ÁLVARO PARAFITA

SENTIMENT ANALYSIS WITH GLOVE

TRIPADVISOR RESTAURANT REVIEW DATASET

▸ Barcelona Restaurant Reviews:

  ▸ 7,343 restaurants

  ▸ 861,591 reviews (text + 1-5 rating)

▸ Sentiment Analysis model:

  ▸ Input: text of a review

  ▸ Predict: rating/sentiment of a review
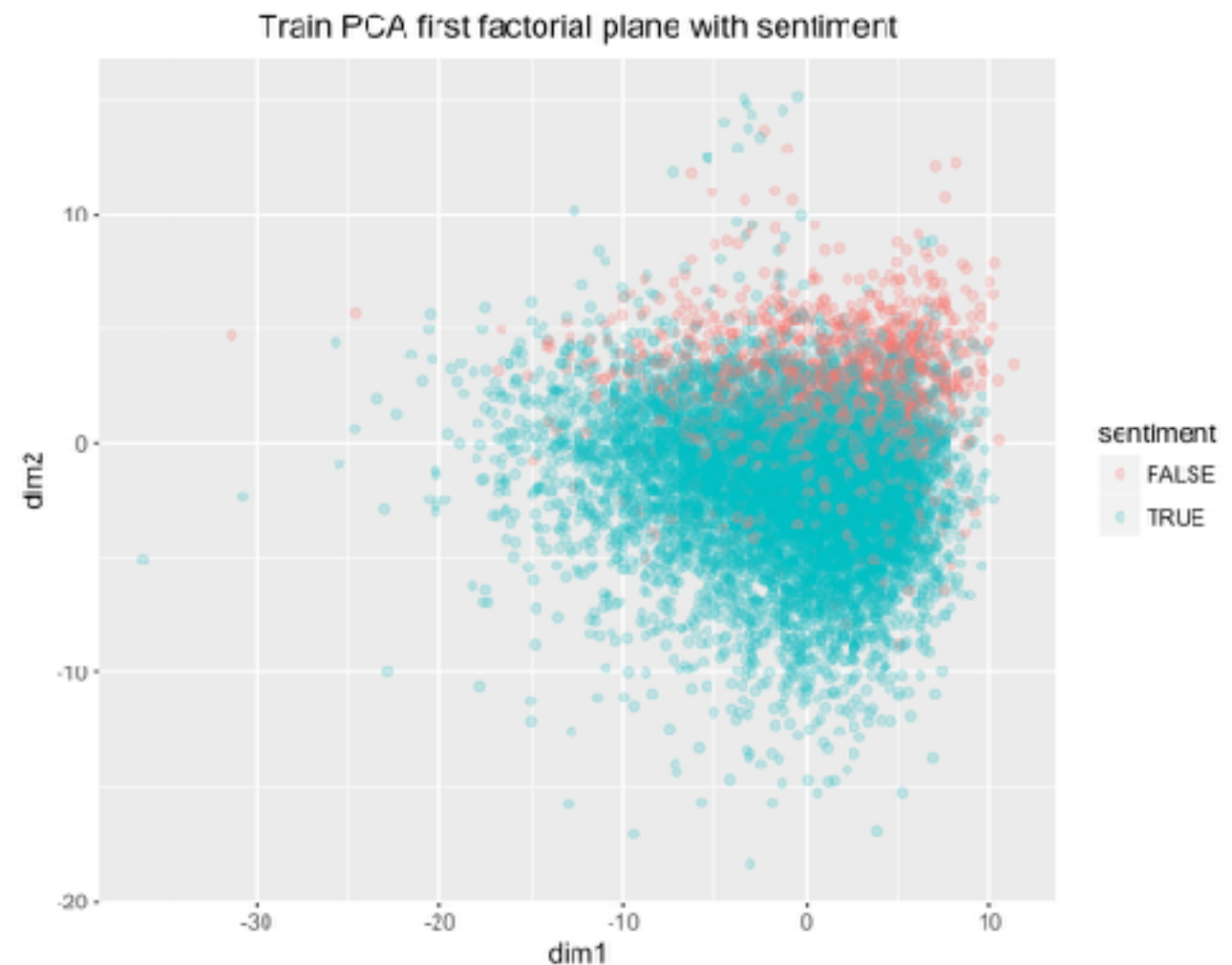
  ▸ Model: LDA / QDA / SVM

tripadvisor®

ENGLISH DATASET

▸ Language Classifier: English/Not English

▸ Input of the classifier: trigrams of characters

▸ Random Forest classifier: 100 trees and max. depth of 4

▸ Results on a test set: 100% accuracy

▸ Resulting dataset: 151,014 English reviews

WORD & DOCUMENT EMBEDDINGS

▸ Embeddings = numerical vectors that encode the meaning of words/documents

▸ GloVe: state of the art method to create word embeddings

    ▸ GloVe dataset: word-vector dictionary created by GloVe

▸ Document Embeddings: weighted average of GloVe word vectors (using TF-IDF as weights)

Train PCA first factorial plane with sentiment

RATING / SENTIMENT CLASSIFICATION

model	accuracy	MSE
LDA	**0.598**	1.018
QDA	0.531	1.471
SVM	0.553	**0.977**

Rating classification results

model	acc.	+prec.	-prec.	sens.	spec.
LDA	**0.932**	0.944	**0.845**	**0.978**	0.671
QDA	0.872	**0.965**	0.548	0.881	**0.819**
SVM	0.931	0.962	0.761	0.957	0.786

Sentiment classification results