

Détection du mildiou de la pomme de terre par imagerie grâce aux méthodes de Machine Learning

Yasmine BOUCHIBTI

Leslie CIETERS

Meryem GRIMAJ

Sommaire

1

Introduction

2

Le jeux de données

3

Modèles de
Machine Learning

4

Modèles de Deep
Learning

5

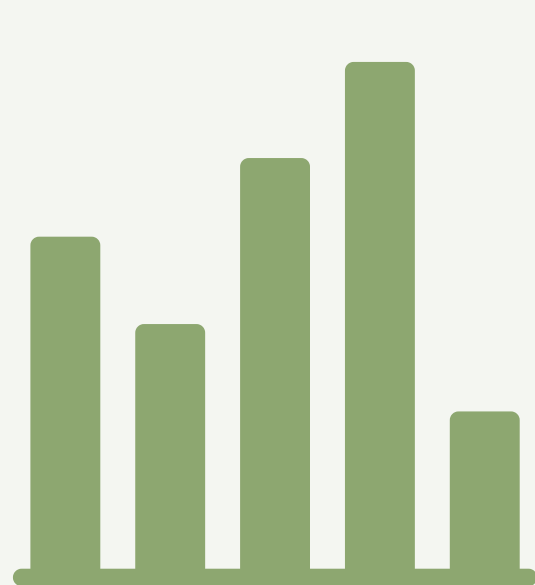
Limites et
perspectives

6

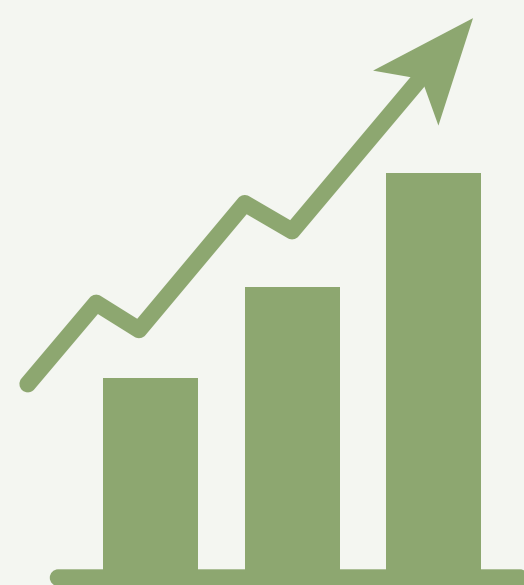
Conclusion

Introduction

- *Alternaria solani* (mildiou précoce)
- *Phytophthora infestans* (mildiou tardif)



