Mass Valuation of Real Estate Using GIS-based Nominal Valuation and Machine Learning Methods

Mete MO[*1], Yomralioglu T[†1]

[1]Department of Geomatics Engineering, Istanbul Technical University, Istanbul, Turkey

Abstract

Geographic Information Systems (GIS) and Machine Learning methods are widely used in mass real estate valuation practices. Focusing on the physical attributes of properties, locational criteria are insufficiently used during the price prediction process. Whereas, locational criteria like proximity to important places, sea or forest views, flat topography are some of the spatial factors that extremely affect the real estate value. In this study, a hybrid approach is developed by integrating GIS and Machine Learning for automated mass valuation of residential properties in Turkey and the United Kingdom. GIS-based Nominal Valuation Method was applied to produce a land value map by carrying out proximity, terrain, and visibility analyses. Besides, ensemble regression methods like XGBoost, CatBoost, LightGBM, and Random Forest are built for price prediction. Spatial criteria scores obtained from GIS analyses were included in the price prediction data for feature enrichment purpose. Results showed that adding locational factors to the real estate price data increased the prediction accuracy dramatically. It also demonstrated that Random Forest was the most successful regression model compared to other ensemble methods.

KEYWORDS: real estate valuation; property valuation; mass valuation; GIS, Machine Learning

1. Introduction

Real estate value is an important component of the global land administration model required to achieve sustainable development goals. (Enemark, 2004). It supports the formation of an effective property market and is needed in many transactions such as purchase and sale, taxation, zoning applications, capital market activities (McDermott, Myers and Augustinus, 2018; UN-GGIM, 2019). Therefore, it is very important to evaluate the property values with an objective approach in accordance with the standards.

Numerous valuation reports are created each year due to the large volume of property transactions all over the world. However, there is a need for automated and faster valuation methodology in applications such as taxation, urban transformation, investment management, and planning. Mass property valuation methods like Hedonic Pricing, Nominal Valuation, Multiple Regression Analysis (MRA), Ensemble Regression Methods, Artificial Neural Networks (ANN) are widely used for the above-mentioned purposes in the literature (Jahanshiri et al., 2011; Mohd et al., 2020; Pagourtzi et al., 2003; Wang and Li, 2019).

Open data phenomenon has increased substantially with the determinant political will of countries recently (Open Data Barometer, 2017). Open data has many benefits like economic value, transparency, efficiency, and public service improvement (The European Data Portal, 2021). Thanks to open government and municipality data portals, users can reach updated and accurate data easily. The United Kingdom has made numerous data openly accessible on several topics through a data portal since 2010. HM Land Registry - Price Paid Data (PPD) is an open governmental data that contains sale records and several physical attributes of the residential properties in England and Wales. On the other hand, Energy Performance Certificates (EPC) data is also open data which holds many informative attributes such as

[*] metemu@itu.edu.tr
[†] tahsin@itu.edu.tr

total floor area, energy ratings, carbon dioxide emissions, heating costs of the buildings in England and Wales. In Turkey, it is also observed that open data portals have become widespread recently and that geographical data is shared with citizens creating an important value on the way to geospatial information industrialization.

Depending on the developments in information technologies, Geographic Information Systems (GIS) and Machine Learning have been used widely in real estate valuation activities. GIS provides an effective database management system, robust spatial analysis capability, and map-based visual outputs for real estate valuation activities. On the other hand, Machine Learning offers an accurate price prediction for mass valuation studies. The objective of this research is to create a mass valuation model with GIS and Machine Learning hybrid approach. In this sense, the study covers the use of the GIS-based Nominal Valuation Method for creating a land value map based on scientific and objective evaluation. Besides, it compares performances of various Machine Learning regression models for mass valuation of residential properties in Istanbul, Turkey and Greater London, England. It also discusses the feature enrichment of real estate price data by adding proximity, terrain, and visibility analysis scores and examines the contribution of the spatial features on the prediction accuracy.

