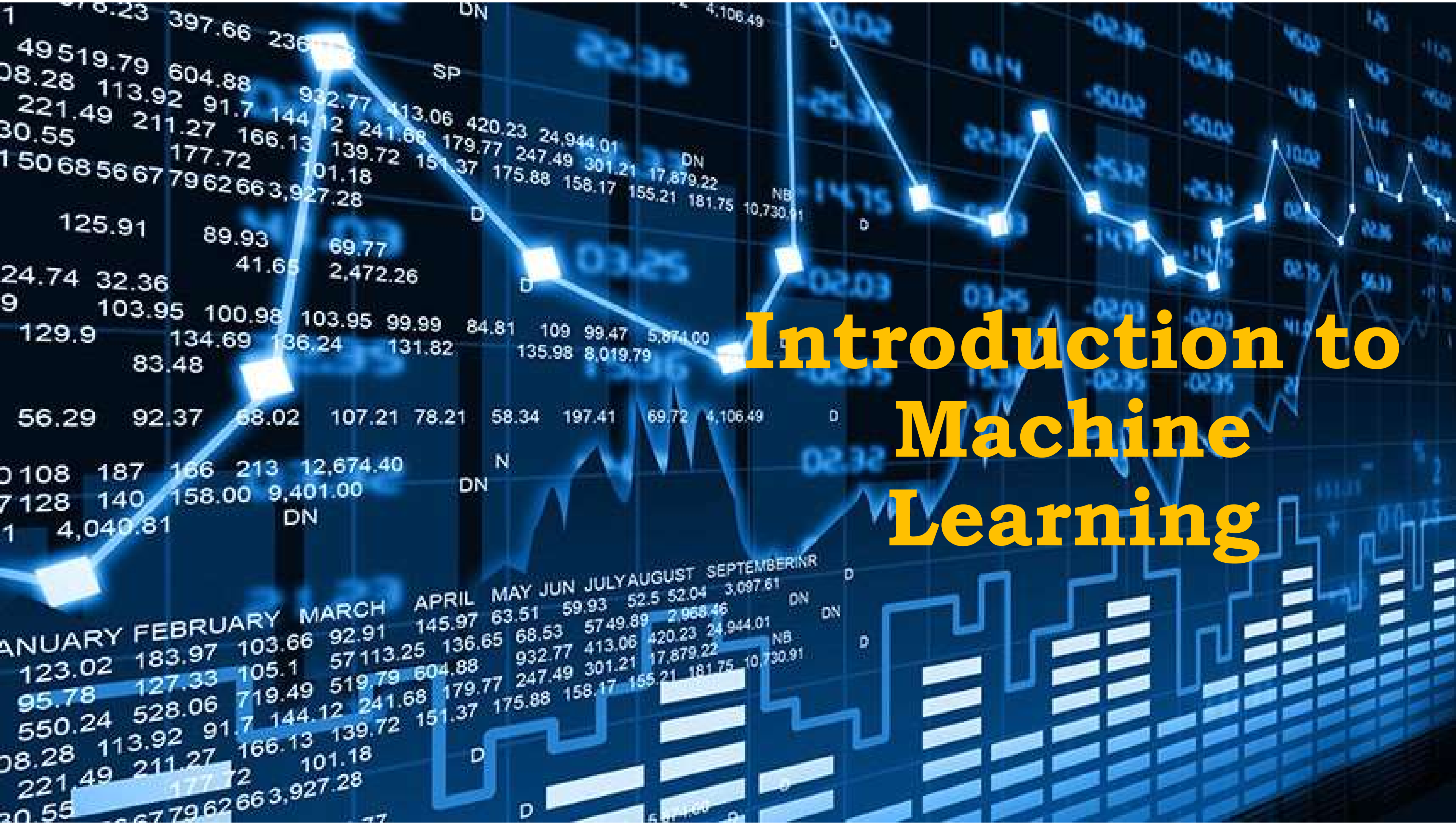
