Examining Machine Learning Techniques to Improve the Accuracy of Mass Valuation Prediction on Big Geographic Data: a Case Study of Tuzla, Istanbul

Arif Cagdas Aydinoglu and Suleyman Sisman (Türkiye)

Key words: Land management; Real estate development; Valuation; GIS, Machine Learning, Mass Valuation, RF, GBM, LightGBM.

SUMMARY

Considering today's rich data sources, preparing the dataset appropriately and using it in Machine Learning (ML)-based mass valuation models is crucial for accuracy. In this context, a case study was conducted in the Tuzla district of Istanbul city. 42.069 market samples were obtained covering in the period 2020-2023 and outliers were removed with pre-processing process techniques. Big geo-datasets representing the factors affecting real estate value were produced by utilizing spatial analysis techniques in GIS. In this way, a big geo-enriched dataset containing 200 features was prepared and 120 features were selected for modelling. The prepared dataset was divided into training/test datasets. Prediction models were developed with Gradient Boosting Machine (GBM), Light Gradient Boosting Machine (LightGBM), and Random Forest (RF) techniques The performance of the models was evaluated on the test dataset with used ML metrics such as MAE, MAPE, RMSE, R2 and mass valuation metrics such as COD and PRD. When the results were evaluated comparatively, it was determined that while the RF and GBM models showed very close results, the LightGBM model showed superior performance with an R2 value of 0.86. The results clearly demonstrate that big geo-enriched datasets and ML techniques can be used effectively in mass valuation.
