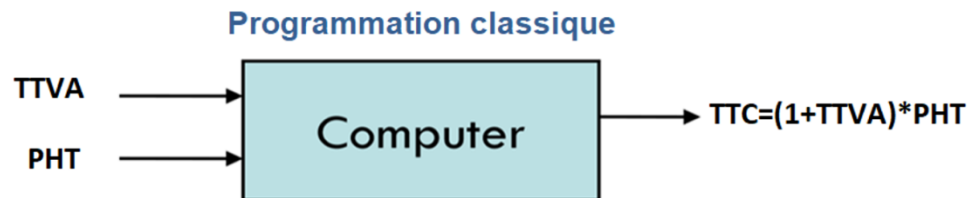


Machine Learning

Introduction

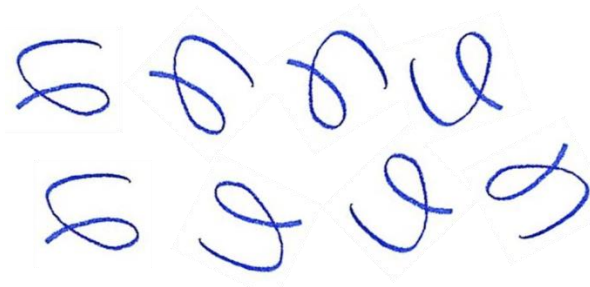
Introduction

- Pour résoudre un problème sur un ordinateur, nous avons besoin d'un algorithme.
- Un algorithme est une séquence d'instructions qui devrait être effectuée pour transformer une entrée en une sortie.
- Dans ce type de programmation (**qualifiée de programmation classique**), on essaye de trouver un algorithme qui permet de trouver les résultats à partir des données saisies par l'utilisateur.



Introduction

- Mais pour certaines tâches nous n'avons pas d'algorithmes.
- En d'autres termes, on ne dispose pas d'une formule ou d'une méthode analytique qui permet de trouver les résultats à partir des données.
- Par exemple :
 - Comment distinguer des spams des emails légitimes.
 - Comment reconnaître l'une des images suivantes ? S'agit-il du chiffre 6 ou 9 ou la lettre v ?

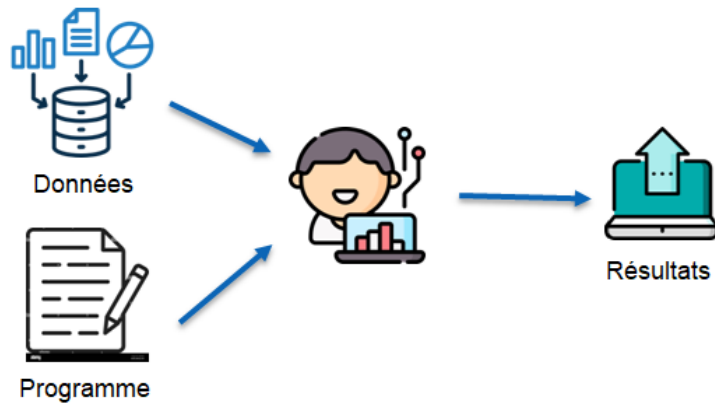


Introduction

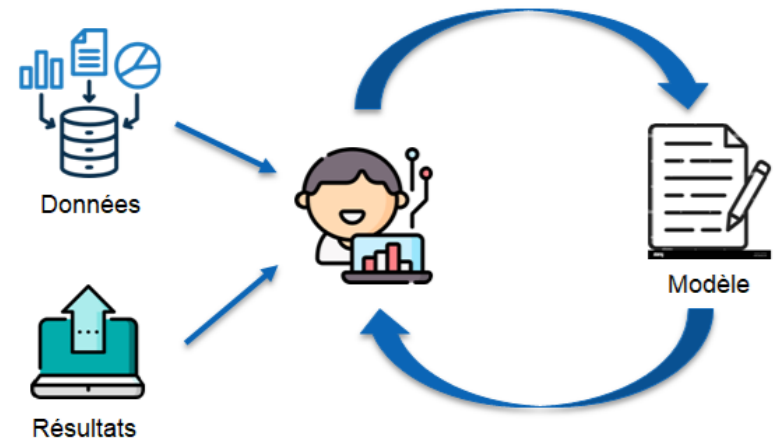
- Ce qui nous manque dans la connaissance (algorithme), nous le compensons dans les données.
- Par exemple pour le cas des emails, nous pouvons passer à la machine des milliers de messages d'exemple dont certains sont des spam et d'autres des ham, et demander à la machine d' "apprendre".
- En d'autres termes, nous aimerions que l'ordinateur (machine) extraie automatiquement l'algorithme pour réaliser cette tâche.

