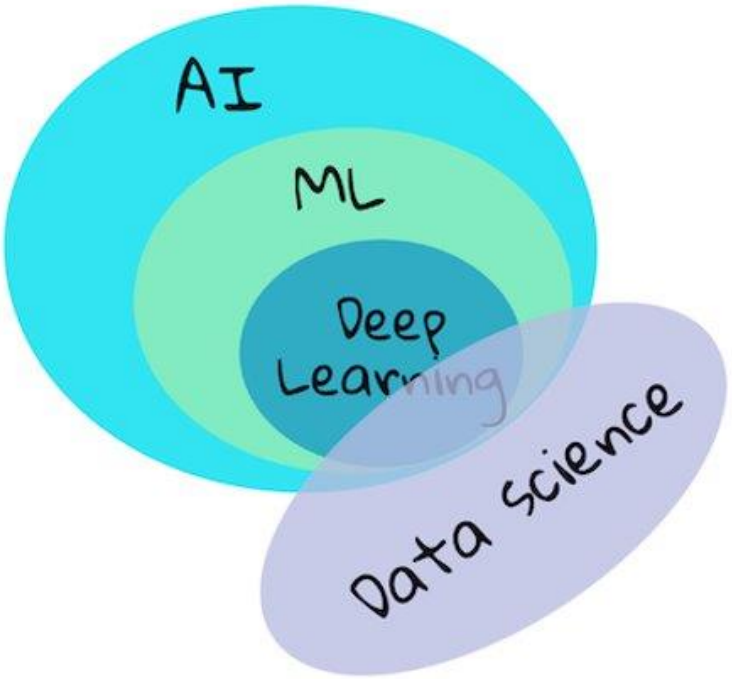
An Introduction to Machine Learning

NASTARAN SHAHPARIAN |SHARCNET | COMPUTE ONTARIO | COMPUTE CANADA |YORK UNIVERSITY

Difference between ML, DL, AI, and Data science

- Artificial Intelligence
  - Any techniques that enables computers to mimic human behaviour
- Machine Learning
  - Ability to learn without explicitly being performed
- Deep learning
  - Extract pattern from data using neural networks