2. Material and Methods

Mass valuation can be defined as analysing a set of factors to estimate the value of a large number of properties by utilizing statistical methods and standards. Unlike single real estate valuation, mass valuation approach utilizes statistics and automated valuation methods for maintaining and valuing a group of properties. There are several studies that implement different mass valuation methods like Multiple Regression Analysis (Benjamin et al., 2020; Yilmazer and Kocaman, 2020; Zurada et al., 2011), Hedonic Pricing (Lisi, 2019; Peterson and Flanagan, 2009; Yamani et al., 2019), Nominal Valuation (Mete and Yomralioglu, 2019; Yomralioglu, 1993; Yomralioglu et al., 2004), Geographically Weighted Regression (GWR) (Dimopoulos and Moulas, 2016; B. Huang et al., 2010; Wang et al., 2020), Ensemble Methods (Alfaro-Navarro et al., 2020; Aydinoglu et al., 2021; Gnat, 2021), and ANN (Demetriou, 2017; Lee, 2021; Yalpır, 2018). With the improvements in computing power, GIS and Artificial Intelligence applications have become widespread in Computer Assisted Mass Appraisal (CAMA) and Automated Valuation Models (AVMs) have been adopted in many countries recently (Renigier-Biłozor et al., 2022; Wang and Li, 2019).

Nominal Valuation is a stochastic method based on the Weighted Linear Combination approach which is used to calculate aggregation of criteria with different importance (Yomralioglu, 1993). It can be applied for both land and buildings by assigning scores for each affecting criterion. The total nominal value of a real estate is calculated as the weighted sum of criterion scores multiplied with the parcel or pixel area (1).

$$V_i = S_i * \sum_{j=1}^{k} (f_{ji} * w_j) \qquad (1)$$

V: Total nominal value
S: Parcel or pixel area
f: Factor score
w: Factor weight
k: Total number of the factors

Nominal Valuation method has many advantages in comparison with the other mass valuation methods. Since it is not directly dependent on market prices, it can be easily applied to real estate within large areas to reveal value differences based on spatial analyses. Besides, using pixel-based valuation, nominal value maps can be created with high granularity like sub-parcel level. It is also possible to calculate market prices from nominal values by using reliable real estate prices.

In this study, Nominal Valuation Method is used to create a land value map based on scientific and objective evaluation. Besides, prediction accuracies of various Machine Learning regression models are compared for mass valuation of residential properties in Istanbul and London cities (**Figure 1**).

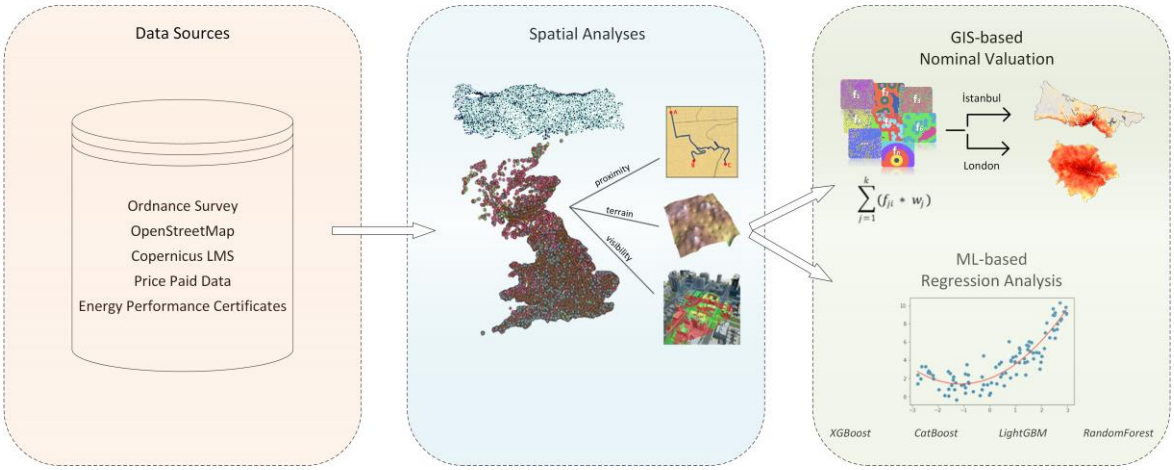

Figure 1 Workflow diagram of the study.

