

МОДЕЛИРОВАНИЕ АКАДЕМИЧЕСКОЙ УСПЕВАЕМОСТИ СТУДЕНТОВ НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ОБРАЗОВАТЕЛЬНЫХ ДАННЫХ

Виктория Викторовна БОБРОВА ^a,

Ольга Игоревна БАНТИКОВА ^{b,*},

Влада Александровна НОВИКОВА ^c

^a доктор экономических наук, доцент, директор Института менеджмента,
заведующая кафедрой таможенного дела,
Оренбургский государственный университет (ОГУ),
Оренбург, Российская Федерация
bobrova1971@mail.ru
<https://orcid.org/0000-0003-3558-5662>
SPIN-код: 7357-1479

^b кандидат экономических наук, доцент кафедры физико-математических дисциплин,
Российский биотехнологический университет (РОСБИОТЕХ),
Москва, Российская Федерация
bantikova777@mail.ru
<https://orcid.org/0000-0002-0577-6276>
SPIN-код: 6682-0574

^c студентка института кибербезопасности и цифровых технологий,
МИРЭА – Российский технологический университет,
Москва, Российская Федерация
vlanovickova@gmail.com
ORCID: отсутствует
SPIN-код: отсутствует

* Ответственный автор

История статьи:

Рег. № 565/2022
Получена 13.12.2022
Получена в
доработанном виде
24.12.2022
Одобрена 09.02.2023
Доступна онлайн
28.02.2023

УДК 330.43:519.254
JEL: C01, C38, C51,
C53, C55

Ключевые слова:

образовательные
данные,
интеллектуальный
анализ, академическая
успеваемость,
моделирование,
машинное обучение

Аннотация

Предмет. Применение методов машинного обучения для анализа академической успеваемости студентов.

Цели. Выявление факторов, оказывающих влияние на академическую успеваемость студентов, обнаружение скрытых закономерностей, полезных и интерпретируемых знаний о результатах учебного процесса и его участниках с помощью интеллектуального анализа образовательных данных.

Методология. Исследование базируется на методах эконометрического моделирования, многомерной классификации и кластеризации больших данных.

Результаты. Разработаны модели интеллектуального анализа образовательных данных, позволяющие провести сравнительный анализ обучающихся и прогнозирование уровня освоения образовательной программы в зависимости от таких факторов, как суммарный балл вступительных испытаний, средний балл академической успеваемости, основа и форма обучения, курс, уровень подготовки, пол и возраст студента.

Выводы. Результаты применения методов машинного обучения для анализа успеваемости позволят дифференцированно подбирать эффективные методы и технологии обучения к группам студентов с различными уровнями академических результатов, своевременно

принимать корректирующие действия относительно студентов из группы риска, что в итоге будет способствовать сохранности контингента обучающихся и повышению качества образовательного процесса.

© Издательский дом ФИНАНСЫ и КРЕДИТ, 2022

Для цитирования: Боброва В.В., Бантикова О.И., Новикова В.А. Моделирование академической успеваемости студентов на основе интеллектуального анализа образовательных данных // *Экономический анализ: теория и практика*. – 2023. – Т. 22, № 2. – С. 235 – 253.
<https://doi.org/10.24891/ea.22.2.235>

Проблема качества обучения на протяжении многих лет была и остается одной из центральных проблем в образовании. Качество образования высшего учебного заведения представляет собой комплексную интегральную категорию, которая определяется уровнем успеваемости обучающихся, их базовой подготовкой и заинтересованностью в учебном процессе, квалификацией профессорско-преподавательского состава, уровнем образовательных технологий, состоянием учебно-материальной базы. В рамках решения задач, связанных с достижением высокого уровня академических результатов, возникает необходимость регулярного контроля уровня успеваемости студентов, выявления факторов, оказывающих влияние на академическую успеваемость, и прогнозирования результатов освоения образовательной программы.

Развитие дистанционного образования и электронных форм обучения, внедрение цифровых технологий в образовательный процесс привели к формированию и интенсивному накоплению информации о различных аспектах образовательного процесса, включающей данные об успеваемости, посещаемости и достижениях обучающихся.

В связи с формированием больших объемов образовательных данных актуальными становятся задачи их обработки и интеллектуального анализа для выявления новых, порою скрытых, взаимосвязей в данных, получения практически полезных знаний и их интерпретации для принятия эффективных управленческих решений.

Анализу и моделированию образовательных данных, в том числе с применением методов машинного обучения, занимались как отечественные, так и зарубежные ученые, среди которых R. Baker, P. Inventado [1], G. Siemens¹, S. Mohamad, Z. Tasir [2], О.О. Замков, А.А. Пересецкий [3], И.Г. Захарова [4, 5], С.В. Русаков [6], С.В. Помян [7], Я.К. Салал, С.М. Абдуллаев [8], О.А. Фиофанова [9], В.А. Частикова [10].