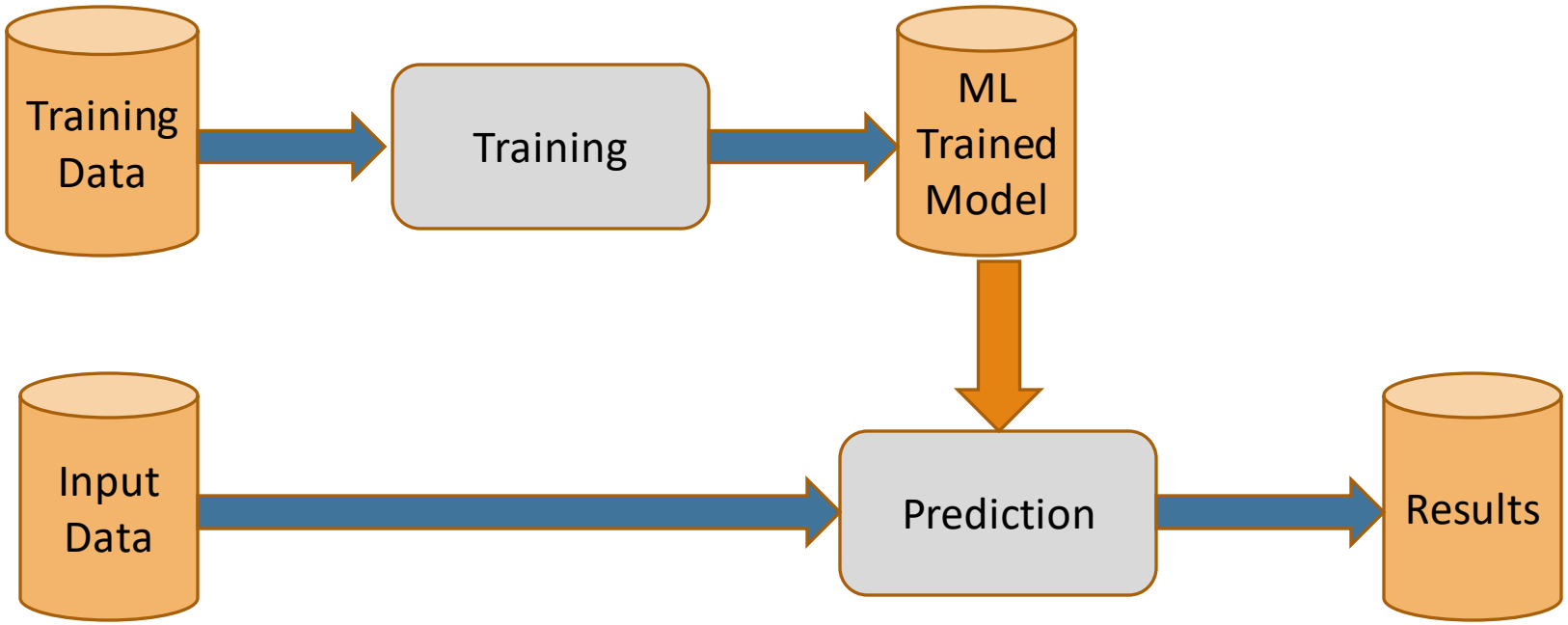
Machine Learning in simple words

- Training a machine learning algorithm on a set of data, allowing it to identify patterns and make predictions or decisions based on that data.
- A type of artificial intelligence that enables machines to learn from experience without being explicitly programmed.
- Has many practical applications, including image and speech recognition, natural language processing, fraud detection, and recommendation systems.

Data is in different forms

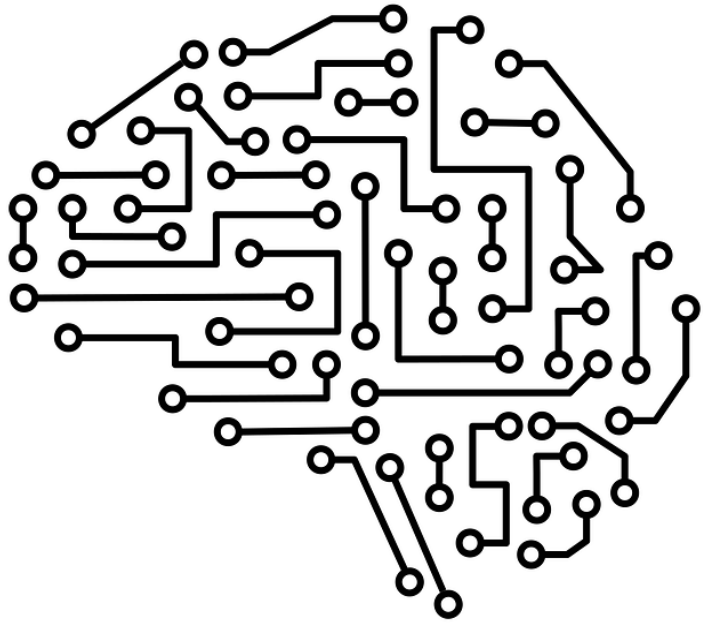
- Numerical data (Marketing Analytics)
- Image data (Face recognition)
- Video data (Object recognition)
- Sound data (Music generation)
- Text data (Sentiment analysis)

