

2025, № 09 (сентябрь)

Раздел 5.8. Педагогика

ART 251173

DOI 10.24412/2304-120X-2025-11173

УДК 378.147:004.9

Реализация компьютерного эксперимента в лабораторных работах по машинному обучению для студентов медицинских специальностей вузов

Implementation of a computer experiment in laboratory work on machine learning for university medical students

Авторы статьи

Фирсова Светлана Анатольевна,
 кандидат физико-математических наук, доцент ка-
 федры информационных технологий и программиро-
 вания ФГБОУ ВО «Национальный исследовательский
 Мордовский государственный университет
 им. Н. П. Огарёва», г. Саранск, Российская Федерация
 karpushkinasa@yandex.ru
 ORCID: 0000-0003-1831-684X

Рябухина Елена Александровна,
 кандидат педагогических наук, доцент кафедры ин-
 формационных технологий и программирования
 ФГБОУ ВО «Национальный исследовательский Мор-
 довский государственный университет им. Н. П. Ога-
 рёва», г. Саранск, Российская Федерация
 colomentsevaa@yandex.ru
 ORCID: 0009-0007-6533-3497

Authors of the article

Svetlana A. Firsova,
 Candidate of Physical and Mathematical Sciences, Asso-
 ciate Professor, Information Technology and Program-
 ming Chair, National Research Mordovia State University
 named after N.P. Ogarev, Saransk, Russian Federation
 karpushkinasa@yandex.ru
 ORCID: 0000-0003-1831-684X

Elena A.I. Ryabukhina,
 Candidate of Pedagogical Sciences, Associate Professor,
 Information Technology and Programming Chair, Na-
 tional Research Mordovia State University named after
 N.P. Ogarev, Saransk, Russian Federation
 colomentsevaa@yandex.ru
 ORCID: 0009-0007-6533-3497

Конфликт интересов

Конфликт интересов не указан

Conflict of interest statement

Conflict of interest is not declared

Для цитирования

Фирсова С. А., Рябухина Е. А. Реализация компью-
 терного эксперимента в лабораторных работах по ма-
 шинному обучению для студентов медицинских спе-
 циальностей вузов // Научно-методический электрон-
 ный журнал «Концепт». – 2025. – № 09. – С. 40–62. –
 URL: <https://e-koncept.ru/2025/251173.htm> – DOI:
 10.24412/2304-120X-2025-11173

For citation

S. A. Firsova, E. A.I. Ryabukhina, Implementation of a com-
 puter experiment in laboratory work on machine learn-
 ing for university medical students // Scientific-method-
 ological electronic journal "Koncept". – 2025. – No. 09. –
 P. 40–62. – URL: <https://e-koncept.ru/2025/251173.htm>
 – DOI: 10.24412/2304-120X-2025-11173

Поступила в редакцию <i>Received</i>	16.05.25	Получена положительная рецензия <i>Received a positive review</i>	10.08.25
Принята к публикации <i>Accepted for publication</i>	10.08.25	Опубликована <i>Published</i>	30.09.25



Аннотация

В связи с введением в учебные планы медицинских вузов дисциплины «Введение в современные информационные и интеллектуальные технологии», включающей раздел по изучению основных принципов и алгоритмов машинного обучения и искусственного интеллекта (далее – ИИ), возникает необходимость разработки соответствующего методического обеспечения. Актуальность темы исследования обусловлена тем, что студенты-медики в подавляющем большинстве не обладают необходимыми знаниями в области математики и программирования, являющимися теоретической основой современных интеллектуальных технологий, что может привести в их дальнейшей профессиональной деятельности к неверной интерпретации полученных с помощью ИИ результатов; поэтому при разработке методических материалов следует учитывать специфику данной категории обучающихся. Целью статьи является описание предлагаемого авторами подхода по реализации компьютерных экспериментов в лабораторном практикуме вышеуказанной дисциплины, позволяющих студентам более глубоко понять суть изучаемых методов машинного обучения и условий их применения. В качестве методов исследования выступили: методы классификации, применяемые в машинном обучении, методы планирования эксперимента, а также двухфакторный дисперсионный анализ, используемый в качестве доказательной базы эффективности предлагаемого подхода. Базу для проведения опытно-экспериментальных работ составил учебный контингент Медицинского института Мордовского университета. Результатом является разработанный авторами подход к выполнению лабораторных работ студентами медицинских специальностей вузов. Данный подход включает: изучение основных математических формул, лежащих в основе того или иного метода машинного обучения, выполнение несложных вычислительных примеров на небольшом объеме данных, изучение основных функций языка Python для реализации конкретной практической задачи, проведение компьютерных экспериментов для более углубленного изучения выбранной математической модели и повышения ее точности. Эффективность предложенного подхода была подтверждена результатами проведенного авторами педагогического эксперимента. Теоретическая значимость работы состоит в обобщении и систематизации материалов по данной теме, полученные результаты дополняют имеющиеся научные и педагогические разработки, углубляя понимание процессов, связанных с данной темой, и могут быть использованы в дальнейших исследованиях. Практическая значимость заключается в том, что представленный подход может быть использован в качестве основы при разработке методических материалов по разделу «Технологии машинного обучения и искусственного интеллекта в здравоохранении».

Ключевые слова

медицинский вуз, компьютерный эксперимент, машинное обучение, категориальный байесовский классификатор, язык программирования Python, диагностика заболеваний

Благодарности

Авторы выражают глубокую благодарность коллективу кафедры информационных технологий и программирования ФГБОУ ВО «Мордовский государственный университет имени Н. П. Огарёва» за помощь в организации и проведении тестирования студентов Медицинского института при реализации педагогического эксперимента.

Abstract

Due to the introduction of the discipline "Introduction to Modern Information and Intellectual Technologies" into the curricula of medical universities, which includes a section devoted to the basic principles and algorithms of machine learning and artificial intelligence (AI), there is a need to develop appropriate methodological support. The relevance of the research topic is due to the fact that the vast majority of medical students do not have the necessary knowledge in mathematics and programming, which are the theoretical foundations of modern intelligent technologies. This can lead to incorrect interpretation of AI-generated results in their future professional work. Therefore, it is important to consider the specific needs of this category of students when developing teaching materials. The aim of the article is to describe the authors' proposed approach to implementing computer experiments in the laboratory practice of the above-mentioned discipline, which allows students to gain a deeper understanding of the essence of the studied machine learning methods and the conditions for their application. The research methods used include classification methods used in machine learning, experimental design methods, and two-factor analysis of variance, which serves as a basis for evaluating the effectiveness of the proposed approach. The study was conducted with a participation of students from the Medical Institute of Mordovian University. The result is an approach developed by the authors for conducting laboratory work by medical students. This approach includes: studying the basic mathematical formulas underlying a particular machine learning method, performing simple computational tasks on a small amount of data, studying the basic functions of the Python language to implement a specific practical task, and conducting computer experiments to further explore the chosen mathematical model and improve its accuracy. The effectiveness of this approach has been confirmed by the results of a pedagogical experiment conducted by the authors. The theoretical significance of this work lies in the generalization and systematization of materials on this topic, and the results obtained complement existing scientific and pedagogical developments, deepening our understanding of the processes related to this topic, and they can be used in further research. The practical significance lies in the fact that the presented approach can be used as a basis for developing methodological materials on the topic "Machine Learning and Artificial Intelligence Technologies in Healthcare."

Key words

medical university, computer experiment, machine learning, categorical Bayesian classifier, Python programming language, diagnosis of diseases.

Acknowledgements

The authors express their deep gratitude to the staff of the Department of Information Technologies and Programming at the Ogarev Mordovian State University for their assistance in organizing and conducting testing of medical students during the implementation of the pedagogical experiment.

Введение / Introduction

В условиях широкого внедрения технологий машинного обучения и искусственного интеллекта в различные сферы производства и социальной сферы возникает необходимость актуализации образовательных программ высшего образования по всем специальностям и направлениям подготовки, связанным с разработкой и применением указанных технологий с целью повышения качества подготовки специалистов и в соответствии с современными требованиями рынка труда.

Данное требование отражено в перечне поручений Президента РФ по итогам конференции «Путешествие в мир искусственного интеллекта», состоявшейся 4 декабря 2020 года [1], согласно которому в срок до 1 сентября 2021 года необходимо дополнить содержание учебных дисциплин соответствующими разделами.

С целью выполнения данного требования в Мордовском государственном университете им. Н. П. Огарёва начиная с 2022/2023 учебного года в учебный процесс медицинских специальностей была введена дисциплина «Введение в современные информационные и интеллектуальные технологии», заменившая изучаемую ранее дисциплину «Математика, медицинская информатика». Основное отличие этих дисциплин состоит во включении в рабочую программу новой дисциплины раздела, посвященного применению технологий ИИ и машинного обучения в здравоохранении.

Однако, несмотря на то что профессиональные компетенции выпускников медицинских вузов должны в том числе обеспечивать возможность использования искусственного интеллекта в практической деятельности, освоение данного раздела вызывает значительные трудности, связанные, во-первых, с отсутствием в учебном плане логически предшествующих дисциплин, закладывающих основу для применения ИИ в медицине, и, во-вторых, с недостаточной разработкой методических материалов, способствующих успешному освоению теоретической и программной составляющих данной технологии.