¹ Siemens G., Baker R. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. Association for Computing Machinery, New York, 2012, pp. 252–254.
URL: <https://doi.org/10.1145/2330601.2330661>

В исследовании Е.Е. Котовой [11] предложен метод автоматизированного прогнозирования результатов обучения студентов с учетом результатов сдачи ЕГЭ по русскому языку, математике, физике и информатике, оценок академической успеваемости и индивидуальных параметров когнитивно-познавательной сферы учащихся. В свою очередь П.П. Белоножко, А.П. Карпенко, Д.А. Храмов [12] проанализировали направления приложения анализа образовательных данных и те преимущества, которые дает его использование всем участникам образовательного процесса. В работе наряду с традиционными для анализа данных методами рассмотрены специфические методы анализа образовательных данных, среди которых открытие с помощью моделей и перегонка данных для принятия решений человеком. В работе Н.А. Чернышовой [13] выявлено наличие связи между результатами ЕГЭ и успеваемостью студентов в сельскохозяйственном вузе на основе многофакторного регрессионного анализа по факультетам и укрупненным группам направлений подготовки. Автором работы [14] Е.В. Ширинкиной обосновано, что учебная аналитика в совокупности с интеллектуальным анализом образовательных данных дает возможность разработать точные модели, характеризующие поведение обучающихся, их свойства, слабые и сильные места контента и взаимодействия с ним, командную и групповую динамику. В исследовании представлены метрики эффективности образовательного продукта. Эти показатели помогают оценить качество учебного контента и отдельных его частей, уровень подготовки преподавателей и ведущих, прогресс слушателей. В исследовании И.Ю. Харламовой [15] на примере физико-технического факультета Саратовского государственного технического университета им. Ю.А. Гагарина с помощью корреляционного анализа выявлена значимая статистическая связь между параметрами, характеризующими результаты сдачи ЕГЭ и результатами сдачи сессии. В исследовании В.И. Токтаровой, О.Г. Поповой [16] продемонстрирована модель взаимосвязи итоговой оценки студента и его активности (поведения) при обучении на примере онлайн-курса по программированию, размещенного на платформе LMS Moodle, в результате чего были выделены действия студента, в наибольшей степени влияющие на итоговую оценку.

Актуальность и уровень проработанности темы определили цель нашего исследования – разработать модели интеллектуального анализа образовательных данных для выявления взаимосвязей между индивидуальными характеристиками обучающихся и их академической успеваемостью.

Процесс применения методов машинного обучения для анализа академической успеваемости студентов состоит из следующих этапов [12, 17]:

- предварительной обработки данных;
- выявления закономерностей в данных;
- кластеризации объектов наблюдений;

- применения выявленных закономерностей и результатов кластеризации для прогнозирования будущих событий в обучающей среде;
- использования построенных прогнозов при поддержке принятия решений и выработки образовательной политики.

Методология данного исследования базируется на совокупности методов интеллектуального анализа образовательных данных на каждом из перечисленных этапов.

Предварительная обработка образовательных данных включает отбор признаков, характеризующих участников образовательного процесса, сбор статистических данных, восстановление пропущенных значений, выявление аномальных наблюдений, а также визуализацию исходных данных. Предварительный анализ природы статистических данных предполагает проверку гипотез стационарности, нормальности, независимости, однородности, оценку вида функции распределения, ее параметров. Обнаружение отклонений и выбросов в данных осуществляется с помощью построения диаграммы «ящик с усами» и методов робастного оценивания [18].

Для выявления закономерностей в данных в зависимости от их типа в нашем исследовании применялись методы корреляционно-регрессионного анализа, построение и анализ таблиц сопряженности, а также методы непараметрической статистики [18].

В задачу кластеризации обучающихся по сформированному набору признаков положен принцип максимизации межклассового и минимизация внутриклассового расстояния, для решения которой применяются иерархические и итерационные методы кластерного анализа².

Заключительный этап интеллектуального анализа образовательных данных сводится к применению выявленных закономерностей и результатов кластеризации для прогнозирования будущих событий в обучающей среде, для реализации которого используется метод классификации с помощью деревьев решений [17].

Исходной информационной базой для проведения интеллектуального анализа образовательных данных выступила выборочная совокупность, содержащая информацию о 1 095 обучающихся института менеджмента Оренбургского государственного университета за учебный 2021/22 г. по следующим показателям:

- средний балл академической успеваемости;

² Бантикова О.И., Раменская А.В. Математические и инструментальные методы сравнительного анализа высших учебных заведений г. Оренбурга // Университетский комплекс как региональный центр образования, науки и культуры: материалы Всероссийской научно-методической конференции (с международным участием), Оренбург, 23–25 января 2020 г.. Оренбург: Оренбургский государственный университет, 2020. С. 2363–2368.

- суммарный балл вступительных испытаний³;
- курс;
- возраст;
- основа обучения (1 – бюджетная, 0 – платная);
- форма обучения (1 – очная, 0 – заочная);
- уровень подготовки (1 – бакалавриат/специалитет, 0 – магистратура);
- пол (1 – женский, 0 – мужской);
- населенный пункт прописки (1 – Оренбург, 0 – другой город/село);
- условия проживания (1 – проживает в общежитии, 0 – не проживает в общежитии);
- изучаемый язык (1 – английский, 0 – другой);
- направление подготовки.

Показатель «направление подготовки» оцифрован следующим образом: 1 – бизнес-информатика; 2 – государственное и муниципальное управление; 3 – менеджмент; 4 – прикладная информатика; 5 – сервис; 6 – таможенное дело; 7 – торговое дело; 8 – туризм; 9 – управление персоналом.