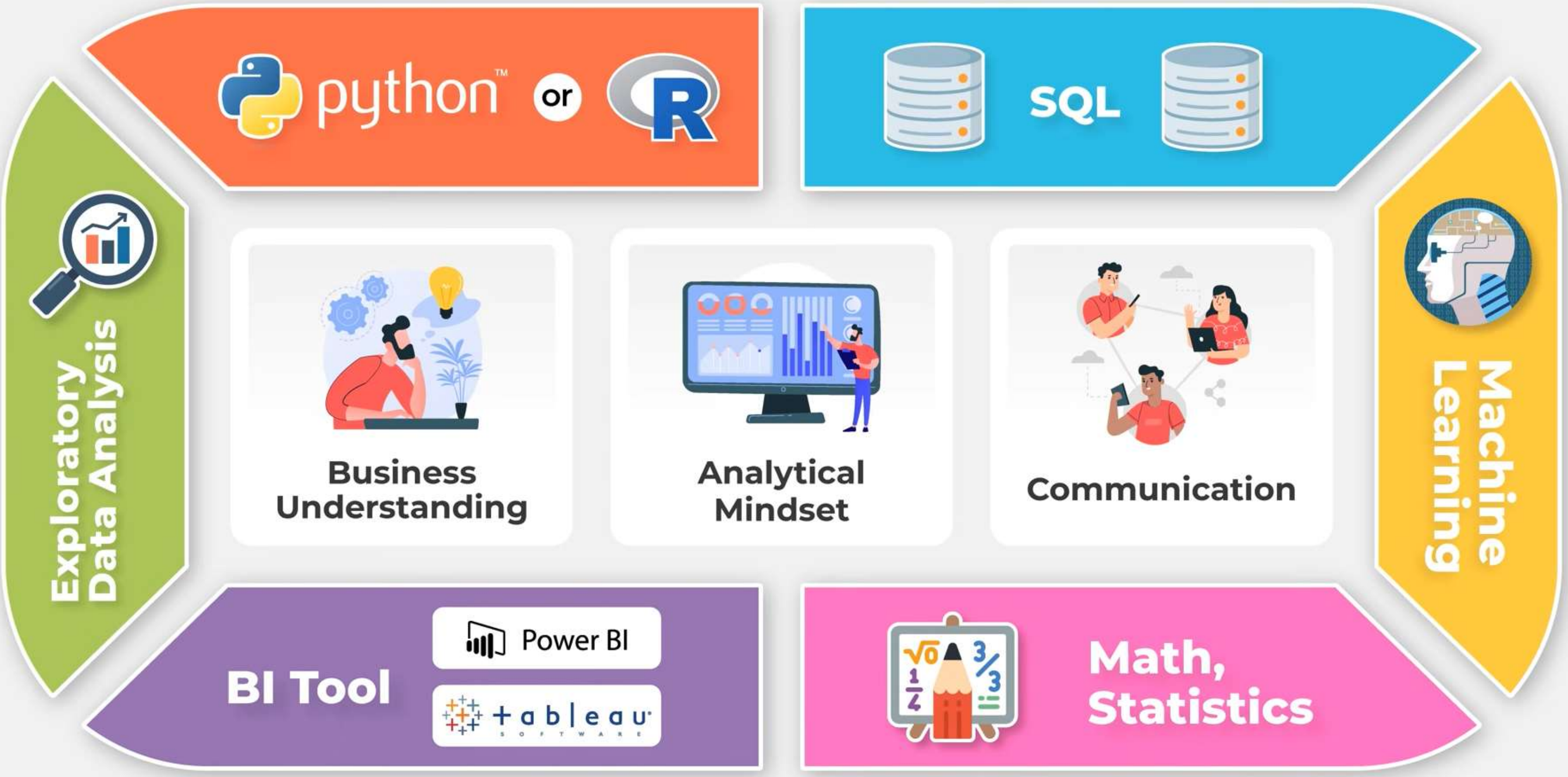


