

Prediksi Harga Kos di Wilayah Jimbaran Menggunakan Algoritma Decision Tree Regression Berbasis Web

I Putu Krishna Paradipta¹⁾; Putu Manik Prihatini²⁾; Ni Nyoman Harini Puspita³⁾

^{1,2,3)} Program Studi Sarjana Terapan Teknologi Rekayasa Perangkat Lunak, Jurusan Teknologi Informasi, Politeknik Negeri Bali
Email: krishnaparadipta@gmail.com

How to Cite :

Paradipta, I, P, K., Prihatini, P, M., Puspita, N, N, H. (2026). Prediksi Harga Kos di Wilayah Jimbaran Menggunakan Algoritma Decision Tree Regression Berbasis Web. Jurnal Media Computer Science, 5(3).

ARTICLE HISTORY

Received [05 Juni 2026]

Revised [15 Juli 2026]

Accepted [22 Juli 2026]

KEYWORDS

Decision Tree Regression, Prediksi Harga Kos, Machine Learning, Data Mining, Streamlit.

This is an open access article under the
[CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Kecamatan Jimbaran merupakan salah satu kawasan strategis di Kabupaten Badung, Bali, yang mengalami perkembangan pesat pada sektor pendidikan, pariwisata, dan ekonomi. Kondisi tersebut menyebabkan meningkatnya kebutuhan tempat tinggal sementara, khususnya rumah kos bagi mahasiswa dan pekerja pendatang. Banyaknya pilihan kos dengan variasi fasilitas dan harga seringkali menyulitkan calon penyewa dalam menentukan kos yang sesuai dengan kebutuhan dan kemampuan finansial. Penelitian ini bertujuan untuk membangun aplikasi prediksi harga kos berbasis web menggunakan algoritma *Decision Tree Regression* untuk membantu pengguna memperkirakan harga kos berdasarkan fasilitas yang diinginkan. Penelitian ini menggunakan pendekatan kuantitatif dengan metode *machine learning* dan *data mining*. Dataset yang digunakan berupa data kos di wilayah Jimbaran yang terdiri dari atribut luas kamar, jenis kamar mandi, fasilitas kamar, fasilitas umum, jarak ke kampus, dan harga kos per bulan. Tahapan penelitian meliputi preprocessing data, transformasi fitur, pembagian dataset menggunakan metode *Train-Test Split*, pelatihan model, dan evaluasi model menggunakan metrik MAE, RMSE, dan R^2 . Hasil penelitian menunjukkan bahwa algoritma *Decision Tree Regression* mampu menghasilkan performa yang cukup baik dengan nilai R^2 sebesar 0,8332, MAE sebesar Rp225.462, dan RMSE sebesar Rp302.509.

ABSTRACT

Jimbaran District is one of the strategic areas in Badung Regency, Bali, experiencing rapid growth in the education, tourism, and economic sectors. This condition has increased the demand for temporary housing, especially boarding houses for students and migrant workers. The large number of boarding house options with various facilities and rental prices often makes it difficult for prospective tenants to choose accommodations that match their needs and financial capabilities. This study aims to develop a web-based boarding house price prediction application using the *Decision Tree Regression* algorithm to assist users in estimating boarding house prices based on the desired facilities. The research applies a quantitative approach using *machine learning* and *data mining* methods. The dataset consists of boarding house data in the Jimbaran area, including room size, bathroom type, room facilities, public facilities, distance to campus, and monthly rental price. The research stages include data preprocessing, feature transformation, dataset splitting using the

Train-Test Split method, model training, and model evaluation using MAE, RMSE, and R^2 metrics. The results show that the Decision Tree Regression algorithm achieved good performance with an R^2 value of 0.8332, an MAE value of IDR 225,462, and an RMSE value of IDR 302,509.

PENDAHULUAN

Kecamatan Jimbaran merupakan salah satu wilayah strategis yang berada di Kabupaten Badung, Provinsi Bali. Wilayah ini mengalami perkembangan yang cukup pesat setiap tahunnya, baik pada sektor infrastruktur, ekonomi, pendidikan, maupun pariwisata. Keberadaan berbagai fasilitas umum, pusat aktivitas wisata, dan institusi pendidikan menjadikan Jimbaran sebagai kawasan dengan tingkat mobilitas penduduk yang tinggi (Rekayasa et al., 2021). Selain itu, Jimbaran juga menjadi salah satu daerah yang banyak diminati oleh mahasiswa, pekerja, serta pendatang dari berbagai wilayah. Kehadiran beberapa perguruan tinggi seperti Politeknik Negeri Bali, kawasan industri pariwisata, serta akses yang dekat dengan pusat Kota Denpasar dan Bandara Internasional I Gusti Ngurah Rai menjadikan Jimbaran sebagai lokasi yang sering dipilih untuk tempat tinggal sementara maupun jangka panjang. Kondisi tersebut menyebabkan kebutuhan akan tempat tinggal, khususnya rumah kos, terus mengalami peningkatan.

