INTERNATIONAL REAL ESTATE REVIEW

2026 Vol. 29 No. 2: pp. 319 – 336

# Use of Urban Lighting Data for Real Estate Price Forecasting

**Dmitriy Kravtsov***
Odessa National Maritime University, Odessa, Ukraine,
Email: dmkravtsov@gmail.com

**Nikolay Poletaev**
Odessa National Maritime University, Odessa, Ukraine,
Email: poletaev@ukr.net

This paper presents an improved model for forecasting rental property values in large metropolitan areas by incorporating features derived from the spatial distribution of urban lighting. The study uses rental housing data from Houston and Los Angeles (USA) together with high-resolution nighttime satellite imagery to generate additional explanatory variables. Light clusters are identified from satellite images and processed to determine their geographic location and spatial relationships. The clusters are georeferenced by using the Quantum Geographic Information System, thus enabling integration with other spatial datasets and improving modelling accuracy. The paper describes the methodology for feature extraction, spatial clustering, and integration into machine learning workflows. A Light Gradient-Boosting Machine predictive model is developed and compared with baseline models. The experimental results show that the proposed approach reduces the mean squared error by 11.8% for Houston and 9.37% for Los Angeles relative to conventional models. The findings demonstrate the usefulness of nighttime illumination features for capturing socio-economic and spatial patterns relevant to urban rental markets. The proposed methodology highlights the potential of combining geospatial data and machine learning techniques to improve automated valuation models and support urban analytics and smart city planning.

**Keywords**
Real estate price forecasting, Rental valuation, Machine learning, Nighttime satellite imagery, Geospatial analysis

---

* Corresponding author

**DOI:** https://doi.org/10.53383/100424

# 1. Introduction

Machine learning (ML) is a set of methods that allow models to learn and make predictions from data without the explicit programming of rules (Goodfellow et al., 2016; Bishop, 2006). Supervised learning methods are most frequently employed in real estate price forecasting, such as linear regressions, decision trees, ensemble methods (e.g., random forests, XGBoost, Light Gradient-Boosting Machine (LightGBM)), as well as deep neural networks (Bhistannavar et al., 2022; Peterson and Flanagan, 2009). These methods enable models to identify relationships between features (e.g., geographical location, property size, number of rooms) and the target variable, such as the price of a property. A key stage of forecasting is training the model on historical data, which include both features and known values of the target variable, followed by testing on new, unseen data to evaluate prediction accuracy.

The main advantage of ML methods lies in their ability to integrate large volumes of diverse data and to capture complex linear and non-linear dependencies between features. For example, the integration of textual property descriptions, visual data (satellite imagery, façade photographs), and geospatial information can improve predictive accuracy (Eman Ahmed and Moustafa, 2016; Wang et al., 2021). However, an important limitation is the dependence of model quality on the volume and diversity of data. A lack of data or limited feature sets can significantly reduce model effectiveness.

One of the key directions for improving forecasting accuracy is the enrichment of the dataset with new features. For example, the integration of geospatial data such as infrastructure accessibility, distance to key facilities, or the level of urbanisation has proven promising in real estate price forecasting (Orford, 2002; Das et al., 2021). This research direction opens opportunities to include new types of features that have not yet been extensively explored in this field. Among such features are data on nighttime illumination in large metropolitan areas, which can be obtained from high-resolution satellite imagery. In the scientific literature, the intensity and spatial distribution of illumination are considered indicators of economic activity, since these features correlate with population density, income level, and industrial activity (Henderson et al., 2012; Ghosh et al., 2013). Satellite-derived nighttime light data are widely used for monitoring urbanisation and changes in the urban environment, thus allowing researchers to identify areas of city growth, examine the extent of urban sprawl, and detect changes in urban structure (Small et al., 2005; Sutton et al., 2007). These data are also applied for assessing energy consumption, enabling analyses of the effectiveness of energy-saving technologies in cities (Falchi et al., 2016; Bennie et al., 2014), as well as understanding the spatial distribution of resources and its impact on socio-economic development (Levin et al., 2020; Ma et al., 2014).

Building on this experience, as well as the preliminary results of research on the enrichment of property price forecasting models with nighttime illumination data for the city of Houston (Bushuyev et al., 2024), a significant improvement in prediction accuracy can be expected when light-based features are used. Several arguments support this assertion. First, nighttime brightness can indicate the level of activity in an area. Well-illuminated places are often associated with business and cultural centres, popular residential districts, or areas with well-developed infrastructure (Henderson et al., 2012; Ghosh et al., 2013). Second, light data allow the dynamics of neighbourhood changes to be captured. For example, a reduction in illumination may indicate declining activity, which can influence the housing market. Finally, light-based features complement traditional data well, such as property size, location, number of bedrooms, bathrooms, or year of construction. They provide a more comprehensive representation of the surrounding environment, thus helping models to better understand the influence of neighbourhood characteristics on property value.

The use of light data in real estate price forecasting is a novel approach that can significantly enhance the accuracy and practical value of predictive models. The aim of this study is to develop the methodological foundations for the application of this approach, by using two US metropolitan areas (Houston and Los Angeles) as examples, which differ in their spatial distribution of light intensity. The study also seeks to conduct a comparative analysis of the impact of various light-based features on prediction accuracy and interpret the obtained results.

