Е. В. БАРАНОВА, Г. В. ШВЕЦОВ

Методы и инструменты для анализа цифрового следа студента при освоении образовательного маршрута

Сбор и анализ цифрового следа студента является неотъемлемой частью развития российского образования в условиях цифровой экономики. Такая деятельность связана, в том числе, с оценкой формирования компетенций студентов при реализации смешанного обучения в образовательном процессе. Цель исследования – разработка подходов, методов и инструментов для анализа данных, сформированных в цифровой образовательной среде (ЦОС), с целью выявления связей между структурой и содержанием образовательных программ и результатами учебной деятельности студентов по их освоению.

Методология работы базируется на методах анализа образовательных данных Educational Data Mining (EDM), методах проектирования и разработки информационных систем и баз данных. В качестве методов обработки данных использовались: корреляционный анализ и «перегонка данных для принятия решение человеком» (Distillation of Data for Human Judgment). Для анализа использовался большой объем детальных данных о различных аспектах образовательного процесса, накопленных более чем за 10 лет на базе Российского государственного педагогического университета им. А. И. Герцена.

Авторами разработан инструмент для сбора и анализа методами EDM цифрового следа обучающихся в условиях смешанного обучения. В соответствии с методами EDM, были произведены расчеты, которые позволяют сделать предварительные выводы о характере взаимосвязи выделенных показателей (количество действий на курсе, активность студентов в чате(-ах) курса, итоговая оценка за курс, время на курсе), полученных в рамках работы студентов в системе дистанционного обучения (СДО) Moodle и результатов промежуточной аттестации. Например, значение зависимости между показателем итоговой оценки за ЭУК и результатами промежуточной аттестации равно 0,7108, что говорит о сильной взаимосвязи. Значение зависимости между количеством действий на курсе и оценкой за промежуточную аттестацию (0,314) показывает слабую взаимосвязь. На основе таких расчетов можно формулировать предположения: об оценке влияния видов деятельности студентов на успешность освоения дисциплин, об эффективности электронных учебных курсов (ЭУК) и рекомендациям по изменению образовательных ресурсов с целью повышения их качества. Представленный авторами опыт, подходы, результаты свидетельствуют о целесообразности использования методов EDM для анализа образовательного процесса.

Ключевые слова: цифровой след студента, Data Mining, Educational Data Mining, Big Data, онлайн-обучение, цифровая образовательная среда

Ссылка для цитирования:

Баранова Е. В., Швецов Г. В. Методы и инструменты для анализа цифрового следа студента при освоении образовательного маршрута // Перспективы науки и образования. 2021. № 2 (50). С. 415–430. doi: 10.32744/pse.2021.2.29

E. V. BARANOVA, G. V. SHVETSOV

Methods and tools for analysing students' digital footprint in the course of work under educational programmes

Collecting and analysing a student's digital footprint is an integral part of the Russian education development in the conditions of the digital economy. This activity is connected, among the other things, with the assessment of the students' formed competencies in the course of blended learning in the educational process. The purpose of the research is development of approaches, methods and tools for analysing the data generated in the digital educational environment (DEE), with a view to assessing the correlation between the structure and content of educational programmes and the results of the students' educational activities under these programmes.

The methodology of the essay is based on the Educational Data Mining (EDM), along with the methods for design and development of information systems and databases. The following data processing methods were used: correlation analysis and Distillation of Data for Human Judgment. The analysis used extensive detailed data on various aspects of the educational process, accumulated for over more than 10 years on the basis of Herzen State Pedagogical University of Russia.

The authors have developed a tool for collecting and analysing the students' digital footprint in a blended learning environment, using EDM methods. A series of calculations was made in accordance with the EDM methods, that make it possible to draw preliminary conclusions about the nature of relationship of some selected indicators (number of actions during the course, students' activity in the course chat(s), final grade for the course work, attendance time at the course), obtained as part of the student work in the Moodle distance teaching system (DTS), with the interim certification results. For instance, the extent of interrelation between the final e-learning course grade and the results of interim certification is 0.7108, which is evidence of close interdependence. The extent of interrelation between the scope of course activities and the interim certification grade (0.314) shows only slight interdependence. Such calculations serve a basis for the assumptions in respect of assessing the impact of student activities on academic success, on the efficiency of e-learning courses and recommendations for modifying the educational resources with a view to improving their quality. The experience, approaches and results presented by the authors testify to expediency of using EDM methods for the analysis of the educational process.

Keywords: student digital footprint, Data Mining, Educational Data Mining, Big Data, online learning, digital educational environment

For Reference:

Baranova, E. V., & Shvetsov, G. V. (2021). Methods and tools for analysing students' digital footprint in the course of work under educational programmes. *Perspektivy nauki i obrazovania – Perspectives of Science and Education*, 50 (2), 415-430. doi: 10.32744/pse.2021.2.29

Введение

Переход к цифровой экономике – серьезный вызов для всей системы образования. Цифровая экономика требует от системы образования не просто «оцифровки» отдельных процессов, а комплексного подхода, который бы ставил новые цели, менял структуру и содержание образовательного процесса.

Эффективное образование включает не только передачу информации от педагога к ученику, оно требует сложных социальных взаимодействий и адаптации к потребностям каждого учащегося и их культурно-социальному контексту. Цифровая трансформация изменяет содержание, формы и методы образования. «Оцифровка» затрагивает прежде всего системы школьного и университетского образования. Массовые открытые онлайн-курсы, преподавание, с использованием чат-ботов и планы занятий, основанные на использовании искусственного интеллекта – вот только несколько примеров цифровой трансформации в высшем образовании [24].

Активное внедрение информационных технологий в различные сферы жизни, в том числе образовательный процесс, приводит к частичному или полному переносу деятельности человека в виртуальную среду. Информацию, размещаемую пользователем на своей странице в цифровой среде, можно рассматривать как продукт деятельности, который позволяет выделить особенности конкретного человека, строить предположения о его психологических особенностях и когнитивных способностях [31].

