

Гусейнов Магомед Мурадович
Московская международная академия

Использование технологий big data при финансовом анализе деятельности предприятия

Аннотация. В работе предлагается расширенное определение «больших данных» как междисциплинарной и относительной категории, объединяющей свойства массивов и потоков данных, требования к их обработке и критерии практической ценности результатов. Показано, что эволюция подходов от триады «объём–скорость–разнообразие» к включению аспектов качества, изменчивости и воспроизводимости обусловлена усложнением источников (сенсоры, транзакции, мультимедиа) и ростом институциональных требований. Обосновано, что статус «больших данных» определяется соответствием архитектуры хранения и вычислений контекстным ограничениями по пропускной способности, латентности и надёжности, а не абсолютным размером наборов. Предложена трёхкомпонентная рамка определения (свойства данных, требования к обработке, критерии полезности) и методология из трёх этапов (единая модель данных, конвейеры обработки, моделирование и верификация). Отдельно рассмотрены вопросы качества и этико правового соответствия: происхождение данных, контроль смещений, валидация моделей и документирование преобразований. Сделан вывод, что институционально оформленные и технологически поддержанные контуры данных создают проверяемый прирост управлеченческой ценности и основу долгосрочной устойчивости финансовых решений.

Ключевые слова: большие данные, финансовый анализ, управление активами, клиентская аналитика, потоковая обработка, качество данных, регуляторные требования.

Guseinov Magomed Muradovich
Moscow International Academy

Using big data technology in financial analysis of enterprise activities

Annotation. The paper proposes an extended definition of "big data" as an interdisciplinary and relative category that combines the properties of data arrays and streams, the requirements for their processing, and the criteria for the practical value of the results. It is shown that the evolution of approaches from the "volume-speed-variety" triad to the inclusion of aspects of quality, variability, and reproducibility is driven by the complexity of sources (sensors, transactions, and multimedia) and the growing institutional requirements. For the financial sector of Russia, three key blocks of application are described: asset management, risk management and work with the client base, where the integration of heterogeneous sources, streaming and batch processing, standardization and verification of indicators provide an increase in the accuracy of forecasts, a reduction in response time and an improvement in the client experience. The issues of using the technology of "big data" in the financial analysis of the enterprise activity are considered. A three-component definition framework (data properties, processing requirements, and utility criteria) and a three-stage methodology (unified data model, processing pipelines, and modeling and verification) are proposed. The article also addresses quality and ethical considerations, including data origin, bias control, model validation, and documentation of transformations. The authors conclude that institutionalized and technologically supported data frameworks create verifiable management value and a foundation for long-term sustainability in financial decision-making.

Keywords: big data, financial analysis, asset management, customer analytics, streaming processing, data quality, and regulatory requirements.

Понятие BIG DATA «большие данные» целесообразно рассматривать как междисциплинарную категорию, описывающую массивы и потоки данных, обработка которых традиционными методами становится затруднительной из-за сочетания масштабов, скорости обновления, разнородности структуры и требований к достоверности и ценности результатов. В научной литературе первоначально акцент делался на совокупности характеристик объёма, быстроты и разнообразия, что позволило сформировать операциональный подход к идентификации явления. Позднее к ним были добавлены аспекты качества, изменчивости и экономической полезности, отражающие связь данных с управленческими решениями и исследованиями [1].

Такой переход от минимального набора признаков к расширенной системе критериев объясняется усложнением источников данных (сенсорные сети, транзакционные логи, мультимедийные потоки) и институциональными требованиями к воспроизводимости результатов.

В методологическом плане «большие данные» не тождественны простой аккумуляции записей. Структурообразующим признаётся сочетание технических параметров и контекстуальных ограничений: например, пределы пропускной способности коммуникаций, асимметрия доступов к реестрам, а также нормативные рамки обработки персональной информации. Тем самым определение опирается на относительный критерий: одни и те же массивы могут не считаться «большими» при наличии адекватной архитектуры хранения и вычислений, но приобретать свойства «больших» при дефиците ресурсов или росте требований к латентности и воспроизводимости. На это указывают как зарубежные, так и российские исследования, рассматривающие соотношение технологической готовности и статистических процедур [1, 2, 3].

Содержательно «большие данные» характеризуются многоуровневой гетерогенностью. На уровне источников это объединение формально структурированных таблиц, полуструктурных журналных записей и неструктурных корпусов изображений и текста; на уровне процессов, это конвейеры сбора, очистки, анонимизации и агрегации; на уровне выводов — оценивание моделей, устойчивых к выбросам и неполному наблюдению. В этом смысле общая дефиниция включает три компонента: а) свойства данных (масштаб, разнородность, динамика поступления), б) требования к обработке (распределённые хранилища, потоковые вычисления, воспроизводимые экспериментальные схемы), в) критерии полезности (проверяемые улучшения качества прогнозов и решений). Такой трёхчастный подход используется в аналитических работах, где сопоставляются эмпирические результаты и технические ограничения [4-5].

