

## Большие данные как новый статистический подход в развитии науки и образования

Джавадов Рамиз Джавад оглы

Азербайджанский Государственный Экономический Университет, г. Баку, Азербайджанская Республика

Для цитирования: Джавадов Р.Д. Большие данные как новый статистический подход в развитии науки и образования. Вопросы статистики. 2023;30(6):74–76.

## Big Data as a New Statistical Approach in the Development of Science and Education

Ramiz J. Javadov

Azerbaijan State Economic University (UNEC), Baku, Republic of Azerbaijan

For citation: Javadov R.J. Big Data as a New Statistical Approach in the Development of Science and Education. *Voprosy Statistiki*. 2023;30(6):74–76. (In Russ.)

### Статистический анализ и большие данные.

В статистических исследованиях чаще всего применяют выборочный метод, но при этом одна из проблем статистики — недостоверные результаты при маленьких выборках. Повысить достоверность статистики помогают большие данные.

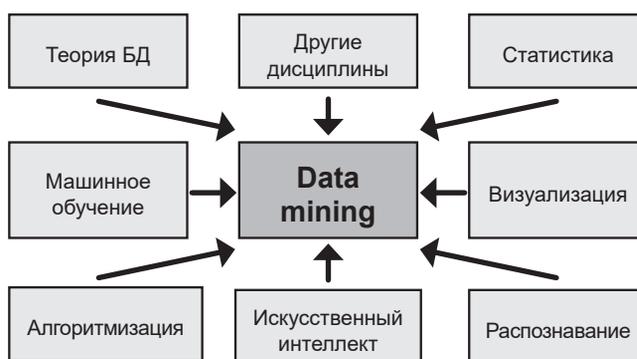
Мы живем в эпоху больших данных (*big data*). Скорость и объемы ежедневно производимых и собираемых данных в различных ресурсах во всем мире создают огромный потенциал для их использования в науке.

В науке о данных (*data science*) большие данные обрабатывают и анализируют, для того чтобы найти в них закономерности и использовать их для повышения эффективности компании или системы. Например, анализ больших данных в социальных сетях помогает делать рекламу релевантной для пользователей. На практике в работе с большими данными одним из ее этапов является интеллектуальный анализ данных (*data mining*), или, другими словами, поиск закономерностей методами интеллектуального анализа данных.

**Интеллектуальный анализ данных: поиск закономерностей.** *Data mining* применяется везде, где необходимо в больших данных найти какие-то тенденции и закономерности. Решение большинства задач компании, связанных с данными, сводится к той или иной задаче *data mining* или их комбинации.

Например, можно оценить риски, используя регрессионный анализ, сегментировать инфор-

мацию о покупателях с помощью метода кластеризации, предсказать спрос путем выявления ассоциаций в данных и так далее. Таким образом, *data mining* объединяет разные методики и технологии работы с данными, что проиллюстрировано на рисунке ниже.



**Использование больших данных в официальной статистике.** Большие данные — это изначально не систематизированные данные о социально-экономическом развитии и окружающей среде, генерируемые на основе информационно-коммуникационных технологий за пределами официальной статистики.

Главными источниками больших данных можно считать:

- интернет вещей (*IoT*) и подключенные к нему устройства;
- социальные сети, блоги и СМИ;
- данные компаний (транзакции, заказы товаров и услуг, поездки на такси, профили клиентов);

– показания приборов (метеорологические станции, измерители состава воздуха и водоемов, данные со спутников);

– статистика городов и государств (данные о перемещениях, рождаемости и смертности населения);

– медицинские данные (анализы, заболевания, диагностические снимки).

Уже несколько лет в мире серьезно рассматривается проблема взаимодействия официальной статистики и владельцев больших данных. В числе основных ведущих мировых центров (основных игроков) выделим следующие:

– Статистическую комиссию ООН;

– Глобальную рабочую группу по большим данным в официальной статистике;

– проект ООН «Глобальный пульс».

С помощью перечисленных ниже характеристик можно определить стратегическую ценность больших данных для официальной статистики:

– огромный массив данных может способствовать получению более детализированных сведений по конкретным направлениям статистики;

– высокая скорость обновления информации приводит к увеличению частоты статистических оценок;

– разнообразие больших данных открывает возможности для производства статистики в новых областях, получения информации о сферах, не наблюдаемых официальной статистикой;

– разнородность происхождения больших данных обеспечивает многоаспектность измерений, повышая тем самым надежность статистических оценок.

### **Мировой опыт применения больших данных.**

В современном мире развитые страны широко используют большие данные. Приведем некоторые примеры:

– Австралия – аналитика *big data* используется для устранения последствий чрезвычайных ситуаций;

– Испания – применяется сервис «Умный Мадрид»;

– Канада – большие данные используются для развития здравоохранения;

– Россия – Единое информационное пространство и технология доступа к налоговым данным;

– Япония – с апреля 2016 г. Министерство внутренних дел Японии применяет большие данные для своевременного предоставления информации об эвакуации во время стихийных бедствий;

– Китай – создание смарт-городов;

– Азербайджан – создание «умного» села.

