

---

---

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

УДК 37.012, 004.891, 303.722

DOI: 10.17853/1994-5639-2022-1-163-190

## СОПРОВОЖДЕНИЕ ИНДИВИДУАЛЬНЫХ ОБРАЗОВАТЕЛЬНЫХ ТРАЕКТОРИЙ НА ОСНОВЕ КОНЦЕПЦИИ ОБЪЯСНИМОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

И. Г. Захарова<sup>1</sup>, М. С. Воробьева<sup>2</sup>, Ю. В. Боганюк<sup>3</sup>

Тюменский государственный университет, Тюмень, Россия.

E-mail: <sup>1</sup>i.g.zakharova@utmn.ru, <sup>2</sup>m.s.vorobeveva@utmn.ru, <sup>3</sup>y.v.boganyuk@utmn.ru

**Аннотация.** *Введение.* Получение профессионального образования в рамках индивидуальных образовательных траекторий (ИОТ) отвечает запросам как самих студентов, так и рынка труда в плане актуальности содержания, гибкости образовательного процесса и технологий обучения. Однако в условиях цифровизации сопровождение ИОТ, включая их планирование и последующее управление обучением, влечет появление новых требований к информационным системам, предназначенных для решения задач управления образовательным процессом вуза. Проблема настоящего исследования определяется существованием противоречия между закономерным для цифровизации интенсивным ростом объемов и разнообразия типов собираемых данных, которые можно и нужно использовать для сопровождения ИОТ, с одной стороны, и отсутствием в соответствующих информационных системах адекватных аналитических инструментов, с другой стороны.

*Цель статьи* – исследование и апробация методологии цифровизации сопровождения ИОТ, основанной на применении концепции объяснимого искусственного интеллекта для анализа данных цифрового следа студента, содержания документов, регламентирующих образовательный процесс, а также запросов рынка труда.

*Методология и методы исследования.* Теоретическую основу исследования составляет применение принципов объяснимого искусственного интеллекта к интерпретации данных об образовательном процессе и прогнозированию его результатов для выработки управляющих решений. Для предварительной обработки исходных документов используются методы интеллектуального анализа текстов на естественном языке. Для прогнозирования образовательных результатов применяются модели кластеризации, классификации и регрессии, созданные с помощью методов машинного обучения.

**Результаты.** Были разработаны и исследованы модели прогнозирования с последующим формированием рекомендаций для задач выбора образовательной программы абитуриентами, выбора элективной дисциплины, формирования команды для группового проекта и трудоустройства в соответствии с профессиональными компетенциями. Разработана компьютерная программа, которая автоматически формирует объективные и объяснимые рекомендации, основанные на экспертных знаниях и результатах прогнозирования. Алгоритм построения рекомендаций разделен на этапы и предусматривает вариативность в принятии решений.

**Научная новизна.** Предложена методология цифровизации сопровождения ИОТ, отвечающая принципам объяснимого искусственного интеллекта, когда по результатам анализа данных об образовательном процессе выполняется прогноз образовательных результатов и автоматически формируются персонализированные рекомендации. Разработанный подход подтвердил свою результативность при апробации на примере направлений подготовки бакалавриата и магистратуры в области компьютерных наук, информационных технологий и информационной безопасности.

**Практическая значимость.** Предварительный анализ значительных объемов исходных данных позволил получить объективную информацию о качестве данных, включая содержание и структуру документов, представленных в различных информационных системах вуза. На основании теоретических результатов исследования была разработана рекомендательная система. Она включила специальные сервисы для обучающихся, преподавателей, тьюторов и административных работников, предоставляющие наглядные и ориентированные на конкретного пользователя результаты прогнозов и рекомендации. Тестирование сервисов в Институте математики и компьютерных наук Тюменского государственного университета подтвердило целесообразность развития функциональных возможностей информационных систем вуза в направлении сбора и анализа данных цифрового следа студента и востребованность результатов этого анализа как субъектами образовательного процесса, так и представителями рынка труда.

**Ключевые слова:** индивидуальная образовательная траектория, цифровой след студента, интерпретация данных, прогнозирование, рекомендательная система, машинное обучение, объяснимый искусственный интеллект.

**Благодарности.** Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-37-51028 и Научно-технологического университета «Сириус». Авторы выражают искреннюю признательность анонимным рецензентам, сделавшим ряд ценных замечаний.

**Для цитирования:** Захарова И. Г., Воробьева М. С., Боганюк Ю. В. Сопровождение индивидуальных образовательных траекторий на основе концепции объяснимого искусственного интеллекта // Образование и наука. 2022. Т. 24, № 1. С. 163–190. DOI: 10.17853/1994-5639-2022-1-163-190

## SUPPORT OF INDIVIDUAL EDUCATIONAL TRAJECTORIES BASED ON THE CONCEPT OF EXPLAINABLE ARTIFICIAL INTELLIGENCE

I. G. Zakharova<sup>1</sup>, M. S. Vorobeva<sup>2</sup>, Yu. V. Boganyuk<sup>3</sup>

*University of Tyumen, Tyumen, Russia.*

*E-mail: <sup>1</sup>i.g.zakharova@utmn.ru, <sup>2</sup>m.s.vorobeva@utmn.ru, <sup>3</sup>y.v.boganyuk@utmn.ru*

**Abstract.** *Introduction.* Professional education in the context of individual educational trajectories (IET) meets the needs of both students themselves and the labour market due to the relevance of the content, flexibility of the educational process and learning technologies. However, in the context of digitalisation, IET support, including their planning and subsequent management of learning, entails the emergence of new requirements for information, analytical and methodological support of information systems designed to manage the educational process of the university. The problem of this study is determined by the contradiction between the intensive growth (natural for digitalisation) in the volume and variety of types of collected data, which can and should be used to support IET. In addition, there is also a lack of adequate analytical tools in educational information management systems.

*Aim.* The present research *aimed* to study and test the digitalisation methodology for IET support, based on the application of the concept of explainable artificial intelligence for analysing student digital footprint data, the content of documents regulating the educational process, as well as labour market demands.

*Research methodology and methods.* As a theoretical basis for the study, the authors relied on the principles of explainable artificial intelligence and their application to the interpretation of data from the educational process and the prediction of educational outcomes. The methods of intellectual analysis of texts in natural language were employed for preliminary processing of source documents. To predict educational outcomes, the authors used clustering, classification and regression models created through applying machine learning methods.

*Results.* The authors developed and studied predictive models with the subsequent formation of recommendations for the tasks of choosing an educational programme by applicants, choosing an elective discipline, forming a team for a group project and employment in accordance with professional competencies. The developed computer program automatically generates objective and explainable recommendations based on expert knowledge and predicting results. The algorithm for constructing recommendations is divided into stages and provides for variability in decision making.

*Scientific novelty.* The authors proposed a methodology for digital support of IET, corresponding to the principles of explainable artificial intelligence, i.e. machine learning models predict educational outcomes, and a special algorithm automatically generates personalised recommendations based on the results of the analysis of data on the educational process. The developed approach confirmed its effectiveness in testing on the example of bachelor's and master's degree programmes in the field of computer science, information technology and information security.

*Practical significance.* A preliminary analysis of significant volumes of initial data made it possible to obtain objective information about the data quality, including the content and structure of documents presented in various university information systems. Based on the the-

oretical results of the research, the authors developed a recommendation system. It included special services for students, teaching staff, tutors, and administrators, providing visual and user-oriented predictive results and recommendations. Testing of services at the Institute of Mathematics and Computer Science of University of Tyumen confirmed the feasibility of developing the functionality of the university information systems in the direction of collecting and analysing data from a student's digital footprint and the relevance of this analysis results both by subjects of the educational process and by the labour market.

**Keywords:** individual educational trajectory, student digital footprint, data interpretation, prediction, recommendation system, machine learning, explainable artificial intelligence.

**Acknowledgements.** The reported study was funded by the Russian Foundation for Basic Research (RFBR), project number 19-37-51028, and Sirius University of Science and Technology. The authors would like to express their sincere gratitude to the anonymous reviewers, who provided valuable comments.

