🏠🤖

Real Estate Price Prediction

Machine Learning Comparison Study

UPM UNIVERSITI PUTRA MALAYSIA BERILMU BERBAKTI PUTRA PERTANIAN UNTUK RAKYAT

Our Mission

The Project

Compare two global real estate markets using the same ML algorithms. Compare that markets follow different rules and require market-specific models.

The Goal

Build ML models for real-world price prediction and demonstrate why a universal model cannot work across different markets.

Markets: Malaysia vs USA
Samples: 4,930 total house transactions
Algorithms: Lasso, Random Forest, XGBoost

❓ The Problem We Solved

- Real estate markets are different worldwide
- Investors need market-specific predictions
- Question: Do housing price patterns TRANSFER across countries?
- Answer: Let's build models and find out!

⚡ **Core Hypothesis: Markets have different price drivers. A single model may not generalize**

Research Strategy

Questions We Asked

- What predicts prices best?
- Are the predictors the same?
- How accurate can we get?

What We Searched For

- Key price drivers
- Market differences
- Model performance gaps

Methodology: Separate ML pipelines → train on market data → compare results

Same preprocessing pipeline and evaluation metrics applied to both markets; models tuned separately per dataset.

The Datasets

Ames Housing

Samples:	2,930 houses
Price Range:	$12.9K - $755K
Average:	$180,921
Features:	80+ attributes
Time Period:	2006-2010

Malaysia Market

Samples:	2,000 properties
Price Range:	MYR 27K - 11.4M
Average:	MYR 490,685
Features:	Location-based
Time Period:	2025 Market

Three Algorithms, One Goal

Lasso

Linear regression with regularization. Simple. Interpretable. Baseline model.

Random Forest

Ensemble of decision trees. Handles non-linearity. Captures feature importance.

⭐ XGBoost

Gradient boosting. Most powerful. Best accuracy. Best-performing model

Why test three? Because the best algorithm may be different for each market!

The Technical Takeaway

Same algorithm. Completely different results. Conclusion: Market-specific models required

The Results That Matter

Ames (USA) - XGBoost Winner

R² Score

0.9067

~91% variance explained!

Prediction Error

±7.11%

Very accurate ✓

Malaysia - XGBoost Winner

R² Score

0.4271

~43% variance explained

Prediction Error

±23.33%

Moderate accuracy

Key Findings:

- Results differ strongly across markets, largely due to data granularity
- Malaysia performance is constrained by sparse features

🔍 What Drives Prices? (XGBoost Feature Importance)

USA: Quality Rules

Top Driver (32.8%):
Overall Qual - House quality
2nd Driver (9.8%):
Total Area - Area of the housing
Insight:
Americans pay for craftsmanship and practicality!

Malaysia: Location Rules

Top Driver (8.4%):
Type Flat - Flat as a type of housing
2nd Driver (3.2%):
State Selangor - Proximity to Selangor
Insight:
Malay pay for type of housing and location access!

🎯 Our Interactive App

📊 5 Interactive Tabs

- Dashboard for USA market
- Dashboard for Malaysia market
- Key insights for USA market
- Key insights for Malaysia market
- Cross-market comparison

🧮 What Can You Calculate?

USA: Quality + Area + Garage → Price

Malaysia: Price/Sq Ft + Location → Price

Error ranges included!

✨ **Real-time predictions with accuracy margins shown**

⚡ Limitations

- Data Granularity differs substantially across markets
- Aggregated features limit predictive precision
- Model performance is bounded by available information
- Results are markes-specific and not universally transferable

The Summary

91%
Ames Accuracy
(Highly Predictable)

43%
Malaysia Accuracy
(Context-Dependent)

Feature-detail gap
Market Differences
(Fundamental)

🌍 Market-specific models are more reliable
🏠 Market data structure changes what the model can learn
🤖 Data-Driven Decisions Work Best