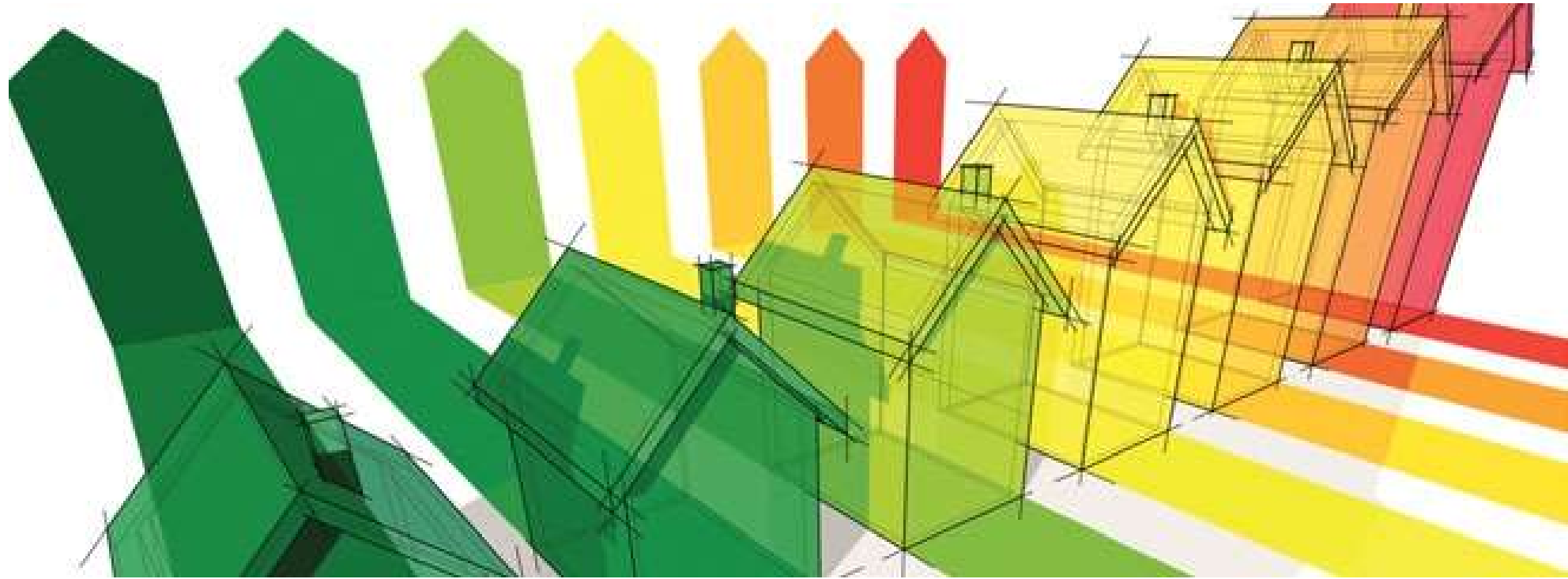


Energy Efficiency of Buildings Prediction



Presented by
Artem Ramus

Background

Source:

The dataset was created by Angeliki Xifara (angxifara '@' gmail.com, Civil/Structural Engineer) and was processed by Athanasios Tsanas (tsanasthanasis '@' gmail.com, Oxford Centre for Industrial and Applied Mathematics, University of Oxford, UK).

Heysem Kaya, Department of Computer Engineering, Boğaziçi University, TR-34342, Beşiktaş, İstanbul, Turkey Email: heysem '@' boun.edu.tr

<https://archive.ics.uci.edu/ml/datasets/Energy+efficiency#>

Introduction

Energy analysis was performed using 12 different building shapes simulated in Ecotect. The buildings differ with respect to the glazing area, the glazing area distribution, and the orientation, amongst other parameters. We simulate various settings as functions of the before-mentioned characteristics to obtain 768 building shapes. The data set comprises 768 samples and 8 features, aiming to predict two real valued responses.

