



UNIVERSIDADE
CATÓLICA
PORTUGUESA

BRAGA

Machine Learning

Session 1 - T

Introduction to Machine Learning

Degree in Applied Data Science

2024/2025

Objectives

- Learn the theoretical foundations of advanced machine learning concepts;
- Implement advanced machine learning algorithms using Python;

- Typically, classes will be divided in two parts:
 - Foundations of Data Science and Machine Learning (1 hour).
 - Hands-on Implementation in Python (1 hour).

Evaluation

- Exam (35%)
 - Minimum passing grade of 9.5
- Group project (35%)
 - Groups of 3 students;
 - Presentation;
 - Jupyter Notebook;
 - Minimum passing grade of 9.5.
- ML package (30%)
 - Group development of an ML package from scratch on GitHub.



UNIVERSIDADE
CATOLICA
PORTUGUESA
BRAGA

Introduction to Machine Learning

Terminology



FinancialForecasting
PredictiveMaintenance
DataPreprocessing CloudComputing
AutonomousVehicles EnsembleMethods NeuralNetworks
HealthcareAnalytics FeatureEngineering ReinforcementLearning
ROCCurve KnearestNeighbors SpeechRecognition
SupportVectorMachines GradientBoosting
Precision RandomForest NaiveBayes LogisticRegression
PatternRecognition AUC
ModelEvaluation Classification Kmeans FeatureSelection
DeepLearning Clustering MachineLearning RecommendationSystems
AnomalyDetection DecisionTrees Recall Regression F1Score DataMining
TimeSeriesAnalysis SupervisedLearning CrossValidation
ConfusionMatrix UnsupervisedLearning ComputerVision
FraudDetection DimensionalityReduction
ImageProcessing BiasVarianceTradeoff BigData
HyperparameterTuning
NaturalLanguageProcessing
CustomerSegmentation

Data everywhere!

- 328.77 million terabytes of data are created each day.*
- By 2025, it is expected to skyrocket to a staggering 181 zettabytes.*
- Last two years alone, account for 90% of the world's data.*

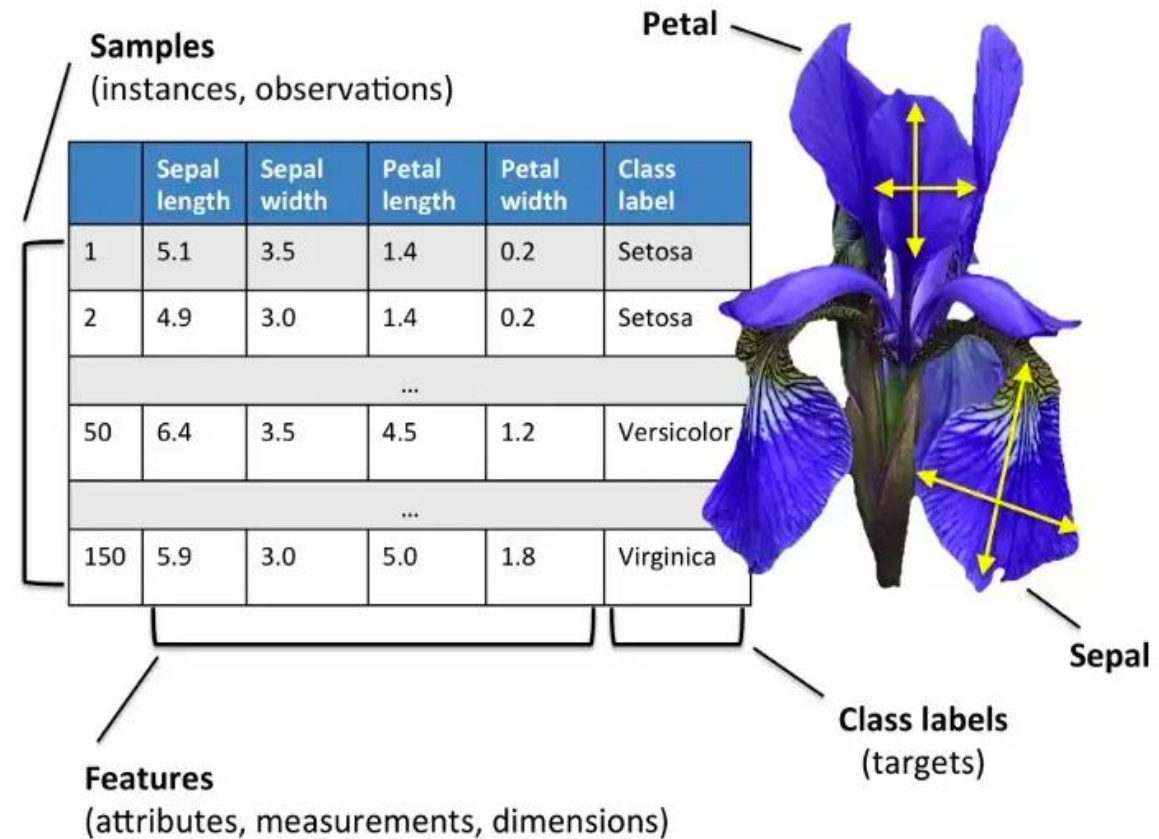
* <https://news.wildintelligence.xyz/p/328-million-terabytes-of-data>