Диаграмма «ящик с усами» по показателю «возраст» (*рис. 1*) позволила выявить аномальные наблюдения (выбросы) в выборке, однако, учитывая, что данные объекты – это реальные студенты, обучающиеся заочно, возраст которых варьирует от 30 лет до 51 года, и завышенные значения показателя не являются опечатками или ошибками, данные объекты были учтены для дальнейшего исследования.

Результаты описательной статистики по показателям, измеренным в интервальной шкале, а также диаграммы распределения студентов по номинальным признакам, позволили охарактеризовать контингент обучающихся института менеджмента Оренбургского государственного университета. Так, средний балл академической успеваемости студентов составляет 4,19, в среднем обучающиеся поступают в институт менеджмента с суммарным баллом вступительного испытания около 62, что соответствует 186 баллам по системе ЕГЭ. Среди обучающихся присутствуют студенты с первого по пятый курс, возраст которых варьирует от 17 до 51 года, средний возраст обучающихся составляет приблизительно 22 года. По состоянию на

³ Учитывая, что для бакалавриата/специалитета и магистратуры используются разные системы оценок вступительных испытаний (для бакалавриата/специалитета – это суммарный балл ЕГЭ от 0 до 300, для магистратуры – результаты вступительных испытаний от 0 до 100), то для дальнейшего анализа была произведена нормализация данных в интервале [0–100].

учебный 2021/22 г. в институте менеджмента обучалось 70% девушек и 30% юношей, чуть больше половины обучаются на платной основе (57%) и на очной форме обучения (55%), большинство студентов (76%) осваивают первую ступень образования на бакалавриате или специалитете, 40% студентов проживают в Оренбурге, где и расположен вуз, остальная часть – приезжие из других регионов или муниципальных образований Оренбургской области, 9% обучающихся проживают в общежитиях университета, более 90% изучают английский язык в стенах университета. Среди обучающихся института менеджмента в выборке преобладают студенты направлений «менеджмент» и «государственное и муниципальное управление», на долю которых приходится 21 и 19% соответственно. Самыми малочисленными являются представители таких направлений, как «торговое дело» и «бизнес-информатика».

Важной задачей интеллектуального анализа образовательных данных является выявление взаимосвязей между признаками. На основе визуализации данных были предположены зависимости между признаками, характеризующими обучающихся института менеджмента. В частности, по *рис. 2* можно предположить умеренную прямую зависимость между показателями «средний балл академической успеваемости» и «суммарный балл вступительных испытаний».

На *рис. 3* видно, что средний балл академической успеваемости студенток-девушек выше среднего балла академической успеваемости студентов-юношей, это позволяет предположить, что пол обучающегося оказывает влияние на его успеваемость.

Предположения о наличии зависимости между признаками, характеризующими обучающихся, были проверены с помощью различных методов в зависимости от шкал их измерения, в частности, на основе анализа таблиц сопряженности с использованием непараметрических методов статистики, а также аппарата корреляционно-регрессионного анализа [15].

В результате было выявлено, что средний балл студентки (4,27) выше среднего балла студента (4); средний балл студента, обучающегося на бюджетной основе (4,37), а также на очной форме обучения (4,36) выше, чем студента, обучающегося на платной основе (4,06) и заочной формах обучения (3,99); средний балл академической успеваемости у студентов магистратуры (4,35) выше, чем у студентов, обучающихся на бакалавриате или специалитете (4,14); чуть лучше учатся студенты, проживающие в общежитии, изучающие английский язык и проживавшие до поступления в вуз за пределами Оренбурга. Незначительно отличается средний балл успеваемости по различным направлениям подготовки, наиболее высокий средний балл академической успеваемости (4,37) наблюдается на специалитете «таможенное дело», наименьший (4,13) – на бакалавриате направлений подготовки «менеджмент» и «прикладная информатика».

Для количественной оценки влияния факторов на средний балл академической успеваемости разработана линейная модель множественной регрессии, оценка которой представлена в табл. 1. Оценка уравнения регрессии статистически значима, регрессионные остатки представляют собой белый шум, вариация общего коэффициента смертности на 85% объясняется вариацией вошедших в модель факторов, что свидетельствует о хорошем качестве модели регрессии.

Одним из существенных факторов, оказывающих влияние на средний балл академической успеваемости, является форма обучения, так, для студентов очной формы средний балл академической успеваемости выше в среднем на 0,293 балла по сравнению со студентами заочной формы при прочих равных условиях. Регулярные аудиторные занятия, закрепление теоретических знаний на практике, постоянный контакт с преподавателями, четкий график, возможность посещения библиотек, конференций, контроль со стороны и здоровая конкуренция в коллективе, доступные в большей степени студентам очной формы обучения, позволяют получить более глубокие и разносторонние знания, стимулируют к более усердной работе студента и положительно сказываются на его успеваемости.

Отмечаются различия в успеваемости в зависимости от пола обучающегося. Так, девушки имеют средний балл академической успеваемости в среднем на 0,25 балла выше по сравнению с юношами.

Для студентов, обучающихся на бакалавриате или специалитете, средний балл академической успеваемости ниже в среднем на 0,248 балла по сравнению с магистрантами, отчасти это можно объяснить тем, что в магистратуру поступают более мотивированные на учебу студенты.

Наблюдается динамика успеваемости от первого курса к старшим: с переходом на более старший курс средний балл академической успеваемости увеличивается в среднем на 0,019 балла, что объясняется, с одной стороны, адаптацией первокурсников к учебному процессу, с другой стороны, ростом мотивации к учебе старшекурсников.