Land value maps provide an understanding of price variations in a region and facilitate management, planning, and taxation activities. Using open data sources, land value maps of Istanbul and London cities are produced with the GIS-based Nominal Valuation method (**Figure 2, 3**). Firstly, a valuation database is created on Amazon Aurora PostgreSQL Serverless (Mete and Yomralioglu, 2021) and all raster and vector data are imported into the spatial database. In order to carry out proximity, terrain, and visibility analyses, open source QGIS software is used.

Regression analysis is one of the fundamental algorithms of Machine Learning and it is widely used for real estate price prediction. In this study, several ensemble regression methods like XGBoost, CatBoost, LightGBM, Random Forest are built for price prediction. In order to measure the model performances, several accuracy metrics are used such as R^2, Adjusted R^2, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) (**Table 1**).

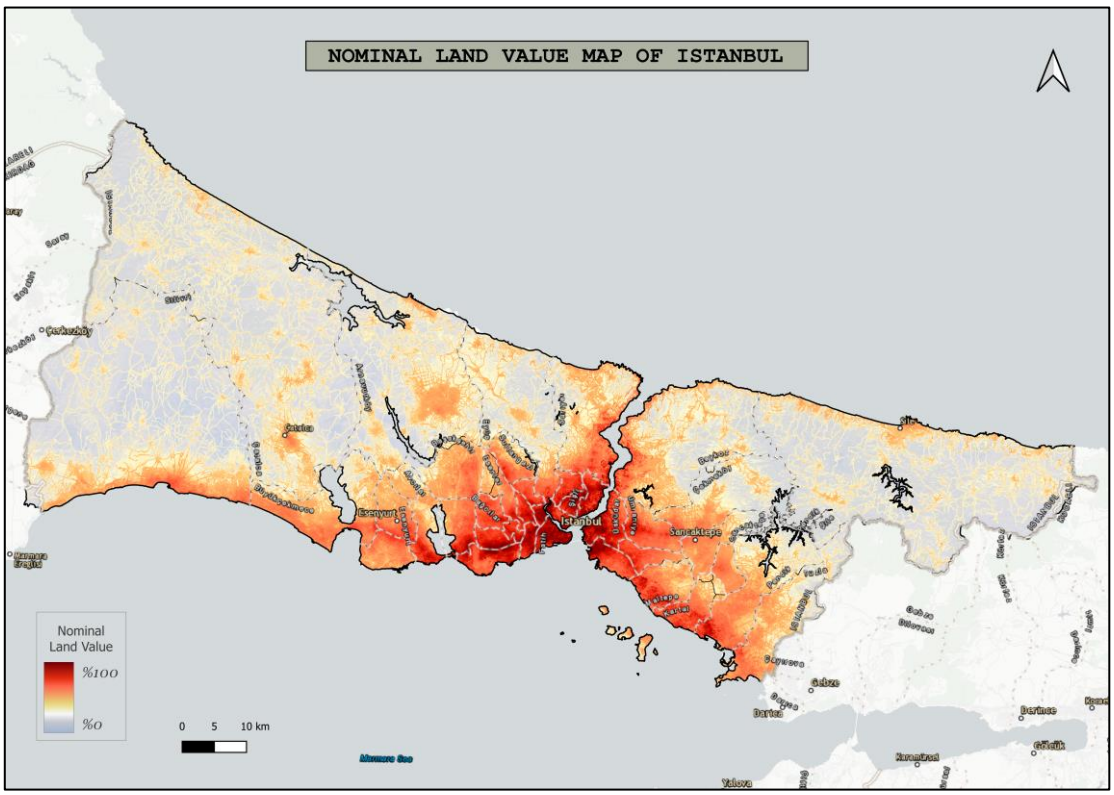

Figure 2 Nominal land value map of Istanbul, Turkey.

