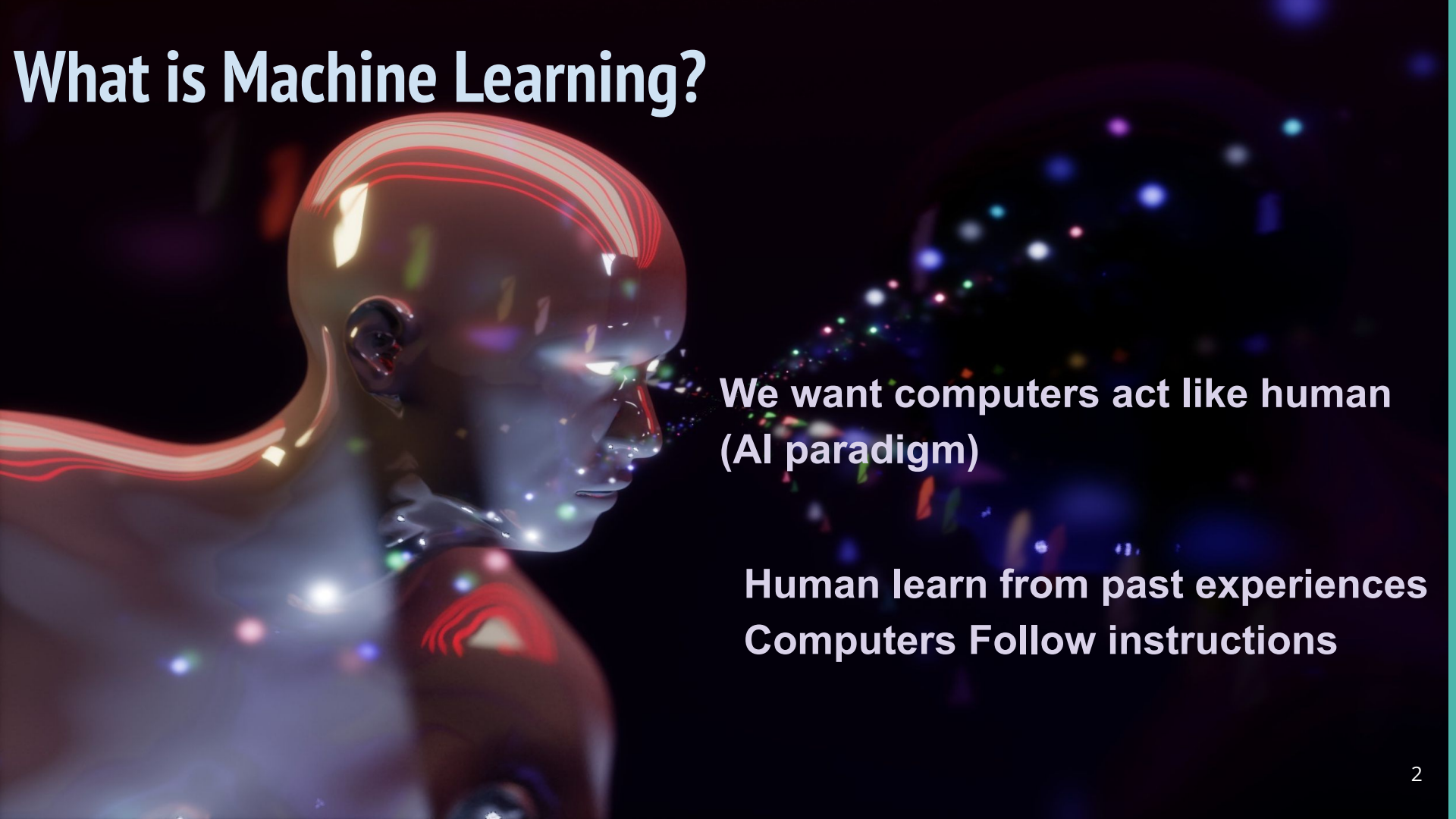
Introduction to Machine Learning and Python Programming

Farnoosh Khodakarami

This material is made by

Farnoosh Khodakarami and Ali Madani

What is Machine Learning?



We want computers act like human (AI paradigm)

Human learn from past experiences
Computers Follow instructions

https://www.ibm.com/ibm/history/ibm100/us/en/icons/ibm700series/impacts/

Field of study that gives computers the ability to learn without being explicitly programmed

(Arthur Samuel 1951)

The Samuel Checkers-playing Program was among the world's first successful self-learning programs.

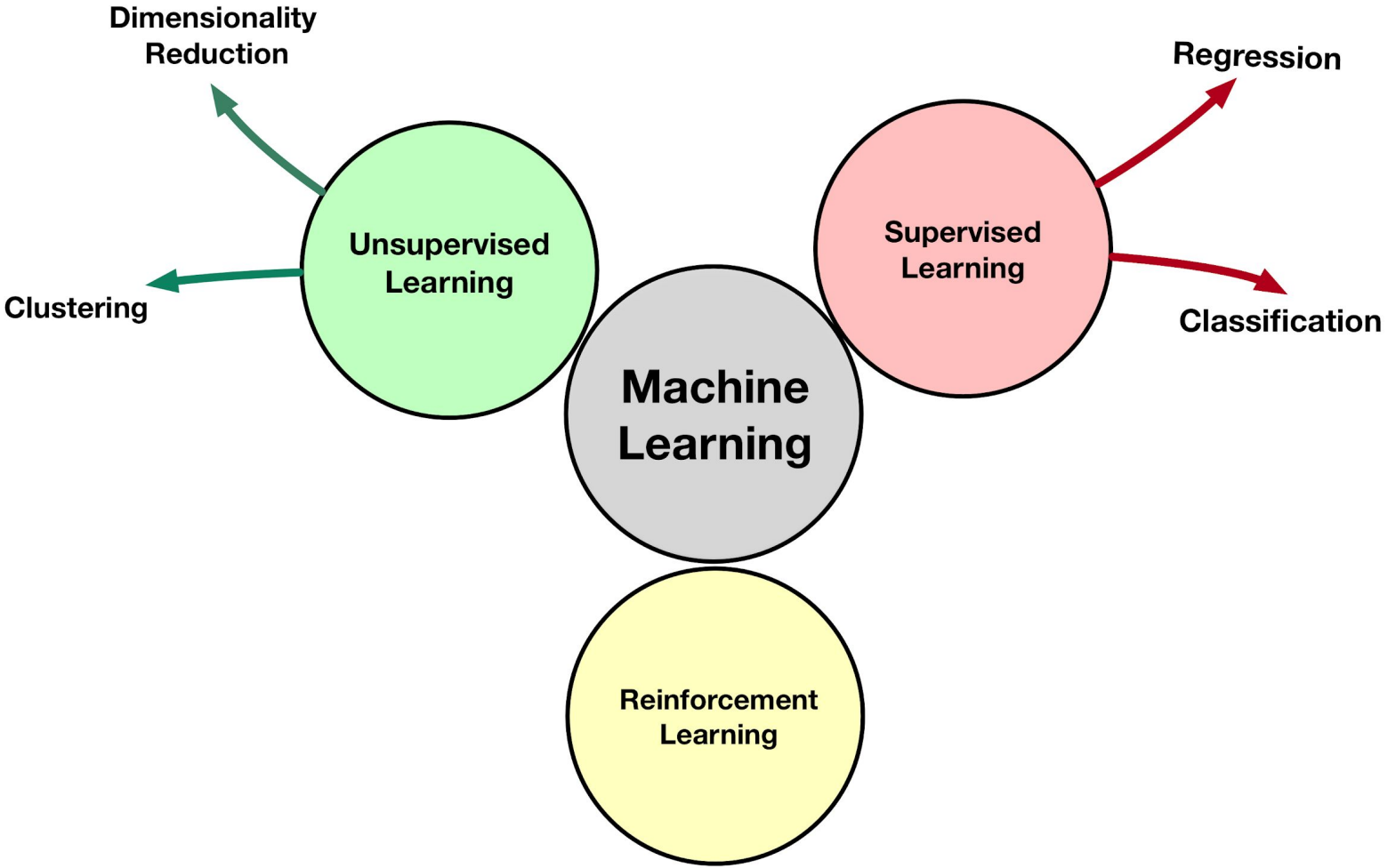
He coined the term "machine learning" in 1959

Identifying relationship (learning) between
Features & output : supervised learning
Features themselves: dimensionality reduction and feature engineering
Datapoints: clustering

