

5. ТЕХНОЛОГИИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ DATA MINING

Драб А.А., студент гр.374002

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Канах А.В. – старший преподаватель каф. ЭИ

Аннотация. Современное информационное общество характеризуется интенсивным накоплением и обменом данными, что делает традиционные методы анализа неэффективными. В данной работе рассматриваются теоретические и прикладные аспекты интеллектуального анализа данных (Data Mining): определение понятия, этапы анализа, ключевые методы и алгоритмы, практические области применения, современные инструменты и перспективы развития. Подчеркивается значимость интеграции Data Mining в научные исследования, бизнес-процессы, здравоохранение и другие сферы. Особое внимание уделяется русскоязычным источникам и отечественным разработкам.

Ключевые слова. интеллектуальный анализ данных; Data Mining; извлечение знаний; большие данные; машинное обучение; обработка информации; интерпретация данных; автоматизация анализа; инструменты Data Mining; искусственный интеллект.

Data mining (интеллектуальный анализ данных) – процесс использования компьютеров и автоматизации для поиска в больших наборах данных закономерностей и тенденций, превращение этих результатов в бизнес-идеи и прогнозы.

Концепция интеллектуального анализа данных существовала ещё до появления компьютеров. Статистическое начало интеллектуального анализа данных было положено теоремой Байеса в 1763 году и открытием регрессионного анализа в 1805 году. Термин «интеллектуальный анализ данных», был введён в 1989 году Григорием Пятецким-Шапиро. Благодаря статистике (числовое исследование взаимосвязей данных), искусственному интеллекту (человеческий интеллект, отображаемый программным обеспечением и/или машинами) и машинному обучению (алгоритмы, которые могут учиться на данных, чтобы делать прогнозы) была создана основа для нашего современного понимания того, что такое интеллектуальный анализ данных сегодня. А по мере бурного роста компьютерных процессоров, систем хранения данных и технологий в 1990-х и 2000-х годах интеллектуальный анализ данных стал не только более мощным, но и более продуктивным во всех ситуациях.

В 2003 году книга «Moneyball» познакомила гораздо более широкую аудиторию с интеллектуальным анализом данных, рассказывая об аналитическом подходе профессиональной бейсбольной команды к составлению состава. Сегодня, когда компании используют решения для больших данных во все большем разнообразии ситуаций, интеллектуальный анализ данных играет решающую роль во многих отраслях.

Методы интеллектуального анализа данных

В интеллектуальном анализе данных используются алгоритмы и различные другие методы для преобразования больших коллекций данных в полезные выходные данные.

Среди самых распространенных методов интеллектуального анализа данных можно выделить следующие:

Классификационный анализ

С помощью этого метода точки данных распределяются по группам или классам на основе конкретного вопроса или проблемы, которую необходимо решить. Например, если компания, производящая потребительские товары, хочет оптимизировать свою стратегию скидок по купонам для конкретного продукта, она может проанализировать уровни запасов, данные о продажах, коэффициенты погашения купонов и данные о поведении потребителей, чтобы принять наилучшее решение.

Изучение правил ассоциации

Эта функция направлена на раскрытие взаимосвязей между точками данных; он используется для определения того, имеет ли конкретное действие или переменная какие-либо характеристики, которые можно связать с другими действиями.

Обнаружение аномалий или выбросов

Помимо поиска закономерностей, интеллектуальный анализ данных направлен на обнаружение необычных данных в наборе. Обнаружение аномалий — это процесс поиска данных, которые не соответствуют шаблону. Этот процесс может помочь выявить случаи мошенничества и помочь ритейлерам узнать больше о скачках или падениях продаж определенных продуктов.

Кластерный анализ

Кластеризация ищет сходства в наборе данных, разделяя точки данных, имеющие общие черты, на подмножества. Это похоже на классификационный тип анализа в том, что он группирует точки данных, но при кластерном анализе данные не относят к ранее определенным группам. Кластеризация полезна для определения характеристик в наборе данных, таких как сегментация клиентов на основе покупательского поведения, состояния потребностей, жизненного этапа или вероятных предпочтений в маркетинговых коммуникациях.

