

УДК 336.76 (075.8)

Е. А. Локтионова*Байкальский государственный университет,
г. Иркутск, Российская Федерация***А. В. Рагозина***Байкальский государственный университет,
г. Иркутск, Российская Федерация*

ОСОБЕННОСТИ ПРИМЕНЕНИЯ СИСТЕМ АНАЛИЗА БОЛЬШИХ ДАННЫХ В ДЕЯТЕЛЬНОСТИ КОММЕРЧЕСКОГО БАНКА

АННОТАЦИЯ. В представленной статье рассматриваются особенности использования в банковской сфере современных информационных технологий для создания условий, способствующих эффективному управлению банком: больших данных и аналитических систем типа «Data Mining». Раскрываются возможности интеллектуального анализа данных банками. Отмечается, что с позиций банковского маркетинга, технология больших данных помогает обнаружить новые источники информации, улучшить оперативность и качество решений, принимаемых для привлечения клиентов, персонализации специальных предложений для клиентов, разработки новых услуг. Проводится анализ состояния рынка технологии больших данных в России. Делается вывод о том, что ключевыми потребителями технологии «Big Data» в России являются банки. Обобщается опыт использования технологии больших данных в российской банковской практике. На основе анализа проблем и перспектив использования больших данных в банковской практике делается вывод о необходимости использования современных информационных технологий крупными финансовыми организациями в целях укрепления их конкурентных преимуществ на рынке.

КЛЮЧЕВЫЕ СЛОВА. Информация; информационные технологии; большие данные; Big Data; банки.

ИНФОРМАЦИЯ О СТАТЬЕ. Дата поступления 28 марта 2017 г.; дата принятия к печати 18 апреля 2017 г.; дата онлайн-размещения 19 июня 2017 г.

Ye. A. Loktionova*Baikal State University,
Irkutsk, Russian Federation***A. V. Ragozina***Baikal State University,
Irkutsk, Russian Federation*

FEATURES OF USING SYSTEMS OF BIG DATA ANALYSIS SYSTEMS IN COMMERCIAL BANK'S ACTIVITY

ABSTRACT. The article examines the features of using modern information technologies in the banking sector for creating the conditions that contribute to effective bank management: big data and analytical systems of the «Data Mining» type. It reveals the possibilities of intellectual analysis of the data by banks. It notes, that from the point of view of banking marketing, the Big Data technology helps to discover new sources of information, to improve the efficiency and quality of decisions taken to attract clients, personalization of special offers for customers, development of new services. It draws a conclusion that the key consumers of the «Big Data» in Russia are banks. It generalizes the experience of using the big data technology in Russian banking practice. The analysis of the Russian Big Data technology market is being conducted. It is concluded that banks are the main consumers of the technology in Russia. Based on the analysis of the problems and prospects of using big data in banking practice, a conclusion is made about the need of using modern information

© Е. А. Локтионова, А. В. Рагозина, 2017

technologies by large financial organizations in order to strengthen their competitive advantages on the market.

KEYWORDS. Information; information technologies; Big Data; banks.

ARTICLE INFO. Received March 28, 2017; accepted April 18, 2017; available online June 19, 2017.

Постоянные изменения внешней среды, конъюнктуры рынка, потребностей и поведения клиентов, политики регуляторов и многих других внешних и внутренних факторов приводят к тому, что в настоящее время все участники экономических отношений работают в условиях высокой неопределенности [1–4]. Одним из факторов неопределенности является большой объем разнородной поступающей информации, влияющей на принятие решений, вынуждающий организации использовать современные информационные технологии. Одной из технологий, дающих возможность не только обработки накопленной и вновь поступающей информации, но и использования ее для получения дополнительных доходов, стала технология «Большие данные» («Big Data») [5].

Ценность технологии больших данных состоит в том, что она позволяет организациям сохранять большие объемы разнообразной информации, создавая соответствующие базы данных, а также управлять ими и обрабатывать их в требуемое время с требуемой скоростью в соответствии с поставленными задачами [6]. По типу хранимой информации базы данных подразделяются на три группы, представленные на рис. 1.

