

УДК 378:311

ББК 74.484к

Е. Ю. Огурцова, Р. Н. Фадеев

БОЛЬШИЕ ДАННЫЕ И ЦИФРОВАЯ АНАЛИТИКА В УНИВЕРСИТЕТСКОМ ОБРАЗОВАНИИ

В статье затрагивается проблема использования больших данных и цифровой аналитики в университетском образовании. Анализируются характеризующие признаки понятия «Big Data», что важно для представления о сложности задач обработки и анализа этих данных. Рассматриваются четыре основных вида аналитики данных: описательная, диагностическая, прогнозная и предписывающая. Обсуждается вопрос источников больших данных в университетах. Отмечается, что большие данные стали применяться для решения задач в рамках учебной аналитики. Приводятся примеры использования Big Data и цифровой аналитики в зарубежной и российской системах высшего образования. Анализ научных и публицистических работ позволяет выделить риски использования в образовательном процессе больших данных.

Ключевые слова: большие данные, цифровая аналитика, образовательный процесс, университет.

E. Yu. Ogurtsova, R. N. Fadeev

BIG DATA AND DIGITAL ANALYTICS IN THE UNIVERSITY EDUCATION

The article touches upon the problem of using big data and digital analytics in university education. The characterizing features of the concept of "big data" are analyzed, which is important for understanding the complexity of the tasks of processing and analyzing this data. Four main types of data analytics are considered: descriptive, diagnostic, predictive and prescriptive. The issue of big data sources in universities is discussed. It is noted that big data began to be used to solve problems in the framework of educational analytics. Examples of the use of Big Data and digital analytics in foreign and Russian systems of higher education are given. Analysis of scientific and publicistic works allows us to highlight the risks of using big data in the educational process.

Key words: big data, digital analytics, educational process, university.

DOI: 10.46724/NOOS.2021.4.37-44

Ссылка для цитирования: Огурцова Е. Ю., Фадеев Р. Н. Большие данные и цифровая аналитика в университетском образовании // Ноосферные исследования. 2021. Вып. 4. С. 37—44.

Citation Link: Ogurtsova, E. Yu., Fadeev, R. N. (2021) Bol'shiye dannyye i tsifrovaya analitika v universitetskom obrazovanii [Big data and digital analytics in the university education], *Noosfernyye issledovaniya* [*Noospheric Studies*], vol. 4, pp. 37—44.

Началом активного использования термина «большие данные» (Big Data) стал спецвыпуск в 2008 году журнала «Nature» с темой «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объемами данных?». В нем были представлены материалы, посвященные проблеме взрывного роста объемов информации [7].

В настоящее время большие данные характеризуют совокупностью семи признаков («семь V»), существенных для представления о сложности задач обработки и анализа этих данных: Volume, Velocity, Variety, Variability, Veracity, Value, Visualization. Рассмотрим подробнее эти признаки:

1. Volume (объем) свыше 150 Гб в сутки.
2. Velocity (обновление) информации происходит с большой скоростью и для ее обработки в режиме реального времени необходимы интеллектуальные технологии.
3. Variety (разнообразие) данных, которые могут быть структурированными, неструктурированными или структурированными частично. Каждый из этих типов данных требует различных видов анализа и подходящих инструментов.
4. Variability (переменчивость) потоков данных, которые могут иметь пики и спады, что приводит к сложностям в управлении и требует мощных технологий обработки.
5. Veracity (достоверность) набора данных и результатов его анализа.
6. Value (ценность или значимость) больших данных определяется эффективностью использования результатов их анализа для принятия решений.
7. Visualization (визуализация) является частью анализа больших данных и делает их доступными для человеческого восприятия.

