

1 Введение

2 Data Mining как KDD

Источники [1] [2] [3]

1. Определение, данное на слайде часто цитируется во многих источниках. Оно было предложено в статье Fayyad et.al в 1996 году, но сам термин KDD появился на 7 лет раньше.
2. KDD - название, предложенное Григорием Пятецким-Шапиро для семинара, который он организовывал в рамках конференции IJCAI-89. Про воркшоп поговорим чуть позже в рамках исторической справки.
3. Хотя это определение часто цитируется в литературе, оно имеет ряд существенных неточностей.
4. Если интерпретировать valid, как “точный”, то зачастую возникает противоречие с остальными перечисленными качествами.
5. Рассмотрим, например, задачу анализа переписи населения. Одна из очень точных закономерностей, которую можно извлечь - женщины не служат в армии. Несмотря на высокую точность, у этой закономерности нет ни новизны, ни практической полезности.
6. Есть также и проблемы с интерпретируемостью закономерностей. В случае, когда решение принимается автоматической системой, она не нужна. А для человека интерпретируемость - слишком субъективное понятие.
7. Учитывая перечисленные недостатки данного определения, делаем вывод, что нам понадобится что-то лучше.

3 Data Mining как моделирование

Источники [4]

4 Некоторые важные события в истории Data Mining

Источники [1]

5 Data Mining – дисциплина тысячи имен

Источники [1]

- 1.

Список литературы

- [1] Journeys to Data Mining: The Journey of Knowledge Discovery
- [2] Knowledge Discovery in Databases: 10 years after
- [3] Are We Really Discovering “Interesting” Knowledge From Data?
- [4] Mining of Massive Datasets