Introduction

Programmation Classique



Apprentissage Automatique

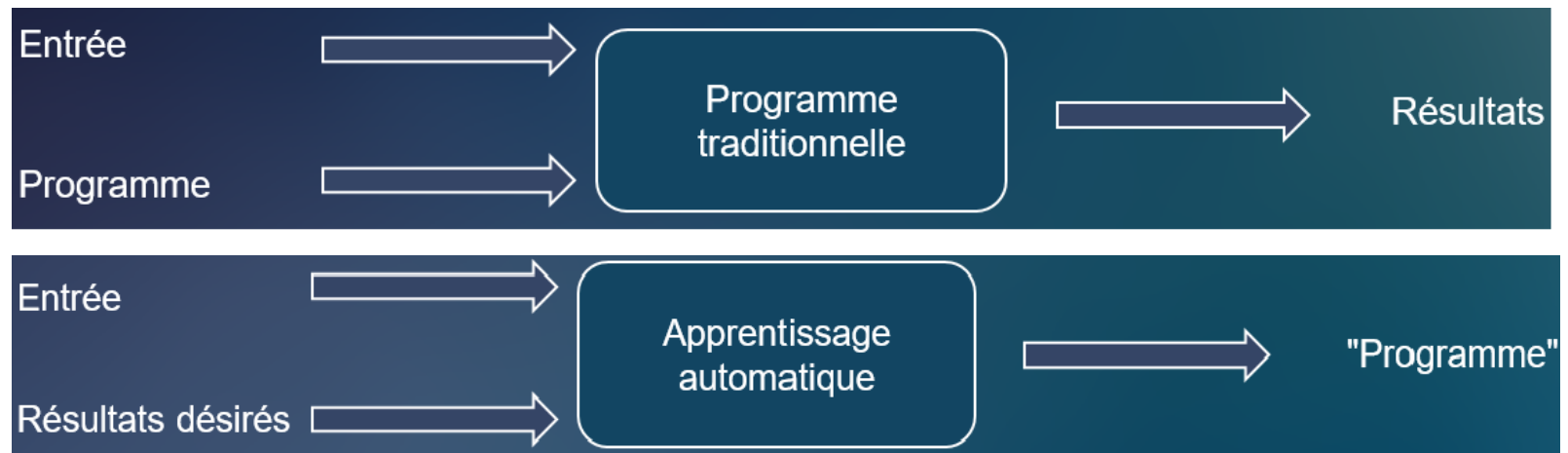


Définition de Machine Learning

- Le **Machine Learning (ML)** est une branche de l'intelligence artificielle (AI) qui permet aux ordinateurs d'apprendre à partir des données et de s'améliorer automatiquement avec l'expérience, sans avoir été explicitement programmés pour une tâche spécifique.
- Il repose sur des algorithmes capables d'identifier des motifs dans les données, de faire des prédictions ou de prendre des décisions basées sur ces analyses.

Définition de Machine Learning

- "Field of study that gives computers the ability to learn without being explicitly programmed" (Arthur Samuel 1959)



Définition de 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 1990)

Exemple : Classification des emails

- **T (tâche)** : C'est la tâche spécifique que le programme doit accomplir. Par exemple, classer des emails comme "spam" ou "non-spam".
- **P (performance)** : C'est la manière dont on mesure la réussite du programme sur la tâche. Par exemple, le pourcentage d'emails correctement classés.
- **E (expérience)** : C'est l'expérience d'apprentissage, comme un ensemble de données d'emails étiquetés (spam/non-spam), que le programme utilise pour s'améliorer.

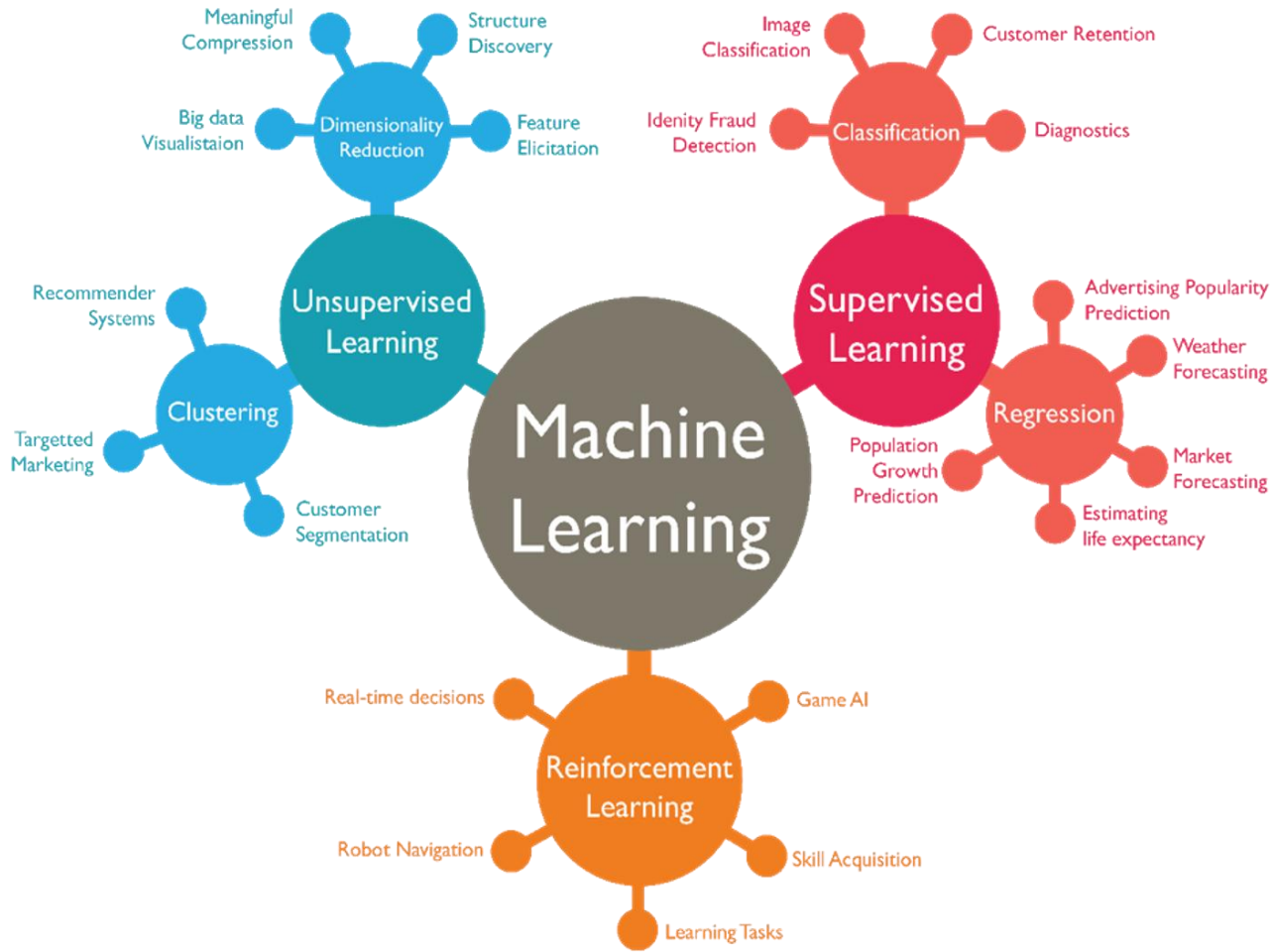
Définition de Machine Learning

"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 1990)

Exemple : système de recommandation de films

- **Tâche (T)** : Recommander des films à un utilisateur en fonction de ses préférences passées.
- **Expérience (E)** : L'expérience est constituée des historiques de visionnage et des évaluations que l'utilisateur a fournies pour différents films.
- **Mesure de Performance (P)** : La mesure de performance peut être la précision du système à recommander des films que l'utilisateur aime (par exemple, en calculant le pourcentage de recommandations correctes).