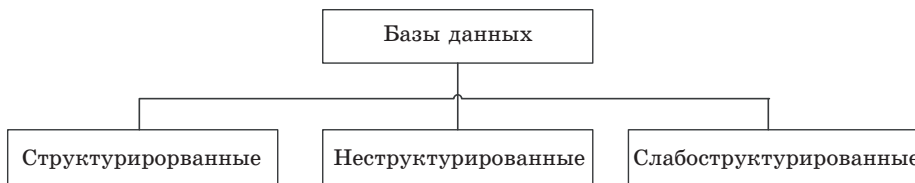


Рис. 1. Классификация баз данных по типу хранимой информации

Структурированные базы данных содержат информацию, имеющую определенную длину и формат (например, комбинации слов и чисел, даты, адреса клиентов, имена, и т. д.). По мнению экспертов, к информации, хранящейся в структурированных базах данных, относится порядка 20 % от общего объема данных¹. Структурированные данные являются наиболее привычным форматом данных. Источниками структурированных данных являются:

- данные с датчиков (например, GPS);
- финансовые данные (например, данные, относящиеся к торговле на бирже);
- данные о посещениях веб-сайтов;
- любые данные, вводимые человеком на различных сайтах, например, при участии в социологических опросах.

В неструктурированных базах данных содержится информация, не имеющая определенного формата. На долю неструктурированной информации относится большая часть всего объема данных (порядка 70 %). Таким образом, большая часть информации, с которой сталкивается организация, является неструктурированной. Источниками неструктурированных данных являются:

- спутниковые снимки;
- фото и видеодокументы (записи с видеорегистраторов);
- данные социальных сетей (YouTube, Twitter и т. д.);

¹ Аналитический обзор рынка Больших данных. URL: <http://www.ipoboard.ru/files/cms/5e3af134b9942559eb802ea93a1c9050>.

– мобильные данные (текстовые сообщения, информация о месте нахождения).

Слабоструктурированная информация обладает признаками как структурированной, так и неструктурированной информации и не соответствует четким стандартам. К слабоструктурированной относится информация в форматах стандартов EDI, SWIFT и XML. Информацию подобного типа можно рассматривать как промежуточный результат обработки сложных данных.

Самыми крупными пользователями технологии больших данных в России можно назвать банки [7]. Основываясь на использовании открытых данных, технология «Big Data» применяется в таких банках, как «Банк ВТБ (ПАО)», ПАО «Сбербанк России», АО «АЛЬФА-БАНК», ПАО «БАНК УРАЛСИБ», АО «ОТП Банк», АО «Газпромбанк», АО «Райффайзенбанк»². В целом, поступающие и имеющиеся в наличии данные можно охарактеризовать как потенциальную «золотую жилу» любого банка, но чтобы добраться до настоящего золота, нужно качественно обработать внушительный объем разнородного материала и извлечь из него информацию о закономерностях и тенденциях, которые наиболее важны для принятия решений, связанных с деятельностью банка. На рис. 2 показана общая схема получения необходимых для организации проанализированных данных, построенная на использовании технологии больших данных и различных аналитических систем.



Рис. 2. Схема получения данных, основанная на использовании технологии больших данных и различных аналитических систем

На рис. 2 видно, что вся имеющаяся и вновь поступающая информация попадает на так называемую ВИ Платформу. ВИ Платформа (Платформа «Business intelligence») представляет собой совокупность разнообразных методов и инструментов для перевода исходной информации в форму, оптимальную для использования организацией в соответствии с имеющимися целями. Отметим, что структурированная информация передается на ВИ Платформу напрямую, а неструктурированная и слабоструктурированная информация проходит предварительную обработку посредством технологии «Big Data» и трансформируется в интеллектуальный интегрированный архив.

Результатом работы ВИ Платформы является качественно иная информация, представленная в удобном и привычном формате в виде отчетов, графиков, диаграмм, предупреждений систем мониторинга, а также в виде хранилищ баз дан-

² Большие данные (Big Data) в России URL: [http://www.tadviser.ru/index.php/Статья:Большие_данные_\(Big_Data\)](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data)).

ных, имеющих существенные отличия от моделей, используемых в традиционных реляционных системах управления базами данных, делающих возможным обнаружение новых трендов посредством применения метода свободного поиска.

В процессе практического использования технологии больших данных очень часто возникает необходимость интеграции разных источников [8]. Данные могут поступать в банк из различных внутренних систем, либо одновременно из внешних и внутренних источников или только внешних, и в целях извлечения из них максимальной пользы разные данные необходимо интегрировать. Для этого в мире информационных технологий существуют метаданные и соединители.

