

СТАТЬИ

УДК 378:004.853

DOI 10.17513/snt.40464

**ОСНОВНЫЕ ЗНАНИЯ, УМЕНИЯ И НАВЫКИ, НЕОБХОДИМЫЕ
ДЛЯ СПЕЦИАЛИСТА В ОБЛАСТИ МАШИННОГО ОБУЧЕНИЯ**¹Беспалько А.А., ²Строганова Н.В., ¹Тутынина О.И.¹ФГБОУ ВО «Национальный исследовательский Нижегородский государственный университет имени Н.И. Лобачевского», Нижний Новгород;²Волго-Вятский филиал ФГБОУ ВО «Московский институт связи и информатики», Нижний Новгород, e-mail: snv_nki@mail.ru

Современная подготовка специалистов по машинному обучению требует сочетания глубоких теоретических знаний, определенных практических навыков и развития целого набора общих компетенций. Целью данного исследования является выделение в рамках существующих компетенций тех знаний, умений и навыков, которыми должен обладать специалист в области машинного обучения. Конкретный набор этих знаний, умений и навыков определяется выбранной специализацией в области машинного обучения (например, программист, исследователь машинного обучения, аналитик данных, архитектор машинного обучения и т.д.) и основан на обучении решению основных типов задач в области искусственного интеллекта. В статье представлен базовый структурированный план обучения специалиста по искусственному интеллекту, основанный на актуальных требованиях рынка и современных образовательных стандартах. В нем выделены все основные виды и задачи машинного обучения и сформулированы конкретные знания, умения и навыки, которыми должны овладеть обучающиеся, получающие образование в этой области. Такой структурированный план охватывает ключевые теоретические и практические аспекты подготовки специалиста в сфере машинного обучения с учетом выбранного направления, а также обеспечивает развитие универсальных компетенций, востребованных в современной индустрии искусственного интеллекта. Этот план может быть взят за основу обучения на любом уровне образования: от бакалавриата до курсов повышения квалификации или переподготовки, поскольку он гибко адаптируется под индивидуальные потребности и может служить базой для дальнейшей специализации в конкретных областях машинного обучения.

Ключевые слова: машинное обучение, искусственный интеллект, специалист, Python, обработка данных, алгоритм

**THE BASIC KNOWLEDGE, SKILLS AND ABILITIES REQUIRED
FOR A SPECIALIST IN THE FIELD OF MACHINE LEARNING**¹Bespalko A.A., ²Stroganova N.V., ²Tutynina O.I.¹Lobachevsky National Research Nizhny Novgorod State University, Nizhny Novgorod;²Volga-Vyatka Branch of the Moscow Institute of Communications and Informatics, Nizhny Novgorod, e-mail: snv_nki@mail.ru

Modern training of machine learning specialists requires a combination of deep theoretical knowledge, certain practical skills and the development of a whole set of general competencies. The purpose of this study is to identify, within the framework of existing competencies, the knowledge, skills and abilities that a specialist in the field of machine learning should possess. The specific set of these knowledge, skills and abilities is determined by the chosen specialization in the field of machine learning (for example, programmer, machine learning researcher, data analyst, ML architect, etc.) and is based on training in solving basic types of tasks in the field of artificial intelligence. The article presents a basic structured training plan for an AI specialist based on current market requirements and modern educational standards. It highlights all the main types and tasks of machine learning and formulates specific knowledge, skills and abilities that students who receive education in this field should master. Such a structured plan covers the key theoretical and practical aspects of training a specialist in ML, taking into account the chosen direction, and also ensures the development of universal competencies in demand in the modern AI industry. This plan can be used as the basis for training at any level of education: from bachelor's degree to advanced training or retraining courses, as it flexibly adapts to individual needs and can serve as a basis for further specialization in specific areas of machine learning.

