

УДК 004.8

Научно-исследовательская статья

DOI: 10.35330/1991-6639-2023-3-113-11-19

EDN: AQTQHZ

Применение методов машинного обучения при оценке деятельности образовательной организации высшей школы

А. М. Бозиева¹, Ф. М. Цеева², Д. В. Хатухова²

¹ Научно-образовательный центр

Кабардино-Балкарского научного центра Российской академии наук
360010, Россия, г. Нальчик, ул. Балкарова, 2

² Институт информатики, электроники и робототехники ФГБОУ ВО КБГУ
360004, Россия, г. Нальчик, ул. Чернышевского, 173

Аннотация. В данной статье решается задача разработки программного кода для оценки деятельности образовательных организаций высшей школы на основе совокупности показателей. В качестве входных данных используются показатели предыдущих оценок вузов и их итоговые результаты. Для успешного решения задачи используется машинное обучение с учителем на основе алгоритма множественной линейной регрессии, что позволяет выявить закономерности для адекватной оценки. Данные закономерности выявляются на основе данных, накопленных в ходе деятельности вузов, и существующего в образовательной практике опыта по оценке вузов. В результате разработанный программный код на основе имеющихся данных дает оценку вузу с определенной точностью.

Ключевые слова: образовательные организации, прогнозирование, оценка деятельности, машинное обучение, регрессия

Поступила 15.05.2023, одобрена после рецензирования 02.06.2023, принята к публикации 05.06.2023

Для цитирования. Бозиева А. М., Цеева Ф. М., Хатухова Д. В. Применение методов машинного обучения при оценке деятельности образовательной организации высшей школы // Известия Кабардино-Балкарского научного центра РАН. 2023. № 3(113). С. 11–19. DOI: 10.35330/1991-6639-2023-3-113-11-19

MSC 62J05 + 68T09

Research article

Application of machine learning methods in assessing the activities of an educational organization of a higher school

A.M. Bozieva¹, F.M. Tseeva², D.V. Khatukhova²

¹ Scientific and Educational Center

Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences
360010, Russia, Nalchik, 2 Balkarov street

² Institute of Informatics, Electronics and Robotics of KBSU
360004, Russia, Nalchik, 173 Chernyshevsky street

Abstract. This article solves the problem of developing a program code for evaluating the activities of educational institutions of higher education on the basis of a set of indicators. The indicators of previous university assessments and their final results are used as input data. Machine learning with a teacher based on a multiple linear regression algorithm is used to successfully solve the problem, which allows us to identify patterns for an adequate assessment. These patterns are revealed on the basis of data accumulated

during the activities of universities, and the experience existing in educational practice in assessing universities. As a result, the developed program code based on the available data gives an assessment of the university with a certain accuracy.

Keywords: educational organizations, forecasting, performance evaluation, machine learning, regression

Submitted 15.05.2023,

approved after reviewing 02.06.2023,

accepted for publication 05.06.2023

For citation. Bozieva A.M., Tseeva F.M., Khatukhova D.V. Application of machine learning methods in assessing the activities of an educational organization of a higher school. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2023. No. 3(113). Pp. 11–19. DOI: 10.35330/1991-6639-2023-3-113-11-19

ВВЕДЕНИЕ

Оценка деятельности высшего учебного заведения (вуза) является достаточно известной задачей. В основном для ее решения применяются методы группы экспертных оценок с дальнейшей обработкой экспертных мнений статистическими инструментами. Современные вузы ежедневно генерируют большие объемы данных. Анализ этих данных может предоставить администраторам и преподавателям ценную информацию об эффективности деятельности вуза или его отдельных подсистем и процессов. В этом контексте все более актуальным становится создание интеллектуальных систем, которые могут помочь высшим учебным заведениям в оценке своей деятельности.

Для решения данной задачи наибольший интерес представляет применение методов нечеткой логики и методов машинного обучения. Применение первой группы методов рассмотрено в [1].

Целью данной работы являются выбор метода анализа данных и разработка адекватного программного кода для оценки вуза.

