

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ БОЛЬШИХ ДАННЫХ: ИССЛЕДОВАНИЕ ПРОЦЕССОВ СБОРА И ОБРАБОТКИ НА ПРЕДПРИЯТИЯХ

Кулдашева Наргиза Комолиддин кизи
докторант ТГЭГ,

Аннотация

В этой статье рассматриваются процессы сбора и обработки больших данных на предприятиях. В нем представлен подробный анализ современной литературы, определены ключевые методы, используемые в интеллектуальном анализе данных, и рассмотрено влияние этих процессов на бизнес-операции. В исследовании также предлагаются практические рекомендации по повышению эффективности интеллектуального анализа данных в корпоративной среде.

Ключевые слова: Большие данные, Интеллектуальный анализ данных, Сбор данных, Обработка данных, Предприятия, Машинное обучение, Бизнес-аналитика, Анализ данных

Annotatsiya

Ushbu maqola korxonalarda katta ma'lumotlarni yig'ish va qayta ishlash jarayonlarini ko'rib chiqadi. Qo'yilgan muammo bo'yicha adabiyotlar batafsil tahlil qilinaadi, ma'lumotlarni olishda qo'llaniladigan asosiy texnikalar aniqlanadi va ushbu jarayonlarning biznes operatsiyalariga ta'siri o'rganiladi. Tadqiqot shuningdek, korporativ muhitda ma'lumotlarni qidirish samaradorligini oshirish bo'yicha amaliy tavsiyalarni taklif qiladi.

Kalit so'zlar: Katta ma'lumotlar, ma'lumotlarni qazib olish, ma'lumotlarni qazib olish, ma'lumotlarni qayta ishlash, korxona, mashinani o'rganish, biznes razvedkasi, ma'lumotlar tahlili

Abstract

This paper examines the processes of collecting and processing big data in enterprises. It provides a detailed analysis of the current literature, identifies key techniques used in data mining, and examines the impact of these processes on business operations. The study also offers practical recommendations for improving the effectiveness of data mining in an enterprise environment.

Keywords: Big Data, Data Mining, Data Collection, Data Processing, Enterprises, Machine Learning, Business Intelligence, Data Analysis

ВВЕДЕНИЕ

В современных условиях ведения бизнеса способность использовать и анализировать большие объемы данных стала важнейшим преимуществом. Интеллектуальный анализ больших данных (Big data mining) - процесс извлечения ценной информации из обширных массивов данных - позволяет предприятиям принимать обоснованные решения, оптимизировать операции и сохранять конкурентные преимущества. По мере того как предприятия генерируют и собирают огромные объемы данных из различных источников,

потребность в эффективных методах интеллектуального анализа данных становится все более актуальной. В этой статье рассматриваются процессы, связанные со сбором и обработкой больших данных на предприятиях, освещаются их значение, проблемы и лучшие практики.

АНАЛИЗ ЛИТЕРАТУРЫ

Концепция больших данных возникла в ответ на растущий объем, скорость и разнообразие данных, генерируемых цифровыми технологиями. Ранние исследования, посвященные большим данным, были сосредоточены на технических аспектах, таких как возможности хранения и обработки. Однако в последнее время появилось больше литературы, посвященной изучению стратегического значения больших данных в бизнес-контексте.

В литературе можно выделить несколько ключевых тем:

1. Методы сбора данных: Для сбора данных используются различные методы, включая веб-анализ, данные датчиков, записи транзакций и аналитику в социальных сетях. Выбор метода зависит от типа данных и бизнес-целей.

2. Инструменты обработки данных : Hadoop, Apache Spark и другие платформы распределенных вычислений широко используются для обработки больших массивов данных. Эти инструменты позволяют компаниям управлять данными и анализировать их в больших масштабах.

3. Машинное обучение и искусственный интеллект: Алгоритмы машинного обучения все чаще применяются в интеллектуальном анализе данных для выявления закономерностей и прогнозирования результатов. Интеграция искусственного интеллекта с аналитикой больших данных еще больше расширила возможности предприятий в принятии решений, основанных на данных.

МЕТОДОЛОГИЯ

В этом исследовании используется смешанный подход, сочетающий качественные и количественные методы исследования для анализа сбора и обработки больших объемов данных на предприятиях. Качественный аспект включает интервью со специалистами по обработке данных, ИТ-менеджерами и бизнес-аналитиками, чтобы получить представление о проблемах и практиках интеллектуального анализа данных. Количественный аспект включает опрос 50 предприятий для оценки эффективности их стратегий интеллектуального анализа данных.

Сбор данных:

Интервью: Проведено с ключевыми заинтересованными сторонами на 10 предприятиях из разных отраслей.

Опрос: Проводится среди 50 предприятий, основное внимание уделяется инструментам, методам и результатам анализа данных.