Relationship identification with delayed rewarding: reinforcement learning

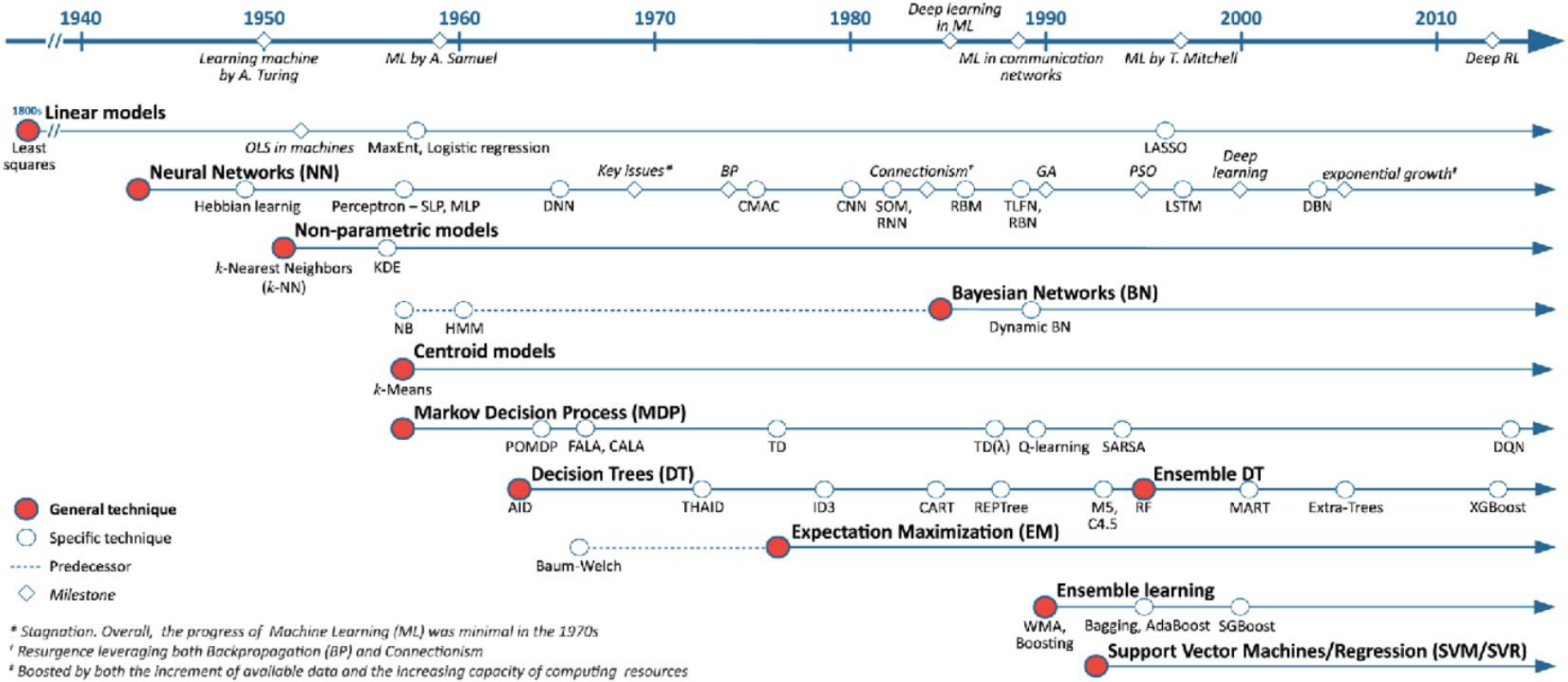
Different variable types

- Continuous
- Discrete
  - Nominal
  - Ordinal
- ...


Dimensionality Reduction
Unsupervised Learning
Clustering
Regression
Supervised Learning
Classification
Machine Learning
Reinforcement Learning

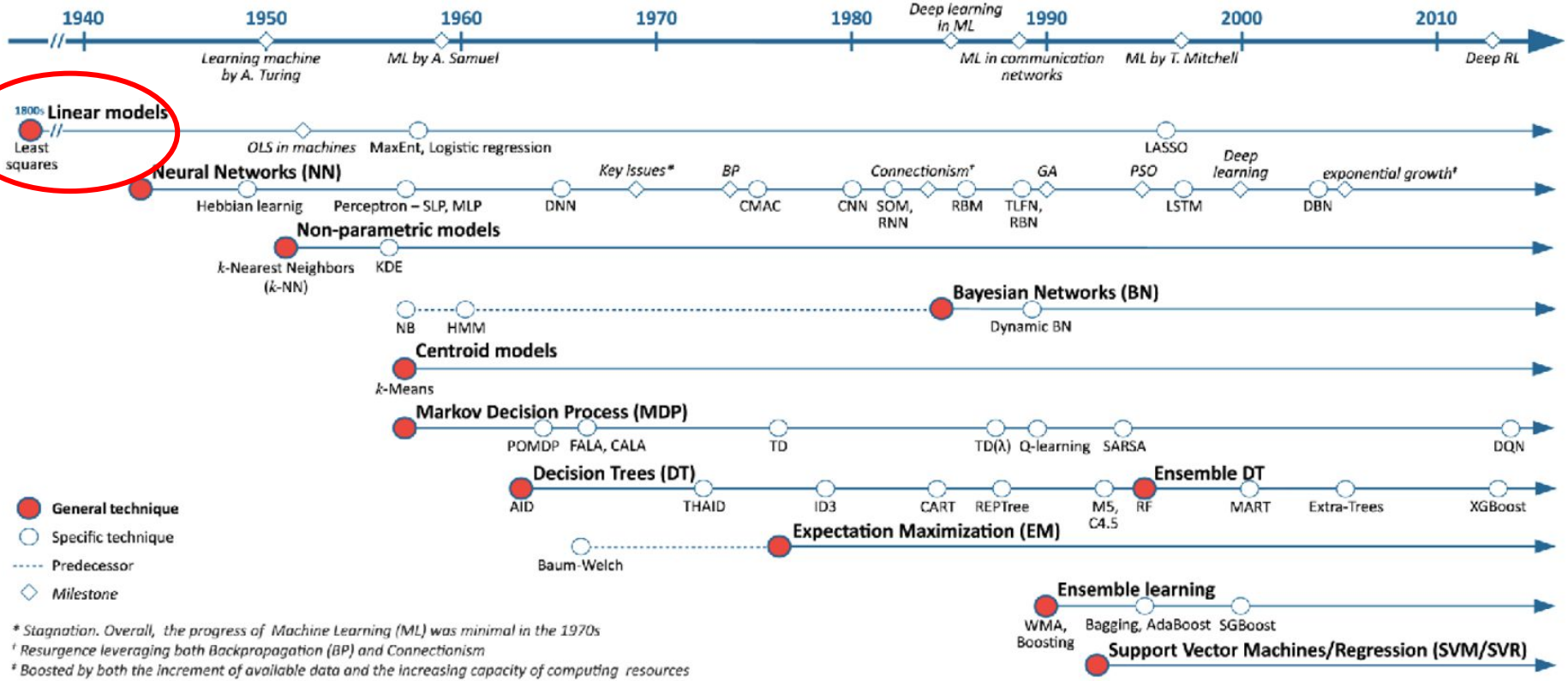
Evolution of machine learning

Evolution of machine learning



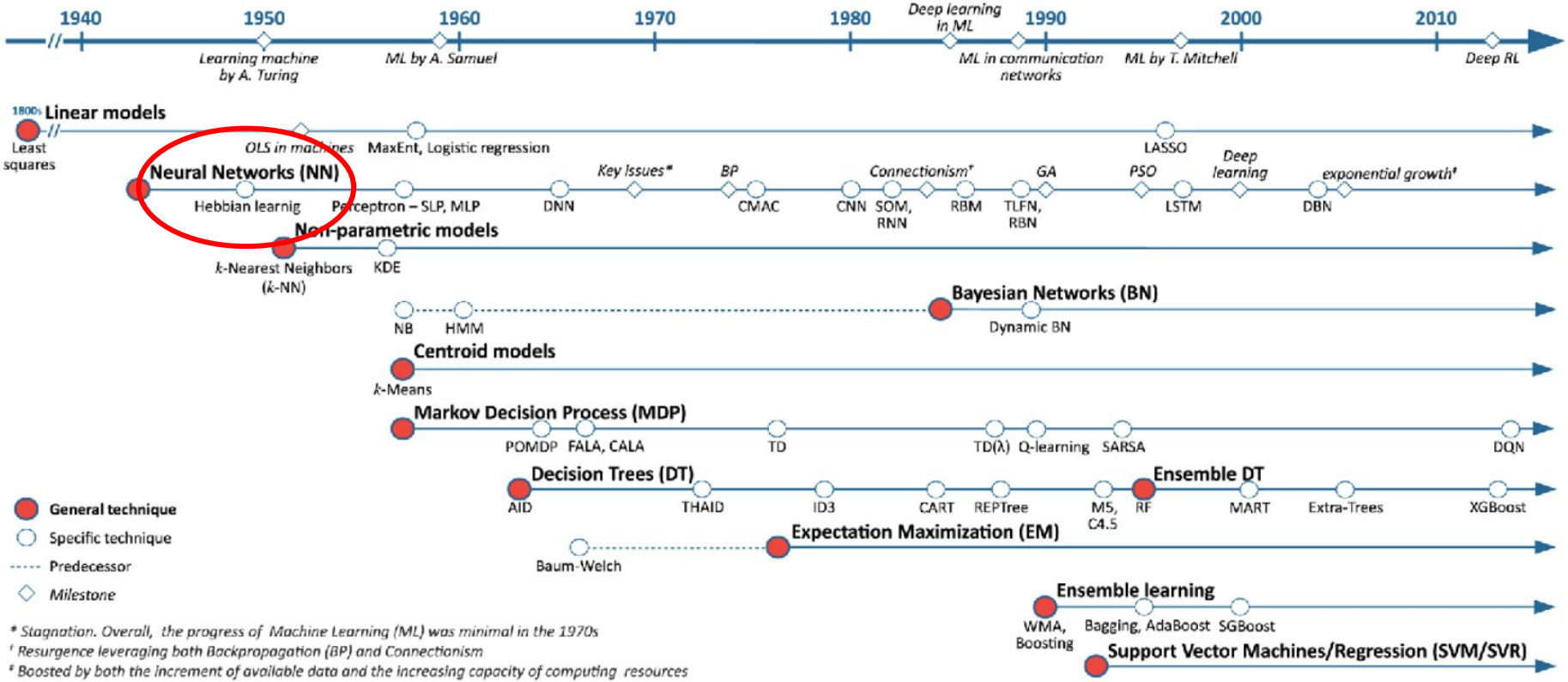
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



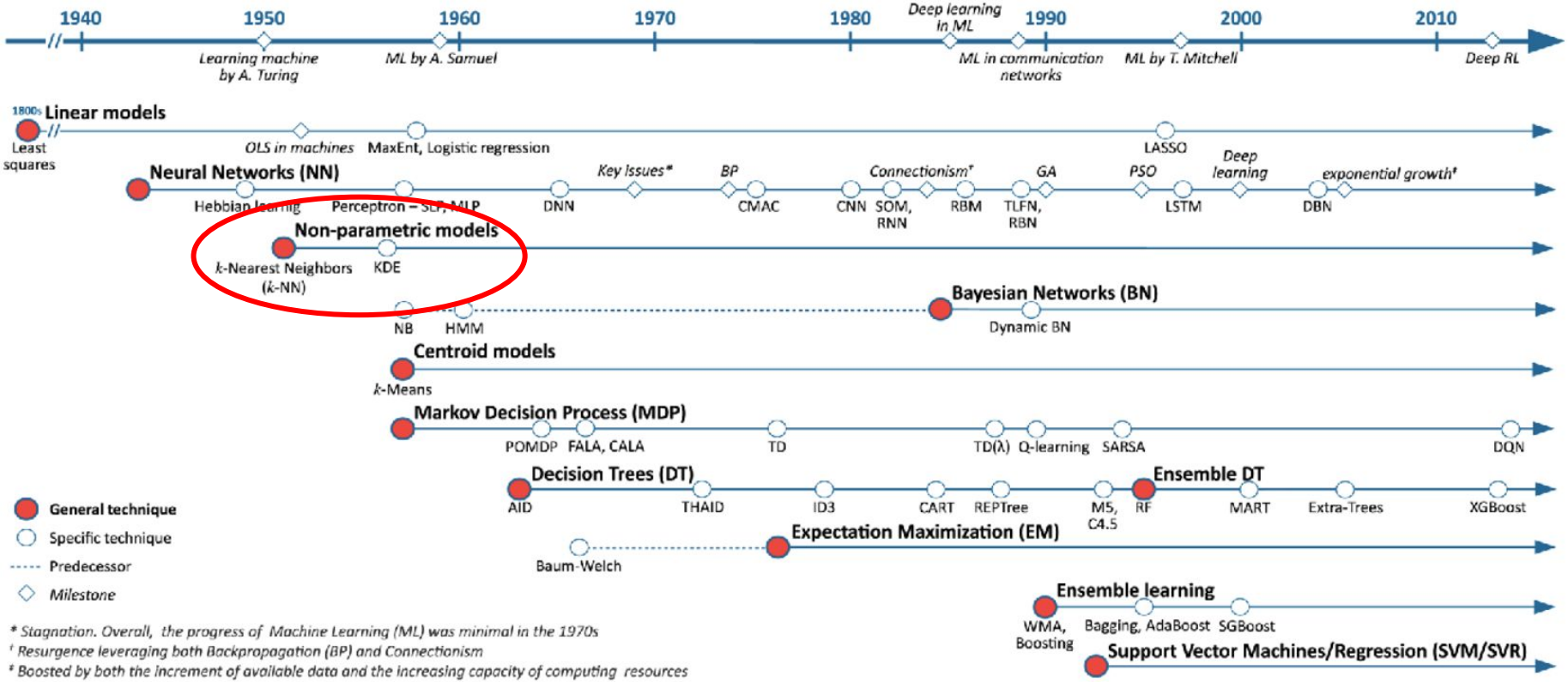
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



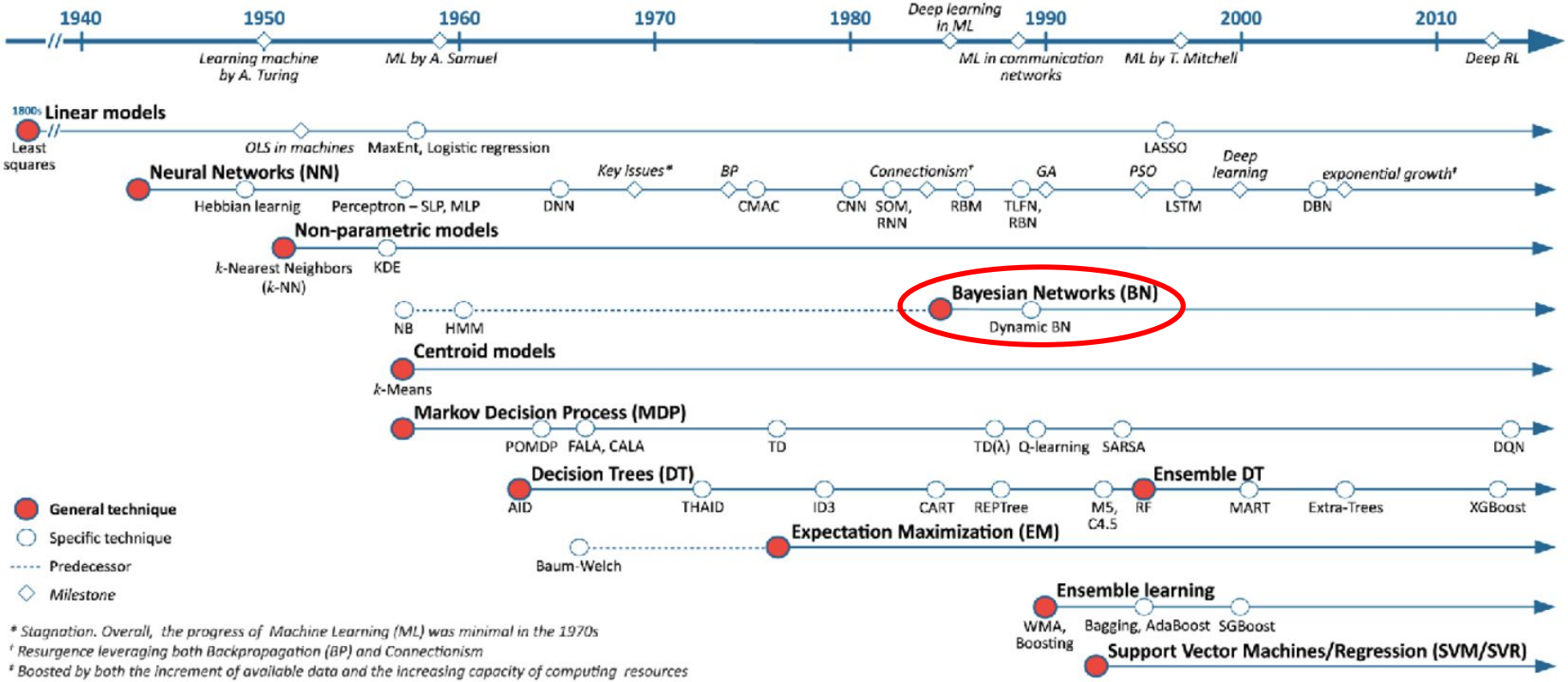
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



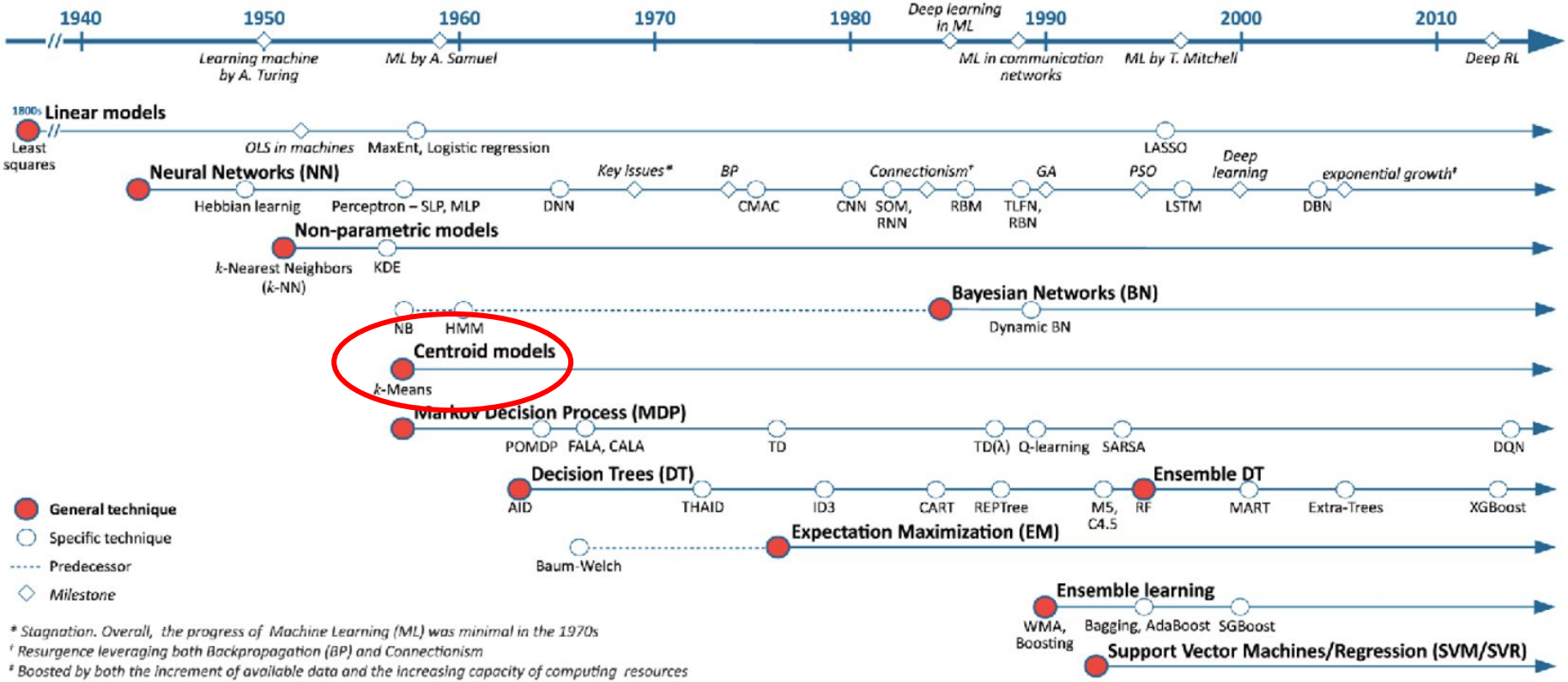
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



