

*Эта часть работы выложена в ознакомительных целях. Если вы хотите получить работу полностью, то приобретите ее воспользовавшись формой заказа на странице с готовой работой:*

<https://stuservis.ru/referat/377781>

**Тип работы:** Реферат

**Предмет:** Информатика

Оглавление

Введение 3

История 4

Сущность интеллектуального анализа данных 7

Типы интеллектуального анализа данных 11

Использование интеллектуального анализа данных 13

Заключение 17

Список литературы 18

Введение

Чем больше данных мы собираем, тем сложнее становится разобраться во всех этих данных и извлечь из них значимую информацию. Представьте себе, что вы стоите среди триллионов деревьев; с чего начать анализ леса?

Интеллектуальный анализ данных предлагает решение этой проблемы, которое определяет способы принятия решений компаниями, сокращения затрат и увеличения доходов. В результате различные специалисты по обработке и анализу данных используют майнинг как часть своих повседневных обязанностей.

Интеллектуальный анализ данных часто воспринимается как сложный для понимания процесс. Однако изучение этой важной дисциплины науки о данных не так сложно, как кажется.

Интеллектуальный анализ данных чаще всего определяется как процесс использования компьютеров и автоматизации для поиска в больших наборах данных закономерностей и тенденций, превращая эти результаты в бизнес-идеи и прогнозы. Интеллектуальный анализ данных выходит за рамки процесса поиска, поскольку он использует данные для оценки будущих вероятностей и разработки действенного анализа.

История

Знаете ли вы, что концепция интеллектуального анализа данных существовала до появления компьютеров?

Статистическое начало интеллектуального анализа данных было положено в ход теоремой Байеса в 1763 году и открытием регрессионного анализа в 1805 году. Благодаря универсальной машине Тьюринга (1936 год), открытию нейронных сетей (1943 год), развитию баз данных (1970-е годы) и генетические алгоритмы (1975) и обнаружение знаний в базах данных.(1989), была подготовлена почва 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.

В 1990-х годах был введен термин «Интеллектуальный анализ данных», но интеллектуальный анализ данных — это эволюция сектора с обширной историей.

Истоки интеллектуального анализа данных восходят к трем семейным линиям: классическая статистика, искусственный интеллект и машинное обучение.

Классическая статистика:

Статистика является основой большинства технологий, на которых построен интеллектуальный анализ

данных, таких как регрессионный анализ, стандартное отклонение, стандартное распределение, стандартное отклонение, дискриминационный анализ, кластерный анализ и доверительные интервалы. Все они используются для анализа данных и подключения к данным.

Искусственный интеллект:

ИИ или искусственный интеллект основан на эвристике, а не на статистике. Он пытается применить человеческое мышление, например обработку данных, к статистическим задачам. Определенная концепция искусственного интеллекта была принята в некоторых высокопроизводительных коммерческих продуктах, таких как модули оптимизации запросов для системы управления реляционными базами данных (СУБД).

Машинное обучение:

Машинное обучение — это сочетание статистики и искусственного интеллекта. Его можно рассматривать как эволюцию ИИ, поскольку он сочетает в себе эвристику ИИ со сложным статистическим анализом. Машинное обучение пытается дать компьютерным программам возможность узнать об изучаемых данных, чтобы программы могли принимать четкое решение на основе характеристик исследуемых данных. Он использует статистику для основных концепций и добавляет дополнительные эвристики и алгоритмы ИИ для достижения своей цели.

Различия между интеллектуальным анализом данных и машинным обучением.

Интеллектуальный анализ данных и машинное обучение — это уникальные процессы, которые часто считаются синонимами. Однако, хотя оба они полезны для обнаружения закономерностей в больших наборах данных, они работают по-разному.

Интеллектуальный анализ данных — это процесс поиска закономерностей в данных. Прелесть интеллектуального анализа данных в том, что он помогает ответить на вопросы, которые мы не знали, задавать, путем активного выявления неинтуитивных шаблонов данных с помощью алгоритмов (например, потребители, покупающие арахисовое масло, с большей вероятностью покупают бумажные полотенца). Однако интерпретация этих идей и их применение к бизнес-решениям по-прежнему требуют участия человека.

