

Научная статья

УДК 376.1

DOI: 10.25688/2072-9014.2022.62.4.06

ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗА УСПЕВАЕМОСТИ УЧАЩИХСЯ ОСНОВНОЙ ШКОЛЫ

Татьяна Васильевна Побединская¹,

Ольга Юрьевна Заславская² ✉

¹ ООО «Мобильное электронное образование», Москва, Россия,
0724tatiana@gmail.com 0724tatiana@gmail.com

² Московский городской педагогический университет, Москва, Россия,
zaslavskaya@mgpu.ru ✉, <https://orcid.org/0000-0002-6119-8271>

Аннотация. В статье рассматривается применение алгоритмов машинного обучения для прогнозирования успеваемости учащихся основной школы. Сравниваются результаты работы двух алгоритмов машинного обучения и определяются необходимые данные об учащихся, которые целесообразно собрать для получения высокой точности прогнозирования успеваемости. *Цель исследования:* изучить эффективность использования алгоритмов машинного обучения для решения задачи прогнозирования успеваемости учащихся основной школы. *Задачи исследования:* собрать данные об учащихся для прогнозирования успеваемости; исследовать практическое применение алгоритмов машинного обучения для решения задачи прогнозирования успеваемости учащихся основной школы. Ведущим методом для решения первой задачи был выбран метод анкетирования. Для решения второй задачи использовался метод эксперимента.

Ключевые слова: искусственный интеллект в образовании; большие данные; прогнозирование успеваемости учащихся; алгоритмы машинного обучения; информатизация образования.

Original article

UDC 376.1

DOI: 10.25688/2072-9014.2022.62.4.06

THE USE OF MACHINE LEARNING ALGORITHMS TO PREDICTION THE PERFORMANCE OF BASIC SCHOOL STUDENTS

Tatiana V. Pobedinskaya¹,
Olga Yu. Zaslavskaya² ✉

¹ LLC «Mobile e-education», Moscow, Russia,
0724tatiana@gmail.com

² Moscow City University, Moscow, Russia,
zaslavskaya@mgpu.ru ✉, <https://orcid.org/0000-0002-6119-8271>

Abstract. The article discusses the use of machine learning algorithms for predicting the performance of students in a primary school. The results of the work of two machine learning algorithms are compared and the necessary data about students are determined, which are advisable to collect in order to obtain high accuracy in predicting academic performance. *The purpose of the study:* to study the effectiveness of using machine learning algorithms to solve the problem of predicting the progress of primary school students. *Research objectives:* to collect data on students to predict academic performance; to explore the practical application of machine learning algorithms for solving the problem of predicting the progress of students in a primary school. The method of questioning was chosen as the leading method for solving the first problem. The experimental method was used to solve the second problem.

Keywords: artificial intelligence in education; big data; predicting student performance; machine learning algorithms; informatization of education.

Для цитирования: Побединская, Т. В., Заславская, О. Ю. (2022). Использование алгоритмов машинного обучения для прогноза успеваемости учащихся основной школы. *Вестник МГПУ. Серия «Информатика и информатизация образования»*, 4(62), 75–82. DOI: 10.25688/2072-9014.2022.62.4.06

For citations: Pobedinskaya, T. V., Zaslavskaya, O. Yu. (2022). The use of machine learning algorithms to prediction the performance of basic school students. *MCU Journal of Informatics and Informatization of Education*, 4(62), 75–82. <https://doi.org/10.25688/2072-9014.2022.62.4.06>

Введение

Одной из актуальных проблем, стоящих перед образовательной организацией на сегодня, остается достаточно низкая успеваемость школьников. Многообразие причин затрудняет деятельность учителя по выявлению факторов, влияющих на такое снижение успеваемости школьников, и в большинстве случаев учитель, как правило, выбирает

традиционный способ работы с учащимися, имеющими определенные затруднения в освоении учебной программы, — дополнительные занятия, основная задача которых заключается в организации повторения пройденного учебного материала [1, с. 197].

Однако современные технологии информатизации системы образования приводят к появлению большого количества разнообразных систем сбора, обработки и хранения больших массивов информации [2], которые можно использовать для эффективного прогнозирования успеваемости учащихся. Также в настоящее время существуют методы интеллектуального анализа данных, позволяющие из необработанных данных извлекать высокоуровневые знания, которые могут быть полезны в сфере образования [3].

