Data Mining for Business

Introduction to Data Mining Concepts

Dr. Shipra Maurya
Department of Management Studies
IIT (ISM) Dhanbad
Email: shipra@iitism.ac.in

What is Data Mining?

- “Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes”.

- The foundation of data mining consists of three intertwined disciplines:
  - **Statistics**: numeric study of data relationships
  - **AI**: human-like intelligence displayed by machines
  - **ML**: algorithms learn from data to make predictions

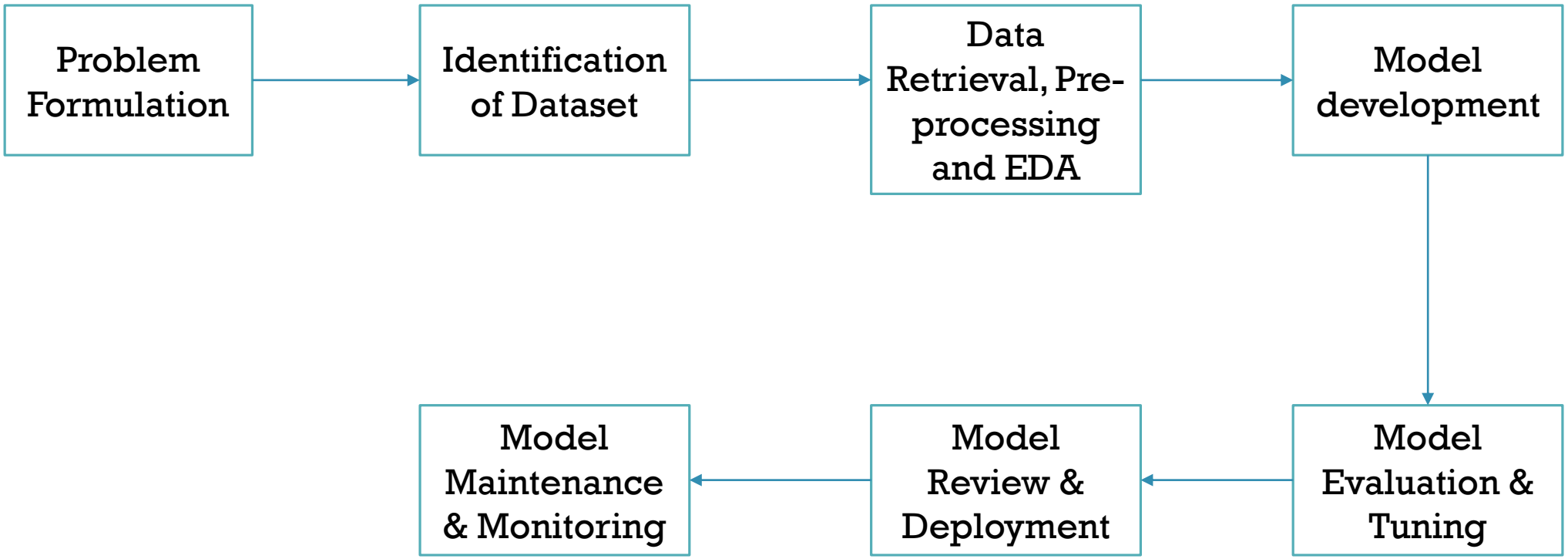
Image source : Internet

Why Data Mining?

- Making smart business decisions is not driven by how much data you have – but by how quickly you can discover insights from all that data

- Data mining helps to achieve above objective with the help of descriptive, predictive and prescriptive modeling

Data Mining Pipeline/ Process



Key concepts (1/3)

- Target variable / Dependent variable / Response variable / Outcome variable / Output variable
- Feature / Input variable / Independent variable / Predictor / Attribute / Input feature
- Categorical variable / Factor variable
- Holdout set / Holdout data / Test set / Test data
- Training data / Training set
- Validation data / Validation set

Key concepts (2/3)

- Algorithm
- Model
- **Prediction / Estimation**
- Observation / Record / Case
- **Sample** – single observation in ML / collection of observations in Statistics
- **Score** – a predicted value or class
- **Scoring** - using a model developed with training data to predict output values in new data
- **Variable** - Any measurement on the records, including both the input (X) and the output (Y) variables