Data Types



- Tabular data:



<https://eminebozkus.medium.com/exploring-the-iris-flower-dataset-4e000bcc266c>

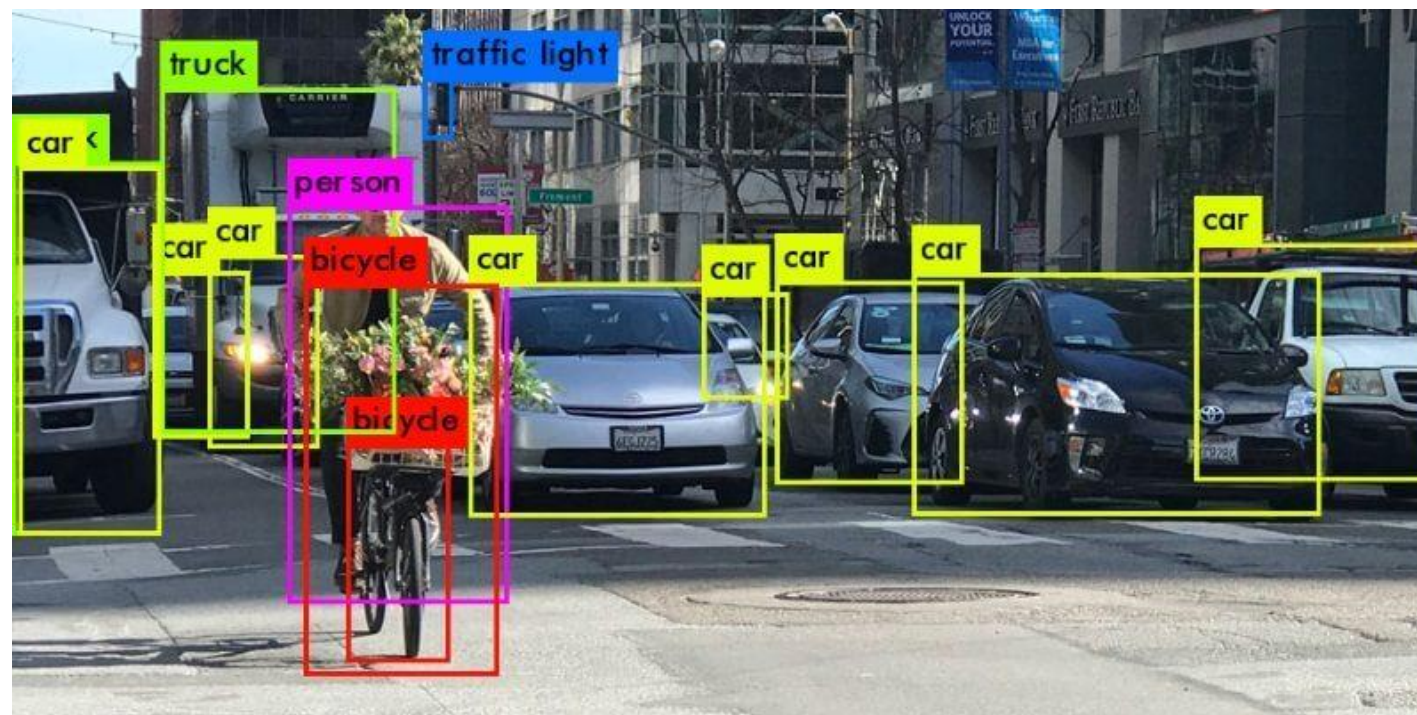
Data Types



UNIVERSIDADE
CATOLICA
PORTUGUESA

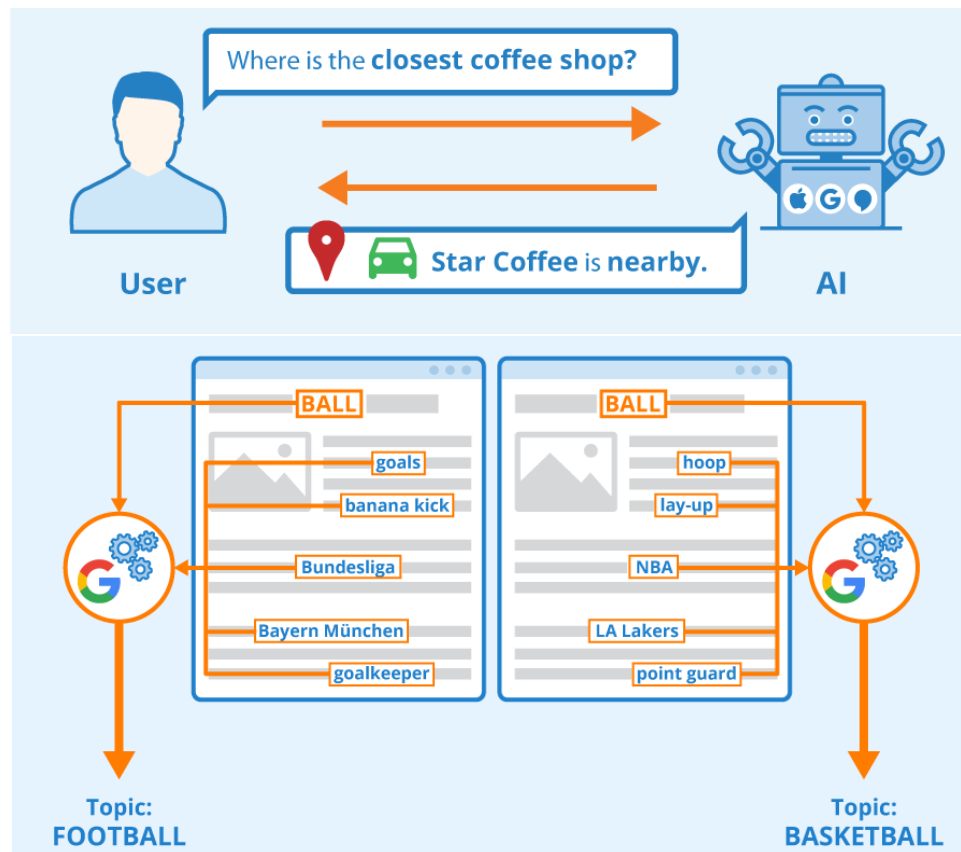
BRAGA

- Image data:



Data Types

- Text data:



Data Types

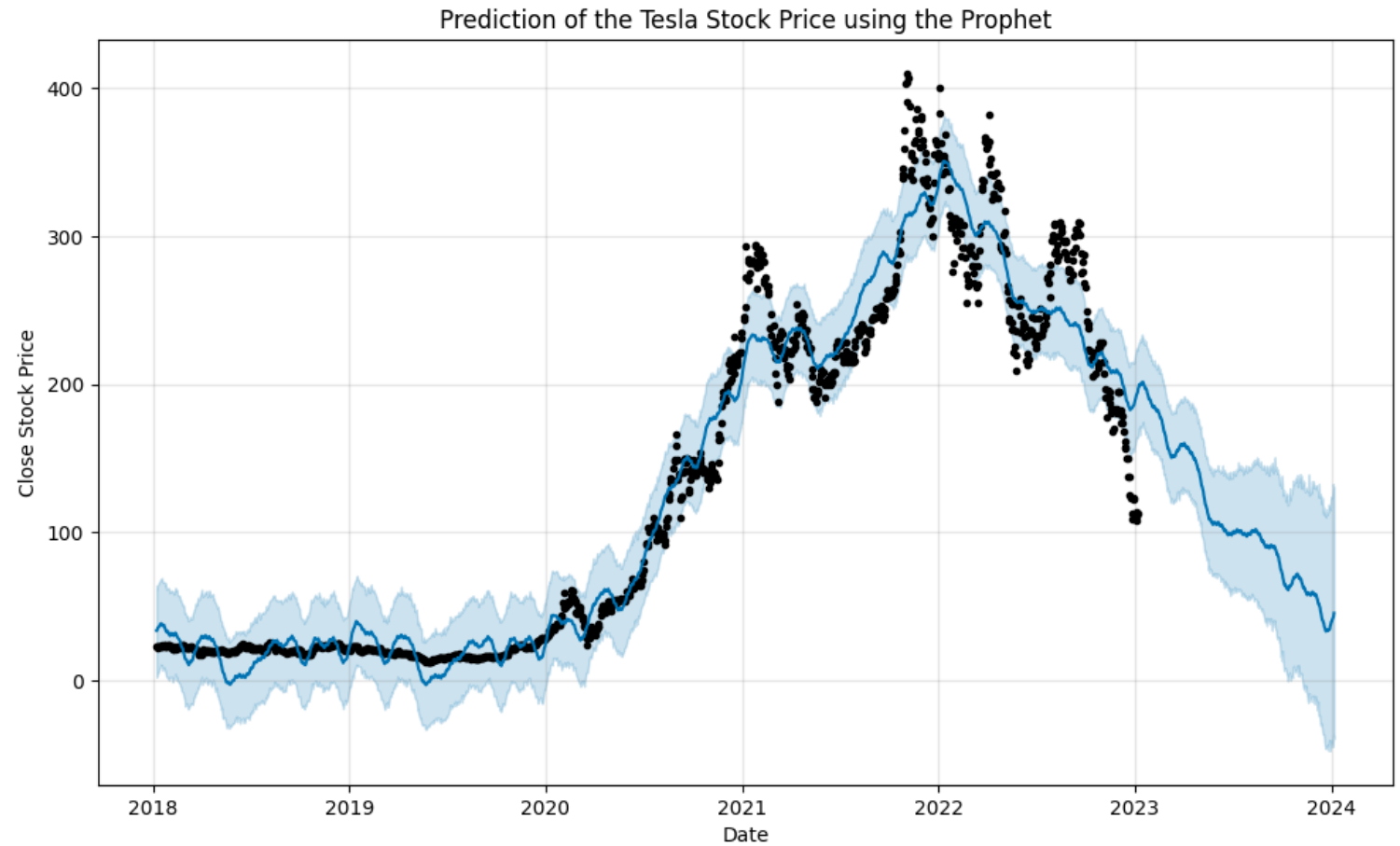


- Audio data:



Data Types

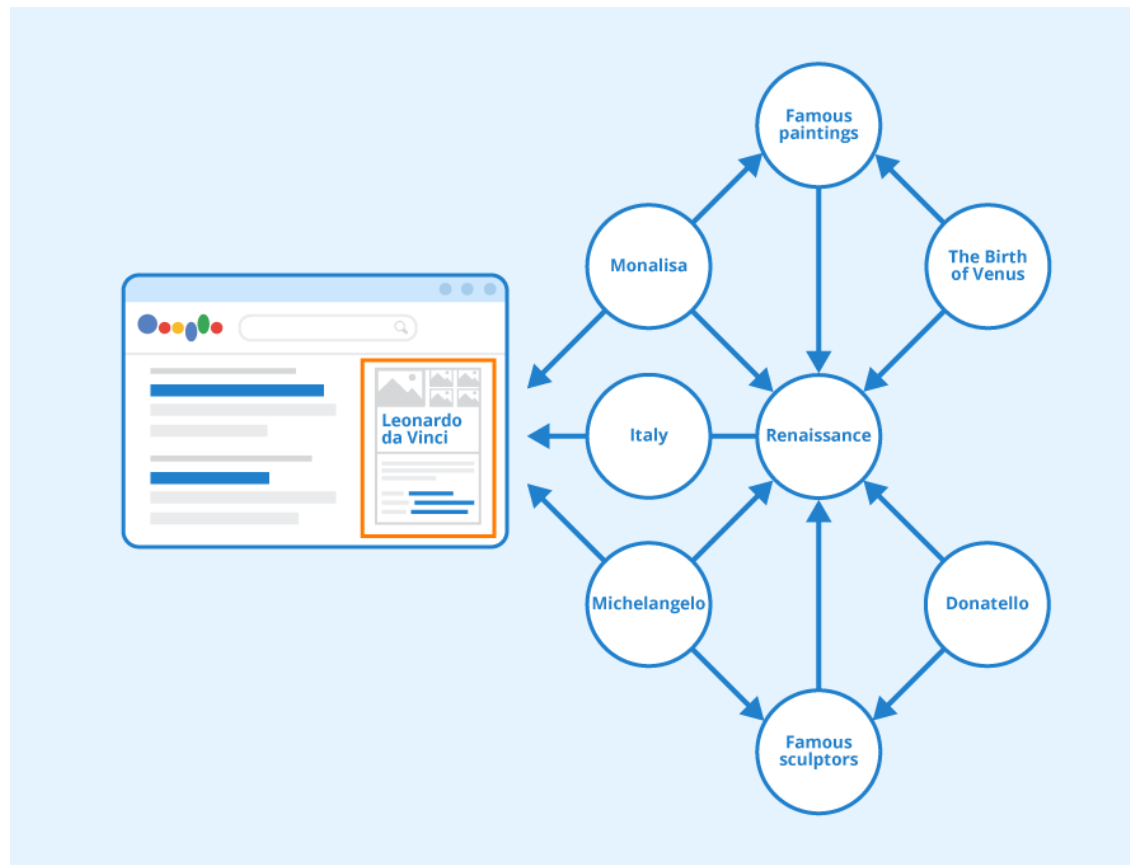
- Time series data:



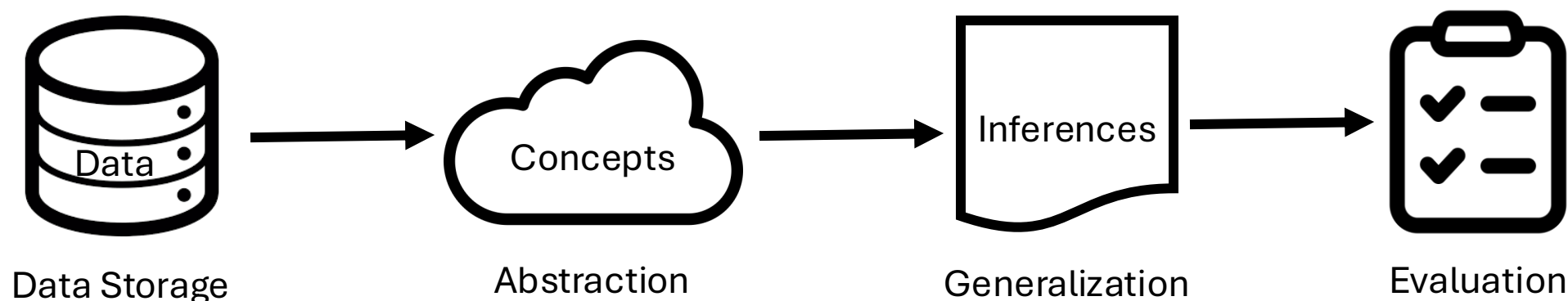
Data Types



- Graph data:



The Components of Learning



- **Data storage:** every learning process starts with data. Both humans and computers rely on data storage as a fundamental basis for more advanced reasoning..
- **Abstraction:** involves extracting knowledge from stored data by forming general concepts that represent the data as a whole.
- **Generalization** involves leveraging abstracted data to generate knowledge and inferences, which subsequently inform actions in novel situations.
- **Evaluation:** serves as a feedback mechanism, measuring the effectiveness of acquired knowledge and offering insights for possible enhancements.