**For citation:** Zakharova I. G., Vorobeva M. S., Boganyuk Yu. V. Support of individual educational trajectories based on the concept of explainable artificial intelligence. *The Education and Science Journal*. 2022; 24 (1): 163–190. DOI: 10.17853/1994-5639-2022-1-163-190

## Введение

В настоящее время вузы сталкиваются с множеством вызовов самого разного характера, но одними из наиболее актуальных и при этом тесно связанных являются цифровизация и индивидуализация образования. В контексте образовательного процесса цифровизация – это методология использования цифровых ресурсов для достижения вполне конкретных целей (качества, доступности, рентабельности, интенсивности и т. д.). В любом случае цифровизация предполагает пересмотр бизнес-процессов, связанных с самыми разнообразными задачами, решаемыми всеми участниками образовательного процесса – студентами, преподавателями, тьюторами, административными работниками. В основе этих изменений лежат преобразование образовательной среды вуза, совершенствование взаимодействия не только между субъектами образовательного процесса, но и с рынком труда.

Цифровизация высшего образования является необходимым условием реализации индивидуальных образовательных траекторий (ИОТ). Действительно, только благодаря цифровизации появляется возможность собирать, а затем и анализировать самые разнообразные данные об образовательном процессе, зачастую имеющие произвольную, необработанную форму. При этом основная цель анализа данных – это установление связей и закономерностей для выявления проблем как студентов, так и преподавателей, прогнозирование результатов и выработка рекомендаций всем участникам образовательного процесса. Без такого анализа невозможно качественно со-

проводить ИОТ для конкретного студента, оперативно реагируя на изменения и быстро принимая необходимые решения.

Цифровизация образования малопродуктивна без понимания того, что необходимо обеспечить, по сути, замкнутый цикл – от извлечения четкой информации из данных, продуцируемых обучающимися, до выработки понятных и адекватных ситуации рекомендаций, которые получают студенты, преподаватели, тьюторы и другие заинтересованные лица. Тем самым данные, продуцируемые в ходе образовательного процесса, превращаются в его двигатель. Разумеется, вечных двигателей не существует, поэтому использование данных может дать эффект, только если из них будут получены новые знания, использование которых может расширить возможности всех субъектов образовательного процесса, повысив его результативность. При этом такие знания можно извлечь только с опорой на большие объемы данных, отражающих актуальное содержание деятельности студентов и преподавателей. И в этом смысле толчком для получения таких данных послужил переход к онлайн-обучению в начале 2020 г., который был вызван вполне объективными обстоятельствами.

Необходимо отметить, что информационные системы, используемые в учреждениях высшего образования, обеспечивают сбор и хранение значительных объемов информации, часть которой представлена на сайтах вузов или доступна в корпоративной сети, а какая-то информация носит сугубо служебный или персональный характер. Данные отражают, как правило, распределение учебной нагрузки преподавателей, их научные интересы и достижения, содержание учебных планов и рабочих программ дисциплин, оценочные материалы и соответствующие результаты аттестации, тексты отчетов, рефератов и других работ студентов. Кроме того, могут сохраняться и менее формальные данные, непосредственно связанные с онлайн-обучением: переписка в чатах на внутренних форумах и платформах онлайн-обучения, аудио- и видеозаписи онлайн-лекций, семинаров, экзаменов, защит проектов. Безусловно, все эти данные очень важны для управления образовательным процессом, но используется только малая часть из них – для формирования традиционных отчетов и в лучшем случае в отдельных психолого-педагогических и социологических исследованиях.

Это вполне объяснимо: традиционные вузовские информационные системы разрабатывались как учетные, поэтому в них не предусмотрены возможности интеллектуального анализа для интерпретации данных и тем более для управления образовательным процессом в условиях индивидуализации. Но это только одна сторона медали. Недостаточно просто привлечь современные методы искусственного интеллекта (машинное обучение, искусственные нейронные сети). Их применение в сфере образования возможно только при условии обоснованности и объяснимости соответствующих

решений и рекомендаций, будь то предлагаемый студенту дополнительный курс или оценка достижений.

Налицо явное противоречие: внедрение технологий онлайн- и смешанного обучения, индивидуализация образования способствуют росту объема и видов данных, отражающих различные стороны образовательного процесса, в отличие от формальных обобщенных показателей, традиционно использующихся при принятии управленческих решений. Именно необходимость обоснованного и прозрачного сопровождения ИОТ не только придает этим данным особую ценность, но и требует развития методологии получения и понятного представления новых знаний, что определяет проблему настоящего исследования.

**Цель** настоящей работы заключается в исследовании и апробации методологии цифровизации сопровождения ИОТ на основе результатов интеллектуального анализа больших объемов данных из различных источников, которые отражают содержание и структуру образовательных программ и отдельных учебных курсов, текущие, промежуточные и итоговые образовательные результаты, а также актуальные требования рынка труда.

В соответствии с целью можно поставить следующие **исследовательские вопросы**:

1. Какие именно данные цифрового следа студента можно получить из различных информационных систем вуза и как их использовать для цифровизации сопровождения ИОТ?
2. Что должно послужить основанием для выбора моделей и методов искусственного интеллекта, предназначенных для построения прогнозов образовательных результатов с достаточной для практического использования точностью?
3. Каким образом можно построить персонализированную рекомендацию в ответ на конкретный вопрос сопровождения ИОТ, чтобы она в доступной форме отражала связь между данными цифрового следа, результатами прогнозирования и предлагаемыми вариантами решения?

В основе **гипотезы исследования** лежит предположение о возможности использования методологии объяснимого искусственного интеллекта для создания инструментов, позволяющих устранить объективно существующие разрывы между продуцированием данных участниками образовательного процесса, их интерпретацией, прогнозированием возможных последствий (образовательных результатов) и выработкой персональных объяснимых решений и рекомендаций, которые получают все заинтересованные лица.

**Ограничения** настоящего исследования определяются его эмпирической базой – тем, что мы сосредоточились на изучении вопросов сопровождения ИОТ, опираясь на данные цифрового следа студентов Института матема-

тики и компьютерных наук, запросы рынка труда в области информационных технологий и формальные описания соответствующей предметной области.

## **Обзор литературы**

### ***ИОТ и прогнозирование образовательных результатов***

В последние годы наряду с изучением общих проблем индивидуализации и персонализации образования многие исследователи уделяют особое внимание задачам проектирования и планирования ИОТ профессионального развития. Традиционно их решение опирается на мониторинг и прогноз изменения целевых показателей. Поэтому принципиально важной для нашего исследования представляется сама возможность такого прогнозирования в условиях открытой системы, обоснованная в работе Э. Ф. Зеера, Э. Э. Сыманюк [1]. При этом целевыми показателями для ИОТ, ориентирами профессионального развития выступают соответствующие компетенции [2].

Ключевая роль отводится прогнозированию образовательных результатов в условиях ИОТ и при различном понимании этого термина – глобальном или локальном. Так, рассматривая траектории, или пути (*pathways*), персонального развития студентов колледжей и университетов в глобальном понимании, R. Andrews, J. Li и M. F. Lovenheim отмечают необходимость выявления факторов, связанных с особенностями получения образования, которые позволяют прогнозировать его итоговые результаты вплоть до ожидаемых будущих доходов [3]. При этом анализ, проведенный в исследовании С. Наас и А. Hadjar [4], показывает, что не менее важную роль играет прогноз и при выборе модели изучения отдельного учебного курса конкретным обучающимся, то есть при самостоятельном планировании «локальной» ИОТ.

Однако несмотря на существование принципиальной возможности прогнозирования своего профессионального развития, обучающиеся (особенно студенты младших курсов) нуждаются в поддержке. На этом делается очень важный акцент в выводах исследования Э. Ф. Зеера, Н. Г. Церковниковой, В. С. Третьяковой: «Существует необходимость помощи представителям цифрового поколения в развитии способности к прогнозированию своего будущего, ..., в том числе профессионального» [5, с. 177]. Поскольку речь идет действительно о цифровом поколении, то и сопровождение ИОТ должно, в первую очередь, опираться на достижения цифровизации.