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Машинное обучение – это направление искусственного интеллекта, включающее алгоритмы обучения на больших наборах данных для прогнозирования или классификации данных¹ [2, 6, 7].

В контексте высшего образования машинное обучение можно использовать для прогнозирования успеваемости учащихся, выявления учащихся из групп риска, предоставления рекомендаций по курсам и т.п.

Одной из ключевых концепций машинного обучения является понятие функции потерь, которая измеряет разницу между прогнозируемыми и истинными значениями. В процессе обучения модели функция потерь минимизируется, чтобы повысить точность прогнозирования [9].

Другой важной концепцией машинного обучения является выборка данных. В процессе обучения модель использует обучающий набор данных для настройки своих параметров и

¹ Федеральный закон от 24.04.2020 № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации – городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных». [Электронный ресурс] URL: http://www.consultant.ru/document/cons_doc_LAW_351127 (дата обращения: 23.01.2023).

улучшения точности прогнозирования. Это может быть сделано с помощью ручной куратизации данных или использования алгоритмов отбора, основанных на данных.

Существует несколько методов машинного обучения, включая обучение с учителем, обучение без учителя и обучение с подкреплением. Обучение с учителем предполагает использование маркированных данных для обучения модели, тогда как обучение без учителя использует немаркированные данные, чтобы обнаруживать скрытые структуры в данных. Обучение с подкреплением рассматривает процесс обучения модели, которая принимает решения на основе опыта, полученного через взаимодействие с окружающей средой [3].

Приведем классификацию известных алгоритмов анализа данных для каждого метода [3–6] (рис. 1).

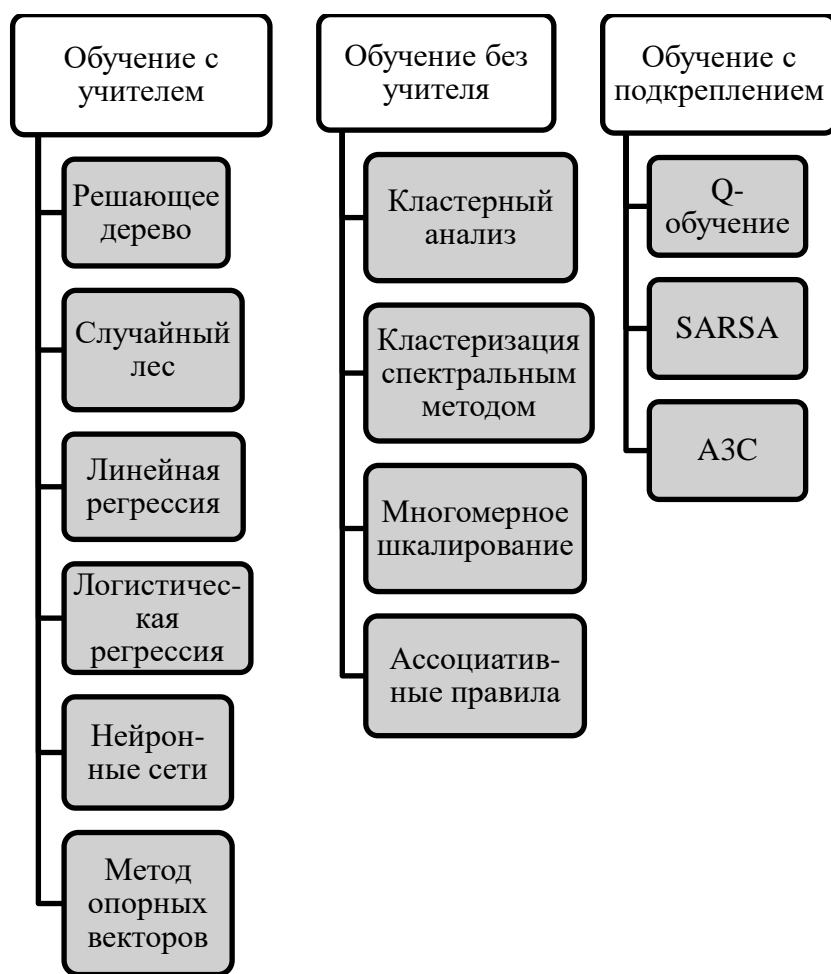


Рис. 1. Алгоритмы анализа данных на основе метода обучения с учителем

Fig. 1. Data analysis algorithms based on the supervised learning method

Для оценки деятельности вуза необходимо иметь некоторую выборку данных, которая содержит информацию об определенном наборе показателей (переменных), связанных с деятельностью данного вуза, а также информацию о целевой переменной, которую необходимо оценить (например, количество выпускников, общая оценка качества образования и т.д.). Предполагается, что есть зависимость между значениями показателей и целевой переменной.