В представленной статье предлагается возможное решение данной проблемы с примером одной из работ, входящих в лабораторный практикум вышеуказанной дисциплины.

Обзор литературы / Literature review

Проблема методического обеспечения раздела «Технологии машинного обучения и искусственного интеллекта в здравоохранении» рассматривалась в разных аспектах.

В результате педагогического эксперимента, проведенного Э. Т. Бихатовой и О. В. Иванчук, была выявлена проблема недостатка знаний, умений и навыков студентов медицинских вузов для применения ими технологий цифрового здравоохранения в дальнейшей профессиональной деятельности [2]. К тому же выводу в своих исследованиях приходят Д. А. Забелин и Е. В. Плащевая, подчеркивая при этом необходимость обучения студентов-медиков критически оценивать результаты взаимодействия с ИИ, а также соблюдения этических и правовых норм [3].

В исследовании Лукаса Вайденера и Майкла Фишера, основанном на опросе 487 студентов-медиков в Германии, Австрии и Швейцарии, было выявлено выраженное несоответствие между использованием студентами ИИ-чатов, например ChatGPT, и преподаванием ИИ в медицинском образовании. Данное несоответствие означает применение студентами ChatGPT при решении конкретных учебных задач с целью получения быстрого ответа без осмысления алгоритмов ИИ-технологии и отсутствие

в учебных планах целостного раздела, подробно освещающего различные аспекты ИИ-технологии. Для адекватной подготовки будущих медицинских работников существует острая необходимость интеграции преподавания ИИ, а также его этических и правовых аспектов в учебные программы медицинских вузов [4].

Безусловный интерес представляет исследование Е. И. Горюшкина, в котором классифицируются методы и технологии обработки больших наборов данных, являющихся основой ИИ, и совершенно, на наш взгляд, справедливо предлагается изучать данный раздел только после освоения базовых дисциплин, результат которого – освоение цифровых технологий и основ математической статистики [5].

Проведенное авторами Т. В. Кормилицыной, Е. А. Бакулиной, С. И. Проценко, О. И. Пауткиной исследование проблемы включения в образовательные программы вузов модуля «Системы искусственного интеллекта» для студентов специальностей, не связанных непосредственно с ИТ-технологиями, показало значимость и своевременность введения изучения данного модуля именно для указанной категории студентов; также интерес представляют описанные в работе методические аспекты изучения дисциплины «Основы программирования на Python», акцентирующие внимание на том, что на сегодняшний день Python является стандартом индустрии анализа данных [6].

Конкретную практическую реализацию применения ИИ в медицине на уровне, доступном студенту, предлагает в своем исследовании В. Н. Русакова. Ценность предложенной методики заключается в двухуровневом подходе к решению задач диагностики заболевания: сначала студент работает с готовым сервисом, отвечая на вопросы о наличии определенных симптомов, а затем, самостоятельно изменяя код программы, оценивает результативность обученной модели [7].

В рамках исследования, представленного Е. М. Антоновой в [8], был рассмотрен разработанный автором и внедренный в практику обучения будущих врачей курс «Искусственный интеллект в медицине», а также произведена оценка готовности применения студентами-медиками искусственного интеллекта в профессиональной деятельности с помощью шкалы MAIRS-MS. Данная шкала, разработанная на основе применения многофакторного анализа авторами О. Карача, А. Чалышхан, К. Демир, представляет собой валидный инструмент оценки уровней готовности студентов-медиков к внедрению технологий и приложений ИИ в практическую работу врача [9].

Авторы В. П. Мудров, С. С. Иванов, М. Йовичич в своей статье [10] подчеркивают, что традиционная статистика является фундаментальной основой машинного обучения, алгоритмы которого выводятся из базовой математики, но усовершенствованы с учетом применения автоматизированного анализа, обработки больших данных и предоставления интерактивных визуализаций результатов. Следует отметить, что предмет «Вероятность и статистика» был введен в программу общеобразовательных учреждений практически одновременно с курсом «Введение в современные информационные и интеллектуальные технологии» в вузах, что способствовало установлению логической взаимосвязи между школьным обучением основам статистики и вузовскими курсами по искусственному интеллекту.

Другим аспектом вышеуказанной проблемы является обзор возможных областей применения ИИ в медицине и определение наиболее распространенных ситуаций, ограничивающих эффективность внедрения систем ИИ в практическую деятельность врача.

В обзорной статье Д. Р. Нурлыгаяновой, Р. З. Нурлыгаянова, А. М. Нурлыгаяновой рассматривается применение ИИ в различных областях медицины, включая диа-

гностику, прогнозирование, персонализированную терапию и клинические исследования. Для проведения обзора проводился поиск научных публикаций за последние 10 лет в базах данных PubMed, IEEE Xplore, ACM Digital Library и Science Direct [11].

В работе Кевина Чаморро, Рикардо Кальдерона Альвареса, Мелани Карвахаль Ахтти, Мойсеса Куинга приведен всесторонний библиометрический анализ достижений ИИ в медицине в период с 2017 по 2024 год, представленных научными работами из базы данных Scopus. С помощью библиометрических инструментов Bibliometrix и VOSviewer были проанализированы 6900 статей; результаты анализа демонстрируют возрастающий интерес к применению ИИ в медицине, что подтверждается значительным количеством публикаций, особенно в технологиях глубокого обучения и машинного обучения, которые оказывают существенное влияние на диагностику заболеваний, персонализированную медицину и стратегию лечения. По мнению авторов, безусловно перспективным представляется международное сотрудничество в данном направлении [12].

В работе К. С. Итинсона проведена классификация систем ИИ с описанием возможностей их применения при решении типовых задач анализа, диагностики и прогнозирования лечения [13].

В то же время авторы М. А. Шмонова и Т. Г. Авачева в работе [14] выделяют основные проблемы, препятствующие эффективному внедрению систем ИИ в медицинскую практику, такие как некорректный сбор данных, некомпетентность разработчиков, приверженность многих медицинских работников традиционным методикам диагностики и лечения, и предлагают возможные варианты решения указанных проблем.

Подобный обзор желательно излагать студенту-медику в форме алгоритма и с максимальной визуализацией информации.

Третьим аспектом методического обеспечения изучения современных интеллектуальных технологий является выбор способа их реализации.

В настоящее время для решения задач машинного обучения и искусственного интеллекта в медицине широкое распространение получило применение языка Python, что подтверждается многочисленными работами исследователей, работающих на стыке медицины и информационных технологий.

Например, в работе Г. Набиевой, Н. Абуталиповой, А. Калижановой представлены модели и методы ИИ для задач распознавания и сегментации медицинских изображений в области онкологии для ранней диагностики заболеваний, реализованные на языке Python [15]. Авторы Н. Равшанов, Ф. А. Мурадов, З. Нарзикулов предлагают использовать библиотеки языка Python для распознавания лиц с целью идентификации пациентов в лечебных учреждениях, что позволяет упростить процесс регистрации и ускорить доступ к критически важной информации о состоянии здоровья пациента [16]. Также можно отметить работу А. Д. Кубегеновой, Ж. С. Иксебаевой, Е. С. Кубегенова, в которой рассматривалось решение задачи определения групп ВИЧ-инфицированных пациентов методом кластеризации [17]. В статье Т. В. Золотовой, А. С. Марунько были проведены классификации социально значимых заболеваний на основе экологических показателей места проживания с помощью моделей машинного обучения, математическую основу которых составляют метод k-ближайших соседей, многослойный перцептрон, градиентный бустинг; при этом для построения моделей был использован программный инструмент Jupyter Notebook, поддерживающий язык программирования Python [18]. В своем исследовании И. А. Мишкин, А. В. Концевая, А. В. Гусев предложили новые методические подходы к прогнозированию сердечно-сосудистых событий с использованием алгоритмов классификации

RandomForestClassifier, GradientBoostingClassifier, ExtraTreesClassifier, XGBClassifier, LGBMClassifier, реализованных на языке Python; для оценки эффективности моделей прогноза использовали ROC-анализ [19].

Следует отметить, что сообщество зарубежных исследователей при обсуждении вопросов применения методов машинного обучения в медицинской практике уделяет особое внимание взаимодействию практикующих врачей и ИИ-специалистов.

В монографии Бруно Карпентьери и Паолы Лекка представлены проведенные авторами исследования в области анализа больших данных и искусственного интеллекта для медицинских задач. Авторы подчеркивают, что данная книга является результатом многолетнего междисциплинарного сотрудничества с клиницистами, программистами, математиками и инженерами. Также подробно рассмотрены различия между традиционными вычислительными подходами к обработке данных и циклом «эксперимент-данные-теория-модель-валидация» [20].

Всесторонний обзор новейших медицинских приложений, использующих генеративный искусственный интеллект, провели А. Хампария и Д. Гупта в монографии [21], важной особенностью которой является включение отзывов и впечатлений врачей и медицинского персонала о целесообразности использования и эффективности применения предлагаемых решений.