Метаданные, представляющие собой схемы расположения и другие характеристики, указывающие путь к программным компонентам и данным организации, описывающие условия доступа к информации и особенности ее использования, являются важнейшим компонентом для объединения любого вида данных. Данные о номере банковского счета можно назвать примером метаданных, ведь он может содержать сам номер, описание, указание на тип данных, имя, телефон и адрес владельца и уровень доступа к этой информации. Использование метаданных позволяет упорядочить хранилища данных и реагировать на появление и изменения источников данных. Традиционно метаданные используются в качестве каталога, обеспечивающего представление обо всех источниках информации, но для использования в технологии больших данных каталог должен быть устроен иначе, с использованием дополнительного аналитического инструментария — соединителей. Соединители — программные средства, — позволяющие получать и объединять большие данные из различных источников. С помощью соединителей возможно, например, объединить внешние источники больших данных (Facebook, Twitter и т. д.) с внутренними базами данных организации, чтобы иметь возможность проанализировать эти данные совместно.

Говоря о способах организации интегрированных данных, следует отметить, что еще несколько десятилетий назад большие массивы информации обрабатывались только с помощью суперкомпьютеров, что требовало огромных денежных затрат. Постепенное снижение цен на программно-аппаратные системы привело к появлению и распространению новых методов распределительных вычислений. Но настоящий прорыв в этой области произошел тогда, когда крупные компании (такие как Yahoo, Facebook, Google), начали искать способы получения прибыли из огромного количества данных, создающихся при использовании информационных продуктов. Именно тогда впервые потребовались технологии, которые были бы способны обеспечивать доступ к большим объемам данных, их хранение и анализ в режиме реального или приближенного к реальному времени. Это и стало толчком для создания таких систем хранения и управления данными, как «Hadoop», «MapReduce» и «Big Table», которые впоследствии стали активно использоваться банками.

Платформа «Hadoop» организует параллельную обработку данных на множестве вычислительных узлов, что позволяет сократить задержки за счет ускорения вычислений. «Hadoop» включает в себя два основных компонента: высокомасштабируемую распределительную файловую систему, способную сохранять петабайты данных, и высокомасштабируемый механизм «MapReduce», осуществляющий вычисления в пакетном режиме.

«MapReduce» — это технология, разработанная компанией «Google» для эффективного выполнения набора функций, охватывающих большие объемы данных в пакетном режиме. Компонент «map» распределяет задачи между большим количеством систем, распределяя их нагрузку, а также восстанавливает систему после сбоев, а функция «reduce» собирает все элементы воедино и формирует результат.

Система распределительного хранения данных «Big Table» также разработана компанией «Google». Она предназначена для управления масштабируемыми структурированными данными. «Big Table» представляет собой разреженную распределительную постоянную и многомерную сортировочную карту, предназначенную для хранения больших объемов данных на стандартных серверах.

Мировыми лидерами в разработке систем анализа больших данных являются компании «IBM», «SAS», «Tableau» и «Oracle». «IBM» использует корпоративный подход к большим данным и межплатформенной интеграции, в том числе встраивание и пакетное предоставление возможностей аналитики. Продукты этой компании включают хранилище данных, набор технологий для углубленной аналитики, статистический пакет для общественных наук, а также прогнозный анализ в режиме реального времени. «SAS» предоставляет несколько возможностей распределенной обработки непосредственно в базе данных, сетевые распределительные вычисления, аналитику в памяти. «Tableau» выпускает программное обеспечение для визуализации данных и различной бизнес-аналитики. «Oracle» разработал целый ряд инструментов, дополняющих основную платформу больших данных, которая называется «Oracle Exadata». В их число входит углубленная аналитика и возможность размещения базы данных в памяти благодаря решению «компьютер в памяти» («Oracle Exalytics»).

Одной из самых востребованных в мире систем для работы с большими данными является «Data Mining». «Data Mining» — технология, позволяющая обрабатывать информацию с целью автоматического поиска шаблонов, способных охарактеризовать какие-либо фрагменты неоднородных многомерных данных. Рассмотрим возможности применения систем анализа больших данных в банке на примере использования технологии «Data Mining». Область применения «Data Mining» не имеет ограничений. В методах работы этой системы в первую очередь заинтересованы именно банки, так как она позволяет выявить характеристики наиболее успешных для работы регионов, филиалов и клиентов для планирования дальнейшей деятельности банка [8; 9]. Применяемая в этой системе программа профилирования используется для оценки риска при выдаче кредитов клиентам банка [10].

На рис. 3 представлены задачи, решаемые с помощью «Data Mining» в банковской сфере.