Показателями могут быть, например, количество выпускников, которые находят работу в своей профессиональной области, количество часов, отведенных на преподавание конкретных дисциплин, количество научных публикаций вуза, количество студентов, получивших гранты и т.п.

ПОСТАНОВКА ЗАДАЧИ

Формальная общая постановка задачи машинного обучения с учителем имеет следующий вид. Пусть дана образовательная организация высшей школы, имеющая набор показателей $X = \{X_1, X_2, \dots, X_n\}$, которые влияют на оценку ее деятельности Y . Пусть также имеется обучающая выборка из N пар x_i, y_i . Каждый обучающий пример определяется в следующем виде: $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, где $x_i \in \{1 \dots n\}$ – вектор входных признаков i -го примера, $y_i \in \{1 \dots n\}$ – результирующее значение i -го примера.

Задача заключается в том, чтобы разработать модель, описывающую функциональную зависимость $f: X \rightarrow Y$, которая позволила бы прогнозировать оценку деятельности образовательной организации высшей школы на основе ее показателей.

Таким образом, задача сводится к поиску наилучшей функциональной зависимости f , которая минимизирует ошибку прогнозирования оценки деятельности Y по ее характеристикам X .

Математически задача может быть записана следующим образом:

$$Y = f(X) + \varepsilon,$$

где Y – итоговая оценка деятельности образовательной организации высшей школы, X – набор показателей образовательной организации, f – функциональная зависимость, является элементом пространства гипотез (функций) F , ε – ошибка, возникающая при прогнозировании.

МЕТОДЫ РЕШЕНИЯ

Процедура оценки деятельности вуза включает в себя следующие этапы:

1. Формирование списка показателей оценки, квалиметрическая оценка шкал.
2. Сбор данных. Собираются данные из различных источников, таких как оценки студентов, материалы преподавателей, статистика посещаемости лекций и семинаров, успеваемость студентов и т.д.

2. Обработка данных. Здесь решается вопрос, какие методы будут использованы для обработки данных, – нормирование, стандартизация, преобразование, агрегирование, а также создаются новые показатели или выбираются самые значимые показатели для использования в модели машинного обучения.

3. Выбор алгоритма машинного обучения. В данной работе применен алгоритм машинного обучения с учителем – множественная линейная регрессия. Выбор алгоритма зависит от характеристик данных и условий задачи.

4. Создание модели. На этом этапе строится модель с использованием выбранного алгоритма машинного обучения. Модель определяет зависимость между оценкой деятельности образовательной организации и ее характеристиками.

5. Обучение модели. Модель обучается на собранных данных, чтобы определить зависимость между показателями деятельности образовательной организации и ее оценкой.

6. Проверка модели на новых данных. После обучения модели проводится проверка ее работоспособности на тестовых данных. Это позволяет оценить точность модели.

7. Применение модели. Разработанная модель может быть использована для оценки деятельности образовательных организаций высшей школы. Более конкретно модель может быть использована для прогнозирования того, насколько хорошо высшее учебное заведение выполнит свои цели и задачи.

Пример. Предположим, что у нас есть несколько вузов. Для построения модели машинного обучения мы можем использовать имеющиеся данные об оценке деятельности данных вузов (табл. 1).

