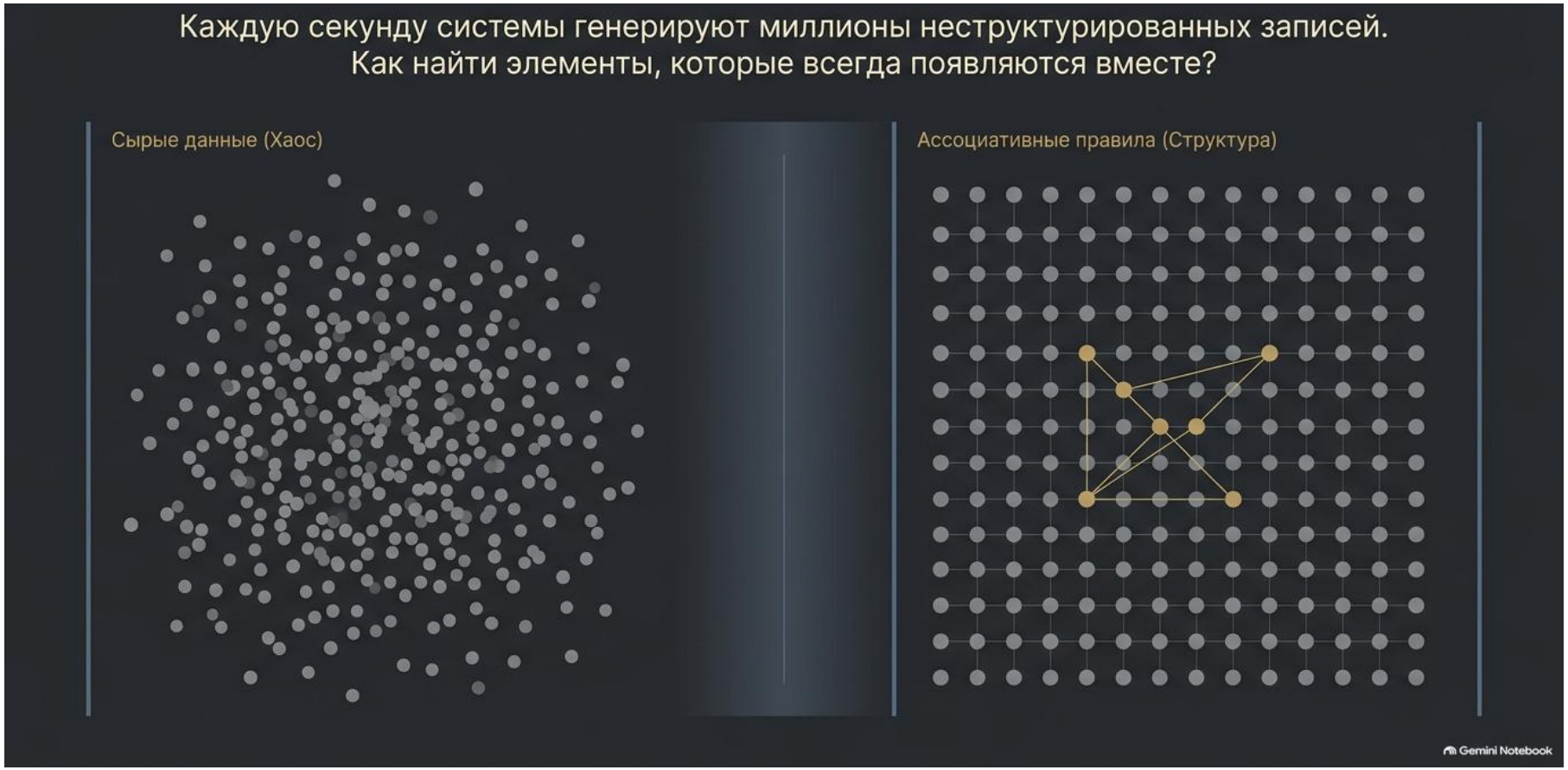
# ЛЕКЦИЯ 5. ОТ ХАОСА К ЗАКОНОМЕРНОСТЯМ: АНАЛИЗ АССОЦИАТИВНЫХ ПРАВИЛ (ASSOCIATION RULE MINING)

