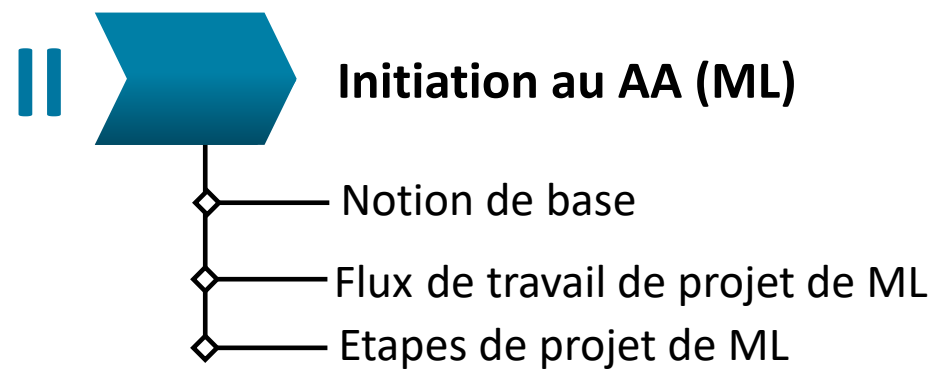
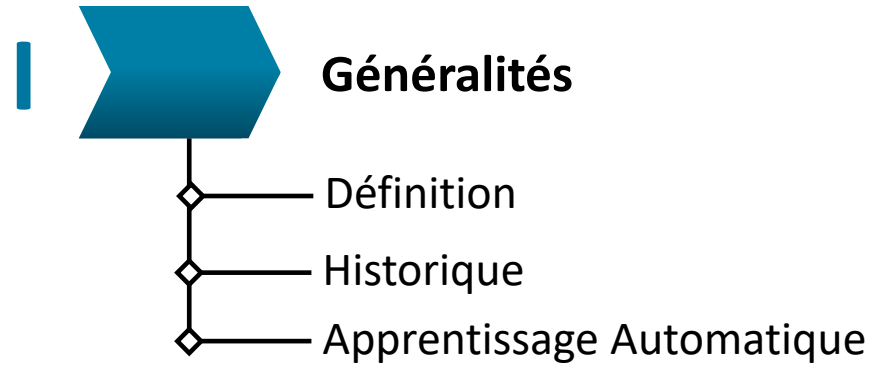




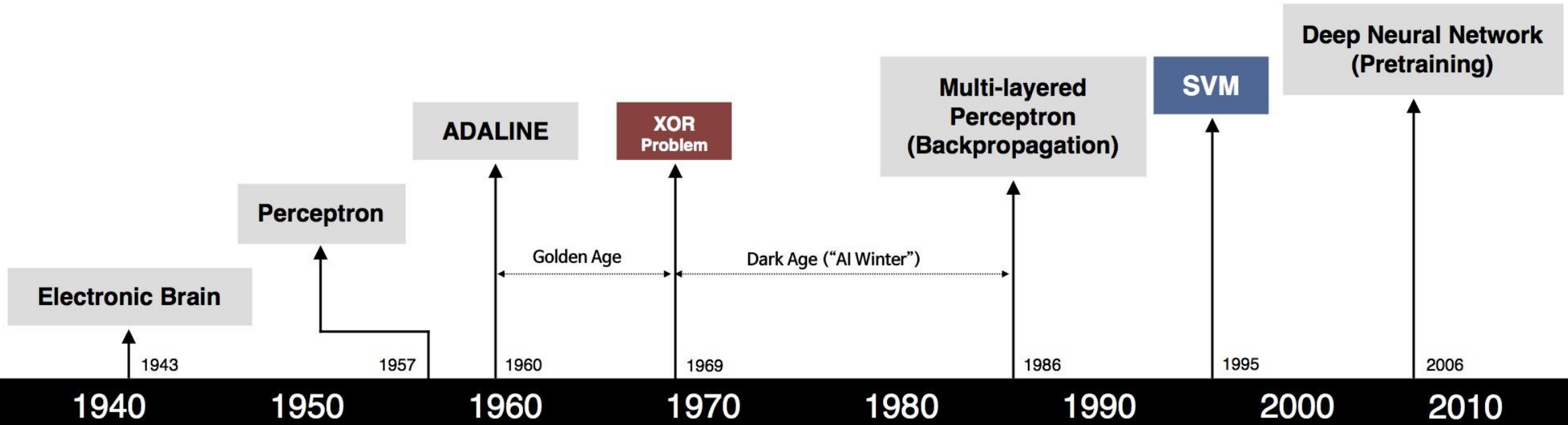
# Organisation de la présentation



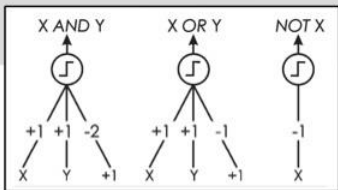
« *L'intelligence artificielle (IA) désigne **la simulation de l'intelligence humaine** dans des machines programmées pour **penser et apprendre comme les humains**. L'IA englobe diverses technologies et approches, notamment **l'apprentissage automatique, le deep learning, le traitement du langage naturel, la vision par ordinateur, etc.**, afin de permettre aux ordinateurs d'accomplir des tâches qui nécessitent généralement une intelligence humaine, comme **comprendre le langage naturel, reconnaître des modèles, prendre des décisions et résoudre des problèmes complexes** ».*

- Titre : "Intelligence Artificielle : Une Approche Moderne"  
Auteurs : Stuart Russell et Peter Norvig Édition : Troisième édition  
Éditeur : Pearson Année de publication : 2009.





S. McCulloch – W. Pitts



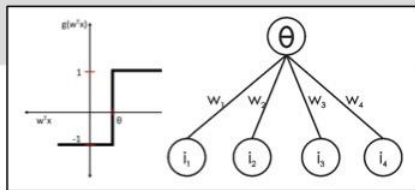
- Adjustable Weights
- Weights are not Learned



