

Training Multiple Models: Case Study

Download the dataset below to solve this Data Science case study on training multiple models.

In predictive modeling, the selection of an appropriate machine learning algorithm is crucial for achieving high accuracy and generalizability. However, given the complexity and unique characteristics of various datasets, no single model universally outperforms others across all tasks.

The provided dataset is based on real estate prices. The primary objective of this problem is to conduct a comprehensive analysis and comparison of multiple machine learning models to predict real estate prices. By evaluating a diverse set of algorithms, you need to identify the model that provides the highest accuracy and reliability in forecasting property prices based on features such as house age, distance to the nearest MRT station, number of convenience stores, geographic coordinates, and transaction date.

Your key tasks include:

- Train a variety of machine learning models, including Linear Regression, Decision Tree, Random Forest, and Gradient Boosting, to predict the real estate prices.
- Evaluate each model's performance using metrics such as Mean Absolute Error (MAE) and R-squared (R^2) to assess prediction accuracy and the variance explained by the model.
- Identify the best-performing models based on the evaluation metrics.