Таблица 1. Обучающая выборка (пример)

Table 1. Training sample (example)

Показатели \ Вузы	1	2	3	4	5	6	7	8
Система управления вузом	5	8	9	4	4	8	5	6
Качество обучения	7	–	–	–	6	–	–	–
Средний балл ЕГЭ	58	63	82	–	53	82	–	76
Качество инфраструктуры	4	–	6	6	–	7	–	–
Публикационная активность	253	852	765	402	–	–	303	–
Материально-техническое обеспечение	4	–	–	4	–	–	5	–
Репутация	4	–	–	–	6	7	–	–
Трудоустроенность выпускников	6	–	3	–	–	–	–	–
Социальная и воспитательная работа	7	–	–	–	–	5	–	6
Стоимость обучения	2	–	–	4	–	–	7	–
Итоговая оценка	4,3	3,8	4,1	4,4	4,2	4,4	4,5	3,9

Для обучения модели, как было обозначено выше, применен алгоритм множественной линейной регрессии. Для этого мы разделим наш набор данных на две части: обучающую и тестовую выборки. Обучающая выборка будет использоваться для обучения модели, а тестовая выборка – для проверки ее точности [7].

Алгоритм множественной линейной регрессии – это модель машинного обучения, которая используется для предсказания значения некоторой целевой переменной на основе значений набора независимых переменных (признаков). Цель состоит в том, чтобы найти линейную связь между независимыми переменными и целевой переменной. Математическая модель алгоритма множественной линейной регрессии может быть описана следующим образом:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon,$$

где x_1, x_2, \dots, x_n – значения признаков объекта, $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ – коэффициенты уравнения регрессии, которые соответствуют каждому признаку, ε – ошибки модели, которые являются случайными и связаны с невозможностью предсказания всех факторов, которые могут влиять на целевую переменную.

Модель будет описываться гиперплоскостью.

Целью линейной регрессии является нахождение оптимальных значений параметров $\beta_0, \beta_1, \beta_2, \dots, \beta_n$, которые минимизируют ошибку модели. Одним из распространенных

методов для нахождения оптимальных параметров является метод наименьших квадратов. Для этого минимизируется квадратичная функция ошибки методом градиентного спуска:

$$J(\beta) = \frac{1}{2m} \sum_{i=1}^m (h_{\beta}(x^i) - y^i)^2,$$

где $h_{\beta}(x^i)$ – это гипотеза модели, предсказывающая значение целевой переменной y , основываясь на значениях признаков x^i и параметрах модели β ; y^i – это фактическое значение целевой переменной для i -го примера; m – это количество примеров в обучающей выборке.

После обучения модели можно использовать ее для прогнозирования оценки деятельности образовательной организации на основе ее показателей. Например, если у нас есть данные о новом вузе, мы можем использовать нашу модель для оценки деятельности данного вуза.

Ввиду неравнозначности показателей оценки, приведенных в таблице 1, необходимо производить расчет их весов. Для этой цели целесообразно применение метода анализа иерархий [8].

Для создания компьютерной модели машинного обучения на основе алгоритма линейной регрессии воспользуемся библиотеками Scikit-learn, Pandas, Numpy для Python. Фрагмент программного кода для обучения модели и прогнозирования оценки деятельности образовательной организации на основе показателей (табл. 1) представлен на рисунке 2.

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression

# создание датафрейма с данными о вузах и весовыми коэффициентами
data = pd.DataFrame({
    'system': [5, 8, 9, 4, 4, 8, 5, 6],
    'exam': [58, 63, 82, 80, 53, 82, 73, 76],
    'cost': [2, 5, 6, 4, 5, 6, 7, 4],
    'reputation': [4, 7, 3, 8, 6, 7, 8, 7],
    'mater_teh_obes': [4, 6, 7, 4, 6, 5, 5, 4],
    'students_trud': [6, 6, 3, 5, 9, 8, 4, 7],
    'publications': [253, 852, 765, 402, 800, 702, 303, 130],
    'soucial': [7, 2, 2, 5, 1, 5, 4, 6],
    'infrastructure': [8, 7, 9, 7, 8, 9, 10, 7],
    'teaching_quality': [7, 8, 9, 7, 8, 9, 10, 7],
    'rating': [4.3, 3.8, 4.1, 4.4, 4.2, 4.4, 4.5, 3.9],
})

