

**Machine learning** is about extracting knowledge from data. It is a research field at the intersection of statistics, artificial intelligence, and computer science and is also known as predictive analytics or statistical learning. The application of machine learning methods has in recent years become ubiquitous in everyday life.

### Problems Machine Learning Can Solve

The most successful kinds of machine learning algorithms are those that automate decision-making processes by generalizing from known examples. In this setting, which is known as supervised learning, the user provides the algorithm with pairs of inputs and desired outputs, and the algorithm finds a way to produce the desired output given an input.

### Types of machine learning

- A. **Supervised Learning**-Machine learning algorithms that learn from input/output pairs are called supervised learning algorithms because a "teacher" provides supervision to the algorithms in the form of the desired outputs for each example that they learn from.
- B. **Unsupervised learning**-In unsupervised learning, only the input data is known, and no known output data is given to the algorithm.

### Why Python?

One of the main advantages of using Python is the ability to interact directly with the code, using a terminal or other tools like the Jupiter Notebook, which we'll look at shortly.

### scikit-learn

scikit-learn is an open source project, meaning that it is free to use and distribute, and anyone can easily obtain the source code to see what is going on behind the scenes.

### Essential Libraries and Tools

- ❖ The Jupyter Notebook is an interactive environment for running code in the browser
- ❖ NumPy. NumPy is one of the fundamental packages for scientific computing in Python. It contains functionality for multidimensional arrays, high-level mathematical functions such as linear algebra operations and the Fourier transform, and pseudorandom number generators.
- ❖ SciPy. SciPy is a collection of functions for scientific computing in Python. It provides, among other functionality, advanced linear algebra routines, mathematical function optimization, signal processing, special mathematical functions, and statistical distributions.
- ❖ Matplotlib. matplotlib is the primary scientific plotting library in Python. It provides functions for making publication-quality visualizations such as line charts, histograms, scatter plots, and so on.
- ❖ Pandas. pandas is a Python library for data wrangling and analysis.

### Python 2 Versus Python 3

Python 2 (more precisely, 2.7) and Python 3 (with the latest release being 3.5 at the time of writing).