F. Rosenblatt



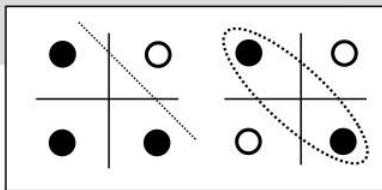
B. Widrow – M. Hoff



- Learnable Weights and Threshold



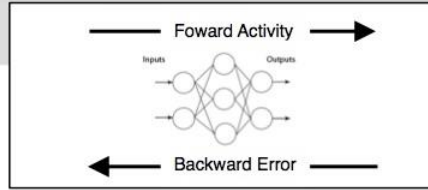
M. Minsky – S. Papert



- XOR Problem



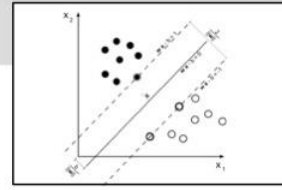
D. Rumelhart – G. Hinton – R. Williams



- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting



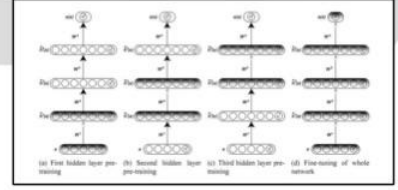
V. Vapnik – C. Cortes



- Limitations of learning prior knowledge
- Kernel function: Human Intervention

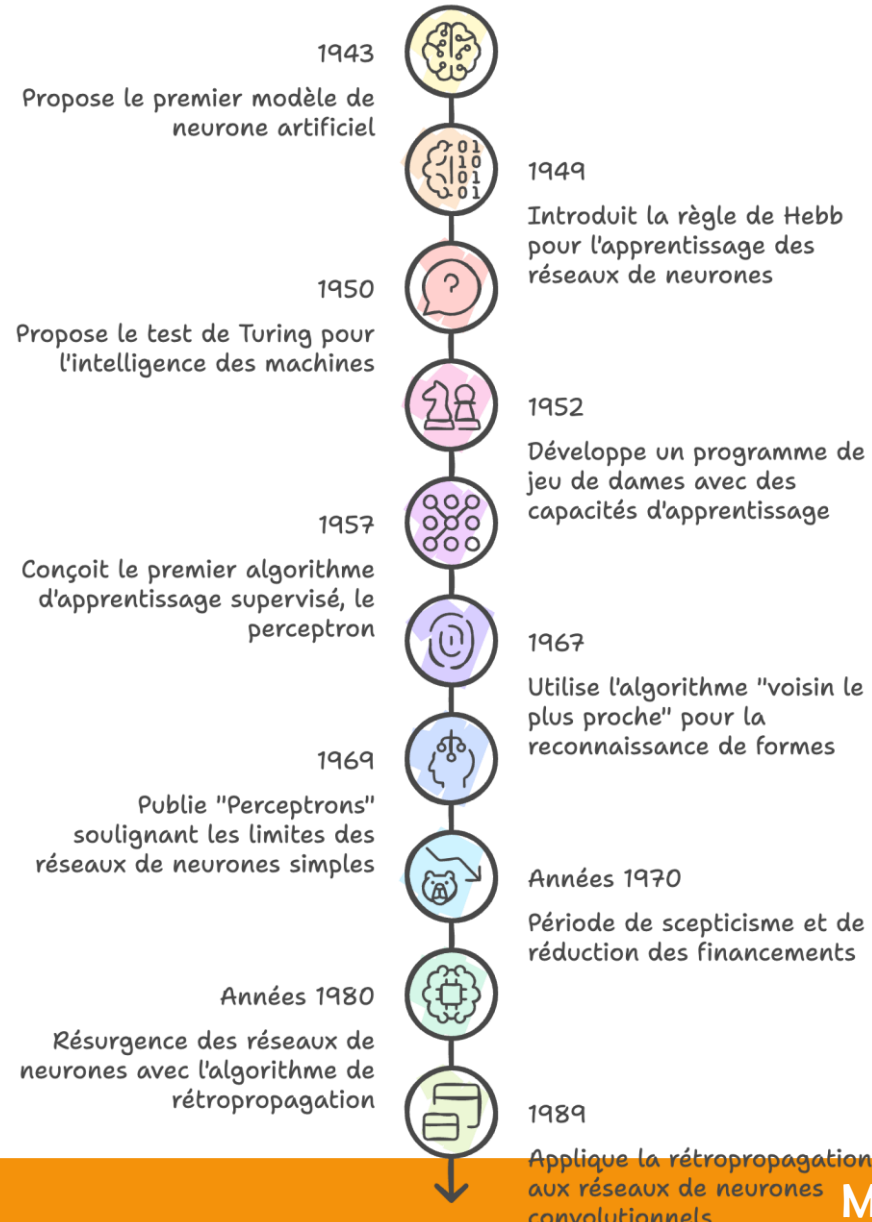


G. Hinton – S. Ruslan

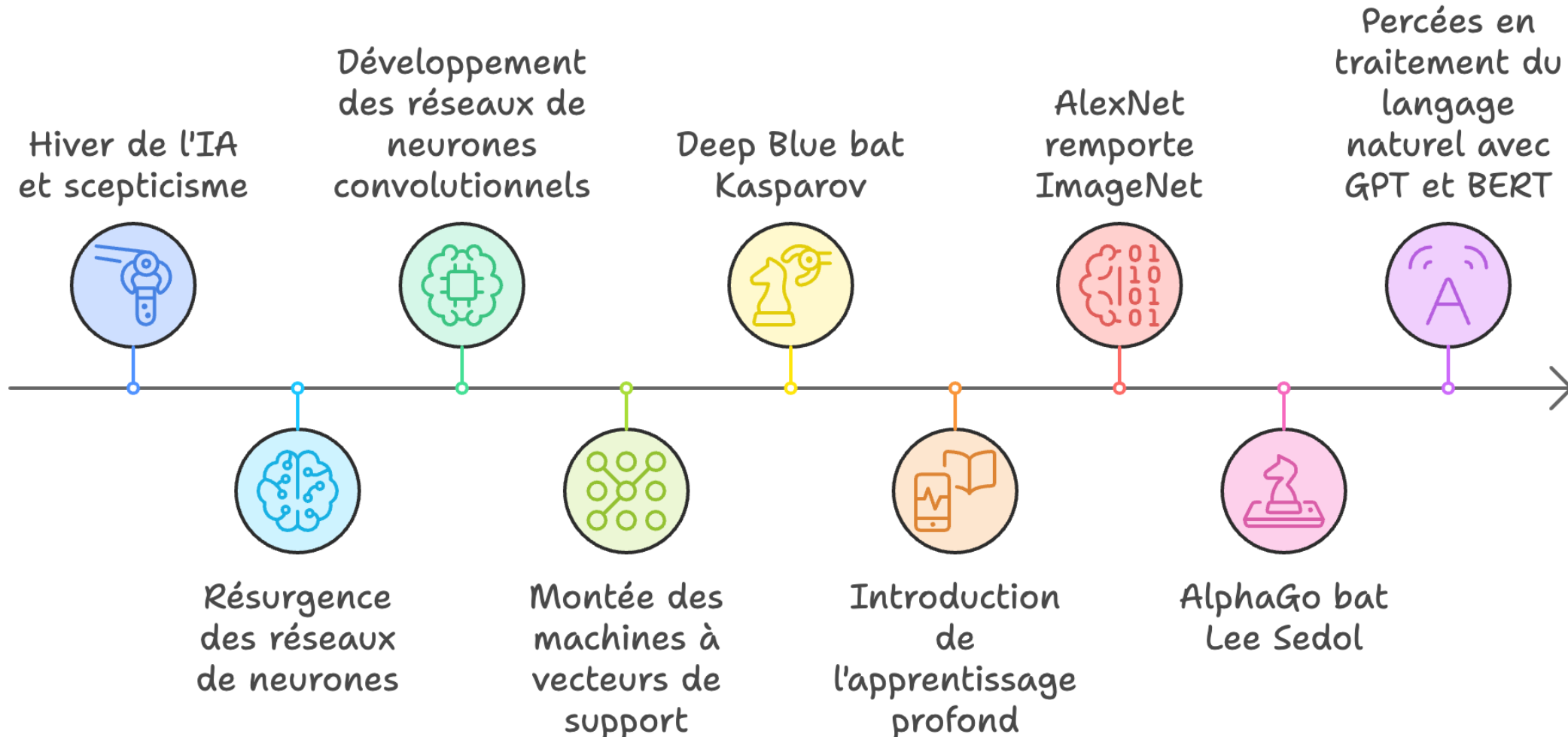


- Hierarchical feature Learning

## Histoire de l'intelligence artificielle



## Renaissance et nouvelles approches (1980-2000)

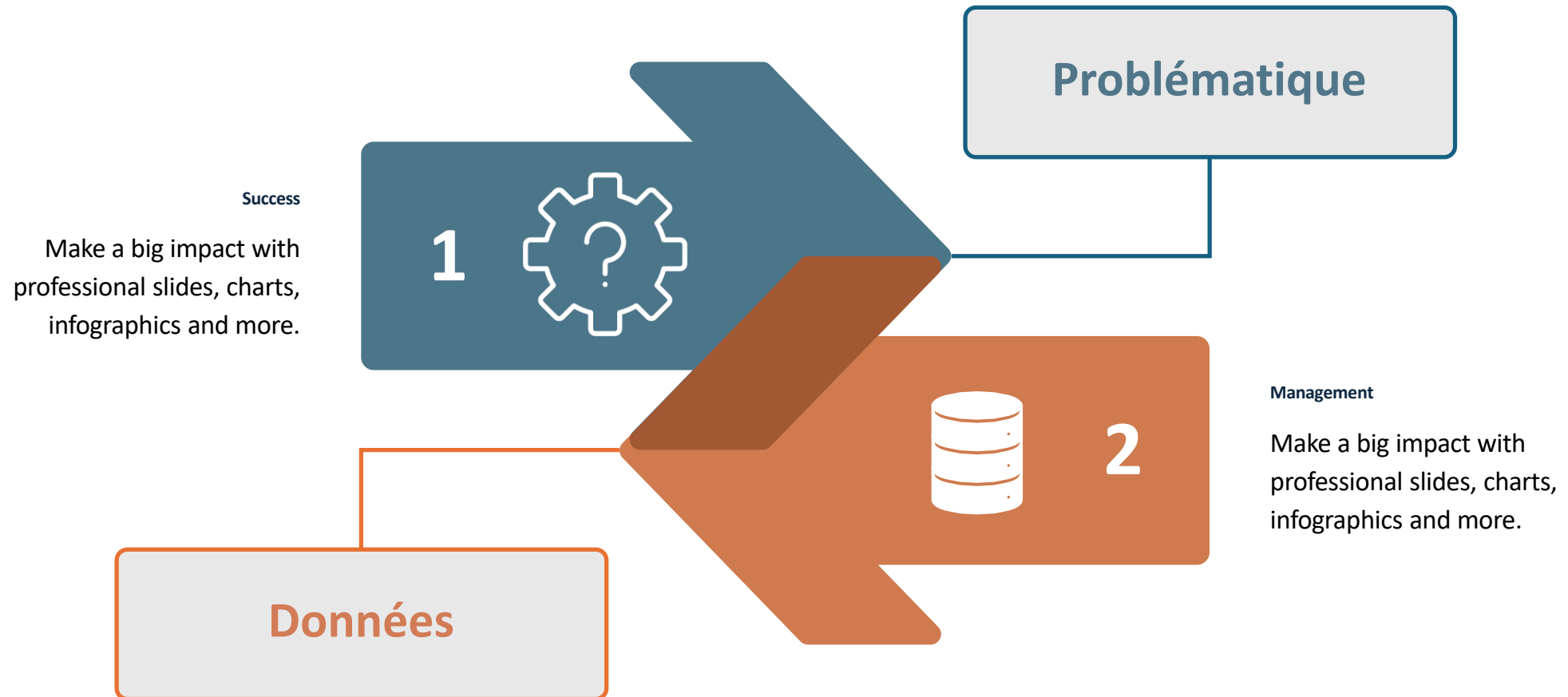


Le premier objectif du Data Scientist est de produire des méthodes (automatisées, autant que possible) de tri et d'analyse de données, afin d'en extraire des informations utiles.

Le métier de Data Scientist est apparu et a pu émerger pour trois raisons principales:

1. Explosion de données
2. Algorithmes améliorés et plus accessibles
3. Ordinateurs puissants (capacité augmente de manière exponentielle)

Pour résoudre un problème avec **la science des données** on a besoin de :



Quelques exemples

1. Prédire les ventes
2. Segmenter des utilisateurs
3. Identifier une image

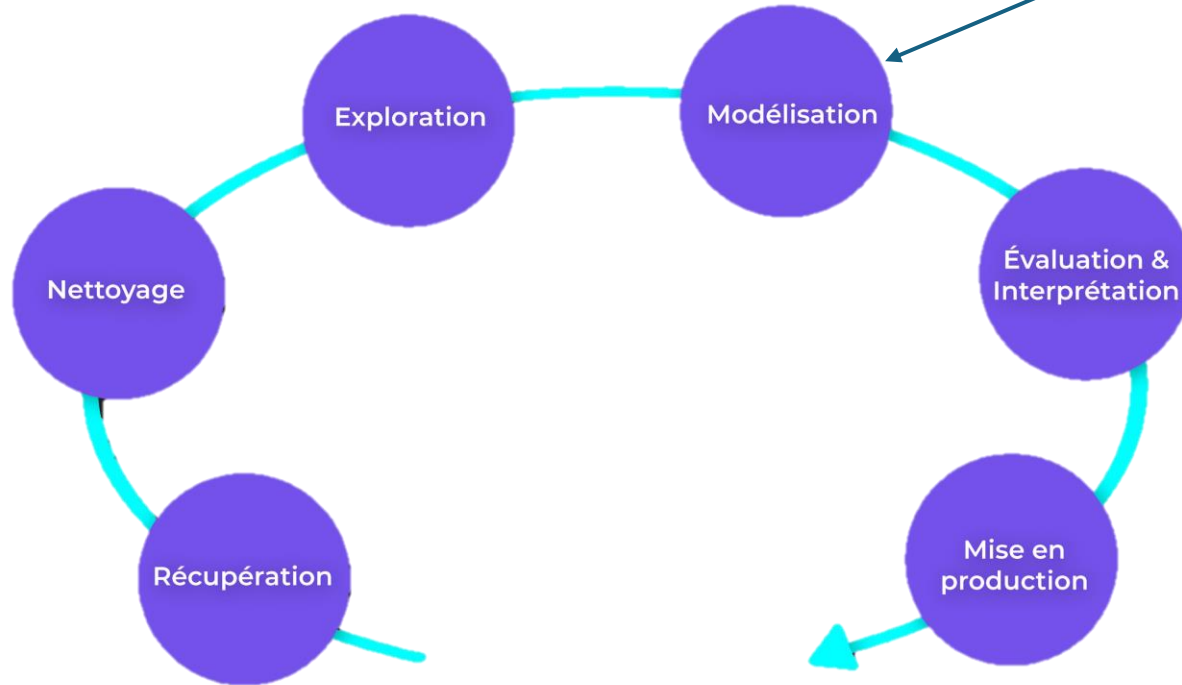
**En quoi la data science se distingue-t-elle de l'analyse de données classique ? A-t-on vraiment besoin de ce nouveau domaine ?**

## Data Science vs data analysis

- La **science de données** se distingue dans un premier temps par l'approche utilisée pour résoudre la problématique - **la notion d'apprentissage** - que l'on va étudier dans la suite de ce cours.
- **L'analyse de données** ("data analysis"), elle, utilise des méthodes issues des statistiques classiques comme les estimateurs, les analyses de corrélations ou encore la régression linéaire.

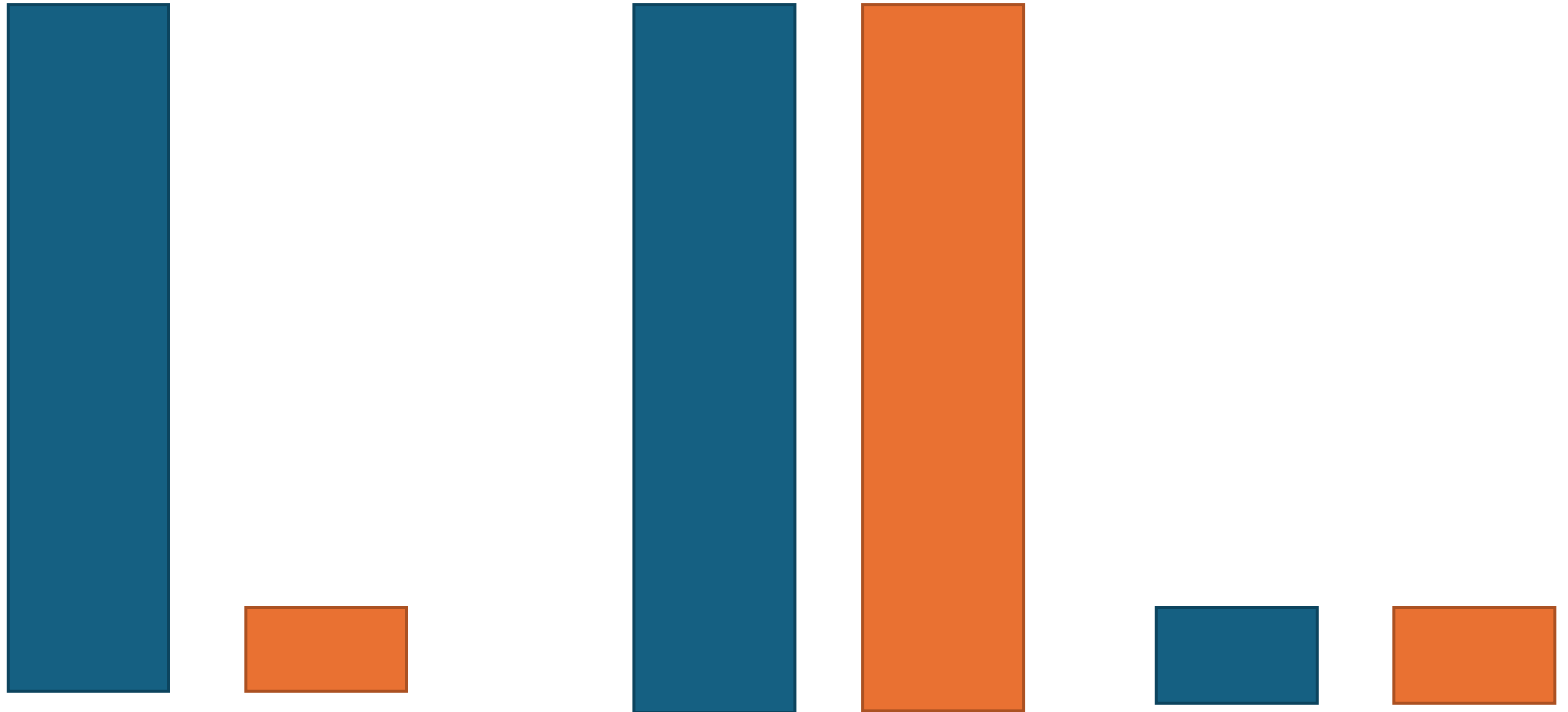
## Data Science vs data analysis

La data science est un nouveau domaine de travail, qui augmente les capacités d'analyse classique, afin d'aider les entreprises à prendre des décisions informées. Elle s'appuie pour cela sur des données utiles et ne peut s'appliquer que dans certaines problématiques précises.

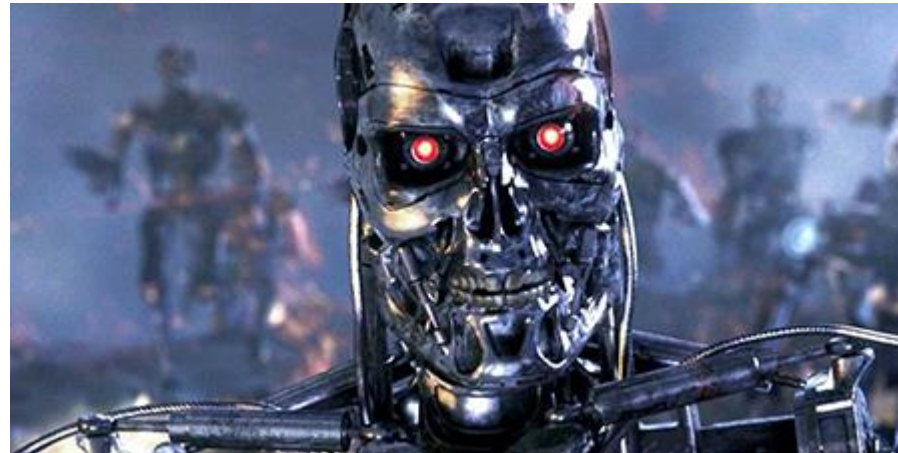


Au sein du cycle de travail du data scientist, le **machine learning** désigne l'ensemble des méthodes de **modélisation statistique à partir des données**.

Les meilleurs data scientists ne sont pas ceux qui connaissent les algorithmes les plus complexes, mais **ceux qui ont une très bonne connaissance des données** et ont préparé le terrain avec soin en amont.