Машинное обучение, между тем, — это процесс обучения компьютера обучению, как это делают люди. Благодаря машинному обучению компьютеры учатся определять вероятности и делать прогнозы на основе анализа данных. И хотя машинное обучение иногда использует интеллектуальный анализ данных как часть своего процесса, в конечном итоге оно не требует частого постоянного участия человека (например, беспилотный автомобиль использует интеллектуальный анализ данных, чтобы определить, где остановиться, ускориться и повернуть).[7]

Сущность интеллектуального анализа данных

Чтобы полностью ответить на вопрос «Что такое интеллектуальный анализ данных?» необходимы практические знания всего процесса. Интеллектуальный анализ данных основан на довольно структурированном шестиэтапном методе, известном как Межотраслевой стандартный процесс интеллектуального анализа данных (CRISP-DM).

Схема - Этапы интеллектуального анализа данных

Этот процесс предполагает поэтапную работу и повторение шагов при необходимости. Фактически, повторение шагов часто необходимо для учета изменения данных или введения различных переменных.[10]

Этапы интеллектуального анализа данных

Давайте подробнее рассмотрим каждый этап CRISP-DM:

Понимание бизнеса

Для начала задайте следующие вопросы: Какова наша цель? Какую проблему мы пытаемся решить? Какие данные нам нужны для ее решения?

Без четкого понимания того, какие данные нужно собирать, проект может привести к ошибкам, неточным результатам или результатам, которые не отвечают на правильные вопросы.

Понимание данных

После того как общая цель определена, необходимо собрать соответствующие данные. Данные должны соответствовать теме и обычно поступают из различных источников, таких как отчеты о продажах, опросы клиентов и данные геолокации. Цель этого этапа — убедиться, что данные правильно включают все необходимые наборы данных для достижения цели.

Подготовка данных

Самый трудоемкий этап — этап подготовки — состоит из трех этапов: извлечение, преобразование и загрузка — также называемый ETL. Сначала данные извлекаются из различных источников и помещаются в промежуточную зону. Далее, на этапе преобразования: данные очищаются, заполняются нулевыми наборами, удаляются повторяющиеся данные, устраняются ошибки и все данные распределяются по таблицам. На последнем этапе загрузки отформатированные данные загружаются в базу данных для использования.

#### Моделирование

Моделирование данных обращается к соответствующему набору данных и учитывает лучший статистический и математический подход к ответу на объективный вопрос(ы). Доступны различные методы моделирования, такие как классификация, кластеризация и регрессионный анализ (подробнее о них позже). Также нередко используются разные модели для одних и тех же данных для решения конкретных задач.

#### Оценка

После того как модели построены и протестированы, пришло время оценить их эффективность при ответе на вопрос, поставленный на этапе понимания бизнеса. Это этап, управляемый человеком, поскольку человек, выполняющий проект, должен определить, достаточно ли результаты модели соответствуют его целям. В противном случае можно создать другую модель или подготовить другие данные.

#### Развертывание

Как только модель интеллектуального анализа данных будет признана точной и успешной в ответе на объективный вопрос, пришло время применить ее. Развертывание может происходить в форме визуальной презентации или отчета, в котором представлены аналитические сведения. Это также может привести к таким действиям, как разработка новой стратегии продаж или реализация мер по снижению рисков.

Также можно представить в виде следующих этапов:

1. Изучение предметной области приложения.
2. Создание набора данных.
3. Очистка и обработка данных.
4. Сокращение и проецирование данных.
5. Выбор функции интеллектуального анализа данных.
6. Выбор алгоритма интеллектуального анализа данных.
7. Интеллектуальный анализ данных.
8. Интерпретация.
9. Использование полученных знаний.

Шаг 1. На первом этапе необходимо получить представление о предметной области приложения и соответствующих предварительных знаниях, после чего следует определить цель процесса с точки зрения клиента.

Шаг 2. Второй шаг включает в себя выбор набора данных с упором на подмножество переменных или выборок данных, на которых должно быть выполнено обнаружение.

Шаг 3. На третьем этапе выполняются основные операции по удалению шума или выбросов. Также рассматривается сбор необходимой информации для моделирования или учета шума, принятие решения о стратегиях обработки отсутствующих полей данных, учет типов данных, схемы и сопоставление отсутствующих и неизвестных значений.