Регрессионный анализ

Регрессионный анализ позволяет понять, какие факторы в наборе данных являются наиболее важными, а какие можно игнорировать, и как эти факторы взаимодействуют. Использование регрессионного анализа максимизирует продажи, сводит к минимуму случаи отсутствия товара на складе, а также помогает избежать его излишки.

Дерево решений

Этот процесс классифицирует или прогнозирует потенциальные результаты, используя методы классификации или регрессии. Древовидные структуры используются для представления потенциальных результатов решений.

Нейронные сети

Нейронная сеть — это набор алгоритмов, моделирующих деятельность человеческого мозга, где данные обрабатываются с помощью узлов. Нейронные сети особенно полезны в сложных приложениях распознавания образов, включающих глубокое обучение — более продвинутое ответвление машинного обучения.

K-Nearest Neighbor (KNN)

Этот метод интеллектуального анализа данных классифицирует данные на основе их близости к другим точкам данных. Предполагая, что близлежащие точки данных больше похожи друг на друга, чем другие точки данных, KNN используется для прогнозирования групповых признаков.

Применение интеллектуального анализа

Предприятия используют интеллектуальный анализ данных, чтобы получить конкурентное преимущество, используя данные, которые они собирают о своих клиентах, продуктах, продажах, а также рекламных и маркетинговых кампаниях. Интеллектуальный анализ данных помогает им совершенствовать операции, улучшать отношения с текущими клиентами и приобретать новых клиентов.

Основные способы использования интеллектуального анализа данных компаниями:

Анализ корзины

В самом базовом применении розничные торговцы используют анализ корзин для анализа того, что потребители покупают. Это форма метода ассоциации, позволяющая ритейлерам понять покупательские привычки и рекомендовать другие покупки.

Прогноз продаж

Прогнозирование продаж — это форма прогнозного анализа, которому предприятия посвящают большую часть своих бюджетов. Интеллектуальный анализ данных может помочь предприятиям спрогнозировать продажи и установить цели путем изучения исторических данных, таких как записи продаж, финансовые показатели, привычки потребительских расходов, продажи, относящиеся к определенному времени года, и тенденции, которые могут повлиять на стандартные предположения о бизнесе.

Маркетинг баз данных

Компании создают большие базы данных о потребителях, которые они используют для формирования и направления своих маркетинговых усилий. Этим предприятиям нужны способы управления и использования этих данных для разработки целевых, персонализированных маркетинговых коммуникаций. Интеллектуальный анализ данных помогает компаниям понимать поведение потребителей, отслеживать контактную информацию и потенциальных клиентов, а также привлекать больше клиентов к своим маркетинговым базам данных.

Планирование запасов

Интеллектуальный анализ данных может предоставить предприятиям актуальную информацию о запасах продукции, графиках поставок и производственных требованиях. Интеллектуальный анализ данных также может помочь устранить некоторую неопределенность, связанную с простыми проблемами спроса и предложения в цепочке поставок. Скорость, с которой выявляются закономерности и составляются прогнозы, помогает компаниям лучше управлять своими запасами продукции и работать более эффективно.

Лояльность клиентов

Предприятия, особенно розничные торговцы, генерируют огромное количество данных с помощью программ лояльности. Интеллектуальный анализ данных позволяет этим предприятиям строить и улучшать отношения с клиентами с помощью этих данных. Например, группируя клиентов по

общей сумме корзин, частоте покупок и вероятным расходам на продукты в неделю, розничные продавцы могут предлагать покупателям скидки, чтобы «довести» их до уровня расходов.