Introduction to Machine Learning

Data Scientist Core Skills Surrounded By Tool Skills



python or R

SQL

Exploratory Data Analysis

Business Understanding

Analytical Mindset

Communication

Machine Learning

BI Tool

Power BI

tableau

Math, Statistics

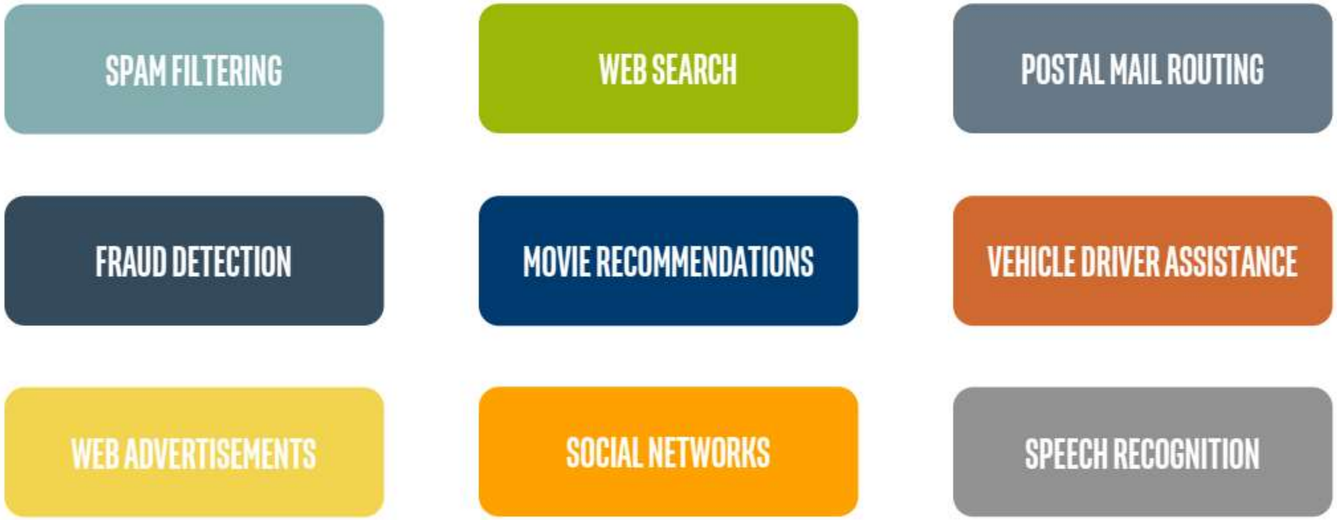
Machine Learning

- Definition: The use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data
- In short, machine learning allows computers to learn to recognize patterns and infer/predict from data i.e. being explicitly programmed where to look!
- Let’s understand this with a simple example:

0	**1**	**1**	**2**	**3**	**5**	**8**	**?**	**?**	**?**

Could we make computers do the guess?