Jusqu'à 80% de
pertes

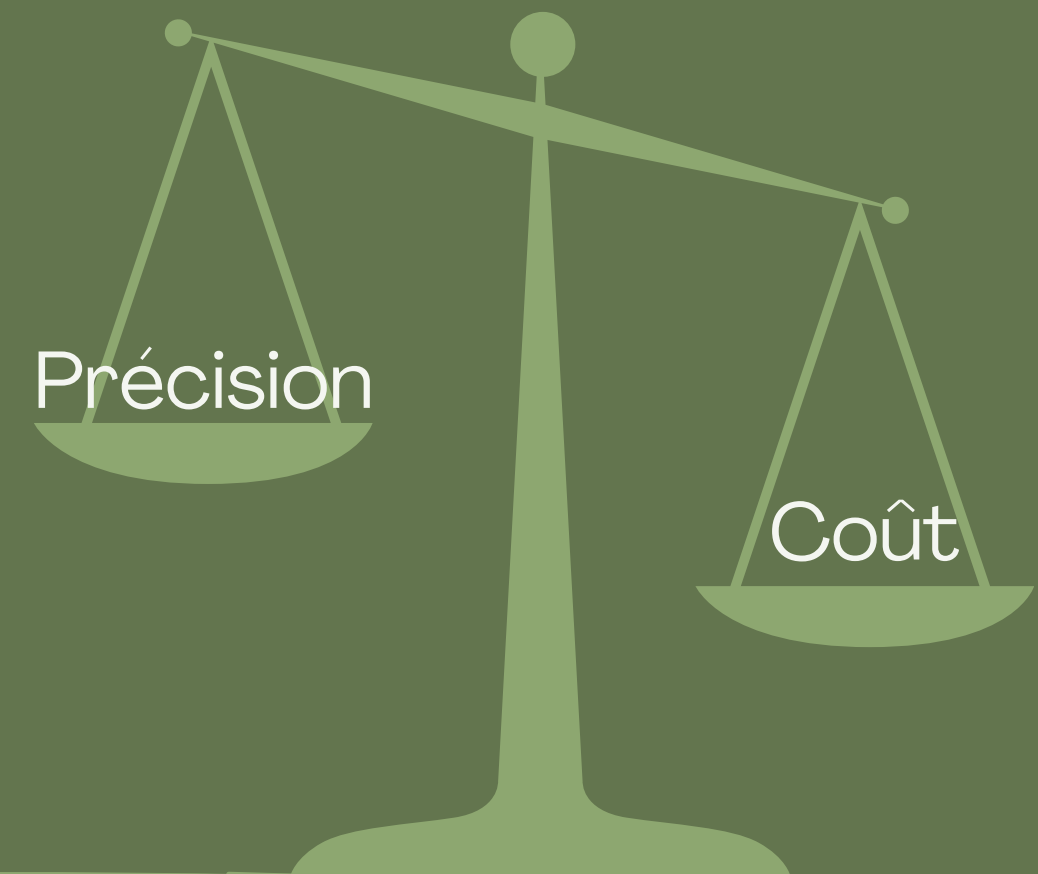
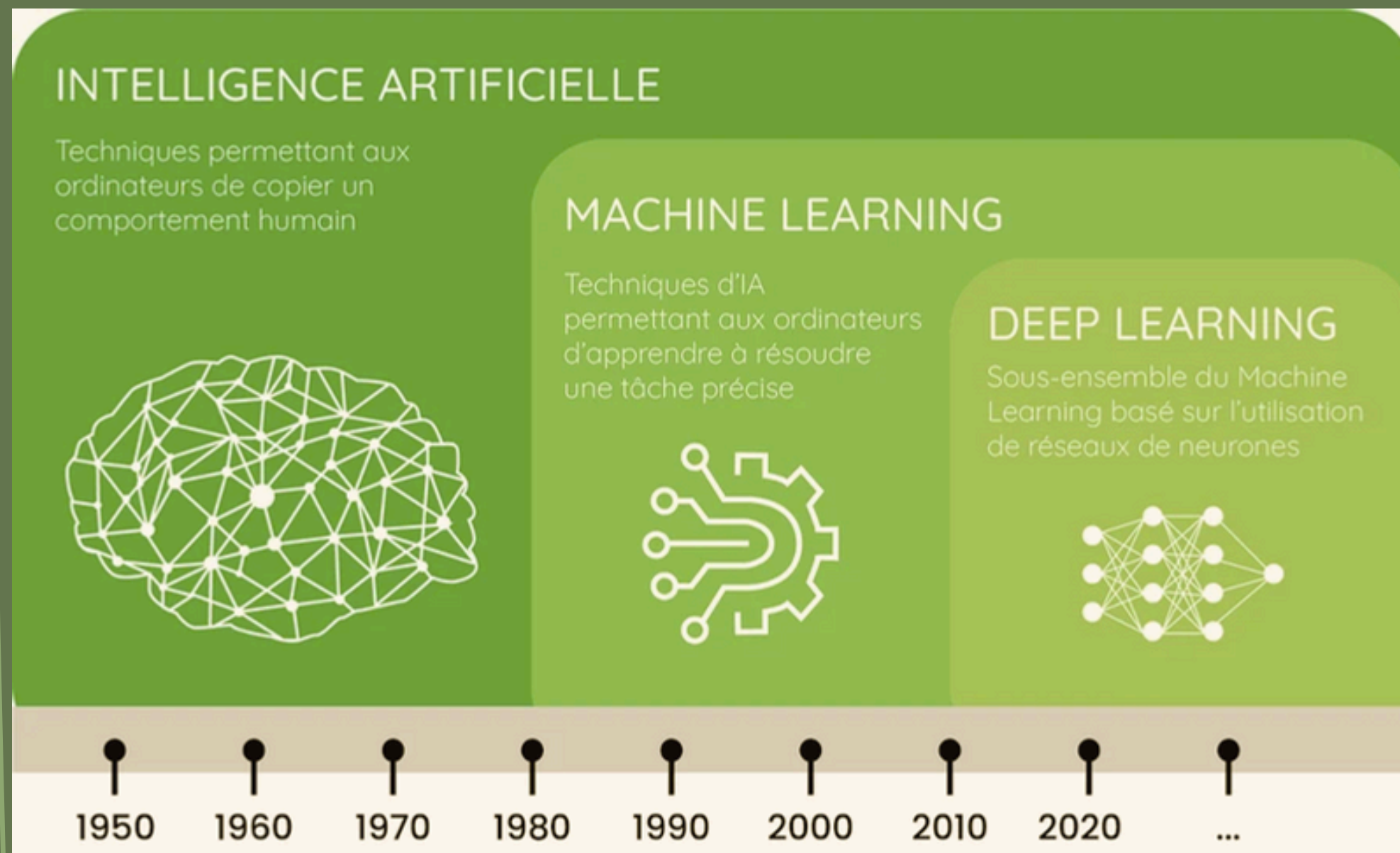