Положительная корреляция между суммарным баллом вступительных испытаний (ЕГЭ) и средним баллом академической успеваемости объясняется уровнем знаний и общим отношением студента к учебному процессу, так, увеличение суммарного балла вступительных испытаний на 10 ед. приводит к росту среднего балла академической успеваемости на 1,1 балла.

Мотивацией к получению стипендии можно объяснить то, что средний балл академической успеваемости в среднем выше на 0,114, если студент обучается на бюджетной основе, по сравнению со студентами платной основы обучения.

Можно предположить, что оставшиеся 15% вариации среднего балла академической успеваемости объясняются вариацией таких неучтенных в модели факторов, как посещаемость студентов, успеваемость в разрезе дисциплин и т.д.

Учитывая, что успешность освоения образовательной программы формируется под воздействием различных факторов, таких как суммарный балл вступительных испытаний, курс, пол, форма и основа обучения, уровень подготовки, для сравнительного анализа обучающихся вуза осуществлена их кластеризация по сформированному ранее набору показателей, результаты которой представлены в *табл. 2*.

В состав первого кластера вошли 351 студент, которые являются наименее успешными в плане освоения образовательной программы, поскольку обучающиеся данного кластера характеризуются самыми низкими (по сравнению с представителями других кластеров) значениями среднего балла академической успеваемости (3,95) и суммарного балла вступительных испытаний (47,34). Данный кластер представляют студенты более старшего возраста (в среднем около 24 лет), преимущественно женского пола, обучающиеся на бакалавриате или специалитете заочной формы обучения на платной основе.

Второй кластер самый многочисленный, состоит из 443 студентов, преимущественно из девушек, средний возраст которых чуть более 21 года, обучающихся на бакалавриате или специалитете, как очной, так и заочной форм обучения, на платной основе. Данный кластер студентов характеризуется средним (по сравнению с другими кластерами) уровнем успешности в освоении образовательной программы, средний балл академической успеваемости 4,17, а суммарный балл вступительных испытаний составляет 62,56.

Третий кластер, состоящий из 301 студента, можно охарактеризовать как кластер наиболее успешных студентов, средний балл академической успеваемости которых составляет 4,51, а суммарный балл вступительных испытаний – 78,16, что значительно выше, чем у представителей предыдущих двух кластеров. Обучающиеся третьего кластера являются самыми молодыми, средний возраст которых составляет 21 год, преимущественно это студенты женского пола, очной формы обучения, обучающиеся на бакалавриате или специалитете на бюджетной основе.

Следующий этап интеллектуального анализа образовательных данных заключается в прогнозировании успеваемости студентов в вузе, для реализации которого был применен метод классификации с помощью деревьев решений.

С помощью процедуры кросс-валидации исходная совокупность была разделена на обучающую и тестовую выборки, объем которых составил 986 и 109 студентов соответственно. Качество классификации с помощью дерева решений оценено на

основе анализа матрицы запутанности (табл. 3) и с помощью показателей полноты, чувствительности, точности и специфичности:

- модель полностью распознает третий (наиболее успешный с точки зрения успеваемости) класс студентов, в пользу чего свидетельствуют все показатели качества классификации;
- для первого класса показатели полноты и чувствительности составили 0,997, для второго класса показатели точности и специфичности составили 0,998, в связи с тем, что один студент из первого класса по методу деревьев решений был ошибочно отнесен ко второму классу;
- общая точность классификации составила 99,9%.

Результаты классификации обучающихся методом дерева решений позволяют представить в структурированном виде следующие правила классификации вида «если..., то...»:

- если студент обучается на платной основе бакалавриата или специалитета заочной формы обучения, старше 22,5 лет, его суммарный балл вступительного испытания варьирует в пределах от 54,5 до 55,5, средний балл академической успеваемости менее 3,76, то объект относится к первому классу;
- если студент обучается на платной основе бакалавриата или специалитета очной формы обучения, младше 22,5 лет, его суммарный балл вступительного испытания более 55,5, но менее 70,5, средний балл академической успеваемости ниже 3,76, то объект относится ко второму классу;
- если студент женского пола, обучается на бюджетной основе бакалавриата или специалитета очной формы обучения, младше 22,5 лет, его средний балл академической успеваемости выше 3,76, суммарный балл вступительного испытания более 70,5, то объект относится к третьему классу.

В результате проведенного исследования, связанного с применением методов машинного обучения для анализа академической успеваемости студентов, методами корреляционно-регрессионного анализа и непараметрической статистики, было выявлено, что на средний балл академической успеваемости статистически значимое влияние оказывают такие факторы, как суммарный балл вступительных испытаний, курс, основа и форма обучения, уровень подготовки и пол обучающегося.

Многомерная классификация обучающихся с помощью кластерного анализа и деревьев решений позволила сформировать правила отнесения студента к классу с тем или иным уровнем освоения образовательной программы в зависимости от его пола, возраста, уровня, формы и основы обучения, результатов вступительных

испытаний и среднего балла успеваемости. Полученные правила классификации можно использовать для выявления студентов, являющихся группой риска, с низким уровнем успеваемости и высокой вероятностью отчисления; принятия корректирующих и предупреждающих действий в рамках учебного процесса; подбора для каждой группы оптимальной технологии и наиболее эффективных методов обучения, что будет способствовать сохранности контингента обучающихся и повышению качества образовательного процесса.