Анализ данных:

Качественный анализ: Тематический анализ стенограмм интервью для выявления общих проблем и стратегий.

Количественный анализ: Статистический анализ ответов на опросы для оценки эффективности различных методов интеллектуального анализа данных.

РЕЗУЛЬТАТЫ

Интеллектуальный анализ больших данных - это сложный процесс, который включает в себя сбор, обработку, анализ и извлечение информации из огромных объемов данных. На предприятиях этот процесс необходим для принятия решений, основанных на данных, оптимизации операций и получения конкурентных преимуществ. Вот обзор ключевых компонентов интеллектуального анализа больших данных в корпоративном контексте:

1. Сбор данных

Источники: Предприятия собирают данные из различных источников, включая данные о транзакциях, взаимодействии с клиентами, социальные сети, датчики и многое другое.

Технологии: Сбор данных осуществляется с использованием таких технологий, как устройства Интернета вещей, CRM-системы, веб-аналитика, API и решения для хранилищ данных.

Проблемы: Сбор данных в больших масштабах может быть сложной задачей из-за огромного объема, разнообразия и скорости передачи данных. Обеспечение качества, актуальности и точности данных имеет решающее значение.

2. Хранение данных

Хранилища данных: Структурированные данные часто хранятся в хранилищах данных, которые оптимизированы для повышения производительности запросов и создания отчетов.

Хранилища данных: Неструктурированные или полуструктурированные данные обычно хранятся в хранилищах данных, которые предоставляют гибкие решения для хранения.

Облачное хранилище: Многие предприятия используют облачные платформы, такие как AWS, Google Cloud или Azure, для хранения больших объемов данных и управления ими, используя преимущества масштабируемых решений для хранения данных.

Проблемы: Эффективное управление массивными наборами данных и обеспечение их безопасности при одновременном обеспечении быстрого доступа и извлечения является серьезной проблемой.

3. Обработка данных

Пакетная обработка: При этом данные обрабатываются большими блоками, что полезно для задач, не требующих обработки данных в реальном времени. Широко используются такие технологии, как Hadoop и Apache Spark.

Потоковая обработка: Для обработки данных в режиме реального времени используются инструменты потоковой обработки, такие как Apache Kafka, Apache Flink и AWS Kinesis, которые обрабатывают непрерывные потоки данных.

ETL/ELT: Процессы извлечения, преобразования, загрузки (ETL) или Извлечения, загрузки, преобразования (ELT) используются для очистки, преобразования и загрузки данных в соответствующие системы хранения.

Аналитика больших данных
Big data analytics

Особенности больших данных	Проблемы в непрерывном аудите	Большой разрыв данных
• объем • разнообразие • скорость • достоверность	• конфликтующие данные, • неполные данные, • данные с различными идентификаторами, • данные в разных форматах, • недостоверные данные, • асинхронные данные, • поиск зашифрованных данных, • аудит зашифрованных данных, • аудит агрегированных данных	• идентификация, • целостность, • согласованность, • конфиденциальность, • агрегирование данных

4. Анализ данных

Методы интеллектуального анализа данных : Такие методы, как кластеризация, классификация, регрессия и изучение правил ассоциации, используются для выявления закономерностей и тенденций в данных.

Машинное обучение: Алгоритмы машинного обучения, включая контролируемое, неконтролируемое и подкрепляющее обучение, применяются для построения прогностических моделей и получения аналитической информации.

Инструменты: Предприятия используют такие инструменты, как Python (с такими библиотеками, как Pandas, Scikit-learn), R, SAS и Apache Mahout для анализа данных.

5. Визуализация данных

Инструменты: Инструменты визуализации, такие как Tableau, Power BI и D3.js помогают представить аналитические данные в визуальном формате, что облегчает лицам, принимающим решения, понимание сложных наборов данных.

Информационные панели: Предприятия часто создают информационные панели для мониторинга ключевых показателей и KPI в режиме реального времени.

6. Безопасность и конфиденциальность данных

Соответствие требованиям: Обеспечение соблюдения таких нормативных актов, как GDPR, CCPA и HIPAA, имеет решающее значение при обработке конфиденциальных данных.

Меры безопасности: Для защиты данных от несанкционированного доступа и взломов используются методы шифрования, контроля доступа и анонимизации.

7. Проблемы и рекомендации.

Масштабируемость: Управление масштабом больших данных является сложной задачей, и предприятиям необходимо инвестировать в масштабируемую инфраструктуру и инструменты.

Управление данными: Разработка четких политик управления данными помогает обеспечить качество данных, безопасность и соответствие требованиям.

Таланты: Привлечение и удержание квалифицированных специалистов по обработке данных, инженеров и аналитиков имеет важное значение для успешного анализа больших объемов данных.

8. Приложения

Анализ потребностей клиентов : Предприятия используют интеллектуальный анализ больших объемов данных для понимания поведения, предпочтений и настроений клиентов.

