

Преподавание анализа данных и машинного обучения в вузе: обобщение опыта и перспективы

В. И. Дорофеева^{1*}, С. П. Строев¹, Д. Ю. Дорофеев²

¹ ФГБОУ ВО «Орловский государственный университет имени И.С. Тургенева», г. Орёл, Российская Федерация

Адрес: 302001, Российская Федерация, г. Орёл, ул. Комсомольская, д. 95

* vdorofey@mail.ru

² ФГАОУ ВО «Национальный исследовательский технологический университет «МИСИС», г. Москва, Российская Федерация

Адрес: 119049, Российская Федерация, г. Москва, Ленинский пр., д. 4, стр. 1.

Аннотация

В работе обосновывается вывод об актуальности подготовки профессиональных кадров, которые могут быть вовлечены в выполнение национального проекта «Цифровые технологии», в частности, как специалисты в области технологий, связанных с анализом данных, обработкой больших данных и машинным обучением. Очевидно, что особенно важным является изучение технологий, связанных с анализом данных, обработкой больших данных и машинным обучением, для студентов физико-математических и ИТ-направлений подготовки. Однако знакомство с задачами, которые возникают в анализе данных, приводит к мысли, что этот раздел информатики может быть особенно полезен и студентам естественно-научных, медицинских направлений подготовки, а также социологам, историкам, психологам и пр. Это определяется накоплением значительного объема информации по различным прикладным задачам в области медицины, социологии, психологии, где появляется возможность обрабатывать и исследовать данные для исследования и прогнозирования развития различных ситуаций. С учетом междисциплинарной направленности рассматриваемых прикладных задач, становится очевидным, что при освоении соответствующих компетенций студентам придется столкнуться с серьезными проблемами изучения абсолютно новых для них разделов науки. В статье предпринята попытка систематизации опыта работы по обучению студентов ФГБОУ ВО «Орловский государственный университет имени И.С.Тургенева» технологиям анализа данных и машинного обучения, который был получен в течение последних четырех лет в рамках обучения по основным и дополнительным образовательным программам. Целью проводимого исследования является обобщение полученного опыта и поиск решений по преодолению трудностей, связанных с разработкой и модификацией дисциплин, направленных на изучение технологий анализа данных и машинного обучения. Демонстрируются разработанные дисциплины (модули) по изучению анализа данных и машинного обучения для студентов различных направлений подготовки. Отмечается, что для выполнения современных задач в области информационных технологий, подготовленным специалистам необходимо иметь навыки не только в программировании, но и быть успешными в анализе междисциплинарных проблем, что требует от обучающихся развития широкого кругозора и стремления к получению современной информации.

Ключевые слова: технологии анализа данных, машинное обучение, язык программирования Python

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Дорофеева В. И., Строев С. П., Дорофеев Д. Ю. Преподавание анализа данных и машинного обучения в вузе: обобщение опыта и перспективы // Современные информационные технологии и ИТ-образование. 2024. Т. 20, № 1. С. 251-259. <https://doi.org/10.25559/SITITO.020.202401.251-259>

© Дорофеева В. И., Строев С. П., Дорофеев Д. Ю., 2024



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Teaching Data Analysis and Machine Learning at University: Generalization of Experience and Perspectives

V. I. Dorofeyeva^{a*}, S. P. Stroev^a, D. Yu. Dorofeyev^b

^a Orel State University named after I.S. Turgenev, Orel, Russian Federation

Address: 95 Komsomolskaya St., Orel 302026, Russian Federation

* vdorofey@mail.ru

^b National University of Science and Technology "MISIS", Moscow, Russian Federation

Address: 4 Leninsky pr., building 2, Moscow 119049, Russian Federation

Abstract

The work substantiates the conclusion about the relevance of training professional personnel who can be involved in the implementation of the national project "Digital Technologies", in particular, as specialists in the field of technologies related to data analysis, big data processing and machine learning. Obviously, it is especially important to study technologies related to data analysis, big data processing and machine learning for students of physics, mathematics and IT training. However, familiarity with the tasks that arise in data analysis leads to the idea that this section of computer science can be especially useful for students of natural sciences, medical fields of training, as well as sociologists, historians, psychologists, etc. This is determined by the accumulation of a significant amount of information on various applied tasks in the field of medicine, sociology, psychology, where it becomes possible to process and study data to study and predict the development of various situations. Taking into account the interdisciplinary focus of the applied problems under consideration, it becomes obvious that when mastering the relevant competencies, students will have to face serious problems in studying branches of science that are completely new to them. The article makes an attempt to systematize the experience of training students of the Federal State Budgetary Educational Institution of Higher Education "Oryol State University named after I.S. Turgenev" in data analysis and machine learning technologies, which was obtained over the past four years as part of training in basic and additional educational programs. The purpose of the research is to generalize the experience gained and find solutions to overcome difficulties associated with the development and modification of disciplines aimed at studying data analysis and machine learning technologies. The developed disciplines (modules) for studying data analysis and machine learning are demonstrated for students in various areas of training. It is noted that in order to fulfill modern tasks in the field of information technology, trained specialists need to have skills not only in programming, but also to be successful in analyzing interdisciplinary problems, which requires students to develop a broad outlook and the desire to obtain modern information.

