PROJECT REPORT

COMPARISON OF LINEAR REGRESSION, RIDGE REGRESSION, AND XGBOOST ALGORITHMS IN HOUSE PRICE PREDICTION IN SOUTH JAKARTA

PRILLY PUTRI MAHARANI
22.K1.0040

Faculty of Computer Science
Soegijapranata Catholic University
2026

ABSTRACT

This study aims to predict house prices in South Jakarta, which exhibit high variability and are influenced by various factors. Linear regression-based approaches are often unable to fully capture the complexity of relationships within the data. Therefore, this study compares three algorithms – Linear Regression, Ridge Regression, and XGBoost Regression – to determine the most optimal model for accurate price prediction. The dataset consists of housing data in South Jakarta that has undergone preprocessing, feature engineering, and normalization. The features used include land area, building area, and derived variables such as price per square meter and room density. The results show that Linear Regression achieved an R^2 value of 0.8662 and Ridge Regression achieved 0.8658, while XGBoost Regression produced the best performance with an R^2 value of 0.9545 and lower error values. This indicates that XGBoost is more effective in capturing non-linear relationships and data complexity. Therefore, XGBoost Regression is the most optimal model for predicting house prices in South Jakarta.

Keywords: house price prediction, regression, XGBoost, machine learning, South Jakarta