Practical Business Python

Lecture 11: Intro to Machine Learning in Python

Iegor Vyshnevskyi

Woosong University

November 30, 2023

Agenda

1. Intro to Machine Learning
2. Machine Learning Process
3. Machine Learning Applications
4. Machine Learning Algorithms / Methods
5. Machine Learning in Practice
6. In-class Assignment

Acknowledgment: Used a number of open sources and materials from the web.

1. Intro to Machine Learning

*What is **Machine Learning?***

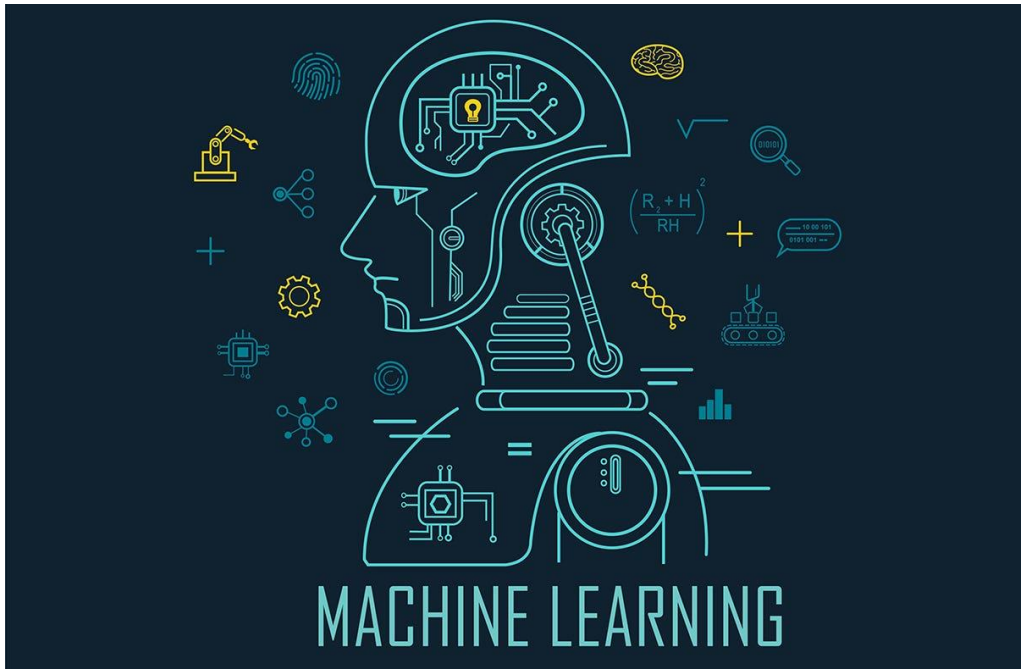
Machine Learning is a branch in computer science that studies the design of algorithms that can learn.

Arthur Samuel, a pioneer in the field of artificial intelligence and computer gaming, coined the term “Machine Learning”. He defined machine learning as – “Field of study that gives computers the capability to learn without being explicitly programmed”.

AI vs. ML

Although Artificial intelligence (AI) and machine learning (ML) are used interchangeably, but they differ with uses, data sets, and more.

- While AI encompasses the idea of a machine that can mimic human intelligence, ML does not.
- ML aims to teach a machine how to perform a specific task and provide accurate results by identifying patterns.



*What is **Machine Learning?** (cont.)*

- In general, *Machine Learning(ML)* can be explained as automating and improving the learning process of computers based on their experiences without being actually programmed i.e. without any human assistance.

- ML can be seen as a form of artificial intelligence (AI).

- “Machine learning is a scientific discipline that is concerned with the design and development of algorithms that allow computers to evolve behaviours based on empirical data, such as from sensor data or databases.” (Wikipedia)

- Primary goal of a ML implementation is to develop a general purpose algorithm that solves a practical and focused problem.

- Important aspects in the process include *data*, *time*, and *space* requirements.

- The goal of a learning algorithm is to produce a result that is a rule and is as accurate as possible.

How Do Machines Learn?

- Traffic prediction, face recognition, product recommendations, self-driving cars—machine learning is at the core of the future of technology.

- With machine learning, machines can match and even transcend human information processing and learning capabilities to solve complex problems.

- Machines and their learning accuracy determine how good they are at solving everyday problems. Understanding how machines learn should be a priority for organizations worldwide, so they can leverage their ability and take their operations to the next level.