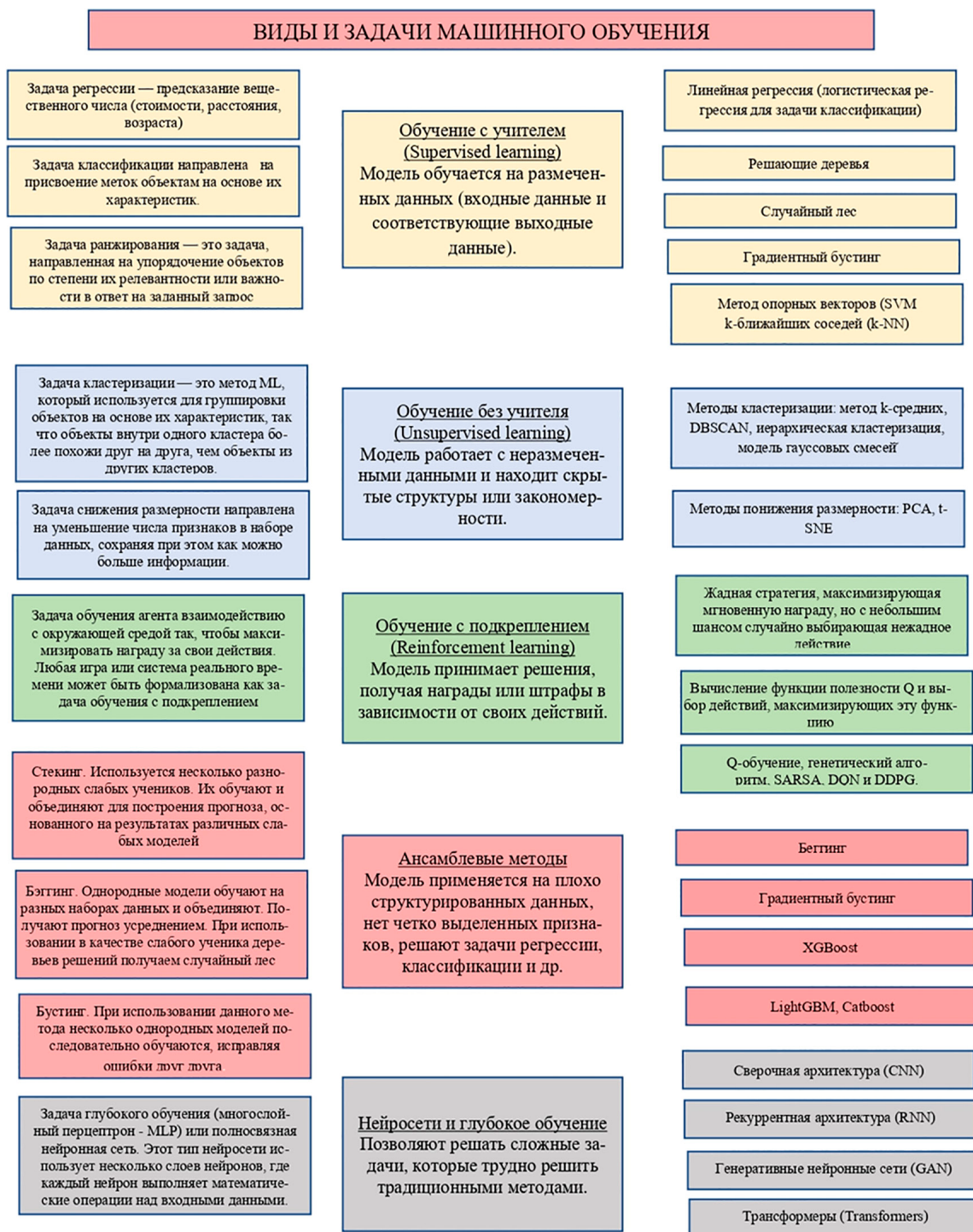
Keywords: machine learning, artificial intelligence, specialist, Python, data processing, algorithm

Введение

Машинное обучение (ML) можно назвать одним из этапов реализации идеи искусственного интеллекта (AI). Технологии машинного обучения позволяют компьютерным системам учиться на основе данных, предоставленных для обучения без непосредственного участия человека, то есть на основе некоторых алгоритмов. За последние десятилетия данная технология вошла

во многие сферы человеческой жизни – это рекомендательные системы, прогнозирование, машинное зрение, генерация текста, изображения, музыки и др. [1, 2].

