

# Apprentissage supervisé : Classification

## Mini-projets



**Université Abdelmalek Essaadi  
Faculté des Sciences et Techniques  
Tanger**

Cours du Pr. Abdellah AZMANI

**Cours du Pr. Abdellah AZMANI**  
Département Génie Informatique  
Cours pour les étudiants des Masters  
et des Cycles d'Ingénieurs



# Machine Learning

Apprentissage supervisé : Classification

Consignes générales



## ➤ Objectif du mini-projet

Mettre en œuvre un processus complet de classification supervisée, depuis des données imparfaites jusqu'à l'évaluation de plusieurs modèles, en appliquant cette démarche à un seul projet choisi par chaque groupe et en justifiant chaque étape de prétraitement, de modélisation et d'analyse des résultats

### Mini-projets :

- **Mini-projet 1** : Classification de l'occupation du sol
  - Jeu de données : ***geo\_lulc\_data.csv***
- **Mini-projet 2** : Classification des zones à risque d'inondation
  - Jeu de données : ***geo\_flood\_risk\_data.csv***
- **Mini-projet 3** : Classification de zones urbaines selon leur densité
  - Jeu de données : ***geo\_urban\_density\_data.csv***
- **Mini-projet 4** : Classification de l'aptitude des sols agricoles
  - Jeu de données : ***geo\_agri\_aptitude\_data.csv***
- **Mini-projet 5** : Classification des zones selon leur sensibilité environnementale
  - Jeu de données : ***geo\_sensibilite\_env\_data.csv***
- **Mini-projet 6** : Classification de la stabilité des terrains (risque de glissement)
  - Jeu de données : ***geo\_stabilite\_terrain\_data.csv***

# Machine Learning

Apprentissage supervisé :  
Classification

Consignes générales

## ➤ Travail demandé

### 1. Compréhension des données

- Comprendre le contexte métier
- Identifier les variables explicatives et la variable cible
- Analyser la structure du jeu de données

### 2. Analyse exploratoire et visualisation des données (EDA)

- Réaliser une **visualisation des données avant le prétraitement** afin de :
  - comprendre la distribution des variables, géographique
  - identifier les valeurs manquantes, aberrantes et les déséquilibres de classes
  - ...
- Réaliser une **visualisation des données après le prétraitement** afin de :
  - vérifier l'impact des traitements appliqués
  - comparer l'état des données avant/après nettoyage
  - valider la cohérence des transformations effectuées
  - ...

# Machine Learning

Apprentissage supervisé : Classification

Consignes générales



## ➤ Travail demandé

### 3. Prétraitement des données

- Harmonisation des noms de colonnes et des modalités
- Détection et traitement des valeurs manquantes
- Suppression des colonnes trop manquantes et des lignes quasi vides
- Détection et suppression des doublons
- Normalisation / standardisation des variables numériques
- ....

### 4. Modélisation

- Séparation des données en jeu d'entraînement et de test (Train/Test)
- Implémentation de plusieurs modèles de classification (ex. : Régression Logistique, KNN, Arbre de Décision, Random Forest, SVM, etc.)
- Utilisation initiale des modèles avec les hyperparamètres par défaut de scikit-learn (ex. : *KNN avec KNeighborsClassifier() sans modification*)

[https://scikit-learn.org/stable/supervised\\_learning.html](https://scikit-learn.org/stable/supervised_learning.html)

- Optimisation des hyperparamètres des modèles utilisés afin d'améliorer les performances (tester plusieurs configurations d'un même modèle et sélectionner la meilleure)

# Machine

## Learning

Apprentissage  
supervisé :  
Classification

Consignes  
générales



### ➤ Travail demandé

#### 5. Évaluation des modèles

- Calcul des métriques de performance : Matrice de confusion, Accuracy, Precision, Recall et F1-score
- Comparaison des performances :
  - entre les différents modèles de classification
  - entre les modèles avant et après l'optimisation des hyperparamètres

#### 6. Analyse et interprétation

- Identifier le meilleur modèle
- Justifier les choix effectués
- Analyser les erreurs de classification

#### ⚠ Règles importantes

- ✗ Pas d'imputation de la variable cible
- ✓ Les variables spatiales sont utilisées uniquement pour les cartes et les visualisations, et ne sont pas intégrées dans les modèles
- ✓ Même jeu de données et même split pour tous les modèles
- ✓ Justification obligatoire de chaque décision de prétraitement