В книге А. Кумара, С. Рани, С. Рати, Н. Хемрадхани рассматриваются системы для диагностики и прогнозирования заболеваний, обработки клинических данных и медицинских услуг, лечения и мониторинга пациентов, основанные на ИИ. При этом авторы подчеркивают важность технологий мобильной связи 5G и 6G для применения в телемедицине, предоставляющих данные для подобных интеллектуальных систем [22].

Основы глубокого обучения и новейшие DL-алгоритмы для анализа медицинских изображений рассмотрели в своей монографии Р. Раджаникант Алувалу, М. Мехта, С. Сиарри [23]. Особое внимание в данной работе уделяется объяснимому искусственному интеллекту (XAI), который раскрывает «черный ящик» ML и устраняет семантический разрыв между ИИ-специалистами и медицинским сообществом.

Основной целью применения интеллектуальных технологий в высшем образовании является развитие у студентов системного и критического мышления. Поэтому после этапа разработки и реализации ИИ-модели для конкретной медицинской задачи необходимо осуществить серию компьютерных экспериментов для углубленного понимания работоспособности и адекватности предлагаемой модели.

Так, К. М. Бородина в работе [24] подчеркивает, что заинтересованность студентов в проведении экспериментов с целью разработки и модернизации существующих методов лечения является неотъемлемой частью учебного процесса в медицинском образовании.

Л. Л. Гатиятуллина отмечает, что в современном медицинском вузе осуществляется активный процесс симуляционного обучения, при участии в котором студент связывает воедино свои теоретические и прикладные знания путем искусственного воссоздания конкретной ситуации в комплексной интерактивной модели. При этом в качестве показателя, позволяющего оценить уровень сформированности практических умений, предлагается рассматривать способность студента оценивать принятые решения и перспективы его развития [25].

В статье Е. А. Уточкиной представлены результаты педагогического эксперимента, свидетельствующие об эффективности формирования исследовательских

умений у студентов-медиков за счет включения в образовательный процесс этапов исследовательской деятельности и методики их реализации [26].

В статье В. Е. Евдокимовой на основе обработки данных педагогического эксперимента было доказано, что применение предложенных авторами специальных заданий с профессиональным уклоном в большей степени способствует стимулированию познавательного интереса у студентов-медиков к современным информационным и интеллектуальным технологиям, чем задания, не имеющие отношения к врачебной деятельности [27].

Коллективная монография Н. В. Богачевой, О. П. Дуяновой, Е. П. Колеватых, С. В. Потехиной, М. Л. Смертиной, З. Р. Хасаншиной посвящена современным достижениям и перспективам интеграции искусственного интеллекта и биомедицинских наук [28].

Изданные за последнее время учебники и учебные пособия в области применения ИИ в здравоохранении, например учебник для вузов Ю. А. Антохиной, Т. М. Татарниковой [29], учебные пособия А. Л. Золкина, В. Д. Мунистера [30], С. А. Игнатьева, А. В. Кулигина, А. А. Игнатьева, В. В. Шалунова [31], Е. С. Митякова, А. Г. Шмелевой, А. И. Ладынина [32], М. П. Потапова, А. В. Павлова, М. Ю. Котловского [33], рассматривают возможности и примеры использования данной технологии в различных областях медицины и ориентированы на ознакомление студентами с теоретическим материалом.

Таким образом, проблема обучения студентов медицинских специальностей современным интеллектуальным технологиям представляется весьма актуальной и находит свое отражение в рассмотренных выше исследованиях. Вместе с тем следует отметить, что в представленных работах недостаточно внимания уделено именно методическим аспектам обучения данной категории студентов применению ИИ-технологий, их реализации с помощью свободно распространяемого программного обеспечения и проведению обучающего компьютерного эксперимента. Представленная авторами разработка в определенной мере способствует решению данной проблемы.

Материалы и методы / Materials and methods

Базу для проведения данного исследования составил учебный контингент Медицинского института ФГБОУ ВО «Мордовский государственный университет им. Н. П. Огарёва» в количестве 634 студентов.

При написании данной статьи использовались следующие материалы:

- общедоступные наборы файлов, необходимые для работы в области науки о данных Kaggle [34];
- документация по библиотекам и функциям языка Python, используемым при реализации алгоритмов машинного обучения [35];
- результаты тестирования студентов по разделу «Технологии машинного обучения и искусственного интеллекта в здравоохранении», проведенного с использованием электронной образовательной среды МГУ им. Н. П. Огарёва.

Основным методом исследования является предлагаемый авторами подход по реализации компьютерных экспериментов в лабораторном практикуме указанного раздела, позволяющий студентам более глубоко понять суть изучаемых методов машинного обучения и условий их применения. Данный подход включает:

- изучение основных математических формул, лежащих в основе того или иного метода машинного обучения;

- выполнение несложных вычислительных примеров на небольшом объеме данных с применением популярных табличных процессоров;
- изучение основных функций и библиотек языка Python для реализации конкретной практической задачи;
- проведение компьютерных экспериментов для более углубленного изучения выбранной математической модели и повышения ее точности.

Также методами исследования являлись:

- методы классификации объектов, применяемые в машинном обучении, в частности категориальный байесовский классификатор;
- методы планирования эксперимента, в частности полный факторный эксперимент, основанный на использовании наиболее распространенных способов повышения точности построенных классификационных моделей;
- двухфакторный дисперсионный анализ, использованный для обработки результатов педагогического эксперимента, проведенного с целью подтверждения преимуществ использования предлагаемого авторами подхода при изучении студентами медицинских вузов дисциплины «Введение в современные информационные и интеллектуальные технологии».

Результаты исследования / Research results

Результатом исследования является разработанный авторами подход к изучению базовых методов машинного обучения и выполнению лабораторных работ студентами непрофильных специальностей вузов по дисциплине «Введение в современные информационные и интеллектуальные технологии».

Предложенный подход включает: изучение математических формул, лежащих в основе выбранного метода машинного обучения, выполнение несложных вычислительных примеров на небольшом объеме данных для демонстрации математической модели, изучение основных конструкций и библиотек языка Python для реализации конкретной практической задачи, проведение компьютерных экспериментов для углубленного изучения выбранной математической модели и повышения ее точности.

Алгоритм реализации предлагаемого подхода представлен на рис. 1.

В качестве типичного примера применения данного подхода рассмотрим проведение компьютерного эксперимента и анализ его результатов для построения классификационной модели и повышения ее точности в процессе выполнения лабораторной работы «Применение категориального байесовского классификатора в диагностике заболеваний».

Приведем пошаговое описание этой схемы в данной лабораторной работе.

1. Целями данной лабораторной работы являются изучение студентами категориального байесовского классификатора и его реализация на примере диагностики диабета. Байесовский классификатор в данном случае выбран как один из самых несложных для понимания и реализации методов классификации категориальных данных в машинном обучении.

2. В теоретической части лабораторной работы приводятся основные сведения о категориальном байесовском классификаторе, областях его применения, достоинствах и недостатках данного метода классификации. Также описывается необходимый для реализации метода математический аппарат – теорема Байеса, позволяющая найти вероятность принадлежности объекта к определенному классу при определенном сочетании значений признаков, характерных для данного класса.



Рис. 1. Алгоритм реализации предлагаемого подхода

3. Данный шаг предполагает выполнение простейшего примера на применение теоремы Байеса, позволяющего студентам освоить математический аппарат и его реализацию в табличном процессоре Microsoft Excel, изученном в рамках выполнения предыдущих лабораторных работ. Такой подход позволяет студентам понять принципы работы систем искусственного интеллекта и избежать проблемы «черного ящика» при использовании готовых библиотек алгоритмов машинного обучения. Тем самым формируется критическое мышление, позволяющее оценить результаты применения ИИ для решения задач диагностики и прогнозирования.

В качестве примера, на котором можно ознакомиться с алгоритмом работы категориального байесовского классификатора, рассмотрим выборку из 20 пациентов (см. табл. 1), у которых были зафиксированы четыре показателя: артериальное давление (0 – нормальное, 1 – повышенное), уровень сахара в крови (0 – нормальный, 1 – повышенный), абдоминальное ожирение (0 – нет, 1 – есть), наличие диабета 2-го типа (0 – диабет не диагностирован, 1 – диабет диагностирован). Разделим исходные данные на обучающую и тестовую выборки согласно табл. 1.

Для обучающей выборки выполним расчеты вероятностей наличия каждой комбинации значений признаков согласно теореме Байеса, которой для данного примера соответствует формула:

$$P(Y|X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{P(Y) \cdot P(X_1 = x_1, X_2 = x_2, X_3 = x_3|Y)}{P(X_1 = x_1, X_2 = x_2, X_3 = x_3)}$$

Здесь:

– $P(Y)$ – вероятность того, что у человека не диагностирован диабет, определяется как отношение количества пациентов без диабета к общему количеству пациентов в обучающей выборке;

– $P(X_1=x_1, X_2=x_2, X_3=x_3)$ – вероятность появления определенной комбинации признаков, определяется как отношение количества повторений комбинации к общему количеству пациентов в обучающей выборке;

– $P(X_1=x_1, X_2=x_2, X_3=x_3 | Y)$ – вероятность появления определенной комбинации признаков при условии, что у пациента не диагностирован диабет, определяется как отношение количества повторений комбинации для пациентов без диабета к общему количеству пациентов в обучающей выборке;

– $P(Y | X_1=x_1, X_2=x_2, X_3=x_3)$ – вероятность того, что у пациента не будет диагностирован диабет, если у него появляется определенная комбинация признаков, определяется по вышеприведенной формуле.