Keywords: data analysis technologies, machine learning, Python programming language

Conflict of interests: The authors declare no conflict of interest.

For citation: Dorofeyeva V.I., Stroev S.P., Dorofeyev D.Yu. Teaching Data Analysis and Machine Learning at University: Generalization of Experience and Perspectives. *Modern Information Technologies and IT-Education*. 2024;20(1):251-259. <https://doi.org/10.25559/SITITO.020.202401.251-259>



Введение

Национальная программа «Цифровая экономика Российской Федерации», утвержденная в 2018 г., содержит в своем составе федеральный проект «Цифровые технологии»¹. Среди дорожных карт этого проекта имеются разработанные карты по технологиям обработки больших данных, систем распределенного реестра². В ноябре 2021 г. в России утвердили первые национальные стандарты в области больших данных, которые разработаны Центром компетенций НТИ по большому данным МГУ имени М.В. Ломоносова и Институтом развития информационного общества³. Стандарты призваны обеспечить взаимодействие и взаимопонимание в предметной области «большие данные» между органами власти, коммерческими компаниями и научно-образовательным сообществом. Наряду с большими данными, речь так же идет о технологиях анализа данных и машинного обучения в целом. В ближайшем будущем планируется обсуждение нацпроекта «Экономика данных», в рамках которого говорится о сборе и передаче данных, создании инфраструктуры вычисления и хранения данных с использованием отечественных оборудования, технологий и программного обеспечения, обеспечении безопасности данных⁴.

Все вышеизложенное позволяет сделать вывод о необходимости формирования кадров, которые могли бы быть вовлечены в выполнении задач национальных проектов, а значит обосновать заключение об актуальности подготовки специалистов в области технологий, связанных с анализом данных, обработкой больших данных и машинным обучением и т.п.

Кроме того, становится очевидным, что для выполнения современных задач в области информационных технологий, подготовленным специалистам необходимо иметь навыки не только в программировании, но и быть успешными в анализе междисциплинарных проблем, что требует от обучающихся развития широкого кругозора и стремления к получению современной информации [1-4].

Очевидно, что особенно актуальным является изучение технологий, связанных с анализом данных, обработкой больших данных и машинным обучением, для студентов физико-математических и ИТ-направлений подготовки. Однако знакомство с задачами, которые возникают в анализе данных, приводит к мысли, что этот раздел информатики может быть особенно полезен и студентам естественно-научных, медицинских направлений подготовки, а также социологам, историкам, психологам и пр. Это определяется накоплением значительного объема информации по различным прикладным задачам медицины, социологии, психологии, где появляется возмож-

ность обрабатывать и исследовать данные для прогнозирования развития различных ситуаций [5-10].

Соответственно, возникает необходимость преподавания студентам различных факультетов и институтов дисциплин, связанным с основами программированием, статистикой, математикой и информационными технологиями для анализа данных. С учетом междисциплинарной направленности рассматриваемых прикладных задач, становится очевидным, что при освоения соответствующих компетенций студентам придется столкнуться с серьезными проблемами изучения абсолютно новых для них разделов науки.

При этом следует отметить, что вопросы преподавания дисциплин, связанных с анализом данных, являются актуальными еще и потому, что эти дисциплины являются сравнительно новыми для многих вузов, и появились в учебных планах образовательных программ буквально в последнее десятилетие. В статье предпринята попытка систематизации опыта работы по обучению студентов ФГБОУ ВО «Орловский государственный университет имени И.С.Тургенева» технологиям анализа данных и машинного обучения, который был получен в течение последних четырех лет в рамках обучения по основным и дополнительным образовательным программам.

Цель исследования

Целью проводимого исследования является обобщение полученного опыта и поиск решений по преодолению трудностей, связанных с разработкой и модификацией дисциплин, направленных на изучение технологий анализа данных и машинного обучения.

Анализ ситуации и выработка рекомендаций могут быть направлены на решение следующих задач:

- разработка образовательных курсов для различных групп обучающихся с учетом их профильной подготовки;
- формирование рабочих программ дисциплин в объеме, наиболее подходящем для обучения студентов с учетом их базового образования;
- разработка дисциплин базового и продвинутого уровня для студентов одного направления подготовки бакалавриата и магистратуры;
- выработка рекомендаций по используемым языкам программирования и программному обеспечению, учебникам и онлайн-ресурсам;
- обучение действиям в различных ситуациях: принятию решений на основе анализа данных и машинного обучения, оптимальному использованию данных из различных источников информации, применению искусственного интеллекта и машинного обучения для решения прикладных задач, выстраиванию процесса работы с данными

¹ Паспорт национальной программы «Цифровая экономика Российской Федерации»: утв. президиумом Совета при Президенте РФ по стратегическому развитию и нац. проектам (протокол от 24 декабря 2018 г. № 16). Москва, 2018 [Электронный ресурс]. URL: <http://static.government.ru/media/files/urKHm0gTPPnzJlaKw3M5cNLo6gczMkPF.pdf> (дата обращения: 15.11.2023).

