Université Constantine 2
جامعة قسنطينة 2

Module : Machine Learning (ML – SDSI)

– Course 0 –

Chapter 0 : Introduction to Machine Learning

Habib-Ellah GUERGOUR
Faculty of NTIC / TLSI Department
Contact: habib.guergour@univ-constantine2.dz

Université Constantine 2
جامعة قسنطينة 2

Module : Machine Learning (ML – SDSI)

– Course 0 –

Chapter 0 : Introduction to Machine Learning

Habib-Ellah GUERGOUR
Faculty of NTIC / TLSI Department
Contact: habib.guergour@univ-constantine2.dz

Etudiants concernés

Faculté/Institut	Département	Niveau	Spécialité
NTIC	TLSI	M1	SDSI

Goals of the Chapter

- Understand what machine learning is and its importance.
- Discover Machine Learning Key Definitions
- How to Build a Machine Learning Project
- Challenges and Limitations

Main Titles

- Introduction to Machine Learning
- History of Machine Learning
- Definitions and Key Concepts
- The Lifecycle of a Machine Learning Project
- Challenges and Limitations of Machine Learning

Introduction to Machine Learning

What is Machine Learning?

- Machine Learning (ML) is a branch of Artificial Intelligence (AI) that focuses on developing systems capable of learning from data and improving their performance over time without being explicitly programmed for specific tasks.
- At its core, machine learning is about building algorithms that:
  - Take in data as input.
  - Identify patterns or relationships in that data.
  - Make predictions or decisions based on those patterns.

What is Machine Learning?

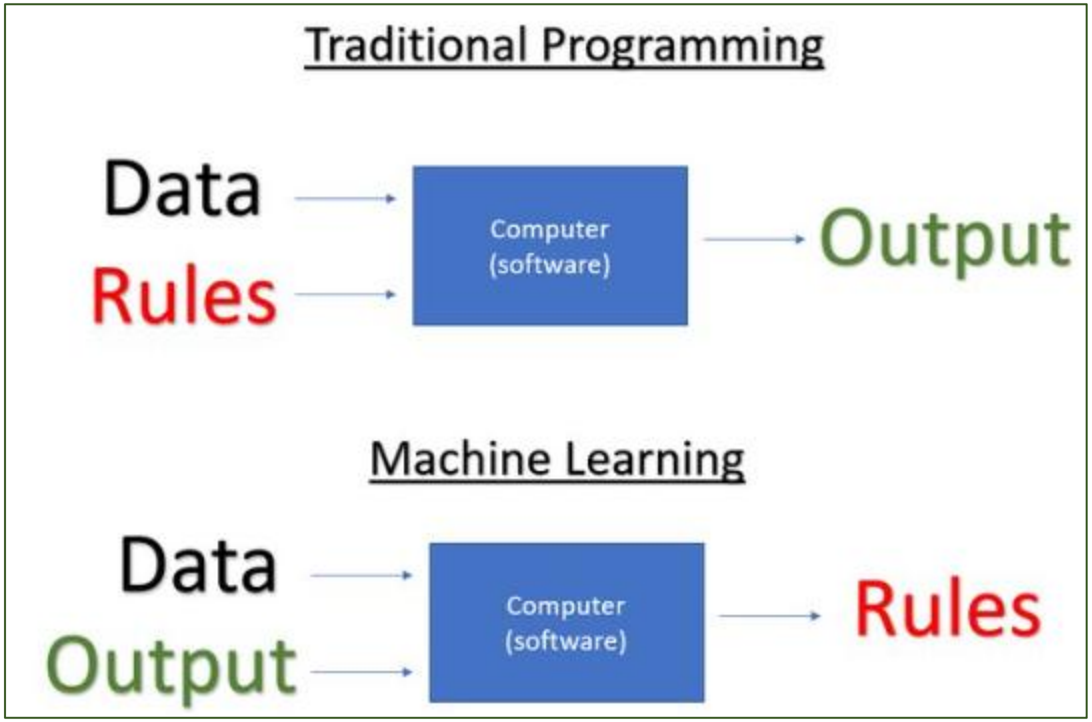
In simple terms, machine learning allows computers to "learn" from past experiences (data) and generalize to unseen situations..

Why is Machine Learning Important?

- Machine learning is transforming the way problems are solved in various domains. Its importance lies in its ability to:
  - Automate Decision-Making
  - Handle Complex Problems
  - Adapt Over Time
  - Unlock Value from Data

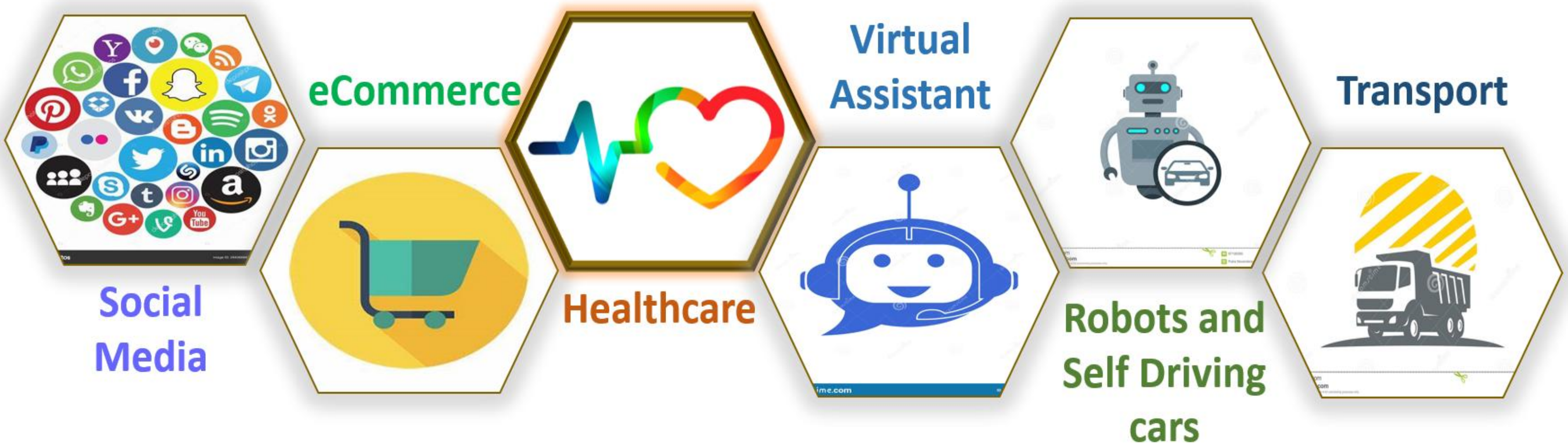
Key Idea Behind Machine Learning

- The central idea of machine learning is that instead of hardcoding rules for solving a specific task, a model is trained using data to automatically derive rules or patterns.
- For instance:



