

**ИНФОРМАТИКА. КРАТКИЕ СООБЩЕНИЯ**

doi: 10.51639/2713-0576\_2023\_3\_2\_203

УДК 0.00.005.22

ГРНТИ 28.29.01

ВАК 20.00.00

**Анализ данных**

Юрьева Д. Ю.

*Институт Компьютерных Технологий и Информационной Безопасности ЮФУ  
347922, Россия г. Таганрог, ул. Чехова 2*

email: [dyureva@sfedu.ru](mailto:dyureva@sfedu.ru)

В век цифровых технологий спрос на обработку и анализ данных стремительно повышается, с каждым днем объем данных возрастает в геометрической прогрессии. Анализ данных играет важную роль в нашей жизни. Он является незаменимой частью повседневности. Люди постоянно используют его, но не задумываются и не осознают это. Каждый день эта сфера увеличивает свои размеры и значение. Для того чтобы заниматься анализом данных у человека должны быть знания основ. Проблематика данной темы в том, что анализ данных всегда рассматривается по большей части только для одной сферы, но общее понимание тематики, использования и применения не представляется. Опираясь на учебные материалы и открытые источники, было принято решение собрать все воедино и представить это в одной статье. В статье представлены общие понятия, виды, типы и использование анализа данных. Статья ориентирована на учеников и студентов, интересующихся современными методами анализа данных.

*Ключевые слова:* анализ данных, этапы анализа данных, типы методов анализа данных, виды анализа данных.

Анализ данных является важным аспектом принятия обоснованных решений в различных отраслях. С развитием технологий это стало динамичной и захватывающей областью. Но что это, если говорить простыми словами?

**Что такое анализ данных?**

Анализ данных — это процесс сбора, моделирования и обработки данных с использованием различных статистических и логических методов и приемов. Методы анализа развивались вместе с увеличением объема данных. Все начиналось с обработки небольших баз данных, но сейчас эта область многогранна она строится на таких науках как прикладная статистика, распознавание образов, искусственный интеллект, теория баз данных и другое.

**Зачем нужен анализ данных?**

Анализ данных — это широкий термин, который охватывает множество сторон. Любой тип информации может быть подвергнут методам анализа данных. Хотя бы один раз в месяц каждый человек проводит анализ данных. Это может быть, как и большой объем информации, так и не очень. Дети анализируют все вокруг себя, и тем самым получают новую информацию. Ученики и студенты обрабатывают большое количество данных интернета и других источников. Взрослые люди же анализируют свою работу и не только. Внедрение анализа в бизнес-модель, помогает компаниям оптимизировать свою работу. К примеру, с помощью анализа тенденций и степени удовлетворенности клиентов, компания может понять какой продукт востребован, а какой нужно улучшить.

### Виды анализа данных

Данные можно использовать для ответов на вопросы и принятия решений разными способами. Чтобы определить лучший способ анализа, нужно познакомиться с каждым из них.

#### 1. Исследовательский анализ (Exploratory analysis)

После того как данные изучены, исследовательский анализ помогает найти связи и сгенерировать гипотезы и решения для конкретных проблем. Исследовательский анализ отвечает на вопрос «Как исследовать связь данных?»

#### 2. Статистический анализ (Statistical Analysis)

Статистический анализ использует прошлые данные в виде информационных панелей. Статистический анализ включает сбор, анализ, интерпретацию, представление и моделирование данных. Он анализирует набор данных или выборку данных. Существует две категории этого типа анализа — описательный анализ и логический анализ. Статистический анализ отвечает на вопрос «Что случилось?»

#### 3. Описательный анализ (Descriptive analysis)

Описательный анализ используется для обобщения и описания основных характеристик набора данных. С помощью него можно упорядочить, обработать и интерпретировать данные из различных источников, чтобы превратить их в целостную информацию. Описательный анализ обеспечивает всесторонний обзор данных и понимание их свойств и структуры. Описательный анализ отвечает на вопрос «Что это?»