Utilisation de
produits
phytosanitaires



la grande famine
1845-1849

Machine Learning et Deep Learning





Le jeu de données

Imagerie

Le jeu de données

2152 images réparties en
trois classes :

- Mildiou précoce
- Mildiou tardif
- Sain

Mildiou précoce



Mildiou précoce



Mildiou précoce



Mildiou précoce



Mildiou précoce



Mildiou tardif



Mildiou tardif



Mildiou tardif



Mildiou tardif



Mildiou tardif



Sain



Sain



Sain

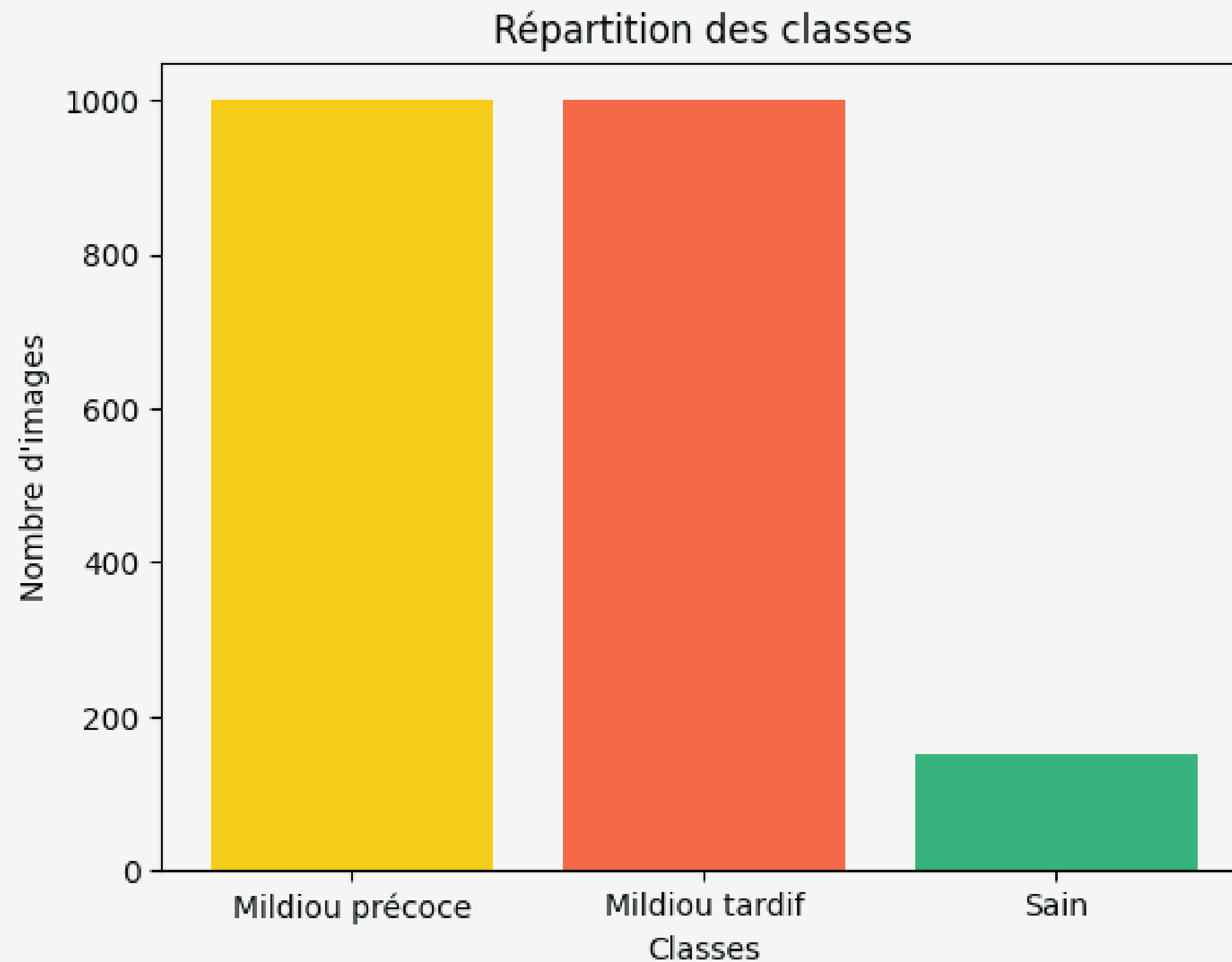


Sain



Sain

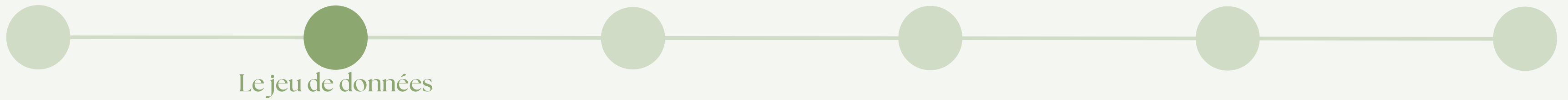




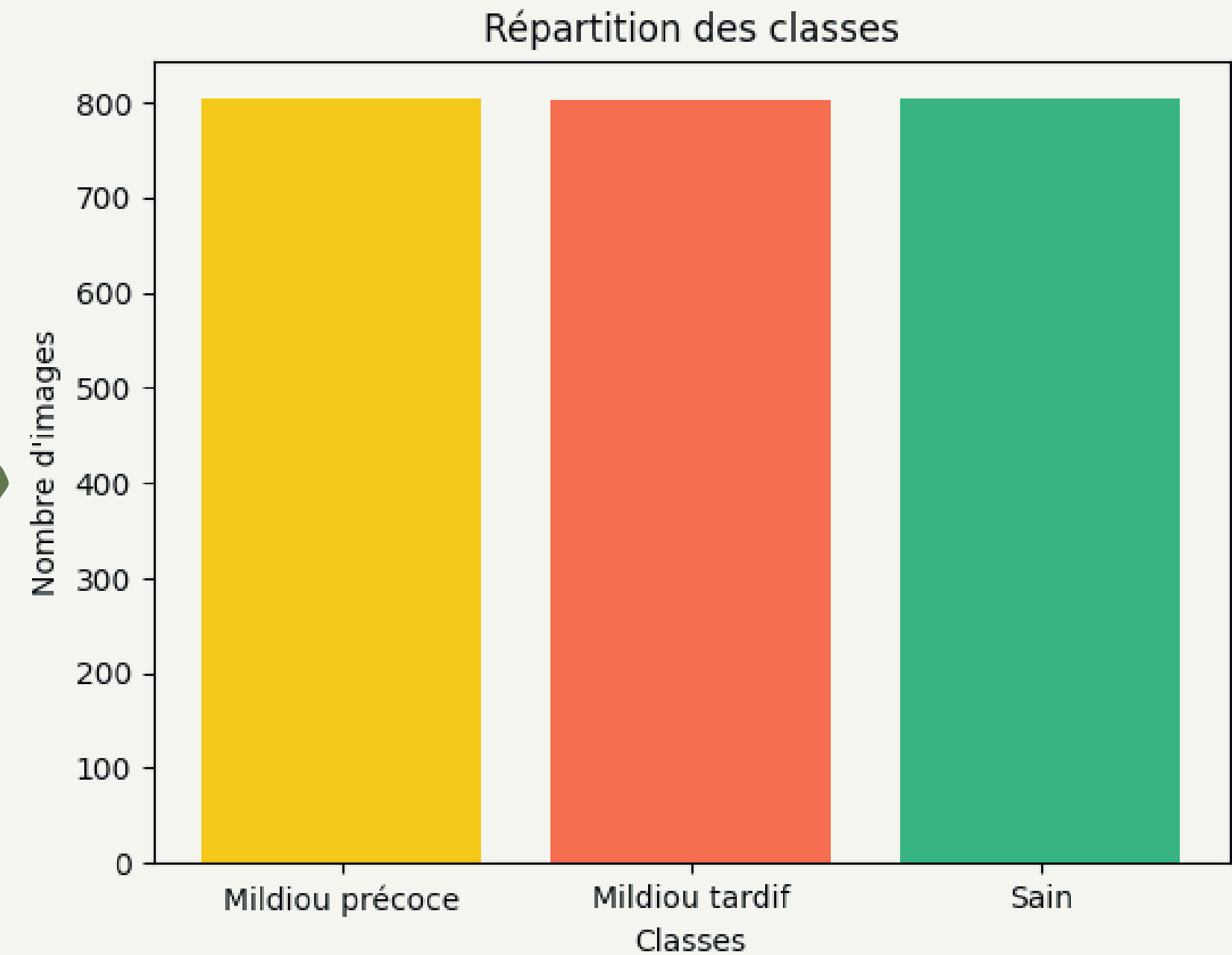
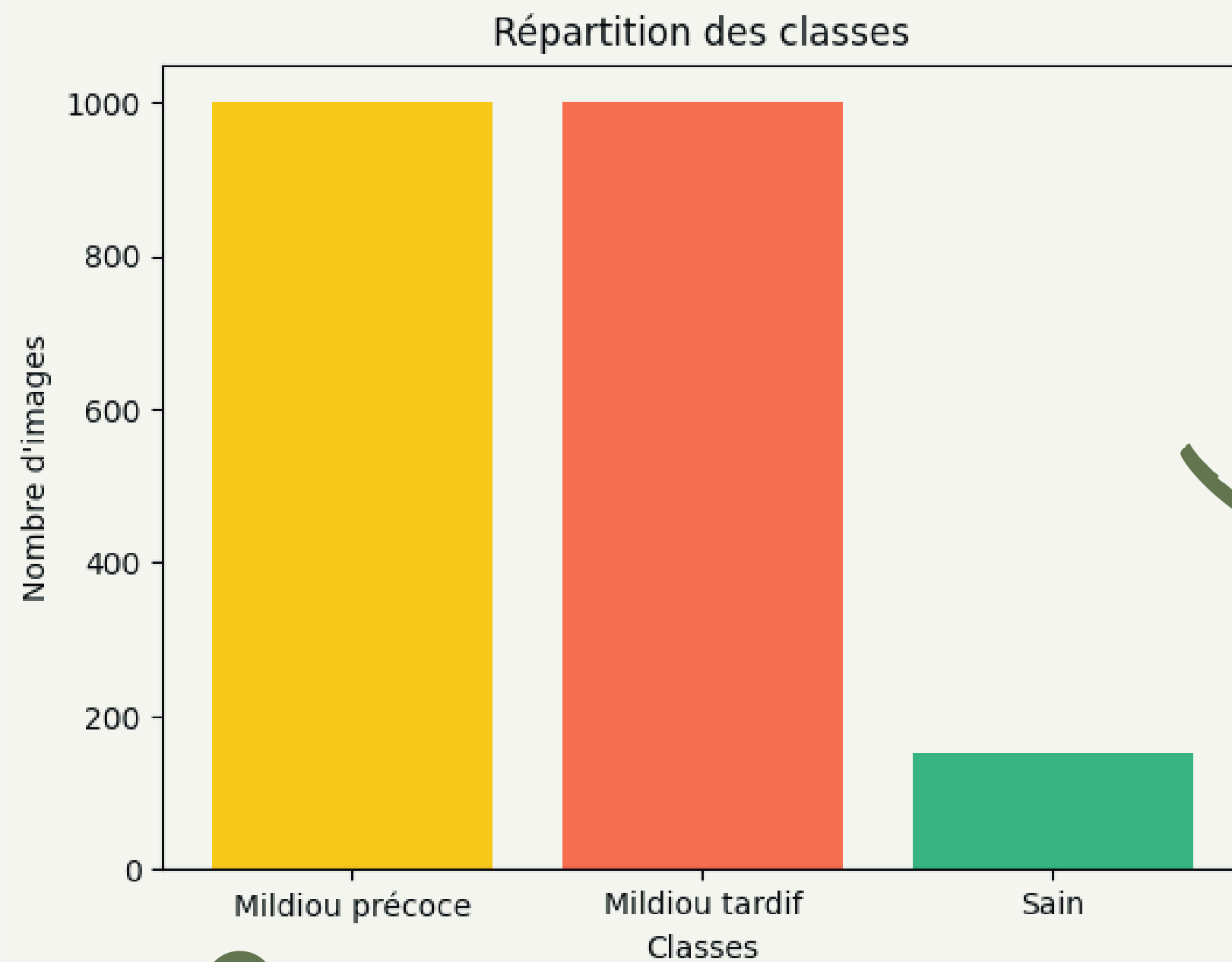
Déséquilibre important des classes