К основным сферам использования относятся:

Телекоммуникация, СМИ

Телекоммуникационные и медиа компании могут использовать аналитические модели для анализа огромного количества данных о клиентах, помогая им прогнозировать поведение клиентов и предлагать узконаправленные и актуальные кампании.

Страхование

Страховые компании решают сложные проблемы, связанные с мошенничеством, соблюдением требований, управлением рисками и оттоком клиентов. Компании используют методы интеллектуального анализа данных, чтобы более эффективно оценивать продукты по всем направлениям бизнеса и находить новые способы предлагать конкурентоспособные продукты своей существующей клиентской базе.

Образование

Благодаря унифицированному, основанному на данных представлению об успеваемости учащихся преподаватели могут прогнозировать успеваемость учащихся еще до того, как они вступят в класс, и разрабатывать стратегии вмешательства, которые помогут им не сбиться с курса. Интеллектуальный анализ данных помогает преподавателям получать доступ к данным учащихся, прогнозировать уровни успеваемости и выявлять учащихся или группы учащихся, нуждающихся в дополнительном внимании.

Производство

Важное значение имеет согласование планов поставок с прогнозами спроса, а также раннее выявление проблем, обеспечение качества и инвестиции в капитал бренда. Производители могут прогнозировать износ основных средств и планировать техническое обслуживание, что позволяет максимально увеличить время безотказной работы и поддерживать график производственной линии.

Банковское дело

Автоматизированные алгоритмы помогают банкам понять свою клиентскую базу, а также миллиарды транзакций, лежащих в основе финансовой системы. Интеллектуальный анализ данных помогает компаниям, предоставляющим финансовые услуги, лучше понимать рыночные риски, быстро обнаруживать и предотвращать мошенничество, выполнять обязательства по соблюдению нормативных требований и получать оптимальную прибыль от своих маркетинговых инвестиций.

Розничная торговля

Большие базы данных клиентов содержат скрытую информацию о клиентах, которая может помочь улучшить отношения, оптимизировать маркетинговые кампании и прогнозировать продажи. Благодаря более точным моделям данных и маркетинговой аналитике розничные компании могут предлагать более целевые направления и находить предложения, которые оказывают наибольшее влияние на клиента.

Управление персоналом

Отделы кадров обычно работают с большими объемами данных. Сюда входят данные об удержании, продвижении по службе, зарплате и льготах. Интеллектуальный анализ данных сравнивает эти данные, чтобы лучше помочь процессам управления персоналом.

Социальные сети

Компании, занимающиеся социальными сетями, используют интеллектуальный анализ данных для сбора больших объемов данных о пользователях и их действиях в Интернете.

Преимущества:

Прибыльность и эффективность: интеллектуальный анализ данных гарантирует, что компания собирает и анализирует надежные данные. Часто это более жесткий, структурированный процесс, который формально идентифицирует проблему, собирает данные, связанные с проблемой, и пытается сформулировать решение. Таким образом, интеллектуальный анализ данных помогает бизнесу стать более прибыльным, более эффективным, более сильным.

Широкие возможности применения: интеллектуальный анализ данных может выглядеть по-разному в разных приложениях, но в целом процесс можно использовать практически с любым новым или устаревшим приложением. По сути, можно собирать и анализировать любые типы данных и почти каждую бизнес-задачу, которая опирается на поддающиеся проверке доказательства.

Скрытая информация и тенденции: конечная цель интеллектуального анализа данных — взять необработанные биты информации и определить, существует ли связность или корреляция между данными. Это преимущество позволяет компании создавать ценность с помощью имеющейся у нее информации, которая в противном случае не была бы слишком очевидной. Хотя модели данных могут быть сложными, они также могут давать потрясающие результаты, выявлять скрытые тенденции и предлагать уникальные стратегии.

Недостатки:

Сложность: один из его самых больших недостатков. Аналитика данных часто требует набора технических навыков и определенных программных инструментов. Небольшие компании могут счесть этот барьер входа слишком трудным для преодоления.