Наконец, определение «больших данных» должно учитывать управленческие и этико-правовые аспекты. В условиях расширяющегося обмена данными критически важны механизмы прозрачности, документирование происхождения наборов и верификация методик. Научная и прикладная литература подчёркивает, что статус «больших данных» оправдан постольку, поскольку обеспечивается воспроизводимость результатов и контролируется неопределенность, возникающая из смещений выборки, дрейфа процессов и ошибок измерения [6, 7].

Таким образом, под «большими данными» следует понимать не только крупные по объёму и скоростные массивы, но и институционально оформленные потоки информации, для которых требуются специализированные архитектуры, формальные процедуры контроля качества и обоснование социальной и экономической ценности получаемых знаний.

Современная финансовая индустрия в России демонстрирует устойчивую интеграцию методов обработки больших данных в ключевые управленческие и аналитические контуры. Практика ведущих банков и небанковских финансовых организаций показывает, что именно инструменты работы с крупными, высокоскоростными и разнородными массивами информации позволяют одновременно решать задачи

управления активами, измерения и контроля рисков, а также удержания и расширения клиентской базы. Эта тенденция согласуется с международной исследовательской повесткой, где «большие данные» трактуются как совокупность технологических и методологических решений, обеспечивающих извлечение экономически значимой информации из потоков данных со строгими требованиями к качеству, воспроизводимости и своевременности результатов [1, 6].

Содержательно использование больших данных в финансовом секторе опирается на три взаимосвязанных блока. Во-первых, управление активами опирается на интеграцию разнородных источников (рыночных котировок, транзакционных журналов, альтернативных индикаторов) с последующей стандартизацией и верификацией данных. Такой подход повышает устойчивость оценочных моделей к шуму и структурным сдвигам, а также позволяет оперативно обновлять параметры стратегий в условиях изменчивой ликвидности и волатильности [8]. Во-вторых, оценка и управление рисками получают качественный эффект за счёт объединения кредитных, рыночных и операционных показателей с поведенческими и макроэкономическими переменными, что расширяет класс выявляемых зависимостей и поддерживает раннее предупреждение о накапливающихся уязвимостях [9]. В-третьих, работа с клиентской базой строится на анализе жизненного цикла клиента, сегментации и детекции событий, влияющих на спрос, при обязательном соблюдении требований к защите и локализации данных [4-6].

Российская специфика проявляется в активном развитии институциональных и технологических оснований для применения больших данных. Публикации Национального исследовательского университета «Высшая школа экономики» по цифровой экономике фиксируют распространение аналитических платформ, объединяющих транзакционные данные, телеком-сигналы и открытые источники в целях повышения точности моделей скоринга, антифлага и ценообразования финансовых продуктов [5]. Работы Института проблем передачи информации РАН и отечественные исследования по высокопроизводительным вычислениям показывают, что практическая применимость методов обработки больших массивов определяется не столько абсолютным объёмом данных, сколько соответствием вычислительной архитектуры требованиям к пропускной способности, задержкам и воспроизводимости экспериментов [10].

Методологически «большие данные» в финансах следует трактовать как относительную категорию, так как один и тот же набор может не представлять сложности при наличии адекватной инфраструктуры, но становиться «большим» в условиях дефицита вычислительных ресурсов или ужесточения нормативных требований к качеству и времени отклика. Такой взгляд согласуется с зарубежными и российскими источниками, подчеркивающими роль контекстных ограничений — пропускной способности систем, асимметрии доступа к реестрам, регуляторных рамок обработки персональной информации.

С практической точки зрения ценность больших данных для российских финансовых институтов проявляется в четырех направлениях. Первое — повышение точности оценок: объединение традиционных финансовых индикаторов с поведенческими метриками снижает дисперсию прогнозов и улучшает калибровку моделей рисков [6]. Второе — ускорение принятия решений: потоковая обработка и обновляемые витрины признаков сокращают лаг между событием и управлением откликом, что критично для мониторинга ликвидности и лимитов. Третье — улучшение клиентского опыта: персонифицированные предложения и превентивные меры против мошенничества уменьшают отток и повышают маржинальность сегментов. Четвёртое — институциональная устойчивость: формализация процессов управления данными, ведение журналов происхождения наборов и регулярная валидация моделей обеспечивают воспроизводимость и соответствие регуляторным ожиданиям [2-6].

