

**PREDIKSI HARGA PERUMAHAN MENGGUNAKAN LGBM
REGRESSOR DAN XGB REGRESSOR**

RISET INFORMATIKA



OLEH :

MIA DUWI UMIATI (20081010063)

DOSEN PENGAMPU :

DR. BASUKI RAHMAT, S.SI, MT.

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2023

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era teknologi dan informasi, analisa dan prediksi harga rumah menggunakan metode kecerdasan buatan atau artificial intelligence telah menjadi fokus penelitian yang semakin relevan. Harga rumah bukan hanya mencerminkan nilai properti, tetapi juga menjadi indikator penting bagi pasar properti suatu daerah. Penelitian ini untuk menganalisa dan memprediksi harga rumah di masa depan dengan mengambil pendekatan inovatif dengan memanfaatkan berbagai teknik dan model machine learning, yaitu LGBM Regressor dan XGB Regressor. Data yang digunakan dalam penelitian ini mencakup berbagai variabel, seperti luas bangunan, jumlah kamar mandi dan kamar tidur, lingkungan, serta tahun pembangunan. Beberapa variabel tersebut secara komprehensif mencerminkan beragam faktor yang dapat mempengaruhi harga rumah. Dengan demikian, penelitian ini bertujuan untuk tidak hanya memprediksi tren harga, tetapi juga memberikan wawasan mendalam mengenai faktor-faktor yang mempengaruhi pasar properti, merambah ke dimensi kecerdasan buatan sebagai alat analisis yang canggih dan inovatif dalam ranah real estate.

1.2 Rumusan Masalah

Dari latar belakang di atas dapat dirumuskan masalah yaitu sebagai berikut.

1. Bagaimana teknik dan model machine learning, khususnya LGBM Regressor dan XGB Regressor, dapat diterapkan untuk menganalisis dan memprediksi harga rumah?
2. Apa saja variabel-variabel yang paling signifikan dalam mempengaruhi harga rumah, dan bagaimana variabel-variabel ini dapat dioptimalkan menggunakan metode kecerdasan buatan?
3. Sejauh mana keakuratan dan keandalan model LGBM Regressor dan XGB Regressor dalam melakukan prediksi harga rumah di masa depan?

1.3 Batasan Masalah

Dalam menjaga fokus terhadap relevansi, penelitian ini membatasi beberapa masalah yaitu sebagai berikut.

1. Variabel yang digunakan terbatas pada luas bangunan, jumlah kamar mandi dan kamar tidur, lingkungan, serta tahun pembangunan rumah.
2. Analisis hanya difokuskan pada prediksi harga rumah dan tidak mempertimbangkan faktor eksternal yang mungkin mempengaruhi pasar properti.

1.4 Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan yaitu sebagai berikut.

1. Mengembangkan pendekatan inovatif dalam analisis dan prediksi harga rumah dengan menggunakan teknik dan model machine learning.
2. Mengidentifikasi variabel-variabel yang paling berpengaruh dalam menentukan harga rumah.
3. Mengevaluasi dan membandingkan kinerja model LGBM Regressor dan XGB Regressor dalam konteks prediksi harga rumah.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat yaitu sebagai berikut.

1. Bagi pengembang properti, untuk mendukung pengambilan keputusan yang lebih tepat dalam menentukan harga jual properti.
2. Bagi konsumen, sebagai panduan yang lebih akurat dalam melakukan pembelian atau investasi properti.
3. Sebagai sumbangan terhadap literatur penelitian mengenai analisis dan prediksi harga rumah menggunakan pendekatan kecerdasan buatan.

BAB II

KAJIAN PUSTAKA

2.1 Rumah

Rumah memiliki peran penting bagi kehidupan manusia, tidak hanya sebagai tempat berlindung, tetapi juga sebagai aset dan investasi yang memiliki nilai ekonomi yang signifikan. Harga rumah seringkali melibatkan analisis faktor-faktor seperti lokasi, ukuran, fasilitas, dan kondisi ekonomi. Rumah juga mencerminkan gaya hidup, kebutuhan keluarga, dan merupakan investasi jangka panjang. Dalam konteks penelitian ini, pemahaman mendalam tentang variabel-variabel rumah menjadi krusial untuk memprediksi harga dengan akurat.