Machine Learning in Daily Lives



SPAM FILTERING

WEB SEARCH

POSTAL MAIL ROUTING

FRAUD DETECTION

MOVIE RECOMMENDATIONS

VEHICLE DRIVER ASSISTANCE

WEB ADVERTISEMENTS

SOCIAL NETWORKS

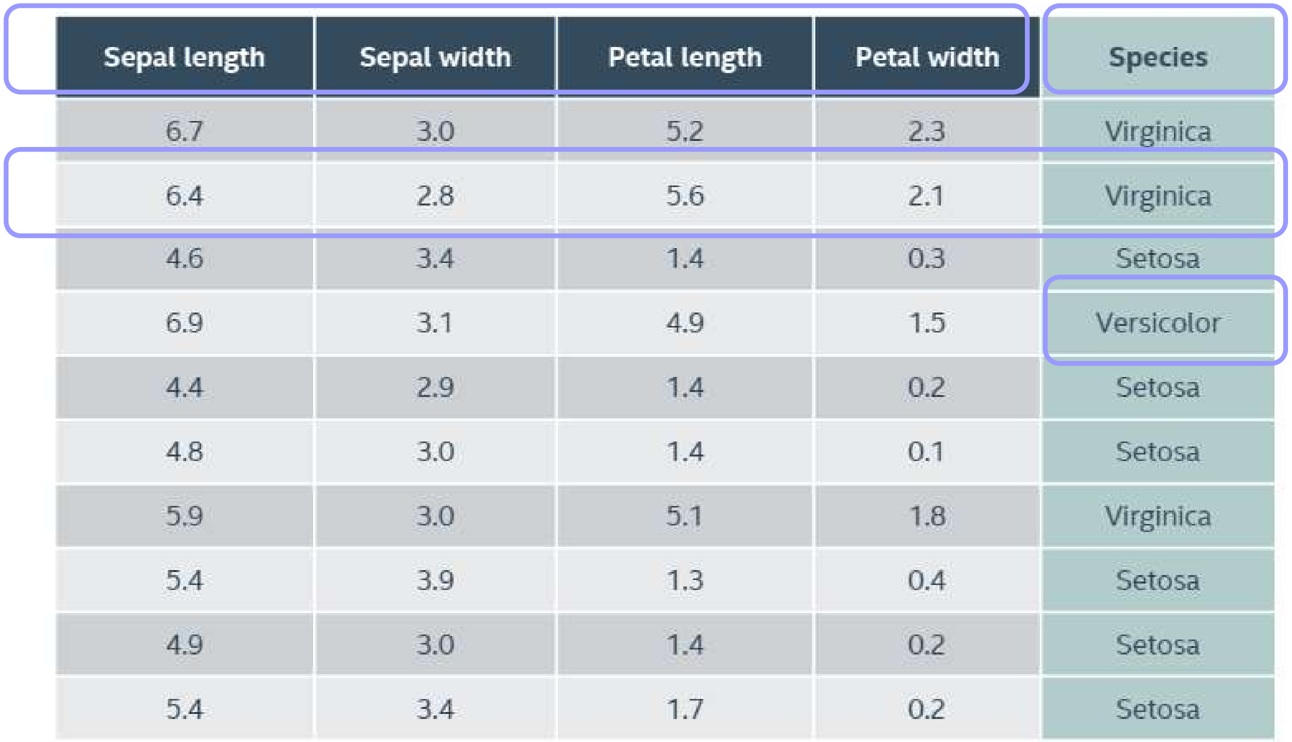
SPEECH RECOGNITION

Machine Learning Vocabulary

Dataset: Tabulated data

Features: properties of the data

Example: one row

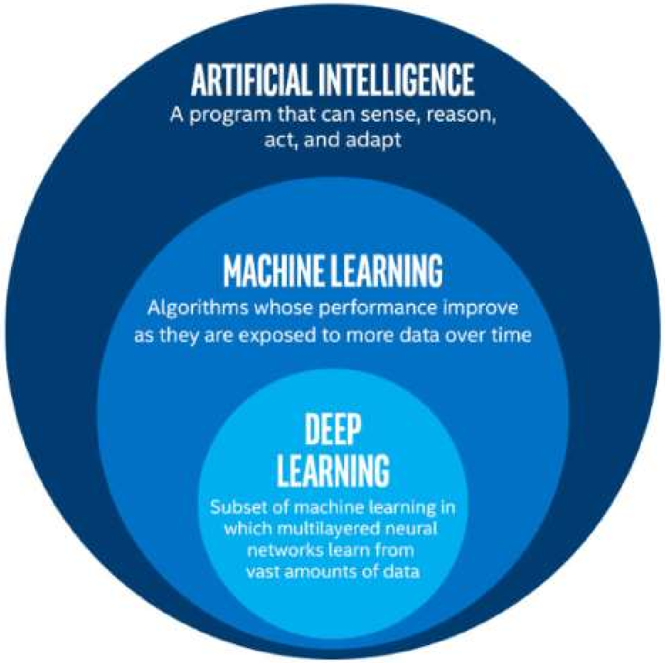


Sepal length	Sepal width	Petal length	Petal width	Species
6.7	3.0	5.2	2.3	Virginica
6.4	2.8	5.6	2.1	Virginica
4.6	3.4	1.4	0.3	Setosa
6.9	3.1	4.9	1.5	Versicolor
4.4	2.9	1.4	0.2	Setosa
4.8	3.0	1.4	0.1	Setosa
5.9	3.0	5.1	1.8	Virginica
5.4	3.9	1.3	0.4	Setosa
4.9	3.0	1.4	0.2	Setosa
5.4	3.4	1.7	0.2	Setosa