Lorsqu'on entend parler de machine learning – ou plus généralement de l'intelligence artificielle, dont le machine learning est un sous-domaine – on pense généralement à ça :

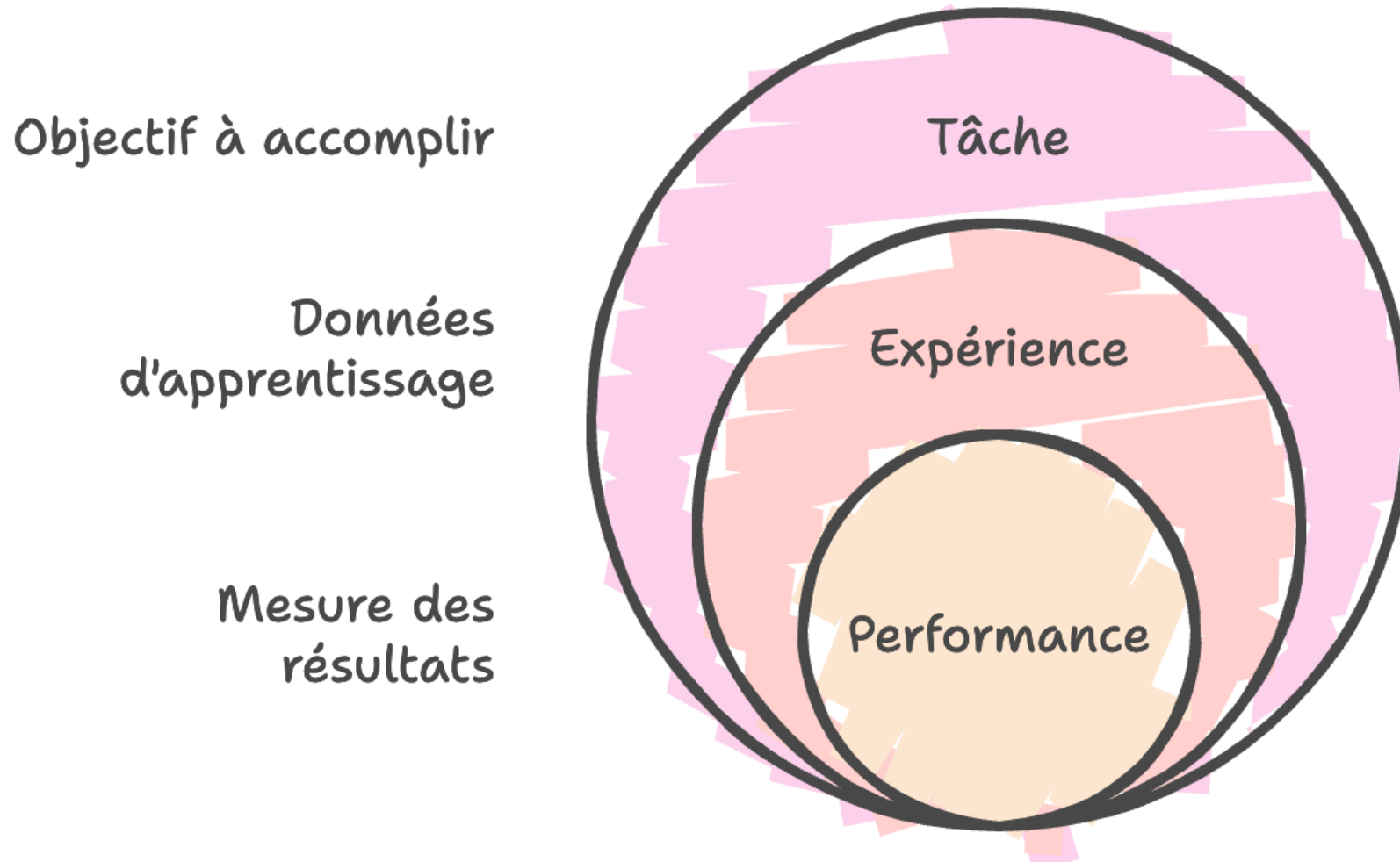


**Démystifions le machine learning**

Le machine learning est l'apprentissage d'un modèle statistique par la machine grâce à des données d'entraînement. Un problème de machine learning comporte plusieurs éléments spécifiques :

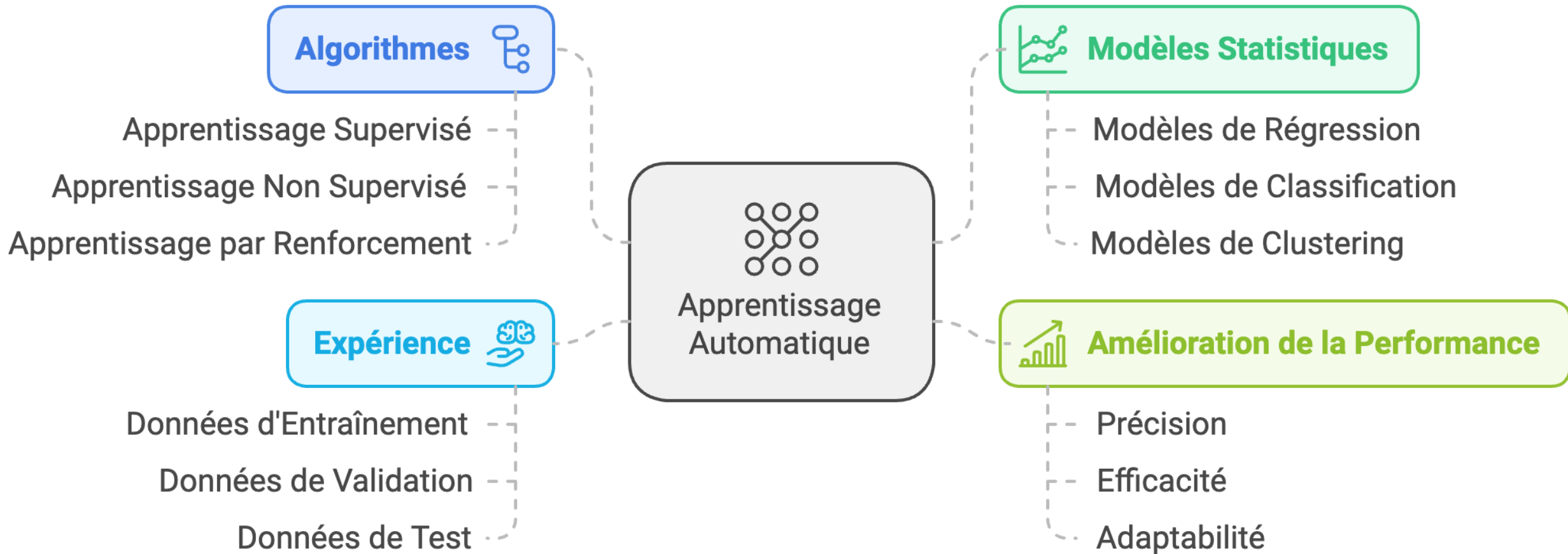
- Des données (les exemples) ;
- Une tâche à accomplir [Un objectif (ce qu'on veut prédire ou classifier)];
- Un modèle/ un algorithme d'apprentissage (la façon d'apprendre)
- Une mesure des performances.

# Composants de l'apprentissage automatique

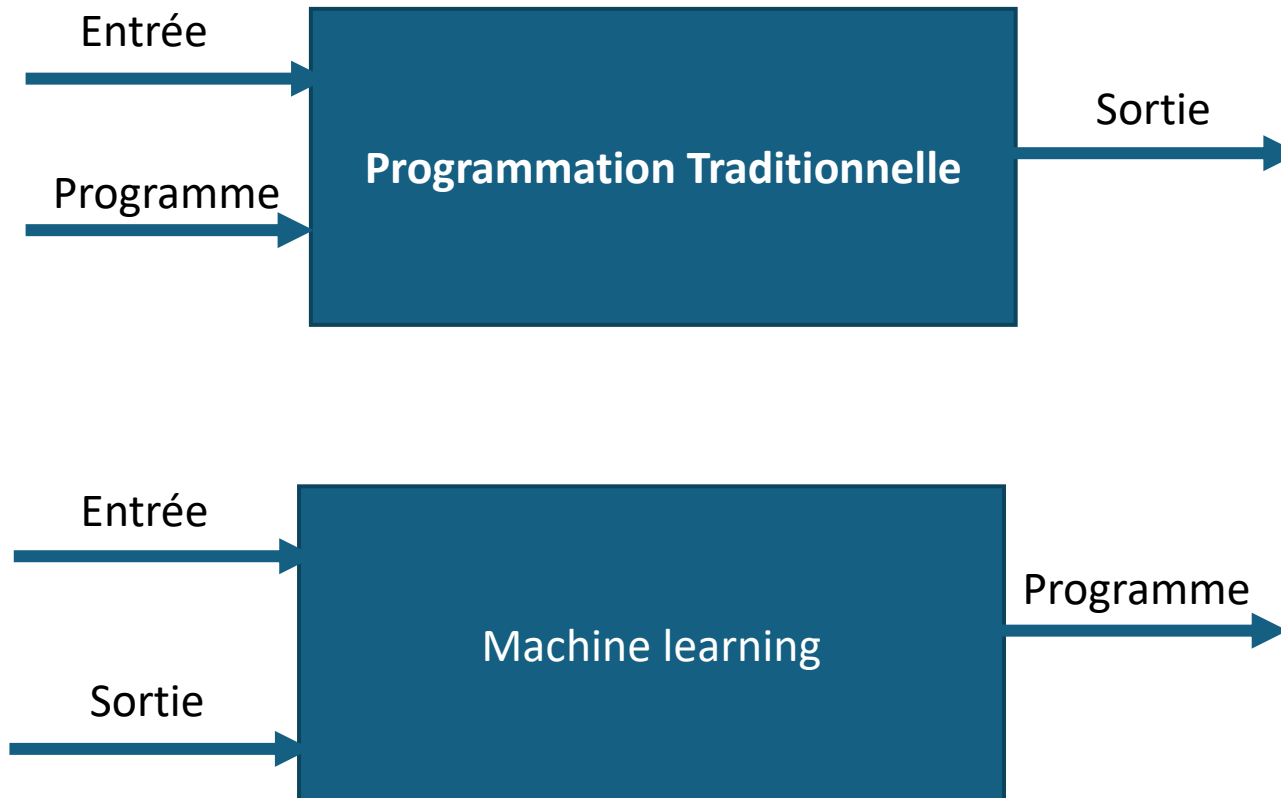


En une phrase, un ordinateur **apprend** à partir de **données** pour résoudre une **tâche (T)** en faisant attention à **mesurer les performances (P)**. S'il **améliore** les performances sur cette tâche lorsqu'on lui fournit les données d'entraînement (**Expérience E**), on dit alors qu'il **apprend**.

L'apprentissage automatique est un domaine de l'informatique qui se concentre sur le développement d'**algorithmes** et de **modèles statistiques** permettant à des systèmes informatiques **d'améliorer leurs performances** sur une tâche spécifique à travers **l'expérience**, sans être explicitement programmés pour chaque scénario.



## Programmation traditionnelle Vs Programmation Machine Learning

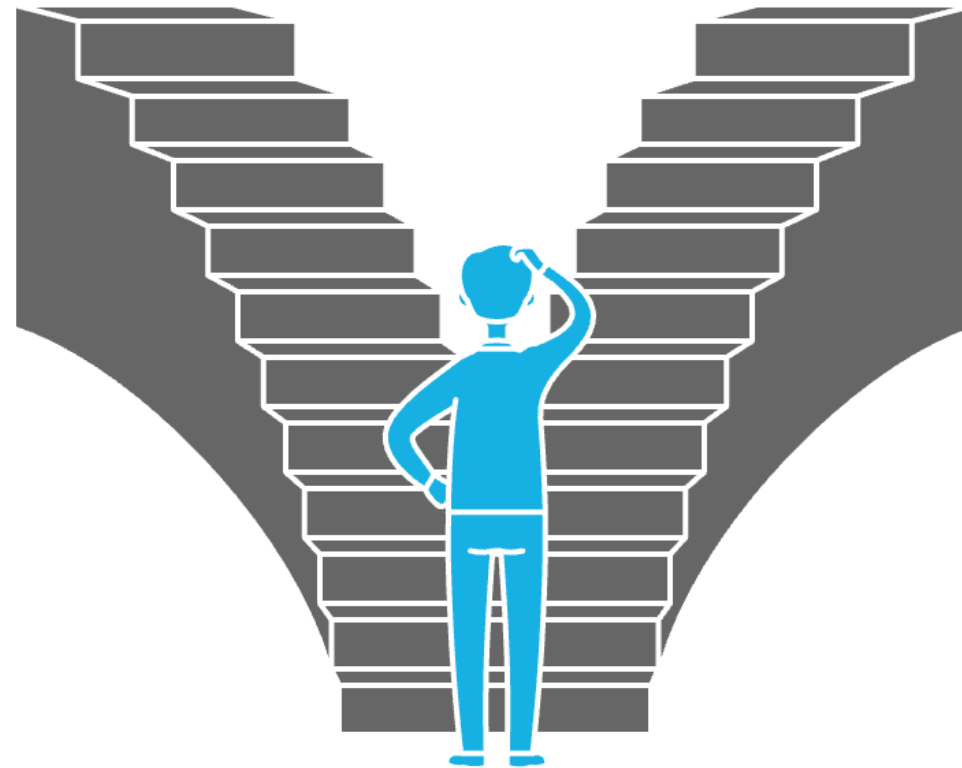


## Programmation traditionnelle

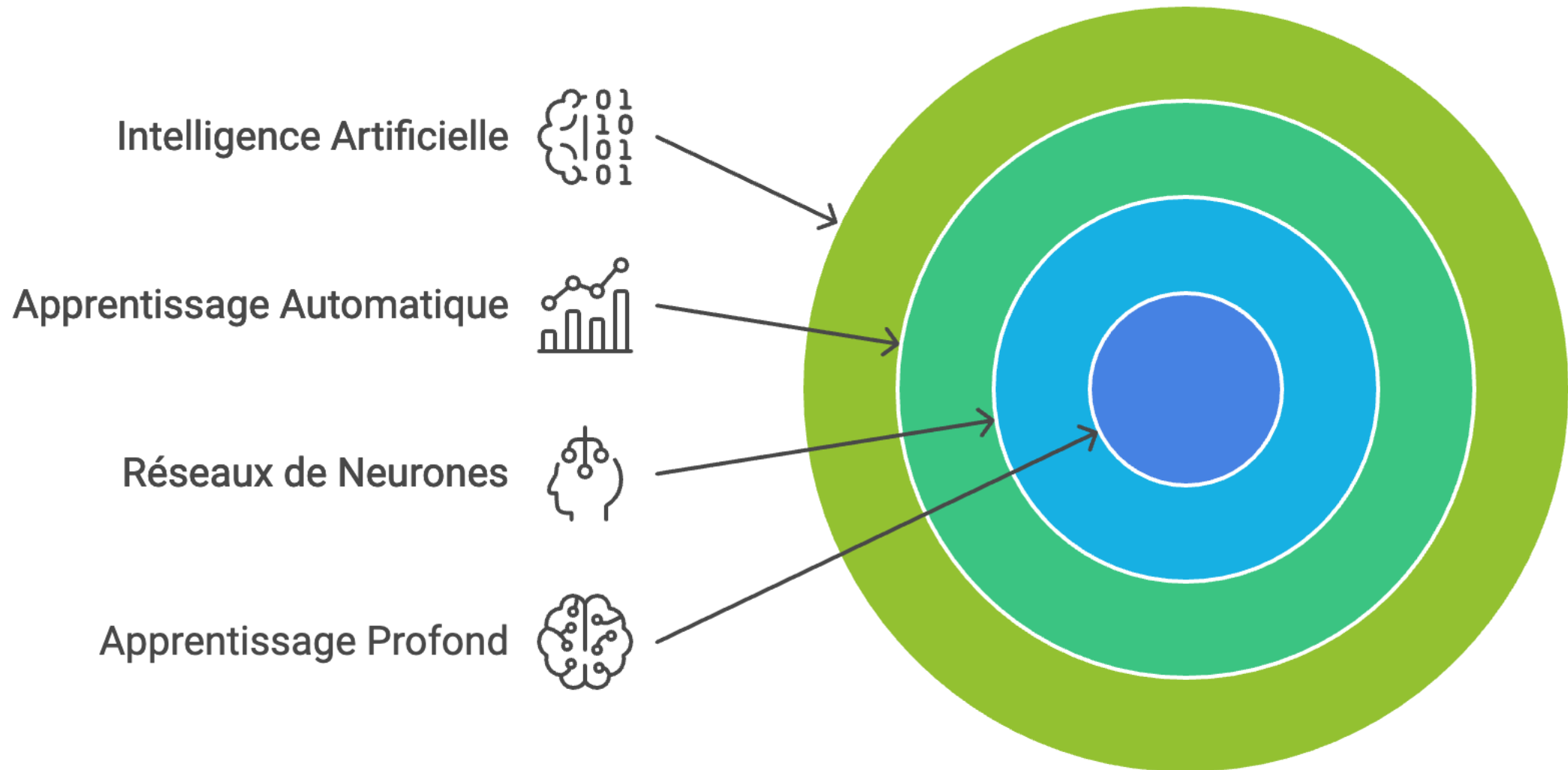
On dit à l'ordinateur  
"COMMENT" faire.

