

ML в модерации

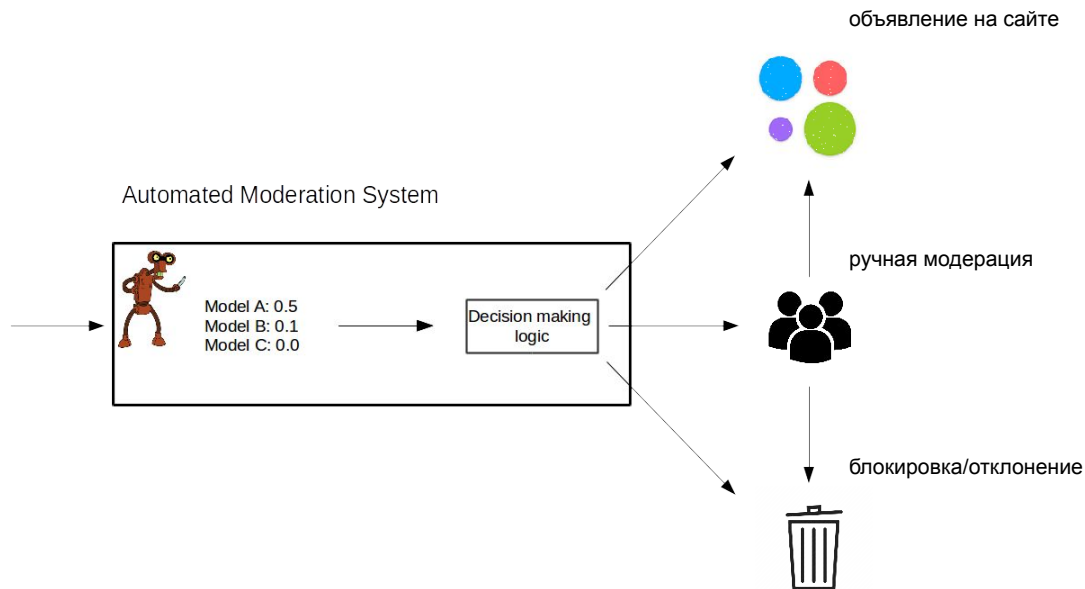
Павел Гладков

Unit Leader @ Automated Moderation

Зачем нужна модерация

- Поддержание порядка на сайте
- Снижение юридических рисков
- Защита интересов бизнеса

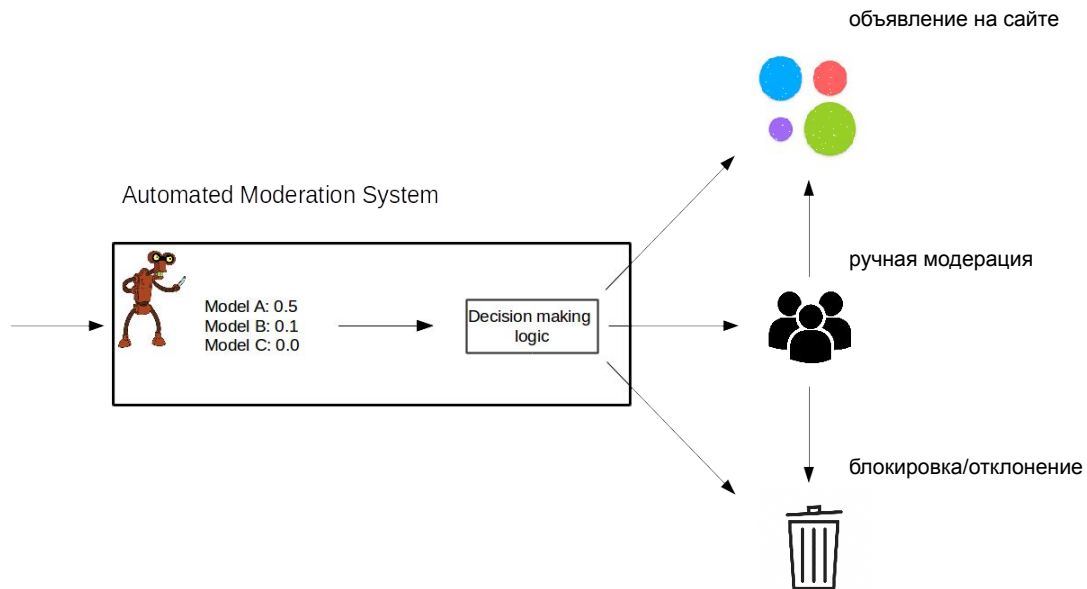
Модерация в Авито



Основные нарушения:

- Повторная подача объявлений
- Неправильная категория
- Запрещенный товар
- Контакты/ссылки на фото и в тексте
- Спам
- Другое

Автоматизация



Модели:

- Деревья решений
- Экспертные правила
- Линейные модели
- RNN, CNN

Пример: Повторная подача объявлений

Опорное 1331810560

Продаю iPhone 6s

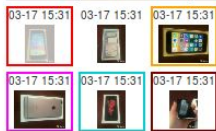
20 000 руб.