Непрерывная цифровая трансформация общества вносит свои коррективы в традиционные модели обучения, в том числе – и в модели электронного образования [33]. Один из новых трендов последнего времени – построение цифрового профиля обучаемых, основанного на анализе так называемого цифрового следа. В исследованиях современных авторов понятие цифрового следа обучаемых трактуется по-разному, но при этом прослеживается общая перспектива – переход от узкого, технического понимания следа как набора IP-адресов и отчетов о сетевой активности обучаемых, к более широкому рассмотрению.

О. П. Жигалова отмечает, что «Цифровые следы, рассматриваемые как результаты учебной и профессиональной деятельности в цифровом формате, представляют собой данные, позволяющие определить уровень компетенций, сформировать траекторию обучения, оценить возможности и стратегию дальнейшего развития и профессионального становления в определённой сфере» [22]. По мнению В. Н. Курбацкого: «Цифровой след представляет собой массив данных о результатах образовательной и проектной деятельности студента, включая все материалы, которые учащийся создает: презентации, прототипы, аудио- и видеозаписи, дорожные карты и т.д.» [24].

Распространенность и влияние социальных сетей на цифровой облик обучающегося увеличивается. В работах (С. А. Вартанов [21]; Г. А. Николаенко [27]; Т. В. Тулупьева [32] и др.) говорится, что подобная информация может быть очень разной и включать в себя сотни видов «следов»: от фотографий и видеороликов до комментариев, лайков, репостов и прочей виртуальной активности.

Цифровой след студента начинает формироваться с момента поступления в университет и содержит персональные данные, информацию о поступлении, направлении подготовки, образовательной программе, успеваемости в течение всего периода обучения, а также данные учебной аналитики, собираемые автоматически при работе

студентов в системах дистанционного обучения и на платформах открытого образования при прохождении студентами онлайн-курсов.

Цифровой след в образовательном пространстве – отчуждаемый результат образовательной деятельности человека. Всё образовательное пространство, которое собирает цифровой след, должно быть особым образом спроектировано с применением педагогического дизайна.

К методам обработки информации об образовательном процессе, в частности цифрового следа его субъектов, можно отнести методы Educational Data Mining (EDM), предоставляющих обширные возможности для анализа и интерпретации полученных результатов. Однако, несмотря на очевидные перспективы применения методов EDM, исследований, связанных с их использованием для моделирования образовательного процесса в вузе, явно недостаточно.

Отсутствуют разработанные методики, инструменты анализа и интерпретации данных, базирующиеся на этих методах и предназначенные для:

- выявления критических точек в процессе освоения образовательных программ, дисциплин и модулей, вызывающих наибольшие затруднения у студентов,
- определения взаимосвязи и взаимовлияния компонентов традиционного и дистанционного обучения на успешность освоения студентами образовательных программ
- выявления на раннем этапе контингента студентов, находящихся в группе «риска», с большой вероятностью отчисления т .д.

Цель проводимого авторами исследования состоит в разработке методов и инструментов анализа данных, сформированных в цифровой образовательной среде (ЦОС), для выявления связей между структурой и содержанием образовательных программ и результатами учебной деятельности студентов по их освоению с целью оптимизации образовательного процесса.

Анализ литературы

Интеллектуальный анализ данных – английский синоним Data Mining (DM) – трактуется в отечественной и зарубежной литературе, как собирательное название целой совокупности методов, для обозначения которых в литературе также используются термины: «извлечение информации», «раскопка данных», «интеллектуальный анализ данных», «извлечение знаний», «анализ шаблонов», «обнаружение знаний в базах данных» [29].

В основе подходов DM лежат методы классификации, кластеризации, моделирования и прогнозирования, построения деревьев решений, эволюционного программирования и нечеткой логики. Первоначально методы DM базировались на результатах исследований в области прикладной статистики. В настоящий момент развитие методов DM существенно связано с исследованиями в области искусственного интеллекта (распознавание образов, поиск моделей представления знаний, нейронные сети, генетические алгоритмы), работами по визуализации информации, машинному обучению и т.д. В основу технологии DM положена концепция шаблонов (patterns), которые описывают собой закономерности, выраженные в понятных человеку формах.

Современное направление исследований, связанных с применением методов интеллектуального анализа данных, машинного обучения и статистики к разного

рода информации об образовательном процессе получило название Educational Data Mining (EDM), в русскоязычной литературе переводится как «анализ образовательных данных» [20]. В рамках этого направления разрабатываются модели и методы обработки образовательных данных, направленные на совершенствование образовательного процесса. Направление EDM конкретизирует область источника больших массивов данных (Big Data, далее BD) применительно к образовательному процессу, ставит перед собой цель поиска паттернов (образцов, шаблонов, схем, закономерностей), характерных для этой предметной области. Подробный обзор ключевых компонентов EDM был представлен в [9]. Реña-Ayala, A. выделил три категории – основные блоки исследований, связанных с EDM: задачи, методы и алгоритмы. Использование EDM также рассматривается в исследованиях, цель которых состоит в том, чтобы понять, как учащиеся взаимодействуют с образовательными технологиями [2].

В работе [8] утверждается, что на основе методов EDM можно построить модели для описания различных аспектов образовательного процесса для выявления факторов, влияющих на успешность освоения образовательных маршрутов обучающимися, в том числе, преподавательская, административная и организационная деятельность. В работе [6] говорится о том, что методы кластеризации и визуализации могут использоваться в целях анализа образовательных данных и на основе этого улучшать процесс электронного обучения. В исследовании [10] с помощью метода EDM – правила ассоциаций на основе данных, полученных в результате проведения опроса, выявляются связи между категориальными переменными опроса (эффективность окружающей среды, влияние окружающей среды, полезность окружающей среды, контроль над окружающей средой, легкость обучения в окружающей среде) для дальнейшего улучшения иммерсивной среды (технология расширенной реальности, обеспечивают эффект полного или частичного присутствия в альтернативном пространстве). В исследовании [12] с помощью методов EDM определяется влияние внешних факторов (отношения, употребление алкоголя, образование родителей, частота встреч с друзьями, место проживания) на успеваемость студентов.