Таким образом для подготовки специалистов в сфере машинного обучения необходима актуализация знаний, умений и навыков, дающих непрерывное и эффективное движение по образовательной траектории [3, 4].



Виды и задачи машинного обучения
Источник: составлено авторами

Цель исследования – выделение в рамках существующих компетенций тех знаний, умений и навыков, которыми должен обладать специалист в области ML.

Применение этого набора целесообразно при формировании учебных планов, на-

правленных на подготовку специалистов в области ML. На практике перед авторами стояла задача разработать план обучения в рамках дополнительного профессионального образования по направлению «Анализ данных и машинное обучение». Особен-

ность заключается в том, что на курс попадают обучаемые самого разного уровня подготовки, поэтому учебный план должен включать максимально широкий набор умений, знаний и навыков (УЗН). Вторая особенность курса – его назначение. Он подразумевает некий старт для обучающихся в будущей профессиональной деятельности в сфере ML, с последующим выбором направления для развития. Поэтому важная задача – выработать общий базовый набор УЗН, который даст основу студентам в дальнейшей профессиональной траектории.

Материалы и методы исследования

Материалом исследования служит набор компетенций, выставляемых на рынке труда к специалистам различной направленности в областях применения искусственного интеллекта. Методика исследования включает в себя систематизацию и анализ предъявляемых требований к соискателям на позиции различного рода специалистов по машинному обучению и обработке данных.

Суть машинного обучения

Машинное обучение берет свое начало в середине XX в. из работ таких ученых, как Алан Тьюринг, Артур Самуэль, Фрэнк Розенблатт, Марвин Минский, Сеймур Пейперт и др. С развитием вычислительных мощностей и возможностью копить большие объемы информации в 1990-х гг. наблюдался прогресс в области ML. В начале 2000-х гг. активно развиваются биоинформатика, обработка естественного языка и компьютерное зрение [5, 6]. В 2010-е гг. появляются технологии распознавания изображений и речи. В последние годы машинное обучение продолжает развиваться быстрыми темпами, сейчас ML способен генерировать тексты, изображения и видео [5, 6].

Интересным направлением будущего развития технологий машинного обучения является разработка методов обучения с малым количеством данных [8, 9].

Общие этапы процесса машинного обучения включают следующие этапы: сбор данных, предобработка данных, выбор модели, обучение модели, оценка производительности и внедрение. Все задачи машинного обучения принято делить на классы, которые представлены на рисунке.

Обучение будущих специалистов в области ML

Бурное развитие технологий машинного обучения за последние десятилетия и практическое их применение в повседневной деятельности человека привело к возрастанию интереса к этой области информационных технологий со стороны молодых специалистов. Для высшего образования остро встали вопросы о компетенциях, знаниях, умениях и навыках будущих специалистов в сфере ML и Data Scientist. Вузы встали перед проблемой построения траектории роста. Альянс в сфере искусственного интеллекта, который объединяет ведущие технологические компании для совместного развития их компетенций и ускоренного внедрения искусственного интеллекта в образовании, научных исследованиях и в практической деятельности бизнеса, разработал базовую модель компетенций специалистов в сфере AI [10, 11]. Она служит основой для построения образовательных программ соответствующего направления. Однако вузам необходимо самостоятельно определять перечень предметов и спецкурсов, а также их содержание.

Авторами был выделен набор специальностей в сфере ML и составлены требования к их УЗН, представленные в табл. 1–6.

