

Предсказание ухода слушателей MOOC курса на платформе Stepic

выполнил: Рауф Курбанов
руководитель: Николай Вяххи

Введение

- Хотим знать, кто не окончит курс
- Дать преподавателю возможность адаптироваться под слушателей
- Обнаружить скрытые проблемы в курсе
- Найти желаемую планку по сложности

Постановка задачи

- Спроектировать модель для классификации слушателей в зоне риска
- Реализовать django app для предсказания ухода слушателей с курса
- Попасть в релиз

Существующие решения

- Likely to stop? Predicting Stopout in Massive Open Online Courses

Colin Taylor, Kalyan Veeramachaneni, Una-May O'Reilly

- компендиум статей
- нельзя переиспользовать напрямую

Примеры признаков ухода

- Зависящие от курса
 - продолжительность
 - автор
- Зависящие от времени
 - доля успешных попыток в решениях
 - количество параллельно проходимых курсов
- Не зависящие от времени
 - как давно пользуется платформой
 - сова или жаворонок

Модель

- Ансамбль решающих деревьев
- Обучается автоматически

- precision 0.68
- recall 0.74
- F1 score 0.68

Технологии

- Python 3
- Django, Numpy, Scipy, Pandas, Scikit-learn
- Vagrant

Результаты

Что получилось

- Реализован django app для предсказания ухода слушателей
- Реализован сбор данных о пользователях
- Интеграция в backend

Что не получилось

- Попасть в релиз
- Протестировать на актуальных данных
- Вынести во frontend

Что узнал

- Опыт работы в большом проекте
- Изучил ряд популярных фреймворков
- В курсе последних публикаций в области dropout'a