Meningkatnya jumlah pendatang, terutama mahasiswa dan pekerja, berdampak pada bertambahnya jumlah rumah kos yang tersedia di wilayah Jimbaran. Rumah kos yang tersedia memiliki beragam fasilitas, lokasi, serta tingkat kenyamanan yang berbeda-beda (Labib Mu'tashim et al., 2021). Namun, keberagaman tersebut seringkali menyulitkan calon penyewa dalam menentukan pilihan kos yang sesuai dengan kebutuhan dan kemampuan finansial mereka. Fasilitas seperti ukuran kamar, kamar mandi dalam, akses internet, area parkir, sistem keamanan, hingga jarak menuju pusat aktivitas memiliki pengaruh terhadap harga sewa yang ditetapkan oleh pemilik kos (Zahra, 2023). Akan tetapi, harga yang ditawarkan tidak selalu sebanding dengan fasilitas yang tersedia. Di sisi lain, proses pencarian dan pemilihan kos yang masih dilakukan secara manual membutuhkan waktu yang cukup lama dan berpotensi menghasilkan keputusan yang kurang tepat, khususnya bagi calon penyewa dengan keterbatasan anggaran. Oleh karena itu, diperlukan sebuah aplikasi yang dapat membantu calon penyewa dalam menganalisis dan memprediksi harga kos berdasarkan fasilitas yang diinginkan.

Perkembangan teknologi informasi memberikan peluang dalam pemanfaatan metode data mining sebagai solusi untuk membantu proses penentuan harga kos sesuai dengan fasilitas yang tersedia (Muharrom, 2023). Metode data mining memungkinkan pengolahan data kos dalam jumlah besar untuk menemukan pola dan hubungan antara fasilitas dengan harga sewa yang sebelumnya sulit dianalisis secara manual. Salah satu algoritma yang dapat digunakan dalam proses prediksi adalah Decision Tree Regression. Algoritma ini mampu menghasilkan model keputusan yang mudah dipahami, memiliki struktur yang jelas, serta dapat mengolah data numerik maupun kategorikal (Ramadhan et al., 2023). Melalui penerapan algoritma Decision Tree Regression, data fasilitas rumah kos dapat dianalisis untuk menghasilkan prediksi harga sewa yang lebih akurat sehingga dapat membantu calon penyewa dalam menentukan pilihan kos sesuai kebutuhan fasilitas yang diinginkan.

Beberapa penelitian terdahulu telah membahas penerapan machine learning dan data mining dalam proses prediksi maupun pengambilan keputusan. Penelitian oleh (Saputra & Pratama, 2025) menunjukkan bahwa algoritma Decision Tree Regression mampu digunakan untuk memprediksi data numerik dengan struktur model yang mudah dipahami dan memiliki performa yang cukup baik. Selain itu, penelitian oleh (Wijaya & Triayudi, 2023) menjelaskan bahwa metode data mining dapat membantu proses analisis data secara lebih efektif. Meskipun demikian, penelitian terdahulu masih memiliki beberapa keterbatasan, seperti belum tersedianya implementasi aplikasi berbasis web yang dapat digunakan langsung oleh pengguna, belum

optimalnya penerapan preprocessing data dan feature engineering, serta belum adanya penelitian yang secara khusus menggunakan data rumah kos di wilayah Jimbaran.

LANDASAN TEORI

Data Mining

Data mining merupakan bidang ilmu yang mempelajari proses penggalian informasi atau pengetahuan yang bernilai dari kumpulan data berukuran besar. Proses ini dilakukan dengan menerapkan berbagai metode, seperti statistik, matematika, dan kecerdasan buatan, untuk menemukan pola, hubungan, tren, maupun struktur tersembunyi yang sulit dikenali secara langsung melalui pengamatan biasa. Pemanfaatan data mining telah banyak diterapkan dalam berbagai bidang karena mampu membantu proses pengambilan keputusan melalui analisis data historis yang tersedia (Rahmadini, 2023). Selain itu, data mining juga digunakan untuk membantu proses prediksi dan analisis pola data sehingga informasi yang dihasilkan dapat dimanfaatkan sebagai dasar pengambilan keputusan yang lebih baik (Nuruzzaman et al., 2023).

Machine Learning

Machine learning adalah salah satu cabang dari kecerdasan buatan (Artificial Intelligence) yang berfokus pada pengembangan algoritma dan metode agar komputer mampu mempelajari pola dari data secara otomatis tanpa perlu diprogram secara rinci (Agbulut et al., 2021). Dalam prosesnya, machine learning memanfaatkan data latih untuk membentuk model matematis yang dapat merepresentasikan hubungan antarvariabel dalam data. Model tersebut kemudian digunakan untuk melakukan prediksi, klasifikasi, maupun pengambilan keputusan terhadap data baru yang sebelumnya belum pernah diproses (Narayana Raju et al., 2024).

Preprocessing

Preprocessing atau pengolahan data merupakan tahapan awal yang memiliki peranan penting dalam proses machine learning. Tahapan ini dilakukan untuk mempersiapkan data mentah agar berada dalam kondisi yang baik sebelum digunakan dalam proses pelatihan model. Data yang diperoleh dari berbagai sumber sering kali masih mengandung permasalahan, seperti nilai kosong (missing value), inkonsistensi data, kesalahan format, maupun ketidaksesuaian tipe data. Jika data tersebut langsung digunakan tanpa proses preprocessing, maka kualitas model yang dihasilkan dapat menurun (Vinita Yadav & Kumar, 2025).

Feature Engineering

Feature engineering merupakan proses membentuk atribut baru dari data asli untuk meningkatkan kemampuan model dalam mengenali pola. Pada penelitian ini dilakukan pembentukan fitur seperti `total_fasilitas`, `premium_score`, `fasilitas_per_m2`, dan `premium_per_m2` untuk membantu model memprediksi harga kos secara lebih baik.