ML Workflow


Training Data
Training
ML Trained Model
Input Data
Prediction
Results

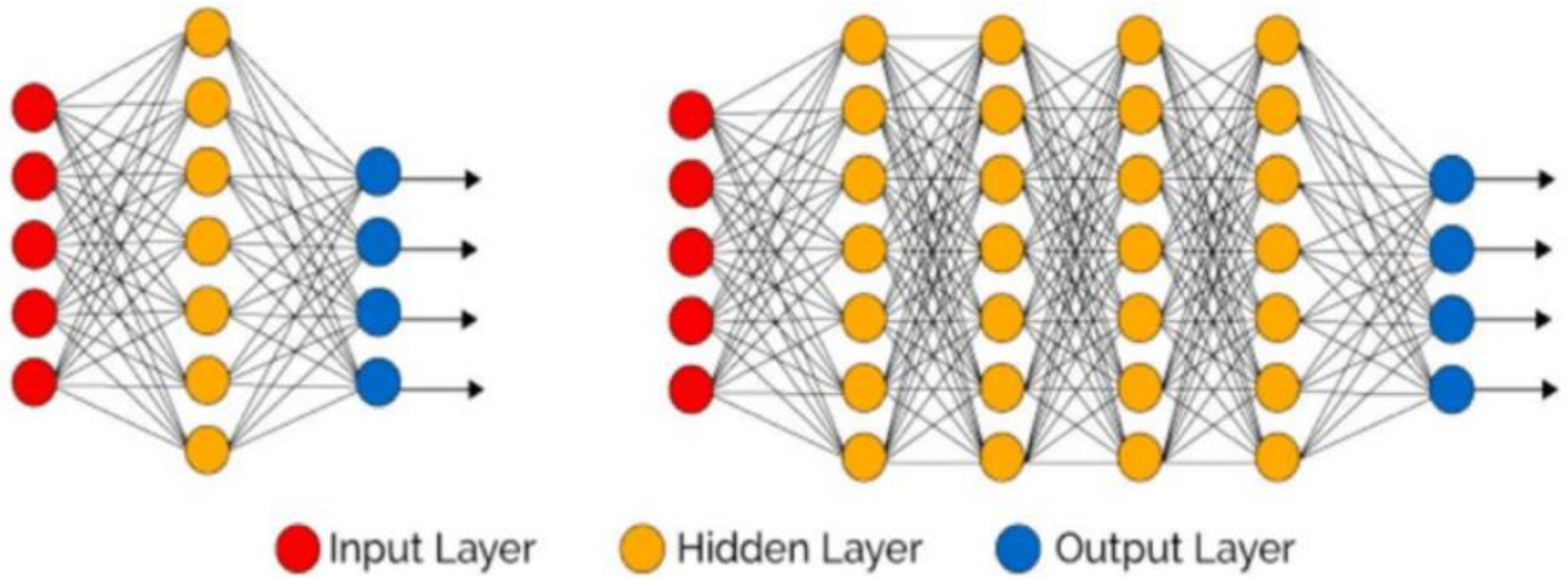
What is Deep Learning

- Deep Learning (DL) is a subset of machine learning
- Multiple layers of nonlinear processing units are used for feature extraction and transformation.
- A computational approach that involves the use of multiple layers of artificial neural networks to model and solve complex problems.