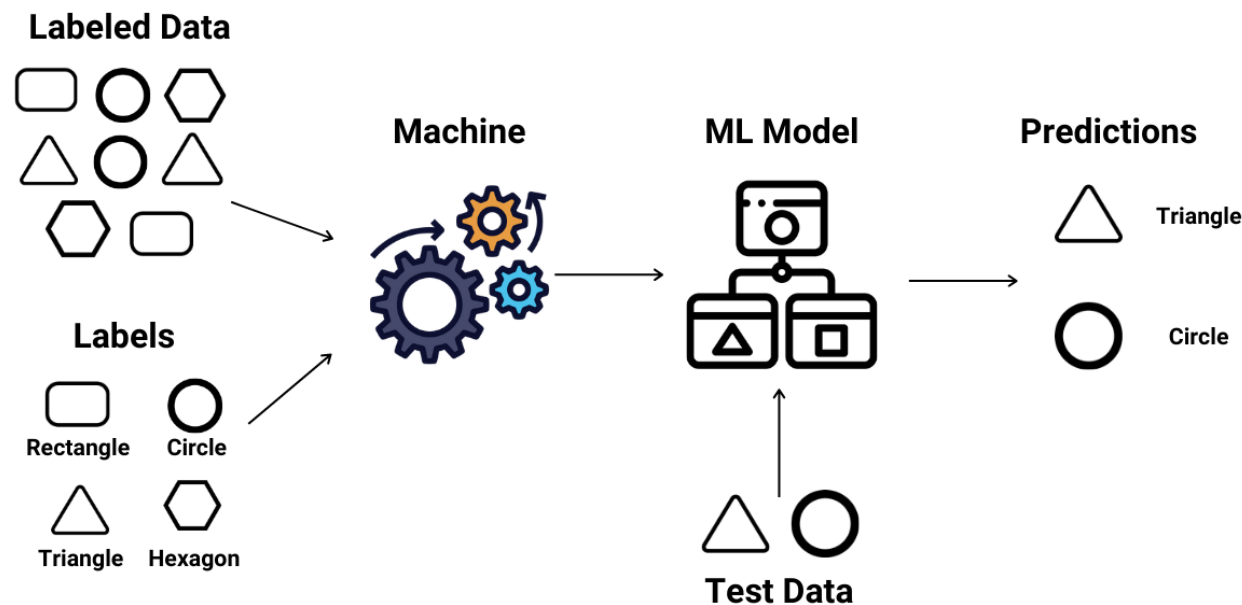
Machine Learning

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E . ”

Tom Mitchell. Machine Learning 1997.

Machine Learning Paradigms

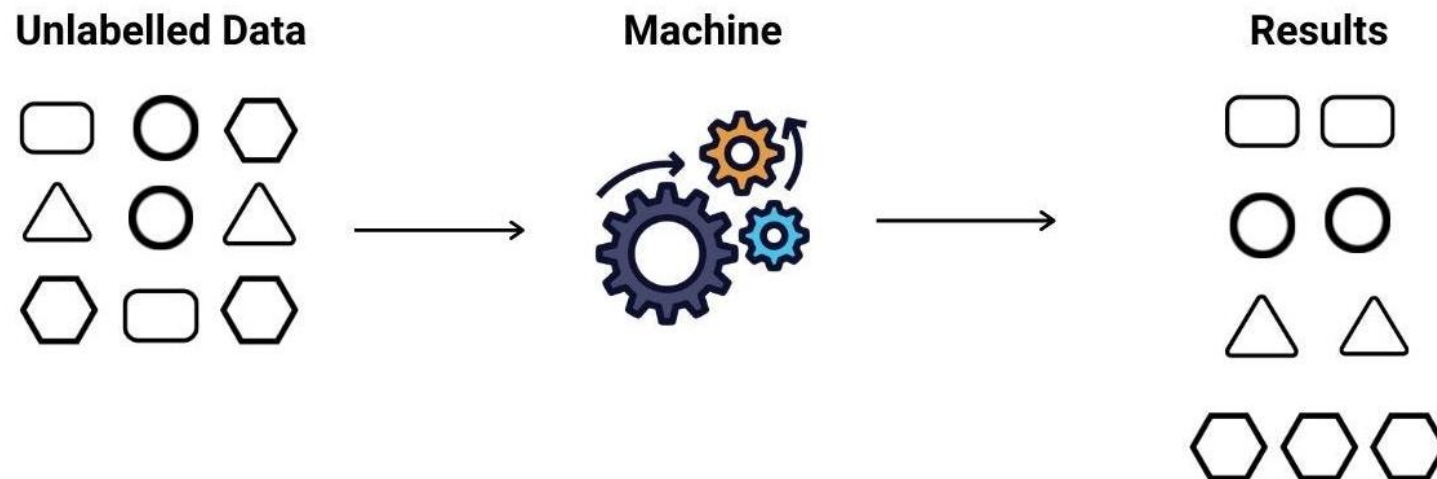
- **Supervised:** the algorithm is trained on a labeled dataset, where the input data is paired with corresponding output labels. The goal is to learn a mapping from inputs to outputs, allowing the algorithm to make predictions on new, unseen data.