weights = pd.DataFrame({
    'system': 0.1,
    'exam': 0.1,
    'cost': 0.1,
    'reputation': 0.1,
```

Рис. 2. Фрагмент кода оценки деятельности вуза

Fig. 2. Fragment of the code for assessing the activities of the institute or university

Весь процесс разработки кода состоит из следующих этапов:

1. Создание датафрейма с данными о вузах и весовыми коэффициентами.
2. Объединение данных и весовых коэффициентов в один датафрейм.
3. Разделение данных на признаки и целевую переменную.

4. Обучение модели линейной регрессии.
5. Предсказание рейтинга для новых данных.
6. Оценка точности модели.
7. Вывод результатов.

В результате получаем следующий вывод (рис. 3):

Прогнозируемая оценка деятельности: [3.58] Точность модели: 0.86

Рис. 3. Вывод программы

Fig.2. Program output

Это означает, что модель смогла прогнозировать оценку деятельности образовательной организации на основе ее характеристик с точностью 86 %.

ЗАКЛЮЧЕНИЕ

Прогнозирование оценки деятельности образовательных организаций является актуальной задачей в современных условиях. Существуют различные методы, которые могут помочь в ее решении. В данной работе использован метод машинного обучения с учителем. Из существующих алгоритмов данного метода для разработки модели анализа данных о деятельности вузов использован алгоритм линейной регрессии. Модель разработана в виде программного кода, написанного на языке программирования высокого уровня Python с использованием прикладных библиотек для анализа данных. В результате получена программа, позволяющая оценивать вузы с определенной точностью.

Предлагаемый подход позволяет решать многофакторную слабоформализуемую задачу в условиях необходимости работы с нечеткой информацией. Однако, как и любой другой метод, линейная регрессия не является универсальным решением и может иметь свои ограничения. Поэтому для достижения наилучших результатов необходимо использовать несколько методов и сравнить их результаты.

СПИСОК ЛИТЕРАТУРЫ

1. Бозиева А. М., Дзамихова Ф. Х. Модель интеллектуальной системы, основанной на нечеткой логике, в задачах оценки деятельности высшего учебного заведения // Известия Кабардино-Балкарского научного центра РАН. 2023. № 1(111). С. 11–17. DOI: 10.35330/1991-6639-2023-1-111-11-17, EDN: VGNLBT
2. Котлярова И. О. Технологии искусственного интеллекта в образовании // Вестник Южно-Уральского государственного университета. Серия: Образование. Педагогические науки. 2022. Т. 14. № 3. С. 69–82. DOI: 10.14529/ped220307, EDN: JADHNV
3. Полетаева Н. Г. Классификация систем машинного обучения // Вестник Балтийского федерального университета им. И. Канта. Серия: Физико-математические и технические науки. 2020. №1. С. 5–22. EDN: RCHVEU
4. Шай Ш.-Ш., Шай Б.-Д. Идеи машинного обучения. Москва: ДМК-Пресс, 2019. 436 с. ISBN: 978-5-97060-673-5
5. Бринк Х., Ричардс Дж., Феверолф М. Машинное обучение / Пер. И. Рuzмайкина. Санкт-Петербург: Питер, 2017. 336 с. ISBN: 978-5-496-02989-6

6. Вьюгин В. В. Математические основы машинного обучения и прогнозирования. Москва: МЦНМО, 2022. 400 с. ISBN: 978-5-4439-1249-3
7. Гельман Э., Хилл Дж., Вехтари А. Регрессия: теория и практика. С примерами на R и Stan / Пер. с англ. В. Яценкова. Москва: ДМК Пресс, 2022. 748 с. ISBN: 978-5-97060-987-3
8. Саати Т. Принятие решений: метод анализа иерархий / Пер. с англ. Р. Г. Вачнадзе. Москва: Радио и связь, 1993. 278 с.
9. Воронов М. В., Пименов В. И., Небаев И. А. Системы искусственного интеллекта: учебник и практикум для вузов. Москва: Юрайт, 2022. 256 с.

REFERENCES