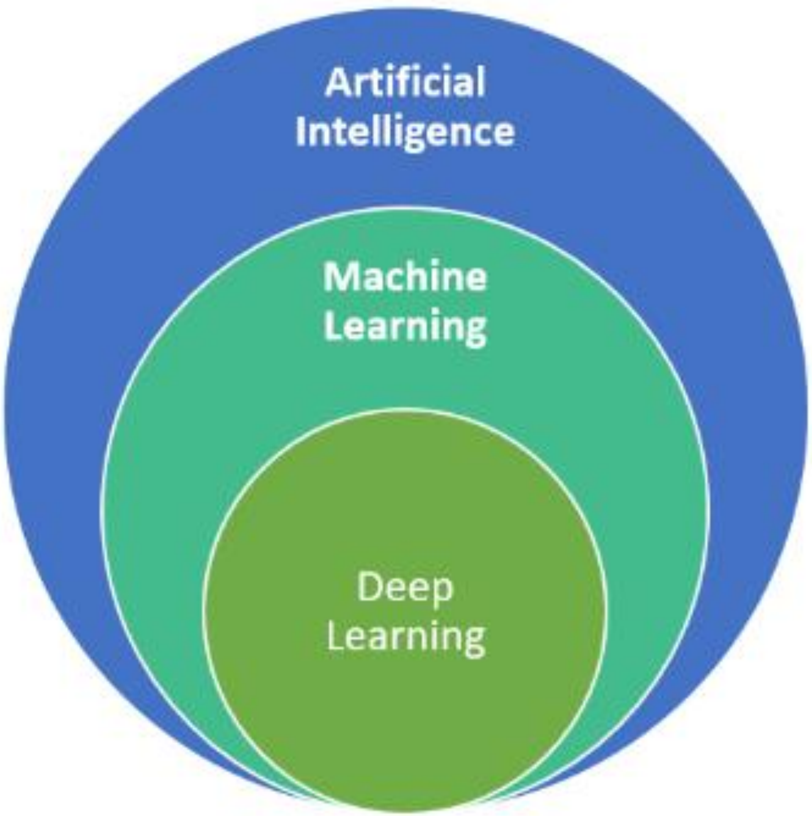
Machine Learning in Our Daily Lives

- Machine Learning in Our Daily Lives



Relation Between AI, Machine Learning, and Deep Learning

- **Artificial Intelligence (AI)**: The broad field of creating systems that exhibit intelligent behavior.
- **Machine Learning (ML):** A subset of AI focusing on algorithms that learn from data.
- **Deep Learning (DL):** A further subset of ML that uses neural networks with many layers to model complex patterns in data.



Main Titles

- Introduction to Machine Learning
- History of Machine Learning
- Definitions and Key Concepts
- The Lifecycle of a Machine Learning Project
- Challenges and Limitations of Machine Learning

Main Titles

- Introduction to Machine Learning
- History of Machine Learning
- Definitions and Key Concepts
- The Lifecycle of a Machine Learning Project
- Challenges and Limitations of Machine Learning

History of Machine Learning

Key Milestones in Machine Learning

Year	Event
1943	McCulloch and Pitts propose the first mathematical model of artificial neurons.
1950	Alan Turing publishes *"Computing Machinery and Intelligence"* and introduces the **Turing Test**.
1957	Frank Rosenblatt invents the **Perceptron**, the first algorithm to learn from data.
1969	Minsky and Papert identify the limitations of the Perceptron, leading to the **first AI Winter**.
1980	The **Backpropagation** algorithm is rediscovered, reigniting interest in neural networks.
1987–1993	The **Second AI Winter** occurs as expert systems fail to deliver on high expectations, causing funding cuts.
1995	Vladimir Vapnik introduces **Support Vector Machines (SVMs)**, a breakthrough in classification.
..	...

Key Milestones in Machine Learning

Year	Event
..	…
2012	**AlexNet** wins the ImageNet competition, marking the rise of **deep learning**.
2016	**AlphaGo**, developed by DeepMind, defeats the Go world champion, showcasing the power of reinforcement learning.
2020	Advanced natural language models like **GPT-3** revolutionize text generation and NLP tasks.
..	…

Main Titles

- Introduction to Machine Learning
- History of Machine Learning
- Definitions and Key Concepts
- The Lifecycle of a Machine Learning Project
- Challenges and Limitations of Machine Learning

Main Titles

- Introduction to Machine Learning
- History of Machine Learning
- Definitions and Key Concepts
- The Lifecycle of a Machine Learning Project
- Challenges and Limitations of Machine Learning

Definitions and Key Concepts

Definition of Machine Learning

- **"Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.“**

(Arthur Samuel,1959)

« Un domaine d'études qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés "

Definition of Machine Learning

- **"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."**

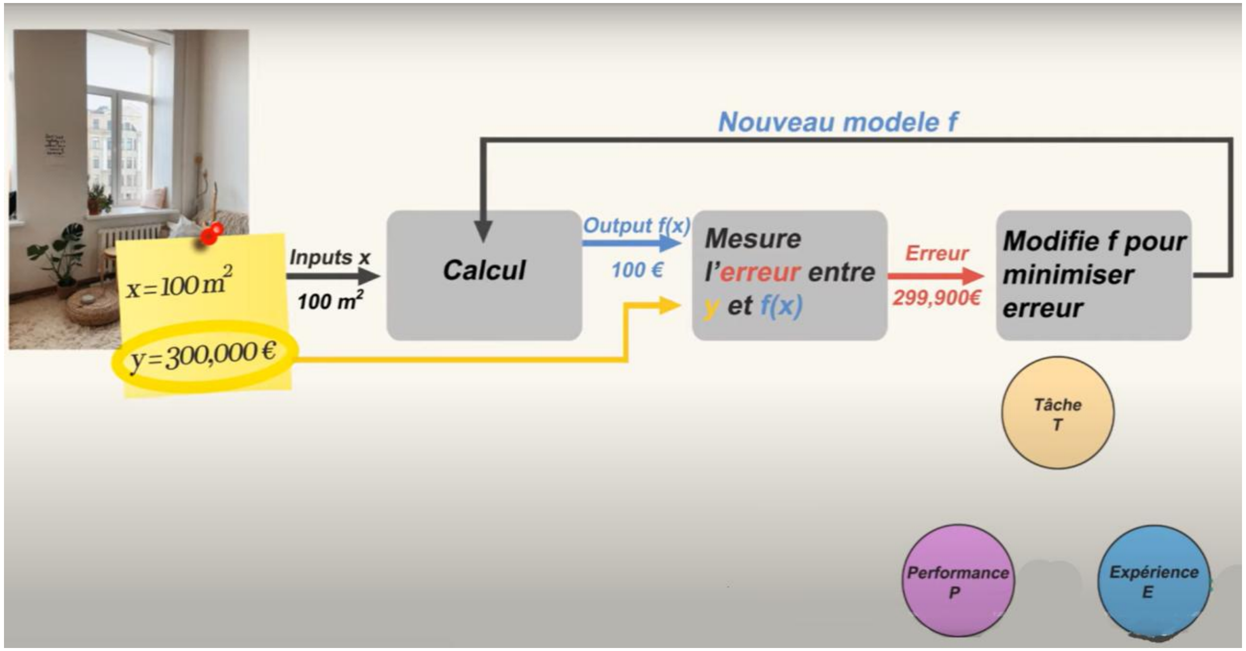
" Un programme informatique se dit d'apprendre de l'expérience **E** par rapport à une catégorie de tâches **T** et mesure de la performance **P**, si sa performance à des tâches **T**, telle que mesurée par **P**, s'améliore avec l'expérience **E.**"

(Tom Mitchell, 1997)