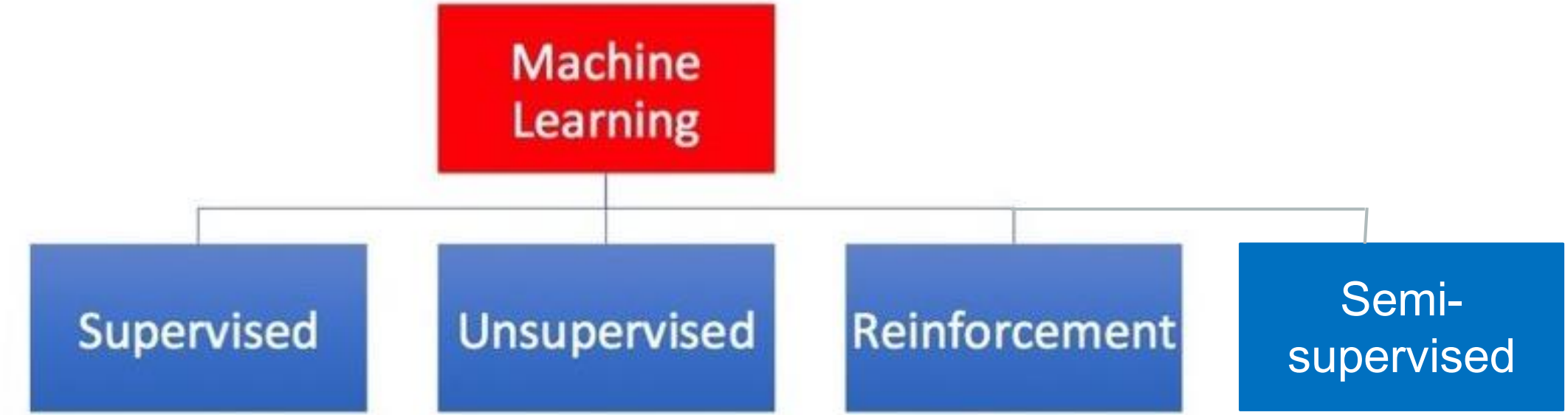
Key concepts (3/3)

- Fitting a model
- Re-fit a model
- **Supervised learning –** Regression and Classification Methods
- **Unsupervised learning –** Cluster analysis, Collaborative filtering and association rules
- **Semi-supervised learning**
- **Reinforcement learning**

Types of data

- Qualitative – Nominal, Ordinal
- Quantitative – Discrete, Continuous

Types of Machine Learning



Task Driven
(Predict next value)

Target variable is defined (also known as labeled data)

Data Driven
(Identify Clusters)

Target variable is not known (also known as unlabeled data)

Learn from
Mistakes

Both data driven and task driven

Image source : Internet

Reinforcement learning

- A technique that enables an agent to learn in an interactive environment by **trial and error using feedback** from its own actions and experiences

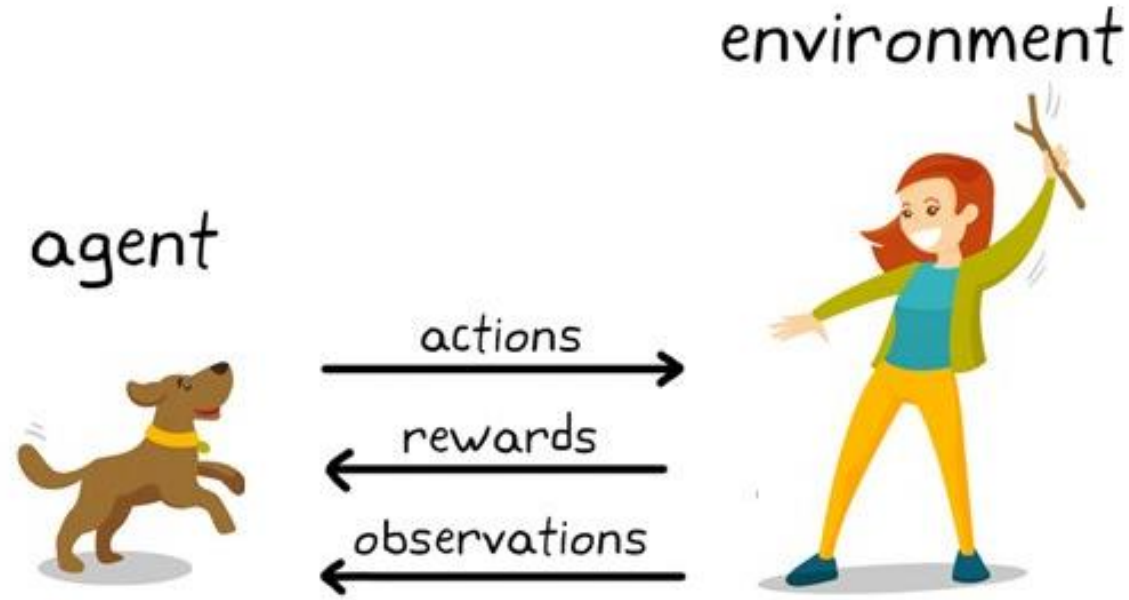
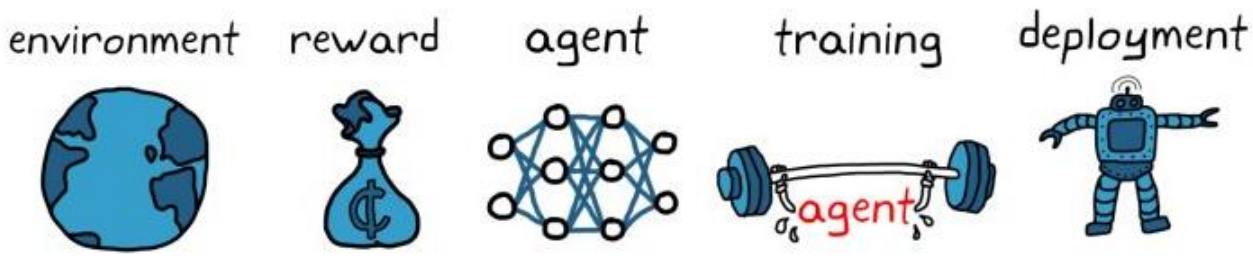