Результаты вычислений представлены в табл. 2. В последнем столбце таблицы классификация диабета происходит следующим образом: если значение рассчитанной вероятности превышает 0,5, то для данной комбинации признаков можно считать, что у пациента диабет не диагностирован; в противном случае можно предполагать наличие диабета.

Таблица 1

Исходные данные

№	Артериальное давление	Уровень сахара в крови	Абдоминальное ожирение	Диабет	Выборка
1	0	0	0	0	Обучающая
2	0	0	1	0	
3	0	1	0	0	
4	0	1	1	1	
5	1	0	0	0	
6	1	0	1	0	
7	1	1	0	1	
8	1	1	1	1	
9	0	0	0	0	
10	0	0	1	1	
11	0	1	0	0	
12	0	1	1	0	
13	1	0	0	0	
14	1	0	1	0	Тестовая
15	1	1	0	0	
16	1	1	1	1	
17	1	0	0	0	
18	1	0	1	0	
19	1	1	0	0	
20	1	1	1	1	

Таблица 2

Расчет вероятностей по формуле Байеса

$X_1X_2X_3$	$P(X_1=x_1, X_2=x_2, X_3=x_3)$	$P(X_1=x_1, X_2=x_2, X_3=x_3 Y)$	$P(Y X_1=x_1, X_2=x_2, X_3=x_3)$	Диабет
000	0,15	0,15	0,69	0
001	0,15	0,08	0,35	1
010	0,15	0,15	0,69	0
011	0,15	0,08	0,35	1
100	0,15	0,15	0,69	0
101	0,08	0,08	0,69	0
110	0,08	0,00	0,00	1
111	0,08	0,00	0,00	1

Для оценки качества построенной модели выполним прогнозирование наличия диабета на данных тестовой выборки и сравним полученные значения с фактическими (табл. 3).

Таблица 3

Прогнозирование диабета для тестовой выборки

№	$X_1X_2X_3$	Диабет (фактическое значение)	Диабет (предсказанное значение)	Верно ли предсказанное значение?
14	101	0	0	Да
15	110	0	1	Нет
16	111	1	1	Да
17	100	0	0	Да
18	101	0	0	Да
19	110	0	1	Нет
20	111	1	1	Да

Для оценки точности построенной модели вычислим долю пациентов в тестовой выборке, для которых фактическое значение показателя «Диабет» совпадает с предсказанным, в результате получим значение 71,43%. В табл. 4 представлена матрица ошибок.

Таблица 4

Матрица ошибок

Матрица ошибок	T	F
P	43%	0%
N	29%	29%

Здесь:

- true positive (T&P) – количество правильных предсказаний отсутствия диабета;
- false positive (F&P) – количество неправильных предсказаний отсутствия диабета;
- true negative (T&N) – количество правильных предсказаний наличия диабета;
- false negative (F&N) – количество неправильных предсказаний наличия диабета.

Реализация данного примера в табличном процессоре Microsoft Excel предполагает наличие у студентов базовых умений работы с формулами и простейшими статистическими функциями.

Дальнейшее выполнение лабораторной работы осуществляется на языке Python, предоставляющего обширный набор библиотек для реализации алгоритмов машинного обучения, с применением сервиса Google Colab.

4. В качестве исходных данных для решения задачи диагностики диабета используется готовый набор данных “Diabetes Health Indicators Dataset” [34]. Расшифровка заголовков столбцов и их значений приведена в табл. 5.

5. Далее приводится последовательность операций на языке Python, реализующая модель категориального байесовского классификатора:

а) загрузка и описание набора данных (датасета), на основе которого будет проводиться классификация:

- считывание датасета из файла;
- определение столбцов-признаков, описывающих названия учтенных показателей пациента, и столбца, являющегося целевой переменной;
- определение типов столбцов и количества уникальных значений в них;

Таблица 5

Расшифровка заголовков столбцов и их значений в датасете

№	Признак	Расшифровка
0	Diabetes_binary	Целевая переменная, определяющая наличие диабета и принимающая два значения: 0 – диабет не диагностирован, 1 – диабет диагностирован
1	HighBP	Стабильно повышенное артериальное давление (0 = не наблюдается, 1 = наблюдается)
2	HighChol	Стабильно повышенный уровень холестерина (0 = не наблюдается, 1 = наблюдается)
3	CholCheck	Динамика уровня холестерина за последние 5 лет (0 = не проводилась, 1 = проводилась)
4	BMI	Индекс массы тела
5	Smoker	Курение (0 = не курит или курит от случая к случаю, 1 = курит постоянно не менее 1 пачки в день)
6	Stroke	Был ли у вас инсульт (0 = нет, 1 = да)
7	HeartDiseaseor Attack	Ишемическая болезнь сердца или инфаркт миокарда (0 = нет, 1 = да)
8	PhysActivity	Посильная физическая работа или регулярные занятия спортом в последние 30 дней (0 = нет, 1 = да)
9	Fruits	Употребление фруктов ежедневно не менее 500 г (0 = нет, 1 = да)
10	Veggies	Употребление овощей ежедневно не менее 500 г (0 = нет, 1 = да)
11	HvyAlcohol- Consump	Злоупотребление алкоголем (взрослые мужчины более 14 доз в неделю, взрослые женщины более 7 доз в неделю, одна доза по рекомендации ВОЗ соответствует 10 мл чистого спирта) (0 = нет, 1 = да)
12	GenHlth	Общее состояние здоровья (excellent = отлично, very good = очень хорошо, good = хорошо, fair = удовлетворительно, poor = плохо)
13	DiffWalk	Трудности при ходьбе или подъеме по лестнице (0 = нет, 1 = да)
14	Sex	Пол (0 = женский, 1 = мужской)
15	Age	Возраст: от 18 до 100 лет

b) предварительная обработка данных:

- удаление строк, содержащих пропущенные значения;
- преобразование столбцов, которые содержат значения, отличные от категориальных, к категориальному виду;

c) разделение датасета на признаки и целевую переменную;

d) деление признаков и целевой переменной на обучающую и тестовую выборки;

e) применение наивного байесовского классификатора для создания и обучения модели;

f) реализация прогнозирования на тестовой выборке с использованием обученной модели;

g) оценка точности разработанной модели и построение для нее матрицы ошибок.

Указанной последовательности соответствуют следующие фрагменты кода.

Для загрузки и определения размерности датасета используется следующий код:

```
import pandas as pd
# Загрузка данных
data = pd.read_csv('data.csv')
# Просмотр размера датасета
print(data.shape)
```


Полученный размер датасета: 253680 строк и 16 столбцов. Определение типов данных столбцов и количества уникальных значений в каждом столбце реализуется следующим кодом:

```
# Определение типов данных столбцов
print (data.dtypes)
# определение количества уникальных значений в каждом столбце
data.nunique ()
```

Анализ полученных результатов показал, что столбец *GenHlth* имеет тип данных *object*, отличный от целочисленного, а столбцы *BMI* и *Age* содержат большое количество уникальных значений (64 значения), что приводит к значительному увеличению машинного времени при реализации алгоритма, тем самым снижая эффективность применения категориального байесовского классификатора.

Первым этапом предварительной обработки данных является удаление строк, содержащих пропущенные значения:

```
# Удаление строк, содержащих пропущенные значения
data.dropna()
print(data.shape)
```

После применения функции *dropna()* размер датасета остался прежним: (253680, 16), то есть в нем не было пропущенных значений.

Выполним преобразование фиксированных ранее столбцов *GenHlth*, *BMI* и *Age* к категориальному виду.

Так как в столбцах *BMI* и *Age* находится большое количество уникальных значений, то необходимо разбить эти значения на меньшее число интервалов. Известно, что по индексу массы тела *BMI* пациента можно отнести к одной из следующих категорий:

- $BMI < 19$ – недостаточный вес;
- $BMI \in [19, 25)$ – нормальный вес;
- $BMI \in [25, 30)$ – избыточный вес;
- $BMI \in [30, 35)$ – ожирение 1-го класса;
- $BMI \in [35, 40)$ – ожирение 2-го класса;
- $BMI \geq 40$ – ожирение 3-го класса.

Каждый интервал закодируем соответственно целыми числами от 0 до 5:

```
# Интервалы и категории для BMI
bmi_intervals = [-float('inf'), 19, 25, 30, 35, 40, float('inf')]
bmi_labels_ordered = [0, 1, 2, 3, 4, 5]
# Замена значений столбца BMI на категории
data['BMI'] = pd.cut(data['BMI'], right=False, include_lowest=True, bins=bmi_intervals, labels=bmi_labels_ordered)
```

Аналогично осуществляется разбиение на категории столбца *Age*.