Definition of Machine Learning

- "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

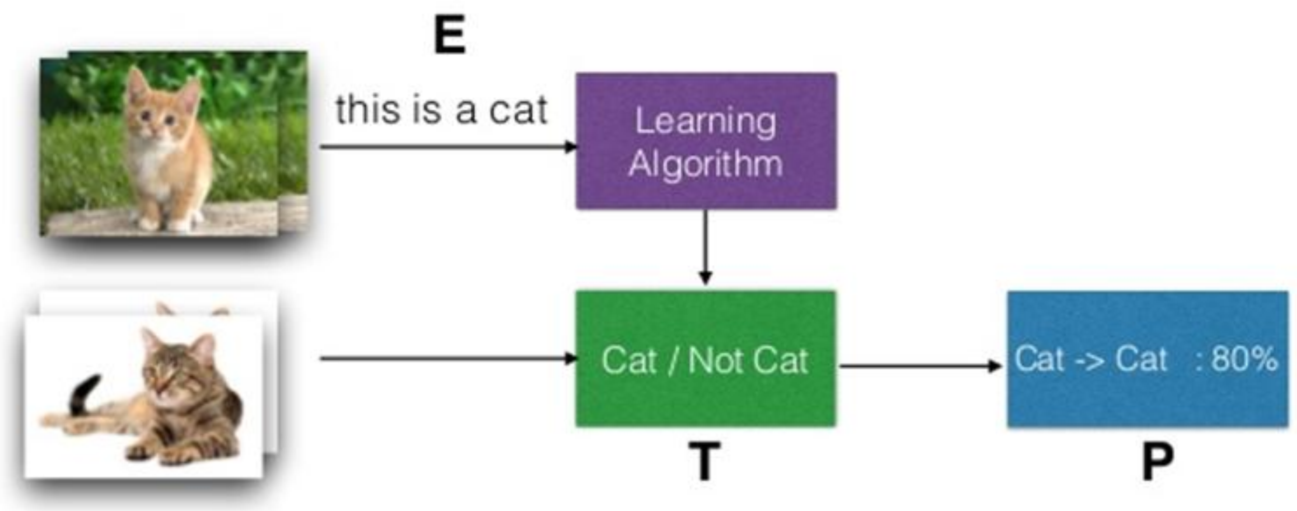
(Tom Mitchell, 1997)



Definition of Machine Learning

- "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

(Tom Mitchell, 1997)



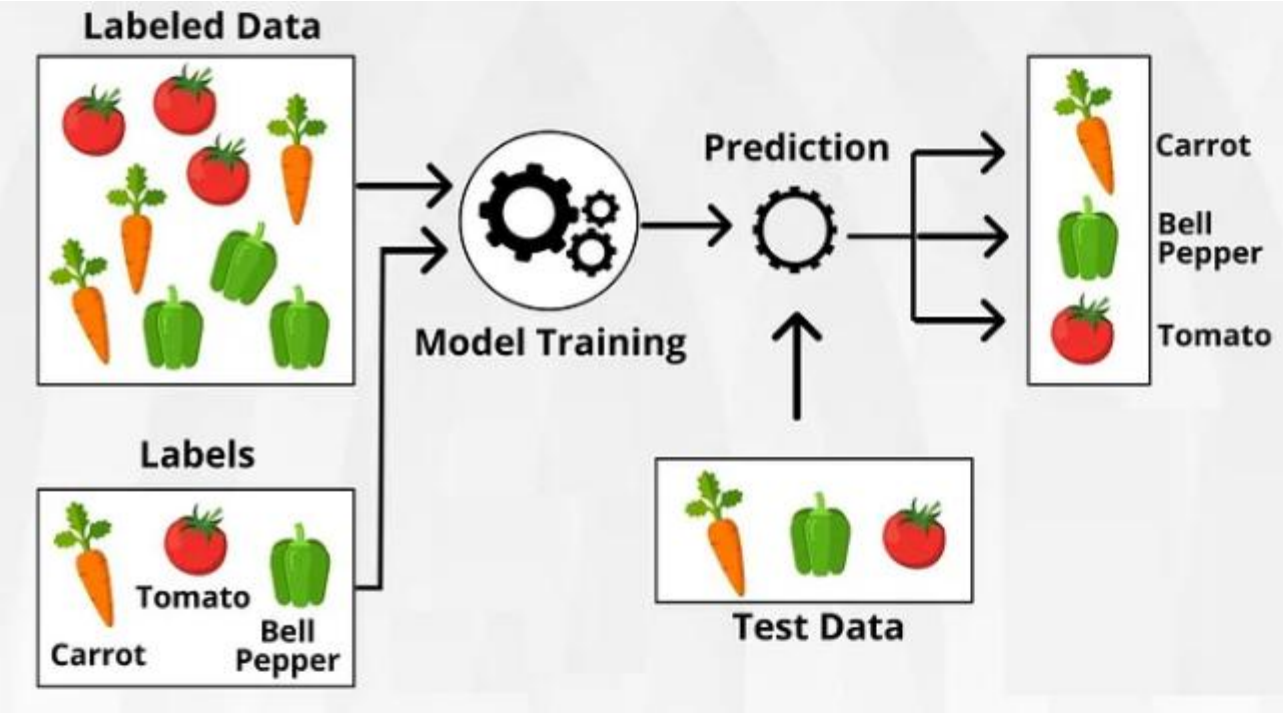
Key Concepts : Learning Types

Learning Types

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Key Concepts : Learning Types

- **Supervised Learning:** The algorithm learns from labeled data (input-output pairs).