Image source : Internet

Reinforcement Learning in ML

Reinforcement Learning Workflow



Reinforcement Learning Workflow – KDNuggets

Reinforcement Learning example

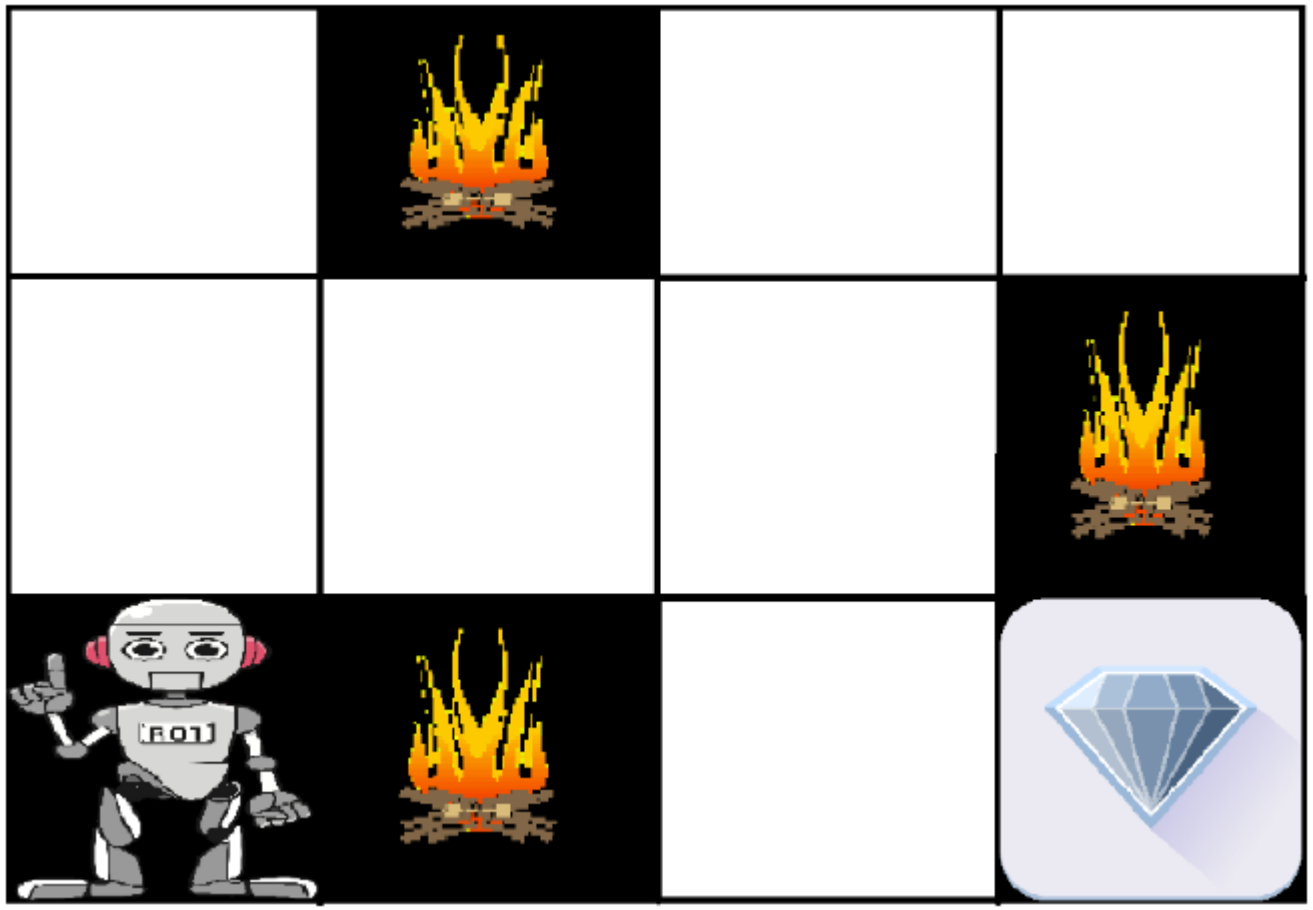
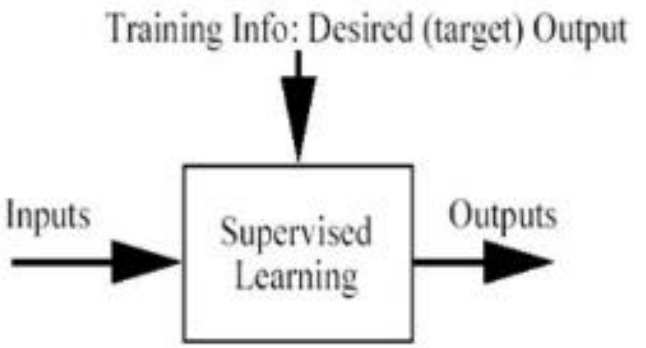