## Machine Learning

On montre à l'ordinateur  
"QUOI" faire à travers  
des exemples.



## Hiérarchie de l'Intelligence Artificielle



### Applications Diverses de l'IA

Reconnaissance  
d'images et de la  
parole



Recommandations  
personnalisées



Véhicules  
autonomes



Prédiction de  
tendances

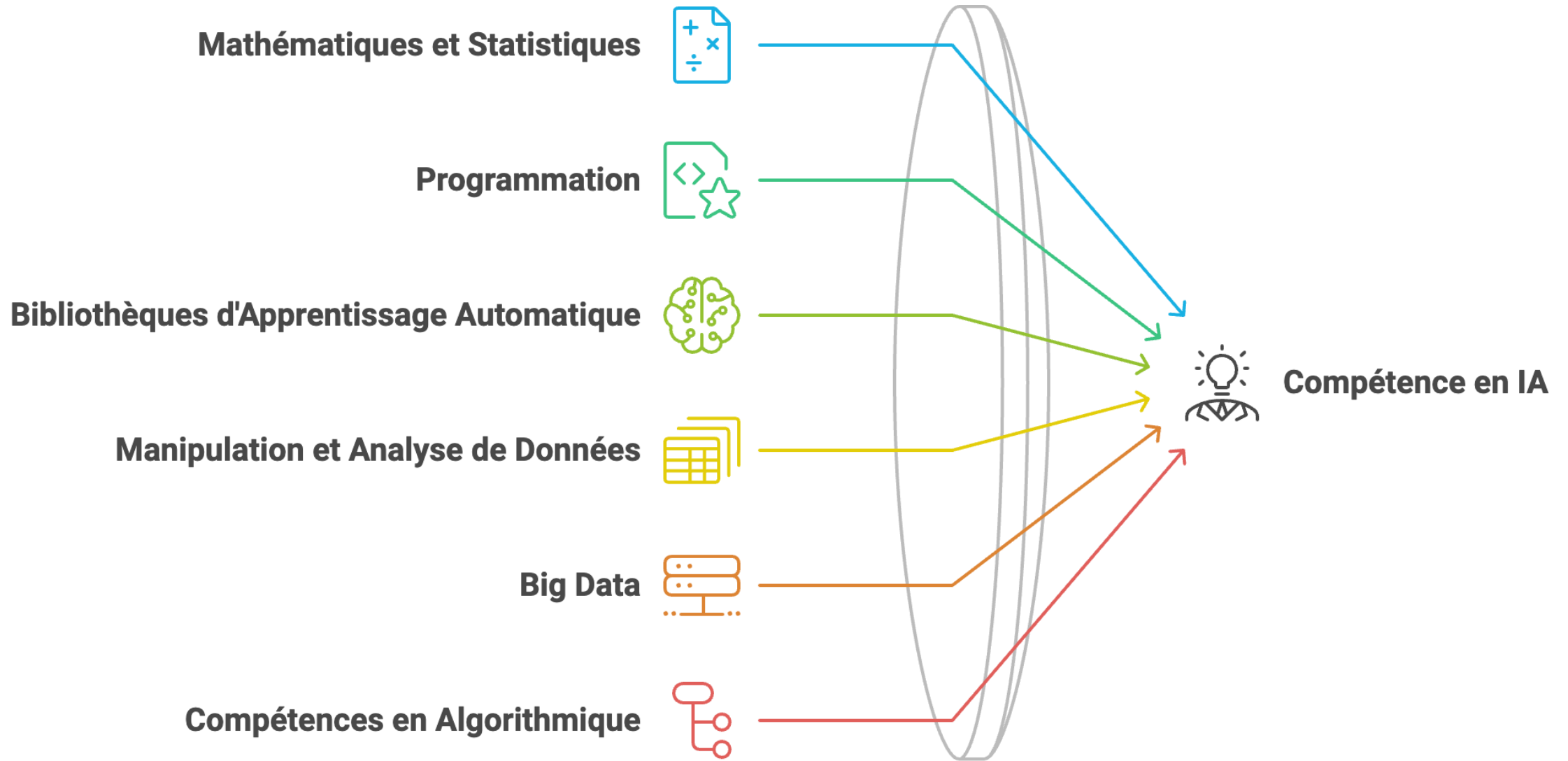


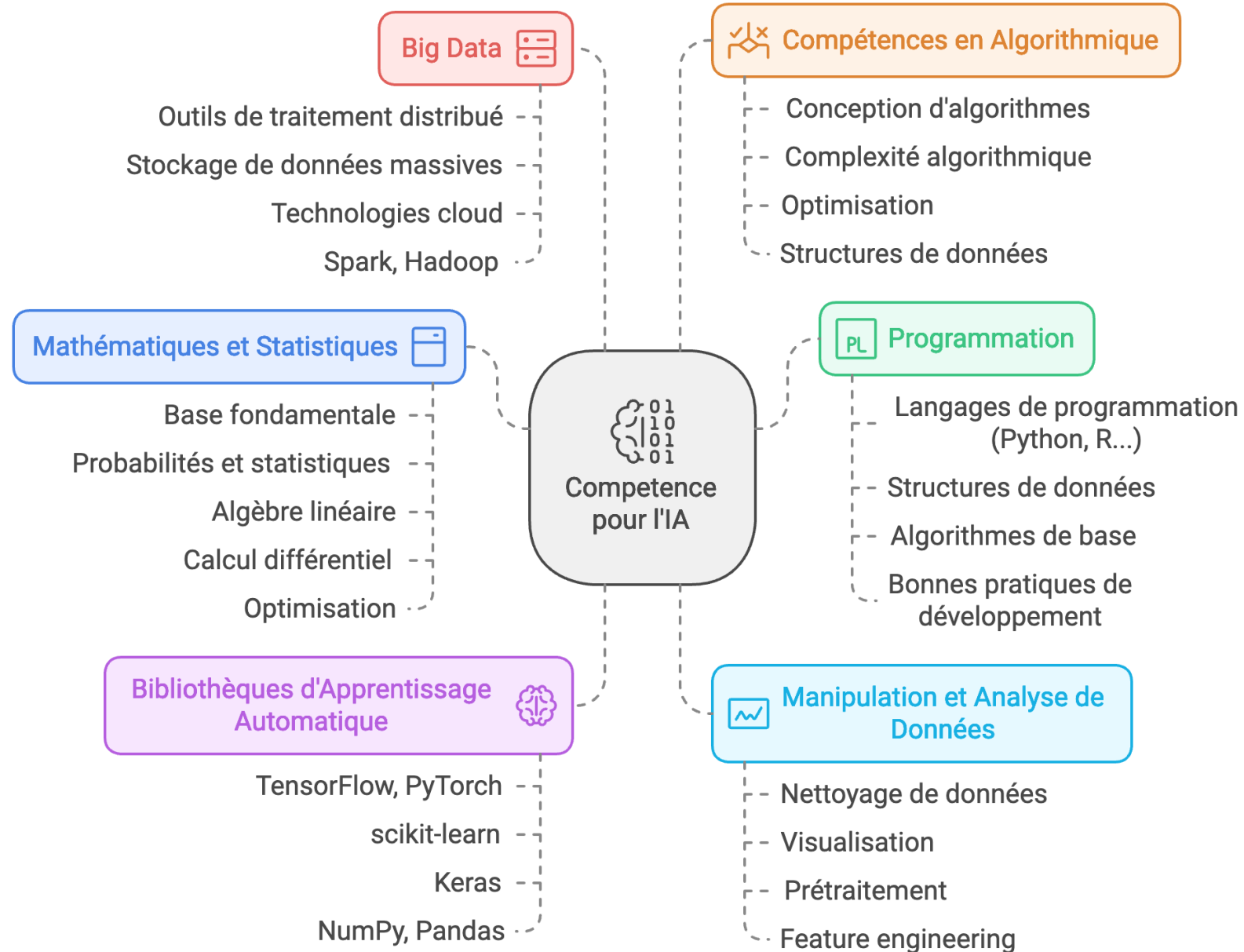
Diagnostic  
médical



Agriculture







## Flux de travail du projet de Machine Learning



### Flux de travail de projet de ML

#### Problem definition

Problem statement

Data science problem formulation

#### Data collection

Raw data

Data storage



#### Data preparation

Transformation

Data cleaning

Normalization

Feature engineering

#### Model training

Training

Model selection

Hyper-parameter tuning

Model evaluation

#### Model deployment

Deployment

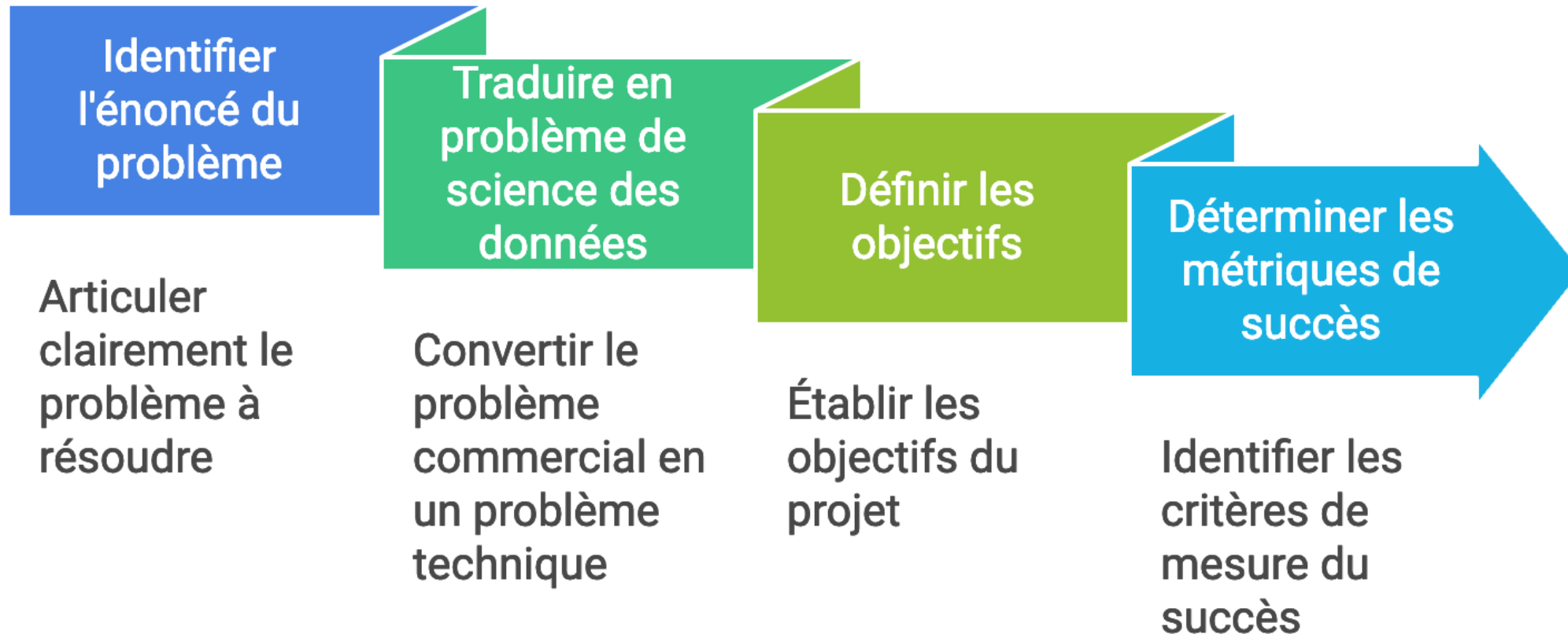
Prediction

Performance monitoring

Retraining

# Flux de travail de projet de ML : Etape 1

## Processus de définition du problème



## Flux de travail de projet de ML : Etape 2

### Collecte et stockage des données complets



# Flux de travail de projet de ML : Etape 2 Sources de collecte de données

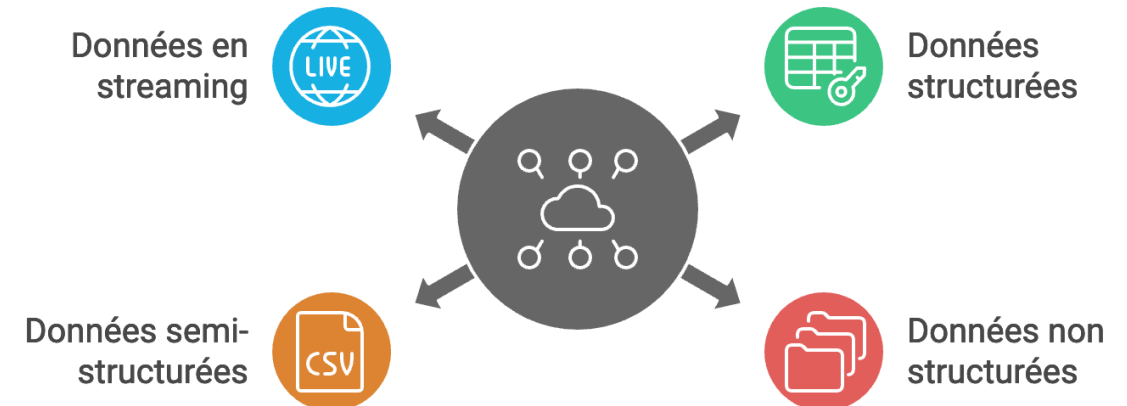
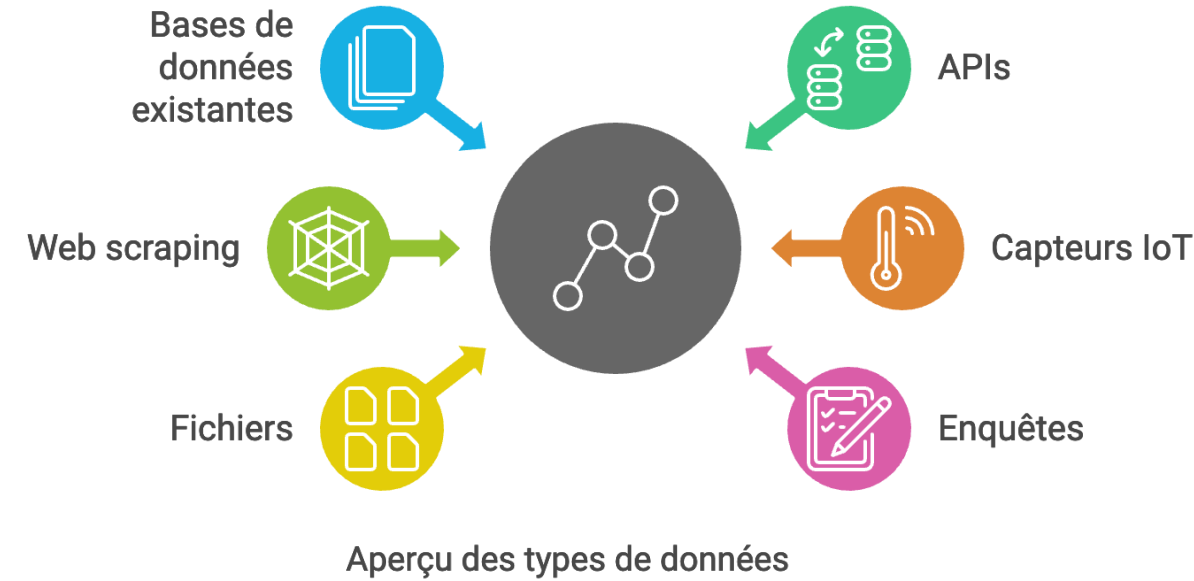
## Raw Data (Données brutes)

### • Sources possibles :

- Bases de données existantes
- APIs publiques ou privées
- Web scraping
- Capteurs IoT
- Fichiers (CSV, JSON, XML...)
- Enquêtes et formulaires
- Données open source

### • Types de données :

- Structurées (tables, bases de données)
- Non structurées (textes, images, vidéos)
