

УДК 004.048

БОЛЬШИЕ ДАННЫЕ В СФЕРЕ ОБРАЗОВАНИЯ, ВОЗМОЖНОСТИ И ПРОБЛЕМАТИКА

Андреев С. В.

Магистр,

Национальный исследовательский ядерный университет «МИФИ»,

Москва, Россия

Фазульянов Д. В.

Аспирант,

Национальный исследовательский ядерный университет «МИФИ»,

Москва, Россия

Аннотация

Статья посвящена актуальной теме применения технологии больших данных (Big Data) в образовании, с помощью которой возможно провести анализа этих данных и повысить качество образования его ускорение, а также развитие новых сервисов и подходов к обучению.

Целью исследования является анализ областей применения Big Data в образовательном процессе для повышения качества и эффективности обучения. В результате проведенного анализа выявлены основные области применения технологии Big Data в образовании, а также обозначены проблемы, связанные с ее использованием.

Результаты исследования могут быть полезны для разработки новых образовательных сервисов и улучшения качества образовательного процесса.

Ключевые слова: Big Data, образование, перспективы, проблематика, анализ образовательных данных, область применения, Educational Data Mining

BIG DATA IN EDUCATION, OPPORTUNITIES AND ISSUES

Andreev S.V.

Undergraduate,

National Research Nuclear University "MEPhI",

Moscow, Russia

Fazulianov D.V.

Post-graduate student,

National Research Nuclear University "MEPhI",

Moscow, Russia

Abstract

The article is dedicated to the topical issue of applying Big Data technology in education, which allows for the analysis of this data and the improvement of education quality, its acceleration, as well as the development of new services and approaches to learning.

The aim of the research is to analyze the areas of Big Data application in the educational process and evaluate its impact on the quality and effectiveness of learning.

As a result of the conducted analysis, the main areas of Big Data technology application in education were identified, and the problems associated with its use were outlined.

The results of the research can be useful for the development of new educational services and the improvement of the quality of the educational process.

Keywords: Big Data, education, prospects, problems, analysis of educational data, scope, Educational Data Mining

На сегодняшний момент информационные системы очень плотно вошли в нашу жизнь и крайне сложно найти человека, который не сталкивался с ними с той или иной стороны. Информационные технологии помогают нам в

повседневной жизни, управляют нашим домом, помогают нам общаться между собой, узнавать и делиться информацией, информационные технологии занимают особое место в самых разных отраслях и даже сами по себе создают целые отрасли. Не обошли информационные технологии и образование. Если по началу информационные технологии были интересны как объект изучения, то на сегодняшний день они плотно вошли в область образования и помогают планировать и обеспечивать учебные процессы, коммуницировать преподавателю и обучающемуся, находить информацию, проводить исследования и многое другое.

Определение информационных технологий (совокупность методов, производственных и программно-технологических средств, объединенных в технологическую цепочку, обеспечивающую сбор, хранение, обработку, вывод и распространение информации) говорит нам о том, что информационные технологии работают прежде всего с информацией, помогая нам в ее хранении и обработке. Само понятие информация произошла от лат. «informātiō» и означает разъяснение, представление, понятие о чём-либо, то есть это сведения об окружающем нас мире, которые являются объектом хранения, накопления, обработки, преобразования, передачи и использования. И с развитием информационных технологий этих объектов становится все больше и их создание подчас даже не заметно для нас. С развитием мобильных устройств, когда у нас в кармане, по сути своей, стал лежать не просто мобильный телефон, а полноценный компьютер, живущий «собственной жизнью» и порождение данных и информации стало происходить даже без нашего участия [7].

Стоит отметить, что «информация» и «данные» хотя и очень близкие по своей сути определения, все же между ними есть различия. Так, «данные» сами по себе могут не нести какой-либо ценности, так как они дискретны и лишь говорят нам о каком-либо факте, событии или значении, так как они лишены контекста, а данные связанные с контекстом, смысл и значение которых понятен, уже превращаются в информацию. Например, если мы видим цифру «5» это Вектор экономики | www.vectoreconomy.ru | СМИ Эл № ФС 77-66790, ISSN 2500-3666

мало, о чем нам может сказать, вот он наш пример данных. А когда цифра 5 размещена на дисплее термометра, замеряющего температур на улице, мы уже получаем информацию, так как уже понимаем, что на улице прохладно и надо одеваться по погоде, у нас появился контекст и данные стали информацией [7].