Согласно В. А. Ларионовой, А. А. Карасик, интеллектуальный анализ данных позволяет изучить поведенческие шаблоны студентов, узнать, как учесть индивидуальные особенности каждого при формировании образовательных траекторий [25].

В исследовании Abdulmohsen A. Говорится, что к основным источникам данных, используемых в EDM на сегодняшний день, можно отнести следующие категории:

- Традиционное обучение – знания передаются учащимся на основе личного контакта. Данные можно собирать традиционными методами, такими как наблюдение и анкетирование, которые позволяют изучить когнитивные навыки учащихся и определить успешность их обучения. К данным можно применить статистический метод и психометрию.
- Электронное обучение и системы управления обучением (LMS) предоставляют студентам материалы, инструкции, средства коммуникации и отчетности, которые позволяют им учиться самостоятельно. Методы интеллектуального анализа данных могут применяться к данным, хранимым в базах данных.
- Интеллектуальные обучающие системы и адаптивные образовательные гипермедийные системы предоставляют контент студентам, на основе их профилей, поэтому для анализа созданных профилей пользователей необходимо применять методы интеллектуального анализа данных [ссылка].

В частности, в исследованиях (Ezekiel U. O. [7], Rawad C. [11] и др.) рассматривалось применение методов интеллектуального анализа образовательных данных для выявления зависимостей во взаимодействии студентов с элементами СДО (файлы, тестовые задания, форумы) и их итоговой успеваемостью; взаимосвязи между данными обучающимися массового открытого онлайн-курса (MOOC) – возраста, пола, географического региона, образования и социально-профессионального статуса и их успеваемостью.

Анализ зарубежных и отечественных источников [3-5] позволяет выделить следующие направления использования методов EDM в образовательной практике, наиболее важные для высшего образования:

- моделирование поведения студентов в процессе обучения, с целью прогнозирования развития их познавательных возможностей, выявления контингента студентов, входящих в группы «риска» с высокой вероятностью быть отчисленными;
- разработка новых моделей и способов представления знаний в предметной области, которые бы соответствовали разнообразным стилям обучения и познавательным возможностям, выявленным у студентов при помощи методов EDM;
- изучение процессов взаимодействия студентов с цифровой средой, выявление компонентов среды, изучения эффектов, которые оказывает образовательная среда на результаты этого обучения.

Материалы и методы

В рамках данного исследования основной интерес представляют компоненты цифрового следа, связанные с освоением обучающимися образовательных маршрутов.

Объектом исследования являются компоненты цифрового следа, представленные промежуточной успеваемостью студентов в виде оценок, полученных в ходе изучения дисциплин, практик, модулей, результатами учебной деятельности студентов в СДО Moodle в процессе освоения образовательной программы и их взаимосвязь. Исследование осуществляется с целью выявления критических для студентов этапов с целью оптимизации структуры, содержания и методов реализации образовательных программ высшего образования.

К средствам исследования относятся информационные системы и базы данных для формирования и сбора данных, методы EDM – для анализа, обработки и измерения данных.

Разработанная авторами методика анализа цифрового следа студента базируется на сформулированных в [26] этапах использования методов EDM для анализа данных и состоит в содержании этапов, уточненных применительно к предметной области в соответствии с поставленной целью исследования, характеристиках средств и инструментов для получения, обработки, анализа данных и представления результатов.

Первый этап включает постановку задачи и выявление сущностей рассматриваемой предметной области. Современные условия определяют необходимость интеграции различных технологий дистанционного обучения и традиционных форм реализации учебного процесса. В этих условиях актуальной представляется задача анализа и выявления связей показателей успеваемости студентов при реализации различных форм обучения. Выделим основные информационные источники данных:

- образовательная программа – комплекс основных характеристик образования (объем, содержание, планируемые результаты), организационно-педагогических условий и форм аттестации, который представлен в виде учебного плана, календарного учебного графика, рабочих программ учебных предметов, курсов, дисциплин (модулей), иных компонентов, а также оценочных и методических материалов [28];
- деятельность профессорско-преподавательского состава (ППС), включающая: разработку учебно-методических материалов (лекции, практические занятия, тесты), в том числе в цифровом виде, оценивание студентов в рамках текущей, промежуточной успеваемости и т.д.;
- учебная деятельность студентов, которая оценивается в ходе текущей, промежуточной и итоговой успеваемости.

Детализация взаимосвязей между указанными информационными источниками образовательного процесса позволила разработать схему информационных потоков, представленную на рисунке 1.

Рисунок 1 Схема информационных потоков

Второй этап предполагает подготовку данных для EDM и базируется на методах проектирования и разработки информационных систем и БД [30]. На этом этапе происходит моделирование данных, т.е. осуществляется определение и анализ требований к данным, которые необходимы для осуществления EDM, разработка структур оперативных, справочных и архивных базы данных (БД), для представления источников данных, формируемых с использованием внутренних и внешних информационных систем.

Для реализации представленной на рисунке 1 схемы информационных потоков в РГПУ им. А. И. Герцена разработана и полномасштабно внедрена в практику образова-

тельного процесса интегрированная информационная система управления учебным процессом – ИСУП (<http://oio.herzen.edu.ru>) [17]. ИСУП интегрирована с различными компонентами ЦОС, в том числе с СДО Moodle. ИСУП включает совокупность информационных систем (ИС) и веб-ресурсов, базируется на распределенной БД HERZEN, которая обеспечивает хранение целостной информации о различных аспектах образовательного процесса, в том числе:

- данные о реализуемых основных профессиональных образовательных программах;
- успеваемость студентов;
- информацию об учебной нагрузке профессорско-преподавательского состава;
- расписание занятий и экзаменов студентов;
- данные об ЭУК преподавателей, используемых в рамках реализации образовательных программ (дисциплин, модулей, практик) и т.д.

Формирование данных в БД для последующего анализа реализуется с использованием разработанных с участием авторов инструментов (информационных систем) – компонентов ИСУП.