Train-Test Split

Train-Test Split merupakan metode validasi dalam machine learning yang digunakan untuk membagi dataset menjadi dua bagian, yaitu data latih (training data) dan data uji (testing data). Data latih digunakan dalam proses pembentukan model, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data baru yang belum pernah dipelajari sebelumnya (Salazar et al., 2022). Tujuan utama penggunaan metode Train-Test Split adalah untuk mengukur kemampuan generalisasi model sehingga hasil evaluasi yang diperoleh tidak bersifat bias. Pembagian dataset biasanya dilakukan dengan proporsi tertentu, misalnya 80% data latih dan 20% data uji, meskipun proporsi tersebut dapat disesuaikan dengan jumlah serta karakteristik data yang digunakan (Bichri et al., 2024).

Regresi

Regresi merupakan salah satu metode dalam machine learning yang digunakan untuk menganalisis hubungan antara variabel independen (X) dan variabel dependen (Y) yang memiliki nilai numerik atau kontinu. Metode regresi bertujuan untuk membentuk suatu model matematis yang dapat menggambarkan hubungan antarvariabel sehingga nilai variabel dependen dapat diprediksi berdasarkan variabel independen yang tersedia (Hasanah, 2023). Dalam penerapannya, regresi banyak dimanfaatkan pada berbagai bidang untuk keperluan analisis data, prediksi, serta mendukung proses pengambilan keputusan berbasis data (Muttaqin & Srihartini, 2022).

Decision Tree Regression

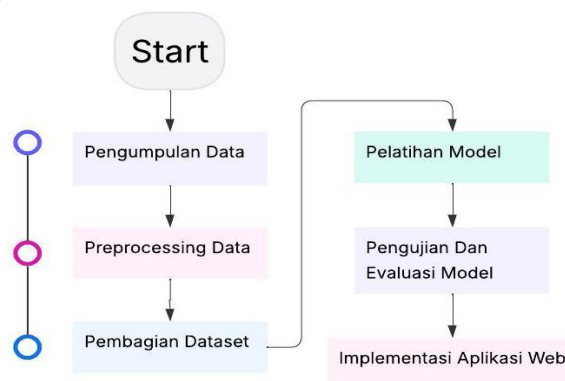
Decision Tree Regression merupakan salah satu algoritma regresi dalam machine learning yang menggunakan struktur pohon keputusan untuk memodelkan hubungan antara variabel independen dengan variabel dependen yang bersifat numerik. Algoritma ini bekerja dengan membangun model prediksi berbentuk pohon yang terdiri atas root node, internal node, dan leaf node. Root node merupakan titik awal pembentukan pohon, internal node berfungsi sebagai pemisah data berdasarkan atribut tertentu, sedangkan leaf node menunjukkan hasil prediksi atau nilai keluaran (Karim et al., 2021). Metode Decision Tree Regression melakukan proses pembagian data secara bertahap berdasarkan atribut yang dianggap paling optimal dalam mengurangi kesalahan prediksi (Nurani et al., 2023). Setiap percabangan pada pohon merepresentasikan aturan keputusan tertentu hingga data terbagi ke dalam kelompok-kelompok yang memiliki karakteristik serupa. Salah satu keunggulan algoritma ini adalah kemampuannya dalam menghasilkan model yang mudah dipahami dan divisualisasikan, sehingga memudahkan interpretasi hasil prediksi. Selain itu, Decision Tree Regression juga mampu menangani data numerik maupun kategorikal dengan baik serta tidak memerlukan asumsi hubungan linear antarvariabel (Hidayah et al., 2024).

METODE PENELITIAN

Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan jenis penelitian terapan (applied research) yang bertujuan menghasilkan solusi praktis terhadap permasalahan estimasi harga kos di wilayah Jimbaran melalui pembangunan aplikasi prediksi harga kos berbasis web menggunakan algoritma Decision Tree Regression. Penelitian ini juga termasuk penelitian prediktif karena berfokus pada pembangunan model yang mampu memprediksi nilai numerik berupa harga kos berdasarkan fasilitas kos.

Pendekatan yang digunakan adalah pendekatan kuantitatif dengan metode data mining dan machine learning. Pendekatan kuantitatif dipilih karena penelitian memanfaatkan data numerik seperti harga kos, luas kamar, dan jarak ke kampus, serta menggunakan proses perhitungan matematis dalam pembangunan model prediksi. Evaluasi model dilakukan menggunakan metrik regresi berupa *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R-Squared* (R^2).



Gambar 1 Tahapan Penelitian

HASIL DAN PEMBAHASAN

Hasil

Pengumpulan Data

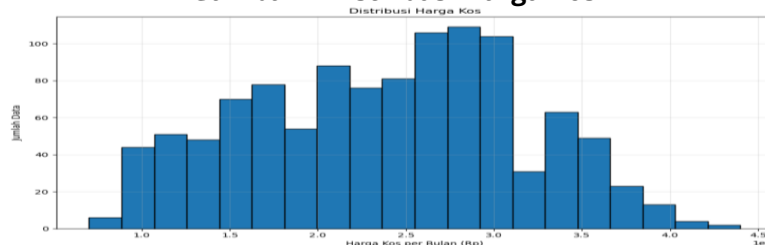
Hasil pengumpulan data menunjukkan bahwa penelitian berhasil memperoleh sekitar 1100 data kos di wilayah Jimbaran yang terdiri dari atribut lokasi, tipe kos, ukuran kamar, fasilitas kamar, fasilitas kamar mandi, fasilitas umum, jarak ke kampus, dan harga kos per bulan sebagai variabel target. Data tersebut kemudian digunakan sebagai dasar dalam proses pembangunan model prediksi harga kos.