Как уже стало понятно, данных нас окружает огромное количество, но для того, чтобы они имели ценность, они должны превратиться для нас в информацию, то есть необходимо их собрать, обработать и интерпретировать и прежде всего такая задача встала перед предприятиями.

На просторах интернета можно найти весьма показательный пример: Walmart проведя анализ накопившихся у них данных о продажах заметили, что мужчины 30-40 лет по вечерам в пятницу часто покупают подгузники и пиво [8]. На первый взгляд странное сочетание, но оказывается все логично: у мужчин в этом возрасте скорее всего есть маленькие дети и в конец недели подгузники заканчиваются, а посещая магазин вечером в пятницу, когда впереди выходные, он приобретает эти, казалось бы, не сочетаемые товары. В магазинах поставили стойку подгузников рядом со стойкой пива и продажи этих товаров выросли на 30%. Данный пример показывает нам, что благодаря анализу данных можно выявлять различные закономерности и причины событий и прогнозировать их последствия.

Сам по себе термин «большие данные» или Big Data появился в 2008 году, благодаря редактору журнала Nature Клиффорд Линч подготовивший выпуск с темой «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объемами данных», в котором рассматривались материалы о набравших колоссальные темпы роста, объема и многообразия данных и перспективах развития технологий с ними связанных. И если поначалу термин big data использовался преимущественно в научной среде и статистике, то с 2014 года сбором и анализом данных занялись ведущие вузы мира и IT-компании: IBM, Google, Microsoft.

Источников данных огромное количество - это могут быть данные из внутренних информационных систем предприятия, сообщения из социальных сетей, данные о заказах и посетителях как вашего сайта и других сайтов, содержимое веб-страниц и мобильных приложений, данные с различных сенсоров и прочие данные которая есть в мире. Основная задача после того как они получены – это их приведение в необходимый нам вид и правильная их интерпретация [1].

Анализ больших данных, в настоящее время, является крайне актуальной технологией, так как она не ограничивается какой-то узкой сферой применения, а применима в совершенно разных отраслях. Не случайно, в нашей стране большие данные вошли в Национальные технологические инициативы (НТИ) сквозные технологии, определенные как ключевые научно-технические направления, оказывающие существенное влияние на развитие рынков. Иными словами, к сквозным технологиям относятся технологии, которые охватывают несколько трендов или отраслей.

В рамках федерального проекта «Цифровые технологии» были разработаны дорожные карты по технологиям, куда вошли «большие данные» наравне с искусственным интеллектом, робототехникой, систем распределенного реестра, квантовых технологий, новых производственных технологий, промышленного интернета, беспроводной связи, виртуальной и дополненной реальностей. Прогнозируется, что внедрение технологий больших данных наибольшее влияние окажет на информационные технологии в производстве, здравоохранении, торговле, государственном управлении и т.д. [2]

Big Data представляет из себя не просто огромные массивы различных данных, а объединяет в свое понятие инструменты и методы обработки структурированных и неструктурированных данных огромных объемов и многообразия в целях получения результатов, достаточных для интерпретации и принятия управленческих решений.

В качестве характеристик для больших данных обычно выделяют так называемые «пять V»:

- **Volume (объем)** – речь идет о величине физического объема данных. Но эта величина во многом зависит от времени, точнее о том, что понятие «много данных» для разного времени отлично. Например, 30 лет назад объем жесткого диска в 10 Гб считался вполне приличным и на него помещалось много данных, в настоящее же время объем постоянной памяти компьютеров измеряется уже терабайтами.

- **Velocity (скорость)** – речь идет как про скорости прироста данных, так и про потребность в скоростной обработке данных и их интерпретации. Важно отметить что прирост наблюдается не только в количестве данных, но и в источниках данных.

- **Variety (многообразие)** – данные многообразны и их обработка должна вестись одновременно, причем как структурированных, так и неструктурированных (разнообразие).

- **Veracity (достоверность)** — это концепция отбора действительно важных и достоверных данных.