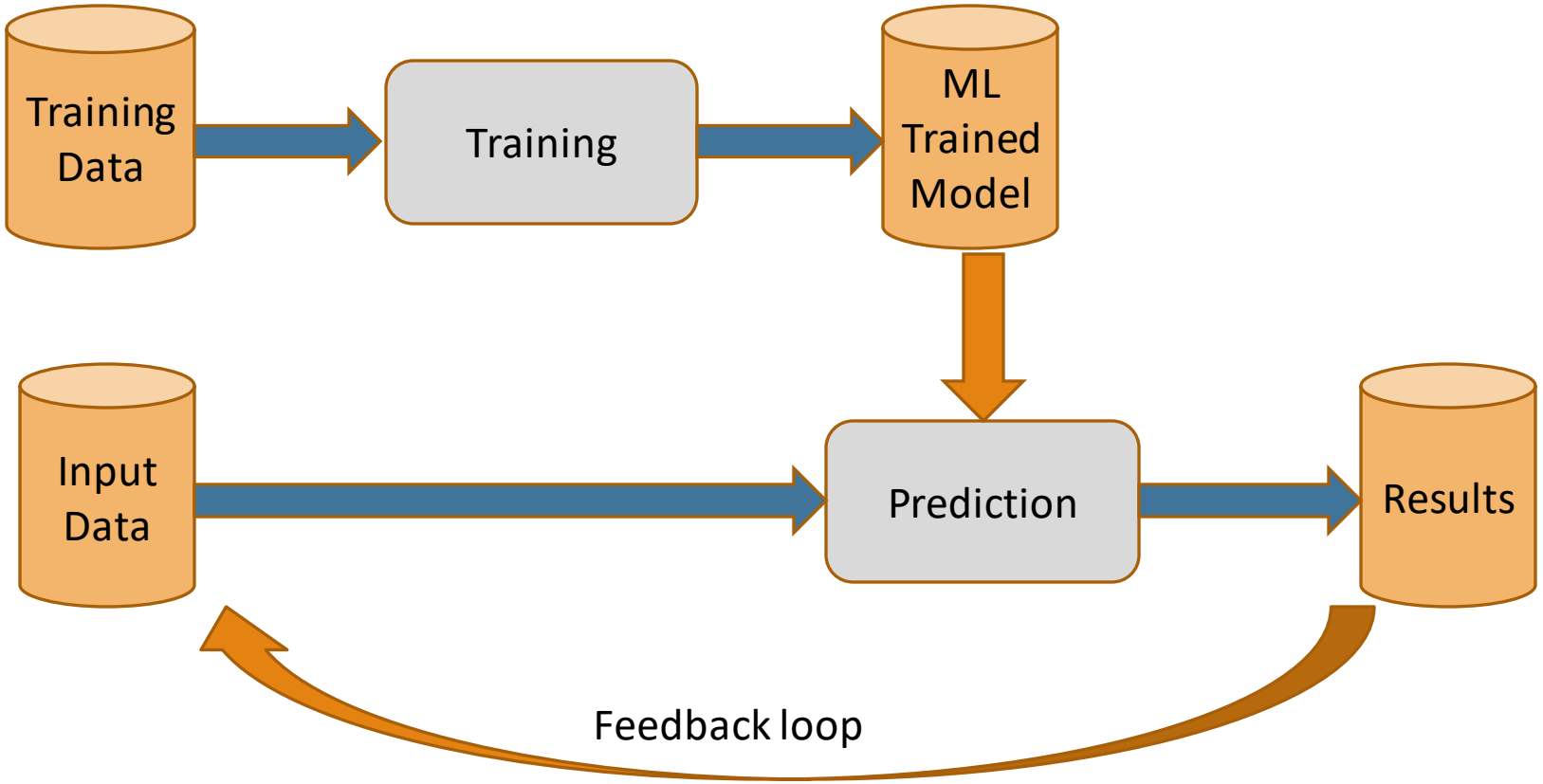
A shallow neural network with single hidden layer

A deep neural network with more than one hidden layer



https://sourcemeridian.com/machine-learning-as-a-tool-to-grow-sales-in-telemarketing-operations/

What is AI?


Training Data
Training
ML Trained Model
Input Data
Prediction
Results
Feedback loop

What is Data Science

- Data Driven Decision making
- Making sense out of data
- Uncovering the hidden insights and patterns in data
- Using machine learning models, data visualizations and intelligent reports