The central research question addressed in this study is how can the spatial structure of nighttime illumination serve as a measurable proxy for urban economic activity and neighbourhood attractiveness, thus enhancing automated valuation models. The hypothesis is that light-cluster proximity reflects socio-economic density, which directly influences rental prices.

## 2. Real Estate Price Forecasting Model with Use of Light-Based Features

### 2.1 Methodology for Extracting Light-Based Features

The foundation of the proposed ML model lies in the enrichment of the dataset with additional features that characterise the spatial distribution of nighttime urban illumination. To obtain these features, high-resolution nighttime satellite images of cities are employed and georeferenced with the Quantum Geographic Information System (QGIS) system (QGIS, n.d.) by aligning four control points of each city to the geographic grid, thereby ensuring basic geodetic correction for subsequent spatial calculations. High-resolution nighttime images are obtained from the Gateway to Astronaut Photography of Earth (National

Aeronautics and Space Administration, 2026; International Space Station digital camera, image example: Houston, February 4, 2017, with a comparable image for Los Angeles). Following preprocessing, the black-and-white images (Figures 1 and 2) are used for the visual analysis.

**Figure 1** **Satellite Image of Houston (Black-and-White)**

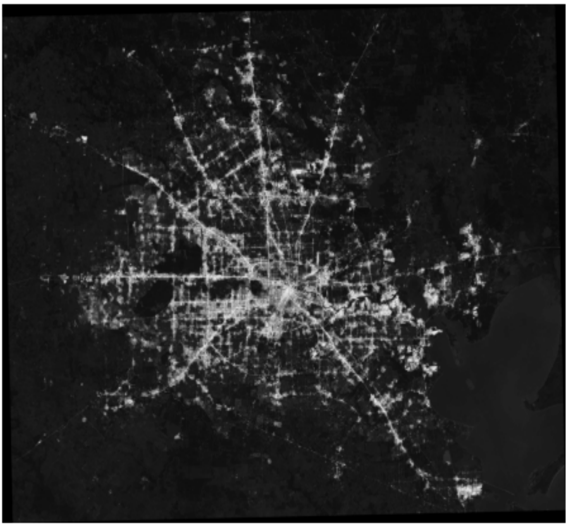


**Figure 2** **Satellite Image of Los Angeles (Black-and-White)**

The details for the satellite image processing are found in Bushuyev et al. (2024) and include the following stages:

1. Adjustment of image brightness and contrast by using OpenCV tools to highlight particularly bright light areas (Bradski, 2000). This process enhances the visibility of key sections, such as luminous spots, and facilitates their identification against less illuminated regions of the image. This constitutes an important stage in data preparation for further analysis.

2. Conversion of images to grayscale, which simplifies their analysis, as each pixel is represented by a single brightness value instead of three colour components (red, green and blue or RGB). This reduces computational load and simplifies operations such as identifying key regions, contours, and other structures within the image.

3. Application of threshold binarization to highlight light spots. This technique converts the grayscale image into a binary format, where pixels are classified as light or dark depending on a specified brightness threshold. The method isolates key regions with high light intensity, thereby simplifying the subsequent identification of contours and centres of luminous spots. This approach emphasises important sections of the image for further analysis. A higher binarization threshold means fewer light regions remaining in the image, whereas a lower threshold preserves more detail, including noise.

4. Contour detection of luminous regions and identification of their centres. The contours of light spots are determined by using OpenCV algorithms, which detect the boundaries of objects in a binarized image by grouping adjacent pixels of equal brightness into closed regions. The centres are calculated as the centroids of contours, by using image moments to define the mean point of pixel distribution within each region.

5. Transformation of the pixel coordinates of centres into geographical coordinates by georeferencing the corners of the images.

As a result of carrying out Steps 1–5, the local centres of luminous spots are identified on the satellite images. The centres of the light spots are then converted from pixel coordinates into geographical coordinates by using the information on the image boundaries. This accurately determines the spatial location of the illuminated areas (Figures 3 and 4).

Subsequently, a density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm (Ester et al., 1996) is applied to group closely located centres, thereby allowing the identification of global geographical centres of light clusters. The data provide a visual representation of the spatial structure of illumination, which is employed for further analysis (Figures 5 and 6).