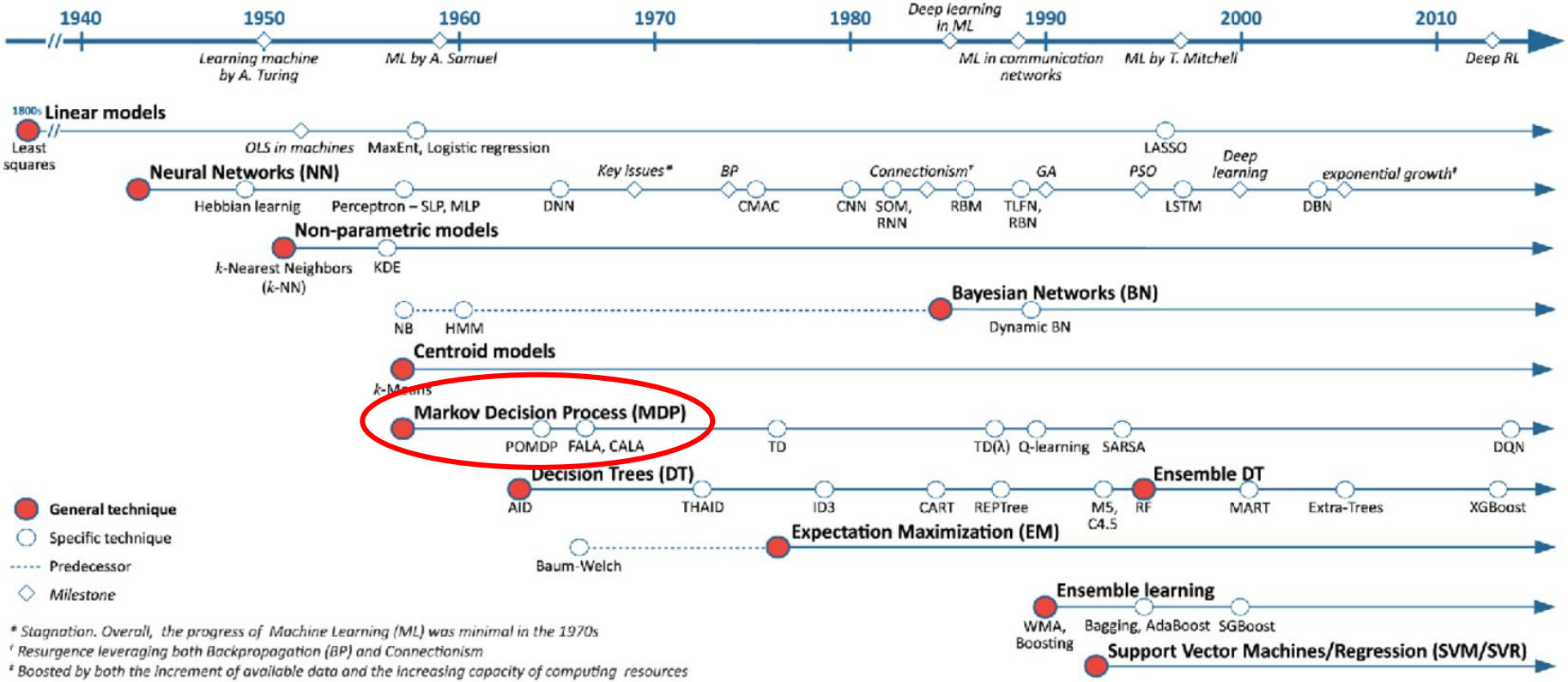
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



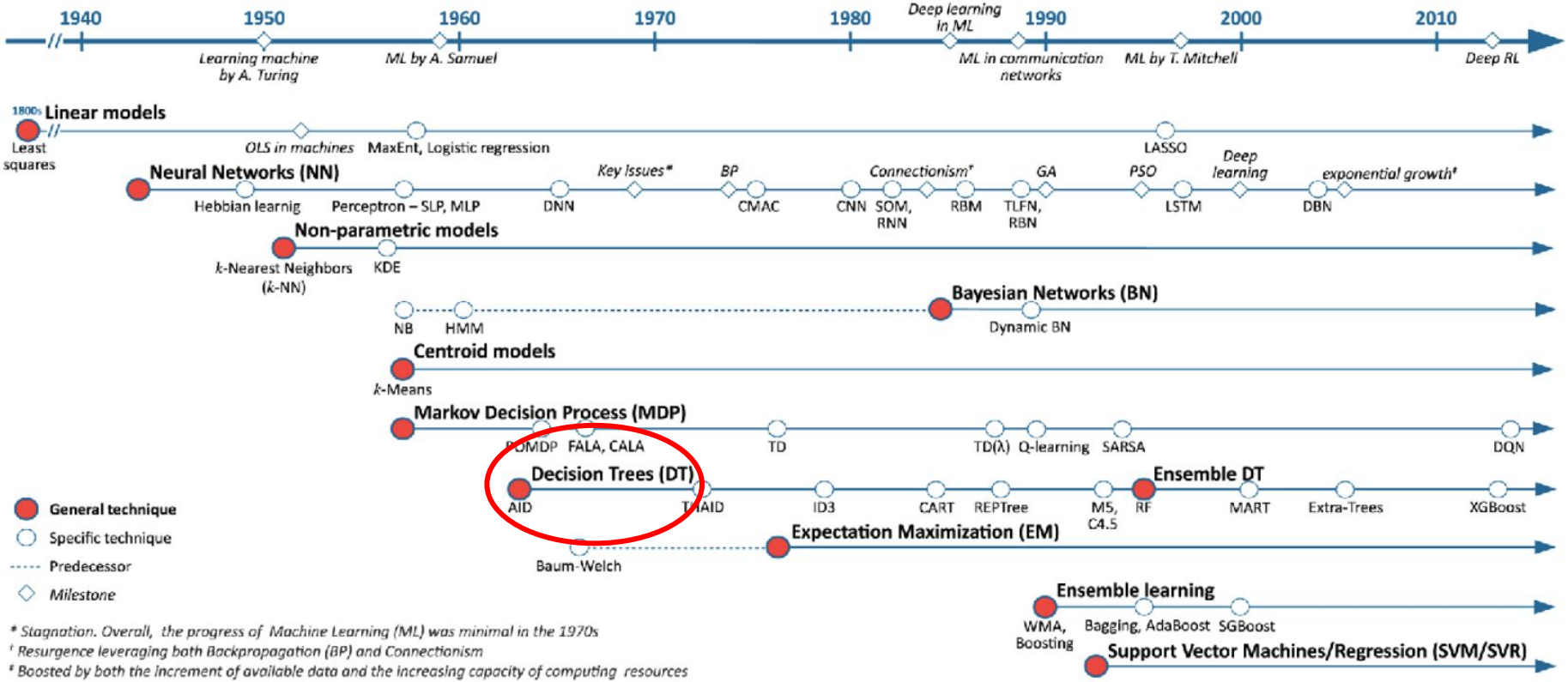
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



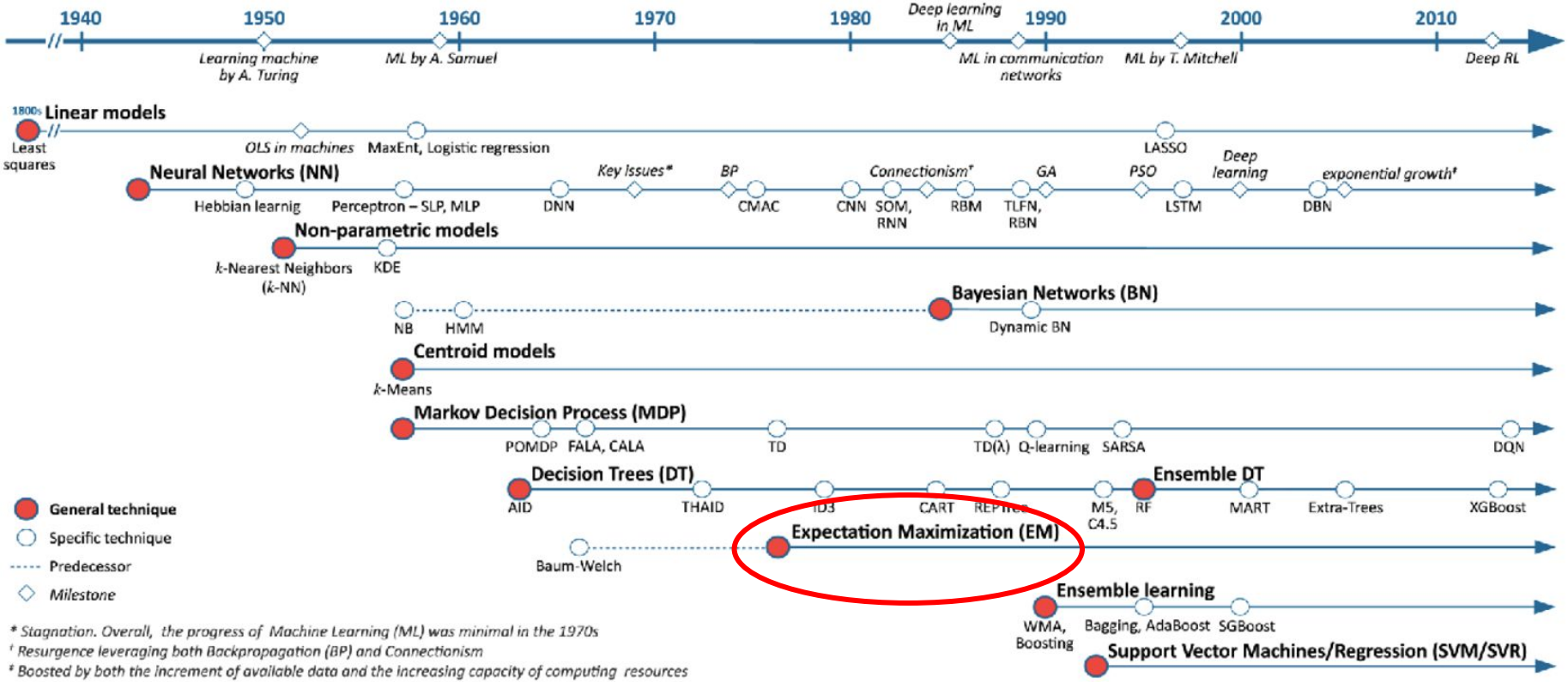
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