К таким методам интеллектуального анализа относятся: метод ассоциации, метод классификации, метод кластеризации, метод деревьев решений и метод случайного леса, — используя которые можно ответить на следующие вопросы: можно ли предсказать успеваемость школьников и студентов и какие факторы влияют на успеваемость обучающихся [4]?

Прогнозирование успеваемости учащихся — важный инструмент оценки эффективности применяемых педагогических технологий и методических приемов для преподавателей, и проведения самооценки — для учащихся. Благодаря полученным результатам прогнозирования успеваемости учащихся преподаватели могут вовремя принять корректирующие меры для работы с учениками, имеющими затруднения в освоении учебной программы, например при планировании компенсирующих занятий или системы уроков [5].

В рамках исследования проанализированы данные двух образовательных организаций за три последних учебных года, с 2019 по 2022 год. Использовались два разных источника: школьные отчеты, предоставленные администрацией образовательных организаций, и анкеты, на вопросы которых отвечали обучающиеся и их родители. Поскольку школьные отчеты содержали ограниченное количество информации (т. е. были доступны только оценки и количество пропусков занятий), было принято решение провести дополнительное анкетирование учащихся и их родителей, что позволило собрать демографические, социальные и связанные со школой атрибуты (например, возраст ученика, наличие дополнительного образования, образование родителей и т. д.). Пример фрагмента анкеты представлен на рисунке 1.

Цель настоящего исследования состоит в том, чтобы определить реальность осуществления прогнозирования достижений учащихся и по возможности выявить ключевые индикаторы (переменные), которые оказывают существенное влияние на успеваемость учащихся. Для реализации этой цели были выбраны два алгоритма машинного обучения: случайный лес и дерево решений [6]. Для каждого из этих подходов будут проверены три настройки ввода (например, с четвертными отметками и без них).

Анкета для родителей

Анкетирование в рамках исследования возможности прогнозирования успеваемости учащихся

1. Сколько полных лет вашему ребенку?

2. Посещает ли ваш ребенок какие-нибудь занятия в рамках дополнительного образования?

Отметьте только один овал.

Да, да посещает

Нет, не посещает

3. Какой у вас уровень образования?

Отметьте только один овал.

Неполное среднее или ниже

Среднее общее (школа)

Начальное профессиональное (ПТУ, лицей и т.п.)

Среднее специальное (ссуз, колледж, техникум, медицинское училище и т.п.)

Незаконченное высшее (обучение в вузе без получения диплома)

Высшее (диплом специалиста, бакалавра, магистра и т.п.)

Рис. 1. Фрагмент анкеты учащихся и их родителей

Методы исследования

Ведущими методами были выбраны метод анкетирования (письменный опрос) и метод эксперимента.

Этапы эксперимента:

1. Констатирующий этап — выбор алгоритмов машинного обучения для решения задачи прогнозирования отметок учащихся. На данном этапе осуществили анализ существующих алгоритмов машинного обучения и выбрали из них два алгоритма для проведения эксперимента.

2. Формирующий этап — работа с данными учащихся и использование алгоритмов машинного обучения, заключался в обработке и подготовке данных, запуске алгоритмов «дерево решений» и «случайный лес», фиксации результатов эксперимента.

3. Контрольный этап — проведение анализа результатов работы алгоритмов для решения задачи прогнозирования отметок учащихся, который заключался в сравнении метрик прогнозирования по результатам работы алгоритмов «случайный лес» и «дерево решений».

Результаты исследования

Поскольку в изначальной гипотезе предполагалось, что значения четвертных отметок будут иметь большое влияние, для каждого алгоритма машинного обучения были протестированы три входные конфигурации данных (рис. 2):

- вариант А — со всеми переменными из таблицы, кроме итоговой отметки;
- вариант Б — аналог конфигурации А, но без отметок за третью и четвертую четверть;
- вариант В — аналогично Б, но без отметок за все четверти.

