



Data:

For training the classifier and building the customized embeddings we used an external dataset "Trip Advisor Hotel Reviews" from Kaggle.

The original data can be found in Kaggle:

<https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews>

The preprocessed (cleaned filtered and embedded) data can be found here:

https://drive.google.com/file/d/1zP0f_ojDudqZHO0ypDs4lsMr60Yah1eX/view?usp=sharing

We use the preprocessed data for training a small classifier:

<https://drive.google.com/file/d/1k0zEft0s0sZxvE7oag1FuiNHMvmxWWmH/view?usp=sharing>

Preprocessing code:

https://drive.google.com/file/d/1wS2d_yrpm4iLwzGBfdSPtkHi_NHJ6C22/view?usp=sharing