

# Краткое введение в data mining

Николай Анохин

# Data Mining как KDD

*Knowledge Discovery in Databases (KDD) – это процесс получения точных, неизвестных, потенциально полезных и интерпретируемых закономерностей из данных.<sup>1</sup>*

---

<sup>1</sup>U. Fayyad, G. Piatetsky-Shapiro, P. Smyth. From data mining to knowledge discovery: an overview. 1996

# Data Mining как моделирование

*Data Mining – процесс построения модели, хорошо описывающей закономерности, которые порождают данные.*

Подходы к построению моделей

- ▶ статистический
- ▶ машинное обучение
- ▶ вычислительный

## Пример 1. Книга – лучший подарок

PRML / C. Bishop

---

Ulysses / J. Joyce

# Некоторые важные события в истории Data Mining

1989 IJCAI-89 Workshop on Knowledge Discovery in Databases

# Data Mining – дисциплина тысячи имен

1960-e Data Fishing, Data Dredging

1980-e Knowledge Discovery in Databases

1990-e Data Mining, Database mining<sup>TM</sup>

2000-e Data Analytics, Data Science<sup>23</sup>

---

<sup>2</sup>Data Scientist is a Data Analyst who lives in California

<sup>3</sup>A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician.