Hedonic Forecast: ML Housing Prediction in Japan

CSE - 6242 FINAL PROJECT - TEAM 090

Jianhuang Li, Shin Ying Chua, Wei Qi Thong, Ryan Kai Yan Seet

Georgia Institute of Technology

Problem Definition

Japan's official price indices (JRPPI) lag by three months and report only broad regional aggregates, making them insufficient for timely, local decision-making. Investors, developers, and urban planners increasingly need ward-level and neighborhood-scale forecasts to identify emerging opportunities. However, current approaches either ignore spatial patterns (time-series models) or temporal dynamics (spatial ML), leaving no unified and interpretable solution. Besides, analysts require transparent, explainable predictions they can justify to investors and policymakers.

ML Model

Model Choices

- Linear Regression – fast, interpretable baseline for price momentum and structural signals.
- Random Forest (300 trees) – captures nonlinear patterns; strong performance on heterogeneous mesh data.
- LightGBM (600 trees) – handles sparse features and missingness; best accuracy–speed balance.
- LSTM (8-quarter window) – learns temporal patterns; reduced feature set (7 vars) for stability.

Data Collection & Overview

Source: Ministry of Land, Infrastructure, Transport and Tourism (MLIT) Real Estate Information Library

- Access Method: Retrieval via the MLIT public API
- Type: Residential transaction records (price, area, building year, coordinates)
- Fields Provided: Price, area, building year, latitude/longitude, transaction date

Coverage & Span:

- Geographic Scope: Tokyo's 23 wards + Sendai's main wards
- Availability: 2005-Q3 to 2025-Q1 (79 quarters; quarterly releases)

Scale & Structure:

- 485,093 transactions (~150MB)
- 3,000+ unique 250m × 250m JIS X 0410 mesh cells
- Ward panel: ~2,100 ward×quarter rows
- Mesh panel: ~37,900 mesh×quarter rows

Key Attributes:

Price per sqm, floor area, building age, lat/long, quarter/time trend, ward/mesh aggregates.

Key Innovations

- Direct 250m mesh encoding (no geocoding APIs)
- Hierarchical ward→mesh features to stabilize sparse grids
- Cold-start smoothing (Moving avg + volatility).
- SHAP explanations for all models, including LSTM
- Tokyo–Sendai metro vs. regional comparison

Comparison to Prior Work

Method	Resolution	Frequency	Our Contribution
JRPPI (official)	City	Monthly +3mo lag	250m, quarterly, no lag
Yoshida 2024	Building	Cross-sectional	Time-series forecasting
Chen 2017 LSTM	City	Monthly	Finer spatial granularity

Evaluation

Below are the 5 evaluation questions:

1) Feature importance: Which features drive prices?

2) Accuracy across scales: How accurate at ward vs mesh scales?

3) Granularity trade-off: insight vs noise

4) Efficiency and scaling: How does accuracy scale with more data?

5) Efficiency and scaling.: Which model balances accuracy and latency?

Key findings

1. Momentum drives forecasts: 1-quarter lags and 4-quarter moving averages dominate SHAP importance; structural features matter far less.
2. Performance varies by resolution: Linear Regression is best for wards, while Random Forest is clearly strongest for meshes.
3. Granularity trade-off: Mesh forecasts reveal neighbourhood hotspots but come with heavier error tails.
4. Efficiency: Linear Regression offers sub-millisecond inference; Random Forest gives the best accuracy–time balance once data coverage >50%.
5. Model accuracy: Linear Regression is strongest on wards and offers sub-millisecond latency; meshes require nonlinear models and Random Forest leads, LightGBM follows, and the LSTM struggles to generalise beyond momentum despite higher inference cost.

Methodology

Hedonic Price Index

We estimate municipality-level indices using two-way fixed effects (PanelOLS):

t

$$\ln(P_{it}) = \beta_0 + \beta_1 \ln(\text{Area}_{it}) + \beta_2 \text{Age}_{it} + \beta_3 \text{AgeUnknown}_{it} + \alpha_i + \gamma_t + \varepsilon_{it}$$

Forecasting Workflow

1. Build quarterly ward and mesh indices from the hedonic regression.
2. Derive engineered features: lags, QoQ/YoY growth, moving averages, and volatility.
3. Train forecasting models on 2005–2019, with 2020–2021 as the validation period.
4. Generate one-quarter-ahead forecasts for the 2022–2025 test period.