В совокупности эти наблюдения позволяют сформулировать обобщённое определение, релевантное финансовой сфере: под «большими данными» разумно понимать

институционально организованные и технологически поддержанные потоки и массивы информации, для обработки которых требуются специализированные архитектуры хранения и вычислений, формализованные процедуры контроля качества и воспроизводимые аналитические методики, обеспечивающие доказуемое улучшение точности и своевременности управлеченческих решений. В этом смысле активное использование больших данных российскими финансовыми институтами не только отвечает текущим потребностям управления активами, рисками и клиентскими отношениями, но и формирует основу для долгосрочной эффективности и устойчивости сектора.

Использование технологий обработки больших данных при финансовом анализе деятельности предприятия следует рассматривать как формирование устойчивого контура данных, в котором первичное наблюдение хозяйственных операций последовательно преобразуется в проверяемые выводы для управлеченческих решений. В таком контуре «большие данные» выступают не просто как совокупность объёмных и разнородных массивов, но как институционально организованная система сбора, очистки, сопоставления и интерпретации сведений из внутренних и внешних источников. К внутренним источникам относятся бухгалтерские и управлеченческие регистры, транзакционные журналы, журналы заявок на закупки и продажи, производственные датчики; к внешним — открытые государственные реестры, отраслевые индикаторы, ценовые потоки, сообщения надзорных органов и публичные цифровые следы контрагентов. Доказано, что именно согласование этих слоёв данных повышает точность оценок и устойчивость показателей к шуму и пропускам наблюдений [11, 12].

Методологически применение больших данных в финансовом анализе целесообразно строить вокруг трех взаимосвязанных этапов. Во-первых, формируется единая модель данных предприятия: описываются сущности, их связи, правила происхождения показателей и уровня качества. Это позволяет сопоставлять показатели из разных систем и фиксировать все преобразования в журнале происхождения данных, что повышает воспроизводимость анализа. Во-вторых, организуется потоковая и пакетная обработка: текущие транзакции и события агрегируются до аналитических витрин с заданной периодичностью, а исторические массивы используются для построения эталонных распределений и порогов аномалий. В-третьих, проводится моделирование и верификация: сравниваются альтернативные спецификации, оценивается устойчивость выводов к изменению выборок, проводится перекрёстная проверка с публичными источниками.

Содержательный вклад технологий больших данных проявляется по ключевым направлениям финансового анализа. При анализе ликвидности и платежеспособности высокочастотные данные о поступлениях и списаниях позволяют выявлять скрытую сезонность и уточнять потребность в оборотном капитале с учётом циклов поставщиков и покупателей. При оценке рентабельности объединение данных о структуре себестоимости, производственных параметрах и логистике даёт возможность рассчитать вклад отдельных факторов в изменения маржи и выявлять устойчивые резервы экономии. При анализе инвестиционных проектов сопоставление внутренних планов с внешними рыночными и технологическими индикаторами улучшает качество прогнозов денежных потоков и снижает неопределённость в сценарных расчётах. В управлении рисками объединение кредитных, операционных и комплаенс событий с внешними сигналами о состоянии контрагентов ускоряет обнаружение нарастающих уязвимостей и поддерживает ранние предупредительные меры [13].

Российская исследовательская и прикладная база подтверждает значимость такого подхода. Публикации Национального исследовательского университета «Высшая школа экономики» по цифровой трансформации финансового сектора показывают, что интеграция транзакционных данных, телеком показателей и открытых реестров повышает точность

скоринговых и прогностических моделей, а также уменьшает задержку между событием и управлением откликом [5].

Работы Института проблем передачи информации имени А. А. Харкевича Российской академии наук демонстрируют, что практическая эффективность обработки больших массивов определяется согласованием архитектуры хранения и вычислений с требованиями к пропускной способности, задержкам и воспроизводимости экспериментов, что критично для корпоративной аналитики [6].

Исследования в области высокопроизводительных вычислений подчёркивают, что выбор распределённых решений оправдан тогда, когда он обеспечивает стабильное время отклика при растущем объёме данных и сложной логике запросов [14].

Отдельного внимания требуют вопросы качества и этико правовой соответствия. Для финансового анализа, основанного на больших данных, базовыми являются следующие процедуры: прозрачная фиксация происхождения показателей; регулярная валидация моделей на вневыборочных данных; контроль смещений выборки и документирование всех преобразований. Международные руководства указывают, что без этих практик повышается риск неверной интерпретации корреляций и подмены причинно-следственных связей статистическими совпадениями [6]. Российские методические материалы по качеству официальной статистики и управлению данными в государственном секторе содержат применимые к корпоративной среде нормы по полноте, сопоставимости и актуальности показателей [7].

Проведённый анализ показывает, что «большие данные» целесообразно трактовать как относительную и междисциплинарную категорию, объединяющую свойства самих информационных массивов, требования к их обработке и критерии практической полезности результатов. Эволюция определения от акцента на объёме, скорости и разнородности к включению компонентов качества, изменчивости и доказуемой ценности отражает усложнение источников и усиление институциональных требований к воспроизводимости. В таком понимании статус «больших данных» обусловлен не только абсолютными размерами, сколько соответствием архитектуры хранения и вычислений целевым ограничениям по латентности, надёжности и прозрачности процедур.

