

УДК 004.658.2

DOI 10.21661/r-559532

Лихоузов К.И.

ПРИМЕНЕНИЕ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ НА ПЛАТФОРМЕ РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ БОЛЬШИХ ДАННЫХ В БАНКОВСКОЙ СФЕРЕ

***Аннотация:** в настоящее время наблюдается колоссальный рост количества цифровых данных, и, по данным экспертов, этот объем будет увеличиваться на 30–40% ежегодно. Объем данных, созданных в 2020 году, более чем в 3 раза превышает объем данных за 2015 год, и составляет 64 Зеттабайт (Зб). В 2022 году объемы всех данных, накопленных человечеством, составили примерно 97 Зб, а к 2025 году это число возрастет до 180 Зб. В связи с этим многие компании ищут новые способы, методы, автоматизированные системы по обработке таких данных. В статье рассмотрены платформы по обработке данных, которые позволяют финансовым организациям выйти на новый уровень как по качеству оценки кредитоспособности, риск-менеджменту, так и по прибыли.*

***Ключевые слова:** большие данные, банки, искусственный интеллект, машинное обучение.*

Введение

Банковская сфера является одной из самых важных и ответственных отраслей экономики. Каждый день банки обрабатывают огромное количество данных, связанных с финансовыми операциями, клиентами, кредитами и многим другим. В связи с этим, задачи машинного обучения на платформе распределенных вычислений больших данных становятся все более актуальными для банковской сферы.

Способность коммерческого банка эффективно использовать большие данные и ускорить внедрение инноваций для достижения AI-трансформации (Artificial Intelligence) определяет его будущие возможности устойчивого развития и сильно влияет на конкурентоспособность. Традиционные платформы

анализа данных, такие как SAS EG, MS Excel, MS SQL, основанные на реляционных хранилищах данных, не может удовлетворить потребности текущего развития бизнеса.

В крупных банках активно используют технологии big data, ведь чем больше данных из различных источников можно получить о клиенте, тем более точный прогноз можно сделать по его дальнейшему поведению. Особенно популярны становятся модели оффлайн расчета – предрасчет по всему возможному базису клиентов – такие пайплайны должны обрабатывать десятки – сотни Гб данных.

Применение распределенных вычислений в банках

Big Data – это огромные объемы информации, которые могут быть использованы для анализа и принятия решений. В банковской сфере, где каждый день обрабатывается огромное количество данных, использование Big Data может стать ключевым фактором для повышения эффективности бизнес-процессов и увеличения прибыли.

Одной из наиболее важных областей применения Big Data в банковской сфере является анализ кредитного риска. Банки могут использовать данные о клиентах, такие как история платежей, кредитный рейтинг, уровень дохода и другие параметры, чтобы определить вероятность того, что клиент не сможет вернуть кредит. Это позволяет банкам принимать более обоснованные решения при выдаче кредитов и уменьшать риски дефолта.

Другой областью применения Big Data является улучшение маркетинговых стратегий. Банки могут использовать данные о клиентах, такие как их возраст, пол, местоположение и покупательское поведение, чтобы создавать персонализированные предложения и рекламу. Это позволяет банкам увеличить эффективность своих маркетинговых кампаний и привлечь больше клиентов.

Big Data также может быть использована для улучшения управления финансами. Банки могут анализировать данные о своих операциях, чтобы определить, где можно сократить расходы и увеличить прибыль. Они также могут использовать данные о рынке и экономике, чтобы прогнозировать будущие тенденции и принимать соответствующие решения.

Однако, использование Big Data в банковской сфере также создает некоторые вызовы. Один из главных вызовов – это безопасность данных. Банки должны обеспечивать защиту конфиденциальности данных своих клиентов и защищать их от хакерских атак. Другой вызов – это сложность обработки огромных объемов данных. Банки должны иметь достаточно мощных вычислительных мощностей, чтобы обрабатывать такие объемы данных.

Заключение

Коммерческим банкам необходимо активно использовать финтех, следить за последними разработками в отрасли и изучать передовые технологии, продолжать улучшать возможности платформ больших данных и обеспечивать сильное движение в направлении цифровой трансформации.

В заключение, использование Big Data в банковской сфере может привести к значительному улучшению эффективности бизнес-процессов и увеличению прибыли. Банки должны обеспечивать безопасность данных и иметь достаточно мощных вычислительных мощностей, чтобы успешно использовать Big Data.

Список литературы

1. Сысоева А.А. Анализ рынка больших данных в банковской сфере / А.А. Сысоева, З.В. Эюбов, А.Е. Фошкин // Финансовая жизнь. – 2017. – №1. – С. 32–35.
2. Zhang Boyu. Research on Data Audit Mode of Commercial Banks Based on Hadoop Technology / Zhang Boyu // China Internal Audit. – 2019. – №(5). – P. 33–37.
3. Zhang Dengyao. Big Data Analysis of Commercial Banks Based on Hadoop Distributed File System / Zhang Dengyao // Journal of Shandong Agricultural University (Natural Science Edition). – 2018. – №49 (5). – P. 884–888.
4. Han Jian. Architecture Design of Bank Big Data Platform Based on Hadoop Technology / Han Jian // Electronic World. – 2017. – №(22). – P. 162–163.
5. Chen Yan. Research on Online Banking Historical Data Based on Hadoop / Chen Yan // Computer Knowledge and Technology. – 2017. – №13 (16). – P. 21–23.

Лихоузов Кирилл Игоревич – аспирант, АНО ВО «Московский информационно-технологический университет – Московский архитектурно-строительный институт», Россия, Москва.
