International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

A Hybrid Spatio-Temporal Decision Tree Architecture for Micro-Localized Land Valuation: A Case Study in Eastern Side of Godhra Town

Ansh D. Shah[1], Krishna D. Sheth[2], Harishkumar S. Sheth[3], Amit K. Yadav[4], Dr. Sanskriti S. Mujumdar[5]

[1]Student, Civil Engineering, Dharmsinh Desai University, Nadiad
[2]Student, Civil Engineering, The Maharaja Sayajirao University of Baroda, Vadodara
[3]Registered Valuer, Civil Engineering, M/S Harishkumar Sheth, Godhra
[4]Assistant Professor, Civil Engineering, Dharmsinh Desai University, Nadiad
[5]Associate Professor, Civil Engineering, The Maharaja Sayajirao University of Baroda, Vadodara

Abstract
Conventional global regression models frequently do not adequately represent the discrete, non-linear spatial boundaries intrinsic to real estate valuation. This paper introduces an innovative hybrid machine learning architecture aimed at forecasting localized land rates in Eastern Side of Godhra Town, Gujarat. The proposed model removes Runge's oscillation by replacing continuous global polynomial regressions with localized Spatial Indicator Functions (Decision Trees) and Core Algorithmic Approaches (Engines) and Functional Synonyms (Process-Oriented). It also captures exact micro-neighborhood pricing behaviors. The architecture that came out of this had a continuous-time R^2 of 0.995. Also, to make sure that the model could be used in legacy computing environments without relying on Python, its mathematical graph was successfully turned into a native, macro-free spreadsheet logic architecture.

Keywords: Real Estate Valuation, Land Rate Prediction, Spatial Modeling, Non-Linear Regression, Machine Learning in Real Estate, Property Price Estimation, Hybrid Machine Learning Architecture, Spatial Heterogeneity, etc.

1. Introduction

The precise estimation of land value is significantly dependent on spatial attributes (latitude, longitude, and proximity to infrastructure) and temporal attributes (appreciation over time). Historically, Hedonic Pricing Models (HPM) have been used to figure out how much something is worth based on its underlying features. But when you use standard linear or polynomial regression on geospatial data, it assumes that the value changes smoothly and continuously. Real estate markets work in separate areas. For example, a commercial property that is only on a main road costs a lot more than one that is only a few meters away. Standard polynomial models overfit this, which makes predictions very unstable.

This study introduces a Hybrid Spatio-Temporal Decision Tree model. It divides the geographic space in to separate "buckets" to create a very localized baseline value. Then, it uses a separate linear growth multiplier for that micro-region.

2. Literature Review

The Hedonic Pricing Model (Rosen, 1974) is the basis for spatial property valuation. It says that a property's value is based on the specific combination of its internal and external features. Spatial Econometrics (Anselin, 1988) built on this by showing that properties have spatial autocorrelation, which means that properties that are close to each other tend to have similar prices.

Recent progress has moved toward Machine Learning methods (Antipov & Pokryshevskaya, 2012), which use ensemble methods like Random Forests and Gradient Boosting to find non-linear relationships. But these models are "black boxes" that need special computing environments to work. The Classification and Regression Tree (CART) algorithm (Breiman et al., 1984) provides a solution by recursively dividing data into subsets that are easy to understand and based on rules. This paper enhances the CART methodology by separating spatial variables from temporal variables, facilitating continuous-time interpolation while preserving the rigid spatial constraints of a decision tree.

3. Methodology

3.1 Data Transformation and Centering

To prevent floating-point overflow and keep the variance stable during recursive partitioning, the Geographic Coordinate System (GCS) values were centered on the dataset's regional mean.

Also, since the economic cost of being far away from a major infrastructure artery follows a power-law decay instead of a straight line, the distance variable (D) was changed to a natural logarithm.

