

УДК 004.6

АНАЛИЗ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ В СИСТЕМЕ ЭЛЕКТРОННОГО ОБУЧЕНИЯ

Антипова Е.С. (федеральное государственное автономное образовательное учреждение высшего образования Университет ИТМО)

Научный руководитель – к.т.н., доцент Волчек Д. Г.

(федеральное государственное автономное образовательное учреждение высшего образования Университет ИТМО)

В данной работе на основе файлов журнала активности пользователей онлайн-курса на платформе «Open edX» проводится прогнозирование итоговых баллов студентов с помощью таких методов машинного обучения, как gradient boosting и случайный лес. Также рассматривается проблема списывания студентов в онлайн-курсе при условии невозможности контроля со стороны преподавателя.

Введение. Учебные заведения стали все чаще внедрять в свои образовательные программы массовые открытые онлайн-курсы (МООК). Это связано, например, с желанием повысить престиж университета, а также в условиях сегодняшней пандемии это может быть просто необходимо. Онлайн-курсы являются удобным инструментом для приобретения новых знаний. Однако, у такого формата обучения есть и свои недостатки. Одним из таких недостатков является сложность отслеживания преподавателями самостоятельности сдачи контрольных мероприятий обучающимися.

Исследуя поведение студентов во время прохождения курса, можно выявить группу студентов, которые не изучали материалы и прибегали к нечестной сдаче работ. Важно выявлять такие случаи, так как они приводят к снижению уровня компетентности будущих выпускников. Также, исходя из траектории прохождения, можно предсказать итоговые баллы, что может позволить на более ранних сроках улучшить успеваемость студентов.

В представленной работе в качестве данных, используемых для анализа поведения пользователей, был использован log-файл курса «Продвинутое машинное обучение», содержащий аудит студентов весеннего семестра 2020 года обучения университета ИТМО.

Основная часть. Для выполнения данной работы был выбран язык программирования Python и интерактивная среда разработки Google Colaboratory (Colab).

Было принято решение сформировать пространство признаков, характеризующее поведение пользователей по каждому конкретному элементу из следующих блоков курса: видео-лекции, текстовые материалы, опросы и упражнения. Была создана итоговая таблица, содержащая сведения по всем студентам. Для каждого студента была выделена информация о том, сколько времени он посмотрел каждое видео и смотрел ли он его или нет. Аналогичная информация была получена и для текстовых материалов — посчитано время просмотра документа и указан факт того, был ли материал изучен или нет. Для упражнений и опросов каждому студенту ставились в соответствие количество попыток сдачи задания, итоговый балл и время прохождения с учетом всех попыток.

Также студентам были сопоставлены балл за прохождение курса и балл за прохождение экзамена. Интерес представляют те студенты, которые имеют большую разницу в этих баллах, а именно студенты, которые получили за курс более 90 баллов, а за экзамен менее 50. Такие обучающиеся предположительно могли прибегнуть к списыванию во время прохождения курса.

Итого, подготовлен набор данных, в котором каждый объект (студент) описывается 173 признаками. Дополнительно итоговые баллы студентов за курс были переведены из 100-балльной шкалы в 5-балльную. Полученный признак является откликом в задаче прогнозирования итоговых баллов.

Для ее решения были применены алгоритмы классификации gradient boosting и случайный лес. Некоторые параметры были подобраны с помощью GridSearch.

Выводы. Таким образом, был проведен анализ поведения пользователей в системе электронного обучения. Выяснилось, что большинство студентов не изучали материалы курса, а только лишь решали оцениваемые упражнения. Оказалось, что 25 студентов прошли курс на 90 и более баллов, а экзамен сдали плохо: на 50 баллов и менее. Можно сделать вывод, что нельзя отрицать возможность списывания во время прохождения онлайн-курсов и необходимо проводить более глубокое исследование для выявления таких случаев.

Лучшим классификатором в задаче прогнозирования баллов оказался алгоритм градиентного бустинга. После его применения на тестовом множестве метрика accuracy составила 91,827. Это говорит о том, что алгоритм хорошо справился со своей задачей и по активности пользователей действительно можно предсказать итоговый балл за курс.

Антипова Е.С. (автор)

Подпись

Волчек Д.Г. (научный руководитель)

Подпись