ДИСЦИПЛИНА: ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

# МОРЕ ТРАНЗАКЦИЙ: ПРОБЛЕМА СКРЫТЫХ СВЯЗЕЙ



# АНАТОМИЯ АССОЦИАТИВНОГО ПРАВИЛА



# МЕТРИКА 1: ПОДДЕРЖКА (SUPPORT)

## Метрика 1: Поддержка (Support)

Как часто данный набор элементов встречается во всем массиве данных?

Формула: $\text{Support}(X \Rightarrow Y) = \frac{\text{freq}(X, Y)}{N}$

**Цель:** Отфильтровать информационный шум. Мы исключаем из анализа редкие комбинации, не имеющие статистической значимости.

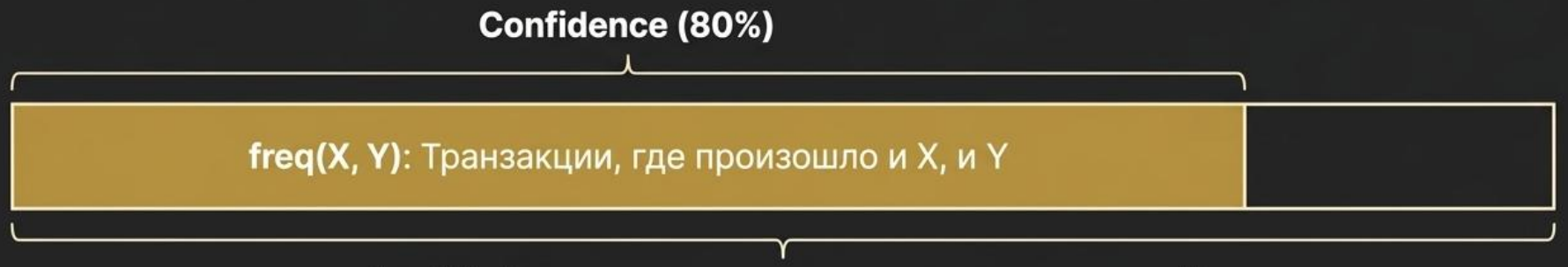



Gemini Notebook

# МЕТРИКА 2: ДОСТОВЕРНОСТЬ (CONFIDENCE)

Насколько сильна эта связь? Какова вероятность появления Y, если X уже произошло?

Формула: $Confidence(X \Rightarrow Y) = \frac{freq(X,Y)}{freq(X)}$




Смысл: Это условная вероятность P(Y|X). Показывает надежность правила, но не учитывает базовую частоту самого следствия Y.

Gemini Notebook

# МЕТРИКА 3: ЛИФТ (LIFT)

## Метрика 3: Лифт (Lift)

Истинная связь или случайность? Учитывает базовую популярность Y.