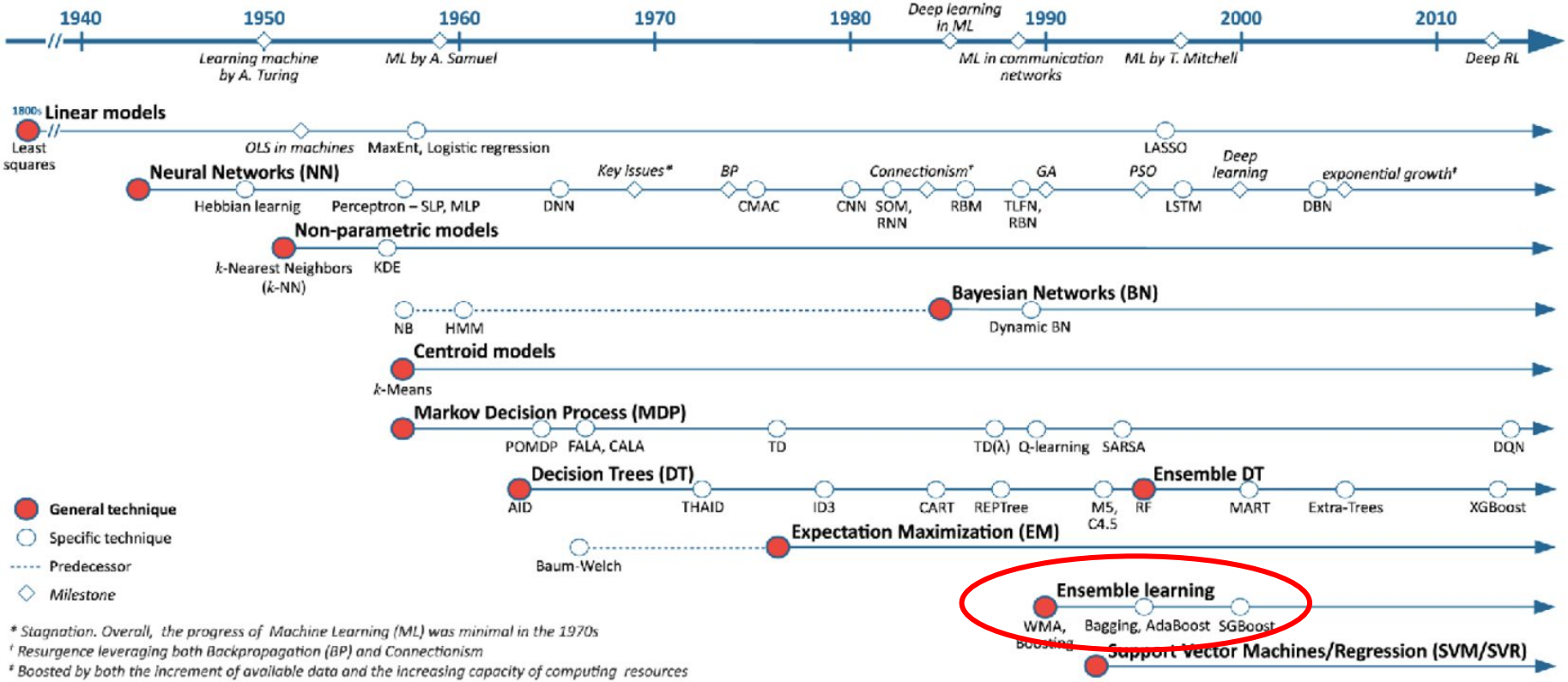
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



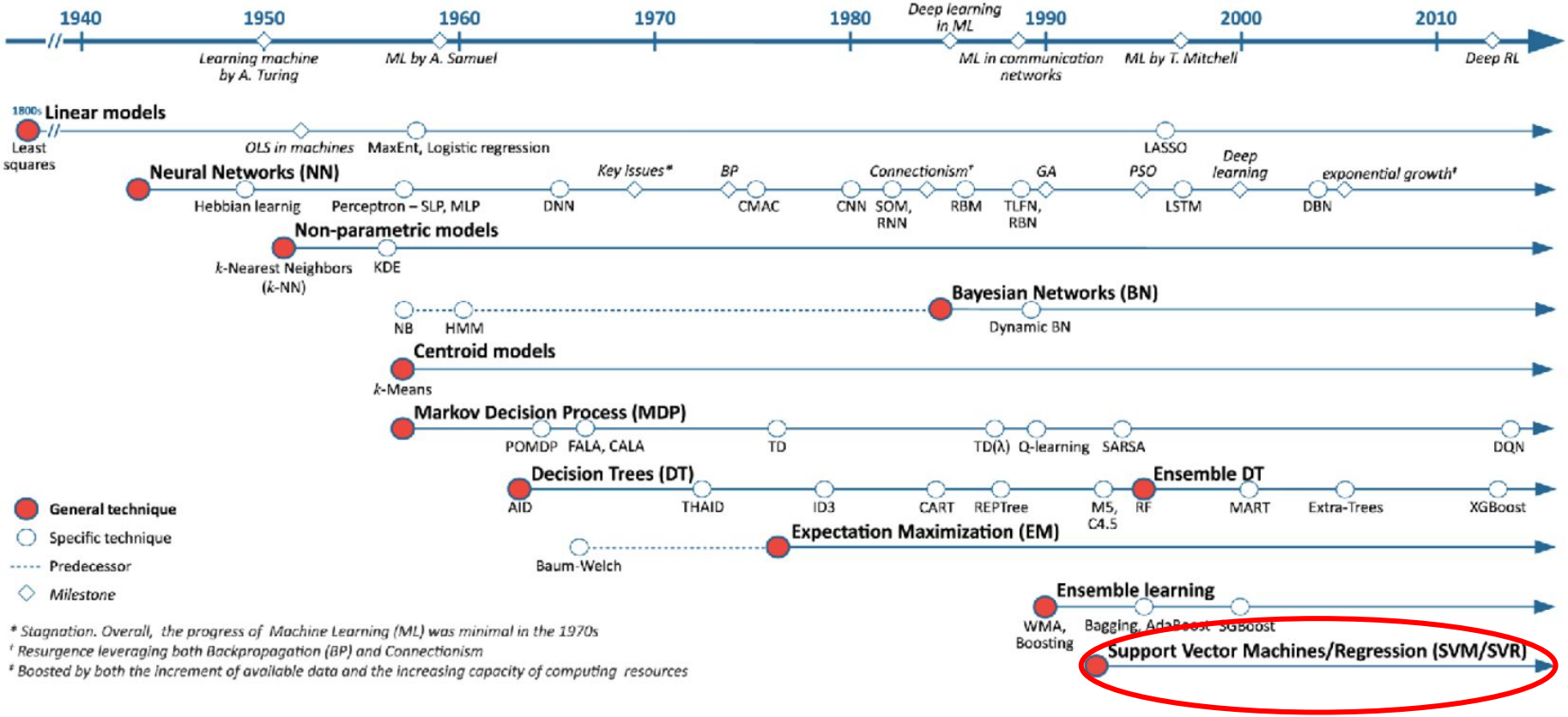
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

Evolution of machine learning



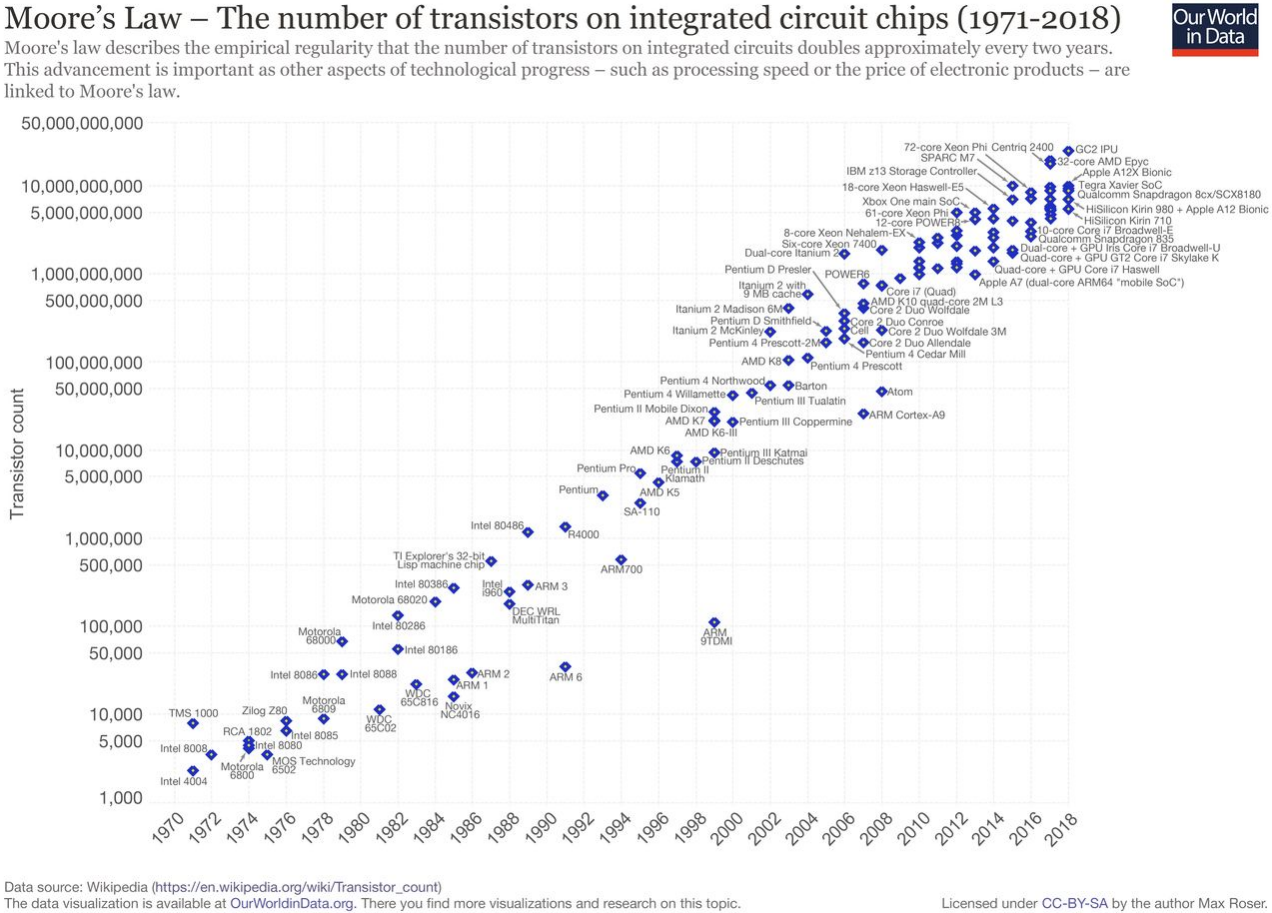
Boutaba, Raouf, et al. *Journal of Internet Services and Applications* 9.1 (2018): 16.

