Modelling the energy demand of individual buildings in Switzerland

Author: Andy Gubser
Quartnerstrasse 2
CH-8882 Unterterzen
andy.gubser@stud.hslu.ch

Lecturer: Prof. Dr. Philipp Schütz
Lucerne School of Engineering and Architecture
Technikumstrasse 21
CH-6048 Horw
philipp.schuetz@hslu.ch

Co-Lecturer: David Suter
geoimpact AG
Heinrichstrasse 267
CH-8005 Zürich
david.suter@geoimpact.ch

Lucerne University of Applied Science and Arts
Master in Applied Information and Data Science (MScIDS)
Autumn Semester 2021

Heitenried, 23 December 2021

Management Summary

Climate change is in full swing, and to avoid the worst consequences, we must eliminate our greenhouse gas emissions. One of the most significant contributors is the building sector, and sustainable solutions are ready and competitive. However, most buildings are still heated with fossil fuels, and many of them urgently need energy retrofits. Geoimpact aims to change this and provide building owners and other stakeholders with fact-based information about sustainable alternatives. To improve their consulting, they need a reliable estimate of the heating energy consumption of each building.

This is the question we are dealing with in this master thesis. We have investigated which models achieve the highest accuracy in predicting the energy performance of individual buildings. Which target variables, which feature variables, and which regressors are appropriate. And whether we can increase accuracy by training multiple models on specific subsets rather than using one model for all buildings.

To this end, we merged energy consumption data from Biel, Geneva, and St. Gallen with data from our industry partner geoimpact and data from the Federal Register of Buildings and Dwellings (RBD) and Minergie. We trained models with different regressors and target variables. We evaluated their predictions with performance evaluation indicators such as the Coefficient of Variance (CV) and the Mean Absolute Percentage Error (MAPE). Thereby, we assessed the Multiple Linear Regressor (MLR), the Random Forest Regressor (RFR), the XGBoost Regressor (XGB), and the Support Vector Regressor (SVR) on the target variables of heating energy consumption (HEC), heating energy consumption per square meter of area (HEPI), and both log-transformed (log-HEC, log-HEPI).

We found that the RFR and XGB are the best fitting regressors and that it does not matter so much on the target variable. We get leverage when we divide the buildings into more homogeneous groups, such as residential buildings and old residential buildings, thus reducing the data variance. We have shown that this increases the accuracy of the models. This finding may be generalizable to other data science projects.

Based on our results, geoimpact can use the prototype for modelling heating energy consumption, which can be further improved with additional information (additional features and observations). In addition, more extensive studies with higher geographic coverage can be conducted. For this purpose, we recommend including more information on occupant behaviour, climate conditions, and building characteristics, such as the insulation quality of facades, roof, windows, and their solar orientation.

Table of Content

List of Tables 4
List of Illustrations 5
List of Abbreviations........ 6
1. Introduction 8
1.1. Geoimpact and Swiss Energy Planning........ 9
1.2. Research Question and Hypothesis 9
1.3. Structure of the thesis 10
2. Background of Estimating Building's Heat Energy Consumption........ 11
2.1. Energy Consumption Discussion 11
2.1.1. Functional form 11
2.1.2. Available datasets for building heat energy estimations........ 12
2.2. Energy Quantification Methods (EQM) 12
2.2.1. Top-Down Approaches........ 12
2.2.2. Bottom-Up Approaches 13
2.3. Machine Learning Methods........ 14
2.3.1. Simple Linear Regressor (SLR)........ 15
2.3.2. Multiple Linear Regression (MLR) 15
2.3.3. Decision Tree Regression (DTR) 16
2.3.4. Random Forest Regression (RFR)........ 17
2.3.5. Gradient Boosting Regressor (GBR) and XGBoost Regressor (XGB) 18
2.3.6. Support Vector Regressor (SVR) 18
2.4. Performance Evaluation Measures 19
3. Methodology........ 20
3.1. Business Understanding & Benchmarking Problem 20
3.2. Data Understanding........ 20
3.3. Data Cleaning and Preparation........ 22
3.4. Modelling and Evaluation 22
3.5. Deployment 24
4. Results 25
4.1. Descriptive Analysis........ 25
4.2. Evaluate the best-fitting regressor and target variable 29
4.3. Evaluate prediction performance of building subsets........ 31
5. Discussion and Conclusion........ 33
6. Bibliography 35
7. Appendix 45
7.1. Scale effects of performance evaluation measures........ 45
7.1.1. Coefficient of Variation (CV)........ 45
7.1.2. Mean Absolute Percentage Error (MAPE) 45
7.1.3. Root Mean Squared Error........ 45
7.1.4. Mean Absolute Error 45
7.2. Results of Regressor Validation (on Val)........ 46
7.4. Results of Subset Validation (on Test)........ 47
7.5. Survey Questionnaire 49