Tabel 1 Hasil Pengumpulan Data

Atribut	Keterangan
Lokasi	Lokasi Kos
tipe_kos	Jenis kos
ukuran_kamar	Ukuran kamar kos
fasilitas_kamar	Fasilitas kamar kos
fasilitas_kamar_mandi	Fasilitas kamar mandi kos
fasilitas_umum	Fasilitas umum kos
jarak_ke_kampus	Jarak ke kampus Politeknik Negeri Bali
harga_bulan	Harga kos per bulan

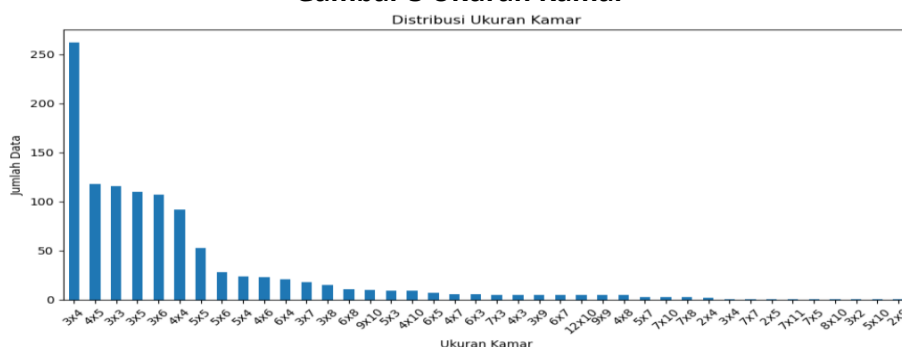
Hasil pengumpulan data penelitian terdiri dari beberapa atribut yang digunakan untuk memprediksi harga kos, yaitu lokasi, tipe kos, ukuran kamar, fasilitas kamar, fasilitas kamar mandi, fasilitas umum, dan jarak ke kampus. Atribut-atribut tersebut dipilih karena dapat menggambarkan karakteristik dan kualitas kos yang berpengaruh terhadap harga sewa. Sementara itu, atribut harga_bulan digunakan sebagai variabel target yang akan diprediksi oleh model Decision Tree Regression.

Gambar 2 Distribusi Harga Kos



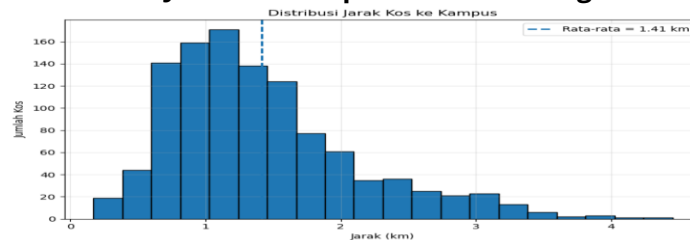
Berdasarkan Gambar 2, sebagian besar data kos berada pada rentang harga Rp1.500.000 hingga Rp3.000.000 per bulan. Hal ini menunjukkan bahwa mayoritas kos di wilayah Jimbaran termasuk dalam kategori harga menengah. Selain itu, jumlah kos dengan harga di atas Rp5.000.000 relatif lebih sedikit sehingga distribusi data tidak merata pada seluruh rentang harga.

Gambar 3 Ukuran Kamar



Berdasarkan Gambar 3, sebagian besar kos memiliki luas kamar antara 9 m² hingga 16 m². Kondisi ini menunjukkan bahwa ukuran kamar sedang merupakan tipe kamar yang paling banyak tersedia pada dataset penelitian.

Gambar 4 Jarak Ke Kampus Politeknik Negeri Bali



Berdasarkan Gambar 4, sebagian besar kos berada pada jarak kurang dari 3 km dari kampus. Hal ini menunjukkan bahwa lokasi yang dekat dengan kampus menjadi karakteristik dominan pada data kos yang digunakan dalam penelitian.

Hasil analisis distribusi data menunjukkan bahwa dataset memiliki variasi karakteristik yang cukup beragam pada aspek harga, luas kamar, dan jarak ke kampus. Variasi tersebut memberikan informasi yang cukup bagi model Decision Tree Regression untuk mempelajari pola hubungan antara atribut kos dan harga sewa sehingga mendukung proses pembangunan model prediksi.

Preprocessing Data

Hasil preprocessing menunjukkan bahwa seluruh data berhasil dibersihkan dan ditransformasikan ke dalam format yang sesuai untuk proses machine learning. Proses yang dilakukan meliputi pembersihan nama kolom, penghapusan data duplikat, validasi tipe data, filtering harga, pengolahan ukuran kamar dan jarak kampus, pengisian nilai kosong, encoding fasilitas, feature engineering, transformasi logaritma, stratifikasi data, dan One Hot Encoding, hasil preprocessing dapat dilihat pada tabel 5-7 dibawah ini.

