

ВК

ИИ по следам пользователей

Банда пяти 2.0

Наша команда



Даня
ML Dev



Саша
Analytic



Мила
Algorithmist/
Data engineer



Антоша
Python Dev



Артём
Social
Influencer

the problems



Низкая доля
положительного
класса



Необходимость
экстракции признаков



Пропуски и
шумы



Падающ
VK Clo

data preprocessing



Экстракция признаков: статистика по каждому юзеру и временным интервалам



Детекция пропусков и шумов



Приведение токенов к однородному виду (транслит + приведение к норм. форме)

our solution



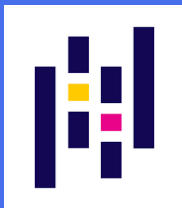
Стек

Pandas — для обработки

Numpy, Matplotlib — для
визуализации

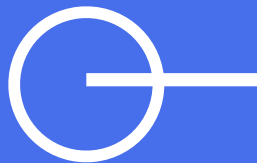
Sklearn — для построения
моделей

Pymorphy2 — для анализа
тегов

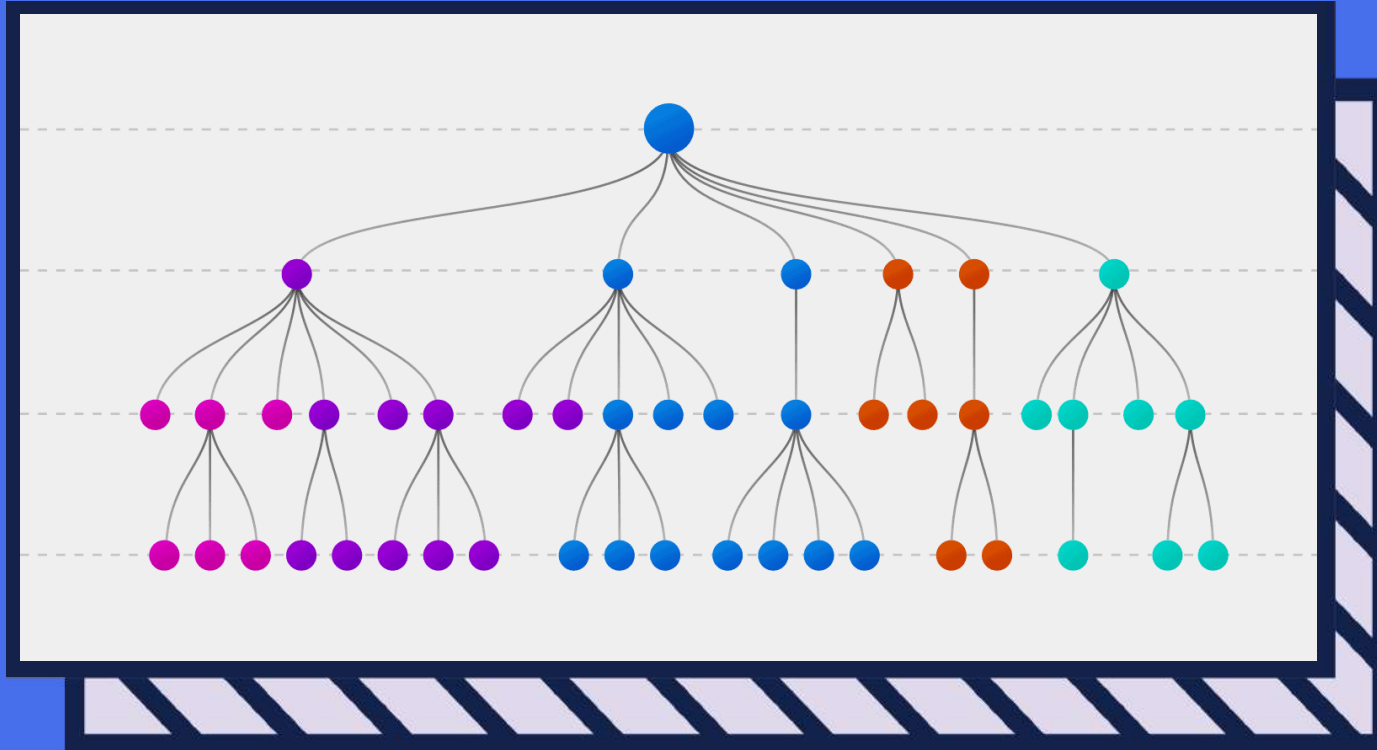


Score:
99.98

ROC AUC

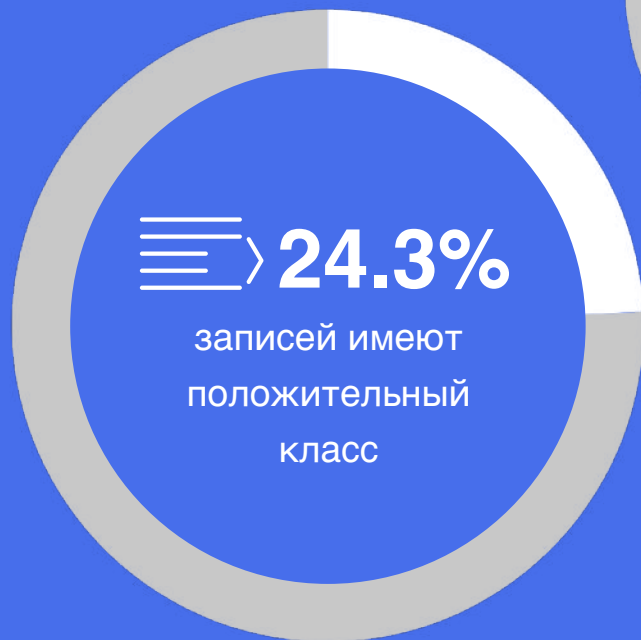


decision tree





the facts



>  **0.23%**
невалидных
записей

Самые популярные токены:
новости, займы, кредиты,
смотреть, купить

Самые непопулярные токены:
фамилии,
токены с грамматическими
ошибками,
жаргонизмы



02



applications



Реклама



Тест идей для функционала сайта



Рекомендательная система

thank you

for your
consideration



band of five 2.0



Если хочешь быть королем джунглей, мало просто вести себя как король, нужно быть королем. И никогда нельзя сомневаться, ведь сомнения порождают хаос и ведут к краху. Этому научила меня моя королева.

— Mickey Pearson

БК