#### 4. Логический анализ (Inferential Analysis)

Он анализирует выборку из базы данных. В этом типе анализа можно сделать разные выводы из одних и тех же данных, выбрав разные параметры. Логический анализ отвечает на вопрос «А что если?»

#### 5. Диагностический анализ (Diagnostic analysis)

Этот анализ полезен для выявления моделей поведения данных. Аналитика диагностических данных расширяет возможности аналитиков и руководителей, помогая им получить четкое контекстуальное понимание того, почему что-то произошло. Если знать, почему что-то произошло, а также как это произошло, можно точно определить способы решения проблемы или причины. Диагностический анализ отвечает на вопрос «Почему это произошло?»

#### 6. Прогнозный анализ (Predictive analysis)

Прогнозная аналитика использует доступные данные для формирования прогнозов на будущее. Данный анализ может опираться на текущие или прошлые данные. Прогнозирование — это всего лишь оценка. Его точность зависит от многих факторов. Предиктивный анализ отвечает на вопрос: «Что может произойти в будущем?»

#### 7. Предписывающий анализ (Prescriptive analysis)

Предписывающий анализ — это анализ принятия решений, который использует математическое моделирование, алгоритмы оптимизации и другие методы. На основе текущих ситуаций или проблем он анализирует данные и принимает наилучшие решения. Предписывающий анализ отвечает на вопрос: «Что нам с этим делать?»

### Типы методов анализа данных

Методы анализа данных можно разделить на две большие категории: количественные и качественные. Каждая из этих категорий имеет мощную аналитическую ценность, которая меняется в зависимости от сценария и типа данных, с которыми предстоит работать.

#### 1. Количественный метод

Количественный анализ относится ко всем методам, которые используют числовые данные или данные, которые могут быть преобразованы в числа (например, переменные категории, такие как пол, возраст и т. д.), чтобы извлечь ценную информацию.

А) Кластерный анализ: проще говоря, распределение по группам. Разделение данных по схожим параметрам.

Б) Винтажный анализ: этот тип подхода к анализу данных использует исторические данные для изучения и сравнения определенного сегмента поведения пользователей, который затем может быть сгруппирован с другими с аналогичными характеристиками.

В) Регрессионный анализ: регрессия использует исторические данные, чтобы понять, как влияет на значение зависимой переменной, когда одна (линейная регрессия) или несколько независимых переменных (множественная регрессия) изменяются или остаются неизменными. Понимая взаимосвязь каждой переменной и то, как она развивалась в прошлом, можно предвидеть возможные результаты и принимать более эффективные решения в будущем.

Г) Нейронные сети: нейронная сеть составляет основу интеллектуальных алгоритмов машинного обучения. Это форма аналитики, которая пытается с минимальным вмешательством понять, как человеческий мозг будет генерировать идеи и прогнозировать значения. Нейронные сети учатся на каждой транзакции данных, а это означает, что они развиваются и совершенствуются с течением времени.

Д) Интеллектуальный анализ данных: метод анализа данных, который является общим термином для инженерных показателей и идей для получения дополнительной ценности, направления и контекста. Используя исследовательскую статистическую оценку, интеллектуальный анализ данных направлен на выявление зависимостей, отношений, закономерностей и тенденций для получения передовых знаний. При рассмотрении вопроса о том, как анализировать данные, принятие мышления интеллектуального анализа данных имеет важное значение для успеха — как таковая, это область, которую стоит изучить более подробно.

Е) Анализ временных рядов: анализ временных рядов используется для анализа набора точек данных, собранных за определенный период времени. Хотя аналитики используют этот метод для наблюдения за точками данных в определенный интервал времени, а не просто для их периодического наблюдения, анализ временных рядов используется не только для сбора данных во времени. Вместо этого он позволяет исследователям понять, изменились ли переменные в ходе исследования, как они зависят друг от друга и как это привело к конечному результату.