Usefulness of ML for Business Analytics

- In today's quickly-evolving corporate landscape, companies must often engage in intense competition to secure users and customers.
- In the age of big data and in-depth analysis of customer behavior, machine learning (ML) solutions are emerging as the de facto way for companies to gain a competitive edge.
- ML was discovered to be a good fit for the corporate landscape, providing cost-effective solutions to problems that previously required a lot of resources.
- Driven by market leaders, such as Google, Microsoft, and Amazon, the corporate landscape as a whole has gravitated towards using ML.
- Smart automation has enabled businesses to effectively deploy low-cost, high-accuracy ML solutions to replace low-skilled workers.
- Machine learning has a variety of applications in the corporate sector, as its capabilities have made it a natural fit for the requirements of an ever-increasing market.

Usefulness of ML for Business Analytics

- 56% of organizations today are using machine learning in at least one business function, according to a recent McKinsey survey. That means that ML (Machine Learning) will benefit more than half of companies in 2022.

Overall, ML is used in business to improve scalability and develop company operations for organisations worldwide and in any industry.

Advantages of using Python for ML

- ***Ease of Learning and Readability :***
  - Python's clear syntax and readability make it beginner-friendly.
  - Quick onboarding for newcomers to machine learning.
- ***Extensive Library Support :***
  - Extensive libraries empower efficient model development.
  - Access to cutting-edge techniques and a vibrant community.
- ***Integration Versatility and Integration:***
  - Seamless integration with other languages and technologies.
  - Python's adaptability across web development, data analysis, and big data tools.

2. Machine Learning Process

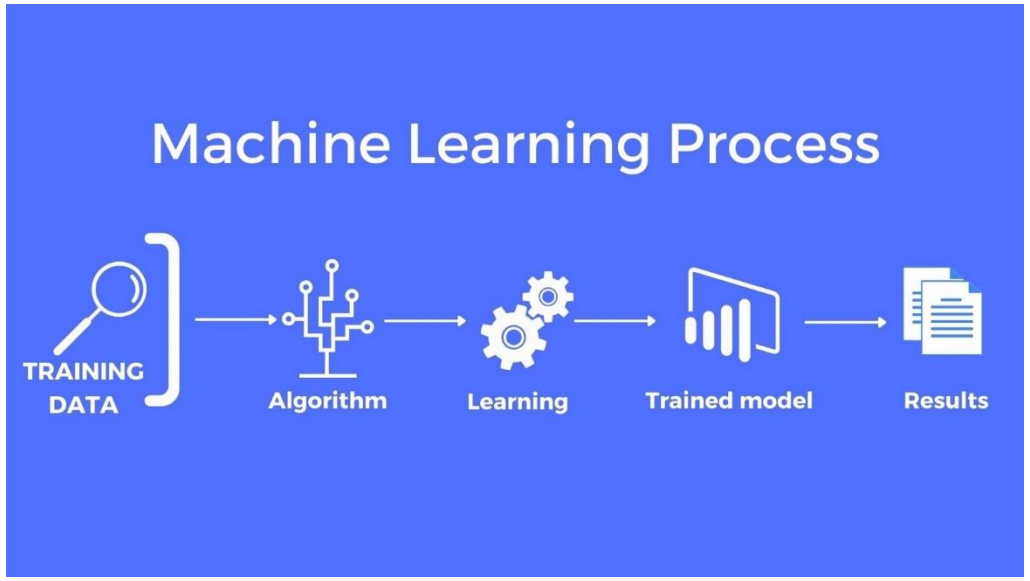
Machine Learning Process

The process starts with feeding good quality data and then training our machines(computers) by building machine learning models using the data and different algorithms.

The choice of algorithms depends on what type of data do we have and what kind of task we are trying to automate.

As such, there are three main phases in an ML process:

- Training Phase: Training Data is used to train the model by using expected output with the input. Output is the learning model.
- Validation/Test Phase: Measuring the validity and fit of the model. How good is the model? Uses validation dataset, which can be a subset of the initial dataset.
- Application Phase: Run the model with real world data to generate results



3. Machine Learning Applications

ML Applications

- ***Image Classification:*** the process by which algorithms are trained to analyze images and find out what they contain. Image classification as a machine learning solution has become highly popular in the enterprise sector, mainly due to its capability to disrupt existing systems created for the same purpose. Previously, human labor was required to go through vast amounts of data and label them. Today, giants like Facebook, Twitter, and Google are using image classification to prevent unwanted content from going viral.
- ***Predictive Modeling:*** is a category of ML solutions that mines large amounts of data to predict the outcomes of potential scenarios. These predictions can then be utilized to make informed business decisions. Predictive modeling algorithms essentially provide a forecast of the future based on past data, allowing companies to make business moves in preparation for these predictions.
- ***Real-time chatbot systems:*** one of the foremost forms of automation. ML enables chatbots to be more productive and more interactive (Alexa, Google Assistant, Siri, Watson Assistant).
- ***Customer recommendation engines:*** ML powers the customer recommendation engines built to deliver customized experiences and improve the overall customer experience. Here, algorithms analyze data points about each customer, including the customer's previous purchases, and other data sets like demographic trends, an organization's present inventory, and the purchase histories of other customers in order to know what services and products to offer as recommendations to each individual customer (Amazon, Walmart, Netflix, and YouTube).