<https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning>

Machine Learning Paradigms

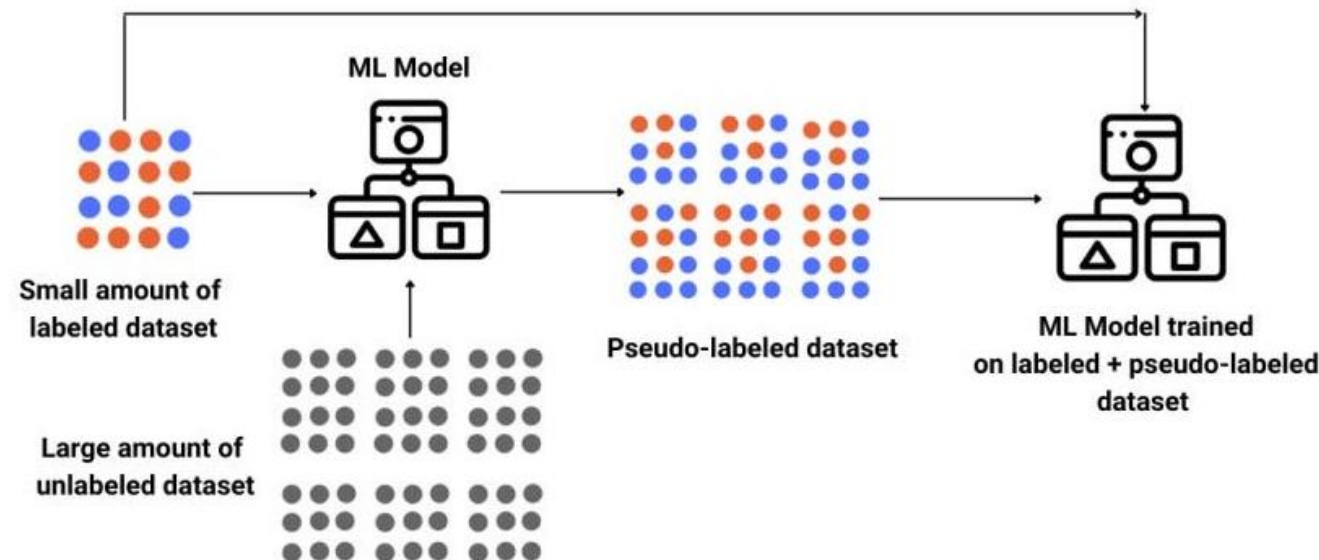
- **Unsupervised:** involves working with unlabeled data, where the algorithm explores the inherent structure and patterns within the input without explicit output guidance.



<https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning>

Machine Learning Paradigms

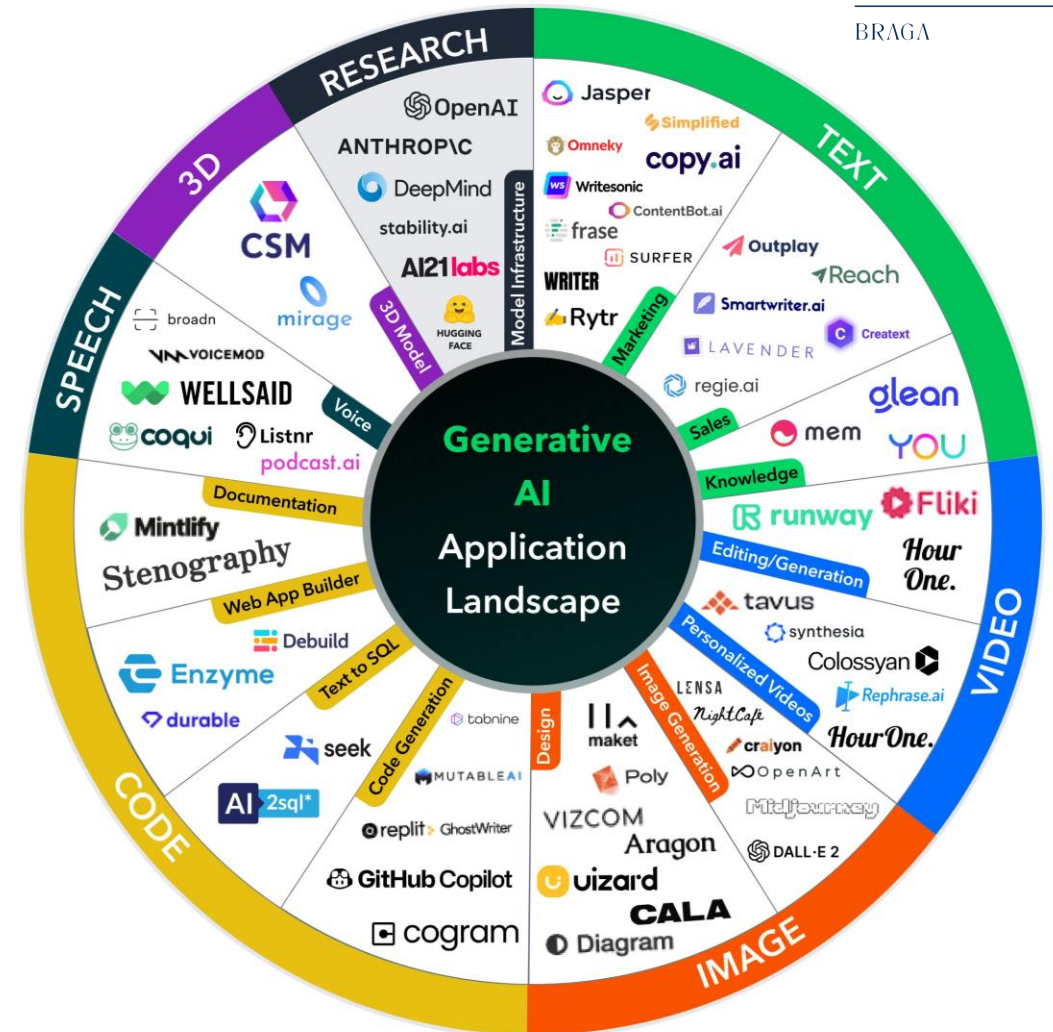
- **Semi-supervised:** combines labeled and unlabeled data for training. It aims to leverage the benefits of both supervised and unsupervised learning, often useful when obtaining labeled data is costly or time-consuming.