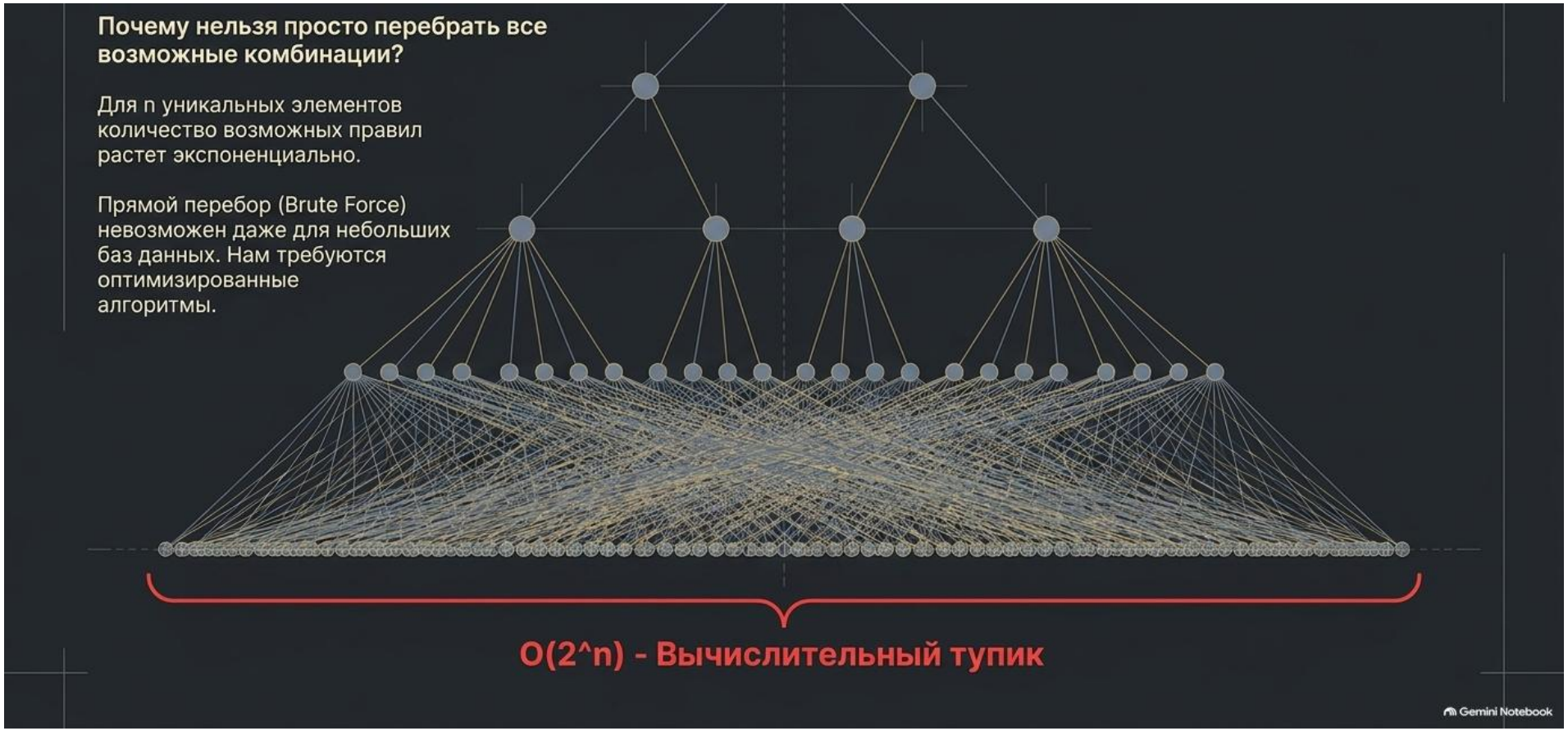
Формула:

$$Lift(X \Rightarrow Y) = \frac{Support(X \cup Y)}{Support(X) \times Support(Y)}$$