**Figure 3** Image of Houston with Centres of Bright Spots (4,815 Centres)

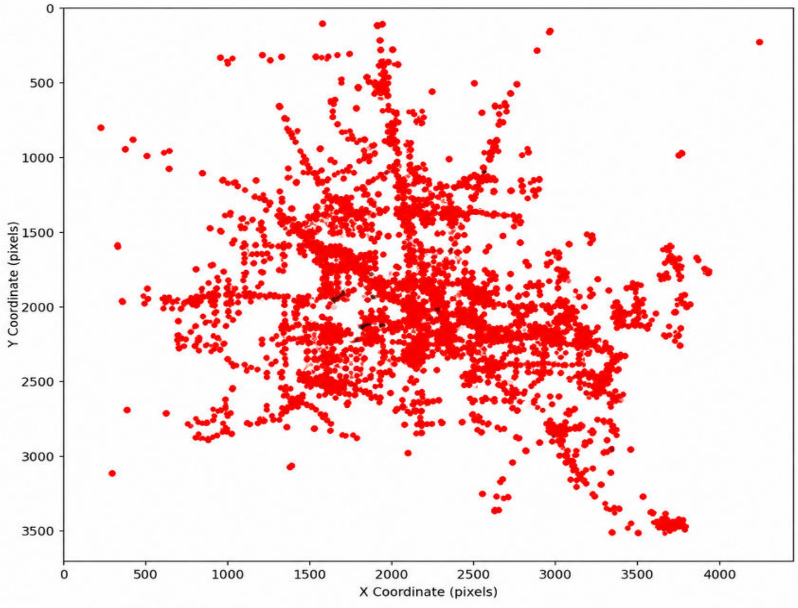


**Figure 4** Image of Los Angeles with Centres of Bright Spots (3,775 Centres)

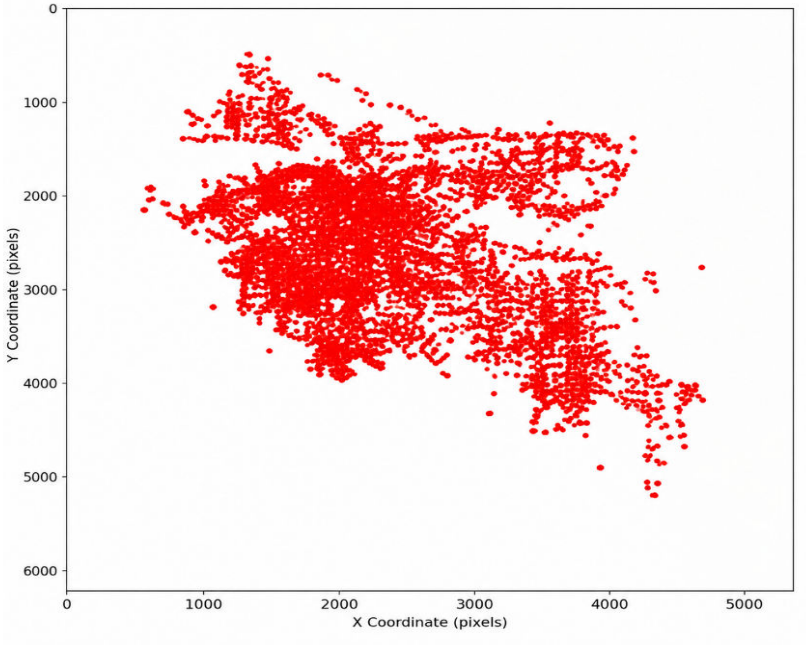


**Figure 5** **Results of DBSCAN Clustering for Houston**

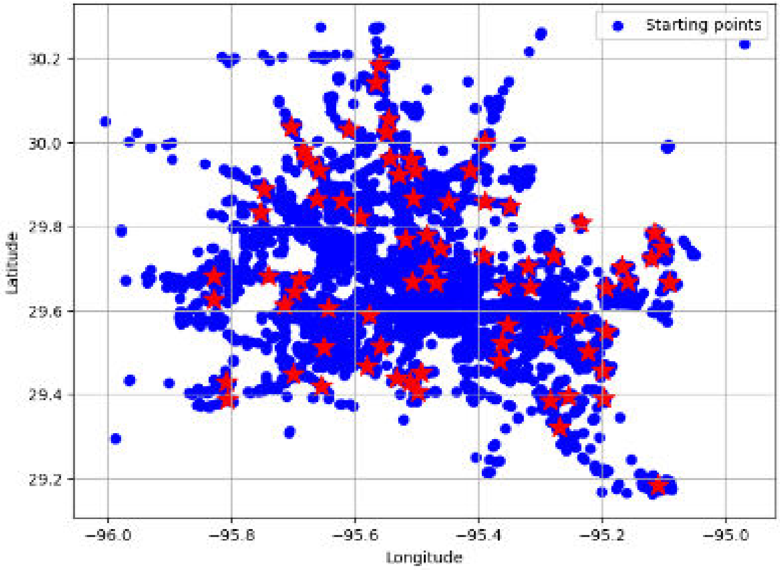
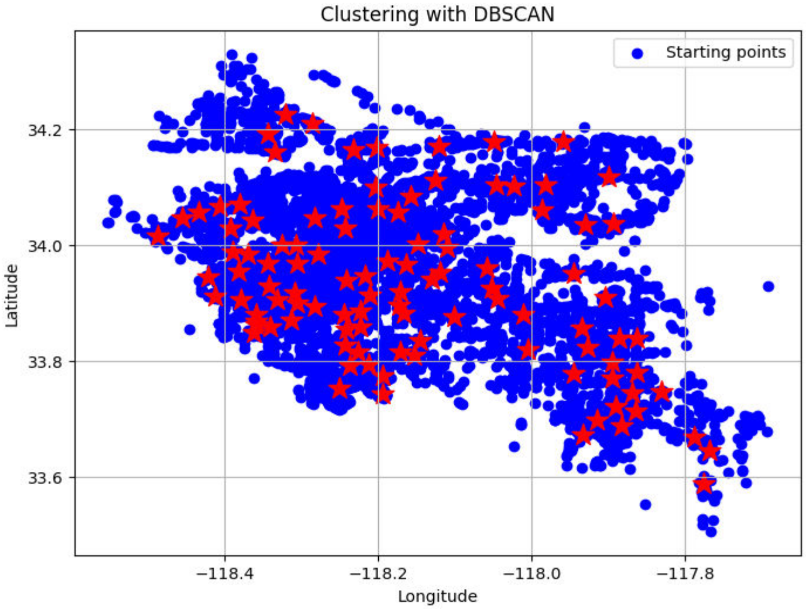