1. Bozieva A.M., Dzamikhova F.Kh. A model of an intelligent system based on fuzzy logic in the tasks of evaluating the activities of a higher educational institution. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2023. No. 1(111). Pp. 11–17. DOI: 10.35330/1991-6639-2023-1-111-11-17. EDN: VGNLBT (In Russian)
2. Kotlyarova I.O. Artificial intelligence technologies in education. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Obrazovanie. Pedagogicheskie nauki* [Bulletin of the South Ural State University. Series: Education. Pedagogical Sciences]. 2022. Vol. 14. No 3. Pp. 69–82. EDN: JADHNV (In Russian)
3. Poletaeva N.G. Classification of machine learning systems. *Vestnik Baltiyskogo federal'nogo universiteta im. I. Kanta. Seriya: Fiziko-matematicheskie i tekhnicheskie nauki* [Bulletin of the Baltic Federal University n.a I. Kant. Series: Physical, mathematical and technical sciences]. 2020. No 1. Pp. 5–22. EDN: RCHVEU (In Russian)
4. Shai Sh.-Sh., Shai B.-D. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press, 2014. 410 p.
5. Brink H., Joseph W. Richards, Fetherolf M. *Real-World Machine Learning*. New-York: Manning Publications. 2014. 264 p.
6. V'yugin V.V. *Matematicheskie osnovy mashinnogo obucheniya i prognozirovaniya* [Mathematical Foundations of Machine Learning and Prediction]. Moscow: MTsNMO, 2022. 400 p. (In Russian)
7. Gelman A., Hill J., Vehtari A. *Regression and Other Stories (Analytical Methods for Social Research)*. Cambridge: Cambridge University Press, 2020. 548 p.
8. Saati T. *Prinyatie resheniy: metod analiza ierarkhiy* [Decision Making: Hierarchy Analysis Method]. Transl. from Eng. R. G. Vachnadze. Moscow: Radio i svyaz', 1993. 278 p.
9. Voronov M.V., Pimenov V.I., Nebaev I.A. *Sistemy iskusstvennogo intellekta: uchebnik i praktikum dlya vuzov* [Artificial intelligence systems: Tutorial and practices for High School]. Moscow: Yurayt, 2022. 256 p. (In Russian)

Информация об авторах

Бозиева Асият Мухтаровна, аспирант, Научно-образовательный центр Кабардино-Балкарского научного центра РАН;

360010, Россия, г. Нальчик, ул. Балкарова, 2;

bozieva.asya@mail.ru, ORCID: <https://orcid.org/0000-0002-1124-2289>

Цеева Фатимат Мухамедовна, доцент кафедры мехатроники и робототехники, Институт информатики, электроники и робототехники КБГУ;

360000, Россия, г. Нальчик, ул. Чернышевского, 173;

mfmkbsu@mail.ru, ORCID: <https://orcid.org/0000-0001-7203-3571>

Хатухова Дана Владимировна, ст. преподаватель кафедры информационных технологий в управлении техническими системами, Институт информатики, электроники и робототехники КБГУ; 360000, Россия, г. Нальчик, ул. Чернышевского, 173; dkhatukhova@list.ru, ORCID: <https://orcid.org/0009-0009-0190-8823>

Information about the authors

Bozieva Asiyat Mukhtarovna, postgraduate student, Scientific and Educational Center of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences; 360000, Russia, Nalchik, 2 Balkarov street; bozieva.asya@mail.ru, ORCID: <https://orcid.org/0000-0002-1124-2289>

Tseeva Fatimat Mukhamedovna, Associate Professor of the Department of Mechatronics and Robotics of the Institute of Informatics, Electronics and Robotics of the KBSU; 360004, Russia, Nalchik, 173 Chernyshevsky street; mfmkbsu@mail.ru, ORCID: <https://orcid.org/0000-0001-7203-3571>

Khatukhova Dana Vladimirovna, Senior Lecturer of the Department of Information Technologies in the Management of Technical Systems of the Institute of Informatics, Electronics and Robotics of the KBSU;

360004, Russia, Nalchik, 173 Chernyshevsky street; dkhatukhova@list.ru, ORCID: <https://orcid.org/0009-0009-0190-8823>