Using Machine Learning to Create High Performance Models for AVM Without Linearity Constraints

Luke Jorgensen and Joshua Jorgensen (USA)

Key words: Education; Real estate development; Valuation; Young surveyor; Large spatial data; Machine learning; Mass appraisal; Property evaluation

SUMMARY

The current standard of utilizing statistical regression in automated valuation models (AVM) has been a critical leap forward in the mass appraisal of real estate. Regression models have been able to provide consistent evaluation to an entire population utilizing the inherent ability to incorporate a multitude of factors in a systematic way with easily explained outcomes. This paves the way for predicted outcomes in real-time for millions of properties in an efficient iterative process. The compact iterative process allows for an individual to quickly make changes while tracking those changes. We build on this traditional process by adding machine learning to the modeling process. While there are many benefits to having linear regression models in AVM to predict values there are some limitations. For example, many statistical models, including linear regression, are only able to capture linear relationships or require all the predictors to be independent of one another. This can lead to difficulties in creating an accurate AVM when utilizing these techniques. Many features in real estate do not follow linear trends e.g. living area, distance to water, etc., to name a few. Employing machine learning in the AVM process significantly changes the efficacy of the models. Machine learning consists of algorithms to identify, "understand" and "learn" latent patterns and relationships from data without hard coding fixed rules. Machine learning models, by design, pick up and understand nonlinear trends, as well as many other nuances. This provides a multitude of benefits from having a more accurate model to not having to try to break apart variables for different purposes since linear regression cannot handle non-linearity. Machine learning models are particularly suited to addressing many of the inadequacies of a traditional linear regression AVM. In a machine learning model, real-time global and local predictions are maintained with a higher level of efficiency. In addition, techniques can be applied to the model in order to provide a high level of "explainability" to the outcomes that capture non linearities and other nuances. The benefits derived from this technique are significant. This paper will explore the differences and demonstrate how machine learning models can be a higher performing alternative to

traditional regression models in AVM, an alternative that results in significantly better outcomes as measured through accuracy, uniformity and explainability.