Technique SMOTE pour le
suréchantillonnage de la classe
minoritaire



- Séparation du jeu de données (80/20) pour l'entraînement et la validation

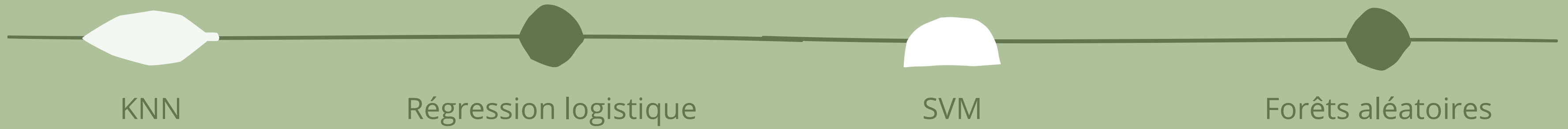


Uniquement sur le jeu de données d'entraînement



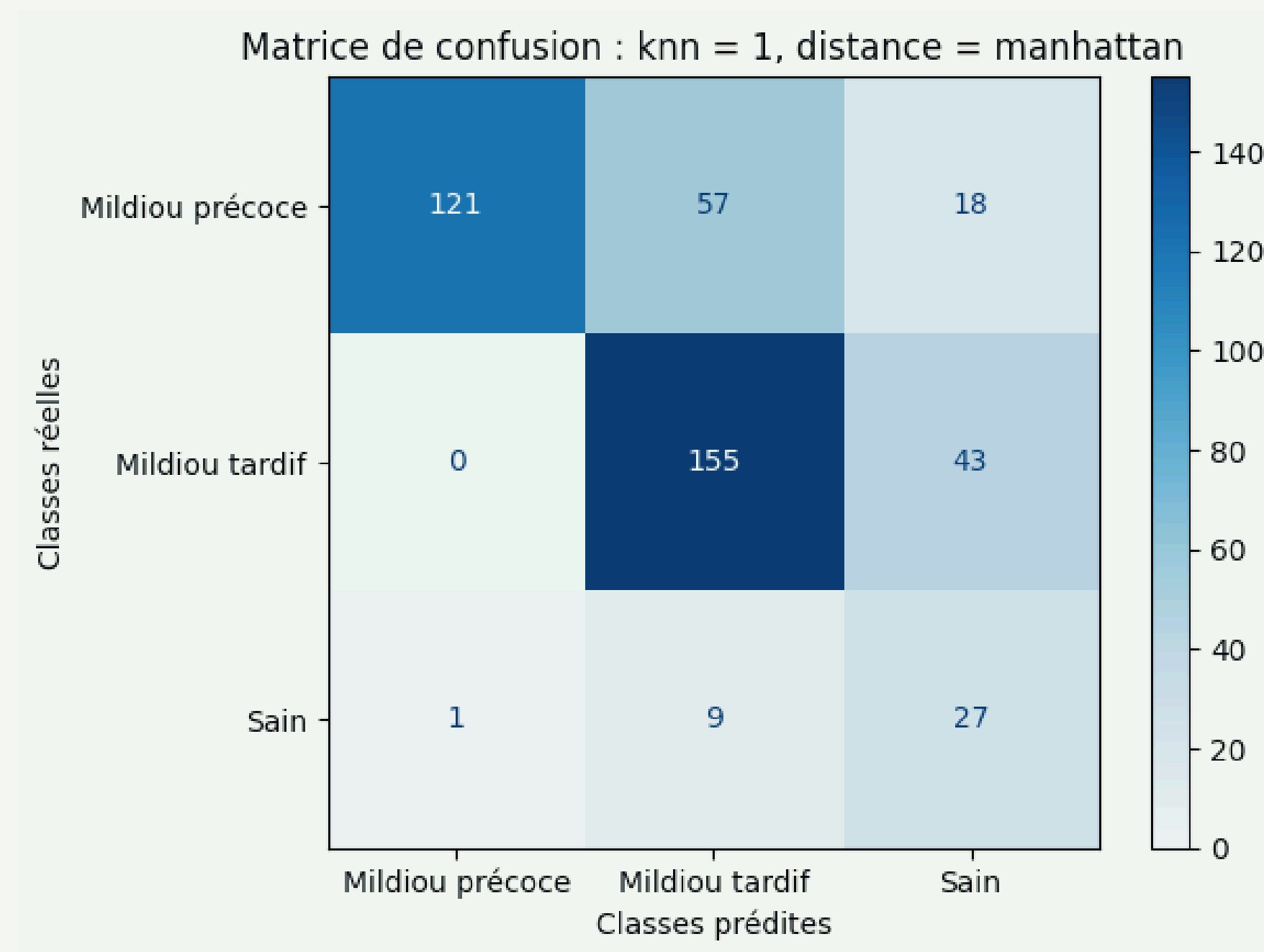


Modèles de Machine Learning

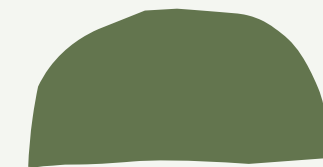


KNN

Comparaison de plusieurs modèles afin de définir les meilleurs paramètres



- Accuracy = 0.70
- Taux de faux positifs = 0.15