### ***Цифровизация в образовании и объяснимый искусственный интеллект***

Цифровизация управления образовательным процессом в целом, а не только сопровождение ИОТ имеет непосредственное отношение к общей

проблеме получения пользы от данных теми, кто эти данные продуцирует. В нашем случае это в первую очередь обучающиеся и данные их цифрового следа. То, что именно к этим «поставщикам» данных по результатам их анализа должны вернуться новые знания и идеи, помогающие им в принятии решений, подчеркивает D. Clow в своих работах [6], [7]. И если решение задач извлечения полезной информации и представления ее в удобном виде не вызывают трудностей в технологическом плане, то само использование результатов, полученных с помощью методов искусственного интеллекта, порождает определенную настороженность студентов и преподавателей. О важности учета психологических и этических аспектов использования данных, продуцируемых обучающимися (выполненные задания, виды активности, временной режим работы с электронными ресурсами и др.), свидетельствуют результаты исследований К. М. Jones [8, 9], Т. С. Pargman и С. McGrath [10]. При этом, как отмечается в работах Y. S. Tsai, O. Poquet, D. Gašević и др. [11], К. Kitto и S. Knight [12], не только обучающиеся, но и педагоги-практики могут не доверять рекомендациям систем, анализирующих данные образовательного процесса.

Эти проблемы не являются уникальными и свойственными именно системе образования – они возникают в любой сфере деятельности, где конечное решение принимается человеком. И если есть сложности с интерпретацией прогнозов и рекомендаций, пониманием механизма и правил, по которым построено решение, оно воспринимается с недоверием и отвергается. В лучшем же случае может возникнуть закономерное желание «заглянуть внутрь черного ящика», как отмечает D. Castelvechi [13, p. 21]. Именно поэтому внедрение методов искусственного интеллекта поставило два важнейших вопроса: доверие к предлагаемым решениям и их безопасность. Так, в одной из ключевых работ по этой теме Z. C. Lipton показывает их связь с общей проблемой интерпретируемости, для решения которой необходимо не только обладать необходимыми технологическими компетенциями, но и стремиться понять особенности восприятия конкретного пользователя [14]. Последний тезис получил свое развитие в исследовании Т. Miller [15], в котором показаны перспективы привлечения достижений философии, психологии, социологии для развития методологии объяснимого искусственного интеллекта – проекта DAPRA eXplainable Artificial Intelligence (XAI), представленного в статье D. Gunning и D. Aha [16].

Таким образом, анализ литературы подтверждает актуальность исследования проблемы цифровизации образования, связанной с сопровождением ИОТ на основе интеллектуального анализа данных цифрового следа студента с помощью методологии XAI.

## **Методология, материалы и методы**

В качестве объекта исследования выступает сопровождение ИОТ студентов вуза. Предметом исследования является обоснование и практическая реализация методологии цифровизации этого процесса на основе концепции объяснимого искусственного интеллекта применительно к интеллектуальному анализу больших объемов числовых и текстовых данных.

Исследование базировалось на анализе текстов 675 рабочих программ по дисциплинам базового цикла 14 образовательных программ по направлениям подготовки Института математики и компьютерных наук Тюменского государственного университета и 424 описаний элективных дисциплин. Также было проанализировано более 950 текстов курсовых и выпускных квалификационных работ, отчетов по научно-исследовательской работе, производственной и преддипломной практике, а также соответствующих отзывов – всего для 865 студентов данных направлений за 2017–2021 гг. Кроме того, в исследовании использовались данные об успеваемости и движении контингента по всем направлениям за указанный период.

Для сопоставления содержания образовательных программ (ОП) и результатов их освоения с требованиями рынка труда были проанализированы 937 текстов вакансий с сайта компании HeadHunter (hh.ru).

В качестве источников данных, отражающих цифровой след студента в контексте образовательного процесса, выступили следующие информационные системы: 1С: Университет (учебные планы, успеваемость, движение контингента), официальный портал университета<sup>1</sup> (тексты рабочих программ), Электронная библиотека ТюмГУ, корпоративное хранилище системы Microsoft Teams и электронные архивы кафедр (тексты курсовых и выпускных квалификационных работ, отчеты по практике, соответствующие отзывы, рецензии и справки о внедрении).

Применительно к конкретному студенту перечисленные выше данные отражают его цифровой след. В ряде документов присутствуют персональные данные. Их использование на исследовательских этапах не требуется. Поэтому вся информация такого рода предварительно была извлечена из текстов и обезличена. То есть упоминаемые фамилия, имя, отчество студентов и преподавателей заменены условными идентификаторами. При этом для получения индивидуальных рекомендаций был предусмотрен доступ через личный кабинет.

Отобранные документы изначально были представлены в различных форматах. Такие данные, как учебные планы, успеваемость, движение кон-

---

<sup>1</sup> Об описании образовательной программы. Режим доступа: <https://www.utmn.ru/sveden/education/eduOp> (дата обращения: 06.12.2021).

тингента, имеющие формат таблиц Excel с фиксированной структурой, не требовали особых преобразований. В то же время в числе прочих категорий были и сканированные документы (отзывы, справки, направления на практику и т. п.), для которых потребовалось приведение к текстовому формату. Далее проводилась автоматическая предварительная обработка данных, зависящая от вида документа. Так, табличные документы проверялись на предмет наличия недопустимых и пропущенных числовых значений. Для текстовых документов проводились автоматические преобразования, традиционные для так называемой обработки текстов на естественном языке: удаление часто встречающихся слов, не несущих смысловой нагрузки (стоп-слов), приведение слов к словарной форме. Затем выполнялась разметка документов в соответствии с их структурой и извлечение метаданных – информации, которая является определяющей для последующего анализа данной категории (табл. 1).

Таблица 1

Разметка документов и метаданные

Table 1

Document markup and metadata

Категория документа / Document category	Элементы структуры / Elements of the structure	Метаданные / Metadata
Рабочая программа / Curriculum	Аннотация, образовательный результат, содержание по темам, вопросы к зачету или экзамену, список литературы / Abstract, educational result, content by topic, questions for the test or exam, references	Направление подготовки, профиль/специализация, название дисциплины, идентификатор(ы) автора, дата создания / Field of study, profile/specialisation, discipline name, author ID(s), date of creation
Курсовая работа, выпускная квалификационная работа, отчет по НИР, реферат/эссе / Term paper, final qualifying work, research report, abstract/essay	Аннотация, введение, содержание по главам, заключение, список литературы, программный код (при наличии) / Abstract, introduction, chapter by chapter, introduction, references, programme code (if available)	Идентификатор(ы) исполнителя, идентификатор руководителя, тема, наличие сведений о внедрении/публикации, дата создания / The ID(s) of the executive, the ID of the director, the subject, the availability of information about the implementation/publication, the date of the building

Отчет по практике / Practice Report	Введение, содержание по главам, заключение, список литературы, программный код (при наличии) / Introduction, contents by chapters, conclusion, references, programme code (if available)	Идентификатор(ы) исполнителя, идентификатор руководителя, база практики, наличие сведений о внедрении/публикации, дата создания / The ID(s) of the executor, the ID of the supervisor, the practice base, the availability of information about the implementation/publication, the date of the building
--	--	--

Важнейшими характеристиками для всех приведенных выше разновидностей текстов выступают ключевые слова, связанные с описанием профессиональных компетенций. Поэтому по аналогии с тем, как это было реализовано нами ранее [17], на основании ключевых слов, выделенных из всей совокупности текстов, был сформирован терминологический словарь. В него вошли ключевые слова на русском языке (алгоритм, база данных, компьютерная графика, машинное обучение и т. д.), а также оригинальные названия профессиональных терминов, в которых могут использоваться цифры, латинские буквы и другие символы (например, C#, Python 3.8, C/C++ и др.). Для элементов словаря была установлена связь с документами, что позволило в последующем выполнять быстрый поиск по ключевым словам.

Для достижения конечной цели – построения системы рекомендаций – последовательно использовалась методология ХАИ, опирающаяся на методы математической статистики и машинного обучения. В частности, применялись методы описательной статистики, методы регрессионного и кластерного анализа, методы классификации и построения рекомендательных систем, отвечающие принципам прозрачности и интерпретируемости [16]. Количественные методы сочетались с использованием качественных, принципиально важных для концепции объяснимого искусственного интеллекта, а именно опросов студентов и профессиональных экспертов. Эксперты оценивали способы агрегации и визуализации данных, облегчающие их интерпретацию, а также прозрачность и объяснимость построенных прогнозных моделей. К оценке структуры и содержания рекомендаций были привлечены 53 студента. В качестве экспертов выступили руководители ОП, преподаватели, тьюторы Института математики и компьютерных наук, представители рынка труда (в том числе руководители практических ИТ-проектов, в которых непосредственно участвуют студенты института), всего 14 человек.

Разработка информационного и программного обеспечения системы рекомендаций основана на использовании свободно распространяемых инструментальных средств и специализированных библиотек для интеллектуального анализа данных Natural Language Tool Kit [18] и SciKit Learn [19].