Шаг 4: Здесь проводится работа по поиску полезных функций для представления данных, в зависимости от цели задачи, применение методов преобразования для поиска оптимального набора функций для данных.

Шаг 5. На пятом этапе определяются целевой результат (например, обобщение, классификация, регрессия, кластеризация).

Шаг 6. Шестой шаг касается выбора метода для поиска закономерностей в данных, принятия решения о том, какие модели и параметры подходят, и сопоставления конкретного метода интеллектуального анализа данных с общими критериями процесса.

Шаг 7: На седьмом этапе проводится работа по анализу данных, то есть поиск интересующих закономерностей в конкретной репрезентативной форме или наборе таких представлений: правила классификации или дерева, регрессия, кластеризация.

Шаг 8: На этом этапе избыточные и нерелевантные шаблоны отфильтровываются, соответствующие шаблоны интерпретируются и визуализируются таким образом, чтобы сделать результат понятным для пользователей.

Шаг 9: На последнем этапе результаты включаются в систему производительности, документируются и сообщаются заинтересованным сторонам и используются в качестве основы для принятия решений[12].

## Типы интеллектуального анализа данных

Интеллектуальный анализ данных наиболее полезен при выявлении закономерностей данных и извлечении полезной бизнес-информации из этих закономерностей. Для выполнения этих задач майнеры данных используют различные методы для получения разных результатов. Вот пять распространенных методов интеллектуального анализа данных.

### 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 (например, выбором номера деловыми путешественниками и привычками питания). Владелец отеля может использовать информацию о правилах ассоциации, чтобы предлагать повышение категории номера или рекламные акции на еду и напитки, чтобы привлечь дополнительных деловых путешественников.

### 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.

### Регрессивный анализ

Регрессионный анализ позволяет понять, какие факторы в наборе данных являются наиболее важными, а какие можно игнорировать, и как эти факторы взаимодействуют. С помощью этого метода майнеры данных могут проверять такие теории, как «когда прогнозируется много снега, перед ураганом будет продано больше хлеба и молока». Хотя это кажется достаточно очевидным, существует ряд переменных, которые необходимо проверить и количественно оценить, чтобы менеджер магазина мог убедиться в наличии достаточного количества товара.

Посредством регрессионного анализа можно рекомендовать конкретные уровни запасов молока и хлеба (в единицах/ящиках) для конкретных прогнозируемых уровней снега в определенные моменты времени (за несколько дней до урагана). Таким образом, использование регрессионного анализа максимизирует продажи, сводит к минимуму случаи отсутствия товара на складе и помогает избежать затоваривания запасов, которое приводит к порче продукта после урагана[5].

## Использование интеллектуального анализа данных

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.

Предприятия, которые не используют методы интеллектуального анализа данных, могут отставать от своих конкурентов. Это некоторые из основных способов использования интеллектуального анализа данных компаниями, чтобы избежать подобных недостатков.

### Анализ корзины

В самом базовом применении розничные торговцы используют анализ корзин для анализа того, что потребители покупают (или кладут в свои «корзины»). Это форма метода ассоциации, позволяющая ритейлерам понять покупательские привычки и рекомендовать другие покупки. Менее известное

приложение используется правоохранительными органами, где анализируются огромные объемы анонимных потребительских данных в поисках комбинаций продуктов, которые можно использовать при изготовлении бомб или производстве метамфетамина.

#### 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 отчеты о продажах, финансовые показатели (например, индекс потребительских цен, маркеры инфляции), 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 интеллектуальный анализ данных позволяет выявить закономерности и составить 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 «поднять» их до уровня расходов (например, потратьте 50 долларов, получите скидку 5 долларов; потратьте 75 долларов, получите 10 долларов). Это не только дает покупателю стимул совершать покупки, но также помогает удерживать доллары, на которые нацелены конкуренты.[11]

#### Профессии связанные с применением анализа данных.[14]

Администратор базы данных Администраторы играют жизненно важную роль в хранении, защите и потенциальном восстановлении данных компании; они гарантируют, что аналитики смогут получить доступ к нужным данным, когда они им понадобятся. Администрирование баз данных — это расширяющаяся область с высоким потенциалом заработной платы.