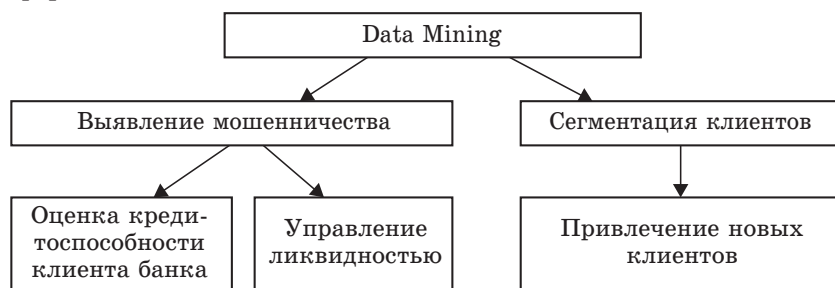


Рис. 3. Основные направления использования технологии «Data Mining» в банковской сфере

Определение некредитоспособности клиента без применения рассматриваемой аналитической технологии основывается преимущественно на интуиции, опыте, субъективных представлениях сотрудников банка о благонадежности или неблагонадежности клиентов. Инструменты «Data Mining» профилируют клиентов, разделяя их на добросовестных и неблагонадежных заемщиков. Кроме того, становится возможным классифицировать заемщика по группам риска, то есть не

только решить вопрос о возможности кредитования, но и установить лимит кредита, проценты по нему и срок возврата [11].

Выявление мошенничества осуществляется с помощью применения «подозрительных стереотипов поведения», определяемых системой в результате анализа банковских транзакций, которые впоследствии были признаны банком мошенническими [12]. Определение подозрительных случаев осуществляется путем использования совокупности последовательных операций на определенном временном интервале. Если в дальнейшем система воспринимает текущую операцию подозрительной, сотрудник банка, ориентируясь на полученную с помощью «Data Mining» информацию, имеет право, например, заблокировать операции с определенной картой. С помощью средств «Data Mining» банк также осуществляет сегментацию клиентов, что позволяет найти «профиль» наиболее выгодных клиентов и далее акцентировать свою маркетинговую политику для привлечения клиентов, соответствующих найденному «профилю». «Data Mining», интегрируя поступающие из внешних источников данные с имеющимися базами данных банков и обрабатывая огромное количество информации, помогает банкам строить прогнозные модели ценностей своих клиентов и соответствующим образом обслуживать каждую категорию [13]. В целях привлечения новых клиентов, банк классифицирует клиентов на «более выгодных» и «менее выгодных». После определения наиболее выгодного сегмента клиентов банк проводит более активную маркетинговую политику по привлечению клиентов именно среди группы «более выгодных» клиентов.

Примером наиболее активного использования аналитической системы «Data Mining» в российском банковском секторе является ПАО «Сбербанк России». Тема важности анализа больших массивов информации (данных) затрагивается в стратегии банка на 2014–2016 гг. Банк в полной мере осознает маркетинговые преимущества использования современных технологий: посредством использования системы «Data Mining» в банке ведется постоянная работа по борьбе с мошенничеством, повышению эффективности продаж и совершенствованию технологии привлечения новых и удержания уже имеющихся клиентов банка. В банке работает лаборатория по большим данным, формирующая личные профили клиентов на основе анализа банковских транзакций, данных, поступающих от банкоматов и другого оборудования банка, данных программ лояльности, информации госструктур, социальных сетей и коллекторских агентств.

В октябре 2016 г. Сбербанк заявил о внедрении новой системы кредитования малого бизнеса без залога, основанной на анализе больших данных о потенциальном заемщике. В ноябре 2016 г. Сбербанк расширил текущую функциональность используемой платформы «Informatica Intelligent Data Platform» путем приобретения дополнительного компонента «Informatica Big Data Management», что позволило аналитикам осуществлять всесторонний предварительный анализ данных и их профилирование в целях обеспечения высокой скорости внесения изменений и передачи на поддержку разработанных процессов. В январе 2017 г. Сбербанк запустил проект «Открытые данные». К основным задачам проекта относятся сбор информации о частных и корпоративных клиентах банка, ее систематизация, обобщение и предоставление всем желающим.

Следует отметить, что широкое применение технологии больших данных в банковском секторе осложняется необходимостью значительных финансовых вложений, недостаточным числом грамотных специалистов, а также потребностью в дополнительных программных решениях, позволяющих осуществлять оценку рисков по различным направлениям банковской деятельности.