# Machine Learning

Apprentissage supervisé :  
Classification

Consignes générales



## ➤ Livrables attendus

### 1. Notebook Jupyter :

- ❖ Le notebook doit être clair, structuré et commenté, et contenir obligatoirement :
  - Titre du projet
  - Noms des membres du groupe
  - Description du jeu de données
  - Analyse exploratoire et visualisation des données
    - visualisations avant le prétraitement
    - visualisations après le prétraitement
  - Détail du prétraitement des données
  - Implémentation des modèles de classification
  - Évaluation et comparaison des modèles
- ❖ Chaque cellule de code doit être accompagnée d'un commentaire ou d'une interprétation (***cellule Markdown ou commentaire dans le code***)
- ❖ Nom du fichier obligatoire : projetX\_geo\_info.ipynb (X= {1, 2, 3, 4, 5})

**Le notebook doit s'exécuter sans erreur du début à la fin**

# Machine Learning

Apprentissage supervisé :  
Classification

Consignes générales



## ➤ Livrables attendus

### 1. Notebook Jupyter :

Bonus :

- Menu interactif permettant de choisir :
  - le modèle de classification
  - les hyperparamètres du modèle
  - ....
- Interaction utilisateur via des entrées clavier (input)
- Visualisations supplémentaires (courbes, graphiques comparatifs) + Interprétation
- Test du modèle sur de nouvelles données
- Commentaires et interprétation détaillée des résultats
- Notebook bien structuré et lisible
- ...

# Machine Learning

## Apprentissage supervisé : Classification

### Consignes générales



#### ➤ Livrables attendus

##### 2. Rapport synthétique (PDF) (un seul rapport par groupe)

Le rapport doit être clair, structuré et analytique (max. 20 pages) et ne doit pas se limiter à un copier-coller du notebook Jupyter. Il se compose des sections suivantes :

- **Page de garde** (Titre du projet, Noms des membres du groupe, Filière / année, ...)
- Table de matière, Listes des figures et des tableaux, des abréviations, ...
- **Introduction** : Contexte du projet, domaine d'application, objectif, ...
- **Description du Data**: Présentation des variables, explication des variables, ...
- **Prétraitement des données** : Problèmes identifiés (valeurs manquantes, doublons, incohérences), solutions appliquées et justification des choix, Impact du prétraitement sur les données, ...
- **Modélisation** : Présentation et explication des modèles de classification utilisés, les hyperparamètres des modèles ...
- **Évaluation et résultats** : Métriques utilisées, tableau comparatif des performances, comparaison avant / après optimisation, ...
- **Discussion** : Analyse des résultats obtenus, Forces et limites des modèles, Difficultés rencontrées, ...
- **Conclusion et perspectives** : Synthèse du travail, Améliorations possibles

# Machine Learning

## Apprentissage supervisé : Classification

### Mini-projet 1



#### Mini-projet 1 : Classification de l'occupation du sol (Land Use / Land Cover)

- Data *geo\_lulc\_data.csv*
- Objectif : Classer des zones géographiques selon leur type d'occupation du sol
- Variables explicatives
  - **NDVI** : indice de végétation (couverture végétale)
  - **altitude\_m** : altitude du terrain (mètres)
  - **pente\_pct** : pente du terrain (%)
  - **dist\_route\_km** : distance aux routes (km)
  - **texture\_sol** : indice de texture du sol
  - **chlorophyll\_index\_lab** : indice de chlorophylle issu d'analyses ponctuelles
- Variables spatiales
  - **latitude** : coordonnée géographique
  - **longitude** : coordonnée géographique
  - **zone\_id** : identifiant de la zone spatiale
- Variable cible
  - **classe\_lulc** : type d'occupation du sol (*Forêt, Urbain, Agricole, Eau*)

# Machine

## Learning

Apprentissage  
supervisé :  
Classification

Mini-projet 2



## Mini-projet 2 : Classification des zones à risque d'inondation

- **Data** *geo\_flood\_risk\_data.csv*
- **Objectif** : Identifier le niveau de risque d'inondation d'une zone.
- **Variables explicatives**
  - **altitude\_m** : altitude du terrain
  - **pente\_pct** : pente (%)
  - **distance\_riviere\_m** : distance au cours d'eau
  - **precip\_mm** : cumul des précipitations (mm)
  - **perm\_sol** : perméabilité du sol
  - **debit\_station\_amont** : débit mesuré à une station hydrométrique
- **Variables spatiales**
  - **latitude** : coordonnée géographique
  - **longitude** : coordonnée géographique
  - **zone\_id** : identifiant de la zone spatiale
- **Variable cible**
  - **niveau\_risque** : niveau de risque d'inondation (*Faible, Moyen, Élevé*)