Key Concepts : Learning Types

Example 2 : House Price Prediction

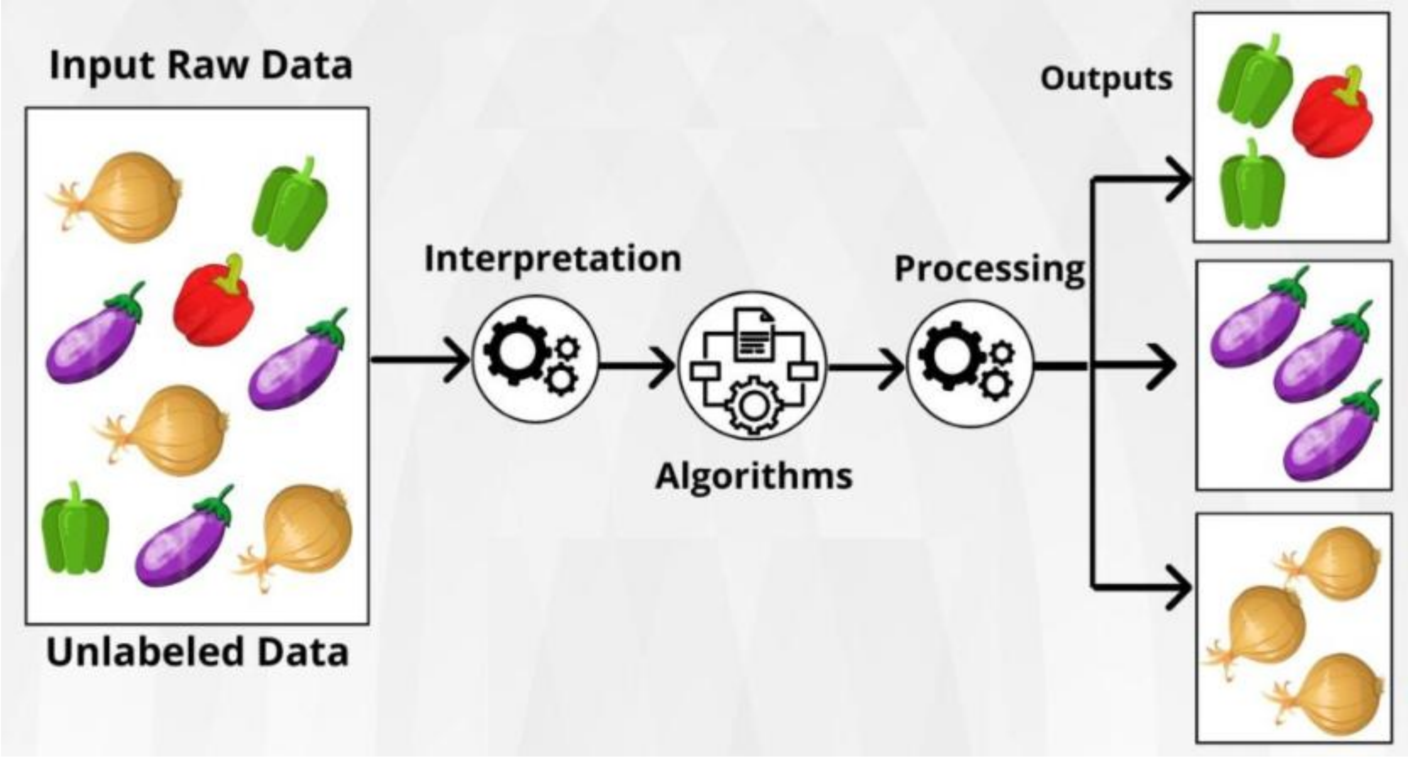
Features | Label

A data: first row

Size ($feet^2$)	Number of bedrooms	Number of floors	age of home (years)	Price($1000)
2104	5	1	45	460
1416	3	2	40	232
1534	2	2	30	315
...	...	...	...	...

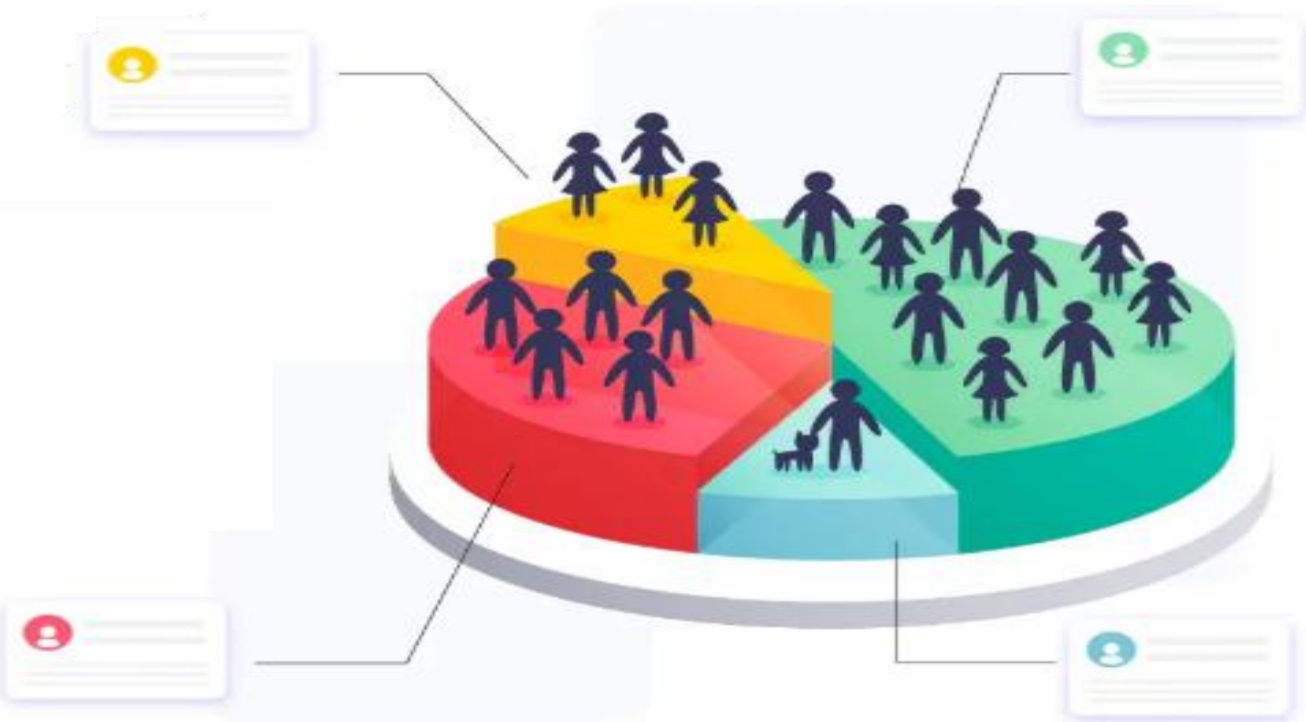
Key Concepts : Learning Types

- **Unsupervised Learning:** The algorithm identifies patterns in unlabeled data.



Key Concepts : Learning Types

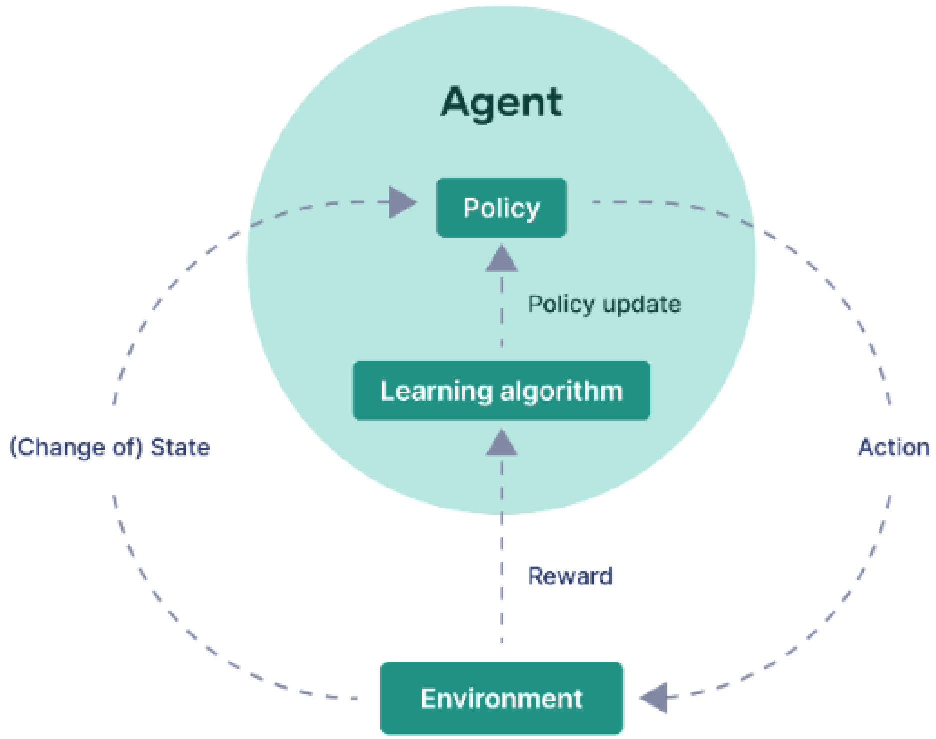
Example 2: Customer segmentation for marketing campaigns