ML problems

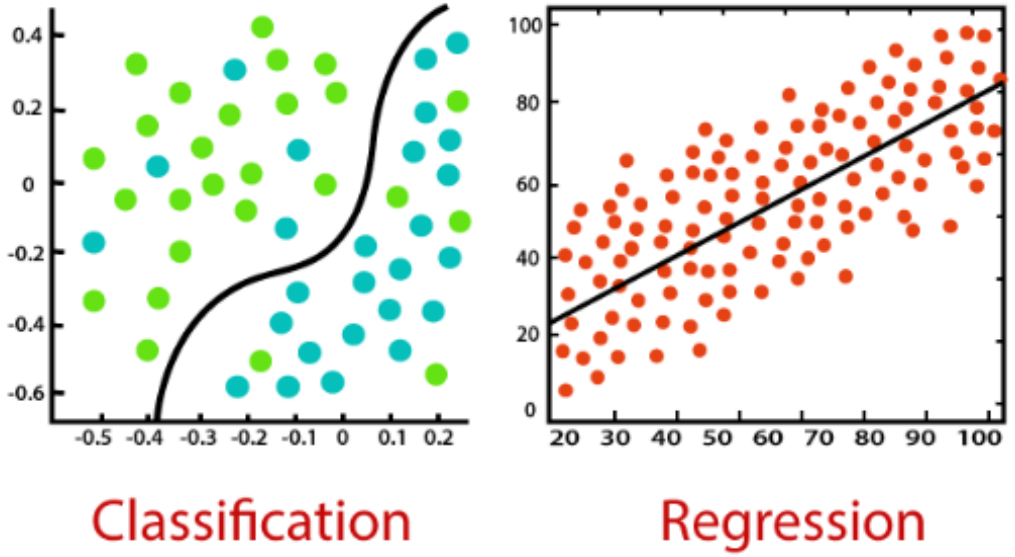
Types of Machine learning

	Supervised	Unsupervised	Reinforcement
Discrete	Classification	Clustering	Rewarding/punishing behaviour
Continuous	Regression	Dimensionality reduction	Rewarding/punishing behaviour

Supervised learning

The algorithm is trained on a labeled datasets to predict unseen data

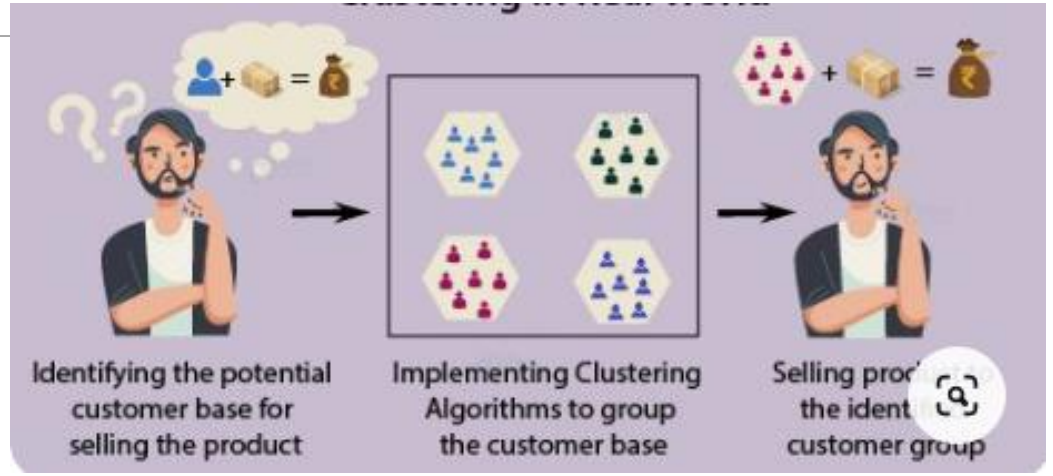
- Regression
  - Predict a continuous output variable. (The price of a house)
- Classification
  - The algorithm learns to predict a categorical output variable (classifying an email as spam or not spam)



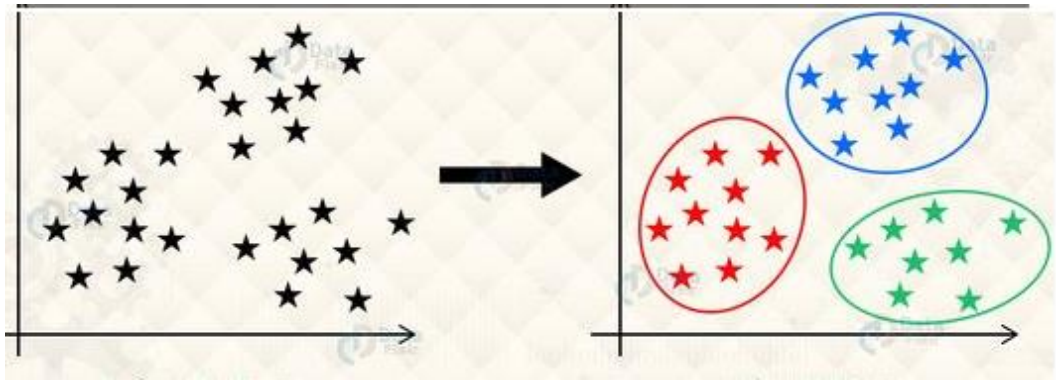
https://www.projectpro.io/article/classification-vs-regression-in-machine-learning/545