² Паспорт федерального проекта «Цифровые технологии»: утв. президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности (протокол от 28 мая 2019 г. № 9). Москва, 2019 [Электронный ресурс]. URL: https://www.consultant.ru/document/cons_doc_LAW_328937/ (дата обращения: 15.11.2023).

³ ГОСТ Р 59925-2021 Информационные технологии. Большие данные. Техническое задание. Требования к содержанию и оформлению: национальный стандарт Российской Федерации: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 2 декабря 2021 г. № 1684-ст: введен впервые: дата введения 2022-03-01. М.: Российский институт стандартизации, 2022.

⁴ В России появится новый нацпроект – «Экономика данных» [Электронный ресурс] // Минцифры России. 13.07.2023. URL: <https://digital.gov.ru/ru/events/45686/> (дата обращения: 15.11.2023).



- ми между техническими и профильными учреждениями;
- разработка банка заданий для практических и лабораторных занятий, а так же фонда оценочных средств для оценки освоения дисциплин;
- организация проектной, научно-исследовательской деятельности и междисциплинарных исследований студентов на основе цифровых технологий.

Основная часть

При значительной развитости статистики и статистического анализа данных, именно наука о данных, как отдельный раздел, сравнительно молодая. Считается, что в современном варианте она получила развитие не более 20 лет назад [11]. В связи с тем, что наука о данных является довольно молодой, то это же можно сказать и об образовании в области науки о данных. Ряд публикаций и учебников, в которых рассматриваются вопросы преподавания анализа данных, датированы в большинстве случаев, 2015-2018 годами [12-15]. Фактически, именно в это время начинают формироваться отдельные положения и разделы дисциплин, которые формируют знания, умения и навыки у студентов. В 2018 году Хикс и Иризарри [16] предложили основные руководящие принципы преподавания науки о данных: разнообразие тематических исследований; использование компьютерных вычислений в каждый этап изучения; изучение теории, но без значительного погружения в математические термины; решение прикладных задач по аналогии с опытом реального аналитика данных; поддержка и демонстрация значимости критического мышления во время решения задач.

Данные рекомендации позволили выделить основные разделы, подлежащие изучению, а именно, исследование и вычисление данных, обработка данных, визуализация данных, моделирование данных и передача данных. Однако разработка отдельных курсов продолжала и продолжает оставаться актуальной задачей. В работах [17-20] рассматриваются различные концепции преподавания науки о данных, причем авторы так же анализируют современные на тот момент разработанные новые курсы, выделяя их сильные и слабые стороны.

В работе [11] приводятся результаты анкетирования, которые проводили авторы исследования среди разработчиков дисциплин по анализу данных. Были выявлены основные проблемы, с которыми столкнулись разработчики новых дисциплин, а также сделаны выводы и рекомендации, позволяющие в дальнейшем оказать помощь в модификации уже существующих курсов и в создании новых.

В сентябре 2022 г. вопросы преподавания анализа данных рассматривались в рамках круглого стола «Опыт преподавания анализа данных и Data Science в вузах России. Тенденции, перспективы, направления развития», где подчеркивалась значимость изучения этих дисциплин, а также анализировались возникающие проблемы⁵.

Среди основных характеристик, которые, как усложняют, так и упрощают задачу разработки новой дисциплины, есть ее междисциплинарный характер. С одной стороны, это дает возмож-

ность использования науки о данных в самых разнообразных науках, как то, математика, физика, медицина, социологи, банковская деятельность, коммерческая деятельность, военное дело и пр. Но, с другой стороны, любой обучающийся, который начинает осваивать науки о данных, понимает, что ему необходимо получить знания как в математике и статистике, так и знания языков программирования и программного обеспечения, а также специальные знания в исследуемой области. Все вместе это существенно усложняет задачу качественного изучения предмета, поскольку у обучающегося должна быть мотивация изучить абсолютно новые для него дисциплины. Среди основных проблем, с которыми сталкиваются разработчики дисциплин по анализу данных можно выделить следующие:

- проблема ограничения объема дисциплины по количеству часов и семестров;
- отсутствие четкого понимания начального уровня подготовки обучающихся по математике, статистике, информационным технологиям и программированию;
- отсутствие четких определений для некоторых понятий, иногда конкретных алгоритмов и рекомендаций к действию;
- разнообразие информационных технологий, языков программирования и программного обеспечения, что приводит в некоторое замешательство начинающих преподавателей;
- сложность определения заранее уровня прикладных задач по реальным датасетам (наборам данных);
- проблема уровня знаний конкретного преподавателя, а именно наличие у него образования, опыта решения, получение дополнительного образования;
- отсутствие единых учебно-методических рекомендаций при разработке нового курса и др.

В Орловском государственном университете имени И.С. Тургенева к вопросу о необходимости преподавания анализа данных и машинного обучения пришли в 2020 г. Было принято решения о внесении изменений в основные образовательные программы с целью введения в обучение информационных технологий, дающих обучающимся компетенции в области анализа данных и машинного обучения. Для обучения студентов современным технологиям, преподаватели физико-математического факультета прошли цифровые стажировки по организации преподавания дисциплин, связанных с Data science, а также повышение квалификации в НИУ «Высшая школа экономики» и в Школе анализа данных компании «Яндекс.Практикум».