Ученый по компьютерам и информатике Ученые в области компьютерных технологий и информатики разрабатывают новые технологии (компьютерные языки, операционные системы, программное обеспечение и т. д.) в быстро расширяющемся пространстве и всегда находятся в поиске новых идей. Они работают в таких областях, как финансы, технологии, здравоохранение и научные исследования.

#### Аналитик по исследованию рынка

Аналитики-исследователи проводят маркетинговые исследования, чтобы помочь компаниям привлечь новых клиентов, увеличить продажи и определить потенциал продаж новых продуктов. Рост электронной коммерции способствует развитию этой области;

Архитектор компьютерных сетей Сетевые архитекторы проектируют, создают и обслуживают сеть передачи данных компании, которая может варьироваться от нескольких компьютеров до большого облачного центра обработки данных. Здравоохранение способствует расширению возможностей трудоустройства в этой профессии.

Аналитик информационной безопасности Эксперты по цифровой безопасности стали незаменимы практически в любой организации, которой необходимо защитить конфиденциальные данные и предотвратить кибератаки. Фактически, учитывая прогнозируемый рост занятости, в будущем, вероятно, появится еще больше рабочих мест в этой области. Месторождение также достаточно доступно для тех, кто приезжает из других отраслевых концентраций. Например, администраторы баз данных могут быть

сильными кандидатами на роли в сфере безопасности баз данных.

#### Заключение

Таким образом, интеллектуальный анализ данных предлагает решение этой проблемы, которое помогает компаниям принимать решения, снижает затраты и увеличивает доходы. В результате различные исследователи данных используют майнинг как часть своей повседневной работы.

Интеллектуальный анализ данных часто рассматривается как сложная задача. Но изучить эту важную дисциплину науки о данных не так сложно, как кажется.

Интеллектуальный анализ данных далее определяется как процесс использования компьютеров и автоматизации для обнаружения больших наборов данных о закономерностях и тенденциях и превращения этих результатов в бизнес-понимание и прогнозы. Интеллектуальный анализ данных выходит за рамки процесса поиска, поскольку он использует данные для оценки будущих вероятностей и разработки действенного анализа.

#### Список литературы

1. Макшанов А.В. Технологии интеллектуального анализа данных: Учебное пособие / А.В. Макшанов, А.Е. Журавлев. — СПб.: Лань, 2018. — 212 с.
2. Макшанов А.В. Технологии интеллектуального анализа данных. — М.: Лань. 2019. 212 с.
3. Миркин Б. Г. Введение в анализ данных. — М.: Юрайт. 2020. 175 с.
4. Мхитарян В. С. Теория планирования эксперимента и анализ статистических данных. — М.: Юрайт. 2020. 491 с.
5. Нгай, Эрик В.Т., Ли Сю и Доротхай СК Чау.«Применение методов интеллектуального анализа данных в управление взаимоотношениями: обзор литературы и Классификация.» Экспертные системы с приложениями 36.2 (2009)
6. Форман, Д. Много цифр: Анализ больших данных при помощи Excel / Д. Форман. - М.: Альпина Пабlishер, 2017. - 461 с.
7. Форман, Дж. Много цифр: Анализ больших данных при помощи Excel / Дж. Форман. - М.: Альпина Пабlishер, 2019. - 461 с.
8. J. Ross Quinlan. C4.5: Programs for Machine learning. Morgan Kaufmann Publishers
9. S.Murthy. Automatic construction of decision trees from data: A Multi-disciplinary survey
10. W. Buntine. A theory of classification rules.
11. Machine Learning, Neural and Statistical Classification. Editors D. Mitchie et.al.
12. Ling, Charles X., and Chenghui Li. "Data Mining for Direct Marketing: Problems and Solutions." KDD.
13. Džeroski, Sašo. Relational data mining. Springer US, 2009.
14. Gorunescu, Florin. Data Mining: Concepts, models and techniques. Vol. 12. Springer Science & Business Media, 2011.

*Эта часть работы выложена в ознакомительных целях. Если вы хотите получить работу полностью, то приобретите ее воспользовавшись формой заказа на странице с готовой работой:*

<https://stuservis.ru/referat/377781>