Tabel 2 Hasil Filtering Harga

Kriteria	Nilai
Harga Minimum	Rp 1.000.000
Harga Maksimum	Rp 6.000.000

Hasil proses filtering harga, menunjukkan bahwa data harga kos yang digunakan dalam penelitian berada pada rentang Rp1.000.000 hingga Rp6.000.000 per bulan. Pembatasan rentang harga ini dilakukan untuk mengurangi pengaruh data ekstrem (outlier) yang dapat menyebabkan model menghasilkan pola prediksi yang kurang stabil. Dengan mempertahankan data pada rentang harga yang umum ditemukan di wilayah Jimbaran, model dapat lebih fokus mempelajari karakteristik kos yang relevan sehingga meningkatkan kualitas proses pelatihan dan akurasi prediksi.

Tabel 3 Hasil Encoding Fasilitas

Data Awal	AC	Wifi	Lokasi
AC, Wifi, Lokasi	1	1	1
Wifi, Lokasi	1	0	1
AC	1	0	0

Hasil proses encoding menunjukkan bahwa data fasilitas yang sebelumnya berbentuk teks berhasil diubah menjadi data numerik biner bernilai 0 dan 1. Proses ini memungkinkan algoritma Decision Tree Regression mengenali keberadaan setiap fasilitas secara individual. Selain itu, teknik ini mampu mempertahankan informasi fasilitas secara lebih detail dibandingkan jika seluruh fasilitas digabungkan menjadi satu atribut.

Tabel 4 Hasil Feature Engineering

Nama Fitur	Fungsi
total_fasilitas	Menghitung jumlah fasilitas kos
premium_score	Mengukur kualitas fasilitas premium
kamar_besar	Menandai kamar berukuran besar
super_dekat	Menandai kos yang sangat dekat dengan kampus
fasilitas_per_m ²	Rasio fasilitas terhadap luas kamar
premium_per_m ²	Rasio fasilitas premium terhadap luas kamar

Hasil proses feature engineering menghasilkan beberapa fitur baru yang tidak tersedia pada data asli. Fitur-fitur tersebut dirancang untuk merepresentasikan kualitas kos secara lebih baik dibandingkan atribut dasar. Sebagai contoh, fitur premium_score mampu menggambarkan tingkat eksklusivitas fasilitas yang dimiliki kos, sedangkan fitur fasilitas_per_m2 dan premium_per_m2 menggambarkan kualitas fasilitas relatif terhadap ukuran kamar. Berdasarkan hasil pengujian model, kedua fitur tersebut memberikan kontribusi yang signifikan terhadap peningkatan performa prediksi karena mampu menangkap hubungan yang lebih kompleks antara fasilitas dan harga kos.

Hasil Pembagian Dataset

Dataset dibagi menggunakan metode Train-Test Split dengan komposisi 80% data training dan 20% data testing. Selain itu, penelitian ini menggunakan stratifikasi data berdasarkan kategori harga kos agar distribusi data pada data training dan testing tetap seimbang, hasil pembagian dataset dapat dilihat pada tabel 8.

Tabel 5 Pembagian Dataset

Jenis Data	Persentase
Data Training	80%
Data Testing	20%

Hasil pembagian dataset menunjukkan bahwa 80% data digunakan sebagai data training dan 20% data digunakan sebagai data testing. Data training digunakan untuk melatih model Decision Tree Regression agar dapat mempelajari pola hubungan antara karakteristik kos dan harga kos. Sementara itu, data testing digunakan untuk menguji kemampuan model dalam melakukan prediksi pada data yang belum pernah digunakan selama proses pelatihan

Pelatihan Model

Hasil Pelatihan model yang dilakukan menggunakan algoritma Decision Tree Regression dengan bantuan library Scikit-learn. Penelitian ini menggunakan Pipeline dan ColumnTransformer agar preprocessing dan pelatihan model berjalan secara terintegrasi. Selain itu, dilakukan proses hyperparameter tuning menggunakan GridSearchCV dan StratifiedKFold Cross Validation sebanyak 5 fold untuk mencari kombinasi parameter terbaik. Hasil hyperparameter tuning dan evaluasi model dapat dilihat pada tabel 9 dibawah ini.

Tabel 6 Hasil Hyperparameter Tuning

Hyperparameter	Nilai Terbaik
max_depth	8
min_samples_split	5
min_samples_leaf	2
max_leaf_nodes	80

Berdasarkan tabel hasil tuning, diperoleh parameter terbaik yaitu max_depth = 8, min_samples_split = 5, min_samples_leaf = 2, dan max_leaf_nodes = 80 seperti yang ditunjukkan pada Tabel 9. Nilai max_depth = 8 menghasilkan kompleksitas pohon yang cukup untuk menangkap

pola data tanpa menyebabkan overfitting, sedangkan parameter lainnya membantu meningkatkan kemampuan generalisasi model. Oleh karena itu, kombinasi parameter tersebut dipilih sebagai model akhir pada penelitian ini.

Pengujian dan Evaluasi

Hasil pengujian dan evaluasi model menunjukkan tingkat performa algoritma Decision Tree Regression dalam memprediksi harga kos berdasarkan data testing. Evaluasi dilakukan menggunakan metrik R-Squared (R^2), Mean Absolute Error (MAE), dan Root Mean Squared Error (RMSE) untuk mengukur akurasi dan tingkat kesalahan prediksi model. Hasil Evaluasi Model dapat dilihat pada tabel 10 dibawah ini.

