

igorpejic / personalize_expedia_hotel_searches_2013

 github.com/igorpejic/personalize_expedia_hotel_searches_2013

igorpejic

This is an implementation of a model which achieves 4th (NDCG@38=0.534) place on: Personalize Expedia Hotel Searches - ICDM 2013 challenge
<https://www.kaggle.com/c/expedia-personalized-sort/overview>

First, you need to install the packages:

```
pip3 install -r requirements.txt
```

Afterwards you can run the training and generation of the submission file like this:

```
python3 run.py /home/i/data/train.csv /home/i/data/test.csv  
/home/i/results_directory/
```

The results (in kaggle submission format) will be saved to `submission.csv` and the trained model to `model.dat`

The process of training will take long (on VM with 32 vCPU it took 2.5 hours with the default parameters). To make training faster, but less accurate change 'dart' to 'goss' and decrease `n_estimators`.

It is recommended to have 8GB RAM free otherwise you might run into memory issues.

This project was done in the scope of the course Data Mining Techniques taught at Vrije Universiteit, Amsterdam in 2019. Special thanks to team members @P4ppenheimer and @ijanerik.