Значения столбца *GenHlth* являются строковыми, поэтому для их категоризации применяется порядковая шкала, реализованная в виде словаря:

```
# Создание словаря для замены значений в столбце GenHlth
replacement_dict = {'poor': 0, 'fair': 1, 'good': 2, 'very good': 3, 'excellent': 4}
# Замена значений
data['GenHlth'] = data['GenHlth'].map(replacement_dict).astype(int)
```

Разделим датасет на признаки X, к которым относятся столбцы с индексами 1–15, и целевую переменную y – столбец *Diabetes_binary*:

```
# Разделение на признаки и целевую переменную
y = data['Diabetes_binary']
X = data.drop('Diabetes_binary', axis=1)
```

Для разделения данных на обучающую и тестовую выборки используется функция `train_test_split()` из библиотеки Scikit-learn:

```
from sklearn.model_selection import train_test_split
# Разделение данных на обучающую и тестовую выборки
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2, random_state=13)
```

Используя модель категориального байесовского классификатора `CategoricalNB()`, с помощью функции `fit()` проведем ее обучение:

```
from sklearn.naive_bayes import CategoricalNB
# Создание и обучение модели
model = CategoricalNB()
model.fit(X_train, y_train)
```

С помощью функции `predict()` выполним прогнозирование значений для тестовой выборки на основе обученной модели:

```
# Прогнозирование на тестовой выборке
predicted = model.predict(X_test)
```

Используя функции `accuracy_score()` и `confusion_matrix()` из `sklearn.metrics`, вычислим точность модели и построим матрицу ошибок:

```
from sklearn.metrics import accuracy_score, confusion_matrix
# Оценка точности модели
accuracy = accuracy_score(y_test, predicted)
# Построение матрицы ошибок
conf_matrix = confusion_matrix(y_test, predicted)
```

В результате полученная точность составила 82,62%, а матрица ошибок, представленная в виде тепловой карты, изображена на рис. 2:

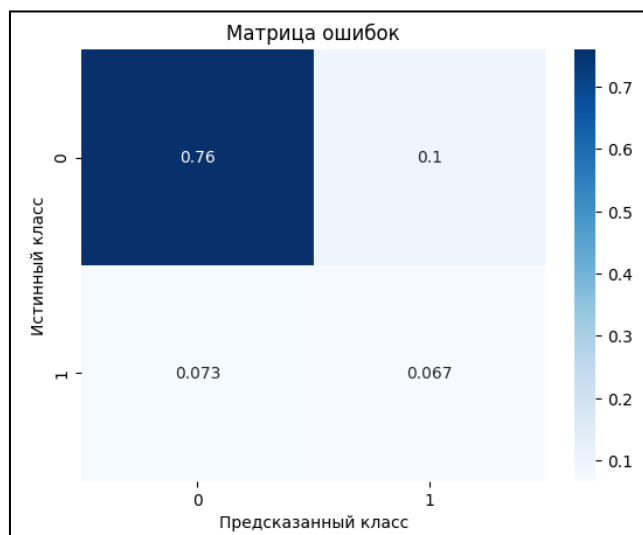


Рис. 2. Тепловая карта матрицы ошибок

Таким образом, студенты изучают основные этапы алгоритма проведения диагностики заболевания с помощью категориального байесовского классификатора и их реализацию на языке Python.

6. Точность полученной модели оказалось достаточно большой с точки зрения машинного обучения, но недостаточной с точки зрения практикующего врача, для которого приемлемой считается погрешность 3–5%.

Удаление дубликатов

Наличие дубликатов в датасете может негативно сказываться на обучении ИИ-моделей, поэтому перед проведением обучения их необходимо удалить, что представлено в коде ниже:

```
# Проверка на дубликаты
print('Количество одинаковых записей:')
print(data.duplicated().sum())
# Удаление дубликатов
data=data.drop_duplicates()
```

Удаление столбцов, имеющих наименьшую корреляцию с целевой переменной

Представленный ниже код рассчитывает корреляцию между признаками и целевой переменной (в данном случае – *Diabetes_binary*), удаляет наименее важные признаки и строит столбчатую диаграмму со значениями корреляции (рис. 3):

```
from matplotlib import pyplot as plt
# Оценка корреляции
corr_data = data.corrwith(data.Diabetes_binary)
# Построение диаграммы корреляции
corr_data.plot(kind='bar', grid=True)
plt.show()
# Отбор 10-ти наиболее влияющих признаков
corr_data = corr_data.abs()
cols=corr_data.nlargest(n=10,keep='first').index.to_list()
```

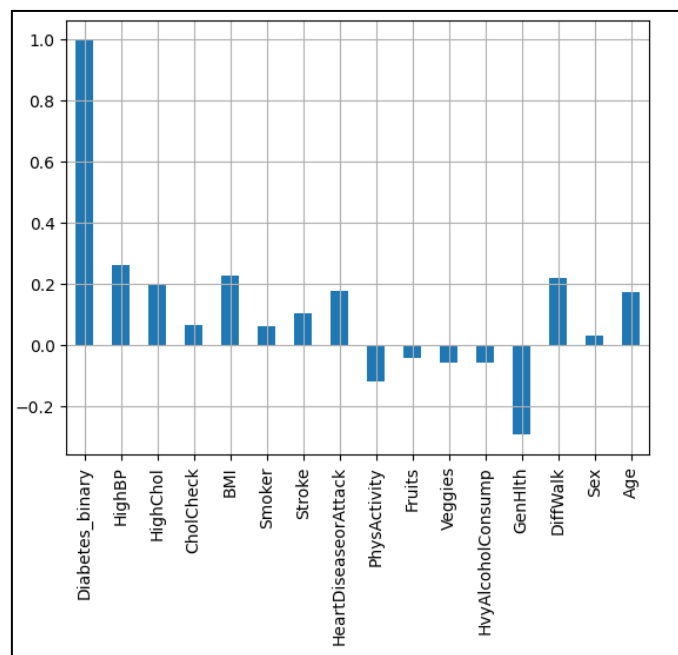


Рис. 3. Диаграмма коэффициентов корреляции признаков с целевой переменной

Из рис. 3 следует, что наибольшая по абсолютному значению взаимосвязь целевой переменной *Diabetes_binary* наблюдается со следующими признаками: *GenHlth*, *HighBP*, *BMI*, *DiffWalk*, *HighChol*, *HeartDiseaseorAttack*, *Age*, *PhysActivity*, *Stroke*.

Балансировка классов

Если в датасете количество записей одного класса значительно больше количества записей альтернативного класса, то построенная модель может хуже распознавать менее представленный класс. Например, при нахождении количества записей в каждом классе представленного датасета с помощью следующего кода:

```
print('Распределение классов в датасете:')
print(y.value_counts())
```

было получено, что класс «Люди с диабетом» представлен значительно меньшим количеством записей – 35346 пациентов, чем альтернативный класс «Здоровые люди» – 218334 пациента, что соответствует реальной ситуации. Поэтому рекомендуется провести балансировку классов, например, с использованием метода случайного уменьшения выборки (*undersampling*), сокращающего количество записей в более крупном классе до уровня меньшего класса.

```
from imblearn.under_sampling import RandomUnderSampler
#балансировка классов
rus = RandomUnderSampler(random_state=13)
X, y = rus.fit_resample(X, y)
```

8. После изучения студентами возможных методов повышения точности классификационной модели приступают к построению плана полного факторного эксперимента для определения наиболее эффективного метода или их сочетания для увеличения точности классификации. Так как в данном случае были представлены три

таких метода (фактора), то план полного факторного эксперимента будет включать следующие семь экспериментов: А, В, С, АВ, АС, ВС, АВС, где А – удаление дубликатов, В – удаление признаков, имеющих наименьшую корреляцию с целевой переменной, С – балансировка классов. Для каждого эксперимента будет применен приведенный выше алгоритм категориального байесовского классификатора (табл. 6).

Таблица 6

Результаты полного факторного эксперимента

	Число строк	Число столбцов	Точность модели, %	T&P, %	T&N, %	F&P, %	F&N, %
Исходный вариант	253680	16	82,62	75,94	6,69	7,29	10,08
Удаление дубликатов (А)	62828	16	70,82	59,72	11,10	17,30	11,88
Удаление слабо коррелированных признаков (В)	253680	10	82,96	76,63	6,33	7,65	9,39
Балансировка классов (С)	35346	16	74,08	35,99	38,09	12,30	13,62
АВ	11474	10	60,35	51,24	9,11	29,32	10,33
ВС	35346	10	73,82	35,94	37,87	12,51	13,67
АС	17900	16	63,92	30,14	33,78	14,78	21,30
АВС	4498	10	58,78	24,83	33,94	15,33	25,89

Примечание. Обозначения T&P, F&P, T&N, F&N приведены в пункте 3; двух- и трехбуквенные обозначения в первом столбце таблицы (АВ, ВС, АС, АВС) соответствуют комбинациям методов.

9. Последним этапом выполнения лабораторной работы является представление студентом аргументированных выводов о целесообразности и эффективности применения модели категориального байесовского классификатора к конкретной задаче и способах улучшения ее точности. Например, результаты, приведенные в табл. 6, позволяют сделать вывод, что в данном случае метод В оказался самым эффективным, остальные методы и их сочетания, приводящие к заметному сокращению числа записей в датасете, показывают явно неудовлетворительные результаты для их практического применения. Кроме того, для методов А и АВ наблюдаются наибольшие значения ложноположительных предсказаний, а также их преобладание над ложноотрицательными значениями, что является критичной ситуацией для пациента, поскольку не диагностированное своевременно заболевание в большинстве областей медицины способно в дальнейшем значительно ухудшить прогноз его течения; ложноотрицательный же результат, как правило, приводит к более углубленному обследованию пациентов, и в определенных случаях предварительно поставленный диагноз может быть снят.

