Investment Property Tool

Sai Dharmayogan , Anish Abraham

Overview

- Compare different methods to effectively classifly property descriptions into
  - Suitable for investment purposes (add value)
  - Ready for occupation

- Training data obtained manually from
  - Zoopla (UK)

Tool Setup

- Software dependency
  - Python 3
  - Jupyter notebook
  - metapy
  - pytoml
  - Numpy
- Code location
  - https://github.com/SDharmayogan/investment_property
- Instructions
  - https://github.com/SDharmayogan/investment_property/blob/master/README.md

Classification

- Challenges
  - Small training data (30 with 15 in each)

- Techniques compared
  - Naïve Bayes
  - SVM
  - Logistic Regression

Retrieval

- Techniques compared
  - BM25
  - Dirichlet Prior
  - Jelinek Mercer
- Optimal setting
  - Dirichlet Prior (Mu = 25)

Conclusion

- Retrieval displays better accuracy / precision than classification
  - Dirichlet Prior Retrieval: (MAP): 0.879
  - SVM (Avg Accuracy): 0.816
  - Logistic Regression (Avg Accuracy): 0.783
  - Naive Bayes (Avg Accuracy): 0.773
- Language model not necessary as terms used to identify suitable documents are more effective in a search