Tabel 7 Hasil Pengujian dan Evaluasi Model

Data	R^2 Score	MAE	RMSE
Training	0.8871	Rp 189.898	Rp 251.789
Testing	0.8332	Rp 225.462	Rp 302.509

Berdasarkan tabel 7 model memperoleh nilai R^2 sebesar 0,8871 pada data training dan 0,8332 pada data testing. Hasil tersebut menunjukkan bahwa model mampu menjelaskan sekitar 83,32% variasi harga kos berdasarkan karakteristik fisik kamar, fasilitas yang disediakan dan lokasi kos. Hasil ini menunjukan bahwa atribut-atribut yang digunakan memiliki hubungan yang kuat terhadap harga sewa kos di wilayah Jimbaran sehingga model mampu menghasilkan prediksi yang cukup baik. Selain itu, nilai MAE sebesar Rp225.462 dan RMSE sebesar Rp302.509 menunjukkan bahwa rata-rata kesalahan prediksi masih berada dalam batas yang dapat diterima. Selisih nilai R^2 antara data training dan testing sebesar 0,0539 menunjukkan bahwa model tidak mengalami overfitting yang signifikan dan masih mampu melakukan generalisasi dengan baik terhadap data baru.

Tabel 8 Tabel Feature Importance

Fitur	Importance
luas_kamar	0.068457
jarak_ke_kampus	0.041765
fasilitas_per_m ²	0.036440
total_fasilitas	0.437743

Tabel 8 menampilkan empat fitur dengan nilai feature importance tertinggi yang paling berpengaruh terhadap prediksi harga kos. Berdasarkan hasil feature importance menunjukkan bahwa fitur total_fasilitas memberikan kontribusi terbesar terhadap proses prediksi harga kos dengan nilai importance sebesar 43,77%. Hal ini mengindikasikan bahwa jumlah fasilitas yang tersedia merupakan faktor dominan yang memengaruhi harga sewa kos. Sementara itu, fitur luas_kamar (6,85%), jarak_ke_kampus (4,18%), dan fasilitas_per_m2 (3,64%) juga berkontribusi dalam pembentukan model, meskipun pengaruhnya relatif lebih kecil. Hasil ini menunjukkan bahwa kelengkapan fasilitas memiliki hubungan yang lebih kuat terhadap harga kos dibandingkan faktor ukuran kamar maupun jarak ke kampus.

Implementasi Aplikasi

Model terbaik hasil pelatihan disimpan menggunakan library joblib dengan format .pkl dan diimplementasikan ke dalam aplikasi berbasis web menggunakan framework Streamlit. Aplikasi memungkinkan pengguna memasukkan data fasilitas dan karakteristik kos untuk memperoleh estimasi harga kos secara otomatis. Hasil implementasi pada aplikasi bisa dilihat pada gambar 2 dibawah ini

Gambar 5 Input Data


Prediksi Harga Kos


Informasi Utama


Status Kos

Lokasi
kecamatan jimbaran

Tipe Kos
campur

Jarak ke Kampus (meter)
500

Luas Kamar (m²)
12

☐ KM Dalam

☒ Kosongan / tanpa isi

Kos kosongan dipilih. Semua fasilitas otomatis nonaktif.


Parkiran

☒ Ada Parkiran

Jenis Parkiran
☐ Motor
☒ Mobil


Tipe Kloset

Jenis Kloset
☒ Duduk
☐ Jongkok

Gambar 5 menunjukkan halaman input data pada aplikasi prediksi harga kos. Pada halaman ini pengguna dapat memasukkan karakteristik kos, seperti lokasi, luas kamar, jarak ke kampus, serta fasilitas yang tersedia. Data yang dimasukkan pengguna akan diproses secara otomatis melalui tahapan preprocessing sebelum digunakan oleh model Decision Tree Regression untuk menghasilkan prediksi harga kos.

Gambar 6 Ringkasan Input

Sistem Prediksi Harga Kos

Aplikasi prediksi harga kos menggunakan Decision Tree Regression.

Ringkasan Input

 Lokasi: kecamatan jimbaran	 Jarak Kampus: 500 meter	 Parkiran: Mobil
 Tipe Kos: campur	 Total Fasilitas: 2	 Kosongan: Ya
 Luas Kamar: 12 m ²	 Premium Score: 2	 KM Dalam: Tidak

Gambar 6 menunjukkan halaman ringkasan data input yang berfungsi menampilkan kembali seluruh informasi yang telah dimasukkan oleh pengguna. Fitur ini membantu pengguna melakukan verifikasi data sebelum proses prediksi dilakukan sehingga dapat mengurangi kesalahan input yang berpotensi memengaruhi hasil prediksi.

Gambar 7 Hasil Prediksi Harga Kos

☒ Prediksi Berhasil

Estimasi Harga Kos

Rp 2,097,275 / bulan

Gambar 7 menunjukkan hasil prediksi harga kos yang dihasilkan oleh model Decision Tree Regression. Setelah data diproses, aplikasi menampilkan estimasi harga kos berdasarkan karakteristik dan fasilitas yang dipilih pengguna. Hasil prediksi ini dapat digunakan sebagai referensi dalam memperkirakan harga kos yang sesuai dengan kebutuhan pengguna.