We are lucky

Because of the following reasons, learning how to implement ML and use it in industrial setting became much easier

- Programming became easier
- Computational power
- Data availability

Computational power and Moore's law



Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

We have huge responsibility

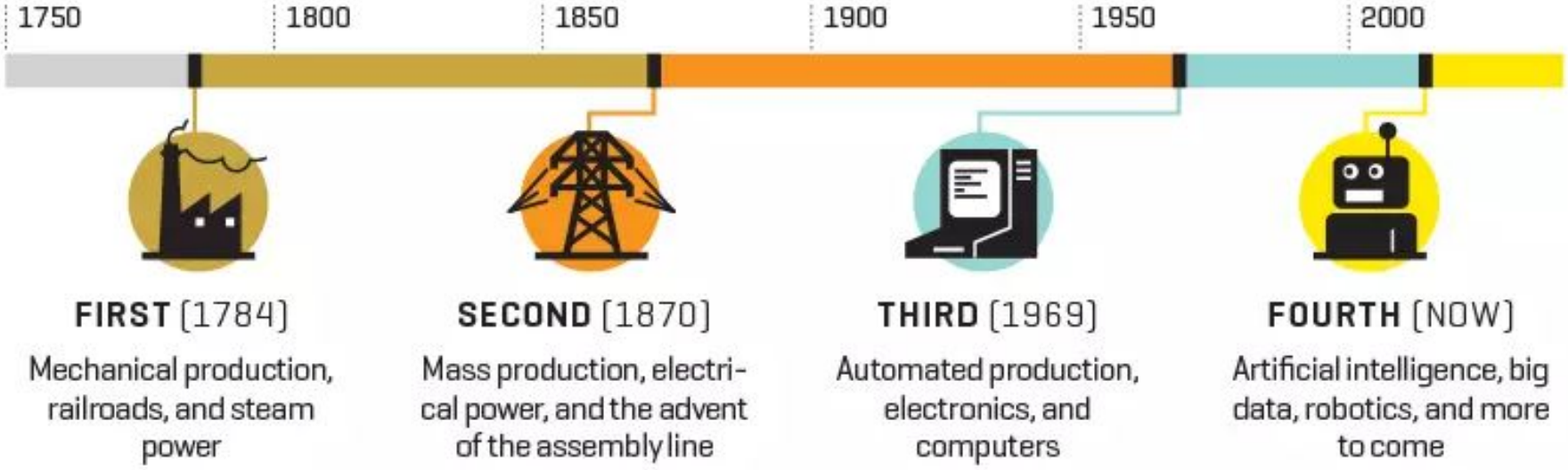
- Do not overpromise
  - It is not magic if
    - we are not magicians
    - the proper ingredient (data) is not available
    - the goal is properly defined

We have huge responsibility

- Do not overpromise
  - It is not magic if
    - we are not magicians
    - the proper ingredient (data) is not available
    - the goal is properly defined
- Proper implementation of ML models
- Proper interpretation

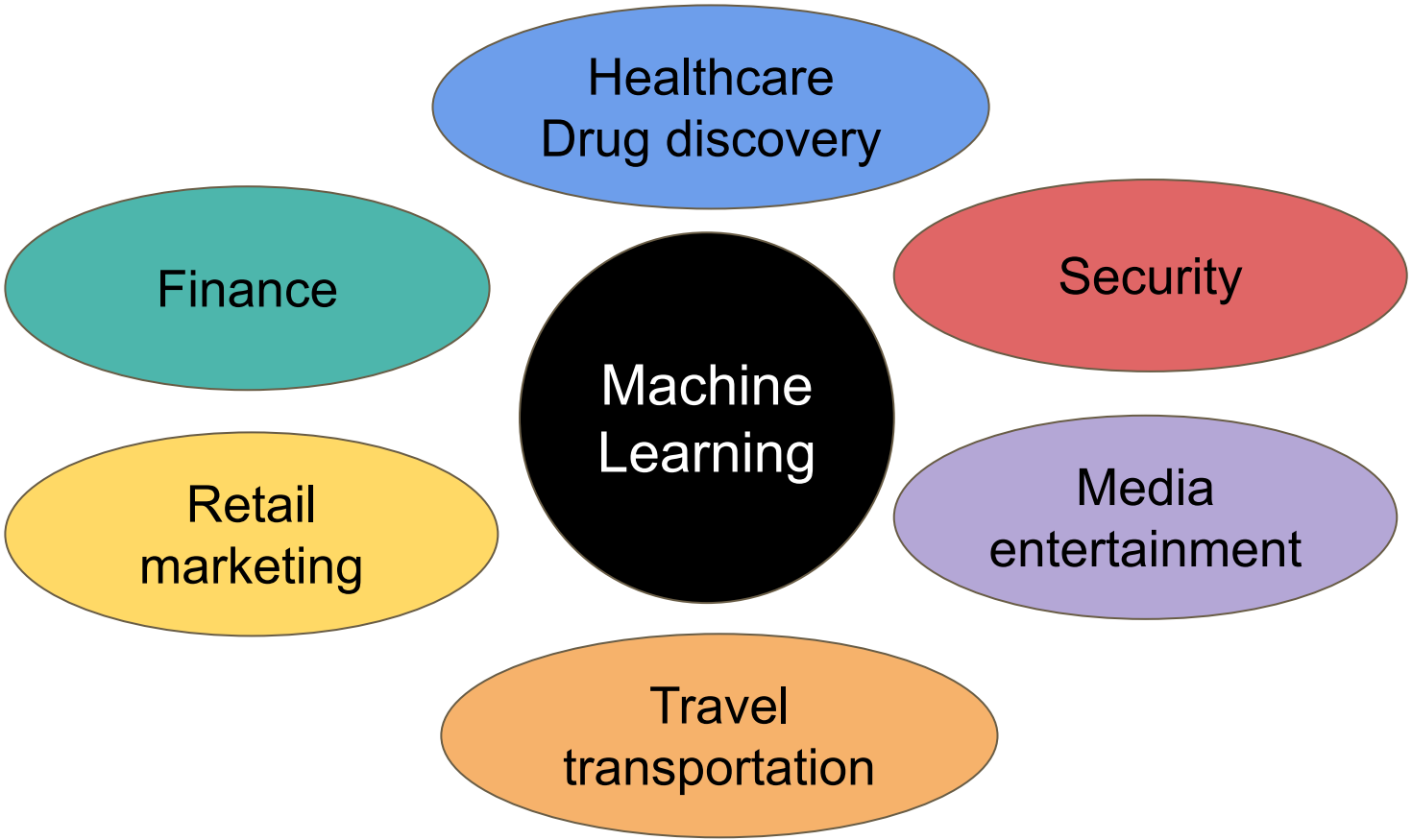
Machine learning application

Four industrial revolutions



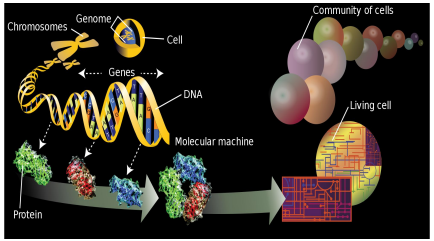
https://www.skynettoday.com/editorials/ai-automation-job-loss

ML had big impact in some of the industries



Example of ML in healthcare and drug discovery

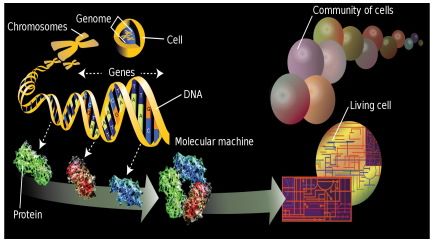
Example of ML in healthcare and drug discovery



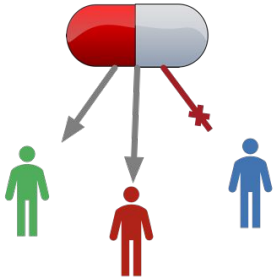
Drug Discovery and Manufacturing

- New drugs (chemical compounds)
- Identifying new targets
- Optimizing manufacturing pipeline

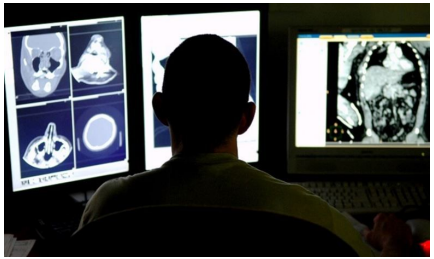
Example of ML in healthcare and drug discovery



Drug Discovery and Manufacturing



Personalized Medicine

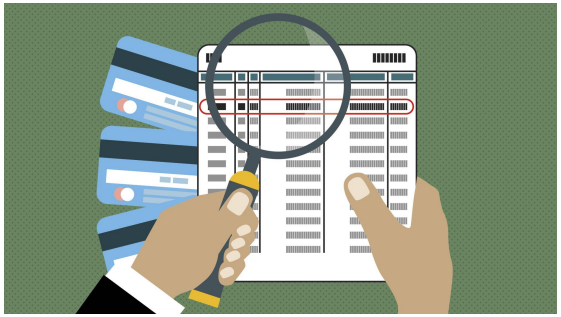


Medical Imaging (Diagnosis; Prognosis)