	B	C	D	E	F	G	H	I	J	K
1	age	Medu	Fedu	Mjob	Fjob	G1	G2	G3	G4	G5
4	15	1	1	at_home	other	4	5	5	5	5
5	14	4	2	health	services	4	4	4	4	4
10	13	3	2	services	other	3	4	3	4	4
11	15	3	4	other	other	3	3	3	3	3
12	14	4	4	teacher	health	4	4	5	5	5

Рис. 2. Данные для прогнозирования успеваемости учащихся

Как и ожидалось, наилучшие результаты дает входная конфигурация данных А. Прогностическая эффективность снижается, когда отсутствуют данные с отметками за третью и четвертую четверть (Б), а худшие результаты получаются тогда, когда не используются отметки учащихся (В). По расчетам метрики точности, алгоритм «случайный лес» показал лучший результат для решения задачи прогнозирования отметок учащихся (точность прогнозирования — 81,4 %), за ним следует алгоритм «дерево решений», который дает результат 64,1 % [7–13].

Тем не менее существуют и другие важные факторы, которые влияют на точность прогнозирования:

- организационные, например количество пропусков, дополнительная поддержка в школе или время в пути;
- демографические, например работа родителей;
- социальные, например прогулки с друзьями, вредные привычки и др.

Заключение

Алгоритмы машинного обучения имеют большой потенциал для их применения в области образования. В данном исследовании рассмотрено прогнозирование

отметок учащихся средних школ по двум предметам: математика и русский язык — с использованием оценок, демографических, социальных и других данных, связанных с учащимися.

Были протестированы два алгоритма машинного обучения: «дерево решений» и «случайный лес». Также были изучены отдельные варианты ввода (например, с четвертными отметками и без них). Полученные результаты показывают, что можно достичь высокой точности прогнозов при условии, что известны четвертные отметки учащихся. Тем не менее анализ результатов работы алгоритмов машинного обучения показывает, что в некоторых случаях существуют и другие важные характеристики, такие как: количество пропусков, причина выбора школы, наличие дополнительного образования, возраст учащегося, работа и образование родителей.

Существует значительный потенциал для формирования среды обучения в режиме онлайн с использованием механизма прогнозирования учащихся. Это позволит собрать дополнительные характеристики (например, отметки за предыдущие школьные годы), а также получить ценные отзывы от школьных специалистов (педагогов и психологов). Кроме того, необходимо расширить эксперимент на большее количество образовательных организаций и временной диапазон, чтобы обогатить базы данных учащихся. Также необходимы дополнительные социологические исследования, чтобы понять, почему такие переменные, как причина выбора школы и работа родителей, влияют на успеваемость учащихся.

Список источников

1. Удодова, О. А. (2014). Организация работы со слабоуспевающими и неуспевающими учащимися на уроке. *Концепт*, 17, 197–204.
2. Гриншкун, В. В. Заславская, О. Ю. (2011). История и перспективы развития программ информатизации образования. *Вестник МГПУ. Серия «Информатика и информатизация образования»*, 21, 5–13.
3. Смолина, Е. М. (2021). Методы интеллектуального анализа данных в задачах оценки качества дистанционного образования. *Наука и бизнес: пути развития*, 3(117), 72–75.
4. Шухман, А. Е. (2021). Анализ и прогнозирование успеваемости обучающихся при использовании цифровой образовательной среды. *Высшее образование в России*, 30(8–9), 125–133.
5. Муханов, Д. А. (2021). Применение машинного обучения для прогнозирования успеваемости учащихся. *Студенческая наука Подмосковью*. Сборник материалов Международной научной конференции молодых (с. 321–324.). Орехово-Зуево: Гуманитарно-технологический университет.
6. Гу, Ш. (2022). Основные виды и сферы применения машинного обучения. *Научный аспект*, 3(3), 266–271.
7. Michie, D., Spiegelhalter, D. J., Taylor, C. C. (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood Series in Artificial Intelligence.
8. Kotsiantis, S., Piarrekeas, C., Pintelas, P. (2007). Predicting Students performance in Distance Learning using Machine Learning Techniques. *Applied Artificial Intelligence*, 18, 411–426.

9. Romero, C., Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 135–146.

10. Канащ, А. В. (2021). Интеллектуальный анализ данных для построения моделей машинного обучения в образовании. *Цифровая трансформация — шаг в будущее*. Материалы II Международной научно-практической конференции молодых ученых, посвященной 100-летию Белорусского государственного университета. Минск: Белорусский государственный университет.

11. Хайруллин, А. М. (2020). Машинное обучение как способ решения проблем в сфере образования. *Моя профессиональная карьера*, 1(15), 102–105.