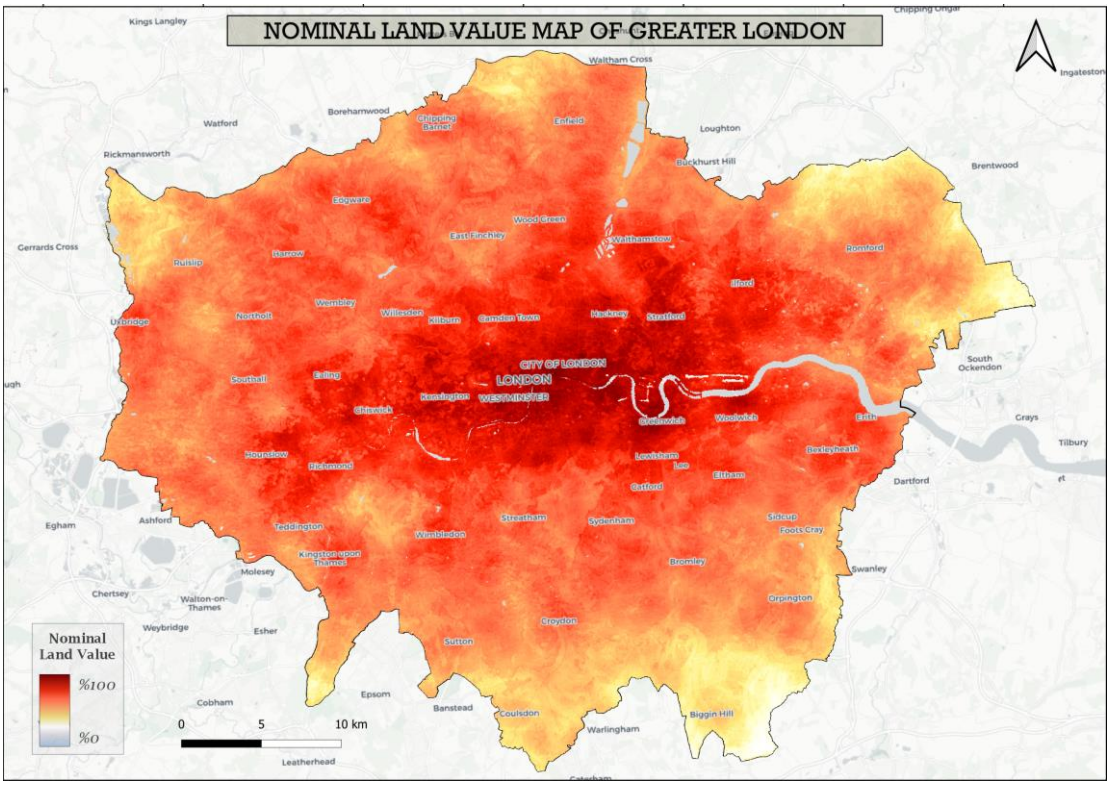

Figure 3 Nominal land value map of Greater London, England.

Table 1 Overall accuracy metrics of the regression models.

Model	R^2	Adjusted R^2	MAE	MAPE	RMSE
XGBoost	0.4341	0.4340	98505.9861	0.4689	204625.5031
CatBoost	0.4551	0.4550	98811.1141	0.4737	200791.2485
LightGBM	0.4366	0.4365	101291.7355	0.4937	204170.2015
RandomForest	0.4690	0.4689	95934.5759	0.4614	198211.9713

Real estate price data contain many physical attributes about properties that are beneficial for price prediction in Machine Learning based valuation process. However, there is a need for the inclusion of locational factors, which are highly correlated with land price, in order to build more accurate regression models (Wyatt, 1997; Kiel and Zabel, 2008; Mete and Yomralioglu, 2019). GIS provides numerous spatial analysis tools that can be utilized for revealing locational criteria effects on the real estate price. Thus, nominal scores of proximity, terrain, and visibility analyses are extracted for feature enrichment of regression models built for price data. After having the enriched data, the same workflow is followed for building the Machine Learning regression models. **Table 2** shows the accuracy metrics of the models.