## **Результаты исследования**

В рамках сформулированной гипотезы мы исследовали возможность создания непрерывной цепочки, включающей получение и интерпретацию данных, прогнозирование образовательных результатов и выработку персональных рекомендаций конкретному субъекту образовательного процесса.

В частности, в контексте цифровизации сопровождения ИОТ речь идет о рекомендациях, отличающихся по своему характеру. Например, абитуриента интересует выбор ОП, студенту нужна подсказка по выбору элективных дисциплин, дополнительных МООК, профиля подготовки, научного руководителя или команды для группового проекта и т. д. Рекомендации для преподавателей и тьюторов в первую очередь направлены на решение оперативных задач управления ИОТ. Но при этом для тех же преподавателей и руководителей ОП они касаются как актуализации содержания курсов и ОП в соответствии с текущими и перспективными требованиями рынка труда, так и согласования между собой отдельных дисциплин и формируемых компетенций. Наконец, представители рынка труда заинтересованы в точной информации не только о том, какие именно выпускники соответствуют требованиям конкретных вакансий, но и о том, студенты каких направлений и каких курсов могут составить команду под определенный проект.

Помимо формирования эмпирической базы исследования, включая сбор и предварительный анализ данных из различных информационных систем вуза, а также текстов вакансий, исследование включило следующие основные этапы:

1. Построение моделей, позволяющих прогнозировать особенности образовательного процесса в рамках ИОТ для определенных студентов.
2. Создание системы рекомендаций для принятия решений в процессе сопровождения ИОТ.

Содержание этапов представлено на примере решения конкретной задачи, а именно планирования ИОТ, начиная с выбора абитуриентом образовательной программы и заканчивая трудоустройством в соответствии с достигнутым уровнем профессиональной подготовки.

### ***Построение прогнозных моделей***

Мы исходили из того, что результаты прогнозирования, полученные с помощью методов, которые не поддаются прямой интерпретации, ско-

рее всего, будут восприниматься с недоверием. Поэтому круг возможных моделей изначально был ограничен наиболее понятными по своей сути и подходящими для представления результатов в простой и наглядной форме.

Итак, для рассматриваемой задачи планирования ИОТ были построены следующие прогнозные модели:

«Образовательная программа» – прогнозирует возможность освоения ОП, отвечающей запросу абитуриента.

«Элективная дисциплина» – подбирает дисциплины, соответствующие запросу и доступные для освоения обучающимся на конкретной ОП.

«Команда» – подбирает исполнителей-обучающихся и руководителя-преподавателя для выполнения группового проекта.

«Трудоустройство» – планирует ИОТ для профессионального развития обучающегося в соответствии с актуальными требованиями рынка труда.

Остановимся более подробно на особенностях построенных решений на примере моделей «Элективная дисциплина» и «Команда» для направлений подготовки Института математики и компьютерных наук Тюменского государственного университета. При построении моделей учитывались особенности реализуемых образовательных программ и ИОТ. А именно: при обязательных учебных курсах по прикладной математике, компьютерным наукам, информационным технологиям и т. п. у обучающихся есть возможность выбора элективных дисциплин из различных предметных областей – как профильных, усиливающих профессиональную подготовку, так и направленных на расширение кругозора и развитие общекультурных компетенций.

*Примечание. В рамках данного исследования мы сосредоточились на профессиональных компетенциях, поскольку их содержание в большинстве случаев легко формализуется путем сведения к перечню конкретных методов и/или технологий и уровням их освоения. Поэтому в дальнейшем изложении мы будем придерживаться этого ограничения.*

### **Элективная дисциплина**

Планирование ИОТ в части выбора элективных дисциплин опирается не только на пожелания студента, но и на возможность освоения соответствующего материала. Поэтому для нахождения оптимального решения был разработан специальный алгоритм по принципу дерева решений. На первом этапе на основании заданного студентом (или тьютором) набора ключевых слов автоматически определяются наиболее релевантные дисциплины. Поиск можно осуществлять на основании полного текста рабочей программы, аннотации, образовательных результатов, содержания по темам и/или вопросов к зачету/экзамену. Затем выполняется сравнение текста программы (или отдельных компонентов по выбору студента) с соответствующими

текстами рабочих программ для учебного плана, по которому обучается студент. Это делается для информирования обучающихся о том, что «похожая» дисциплина (одна или несколько) есть в базовой части учебного плана. То есть она уже изучалась или будет изучаться как обязательная. Студент может принять альтернативное решение – изменить свой запрос или продолжить выбор дисциплины из предложенных.

Кроме поиска на основе ключевых слов, также была построена модель прогнозирования выбора дисциплины на основе коллаборативной фильтрации, успешно примененной для подобной задачи в работе M. Salehi, I. N. Kamalabadi, M. B. G. Ghouschi [20]. Мы опирались на выявление «похожих» ИОТ на основе набора признаков, характеризующих обучающихся. Для этого была выполнена кластеризация ИОТ студентов. Учитывались признаки, как имеющие общий характер (ОП, курс), так и отражающие предшествующие предпочтения студентов при выборе элективных дисциплин. Прогноз осуществлялся после отнесения студента (на основе ИОТ в предшествующий период) к определенному кластеру. После этого в кластере определялись наиболее близкие траектории, по которым и выбирались дисциплины.

### **Команда**

Образовательный процесс на основе ИОТ подразумевает выполнение разнообразных учебных, учебно-исследовательских, научно-исследовательских и инновационных проектов, для которых необходимо не только сформировать команду исполнителей и выбрать руководителя (куратора) из числа преподавателей. Важно еще и спрогнозировать возможность выполнения такого проекта конкретной командой. Подбор команды и построение прогноза подразумевают определенную последовательность шагов.

Изначально инициатор (в первую очередь, обучающийся) может указать только нужные для выполнения проекта компетенции с уточнением уровня владения (имею представление, имею опыт использования, уверенно владею). Поскольку в настоящем исследовании мы сосредоточились на ИТ-проектах, то предполагалось [17], [21], что содержание компетенции включает название конкретной технологии/метода (например, название языка программирования или метода машинного обучения), которой должен владеть участник на уровне не ниже заданного. По этим исходным данным выполняется поиск по текстам рабочих программ всех направлений института. В случае нахождения соответствия (присутствия названия в содержании курса) для студентов, освоивших данную дисциплину, подразумевается уровень «имею представление». Для поиска кандидатов в команду, имеющих уровень «имею опыт использования», выполняется поиск по текстам отчетов по практике, НИР и курсовых работ. Наконец, уверенное владение подтверждается положительными результатами поиска соответ-

ствия в текстах справок о внедрении разработки и характеристик с места прохождения практики.

Дополнительно для всех найденных кандидатов определяется наличие опыта участия в групповых проектах. Для этого из текстов отчетов по практике, учебных проектов по дисциплинам, курсовых работ автоматически извлекаются идентификаторы студентов. Это позволяет оценить не только количество групповых проектов, в которых каждый из них участвовал, но и, возможно, уже сложившиеся коллективы, совместно решившие не одну задачу.

На основе полученных результатов составляются варианты команд с заданным количеством участников, обладающих необходимыми компетенциями. При этом каждый вариант команды характеризуется долей участников, имеющих опыт выполнения групповых проектов.

Кроме того, для каждой команды определяются значения бинарных (да/нет) признаков:

- 1) все учатся на одной ОП;
- 2) все являются студентами одного курса.

Далее подбирается один или несколько потенциальных руководителей, которые в той или иной степени обладают следующим опытом:

- 1) руководство групповыми проектами;
- 2) руководство учебными проектами, курсовыми работами, связанными с применением технологий, которые предполагается использовать в данном проекте;
- 3) чтение курсов, в которых в том или ином объеме изучаются соответствующие технологии.

Необходимая информация извлекается из текстов работ студентов и рабочих программ дисциплин. В итоге мы получаем количественные признаки, характеризующие потенциального руководителя: количество групповых проектов в целом, количество проектов и дисциплин в разрезе компетенций.

Кроме того, учитываются два бинарных (да/нет) признака, отражающие степень «знакомства» руководителя и команды:

- 1) был ли ранее преподаватель руководителем какого-либо проекта по меньшей мере у одного из предполагаемых членов команды;
- 2) читал ли преподаватель один или несколько курсов, предусматривающих формирование требуемых компетенций, кому-либо из участников проекта.