Ж) Дерево решений: анализ дерева решений призван служить вспомогательным инструментом для принятия разумных и стратегических решений. Визуально отображая потенциальные результаты, последствия и затраты в виде древовидной модели, исследователи и пользователи компании могут легко оценить все задействованные факторы и выбрать наилучший курс действий.

З) Конъюнктивный анализ: этот подход обычно используется в опросах, чтобы понять, как люди оценивают различные атрибуты продукта или услуги, и это один из наиболее эффективных методов для извлечения потребительских предпочтений.

И) Анализ соответствий: этот метод начинается с вычисления «ожидаемого значения», которое выполняется путем умножения средних значений строк и столбцов и деления его на общее исходное значение конкретной ячейки таблицы. Затем «ожидаемое значение» вычитается из исходного значения, в результате чего получается «остаточное число», которое позволяет делать выводы о взаимосвязях и распределении.

К) Многомерное шкалирование (MDS): MDS — это метод, используемый для наблюдения сходства или различия между объектами, такими как цвета, бренды, люди, географические координаты и многое другое. Объекты наносятся с использованием «карты MDS», которая размещает похожие объекты вместе, а несопоставимые — далеко друг от друга.

## 2. Качественный метод

Методы качественного анализа данных определяются как наблюдение за нечисловыми данными, которые собираются и производятся с использованием таких методов наблюдения, как интервью, фокус-группы, анкеты и т. д. В отличие от количественных методов,

качественные данные более субъективны и очень ценны для анализа удержания клиентов и разработки продуктов.

А) Анализ текста: анализ текста, также известный как интеллектуальный анализ текста, работает, беря большие наборы текстовых данных и упорядочивая их таким образом, чтобы ими было легче управлять. Тщательно проработав этот процесс очистки, можно извлечь данные, которые действительно имеют отношение к той или иной организации, и использовать их для разработки практических идей, которые будут продвигать компанию вперед.

Б) Контент анализ: это простой и очень популярный метод, который проверяет наличие и частотность определенных слов, понятий и предметов в различных форматах контента, таких как текст, изображение, аудио или видео.

В) Тематический анализ: тематический метод анализирует большие фрагменты текстовых данных, таких как стенограммы фокус-групп или интервью, и группирует их по темам или категориям, которые часто встречаются в тексте. Это отличный метод, когда можно выяснить взгляды и мнения людей по определенной теме.

Г) Повествовательный анализ: повествовательный анализ используется для изучения смысла историй, которые рассказывают люди, и, что наиболее важно, того, как они их рассказывают. Изучая слова, которые люди используют для описания ситуации, можно сделать ценные выводы об их взглядах на конкретную тему. Общие источники повествовательных данных включают в себя: автобиографии, семейные истории, мнения, отзывы и т. д.

Д) Дискурсивный анализ: анализ дискурса используется для понимания значения любого типа письменного, устного или символического дискурса на основе его политического, социального или культурного контекста. Он смешивает анализ языков и ситуаций вместе. Это означает, что на то, как строится содержание, и на его значение в значительной степени влияют культура и общество, в котором оно происходит.

Е) Анализ обоснования теорий: обоснованная теория – единственный метод, который не требует начального исследовательского вопроса или гипотезы, поскольку его ценность заключается в создании новых теорий.

#### **Этапы анализа данных**

Процесс, связанный с анализом данных, включает несколько различных этапов:

##### **1. Определение сферы исследования**

Начинать свою работу нужно с выбора правильных вопросов. Определение вопроса исследования определяет задачу исследования и обеспечивает четкую цель. Вопросы должны быть логичными, четкими и понятными.

##### **2. Сбор данных**

После того как цель установлена, приходит время сбора данных.

- Поиск данных лучше производить на специализированных сайтах и подтвержденных источниках.

- Если работа производится в команде, самым правильным решением будет, распределение задач. Сокращая количество задач и фокусируясь на одной, увеличивается шанс сбора более качественного материала.