Types de Machine Learning



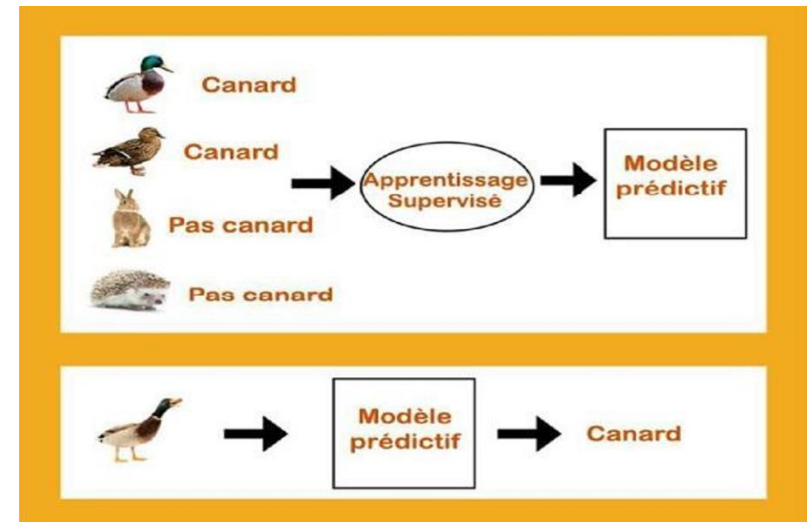
Supervised Learning

- La machine apprend à partir de données étiquetées.
- Chaque exemple de données d'entraînement est composé d'une **entrée** (ensemble de caractéristiques) et d'une **sortie attendue** (ou étiquette).
- L'objectif est que le modèle puisse prédire la sortie pour de nouvelles données inconnues en généralisant les relations apprises.

Supervised Learning

Exemple

Les données d'entrée représentent des images et la cible (ou target en anglais) représente la catégorie de photos.

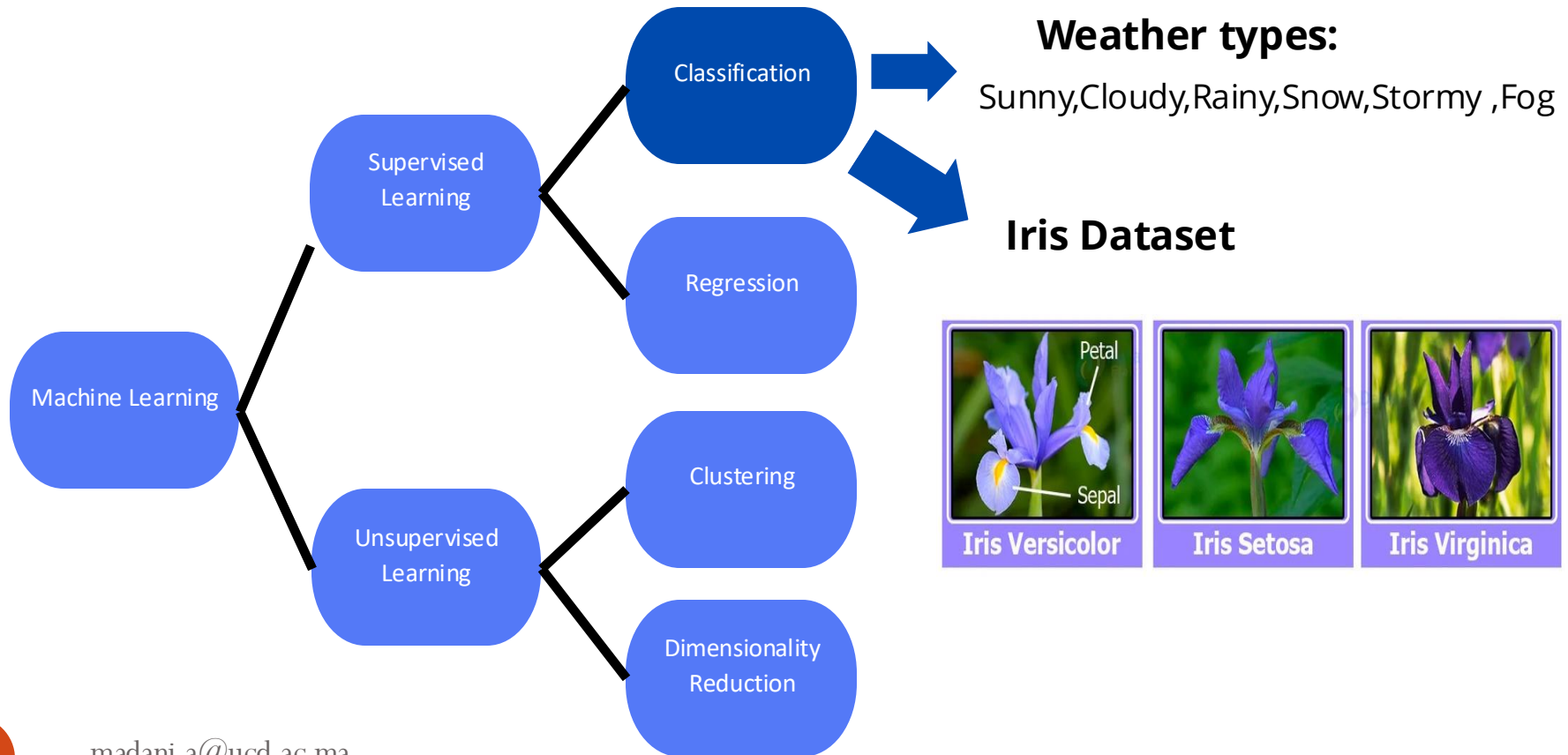


Supervised Learning

- Un apprentissage supervisé peut à son tour être :
 - Une Classification, quand les labels sont discrets
 - Une Régression, quand les labels sont continues