Существуют четыре метода анализа Big Data: описательная (descriptive analytics), диагностическая (diagnostic analytics), прогнозная, или предикативная, аналитика (predictive analytics) и предписывающая (prescriptive analytics). Они представляют собой шаги по обеспечению зрелости аналитики, с каждым из которых расстояние между этапами анализа и принятия действия в отношении конвейера данных сокращается. Описательная аналитика говорит о том, что уже произошло. Диагностическая анализирует не просто произошедшие события, а их причину. Прогнозная аналитика предсказывает вероятные результаты на основе выявленных тенденций и статистических моделей. Предписывающая аналитика позволяет получить оптимальное решение на основе прогнозной аналитики и в значительной степени опирается на аналитику машинного обучения и нейронные сети. Для этого типа аналитики необходима твердая основа на базе трех других видов аналитики.

Сейчас многие крупные компании используют большие данные: собирают их, анализируют, применяют в связке с другими технологиями. Со временем большие данные перестанут быть прерогативой больших компаний. Эксперты отмечают, что быстро и недорого работать с большими объемами неструктурированных данных могут помочь data lake — озера данных. Они в отличие от баз данных могут хранить огромные объемы неструктурированной информации.

Большие данные стали применяться для решения задач в рамках учебной аналитики [2—6].

Источниками больших данных в университетах служат:

1. Персональные данные;

2. Данные о взаимодействии студентов с электронными системами обучения (какие выбирают электронные учебники, онлайн-курсы, показатели отказов, скорости просмотра страниц, возвраты к страницам, количество связей, расстояние связей, количество просмотров страниц одним пользователем и т. д.);

3. Данные об эффективности учебных материалов (какой тип ученика с какой частью контента взаимодействует, результаты взаимодействия, образовательные результаты и т. д.);

4. Административные (общесистемные) данные (посещаемость, пропуски по болезни и т. д.).

Рассмотрим примеры использования Big Data и цифровой аналитики в зарубежном и российском университетском образовании.

На каждого студента в Англии ведется подробное личное дело, которое включает в числе других данных следующее:

- Дату и время каждого опоздания или прогула.
- Дисциплинарные замечания.
- Сведения об инцидентах от университетской полиции, включая список зачинщиков, пострадавших и последствия.
- Участие в олимпиадах, конкурсах, тестах, включая занятое место, тему, призы.
- Участие в учебной и внеучебной деятельности (волонтерская деятельность, занятия по интересам, походы и т. п.).
- Участие в совместных проектах.

В Великобритании Университет Ноттингем Трент (Nottingham Trent University) для снижения показателя отчисления студентов, улучшения посещаемости и формирования чувства принадлежности к университетскому сообществу в 2013 году создал интерактивную систему дескриптивной аналитики. Дэшборд показывает частоту работы с библиотекой, сведения об изучаемых курсах, посещаемость и другие учебные показатели учащегося в сравнении с его одноклассниками. Если система не отмечает признаков активности студента в течение двух недель, то платформа посылает тьютору сообщение. Опрос студентов через 3 года после внедрения системы показал, что 72 % первокурсников использовали эту Big Data панель студенческого мониторинга и считают, что она мотивировала их увеличить количество времени, которое они посвящали учебному процессу.

Американский Университет Пердью (Purdue University) ввел в эксплуатацию систему предиктивной аналитики, которая оценивает вероятность риска отсева для каждого студента. Значение этого показателя периодически направляется координатору курса и обучающемуся. Университет отмечает, что использование интерактивной Big Data системы позволило улучшить результаты обучения и снизить показатели отчисления.

Университет Карнеги – Меллона (Carnegie Mellon University) использует онлайн-платформу с учебными курсами, которая предоставляет студентам подробный отчет об их образовательной активности и успеваемости. Использование учебной аналитики почти в 2 раза ускорило срок прохождения онлайн-курсов.