Инструментом, для формирования данных о реализуемых основных профессиональных образовательных программах является ИС «Учебные и рабочие планы», обеспечивающая создание электронных учебных планов образовательных программ в соответствии с требованиями ФГОС ВПО и ФГОС ВО, автоматического формирования рабочих планов и объемов учебных поручений кафедр для всех уровней образования и форм обучения.

Для сбора данных о промежуточной успеваемости студентов используется ИС «Деканат», обеспечивающая формирование: персональных данных студентов, информации о характеристиках образовательной программы студента, информации о структуре и составе образовательной программы, данных об индивидуальных образовательных маршрутах студентов и успешности их освоения [15; 19].

Распределение учебной нагрузки между преподавателями реализуется с использованием инструмента ИС «Нагрузка кафедры».

Система, предназначенная для создания электронного расписания занятий и экзаменов студентов на основе рабочих планов и распределенной нагрузки преподавателей – ИС «Расписание». Сформированное расписание проверяется на корректность: соблюдение ограничений на количество учебных часов в день для группы и преподавателя; преподавателю не могут назначаться занятия в одно и то же время с разными группами; аудитория в одно и тоже время не может быть назначена для проведения различных занятий; лекционные занятия должны проводить только преподаватели со степенями и много другое.

Совокупность представленных инструментов позволяет формировать информацию в базах данных об основных профессиональных образовательных программах (ОПОП), контингенте студентов, нагрузке ППС для анализа данных методами EDM.

На современном этапе широко используется электронное обучение и дистанционные образовательные технологии. Для получения целостной и разноаспектной информации об образовательном процессе для EDM необходимо формирование данных о контенте и результатах электронного обучения в СДО. На сегодняшний день одной из наиболее популярных систем поддержки учебного процесса в условиях дистанционного образования является СДО Moodle. Важнейшими преимуществами

среды, обеспечивающими ее широкую востребованность, являются: открытость, мобильность, переносимость, расширяемость, широкая распространенность и т.д.

СДО Moodle содержит широкий набор инструментов для создания и сопровождения электронных учебных курсов, различные возможности для представления учебного материала, контроля успеваемости, реализации обратной связи между педагогом и студентами, которая позволяет студентам выкладывать работы в СДО, получать рецензии и консультации преподавателя [23].

Простые по структуре учебные курсы содержат учебно-методические материалы для освоения дисциплины в виде презентаций, текстов лекций, литературы, вопросов к экзамену и т. д. Большая часть курсов, размещенных в СДО Moodle университета, предполагают более широкое использование различных инструментов СДО и включают, как правило: рабочую программу дисциплины; учебно-методические материалы для освоения дисциплины; методические рекомендации для студентов по изучению ЭУК и подготовке к различным видам занятий, аттестации; форум для общения и обсуждения студентами и преподавателем вопросов по курсу; фонды оценочных средств для контроля знаний студентов.

СДО Moodle, базирующаяся на СУБД MariaDB (ответвление СУБД MySQL), обеспечивает хранение различных данных о взаимодействии студентов с цифровой средой в процессе обучения, в том числе текущую успеваемость. Аппарат запросов к БД Moodle позволяет получить различные наборы данных, включая:

- общие данные о курсе, преподавателях и студентах;
- количество проведенного студентом времени в онлайн-курсе времени;
- количество изменений, произведенных студентами в различных модулях курса;
- результаты выполненных заданий студентов на курсе;
- информацию об активности студентов на форумах курса и т.д.

К инструментам, обеспечивающим связь между источниками данных, представленными в ИСУП и СДО Moodle, относятся: ИС «Электронный мониторинг» [18] и веб-ресурс «Электронный атлас» [16] – компоненты для мониторинга деятельности профессорско-преподавательского состава, связанной с онлайн-обучением.

Системы обеспечивают возможность формирования преподавателем заявки на создание ЭУК в системе дистанционного обучения (СДО) Moodle, а также возможность обработки такой заявки. Такие ресурсы позволяют организовать связь между данными о дисциплинах, курсах в СДО Moodle и преподавателях, которые реализуют обучение. Список дисциплин и ЭУК преподавателя отображаются на его личной странице в «Электронном атласе» (см. рис. 2).

Дисциплины, которые ведет преподаватель по основным образовательным программам в 2019/20 учебном году	Форма обучения	Курс	Семестр	Запрос на открытие ЭУК
Дисциплины (модули) по выбору. Проектирование и разработка электронных образовательных ресурсов. Образование и педагогические науки - Теория и методика обучения и воспитания (информатика)	очная	3	1	ЭУК открыт
Модуль "Предметно-содержательный". Алгоритмы и методы программирования. Педагогическое образование - Информатика и информационные технологии в образовании	очная	1	1	ЭУК открыт

Рисунок 2 Веб-ресурс «Электронный атлас». Раздел «Дисциплины преподавателя»

Такая связь позволяет установить соответствие между контингентом (студенты, ППС) и их деятельностью в СДО Moodle, обеспечивает возможность получения более детальной информации о структуре конкретных курсов.

Третий этап включает сбор и анализ итоговых данных, определение методов EDM для их обработки и визуализации полученных результатов. В качестве методов EDM для выявления связей между сущностями предметной области были использованы корреляционный анализ и метод «перегонки данных для принятия решение человеком» (Distillation of Data for Human Judgment).

Термин «перегонки данных» введен Д. Сименсом, Р. Бэйкером и предполагает кластеризацию данных, т.е. автоматическое разбиение элементов на группы по «схожести» набора характеристик с целью сокращения размерности исходного набора данных, и визуализации данных, т.е. представление в виде, способствующем восприятию человеком [13].

На данном этапе предполагается использовать инструменты, предназначенные для получения сводных итоговых данных для анализа качества образовательной программы в целом по различным критериям, таким как успешность освоения программ студентами, квалификационные характеристики профессорско-преподавательского состава, наиболее сложные для освоения составляющие программы, востребованность электронных учебных курсов по дисциплинам и т.д.

Разработанный авторами веб-ресурс «Цифровой паспорт ОПОП» позволяет получить количественные значения показателей приведенных критериев для анализа структуры, состава, различных аспектов реализации образовательных программ.