12. Желябин, Д. В. (2021). Анализ различных моделей машинного обучения при классификации многомерных данных в сфере образования. *Современные проблемы проектирования, производства и эксплуатации радиотехнических систем*. Сборник научных трудов (с. 166–172). Ульяновск: Ульяновский государственный технический университет.

13. Лысенков, А. С. (2020). Технологии машинного обучения и их применение в образовании. *Наука и инновации в XXI веке: актуальные вопросы, открытия и достижения*. Сборник статей XVII Международной научно-практической конференции (с. 58–60.). Пенза.

References

1. Udodova, O. A. (2014). Organization of work with underachieving and underachieving students in the classroom. *Scientific and methodological electronic journal*, 17, 197–204. (In Russ.).

2. Grinshkun, V. V., & Zaslavskaya, O. Y. (2011). History and prospects of development of informatization programs of education. *MCU Journal of Informatics and Informatization of Education*, 21, 5–13. (In Russ.).

3. Smolina, E. M. (2021). Methods of data mining in problems of assessing the quality of distance education. *Science and business: ways of development*, 3(117), 72–75. (In Russ.).

4. Shukhman, A. E. (2021). Analysis and forecasting of students' academic performance using the digital educational environment. *Higher education in Russia*, 30(8–9), 125–133. (In Russ.).

5. Mukhanov, D. A. (2021). The use of machine learning to predict student performance. *Student science in the Moscow region*. A collection of materials of the International Scientific Conference of Young Scientists (pp. 321–324). Orekhovo-Zuyevo. (In Russ.).

6. Gu, Sh. (2022). The main types and fields of application of machine learning. *Scientific aspect*, 3(3), 266–271. (In Russ.).

7. Michie, D., Spiegelhalter, D. J., Taylor, C. C. (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood Series in Artificial Intelligence.

8. Kotsiantis, S., Piarrekeas, C., Pintelas, P. (2007). Predicting Students performance in Distance Learning using Machine Learning Techniques. *Applied Artificial Intelligence*, 18, 411–426.

9. Romero, C., Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 135–146.

10. Kanash, A. V. (2021). Intelligent data analysis for building machine learning models in education. *Digital transformation — a step into the future*. Materials of the II International Scientific and Practical Conference of Young Scientists dedicated to the 100th anniversary of the Belarusian State University (pp. 135–139). Minsk: Belarusian State University. (In Russ.).

11. Khairullin, A. M. (2020). Machine learning as a way to solve problems in the field of education. *My professional career*, 1(15), 102–105. (In Russ.).
12. Zhelyabin, D. V. (2021). Analysis of various machine learning models in the classification of multidimensional data in the field of education. *Modern problems of design, production and operation of radio engineering systems*. A collection of scientific works (pp. 166–172). Ulyanovsk: Ulyanovsk State Technical University. (In Russ.).
13. Lysenkov, A. S. (2020). Machine learning technologies and their application in education. *Science and innovation in the XXI century: current issues, discoveries and achievements*. Collection of articles of the XVII International Scientific and Practical Conference (pp. 58–60). Penza. (In Russ.).

Статья поступила в редакцию: 15.06.2022;
одобрена после рецензирования: 04.08.2022;
принята к публикации: 02.09.2022.

The article was submitted: 15.06.2022;
approved after reviewing: 04.08.2022;
accepted for publication: 02.09.2022.

Информация об авторах:

Татьяна Васильевна Побединская — руководитель продукта, ООО «Мобильное электронное образование», Москва, Россия,
0724tatiana@gmail.com

Ольга Юрьевна Заславская — профессор, доктор педагогических наук, профессор, профессор департамента информатизации образования, Институт цифрового образования, Московский городской педагогический университет, Москва, Россия,
zaslavskaya@mgpu.ru, <https://orcid.org/0000-0002-6119-8271>

Information about authors:

Tatiana V. Pobedinskaya — Product Owner, LLC «Mobile e-education», Moscow, Russia, Moscow, Russia,
0724tatiana@gmail.com

Olga Yu. Zaslavskaya — Doctor of Pedagogical Sciences, Professor, Professor of the Department of Informatization of Education. Institute of Digital Education, Moscow City University, Moscow, Russia,
zaslavskaya@mgpu.ru, <https://orcid.org/0000-0002-6119-8271>

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.