

The background features a dark blue grid with a glowing, multi-colored arrow pointing towards the top right. A wavy, multi-colored ribbon-like shape flows across the lower half of the image. On the right side, there is a vertical axis with tick marks and some faint, illegible text.

# Интеллектуальный анализ данных

Какие группы (например по демографическому признаку, или по уровню доходов, или по социальному статусу) более чувствительны к изменению цен на наши товары и услуги?

На какие товары надо предложить скидку, чтобы это привело к росту доходов за счет продаж товаров-дополнителей (комплементов)?

Как построить CRM-стратегию, чтобы "отсеивать" "непродуктивных" клиентов и удерживать "кормильцев"?

Как наиболее рационально с точки зрения финансового результата расставить товары на полках супермаркета?

На большинство этих вопросов  
руководители и специалисты  
чаще всего отвечают опираясь на свою  
интуицию и опыт

Но если у нас есть данные, описывающие покупателей и их поведение на продолжительном отрезке времени, почему бы их не проанализировать, чтобы подтвердить или опровергнуть свои интуитивные догадки с помощью математического аппарата и логики?

В каких сочетаниях покупают товары клиенты супермаркета; Какие группы покупателей можно выделить исходя из структуры их корзин и чеков?

Как ведут себя различные категории клиентов банков после того, как им предложили увеличить или отказали в увеличении кредитного лимита?

Ответы на эти вопросы можно извлечь из хранящихся в базах данных записей.

# Сферы применения

**Торговля.** Анализ потребительской корзины, исследование временных шаблонов, создание прогнозных моделей, оптимизация складских запасов.

**Банковское дело.** Сегментация клиентов, выявление мошенничества с кредитными картами, прогнозирование изменения клиентуры, анализ финансовых рисков.

**Страховой бизнес.** Сегментация клиентов, выявление фактов мошенничества, анализ страховых рисков, разработка новых продуктов, расчет страховых премий.

**Телекоммуникации.** Анализ лояльности клиентов, сегментирование клиентской базы и услуг, анализ внешних факторов на отказы оборудования, выявление случаев несанкционированного доступа к сети.

**Производственные предприятия.** Оптимизация закупок, диагностика брака на ранних стадиях, диагностика оборудования, маркетинг.

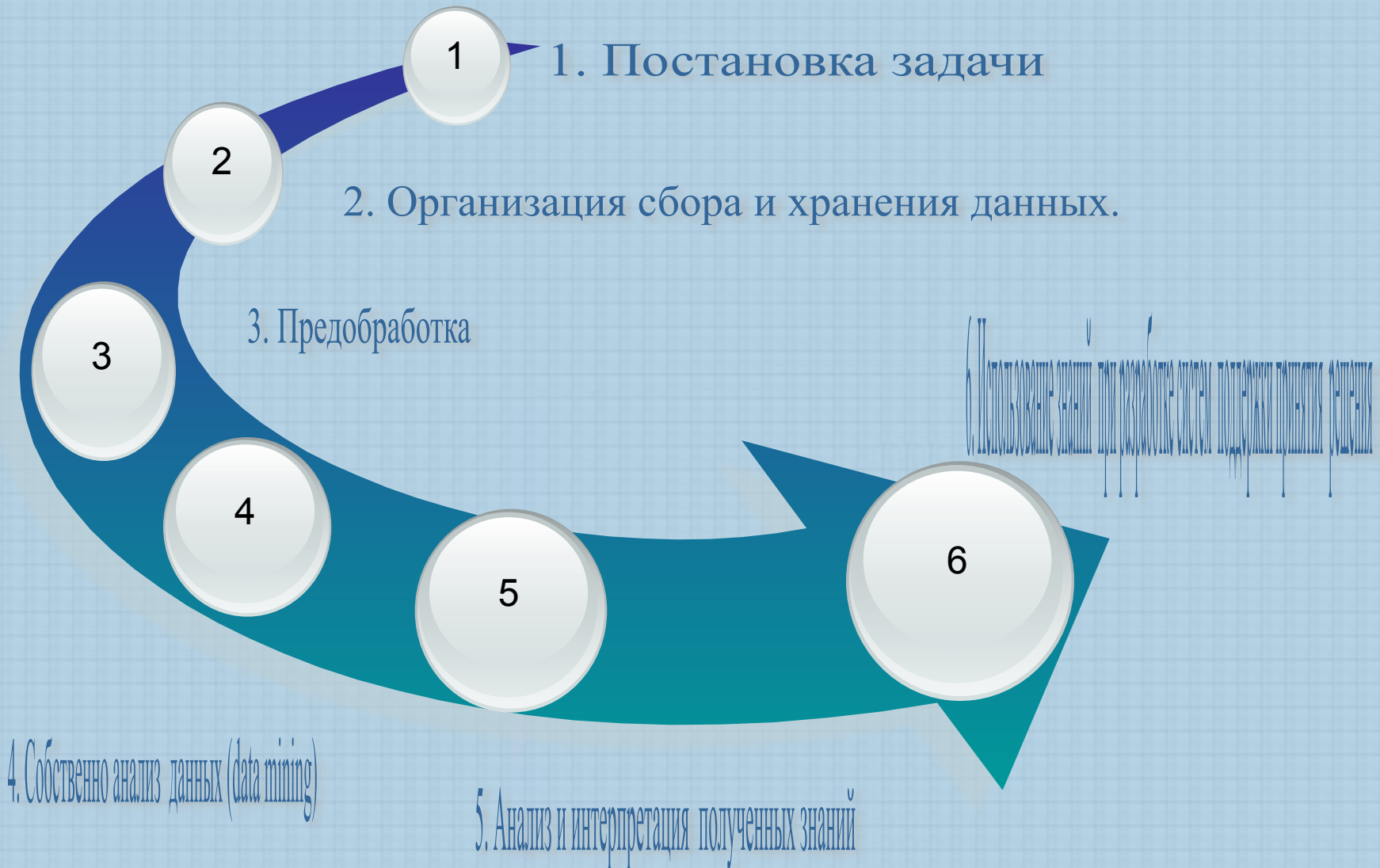
**Нефтегазовая отрасль.** Диагностика оборудования и нефте/газопроводов, прогнозирование цен, разведка месторождений, анализ влияния внешних и внутренних факторов на объемы продаж.

# Определение

Data mining (дословно, “разработка данных”, “добыча данных”) – исследование и обнаружение “машиной” (алгоритмами, средствами искусственного интеллекта) в сырых данных скрытых структур или зависимостей, которые

- ранее не были известны,
- нетривиальны,
- практически полезны,
- доступны для интерпретации человеком.

# Основные стадии добычи данных



# Методы добычи данных

- кластеризация
- ассоциация
- деревья решений
- анализ с избирательным действием
- сети уверенности
- метод ближайших соседей
- нейронные сети
- нечеткая логика
- генетические алгоритмы и эволюционное программирование
- регрессионные методы



Спасибо за внимание!