Unsupervised learning

- The algorithm is trained on unlabelled data
- Tasked with finding patterns on its own, without any feedback
  - Clustering
  - Dimensionality reduction
  - Anomaly detection



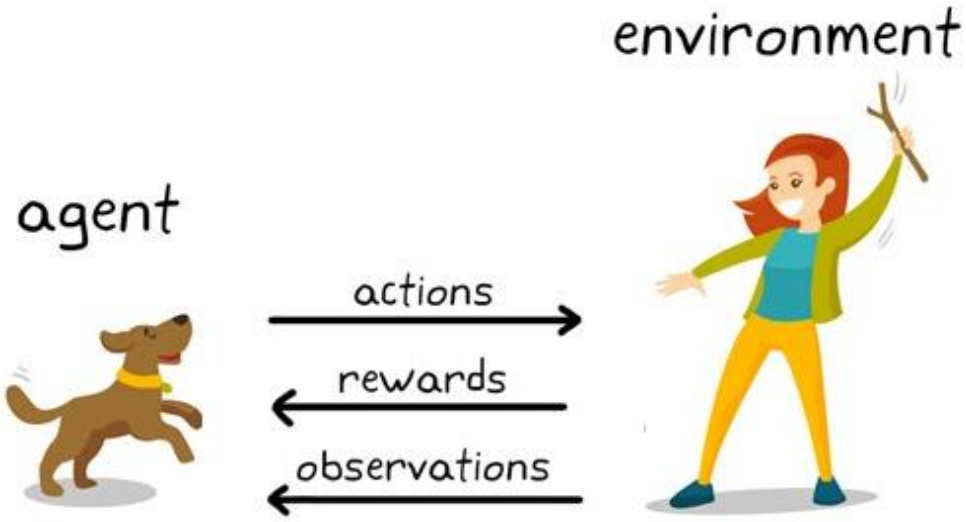
https://data-flair.training/blogs/clustering-in-machine-learning/