Gemini Notebook

# КОМБИНАТОРНЫЙ ВЗРЫВ



# АЛГОРИТМ APRIORI: СВОЙСТВО АНТИМОНОТОННОСТИ

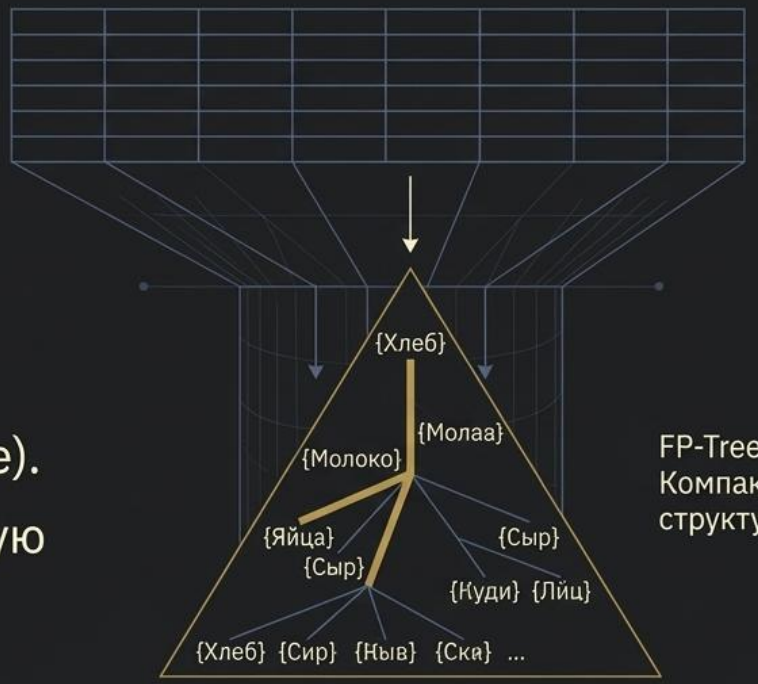


# АЛГОРИТМ FP-GROWTH: СЖАТИЕ ЧЕРЕЗ ПРЕФИКСНЫЕ ДЕРЕВЬЯ

Избавление от этапа генерации кандидатов за счет создания компактной структуры в памяти.

Механика:

1. Сканирует данные 2 раза.
2. Сжимает транзакции в префиксное дерево (FP-Tree).
3. Извлекает правила напрямую из ветвей дерева.




Gemini Notebook

# АРХИТЕКТУРА ВЫБОРА: APRIORI ПРОТИВ FP-GROWTH

Какой алгоритм использовать в Data Mining пайплайне?

| | Apriori | FP-Growth |
|---|---|---|
| Генерация кандидатов | Да (Высокая нагрузка) ✓ | Нет (Прямое извлечение) ✓ |
| Сканирование БД | Многократное (n раз) | Всего 2 раза ✓ |
| Использование памяти | Среднее ✓ | Высокое (для FP-Tree) |
| Лучшее применение | Небольшие / разреженные датасеты | Огромные / плотные датасеты (Big Data) |

Gemini Notebook

# ПАРАДИГМЫ ML: МЕСТО АССОЦИАТИВНЫХ ПРАВИЛ

Интеграция знаний: Как методы интеллектуального анализа дополняют друг друга.

## Классификация (Classification)

- Обучение с учителем (Supervised).
- Цель: Предсказание меток.
- Алгоритмы: Decision Trees, SVM.



## Кластеризация (Clustering)

- Обучение без учителя (Unsupervised).
- Цель: Группировка схожих строк.
- Алгоритмы: K-means, DBSCAN.



