

*Арсентьева Н.В., магистрант
кафедры прикладной информатики и информационных технологий
Скрипин А.А, аспирант кафедры прикладной информатики и
информационных технологий
Скрипина И. И., старший преподаватель кафедры прикладной
информатики и информационных технологий
Белгородский государственный национальный исследовательский
университет, г. Белгород*

DATA MINING: МЕТОДЫ, ЭТАПЫ, ПРИМЕНЕНИЕ И ЗНАЧЕНИЕ В СОВРЕМЕННОМ МИРЕ

Аннотация: Данная статья посвящена интеллектуальному анализу данных (Data Mining) и его роли в современном мире. В работе рассматриваются основные концепции Data Mining, его значение в эпоху цифровой трансформации и экспоненциального роста объемов данных. Описываются ключевые этапы процесса интеллектуального анализа данных, включая постановку задачи, подготовку и изучение данных, построение и проверку моделей, а также их развертывание и обновление. Особое внимание уделяется применению технологий Data Mining в различных сферах, таких как бизнес и экономика, наука и исследования, здравоохранение, финансовый сектор, государственное управление и образование. Статья демонстрирует, как использование методов интеллектуального анализа данных позволяет организациям оптимизировать процессы, улучшить принятие решений и повысить конкурентоспособность.

Ключевые слова: Data Mining, интеллектуальный анализ данных, большие данные, машинное обучение, прогнозирование, оптимизация бизнес-процессов, принятие решений, цифровая трансформация, аналитика, искусственный интеллект.

*Arsentieva N.V., Undergraduate
student of the Department of Applied Informatics and Information
Technology*

*Skripin A.A., post-graduate student of the Department of Applied
Informatics and
Information Technologies*

*I. I. Skripina, Senior Lecturer at the Department of Applied
Informatics and Information Technology
Belgorod State National Research
University, Belgorod*

DATA MINING: METHODS, STAGES, APPLICATION AND SIGNIFICANCE IN THE MODERN WORLD

Abstract: This article is devoted to data mining and its role in the modern world. The paper examines the basic concepts of Data Mining, its importance in the era of digital transformation and exponential growth of data volumes. The key stages of the data mining process are described, including task formulation, data preparation and study, model construction and validation, as well as their deployment and updating.

Special attention is paid to the application of Data Mining technologies in various fields, such as business and economics, science and

В эпоху цифровой трансформации и экспоненциального роста объемов данных, технологии интеллектуального анализа данных, известные как Data Mining, приобретают все большее значение. Data Mining представляет собой междисциплинарную область, объединяющую методы статистики, машинного обучения и баз данных для извлечения ценной информации из больших массивов данных.

Data Mining (добыча данных, интеллектуальный анализ данных, глубинный анализ данных) можно определить, как процесс обнаружения в «сырых» данных ранее неизвестных, нетривиальных, практически полезных

и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Также для этой технологии используется менее популярный термин «обнаружение знаний в данных» или KDD (knowledge discovery in databases). Этот процесс включает в себя не только сбор и анализ данных, но и выявление скрытых закономерностей, построение моделей и формирование гипотез, которые могут быть использованы для прогнозирования и принятия стратегических решений.

В современном мире на момент написания данной статьи Data Mining уже играет ключевую роль в различных аспектах жизни общества и функционирования бизнеса. Данная технология применяется в самых разнообразных сценариях в различных отраслях. Цикл использования технологий представлен на рисунке 1.



Рис. 1. Этапы интеллектуального анализа данных [7]

Процесс Data Mining представляет собой комплексный набор этапов, направленных на систематическое извлечение и анализ данных с целью выявления скрытых закономерностей и получения ценной информации. Этапы интеллектуального анализа данных представлены на рисунке 2.