**Figure 6** **Results of DBSCAN Clustering for Los Angeles**



Based on the obtained data, numerical features associated with the spatial characteristics of the centres of light clusters are developed. To extract these features, the following are determined:
1. The number of cluster centres within various radii from a property,
2. The distance to the nearest cluster centre, and
3. The average distance to several of the nearest cluster centres.

These features are used to enrich the original dataset to increase its informativeness and enhance the accuracy of the real estate price forecasting models.

### 2.2 Integration of Light-Based Features into Machine Learning Model (Property Valuation)

The LightGBM algorithm is used as the baseline forecasting model due to its good performance and efficiency when working with large volumes of data. LightGBM, based on the gradient boosting method, constructs decision trees by using leaves rather than levels, which minimises loss at each stage and ensures high model accuracy. Unlike other boosting models, such as XGBoost, LightGBM provides more rapid training and has the ability to process sparse matrices, which make this framework an ideal choice for tasks that involve large datasets and complex feature dependencies. These characteristics render LightGBM an optimal tool for enhancing both the performance and accuracy of the model (John et al., 2022).

As the initial datasets, rental housing data from the cities of Houston in Texas, and Los Angeles in California, provided on the Redfin platform (Redfin, 2026), are utilised. The datasets contain 9,260 and 5,648 records respectively, with features describing property characteristics such as geographical coordinates, year of construction, number of bedrooms and bathrooms, total floor area, lot size, and monthly rental price (Tables 1 and 2). For data preparation, missing values are removed, outliers are excluded, and value ranges are verified to ensure the accuracy and realism of the data. Since rental prices in both cities contain numerous outliers and exhibit a right-skewed distribution, a logarithmic transformation of the target variable is used to improve model stability and predictive performance; inverse transformation is performed when reporting the final metrics.

After processing the satellite images of Houston and Los Angeles, 16 new features related to the spatial characteristics of light centres are added to the original datasets (Tables 1 and 2) to account for the spatial distribution of illumination. Among these features are:

1. The number of light centres within radii of 0.1, 0.3, 0.5, 1, 2, 4, 6, 8, and 10 km from each property,
2. The distance to the nearest light centre, and
3. The average distance to the 2, 4, 6, and 8 nearest light centres.

**Table 1** **Descriptive Statistics of Dataset Features for Houston**

| Index | count | mean | min | 25% | 50% | 75% | max | std |
|---|---|---|---|---|---|---|---|---|
| Latitude | 9260 | 29.770207 | 29.532642 | 29.722939 | 29.750557 | 29.803406 | 30.135092 | 0.09 |
| Longitude | 9260 | -95.444494 | -95.801611 | -95.527554 | -95.42787 | -95.380369 | -95.074206 | 0.11 |
| Year Built | 9260 | 1952.00 | 1893.0 | 1913.0 | 1940.0 | 1993.0 | 2023.0 | 30.78 |
| Beds | 9260 | 2.0 | 0.0 | 1.0 | 2.0 | 3.0 | 8.0 | 01.08 |
| Baths | 9260 | 1.8 | 0.0 | 1.0 | 2.0 | 2.0 | 8.0 | 0.90 |
| buildingSize | 9260 | 1268.7 | 101.0 | 748.0 | 1088.0 | 1608.0 | 5000.0 | 683.25 |
| lotSize | 9260 | 18726.1 | 0.0 | 0.0 | 0.0 | 5502.0 | 1175401.0 | 74470.52 |
| PostalCode | 9260 | 77050.3 | 77002.0 | 77018.0 | 77044.0 | 77077.0 | 77598.0 | 54.16 |
| Price | 9260 | 1848.2 | 619.0 | 1269.0 | 1698.0 | 2200.0 | 12500.0 | 876.50 |

**Table 2** **Descriptive Statistics of Dataset Features for Los Angeles**

| Index | count | mean | min | 25% | 50% | 75% | max | std |
|---|---|---|---|---|---|---|---|---|
| Latitude | 5648 | 34.062604 | 33.721745 | 34.042198 | 34.059162 | 34.087694 | 36.298804 | 0.05 |
| Longitude | 5648 | -118.340486 | -118.784673 | -118.391407 | -118.334549 | -118.278008 | -115.985656 | 0.08 |
| Year Built | 5648 | 1956 | 1823 | 1910 | 1944 | 1961.0 | 2023.0 | 973.4 |
| Beds | 5648 | 1.70 | 0.0 | 1.0 | 2.0 | 2.0 | 8.0 | 1.2 |
| Baths | 5648 | 1.76 | 0.0 | 1.0 | 1.0 | 2.0 | 8.0 | 1.1 |
| buildingSize | 5648 | 1148.5 | 100.0 | 650.0 | 900.0 | 1310.0 | 8125.0 | 870.8 |
| lotSize | 5648 | 11590.3 | 0.0 | 0.0 | 0.0 | 6003.0 | 17424000 | 234891.6 |
| PostalCode | 5648 | 90104 | 90001 | 90017 | 90028 | 90047 | 92840 | 307 |
| Price | 5648 | 4377.6 | 1306.0 | 2300.0 | 3100.0 | 4395.0 | 50000.0 | 4458.06 |