- **Value (ценность)** — ключевой параметр для оценки эффективности вложенный в обработку больших данных, так как анализ и обработка данных должна представлять собой определенную ценность [9].

Основные принципы работы с Big Data, можно сформулировать так:

- **Горизонтальная масштабируемость** в виду того, что данных много и для их обработки необходимы ресурсы, которые будут доступны при необходимости и вертикальное масштабирование в этом случае проигрывает.

- **Отказоустойчивость** одно из необходимых условий работы с большим данными. В виду того, что машин в кластере может быть много, в процессе работы возможен выход из строя некоторых из них, поэтому архитектура должна быть отказоустойчива к подобным событиям.

- Локальность данных в случае фактического хранения данных на одном сервере, а обработке на другом возникают расходы на передачу данных, поэтому при проектировании BigData-решений стараются локализовать данные и вести их на том же ресурсе, где они и хранятся. В распределённых системах данные так же распределены по машинам.

Методов и техник анализа больших данных существует достаточно большое количество вот некоторые из них:

- методы интеллектуального анализа данных: обучение ассоциативным правилам, классификация, кластерный и регрессионный анализ;
- краудсорсинг способ обогащения и категоризация данных при привлечении неопределённого круга лиц;
- смешение и интеграция данных — набор техник, позволяющих интегрировать разнородные данные из разнообразных источников для возможности глубинного анализа;
- машинное обучение – решение задач или получение прогнозов на основе накопленных статистических данных или машинного обучения;
- искусственные нейронные сети, сетевой анализ, оптимизация, в том числе генетические алгоритмы;
- распознавание образов;
- прогнозная аналитика;
- имитационное моделирование.

Существуют также и аппаратные решения, представляющие из себя готовое решение для работы с большими данными. Например, подобным аппаратным решением для скоростных вычислений в оперативной памяти является Hana от SAP и Exalytics от Oracle.

Целью данного исследования является анализ областей применения Big Data в образовательном процессе для повышения качества и эффективности данной отрасли с выделением перспективных направлений и технологий.

Прогресс так или иначе затрагивает различные сферы, и сфера образования не является исключением. Уже давно стали привычными компьютеры на занятиях не только в ВУЗах, но и в школах, которые в свою очередь уже отошли от некогда привычного дневника для записи домашнего задания.

Как и многие другие отрасли, образование имеет свои, только для него характерные процессы и процедуры что и делает его уникальным. Существуют ряд информационных систем, преимущественное распространение которых наблюдается именно в образовании, речь, например, идет про так называемые LMS (learning management system) – это специализированные продукты, помогающие автоматизировать процесс обучения и его администрирования. Данные программные продукты не только автоматизируют процесс обучения, но также позволяют проводить анализ некоторых показателей.

Основная цель для образовательного учреждения – это передача знаний учащимся и совершенно не важно о какой уровне образования идет речь, меняются лишь методы обучения и подходы, но цель остается неизменной. Информационные системы в образовательных учреждениях помимо оптимизирующих функций, а именно уменьшении каких-то ручных операций, повышения скорости и надежности так же должны и выполнять роль измерительного прибора.

Особый скачек у информационных систем образования случился в период мировой пандемии COVID19 2019-2022 года, когда ранее консервативная отрасль была вынуждена искать новые способы коммуникации со своими учащимися и находить новые инструменты для взаимодействия. Даже школы были переведены на электронные дневники, а многие учителя освоили работу

видео-конференц связи, увеличив таким образом уровень автоматизации данной отрасли [5].

В настоящее время, если смотреть на функциональную принадлежность информационных систем, применяемых в образовательных учреждениях можно выделить как системы, непосредственно занятые в образовательном процессе - это может быть и какая-то LMS-система, так и всевозможные платформы для коммуникаций и совместных работ, так информационно - учетные системы, необходимые для осуществления управленческих, финансово-хозяйственных, аналитической и прочей деятельности необходимые для обеспечения образовательного процесса.

Неправильно будет говорить о том, что анализ анализу образовательных данных не уделяется значение – это далеко не так, более того анализ образовательных данных (Educational Data Mining) выделился в отдельное направление исследований, связанное с применением методов интеллектуального анализа данных и машинного обучения [6].

