**Project – TripAdvisor Rating**

**Автор: Александра Гвоздева**

## Задание

Построение модели для прогнозирования рейтинга ресторана на сайте TripAdvisor.

**Датасет содержит следующие признаки**:

* City: Город
* Cuisine Style: Кухня
* Ranking: Ранг ресторана относительно других ресторанов в этом городе
* Price Range: Цены в ресторане в 3 категориях
* Number of Reviews: Количество отзывов
* Reviews: 2 последних отзыва и даты этих отзывов
* URL\_TA: страница ресторана на 'www.tripadvisor.com'
* ID\_TA: ID ресторана в TripAdvisor
* Rating: Рейтинг ресторана

**В результате работы над данными:**

1. В датасет были внесены дополнительно 15 числовых признаков, не считая тех, которые были получены в результате применения метода get\_dummies. Были внесены следующие признаки: 'Number\_of\_Reviews\_isNAN', 'price\_cat', 'number\_of\_restaurants', 'ranking\_norm\_rest', 'number\_of\_citizens', 'ranking\_norm\_citizens', 'number\_of\_reviews\_norm\_citizens', 'salaries','ranking\_norm\_salaries', 'number\_of\_cuisines', 'last\_review', 'time\_between', 'net\_size', 'ranking\_norm\_net\_size', 'number\_of\_reviews\_norm\_net\_size'. Самым важным из них при обучении модели оказался признак 'ranking\_norm\_rest'. Этот признак показывается ранг ресторана, нормированный на число ресторанов в данном городе.
2. При анализе новых количественных признаков применялись методы работы с выбросами и визуализация. Часто новые признаки содержали много пропусков. Заполнение их проводилось либо медианой, либо случайными значениями из уже имеющихся.
3. Дважды был применен метод get\_dummies (к колонке City и колонке restaurant\_category). Это увеличило общее число количественных признаков до 55.
4. Для Feature Engeneering использовались внешние источники данных (Википедия).
5. Расширение числа количественных признаков позволило получить значение метрики MAE, равное 0.2028.