

Abstract

Building unsupervised Natural Language Processing (NLP) machine learning models to predict whether a business review text is positive or negative. Also, assigns topics (clustering) based on the raw text data to find out the business domains and implement a recommendation system.

Methodology

- EDA (cleaning and visualizing the data).
- Topic models.
- Recommendation system.
- Classification models.

Data

Two separate datasets were imported from [Yelp](#) in json format. The first data contains the user's review text the other one contains the business names. The two datasets were merged and a sample of 1500 observations was selected.

Tools

- Jupyter Notebook
- Seaborn
- Matplotlib
- Numpy
- Pandas
- Sklearn
- Pickle
- Nltk
- Regular Expression

Communication

The slides are provided [here](#).