В Университете Остин Пии (Austin Peay State University) с 2012 года внедрена Big Data рекомендательная система (Degree Compass System), которая помогает студенту выбирать учебные курсы. Она дает свои рекомендации на осно-

ве анализа результатов обучения предыдущих студентов по конкретному курсу, успеваемости каждого студента и сведений об обучающихся с похожими профилями и интересами. Список рекомендуемых курсов каждого студента удобно отображается в веб-интерфейсе на защищенной странице университетского портала. Этот интерактивный интерфейс предоставляет информацию об учебной программе и требованиях каждого рекомендуемого курса, о том, какую роль этот курс играет в программе обучения студента, а также о доступности занятий в предстоящих семестрах. Эта же информация также доступна в PeayMobile, мобильном приложении Austin Peay State University (APSU). В 90 % случаев модель правильно предсказывает курсы, на которых учащиеся достигнут эффективного обучения. Когда сравнивались фактические оценки студентов по их семестровым курсам, оценки на курсах, которые были рекомендованы, в среднем на 0,46 балла были лучше, чем на курсах, которые система не советовала студенту.

В начале 2020 года Университет Северной Каролины (University of North Carolina) анонсировал систему мультизадачного обучения, способную предсказывать вероятность правильного ответа студента на основе его предыдущего поведения в игровом учебном процессе. Эта информация используется для адаптации процесса обучения, например, изменение сюжетной линии, дополнительные инструкции и подсказки.

В Университете Содружества Виргинии (Virginia Commonwealth University) проанализировали данные об отчислениях студентов и построили систему, которая оповещает администрацию, что студент в группе риска и ему нужна помощь. На основе уведомления начинают индивидуальную работу со студентом: предлагают дополнительные занятия, перевод на другой курс или помощь репетитора. По итогам семестра число студентов, закончивших курс, увеличилось на 16 %.

Государственный университет Болл в Индиане использует большие данные, чтобы анализировать участие студентов в разнообразных мероприятиях кампуса. Университет отслеживает частоту посещения кампуса и различных мероприятий с помощью идентификационных карт: если вовлеченность студента снижается, то сотрудники университета выявляют причину и могут предложить помощь.

Университет Сиднея прогнозирует успеваемость (оценки, отношение числа закончивших к числу поступивших и продвижение студентов в получении степени). Было отобрано 77 переменных для тестирования их влияния на успеваемость, например, такие как:

- академическая успеваемость;
- интеграция студентов в образовательный процесс;
- пол;
- социально-экономический статус;
- культурная принадлежность;
- психологические проблемы;
- ограничения физического здоровья;
- ощущение влияния на процесс обучения;
- логистика;
- владение английским языком;

- административные проблемы;
- режим обучения;
- использование доступных средств обучения;
- частота использования учебных ресурсов;
- характеристики изучаемых предметов.

Познакомиться с примерами работы российских университетов с большими данными можно было на Второй Международной конференции «Большие данные в образовании», которую в августе 2021 года организовал Московский городской педагогический университет.

На конференции было отмечено, что любой процесс в университете, от разработки образовательной программы до оценки студенческого проекта, можно улучшить с помощью использования технологии больших данных.

Директор по развитию онлайн-университета Skypro Михаил Свердлов рассказал, как технология больших данных позволяет им проектировать актуальные для рынка труда образовательные программы. Методы сбора и анализа больших данных помогают изучить потребности и тенденции рынка труда и на этой основе сформировать совокупность профессиональных компетенций, которая учитывается при составлении образовательных программ.

В Московском городском педагогическом университете создана система, которая на основе баллов ЕГЭ, результатов предыдущих сессий, информации из университетской библиотеки и электронных библиотечных систем, данных об участии студента в общественной деятельности предсказывает его успеваемость в следующем семестре. Точность прогнозирования составляет 71 %. Использование данного технического решения способствовало снижению процента отчислений за академическую неуспеваемость в два раза.

Вуз разрабатывает собственные алгоритмы для оценки вовлеченности студентов в учебный процесс на занятиях. В настоящий момент технические решения университета позволяют анализировать поведение одного–четырёх человек на видео: фиксируются поза, направление взгляда, выражение лица. В перспективе такие системы окажут помощь преподавателям. Можно будет оценить активность студента, групповую проектную работу. Анализ по данному алгоритму онлайн-лекции или видеозаписи лекции в очном формате определит, насколько удалось поддерживать внимание студенческой аудитории. Результаты анализа могут быть отправлены в личный кабинет студента и преподавателя, чтобы они могли сделать определенные выводы для себя.

