

# Аффинитивный анализ

Автор: Шакиров Тимур, Новицкий  
Ярослав, Абалуев Антон

# Введение

- Аффинитивный анализ — метод для выявления и оценки взаимосвязей между событиями, происходящими совместно.
- Цель: исследование взаимной связи между событиями и количественная оценка ассоциаций.
- Результат: набор ассоциативных правил.

# Применение аффинитивного анализа

- Выявление частых покупок товаров в супермаркете
- Определение доли клиентов, реагирующих на нововведения
- Профилирование посетителей веб-сайта
- Оценка побочных эффектов нового лекарства
- Связь параметров оборудования с характеристиками продукции

# Основные понятия

- Транзакция: Набор событий, происходящих совместно
- Ассоциативное правило: Например,  $A \rightarrow B$  ('Если произошло A, следует ожидать B')
- Характеристики правила:
- Поддержка (Support): Частота появления как условия, так и следствия
- Достоверность (Confidence): Вероятность, что из A следует B

# Поддержка и достоверность

- Поддержка:  $S(A \rightarrow B) = (\text{число транзакций с } A \text{ и } B) / (\text{общее число транзакций})$
- Достоверность:  $C(A \rightarrow B) = (\text{число транзакций с } A \text{ и } B) / (\text{число транзакций с } A)$

# Границы поддержки и достоверности

- Оптимальные значения:
- Высокая поддержка = известные закономерности
- Слишком низкая поддержка = слабая значимость
- Сильные правила: Правила, чья поддержка и достоверность выше установленного порога

# Субъективные меры значимости

- Лифт (Lift): Показывает силу связи ( $L(A \rightarrow B) > 1$  — положительная связь)
- Левередж (Leverage): Разность между наблюдаемой и ожидаемой частотой
- Улучшение (Improvement):  $I(A \rightarrow B) > 1$  — лучше случайного угадывания

# Алгоритмы поиска ассоциативных правил

- AIS: Первый алгоритм, разработанный IBM в 1993 году
- SETM: Использует SQL для генерации кандидатов
- Apriori: Основной алгоритм, использующий этапы формирования и подсчета кандидатов



# Алгоритм Apriori – Основные этапы

- Формирование кандидатов: Генерация множеств для оценки
- Подсчет поддержки: Отсечение кандидатов с низкой поддержкой
- Итог: Часто встречающиеся наборы элементов

# Последовательные шаблоны

- Определение: Учет временного аспекта и последовательности
- Метод: На базе алгоритма Apriori и его модификаций
- Пример применения: Анализ типичных последовательностей в данных

# Примеры алгоритмов для последовательных шаблонов

- AprioriAll: Генерирует и подсчитывает последовательности-кандидаты
- Процесс: Использует частые последовательности предыдущего прохода для новых кандидатов
- Преимущества: Снижение числа кандидатов через параметризацию длины

# Заключение

- Выводы:
- Аффинитивный анализ выявляет полезные взаимосвязи в данных
- Алгоритмы оптимизируют процесс и уменьшают количество правил для анализа
- Применимость: Ритейл, медицина, маркетинг, веб-аналитика