Restaurants antirecommender

by Olga Illarionova




















Where go to lunch? Or do not go

- Reading all reviews takes a lot of time
- Different people care about different things

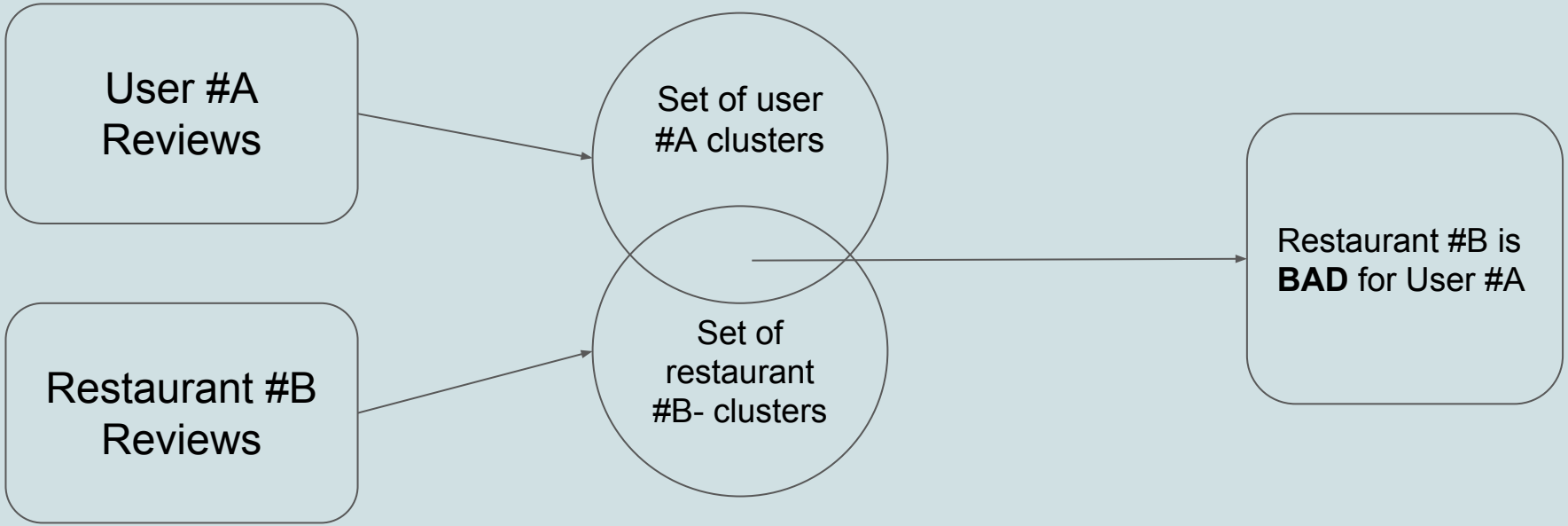
Methods

- Baseline4 - if overall rating less than 4 - stay away of this place
- 1st approach: tried to create recommender with ALS model from pyspark - work almost the same as baseline
- Use NLP and KMeans to cluster review
- Use Random Forest to predict cluster of restaurants