- Semi-structurées (JSON, XML)
- Streaming (données en temps réel)



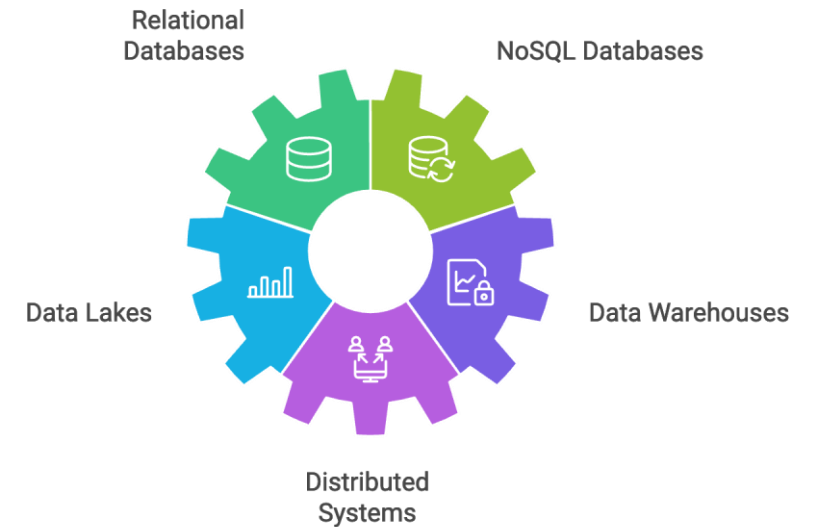
## Flux de travail de projet de ML : Etape 2

### Data Storage (Stockage des données)

#### Solutions de stockage :

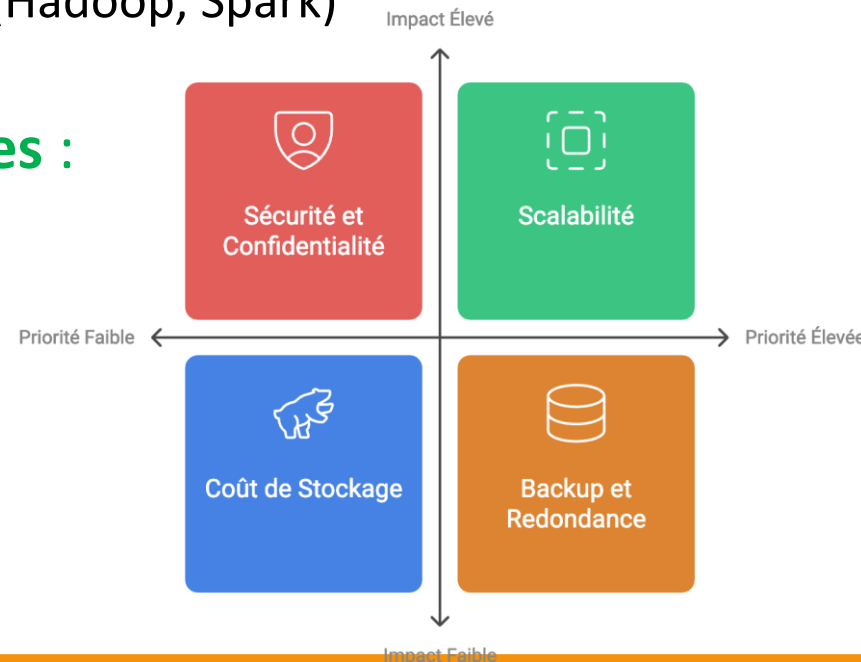
1. Bases de données relationnelles (MySQL, PostgreSQL)
2. Bases NoSQL (MongoDB, Cassandra)
3. Data Lakes (Amazon S3, Azure Blob)
4. Data Warehouses (Snowflake, BigQuery)
5. Systèmes distribués (Hadoop, Spark)

Overview of Data Storage Solutions



### Considérations importantes :

- Volume des données
- Vitesse d'accès nécessaire
- Sécurité et confidentialité
- Coût de stockage
- Scalabilité
- Backup et redondance
- Conformité (RGPD, etc.)

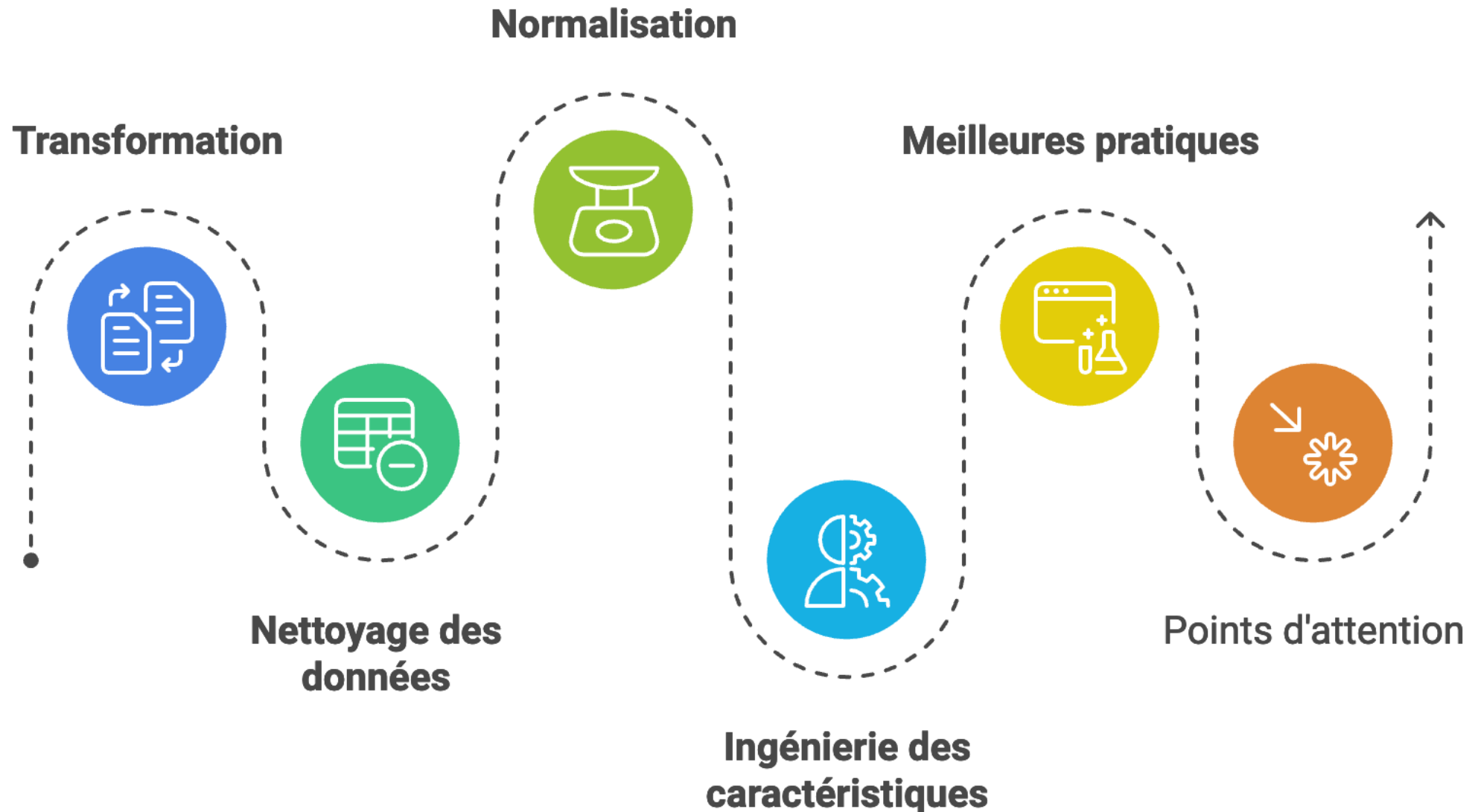


## Flux de travail de projet de ML : Etape 2

### Défis en gestion des données



## Flux de travail de projet de ML : Etape 3



## Flux de travail de projet de ML : Etape 3

### Type casting

Conversion des types de données en formats appropriés



### Restructuration

Restructuration des données par pivotement, dé-pivotement et agrégation



### Conversion de formats

Conversion de texte en formats numériques et normalisation des dates



Transformation

# Flux de travail de projet de ML : Etape 3

## Transformation

- **Conversion de formats:**

- Texte vers numérique
- Dates en format standard
- Encodage des variables catégorielles (one-hot, label encoding)

- **Restructuration:**

- Pivot/Unpivot
- Agrégation
- Jointures de tables

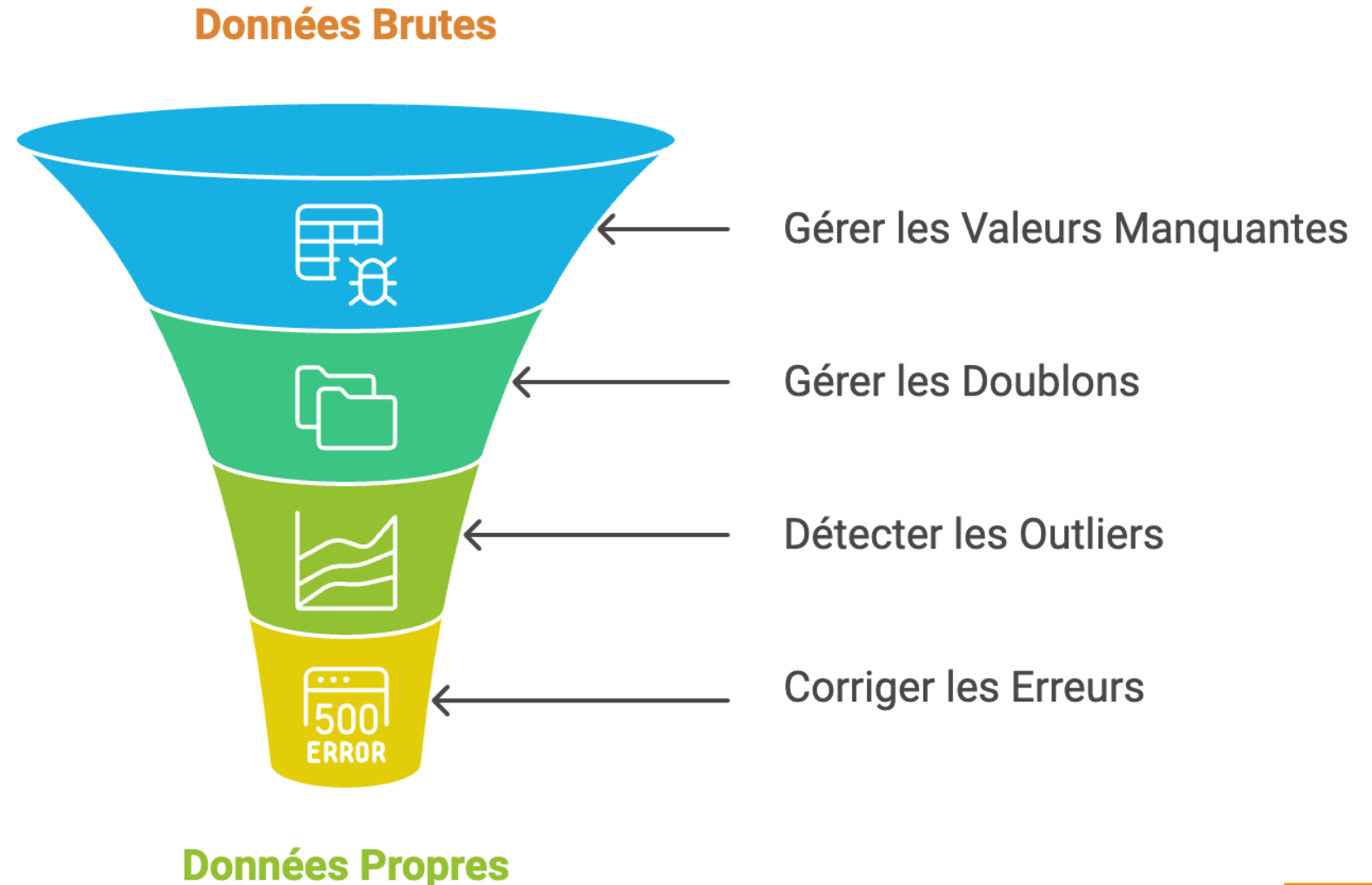
- **Type casting:**

- Conversion des types de données appropriés (int, float, datetime...)