Операционная эффективность : Интеллектуальный анализ данных помогает оптимизировать операции, снизить затраты и улучшить управление цепочками поставок.

Управление рисками: Прогнозная аналитика используется для выявления потенциальных рисков и их упреждающего снижения.

Инновации: Аналитическая информация, основанная на данных, позволяет разрабатывать новые продукты и услуги, способствуя внедрению инноваций.

Подводя итог, можно сказать, что интеллектуальный анализ больших данных на предприятиях - это комплексный процесс, который включает в себя сбор огромных объемов данных, их эффективную обработку, анализ для получения информации и обеспечение их безопасности и соответствия требованиям. Этот процесс является основополагающим для разработки стратегий, основанных на данных, и достижения бизнес-целей в современную цифровую эпоху.

ОБСУЖДЕНИЕ

Результаты этого исследования подчеркивают важнейшую роль интеллектуального анализа больших данных на современных предприятиях. Эффективный сбор и обработка данных необходимы для получения полезной информации, способствующей росту бизнеса. Однако проблемы, выявленные в этом исследовании, особенно связанные с качеством данных и интеграцией новых технологий, подчеркивают необходимость постоянных инвестиций в инфраструктуру данных и развитие навыков.

Исследование также показывает, что, хотя передовые инструменты и алгоритмы могут значительно расширить возможности интеллектуального анализа данных, их успешное внедрение требует комплексного подхода. Это включает в себя не только внедрение правильных технологий, но и развитие культуры, основанной на данных, в организации.

ВЫВОДЫ

Интеллектуальный анализ больших данных - неотъемлемый компонент корпоративной стратегии в эпоху цифровых технологий. Поскольку предприятия продолжают генерировать огромные объемы данных и полагаться на них, способность эффективно собирать, обрабатывать и анализировать эти данные станет ключевым фактором успеха. В исследовании делается вывод о том, что предприятия должны решать проблемы, связанные с качеством данных, конфиденциальностью и нехваткой навыков, чтобы в полной мере реализовать потенциал интеллектуального анализа больших данных.

Инвестируйте в качество данных: Предприятиям следует уделять приоритетное внимание инициативам в области качества данных, чтобы гарантировать точность, согласованность и надежность собираемых данных.

Внедрите передовые средства обработки данных: Компаниям следует изучить возможность внедрения передовых средств обработки данных, таких как Apache Spark, для расширения своих возможностей интеллектуального анализа данных.

Развивать культуру, основанную на данных : Организации должны развивать культуру, в которой ценится принятие решений, основанных на данных, при поддержке постоянного обучения и повышения квалификации сотрудников.

Решение проблем конфиденциальности: Предприятия должны внедрять надежные системы управления данными для решения проблем конфиденциальности и обеспечения соблюдения соответствующих нормативных актов.

СПИСОК ЛИТЕРАТУРЫ

1. Багутдинов Р.А. Исследование новейших информационно-коммуникативных технологий в среднем профессиональном образовании. В сборнике: Научный поиск в XXI веке. Материалы I международной научной конференции по евразийскому научному сотрудничеству. Под редакцией В.А. Должикова. 2014. С. 39–42.

2. Багутдинов Р.А. Проектирование модульной мультисенсорной системы для задач мониторинга окружающей среды на базе Arduino. Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика. 2019. 46 (1): 173–180.

3. Багутдинов Р.А. Подход к обработке, классификации и обнаружению новых классов и аномалий в разнородных и разномасштабных потоках данных. Вестник Дагестанского государственного технического университета. Технические науки. 2018. 45 (3): 85–93.

4. Багутдинов Р.А. Разработка мультисенсорной системы для задач мониторинга и интерпретации разнородных данных. Системный администратор. 2019. 3 (196): 82–85.

5. Островский О.А. Алгоритм мероприятий по анализу ситуации при подозрении в совершении преступлений в сфере компьютерной информации с

учетом специфики источников данных этой информации. Право и политика. 2018. 10: 32–37.

6. Островский О.А. Аспекты современных проблем расследования преступлений, связанных с изъятием цифровых следов и предоставлением соответствующих доказательств. Вестник Алтайской академии экономики и права. 2019. 3: 146–151.

7. Островский О.А., Шевелева И.А. Проблематика формирования и правового регулирования больших данных в исследовании информационных цифровых следов. В сборнике: Уголовное производство: процессуальная теория и криминалистическая практика. Материалы VIII Международной научно-практической конференции. Отв. редакторы М.А. Михайлов, Т.В. Омельченко. 2020. С. 57–59.

8. Kabacoff R. R. Data analysis and graphics with. Manning Publications Co.; 2015 Mar 3.

9. Zhang J, Yang X, Appelbaum D. Toward effective Big Data analysis in continuous auditing. Accounting Horizons. 2015 Jun; 29 (2): 469–76