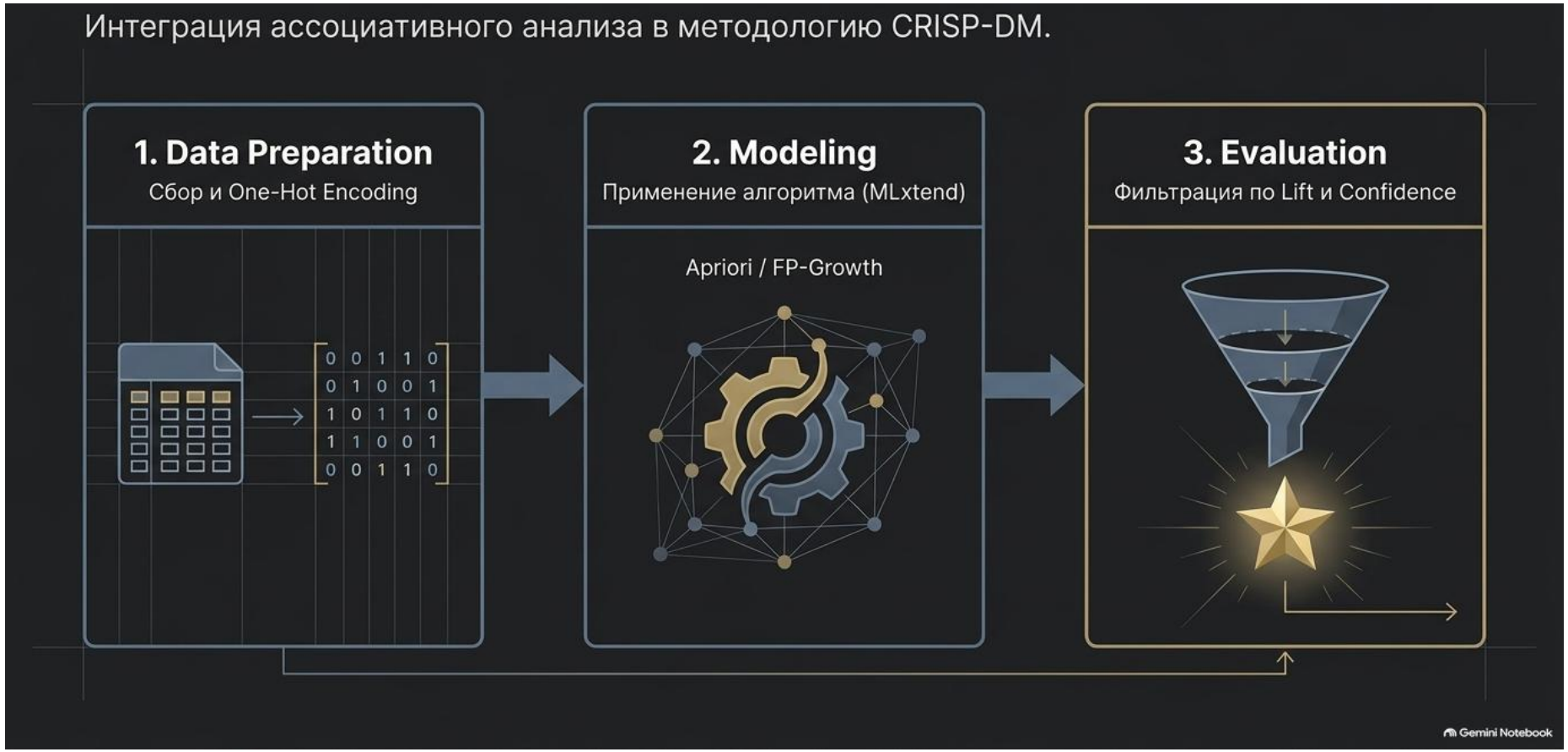
## Ассоциация (Association)

- Обучение без учителя (Unsupervised).
- Цель: Поиск связей между признаками.
- Алгоритмы: Apriori, FP-Growth.



