## Модуль 3. Фидбек

#### *Результат модели*

Первое за, что хочу Вас похвалить MAE: 0.17100. Отличный результат! Это значит, что Вы проделали настоящую исследовательскую работу и вполне успешно. Мои поздравления!

#### *Комментарии*

Отлично, что получилось использовать внешние источники данных. Также очень хорошо, что получилось добавить новые признаки, при этом они ещё и оказались очень полезны для модели в итоге. В целом, удачно и грамотно проведен feature engineering.

Для one-hot encoding признаков, содержащих множественные значения можете попробовать в дальнейшем MultiLabelBinarizer,   
<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MultiLabelBinarizer.html>

Для кодирования переменных по аналогии с Price Range можно использовать <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>

*Как сделать лучше*

Возможно, стоило, попробовать PCA (либо другое снижение размерности) - может какие-то признаки удалось бы оттуда достать - при этом можно как переходить в новое пространство признаков, так и добавлять полученные новые признаки к старым (это делать аккуратно - не совсем математически обосновано, но иногда работает). Возможно стоит попробовать методы feature selection для уменьшения количества используемых признаков (примеры методов можно посмотреть тут <https://scikit-learn.org/stable/modules/feature_selection.html>). Можно было попробовать расширить информацию о городах, добавив признаки близости ресторана к центру города или к основным достопримечательностям и т.д. Можно было попробовать добавить признаки на основе кластеризации объектов обучающей выборки - значением признака выступало бы отношение объекта к тому или иному кластеру. Также, очевидным направлением дальнейшего улучшения является анализ текстов отзывов на тональность (положительную и отрицательную), при этом не обязательно сразу использовать для этого все state of the art достижения в области natural language processing - возможно даже просто подсчёт количества положительных/отрицательных слов мог бы дать прирост. Кроме того, прирост качества, возможно, было бы получить добавив дополнительные данные (именно строки, а не признаки) - они могут быть получены либо из других датасетов/парсинга, либо (так делать очень аккуратно) сгенерированы искусственно, повторяя распределение обучающей выборки.

Это всё небольшие советы - вы отлично справились с заданием!

Отзыв подготовил ментор Леонид Саночкин.  
Если возникнут вопросы, обращайтесь ко мне в канал проекта **#0\_project\_3-о\_вкусной\_и\_здоровой\_пище** в Slack. Постараюсь ответить на ваши вопросы и разобраться с моментами, которые вызывают трудности.

Буду ждать вас на итоговый вебинар по данному проекту, который я провожу по четвергам в 20:00 МСК. На нём мы разберём основные ошибки и рассмотрим как улучшить решение. До встречи на вебинаре! (https://us02web.zoom.us/j/89031549953?pwd=NUpoMmFWenl6ekRkM09aSjcxWVRjZz09)

Удачи в обучении!