Анализ данных в образовательных организациях прежде всего пытается извлечь закономерности в процессах обучения, оценить взаимосвязь итоговой оценки от применяемых подходов в обучении, понять, как сделать процесс обучения более эффективным [4].

Полученная обучающимся оценка хотя и отражает некий результат, но все же не дает нам полноты картины относительно качества выстроенного нами процесса. Без условно на помощь в это нам приходят разнообразные LMS-системы, которые могут фиксировать большое количества различных данных, таких как: длительность изучения курса, количество обращений, скорость прохождения различных тестов, последовательность их выполнения и т.д. Цифровизации обучения все более способствует тому, что данных становится все больше и они становятся все более разнообразнее, открывая возможность для исследования.

В результате применения подходов и методов Big Data в образовательной отрасли позволит оптимизировать уже образовательные подходы, выявить причины имеющихся затруднений у обучающихся, выработать рекомендации как для обучающегося, так и для преподавателя по процессу обучения и применив новые подходы кардинально повысить качество обучения и перейти на новый уровень.

К примеру, изучив статистику обращений студентов к учебным материалам курса и полученную итоговую оценку возможно создать инструмент прогнозирования, который предсказывал обучающемуся на основе анализа его текущего поведения его потенциальную оценку.

Безусловно, цифровизация дала нам богатые возможности для анализа показателей обучения, но в виду большей формализации и необходимости использования информационных систем — это достаточно просто сделать при дистанционном обучении, а в случае традиционного обучения уже возникают сложности в виду сложности получения данных. Но все же способы получения данных для анализа существуют – это могут быть как предоставление учебных материалов и частичное выполнение заданий в информационной среде, так и использование камер видеонаблюдения для оценки учебного процесса. В этой ситуации на помощь может прийти компьютерное зрение. Анализируя данные полученные с камер, установленных в учебном классе можно оценить посещаемость занятий, вовлеченность обучающихся в процесс, концентрацию на учебном материале.

Но большую популярность компьютерное зрение приобрело в качестве инструмента контроля при проведении уделенного экзамена (проктеринга) и тут система может как осуществлять элементарные функции контроля наличия экзаменуемого в поле зрения камеры, так и контролировать направление взгляда, фиксировать посторонних в поле зрения камеры, а также замерять необходимые временные показатели.

Анализ данных так же может способствовать получению обратной связи от обучающихся, что очень актуально для преподавателей, работающих удаленно, так как взаимодействие с аудиторией в этом случае крайне слабое. Источником данных для этого могут служить как те же логи поведения в системах дистанционного обучения, результаты опросов, тестов, анализ текстов работ, обучающихся и т.д. [4].

Анализ данных полученных в процессе обучения может способствовать выявлению предпочтений того или иного обучающегося и выработки рекомендаций для его программы обучения, и предоставления для него рекомендуемых источников информации [4]. В качестве аналога можно привести алгоритмы работы рекомендаций в соцсетях, когда на основе данных о наших запросах, просмотрах, подписок и т.д. формируются рекомендации, которые могли быть нам интересны. Возможно применение аналогичных подходов, но с акцентом на образовательную траекторию.

Анализ данных об успеваемости предыдущих лет может способствовать построению прогноза потенциальной оценки обучающегося в зависимости от его вовлеченности в процесс, скорость и качество выполнения заданий, посещаемость и иных факторах, коррелирующих с итоговой оценкой.

Не стоит забывать и о возможности реактивного устранения проблем. Очень часто о проблемах в обучении становится известно уже после того, как обучающийся окончательно отстал, его мотивация находится на нуле. Особенной актуально это для иностранных студентов у которых помимо учебной нагрузки еще накладывается проблема адаптации в незнакомой языковой и культурной среде и очень часто они просто не говорят о своих проблемах, да и не знают кому именно они могут о них сообщить. Своевременное выявление данных негативных динамик позволит вовремя принять меры как на уровне администрации, так и на уровне системы, которая сможет дать рекомендации относительно той или иной проблемной ситуации у обучающегося [3].

В целом, опыт работы социальных сетей весьма полезен в организации процессов образования, в частности выделение схожих тематик научных работ и интересов студентов и преподавателей возможна выработка рекомендаций по созданию эффективных команд.