Target: column to predict

Label: target value for single data point

Machine Learning Categories



ARTIFICIAL INTELLIGENCE

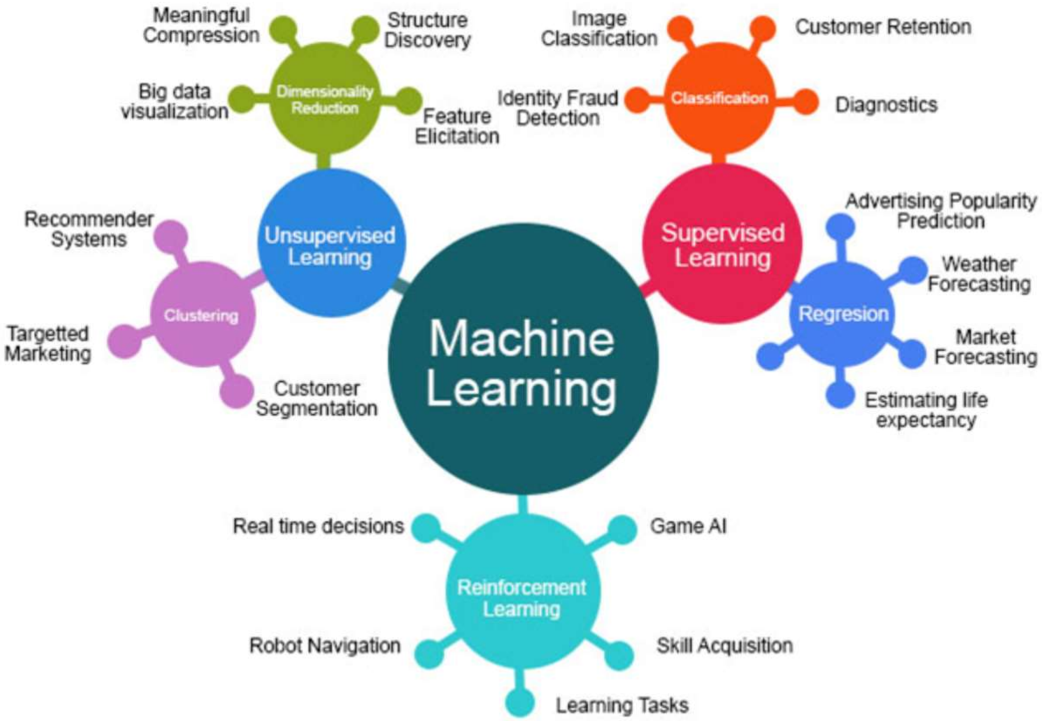
A program that can sense, reason, act, and adapt

MACHINE LEARNING

Algorithms whose performance improve as they are exposed to more data over time

DEEP LEARNING

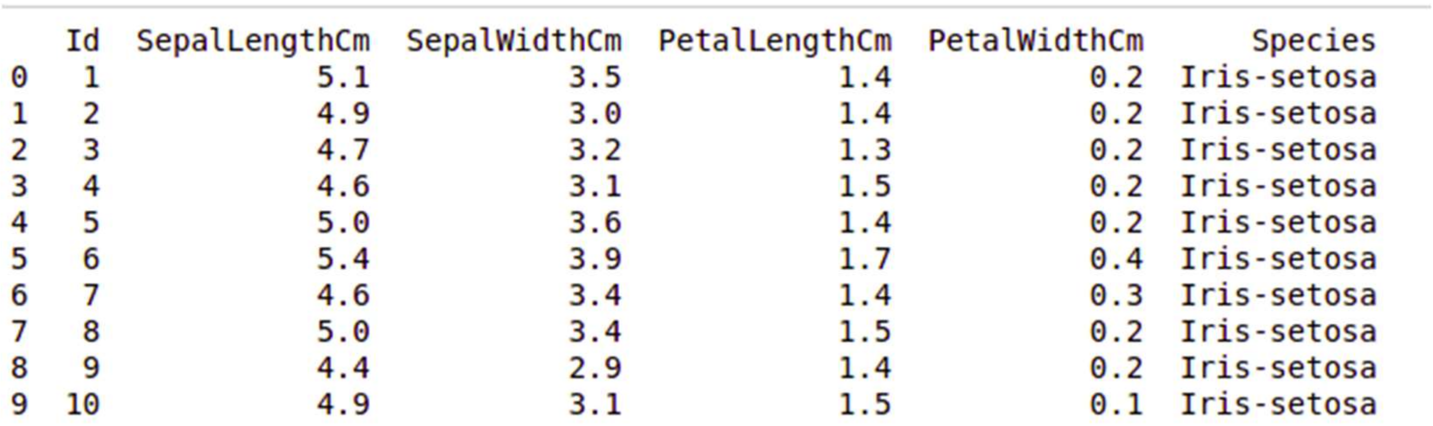
Subset of machine learning in which multilayered neural networks learn from vast amounts of data



Supervised Learning

- For each observation of the predictor measurement(s) xi, i = 1, . . . , n there is an associated response measurement yi.
- We wish to fit a model that relates the response to the predictors, with two aims:
  - Accurately predicting the response for future observations (prediction)
  - Better understanding the relationship between the response and the predictors (inference).
- We could use the supervised learning algorithms to the following scenarios:
  - Anticipate which credit card transactions could be fraudulent
  - Which insurance customer is likely to file a claim
  - Predict the price of the house based on different features