Таблица 1

Профессиональные знания, умения и навыки программиста по машинному обучению

Знания	Умения	Навыки
Математика и статистика. Основы линейной алгебры, статистики и теории вероятностей. Языки программирования: Python, R, основы SQL. Библиотеки и фреймворки: Keras, PyTorch, Scikit-learn, TensorFlow. Основные алгоритмы: регрессия, классификация и кластеризация	Работать с базами данных, извлекать данные, очищать и преобразовывать данные, представлять результаты анализа в понятном виде, проводить оценку производительности моделей с использованием различных метрик (точность, полнота, F1-мера)	Программирование и документирование кода, работа с данными (Pandas, NumPy), применение базовых алгоритмов ML

Источник: составлено авторами.

Таблица 2

Профессиональные знания, умения и навыки исследователя машинного обучения

Знания	Умения	Навыки
Линейная алгебра, математический анализ, статистика и теория вероятности. Python, R, Java, SQL. Библиотеки и фреймворки: Keras, PyTorch, Scikit-learn, TensorFlow. Основные алгоритмы: регрессия, классификация и кластеризация. Глубокое обучение	Способность проводить разведочный анализ данных (EDA), выявлять закономерности и аномалии в наборах данных. Оценивать данные и результаты экспериментов, формулировать гипотезы и проверять их	Выбирать подходящий алгоритм в зависимости от задачи. Писать статьи и отчеты по результатам исследований для публикации в научных журналах, объяснение сложных концепций как техническим, так и нетехническим аудиториям, непрерывное обучение

Источник: составлено авторами.

Таблица 3

Профессиональные знания, умения и навыки инженера по данным

Знания	Умения	Навыки
Теория баз данных, реляционные и нереляционные системы управления базами данных. Принципы работы с большими объемами данных и использование инструментов для их обработки, таких как Hadoop, Spark и Kafka, основные облачные платформы (AWS, Google Cloud, Azure), языки программирования: Python, Java или Scala. Знание инструментов бизнес-аналитики (например, Tableau или Power BI). Понимание основных принципов машинного обучения для эффективного взаимодействия с data scientists	Строить ETL-процессы (процессы извлечения, преобразования и загрузки данных), уметь разрабатывать архитектуру для хранения и обработки данных в облаке, анализировать системы на предмет их производительности и вносить изменения для улучшения работы	Работать с большими данными (Hadoop, Spark), эффективно общаться и объяснять сложные технические концепции, быть внимательным к деталям для предотвращения ошибок в процессе обработки информации.

Источник: составлено авторами.

Таблица 4

Профессиональные знания, умения и навыки аналитика данных

Знания	Умения	Навыки
Математическая статистика, теория вероятностей и линейная алгебра реляционных и нереляционных баз данных, SQL для выполнения запросов, основы программирования: Python или R	Представлять данные в понятной форме через графики и диаграммы, оценивать достоверность данных и выявлять возможные источники ошибок, использовать разнообразные методы анализа данных	Эффективно сотрудничать со специалистами различных направлений (маркетологами, разработчиками и т.д.). Объяснять сложные концепции простыми словами, понятными даже нетехническим специалистам

Источник: составлено авторами.

К общим компетенциям (так называемым soft skills) относятся работа в команде, эффективное взаимодействие со специалистами других направлений, умение четко формулировать вопросы, а также навык работы с англоязычными источниками [12, 13].

Чтобы подготовить специалиста в области машинного обучения, необходимо создать структурированный план обучения, который охватывает теоретические и практические аспекты [14, 15].

Результаты исследования и их обсуждение

В качестве основного результата исследования авторами предложен рассмотренный выше план по базовой подготовке специалистов в области ML, который может служить основой будущей специализации в одном из более узких направлений. В плане для каждой темы указан минимальный набор аудиторных часов, необходимый для ее усвоения.