And many more: https://www.techfunnel.com/information-technology/machine-learning-business-applications/

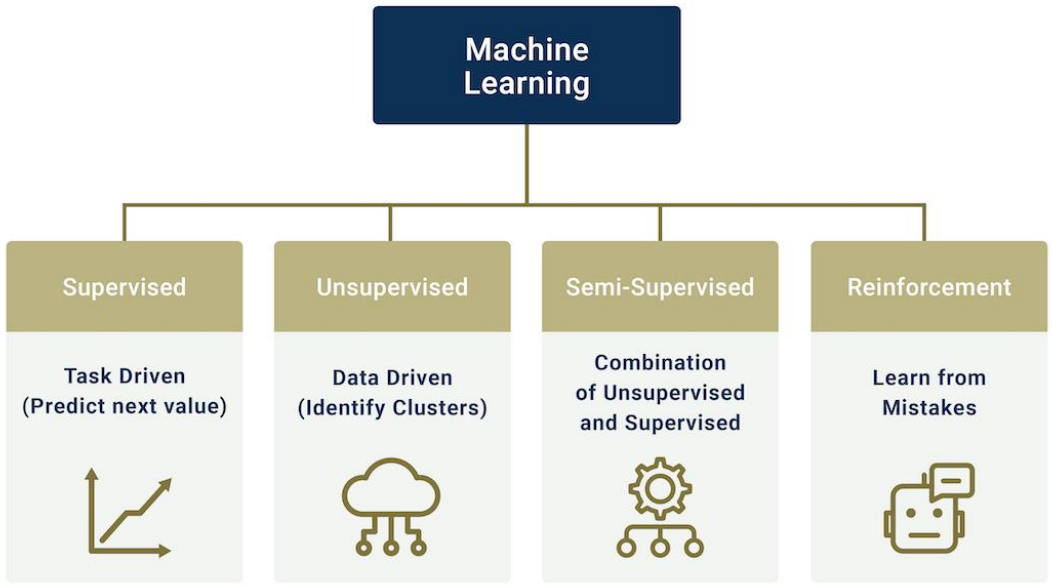
4. Machine Learning Algorithms / Methods

ML Algorithms

- Algorithms are *sets of rules* created to help computers perform problem-solving operations. Machine learning algorithms are systems that run the machine learning process. They are *the key* to how computers interact with data to learn how to interpret it and put it to use.
- Interacting with, interpreting, and using data can happen in many different ways. Machine learning algorithms must be trained to learn how to perform any task. This training often requires the machine *to be fed large volumes of data multiple times over*.
- The machine learns the specific representation after applying an algorithm to a dataset is referred to as a machine learning model or hypothesis.
- Upon multiple instances of using algorithms, computers get better at achieving desirable results. This type of zero interference model makes up the core of all machine learning processes.
    - Therefore, all machine learning relies on the availability of data to train the machines. This data must be well-organized and labeled to ensure the outcomes are accurate.
- There are *many types of machine learning algorithms* based on how the machine interacts with the data and what it needs to accomplish.