Армавир

vasialetkov@mail.ru

Объявлений: 7/0

2018-03-17 15:31:22



Категория: **Телефоны**

Вид телефона: **iPhone**

Продаю iPhone 6s на 16 gb, состояние абсолютно нового телефона. Ни одной коцки, или царапины. Отпечаток, 3 d touch все работает идеально. Зарядка и наушники в комплекте.

99%

iPhone 6s

20 000 руб.

Армавир

vasialetkov@mail.ru

Объявлений: 7/0

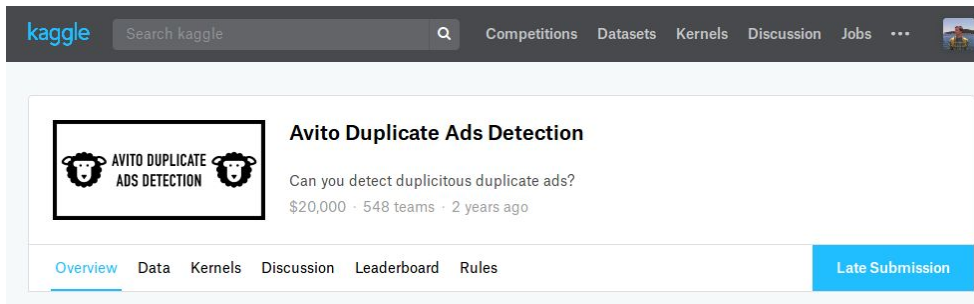
2018-03-16 11:30:57



Категория: **Телефоны**

Вид телефона: **iPhone**

Продаю iPhone 6s 16 gb, состояние абсолютно нового телефона. Все четко работает. Полный комплект: зарядка и наушники (которые не разу не использовались).



<https://www.kaggle.com/c/avito-duplicate-ads-detection>

Пример: Работа с картинками

– Классификация

см. на YouTube “AvitoNet: сервис компьютерного зрения в Avito”

– Поиск запрещенных товаров



[Обмен на скутер](#)

0 руб.

Транспорт / Мотоциклы и мототехника

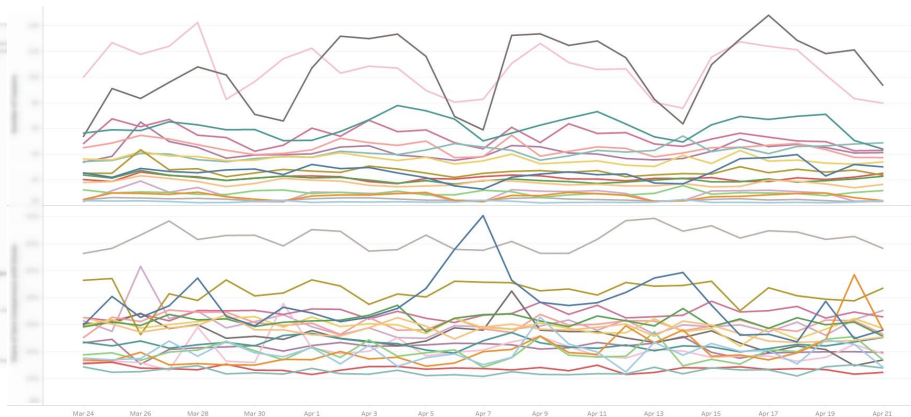
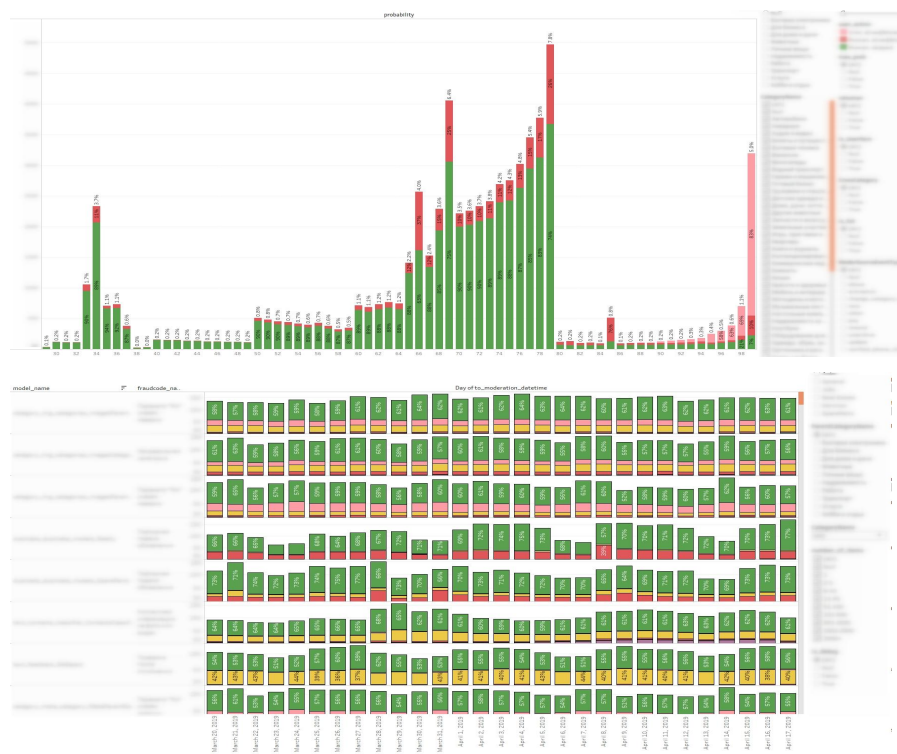
– Контакты на фото