Таблица 1

Результаты оценки линейной модели множественной регрессии методом наименьших квадратов

Table 1

Results of evaluation of the linear model of multiple regression by the least square method

Фактор	Коэффициент	Стандартная ошибка	t-статистика	Уровень значимости
Свободный коэффициент	3,268	0,068	47,726	0***
Суммарный балл вступительных испытаний	0,011	0,001	10,421	0***
Курс	0,019	0,011	1,679	0,094*
Основа обучения	0,114	0,028	4,033	0***
Форма обучения	0,293	0,026	11,116	0***
Уровень подготовки	-0,248	0,036	-6,95	0***
Пол	0,25	0,026	9,575	0***
Торговое дело	-0,173	0,06	-2,903	0,004***

Примечание. * Значимость на 10%-ном уровне; *** значимость на 1%-ном уровне.

Источник: авторская разработка

Source: Authoring

Таблица 2

Результаты кластеризации методом k-средних

Table 2

K-means clustering results

Показатель	Кластер 1	Кластер 2	Кластер 3
Количество объектов	351	443	301
Средний балл академической успеваемости	3,95	4,17	4,51
Суммарный балл вступительных испытаний	47,34	62,56	78,16
Возраст	23,58	21,33	21,12
Основа обучения (1 – бюджетная, 0 – платная)	0,24	0,37	0,75
Форма обучения (1 – очная, 0 – заочная)	0,34	0,58	0,74
Уровень подготовки (1 – бакалавриат/специалитет, 0 – магистратура)	0,74	0,76	0,77
Пол (1 – женский, 0 – мужской)	0,63	0,7	0,77

Источник: авторская разработка

Source: Authoring

Таблица 3**Матрица запутанности классификации обучающихся в вузе методом дерева решений****Table 3****Entanglement matrix for classification of university students under the decision tree method**

Прогнозный класс	Реальный класс			Итого...
	1-й	2-й	3-й	
1-й	350	0	0	350
2-й	1	443	0	444
3-й	0	0	301	301
Итого...	351	443	301	1 095

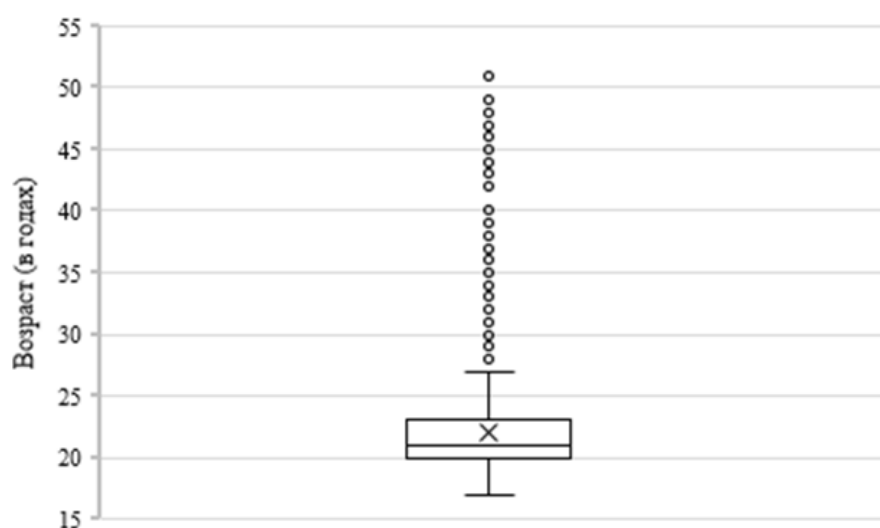
Источник: авторская разработка*Source:* Authoring**Рисунок 1****Диаграмма «ящик с усами» для показателя «возраст»****Figure 1****Box-and-Whisker plot for the Age indicator***Источник:* авторская разработка*Source:* Authoring

Рисунок 2

График зависимости между показателями «средний балл академической успеваемости» и «суммарный балл вступительных испытаний»

Figure 2

Dependency graph for *Average Academic Score* and *Total Score of Entrance Examinations* indicators



Источник: авторская разработка

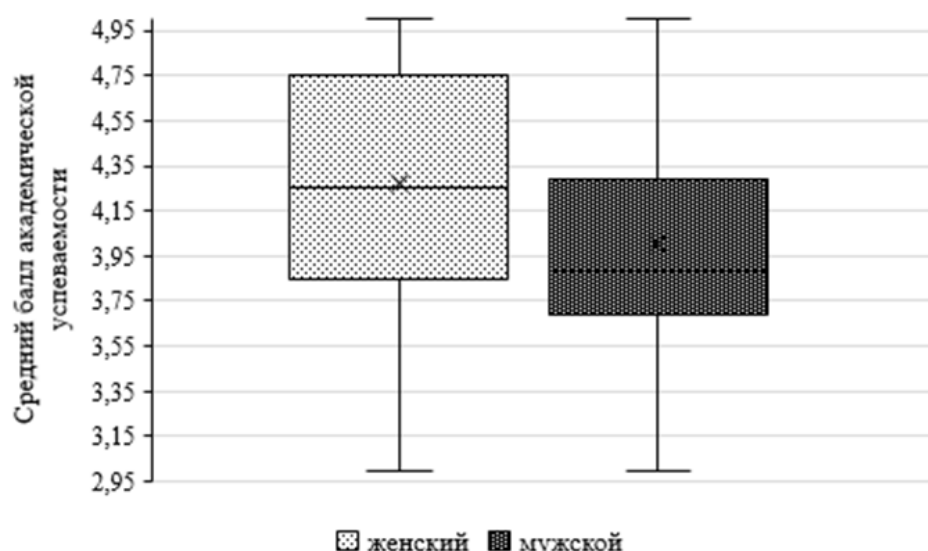
Source: Authoring

Рисунок 3

Диаграмма «ящик с усами» для показателя «средний балл академической успеваемости» в зависимости от пола