В Зангиланском районе, в Восточном Зангезуре, создано «умное село» – Агалы, которое можно считать первой в мире моделью такого рода. «Умное» село – это общая концепция интеграции нескольких информационных и коммуникационных технологий для управления имуществом, включая, в частности, школы, библиотеки, транспорт, больницы, электростанции, системы водоснабжения и управления отходами. На основе совокупности всей этой информации о населенном пункте повышается эффективность расходов и улучшается качество жизни населения.

**Использование больших данных в науке.** Обработка и анализ больших данных могут предоставить ценную информацию для научных исследований и позволить ученым получить новые знания и лучше понять окружающий нас мир.

Большие данные могут быть использованы в различных областях науки, включая медицину, экономику, экологию, сельское хозяйство, биологию и т. д. Приведем несколько примеров использования больших данных в разных сферах:

1. Медицина: использование данных о здоровье пациентов, включая медицинские записи, лабораторные исследования, данные об их весе и физической активности, для разработки новых лекарств и методов лечения различных заболеваний.

2. Экономика: снижение стоимости продукции, оптимизация производства, оценка платежеспособности населения и т. д. Например, *Mastercard* работает не только как платежная система – она собирает данные, которые помогают выявлять неплатежеспособных контрагентов, не возвращающих кредиты. *Mastercard* предупреждает финансовые организации о том, что с ними не стоит вести дела.

3. Сельское хозяйство: *IoT*-решение из области так называемого точного земледелия, когда специальные метеостанции, которые находятся в полях, с помощью сенсоров собирают данные (о температуре, влажности) и с помощью передающих радио-GSM модулей отправляют их на *IoT*-платформу.

4. Биология: использование данных о геномах и белковых структурах для изучения биологических процессов, таких как развитие болезней и размножение, что позволяет разрабатывать новые методы диагностики и лечения.

**Большие данные в образовании.** Образовательная система непрерывно создает и накапливает значительный объем данных, которые трудно собрать и оценить с помощью традиционных методов и программных технологий.

Аналитика больших данных позволяет повысить эффективность обучения учеников. Учитель ежедневно получает все виды данных об учениках, такие как сведения об их посещаемости, результатах экзаменов, личных оценках, здоровье, проблемах и трудностях в обучении, типах вопросов, которые они часто задают, и другие виды информации.

В области высшего и профессионального образования проекты *big data* применяются для профессиональной ориентации студентов – анализа их способностей и оказания им помощи в выборе направления обучения и будущей профессии. Например, изучив опыт американского Университета Остина Пи (Austin Peay State University, APSU) в Азербайджанском государственном экономическом университете (АзГЭУ) разработали рекомендательную систему подбора курсов. Она собирает данные об успеваемости, находит «похожих» студентов и на основе этого подбирает образовательные курсы для конкретного человека.

**Проблемы и перспективы использования больших данных.** В официальной статистике использование больших данных сопряжено с решением таких проблем, как обеспечение конфиденциальности, финансирование работ, а также наличие специалистов и аналитиков.

Так, одной из основных проблем применения технологий больших данных для целей официальной статистики является нехватка специалистов и аналитиков в этой области. Учитывая быстрый рост цифровизации и больших данных в мире, подготовка статистиков и аналитиков, которые способны работать в этой сфере, должна стать приоритетной сферой на государственном уровне.

Для подготовки статистиков-аналитиков, которые способны работать с большими данными, было бы целесообразно усилить взаимосвязи между Государственным комитетом по статистике Азербайджана (ГКС) и выпускниками АзГЭУ по специальности «Статистика». Для этого необходимо:

- создать совместные учебные аудитории и лаборатории;
- сформировать соответствующие условия для участия специалистов ГКС в учебном процессе со студентами;
- организовать студенческую производственную практику, которая бы отвечала требованиям современного образования;
- начиная со второго курса (после изучения основ статистики) привлекать студентов в процесс сбора и анализа данных;
- вовлекать студентов в процесс проведения выборочных обследований различных социально-экономических явлений;
- открыть по крайней мере в одном из вузов Азербайджана факультет, обеспечивающий современную подготовку специалистов в области *data science*;
- создать центр развития *data science* на государственном уровне и т. п.

Практика показывает, что на сегодняшний день технологии больших данных развиваются в мире очень быстро. Это требование времени, и в следующем десятилетии динамика будет расти.

Большие данные помогают снизить затраты на статистическое производство на государственном уровне. Например, в 2019 г. в Германии в бюджет было возвращено около 15 млрд евро, после того как на основе анализа транзакций было выявлено, что часть граждан получают пособия по безработице безо всяких на то оснований.

В ближайшем будущем большие данные станут главным инструментом для принятия решений – начиная с сетевых бизнесов и заканчивая целыми государствами и международными организациями.

### Информация об авторе

*Джавадов Рамиз Джавад оглы* – PhD (экон.), доцент, доцент кафедры экономики, Азербайджанский государственный экономический университет. Республика Азербайджан, AZ1001, г. Баку, ул. Истиглалият, д. 6. E-mail: ramizjavad@gmail.com. ORCID: <https://orcid.org/0000-0002-0666-4022>.

### About the author

*Ramiz J. Javadov* – PhD (Econ.), Associate Professor; Associate Professor of the Department of Economics, Azerbaijan State Economic University (UNEC). 6, Istiglaliyyat Str., Baku, AZ1001, Republic of Azerbaijan. E-mail: ramizjavad@gmail.com. ORCID: <https://orcid.org/0000-0002-0666-4022>.