Pengujian Aplikasi

Pengujian aplikasi dilakukan untuk memastikan bahwa seluruh fungsi yang tersedia pada sistem prediksi harga kos dapat berjalan sesuai dengan kebutuhan pengguna. Pengujian mencakup aspek fungsionalitas aplikasi melalui metode black box testing serta pengujian kinerja aplikasi untuk mengukur waktu respon dan penggunaan sumber daya sistem. Hasil pengujian digunakan untuk mengevaluasi apakah aplikasi mampu menjalankan proses input data, preprocessing, dan prediksi harga kos dengan baik.

Tabel 9 Pengujian Black Box

Skenario Pengujian	Hasil yang Diharapkan	Status
Input data kos	Data berhasil diterima aplikasi	Berhasil
Prediksi harga kos	Harga kos berhasil ditampilkan	Berhasil
Input Kosong	Muncul pesan validasi	Berhasil
Menampilkan hasil prediksi	Hasil prediksi tampil dengan benar	Berhasil

Berdasarkan tabel 12 yaitu hasil pengujian black box, seluruh fungsi utama aplikasi berjalan sesuai dengan kebutuhan aplikasi. Pengguna dapat memasukkan data kos, melakukan proses prediksi, serta memperoleh hasil prediksi tanpa ditemukan kesalahan fungsional selama pengujian dilakukan.

Tabel 1 Pengujian Kinerja Aplikasi

Parameter	Hasil
Waktu inferensi	0,05 detik
Ukuran model	2,3 MB
Penggunaan RAM	80 MB

Berdasarkan Hasil pengujian kinerja menunjukkan bahwa aplikasi mampu menghasilkan prediksi dengan waktu respon yang cepat serta penggunaan sumber daya yang relatif rendah. Waktu inferensi sebesar 0,05 detik menunjukkan bahwa model dapat memberikan hasil prediksi secara real-time sehingga meningkatkan kenyamanan pengguna dalam menggunakan aplikasi.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Decision Tree Regression berhasil diterapkan untuk membangun model prediksi harga kos berbasis web di wilayah Jimbaran. Model yang dikembangkan mampu memprediksi harga kos berdasarkan karakteristik fisik kamar, fasilitas yang tersedia, serta lokasi kos dengan tingkat akurasi yang cukup baik.

Hasil penelitian menunjukkan bahwa karakteristik fisik kamar, fasilitas, dan lokasi memiliki hubungan yang cukup kuat terhadap harga sewa kos. Tahapan preprocessing dan feature engineering yang diterapkan mampu membantu model dalam mengenali pola hubungan antar atribut secara lebih efektif. Berdasarkan analisis feature importance, total fasilitas merupakan faktor yang paling berpengaruh dalam menentukan harga kos, diikuti oleh luas kamar, jarak ke kampus, dan fasilitas per meter persegi.

Evaluasi model menunjukkan bahwa Decision Tree Regression menghasilkan nilai R^2 sebesar 0,8332, MAE sebesar Rp225.462, dan RMSE sebesar Rp302.509 pada data pengujian. Nilai tersebut menunjukkan bahwa model mampu menjelaskan sekitar 83,32% variasi harga kos dan memiliki tingkat kesalahan prediksi yang relatif rendah. Hasil ini mengindikasikan bahwa Decision Tree

Regression mampu memodelkan hubungan non-linear antar atribut dengan baik sehingga dapat dimanfaatkan sebagai pendukung pengambilan keputusan dalam pencarian kos di wilayah Jimbaran.

Selain itu, hasil implementasi dan pengujian aplikasi menunjukkan bahwa seluruh fungsi sistem berjalan sesuai dengan kebutuhan pengguna, mulai dari proses input data, preprocessing, hingga penyajian hasil prediksi. Dengan demikian, aplikasi yang dikembangkan dapat digunakan sebagai alat bantu untuk memperoleh estimasi harga kos secara lebih cepat, mudah, dan berbasis data.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan penelitian selanjutnya. Pertama, jumlah dataset dapat ditingkatkan agar model memiliki lebih banyak data untuk dipelajari sehingga akurasi prediksi dapat ditingkatkan. Kedua, penelitian selanjutnya dapat membandingkan performa algoritma Decision Tree Regression dengan algoritma lain seperti Random Forest Regression, Gradient Boosting, atau XGBoost untuk memperoleh model yang lebih optimal.

Dari sisi pengembangan sistem, aplikasi dapat ditingkatkan dengan menambahkan fitur penyimpanan data pengguna, riwayat prediksi, dan integrasi database sehingga pengelolaan data menjadi lebih baik. Selain itu, aplikasi juga dapat dikembangkan dalam bentuk aplikasi mobile atau sistem berbasis cloud agar lebih mudah diakses oleh pengguna di berbagai perangkat.