Для прогнозирования вероятности успешной работы команды была построена модель бинарной классификации Байеса, положительно зарекомендовавшая себя при решении подобных задач, как отмечено в обзорной

статье O. Zawacki-Richter, V. I. Marín, M. Bond, F. Gouverneur [22]. Эта модель дает не только однозначный ответ (рассматривается бинарная классификация – да/нет), но и дополнительно позволяет определить вероятность положительного исхода. Для ее обучения использовались данные по завершенным групповым проектам, выполненным в рамках курсовых и выпускных квалификационных работ, НИР, практик или при изучении отдельных дисциплин. Успешным считался проект, защищенный в срок с оценкой не ниже «хорошо». Для получения прогноза по каждому варианту команды (включая варианты выбора руководителя) использовались перечисленные выше признаки.

### **Создание системы рекомендаций**

Выработка рекомендаций, связанных с тем или иным прогнозом, – это закономерный завершающий шаг конвейера интеллектуального анализа, когда новые знания возвращаются к «поставщикам» анализируемых исходных данных: студентам, педагогам, представителям рынка труда. При этом мы не отождествляем такие рекомендации с традиционной обратной связью систем электронного обучения. В указанных системах изначально предусматриваются возможные траектории освоения конкретного модуля или дисциплины и обратная связь полностью детерминирована. Это могут быть не только рекомендации обучающимся по изучению определенного материала в случае неверного ответа на вопрос теста, но и информация для преподавателя о том, сколько времени студенты тратят на выполнение тех или иных заданий и т. д. В нашем случае требуется построить понятные рекомендации, в которых необходимо наглядно представить полученный прогноз и дать толкование тем факторам, которые определяют предсказанный результат.

Далее будут показаны примеры рекомендаций, которые автоматически формируются разработанными нами программами для конкретной ситуации в соответствии с описанными выше прогнозными моделями.

### **Элективная дисциплина**

Исходными данными для формирования рекомендаций, которые помогают студенту выбрать элективную дисциплину, служат в первую очередь ключевые слова. При этом мы предполагаем, что студент или тьютор, который должен помочь с выбором, имеет представление о структуре рабочих программ. Поэтому прогнозирование наиболее релевантных дисциплин может выполняться на основе сопоставления ключевых слов со всем текстом или отдельными разделами программ. Соответственно, и шаблоны автоматически формируемых рекомендаций имеют свои отличия.

Результатом прогноза по полному тексту или содержанию тем будут все отвечающие запросу дисциплины, которые будут включены в рекомендацию по шаблону:

«Ключевым словам <перечень ключевых слов> отвечают дисциплины <названия дисциплин>, ранжированные в порядке убывания релевантности».

*Примечание.* Здесь и далее в шаблонах с помощью угловых скобок (“<...>”) выделяются элементы рекомендации, которые заполняются автоматически с помощью соответствующей программы.

Если, с точки зрения студента или тьютора, было найдено слишком большое число дисциплин, то для уточнения прогноза можно дополнительно проанализировать тексты, в которых представлены планируемые образовательные результаты, а также контрольные задания и вопросы к зачету или экзамену. Соответственно, уточняются и рекомендации:

«С учетом раздела программы <название раздела> запросу в наибольшей степени соответствует дисциплина <название дисциплины>».

Указав свою ОП, студент также может получить информацию о том, что и в его учебном плане предусмотрены базовые дисциплины, отвечающие исходному запросу. В рекомендации это будет отражено в следующем виде:

«Запросу также соответствует (ют) дисциплина (ы) базовой части учебного плана <название дисциплины, семестр>, <название дисциплины, семестр>, ...».

В качестве примера приведем расширенную рекомендацию по запросу обучающегося на направлении «Математическое обеспечение и администрирование информационных систем» на основании заданных ключевых слов «веб-технологии, программный код»:

«Ключевым словам <веб-технологии, программный код> отвечают дисциплины <Разработка интерфейсов web-приложений, Современные технологии UX и web-дизайна, Основы программирования>, ранжированные в порядке убывания релевантности. С учетом раздела программы <контрольные задания> запросу в наибольшей степени соответствует дисциплина <Разработка интерфейсов web-приложений>. Запросу также соответствует (ют) дисциплина (ы) базовой части учебного плана <Web-технологии, 4 семестр>».

Если для прогноза выбрана модель, основанная на коллаборативной фильтрации («похожие» ИОТ), то рекомендация основывается на принадлежности ИОТ обучающегося к определенному кластеру, объединяющему выявленные ранее «похожие» траектории, где студенты отдавали явное предпочтение определенным элективным дисциплинам. В этом случае просто перечисляются названия определенного количества дисциплин, наиболее популярных для спрогнозированного кластера и ранее не выбиравшихся данным студентом.

### **Команда**

Исходя из представленного выше способа прогнозирования состава команд для выполнения группового проекта, мы разработали шаблоны для генерирования рекомендаций.

В простейшем случае в качестве исходных данных выступает набор пар «требуемая компетенция, уровень владения». Мы предположили, что уровню «имею представление» отвечают все студенты, освоившие соответствующие дисциплины. Поэтому предусмотрены два шаблона рекомендации:

1. «По результатам анализа текстов работ студентов требованию *<требуемая компетенция, уровень владения>* соответствуют студенты *<список идентификаторов студентов>*».

2. «По результатам анализа текстов рабочих программ требованию *<требуемая компетенция, имею представление>* соответствуют студенты направлений *<список “ОП, курс”>*».

Например, для запроса, включающего пары «мобильное приложение разработка, уверенно владею», «криптография, имею представление», была автоматически построена рекомендация:

«По результатам анализа текстов рабочих программ требованию *<криптография, имею представление>* соответствуют студенты направлений *<Информационная безопасность, 2 курс, Компьютерная безопасность, 2 курс>*. По результатам анализа текстов работ студентов требованию *<мобильное приложение разработка, уверенно владею>* соответствуют студенты *<02.03.03\_2018\_023, 09.03.01\_2017\_019, 10.03.01\_2018\_012>*».

*Примечание. По идентификатору студента автоматически определяется ОП (первые 6 цифр – код направления подготовки), а по году (следующие 4 цифры) – курс.*

Для принятия решения с учетом опыта реализации групповых проектов в рекомендацию для предварительно отобранных кандидатов добавляется соответствующая информация, полученная в результате извлечения данных об исполнителях из текстов работ студентов. Для этого используется шаблон:

«Имеют опыт участия в групповых проектах *<список “идентификатор студента-число групповых проектов”>*».

Для рассмотренного примера дополнение к рекомендации примет вид:

«Имеют опыт участия в групповых проектах *<10.03.01\_2018\_012 – 1, 10.05.01\_2017\_003 – 3, 10.05.01\_2017\_011 – 3>*».

Можно также использовать естественное ограничение: команда формируется только в том случае, если инициатор-студент обладает хотя бы одной из нужных компетенций и имеет опыт выполнения групповых проектов. В рассмотренном примере данному условию отвечают два человека.

Это студент с идентификатором 10.03.01\_2018\_012, который в начале 2021 года учится на 3-м курсе направления «Информационная безопасность», поэтому имеет представление о криптографии из освоенных дисциплин базовой части учебного плана. Кроме того, он выполнял проект (ы), связанный (ые) с разработкой мобильных приложений, и участвовал в одном групповом проекте. Также инициатором может быть студент с идентификатором 10.05.01\_2017\_011 (4-й курс специальности «Компьютерная безопасность»), имеющий представление о криптографии и участвовавший в трех групповых проектах.

Далее автоматически генерируется текст рекомендации по выбору руководителя. Мы полагали, что для прозрачности предлагаемого решения нужно показать, на чем оно основано. Шаблон имеет следующую структуру:

«Преподаватель <идентификатор преподавателя> обладает необходимыми компетенциями: 1) руководство групповыми проектами – <число проектов>; 2) руководство учебными проектами, курсовыми работами, связанными с применением <список “компетенция – количество”>; 3) чтение курсов, в которых изучаются <список “компетенция – количество”>».

Продолжая пример, приведем рекомендации, сформированные автоматически на основе прогнозирования потенциальных руководителей (найденно два потенциальных руководителя):

«Преподаватель <department02\_03> обладает необходимыми компетенциями: 1) руководство групповыми проектами <4>; 2) руководство учебными проектами, курсовыми работами, связанными с применением <криптография – 16>; 3) чтение курсов, в которых изучаются <криптография – 2>».

