

Abstract

In this thesis, we propose alternative ways to apartments' mass appraisal. This work enriches the current literature by combining several techniques of data extraction and price estimation. We are not aware of any similar work providing an in-depth overview of the Czech apartment market.

Throughout the empirical analysis, five different methods (OLS, LASSO, decision tree, random forests, and kNN) are applied to the dataset of 15,848 classifieds. The aim of the study is to find the most accurate method of estimating offering prices, using structured variables as well as data extracted by text mining. We use various accuracy statistics and graphical analysis to validate our results. Tree-based methods, specifically the random forest algorithm, results with the highest accuracy in predicting offering prices. Additionally, text-based variables included in the model cause the reduction of errors on linear models.

The last part of the analysis covers the main determinants of property value in Prague and the rest of the Czech Republic. We show that prices in Prague can be estimated with higher preciseness and with the lower number of independent variables.