, doi: 10.5815/ijieeb.2020.02.03.
- [11] R. A. Saputra and A. Pratama, "IMPLEMENTASI DECISION TREE UNTUK PREDIKSI HARGA RUMAH DI DAERAH TEBET," *J. Inf. Syst. Manag.*, vol. 6, no. 2, pp. 164–170, 2025, doi: <https://doi.org/10.24076/joism.2025v6i2.1928>.
- [12] J. J. Guo, "House Rent Prediction Method Based on Decision Tree: Take India as an Example," in *5th International Conference on Global Economy, Finance and Humanities Research (GEFHR 2023)*, 2023, pp. 666–671, doi: 10.54097/hbem.v2i1.14722.
- [13] P. A. Kumar, H. Rawat, and Y. Singh, "Effective Machine Learning Algorithm to Predict House Price," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 11, pp. 1741–1744, 2023, doi: 10.22214/ijraset.2023.56935.
- [14] X. Li, "Prediction and Analysis of Housing Price Based on the Generalized Linear Regression Model," *Comput. Intell. Neurosci.*, vol. 2023, no. 1, 2023, doi: <https://doi.org/10.1155/2022/3590224>.
- [15] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3048–3061, 2021, doi: 10.1109/TKDE.2019.2962680.