Figure 3

Box-and-Whisker plot for the *Average Academic Performance Score* indicator, depending on gender



Источник: авторская разработка

Source: Authoring

Список литературы

1. *Baker R.S., Inventado P.S.* Educational Data Mining and Learning Analytics. In: Larusson J., White B. (eds) *Learning Analytics*. Springer, New York, 2014, pp. 61–75. URL: https://doi.org/10.1007/978-1-4614-3305-7_4
2. *Siti Khadijah Mohamad, Zaidatun Tasir.* Educational Data Mining: A Review. *Procedia – Social and Behavioral Sciences*, 2013, vol. 97, pp. 320–324. URL: <https://doi.org/10.1016/j.sbspro.2013.10.240>
3. *Замков О.О., Пересецкий А.А.* ЕГЭ и академические успехи студентов бакалавриата МИЭФ НИУ ВШЭ // *Прикладная эконометрика*. 2013. № 2. С. 93–114. URL: <https://cyberleninka.ru/article/n/ege-i-akademicheskie-uspehi-studentov-bakalavriata-mief-niu-vshe-1/viewer>
4. *Захарова И.Г.* Методы машинного обучения для информационного обеспечения управления профессиональным развитием студентов // *Образование и наука*. 2018. Т. 20. № 9. С. 91–114. URL: <https://doi.org/10.17853/1994-5639-2018-9-91-114>
5. *Захарова И.Г.* Big Data и управление образовательным процессом // *Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates*. 2017. Т. 3. № 1. С. 210–219. URL: <https://doi.org/10.21684/2411-197X-2017-3-1-210-219>
6. *Русаков С.В., Русакова О.Л., Посохина К.А.* Нейросетевая модель прогнозирования группы риска по успеваемости студентов первого курса // *Современные информационные технологии и ИТ-образование*. 2018. № 4. С. 815–822. URL: <https://doi.org/10.25559/SITITO.14.201804.815-822>
7. *Помян С.В., Белоконь О.С.* Прогноз результатов успеваемости студентов вуза на основе марковских процессов // *Вестник Вятского государственного университета*. 2020. № 4. С. 63–73. URL: <https://cyberleninka.ru/article/n/prognoz-rezultatov-uspevaemosti-studentov-vuza-na-osnove-markovskih-protsessov>
8. *Салал Я.К., Абдуллаев С.М.* Мониторинг качества образования и внедрение индивидуального обучения: демонстрация подходов и алгоритмов интеллектуального анализа образовательных данных // *Известия ЮФУ. Технические науки*. 2020. № 3. С. 112–122. URL: <https://doi.org/10.18522/2311-3103-2020-3-112-122>
9. *Фиофанова О.А.* Анализ больших данных в сфере образования: методология и технологии. М.: Дело, 2020. 200 с.

10. *Частикова В.А., Псеуш А.Г.* Интеллектуальный анализ данных при построении индивидуальных образовательных траекторий // Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. 2021. № 2. С. 66–71.
URL: <https://doi.org/10.53598/2410-3225-2021-2-281-66-71>
11. *Котова Е.Е.* Прогнозирование успешности обучения в интегрированной образовательной среде с применением инструментов онлайн аналитики // Компьютерные инструменты в образовании. 2019. № 4. С. 55–80.
URL: <https://doi.org/10.32603/2071-2340-2019-4-55-80>
12. *Белоножко П.П., Карпенко А.П., Храмов Д.А.* Анализ образовательных данных: направления и перспективы применения // Наукоедение. 2017. Т. 9. № 4.
URL: <http://naukovedenie.ru/PDF/15TVN417.pdf>
13. *Чернышова Н.А.* Связь результатов ЕГЭ и академических успехов студентов в сельскохозяйственном вузе // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Социальные науки. 2017. № 1. С. 171–177.
URL: <https://cyberleninka.ru/article/n/svyaz-rezultatov-ege-i-akademicheskikh-uspehov-studentov-v-selskohozyaystvennom-vuze>
14. *Ширинкина Е.В.* Методы интеллектуального анализа данных и образовательной аналитики // Современное образование. 2022. № 1.
URL: https://nbpublish.com/library_read_article.php?id=37582
15. *Харламова И.Ю.* Прогнозирование успеваемости студентов первого курса по результатам сдачи единого государственного экзамена // Базис. 2017. № 1. С. 57–59. URL: <https://cyberleninka.ru/article/n/prognozirovanie-uspevaemosti-studentov-pervogo-kursa-po-rezultatam-sdachi-edinogo-gosudarstvennogo-ekzamena>
16. *Токтарова В.И., Попова О.Г.* Анализ образовательных данных взаимосвязи успешности обучения и поведения студентов в цифровой образовательной среде вуза // Информатика и образование. 2022. Т. 37. № 4. С. 54–63.
URL: <https://doi.org/10.32517/0234-0453-2022-37-4-54-63>
17. *Котова Е.Е., Писарев А.С.* Задача классификации учащихся с использованием методов интеллектуального анализа данных // Известия СПбГЭТУ ЛЭТИ. 2019. № 4. С. 32–42.
URL: https://izv.etu.ru/assets/files/izvestiya-4_2019_p032-042.pdf
18. *Бантикова О.И., Туктамышева Л.М.* Региональные демографические процессы: состояние и перспективы // Азимут научных исследований: экономика и управление. 2021. Т. 10. № 1. С. 66–71.
URL: <https://cyberleninka.ru/article/n/regionalnye-demograficheskie-protsessy-sostoyanie-i-perspektivy?ysclid=lc4zko78i3214394833>