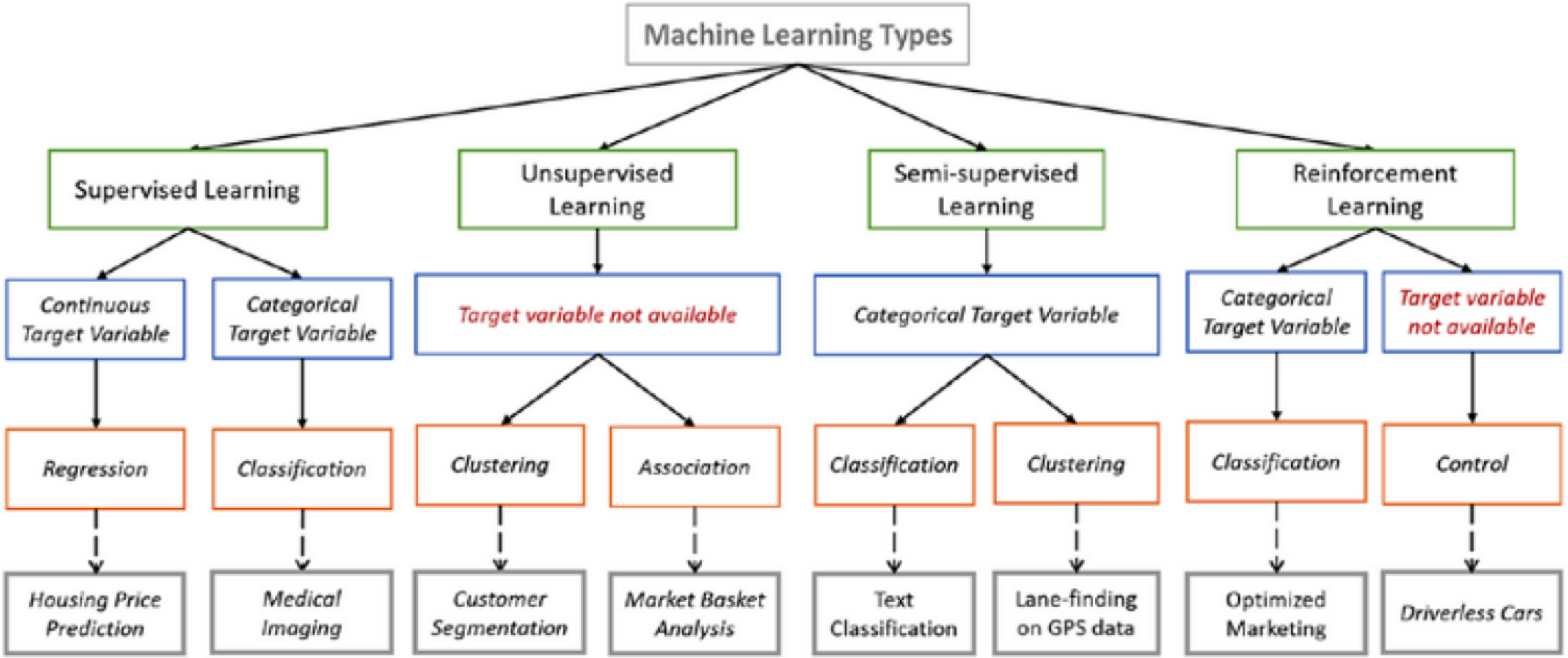
*Overview of **ML Algorithms***

There are mainly four types of learning.