«Преподаватель <department04\_07> обладает необходимыми компетенциями: 1) руководство групповыми проектами <3>; 2) руководство учебными проектами, курсовыми работами, связанными с применением <криптография – 1, мобильное приложение разработка – 5>».

Если полученных рекомендаций все еще недостаточно для принятия решения, автоматически формируются все допустимые варианты «команда + руководитель». Для этого указывается требуемое количество участников по каждой компетенции (кроме инициатора). Затем для каждого варианта определяются дополнительные признаки: доля участников, имеющих опыт выполнения групповых проектов; все учатся на одной ОП (да/нет); все являются студентами одного курса (да/нет); преподаватель был руководителем проекта у кого-либо из членов команды (да/нет); преподаватель читал хоть один курс, предусматривающий формирование требуемых компетенций, кому-либо из участников проекта (да/нет).

По всей совокупности признаков с помощью соответствующей модели (см. выше) строится прогноз: определяется вероятность успешного выполне-

ния проекта конкретной командой. После этого автоматически формируется рекомендация по оптимальному составу команды, для которой соответствующая вероятность имеет максимальное значение.

Для рассматриваемого примера с одним участником по каждой компетенции (всего 3 вместе с инициатором) при условии, что инициатором является студент 4 курса специальности «Компьютерная безопасность» (идентификатор 10.05.01\_2017\_011), автоматически формируется рекомендация следующего вида:

«Проект может быть выполнен успешно с вероятностью <71%> командой в составе: инициатор <10.05.01\_2017\_011>, участники <02.03.03\_2018\_023, 10.03.01\_2018\_012>, руководитель <department04\_07>».

Далее можно получить аналогичные прогнозные оценки для других вариантов. Это очень важно, поскольку итоговое решение должно приниматься всей командой с учетом множества факторов, которые не отражены в данных цифрового следа, но при этом могут играть решающую роль. Это и личные отношения, и заинтересованность в совместной деятельности и т. д.

## **Обсуждение результатов**

Для каждого из определенных нами этапов исследования можно выделить основные результаты, заслуживающие отдельного обсуждения.

В первую очередь необходимо остановиться на сборе, организации хранения и предварительном анализе данных. Проведенное исследование подтвердило существующие оценки сложности и относительной продолжительности этого этапа, представленные в обзорной статье С. Schröer, F. Kruse, J. M. Gómez [23]. Однако более важным явилось понимание того, что априори сложно полностью предусмотреть то, какую именно информацию нужно извлечь из «сырых» данных, в первую очередь текстовых документов, в какую структуру объединить для последующего использования. Поэтому необходимо обеспечить их хранение после минимальной обработки (удаления дубликатов, некачественных сканированных документов и др.). Для построения прогнозных моделей и объяснимых рекомендаций необходимо формировать и сохранять терминологические словари и тезаурусы, описывающие предметную область. Как общее требование это выделяет Н. М. Боргест [24], а конкретизируют для рекомендательных систем в сфере e-learning G. George и F. M. Lal [25]. В нашем случае это компьютерные науки, информационные технологии и информационная безопасность. Наконец, сами модели и рекомендации тоже являются данными для последующего анализа, но уже не «сырыми», а «умными», smart-данными.

Все это привело к необходимости разработки особого информацион-

ного хранилища. Оно было успешно использовано как основа информационно-аналитического обеспечения рекомендательной системы, учитывающей специфику направлений подготовки Института математики и компьютерных наук. Изначально хранилище было ориентировано на решение исследовательских задач. Поэтому для масштабирования системы при существенном увеличении объемов обрабатываемых данных нужно не только изучить специфику других ОП, но и модифицировать технологии хранения путем перехода от локальных баз данных к виртуальным, что, в свою очередь, как отмечают S. Guo, D. Zeng, S. Dong [26], напрямую связано с перспективами Education 4.0.

Результаты, полученные в процессе построения прогнозных моделей и алгоритмов, показали, что они обладают достаточной точностью для использования на практике. В рамках данного исследования мы не проводили сопоставительный анализ решений, полученных с помощью различных моделей и методов. Это связано с тем, что использование принципов ХАИ ограничивает выбор: модели машинного обучения должны быть объяснимы, а используемые для их построения признаки понятны целевой аудитории, что обосновано в работе A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser и др. [27]. В то же время именно благодаря следованию этим принципам мы получили решения, которые носят прозрачный, легко интерпретируемый характер. Мы ограничились моделями классификации Байеса и дерева решений, которые показали точность более 70 %.

Не все предполагаемые решения удалось реализовать в полной мере. Так, нами был разработан алгоритм для определения готовности обучающегося к освоению той или иной дисциплины, которая определяется на основе сопоставления текста входных требований из рабочей программы элективной дисциплины и образовательных результатов ранее изученных дисциплин. Однако при анализе текстов программ выяснилось, что формулировки образовательных результатов не всегда позволяют решить эту задачу с необходимой точностью. Также мы использовали для прогнозирования выбора элективной дисциплины модель коллаборативной фильтрации, основанной на похожести ИОТ. Но надо учесть, что для первокурсников этот подход непригоден, поскольку отсутствует история выбора элективных дисциплин. Поэтому здесь необходимо привлекать дополнительные признаки, в том числе отражающие отзывы студентов о прослушанных курсах (например, ранги или рейтинги), выявлять целевые установки, учитывать динамические показатели и уже по ним определять «похожих» студентов, как показали для систем e-learning W. Zho и W. Han [28].

Безусловно, увеличение количества признаков и привлечение моделей искусственных нейронных сетей в целом гораздо перспективнее с точки зре-

ния повышения точности прогнозирования. Однако проблема объяснения решений, полученных с помощью таких моделей, не менее сложна, чем само построение решения, на чем делают акцент в своих исследованиях Z. C. Lipton [14], T. Miller [15]. Представляется перспективным развивать исследования в этом направлении, и тогда оперативное управление ИОТ будет основано на прогнозе образовательных результатов по всей совокупности структурированных и неструктурированных данных цифрового следа студента.

Основные результаты, связанные с формированием рекомендаций, проиллюстрированы на примере задач по принятию решений в процессе сопровождения ИОТ. В основном мы сосредоточились на рекомендациях, адресованных обучающимся, поскольку в условиях ИОТ предполагается, что именно студент должен проявлять инициативу и принимать ответственные решения по планированию своей траектории. К сожалению, не все обучающиеся готовы это делать самостоятельно, им требуется соответствующая поддержка в форме рекомендаций. Хочется подчеркнуть, что логика их формирования носит поэтапный и вариативный характер. Поэтому при принятии решения остается свобода выбора.

Формирование текстов рекомендаций осуществляется с помощью специальной программы, заполняющей шаблон по определенному алгоритму. Для разработки соответствующих шаблонов были привлечены те, кому рекомендации и адресованы: студенты, тьюторы, преподаватели. Мы не останавливались подробно на предусмотренных возможностях уточнения рекомендаций. В дополнение к этим текстам благодаря организации хранения данных можно получить связанную информацию – как в виде «сырых» данных (например, текст отчета по конкретному проекту), так и агрегированные отчеты и графики (например, частотность упоминания определенных ключевых слов в разрезе дисциплин и ОП).

## **Заключение**

В настоящем исследовании мы показали, как можно извлечь реальную пользу из данных цифрового следа обучающихся, на примере автоматического построения рекомендаций, помогающих планировать ИОТ, и получили ответы на поставленные исследовательские вопросы:

1. Цифровизация образования в целом и широкое использование систем поддержки онлайн-обучения в частности открывают доступ к огромным объемам данных, отражающих активность студента в образовательной среде, через различные числовые признаки, видео- и аудиозаписи, а также текстовые документы. Исходя из этических принципов и требований по защите персональных данных, мы сосредоточили основное внимание на обезличен-

ных текстовых документах, тем более что, в отличие от числовых данных (в первую очередь успеваемости), вопросы их анализа и последующего использования результатов в контексте образования исследованы не в полной мере. Мы опирались на данные, как регламентирующие образовательный процесс (общие описания образовательных программ и рабочие программы дисциплин), так и ориентирующие на достижение определенных целей профессионального развития (профессиональные стандарты, запросы рынка труда), а также тексты, подготовленные студентами в ходе выполнения различных проектов. В результате из текстов были извлечены формализованные характеристики, необходимые для последующего планирования ИОТ.