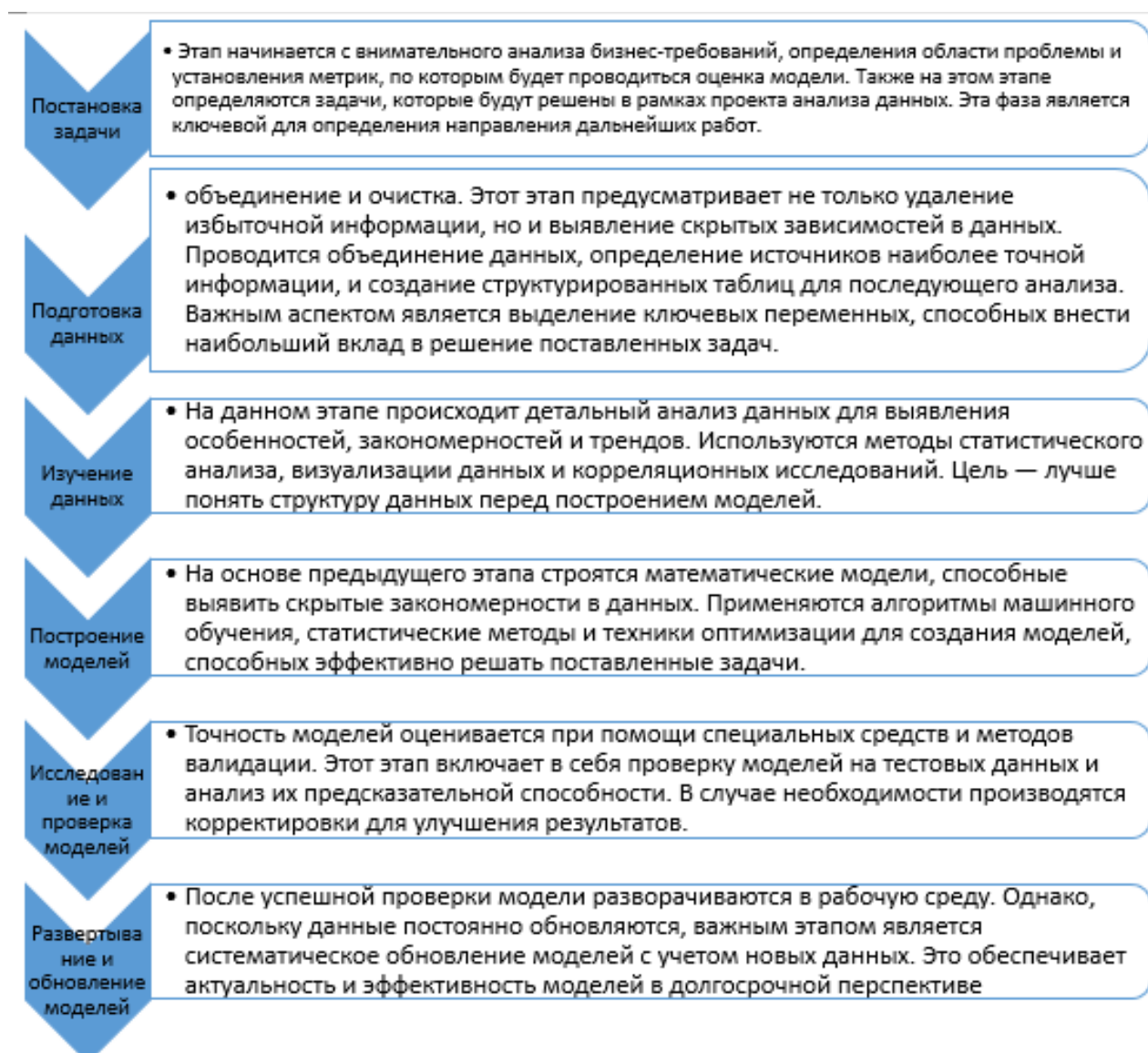


Рис. 2. Этапы интеллектуального анализа данных

Каждый из этих этапов играет важную роль в процессе Data Mining, обеспечивая систематический подход к извлечению знаний из данных и их применению для решения практических задач. Важно отметить, что процесс Data Mining часто является итеративным, где результаты одного этапа могут привести к пересмотру предыдущих этапов для достижения оптимальных результатов. Рассмотрим аналитические методы, которые могут использоваться на каждом из этих этапов (таблица 1).