Подводя итог можно сказать, что в условиях современной экономики описанные выше аналитические системы и системы хранения информации, работающие

в рамках технологии больших данных, просто необходимы крупным финансовым организациям в целях укрепления их конкурентных преимуществ на рынке и получения в будущем дополнительных доходов.

Список использованной литературы

1. Локтионов В. И. Влияние неопределенности исходных данных на варианты долгосрочного развития топливно-энергетического комплекса / В. И. Локтионов // *Экономический анализ: теория и практика*. — 2015. — № 31. — С. 51–60.
2. Локтионов В. И. Учет фактора неопределенности при оценке вариантов использования Ковыктинского газа / В. И. Локтионов, Ю. Д. Кононов, П. В. Ступин // *Энергетика России в XXI веке: стратегия развития — восточный вектор. Энергетическая кооперация в Азии: что после кризиса? — Иркутск : Ин-т систем энергетики им. Л. А. Мелентьева Сиб. отд-ния РАН, 2010. — С. 655–660.*
3. Локтионов В. И. Оценка эффективности инвестиционных проектов в энергетике с учетом предельных цен на энергоносители / В. И. Локтионов // *Экономический анализ: теория и практика*. — 2014. — № 33. — С. 17–22.
4. Демуринов В. Б. Принятие управленческих решений в информационной системе гостиничного комплекса в условиях неопределенности и нечеткости исходных данных / В. Б. Демуринов // *Известия Тульского государственного университета. Экономические науки*. — 2011. — № 1-2. — С. 191–197.
5. Некрасова Е. Большие деньги, большие данные, большие перспективы [Электронный ресурс] / Е. Некрасова. — Режим доступа: <http://www.computerra.ru/cio/487>.
6. Просто о больших данных / Дж. Гурвич, А. Ньюджент, Ф. Халпер, М. Кауфман. — М. : Эксмо, 2015. — 348 с.
7. Локшина Э. А. А ваша компания готова к управлению клиентским опытом? 7 признаков клиент-ориентированной компании [Электронный ресурс] / Э. А. Локшина. — Режим доступа: <http://love-credit.ru/st/kak-banki-upravlyayut-klientskimi-vpechatleniyami>.
8. Рудская Е. Н. Клиентский опыт как инструмент конкурентоспособности коммерческих банков / Е. Н. Рудская, Г. А. Болохова // *Молодой ученый*. — 2016. — № 11. — С. 937–948.
9. Шурчкова Ю. В. Конкурентная среда и ее влияние на стратегию поведения организации: дис. ... канд. экон. наук : 08.00.05 / Ю. В. Шурчкова. — Воронеж, 2004. — 185 с.
10. Рудская Е. Н. Технологии профилирования заемщиков в банковской сфере: инновации в скоринге и минимизации рисков мошенничества / Е. Н. Рудская, Ю. Ю. Полтавская // *Молодой ученый*. — 2015. — № 24. — С. 567–574.
11. Татарнинова Л. В. Методические подходы к оценке финансовой устойчивости коммерческого банка / Л. В. Татарнинова. — Иркутск : Изд-во БГУЭП, 2013. — 130 с.
12. Терехов С. Кроссканальное решение по противодействию мошенничеству [Электронный ресурс] / С. Терехов. — Режим доступа: <http://www.journal.ib-bank.ru/pub/297>.
13. Нестеренко С. Технологии Big Data для клиентской аналитики [Электронный ресурс] / С. Нестеренко. — Режим доступа: https://www.ibm.com/ru/events/presentations/connect2014/12_connect14.pdf.

References

1. Loktionov V. I. Impact of initial data uncertainty on options of long-term development of the fuel and energy complex. *Ekonomicheskii analiz: teoriya i praktika = Economic Analysis: Theory and Practice*, 2015, no. 31, pp. 51–60. (In Russian).
2. Loktionov V. I., Kononov Yu. D., Stupin P. V. Uncertainty factor management in assessing variants of using the Kovykta gas. *Energetika Rossii v XXI veke: strategiya razvitiya — vostochnyi vektor. Energeticheskaya kooperatsiya v Azii: chto posle krizisa?* [Russia's energetics in XXI century: development strategy — Eastern vector. Energy cooperation in Asia: what comes after crisis?]. Irkutsk, Melentiyev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences Publ., 2010, pp. 655–660. (In Russian).
3. Loktionov V. I. Assessing efficiency of investment projects with a glance to marginal prices of energy resources. *Ekonomicheskii analiz: teoriya i praktika = Economic Analysis: Theory and Practice*, 2014, no. 33, pp. 17–22. (In Russian).