# Flux de travail de projet de ML : Etape 3

Processus de Nettoyage des Données

## Data Cleaning (Nettoyage)



## Flux de travail de projet de ML : Etape 3

# Data Cleaning (Nettoyage)

**• Traitement des valeurs manquantes:**

- Suppression
- Imputation (moyenne, médiane, mode)
- Prédiction des valeurs manquantes

**• Gestion des doublons:**

- Identification
- Suppression ou fusion

**• Détection et traitement des outliers:**

- Méthodes statistiques (z-score, IQR)
- Suppression ou transformation

**• Correction des erreurs:**

- Fautes d'orthographe
- Formats incorrects
- Incohérences

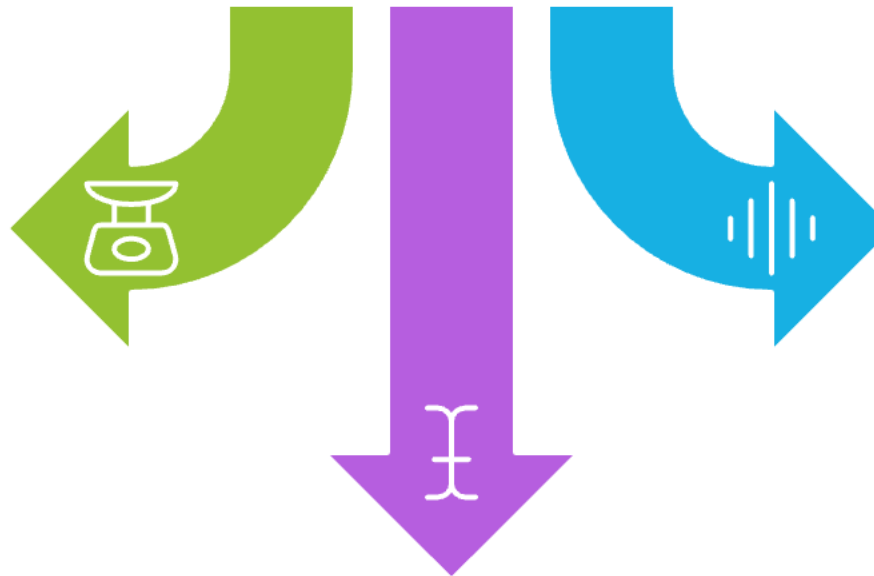
## Flux de travail de projet de ML : Etape 3

## Normalization

Quelle technique de normalisation utiliser ?

**Mise à l'échelle Min-Max**

Met les données à l'échelle d'un intervalle de 0 à 1.



**StandardScaler**

Met les données à l'échelle pour avoir une moyenne de 0 et un écart-type de 1.

**RobustScaler**

Met les données à l'échelle tout en étant résistant aux outliers.

## Flux de travail de projet de ML : Etape 3

• **Mise à l'échelle:**

- Min-Max scaling (0 à 1)
- StandardScaler (moyenne 0, écart-type 1)
- RobustScaler (résistant aux outliers)

• **Autres techniques:**

- Log transformation
- Box-Cox transformation
- Winsorization

## Normalization

## Flux de travail de projet de ML : Etape 3

## Normalization

Quelle technique de normalisation utiliser ?

**Mise à l'échelle Min-Max**  
Met les données à l'échelle d'un intervalle de 0 à 1.



**StandardScaler**

Met les données à l'échelle pour avoir une moyenne de 0 et un écart-type de 1.

**RobustScaler**

Met les données à l'échelle tout en étant résistant aux outliers.

# Flux de travail de projet de ML : Etape 3

Réaliser une Ingénierie des Caractéristiques Efficace

Création de  
Nouvelles  
Caractéristiques



Réduction de  
Dimensionnalité



Sélection de  
Caractéristiques



Feature Engineering

## Flux de travail de projet de ML : Etape 3

## Feature Engineering

**•Création de nouvelles features:**

- Combinaison de variables existantes
- Extraction d'informations (ex: jour de la semaine depuis une date)
- Agrégations temporelles

**•Réduction de dimensionnalité:**

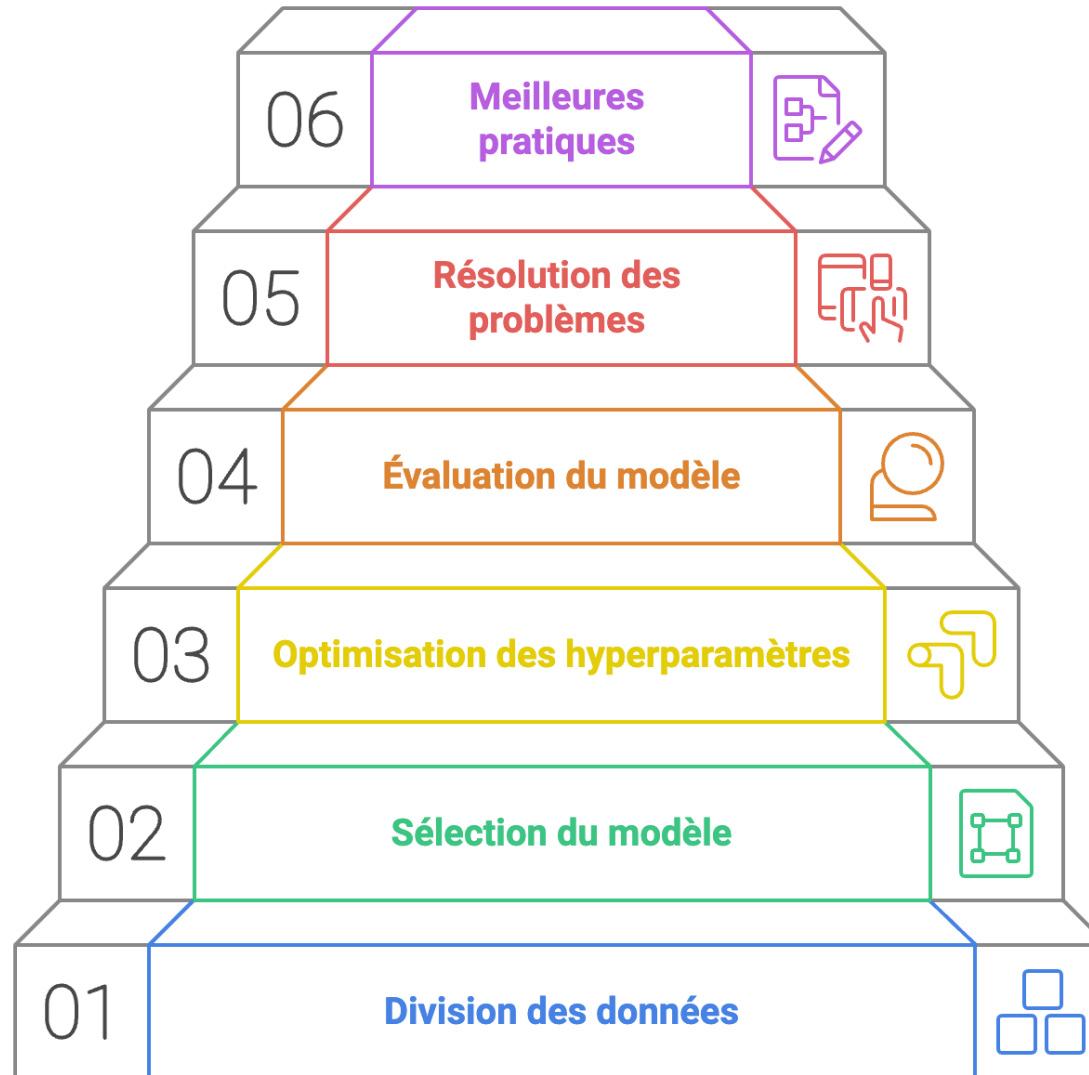
- PCA (Analyse en Composantes Principales)
- t-SNE
- UMAP

**•Sélection de features:**

- Par importance (correlation, mutual information)
- Par méthodes wrapper
- Par méthodes embedded

## Flux de travail de projet de ML : Etape 4

Construction de model

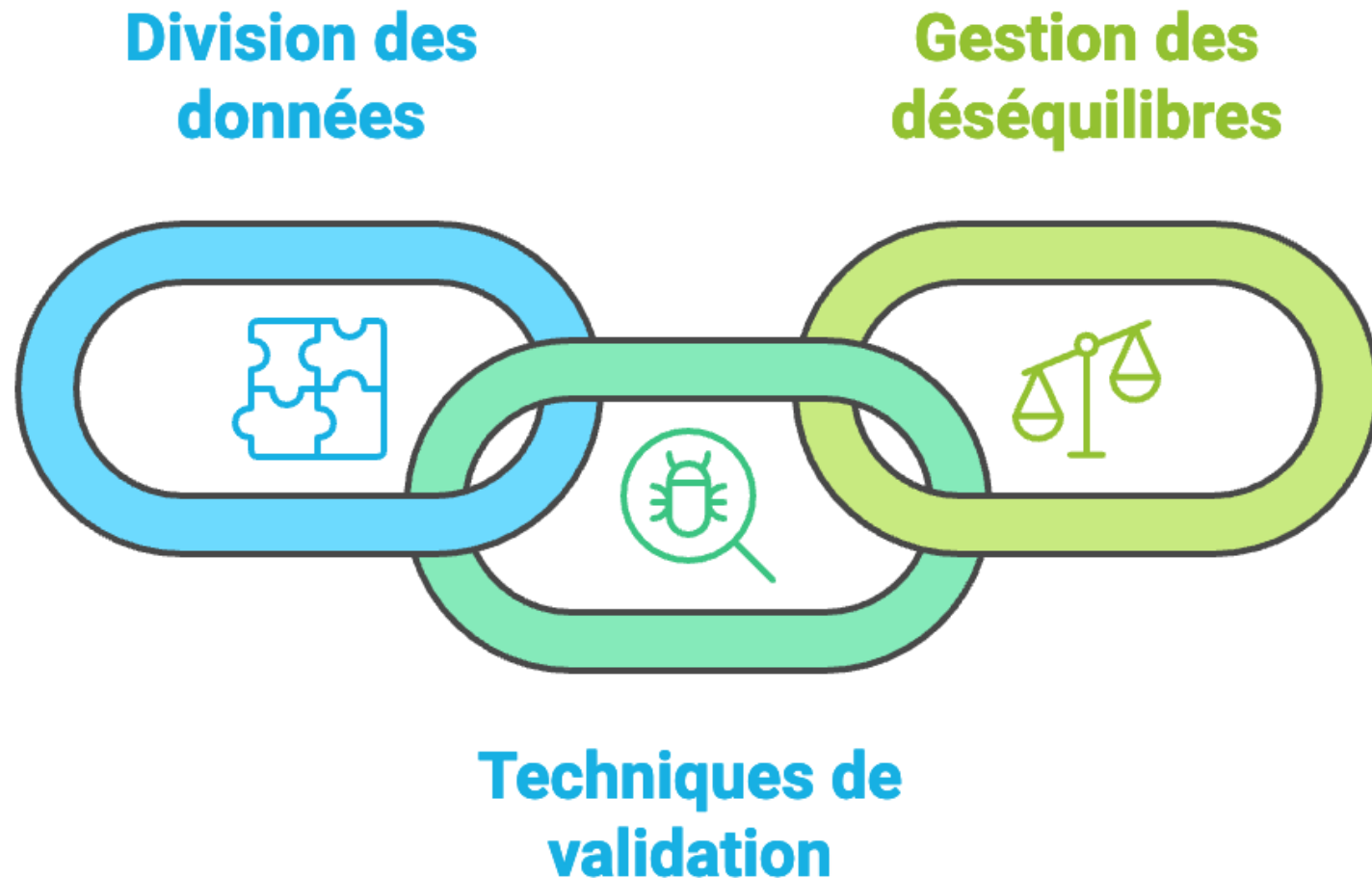


Flux de travail de projet de ML : Etape 4

Processus d'entraînement

Construction de model

1) Training (Entraînement)



## Flux de travail de projet de ML : Etape 4

1) Training (Entraînement)