Таб.1. Аналитические методы

№	Метод	Суть метода
1.	Постановка задачи:	<ul style="list-style-type: none"> — Методы анализа бизнес-требований, такие как SWOT-анализ или анализ целей и задач бизнеса. — Техники определения ключевых метрик и показателей проекта, например, использование KPI (ключевых показателей эффективности).
2.	Подготовка данных	<ul style="list-style-type: none"> — Методы очистки данных, такие как удаление дубликатов, заполнение пропущенных значений и устранение выбросов. — Алгоритмы объединения данных из различных источников, например, методы интеграции данных.
3.	Изучение данных	<ul style="list-style-type: none"> — Аналитические методы, такие как анализ средних, медианы, моды, дисперсии и корреляции. — Визуализация данных с использованием графиков, диаграмм и тепловых карт для выявления особенностей, и трендов.
4.	Построение моделей	<ul style="list-style-type: none"> — Применение алгоритмов машинного обучения, таких как линейная регрессия, деревья решений, метод опорных векторов и нейронные сети. — Использование статистических методов, например, методов кластерного анализа или анализа факторов.
5.	Исследование и проверка моделей	<ul style="list-style-type: none"> — Методы валидации моделей, включая кросс-валидацию, отложенную выборку и бутстрап. — Анализ показателей точности модели, таких как точность, полнота, F1-мера и ROC-кривая.
6.	Развертывание и обновление моделей	<ul style="list-style-type: none"> — Методы мониторинга производительности моделей в реальном времени — Принятие решений об обновлении моделей на основе регулярного анализа новых данных и результатов.

Использование разнообразных аналитических методов на каждом из этих этапов позволяет создать надежные и эффективные модели Data Mining, способные предоставить ценную аналитическую информацию для принятия решений в бизнесе.

Методы анализа данных, направленные на применение искусственного интеллекта, используются в различных областях для разнообразных целей:

- Прогнозирование: оценка перспектив продаж, предсказание нагрузки на сервер или времени его простоя;

- Оценка рисков и вероятностей: выбор подходящих кандидатов для целевой рассылки, определение точек баланса для рискованных сценариев, присвоение вероятностей диагнозам или другим исходам;

- Предоставление рекомендаций: выявление продуктов, которые часто покупают вместе, формирование рекомендательных сообщений;

- Анализ последовательностей: изучение выбора клиентов в процессе покупок, прогноз их поведения;

- Группировка: классификация клиентов или событий в кластеры, анализ и прогнозирование общих характеристик этих кластеров.

Области применения технологий Data Mining для интеллектуального анализа данных:

- Бизнес и экономика. В мире бизнеса и экономики, технологии интеллектуального анализа данных широко используются для анализа рыночных тенденций, прогнозирования спроса, оптимизации производственных процессов и принятия стратегических решений. С помощью этих технологий компании могут классифицировать клиентов, проводить анализ рисков, оптимизировать маркетинговые кампании и повышать эффективность операций.

- Наука и исследования. В области науки и исследований интеллектуальный анализ данных играет ключевую роль в обработке и анализе больших массивов данных, выявлении паттернов и закономерностей,

прогнозировании результатов экспериментов и создании новых научных моделей. Эти технологии помогают ученым и исследователям улучшать качество и точность их исследований, делая процесс открытия новых знаний более эффективным.

— **Здравоохранение.** В здравоохранении интеллектуальный анализ данных применяется для улучшения диагностики и лечения, оптимизации процессов управления медицинскими данными, прогнозирования распространения заболеваний, а также для разработки персонализированной медицины. Анализ данных пациентов и медицинских исследований помогает улучшить качество медицинского обслуживания и принимать обоснованные медицинские решения.

— **Финансовый сектор.** В финансовой отрасли интеллектуальный анализ данных используется для анализа рыночных трендов, прогнозирования финансовых показателей, выявления мошенничества, оптимизации портфеля инвестиций и управления рисками. Эти технологии помогают финансовым учреждениям принимать обоснованные решения, улучшать операционную эффективность и повышать конкурентоспособность на рынке.