2. В основу разработки методов и алгоритмов анализа данных легла методология ХАИ – объяснимого искусственного интеллекта, которая помогла определиться с выбором конкретных моделей машинного обучения для прогнозирования образовательных результатов. Мы показали, что при сохранении достаточной точности этих прогнозов их легко интерпретировать и использовать для формирования персональных рекомендаций, учитывающих особенности целевой аудитории, в первую очередь обучающихся.

3. Нами разработана рекомендательная система, обеспечивающая автоматическую генерацию текстов персонализированных рекомендаций. Алгоритм формирования рекомендации предусматривает использование различного количества признаков-предикторов, от которых зависят прогноз и соответствующие варианты рекомендуемого решения. При этом все данные, на которых основана рекомендация, отображаются в тексте, обеспечивая ее прозрачность.

Нельзя не отметить определенные границы применимости полученных результатов. Мы опирались на данные особой предметной области (компьютерные науки, информационные технологии и информационная безопасность), которые достаточно органично поддаются формализации. Поэтому описание профессиональных компетенций (не затрагивая так называемых «мягких» навыков, *soft skills*) можно представить, опираясь на названия методов, технологий, языков программирования и т. п. В то же время в контексте требований к специалистам для цифровой экономики результаты настоящего исследования можно распространить на планирование ИОТ развития цифровой компетентности студентов самых различных направлений подготовки.

Проведенное исследование заложило определенную основу для своего развития в разных направлениях: адаптация результатов к другим предметным областям, уровням и формам образования; диагностика *soft skills* и выработка рекомендаций по их развитию; масштабирование при увеличении числа источников данных и их объемов.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Зеер Э. Ф., Сыманюк Э. Э. Индивидуальные образовательные траектории в системе непрерывного образования // Педагогическое образование в России. 2014. № 3. С. 74–82.
2. Зеер Э. Ф., Заводчиков Д. П., Зиннатова М. В., Лебедева Е. В. Индивидуальная образовательная траектория как установка субъекта в системе непрерывного образования // Научный диалог. 2017. № 1. С. 266–279.
3. Andrews R., Li J., Lovenheim M. F. Heterogeneous paths through college: detailed patterns and relationships with graduation and earnings // Economics of Education Review. 2014. № 42. P. 93–108. DOI: 10.1016/j.econedurev.2014.07.002
4. Haas C., Hadjar A. Students' trajectories through higher education: a review of quantitative research // Higher Education. 2020. № 79 (6). P. 1099–1118. DOI: 10.1007/s10734-019-00458-5
5. Зеер Э. Ф., Церковникова Н. Г., Третьякова В. С. Цифровое поколение в контексте прогнозирования профессионального будущего // Образование и наука. 2021. Т. 23, № 6. С. 153–184. DOI: 10.17853/1994-5639-2021-6-153-184
6. Clow D. The learning analytics cycle: closing the loop effectively // Proceedings of the 2nd International Conference on Learning Analytics and Knowledge – LAK'12. 29 April – 2 May 2012. Vancouver, British Columbia, Canada. New York: Association for Computing Machinery, 2012. P. 134–138.
7. Clow D. An overview of learning analytics // Teaching in Higher Education. 2013. № 18 (6). P. 683–695. DOI: 10.1080/13562517.2013.827653
8. Jones K. M. Advising the whole student: eAdvising analytics and the contextual suppression of advisor values // Education and Information Technologies. 2019. № 24 (1). P. 437–458. DOI: 10.1007/s10639-018-9781-8
9. Jones K. M. Just because you can doesn't mean you should: practitioner perceptions of learning analytics ethics // Libraries and the Academy. 2019. № 19 (3). P. 407–428. DOI: 10.1353/pla.2019.0025
10. Pargman T. C., McGrath C. Mapping the terrain of ethics in learning analytics: A systematic literature review of empirical research // Journal of Learning Analytics. 2021. № 1. P. 1–17. DOI: 10.18608/jla.2021.1
11. Tsai Y. S., Poquet O., Gašević D., Dawson S., Pardo A. Complexity leadership in learning analytics: Drivers, challenges, and opportunities // British Journal of Educational Technology. 2019. № 50 (6). P. 2839–2854. DOI: 10.1111/bjet.12846
12. Kitto K., Knight S. Practical ethics for building learning analytics // British Journal of Educational Technology 2019. № 50 (6). P. 2855–2870. DOI: 10.1111/bjet.12868
13. Castelveccchi D. Can we open the black box of AI? // Nature News. 2016. № 538 (7623). P. 20–23. DOI: 10.1038/538020a
14. Lipton Z. C. The mythos of model interpretability // Communications of the ACM. 2018. № 61 (10). P. 36–43. DOI: 10.1145/3233231
15. Miller T. Explanation in artificial intelligence: Insights from the social sciences // Artificial Intelligence. 2019. № 267. P. 1–38. DOI: 10.1016/j.artint.2018.07.007
16. Gunning D., Aha D. DARPA's explainable artificial intelligence (XAI) program // AI Magazine. 2019. № 40 (2). P. 44–58. DOI: 10.1609/aimag.v40i2.2850

17. Захарова И. Г. Методы машинного обучения для информационного обеспечения управления профессиональным развитием студентов // Образование и наука. 2018. Т. 23, № 9. С. 91–114. DOI: 10.17853/1994-5639-2018-9-91-114
18. Bird S., Klein E., Loper E. Natural language processing with Python: Analyzing Text with the natural language toolkit. Sebastopol. CA: O'Reilly Media, 2009. 504 p.
19. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Duchesnay E. Scikit-learn: Machine learning in Python // Journal of Machine Learning Research. 2011. № 12. P. 2825–2830. Available from: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf> (date of access: 30.07.2021).
20. Salehi M., Kamalabadi I. N., Ghouschi M. B. G. Personalized recommendation of learning material using sequential pattern mining and attribute based collaborative filtering // Education and Information Technologies. 2014. № 19 (4). P. 713–735. DOI: 10.1007/s10639-012-9245-5
21. Захарова И. Г., Боганюк Ю. В., Воробьева М. С. Павлова Е. А. Диагностика профессиональной компетентности студентов ИТ-направлений на основе данных цифрового следа // Информатика и образование. 2020. № 4. С. 4–11. DOI: 10.32517/0234-0453-2020-35-4-4-11
22. Zawacki-Richter O., Marin V. I., Bond M., Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? // International Journal of Educational Technology in Higher Education. 2019. № 16(1). P. 1–27. DOI: 10.1186/s41239-019-0171-0
23. Schröer C., Kruse F., Gómez J. M. A Systematic literature review on applying CRISP-DM process model // Procedia Computer Science. 2021. Vol. 18. P. 526–534. DOI: 10.1016/j.procs.2021.01.199
24. Боргест Н. М. Стратегии интеллекта и его онтологии: попытка разобраться // Онтология проектирования. 2019. Т. 9, № 4. С. 407–428. DOI: 10.18287/2223-9537-2019-9-4-407-428
25. George G., Lal A. M. Review of ontology-based recommender systems in e-learning // Computers & Education. 2019. Vol. 142. DOI: 10.1016/j.compedu.2019.103642. Available from: <https://www.sciencedirect.com/science/article/pii/S0360131519301952> (date of access: 30.07.2021).
26. Guo S., Zeng D., Dong S. Pedagogical data analysis via federated learning toward Education 4.0 // American Journal of Education and Information Technology. 2020. № 4 (2). P. 56–65. DOI: 10.11648/j.ajeit.20200402.13
27. Arrieta A. B., Diaz-Rodríguez N., Del Ser J., Bennetot A., Tabik S., Barbado A., et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI // Information Fusion. 2020. Vol. 58. P. 82–115. DOI: 10.1016/j.inf-fus.2019.12.012
28. Zhou W., Han W. Personalized recommendation via user preference matching // Information Processing & Management. 2019. № 56 (3). P. 955–968. DOI: 10.1016/j.ipm.2019.02.002

