

# Отчёт по контексту CTR prediction

## 1 Методы

Для предсказания кликов применялось два метода: логистическая регрессия и градиентный бустинг с оптимизацией гиперпараметров.

Логистическая регрессия была обучена по заданию, на тринадцати числовых и двух категориальных факторах для половины датасета. Использовалась регуляризация ElasticNet с коэффициентом 0.01 и долей l2-регуляризации в 0.1.

Градиентный бустинг обучался на тех же данных. Оптимизация выбрала гиперпараметры  $\eta = 0.5$ ,  $\text{num\_round} = 100$ ,  $\text{max\_depth} = 6$ ,  $\text{min\_child\_weight} = 100$ ,  $\gamma = 0.3$ ,  $\text{subsample} = 0.8$ ,  $\text{colsample\_bytree} = 1$ ,  $\alpha = 0.1$ ,  $\lambda = 0.1$ .

На валидационном датасете градиентный бустинг также запускался до оптимизации гиперпараметров. Тогда он показал gsauc 0.725, в то время как после оптимизации - 0.738. Таким образом, метрика улучшилась более, чем на один процент. Кроме того, проводилась оптимизация гиперпараметров логистической регрессии. С подобранными параметрами она показала gsauc 0.703, что является менее значимым улучшением.

## 2 Результаты

В таблице представлены значения метрики gsauc на тестовых датасетах контекста.

Метод	Public score	Private score
Логистическая регрессия	0.69506	0.69740
Бустинг + оптимизация гиперпараметров	0.73285	0.73461

## 3 Выводы

Были применены модели логистической регрессии и градиентного бустинга. Более высокий результат показал градиентный бустинг. Оптимизация гиперпараметров позволила улучшить результат примерно на одну сотую.