В Уральском государственном педагогическом университете разработана модель прогноза посещаемости. В качестве исходных данных в ней используется информация об индивидуальных поведенческих особенностях студентов: темперамент, наличие работы, пересечение сферы интересов студента с учебной дисциплиной, мотивация, умение самоорганизации и многое другое. Дальнейшее обучение нейросети позволит реализовать выработку рекомендаций по развитию качеств студентов и формированию расписания, способствующих улучшению посещаемости занятий.

Академик РАН и РАО Алексей Семёнов в своем докладе обратил внимание на то, что проекты Big Data смогут помочь вузам принимать мотивированных первокурсников. Для большей эффективности приемной компании необходимо использовать цифровой след абитуриентов.

Институт образования НИУ Высшая школа экономики несколько лет подряд проводит исследования, посвященные анализу социального самочувствия студентов, оперируя при этом данными социальной сети «ВКонтакте».

В Чувашском государственном университете им. И. Н. Ульянова используют интеллектуальный анализ данных (Data Mining) при моделировании образовательного процесса. С помощью искусственных нейронных сетей получены вычислительные модели связи между баллами, полученными студентами на ЕГЭ, и успеваемостью студентов на первых курсах. Количество пятерок в первую сессию не зависит (близко к 0) от баллов ЕГЭ по математике до значения 70—75 баллов. При баллах ЕГЭ более 75 количество пятерок, полученных студентами в первую сессию, начинает резко расти, достигая максимума при баллах ЕГЭ более 85. Сделан вывод о том, что средние баллы ЕГЭ не могут служить основанием для определения эффективности вуза с точки зрения «качества образования», понимаемого как успеваемость студентов [1].

Предполагается решение комплекса задач [1]:

- прогнозирование «траектории» учебной и научной работы студентов различных специальностей и направлений подготовки и определение мер, способствующих повышению качества учебной и научной работы студентов;

- кластеризация специальностей и направлений подготовки по качеству приема абитуриентов, качеству учебной и научной работы студентов, уровню востребованности выпускников различных специальностей и направлений подготовки;

- создание многофакторных вычислительных моделей для таких целевых функций, как качество учебы по курсам и качество выпускных квалификационных работ, степень удовлетворенности участников процесса образования и т. п.;

- решение обратных задач, например, какой должен быть абитуриент, чтобы получить «отлично» по всем дисциплинам на первой сессии, учиться на 4 и 5 в первые четыре семестра, устроиться на работу с зарплатой более 30 000 рублей в течение первого года после выпуска, стать кандидатом наук, получить грант на три года после защиты диссертации, какой должна быть зарплата доцента, чтобы 50 % выпускников работали по специальности и т. п.

Программная реализация разработанных моделей и алгоритмов интеллектуального анализа образовательных данных для поддержки принятия решений была внедрена на кафедре интеллектуальных технологий в гуманитарной сфере Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики при разработке основной образовательной программы высшего профессионального образования.

В конце 2018 года на базе РАНХиГС был создан центр по работе с большими данными в системе образования, позволяющий объединить возможности университета и Рособрандзора.

Можно отметить, что в системе высшего образования России использование технологий больших данных находится на стадии одиночных проектов, в зарубежных странах — на этапе перехода от локальных экспериментов к полноценному внедрению.

Анализ научных и публицистических работ позволяет выделить риски использования в образовательном процессе больших данных.

Во-первых, это проблема конфиденциальности, так как часть данных содержит личную информацию.

Во-вторых, прогнозы, основанные на анализе больших данных, могут быть использованы как инструмент влияния и повлечь за собой наказание или поощрение студентов и преподавателей за несовершенные действия.

Но все же очевидным становится то, что сегодня Big Data используются в университетах, которые стремятся улучшить свои стратегические и тактические технологии принятия решений для конструирования образовательных процессов, адекватных будущим потребностям высокоразвитого общества.