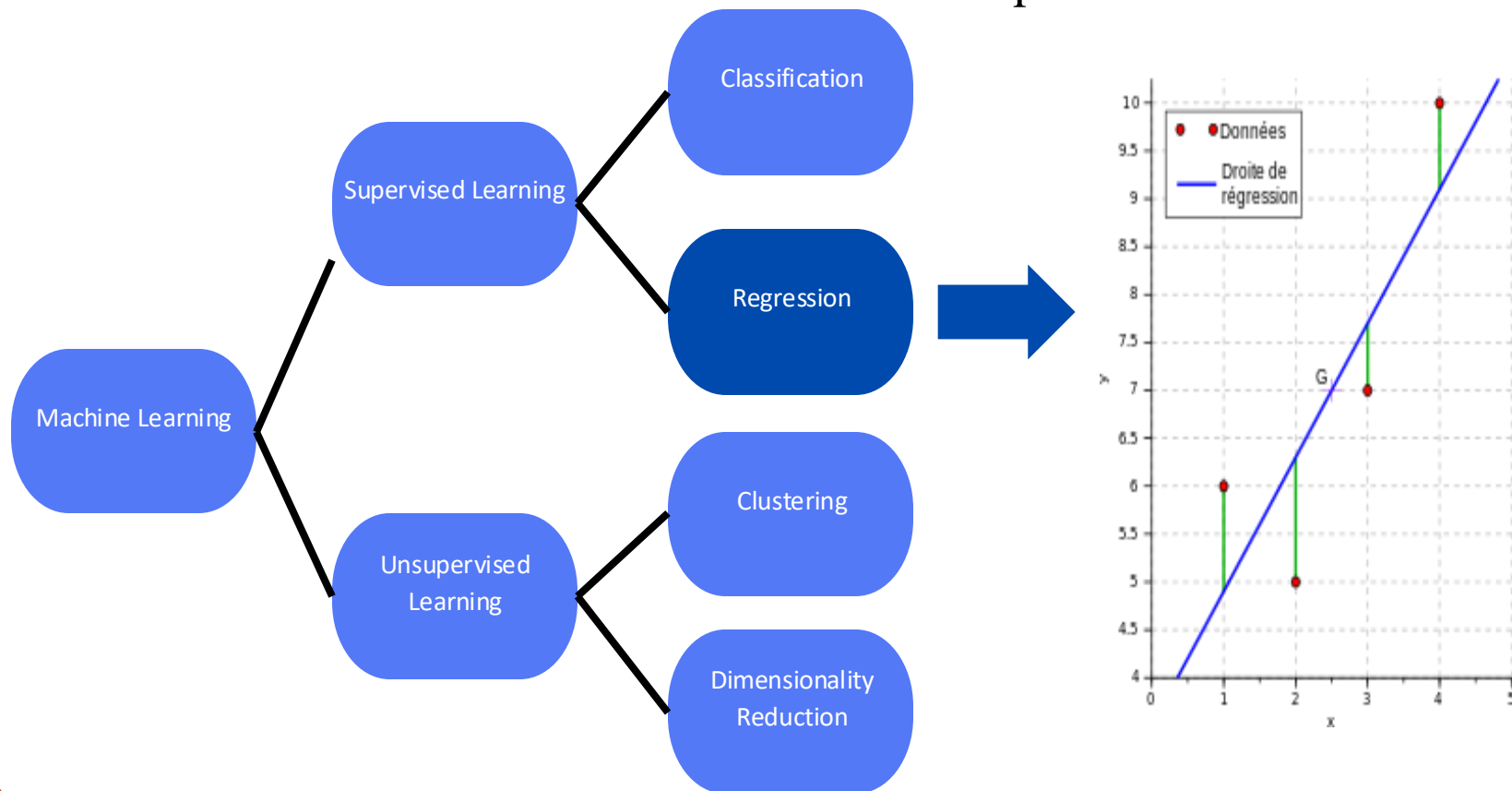
Supervised Learning (Classification)

Permet de prédire des classes ou des catégories pour de nouvelles observations en se basant sur des données étiquetées



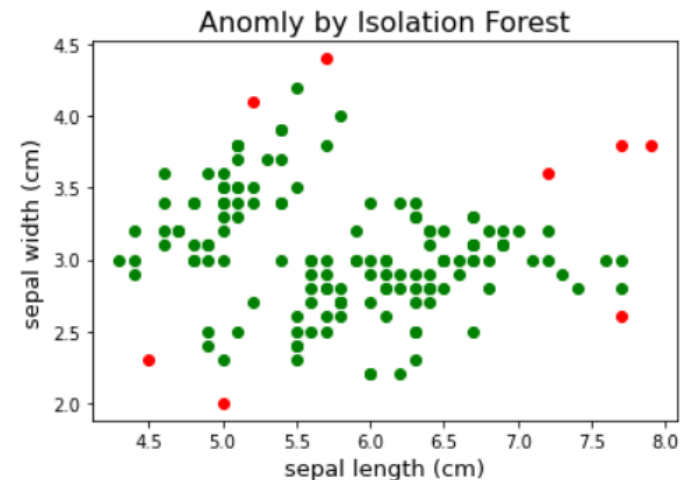
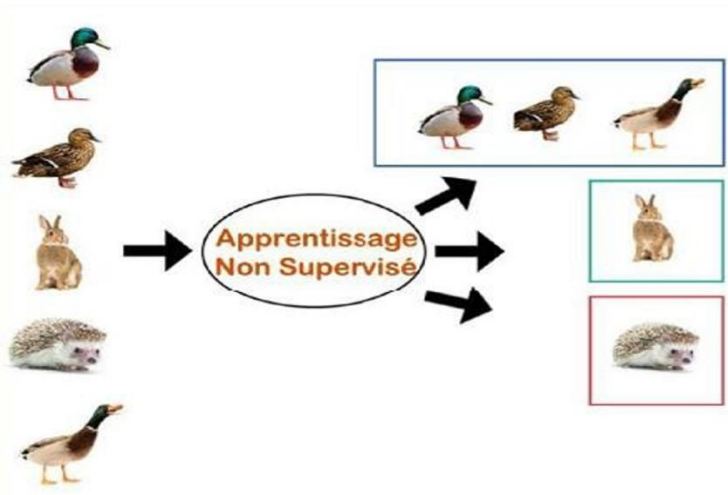
Supervised Learning (Régression)

- La régression implique de prédire des valeurs numériques continues en fonction des caractéristiques d'entrée.