- **Centered Latitude:** $Lat_c = Lat - 22.77$
- **Centered Longitude:** $Lon_c = Lon - 73.63$
- **Distance Decay:** $D_{ln} = \ln(D + 1)$
- **Temporal Normalization:** $Y_c = TargetYear - 2020$

3.2 Model Architecture and Bifurcation

The dataset exhibited a massive, discrete valuation spike at D = 0 (Commercial properties). To prevent these outliers from skewing the interior residential predictions, the model utilizes a piecewise bifurcation framework:

$$Rate(x) = \begin{cases} f_{Comm}(Lat_c, Lon_c, Y_c), & if\ D = 0 \\ f_{Res}(Lat_c, Lon_c, D_{ln}, Y_c), & if\ D > 0 \end{cases}$$

3.3 Phase I: Spatial Indicator Functions (Base Engine)

For interior properties (D > 0), a Decision Tree Regressor is utilized to map the geographic space. A decision tree mathematically represents a sum of spatial indicator functions. The algorithm recursively splits the Eastern Side of Godhra Town grid into M distinct, non-overlapping regions (R_m).

The baseline value function Φ_{Base} for the baseline year (2020) is defined as:

$$\Phi_{Base}(x) = \sum_{m=1}^{M} C_m, base \times I(x \in R_m)$$

Where $I(x \in R_m) = 1$ if the coordinates fall inside the spatial boundary R_m and 0 otherwise. $C_m, base$ is the optimal mean base rate for that specific geographic cluster.

3.4 Phase II: Localized Temporal Detrending (Growth Engine)

Traditional models assume a global appreciation rate. However, data indicates that prime sectors appreciate significantly faster than remote sectors. A secondary Decision Tree was trained specifically on the temporal slope (Annual Growth Rate) of each micro-region.

$$\Phi_{Growth}(x) = \sum_{m=1}^{M} C_m, growth \times I(x \in R_m)$$

Where $C_m, growth$ is the localized annual appreciation extracted via independent Ordinary Least Squares (OLS) regression on that specific coordinate grouping.

The final dynamic prediction function becomes:

$$Rate = \Phi_{Base}(Lat_c, Lon_c, D_{ln}) + [\Phi_{Growth}(Lat_c, Lon_c, D_{ln}) \times Y_c]$$

3.5 Transformation to Spreadsheet Logic

The mathematical graph of the Decision Trees, which had 64 spatial partitions, was turned into strictly nested Boolean algebra (logical IF structures) to make it easier to use in the field. This means that the whole machine learning inference engine can run instantly in a single cell of an old spreadsheet without needing to call any APIs or run any macros.

4. Results

The hybrid spatio-temporal architecture was much better than global linear and polynomial regression baselines.

- **Baseline Linear Regression R^2:** 0.741
- **Global Polynomial (Degree 3) R^2:** 0.950
- **Proposed Hybrid Spatial Tree R^2:** 0.995

The proposed model achieved a Mean Absolute Error (MAE) of just ₹430 across residential data points. More importantly, because time (Y_c) was isolated as a continuous linear multiplier outside of the spatial tree, the model accurately interpolates unseen time periods (e.g., predicting 2025 yields precisely the mid-point between the 2024 and 2026 localized trends).

5. Discussion

The main benefit of using recursive binary partitioning (CART) to model spatial coordinates instead of polynomial equations is that it is more realistic for geography. Zoning laws, road access, and neighbourhood boundaries usually make up hard geographic edges (polygons). A polynomial equation tries to fit a smooth, wavy line across these lines, which leads to predictions that "bleed" value from one zone to another. The indicator functions in this method make rigid geographic "buckets" that are exactly like the real-world economic boundaries of land parcels in the Eastern Side of Godhra Town.

6. Conclusion

This study demonstrates that utilizing hybrid spatial indicator functions coupled with localized temporal detrending yields superior accuracy in micro-local real estate valuation. By transpiling the trained mathematical graph into native spreadsheet logic, advanced machine learning precision can be successfully deployed in constrained, offline operational environments.

Bibliography & References

1. Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.
2. Antipov, E., & Pokryshevskaya, E. (2012). Mass appraisal of residential apartments: An application of Random Forest for flat valuation and a comparison with other predictive techniques. *Expert Systems with Applications, 39*(8), 7255-7259. https://doi.org/10.1016/j.eswa.2012.01.125
3. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.
4. Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.
5. Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34-55. https://doi.org/10.1086/260169
6. Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 4*(2), 234-240. (Reference for Tobler's First Law of Geography, foundational to the spatial decay distance model utilized).