Дополнительным заданием, демонстрирующим возможное последующее использование разработанной модели, является предсказание наличия или отсутствия диабета на показателях конкретного студента.

Для доказательства эффективности предлагаемого авторами подхода был проведен педагогический эксперимент в форме тестирования по разделам изученной темы и выполненной лабораторной работы «Категориальный байесовский классификатор» со студентами I курса Медицинского института Мордовского государственного университета им. Н. П. Огарёва. Из учебного контингента 2024/2025 года обучения случайным образом были отобраны 325 студентов в контрольную группу и 309 студентов в опытную группу. В опытной группе студенты обучались по схеме, представленной на рис. 1, в контрольной группе – по аналогичной схеме, но без включения пунктов 3 «Реализация простейшего примера на математическую модель» и 8

«Построение плана полного факторного эксперимента». Разделы теста представлены в табл. 7, в каждом разделе предлагалось от 5 до 10 вопросов, при этом оценивался средний балл правильных ответов по каждому разделу по 10-балльной шкале.

Примеры вопросов тестирования по разделам:

1. Математическая модель (формула Байеса)
 - При рентгеновском обследовании вероятность обнаружить заболевание туберкулезом у больного равна 0,9, вероятность принять здорового человека за больного равна 0,01. Доля больных туберкулезом по отношению ко всему населению равна 0,001. Найти вероятность того, что случайный человек здоров, если он был признан больным при обследовании (вопрос открытого типа).
2. Предобработка данных
 - Выберите методы, относящиеся к предобработке данных (вопрос закрытого типа с выбором нескольких вариантов ответа);
 - Что означает кодирование категориальных переменных (вопрос закрытого типа с выбором одного варианта ответа)?
3. Разделение данных: 1) на признаки и целевую переменную; 2) на тестовую и обучающую выборки
 - Установите соответствие между двумя множествами (слева термины «тестовая выборка», «обучающая выборка», справа – их определения) (вопрос на установление соответствия).
 - Выберите из списка фрагмент кода, позволяющий провести в Python разделение данных на признаки и целевую переменную (вопрос закрытого типа с выбором одного варианта ответа).
4. Создание и обучение модели классификации
 - Напишите имя функции Python, с помощью которой происходит обучение модели классификации (вопрос открытого типа).
 - Расположите указанные функции Python в порядке их использования при создании и обучении модели классификации (вопрос закрытого типа на сортировку).
5. Оценка точности модели
 - Установите соответствие между двумя множествами (слева термины TP, FP, TN, FN, справа – их определения) (вопрос на установление соответствия).
 - Вычислите метрику Recall по заданным значениям TP и FN (вопрос открытого типа).
6. Методы повышения точности модели
 - Какие из перечисленных методов можно отнести к методам повышения точности модели (вопрос закрытого типа с выбором нескольких вариантов ответа)?

Тестирование проходило с помощью электронной информационно-образовательной среды университета. Результаты педагогического эксперимента приведены в табл. 7.

Таблица 7

Средний балл по разделам тестирования

№	Наименование раздела тестирования	Контрольная	Опытная
1	Математическая модель (формула Байеса)	4,1	7,1
2	Предобработка данных	9,4	9,5
3	Разделение данных: 1) на признаки и целевую переменную; 2) на тестовую и обучающую выборки	9,1	9,5
4	Создание и обучение модели классификации	8,8	8,9

5	Оценка точности модели	4,6	7,8
6	Методы повышения точности модели	4,5	7,2

С помощью двухфакторного дисперсионного анализа (табл. 8) было доказано достоверное различие средних баллов в столбцах (опытная и контрольная группы) и строках (разделы тестирования).

Таблица 8

Двухфакторный дисперсионный анализ

Источник вариации	SS	df	MS	F	P-Значение	F критическое
Строки	33,75	5	6,75	5,79	0,04	5,05
Столбцы	7,52	1	7,52	6,65	0,05	6,61
Погрешность	5,83	5	1,17			
Итого	47,11	11				

Заключение / Conclusion

В данной статье описывается разработанный авторами подход по реализации компьютерных экспериментов в лабораторном практикуме раздела «Технологии машинного обучения и искусственного интеллекта в здравоохранении», входящего в состав дисциплины «Введение в современные информационные и интеллектуальные технологии». Предлагаемый подход продемонстрирован на примере лабораторной работы «Применение категориального байесовского классификатора в диагностике заболеваний», выполняемой студентами Медицинского института. В результате выполнения представленного алгоритма лабораторной работы обучаемый овладевает навыками применения методов машинного обучения при решении задач диагностики, включающими проведение «ручных» расчетов на основе табличного процессора Microsoft Excel, создание программного кода по приведенному образцу, проведение полнофакторного компьютерного эксперимента, формулировку соответствующих выводов на основе полученных результатов. Таким образом, студент получает инструментарий для выполнения самостоятельных исследований в научной и практической деятельности.

Анализ предложенного подхода показал, что его реализация в учебном процессе медицинских специальностей вузов, продемонстрированная на примере выполнения описанной лабораторной работы, не требует глубокого знания вероятностных методов машинного обучения, так как формула Байеса достаточно проста для понимания и применения; используемый программный код универсален с точки зрения применения других методов классификации и представляется относительно несложным для студентов медицинских вузов, не являющихся профессионалами в разработке программного обеспечения. Кроме того, достоинство данного подхода – возможность использования программного кода в сервисе Google Colab, предоставляющего средства работы с кодом на языке Python через Jupyter Notebook без установки дополнительного программного обеспечения.

Применение такого подхода при обучении студентов-медиков I курса Мордовского государственного университета показало его эффективность с точки зрения усвоения и понимания обучающимися достаточно сложного для их уровня материала, заинтересованность в освоении современных интеллектуальных технологий, выработке навыков проведения экспериментальных работ.

Разработанный подход может быть использован в качестве основы при разработке методических материалов по разделу «Технологии машинного обучения и искусственного интеллекта в здравоохранении».

Ссылки на источники / References

1. Перечень поручений по итогам конференции по искусственному интеллекту (утв. Президентом РФ 31.12.2020 № Пр-2242). – URL: <http://www.kremlin.ru/acts/assignments/orders/64859>
2. Бихатова Э. Т., Иванчук О. В. Проблема формирования цифровых компетенций у студентов медицинских вузов // ЦИТИСЭ. – 2021. – № 4(30). – С. 595–605. DOI: 10.15350/2409-7616.2021.4.55.
3. Забелин Д. А., Плащевая Е. В. Искусственный интеллект в системе подготовки медицинских кадров // ЦИТИСЭ. – 2023. – № 3(37). – С. 28–39. DOI: 10.15350/2409-7616.2023.3.03.
4. Weidener L., Fischer M. Artificial Intelligence in Medicine: Cross-Sectional Study Among Medical Students on Application, Education, and Ethical Aspects // JMIR Medical Education. – 2024. – Vol. 10. – P. 51247–51265. DOI: <https://doi.org/10.2196/51247>. – URL: <https://mededu.jmir.org/2024/1/e51247/PDF>
5. Горюшкин Е. И. Интеграция искусственного интеллекта в образовательный процесс медицинских вузов // Азимут научных исследований: педагогика и психология. – 2024. – Т. 13. – № 1(46). – С. 38–41.
6. Кормилицына Т. В., Бакулина Е. А., Проценко С. И., Пауткина О. И. Изучение модуля «Системы искусственного интеллекта» бакалаврами педагогических вузов // Перспективы науки. – 2023. – № 11(170). – С. 344–347.
7. Русакова В. Н. Некоторые методические приемы введения понятия искусственного интеллекта при изучении информационных технологий в медицинском вузе // Международный журнал гуманитарных и естественных наук. – 2022. – № 3-2(66). – С. 231–234. DOI: 10.24412/2500-1000-2022-3-2-231-234.
8. Антонова Е. М. Цифровизация мирового здравоохранения как драйвер стратегии обучения студентов медицинских вузов // ЦИТИСЭ. – 2022. – № 3 (33). – С. 141–151.
9. Karacha O., Chalyshkan S. A., Demir K. Medical Artificial Intelligence Readiness Scale for Medical Students (MAIRMS) – Development, Validity and Reliability Study // BMC Med Education. – 2021. – Vol. 21. – Art. 112. DOI: <https://doi.org/10.1186/s12909-021-02546-6>.
10. Мудров В. П., Иванов С. С., Йовичич М. Машинное обучение в лабораторной медицине // Лабораторная медицина. – 2022. – № 13. – С. 23–34.
11. Нурлыгаянова Д. Р., Нурлыгаянов Р. З., Нурлыгаянова А. М. Возможности использования искусственного интеллекта в медицине // Медицина. Социология. Философия. Прикладные исследования. – 2024. – № 3. – С. 125–131.
12. Chamorro K., Álvarez R. C., Ahtty M. C., Quinga M. Comprehensive bibliometric analysis of advancements in artificial intelligence applications in medicine using Scopus database // Franklin Open. – 2025. – Vol. 10. – P. 100212–100225. DOI: <https://doi.org/10.1016/j.fraope.2025.100212>. – URL: <https://www.sciencedirect.com/science/article/pii/S2773186325000027>
13. Итинсон К. С. Информатизация медицинского образования: системы искусственного интеллекта в обучении студентов и врачей // Балтийский гуманитарный журнал. – 2020. – Т. 9. – № 3(32). – С. 91–93. DOI: 10.26140/bgз3-2020-0903-0021.
14. Шмонова М. А., Авачева Т. Г. Пути повышения эффективности внедрения систем искусственного интеллекта в медицинской практике // Digital Diagnostics. – 2023. – Т. 4. – № S1. – С. 148–150.
15. Набиева Г., Абуталипова Н., Калижанова А. Использование моделей, методов и алгоритмов распознавания и сегментации медицинских изображений в управлении в сфере образования и здравоохранения при разработке образовательных программ по направлению «IT-медицина» // Вестник Казахской академии транспорта и коммуникаций им. М. Тынышпаева. – 2024. – № 2(131). – С. 431–442. DOI: 10.52167/1609-1817-2024-131-2-431-442.
16. Равшанов Н., Мурадов Ф. А., Нарзикулов З. Алгоритм распознавания лиц при регистрации пациентов клиник // Проблемы вычислительной и прикладной математики. – 2023. – № 3(49). – С. 143–152.
17. Кубегенова А. Д., Иксебаева Ж. С., Кубегенов Е. С. Методы интеллектуального анализа и применения технологии data Mining в области медицины // Инновационные научные исследования. – 2021. – № 8-1(10). – С. 48–55. DOI: 10.5281/zenodo.5529228.
18. Золотова Т. В., Марунько А. С. Методы интеллектуальной обработки данных для исследования влияния окружающей среды на заболеваемость населения в Москве // Статистика и Экономика. – 2024. – Т. 21. – № 2. – С. 72–82. DOI: 10.21686/2500-3925-2024-2-72-82.
19. Мишкин И. А., Концевая А. В., Гусев А. В. Разработка и тестирование новых методических подходов прогнозирования сердечно-сосудистых событий у здоровых людей с использованием технологии машинного обучения на базе международного исследования «Интерэпид» // Профилактическая медицина. – 2024. – Т. 27. – № 3. – С. 72–79. DOI: 10.17116/profmed20242703172.