Таблица 5

Профессиональные знания, умения и навыки архитектора машинного обучения

Знания	Умения	Навыки
Линейная алгебра, теория вероятностей и математическая статистика, методы оптимизации. Уверенное владение Python и R, Java или C++. Основные алгоритмы: регрессия, классификация и кластеризация. Знание методов глубокого обучения и нейронных сетей, включая сверточные и рекуррентные сети. Знание принципов DevOps	Работа с библиотеками, такими как TensorFlow, Keras, PyTorch и scikit-learn, необходим для реализации алгоритмов машинного обучения. Проводить исследовательский анализ данных (EDA) для выявления закономерностей и подготовки данных к обучению моделей. Проектировать архитектуру системы машинного обучения, включая выбор подходящих инструментов и технологий для решения конкретных задач принимать решения о технологиях и инфраструктуре. Анализировать бизнес-требования и разрабатывать стратегии внедрения решений на основе машинного обучения анализировать большие объемы информации и делать выводы на основе полученных данных. Находить инновационные решения для сложных задач	Работать с большими данными (Big Data) и распределенными системами обработки данных (например, Hadoop, Spark), визуализации данных с помощью инструментов Matplotlib или Seaborn. Проектировать сложные системы и архитектуры. Ясно и доступно объяснять сложные технические концепции как техническим, так и нетехническим специалистам

Источник: составлено авторами.

Таблица 6

Профессиональные знания, умения и навыки специалиста по развертыванию и эксплуатации ML

Знания	Умения	Навыки
Глубокое понимание линейной алгебры, математического анализа, теории вероятностей и математической статистики. Уверенное владение Python и сопутствующими библиотеками, возможно знание R и Java. Знание современных методов и алгоритмов ML, глубокого обучения и нейросетей	Уметь интегрировать обученные модели в производственные системы и обеспечивать их стабильную работу. Отслеживать их производительность и вносить коррективы для повышения эффективности. Оптимизировать гиперпараметры и обновлять модели на основе новых данных	Четко формулировать задачи и представлять результаты своей работы. Постоянное обучение и адаптация к новым инструментам и методам

Источник: составлено авторами.

Таблица 7

Тематический план для курсов дополнительного профессионального образования «Анализ данных и машинное обучение»

Тема	Часы
<i>Основы программирования на Python</i>	
Знакомство с Python: история и области применения языка, установка и настройка среды разработки, первая программа	2
Основы синтаксиса: переменные и типы данных, арифметические и логические операторы, ввод и вывод данных	2
Условные конструкции: оператор if, elif, else, вложенные условия	4
Циклы: циклы for и while, операторы break и continue, вложенные циклы	6
Структуры данных: операции со списками, индексация, кортежи и множества	8
Функции и модули: определение и вызов функций, аргументы и возвращаемые значения, значения по умолчанию, именованные аргументы, локальные и глобальные переменные, рекурсия	6
Модули и библиотеки: импорт модулей, стандартные и сторонние библиотеки, работа с pip	2
Введение в объектно-ориентированное программирование: классы и объекты, атрибуты и методы	4
<i>Элементы высшей математики и основы теории вероятности</i>	
Математический анализ – основные разделы. Функции и последовательности, их пределы: понятие и основные способы нахождения	4
Производные и их свойства. Теоремы о производных	2