Sample Data for Supervised Learning

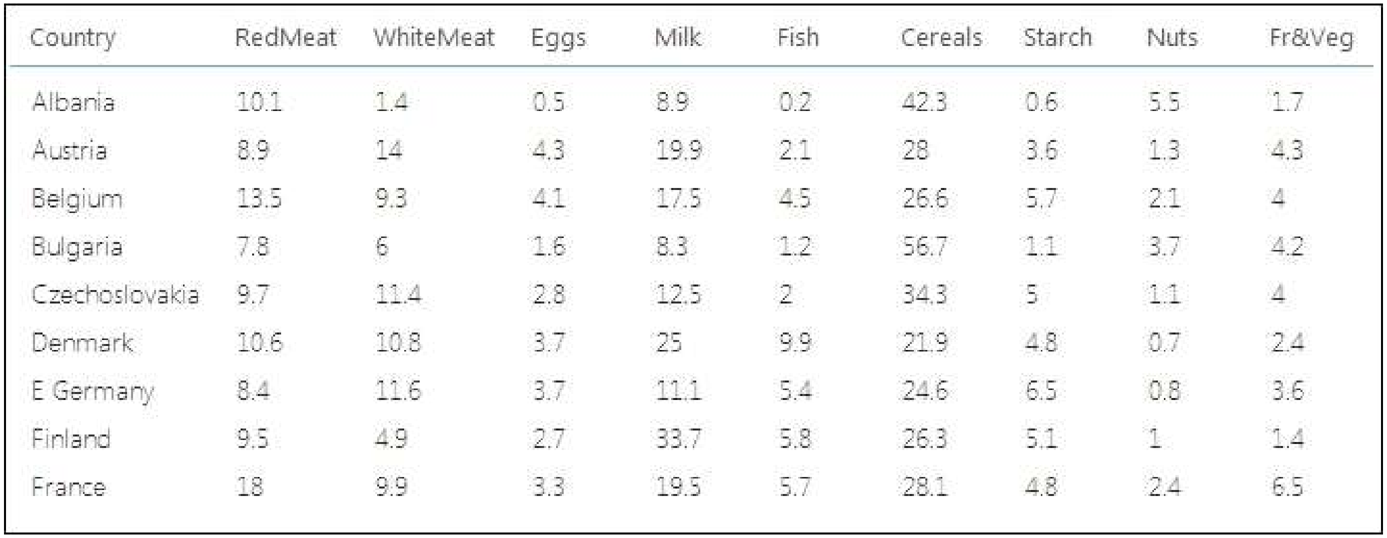


```
   Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm      Species
0   1            5.1           3.5            1.4           0.2  Iris-setosa
1   2            4.9           3.0            1.4           0.2  Iris-setosa
2   3            4.7           3.2            1.3           0.2  Iris-setosa
3   4            4.6           3.1            1.5           0.2  Iris-setosa
4   5            5.0           3.6            1.4           0.2  Iris-setosa
5   6            5.4           3.9            1.7           0.4  Iris-setosa
6   7            4.6           3.4            1.4           0.3  Iris-setosa
7   8            5.0           3.4            1.5           0.2  Iris-setosa
8   9            4.4           2.9            1.4           0.2  Iris-setosa
9  10            4.9           3.1            1.5           0.1  Iris-setosa
```

Unsupervised Learning

- Unsupervised learning describes the somewhat more challenging situation in which for every observation i = 1, . . . , n, we observe a vector of measurements xi but no associated response yi.
- It is not possible to fit a linear regression model, since there is no response variable to predict.
- In this setting, we are in some sense working blind; the situation is referred to as unsupervised because we lack a response variable that can supervise our analysis.
- The goal is to find patterns within the data
- Can be used in scenarios such as:
  - Text segmentation
  - Recommending items

