

**PREDIKSI HARGA RUMAH TINGGAL DI DEPOK, BANDUNG, DAN
JAKARTA SELATAN MENGGUNAKAN METODE *ENSEMBLE
STACKING REGRESSOR***

Oleh:

Gede Aby Sanjiwa Kurniawan

Jurusan Teknik Informatika

Program Studi Sistem Informasi

ABSTRAK

Dinamika pasar properti di wilayah perkotaan Indonesia, terutama di Depok, Bandung, dan Jakarta Selatan, mencerminkan kompleksitas tinggi akibat pengaruh berbagai faktor yang saling berinteraksi, sehingga dibutuhkan model prediksi yang mampu menangkap pola tersebut secara akurat. Penelitian ini bertujuan mengukur perbandingan performa *Stacking Regressor* terhadap model tunggal *Linear Regression*, *Random Forest*, dan *XGBoost* dalam memprediksi harga rumah tinggal di ketiga wilayah tersebut, serta menghasilkan rekomendasi estimasi harga sebagai acuan dalam menentukan harga jual atau memilih properti. Data bersumber dari tiga *dataset* publik di platform *Kaggle* yang digabungkan menjadi satu *dataset* terpadu dengan total 32.412 data, yang setelah proses *data cleaning* menghasilkan 25.518 data bersih untuk digunakan pada tahap pemodelan. Tahapan penelitian meliputi *data cleaning*, *feature engineering*, *exploratory data analysis*, dan *feature selection* menggunakan filter method dengan threshold korelasi Pearson lebih besar sama dengan 0,50, serta preprocessing komprehensif untuk mencegah data leakage. Pelatihan model memanfaatkan *GridSearchCV* dengan *5-Fold Cross-validation* untuk optimasi *hyperparameter*, dan mekanisme *Out-of-Fold* untuk pembentukan meta-features sebagai input *Ridge meta learner*. Berdasarkan evaluasi pada data test 30%, *Stacking Ensemble* menghasilkan performa tertinggi dengan $R^2 = 0,7690$, $RMSE = \text{Rp}694.774.040$, dan $MAE = \text{Rp}356.824.486$. Uji Wilcoxon Signed-Rank

Test membuktikan bahwa keunggulan *Stacking Ensemble* terhadap *Linear Regression* dan *XGBoost* bersifat signifikan secara statistik, sementara perbedaannya terhadap *Random Forest* tidak signifikan. Stabilitas model dikonfirmasi melalui *Repeated K-Fold Cross-validation* dengan 50 evaluasi independen yang menghasilkan R^2 rata-rata 0,7622 (SD = 0,0181). Model diintegrasikan ke dalam antarmuka berbasis Streamlit untuk mendukung estimasi harga properti secara interaktif.

Kata Kunci: Prediksi Harga Rumah; *Ensemble Stacking Regressor*; *Random Forest*; *XGBoost*; Regresi Linear



HOUSING PRICE PREDICTION IN DEPOK, BANDUNG, AND SOUTH JAKARTA USING AN ENSEMBLE STACKING REGRESSOR

By:

Gede Aby Sanjiwa Kurniawan

Informatic Engineering

Information System Study Program

ABSTRACT

Property market dynamics in urban Indonesia, particularly in Depok, Bandung, and South Jakarta, reflect high complexity due to various interacting factors, requiring a predictive model capable of accurately capturing these patterns. This research measures the comparative performance of a Stacking Regressor against single models, namely Linear Regression, Random Forest, and XGBoost, in predicting residential house prices across the three regions, and produces price estimation recommendations as a reference for determining selling prices or selecting properties. Data were sourced from three public Kaggle datasets, combined into a unified dataset totaling 32,412 records, which after data cleaning resulted in 25,518 clean records used for modeling. The research stages include data cleaning, feature engineering, exploratory data analysis, and feature selection using the filter method with a Pearson correlation threshold of ≥ 0.50 , along with comprehensive preprocessing to prevent data leakage. Model training utilized GridSearchCV with 5-Fold Cross-validation for hyperparameter optimization, and an Out-of-Fold mechanism to construct meta-features as input for the Ridge meta learner. Based on evaluation using a 30% test set, the Stacking Ensemble achieved the highest performance with $R^2 = 0.7690$, $RMSE = \text{IDR } 694,774,040$, and $MAE = \text{IDR } 356,824,486$. The Wilcoxon Signed-Rank Test confirmed that the Stacking Ensemble's superiority over Linear Regression and XGBoost was statistically significant, while the difference compared to Random Forest was not. Model stability was confirmed through Repeated K-Fold Cross-validation with 50

independent evaluations, yielding a mean R^2 of 0.7622 ($SD = 0.0181$). The model was integrated into a Streamlit-based interface to support interactive property price estimation.

Keywords: *House Price Prediction; Ensemble Stacking Regressor; Random Forest; XGBoost; Linear Regression*