4. Demurin V. B. Making managerial decisions in information system of hotel complex in terms of uncertainty and fuzzy basic data. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Ekonomicheskie nauki = Bulletin of Tula State University. Economic Sciences*, 2011, no. 1-2, pp. 191–197. (In Russian).

5. Nekrasova Ye. *Bol'shie den'gi, bol'shie dannye, bol'shie perspektivy* [Big money, big data, big prospects]. Available at: <http://www.computerra.ru/cio/487>. (In Russian).

6. Gurvits Dzh., Nyudzhent A., Khalper A., Kaufman M. *Prosto o bol'shikh dannykh* [Simply about Big Data]. Moscow, Eksmo Publ., 2015. 348 p.

7. Lokshina Ye. A. *A vasha kompaniya gotova k upravleniyu klientskim opytom? Sem' priznakov klient-orientirovannoi kompanii* [Is your company prepared for customer experience management? Seven attributes of customer-oriented company]. Available at: <http://love-credit.ru/st/kak-banki-upravlyayut-klientskimi-vpechatleniyami>. (In Russian).

8. Rudskaya E. N., Bolokhova G. A. Customer experience as an instrument of commercial banks' competitiveness. *Molodoi uchenyi = Young Scientist*, 2016, no. 11, pp. 937–948. (In Russian).

9. Shurchkova Yu. V. *Konkurentnaya sreda i ee vliyanie na strategiyu povedeniya organizatsii. Kand. Diss.* [Competitive environment and its impact on strategy of organization's behavior. Cand. Diss.]. Voronezh, 2004. 185 p.

10. Rudskaya E. N., Poltavskaya Yu. Yu. Technologies of profiling loaners in banking sphere: innovations in scoring and minimization of fraud risks. *Molodoi uchenyi = Young Scientist*, 2015, no. 24, pp. 567–574. (In Russian).

11. Tatarinova L. V. *Metodicheskie podkhody k otsenke finansovoi ustoichivosti kommercheskogo banka* [Methodical approaches to assessing commercial bank's financial stability]. Irkutsk, Baikal State University of Economics and Law Publ., 2013. 130 p.

12. Terekhov S. *Krosskanal'noe reshenie po protivodeistviyu moshennichestvu* [Cross-channel decision on counteracting fraud]. Available at: <http://www.journal.ib-bank.ru/pub/297>. (In Russian).

13. Nestrenko S. *Tekhnologii Big Data dlya klientskoi analitiki* [Big Data technologies for client analytics]. Available at: https://www.ibm.com/ru/events/presentations/connect2014/12_connect14.pdf. (In Russian).

Информация об авторах

Локтионова Елена Александровна — кандидат экономических наук, старший преподаватель, кафедра банковского дела и ценных бумаг, Байкальский государственный университет, 664003, г. Иркутск, ул. Ленина, 11, e-mail: loktionova_ea@mail.ru.

Рагозина Александра Вадимовна — магистрант, кафедра банковского дела и ценных бумаг, Байкальский государственный университет, 664003, г. Иркутск, ул. Ленина, 11, e-mail: alex.managarova@yandex.ru.

Authors

Yelena A. Loktionova — PhD in Economics, Senior Lecturer, Chair of Banking and Securities, Baikal State University, 11 Lenin St., 664003, Irkutsk, Russian Federation; e-mail: loktionova_ea@mail.ru.

Alexandra V. Ragozina — Master Degree Student, Chair of Banking and Securities, Baikal State University, 11 Lenin St., 664003, Irkutsk, Russian Federation; e-mail: alex.managarova@yandex.ru.

Библиографическое описание статьи

Локтионова Е. А. Особенности применения систем анализа больших данных в деятельности коммерческого банка / Е. А. Локтионова, А. В. Рагозина // *Baikal Research Journal*. — 2017. — Т. 8, № 2. — DOI: [10.17150/2411-6262.2017.8\(2\).9](https://doi.org/10.17150/2411-6262.2017.8(2).9).

Reference to article

Loktionova Ye. A., Ragozina A. V. Features of using systems of Big Data analysis systems in commercial bank's activity. *Baikal Research Journal*, 2017, vol. 8, no. 2. DOI: [10.17150/2411-6262.2017.8\(2\).9](https://doi.org/10.17150/2411-6262.2017.8(2).9). (In Russian).