— **Государственное управление.** В сфере государственного управления интеллектуальный анализ данных применяется для анализа социально-экономических показателей, мониторинга реализации государственных программ, принятия политических решений и улучшения качества государственных услуг. Эти технологии позволяют государственным органам повысить эффективность своей деятельности, улучшить взаимодействие с гражданами и обеспечить прозрачность власти.

— **Образование.** В образовании интеллектуальный анализ данных используется для анализа успеваемости студентов, оценки эффективности образовательных программ, персонализации обучения и развития инновационных методик обучения. Эти технологии помогают учебным

заведениям улучшить образовательный процесс, повысить качество обучения и подготовить студентов к вызовам современного мира.

— Технологии интеллектуального анализа данных эффективно интегрируются в различные отрасли, обеспечивая компаниям и организациям возможность оптимизировать операции, улучшать принятие решений и повышать конкурентоспособность.

Интеллектуальный анализ данных (Data Mining) представляет собой мощный инструмент для извлечения ценной информации из больших массивов данных. В современном мире, характеризующемся экспоненциальным ростом объемов информации, технологии Data Mining приобретают все большее значение во многих сферах человеческой деятельности.

Применение методов Data Mining позволяет организациям получать конкурентные преимущества, оптимизировать процессы и принимать более обоснованные решения. В бизнесе это выражается в улучшении понимания поведения потребителей, оптимизации маркетинговых стратегий и повышении операционной эффективности. В науке и исследованиях Data Mining способствует выявлению новых закономерностей и формулированию гипотез. В здравоохранении эти технологии применяются для улучшения диагностики и персонализации лечения.

Важно отметить, что процесс Data Mining требует систематического подхода, включающего несколько этапов от постановки задачи до развертывания и обновления моделей. Каждый этап предполагает использование специфических аналитических методов и инструментов.

По мере развития технологий и увеличения доступности данных, роль Data Mining в различных отраслях будет только возрастать. Это открывает новые возможности для инноваций, повышения эффективности и создания ценности в самых разных сферах деятельности.

В заключение можно сказать, что овладение методами и техниками Data Mining становится необходимым навыком для специалистов в различных областях, а организации, эффективно использующие эти технологии, получают значительные преимущества в современном конкурентном мире.

Литература:

1. Барышков Кирилл Васильевич ИСПОЛЬЗОВАНИЕ БОЛЬШИХ ДАННЫХ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ GO-TO-MARKET СТРАТЕГИЙ // Практический маркетинг. 2024. №5. URL: <https://cyberleninka.ru/article/n/ispolzovanie-bolshih-dannyh-dlya-povysheniya-effektivnosti-go-to-market-strategiy-1> (дата обращения: 22.06.2024).

2. Абриков Виктор Сергеевич, Кочаков Валерий Данилович, Смирнов Александр Вячеславович, Абриков Сергей Викторович, Васильев Алексей Иванович Многофакторные вычислительные модели фотодатчиков на основе тонких пленок // Вестник ЧГУ. 2015. №1. URL: <https://cyberleninka.ru/article/n/mnogofaktornye-vychislitelnye-modeli-fotodatchikov-na-osnove-tonkih-plenok> (дата обращения: 24.06.2024).

3. Костюкова Н. И. Система принятия решений по технологии Data Mining // Перспективы систем информатики: материалы Седьмой международной конференции памяти академика А. Е. Ершова. Новосибирск, 2009. С. 72-76.

4. Голиков Виктор Андреевич Применение технологии Big Data в финансовом секторе // Известия ТулГУ. Экономические и юридические науки. 2018. №2-1. URL: <https://cyberleninka.ru/article/n/primenenie-tehnologii-big-data-v-finansovom-sektore> (дата обращения: 20.06.2024).

5. Косоруков А.А., Кшеменецкая М.Н. Большие данные в практике управления современным государством/ Год: 2019 Страницы: 74-81/ eLIBRARY ID: 37146245

6. Мокшанов, Михаил / Михаил Мокшанов. — Текст : электронный // : [сайт]. — URL: <https://www.itweek.ru/bigdata/article/detail.php?ID=229390&ysclid=lxujisdis644865546> (дата обращения: 25.06.2024).