https://data-flair.training/blogs/scipy-clustering/

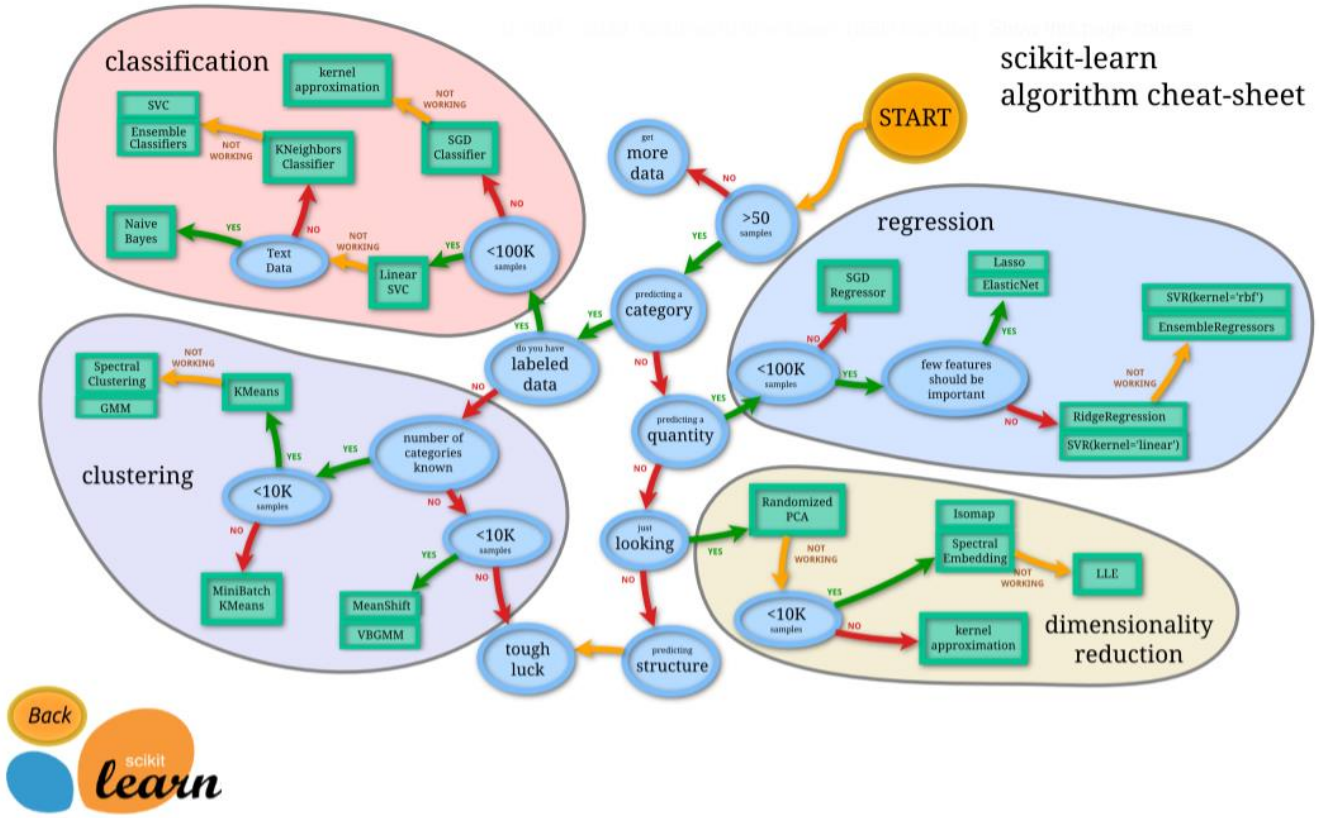
Reinforcement learning

- Rewarding desired/ punishing undesired behaviours
- Able to perceive and interpret its environment, take actions and learn through trial and error



https://medium.com/analytics-vidhya/a-beginners-guide-to-reinforcement-learning-and-its-basic-implementation-from-scratch-2c0b5444cc49

Selecting an Algorithm



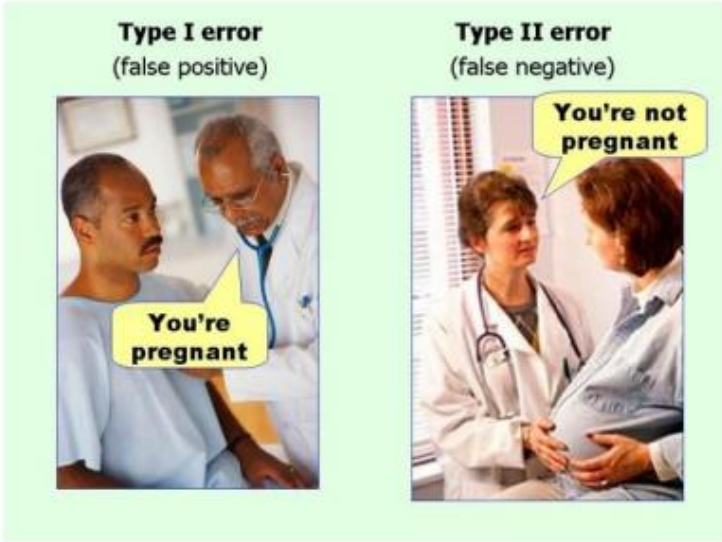
Model Evaluation

Important Metrics:

- Accuracy
- Precision
- Sensitivity
- Specificity

The cost associated with each type of error is different in different situations.

Measured / Perceived	Reality: True	Reality: False
True	Correct ☺	Type I False Positive
False	Type II False Negative	Correct ☺



Data Science Library

NumPy

- N-dimensional array packages for numerical computing with Python

Pandas

- Manipulating and analysing numerical tables and time series

SciPy

- Collection of open source software for scientific computing in Python.