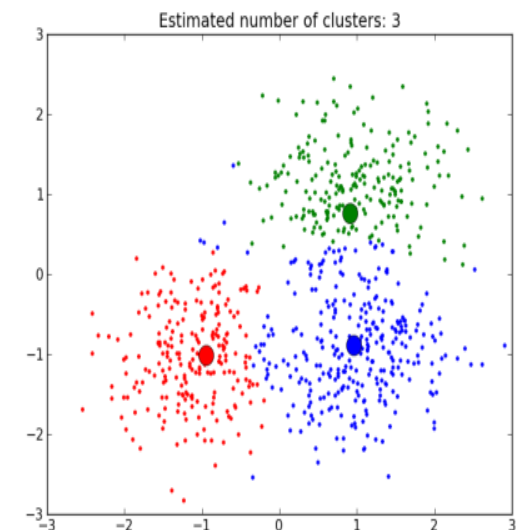
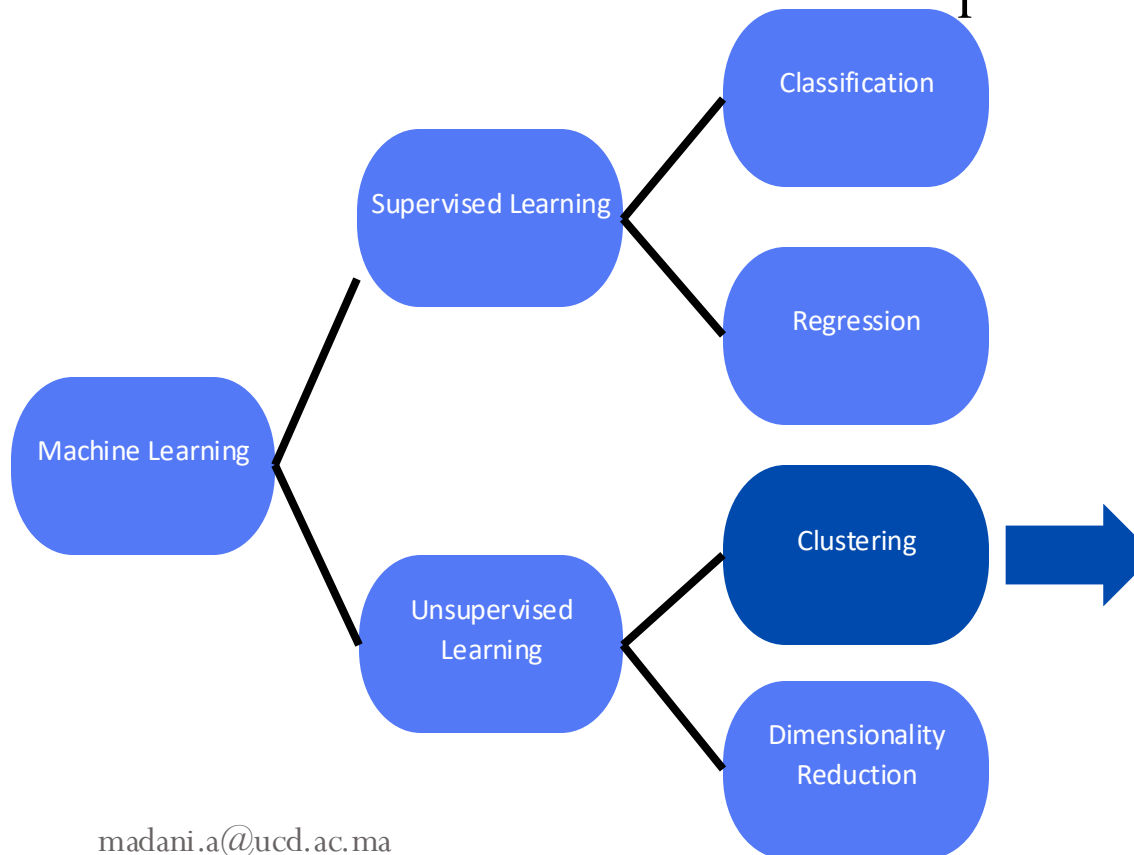
Unsupervised Learning

- L'apprentissage non supervisé est une méthode d'apprentissage automatique où l'algorithme est entraîné sur des données non étiquetées, c'est-à-dire sans sorties attendues ou réponses associées.
- L'objectif est de découvrir des **motifs, structures ou relations cachées** dans les données.