Sample Data for Un-supervised Learning

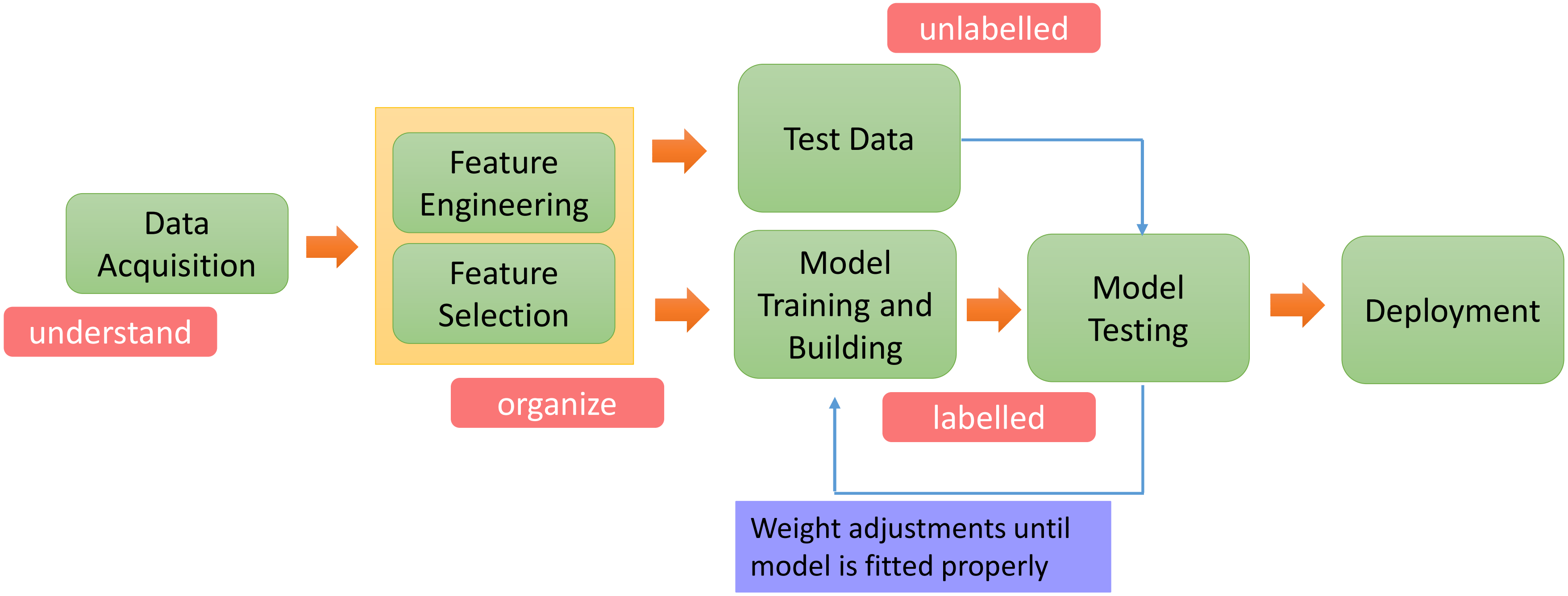


Country	RedMeat	WhiteMeat	Eggs	Milk	Fish	Cereals	Starch	Nuts	Fr&Veg
Albania	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7
Austria	8.9	14	4.3	19.9	2.1	28	3.6	1.3	4.3
Belgium	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4
Bulgaria	7.8	6	1.6	8.3	1.2	56.7	1.1	3.7	4.2
Czechoslovakia	9.7	11.4	2.8	12.5	2	34.3	5	1.1	4
Denmark	10.6	10.8	3.7	25	9.9	21.9	4.8	0.7	2.4
E Germany	8.4	11.6	3.7	11.1	5.4	24.6	6.5	0.8	3.6
Finland	9.5	4.9	2.7	33.7	5.8	26.3	5.1	1	1.4
France	18	9.9	3.3	19.5	5.7	28.1	4.8	2.4	6.5

Reinforcement Learning

- Often used for robotics, gaming and navigation
- With reinforcement learning, the algorithm discovers through trial and error which actions yield the greatest rewards.
- Has three primary components:
  - Agent (the learner or decision maker)
  - Environment (everything the agent interacts with)
  - Actions (what agent can do)
- The objective is for the agent to choose actions that yield expected reward over a given amount of time
- The agent will reach the goal using a good policy, there for a goal is to figure out a good policy

Machine Learning Process


unlabelled
Test Data
Feature Engineering
Feature Selection
Data Acquisition
understand
organize
Model Training and Building
Model Testing
Deployment
labelled
Weight adjustments until model is fitted properly