Никаких гарантий: интеллектуальный анализ данных не всегда означает гарантированные результаты. Компания может проводить статистический анализ, делать выводы на основе достоверных данных, вносить изменения и не получать никаких выгод. Это может быть связано с неточными выводами, изменениями на рынке, ошибками моделей или неподходящим набором данных. Интеллектуальный анализ данных может только направлять решения, но не гарантировать результаты.

Высокая стоимость: интеллектуальный анализ данных также требует затрат. Получение некоторых данных может оказаться дорогостоящим. Проблемы безопасности и конфиденциальности можно решить, хотя дополнительная ИТ-инфраструктура также может оказаться дорогостоящей. Интеллектуальный анализ данных может быть наиболее эффективным при использовании огромных наборов данных; однако эти наборы данных необходимо хранить, а для анализа требуются большие вычислительные мощности.

Технология анализа данных с успехом применяется в мире. В 2021 году мировой рынок услуг по анализу данных составлял \$162,6 млрд, а в 2023 уже \$220,2 млрд, что говорит о том, что данные технологии востребованы во всем мире.

Так как для анализа и обработки данных нужны огромные вычислительные мощности, сейчас в мире активно создаются центры обработки данных (ЦОД), что дополнительно стимулирует развитие наукоёмких отраслей экономики, таких как электроника, микросхемотехника и другие области IT-сферы.

Применительно к Республике Беларусь, технологии интеллектуального анализа данных широко используются в банковской сфере для определения кредитной политики. На основе анализа данных принимается решение о выдаче кредита, так как анализ данных выдает вероятность возвращения кредита. Также в нашей стране технологии анализа данных, кроме банковской сферы, применяются в сельском хозяйстве, в сфере ЖКХ, рекламе. У нашей страны есть высокий потенциал в развитии сферы анализа данных, так как еще со времен СССР мы сохранили серьезную школу математики, которая является основой анализа данных.

Список использованных источников:

1. Что такое интеллектуальный анализ данных (Data Mining) [Электронный ресурс] // Trend Micro. – URL: https://www.trendmicro.com/ru_ru/what-is/machine-learning/data-mining.html (дата обращения: 22.03.2025).
2. Что такое интеллектуальный анализ данных (Data Mining) [Электронный ресурс] // Amazon Web Services. – URL: <https://aws.amazon.com/ru/what-is/data-mining/> (дата обращения: 22.03.2025).
3. Технология Data Mining [Электронный ресурс] // Decosystems. – URL: <https://www.decosystems.ru/tekhnologiya-data-mining/> (дата обращения: 22.03.2025).
4. Зубков Е. В., Белов В. М. Методы интеллектуального анализа данных и обнаружение вторжений [Электронный ресурс] // КиберЛенинка. – URL: <https://cyberleninka.ru/article/n/metody-intellektualnogo-analiza-dannyh-i-obnaruzhenie-vtorzheniy/pdf> (дата обращения: 22.03.2025).
5. 10 основных методов интеллектуального анализа данных [Электронный ресурс] // Astera. – URL: <https://www.astera.com/ru/type/blog/top-10-data-mining-techniques/> (дата обращения: 22.03.2025).

UDC 004.65-047.44

DATA MINING TECHNOLOGIES

Drab A.A., student group 374002

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Kanash A.V. – Senior lecturer at the Department of Economic Informatics

Annotation. The modern information society is characterized by intensive accumulation and exchange of data, which makes traditional methods of analysis ineffective. This paper examines the theoretical and applied aspects of data mining: definition of the concept, stages of analysis, key methods and algorithms, practical applications, modern tools and development prospects. The importance of integrating Data Mining into scientific research, business processes, healthcare and other fields is emphasized. Special attention is paid to Russian-language sources and domestic developments.

Keywords. Data Mining; knowledge extraction; big data; machine learning; information processing; data interpretation; analysis automation; Data Mining tools; artificial intelligence.