Experiments & Results

Experimental Design

Split	Period	Ward	Mesh
Train	2005-Q3 to 2019-Q4	1,507	26,818
Validation	2020-Q1 to 2021-Q4	336	6,337
Test	**2022-Q1 to 2025-Q1**	**252**	**4,762**

Feature Engineering

- Ward: 15 features (momentum, liquidity, structure, time)
- Mesh: 13+ features (adds hedonic indices + missingness flags)
- LSTM: 7 features (reduced set for sequence modeling)

Evaluation Metrics

- MAE (JPY/m²) — Mean absolute error
- RMSE (JPY/m²) — Penalizes outliers
- R^2 — Coefficient of determination
- SHAP values computed on validation set (500-1000 samples)

Model Configuration

- Random Forest: 300 trees | LightGBM: 600 boosting rounds
- LSTM: 8-quarter window, 96 hidden units, dropout 0.3
- Early stopping: patience = 8 epochs on validation loss

Key Findings

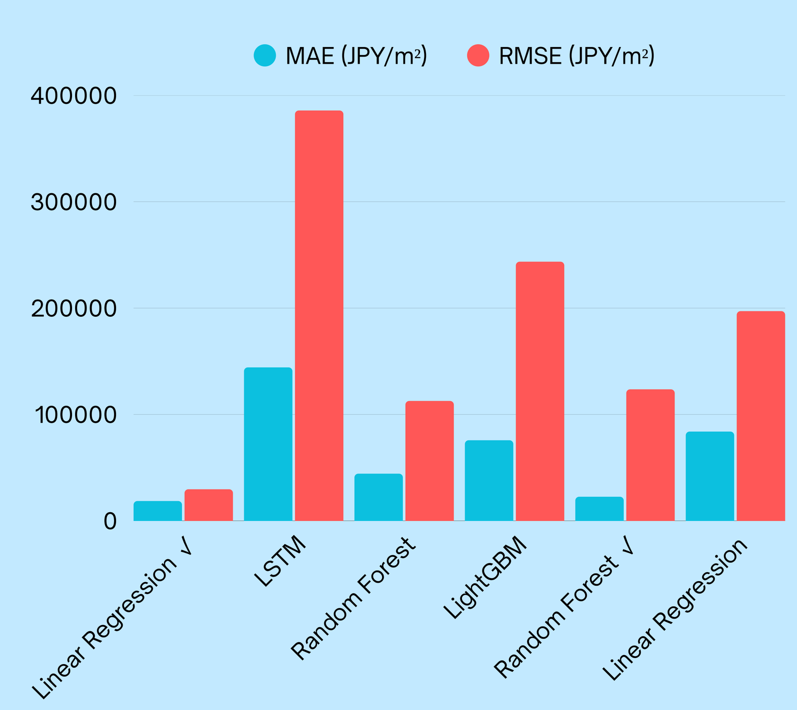
Ward: Linear Regression achieves MAE 18.6k, R^2=0.995 (momentum alone explains 99.5% variance)

Mesh: Random Forest cuts error 73% vs. linear (22.6k MAE, R^2=0.95)—nonlinear interactions matter at fine resolution

SHAP: 4Q moving average dominates (60-70% weight), structure features activate only when data is sparse

Speed: Linear <2ms | RF 40-80ms | LightGBM ~5ms (unstable on small samples)

Tokyo vs. Sendai: Dense markets favor all models; sparse markets perform well with random forest regularization

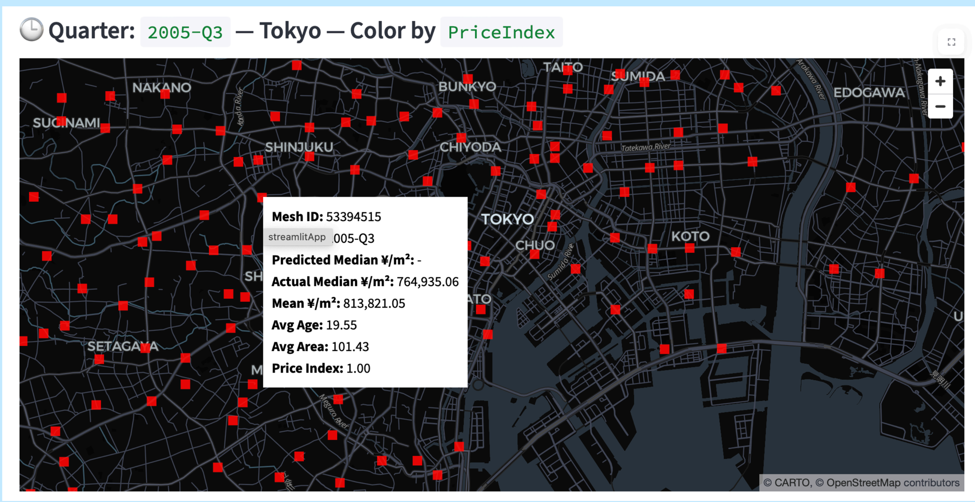


Level	Model	R^2	Inference Time
Ward	Linear Regression ✓	0.995	<2ms
Ward	LSTM	0.98	—
Ward	Random Forest	0.933	~80ms
Ward	LightGBM	0.928	~5ms
Mesh (250m)	Random Forest ✓	0.951	~140ms
Mesh (250m)	LightGBM	0.948	~47ms
Mesh (250m)	Linear Regression	0.876	~3ms
Mesh (250m)	LSTM	0.664	—

