Paper Structure

Contents

1 Introduction

2 Methods

2.1 Preprocessing

2.1.1 Feature Engineering

Images

Reviews

- Description of Sentiment Analysis, stating procedure and results and including **Figure** with Wordcloud, either only English Words or Side-by-Side Wordclouds of English and Norwegian Words
- In addition: Language Detection to include the *number of different languages* and the *fraction of norwegian languages* and Analyzing the reviews lengths to include the *median review length*
- Since there are multiple reviews per apartment the results for each review were averaged for each apartment separately.

Others

- Optionally mention all other features that we added to the dataset
- All self-engineered features from images, from reviews and from existing metric variables were combined into a single dataframe as the foundation of all further analysis

2.1.2 Feature Selection

2.1.3 Price Distribution

- **Figure** of Side-by-Side Histograms of Price and Log-Price Distribution
- Price Distribution right-skewed with some very large outliers
- Explain benefits of normally distributed dependent variable, particular for models with distributional assumptions (e.g. Linear Regression vs. Neural Network)

- State that all classical Machine Learning Models benefitted from log transformation
- Briefly discuss why we did not transform the price variable for the Neural Network

2.2 Models

2.2.1 Classical Models

- serve as benchmark models to better evaluate performance of custom neural network
- selected with increasing degrees of complexity and corresponding decreasing degree of interpretability
- Focus on 4 models: `LinearRegression`, `Ridge`, `RandomForest` and `HistGradientBoosting`
- Describe Model Fitting process and hyperparameter tuning with Randomized Search Cross Validation

2.2.2 Neural Network

3 Results

3.1 Predictive Performance

- **Figure** of performance comparison between selected classical models and neural network for given feature selector (e.g. `RFE`) and different number of selected features
- Interpret Differences in Training and Validation Performance between different models
- Interpret Differences in Performance for different number of selected features
- Compare Performance on Validation Set with Performance on Test Set for the best model of each class by means of a table
  $\Rightarrow$ Models whose hyperparameters were tuned on validation set gen-

eralize worse to test set, e.g. `HistGradientBoosting`, `RandomForest` and `Ridge`

- Include average predictions of top 2/3/4/5 models, where models are selected based on validation set performance and Test Set predictions are averaged
- Potentially mention which models contributed to predictions on new, unseen dataset from challenge

3.2 Explainability / Interpretability

3.2.1 Feature Importance

- **Figure** of Coefficient Plot for Linear Regression with e.g. 25 selected features
- Interpret Figure

3.2.2 Sensitivity of Neural Network Performance on Outliers

- State shortcomings of Neural Net to predict prices in the tails of the distribution, error metrics thus largely impacted by outliers
- State (maybe with a table) drastic increase in predictive performance when excluding largest quantiles of price distribution from dataset
- Discuss if the task itself is theoretically feasible for any kind of model
- **Figure** of latent space representation
- Discuss that the data is not expressive enough to capture all features that determine the price in reality, particularly for apartments with very high prices, that do not differ from lower priced apartments based on their feature set
- Question underlying assumptions that all apartments are reasonably priced, difficult to detect overpriced listings that bias model predictions

4 Conclusion

5 Appendix

- include link to repository with codebase to reproduce all findings

6 References