- Если в задаче поставлен вопрос сбора информации с помощью опроса, наблюдения и т. д., разработка анкеты заранее с четкими вопросами, поможет обеспечить последовательность, а также сэкономить время.

- Поскольку собранный материал чаще всего представляет собой большую базу данных, нужно не забывать вести журнал учета с датой сбора и примечаниями к источнику по мере продвижения.

### 3. Фильтрация

Не все данные, которые собираются, будут полезны. Необработанный материал может содержать дубликаты, ошибки или не актуальную информацию. Поэтому после сбора информации идет её более глубокий статистический анализ.

### 4. Анализ

Программное обеспечение для анализа данных помогает понимать и интерпретировать качественные данные для достижения целей. С помощью различных методов, таких как статистический анализ, регрессия, нейронные сети, анализ текста и т. д., можно начать анализировать данные и манипулировать ими, чтобы делать соответствующие выводы. Использование ПО самый важный шаг в процессе анализа данных, когда происходит извлечение ценности из собранной базы. Все модели анализа, которые можно использовать, будут зависеть от задачи, которая была поставлена изначально.

### 5. Визуализация

Инструменты визуализации данных помогают представить результат анализа данных в привлекательном, ясном и легком для понимания виде. Использование различных источников данных для создания интерактивных диаграмм, отчетов и информационных панелей, приводит к закономерностям и выводам, которые могут быть неочевидны в электронной таблице.

### 6. Интерпретация

Сравнивая результаты проведенной работы с целями, которые были определены в начале процесса, можно получить конечный результат или приближённый к истине ответ.

В статье были рассмотрены только основные составляющие анализа данных. Данная сфера многогранна, необъятна и очень интересна. Анализ данных даёт возможность посмотреть на многие вещи под неизученным углом и узнать что-то новое.

## **Конфликт интересов**

Автор статьи заявляет, что у неё нет конфликта интересов по материалам данной статьи с третьими лицами на момент подачи статьи в редакцию журнала, и ей ничего не известно о возможных конфликтах интересов в настоящем со стороны третьих лиц.

## **Список литературы**

1. Татарникова Т. М. Анализ данных. Учебное пособие. // СПб, СПбГЭУ, 2018. – 82 с.
2. Your Modern Business Guide To Data Analysis Methods And Techniques // datapine: сайт. – URL: <https://www.datapine.com/blog/data-analysis-methods-and-techniques/> (датаобращения: 04.04.2023)
3. Мхитарян В. С. Анализ данных // М, Юрайт, 2017. – 490 с.
4. Рыбальченко Н. Д. Что такое анализ данных? Виды и методы. Применение анализа данных в маркетинге // развитие современной науки и технологий в условиях трансформационных процессов Сборник материалов IV Международной научно-практической конференции. - Москва: ООО "Издательство АЛЕФ", 2022. С. 9.

## **Data analysis**

Yurieva D. Yu.

*Institute of Computer Technologies And Information Security SFEDU  
347922, Russian Federation, Taganrog city, Chekhova str., 2*

email: [dyureva@sfedu.ru](mailto:dyureva@sfedu.ru)

In the age of digital technology, the demand for data processing and analysis is increasing rapidly, with the volume of data increasing exponentially every day. Data analysis plays an important role in our lives. It is an indispensable part of everyday life. People use it all the time, but they don't think about it and don't realize it. Every day this field increases in size and importance. To engage in data analysis, one must have knowledge of the basics. The problematic of this topic is that data analysis is always considered for the most part only for one area, but a general understanding of the topics, uses, and applications is not presented. Relying on tutorials and open sources, it was decided to pull it all together and present it in one article. The article presents general concepts, types, and the use of data analysis. The article is aimed at students and students interested in modern methods of data analysis.

*Keywords:* data analysis, stages of data analysis, types of data analysis methods, types of data analysis