

WB Tech: задачи для аналитиков-разработчиков

Общая информация

Наша цель — подготовить хороших специалистов, которые смогут решать сложные задачи в крупнейшем маркетплейсе.

Хорошо в учении — легко в бою: учебные задачи составлены так, что тот, кто с ними успешно справился, сможет осилить и боевые. Кто справится — того забираем ;-)

Пара советов:

1. Разрешается и приветствуется использование любых справочных ресурсов, привлечение сторонних экспертов и т.д. и т.п.
2. Можно и нужно задавать вопросы, как по условию задач, так и об их решении. Не стесняйтесь спрашивать все у всех в общем чатике.

Как делать задание

Надо взять задачу довести ее до продакшн-состояния.

Для задачи необходимо пройти и показать все этапы разработки ML-продукта:

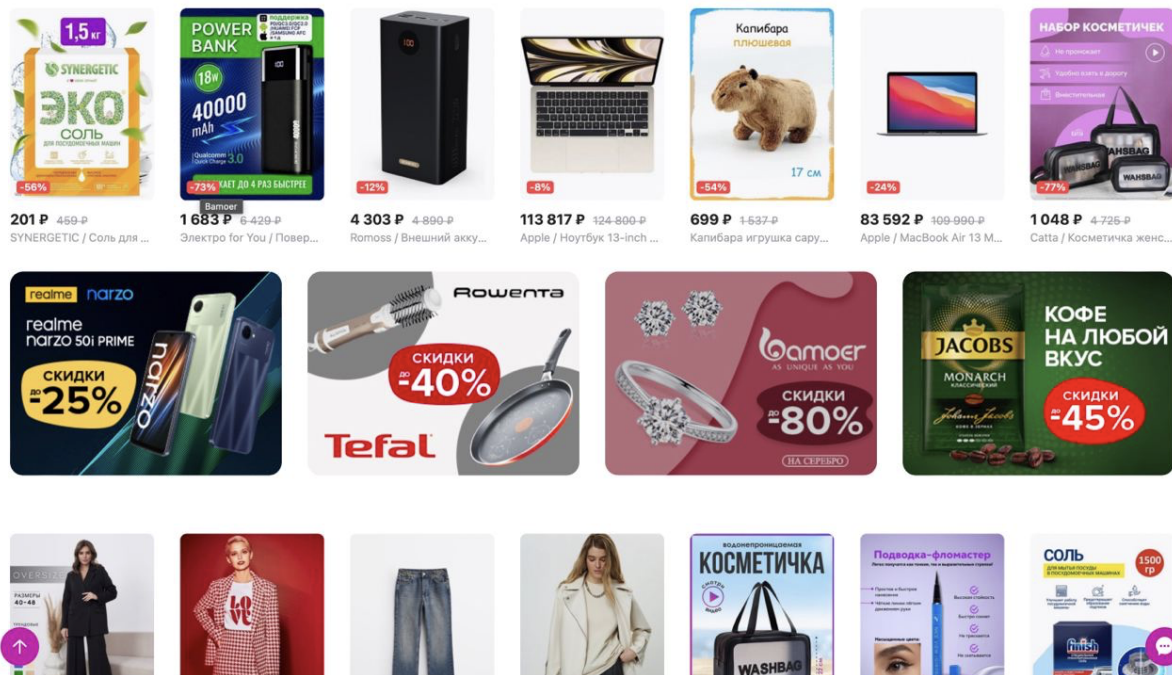
- проработка и формализация задачи, определение таргета;
- первичный анализ данных и их очистка;
- исследование проблематики, выбор подходящих решений;
- формирование обучающей выборки, проектирование валидации;
- разработка бейзлайна и реализация выбранного решения;
- оценка решения, оптимизация алгоритма;
- приведение решения в читабельный вид;
- вывод в продакшен.

Задача: персональные рекомендации

Персональные рекомендации — это тип рекомендаций, построенный на основе активности пользователя и подобранный индивидуально под него. На сайте wildberries их можно встретить на главной странице в блоке «Возможно вам понравится» (см. рис.).

Основная задача этой подборки — увеличение продаж через сайт и приложение.

Возможно, вам понравится



Задача

Необходимо на основе заказов пользователей сформировать предложение для его последующих покупок. Релевантность рекомендации будет оцениваться исходя из последующих покупок пользователей и того, насколько хорошо алгоритм позволяет их предсказать.

Данные

Предоставляем данные пользовательской активности (заказы) по некоторому срезу за определенный период в формате csv (см. файл data).

Результат

Итогом работы должен стать набор скриптов, который позволяет по полученному на вход идентификатору пользователя получить на выходе персональные рекомендации, а также сформированная выдача для всех тестовых пользователей. Должен быть оформлен frontend, где можно вбив в input идентификатор, получить список идентификаторов товаров.