Definitions and Key Concepts

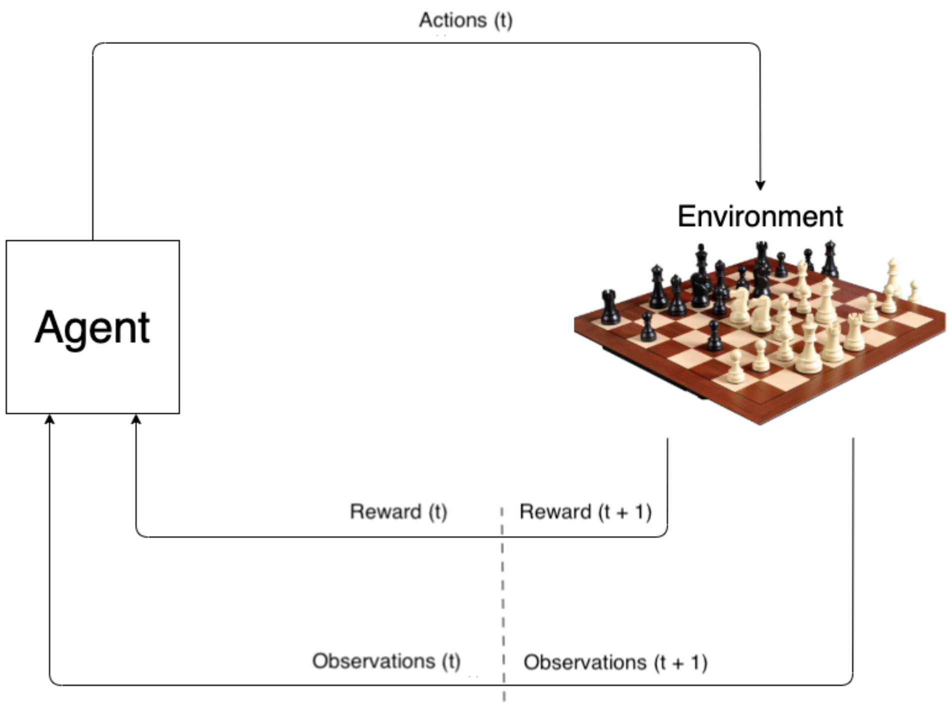
Key Concepts : Learning Types

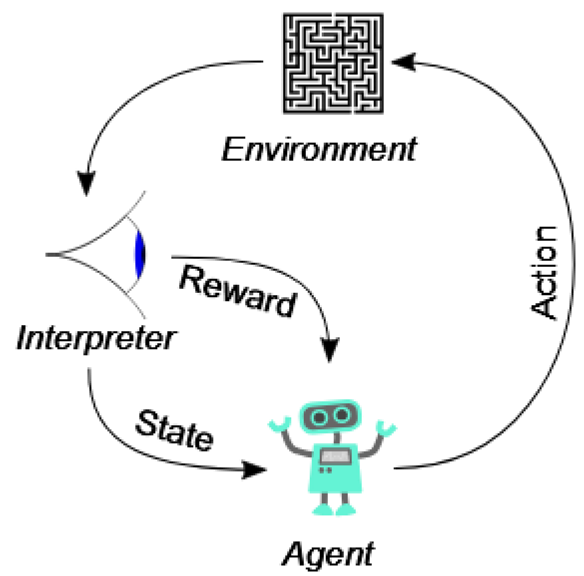
- **Reinforcement Learning:** Learning by interacting with an environment and receiving feedback in the form of rewards or penalties.



Key Concepts : Learning Types

- **Reinforcement Learning:**
  - Examples: Training a robot to walk or an AI to play chess.





Key Concepts : ML Models

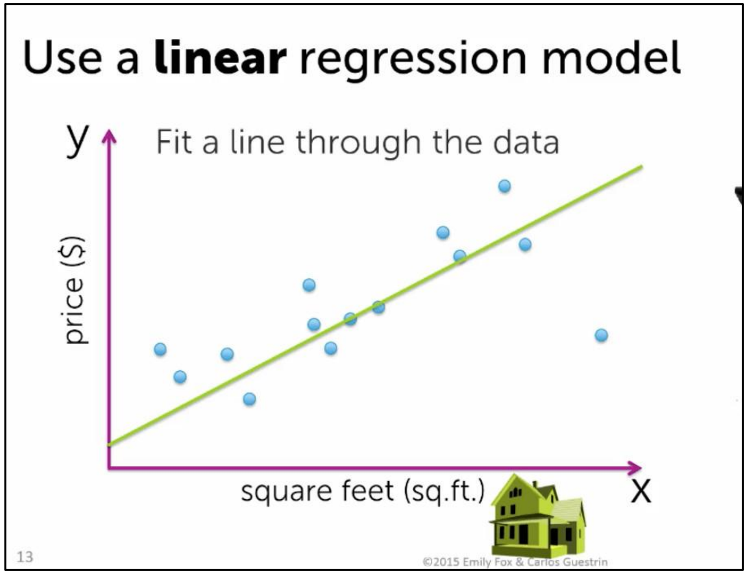
Models of Machine Learning :

- Models of machine learning are algorithms or mathematical frameworks that are used to
  - make predictions (Regression),
  - classify data (Classification)
  - uncover patterns from datasets (Clustering).
- These models can be grouped based on the type of learning they employ (supervised learning, unsupervised learning, reinforcement learning…).

Key Concepts : ML Models

Regression models

- **Regression** is a type of supervised learning in machine learning that focuses on predicting a continuous numerical output based on input data.
- It helps to establish a relationship between dependent (target) and independent (predictor) variables.
- **Example:**
  - Predicting house prices based on features like size, location, and number of bedrooms.
  - Forecasting sales revenue for a company based on past performance.



Key Concepts : ML Models

Types of Regression Models

- Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Ridge and Lasso Regression (Regularized Regression)
- Support Vector Regression (SVR)
- Neural Network Regression

…

Key Concepts : ML Models

Classification models

- **Classification** is a type of supervised learning where the goal is to assign categorical labels (classes) to input data.
- Unlike regression, which predicts a continuous value, classification predicts discrete outcomes.
- **For example:**
- Determining if an email is spam or not (binary classification).
- Predicting the species of a flower based on its features (multiclass classification).

Key Concepts : ML Models

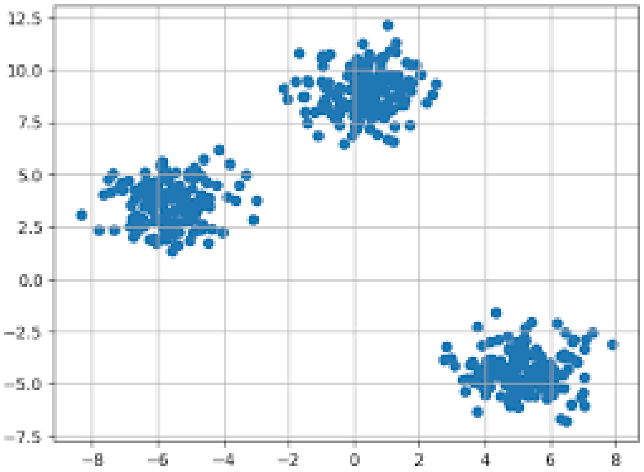
Types of Classification Models

- Logistic Regression
- Naive Bayes
- Support Vector Machines (SVM)
- Decision Trees
- Random Forest
- k-Nearest Neighbors (k-NN)
- …

Key Concepts : ML Models

Clustering models

- **Clustering** is a type of unsupervised learning where the goal is to group data points into clusters (groups) such that points within the same cluster are more similar to each other than to those in other clusters.
- Unlike classification, clustering doesn’t rely on labeled data.
- For example:
  - Grouping customers based on purchasing behavior.
  - Identifying regions of similar vegetation from satellite imagery.