Example of ML in finance

Money-Laundering Prevention

Loan/ Insurance Underwriting



Fraud Detection

Example of ML in finance

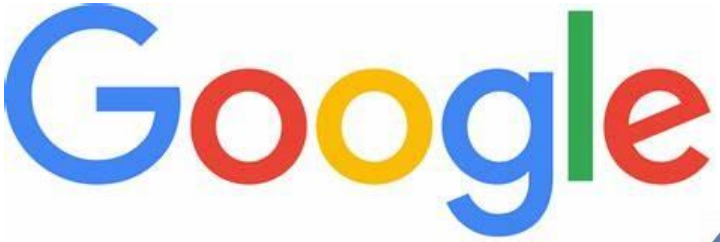
Example of ML in retail and marketing

Make an accurate forecast for many products

Detect fraudulent activities

Offer customer-specific product recommendations

Example of ML in retail and marketing



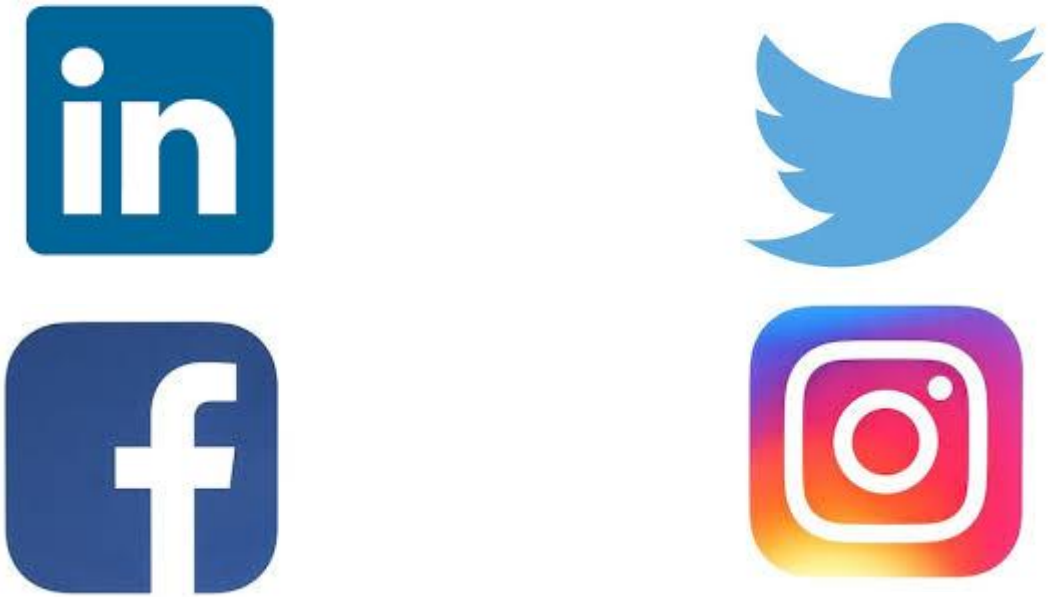
Example of ML in media and entertainment

Recommendation engines

Personalized targeting

Search Optimization

Example of ML in media and entertainment



Example of ML in media and entertainment

Example of ML in transportation

Machine learning Project steps:

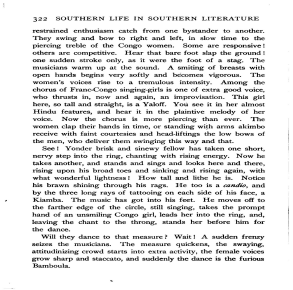
1. Problem Definition.
2. Data Preparation.

    Gathering Data

    Cleaning Data

3. Model Building
4. Algorithms Evaluation.
5. Results Improvement.
6. Results Presentation.

Feature Representation



322 SOUTHERN LIFE IN SOUTHERN LITERATURE

restrained enthusiasm catch from one bystander to another. They swing and bow to right and left, in slow time to the piercing treble of the Congo women. Some are responsive; others are competitive. Hear that bare foot slap the ground! one sudden stroke only, as it were the foot of a stag. The musicians warm up at the sound. A smiting of breasts with open hands begins very softly and becomes vigorous. The women's voices rise to a tremulous intensity. Among the chorus of Franc-Congo singing-girls is one of extra good voice, who thrusts in, now and again, an improvisation. This girl here, so tall and straight, is a Yakoff. You see it in her almost Hindu features, and hear it in the plaintive melody of her voice. Now the chorus is more piercing than ever. The women clap their hands in time, or standing with arms akimbo receive with faint courtesies and head-liftings the low bows of the men, who deliver them swinging this way and that.

See! Yonder brisk and sinewy fellow has taken one short, nervy step into the ring, chanting with rising energy. Now he takes another, and stands and sings and looks here and there, rising upon his broad toes and sinking and rising again, with what wonderful lightness! How tall and lithe he is. Notice his brawn shining through his rags. He too is a *ganako*, and by the three long rays of tattooing on each side of his face, a Kiamba. The music has got into his feet. He moves off to the farther edge of the circle, still singing, takes the prompt hand of an unsmiling Congo girl, leads her into the ring, and, leaving the chant to the throng, stands her before him for the dance.

Will they dance to that measure? Wait! A sudden frenzy seizes the musicians. The measure quickens, the swaying, attitudinizing crowd starts into extra activity, the female voices grow sharp and staccato, and suddenly the dance is the furious Bamboula.

Shape	Size	Color	Texture	Weight

A matrix of color values(pixels)

List of words with their frequency count

DataSet

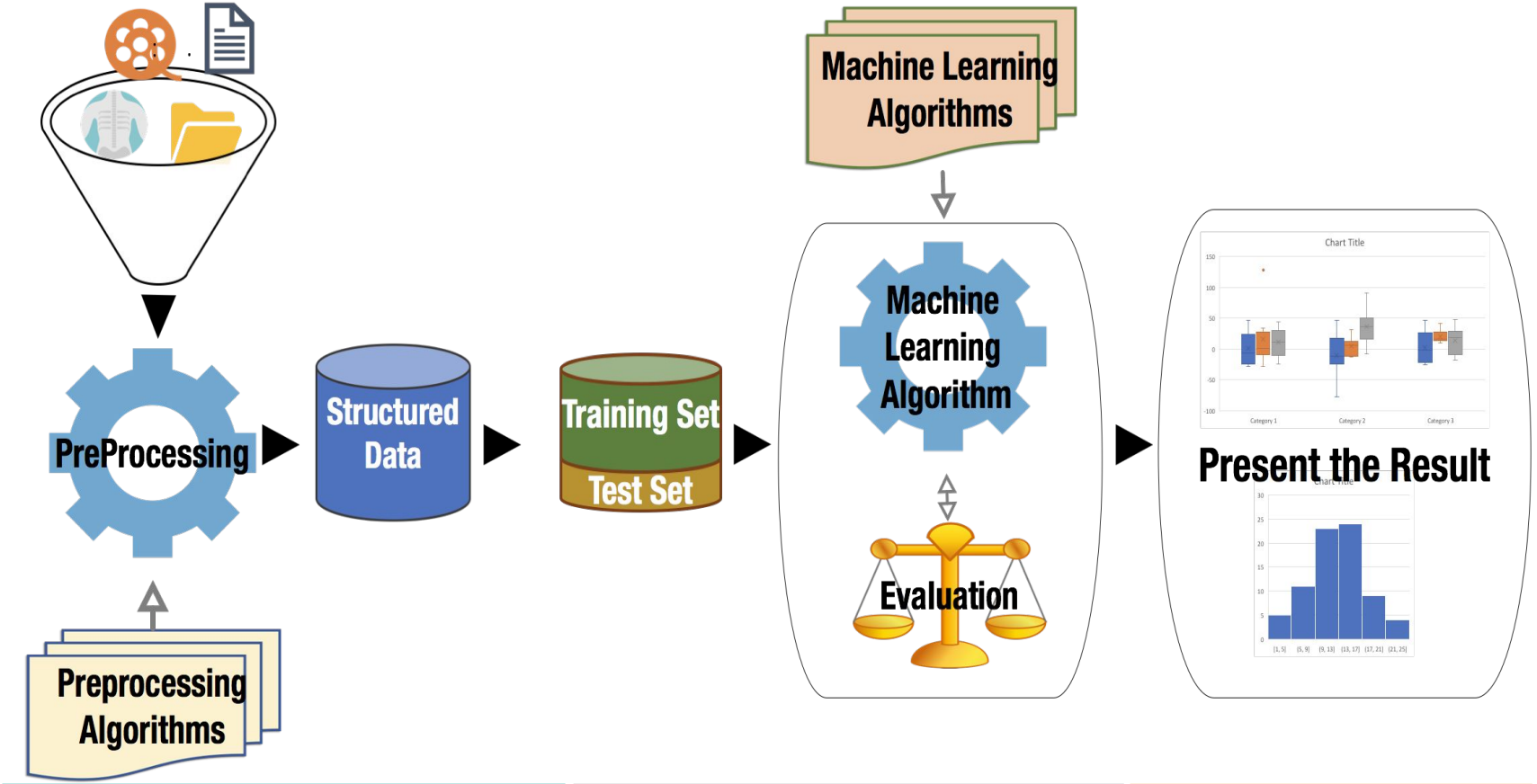
Features(Variable/Attribute)

Data Records (samples)

Size	Location	...	#Rooms	Output

A **data set** (or **dataset**) is a collection of data
Every column of the table represents a particular variable,
Each row corresponds to a given member of the **data set**

The Machine Learning Processes



Gathering and Cleaning data

Model building and Model evaluation

Deploy selected Model and present result

Machine learning in practice

environment setup

Python

Read, understand and prepare data for ML

-Reading: pandas

-Understanding: matplotlib, numpy, etc.

-Preparing: pandas, sklearn

Build ML models and assessing their performances

-sklearn

Preparing the results and saving the outputs

-pandas, matplotlib

Introduction to Python programming

Why Python?

open sourced

easy to use

multi-paradigm

Extensive Support Libraries

Python vs R?


0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Percentage of Matching Job Postings (%)
2012
2013
2014
2015
2016

Environment: IDLE, Shell

- IDLE ,**Integrated DeveLopment Environment**, is the standard Python development environment.
- It works well on both Unix and Windows platforms.
- It has a Python shell window, which gives you access to the Python **interactive mode**.
- It also has a file editor that lets you create and edit existing Python source files.

Environment: Jupyter Notebook

An open-source web application that allows you to create and share documents that contain:

- ❏ live code
- ❏ Equations
- ❏ visualizations
- ❏ narrative text.

Algebraic operations

- Rank, determinant, trace, etc. of an array.
- Eigen values of matrices
- Matrix and vector products (dot, inner, outer,etc. product), matrix exponentiation
- Solve linear or tensor equations and much more!

