ФИНАЛЬНЫЙ ПРОЕКТ

Разработка рекомендательной системы для компании-ритейлера

ЗАДАЧА
Увеличение прибыли от до продаж в интернет-магазине, на 20 %

- РЕШЕНИЕ
- РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ
- БИЗНЕС ЦЕЛИ - УВЕЛИЧЕНИЕ ПРИБЫЛИ НА 20 %
- БИЗНЕС МЕТРИКА - PRECISION@3

ОПИСАНИЕ ДАННЫХ

events — датасет с событиями.

- timestamp — время события
- visitorid — идентификатор пользователя
- event — тип события
- itemid — идентификатор объекта
- transactionid — идентификатор транзакции, если она проходила

category_tree — файл с деревом категорий (можно восстановить дерево).

- category_id — идентификатор категорий
- parent_id — идентификатор родительской категории

item_properties — файл с свойствами товаров.

- timestamp — момент записи значения свойства
- item_id — идентификатор объекта
- property — свойство, кажется, они все, кроме категории, захешированы
- value — значение свойства

ВЫБРАННАЯ МОДЕЛЬ

LightFM

Это реализация на Python ряда популярных алгоритмов рекомендаций как для неявной, так и для явной обратной связи. Будем использовать LightFM не только из-за его гибридных возможностей, но и из-за того, что он позволяет использовать обычные взаимодействия пользователя с элементом для прогнозирования для известных пользователей. Это очень важно, поскольку у нас не было формальных рейтингов для элементов, мы все равно смогли получить рекомендации на основе этих взаимодействий. Поскольку у нас есть только неявная обратная связь от пользователей, гибкость LightFM в этом отношении полезна.

Для построения рекомендательной системы был использован датасет events с выборкой записей по наличию транзакции. В качестве **user** у нас будет выступать **visitorid,** в качестве **item – itemid.**

ПРОВЕДЕННЫЕ ЭКСПЕРИМЕНТЫ

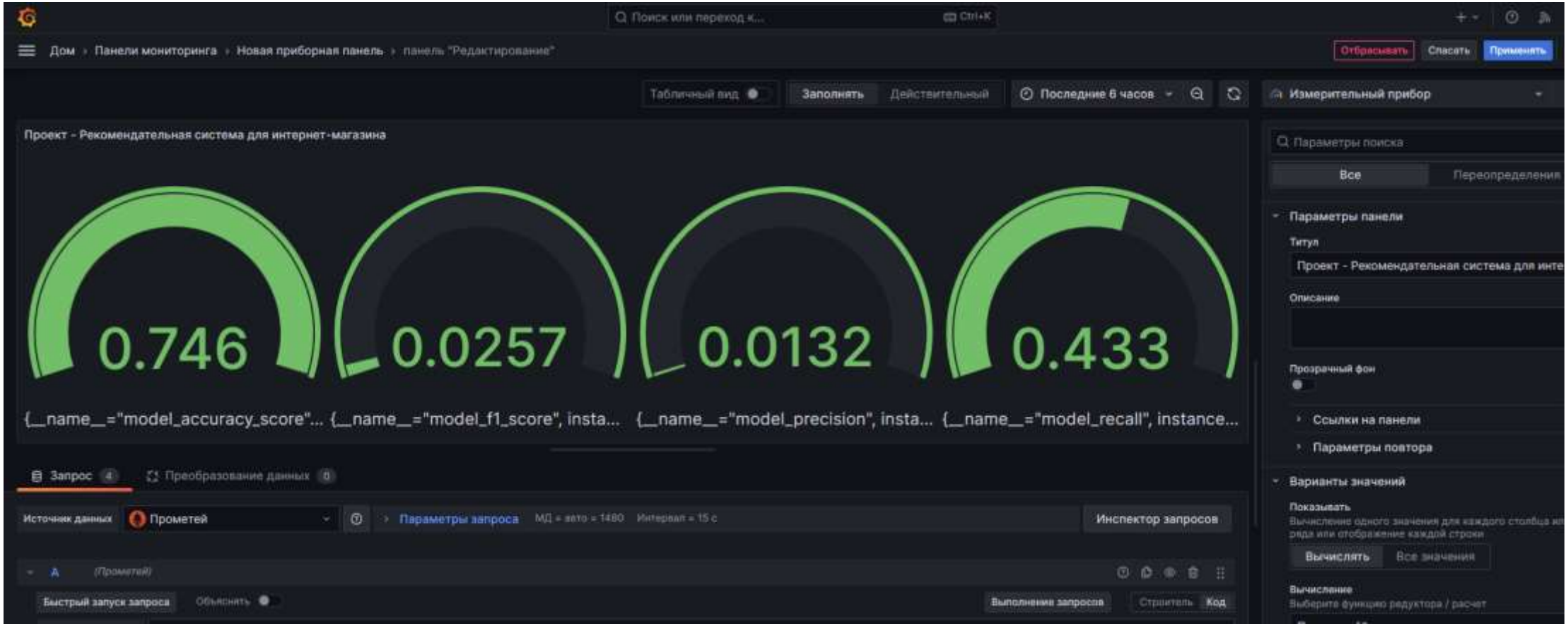
- КОЛОБОРАТИВНАЯ ФИЛЬТРАЦИЯ
- precision@3 = 0.21
- XGBOOST
- precision@3 = 0.0122

- CatBoostClassifier

- precision@3 = 0.0132
- ALS
- precision@3 = 0.13

LightFM = 0.76

МЕТРИКИ *В Grafana*



инструмент для сбора метрик Prometheus