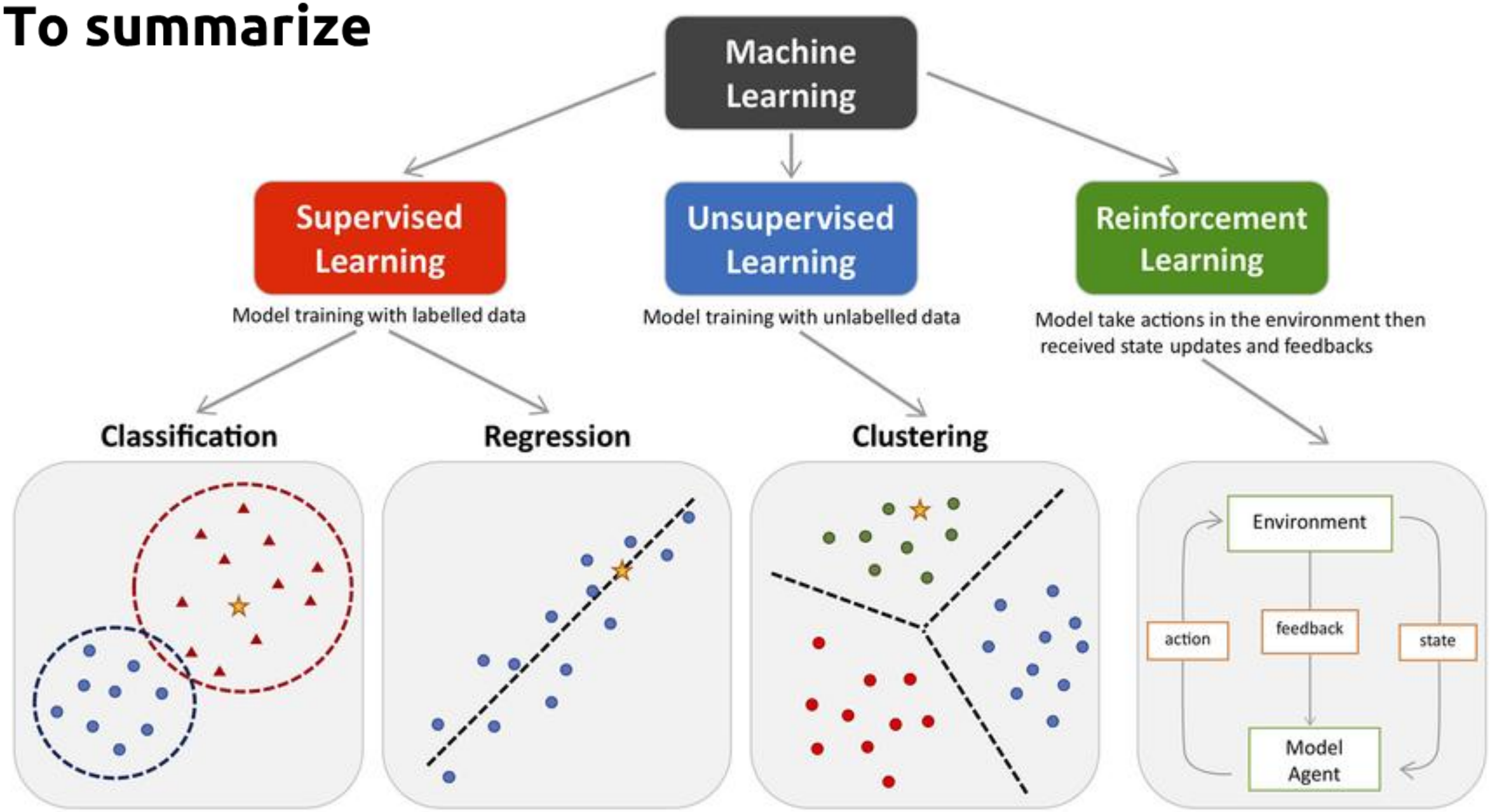
Key Concepts : ML Models

Types of Clustering Models

- K-Means Clustering
- Hierarchical Clustering
- DBSCAN
- Spectral Clustering
- …