Существуют так же и проблемы на пути распространения методов Big data в образовании и прежде всего это непонимание того, какие данные и откуда надо собирать. Можно конечно собирать все что возможно, но тут встает другая проблема – обеспечение хранения этих данных и их обработка, что весьма затратно.

Накладывает свои ограничения и необходимость обеспечения сохранности персональных данных, а также соблюдения этических норм при сборе данных. Необходимо поставить в известность о сборе подобных данных участников процесса, а об обработке персональных данных подписать соглашение.

Основные проблемы применения Big Data тесно связаны с их характеристиками, а именно:

- Volume (объём) – многообразие и данных ставит вопрос о необходимости их хранения, что требует, как грамотного подхода к архитектуре применяемого решения, так и к банальному месту, где эти данные будут храниться.
- Velocity (скорость) – постоянный прирост данных, их источников и требование в быстродействии их обработке и их интерпретации также ставить перед нами задачу обеспечения вычислительных ресурсов.
- Variety (многообразие) – многообразие данных также накладывает свои проблемы при их обработке.

К проблематике можно также отнести и необходимость привлечения компетентных сотрудников для данной работы и если, в качестве научных экспериментов, работа с большими данными может вестись в организациях высшего учебного заведения, то школы такого себе позволить не смогут и им придётся использовать уже готовые решения. Поэтому данные инструменты Вектор экономики | www.vectoreconomy.ru | СМИ Эл № ФС 77-66790, ISSN 2500-3666

должны быть просты в использовании и понятны не слишком подготовленному пользователю [3].

Нельзя забывать и о достоверности данных и их корректности. Очень часто мы можем получать некорректные данные из различных источников, но для их дальнейшего использования нам потребуется дополнительная обработка или верификация.

В заключение можно сказать, что несмотря на имеющиеся проблемы, применение подходов и методов Big Data в образовательной отрасли открывает новые возможности для оптимизации образовательных подходов, выявлении причин затруднений у обучающихся, выработки рекомендаций для обучающегося и преподавателя, повышения качества обучения и перехода на новый уровень. Кроме того, анализ данных может способствовать получению обратной связи от обучающихся, выявлению их предпочтений и выработке рекомендаций для их программы обучения, а также построению прогноза потенциальной оценки обучающегося и предупреждения проблем еще на стадии их зарождения.

Библиографический список

1. Пискорская С.Ю., Гончаров А.Е. Феномен больших данных в социальной философии и профессиональном образовании // Профессиональное образование в современном мире. 2018. № 4. С. 2178-2185.
2. Паспорт федерального проекта "Цифровые технологии" [Электронный ресурс]. URL: https://www.consultant.ru/document/cons_doc_LAW_328937/b000f428c119fc014df6c3a6bc6fd10b188b115b/ (дата обращения: 28.12.2023).
3. Минязова Е.Р., Подымова Л.С. «Большие данные» и персонализированное обучение // Высшее образование сегодня. 2022. № 5-6. С. 41-45.

4. Белоножко П.П., Карпенко А.П., Храмов Д.А. Анализ образовательных данных: направления и перспективы применения // Интернет-журнал «НАУКОВЕДЕНИЕ». 2017. Том 9, №4.— С. 2-6.
5. Морозов А.В., Небродовская-Мазур Е.Ю., Матвеева И.П. Цифровая образовательная среда в период пандемии COVID-19: реалии и перспективы // Телескоп. 2022. №1. С. 30.
6. Веряев А.А., Татарникова Г.В. EDUCATIONAL DATA MINING и LEARNING ANALYTICS - направления развития образовательной квалитологии // Преподаватель XXI в. 2016. № 1(ч.2).— С. 151
7. Муромцев Д.Ю., Тюрин И.В., Белоусов О.А., Курносков Р.Ю. Информационные технологии проектирования радиоэлектронных средств: учебное пособие для вузов. 3-е изд., стер. Санкт-Петербург: Лань, 2024.— С. 20.
8. Митина О.А., Жаров В.В. Технологии и инструментарий машинного обучения: учебное пособие. - Москва: РТУ МИРЭА, 2023. — С. 35.
9. Колмогорова С.С. Обработка данных алгоритмами искусственного интеллекта в системе интернета вещей. - Санкт-Петербург: Лань, 2023.— С. 11.

Оригинальность 81%