20. Bruno Carpentieri B., Lecca P. Big Data Analysis and Artificial Intelligence for Medical Sciences. – Wiley, 2024. – 433 p.
21. Khamparia A., Gupta D. Generative Artificial Intelligence for Biomedical and Smart Health Informatics. – Wiley-IEEE Press, 2025. – 700 p.
22. Kumar A., Rani S., Rathee S., Hemrajani N. Artificial Intelligence in Medicine and Healthcare. – CRC Press, 2025. – 214 p.
23. Rajanikanth Aluvalu R., Mehta M., Siarry S. Explainable AI in Health Informatics. – Springer, 2024. – 287 p.
24. Бородина К. М. Исследовательский процесс в медицинском образовании // Балтийский гуманитарный журнал. – 2021. – Т. 10. – № 1(34). – С. 37–39. DOI: 10.26140/bgj3-2021-1001-0007.
25. Гатиятуллина Л. Л., Нигманова Г. И., Яппарова Р. Р. Симуляционные технологии в обучении как этап формирования практических умений студентов // Педагогическое образование. – 2024. – Т. 5. – № 1. – С. 247–253.
26. Уточкина Е. А. Методологические аспекты формирования исследовательских умений у будущих врачей в процессе изучения химии в медицинском вузе // Мир науки. Педагогика и психология. – 2023. – Т. 11. – № 4. – С. 1–9.
27. Евдокимова В. Е., Каткова А. Л. Стимулирование профессионального интереса студентов к практическим занятиям по информатике // Мир науки. Педагогика и психология. – 2020. – Т. 8. – № 1. – С. 24.
28. Богачева Н. В., Дуянова О. П., Колеватых Е. П. и др. Интеграция искусственного интеллекта и биомедицинских наук: современные достижения и перспективы. – Н. Новгород: Профессиональная наука, 2024. – 127 с.
29. Антохина Ю. А., Татарникова Т. М. Методы и алгоритмы искусственного интеллекта: учеб. для вузов. – СПб.: Лань, 2025. – 304 с.
30. Золкин А. Л., Мунистер В. Д. Реализация принципов организации и использования средств машинного обучения и искусственного интеллекта в медицине: учеб. пособие. – Самара: ЧУОО ВО «Медицинский университет «Реавиз», 2024. – 127 с.
31. Игнатьев С. А., Кулигин А. В., Игнатьев А. А., Шалунов В. В. Интеллектуальные системы и технологии в медицинском образовании: учеб. пособие. – Саратов: Общество с ограниченной ответственностью Издательство «КУБиК», 2024. – 160 с.
32. Митяков Е. С., Шмелева А. Г., Ладынин А. И. Искусственный интеллект и машинное обучение: учеб. пособие. – СПб.: Лань, 2025. – 252 с.
33. Потапов М. П., Павлов А. В., Котловский М. Ю. Искусственный интеллект в здравоохранении: учеб. пособие. – Ярославль: ИПК «Индиго», 2024. – 72 с.
34. Общедоступные наборы файлов, необходимые для работы в области науки о данных. UCI Heart Disease Data. – URL: <https://www.kaggle.com/datasets>
35. Машинное обучение в Python scikit-learn. – URL: <https://scikit-learn.ru/>

1. *Perechen' poruchenij po itogam konferencii po iskusstvennomu intellektu [List of assignments following the conference on artificial intelligence] (utv. Prezidentom RF 31.12.2020 № Pr-2242).* Available at: <http://www.kremlin.ru/acts/assignments/orders/64859> (in Russian).
2. Bihatova, E. T., & Ivanchuk, O. V. (2021). "Problema formirovanie cifrovyyh kompetencij u studentov medicinskih vuzov" [The problem of developing digital competences in medical students], *CITISE*, № 4(30), pp. 595–605. DOI: 10.15350/2409-7616.2021.4.55 (in Russian).
3. Zabelin, D. A., & Plashcheyaya, E. V. (2023). "Iskusstvennyj intellekt v sisteme podgotovki medicinskih kadrov" [Artificial Intelligence in the Medical Personnel Training System], *CITISE*, № 3(37), pp. 28–39. DOI: 10.15350/2409-7616.2023.3.03 (in Russian).
4. Weidener, L., & Fischer, M. (2024). "Artificial Intelligence in Medicine: Cross-Sectional Study Among Medical Students on Application, Education, and Ethical Aspects", *JMIR Medical Education*, vol. 10, pp. 51247–51265. DOI: <https://doi.org/10.2196/51247>. Available at: <https://mededu.jmir.org/2024/1/e51247/PDF> (in English).
5. Goryushkin, E. I. (2024). "Integraciya iskusstvennogo intellekta v obrazovatel'nyj process medicinskih vuzov" [Integration of artificial intelligence into the educational process of medical universities], *Azimut nauchnyh issledovanij: pedagogika i psihologiya*, t. 13, № 1(46), pp. 38–41 (in Russian).
6. Kormilicyna, T. V., Bakulina, E. A., Procenko, S. I., & Pautkina, O. I. (2023). "Izuchenie modulya "Sistemy iskusstvennogo intellekta" bakalavrami pedagogicheskikh vuzov" [Study of the module "Artificial Intelligence Systems" by undergraduates at pedagogical universities], *Perspektivy nauki*, № 11(170), pp. 344–347 (in Russian).
7. Rusakova, V. N. (2022). "Nekotorye metodicheskie priemy vvedeniya ponyatiya iskusstvennogo intellekta pri izuchenii informacionnyh tekhnologij v medicinskom vuze" [Some methodological techniques for introducing the concept of artificial intelligence in the study of information technology in a medical university], *Mezhdunarodnyj zhurnal gumanitarnykh i estestvennykh nauk*, № 3-2(66), pp. 231–234. DOI: 10.24412/2500-1000-2022-3-2-231-234 (in Russian).

