

ADCG WS 2016/17 Challenge 02: Review Spam Data Set.

FILE NAMES

yelp_data_train.dat – training set, label in column named ‘fake’
yelp_data_test.dat - test set
yelp_data_reviewer.dat – data about reviewers
yelp_data_hotel.data – data about hotels

PURPOSE

The database consists of data about hotel reviews in Yelp.

PROBLEM TYPE

Binary Classification

SOURCE

The dataset comes from the paper:

Mukherjee, Arjun, et al. "What yelp fake review filter might be doing?." *ICWSM*. 2013.

Students can find many useful clues there. Also, this data is only allowed to be used in the scope of this exercise research.

SUBMISSION:

In submission, fake (spam) reviews should be given the label ‘1’, while true reviews should have label ‘0’.