A multi-scale approach is adopted in the construction of the light-based features, as the impact of urban environment characteristics on rental prices is expected to manifest across different spatial levels. The selected distance thresholds (0.1–10 km) can be interpreted as approximations of key urban zones: small radii (0.1–0.5 km) capture immediate surroundings and walkability, medium radii (1–4 km) correspond to neighbourhood or submarket scales, while larger radii (6–10 km) reflect access to broader centres of economic activity and commuting zones. In this way, the proposed radii are aligned with standard urban concepts such as walkability, neighbourhood catchment areas, and commuting distances, while avoiding reliance on administrative boundaries. A uniform set of radii is used for all properties to ensure comparability between cities and avoid introducing city-specific assumptions of spatial structure. At the same time, differences in urban form (e.g., between Houston and Los Angeles) are implicitly accounted for through the feature selection procedure. In particular, Recursive Feature Elimination with Cross-Validation (RFECV) is used. which consistently identifies a similar subset of features in both cities, which primarily correspond to larger radii (4–10 km), thus suggesting that macro-level spatial patterns of urban activity play a more significant role in rental price formation than purely local effects.

To optimise the features in both datasets, the RFECV algorithm is also used to facilitate the automatic selection of the most significant characteristics, thus eliminating redundant features or those with weak influence. The iterative selection process is carried out by using a decision tree regressor, which ensures high performance and accuracy (Awad and Fraihat, 2023).

As a result of the optimisation, the dimensionality of the feature space for both cities are reduced from 24 (including baseline features) to 16, which improves the quality of the model training (Table 3).

To ensure a transparent comparison, two model specifications are considered: a baseline model and an enriched model. The former includes standard property-level characteristics, which can be grouped into spatial features (latitude, longitude, and postal code) thus capturing location, and structural attributes (year built, number of bedrooms and bathrooms, building size, and lot size) which describe the physical properties of the dwelling. The latter extends this specification by incorporating features derived from the spatial distribution of light clusters, including the number of clusters within radii of 4, 8, and 10 km, distance to the nearest cluster, and average distance to multiple nearest clusters (2, 4, 6, and 8). More importantly, these features are selected through the RFECV procedure from an initially broader multi-scale feature set, and a consistent subset is identified across both cities. Thus, the only difference between the two model specifications is the inclusion of light-based features, thereby allowing for an isolated assessment of their contribution to the predictive performance.

**Table 3 Optimal Features for Forecasting Model**

| Feature | Description |
|---|---|
| Latitude | Geographical latitude of the property |
| Longitude | Geographical longitude of the property |
| Year Built | Year of construction of the building |
| Beds | Number of bedrooms in the property |
| Baths | Number of bathrooms in the property |
| buildingSize | Floor area of the property |
| lotSize | Land plot area where the building is located |
| PostalCode | Postal code |
| clusters_within_4km | Number of clusters within a radius of 4 km from the property |
| clusters_within_8km | Number of clusters within a radius of 8 km from the property |
| clusters_within_10km | Number of clusters within a radius of 10 km from the property |
| distance_to_nearest_clusters | Distance to the nearest cluster |
| avg_dist_nearest_clusters_2 | Average distance to the 2 nearest clusters |
| avg_dist_nearest_clusters_4 | Average distance to the 4 nearest clusters |
| avg_dist_nearest_clusters_6 | Average distance to the 6 nearest clusters |
| avg_dist_nearest_clusters_8 | Average distance to the 8 nearest clusters |

**Table 4 Results of Model Performance Comparison before and after Inclusion of Light-Based Features**

| City | Metric | Before enrichment | After enrichment | % Improvement |
|---|---|---|---|---|
| Houston | MSE | 139998 | 123431 | 11.85 |
| | RMSE | 374 | 351 | 6.15 |
| | R2 | 0.817752 | 0.839 | 2.61 |
| | MDAPE | 7.830909 | 7.386 | 5.67 |
| Los Angeles | MSE | 389724 | 353336 | 9.37 |
| | RMSE | 624.24 | 594.41 | 4.78 |
| | R2 | 0.837448 | 0.868439 | 3.70 |
| | MDAPE | 10.111689 | 9.28 | 8.22 |

In the final stage, the model hyperparameters are optimised by using the Optuna library. The list of optimised parameters includes tree depth (max_depth), learning rate (learning_rate), number of leaves (num_leaves), maximum number of bins (max_bin), number of trees (n_estimators), and the fraction of random features for splits (colsample_bytree). For each combination of parameters, five-fold cross-validation is performed, which evaluates model performance with the metrics of mean squared error (MSE), root mean squared error (RMSE), coefficient of determination ($R^2$), and median absolute percentage error (MDAPE). The MSE measures the average deviation of predictions from actual values, and gives greater weight to large errors. The RMSE allows errors to be interpreted in the original units of measurement, which makes them more comprehensible. $R^2$ shows the proportion of variance in the target variable explained by the model, thus showing its overall accuracy. The MDAPE is used to evaluate the median percentage error, which provides a robust metric against outliers for analysing typical deviations.