2.2 Machine Learning

Machine learning merupakan sub-bidang dari kecerdasan buatan yang berfokus pada pengembangan algoritma dan model komputer yang belajar dari data yang ada. Machine learning memungkinkan komputer untuk mengidentifikasi pola, membuat suatu prediksi, dan mengambil keputusan tanpa perlu pemrograman yang kompleks.

2.3 LGBM Regressor

Light Gradient Boosting Machine Regressor (LGBM Regressor) merupakan algoritma machine learning yang menggunakan pendekatan penguatan berbasis tree untuk meningkatkan akurasi prediksi. LGBM Regressor memiliki keunggulan dalam kinerja dan kecepatan karena mampu mengatasi masalah seperti overfitting dan penanganan data yang besar. Keunggulan tersebut membuat metode ini menjadi populer dalam regresi yang dimana memiliki tujuan utama yaitu memprediksi nilai numerik.

2.4 XGB Regressor

Extreme Gradient Boosting Regressor (XGB Regressor) merupakan algoritma machine learning yang digunakan untuk tugas regresi dan klasifikasi. XGB Regressor memperkuat model dengan menggabungkan prediksi dari sejumlah model lemah (weak learners). XGB Regressor memiliki keunggulan dalam keandalan prediksi, kecepatan komputasi yang tinggi, dan kemampuan menangani data yang kompleks.

2.5 Catboost Regressor

Catboost Regressor merupakan algoritma machine learning yang dirancang untuk tugas regresi. Catboost menggunakan pendekatan gradient boosting dan memiliki keunggulan dalam menangani data kategorikal tanpa perlu melakukan preprocessing tambahan seperti encoding.

2.6 MedAE

Median Absolute Error (MedAE) merupakan salah satu metrik evaluasi kinerja model regresi. MedAE mengukur kesalahan absolut median antara nilai prediksi dan nilai sebenarnya. MedAE memiliki kelebihan yaitu ketahanannya terhadap outliers karena menggunakan median sebagai nilai tengahnya yang menjadi lebih stabil daripada MAE dalam beberapa kasus.

2.7 R2_Score

R-Squared (R2_Score) merupakan salah satu metrik evaluasi kinerja model regresi. R2_Score mengukur sejauh mana variabilitas dalam variabel dapat dijelaskan oleh model. R2_Score memiliki score antara 0 dan 1 yang dimana 0 menunjukkan bahwa model tidak memberikan kontribusi apapun dalam menjelaskan variasi. Sedangkan 1 menunjukkan bahwa model dapat menjelaskan seluruh variasi.

BAB III

METODE PENELITIAN

3.1 Analisis Data Eksploratif

Analisis data eksploratif dilakukan dengan tujuan untuk mengecek kualitas dan integritas data yang akan digunakan dalam penelitian. Selain itu, analisis data eksplorasi juga dilakukan untuk memperbaiki data yang mungkin memiliki missing value atau outlier, sehingga data dapat digunakan secara optimal untuk proses selanjutnya.

3.2 Seleksi Fitur

Setelah melakukan analisis data eksploratif, langkah selanjutnya yaitu seleksi fitur yang dilakukan dengan menggunakan beberapa model yaitu LGBM Regressor, XGB Regressor, dan Catboost Regressor. Dengan memanfaatkan model-model tersebut, maka dapat mengidentifikasi variabel-variabel yang dapat mempengaruhi harga rumah. Proses seleksi fitur ini menjadi kunci dalam meningkatkan akurasi prediksi pada tahap selanjutnya.

3.3 Modeling

Modeling merupakan tahapan penting dalam penelitian ini karena akan dilakukan evaluasi untuk membandingkan performa model-model yang digunakan untuk memprediksi harga rumah. Modeling digunakan dengan tujuan untuk menentukan model terbaik yang akan digunakan dalam pengujian prediksi.

3.4 Pengujian Prediksi

Pengujian prediksi dilakukan dengan menggunakan model terbaik yang telah dipilih dari beberapa langkah-langkah yang telah dilakukan sebelumnya. Prediksi harga rumah untuk masa depan dievaluasi dengan memanfaatkan metrik pengujian yang sesuai. Beberapa metrik pengujian yang digunakan yaitu MedAE dan R2_Score. Pengujian tersebut dilakukan untuk menguji prediksi dan membandingkan metrik pengujian terbaik yang digunakan.