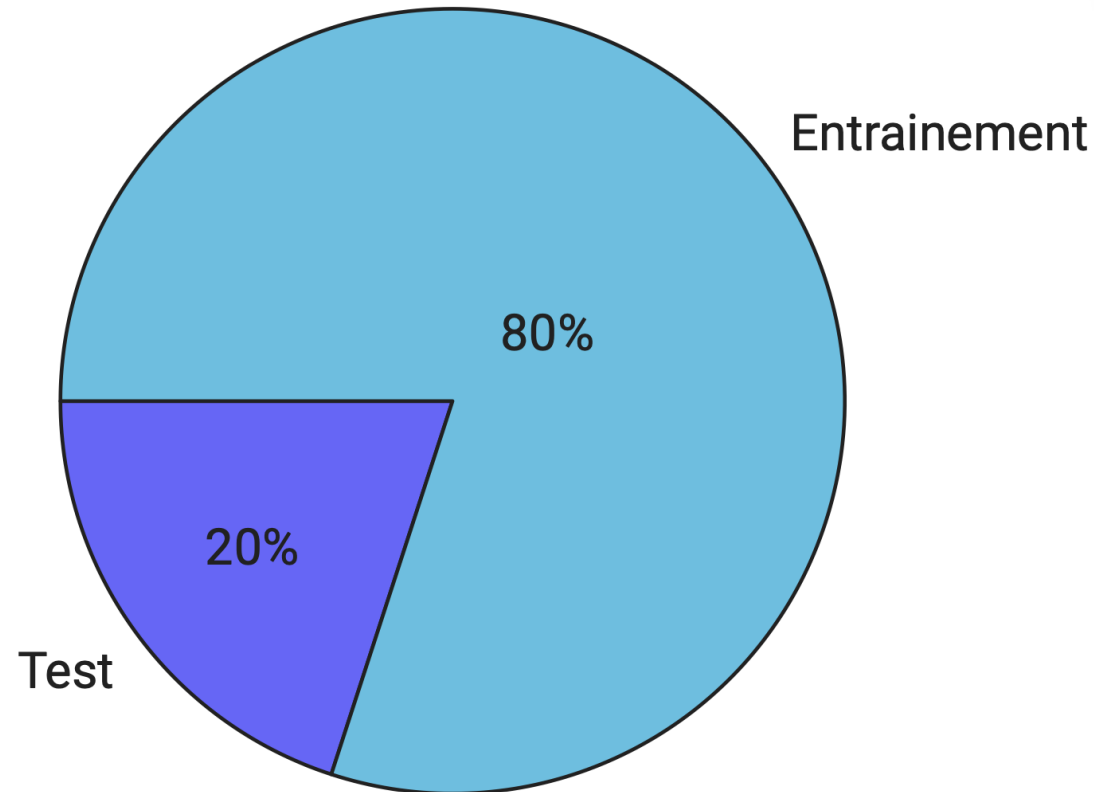
## Construction de model

- **Division des données:**
  - Training set (ensemble d'entraînement)
  - Validation set (ensemble de validation)
  - Test set (ensemble de test)
- **Techniques de validation:**
  - Cross-validation
  - K-fold validation
  - Leave-one-out validation
- **Gestion des déséquilibres:**
  - Over-sampling
  - Under-sampling
  - SMOTE
  - Class weights

## Flux de travail de projet de ML : Etape 4

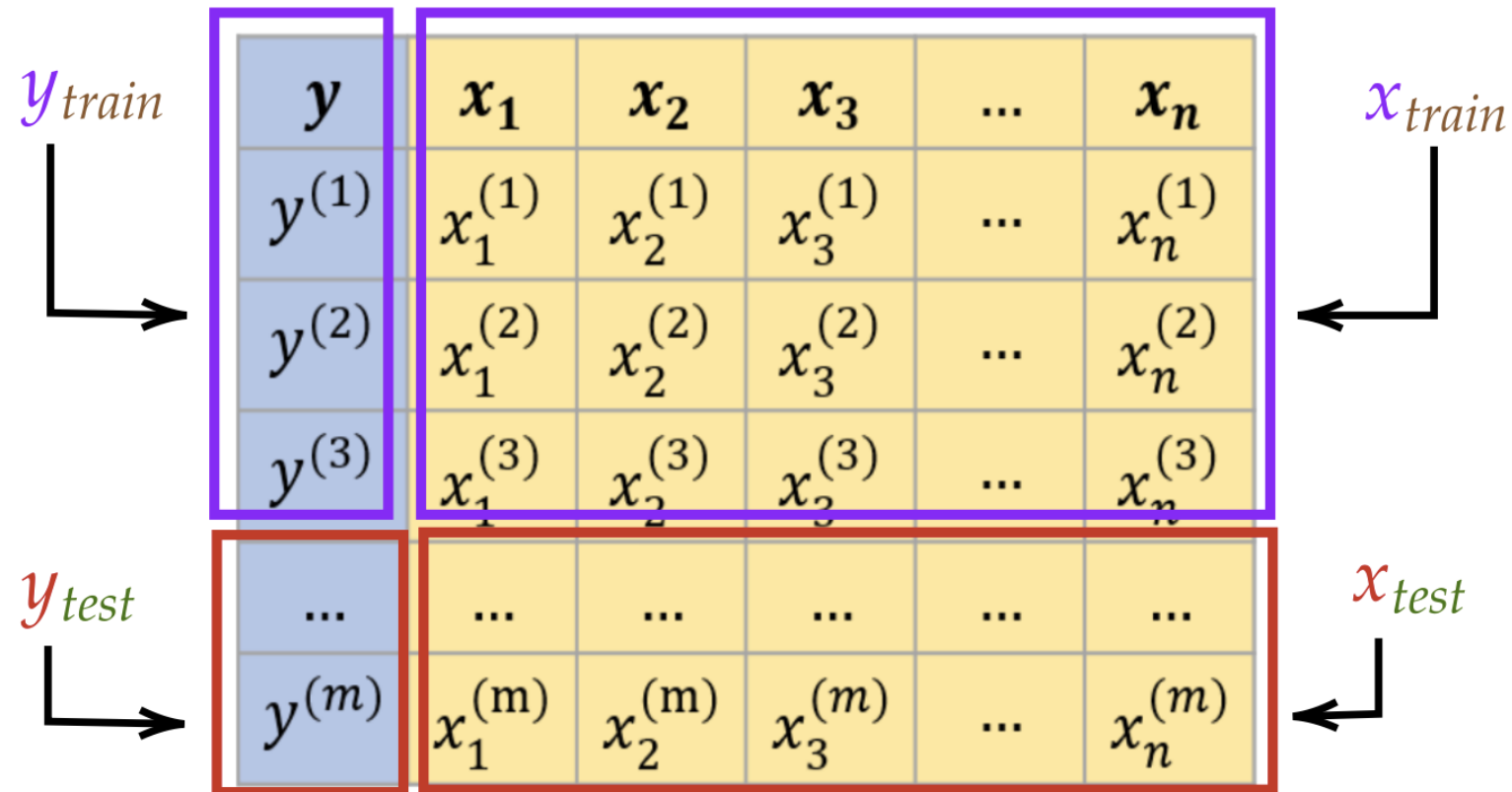
1) Training (Entraînement) : Division des données

## Séparation de dataset



Construction de model

## Flux de travail de projet de ML : Etape 4

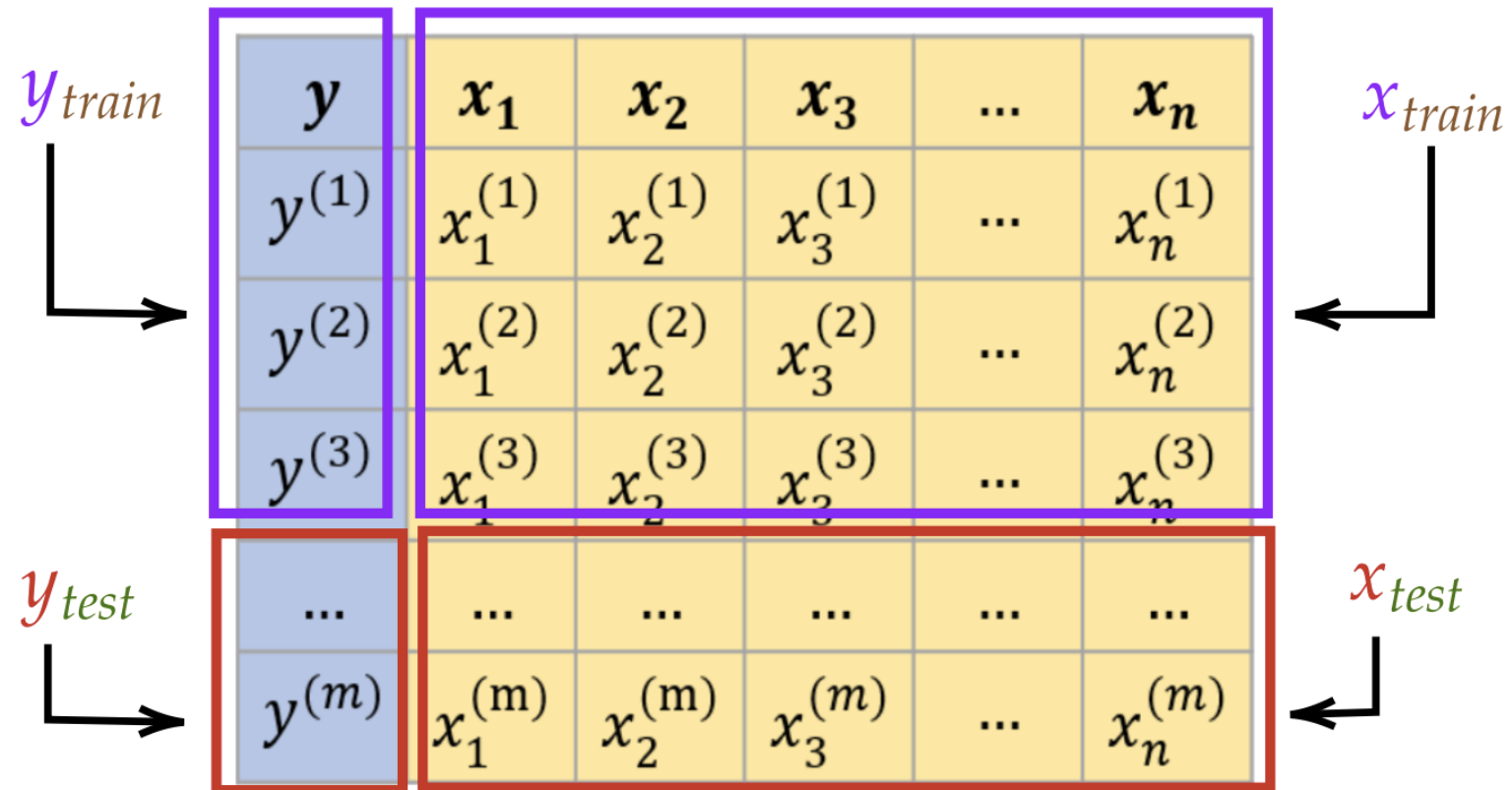
1) Training (Entraînement) : Division des donnéesDataset  $(x, y)$ 

Construction de model

Flux de travail de projet de ML : **Etape 4**

**1) Training (Entraînement) : Division des données**

*Dataset (x, y)*

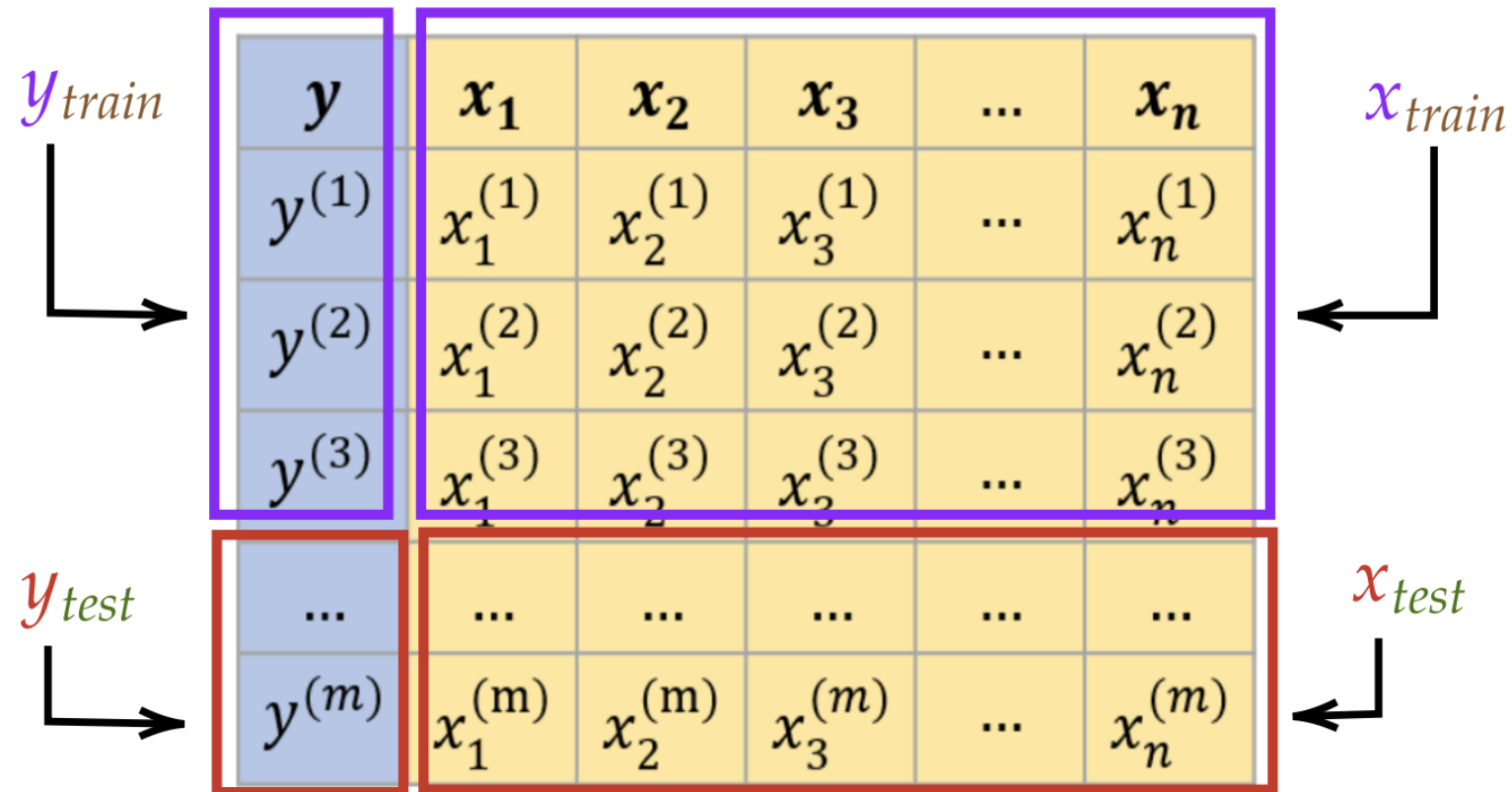


Construction de model

Flux de travail de projet de ML : **Etape 4**

**1) Training (Entraînement) : Division des données**

*Dataset (x, y)*



Construction de model

## Flux de travail de projet de ML : Etape 4

2) Model Selection

## Construction de model

**Types d'algorithmes:**

## •Supervisés:

- Classification (Random Forest, SVM, XGBoost...)
- Régression (Linear, Lasso, Ridge...)

## •Non supervisés:

- Clustering (K-means, DBSCAN...)
- Réduction de dimension (PCA, t-SNE...)

