

Residential Real Estate Price Prediction – Model Selection

Real Estate Investment Trust • Investment Analytics Project

► ISSUE / PROBLEM

- ❖ Selecting a pricing model that balances **predictive accuracy and interpretability** for residential real estate valuation.
- ❖ Linear models offer transparency but may fail to capture non-linear price dynamics, while complex models require clear business justification.

► RESPONSE

- ❖ Multiple regression models were evaluated, including **regularized linear, robust, and tree-based** approaches.
- ❖ Model performance was compared using **RMSE** and **R²** on a held-out test set.
- ❖ Feature sets were tailored to each model's assumptions, and zipcode was encoded categorically to capture neighbourhood-level effects.

► IMPACT

This evaluation enabled the selection of:

- ❖ A **high-accuracy predictive model** suitable for operational pricing.
- ❖ An **interpretable benchmark model** to support business intelligence, transparency, and stakeholder communication.

MODEL PERFORMANCE SUMMARY

Model	RMSE	R ²
XGBoost	0.1580	0.9117
Random Forest	0.1698	0.8980
Ridge Regression	0.1738	0.8926
Huber Regression	0.1740	0.8924
ElasticNet	0.1810	0.8835

FINAL MODEL DECISION

Primary Predictive Model

XGBoost

- ❖ Lowest RMSE and highest R²
- ❖ Captures complex, non-linear price drivers
- ❖ Suitable for operational pricing and forecasting

Interpretable Benchmark

Ridge Regression

- ❖ Stable coefficients
- ❖ Clear economic interpretation
- ❖ Supports BI reporting and stakeholder communication

Tree-based models outperform linear approaches in predictive accuracy, while regularized regression supports interpretability.

Next Step (Stage 6):

Confirm model assumptions and perform diagnostic checks to validate residual behavior, stability, and generalization performance.

► KEY INSIGHTS

- ❖ Non-linear relationships between location, size, and quality significantly influence housing prices.
- ❖ Proper encoding of categorical location features substantially improves linear model performance.
- ❖ Retaining extreme but valid observations preserves real-market behavior without degrading model accuracy.
- ❖ Using complementary models enables both accurate prediction and transparent business interpretation.