A standard five-fold cross-validation scheme is chosen instead of spatial partitioning, as the dataset consists of independent rental listings rather than continuous spatial surfaces; therefore, the geographic proximity between samples does not introduce significant spatial leakage or bias in performance estimation.

The Optuna algorithm automates hyperparameter tuning by using methods such as the tree-structured Parzen estimator and genetic algorithms, thus ensuring an efficient search for optimal values in large parameter spaces. Compared to RandomizedSearchCV (a hyperparameter tuning technique in scikit-learn), and GridSearchCV (a scikit-learn function that finds the optimal parameter values), both of which rely on an exhaustive or random search, Optuna adapts to the results of previous iterations, which makes the process more intelligent. The key advantages of Optuna are its ease of integration, flexibility of settings, and the possibility of parallel task execution, which significantly accelerate optimisation and enhance model accuracy (Akiba et al., 2019).

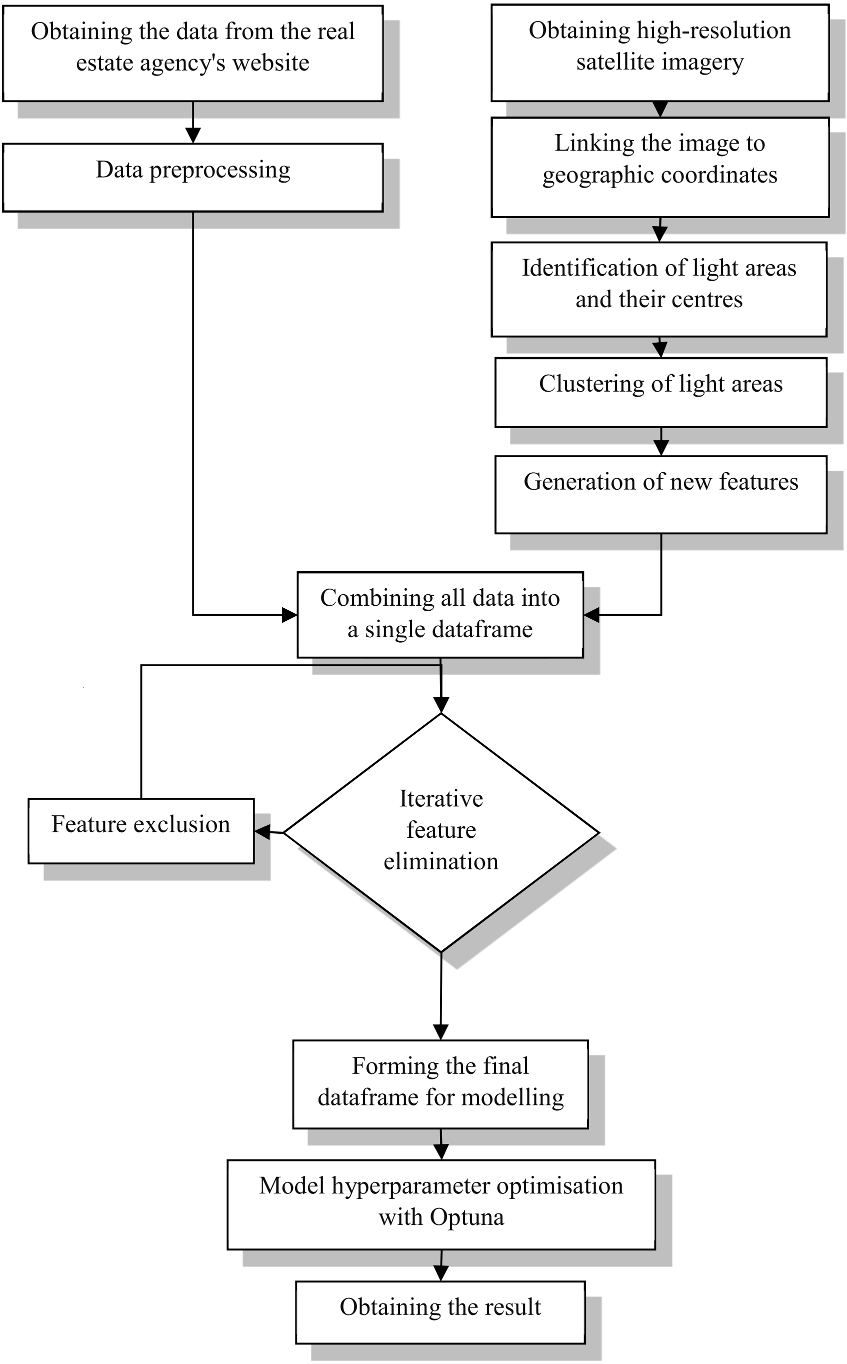
During optimisation, 200 iterations of hyperparameter combinations are carried out for the regressor of each city. The number of iterations is chosen to balance the quality of optimisation with computational cost. The best combination of parameters is selected on the basis of minimising the MSE, so that the optimal model settings can be determined. These results improve the performance and predictive accuracy for both Houston and Los Angeles (Table 4). The integration scheme of the light-based features into the forecasting model is presented in Figure 7.

