

УДК 004.8.

**ВВЕДЕНИЕ В ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ****А.И. ПИРОВИЧ***(Представлено: канд. техн. наук, доц. И.Б. БУРАЧЕНОК)*

*В данной статье рассматривается интеллектуальный анализ данных, описываются области его применения.*

Интеллектуальный анализ данных – это изучение большого количества исходных данных для выделения из них полезной информации, что дает компаниям преимущество перед их конкурентами. Эта информация состоит из значимых формул, цифр, статистики и других необходимых данных, которые уже присутствуют в исходной информации, но ранее были не очевидны для человека.

Наиболее популярный инструмент, который используется при анализе – искусственный интеллект (ИИ). Технологии искусственного интеллекта пытаются работать по примеру человеческого мозга, а именно – путем разумного предположения, например, обучения, и использование дедуктивного мышления. Некоторые из наиболее распространенных методов искусственного интеллекта используются в интеллектуальном анализе данных, например, нейронные сети, кластеризация и деревья решений [1].

Остановимся более подробно на каждом из них.

Правило работы нейронной сети основано на связях между данными или на использовании образца исходных данных. В результате, программное обеспечение непрерывно анализирует значения и сравнивает его с другими значениями, и программное обеспечение сравнивает эти значения до тех пор, пока не найдет новый шаблон. Эти шаблоны могут быть названы правилами. Далее программное обеспечение ищет другие шаблоны, которые основаны на этих правилах, или отправляет оповещение, когда находит необходимое значение.

Кластеризация делит данные на группы, основываясь на схожих чертах или на ограниченных диапазонах данных. Кластеры используются, когда данные не помечены как наилучшими для использования интеллектуального анализа данных. В качестве примера можно рассмотреть страховую компанию, которая пытается выявить случаи мошенничества. В документах компании изначально такие записи не будут храниться помеченными как мошеннические или фальсифицированные. Однако, после анализа модели в рамках кластеров, программное обеспечение, которое проведет анализ, может начать выявлять правила, которые указывают на необходимые записи, а именно на записи с ошибками, либо записи мошеннического характера. Задачу кластеризации формулировали в различных направлениях, например, статистика, распознавание образов, оптимизация процессов, машинное обучение и т.п. Поэтому существует много синонимов кластеру – класс, таксон, сгущение [2].

Деревья решений, так же, как и кластеры, разделяют данные на подмножества, а затем анализируют подмножества, разделяя их еще на более мелкие подмножества, и так далее (еще на несколько уровней при необходимости). Финальные подмножества так малы, что анализирующий процесс может найти неочевидные модели и взаимосвязи внутри данных.

Когда данные для анализа выявлены, они должны быть очищены. Очистка освобождает данные от дубликатов информации и от ошибок. Далее, данные должны быть сохранены в едином формате в необходимых категориях или полях. Инструменты для анализа могут работать со всеми типами хранения данных, от больших хранилищ данных (баз данных) до простых файлов. Хранилища данных и витрины данных это методы хранения, которые включают архивирование больших объемов данных таким образом, что позволяет легко получить доступ в случае необходимости.

Когда процесс завершается, анализирующее программное обеспечение генерирует отчет. Аналитик читает отчет, для того чтобы понять, необходимо ли выполнять дальнейшую обработку данных, такую как очистка параметров, использование других инструментов анализа данных для изучения данных, или даже переработка данных, если они не пригодны для использования. Если не требуется проведение дальнейших работ, отчет передается ответственным людям для принятия соответствующих мер.

Возможности интеллектуального анализа данных могут использоваться для многих областей, таких как здравоохранение, сбор статистики из СМИ, решение сложных задач в производственных процессах, а также анализа последовательностей в человеческих генах. Действительно нет ограничений для бизнеса и областей изучения, где интеллектуальный анализ данных может быть полезным.

В целом, преимущества интеллектуального анализа данных заключаются в способности выявлять скрытые шаблоны и взаимосвязи в данных, которые можно использовать для прогнозирования, влияющего на бизнес. Конкретные преимущества интеллектуального анализа данных зависят от цели и отрасли. Отделы продаж и маркетинга могут собирать данные о клиентах, чтобы повысить коэффициент конверсии

потенциальных клиентов или создать индивидуальные маркетинговые кампании. Данные интеллектуального анализа данных об исторических моделях продаж и поведении клиентов могут использоваться для построения моделей прогнозирования будущих продаж, новых продуктов и услуг.

Компании в финансовой отрасли используют инструменты интеллектуального анализа данных для построения моделей риска и выявления случаев мошенничества. В обрабатывающей промышленности используются инструменты интеллектуального анализа данных для повышения безопасности продукции, выявления проблем с качеством, управления цепочкой поставок и улучшения операций.

Вышеизложенные данные используются для изучения принципов работы интеллектуального анализа данных, которые способствуют разработке мобильного приложения для распознавания одежды из фото и поиска данной одежды в интернет-магазинах.

#### ЛИТЕРАТУРА

1. Интеллектуальный анализ данных [Электронный ресурс]. – Режим доступа: <http://translated-by.com/you/data-mining/into-ru/trans>. – Дата доступа: 20.09.2019.
2. Алгоритмы кластеризации на службе Data Mining [Электронный ресурс]. – Режим доступа: <https://basegroup.ru/community/articles/datamining>. – Дата доступа: 20.09.2019.