8. Antonova, E. M. (2022). "Cifrovizaciya mirovogo zdravooohraneniya kak drajver strategii obucheniya studentov medicinskih vuzov" [Digitalization of global healthcare as a driver of the strategy for training medical students], *CITISE*, № 3 (33), pp. 141–151 (in Russian).
9. Karacha, O., Chalyshkan, S. A., & Demir, K. (2021). "Medical Artificial Intelligence Readiness Scale for Medical Students (MAIRS-MS) – Development, Validity and Reliability Study", *BMC Med Education*, vol. 21, art. 112. DOI: <https://doi.org/10.1186/s12909-021-02546-6> (in English).
10. Mudrov, V. P., Ivanov, S. S., & Jovichich, M. (2022). "Mashinnoe obuchenie v laboratornoj medicine" [Machine learning in laboratory medicine], *Laboratornaya medicina*, № 13, pp. 23–34 (in Russian).
11. Nurlygayanova, D. R., Nurlygayanov, R. Z., & Nurlygayanova, A. M. (2024). "Vozmozhnosti ispol'zovaniya iskusstvennogo intellekta v medicine" [Potential of using artificial intelligence in medicine], *Medicina. Sociologiya. Filosofiya. Prikladnye issledovaniya*, № 3, pp. 125–131 (in Russian).
12. Chamorro, K., Álvarez, R. C., Ahtty, M. C., & Quinga, M. (2025). "Comprehensive bibliometric analysis of advancements in artificial intelligence applications in medicine using Scopus database", *Franklin Open*, vol. 10, pp. 100212–100225. DOI: <https://doi.org/10.1016/j.fraope.2025.100212>. Available at: <https://www.sciencedirect.com/science/article/pii/S2773186325000027> (in English).
13. Itinson, K. S. (2020). "Informatizaciya medicinskogo obrazovaniya: sistemy iskusstvennogo intellekta v obuchenii studentov i vrachej" [Informatization of medical education: artificial intelligence systems in teaching students and doctors], *Baltiyskij gumanitarnyj zhurnal*, t. 9, № 3(32), pp. 91–93. DOI: 10.26140/bg3-2020-0903-0021 (in Russian).
14. Shmonova, M. A., & Avacheva, T. G. (2023). "Puti povysheniya effektivnosti vnedreniya sistem iskusstvennogo intellekta v medicinskoj praktike" [Ways to improve the efficiency of implementing artificial intelligence systems in medical practice], *Digital Diagnostics*, t. 4, № S1, pp. 148–150 (in Russian).
15. Nabieva, G., Abutalipova, N., & Kalizhanova, A. (2024). "Ispol'zovanie modelej, metodov i algoritmov raspoznavaniya i segmentacii medicinskih izobrazhenij v upravlenii v sfere obrazovaniya i zdravooohraneniya pri razrabotke obrazovatel'nyh programm po napravleniyu "IT-medicina" [The use of models, methods and algorithms for recognition and segmentation of medical images in the field of education and health care management for the development of educational programs in the field of "IT medicine"]], *Vestnik Kazahskoj akademii transporta i kommunikacij im. M. Tynyshpaeva*, № 2(131), pp. 431–442. DOI: 10.52167/1609-1817-2024-131-2-431-442 (in Russian).
16. Ravshanov, N., Muradov, F. A., & Narzikulov, Z. (2023). "Algoritm raspoznavaniya lic pri registracii pacientov klinik" [Facial recognition algorithm for registering patients in clinics], *Problemy vychislitel'noj i prikladnoj matematiki*, № 3(49), pp. 143–152 (in Russian).
17. Kubegenova, A. D., Iksebaeva, Zh. S., & Kubegenov, E. S. (2021). "Metody intellektual'nogo analiza i primeneniya tekhnologii data Mining v oblasti mediciny" [Methods of Intelligent Analysis and Application of Data Mining Technology in the Field of Medicine], *Innovacionnye nauchnye issledovaniya*, № 8-1(10), pp. 48–55. DOI: 10.5281/zenodo.5529228 (in Russian).
18. Zolotova, T. V., & Marun'ko, A. S. (2024). "Metody intellektual'noj obrabotki dannyh dlya issledovaniya vliyaniya okruzhayushchej sredy na zaboлеваemost' naseleniya v Moskve" [Methods of intelligent data processing for studying the influence of the environment on the incidence of diseases in Moscow], *Statistika i Ekonomika*, t. 21, № 2, pp. 72–82. DOI: 10.21686/2500-3925-2024-2-72-82 (in Russian).
19. Mishkin, I. A., Koncevaya, A. V., & Gusev, A. V. (2024). "Razrabotka i testirovanie novyh metodicheskikh podhodov prognozirovaniya serdechno-sosudistyh sobytij u zdorovyh lyudej s ispol'zovaniem tekhnologii mashinnogo obucheniya na baze mezhdunarodnogo issledovaniya "Interepid" [Development and testing of new methodological approaches to predicting cardiovascular events in healthy people using machine learning technology based on the international study "Interepid"]], *Profilakticheskaya medicina*, t. 27, № 3, pp. 72–79. DOI: 10.17116/prof-med20242703172 (in Russian).
20. Bruno Carpentieri, B., & Lecca, P. (2024). *Big Data Analysis and Artificial Intelligence for Medical Sciences*, Wiley, 433 p. (in English).
21. Khamparia, A., & Gupta, D. (2025). *Generative Artificial Intelligence for Biomedical and Smart Health Informatics*, Wiley-IEEE Press, 700 p. (in English).
22. Kumar, A., Rani, S., Rathee, S., & Hemrajani, N. (2025). *Artificial Intelligence in Medicine and Healthcare*, CRC Press, 214 p. (in English).
23. Rajanikanth Aluvalu, R., Mehta, M., & Siarry, S. (2024). *Explainable AI in Health Informatics*, Springer, 287 p. (in English).
24. Borodina, K. M. (2021). "Issledovatel'skij process v medicinskom obrazovanii", *Baltiyskij gumanitarnyj zhurnal*, t. 10, № 1(34), pp. 37–39. DOI: 10.26140/bg3-2021-1001-0007 (in Russian).
25. Gatiyatullina, L. L., Nigmanova, G. I., & Yapparova, R. R. (2024). "Simulyacionnye tekhnologii v obuchenii kak etap formirovaniya prakticheskikh umenij studentov" [Simulation technologies in education as a stage of students' practical skills development], *Pedagogicheskoe obrazovanie*, t. 5, № 1, pp. 247–253 (in Russian).

26. Utochkina, E. A. (2023). "Metodologicheskie aspekty formirovaniya issledovatel'skih umenij u budushchih vrachej v processe izucheniya himii v medicinskom vuze" [Methodological aspects of developing research skills among preservice doctors in the process of studying chemistry in a medical university], *Mir nauki. Pedagogika i psihologiya*, t. 11, № 4, pp. 1–9 (in Russian).
27. Evdokimova, V. E., & Katkova, A. L. (2020). "Stimulirovanie professional'nogo interesa studentov k prakticheskim zanyatiyam po informatike" [Stimulating students' professional interest in practical classes in computer science], *Mir nauki. Pedagogika i psihologiya*, t. 8, № 1, p. 24 (in Russian).
28. Bogacheva, N. V., Duyanov, O. P., Kolevatyh, E. P. et al. (2024). *Integraciya iskusstvennogo intellekta i biomeditsinskih nauk: sovremennye dostizheniya i perspektivy* [Integration of Artificial Intelligence and Biomedical Sciences: Current Achievements and Prospects], Professional'naya nauka, N. Novgorod, 127 p. (in Russian).
29. Antohina, Yu. A., & Tatarnikova, T. M. (2025). *Metody i algoritmy iskusstvennogo intellekta* [Methods and algorithms of artificial intelligence]: ucheb. dlya vuzov, Lan', St. Petersburg, 304 p. (in Russian).
30. Zolkin, A. L., & Munister, V. D. (2024). *Realizaciya principov organizacii i ispol'zovaniya sredstv mashinnogo obucheniya i iskusstvennogo intellekta v medicine* [Implementation of principles of organization and use of machine learning and artificial intelligence in medicine]: ucheb. posobie, ChUOO VO "Medicinskij universitet "Reaviz", Samara, 127 p. (in Russian).
31. Ignat'ev, S. A., Kuligin, A. V., Ignat'ev, A. A., & Shalunov, V. V. (2024). *Intellektual'nye sistemy i tekhnologii v medicinskom obrazovanii* [Intelligent systems and technologies in medical education]: ucheb. posobie, Obshchestvo s ogranichennoj otvetstvennost'yu Izdatel'stvo "KUBiK", Saratov, 160 p. (in Russian).
32. Mityakov, E. S., Shmeleva, A. G., & Ladynin, A. I. (2025). *Iskusstvennyj intellekt i mashinnoe obuchenie* [Artificial Intelligence and Machine Learning]: ucheb. posobie, Lan', St. Petersburg, 252 p. (in Russian).
33. Potapov, M. P., Pavlov, A. V., & Kotlovskij, M. Yu. (2024). *Iskusstvennyj intellekt v zdravooohranenii* [Artificial Intelligence in Healthcare]: ucheb. posobie, IPK "Indigo", Yaroslavl', 72 p. (in Russian).
34. *Obshchedostupnye naborы fajlov, neobhodimye dlya raboty v oblasti nauki o dannyh. UCI Heart Disease Data* [Publicly available datasets essential for data science work. UCI Heart Disease Data]. Available at: <https://www.kaggle.com/datasets> (in Russian).
35. *Mashinnoe obuchenie v Python scikit-learn* [Machine Learning in Python scikit-learn]. Available at: <https://scikit-learn.ru/> (in Russian).

Вклад авторов

С. А. Фирсова – постановка проблемы, реализация компьютерного эксперимента на языке Python, статистический анализ результатов.

Е. А. Рябухина – критический обзор и анализ литературы, формулировка заключения.

Contribution of the authors

S. A. Firsova – problem statement, implementation of computer experiment in Python, statistical analysis of results.

E. A. Ryabukhina – critical review and analysis of literature, formulation of conclusion.