

Business Understanding

С чего начинается работа с данными?

Пётр Емельянов

R&D Директор в UBIC Technologies

Skillbox

Business Understanding. С чего начинается
работа с данными?

Skillbox

Что такое Data Mining?

Что такое Data Mining?

CRISP-DM

CRoss

Industry

Standard

Process

for

Data

Mining

Кросс-

индустриальный

стандартный

процесс

для

?

Что такое Data Mining?

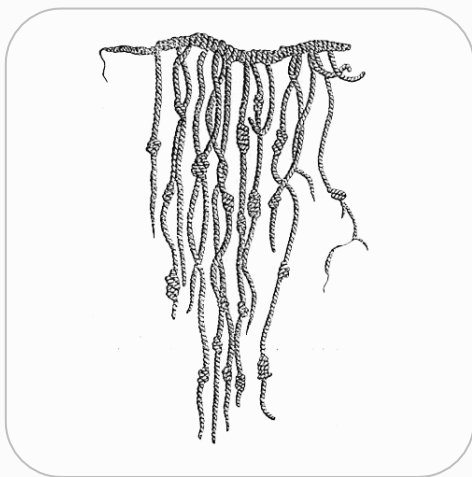
Цель урока

Разобраться в Data Mining — технике анализа данных, которая предназначена для поиска ранее неизвестных закономерностей в больших информационных массивах.

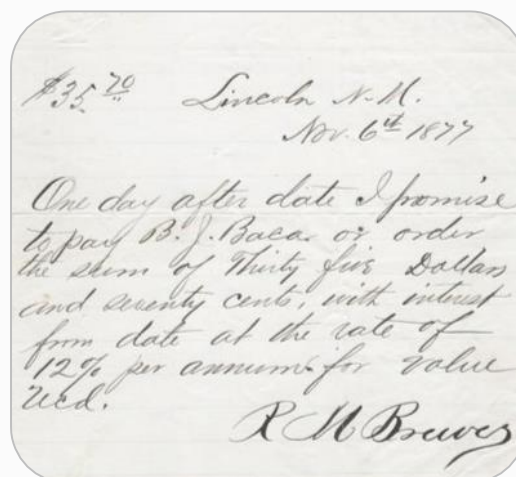
Что такое Data Mining?

Как банки хранили и хранят данные?

Узелковое письмо



Долговые расписки



Базы данных



Проблемы банка

- ✓ Банк теряет прибыль. Почему — непонятно
- ✓ Банк существует давно, за это время накопилось очень много данных
- ✓ В этих данных вся жизнь банка — история прошлых успехов и причины сегодняшних неудач
- ✓ Задача — найти причины неудач. Нужно знать, как искать

Что такое Data Mining?

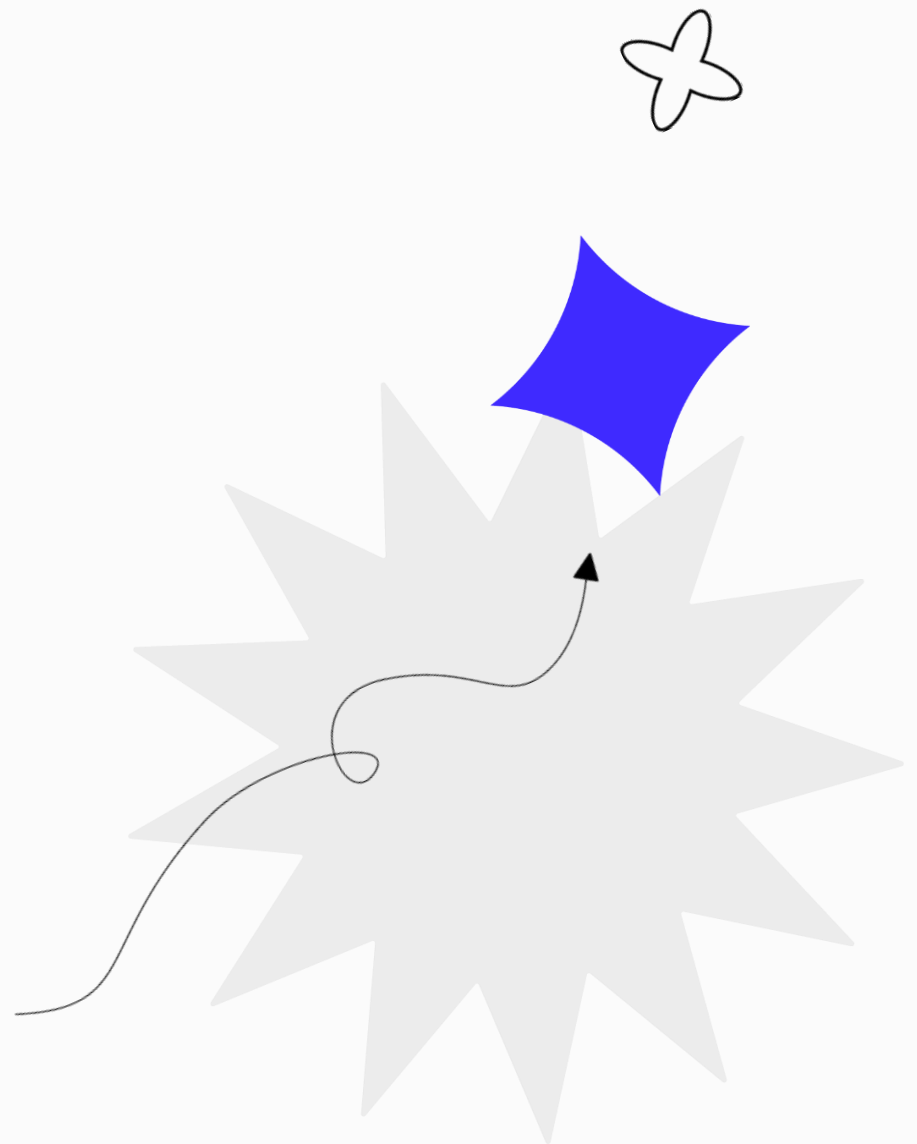
Григорий Пятецкий-Шапиро



Что такое Data Mining?

Data Mining — это

- 1 Интеллектуальный анализ данных
- 2 Посев информации
- 3 Добыча данных
- 4 Обнаружение знаний в базах данных



Что такое Data Mining?

Data Mining — добыча данных



Что такое Data Mining?

Data Mining и статистический анализ

Статистический анализ —

инструмент описания данных и проверки уже сформулированных гипотез.

Отвечает на вопросы вида:
«Сколько в среднем задолжали те, кто пропускает выплаты по кредиту четыре месяца подряд?»

Data Mining —

инструмент для поиска новых, ранее неизвестных знаний.

Отвечает на вопросы вида:
«Есть ли что-то общее у тех, кто не обслуживает свои долги?»

Что такое Data Mining?

Задача Data Mining —
найти закономерности
и взаимосвязи, которые
скрыты в данных.

Что такое Data Mining?

Лекарственные взаимодействия

Антидепрессант



+

Понижает уровень
холестерина



= ?

Что такое Data Mining?

Пиво и подгузники



+



=



Выводы урока

- ✓ Data Mining — это набор методов и практик для поиска в данных новых, ранее неизвестных знаний
- ✓ Data Mining использует методы и алгоритмы машинного обучения
- ✓ Спектр применения Data Mining очень широк — от производства до ритейла, от финансов до фармакологии
- ✓ Data Mining — это не просто проверка гипотез, это, скорее, способ их находить и формулировать