## References

1. Zeer E. F., Symaniuk E. E. Individual educational trajectories in the system of continuous education. *Pedagogicheskoe obrazovanie v Rossii = Pedagogical Education in Russia*. 2014; 3: 74–82. (In Russ.)
2. Zeer E. F., Zavodchikov D. P., Zinnatova M. V., Lebedeva E.V. Individual educational trajectory as intention of subject in continuing education system. *Nauchnyi dialog = Scientific Dialogue*. 2017; 1: 266–279. (In Russ.)
3. Andrews R., Li J., Lovenheim M. F. Heterogeneous paths through college: detailed patterns and relationships with graduation and earnings. *Economics of Education Review*. 2014; 42: 93–108. DOI: 10.1016/j.econedurev.2014.07.002
4. Haas C., Hadjar A. Students' trajectories through higher education: A review of quantitative research. *Higher Education*. 2020; 79 (6): 1099–1118. DOI: 10.1007/s10734-019-00458-5
5. Zeer E. F., Tserkovnikova N. G., Tretyakova V. S. Digital generation in the context of predicting the professional future. *Obrazovanie i nauka = The Education and Science Journal*. 2021; 23 (6): 153–184. DOI: 10.17853/1994-5639-2021-6-153-184 (In Russ.)
6. Clow D. The learning analytics cycle: Closing the loop effectively. In: *Proceedings of the 2<sup>nd</sup> International Conference on Learning Analytics and Knowledge – LAK'12*; 2012 Apr 29–May 2; Vancouver, British Columbia, Canada. New York: Association for Computing Machinery; 2012. p. 134–138.
7. Clow D. An overview of learning analytics. *Teaching in Higher Education*. 2013; 18 (6): 683–695. DOI: 10.1080/13562517.2013.827653
8. Jones K. M. Advising the whole student: eAdvising analytics and the contextual suppression of advisor values. *Education and Information Technologies*. 2019; 24 (1): 437–458. DOI: 10.1007/s10639-018-9781-8
9. Jones K. M. Just because you can doesn't mean you should: practitioner perceptions of learning analytics ethics. *Libraries and the Academy*. 2019; 19 (3): 407–428. DOI: 10.1353/pla.2019.0025
10. Pargman C. T., McGrath C. Mapping the ethics of learning analytics in higher education: A systematic literature review of empirical research. *Learning Analytics* [Internet]. 2021 [cited 2021 July 30]; 1: 1–17. Available from: <https://www.learning-analytics.info/index.php/JLA/article/view/7254>
11. Tsai Y. S., Poquet O., Gašević D., Dawson S., Pardo A. Complexity leadership in learning analytics: Drivers, challenges, and opportunities. *British Journal of Educational Technology*. 2019; 50 (6): 2839–2854. DOI: 10.1111/bjet.12846
12. Kitto K., Knight S. Practical ethics for building learning analytics. *British Journal of Educational Technology*. 2019; 50 (6): 2855–2870. DOI: 10.1111/bjet.12868
13. Castelvechi D. Can we open the black box of AI? *Nature News*. 2016; 538 (7623): 20–23. DOI: 10.1038/538020a
14. Lipton Z. C. The mythos of model interpretability. *Communications of the ACM*. 2018; 61 (10): 36–43. DOI: 10.1145/3233231
15. Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. 2019; 267: 1–38. DOI: 10.1016/j.artint.2018.07.007
16. Gunning D., Aha D. DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*. 2019; 40 (2): 44–58. DOI: 10.1609/aimag.v40i2.2850

17. Zakharova I. G. Machine learning methods of providing informational management support for students' professional development. *Obrazovanie i nauka = The Education and Science Journal*. 2018; 20 (9): 91–114. DOI: 10.17853/1994-5639-2018-9-91-114 (In Russ.)
18. Bird S., Klein E., Loper E. Natural language processing with Python: Analyzing text with the natural language toolkit. Sebastopol: O'Reilly Media; 2009. 504 p.
19. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Duchesnay E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* [Internet]. 2011 [cited 2021 July 30]; 12: 2825–2830. Available from: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
20. Salehi M., Kamalabadi I. N., Ghouschi M. B. G. Personalized recommendation of learning material using sequential pattern mining and attribute based collaborative filtering. *Education and Information Technologies*. 2014; 19 (4): 713–735. DOI: 10.1007/s10639-012-9245-5
21. Zakharova I. G., Boganyuk Yu. V., Vorobeva M. S., Pavlova E. A. Diagnostics of professional competence of IT students based on digital footprint data. *Informatika i obrazovanie = Informatics and Education Journal*. 2020; 4: 4–11. DOI: 10.32517/0234-0453-2020-35-4-4-11 (In Russ.)
22. Zawacki-Richter O., Marin V. I., Bond M., Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*. 2019; 16 (1): 1–27. DOI: 10.1186/s41239-019-0171-0
23. Schröer C., Kruse F., Gómez J. M. A Systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*. 2021; 18: 526–534. DOI: 10.1016/j.procs.2021.01.199
24. Borgest N. M. Strategies of intelligence and its ontology: an attempt to understand. *Ontologii proektirovaniya = Ontology of Designing*. 2019; 9 (4): 407–428. DOI: 10.18287/2223-9537-2019-9-4-407-428 (In Russ.)
25. George G., Lal A. M. Review of ontology-based recommender systems in e-learning. *Computers & Education* [Internet]. 2019 [cited 2021 July 30]; 142. DOI: 10.1016/j.compedu.2019.103642. Available from: <https://www.sciencedirect.com/science/article/pii/S0360131519301952>
26. Guo S., Zeng D., Dong S. Pedagogical data analysis via federated learning toward Education 4.0. *American Journal of Education and Information Technology*. 2020; 4 (2): 56–65. DOI: 10.11648/j.ajeit.20200402.13
27. Arrieta A. B., Díaz-Rodríguez N., Del Ser J., Bennetot A., Tabik S., Barbado A., et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. *Information Fusion*. 2020; 58: 82–115. DOI: 10.1016/j.inffus.2019.12.012
28. Zhou W., Han W. Personalized recommendation via user preference matching. *Information Processing & Management*. 2019; 56 (3): 955–968. DOI: 10.1016/j.ipm.2019.02.002

#### **Информация об авторах:**

**Захарова Ирина Гелиевна** – доктор педагогических наук, профессор, профессор кафедры программного обеспечения Тюменского государственного университета; ORCID 0000-0002-4211-7675; Тюмень, Россия. E-mail: [i.g.zakharova@utmn.ru](mailto:i.g.zakharova@utmn.ru)

**Воробьева Марина Сергеевна** – кандидат технических наук, доцент, заведующая кафедрой программного обеспечения Тюменского государственного университета; ORCID 0000-0002-1508-4089; Тюмень, Россия. E-mail: [m.s.vorobeva@utmn.ru](mailto:m.s.vorobeva@utmn.ru)

**Боганюк Юлия Викторовна** – старший преподаватель кафедры программного обеспечения Тюменского государственного университета; ORCID 0000-0002-3663-0173; Тюмень, Россия. E-mail: y.v.boganyuk@utmn.ru

**Вклад соавторов:**

И. Г. Захарова – постановка проблемы исследования, разработка методологии и методов исследования, обсуждение и обобщение результатов.

М. С. Воробьева – разработка методов исследования и алгоритмов, обсуждение и обобщение результатов.

Ю. В. Боганюк – программная реализация методов и алгоритмов, обсуждение и обобщение результатов.

**Информация о конфликте интересов.** Авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 03.08.2021; поступила после рецензирования 03.12.2021; принята к публикации 10.12.2021.

Авторы прочитали и одобрили окончательный вариант рукописи.

**Information about the authors:**

**Irina G. Zakharova** – Dr. Sci. (Education), Professor, Software Department, University of Tyumen; ORCID 0000-0002-4211-7675; Tyumen, Russia. E-mail: i.g.zakharova@utmn.ru

**Marina S. Vorobeve** – Cand. Sci. (Engineering), Head of the Software Department, University of Tyumen; ORCID 0000-0002-1508-4089; Tyumen, Russia. E-mail: m.s.vorobeve@utmn.ru

**Yulia V. Boganyuk** – Senior Lecturer, Software Department, University of Tyumen; ORCID 0000-0002-3663-0173; Tyumen, Russia. E-mail: y.v.boganyuk@utmn.ru

**Contribution the authors:**

I. G. Zakharova – statement of the research problem, development of research methodology and methods, results discussion, and conclusion.

M. S. Vorobeve – development of research methods and algorithms, results discussion, and conclusion.

Yu. V. Boganyuk – software implementation of methods and algorithms, results discussion, and conclusion.

**Conflict of interest statement.** The authors declare that there is no conflict of interest.

Received 03.08.2021; revised 03.12.2021; accepted for publication 10.12.2021.

The authors have read and approved the final manuscript.