DAFTAR PUSTAKA

- Ağbulut, Ü., Gürel, A. E., & Biçen, Y. (2021). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable and Sustainable Energy Reviews*, 135. <https://doi.org/10.1016/j.rser.2020.110114>
- Anggara, E. D., Widjaja, A., & Suteja, B. R. (2022). Prediksi Kinerja Pegawai sebagai Rekomendasi Kenaikan Golongan dengan Metode Decision Tree dan Regresi Logistik. *Jurnal Teknik Informatika Dan Sistem Informasi*, 8(1). <https://doi.org/10.28932/jutisi.v8i1.4479>
- Bichri, H., Chergui, A., & Hain, M. (2024). Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets. In *IJACSA International Journal of Advanced Computer Science and Applications* (Vol. 15, Number 2). www.ijacsa.thesai.org
- Hasanah, M. (2023). Prediksi Harga Indekost Menggunakan Analisis Regresi Linier Berganda.
- Hidayah, F., Angesti, S. J., & Widyastuti, Y. P. (2024). Prediksi Harga Rumah di Boston Menggunakan Metode Linear Regression, SVR, Decision Tree dan Random Forest Regression.
- Jumin, E., Basaruddin, F. B., Yusoff, Y. B. M., Latif, S. D., & Ahmed, A. N. (2021). Solar radiation prediction using boosted decision tree regression model: A case study in Malaysia. *Environmental Science and Pollution Research*, 28(21), 26571–26583. <https://doi.org/10.1007/s11356-021-12435-6>
- Karim, R., Alam, M. K., & Hossain, M. R. (2021, August 10). Stock market analysis using linear regression and decision tree regression. 2021 1st International Conference on Emerging Smart Technologies and Applications, ESmarTA 2021. <https://doi.org/10.1109/eSmarTA52612.2021.9515762>
- Labib Mu'tashim, M., Damayanti, S. A., Zaki, H. N., Muhayat, T., & Wirawan, R. (2021). Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression. 3.
- Rahmadini, R., L. E. E., P. A., R. Y., & M. T. (2023). PENERAPAN DATA MINING UNTUK MEMPREDIKSI HARGA BAHAN PANGAN DI INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (Vol. 4, Number 4).

- Muharrom, M. (2023). Analisis Penggunaan Orange Data Mining untuk Prediksi Harga USDT/BIDR Binance. *Bulletin of Information Technology (BIT)*, 4(2), 178–184. <https://doi.org/10.47065/bit.v3i1>
- Muttaqin, Z., & Srihartini, E. (2022). PENERAPAN METODE REGRESI LINIER SEDERHANA UNTUK PREDIKSI PERSEDIAAN OBAT JENIS TABLET. *Sistem Informasi* |, 9(1), 12–16.
- Narayana Raju, K., Kumar, P., Satyanarayana, C. S. V., & Pujitha, D. (2024). A Study on “Machine Learning Regression.” In *Journal of Computational Analysis and Application* (Vol. 33, Number 2).
- Nurani, A. T., Setiawan, A., & Susanto, B. (2023). Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma. *Jurnal Sains Dan Edukasi Sains*, 6(1), 34–43. <https://doi.org/10.24246/juses.v6i1p34-43>
- Nuruzzaman, M. T., Yogyakarta, K., & Bagian, E. (2023). Penerapan Data Mining dengan Metode Regresi Linier untuk Memprediksi Data Nilai Hasil Ujian Menggunakan Rapid Miner.
- Ramadhan, A. I., Almajid, N., & Ginting, D. Y. (2023). Penerapan Decision Tree Regression dalam Memprediksi Harga Rumah di Provinsi Jawa Barat. In *Jejaring Penelitian dan Pengabdian Masyarakat (JPPM)* (Vol. 1, Number 3).
- Rekayasa, K. K., Athiyah, U., Hananta, A., Maulidi, T., Meylana, V., Putra, E., Felix, T., Purba, H., Angeline, E., & Bakowatun, W. (2021). Terbit online pada laman web jurnal: <http://journal.ittelkom-pwt.ac.id/index.php/dinda> JOURNAL OF DINDA Sistem Pendukung Keputusan Prediksi Harga Rumah Kost untuk Mahasiswa IT Telkom Purwokerto Menggunakan Metode Fuzzy Tsukamoto Berbasis Web. In *Data Institut Teknologi Telkom Purwokerto* (Vol. 1, Number 2). <http://journal.ittelkom-pwt.ac.id/index.php/dinda>
- Salazar, J. J., Garland, L., Ochoa, J., & Pyrcz, M. J. (2022). Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy. *Journal of Petroleum Science and Engineering*, 209, 109885. <https://doi.org/10.1016/j.petrol.2021.109885>
- Saputra, R. A., & Pratama, A. (2025). IMPLEMENTASI DECISION TREE UNTUK PREDIKSI HARGA RUMAH DI DAERAH TEBET. In *Journal of Information System Management (JOISM)* e-ISSN (Vol. 6, Number 2).
- Vinita Yadav, K., & Kumar, R. (2025). DATA PREPROCESSING TECHNIQUES. *Phoenix: International Multidisciplinary Research Journal*, 1.
- Wardana, F. A., & Juanita, S. (2025). PREDIKSI JUMLAH TENAGA KERJA ASING DI JAWA BARAT MENGGUNAKAN PERBANDINGAN ALGORITMA SUPPORT VECTOR REGRESSION DAN DECISION TREE REGRESSION. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(2), 890–899. <https://doi.org/10.29100/jupi.v10i2.6003>
- Wijaya, Y. F., & Triayudi, A. (2023). Penerapan Data Mining Pada Prediksi Harga Emas dengan Menggunakan Algoritma Regresi Linear Berganda dan ARIMA. *Journal of Computer System and Informatics (JoSYC)*, 5(1), 73–81. <https://doi.org/10.47065/josyc.v5i1.4615>
- Zahra, R. A. (2023). ABSTRACT APPLICATION OF RANDOM FOREST ALGORITHM WITH HYPERPARAMETER TUNING TO PREDICT BOARDING HOUSE RENTAL PRICE IN BANDAR LAMPUNG CITY.