

Основные шаги:

- Пока бигдата готовит входной файл для StarSpace, команда Data Science:
 - Играется с [ноутбуком-примером](#)
 - Изучает [эмбединги](#)
 - Читает про [StarSpace](#)
 - Пытается понять как лучше сделать рекоммендер
- Как только входной файл для старпейса готов
 - Зайти на инстанс через Putty
 - Закачать входной файл от бигдаты
 - `aws s3 cp s3://путь_в_s3 starspace_input.tsv`
 - Закачать шаблон рекоммендера
 - `sudo apt-get install zip`
 - `aws s3 cp s3://nonauth-friendsday-kiur1wee/common/data_science/recommender_template.zip recommender_template.zip`
 - Скомпилировать StarSpace
 - `git clone https://github.com/facebookresearch/StarSpace`
 - `wget https://dl.bintray.com/boostorg/release/1.63.0/source/boost_1_63_0.zip`
 - `unzip boost_1_63_0.zip`
 - `sudo mv boost_1_63_0 /usr/local/bin`
 - `make`
 - Обучить вектора
 - `./starspace train -trainFile 'starspace_input.tsv' -label 'public' -trainMode 1 -dim 64 -model 'starspace_output' -negSearchLimit 300 -maxNegSamples 10`
 - Написать функцию рекоммендера (либо напрямую, либо через KNN)
 - `recommender_template -> recommendation_engine.py -> get_recommendations_from_publics`
 - Оценить качество рекоммендера функцией `evaluate` на сервере оценки
 - `http:// ec2-3-120-132-157.eu-central-1.compute.amazonaws.com:8080/evaluate?team_name=%ИМЯ_КОМАНДЫ%&endpoint=http://%АДРЕС_СЕРВЕРА%:6767/endpoint`
 - Надо помнить что есть таймаут на выполнение запроса– 5 минут
- Опционально если есть время
 - Взять 100% датасет (все данные)
 - Потюнуть параметры starspace
 - Использовать другие алгоритмы по желанию