

# Lab 5: Prediction Challenge

## Machine Learning for Economists, University of Cologne

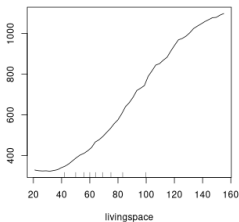
January 8, 2021

# Model Estimation Approach

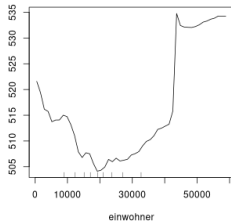
- Additional input used:
  - German federal states to differentiate between rents in East vs. West Germany
  - Number of inhabitants in the city to differentiate between rents in rural vs. urban areas
- Method: Random Forest with hyperparameters determined by grid search
  - no need for input standardization/transformation
  - automatically captures non-linear dependencies
- Procedure for hyperparameter tuning:
  - do a k-fold Cross-Validation for each hyperparameter combination
  - calculate the average RMSE
  - pick combination that minimizes the average RMSE
- RMSE of Estimated Model: 102.8578 (vs. RMSE of Benchmark model: 184.7283)

# Estimated Model - Partial Dependence Plots

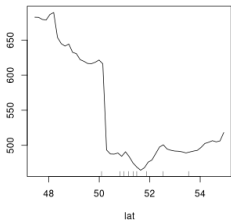
Partial Dependence on livingspace



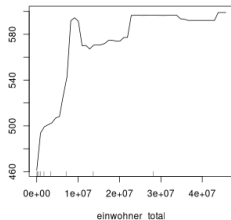
Partial Dependence on einwohner



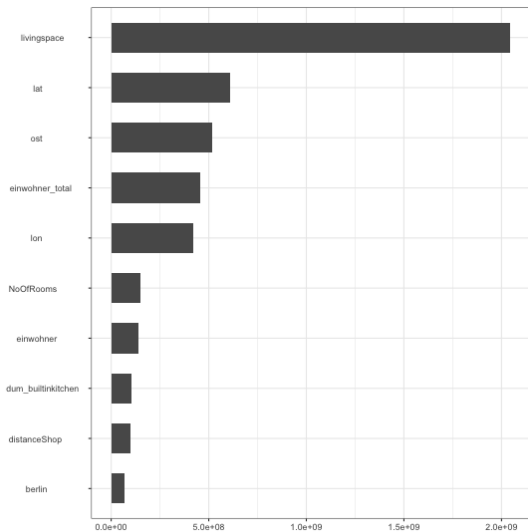
Partial Dependence on lat



Partial Dependence on einwohner\_total



# Estimated Model - Variable Importance



Comprehensive data and code to accompany our model estimation is available here: [https://github.com/FlyingKawasaki74/ml\\_lab5](https://github.com/FlyingKawasaki74/ml_lab5)