Interactive Dashboard

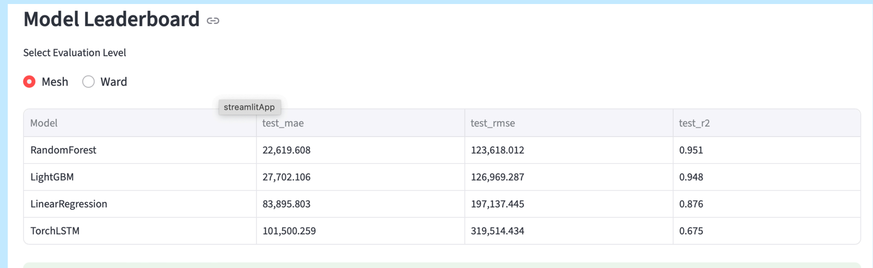
Feature 1:

Interactive 250 m mesh map with city/model/quarter controls and hover tooltips showing predicted vs actual prices, building age, floor area, and hedonic index. Supports quarter-by-quarter animation.



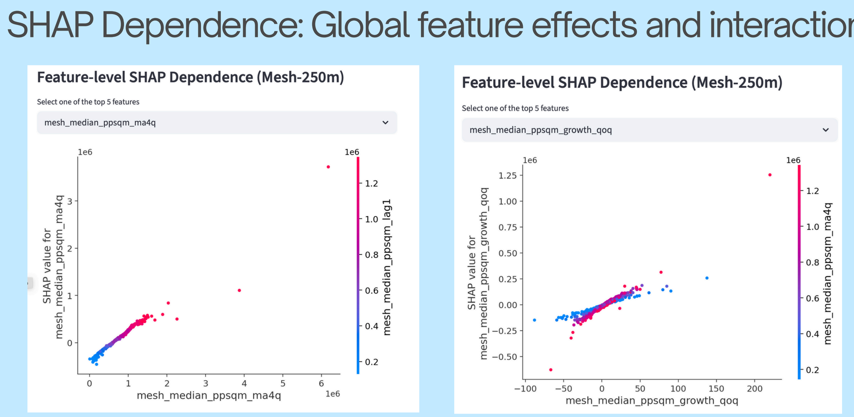
Feature 2:

Model leaderboard with MAE, RMSE, and R^2, highlighting the best model for Mesh/Ward.



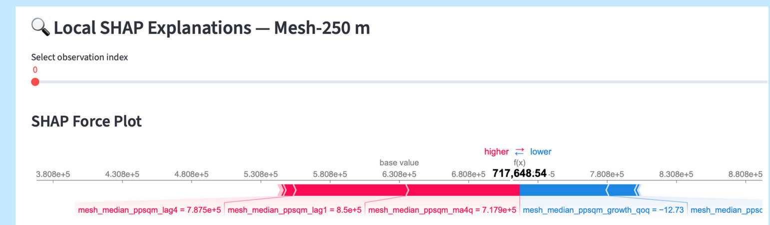
Feature 3:

SHAP Dependence: Global feature effects and interactions.



Feature 4.1 :

SHAP Dependence: Global feature effects and interactions.

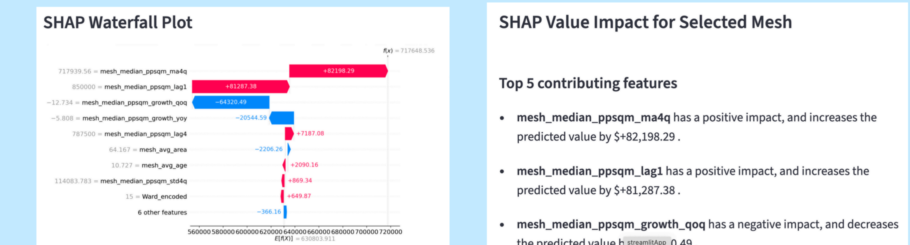


Key insights (1)

- Strong long-term momentum pushes the prediction up (lag-1, lag-4, MA-4q).
- Short-term QoQ growth pushes the prediction down slightly.
- Positive historical effects outweigh the negative short-term signal.
- Result: the mesh's past price strength leads to a higher-than-baseline prediction.

Feature 4.2 :

Local SHAP: Mesh-level explanations via force, waterfall, and impact plots.



Key insights (2)

- Long-term momentum is the main driver; The 4-quarter moving average and lag-1 price strongly increase the predicted value (+82k, +81k).
- Short-term weakness reduces the prediction; QoQ and YoY growth contribute large negative impacts (–64k, –20k).

Scan QR to launch app

Key Reference

[1] MLIT (2005-present). Real Estate Information Library. https://www.reinfolib.mlit.go.jp/
[2] MLIT (2020). Methodology of JRPPI: Japan Residential Property Price Index.
[3] Yoshida & Seya (2024). Spatial prediction of apartment rent. J. Real Estate Finance & Economics.
[4] Chen et al. (2017). House price prediction using LSTM. arXiv:1709.08432.
[5] Lundberg & Lee (2017). A unified approach to interpreting model predictions. NIPS.