

УДК 004.8:519.86:378

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ УСПЕВАЕМОСТИ СТУДЕНТОВ

А. А. Ефремов, Ю. В. Сильванович

УО «Белорусский государственный университет информатики и радиоэлектроники» (Республика Беларусь, 220013, г. Минск, ул. П. Бровки, 6; e-mail: dekief@bsuir.by)

Аннотация. В работе анализируется применение методов машинного обучения для обработки образовательных данных и прогнозирования академической успеваемости студентов. Рассматривается модель на основе логистической регрессии, позволяющая оценивать вероятность успешного обучения и выявлять влияние факторов учебной активности.

Ключевые слова: машинное обучение, академическая успеваемость, логистическая регрессия, прогнозирование.

APPLICATION OF MACHINE LEARNING METHODS FOR PREDICTING STUDENT ACADEMIC PERFORMANCE

A. A. Efremov, Y. V. Silvanovich

EI «Belarusian state university of informatics and radioelectronics» (Republic of Belarus, 220013, Minsk, P. Brovki st., 6; e-mail: dekief@bsuir.by)

Summary. The paper examines the application of machine learning methods for processing educational data and predicting students' academic performance. A model based on logistic regression is considered, allowing the estimation of the probability of successful learning and the identification of the influence of learning activity factors.

Key words: machine learning, academic performance, logistic regression, prediction.

Цифровизация высшего образования и распространение онлайн-обучения приводят к активному использованию образовательных платформ, в которых накапливаются данные о деятельности студентов. При этом управление образовательным процессом зачастую носит реактивный характер, что не позволяет своевременно выявлять проблемы в обучении. В этих условиях актуальным становится применение методов анализа данных и машинного обучения для прогнозирования результатов обучения.

Цель исследования заключается в разработке модели прогнозирования академической успеваемости студентов на основе методов машинного обучения с использованием данных образовательной платформы.

Для достижения цели решаются задачи анализа данных, формирования системы факторов, выбора метода моделирования, построения модели и определения возможностей ее практического применения.

Анализ образовательных данных развивается в рамках направлений learning analytics и educational data mining, направленных на выявление закономерностей обучения и прогнозирование результатов обучающихся [1].

Для прогнозирования успеваемости используются различные методы машинного обучения, такие как нейронные сети, деревья решений и регрессионные модели. Однако использование сложных моделей в образовательной практике часто затруднено из-за их низкой интерпретируемости, поэтому в качестве основного метода выбрана логистическая регрессия, которая представляет собой метод прогнозирования вероятности осуществления конкретного события на основе известных значений ряда признаков. Способность моделировать вероятности событий делает ее ценным инструментом в различных сферах, включая образование и корпоративное обучение [2].

В исследовании используются данные образовательной платформы высшего учебного заведения, отражающие действия студентов в процессе обучения.

В модели используются показатели, характеризующие поведение студентов: время обучения, количество просмотренных материалов, число выполненных заданий, результаты тестирования и доля освоенного контента, что дает возможность установить связь с результатами обучения.

Связь между показателями активности студентов и результатами обучения определяется с использованием логистической регрессии как метода машинного обучения, которая применяется для задач бинарной классификации (успешно / неуспешно), и позволяет оценить вероятность успешного обучения и вклад факторов в результат [3]. Метод обеспечивает количественный анализ влияния характеристик учебной деятельности на результаты обучения.

Вероятность успешного обучения определяется с использованием логистической функции [3]:

$$P(y = 1) = \frac{1}{1 + e^{-z}}.$$

Параметр z представляет собой линейную комбинацию факторов:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5.$$

Коэффициенты модели отражают степень влияния показателей активности (время обучения, просмотры, задания, результаты тестирования и доля освоенного контента) на вероятность успешного обучения.

На основе построенной модели при подстановке значений показателей активности студентов рассчитывается значение линейной комбинации факторов, после чего определяется вероятность успешного обучения.

При значениях показателей активности студентов: $x_1 = 40$ (время обучения), $x_2 = 25$ (просмотры), $x_3 = 8$ (выполненные задания), $x_4 = 7$ (средний балл), $x_5 = 0,8$ (доля освоенного контента), значение линейной комбинации факторов составляет $z = 1,5$, что соответствует вероятности успешного обучения $P(y = 1) \approx 0,82$.

Результаты расчета показывают, что при заданных значениях факторов вероятность успешного обучения является высокой.

В результате проведенного исследования была разработана модель прогнозирования академической успеваемости студентов на основе методов машинного обучения, которая позволяет оценивать вероятность успешного обучения и выявлять факторы, влияющие на результаты образовательного процесса.

Модель может применяться в образовательных организациях для раннего выявления студентов с риском неуспеваемости и автоматизированного мониторинга учебной деятельности, что способствует повышению эффективности управления обучением и развитию интеллектуальных образовательных систем.

ЛИТЕРАТУРА

1. Romero, C. Educational Data Mining: A Review of the State of the Art / C. Romero, S. Ventura // IEEE Transactions on Systems, Man, and Cybernetics. – 2010. – Vol. 40, No. 6. – P. 601-618.
2. Медетова, К. М. Использование логистической регрессии для анализа данных и оценки успеваемости студентов (сотрудников) в образовательных программах / К. М. Медетова // Инженерное образование в цифровом обществе: материалы Междунар. науч.-метод. конф. – 2024. – С. 22-24.
3. Wang, J. Machine learning approach to student performance prediction of online learning / J. Wang, Y. Yu // PLoS ONE. – 2025. – Vol. 20, No. 1. – e0299018.