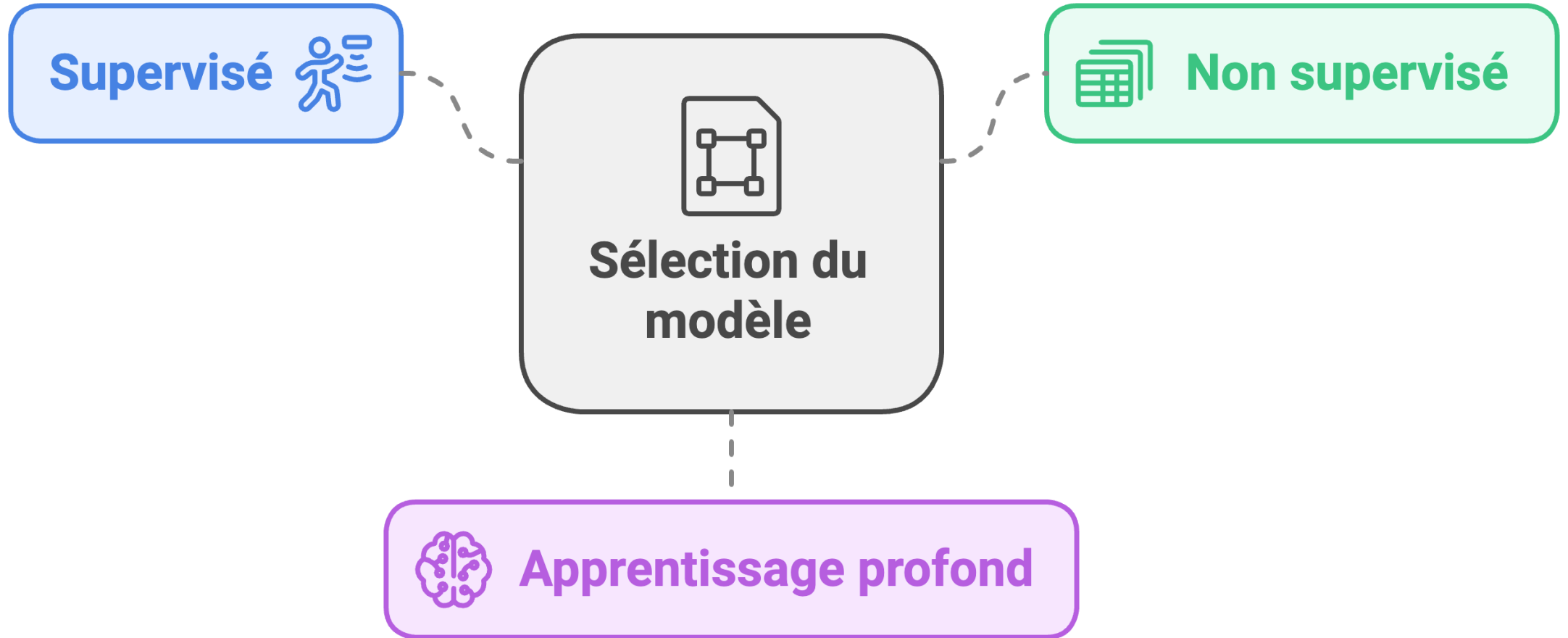
## •Deep Learning:

- CNN pour les images
- RNN/LSTM pour les séquences
- Transformers pour le texte

Flux de travail de projet de ML : Etape 4

2) Model Selection

Construction de model



## Flux de travail de projet de ML : Etape 4

2) Model Selection

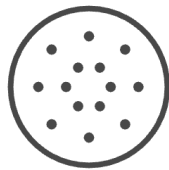
**Hyperparameter Tuning** (Optimisation des hyperparamètres)

• **Méthodes de recherche:**

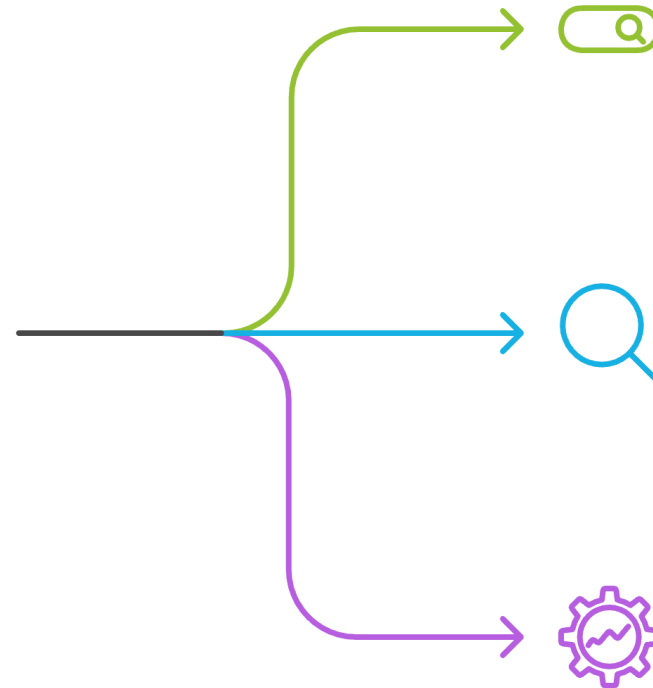
- Grid Search (recherche exhaustive)
- Random Search (recherche aléatoire)
- Bayesian Optimization

• **Paramètres courants:**

- Learning rate
- Nombre de couches/neurones
- Taille des batch
- Nombre d'arbres
- Profondeur maximale

Flux de travail de projet de ML : **Etape 4****2) Model Selection****Construction de model**

Quelle méthode  
d'optimisation des  
hyperparamètres  
utiliser ?

**Recherche exhaustive  
/Grid Search**

Recherche exhaustive,  
adaptée aux petits espaces  
de paramètres.

**Recherche aléatoire/Random Search**

Échantillonnage aléatoire,  
plus efficace pour les grands  
espaces de paramètres.

**Optimisation bayésienne**

Modèle probabiliste, efficace  
pour les évaluations  
coûteuses.

## Flux de travail de projet de ML : Etape 4

2) Model Selection

## Model Evaluation (Évaluation du modèle)

## • Métriques de classification:

- Accuracy
- Precision/Recall
- F1-Score
- ROC/AUC
- Confusion Matrix

## • Métriques de régression:

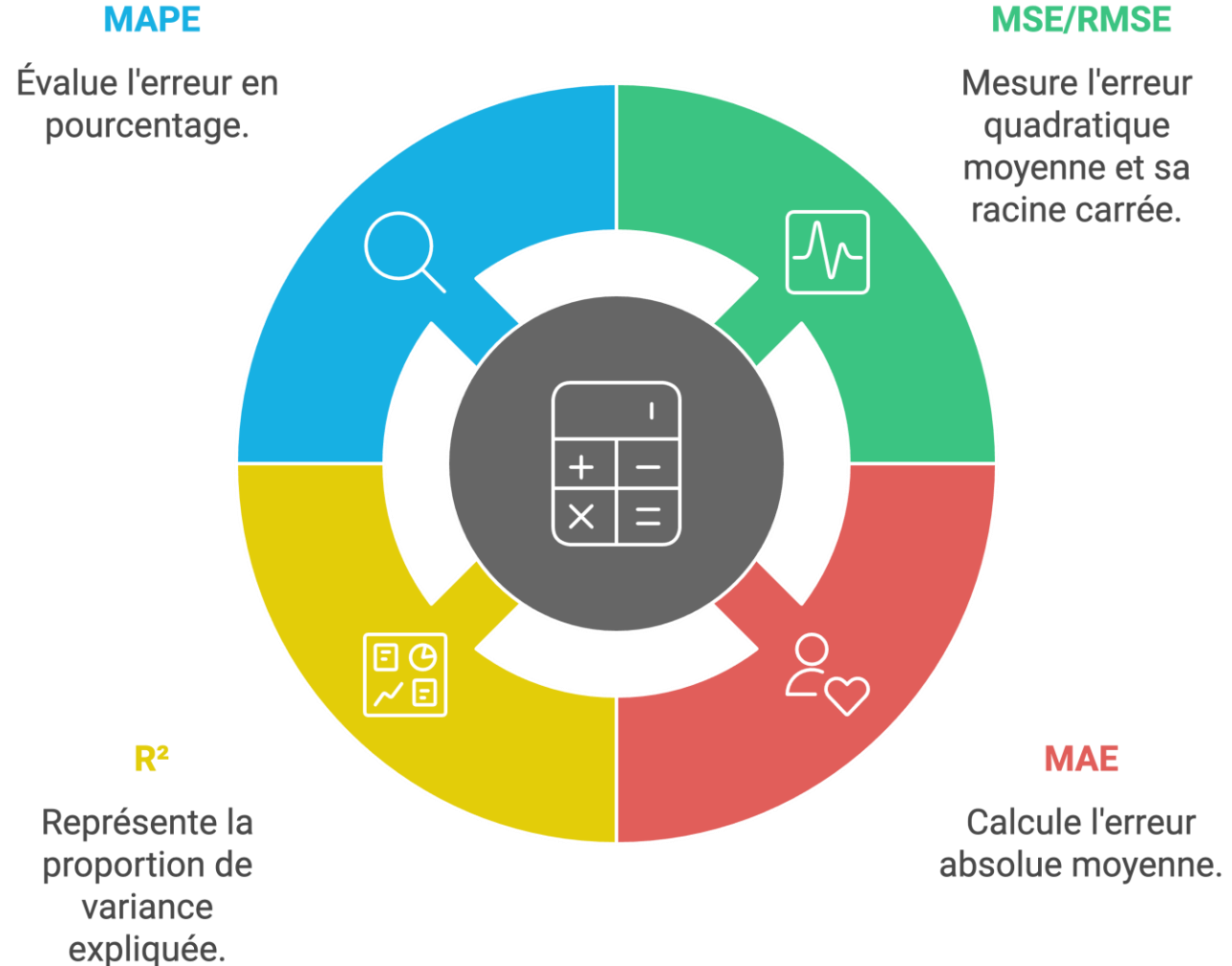
- MSE/RMSE
- MAE
- $R^2$
- MAPE

Flux de travail de projet de ML : **Etape 4**

## Aperçu des Métriques de Régression

**2) Model Selection**

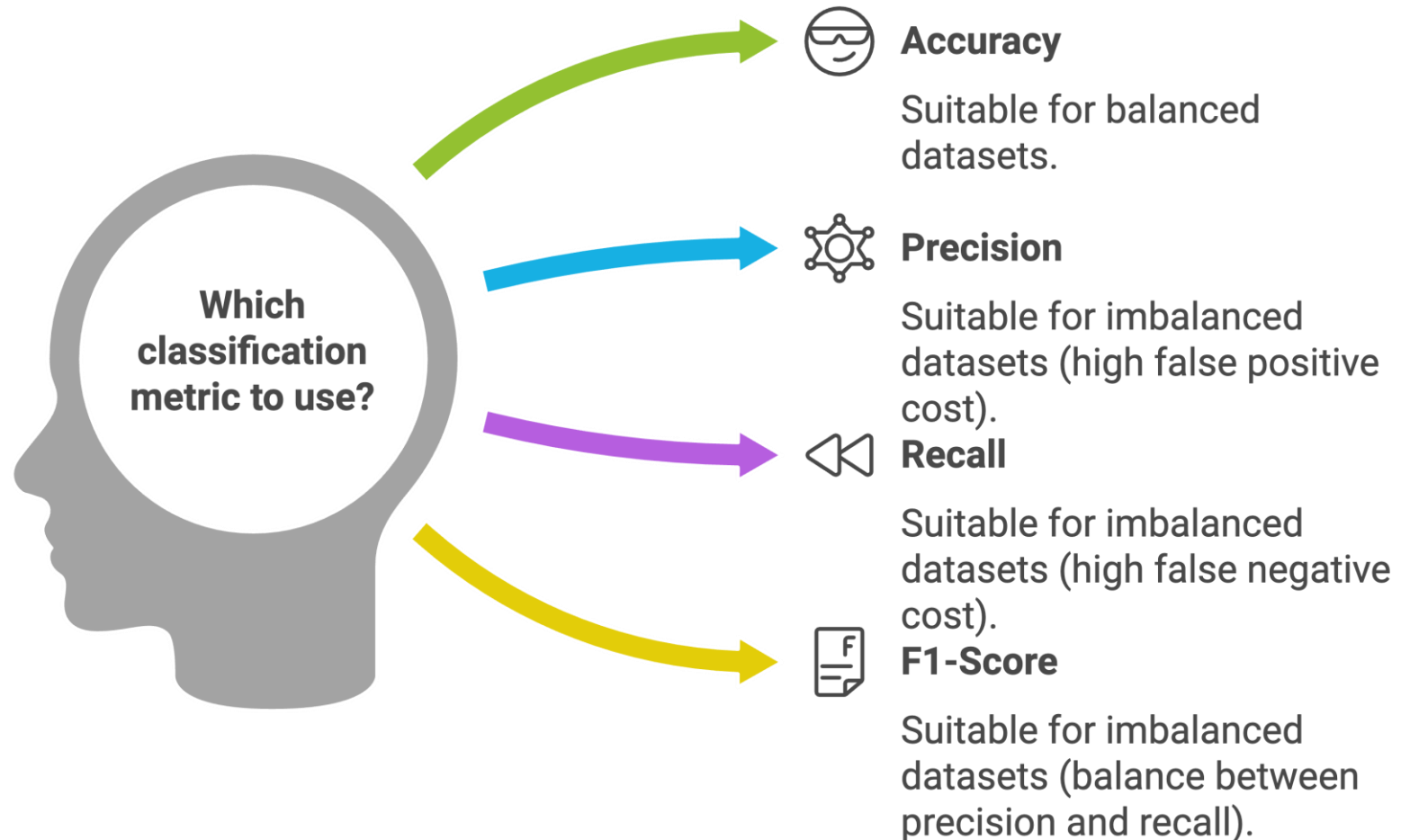
Construction de model



## Flux de travail de projet de ML : Etape 4

2) Model Selection

## Construction de model



## Flux de travail de projet de ML : Etape 4

2) Model Selection

## Problèmes courants:

• **Overfitting** (surapprentissage):

- Régularisation
- Dropout
- Early stopping

• **Underfitting** (sous-apprentissage):

- Augmentation de la complexité du modèle
- Feature engineering supplémentaire
- Plus de données

• **Convergence:**

- Ajustement du learning rate
- Batch normalization
- Gestion des gradients

Flux de travail de projet de ML : Etape 4

## 2) Model Selection

Construction de model

















**MERCI POUR VOTRE ATTENTION**