Cluster bad reviews to find hidden features

1. Filter bad reviews (1 & 2)
2. TF IDF
3. Cluster train set
4. Now we can cluster all bad reviews

Assign cluster to user/ restaurants and predict

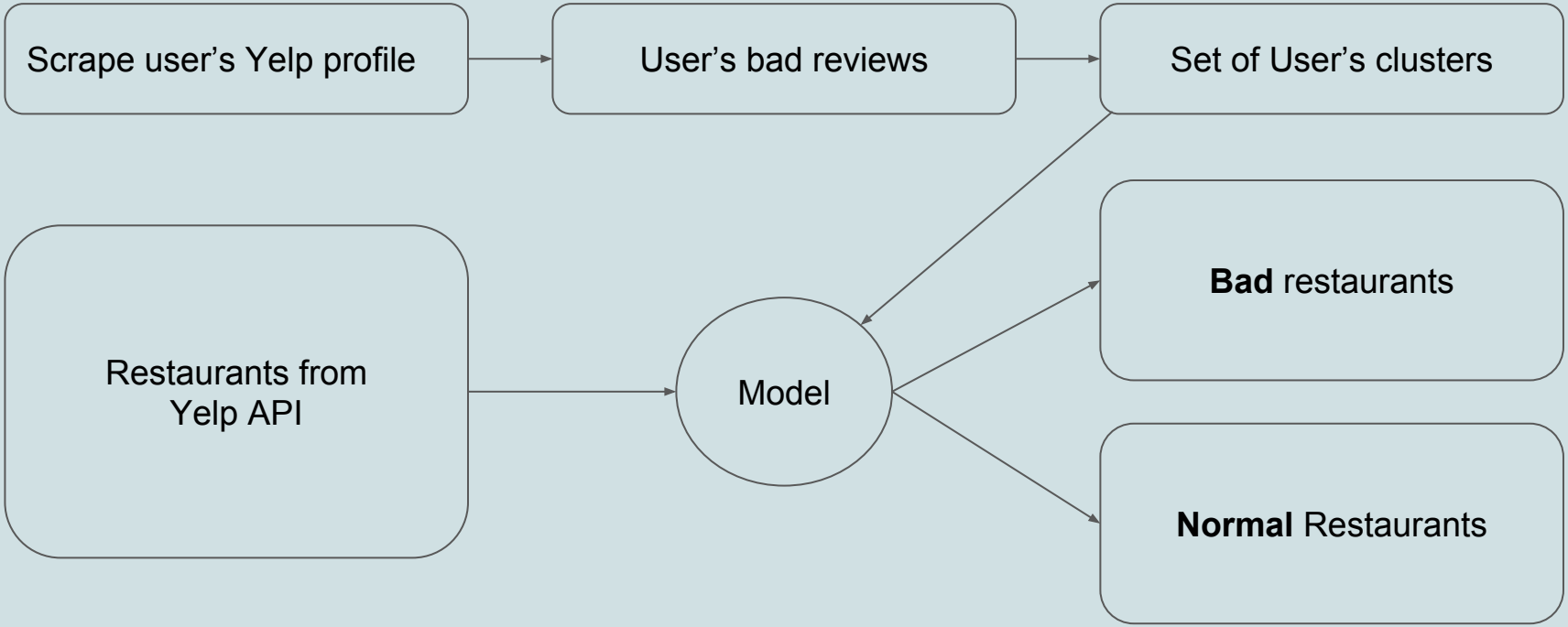


Restaurant cold start

- For every cluster need to predict can it be assigned to this restaurant

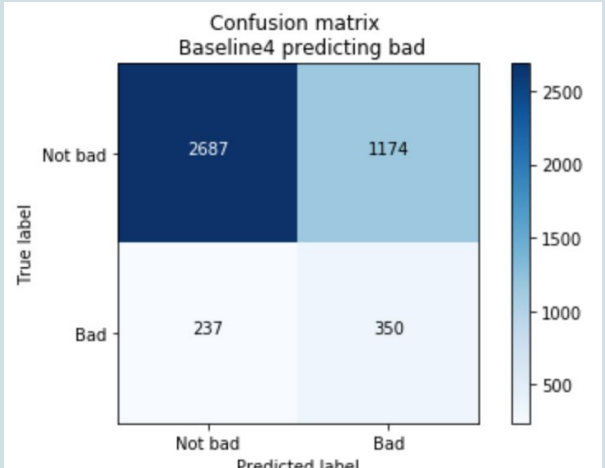
- Train RandomForest for every cluster

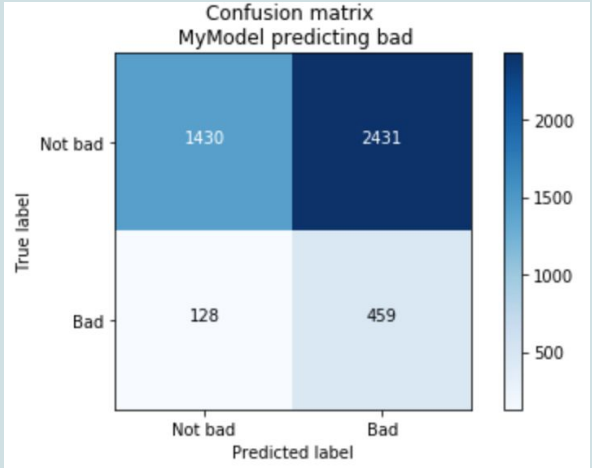
Final algorithm



Validate on real data

- Metric : recall (True positive rate)
- Compare with Baseline4 - if rating less than 4 then predict that restaurant bad
- 0.78 vs 0.59





Further improvements

- Website deployment
- Try different clustering algorithms
- Retrain on whole training dataset
- Tune models parameters
- Tried to find core features for reviews clusters

Thank you!

Olga Illarionova
ola.illari@gmail.com
Dec 2018