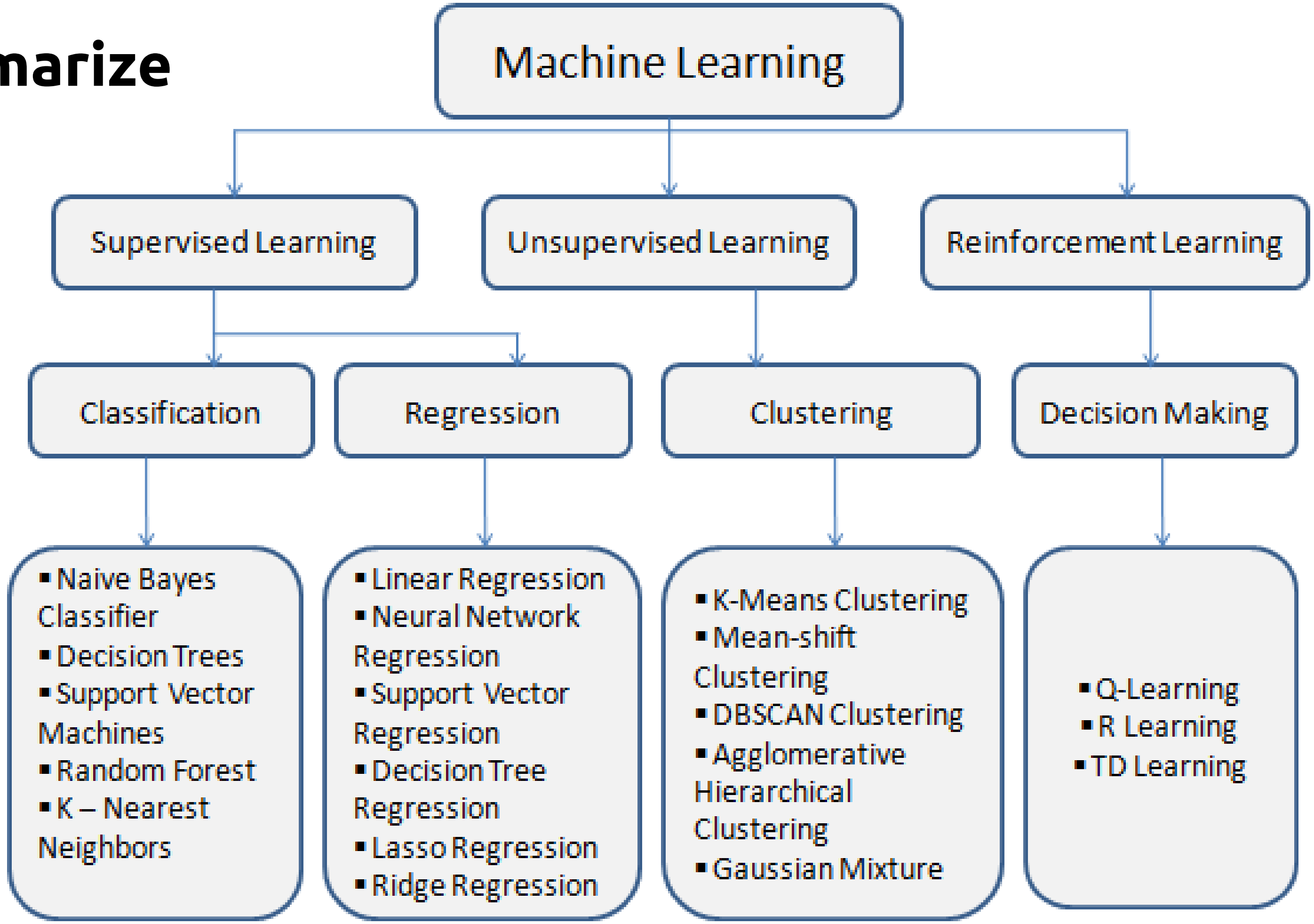
Key Concepts : ML Models

To summarize



Key Concepts : ML Models

To summarize



Key Concepts : Key Components of ML

Key Components of ML

1. **Data**: The raw material for machine learning. Good-quality, representative data is crucial.
2. **Model**: A mathematical representation of the system being studied.
   - Example: Linear regression, neural networks, decision trees…

Key Concepts : Key Components of ML

Key Components of ML

3. **Features**: Input variables or attributes used by the model to make predictions.
   - Example: In predicting housing prices, features could include the number of bedrooms or proximity to schools.
4. **Labels**: Known outputs in supervised learning
   - Example: (e.g., "spam" or "not spam" in an email classifier).
5. **Training and Testing:** Splitting the dataset to train the model on one portion and test its performance on another.

Key Concepts : Core Metrics for Evaluation

Core Metrics for Evaluation

- **Accuracy:** Percentage of correct predictions.
- **Loss:** The difference between predicted and actual values, minimized during training.
- **Precision, Recall, F1 Score:** Metrics for classification tasks, especially in imbalanced datasets.
- **Mean Squared Error (MSE) and Mean Absolute Error (MAE):** Metrics for regression tasks.

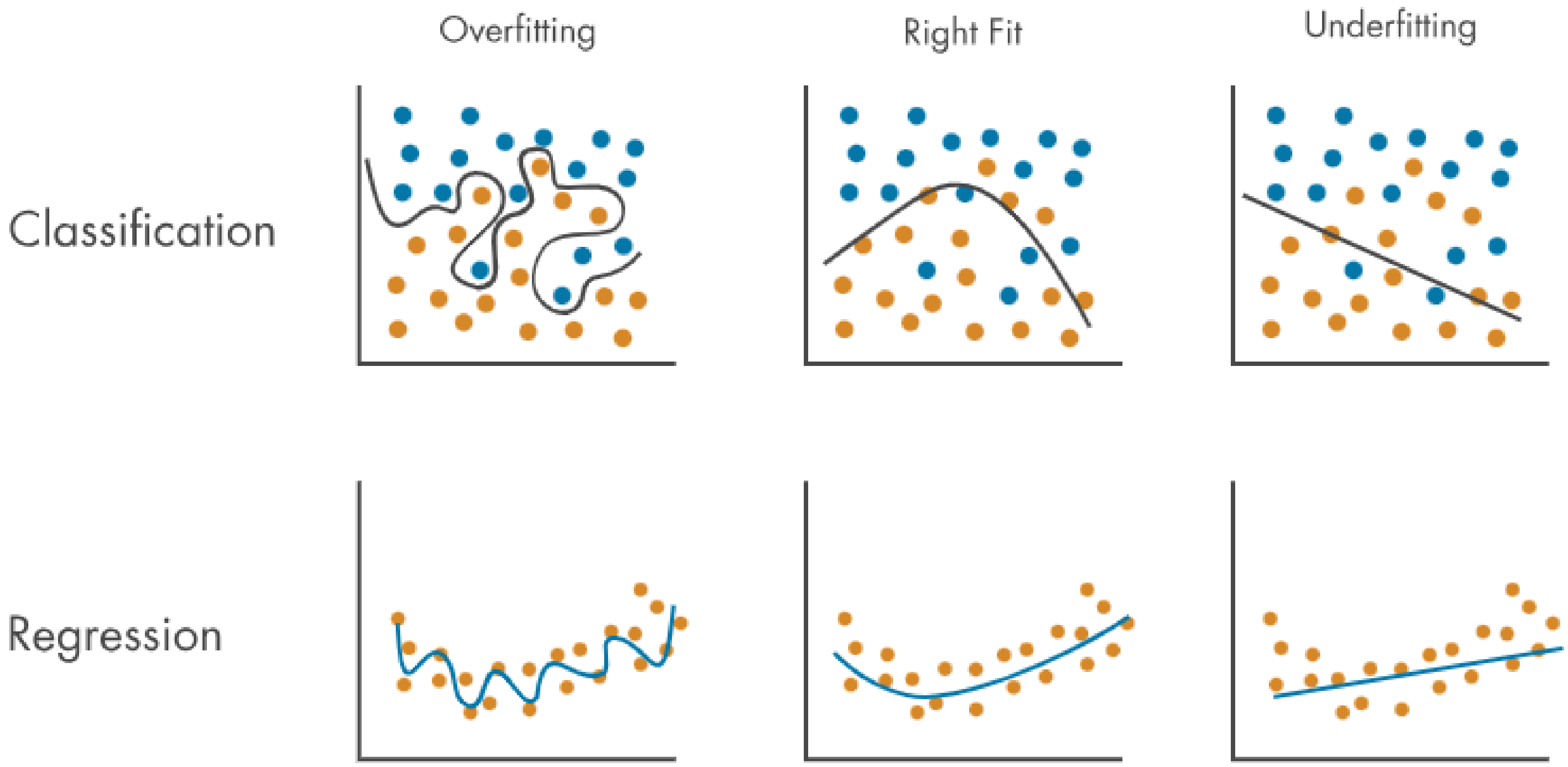
Key Concepts : Overfitting and Underfitting

Overfitting and Underfitting

- **Overfitting:** The model performs well on training data but poorly on unseen data because it learns noise instead of general patterns
  - Solution: Use regularization, more data, or simpler models.
- **Underfitting:** The model fails to capture patterns in the data due to being too simple or poorly trained.
  - Solution: Use a more complex model or tune hyperparameters.

