

CMPT 459 Data Mining
Spring 2020
Martin Ester
TA: Ruijia Mao and Ruchita Rozario

Course project: Final Report

In this course project, we have used the Rental Listing Inquiries dataset from Kaggle. The main goal of this project was to predict how popular an apartment rental listing will be on the basis of available data. In this final report, you will present your work and your results in context and share your lessons learnt.

Note: You are not restricted to present the work that you have done for milestones 1, 2 and 3, but you can improve the classifiers developed for the milestones using the insights gained later.

Your report should contain three to four pages of text (12 point font, one-inch page margins), not including figures and tables, per group member.

Please, make sure that your report clearly states which group member wrote which section.

The report should contain the following sections:

1) Exploratory data analysis

- Distributions for the most important attributes
- Listing trends over time

2) Data pre-processing

- Dealing with missing values, outliers
- Feature extraction from images and text
- Additional features from external datasets
- Feature selection

3) Classifiers

- Choice of classifiers, including choice of libraries
- Optimization of classifiers
- Accuracy in cross-validation
- Accuracy on Kaggle
- Comparison of the results of the different methods

4) Lessons learnt

- Which features were most relevant, and why?
- If you created additional features from external datasets, how much did they help?
- Which classifiers worked best, and why?
- Which classifiers were more efficient to train?
- Was overfitting a problem? If so, how did you address it?

5) Recommendations for rental property owners

Based on the relevance of features for classification, which recommendations can you give to an owner of a rental property?

- When planning a new rental property, what properties should it have to maximize the interest of potential renters?
- When promoting an existing rental property, which properties should one highlight?

6) References

List all the resources that you have used, such as software libraries, blogs, technical reports.

Marking scheme (out of 100 marks)

Questions	Marks
1) Exploratory data analysis	10
2) Data pre-processing	10
3) Classifiers	20
4) Lessons learnt	30
5) Recommendations for rental property owners	20
6) References	10

Please submit on Coursys: A pdf-format report named Report.<GroupName>.pdf

Deadline: April 2, 2020