## 3. Results and Discussion

After tuning the optimal hyperparameters for the LightGBM models for each city, we obtain the following cross-validation performance results.

The presented data show that the proposed feature enrichment method leads to improvements across all key model metrics. The differences in the degree of improvement can be explained by the specific characteristics of each metric.

**Figure 7** **Scheme of Integration of Light-Based Features into Forecasting Model**



The MSE, by squaring errors, places greater emphasis on large deviations, and its notable improvement may indicate a reduction in the number of large errors. The RMSE, as the square root of the MSE, returns values in the original units of measurement, so that they are more interpretable. However, the smaller improvement in the RMSE compared to MSE is explained by its lower sensitivity to large deviations.

The coefficient of determination ($R^2$) shows how effectively the model explains the variability of the dependent variable. Increases of 2.61% and 3.70% respectively indicate more accurate model performance in explaining the data. The MDAPE is robust to outliers. Its reduction indicates an overall improvement in accuracy, although it is less sensitive to large deviations compared to the MSE and RMSE.

As each metric evaluates different aspects of performance, the degree of improvement varies depending on how the new data enrichment method influences the specific characteristics of the model.

An analysis of the results shows that both cities, Houston and Los Angeles, exhibit improved forecasting quality after the enrichment of data with light-based features; however, the improvements are manifested differently.

For Houston, more significant improvements are observed in metrics related to large errors: a reduction in the MSE by 11.85% and RMSE by 6.15% indicates a substantial decrease in large deviations in the forecasts. This may be associated with a more uniform distribution of data in the city.

For Los Angeles, although the decreases in the MSE (9.37%) and RMSE (4.78%) are less pronounced, the model shows a more substantial improvement in the MDAPE of 8.22%. This suggests that the Los Angeles model better handles typical errors, although it remains somewhat more sensitive to extreme values.

In terms of $R^2$, Los Angeles also outperforms Houston, with an increase of 3.70% compared to 2.61%. This indicates that the model for Los Angeles better explains the overall data structure.

Overall, the model for Houston is more effective in reducing large errors, whereas the model for Los Angeles shows more balanced accuracy and robustness to outliers, which may be related to the more complex data structure in Los Angeles.

The identified light clusters can be interpreted as proxies for urban activity centres such as business districts, transport hubs, or entertainment areas, thus reflecting variations in population mobility and economic vitality across the city. In conventional real estate valuation models, spatial effects are typically captured by using variables such as distance to the central business district, accessibility to public transport and infrastructure, and socio-economic or demographic characteristics of neighbourhoods. However, these variables generally reflect specific aspects of the urban environment and often require the integration of multiple heterogeneous data sources. In contrast, light-based features can be interpreted as an integrated proxy that simultaneously captures several dimensions of urban activity, including economic intensity, infrastructure concentration, and spatial organisation. In this sense, the proposed approach is not intended to replace traditional spatial controls, but rather to complement them by providing a unified and scalable representation of the urban environment. This is particularly relevant in contexts where detailed spatial or socio-economic data are limited or not available, such as in developing regions. A comprehensive benchmarking against alternative sets of spatial variables is beyond the scope of the present study but can be a promising direction for future research.

## 4. Conclusions

Unlike previous studies that primarily use low-resolution global composites of nighttime lights, this work shows the practical integration of fine-scale illumination features into ML valuation pipelines, thereby contributing as a methodological advancement rather than merely a data substitution.

The effectiveness of adding light-based features derived from nighttime illumination data to improve real estate price forecasting models has been verified in this study. The conducted analysis shows that the proposed approach significantly enhances prediction accuracy. The successful application of this method to two cities with markedly different distributions of light clusters—Houston and Los Angeles—demonstrates its universality and potential for use in other regions. The simplicity and accessibility of satellite data make it possible to adapt the approach to other aspects of the urban environment, such as proximity to parks, water bodies, or industrial areas. Beyond its technical implementation, this study contributes conceptually to urban science by demonstrating how illumination intensity and spatial clustering can serve as measurable indicators of socio-economic activity and spatial heterogeneity in metropolitan environments. The proposed approach, while validated on two large US metropolitan areas, is expected to be applicable to a broader range of urban contexts. In particular, it can be extended to both monocentric and polycentric cities, as nighttime illumination reflects the spatial distribution of economic activity and built environment density. However, its effectiveness may be reduced in rural areas or regions with low levels of artificial lighting, where spatial variation is less pronounced. At the same time, the approach may be especially valuable in settings where conventional spatial and socio-economic data (e.g., census-based indicators or infrastructure datasets) are limited or unavailable, which is a common challenge in many developing countries. In such cases, light-based features can serve as a scalable and consistent proxy for urban activity and infrastructure intensity across different regions. In addition, the method has limitations associated with the availability of high-quality nighttime imagery for all regions, which necessitates careful work with the source data.