Table 2 Overall accuracy metrics of the regression models after feature enrichment.

Model	R^2	Adjusted R^2	MAE	MAPE	RMSE
XGBoost	0.7907	0.7906	65367.1736	0.3107	123585.9848
CatBoost	0.8380	0.8379	56668.9762	0.2660	108740.2277
LightGBM	0.8058	0.8058	65567.8885	0.3188	119042.0076
RandomForest	0.8726	0.8725	31343.4681	0.1570	71781.2148

3. Conclusion

The adoption of mass valuation techniques in value-based activities has been increasing in recent years. Hedonic Pricing, Nominal Valuation, Spatial Analysis, and Regression Analysis are some of the prominent methods of mass valuation practices. Although there are numerous studies using Machine Learning algorithms for mass real estate valuation, it has been observed that the spatial factors that extremely affect the land value are neglected. In this study, GIS-based Nominal Valuation method and Ensemble Machine Learning regression algorithms were used for mass valuation of residential properties in Turkey and the United Kingdom. Results showed that Random Forest algorithm was the most successful model in price prediction. Countries or valuation organizations can implement this approach while conducting an automated valuation of immovable assets to use as a reference in value-based applications.

4. Acknowledgements

This work was supported by the Scientific Research Projects Coordination Unit of Istanbul Technical University [Grant No. MDK-2021-43080].

References

Alfaro-Navarro, J.-L., Cano, E. L., Alfaro-Cortés, E., García, N., Gámez, M., & Larraz, B. (2020). A Fully Automated Adjustment of Ensemble Methods in Machine Learning for Modeling Complex Real Estate Systems. *Complexity*, 2020, 1–12. https://doi.org/10.1155/2020/5287263

Aydinoglu, A. C., Bovkir, R., & Colkesen, I. (2021). Implementing a mass valuation application on interoperable land valuation data model designed as an extension of the national GDI. *Survey Review, 53*(379), 349–365. https://doi.org/10.1080/00396265.2020.1771967

Benjamin, J. D., Guttery, R. S., & Sirmans, C. F. (2020). Mass Appraisal: An Introduction to Multiple Regression Analysis for Real Estate Valuation. *Journal of Real Estate Practice and Education, 7*(1), 65–77. https://doi.org/10.1080/10835547.2004.12091602

Demetriou, D. (2017). A spatially based artificial neural network mass valuation model for land consolidation. *Environment and Planning B: Urban Analytics and City Science, 44*(5), 864–883. https://doi.org/10.1177/0265813516652115

Dimopoulos, T., & Moulas, A. (2016). A Proposal of a Mass Appraisal System in Greece with CAMA System: Evaluating GWR and MRA techniques in Thessaloniki Municipality. *Open Geosciences, 8*(1), 675–693. https://doi.org/10.1515/geo-2016-0064

Enemark, S. (2004). Building Land Information Policies. *UN, FIG, PC IDEA Inter-regional Special Forum on The Building of Land Information Policies in the Americas*. Aguascalientes, Mexico: FIG.

Gnat, S. (2021). Property Mass Valuation on Small Markets. *Land, 10*(4), 388. https://doi.org/10.3390/land10040388

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science, 24*(3), 383–401. https://doi.org/10.1080/13658810802672469

Jahanshiri, E., Buyong, T. and Shariff, A. R. M. (2011). A review of property mass valuation models. *Pertanika Journal of Science and Technology, 19* (SPEC. ISSUE), 23–30.

Kiel, K. A. and Zabel, J. E. (2008) 'Location, location, location: The 3L Approach to house price determination', *Journal of Housing Economics*, 17(2), pp. 175–190. doi: 10.1016/j.jhe.2007.12.002.

