Investigating Performance Improvement of Machine Learning-Based Mass Valuation Models with Spatially Constrained Multivariate Clustering Algorithm: a Case Study of Istanbul and Kocaeli, Turkiye

Suleyman Sisman and Arif Cagdas Aydinoglu (Türkiye)

Key words: Geoinformation/GI; Land management; Real estate development; Valuation; Clustering; GIS; Machine Learning; Mass Valuation; LightGBM; SCMCA

SUMMARY

Machine Learning (ML) and clustering algorithms can be used effectively mass valuation models. In this context, a case study was conducted in the five districts of Istanbul and Kocaeli cities. Market samples and other datasets were obtained and prepared through data pre-processing steps. 11 features were determined for clustering. The Spatially Constrained Multivariate Clustering Algorithm (SCMCA) was used to improve the modelling performance by identifying similar sub-market regions and five spatial clusters were created. Training/test datasets were prepared for the study area and spatial clusters Mass valuation models were developed with the Light Gradient Boosting Machine (LightGBM) technique, which provides effective performance compared to other ML techniques. The accuracy of the models was measured on test datasets with widely used performance metrics. The performances of the models for the study area and spatial cluster-based models were compared. Considering the MAPE, the study area model exhibited an error of 10.43%, while Cluster-1 exhibited an error of 8.58%, Cluster-2 13.40%, Cluster-3 9.51%, Cluster-4 7.13% and Cluster-5 7.97%. All models except Cluster-2 exhibited higher prediction accuracy than the study area model. The results clearly show that with SCMCA, both sub-market regions can be identified, and modelling performance can be increased in mass valuation.