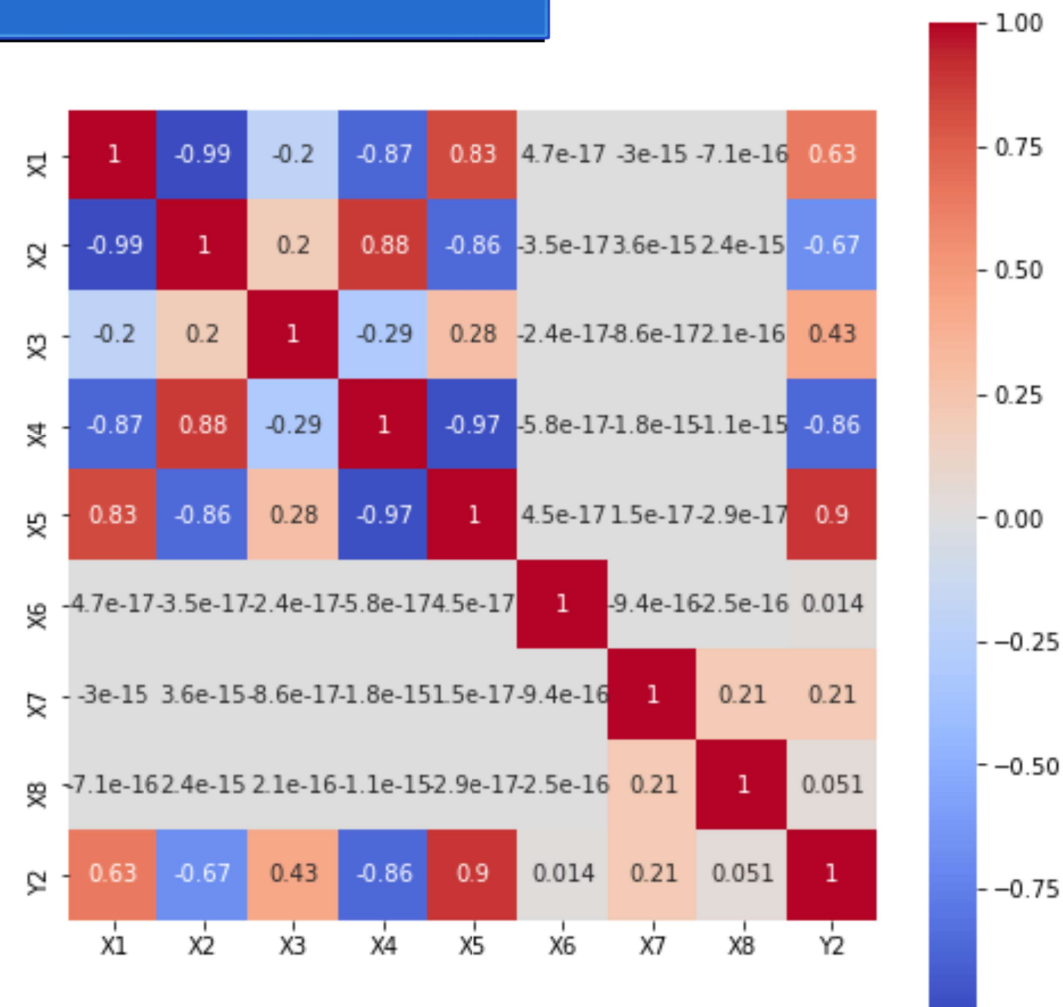
Citation Request:

A. Tsanas, A. Xifara: 'Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools', Energy and Buildings, Vol. 49, pp. 560-567, 2012 (the paper can be accessed from [\[Web Link\]](#))

For further details on the data analysis methodology: A. Tsanas, 'Accurate telemonitoring of Parkinsonâ€™s disease symptom severity using nonlinear speech signal processing and statistical machine learning', D.Phil. thesis, University of Oxford, 2012 (which can be accessed from [\[Web Link\]](#))

Methodology

- The data set is checked for duplicates, null values and homogeneous features.
- Data distribution is checked with histogram, distribution and box plots for skewness and outliers.
- Correlation between features was examined with a correlation matrix. Low-correlated features were removed.



Performance of the Model

Performance of the following models was estimated based on Pearson coefficient:

- Random forest
- XGBoost

Summary and Conclusions

XGBoost regression gave best R square score of 99%,
random forest - 97%

The end

Thank you for your attention!