Окончание табл. 7

Тема	Часы
Интегральное исчисление: неопределенный интеграл как действие, обратное нахождению производной; его свойства	2
Основные способы вычисления неопределенных интегралов. Вычисление интегралов от тригонометрических и иррациональных функций; интегрирование рациональных дробей.	4
Определенный интеграл – свойства и вычисление. Приложения: вычисление площадей, длин дуг, объемов тел вращения и т.д.	2
Линейная алгебра. Алгебра матриц. Решение систем линейных уравнений: метод Гаусса, метод Крамера, матричный метод	4
Ряды, их свойства и приложения	4
Теория вероятностей: основные формулы и теоремы теории вероятностей. Комбинаторика.	4
Случайные величины: свойства и характеристики. Распределения СВ	2
Основные распределения дискретных СВ: Бернулли, геометрическое, Пуассона; их числовые характеристики	2
Непрерывные СВ: их свойства. Равномерное и нормальное распределения, их основные характеристики.	2
Ковариация и корреляция СВ. Метод наименьших квадратов. Метод наибольшего правдоподобия	2
Статистические гипотезы и их проверка	2
Интервальные и точечные оценки параметров	2
<i>Введение в машинное обучение</i>	
Машинное обучение, его виды (обучение с учителем, без учителя и обучение с подкреплением)	4
Базовые алгоритмы: линейная регрессия, логистическая регрессия, деревья решений	4
<i>Работа с данными</i>	
Сбор и предобработка данных: изучение методов очистки данных и работы с пропущенными значениями	4
Визуализация данных: использование библиотек Matplotlib и Seaborn для визуализации данных	4
<i>Алгоритмы машинного обучения</i>	
Обучение с учителем: глубокое изучение алгоритмов классификации (например, SVM, KNN) и регрессии.	6
Обучение без учителя: Кластеризация (K-means, иерархическая кластеризация) и методы понижения размерности (PCA).	6
<i>Углубленное изучение нейронных сетей</i>	
Основы нейронных сетей: изучение архитектуры нейронных сетей и их применения	2
Типы и архитектуры нейронных сетей: однослойные и многослойные ИНС, персептрон, многослойные сети с прямыми связями, теорема Арнольда – Колмогорова, результаты Хехт-Нильсена, процедура обратного распространения ошибки (Backpropagation)	6
Методы обучения нейронных сетей: обучение с учителем и без учителя, метод обучения Хэбба, алгоритм обучения Кохонена, алгоритмы обучения Хопфилда и Хэмминга	6
Глубокое обучение: работа с библиотеками TensorFlow и PyTorch для построения глубоких нейронных сетей	4
Регуляризация и борьба с переобучением	2
Применение современных фреймворков (PyTorch, TensorFlow, Keras)	4
Примеры архитектур: CNN, RNN, GAN	4
Использование предобученных моделей	4
<i>Специализированные темы</i>	
Обработка естественного языка (NLP): изучение методов обработки текстовых данных	4
Компьютерное зрение: основы работы с изображениями и видео	4
<i>Практические проекты</i>	
Предсказание цен на жилье: построение модели регрессии для оценки стоимости недвижимости на основе характеристик объектов	4
Классификация изображений (например, рукописных цифр MNIST или болезней растений по фото). Использование сверточных нейронных сетей для распознавания объектов на изображениях	4
Система рекомендаций. Разработка рекомендательной системы для фильмов, музыки или товаров на основе пользовательских предпочтений	4
Итоговый проект	8

Курс рассчитан на 10 учебных месяцев, 160 ч и планируется к внедрению в 2025–2026 учебном году. Курс сопровождается полным методическим обеспечением.

В настоящее время авторами был проведен опрос среди студентов второго и третьего курсов, обучающихся в Волго-Вятском филиале Московского технического университета связи и информатики по специальностям «Инфокоммуникационные технологии и системы связи». Опрос 42 чел. показал, что более половины (60%) из них заинтересованы в обучении на подобных курсах в качестве дополнительного образования с перспективой дальнейшего углубленного изучения по выбранной специализации.

Заключение

В качестве основного результата исследования авторами предложен план базовой подготовки специалистов в области машинного обучения, который может служить фундаментом для дальнейшей специализации в одном из более узких направлений в этой области. Проведенный опрос показывает высокую востребованность предложенной образовательной траектории среди будущих специалистов в области информационных технологий. В дальнейшем авторы планируют внедрить данный курс в качестве факультатива или спецкурса для студентов технических специальностей. Такой подход позволит гибко интегрировать программу в существующую образовательную структуру и адаптировать ее под потребности различных категорий обучающихся. По результатам внедрения курса будет проведен мониторинг успеваемости и анализ распределения студентов по направлениям специализации. Это позволит оценить эффективность предложенной методики и скорректировать программу в соответствии с реальными образовательными запросами и интересами студентов.