Вследствии, при разработке дисциплины (образовательного модуля) для изучения анализа данных и машинного обучения в решении задач медицины были использованы компетенции в области ИКТ для медиков. В частности, был изучен опыт ведущих медицинских и технических вузов Российской Федерации [21-24].

В 2021 г. были разработаны новые учебные планы, которые включали дисциплины по анализу данных для обучающихся физико-математического факультета. Стало понятно, что при формировании рабочей программы подобной дисциплины (образовательного модуля) обязательно нужно учесть послед-

⁵ Дорофеева В. И. К вопросу преподавания анализа данных студентам физико-математических направлений подготовки // Информатизация образования – 2022 : сб. материалов Межд. НПК, Липецк, 25-27 мая 2022 года. Липецк : ЛГТУ, 2022. С. 59-61. EDN: QNVUXX



ние тенденции развития IT-технологий, связанные с современными языками программирования, библиотеками анализа данных для качественной обработки данных в различных областях естественных наук.

Опыт преподавания дисциплин о данных показал, что совершенно необходимым является изучения основ языка программирования Python. Как правило, студенты физико-математических направлений изучают эту дисциплину на 1-2 курсе, а студентам остальных факультетов можно рекомендовать

включить отдельные разделы в модуль «Введение в информационные технологии» [25].

Эта дисциплина (образовательный модуль) дает базовые знания по программированию, структурам языка, основным алгоритмам и делает акцент на библиотеках, которые будут использоваться в дальнейшем. В таблице 1 приводятся общие сведения об указанной дисциплине (образовательном модуле).

Таблица 1. Общие данные о дисциплине (образовательном модуле) «Язык программирования Python»
Table 1. General information about the discipline (educational module) "Python Programming Language"

Название дисциплины	Язык программирования Python
О дисциплине	Цель-получение навыков программирования на языке Python, включая знания основных структур языка и особенностей разработки алгоритмов
Объем дисциплины	Всего ауд. часов 52: лек. 20 ч., лаб. раб. 32 ч.; сам. раб. 20 ч., экзамен (36 ч.). общая трудоемкость курса 108 ч. (3 з.е.)
Требования к освоению	Для изучения дисциплины слушатели должны обладать базовыми знаниями и умениями в области информационно-коммуникационных технологий
Программа дисциплины	Список тем (разделов) дисциплины: 1. Введение в Python; Циклы и последовательности. Списки; Функции в Python; Организация общения с пользователем; 2. Модули; Массивы; Многомерные массивы; 3. Графика; Рисование в Matplotlib; Обработка файлов

Источник: здесь и далее в статье все таблицы составлены авторами.

Source: Hereinafter in this article all tables were made by the authors.

Помимо изучения разделов программирования, можно было бы рекомендовать студентам изучить некоторые разделы статистики самостоятельно или в рамках обучения выделить несколько часов для изучения основ статистики на аудиторных занятиях.

Далее студенты начинают осваивать дисциплину «Анализ данных и машинное обучение с использованием Python». В соответствии с целями и задачами дисциплины (образовательного модуля) можно выделить следующие разделы, связанные с анализом данных и машинным обучением, в том числе:

1. Основные сведения.
2. Введение в язык программирования Python (краткое повторение).
3. Процесс анализа данных

4. Работа с данными в библиотеке Pandas.
5. Подготовка данных
6. Исследование и визуализация данных
7. Разбор кейсов: проверка гипотез; A/B тестирование.
8. Обзор процесса машинного обучения
9. Задача кластеризации. Задача классификации и регрессии.

В таблице 2 представлена структура предлагаемой дисциплины «Анализ данных и машинное обучение с использованием Python» или отдельных модулей в рамках более общей дисциплины по информационным технологиям. Приводится примерная почасовая аудиторная тематическая разбивка тем.

Таблица 2. Проектирование структуры дисциплины «Анализ данных и машинное обучение с использованием Python»
Table 2. Designing the structure of the discipline "Data Analysis and Machine Learning using Python"

Основные разделы	Подразделы	Общая трудоемкость
1. Базовые сведения	Тема 1. Информационное сопровождение курса. Знакомство с Google Colaboratory, Jupyter Notebook. Язык разметки markdown.	2
	Тема 2. Введение в язык программирования Python: переменные, операторы, условные операторы и циклы, строки, списки, словари, множества, кортежи, диапазоны, пользовательские функции.	4



Основные разделы	Подразделы	Общая трудоемкость
2. Основы анализа данных	Тема 3. Процесс анализа данных: определение проблемы; извлечение данных; подготовка данных; исследование и визуализация данных; предсказательная модель; проверка модели, тестирование; развертывание.	4
	Тема 4. Работа с данными в библиотеке Pandas.	4
	Тема 5. Подготовка данных: описательные статистики, работа с типами данных, поиск дубликатов, обработка аномальных значений, восстановление пропусков.	8
	Тема 6. Исследование и визуализация данных: разведочный анализ данных, визуализация данных.	8
	Тема 7. Разбор кейсов: проверка гипотез; A/B тестирование.	4
3. Основы машинного обучения	Тема 8. Обзор процесса машинного обучения: подготовка данных; моделирование, оценка. Обучение с учителем и без учителя.	2
	Тема 9. Задача кластеризации.	6
	Тема 10. Задача классификации и регрессии.	8
4. Аттестация	Итоговый проект	
	Итого:	52

Если студенты знакомы с основами языка программирования Python, то в рамках темы 2 можно рекомендовать изучить основы математической статистики.