Например, сформировать по каждой образовательной программе:

- данные о структуре кадрового состава, реализующего программу: ученые степени и должности преподавателей, участие преподавателей в реализации ОПОП (дисциплины/модули, практики, руководство ВКР, ГИА), участие кафедр в реализации ОПОП, деятельность преподавателя в СДО Moodle (количество часов, структура разработанных курсов) и т.д.
- данные о структуре контингента обучающихся: процент иностранных граждан, доля обучающихся на договорной основе, движение контингента (прием, отчисление, выпуск), трудоустройство выпускников (наличие работы, работа по специальности, трудоустройство по сферам деятельности);
- сводные данные о посещаемости и успеваемости студентов, рейтингах студентов в группе, учебной деятельности в СДО Moodle и т.д.

Для наглядного представления сформированные данные могут быть агрегированы по выбранным атрибутам: форма обучения, уровень образования, направление/специальность, направленность (профиль), бюджет/внебюджет и т.д..

Предложенная авторами методика позволяет определять характер взаимосвязей между выделенными сущностями предметной области, осуществлять анализ цифрового следа студента с целью внесения изменений в контент образовательных программ и прогнозирования успешности освоения студентами образовательных маршрутов.

Результаты исследования

Авторами разработан инструмент «Цифровой паспорт ОПОП» (ЦП ОПОП). Этот инструмент обеспечивает доступ к аналитическим данным об ОПОП и их реализации, позволяет отследить динамику отчисления контингента студентов, выявить на раннем этапе категории студентов, находящихся в группе «риска», образовательные программы и направления подготовки, сложные для освоения.

Инструмент позволяет осуществлять автоматическую кластеризацию данных об ОПОП (ППС, контингент студентов) по определенным характеристикам (штатный состав ППС, ученые степени, нагрузка по видам договоров, сводная успеваемость студентов, действующие студенты) и осуществлять их визуализацию. Ресурс является компонентом ИСУП, создается на базе современных технологий разработки веб-решений [14], в настоящий момент включает комплекс отчетов для руководства университета, который позволяет сформировать и визуализировать аналитические данные о: реализации ОПОП, ППС, контингенте студентов, посещаемости занятий.

Для оценки интенсивности работы студентов в СДО Moodle были выделены показатели, например, количество действий, выполненных с элементами курса, в том числе: количество переходов по страницам курса, загрузка ресурсов, переход по ссылкам, отправка сообщений на форумах и чатах. Среди показателей также были выделены: активность студентов в чате(-ах) курса (количество отправленных сообщений в чате(-ах)); итоговая оценка за курс (формируется на основе полученных баллов при выполнении заданий, тестов); время, проведенное студентами на курсе.

Далее был выбран ряд дисциплин, играющих ключевую роль в подготовке по различным образовательным программам, освоение которых базируется на широком использовании ЭУК в СДО Moodle: геометрия, теория организации и управление организационной политикой в образовании, систематика растений, история Древнего мира, защита и сохранение культурного наследия и др.

Информация об успешности освоения студентами образовательной программы формируется в виде оценок за промежуточную аттестацию по модулям, дисциплинам и практикам в БД Herzen. Для связи данных из двух баз была разработана модель интеграции (см. рис. 3) и реализована в ЦП ОПОП.

Инструмент позволяет выявлять взаимосвязь между выделенными показателями работы студентов в СДО Moodle и успешностью их обучения. На основе автоматического расчета показателей, значения которых могут быть рассчитаны автоматически с использованием запросов к БД MariaDB.

Приведем данные, полученные с использованием инструмента ЦП ОПОП, об обучении пяти групп, среднее количество студентов в которых, составляет 20 человек. Для каждой дисциплины, группы студентов и выбранному показателю были рассчитаны коэффициенты корреляции, отражающие зависимость значений показателя и оценки студентов в рамках промежуточной аттестации по дисциплинам (см. табл. 1).

Показатели количества действий студентов и их активности в чате(-ах) курса характеризуют побочную деятельность студентов для достижения конечного результата, значения итоговой оценки за курс формируются на основе качества выполнения заданий, предложенных студентам в процессе прохождения онлайн-курса.

Рисунок 3 Модель интеграции БД HERZEN и БД Moodle

Таблица 1

Коэффициенты корреляции (r) для выбранных показателей

Дисциплина	Кол-во студентов в группе	Значения r для показателя			
		Количество действий на курсе	Активность студентов в чате(-ах)	Итоговая оценка за курс	Время на курсе
Геометрия	20	0.2582	0.051	0.4014	0.5269
Теория организации и управление организационной политикой в образовании	27	0.6349	0.0308	0.917	0.5615
Систематика растений	16	0.1388	0.4070	0.9369	0.3882
История Древнего мира	29	0.1935	0.0494	0.6303	0.4290
Защита и сохранение культурного наследия	27	0.3447	0.5205	0.6959	0.4753

Представленные в таблице 1 расчеты, являются частью результатов исследования, базируются на сравнительно небольшом объеме данных, однако, позволяют сформулировать предположения, которые при дальнейшем исследовании, будут проверяться на больших объемах данных методами EDM.

Например, на основе полученных данных можно сделать предположение о том, что значения третьего показателя имеют наиболее стабильно высокую связь, среднее значение которой равно 0,7108. Таким образом, наличие в ЭУК средств, обеспечивающих возможность оценки текущей успеваемости (заданий, тестов) с достаточной степенью вероятности позволяют спрогнозировать успешность освоения дисциплины студентом.

В то же время, первый показатель имеет наименьшую связь, среднее значение равно 0,314. Можно сделать предварительный вывод о том, что данный показатель не является существенным фактором, характеризующим качество изучения дисциплины студентами.

Анализ данных в разрезе дисциплин показал существенные различия в значениях показателей, связанные, по мнению авторов, со спецификой предметной области и структурой ЭУК по дисциплине. На основе анализа результатов взаимодействия пользователей с учебно-методическими материалами ЭУК можно формулировать предположения об эффективности курса и рекомендации по изменению образовательного ресурса с целью повышения его качества.

