NEW USER BOOKINGS

FOR AIRBNB

TING SIT

AUGUST 2020

OBJECTIVE

- Predict in which country a new user will make his or her first booking.

- Benefits:
  - Allow more relevant marketing to new users, which drives booking conversion
  - Better user experience with posting more aligned with users' preference

- Model Evaluation requirements:
  - NDCG (Normalized discounted cumulative gain)

MODEL CONSIDERATION

- Mutli-label classification problem: target variables = “DEST”
- 5 models: SGD Regression, Random Forecast, Extra Tree, Lightbgm, Keras Deep Learning
- Features:
  - 126 categorical variables, coming from train_df and session datasets

RESULTS

Model	NDCG	Time
dummy classifier	0.66094	
extra trees	0.85212	1m 42s
extra trees (tuned)	0.86487	8m 49s
random forecast	0.85788	1m 7s
random forecast (tuned)	0.85359	25s
lightgbm	0.86849	22s
lightgbm tuned	0.86769	41s
Deep learning (keras)	0.86891	44s

- Lightgbm (no tuning) and Deep learning both have competitive performance
  - Recommend lightbgm as the winning model for this problem

CONCLUSION/NEXT STEPS

1. Lightbgm has the highest NDGC as the among base regression models
2. Dataset is unbalance with destination being ~70% US, hence most regression models produce at least 0.8 NDGC scores. Perhaps analysis could be done at a more micro level within US to make the predictions more precise – need extra data points from Airbnb.
3. Deep Learning model (Keras) showed slight improvement from base regression models. Further model development in deep learning area could be beneficial
4. More features can be extracted from session data since only last action information is being considered in current analysis