

Краткое введение в data mining

Николай Анохин

Data Mining как KDD

Knowledge Discovery in Databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.¹

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

Data Mining как моделирование

Процесс построения модели, хорошо описывающей закономерности, которые порождают данные.

Подходы к построению моделей

- ▶ статистический
- ▶ машинное обучение
- ▶ вычислительный

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 miningTM

2000-e Data Analytics, Data Science²³

²Data Scientist is a Data Analyst who lives in California

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