Обсуждение результатов

Таким образом, в ходе анализа образовательных данных авторами были выделены этапы применения методов EDM которые включают: исследование сущностей предметной области, проектирование и разработку инструментов для реализации сбора и формирования данных, выбор подходов к выявлению взаимосвязей между этими данными и осуществление средствами разработанных инструментов обработки данных и наглядного представления результатов. Экспериментальные данные формировались на базе ЦОС Российского государственного педагогического университета им. А. И. Герцена им. А. И. Герцена.

Расчеты зависимостей между выделенными показателями цифрового следа студента подтвердили целесообразность и значимость использования методов EDM, так как позволили сделать выводы о воздействии таких показателей на образовательные результаты студентов в рамках текущей и промежуточной успеваемости. Справедливым будет вывод о том, что применение методов EDM напрямую зависит от подготовленности ЦОС образовательной организации (наличие соответствующих программных сред и навыков их использования), цифровой и IT-компетентности субъектов образовательного процесса.

Результаты дальнейших исследований, направленных на выявление факторов, влияющих на успешность освоения студентами образовательных программ, позволят, по мнению авторов, создавать рекомендации по внесению изменений в структуру и контент образовательных программ с целью повышения качества подготовки.

Заключение

Полученные результаты позволяют авторам перейти к очередному этапу исследования, с опорой на расширенную базу данных для анализа успеваемости контингента студентов в целом по всем образовательным программам, реализуемым в университете.

На основе таких данных будут разрабатываться модели (паттерны), описывающие учебную деятельность студентов в рамках смешанного обучения, что, по мнению авторов, позволит прогнозировать успешность освоения студентами образовательных маршрутов, выявлять студентов, находящихся в группе риска, и «критические точки» образовательных программ, сложные для освоения студентов, разрабатывать рекомендации по улучшению используемых электронных учебных курсов и т.д.

Таким образом, использование методов EDM в рамках предложенной методики и с применением разработанных инструментов для анализа и интерпретации детальных данных по различным аспектам образовательного процесса позволит повысить эффективность и результативность обучения.

Предложенная авторами методика анализа цифрового следа студентов в дальнейшем может использоваться для принятия обоснованных управленческих решений при проектировании, обновлении и реализации образовательных программ высшего образования с целью повышения качества подготовки.

ЛИТЕРАТУРА

1. Abdulmohsen A. Data Mining in Education. (IJACSA) International Journal of Advanced Computer Science and Applications, 2016, vol. 7, no. 6.
2. Angeli, C., Howard, S., Ma, J., Yang, J., Kirschner, P.A. Data mining in educational technology classroom research: can it make a contribution? Computers & Education, 2017. 113, pp. 226–242.
3. Baker R., Yacef K. The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 2009, vol. 1, no. 1, pp. 3–17. DOI: 10.5281/ZENODO.3554657.
4. Bowers A. J., Sprott R., Taff S. A. Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. The High School Journal, 2012, vol. 96, no. 2, pp. 77–100. DOI: 10.1353/hsj.2013.0000.
5. Bowers A. J. Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. Practical Assessment Research and Evaluation, 2010, vol. 15. Article 7. DOI: 10.7275/r4zq-9c31.
6. Castro, F. Applying Data Mining Techniques to e-Learning Problems [Text] / F. Castro, A. Vellido, A. Nebot, F. Mugica. Studies in Computational Intelligence, 2007, vol. 62, pp. 183–221.
7. Ezekiel U. O., Mogorosi, M. Educational Data Mining for Monitoring and Improving Academic Performance at University Levels. (IJACSA) International Journal of Advanced Computer Science and Applications, 2020, vol. 11, no. 11.
8. Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A. Big Data: The Next Frontier for Innovation, Competition, and Productivity [Text] / Ed. McKinsey Global Institute, 2011.
9. Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. Expert systems with applications, 2014, vol. 41(4), pp. 1432–1462.
10. R. Zamora-Musa and J. Velez. Use of Data Mining to Identify Trends between Variables to Improve Implementation of an Immersive Environment. Journal of Engineering and Applied Sciences, 2017, vol. 12, no. 22, pp. 5944–5948.
11. Rawad C., Rémi B. Internationalizing Professional Development: Using Educational Data Mining to Analyze Learners' Performance and Dropouts in a French MOOC. International Review of Research in Open and Distributed Learning, 2020, vol. 21, no. 40.
12. Razaque, Abdul & Alajlan, Abrar. Supervised Machine Learning Model-Based Approach for Performance Prediction of Students. Journal of Computer Science, 2020, 16, pp. 1150–1162.
13. Siemens G., Baker R. S. J. d. Learning analytics and educational data mining: towards communication and collaboration / LAK'12. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. New York: Association for Computing Machinery, 2012, pp. 252–254. DOI: 10.1145/2330601.2330661.
14. Stauffer M. Laravel: Up & Running: A Framework for Building Modern PHP Apps 2nd Edition, O'REILLY, 2019.
15. Баранова Е. В., Верещагина Н. О., Швецов Г. В. Цифровые инструменты для анализа учебной деятельности студентов // Известия Российского государственного педагогического университета имени А. И. Герцена. Санкт-Петербург, 2020. № 198. С. 56–65.
16. Баранова Е. В., Елизарова И. К. Интегрированная информационная система управления образовательными маршрутами в вузе («Электронный атлас»). Свидетельство о государственной регистрации программы для ЭВМ № 2011613872, 18.05.2011 г.
17. Баранова Е. В., Елизарова И. К., Верещагина Н. О. и др. Программный комплекс для управления учебным процессом в вузе (ПК «Герцен»). Свидетельство о государственной регистрации программы для ЭВМ № 2019618795. 04.07.2019 г.
18. Баранова Е. В., Елизарова И. К., Верещагина Н. О., Комарова С. М., Мороз Д. И. Мониторинг электронного обучения. Свидетельство о государственной регистрации программы для ЭВМ № 2016619725, 26.08.2016 г.
19. Баранова Е. В., Елизарова И. К., Слепухина Н. В. Информационная система «Деканат». Свидетельство о государственной регистрации программы для ЭВМ № 2014616013, 09.06.2014 г.
20. Белоножко П. П., Карпенко А. П., Храмов Д. А. Анализ образовательных данных: направления и перспективы применения // Интернет-журнал «НАУКОВЕДЕНИЕ». 2017. Том 9. № 4. URL: <https://cyberleninka.ru/article/n/analiz-obrazovatelnyh-dannyh-napravleniya-i-perspektivy-primeneniya/viewer> (дата обращения: 05.07.2020).
21. Бартанов С.А. Цифровые медиа и Big Data: математический подход к анализу медиасреды // Век информации. 2018. Т. 1. № 2. С. 211–213.
22. Жигалова О.П. Формирование образовательной среды в условиях цифровой трансформации общества //