При проведении итоговой аттестации возможно выполнение проекта, связанного с анализом специальных данных, которые широко представлены в интернет-ресурсах. Это даст возможность студентам получить опыт применения полученных знаний и, в частности, смоделировать реальную ситуацию прогнозирования.

Приведем пример кейса, которые могут быть предложены студентам в рамках итогового выполнения проекта:

Кейс 1. Статистический анализ факторов, влияющих на ожидаемую продолжительность жизни.

Предоставленные данные содержат информацию об иммунизации, смертности, экономических факторах, социальных факторах и других факторах, связанных со здоровьем. Поскольку наблюдения в этом наборе даны для разных стран, необходимо для отдельно взятой страны определить прогностический фактор, который способствует более низкому значению ожидаемой продолжительности жизни. Это поможет подсказать стране, какой области следует уделить особое внимание, чтобы эффективно повысить продолжительность жизни ее населения.

Следует отметить, что на начальном этапе изучения дисциплины студенты очень активно стремятся изучать излагаемые вопросы, так как желание стать аналитиком данных у многих велико, тем более, что он поддерживается информацией о больших зарплатах специалистов по анализу данных. Однако, в процессе обучения ряд студентов сталкивается со сложностями в понимании материала, частичным отторжением некоторых разделов, которые кажутся им незначительными или непонятными. Но, по мере изучения дисциплины и освоения все более увлекательных технологий анализа и визуализации данных, студенты приобретают уверенность и с большим энтузиазмом разбираются в изучаемых вопросах, нежели в середине курса.

Результаты исследования

Таким образом, в результате изучения опыта иностранных и российских авторов, учета проблем и рекомендаций сторонних авторов, анализа собственного опыта преподавания на физико-математическом факультете Орловского государственного университета имени И.С.Тургенева, проведения нескольких программ повышения квалификации по анализу данных и машинному обучению в рамках проекта «Приоритет-2030» авторами были выработаны некоторые рекомендации, которые были использованы для разработки дисциплин (образовательных модулей) по изучению наук о данных для разных уровней базовой подготовки, а именно, для студентов физико-математических специальностей и для студентов, не имеющих специального математического образования. Полученные в результате освоения этих дисциплин знания и навыки позволят будущим специалистам в конкретных областях расширить круг своих возможностей и в дальнейшем успешно применять современные информационные технологии при выполнении научно-исследовательских проектов и выпускных квалификационных работ. Однако следует отметить, что освоение технологий анализа данных и машинного обучения предполагает наличие широкого кругозора у специалиста, а также умения формулировать верные выводы и давать оптимальные рекомендации.

Заключение

Таким образом, в данной работе, на основе изучения стороннего и анализа собственного опыта преподавания анализа данных и машинного обучения, разработаны образовательные дисциплины (модули) для получения компетенций в науках о данных и их последующего использования в профессиональной деятельности.

Сформулированные выводы и рекомендации могут быть на-



правлены на решение задач разработки образовательных курсов для различных групп обучающихся с учетом их профильной подготовки, формирование рабочих программ дисциплин в объеме, наиболее подходящем для обучения студентов с учетом их базового образования; обучение действиям в ситуациях принятия решений на основе анализа данных и машинного

обучения; оптимальному использованию данных из различных источников информации, выстраиванию процесса работы с данными между техническими и профильными учреждениями; организацию проектной, научно-исследовательской деятельности и междисциплинарных исследований студентов на основе цифровых технологий.