Библиографический список

1. Аbruков В. С., Кожин А. Ю., Троешестова Д. А., Петрова М. В., Ануфриева Д. А. Возможности методов интеллектуального анализа данных при моделировании образовательного процесса в вузе и разработке систем поддержки принятия решений // *Современные информационные технологии и ИТ-образование*. 2013. № 9. С. 682—688.
2. Бебенина Е. В., Елкин О. М. Повышение качества управления образованием с использованием технологии обработки больших данных // *Отечественная и зарубежная педагогика*. 2020. № 6 (72). С. 22—29.
3. Волобуева Т. Б. Педагогический форсайт: большие данные // *Педагогическая перспектива*. 2021. № 1. С. 15—22.
4. Мамедова Г. А., Зейналова Л. А., Меликова Р. Т. Технологии больших данных в электронном образовании // *Открытое образование*. 2017. № 6. С. 41—48.
5. Boyd D., Crawford K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon // *Information, Communication & Society*. 2012. Vol. 15, № 5. P. 660—679.
6. Liebowitz J. Thoughts on recent trends and future research perspectives in big data and analytics in higher education // *Big data and learning analytics in higher education: Current theory and practice*. January, 2016. P. 7—17.
7. Lynch C. Big data: how do your data grow? // *Nature*. 2008. Vol. 455. № 7209. P. 28—29.

References

- Abrukov, V. S., Kozhin, A. Yu., Troeshestova, D. A., Petrova, M. V., Anufrieva, D. A. (2013) *Vozmozhnosti metodov intellektual'nogo analiza dannykh pri modelirovanii obrazovatel'nogo protsessa v vuze i razrabotke sistem podderzhki prinyatiya reshenii* [Possibilities of data mining methods in modeling the educational process at a university and developing decision support systems], *Sovremennye informatsionnye tekhnologii i IT-obrazovanie* [Modern information technology and IT education], no. 9, pp. 682—688.
- Bebenina, E. V., Elkin, O. M. (2020) *Povyshenie kachestva upravleniya obrazovaniem s ispol'zovaniem tekhnologii obrabotki bol'shikh dannykh* [Improving the quality of education management using big data processing technology], *Otechestvennaya i zarubezhnaya pedagogika* [National and foreign pedagogy], no. 6 (72), pp. 22—29.
- Volobueva, T. B. (2021) *Pedagogicheskii forsait: bol'shie dannye* [Pedagogical foresight: Big Data], *Pedagogicheskaya perspektiva* [Pedagogical perspective], no. 1, pp. 15—22.
- Mamedova, G. A., Zeinalova, L. A., Melikova, R. T. (2017) *Tekhnologii bol'shikh dannykh v elektronnom obrazovanii* [Technologies of big data in e-learning], *Otkrytoe obrazovanie* [Open education], no. 6, pp. 41—48.
- Boyd, D., Crawford, K. (2012) *Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon*, *Information, Communication & Society*, vol. 15, no. 5, pp. 660—679.

Liebowitz, J. (2016) Thoughts on recent trends and future research perspectives in big data and analytics in higher education, *Big data and learning analytics in higher education: Current theory and practice*, January, pp. 7—17.

Lynch, C. (2008) Big data: how do your data grow?, *Nature*, vol. 455, no. 7209, pp. 28—29.

Статья поступила в редакцию 10.12.2021.

Сведения об авторах

Огурцова Елена Юрьевна — кандидат педагогических наук, доцент, Ивановский государственный университет (Шуйский филиал), г. Шуя, Россия, ogurtsova-elena@mail.ru

Фадеев Роман Николаевич — студент, Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых, г. Владимир, Россия, fadeevroman.shua@gmail.com

Information about the authors

Ogurtsova Elena Yurievna — Cand. Sc. (Pedagogical), Associate Professor, Ivanovo State University (Shuya branch), Shuya, Russian Federation, ogurtsova-elena@mail.ru

Fadeev Roman Nikolaevich — student, Vladimir State University named after Alexander Grigorievich and Nikolai Grigorievich Stoletovs, Vladimir, Russian Federation, fadeevroman.shua@gmail.com