Image source : Internet

Reinforcement Learning vs. Supervised Learning

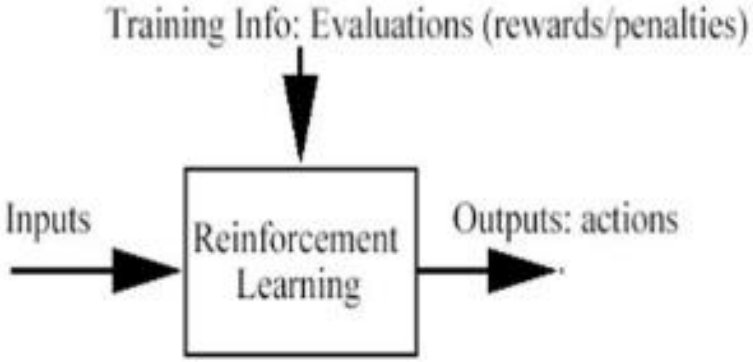
Supervised Learning



Error = (target output - actual output)

Correct set of action for performing a task is supplied to the agent in the form of actual target value

Reinforcement Learning



Objective: Get as much reward as possible

Uses rewards and punishments as signals for positive and negative behavior

Image source : Internet

Reinforcement Learning vs. Unsupervised Learning

- The goal is different in both learning types
- Goal of unsupervised learning is to **find similarities and differences between the data points**
- Goal of reinforcement learning is to find a suitable action model that would **maximize the total cumulative reward** of the agent

Types of Reinforcement

- Positive
- Negative

Mostly used Reinforcement Learning Algorithms

- **Model-based methods** : Example – Markov Decision Processes (MDPs)
- **Model-free methods** :
  - Q-learning (value-based)
  - SARSA (State-Action-Reward-State-Action) (Policy based)
  - Deep-Q-Networks(DQN)
  - Deep Deterministic Policy Gradient (DDPG)

Applications of Reinforcement Learning

- Autonomous self-driving cars
- Robotics for Industrial automation
  - https://youtu.be/ZhsEKTo7V04
- Building AI for computer games
- RL based agents for stock trading
- Chatbots

Reinforcement Learning – Reading material

- Reinforcement Learning-An Introduction by Richard Sutton and Andrew Barto

Thank you!