## Mini-projet 3 : Classification de zones urbaines selon leur densité

- **Data** *geo\_urban\_density\_data.csv*
- **Objectif** : Classer les zones urbaines selon leur densité et usage
- **Variables explicatives**
  - **densite\_batiments** : densité des bâtiments
  - **densite\_population** : densité de population (par km<sup>2</sup>)
  - **taux\_impermeabilisation** : taux de surfaces imperméabilisées (%)
  - **distance\_centre\_ville\_km** : distance au centre urbain
  - **hauteur\_moyenne\_batiments** : hauteur moyenne des bâtiments
  - **wifi\_density\_probe** : densité estimée de signaux Wi-Fi dans la zone
- **Variables spatiales**
  - **latitude** : coordonnée géographique
  - **longitude** : coordonnée géographique
  - **zone\_id** : identifiant de la zone spatiale
- **Variable cible**
  - **classe\_urbaine** : classe de densité urbaine (*faible, moyenne, élevée*)

# Machine Learning

## Apprentissage supervisé : Classification

### Mini-projet 4



#### Mini-projet 4 : Classification de l'aptitude des sols agricoles

- **Data** *geo\_agri\_aptitude\_data.csv*
- **Objectif** : Classifier les parcelles agricoles selon leur niveau d'aptitude à la production
- **Variables explicatives**
  - **pente\_pct** : pente du terrain (%)
  - **humidite\_sol** : niveau d'humidité du sol
  - **humidite\_sol** : niveau d'humidité du sol
  - **precip\_mm** : cumul des précipitations (mm)
  - **temp\_moy\_C** : température moyenne (°C)
  - **texture\_sol** : texture du sol
  - **ph\_sol\_lab** : pH mesuré en laboratoire
- **Variables spatiales**
  - **latitude** : coordonnée géographique
  - **longitude** : coordonnée géographique
  - **parcelle\_id** : identifiant de parcelle agricole
- **Variable cible**
  - **aptitude\_sol** : aptitude du sol (Faible, Moyenne, Élevée)

# Machine Learning

## Apprentissage supervisé : Classification

### Mini-projet 5



## Mini-projet 5 : Classification des zones selon leur sensibilité environnementale

- Data *geo\_sensibilite\_env\_data.csv*
- Objectif : Classifier des zones géographiques selon leur niveau de sensibilité environnementale
- Variables explicatives
  - **ndice\_vegetation** : couverture végétale
  - **distance\_eau\_m** : proximité des ressources en eau
  - **pente\_pct** : pente du terrain (%)
  - **pression\_humaine** : indice de pression anthropique
  - **qualite\_sol** : indice global de qualité du sol
  - **indice\_biodiversite\_lab** : indice issu d'inventaires écologiques
- Variables spatiales
  - **latitude** : coordonnée géographique
  - **longitude** : coordonnée géographique
  - **zone\_id** : identifiant de la zone spatiale
- Variable cible
  - **sensibilite\_env** : sensibilité environnementale (*Faible, Moyenne, Élevée*)

# Machine Learning

## Apprentissage supervisé : Classification

### Mini-projet 6



### Mini-projet 6 : Classification de la stabilité des terrains (risque de glissement)

- **Data** *geo\_stabilite\_terrain\_data.csv*
- **Objectif** : Classifier les zones géographiques selon leur niveau de stabilité du terrain afin d'identifier les zones à risque de glissement
- **Variables explicatives**
  - **pente\_pct** : pente du terrain (%)
  - **altitude\_m** : altitude du terrain (mètres)
  - **texture\_sol** : type / texture du sol
  - **humidite\_sol** : niveau d'humidité du sol
  - **distance\_faille\_km** : distance aux failles géologiques (km)
  - **couverture\_vegetale** : taux de couverture végétale
  - **indice\_geotech\_lab** : indice géotechnique issu de tests ponctuels en laboratoire
- **Variables spatiales**
  - **latitude** : coordonnée géographique
  - **longitude** : coordonnée géographique
  - **zone\_id** : identifiant de la zone spatiale
- **Variable cible**
  - **stabilite\_terrain** : niveau de stabilité du terrain (*stable, moyennement stable, instable*)