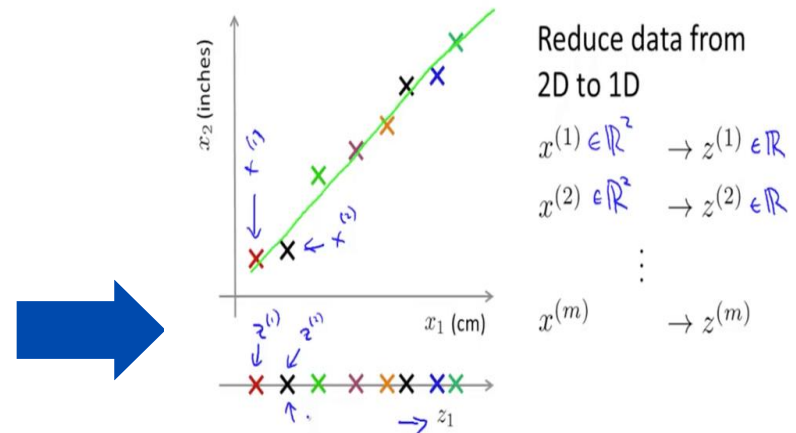
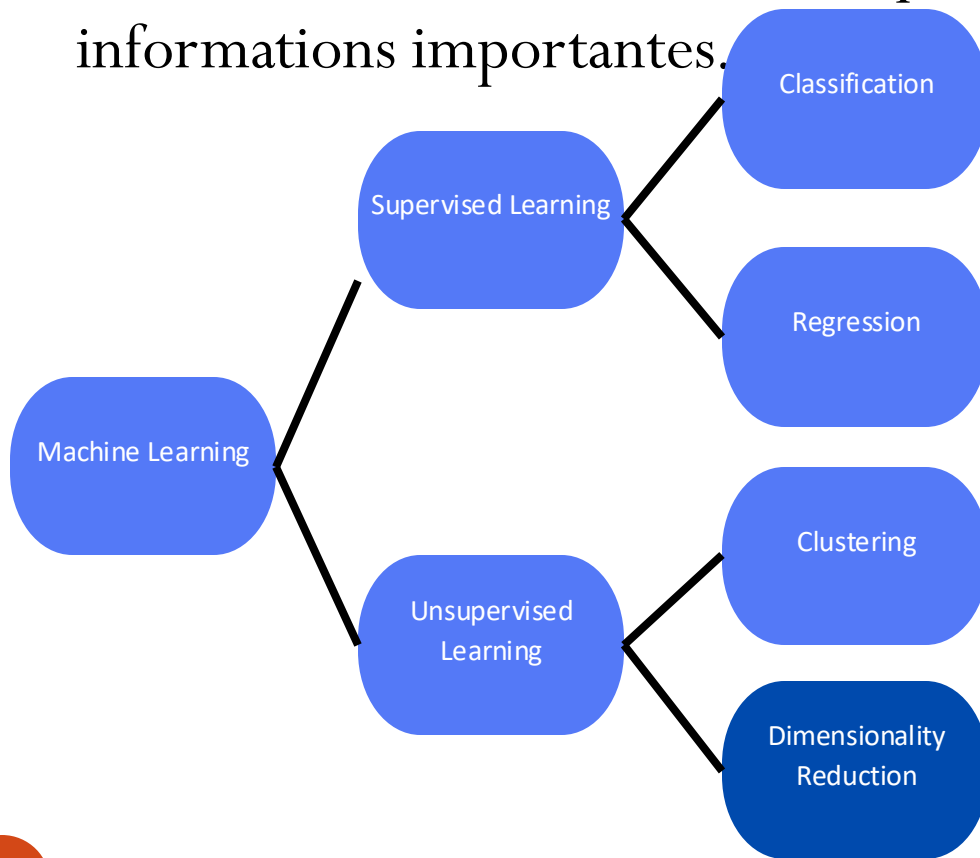
Unsupervised Learning (Clustering)

Permet de regrouper des observations similaires en fonction de leurs caractéristiques, sans étiquettes préalables, afin de découvrir des structures intrinsèques dans les données.



UnSupervised Learning (Dimensionality Reduction)

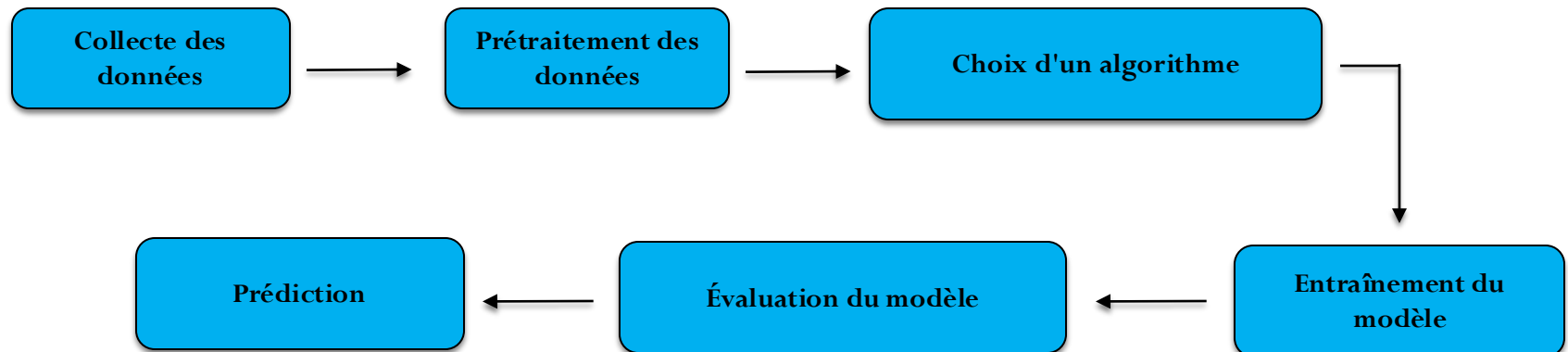
Permet de réduire le nombre de variables ou de caractéristiques dans un ensemble de données tout en préservant au mieux les informations importantes.



Machine Learning : Process Overview

Le process comporte plusieurs étapes :

1. Collecte des données
2. Prétraitement des données
3. Choix d'un algorithme (KNN, KMEANS, SVM, ...)
4. Entraînement du modèle
5. Test du modèle (accuracy, F1-Score, ...)
6. Déploiement



Machine Learning : Process Overview

1. Collecte des données :

1. Les données doivent être étiquetées ou non.
2. Open data, Webscraping, API,

2. Prétraitement des données :

1. Nettoyer les données pour éliminer les valeurs manquantes ou aberrantes.
2. Normaliser ou standardiser les caractéristiques pour que toutes soient sur une échelle comparable.
3. Diviser les données en trois ensembles :
 1. Entraînement : Pour former le modèle.
 2. Validation : Pour ajuster les hyperparamètres.
 3. Test : Pour évaluer les performances finales.