Список литературы

1. Шмагин В.С., Маслова М.А. Обзор и анализ развития искусственного интеллекта // Научный результат. Информационные технологии. 2020. Т. 5. № 4. С. 3–8. DOI: 10.18413/2518-1092-2020-5-4-0-1.
2. Осипенко Л.Е., Козицына Ю.В., Коротков А.В. Исследовательское и машинное обучение: от сопоставления к конвергенции // Психолого-педагогические исследования. 2022. Т. 14. № 4. С. 127–146. DOI: 10.17759/psyedu.2022140408.
3. Сорокина Г.П., Долгих Е.А., Першина Т.А. Анализ профессиональных навыков специалистов в сфере искусственного интеллекта // Вестник университета. 2022. № 4. С. 81–89. DOI: 10.26425/1816-4277-2022-4-81-89.
4. Коротеев М.В. Обзор некоторых современных тенденций в технологии машинного обучения // E-Management. 2018. № 1. С. 26–35. DOI: 10.26425/2658-3445-2018-1-26-35.
5. Гольдберг Й. Нейросетевые методы в обработке естественного языка. М.: ДМК Пресс, 2019. 282 с.
6. Либерман А.И. Решение задачи классификации объектов на изображении методами компьютерного зрения в различных сферах человеческой деятельности // Вестник ВГУ. Серия: Системный анализ и информационные технологии. 2024. № 3. С. 74–91. DOI: 10.17308/sait/1995-5499/2024/3/74-91.
7. Бабкина А.А., Андришечкина Н.А. Применение искусственного интеллекта в математике // Международный журнал гуманитарных и естественных наук. 2023. Т. 11–2 (86). С. 178–181. DOI: 10.24412/2500-1000-2023-11-2-178-181.
8. Михайлова С.С., Гринева Н.С. Разработка модели бинарной классификации на малых данных с использованием методов машинного обучения // Проблемы экономики и юридической практики. 2024. Т. 20. № 1. С. 129–140. DOI: 10.33693/2541-8025-2024-20-1-129-140.
9. Лютикова Л.А. Применение метода машинного обучения для анализа неполных данных // Известия Кабардино-Балкарского научного центра РАН. 2024. Т. 26. № 6. С. 139–145. DOI: 10.35330/1991-6639-2024-26-6-139-145.
10. Ендовицкий Д.А., Гайдар К.М. Университетская наука и образование в контексте искусственного интеллекта // Высшее образование в России. 2021. Т. 30. № 6. С. 121–131. DOI: 10.31992/0869-3617-2021-30-6-121-131.
11. Абашин В.Г., Хасаншин И.Я., Семенов Д.Н., Круглов В.И., Никитин П.В., Курилева Н.Л. Реализация междисциплинарного подхода в обучении студентов анализу данных и машинному обучению на основе метода проекта // Современные наукоемкие технологии. 2021. № 10. С. 113–117. URL: <https://top-technologies.ru/ru/article/view?id=38863> (дата обращения: 24.04.2025). DOI: 10.17513/snt.38863.
12. Рябко Т.В., Гуртов В.А., Степутьев И.С. Анализ показателей подготовки кадров для сферы искусственного интеллекта по результатам мониторинга вузов // Высшее образование в России. 2022. Т. 31. № 7. С. 9–24. DOI: 10.31992/0869-3617-2022-31-7-9-24.
13. Harris H., Murphy S., Vaisman M. Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work. O'Reilly Media, 2013. ISBN 978-1-4493-6840-1.
14. 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. DOI: 10.25559/SITITO.020.202401.251-259.
15. Щеголева Л.В., Аверьянов А.О., Гуртов В.А. Источники компетенций для образовательных программ в сфере искусственного интеллекта // Непрерывное образование: XXI век. 2023. Вып. 2 (42). DOI: 10.15393/j5.art.2023.8466.