Matplotlib

- Plotting library for the Python

Scikit-learn

- an open-source, simple, and efficient tool for data mining and data analysis
- Built on NumPy, SciPy, and Matplotlib

Pandas

Offers data analysis, data cleaning, and data wrangling

Provides a DataFrame object

- Two-dimensional table-like data structure
- Store and manipulate data in a tabular format

Provides a Series object

- One-dimensional labeled array
  - Representing a column or row of data in a DataFrame

Scikit-learn

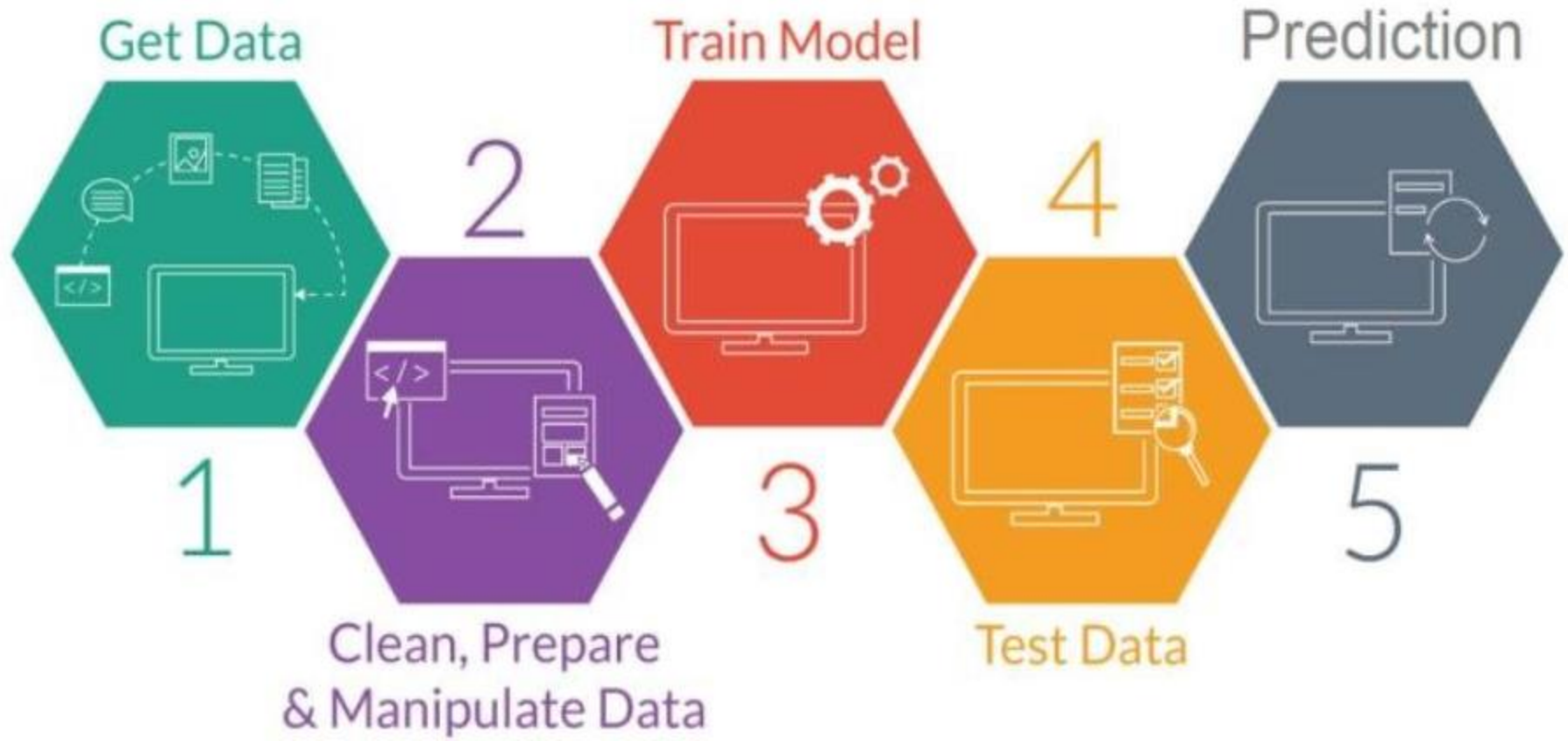
Provides a range of supervised and unsupervised learning algorithms

- Classification
- Regression
- Clustering
- Dimensionality reduction

Provides tools for data preprocessing

- Model selection
- Evaluation

Workflow of Machine learning



https://www.researchgate.net/figure/Predictive-Modelling-Phases_fig1_328216741