## References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework'. *'KDD '19:Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* Anchorage AK USA. August 4 – 8. pp. 2623 – 2631. https://doi.org/10.1145/3292500.3330701

Awad, M., and Fraihat, S. (2023). Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems *Journal of Sensor and Actuator Networks,* 12(5), 67. https://doi.org/10.3390/jsan12050067

Bennie, J., Davies, T. W., Duffy, J. P., Inger, R., and Gaston, K. J. (2014). Contrasting Trends in Light Pollution across Europe Based on Satellite Observed Nighttime Lights. *Scientific Reports,* 4, 3789. https://doi.org/10.1038/srep03789

Bhistannavar, V., Chauhan, N., Matey, V., and Shitole, A. (2022). *Real Estate Price Prediction using Supervised Learning. IEEE Pune Section International Conference*. Pune, India. December 15 to 17, 2022. https://doi.org/10.1109/PuneCon55413.2022.10014818

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Bradski, G. (2000). The OpenCV library. Dr. Dobb's Journal of Software Tools. https://opencv.org/

Bushuyev, S., Bushuiev, D., Poletaev, N., Malaksiano, M., and Kravtsov, D. (2024). 'A Machine Learning Method for Real Estate Operation Projects Forecasting'. *Proceedings of the 5th International Workshop IT Project Management (ITPM 2024),* Bratislava, Slovak Republic, May 22, pp. 77-89. https://doi.org/10.23939/IW_itpm2024.077

Das, S. S., Ali, M. E., Li, Y.-F., Kang, Y.-B., and Sellis, T. (2021). Boosting House Price Predictions Using Geo-Spatial Network Embedding. *Data Mining and Knowledge Discovery,* 35, 2221–2250. https://doi.org/10.1007/s10618-021-00789-x

Eman Ahmed, and Moustafa, M. (2016). 'House Price Estimation from Visual and Textual Features', *Proceedings of the 8th International Joint Conference on Computational Intelligence - (Volume 2)*, Porto, Portugal, November 9-11. pp. 62-68. https://doi.org/10.5220/0006040700620068

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). 'A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise'. *KDD;96 Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, August 2 – 4, pp. 226–231. https://dl.acm.org/doi/10.5555/3001460.3001507

Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M., Elvidge, C. D., Baugh, K., Portnov, B. A., Rybnikova, N. A., and Furgoni, R. (2016). The New World Atlas of Artificial Night Sky Brightness. *Science Advances,* 2(6). https://doi.org/10.1126/sciadv.1600377

Ghosh, T., Anderson, S., Elvidge, C. D., and Sutton, P. C. (2013). Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being. *Sustainability*, 5(12), 4988–5019. https://doi.org/10.3390/su5124988

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. USA: The MIT Press.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review,* 102(2), 994–1028. https://doi.org/10.1257/aer.102.2.994

John, L. M., Shinde, R., Shaikh, S., and Ashar, D. (2022). 'Predicting House Prices using Machine Learning and LightGBM'. *Proceedings of the 7th International Conference on Innovations and Research in Technology and Engineering (ICIRTE-2022),* Mumbai, India, April 8, pp. 5. https://dx.doi.org/10.2139/ssrn.4108744

Levin, N., Kyba, C. C. M., Zhang, Q., Sánchez de Miguel, A., Román, M.O., Li, X., Portnov, B.A., Molthan, A.L., et al. (2020). Remote Sensing of Night Lights: A Review and an Outlook for the Future. *Remote Sensing of Environment,* 237, 111443. https://doi.org/10.1016/j.rse.2019.111443

Ma, T., Zhou, Y., Zhou, C., Haynie, D., and Pei, T. (2015). Night-Time Light Derived Estimation of Spatio-Temporal Characteristics of Urbanisation Dynamics Using DMSP/OLS Satellite Data. *Remote Sensing of Environment*, 158, 453-464. https://doi.org/10.1016/j.rse.2014.11.022

National Aeronautics and Space Administration (2026). *Earth Observatory.* Available at: https://earthobservatory.nasa.gov

Orford, S. (2002). Valuing Locational Externalities: A GIS and Multilevel Modelling Approach. *Environment and Planning B: Planning and Design,* 29(1), 105–127. https://doi.org/10.1068/b2780

Peterson, S., and Flanagan, A. (2009). Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research,* 31(2), 147–164. https://doi.org/10.1080/10835547.2009.12091245

QGIS (n.d.). Spatial without Compromise. Spatial Visualization and Decision-Making Tools for Everyone. Available at: https://qgis.org/

Redfin (2026). Available at: https://www.redfin.com/

Small, C., Pozzi, F., and Elvidge, C. D. (2005). Spatial Analysis of Global Urban Extent from DMSP-OLS Nighttime Lights. *Remote Sensing of Environment,* 96(3–4), 277–291. https://doi.org/10.1016/j.rse.2005.02.002

Sutton, P. C., Elvidge, C. D., and Ghosh, T. (2007). Estimation of Gross Domestic Product at Sub-National Scales Using Nighttime Satellite Imagery. *International Journal of Ecological Economics and Statistics,* 8, 5–21. Available at: https://www.ceser.in/ceserp/index.php/ijees/article/view/1108

Wang, P.-Y., Chen, C.-T., Su, J.-W., Wang, T.-Y., and Huang, S.-H. (2021). Deep Learning Model for House Price Prediction using Heterogeneous Data Analysis Along with Joint Self-Attention Mechanism. *IEEE Access,* 9, 55244–55259. https://doi.org/10.1109/ACCESS.2021.3071306