Список использованных источников

- [1] Khamitova A. Reimagining higher education learning spaces in a new social contract for education // *Prospects*. 2024. Vol. 54, issue 2. P. 427-431. <https://doi.org/10.1007/s11125-024-09680-1>
- [2] Гузь Н. А. Цифровая трансформация высшего образования: глобальные тренды и антитренды // *Мир науки, культуры, образования*. 2022. № 3(94). С. 99-103. <https://doi.org/10.24412/1991-5497-2022-394-99-104>
- [3] Management of Ungraded Educational Organizations in the Digital Transformation of Education / V. I. Abramov [et al.] // 2022 2nd International Conference on Technology Enhanced Learning in Higher Education (TELE). Lipetsk, Russian Federation: IEEE Press, 2022. P. 14-16. <https://doi.org/10.1109/TELE55498.2022.9801052>
- [4] Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges / R. I. Mukhamediev [et al.] // *Mathematics*. 2022. Vol. 10, issue 15. Article number: 2552. <https://doi.org/10.3390/math10152552>
- [5] Саханевич Д. Ю., Кремин А. Е. Систематизация методов машинного обучения в целях внедрения искусственного интеллекта в социально-экономические процессы региона // *Вестник Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых*. Серия: Экономические науки. 2019. № 4(22). С. 57-65. EDN: PQFDPE
- [6] Боровский А. А. Перспективы применения технологий машинного обучения к обработке больших массивов исторических данных // *Кибернетика и программирование*. 2015. № 1. С. 77-114. <https://doi.org/10.7256/2306-4196.2015.1.13730>
- [7] Саханевич Д. Ю. Исследование подходов и методов применения искусственного интеллекта и машинного обучения в социально-экономических процессах // *Вестник Омского университета*. Серия: Экономика. 2020. Т. 18, № 2. С. 65-79. [https://doi.org/10.24147/1812-3988.2020.18\(2\).65-79](https://doi.org/10.24147/1812-3988.2020.18(2).65-79)
- [8] Гребенюк А. А., Максимова А. С., Лемэр Л. Г. Исследование социальной напряженности на основе больших данных социальных сетей // *Цифровая социология*. 2021. Т. 4, № 4. С. 4-12. <https://doi.org/10.26425/2658-347X-2021-4-4-4-12>
- [9] Косников С. Н., Новикова Д. Н., Сухиненко Д. Д. Моделирование и прогнозирование показателей социально-экономического развития // *Естественно-гуманитарные исследования*. 2022. № 41(3). С. 151-154. EDN: JEPSPG
- [10] Седова Н. О., Перцева И. А., Юрьева О. Д. Анализ эффективности профориентационных мероприятий вуза с использованием инструментов машинного обучения // *Современные информационные технологии и ИТ-образование*. 2021. Т. 17, № 4. С. 1032-1046. <https://doi.org/10.25559/SITITO.17.202104.1032-1046>
- [11] Schwab-McCoy A., Baker C. M., Gasper R. E. Data Science in 2020: Computing, Curricula, and Challenges for the Next 10 Years // *Journal of Statistics and Data Science Education*. 2021. Vol. 29, No. sup1. P. S40-S50. <https://doi.org/10.1080/10691898.2020.1851159>
- [12] Baumer B. A Data Science Course for Undergraduates: Thinking With Data // *The American Statistician*. 2015. Vol. 69. P. 334-342. <https://doi.org/10.1080/00031305.2015.1081105>
- [13] Brunner R. J., Kim E. J. Teaching Data Science // *Procedia Computer Science*. 2016. Vol. 80. P. 1947-1956. <https://doi.org/10.1016/j.procs.2016.05.513>
- [14] Albert J., Rizzo M. An Undergraduate Data Science Program // *JSM 2016 Proceedings. Section on Statistical Education*. Chicago, Illinois: American Statistical Association, 2016. P. 1135-1147. URL: <https://ww2.amstat.org/meetings/proceedings/2016/data/assets/pdf/389584.pdf> (дата обращения: 15.11.2023).
- [15] Majumder M., Cheng X. Focusing on the Needs: Experiences of Developing a Data Science Program // *Journal of Computational and Graphical Statistics*. 2017. Vol. 26, issue 4. P. 779-780. <https://doi.org/10.1080/10618600.2017.1385475>
- [16] Hicks S. C., Irizarry R. A. A Guide to Teaching Data Science // *American Statistician*. 2018. Vol. 72, issue 4. P. 382-391. <https://doi.org/10.1080/00031305.2017.1356747>
- [17] Tang R., Sae-Lim W. Data Science Programs in U.S. Higher Education: An Exploratory Content Analysis of Program Description, Curriculum Structure, and Course Focus // *Education for Information*. 2016. Vol. 32. P. 269-290. <https://doi.org/10.3233/EFI-160977>
- [18] Song I.-Y., Zhu Y. Big Data and Data Science: What Should We Teach? // *Expert Systems*. 2016. Vol. 33, issue 4. P. 364-373. <https://doi.org/10.1111/exsy.12130>
- [19] West J. Teaching Data Science: An Objective Approach to Curriculum Validation // *Computer Science Education*. 2018. Vol. 28, issue 2. P. 136-157. <https://doi.org/10.1080/08993408.2018.1486120>
- [20] Лапенюк М. В., Шестакова Л. Г. Формирование умений бакалавров в области машинного обучения и интеллектуального анализа данных // *Ученые записки Забайкальского государственного университета*. 2024. Т. 19, № 3. С. 17-26. <https://doi.org/10.21209/2658-7114-2024-19-3-17-26>



- [21] Медведева Е. И., Крошилилин С.В., Авачёва Т. Г. Необходимость развития информационных компетенций при подготовке студентов в медицинских вузах // Медицинское образование и профессиональное развитие. 2023. Т. 14, № 1. С. 66-78. <https://doi.org/10.33029/2220-8453-2023-14-1-66-78>
- [22] Hazzan O., Mike K. The Birth of a New Discipline: Data Science Education // Guide to Teaching Data Science. Cham: Springer, 2023. P. 59-72. https://doi.org/10.1007/978-3-031-24758-3_4
- [23] Aljawarneh S., Lara J. A. Data science for analyzing and improving educational processes // Journal of Computing in Higher Education. 2021. Vol. 33, issue 3. P. 545-550. <https://doi.org/10.1007/s12528-021-09299-7>
- [24] Dorofeeva V. I., Stroev S. P., Dorofeeva E. Y. On the Issue of Teaching Data Analysis and Machine Learning to Students of Medical Specialties // 2023 3rd International Conference on Technology Enhanced Learning in Higher Education (TELE). Lipetsk, Russian Federation: IEEE Press, 2023. P. 101-103. <https://doi.org/10.1109/TELE58910.2023.10184352>
- [25] Дорофеева В. И. К вопросу преподавания дисциплин, связанных с анализом данных, студентам направления подготовки «Прикладная математика и информатика» // Ученые записки Орловского государственного университета. 2022. № 4(97). С. 244-246. EDN: TBIJRK

Поступила 15.11.2023; одобрена после рецензирования 16.01.2024; принята к публикации 27.01.2024.

