

# Machine Learning and Data Mining

Academic Year 2023 – 2024  
Alma Mater Studiorum · University of Bologna

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Data . . . . .	2
1.1.1	Data sources . . . . .	2
1.1.2	Software . . . . .	2
1.1.3	Insight . . . . .	2
<b>2</b>	<b>Business Intelligence</b>	<b>4</b>

# Acronyms

**BI** Business Intelligence

**DM** Data Mart

**DSS** Decision Support System

**DWH** Data Warehouse

**EIS** Executive Information System

**ERP** Enterprise Resource Planning

**MIS** Management Information System

**OLAP** On-Line Analytical Processing

**OLTP** On-Line Transaction Processing

# 1 Introduction

## 1.1 Data

**Data** Collection of raw values.

Data

**Information** Organized data (e.g. relationships, context, ...).

Information

**Knowledge** Understanding information.

Knowledge

### 1.1.1 Data sources

**Transaction** Business event that generates or modifies data in an information system (e.g. database).

Transaction

**Signal** Measure produced by a sensor.

Signal

**External subjects**

### 1.1.2 Software

**On-Line Transaction Processing (OLTP)** Class of programs to support transaction oriented applications and data storage. Suitable for real-time applications.

On-Line Transaction Processing

**Enterprise Resource Planning (ERP)** Integrated system to manage all the processes of a business. Uses a shared database for all applications. Suitable for real-time applications.

Enterprise Resource Planning

### 1.1.3 Insight

Decision can be classified as:

**Structured** Established and well understood situations. What is needed is known.

Structured decision

**Unstructured** Unplanned and unclear situations. What is needed for the decision is unknown.

Unstructured decision

Different levels of insight can be extracted by:

**Management Information System (MIS)** Standardized reporting system built on existing OLTP. Used for structured decisions.

Management Information System

**Decision Support System (DSS)** Analytical system to provide support for unstructured decisions.

Decision Support System

**Executive Information System (EIS)** Formulate high level decisions that impact the organization.

Executive Information System

**On-Line Analytical Processing (OLAP)** Grouped analysis of multidimensional data. Involves large amount of data.

On-Line Analytical Processing

**Business Intelligence (BI)** Applications, infrastructure, tools and best practices to analyze information. Business Intelligence

**Big data** Large and/or complex and/or fast changing collection of data that traditional DBMSs are unable to process. Big data

**Structured** e.g. relational tables.

**Unstructured** e.g. videos.

**Semi-structured** e.g. JSON.

**Analytics** Structured decision driven by data. Analytics

**Data mining** Discovery process for unstructured decisions. Data mining

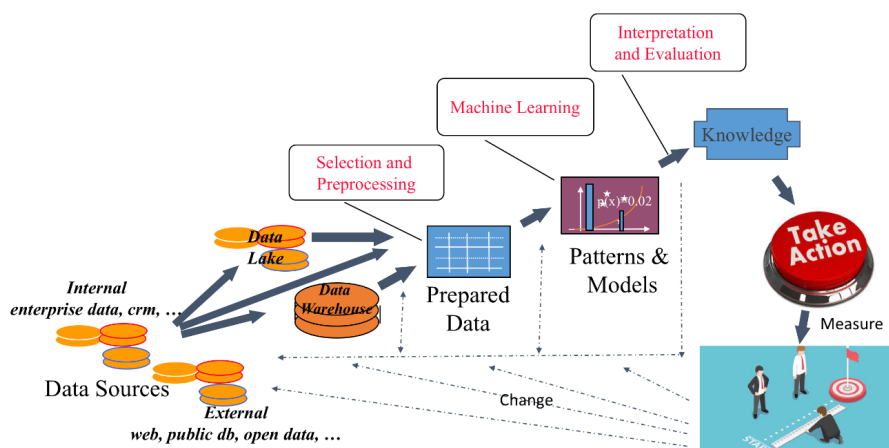


Figure 1.1: Data mining process

**Machine learning** Learning models and algorithms that allow to extract patterns from data. Machine learning

## 2 Business Intelligence

**Business Intelligence** Transform raw data into information. Deliver the right information to the right people at the right time through the right channel. Business Intelligence

**Data Warehouse (DWH)** Optimized repository that stores information for decision making processes. DWHs are a specific type of DSS. Data Warehouse

Features:

- Subject-oriented: focused on enterprise specific concepts.
- Integrates data from different sources and provides an unified view.
- Non-volatile storage with change tracking.

**Data Mart (DM)** Subset of the primary DWH with information relevant to a specific business area. Data Mart