Информация о конфликте интересов

Мы, авторы данной статьи, со всей ответственностью заявляем о частичном и полном отсутствии фактического или потенциального конфликта интересов с какой бы то ни было третьей стороной, который может возникнуть вследствие публикации данной статьи. Настоящее заявление относится к проведению научной работы, сбору и обработке данных, написанию и подготовке статьи, принятию решения о публикации рукописи.

pISSN 2073-039X
eISSN 2311-8725

Analysis of Intellectual Capital

MODELING THE ACADEMIC PERFORMANCE OF STUDENTS BASED ON INTELLIGENT ANALYSIS OF EDUCATIONAL DATA

Viktoriya V. BOBROVA ^a,

Ol'ga I. BANTIKOVA ^{b,*},

Vlada A. NOVIKOVA ^c

^a Orenburg State University,
Orenburg, Russian Federation
bobrova1971@mail.ru
<https://orcid.org/0000-0003-3558-5662>

^b Russian Biotechnological University (BIOTECH University),
Moscow, Russian Federation
bantikova777@mail.ru
<https://orcid.org/0000-0002-0577-6276>

^c MIREA – Russian Technological University,
Moscow, Russian Federation
vlanovickova@gmail.com
ORCID: not available

* Corresponding author

Article history:

Article No. 565/2022
Received 13 Dec 2022
Received in revised form
24 December 2022
Accepted 9 Feb 2023
Available online
28 February 2023

JEL classification: C01,
C38, C51, C53, C55

Keywords: educational
data, data mining,
academic performance,
modeling, machine
learning

Abstract

Subject. The article considers the application of machine learning methods to analyze students' academic performance.

Objectives. The aims are to identify factors influencing the academic performance of students, detect hidden patterns, useful and interpretable knowledge about the results of educational process and its participants, using the intellectual analysis of educational data.

Methods. The study rests on methods of econometric modeling, multidimensional classification, and big data clustering.

Results. The developed models of intellectual analysis of educational data enable to perform a comparative analysis of students and forecast the level of student's mastering an educational program, depending on factors like the total score of entrance tests, average score of academic performance, basis and form of education, course, the level of training, student's gender and age.

Conclusions. The results of the application of machine learning methods to analyze academic performance will help differentiate effective teaching methods and technologies for groups of students with different levels of academic results, timely take corrective actions regarding students from risk group. Eventually, this will contribute to retention of students and improvement of the educational process quality.

© Publishing house FINANCE and CREDIT, 2022

Please cite this article as: Bobrova V.V., Bantikova O.I., Novikova V.A. Modeling the Academic Performance of Students Based on Intelligent Analysis of Educational Data. *Economic Analysis: Theory and Practice*, 2023, vol. 22, iss. 2, pp. 235–253.
<https://doi.org/10.24891/ea.22.2.235>

References

1. Baker R.S., Inventado P.S. Educational Data Mining and Learning Analytics. In: Larusson J., White B. (eds) *Learning Analytics*. Springer, New York, 2014, pp. 61–75. URL: https://doi.org/10.1007/978-1-4614-3305-7_4
2. Siti Khadijah Mohamad, Zaidatun Tasir. Educational Data Mining: A Review. *Procedia – Social and Behavioral Sciences*, 2013, vol. 97, pp. 320–324. URL: <https://doi.org/10.1016/j.sbspro.2013.10.240>
3. Zamkov O.O., Peresetskii A.A. [Russian Unified National Exams (UNE) and academic performance of ICEF HSE students]. *Prikladnaya ekonometrika = Applied Econometrics*, 2013, no. 2, pp. 93–114. URL: <https://cyberleninka.ru/article/n/ege-i-akademicheskie-uspehi-studentov-bakalavriata-mief-niu-vshe-1/viewer> (In Russ.)
4. Zakharova I.G. [Machine Learning Methods of Providing Informational Management Support for Students' Professional Development]. *Obrazovanie i nauka = The Education and Science Journal*, 2018, vol. 20, no. 9, pp. 91–114. (In Russ.) URL: <https://doi.org/10.17853/1994-5639-2018-9-91-114>
5. Zakharova I.G. [“Big Data and Educational Process Management”]. *Vestnik Tyumenskogo gosudarstvennogo universiteta. Gumanitarnye issledovaniya. Humanitates = Tyumen State University Herald. Humanities Research. Humanitates*, 2017, vol. 3, no. 1, pp. 210–219. (In Russ.) URL: <https://doi.org/10.21684/2411-197X-2017-3-1-210-219>
6. Rusakov S.V., Rusakova O.L., Posokhina K.A. [Neural network model of predicting the risk group for the accession of students of the first course]. *Sovremennye informatsionnye tekhnologii i IT-obrazovanie = Modern Information Technology and IT-education*, 2018, no. 4, pp. 815–822. (In Russ.) URL: <https://doi.org/10.25559/SITITO.14.201804.815-822>
7. Pomyan S.V., Belokon' O.S. [Forecast of the results of academic performance of university students based on Markov processes]. *Vestnik Vyatskogo gosudarstvennogo universiteta = Herald of Vyatka State University*, 2020, no. 4, pp. 63–73. URL: <https://cyberleninka.ru/article/n/prognoz-rezultatov-uspevaemosti-studentov-vuza-na-osnove-markovskih-protsessov> (In Russ.)
8. Salal Ya.K., Abdullaev S.M. [Monitoring of the education quality and implementing of individual learning: Demonstration of approaches and educational data mining algorithms]. *Izvestiya YuFU. Tekhnicheskie nauki = Izvestiya SFedU. Engineering Sciences*, 2020, no. 3, pp. 112–122. (In Russ.) URL: <https://doi.org/10.18522/2311-3103-2020-3-112-122>