Об авторах:

Дорофеева Виктория Ивановна, заведующий кафедрой информатики физико-математического факультета, ФГБОУ ВО «Орловский государственный университет имени И.С. Тургенева» (302001, Российская Федерация, г. Орёл, ул. Комсомольская, д. 95), кандидат физико-математических наук, доцент, **ORCID: <https://orcid.org/0000-0001-6116-2511>**, vdorofey@mail.ru

Строев Сергей Павлович, заведующий кафедрой алгебры и математических методов в экономике физико-математического факультета, ФГБОУ ВО «Орловский государственный университет имени И.С. Тургенева» (302001, Российская Федерация, г. Орёл, ул. Комсомольская, д. 95), кандидат экономических наук, **ORCID: <https://orcid.org/0000-0002-2271-5264>**, stroewsp@mail.ru

Дорофеев Дмитрий Юрьевич, студент Института новых материалов, ФГАОУ ВО «Национальный исследовательский технологический университет «МИСИС» (119049, Российская Федерация, г. Москва, Ленинский пр., д. 4, стр. 1), **ORCID: <https://orcid.org/0009-0002-6771-211X>**

Все авторы прочитали и одобрили окончательный вариант рукописи.

References

- [1] Khamitova A. Reimagining higher education learning spaces in a new social contract for education. *Prospects*. 2024;54(2):427-431. <https://doi.org/10.1007/s11125-024-09680-1>
- [2] Guz N.A. The digital transformation of higher education: global trends and anti-trends. *Mir Nauki, Kultury, Obrazovaniya*. 2022;(3):99-103. <https://doi.org/10.24412/1991-5497-2022-394-99-104>
- [3] Abramov V.I., Mindzaeva E.V., Beshenkov S.A., Yalamos G.Y., Alexandrovna A.A., Vadimovna D.Y. Management of Ungraded Educational Organizations in the Digital Transformation of Education. In: 2022 2nd International Conference on Technology Enhanced Learning in Higher Education (TELE). Lipetsk, Russian Federation: IEEE Press; 2022. p. 14-16. <https://doi.org/10.1109/TELE55498.2022.9801052>
- [4] Mukhamediev R.I., Popova Y., Kuchin Y., Zaitseva E., Kalimoldayev A., Symagulov A., Levashenko V., Abdoldina F., Gopejenko V., Yakunin K., Muhamedijeva E., Yelis M. Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges. *Mathematics*. 2022;10(15):2552. <https://doi.org/10.3390/math10152552>
- [5] Sakhanevich D.Yu., Kremin A.E. Machine learning methods for building artificial intelligence in socio-economic processes. *Bulletin of the Vladimir State University named after Alexander G. and Nicholas G. Stoletovs. Series: Economics*. 2019;(4):57-65. (In Russ., abstract in Eng.) EDN: PQFDPE
- [6] Borovskii A.A. Prospects for the use of machine learning techniques in processing large volumes of historical data. *Cybernetics and programming*. 2015;(1):77-114. (In Russ., abstract in Eng.) <https://doi.org/10.7256/2306-4196.2015.1.13730>
- [7] Sakhanevich D.Yu. Research of approaches and methods of applying artificial intelligence and machine learning in socio-economic processes. *Herald of Omsk University. Series "Economics"*. 2020;18(2):65-79. (In Russ., abstract in Eng.) [https://doi.org/10.24147/1812-3988.2020.18\(2\).65-79](https://doi.org/10.24147/1812-3988.2020.18(2).65-79)
- [8] Grebenyuk A.A., Maksimova A.S., Lemer L.G. Study of social tension based on electronic social networks big data. *Digital sociology*. 2021;4(4):4-12. (In Russ., abstract in Eng.) <https://doi.org/10.26425/2658-347X-2021-4-4-4-12>
- [9] Kosnikov S.N., Novikova D.N., Sukhinenko D.D. Modeling and forecasting indicators of socio-economic development. *Bulletin of the Academy of Knowledge*. 2022;(41):151-154. (In Russ., abstract in Eng.) EDN: JEPSPG
- [10] Sedova N.O., Pertseva I.A., Yur'eva O.D. Analysis of the Effectiveness of Vocational Guidance Activities of the University Using Machine Learning. *Modern Information Technologies and IT-Education*. 2021;17(4):1032-1046. (In Russ., abstract in Eng.) <https://doi.org/10.25559/SITITO.17.202104.1032-1046>