<https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning>

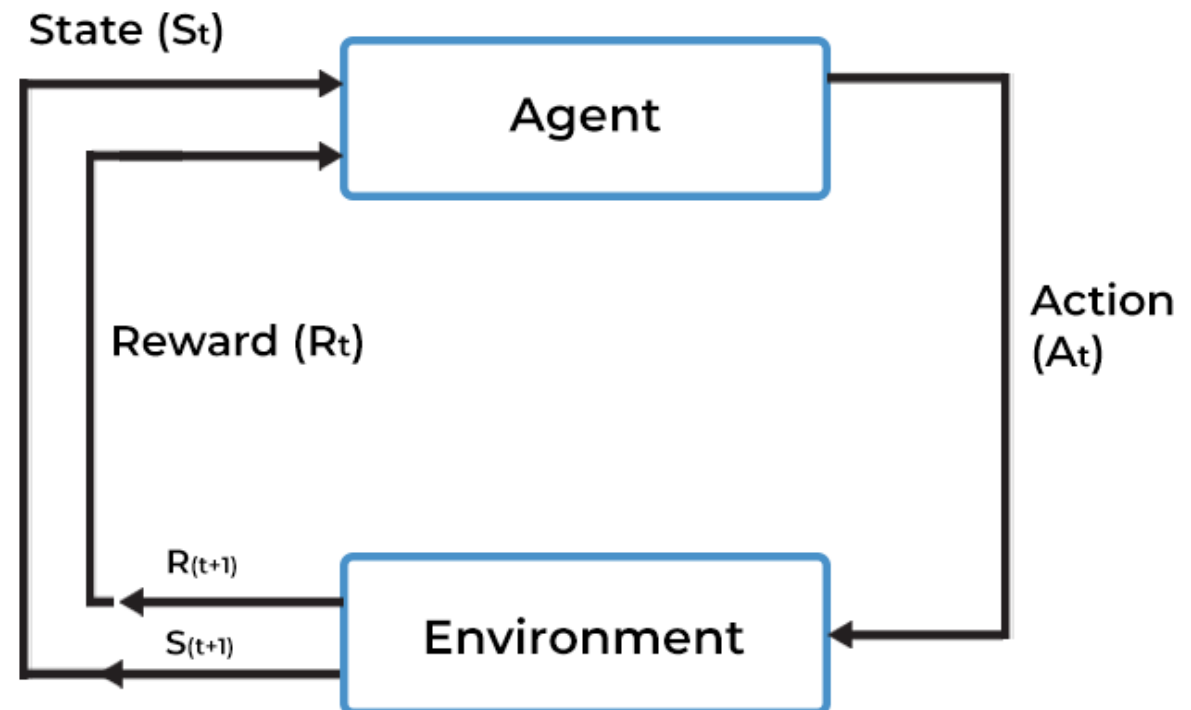
Machine Learning Paradigms

- **Generative:** the focus is on learning the underlying probability distribution of the input data. It can generate new, synthetic samples that resemble the training data.



Machine Learning Paradigms

- **Reinforcement Learning:** an agent learns to make decisions by interacting with an environment. The agent receives feedback in the form of rewards or penalties, enabling it to learn optimal strategies over time.





Machine Learning Applications

- Spam filtering;
- Fraud detection;
- Disease detection;
- Drug discovery;
- Recommendation systems;
- Forecasting;
- Autonomous Vehicles;
- Chatbots;
- Object detection;
- Anomaly detection;
- ...

Resources

- Kelleher, J. D., Namee, B. M., & D'Arcy, A. (2015). Fundamentals of machine learning for predictive data analytics. London, England: MIT Press.
- Guido, S., & Mueller, A. C. (2016). Introduction to machine learning with python. Sebastopol, CA: O'Reilly Media.