Machine Learning : Process Overview

1. Choix d'un algorithme :

1. L'algorithme dépend du type de problème (classification, régression, ...).
2. Exemples courants :
 1. Classification : SVM, k-NN, Random Forest, Réseaux de neurones.
 2. Régression : Régression linéaire, Régression Ridge, Réseaux de neurones.

2. Entraînement du modèle :

1. L'algorithme apprend les relations entre les entrées et les sorties en minimisant une fonction de perte (erreur entre les prédictions du modèle et les vraies valeurs).

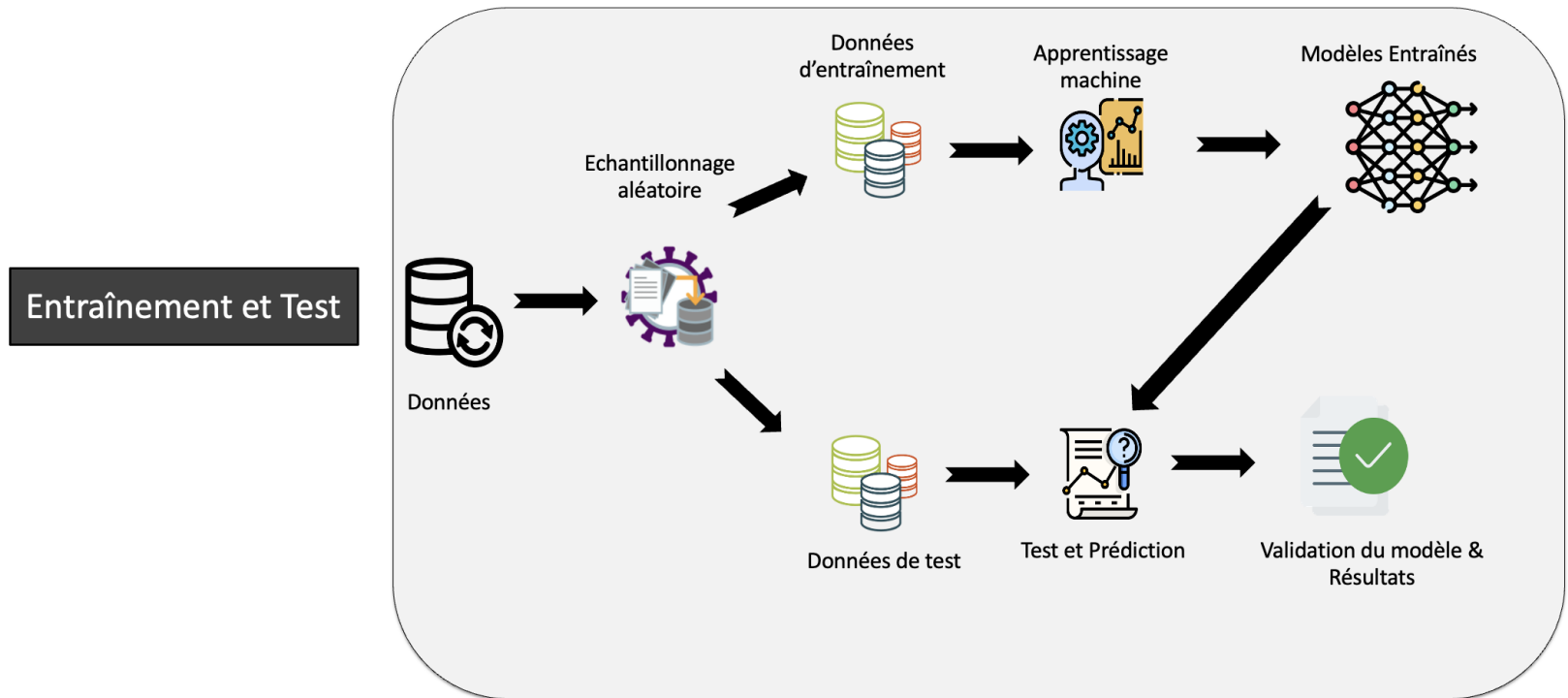
3. Évaluation :

1. Utiliser des métriques comme la précision, le rappel, le F1-score, ou l'erreur quadratique moyenne pour mesurer les performances.

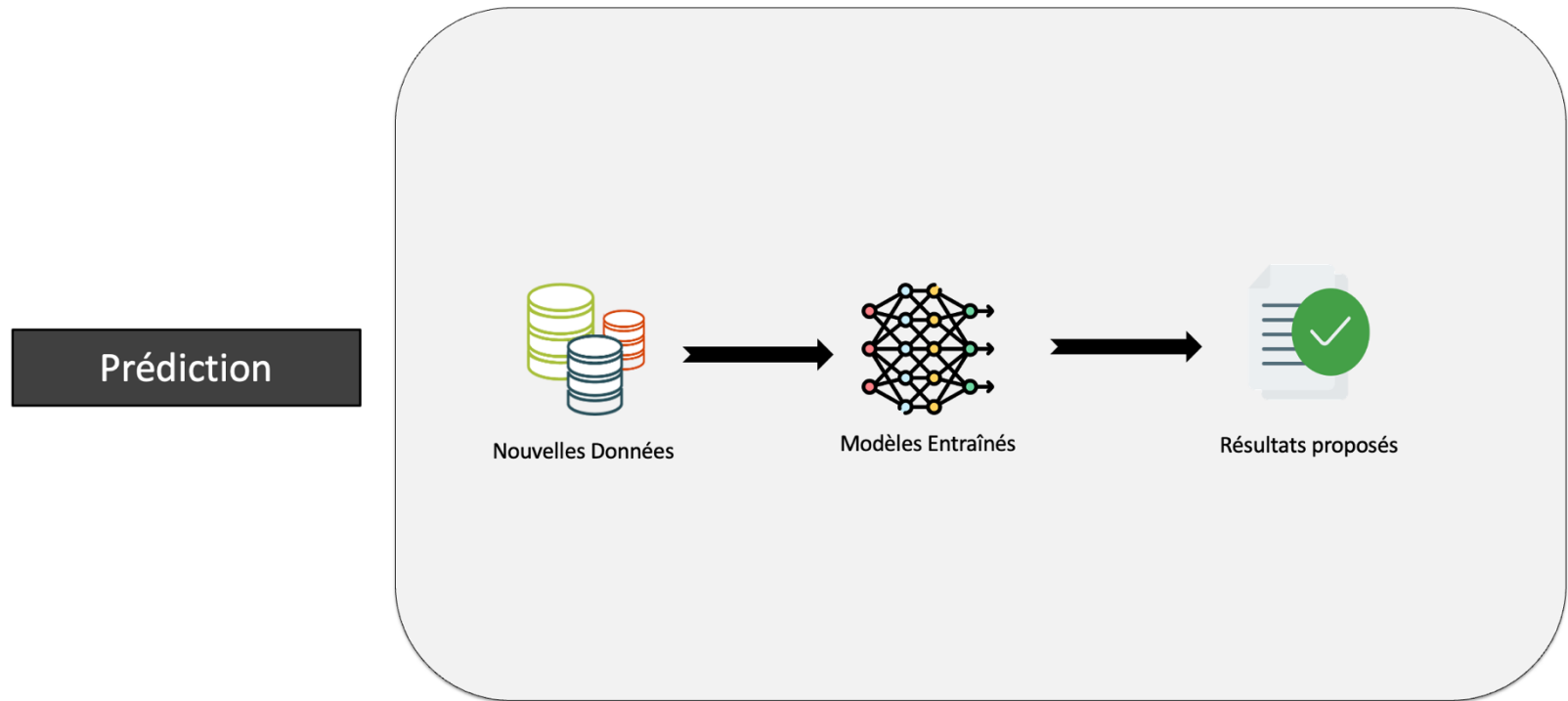
4. Prédiction :