Lee, C. (2021). Enhancing the performance of a neural network with entity embeddings: an application to real estate valuation. *Journal of Housing and the Built Environment*, 1–16. https://doi.org/10.1007/S10901-021-09885-2

Lisi, G. (2019). Property valuation: the hedonic pricing model-location and housing submarkets. *Journal of Property Investment & Finance, 37*(6), 589–596. https://doi.org/10.1108/JPIF-07-2019-0093

McDermott, M., Myers, M. and Augustinus, C. (2018). *VALUATION OF UNREGISTERED LANDS: A Policy Guide.*

Mete, M. O. and Yomralioglu, T. (2019) 'Creation of Nominal Asset Value-Based Maps using GIS: A Case Study of Istanbul Beyoglu and Gaziosmanpasa Districts', *GI_Forum 2019*, 7(2), pp. 98–112. doi: 10.1553/giscience2019_02_s98.

Mete, M. O. and Yomralioglu, T. (2021) 'Implementation of serverless cloud GIS platform for land valuation', *International Journal of Digital Earth*, 14(7), pp. 836–850. doi: 10.1080/17538947.2021.1889056.

Mohd, T., Jamil, N. S., Johari, N., Abdullah, L. and Masrom, S. (2020). An Overview of Real Estate Modelling Techniques for House Price Prediction. N. Kaur and M. Ahmad (Ed.), Charting a Sustainable Future of ASEAN in Business and Social Sciences (ss. 321–338). Springer Singapore. doi:10.1007/978-981-15-3859-9_28

Open Data Barometer (2017) *Open Data Barometer 4th Edition*, *World Wide Web Foundation*. Available at: https://opendatabarometer.org/4thedition/ (Accessed: 28 February 2022).

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance, 21*(4), 383–401.

Peterson, S., & 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

Renigier-Biłozor, M., Źróbek, S., Walacik, M., Borst, R., Grover, R., & D'Amato, M. (2022). International acceptance of automated modern tools use must-have for sustainable real estate market development. *Land Use Policy, 113*, 105876. https://doi.org/10.1016/j.landusepol.2021.105876

The European Data Portal (2021) *The Open Data Maturity (ODM) Report 2021*. Luxembourg. Available at: https://data.europa.eu/sites/default/files/landscaping_insight_report_n7_2021.pdf.

UN-GGIM. (2019). *Framework for Effective Land Administration (FELA).* Available at: http://ggim.un.org/meetings/2018-Deqing-Expert-Group/documents/Concept_Note.pdf

Wang, D., & Li, V. J. (2019). Mass Appraisal Models of Real Estate in the 21st Century: A Systematic Literature Review. *Sustainability, 11*(24), 7006. https://doi.org/10.3390/su11247006

Wang, D., Li, V. J., & Yu, H. (2020). Mass Appraisal Modeling of Real Estate in Urban Centers by Geographically and Temporally Weighted Regression: A Case Study of Beijing's Core Area. *Land, 9*(143). https://doi.org/10.3390/LAND9050143

Wyatt, P. J. (1997) 'The development of a GIS-based property information system for real estate valuation', *International Journal of Geographical Information Science*, 11(5), pp. 435–450. doi: 10.1080/136588197242248.

Yalpır, Ş. (2018). Enhancement of parcel valuation with adaptive artificial neural network modeling. *Artificial Intelligence Review, 49*(3), 393–405. https://doi.org/10.1007/s10462-016-9531-5

Yamani, S. El, Ettarid, M., & Hajji, R. (2019). Building Information Modeling Potential For An Enhanced Real Estate Valuation Approach Based On The Hedonic Method. Building Information Modelling (BIM) in Design, *Construction and Operations III, 1*, 305–316. https://doi.org/10.2495/bim190261

Yilmazer, S., & Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy, 99*(104889). https://doi.org/10.1016/j.landusepol.2020.104889

Yomralioglu, T. (1993) *A Nominal Asset Value-Based Approach For Land Readjustment And Its Implementation Using Geographical Information Systems* (PhD thesis). University of Newcastle upon Tyne, England.

Yomralioglu, T., Nisanci, R., & Yildirim, V. (2004). *Nominal Asset Land Valuation Technique by GIS.* FIG Working Week 2004.

Zurada, J., Levitan, A., & Guan, J. (2011). A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Context. *Journal of Real Estate Research, 33*(3), 349–388. https://doi.org/10.1080/10835547.2011.12091311