Gemini Notebook

# ПАЙПЛАЙН DATA MINING: ОТ СЫРЫХ ДАННЫХ К ИНСАЙТАМ



# МАГИЯ PYTHON: ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ

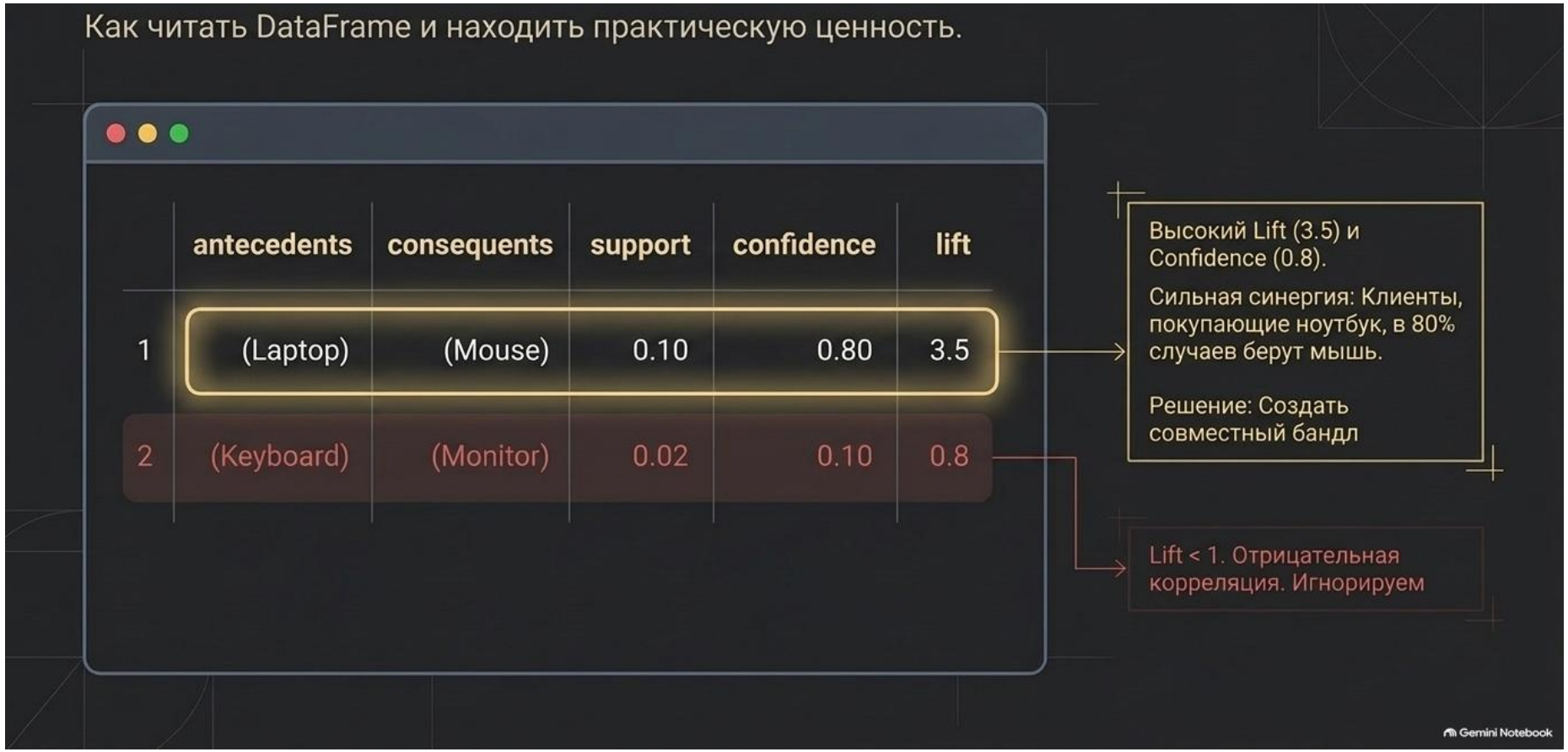
Всего несколько строк кода с использованием библиотек Pandas и MLxtend.

```python
from mlxtend.frequent_patterns import apriori, association_rules

# 1. Поиск частых наборов (Support > 0.05)
frequent_itemsets = apriori(df_encoded, min_support=0.05, use_colnames=True)

# 2. Генерация правил
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1.2)

# 3. Сортировка по силе связи
rules_sorted = rules.sort_values(by=['lift', 'confidence'], ascending=False)
```

Математические пороги отсечения шума

Gemini Notebook

# ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ



# АРХИТЕКТУРА ИНСАЙТОВ: ИТОГИ

Интеллектуальный анализ ассоциативных правил превращает хаос в стратегию.

1 

**Скрытая топология данных:** Алгоритмы извлекают неочевидные связи (X ⇒ Y), выходя за рамки простой статистики.

2 

**Математическая строгость:** Метрики Support, Confidence и Lift отсеивают случайный шум от истинных закономерностей.

3 

**Практическая реализация:** Свойство антимонотонности (Apriori) и префиксные деревья (FP-Growth) позволяют обрабатывать Big Data.

Gemini Notebook