

**Прогнозирование шанса успешного завершения онлайн курса
на основе образовательных данных и алгоритмов машинного обучения**

Лапин.А.В. (ИТМО)

**Научный руководитель – кандидат технических наук, доцент Болдырева Е.А.
(ИТМО)**

Введение. Рынок образовательных онлайн курсов в России стабильно растет. Онлайн курсы, как правило, накапливают данные о студентах: количество просмотренных видео, баллы за тесты, время, потраченное на курс, тип устройства, с которого ученик проходит обучение. Установление зависимости между этими данными и успеваемостью имеет большое практическое значение. С одной стороны, уведомление ученика о вероятности завершения им курса на текущий момент времени может дать ему возможность здраво оценить свои шансы и повысить мотивацию довести своё обучение до конца. С другой стороны, позволяет авторам и преподавателям курса провести анализ текущей и потенциальной успеваемости, оценить ключевые характеристики и разделы, которые оказывают влияние на результат обучения.

Основная часть. Данная работа посвящена разработке и применению моделей машинного обучения и нейросетей на основе алгоритма градиентного бустинга CatdBoost[1] для прогнозирования успешности ученика онлайн-курса (какова вероятность, что ученик завершит этот курс). В рамках исследования был проведен систематический поиск и анализ доступных датасетов данных обратной связи от прохождения онлайн-курсов. Отобранный датасет был использован для обучения различных алгоритмов машинного обучения: логистической регрессии, k-nn, градиентного бустинга над решающими деревьями и полносвязной нейросети MLP[3]. Для демонстрации функциональности разработанной модели был создан веб-интерфейс, позволяющий вводить данные об обучении на онлайн курсе и получать прогноз. Модель машинного обучения интегрирована в веб-интерфейс с помощью Python фреймворка Fast Api[2].

Выводы. Разработана ML модель для прогнозирования вероятности успешного прохождения студентов онлайн-курса на основе данных обратной связи. Разработан веб-интерфейс для демонстрации возможностей модели. Проведен анализ возможности использования таких моделей на практике. Выяснилось, что ML модели способны давать предсказания с точностью превышающую 0.9. Однако, встречаются ситуации, при которых модели машинного обучения, предсказывают ученику меньшую вероятность успешного завершения курса после прохождения им дополнительного материала, чем вероятность, которую модель предсказывала до прохождения этого материала. Таким образом, модели прогнозирования на основе образовательных данных можно использовать на практике, благодаря их высокой точности, но обязательно нужно обрабатывать описанный выше эффект, чтобы избежать проблем с мотивацией у учеников.

Список использованных источников:

1. Дорогуш, А. В., Ершов, В. В., & Гулин, А. И. (2018). CatBoost: градиентный бустинг с поддержкой категориальных признаков. препринт <https://arxiv.org/abs/1810.11363> (Дата обращения: 24.02.2025).
2. Рамирес, С. Документация FastAPI: высокопроизводительный, легкий в изучении, быстрый в разработке, готовый к использованию в продакшене. Получено с <https://fastapi.tiangolo.com/> (Дата обращения: 24.02.2025).
3. Круглов, В. В., Дли, М. И. Интеллектуальные информационные системы: компьютерные нейронные сети. – М.: Физматлит, 2000.