- Ученые записки Забайкальского государственного университета. 2019. Т. 14. № 2. С. 69–74.
23. Зайцева О. Н. Использование LMS Moodle в образовании // Обучение и воспитание: методики и практика. 2012. № 2. С. 59–64.
 24. Курбацкий В. Н. Цифровой след в образовательном пространстве как основа трансформации современного университета // «Высшая школа»: научно-методический и публицистический журнал. 2019. № 5. С. 40–45.
 25. Ларионова В. А., Карасик А. А. Цифровая трансформация университетов: заметки о глобальной конференции по технологиям в образовании Edcrunch Ural // Университетское управление: практика и анализ. 2019. № 23. С. 130–135.
 26. Луньков А.Д., Харламов А.В. Интеллектуальный анализ данных//Учебно-методическое пособие. Саратовский национальный исследовательский университет. URL: http://elibrary.sgu.ru/uch_lit/1141.pdf (дата обращения: 03.10.2020).
 27. Николаенко Г.А. Перспективы использования цифровых следов исследователей для анализа их коммуникативных стратегий (на примере социальной сети ResearchGate) // Социология науки и технологий. 2019.
 28. Об образовании в Российской Федерации: федер. закон от 29.12.2012 № 273-ФЗ (ред. от 31.07.2020).URL: http://www.consultant.ru/document/cons_doc_LAW_140174/b819c620a8c698de35861ad4c9d9696ee0c3ee7a/ (дата обращения: 03.10.2020).
 29. Овсянницкая Лариса Юрьевна Интеллектуальный анализ данных как составляющая педагогического управления // Образование и наука. 2013. № 10. URL: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-kak-sostavlyayushaya-pedagogicheskogo-upravleniya> (дата обращения: 14.10.2020).
 30. Рочев К.В. Информационные технологии. Анализ и проектирование информационных систем. Учебное пособие. Издательство "Лань", 2019 г.
 31. Тулупьева Т. В., Суворова А. В., Азаров А. А., Тулупьев А. Л., Бордовская Н. В. Возможности и опыт применения компьютерных инструментов в анализе цифровых следов студентов-пользователей социальной сети // Компьютерные инструменты в образовании. 2015. № 5. С. 3–13.
 32. Тулупьева Т.В., Тафинцева А.С., Тулупьев А.Л. Подход к анализу отражения особенностей личности в цифровых следах // Вестник психотерапии. 2016. № 60 (65). С. 124–137.
 33. Шамсутдинова Т. М. Когнитивная модель траектории электронного обучения на основе цифрового следа // Открытое образование. 2020. № 2. URL: <https://cyberleninka.ru/article/n/kognitivnaya-model-traektorii-elektronного-obucheniya-na-osnove-tsifrovogo-sleda> (дата обращения: 21.09.2020).

REFERENCES

1. Abdulmohsen A. Data Mining in Education. (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, 2016, vol. 7, no. 6.
2. Angeli, C., Howard, S., Ma, J., Yang, J., Kirschner, P.A. Data mining in educational technology classroom research: can it make a contribution? *Computers & Education*, 2017. 113, pp. 226–242.
3. Baker R., Yacef K. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 2009, vol. 1, no. 1, pp. 3–17. DOI: 10.5281/ZENODO.3554657.
4. Bowers A. J., Sprott R., Taff S. A. Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. *The High School Journal*, 2012, vol. 96, no. 2, pp. 77–100. DOI: 10.1353/hsj.2013.0000.
5. Bowers A. J. Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. *Practical Assessment Research and Evaluation*, 2010, vol. 15. Article 7. DOI: 10.7275/r4zq-9c31.
6. Castro, F. Applying Data Mining Techniques to e-Learning Problems [Text] / F. Castro, A. Vellido, A. Nebot, F. Mugica. *Studies in Computational Intelligence*, 2007, vol. 62, pp. 183–221.
7. Ezekiel U. O., Mogorosi, M. Educational Data Mining for Monitoring and Improving Academic Performance at University Levels. (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, 2020, vol. 11, no. 11.
8. Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A. Big Data: The Next Frontier for Innovation, Competition, and Productivity [Text] / Ed. McKinsey Global Institute, 2011.
9. Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 2014, vol. 41(4), pp. 1432–1462.
10. R. Zamora-Musa and J. Velez. Use of Data Mining to Identify Trends between Variables to Improve Implementation of an Immer-sive Environment. *Journal of Engineering and Applied Sciences*, 2017, vol. 12, no. 22, pp. 5944–5948.
11. Rawad C., Rémi B. Internationalizing Professional Development: Using Educational Data Mining to Analyze Learners' Performance and Dropouts in a French MOOC. *International Review of Research in Open and Distributed Learning*, 2020, vol. 21, no. 40.
12. Razaque, Abdul & Alajlan, Abrar. Supervised Machine Learning Model-Based Approach for Performance Prediction of Students. *Journal of Computer Science*, 2020, 16, pp. 1150–1162.
13. Siemens G., Baker R. S. J. d. Learning analytics and educational data mining: towards communication and collaboration / *LAK'12. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. New York: Association for Computing Machinery, 2012, pp. 252–254. DOI: 10.1145/2330601.2330661.
14. Stauffer M. Laravel: Up & Running: A Framework for Building Modern PHP Apps 2nd Edition, O'REILLY, 2019.
15. Baranova E.V., Vereshchagina N.O., Shvetsov G.V. Digital tools for the analysis of educational activities of students.