- [11] Schwab-McCoy A., Baker C.M., Gasper R.E. Data Science in 2020: Computing, Curricula, and Challenges for the Next 10 Years. *Journal of Statistics and Data Science Education*. 2021;29(sup1):S40-S50. <https://doi.org/10.1080/10691898.2020.1851159>
- [12] Baumer B. A Data Science Course for Undergraduates: Thinking With Data. *The American Statistician*. 2015;69:334-342. <https://doi.org/10.1080/00031305.2015.1081105>
- [13] Brunner R.J., Kim E.J. Teaching Data Science. *Procedia Computer Science*. 2016;80:1947-1956. <https://doi.org/10.1016/j.procs.2016.05.513>
- [14] Albert J., Rizzo M. An Undergraduate Data Science Program. In: JSM 2016 Proceedings. Section on Statistical Education. Chicago, Illinois: American Statistical Association; 2016. p. 1135-1147. Available at: <https://www2.amstat.org/meetings/proceedings/2016/data/assets/pdf/389584.pdf> (accessed 15.11.2023).
- [15] Majumder M., Cheng X. Focusing on the Needs: Experiences of Developing a Data Science Program. *Journal of Computational and Graphical Statistics*. 2017;26(4):779-780. <https://doi.org/10.1080/10618600.2017.1385475>
- [16] Hicks S.C., Irizarry R.A. A Guide to Teaching Data Science. *American Statistician*. 2018;72(4):382-391. <https://doi.org/10.1080/00031305.2017.1356747>
- [17] Tang R., Sae-Lim W. Data Science Programs in the U.S. Higher Education: An Exploratory Content Analysis of Program Description, Curriculum Structure, and Course Focus. *Education for Information*. 2016;32:269-290. <https://doi.org/10.3233/EFI-160977>
- [18] Song I.-Y., Zhu Y. Big Data and Data Science: What Should We Teach? *Expert Systems*. 2016;33(4):364-373. <https://doi.org/10.1111/exsy.12130>
- [19] West J. Teaching Data Science: An Objective Approach to Curriculum Validation. *Computer Science Education*. 2018;28(2):136-157. <https://doi.org/10.1080/08993408.2018.1486120>
- [20] Lapenok M.B., Shestakova L.G. Formation of bachelor's skills in the field of machine learning and intelligent data analysis. *Scholarly Notes of Transbaikal State University*. 2024;19(3):17-26. (In Russ., abstract in Eng.) <https://doi.org/10.21209/2658-7114-2024-19-3-17-26>
- [21] Medvedeva E.I., Kroshilin S.V., Avacheva T.G. The need to develop information competencies in the preparation of students in medical universities. *Meditinskoe obrazovanie i professional'noe razvitiye = Medical education and professional development*. 2023;14(1):66-78. (In Russ., abstract in Eng.) <https://doi.org/10.33029/2220-8453-2023-14-1-66-78>
- [22] Hazzan O., Mike K. The Birth of a New Discipline: Data Science Education. In: Guide to Teaching Data Science. Cham: Springer; 2023. p. 59-72. https://doi.org/10.1007/978-3-031-24758-3_4
- [23] Aljawarneh S., Lara J.A. Data science for analyzing and improving educational processes. *Journal of Computing in Higher Education*. 2021;33(3):545-550. <https://doi.org/10.1007/s12528-021-09299-7>
- [24] Dorofeeva V.I., Stroev S.P., Dorofeeva E.Y. On the Issue of Teaching Data Analysis and Machine Learning to Students of Medical Specialties. In: 2023 3rd International Conference on Technology Enhanced Learning in Higher Education (TELE). Lipetsk, Russian Federation: IEEE Press; 2023. p. 101-103. <https://doi.org/10.1109/TELE58910.2023.10184352>
- [25] Dorofeeva V.I. On the question of teaching disciplines related to data analysis to students of the training direction "Applied Mathematics and Computer Science". *Uchenye zapiski Oryol State University = Scientific Notes of Orel State University*. 2022;(4):244-246. (In Russ., abstract in Eng.) EDN: TBIJRK

Submitted 15.11.2023; approved after reviewing 16.01.2024; accepted for publication 27.01.2024.

About the authors:

Victoria I. Dorofeeva, Head of the Chair of Informatics, Faculty of Physics and Mathematics, Orel State University named after I.S. Turgenev (95 Komsomolskaya St., Orel 302026, Russian Federation), Cand. Sci. (Phys.-Math.), Associate Professor, **ORCID: <https://orcid.org/0000-0001-6116-2511>**, vdorofey@mail.ru

Sergey P. Stroev, Head of the Chair of Algebra and Mathematical Methods in Economics, Faculty of Physics and Mathematics, Orel State University named after I.S. Turgenev (95 Komsomolskaya St., Orel 302026, Russian Federation), Cand. Sci. (Econ.), **ORCID: <https://orcid.org/0000-0002-2271-5264>**, stroewsp@mail.ru

Dmitrii Yu. Dorofeev, student of the Centre of Compositional Materials, National University of Science and Technology "MISIS" (4 Leninsky pr., building 2, Moscow 119049, Russian Federation), **ORCID: <https://orcid.org/0009-0002-6771-211X>**

All authors have read and approved the final manuscript.

