Management Summary

Table of contents

List of Figures

List of Tables

In the face of the global climate crisis, the role of the building sector has become increasingly important - due to its potential to reduce greenhouse gas emissions. Switzerland and the the Canton of Basel-Landschaft have both committed to achieving net-zero emissions by 2050. Two-thirds of the residential buildings in the Canton of Basel-Landschaft are still heated by fossil energy sources. Therefore, a fundamental transformation of the building stock is necessary. This includes switching these heating systems to renewable energy sources within the next decades. The cantonal and municipal energy policies have the task of steering this process with effective measures and a goal-oriented planning. In this context, having a solid and precise data foundation is crucial for monitoring progress and measuring the impact of policy measures. The Canton of Basel-Landschaft has been conducting a cantonal energy statistics since 1990. The statistics involves estimating heat energy consumption for buildings where no measured data is available. At it's core, the estimation relies on measured gas consumption data that is combined with further building information from the Cantonal Register of Buildings and Dwellings (RBD). Although the method offers accurate estimates at the cantonal and, to some extent, municipal levels, the error at the level of individual buildings can be substantial.

The goal of this Master's thesis is to propose and evaluate alternative estimation methods by using various machine learning algorithms that can potentially improve the results on the level of individual buildings and municipalities. To this end, we used the same data basis as the cantonal energy statistics, combining it with additional data sources to integrate information that can have an influence on the heat energy consumption (such as retrofit measures or socio-economic factors). We trained different models on the entire dataset as well as on more homogeneous subsets. Subsequently, the predictive performance of the models was evaluated and compared to the current approach of the energy statistics. In addition, we examined the predictions of the best performing model using model explainability techniques such as Variable Importance Analysis (VIA) and Shapely Additive Explanations (SHAP). The study showed that Gradient Boosting Machines (GBM), XGBoost Regressor (XGB) and Stacked Ensembles provide the most accurate predictions with lower estimation errors than the current approach of the energy statistics. Additionally, we found that splitting the data into more homogeneous subsets based on the building class or municipality led to more precise estimations. The results of this study provide a foundation for further improving the accuracy of the cantonal energy statistics. The developed models can potentially also be used in other Cantons or municipalities where consumption data is not available. Ultimately, the study serves as an example for integrating machine learning methods into official statistics.