Working with numbers, tuples and strings

There are three **numeric** types in Python:
Int, float, complex

String literals in python are surrounded by either single quotation marks, or double quotation marks.

(link to the example)

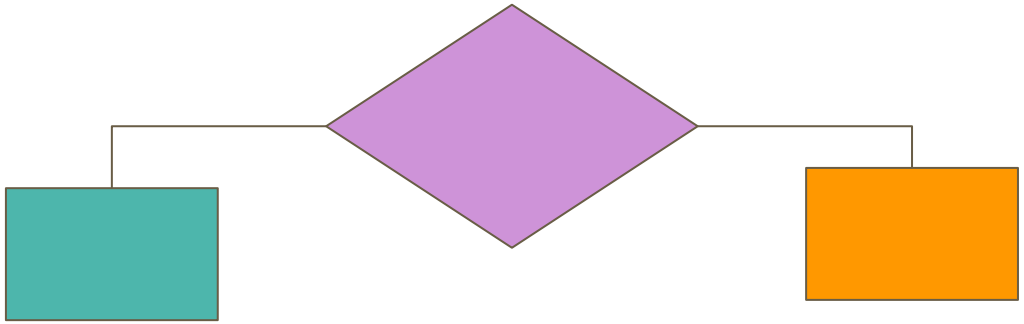
Set, lists, dictionaries in Python

- **List:** ordered and changeable. Allows duplicate members.
- **Tuple:** ordered and unchangeable. Allows duplicate members.
- **Set:** unordered and unindexed. No duplicate members.
- **Dictionary:** unordered, changeable and indexed. No duplicate members.

Conditioning in Python

Python uses boolean variables to evaluate conditions.

The boolean values True and False are returned when an expression is compared or evaluated.



Loops in Python

Python provides three ways for executing the loops.

While all the ways provide similar basic functionality, they differ in their syntax and condition checking time.

```
for val in sequence:
    Body of for
```

```
while test_expression:
    Body of while
```

Writing function in Python

Function is a group of related statements that perform a specific task.

Functions help break our program into smaller and modular chunks. As our program grows larger and larger, functions make it more organized, manageable and readable.

Function avoids repetition and makes code reusable.

```
def function_name(parameters):
    """docstring"""
    statement(s)
```

Fundamental packages in Python: numpy

core library for scientific computing

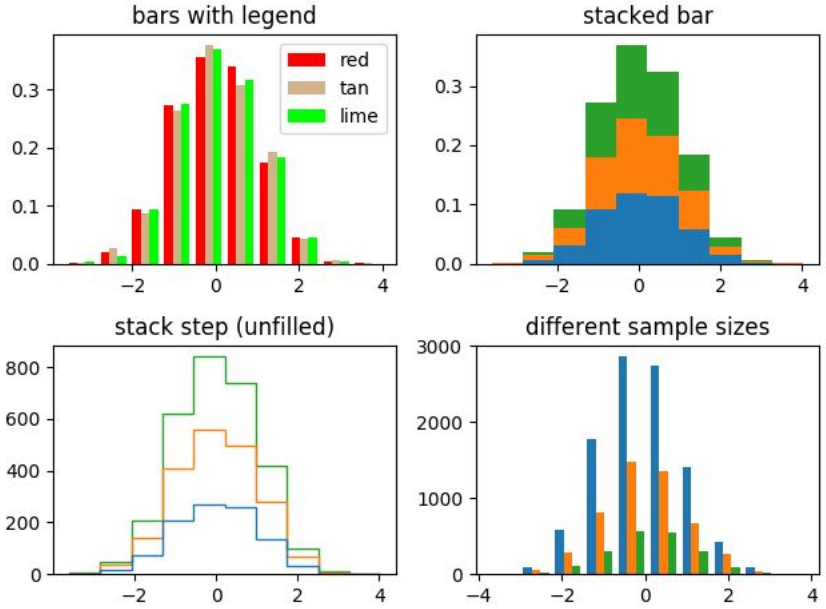
High-performance multidimensional array object

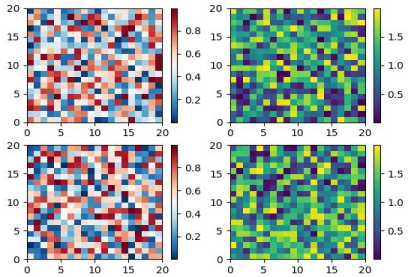
Operations related to linear algebra.

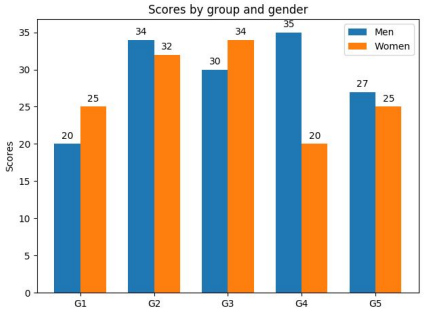
Fundamental packages in Python: matplotlib

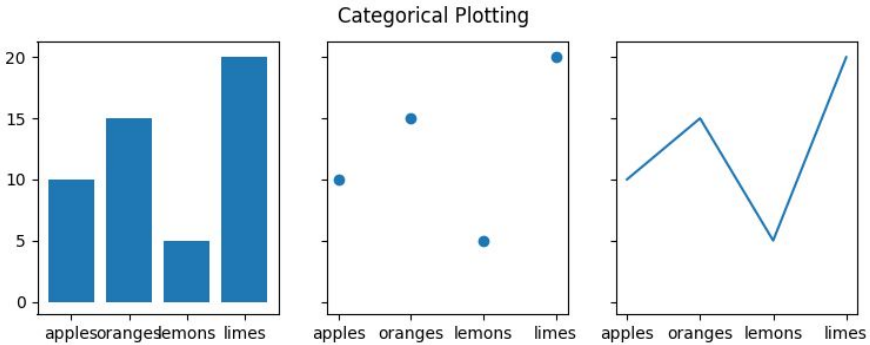
- python library used to create 2D graphs and plots by using python scripts.
- pyplot makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc.
- It supports a very wide variety of graphs and plots namely

Plotting in Python





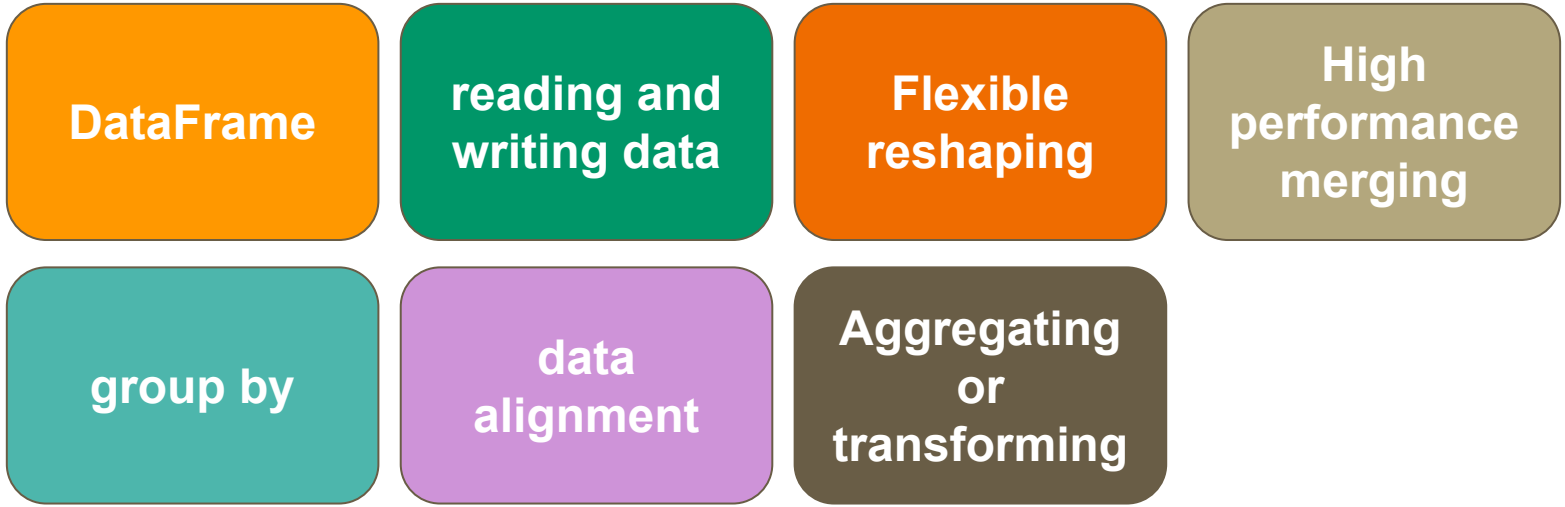




matplotlib
Version 3.1.1

Fundamental packages in Python: pandas

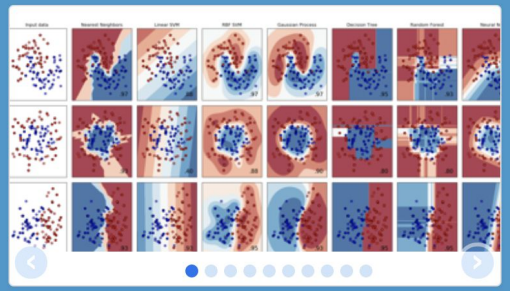
pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.



Working with dataframes in Python

index	col1	col2	col3
1			
2			
3			

Machine learning in Python using sklearn



scikit-learn

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying to which category an object belongs to.

Applications: Spam detection, Image recognition.
Algorithms: SVM, nearest neighbors, random forest, … — Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.
Algorithms: SVR, ridge regression, Lasso, … — Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: k-Means, spectral clustering, mean-shift, … — Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency
Algorithms: PCA, feature selection, non-negative matrix factorization. — Examples

Model selection

Comparing, validating and choosing parameters and models.

Goal: Improved accuracy via parameter tuning
Modules: grid search, cross validation, metrics. — Examples

Preprocessing

Feature extraction and normalization.

Application: Transforming input data such as text for use with machine learning algorithms.
Modules: preprocessing, feature extraction. — Examples