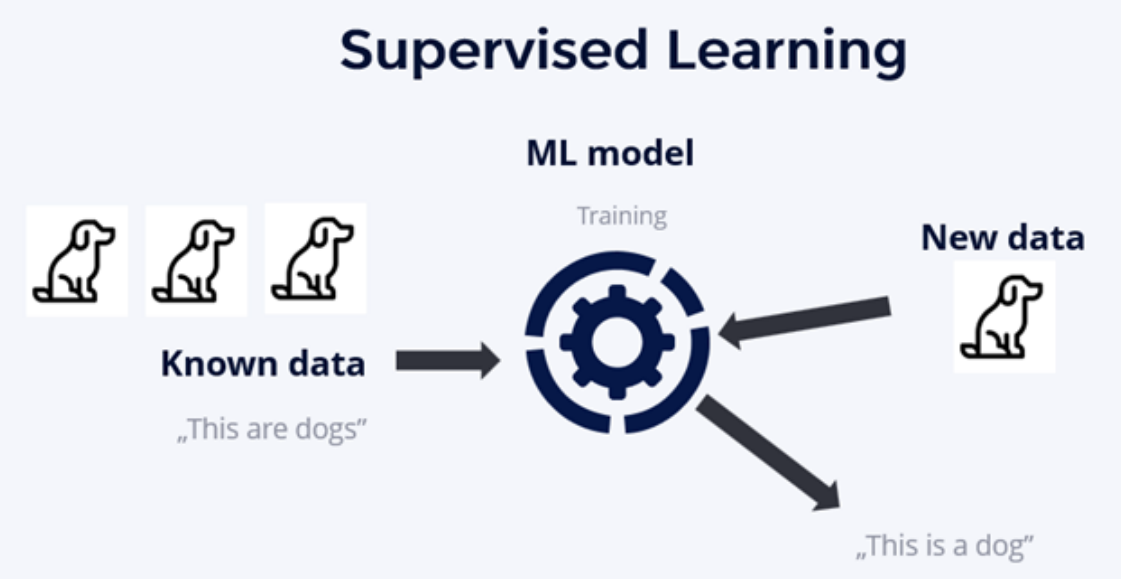


- In **supervised learning** (SML), the learning algorithm is presented with labelled example inputs, where the labels indicate the desired output. SML itself is composed of **classification**, where the output is categorical, and **regression**, where the output is numerical.
- In **unsupervised learning** (UML), no labels are provided, and the learning algorithm focuses solely on detecting structure in unlabeled input data.
- Note that there are also **semi-supervised learning** approaches that use labelled data to inform unsupervised learning on the unlabeled data to identify and annotate new classes in the dataset (also called novelty detection).
- **Reinforcement learning**, the learning algorithm performs a task using feedback from operating in a real or synthetic environment.

*Overview of **ML Algorithms** (cont.)*



Supervised *Learning vs.* ***Unsupervised*** *Learning*

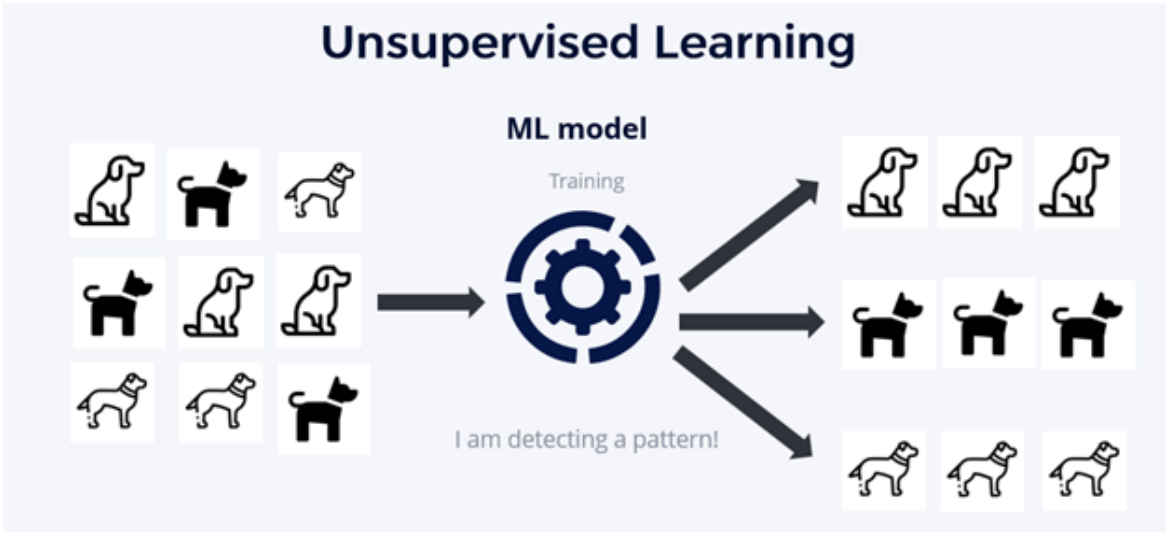


•**Classification**: A classification problem is when the output variable is a category, such as “Red” or “blue” or “disease” and “no disease”.

•**Regression**: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

•**Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

•**Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.



5. Machine Learning in Practice

What we do today

k-Nearest Neighbors (KNN)

- The KNN or k-nearest neighbors algorithm is one of the simplest machine learning algorithms and is an example of instance-based learning, where new data are classified based on stored, labeled instances.
- It is a *Supervised Non-linear classification* algorithm. KNN is a Non-parametric algorithm i.e. it doesn't make any assumption about underlying data or its distribution. It is one of the simplest and widely used algorithm which depends on it's k value(Neighbors) and finds it's applications in many industries like finance industry, healthcare industry etc.
- More specifically, the distance between the stored data and the new instance is calculated by means of some kind of a similarity measure. This similarity measure is typically expressed by a distance measure such as the Euclidean distance, cosine similarity or the Manhattan distance.
- In other words, the similarity to the data that was already in the system is calculated for any new data point that you input into the system. Then, you use this similarity value to perform predictive modeling.
- Predictive modeling is either classification, assigning a label or a class to the new instance, or regression, assigning a value to the new instance. Whether you classify or assign a value to the new instance depends of course on your how you compose your model with KNN.

Example from: https://www.r-bloggers.com/2015/03/machine-learning-in-r-for-beginners/

Example Dataset

Edgar Anderson's Iris Data

- From the *iris* manual page:
  - This famous (Fisher's or Anderson's) *iris* data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica.

6. In-class Assignment