Для финансовой сферы данная рамка имеет прикладной характер. Интеграция внутренних и внешних потоков данных, стандартизация и верификация показателей, а также использование потоковой и пакетной обработки обеспечивают повышение точности оценок ликвидности, рентабельности и денежных потоков, ускоряют управленические реакции и укрепляют контуры риска менеджмента. Практика российских финансовых институтов демонстрирует, что наибольший эффект достигается при сочетании технологической готовности (распределённые хранилища, вычисления в потоках, витрины признаков) с формализованным управлением данными (журналы происхождения, регулярная валидация моделей, контроль смещений). Это согласуется с международными и отечественными исследованиями, подчёркивающими роль контекстных ограничений, нормативных требований и процедур доказательности выводов [1–7; 8–10; 13–14].

С методологической точки зрения применение больших данных в финансовом анализе оправдано тогда, когда выполняются три условия: обеспечена сопоставимость источников и стабильность процессов сбора; построены воспроизводимые конвейеры очистки, агрегирования и моделирования; зафиксированы измеримые улучшения точности прогнозов и управленических решений. Невыполнение любого из этих условий ведёт к росту неопределённости и риску неверной интерпретации статистических зависимостей.

Следовательно, под «большими данными» в контексте финансового анализа деятельности предприятия следует понимать институционально оформленные и технологически поддержанные потоки информации, для обработки которых требуются специализированные архитектуры и формальные процедуры контроля качества, а результаты демонстрируют проверяемый прирост управленической ценности. Такая трактовка обеспечивает согласование научных и прикладных целей, создавая основу для

масштабируемой аналитики, повышения прозрачности и долгосрочной устойчивости финансовых решений.

Список источников

1. Майер-Шенбергер, В. Большие данные Революция, которая измени то, как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер; пер. с англ. Инны Гайдюк. – М.: Манн, Иванов и Фербер, 2014. – 240 с.
2. Атдаева О. Использование технологий big data в финансовой отрасли / О. Атдаева, Х. Сейитгельдиева // Символ науки: международный научный журнал. – 2024. – Т. 1, № 2-2. – С. 36-37.
3. Коноплева Ю. А. Применение технологии Big data на финансовых рынках / Ю. А. Коноплева, О. Н. Пакова, Ю. Р. Дейч // Вестник Северо-Кавказского федерального университета. – 2022. – № 2(89). – С. 58-65. – DOI 10.37493/2307-907X.2022.2.8.
4. Отчёт Европейской комиссии по данным и инновациям «100 радикальных инновационных прорывов будущего». URL:https://commission.europa.eu/publications/annual-activity-report-2024-research-and-innovation_en
5. Цифровая трансформация: ожидания и реальность: докл. к XXIII Ясинской (Апрельской) междунар. науч. конф. по проблемам развития экономики и общества, Москва, 2022 г. [Текст] / Г. И. Абдрахманова, С. А. Васильковский, К. О. Вишневский, М. А. Гершман, Л. М. Гохберг и др.; рук. авт. кол. П. Б. Рудник; Нац. исслед. ун-т «Высшая школа экономики». — М.: Изд. дом Высшей школы экономики, 2022.
6. Руководство ОЭСР по управлению цифровым правительством. URL:<https://u-intosai.org/ru/novosti/rukovodstvo-oesr-po-upravleniyu-cifrovym-pravitelstvom>
7. Статистика и отчёты. Минцифры. URL:<https://digital.gov.ru/activity/statistics-reports>
8. Global statistics at the heart of international cooperation. URL:<https://data.bis.org>
9. Paolo Brandimarte. Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics. URL:<https://onlinelibrary.wiley.com/doi/book/10.1002/9781118593264>
10. Мещеряков Р.В., Шустов С.А. Модель стеганографического встраивания информации в изображения с использованием методов глубокого обучения. – М. ИППИ РАН. Том 25, №1. С.1-13.
11. Доклады Организации экономического сотрудничества и развития о политике в области данных. URL:<https://www.oecd.org/digital/>
12. Отчёты Европейской комиссии по экономике данных. URL:<https://digital-strategy.ec.europa.eu/>
13. Аналитические материалы Банка международных расчётов, русскоязычные и мультиязычные публикации. URL:<https://www.bis.org/>
14. ИППИ РАН. URL:<https://www.ipiran.ru/>
15. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. – Спб.: БХВ-Петербург, 2002.

Сведения об авторе

Гусейнов Магомед Мурадович, аспирант Московской международной академии, г. Москва, Россия

Information about the author

Guseinov Magomed Muradovich, PhD student at the Moscow International Academy, Moscow, Russia