- > 200 моделей
- > 10-20 обновлений каждую неделю

Как поддерживать?

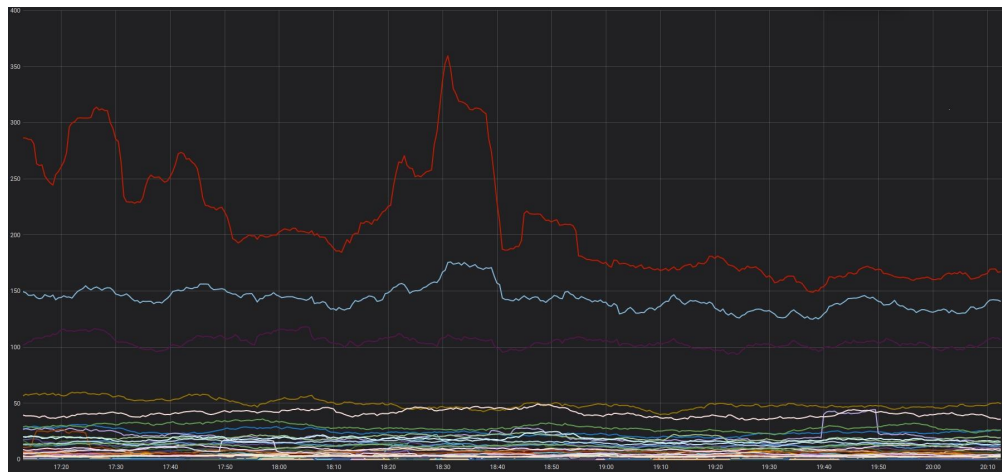
Аналитика по всем моделям



(очень много отчетов в Tableau)

Мониторинг

Realtime графики в Graphan'е



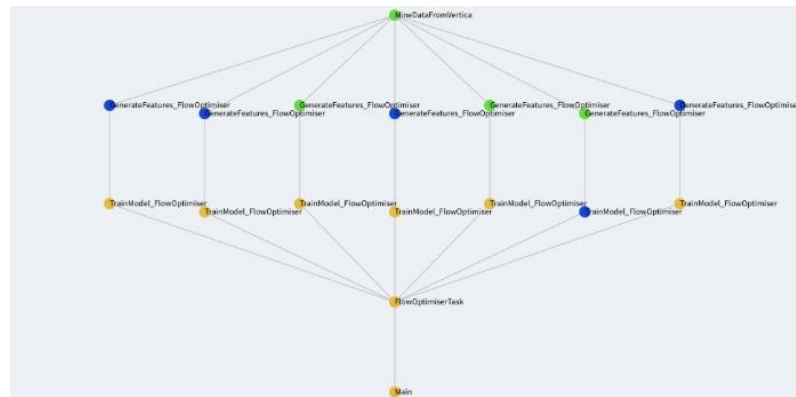
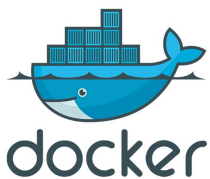
Offline бот в Slack



Воспроизводимость

Внутренний фреймворк Academy

- Весь код для переобучения в одном месте
- Обмен опытом, кодом
- Быстрое создание нужного окружения
- Легче итерироваться



Инструменты

Библиотеки: XGBoost, MXNet, Scikit-learn, Tensorflow и т.д.

Vertica — хранение и извлечение данных

Jupyter Notebook — для быстрых прототипов

Tableau, Grafana — для визуализации

Вопросы?

Павел Гладков

Unit Leader @ Automated Moderation
pgladkov@avito.ru