9. Fiofanova O.A. *Analiz bol'shikh dannykh v sfere obrazovaniya: metodologiya i tekhnologii* [Big data analysis in education: Methodology and technology]. Moscow, Delo Publ., 2020, 200 p.
10. Chastikova V.A., Pseush A.G. [Data mining in the construction of individual educational trajectories]. *Vestnik Adygeiskogo gosudarstvennogo universiteta. Seriya 4: Estestvenno-matematicheskie i tekhnicheskie nauki = Bulletin of Adyghe State University. Series Natural-Mathematical and Technical Sciences*, 2021, no. 2, pp. 66–71. (In Russ.) URL: <https://doi.org/10.53598/2410-3225-2021-2-281-66-71>
11. Kotova E.E. [Prediction of Learning Success in an Integrated Educational Environment Using Online Analytics Tool]. *Komp'yuternye instrumenty v obrazovanii = Computer Tools in Education*, 2019, no. 4, pp. 55–80. (In Russ.) URL: <https://doi.org/10.32603/2071-2340-2019-4-55-80>
12. Belonozhko P.P., Karpenko A.P., Khramov D.A. [Analysis of educational data: Directions and prospects of application]. *Naukovedenie*, 2017, vol. 9, no. 4. (In Russ.) URL: <http://naukovedenie.ru/PDF/15TVN417.pdf>
13. Chernyshova N.A. [Correlation between the results of the unified State examination and achievement of students in an agricultural university]. *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo. Seriya: Sotsial'nye nauki = Vestnik of Lobachevsky State University of Nizhny Novgorod. Social Sciences*, 2017, no. 1, pp. 171–177. URL: <https://cyberleninka.ru/article/n/svyaz-rezultatov-ege-i-akademicheskikh-uspehov-studentov-v-selskohozyaystvennom-vuze> (In Russ.)
14. Shirinkina E.V. [Methods of data mining and educational analytics]. *Sovremennoe obrazovanie*, 2022, no. 1. (In Russ.) URL: https://nbpublish.com/library_read_article.php?id=37582
15. Kharlamova I.Yu. [The prediction of academic performance of students of the first course by results of unified State exam]. *Bazis = Basis*, 2017, no. 1, pp. 57–59. URL: <https://cyberleninka.ru/article/n/prognozirovanie-uspevaemosti-studentov-pervogo-kursa-po-rezultatam-sdachi-edinogo-gosudarstvennogo-ekzamena> (In Russ.)
16. Toktarova V.I., Popova O.G. [An analysis of educational data on the correlation between learning success and students' behavior in the university digital educational environment]. *Informatika i obrazovanie = Informatics and Education*, 2022, vol. 37, no. 4, pp. 54–63. (In Russ.) URL: <https://doi.org/10.32517/0234-0453-2022-37-4-54-63>
17. Kotova E.E., Pisarev A.S. [The problem of classification of students using the methods of intellectual data analysis]. *Izvestiya SPbGETU LETI = LETI Transactions on Electrical Engineering & Computer Science*, 2019, no. 4, pp. 32–42. URL: https://izv.etu.ru/assets/files/izvestiya-4_2019_p032-042.pdf (In Russ.)

18. Bantikova O.I., Tuktamysheva L.M. [Regional demographic processes: Status and prospects]. *Azimut nauchnykh issledovaniy: ekonomika i upravlenie = Azimuth of Scientific Research: Economics and Administration*, 2021, vol. 10, no. 1, pp. 66–71. URL: <https://cyberleninka.ru/article/n/regionalnye-demograficheskie-protsessy-sostoyanie-i-perspektivy?ysclid=lc4zko78i3214394833> (In Russ.)

Conflict-of-interest notification

We, the authors of this article, bindingly and explicitly declare of the partial and total lack of actual or potential conflict of interest with any other third party whatsoever, which may arise as a result of the publication of this article. This statement relates to the study, data collection and interpretation, writing and preparation of the article, and the decision to submit the manuscript for publication.