

DATA MINING FINDS HIDDEN RELATIONSHIPS

FRANK MCDONOUGH

*PRESENTED TO MASTER'S DEGREE CANDIDATES
INFOTEC, MEXICO CITY, UT- DALLAS
JANUARY-MARCH 2016
UPDATED JANUARY 2019*

INTRODUCTION

I delivered an earlier version of this 2019 lecture in 2016 to a graduate class of 30 Master's Degree candidates in Mexico City for the University of Texas (Dallas).

In this 2019 version, there are 25 slides. The first ten are available here for public review. The goal in the original and current version is to acquaint mid-level and senior managers, with sufficient knowledge to manage Data Mining experts and technicians.

- Data Mining has evolved. We used to ask, “What were our unit sales in New England last month? Today, we ask, “What is likely to happen with Boston’s unit sales next month and why?”
- The lecture provides two important examples of Data Mining representing the pharmaceutical industry and NY city’s surveillance system.

To view the complete lecture, contact Frank McDonough at frank@frankamcdonough.com. With your request, please enclose a few sentences describing your current interests and background.

RELATED TERMS UNDER THE UMBRELLA OF BUSINESS INTELLIGENCE

Data Discovery

Data Mining

Big Data

Data Analytics

Data Science

DATA MINING IS DATA DISCOVERY

Uses computers and software to discover patterns in large data sets.

It discovers something “new.”

Then, it transforms what was found into something understandable for further use.

DATA DISCOVERY VS. DATA MINING

- Both are Business/ Intelligence technologies.
- Data Discovery provides summary data.
- Data Mining provides very detailed information often at the individual person level.

USING DD AND DM

Two Steps

Use Data Discovery to identify areas for further investigation
e.g. pinpoint problems with sales in a particular region

Then use Data Mining to drill down into the data to learn
about the behavior of potential buyers in that region.

EXAMPLES

Data Discovery: How do sales this year compare to last year? What can we predict for the coming year?

Data Mining: Who is likely to buy a mutual fund next year? What are the characteristics of each of these likely buyers?

TECHNOLOGIES SUPPORTING DATA MINING

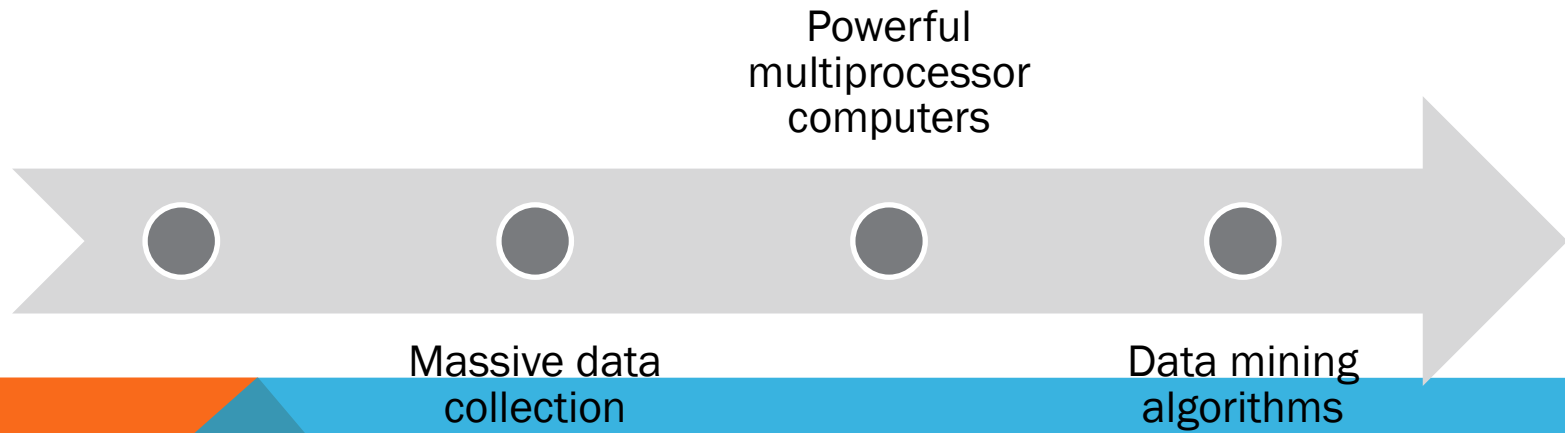


Table 1. Steps in the Evolution of Data Mining.

Evolutionary Step	Business Question	Enabling Technologies	Product Providers	Characteristics
Data Collection (1960s)	"What was my total revenue in the last five years?"	Computers, tapes, disks	IBM, CDC	Retrospective, static data delivery
Data Access (1980s)	"What were unit sales in New England last March?"	Relational databases (RDBMS), Structured Query Language (SQL), ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Retrospective, dynamic data delivery at record level
Data Warehousing & Decision Support (1990s)	"What were unit sales in New England last March? Drill down to Boston."	On-line analytic processing (OLAP), multidimensional databases, data warehouses	Pilot, Comshare, Arbor, Cognos, Microstrategy	Retrospective, dynamic data delivery at multiple levels
Data Mining (Emerging Today)	"What's likely to happen to Boston unit sales next month? Why?"	Advanced algorithms, multiprocessor computers, massive databases	Pilot, Lockheed, IBM, SGI, numerous startups (nascent industry)	Prospective, proactive information delivery

THREE TYPES OF DATA MINING

The first type

Cluster analysis identifies distinct areas or "hot spots" where a similar crime has happened repeatedly over several months.

For example, 24 handbag snatches occurred in the left hand aisle in the National Cathedral on Friday evenings between 5 and 8 pm