- Bulletin of the Herzen Russian State Pedagogical University. St. Petersburg*, 2020, no. 198, pp. 56-65. (in Russ.)
16. Baranova E. V., Elizarova I. K. Integrated information system for managing educational routes in the university ("Electronic Atlas"). Certificate of state registration of a computer program No. 2011613872, 05/18/2011 (in Russ.)
 17. Baranova E. V., Elizarova I. K., Vereshchagina N. O. et al. A software package for managing the educational process at a university (PC "Herzen"). Certificate of state registration of a computer program No. 2019618795. 04.07.2019
 18. Baranova E.V., Elizarova I.K., Vereshchagina N.O., Komarova S.M., Moroz D.I. Monitoring of e-learning. Certificate of state registration of a computer program No. 2016619725, August 26, 2016 (in Russ.)
 19. Baranova E. V., Elizarova I. K., Slepukhina N. V. Information system "Dean's office". Certificate of state registration of a computer program No. 2014616013, 09.06.2014 (in Russ.)
 20. Belonozhko P. P., Karpenko A. P., Khramov D. A. Analysis of educational data: directions and prospects of application. *Scientific Research Internet magazine*, 2017, vol. 9, no. 4. Available at: <https://cyberleninka.ru/article/n/analiz-obrazovatelnyh-dannyh-napravleniya-i-perspektivy-primeneniya/viewer> (accessed 5 June 2020). (in Russ.)
 21. Vartanov S.A. Digital media and Big Data: a mathematical approach to the analysis of the media environment. *Information Age*, 2018, vol. 1, no. 2, pp. 211–213. (in Russ.)
 22. Zhigalova O.P. Formation of the educational environment in the context of the digital transformation of society. *Scientific notes of the Trans-Baikal State University*, 2019, vol. 14, no. 2, pp. 69–74. (in Russ.)
 23. Zaitseva O. N. Using LMS Moodle in education. *Teaching and education: methods and practice*, 2012, no. 2, pp. 59–64. (in Russ.)
 24. Kurbatsky V. N. Digital footprint in the educational space as the basis for the transformation of a modern university. *"Higher school"*, 2019, no. 5, pp. 40–45. (in Russ.)
 25. Larionova V.A., Karasik A.A. Digital transformation of universities: notes on the global conference on technologies in education Edcrunch Ural. *University management: practice and analysis*, 2019, no. 23, pp. 130–135. (in Russ.)
 26. Lunkov A.D., Kharlamov A.V. Data mining. Teaching aid. *Saratov National Research University*. Available at: http://elibrary.sgu.ru/uch_lit/1141.pdf (accessed 3 October 2020). (in Russ.)
 27. Nikolaenko G.A. Prospects for using researchers' digital traces to analyze their communication strategies (on the example of the ResearchGate social network). *Sociology of Science and Technology*, 2019. (in Russ.)
 28. On education in the Russian Federation: Feder. Law of December 29, 2012 No. 273-FZ (as amended on July 31, 2020). Available at: http://www.consultant.ru/document/cons_doc_LAW_140174/b819c620a8c698de35861ad4c9d9696ee0c3ee7a/ (accessed 3 October 2020). (in Russ.)
 29. Ovsyanitskaya Larisa Yurievna. Intelligent data analysis as a component of pedagogical management. *Education and Science*, 2013, no. 10. Available at: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-kak-sostavlyayuschaya-pedagogicheskogo-upravleniya> (accessed 14 October 2020). (in Russ.)
 30. Rochev K.V. Information Technology. Analysis and design of information systems. Tutorial. Lan Publishing House, 2019 (in Russ.)
 31. Tulupyeva T.V., Suvorova A.V., Azarov A.A., Tulupiyev A.L., Bordovskaya N.V. Possibilities and experience of using computer tools in the analysis of digital traces of students-users of a social network. *Computer tools in education*, 2015, no. 5, pp. 3–13. (in Russ.)
 32. Tulupyeva T.V., Tafintseva A.S., Tulupiyev A.L. An approach to the analysis of the reflection of personality traits in digital traces. *Bulletin of psychotherapy*, 2016, no. 60 (65), pp. 124-137. (in Russ.)
 33. Shamsutdinova T.M. Cognitive model of e-learning trajectory based on digital footprint. *Otkrytoe obrazovanie*, 2020, no. 2. Available at: <https://cyberleninka.ru/article/n/kognitivnaya-model-traektorii-elektronnogo-obucheniya-na-osnove-tsifrovogo-sleda> (accessed 21 September 2020). (in Russ.)

Информация об авторах

Баранова Евгения Васильевна
(Россия, Санкт-Петербург)

Профессор, доктор педагогических наук, профессор
кафедры цифрового образования
Российский государственный педагогический
университет им. А. И. Герцена
E-mail: ev_baranova@mail.ru
ORCID ID: 0000-0002-7995-0388
Scopus ID: 57215273771

Швецов Герман Владимирович

(Россия, Санкт-Петербург)

Аспирант
Институт информационных технологий и
технологического образования
Российский государственный педагогический
университет им. А. И. Герцена
E-mail: shvetzoff.german@yandex.ru
ORCID ID: 0000-0003-1908-0525

Information about the authors

Evgeniya V. Baranova

(Russia, St. Petersburg)

Professor,
Doctor of Pedagogy,
Professor of the Department of Digital Education
Herzen State Pedagogical University of Russia
E-mail: ev_baranova@mail.ru
ORCID ID: 0000-0002-7995-0388
Scopus ID: 57215273771

German V. Shvetsov

(Russia, St. Petersburg)

Graduate student
Institute of Information Technology and Technological
Education
Herzen State Pedagogical University of Russia
E-mail: shvetzoff.german@yandex.ru
ORCID ID: 0000-0003-1908-0525