Key Concepts : Overfitting and Underfitting

Overfitting and Underfitting (illustration)



Bias-Variance Tradeoff

- **Bias:** Error due to overly simplistic assumptions in the model. Leads to underfitting.
- **Variance:** Error due to sensitivity to small fluctuations in the training data. Leads to overfitting.
- **The goal is to find the optimal balance.**

Main Titles

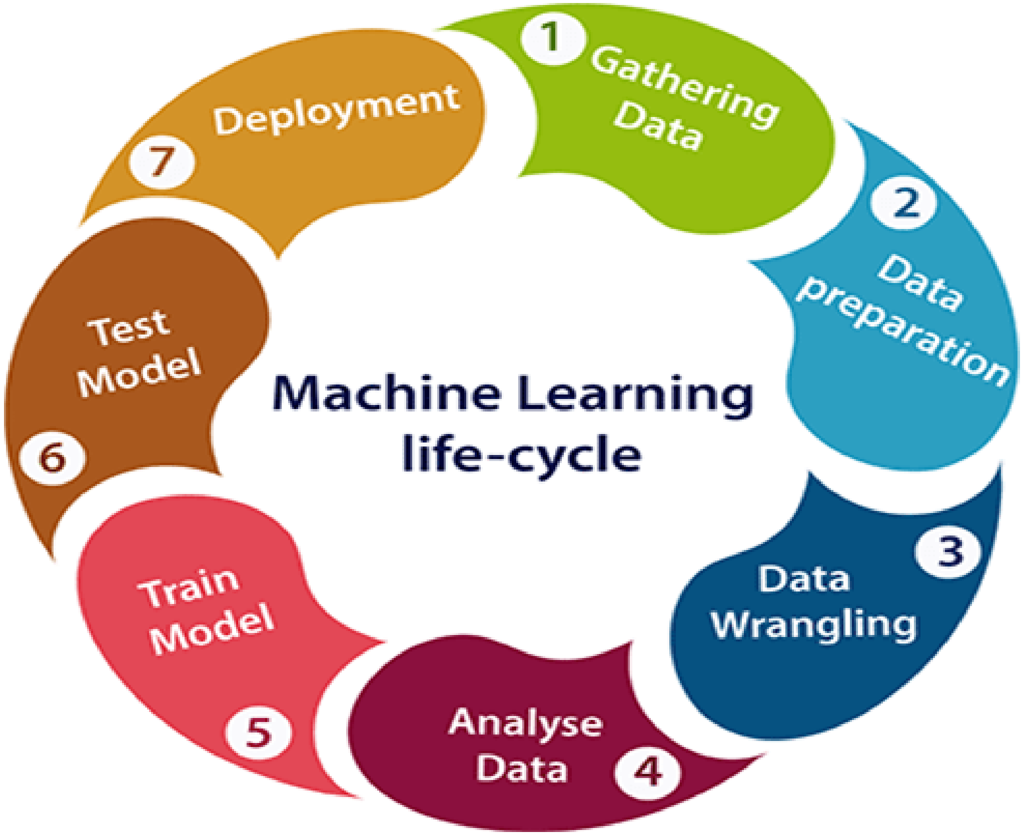
Main Titles

- Introduction to Machine Learning
- History of Machine Learning
- Definitions and Key Concepts
- The Lifecycle of a Machine Learning Project
- Challenges and Limitations of Machine Learning

The Lifecycle of a Machine Learning Project

Lifecycle of a ML Project : Summary

The lifecycle consists of interconnected phases that guide a project from conception to deployment and beyond. Here's a structured breakdown:



Lifecycle of a ML Project : Problem Definition

1- Data Collection

- **Goal:** Collect the data needed to train and evaluate the model.
- **Actions:**
  - Identify data sources (databases, APIs, sensors, CSV files, etc.).
  - Gather relevant data for the problem you want to solve.
  - Ensure the data is sufficient in quantity and quality.
  - Example:
    - For a house price prediction model, collect data about house features (size, number of bedrooms, location) and their prices.

Lifecycle of a ML Project : Data Collection

2- Data Preparation

- **Goal:** Clean and organize the data to make it usable.
- **Actions:**
  - **Data Cleaning**: Remove duplicates, fix errors, and handle missing values.
  - **Formatting**: Convert data into a format compatible with machine learning tools.
  - **Integration**: Combine data from different sources.
  - Example:
    - Remove rows with missing values in the house dataset and convert categorical data (like neighborhood type) into numerical values.

Lifecycle of a ML Project : Data Preprocessing

3. Data Wrangling (Transforming the Data)

- **Goal:** Transform the data to make it suitable for training the model.
- **Actions:**
  - **Feature Engineering**: Create new variables (features) from existing data (e.g., calculate the age of a house based on its construction year).
  - **Normalization/Standardization**: Scale the data to a consistent range (e.g., normalize prices between 0 and 1).
  - **Encoding**: Convert categorical variables into numerical values (e.g., using one-hot encoding).
  - Example: create a new feature like "price per square meter" to improve the model's prediction accuracy.

Lifecycle of a ML Project : Data Preprocessing

4- Analyze Data (Exploring the Data)

- **Goal:** Understand the data and identify patterns.
- **Actions:**
  - **Visualization**: Use graphs (histograms, scatter plots, etc.) to explore the data.
  - **Statistical Analysis**: Calculate metrics like mean, median, correlation, etc.
  - **Anomaly Detection**: Identify outliers or potential errors in the data.
  - Example: Visualize the relationship between house size and price to check for correlation.

Lifecycle of a ML Project : Data Preprocessing

5. Train Model

- **Goal:** Train a machine learning model using the prepared data.
- **Actions:**
  - **Algorithm Selection**: Choose an algorithm suitable for the problem (e.g., linear regression, decision trees, neural networks).
  - **Data Splitting**: Divide the data into training and test sets.
  - **Training**: Use the training set to adjust the model's parameters.
  - Example: Use a linear regression algorithm to predict house prices based on their features.

6. Test Model (Evaluating the Model)

- **Goal:** Evaluate the model's performance on unseen data.
- **Actions:**
  - **Prediction**: Use the test set to make predictions with the trained model.
  - **Performance Metrics**: Calculate metrics like Mean Squared Error (MSE), accuracy, or F1-score.
  - **Adjustment**: If performance is unsatisfactory, revisit earlier steps to improve the model (e.g., tuning hyperparameters or collecting more data).
  - Example: Evaluate the house price prediction model by calculating the average error between predicted and actual prices.

7. Deployment

- **Goal:** Put the model into production so it can be used in real-time.
- **Actions:**
  - **Integration**: Integrate the model into an existing application or system (e.g., a web or mobile app).
  - **Scaling**: Ensure the model can handle a large number of requests (scalability).
  - **Monitoring**: Monitor the model's performance in production and update it if necessary (e.g., in case of data drift).
  - Example: Deploy the house price prediction model on a website where users can input house features and get a price estimate.

Main Titles

- Introduction to Machine Learning
- History of Machine Learning
- Definitions and Key Concepts
- **The Lifecycle of a Machine Learning Project**
- Challenges and Limitations of Machine Learning

Main Titles

- Introduction to Machine Learning
- History of Machine Learning
- Definitions and Key Concepts
- The Lifecycle of a Machine Learning Project
- **Challenges and Limitations of Machine Learning**

Challenges and Limitations of Machine Learning

Technical Challenges

- Insufficient or low-quality data.
- Overfitting and underfitting.
- Interpretability of complex models (e.g., neural networks).

Ethical Challenges

- Bias in data and algorithms.
- Privacy and data security issues

Practical Challenges

- High computational cost. Deployment and maintenance of models in production environments.

Bibliography

Bibliography

Books:

- "Machine Learning Yearning" by Andrew Ng (for understanding project workflows).
- "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron.
- "Pattern Recognition and Machine Learning" by Christopher M. Bishop (for theoretical foundations).

Courses:

- Coursera: "Machine Learning" by Andrew Ng.
- edX: "Principles of Machine Learning" by Microsoft.
- Fast.ai: Practical deep learning courses.

Bibliography

Documentation:

- Scikit-Learn Documentation [Link].
- Google Machine Learning Crash Course [Link].
- Websites:
  - Kaggle: For practice and competitions [Link].
  - Google Dataset Search : Search engine for finding datasets across the web [Link].
  - UCI Machine Learning Repository : A collection of datasets for machine learning experiments [Link].

Thank you for your attention…