1. Une fois le modèle entraîné et validé, il peut être utilisé pour prédire les sorties pour de nouvelles données.

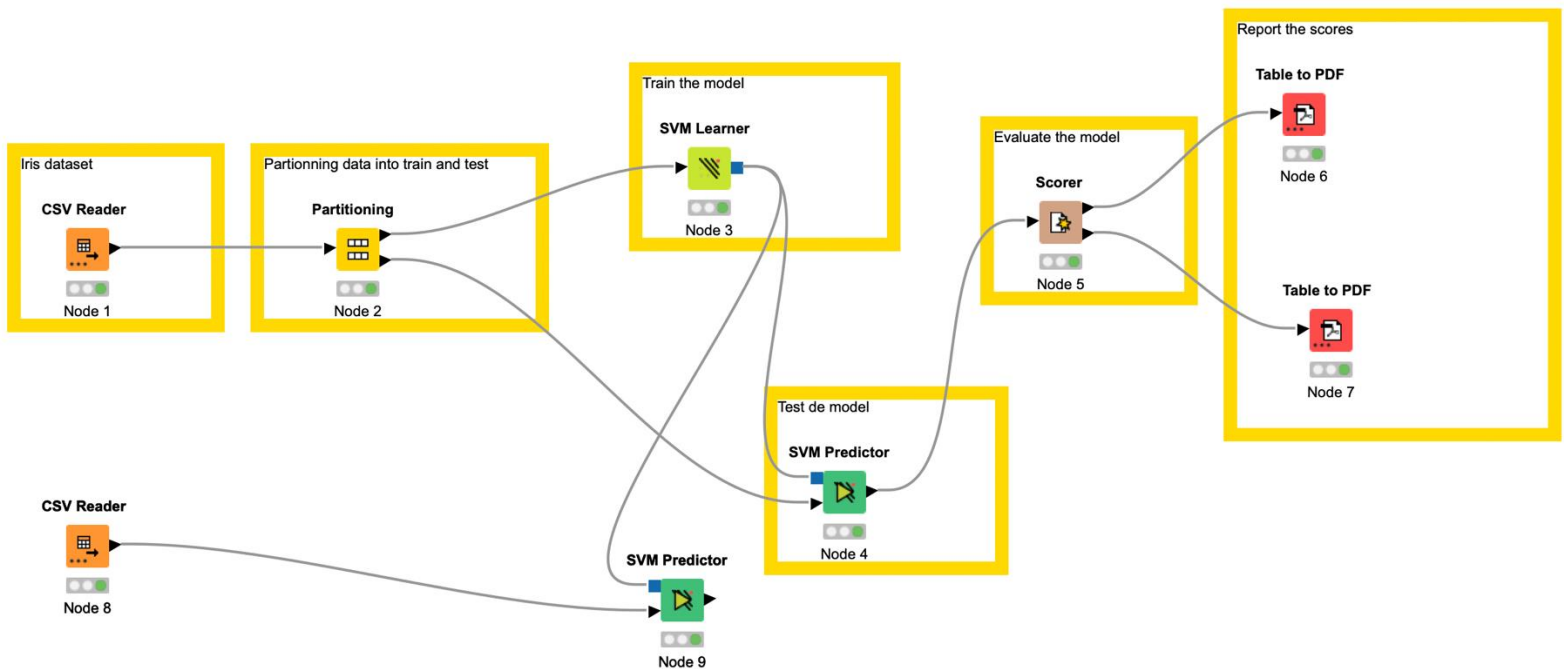
Machine Learning : Process Overview



Machine Learning : Process Overview



Machine Learning : Process Overview



Machine Learning : Process Overview

```
1 from sklearn.datasets import load_iris
2 from sklearn.svm import SVC
3 from sklearn.metrics import accuracy_score
4 from sklearn.model_selection import train_test_split
```

```
1 iris = load_iris()
2 X = iris.data
3 y = iris.target
```

```
1 iris.feature_names
```

```
1 iris.target_names
```

```
1 Xtrain, Xtest, ytrain, ytest = train_test_split(X,y,train_size=0.8)
```

```
1 X.shape
```

```
1 model_SVM = SVC()
```

```
1 y.shape
```

```
1 model_SVM.fit(Xtrain, ytrain)
```

```
1 ypred = model_SVM.predict(Xtest)
2 accuracy_score(ypred, ytest)
```