- Taux de faux négatifs = 0.27

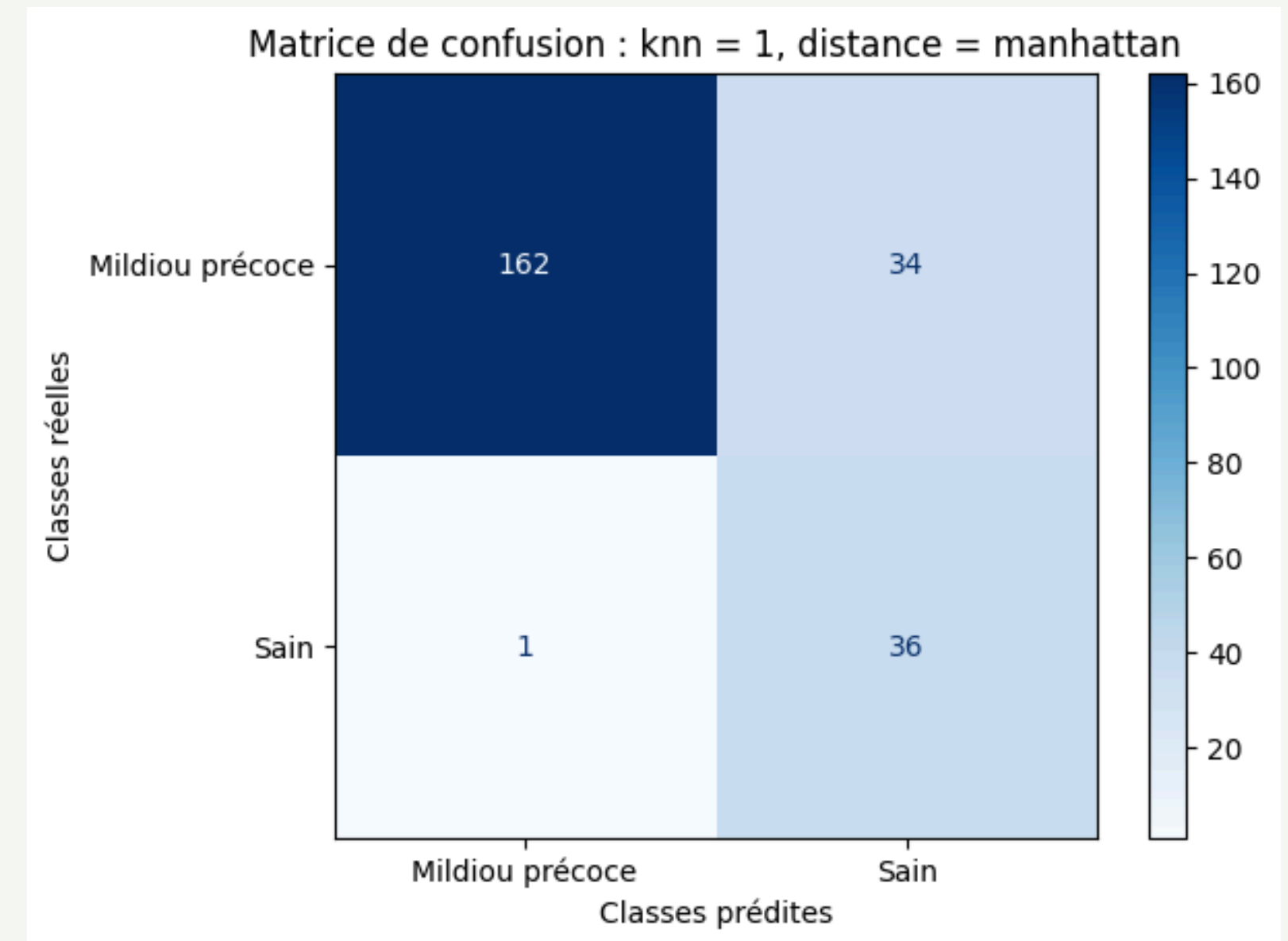
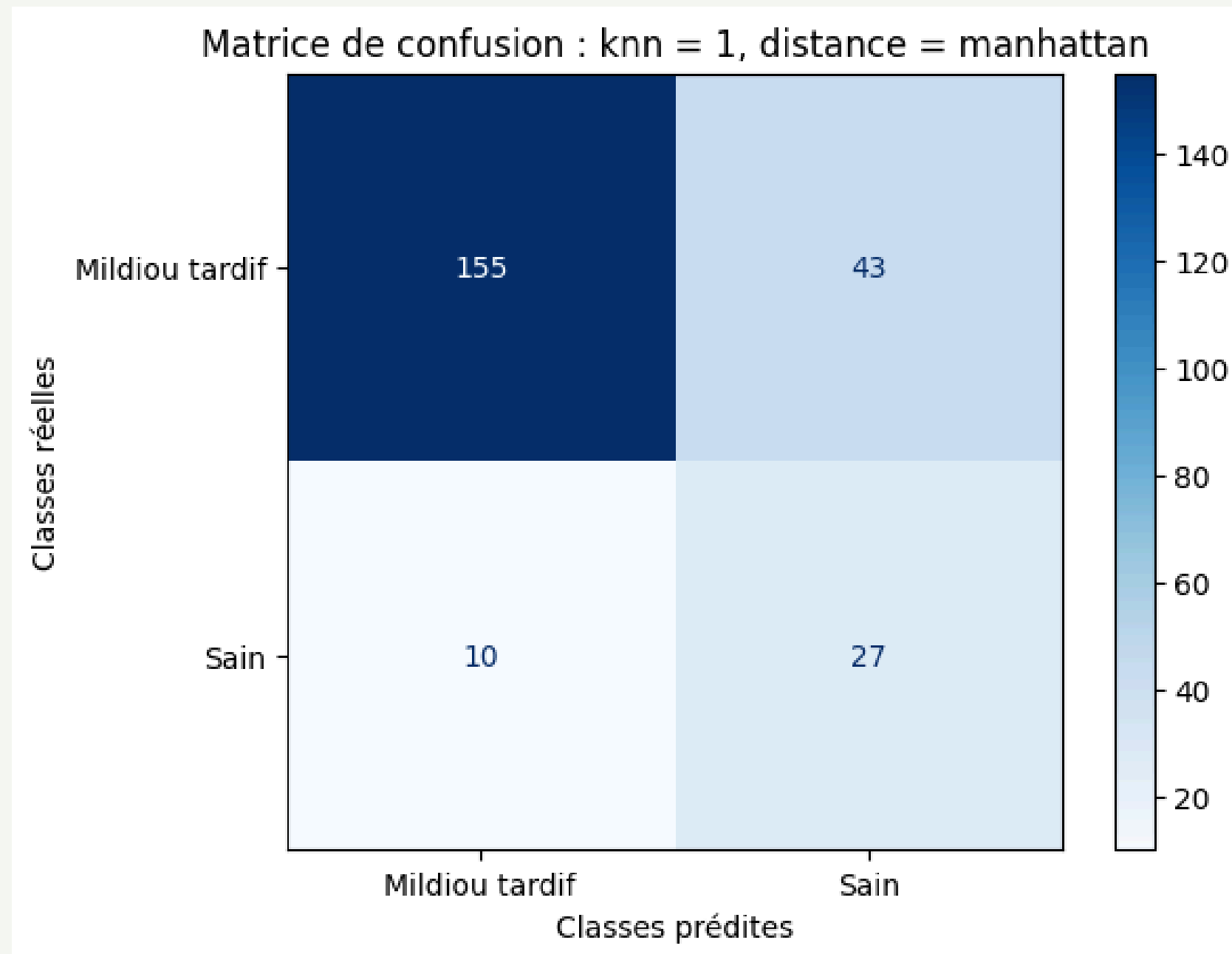


KNN

En séparant les classes

Accuracy = 0.77

Accuracy = 0.85



Régression logistique

1

**Distinction
entre Mildiou
précoce et
sain**

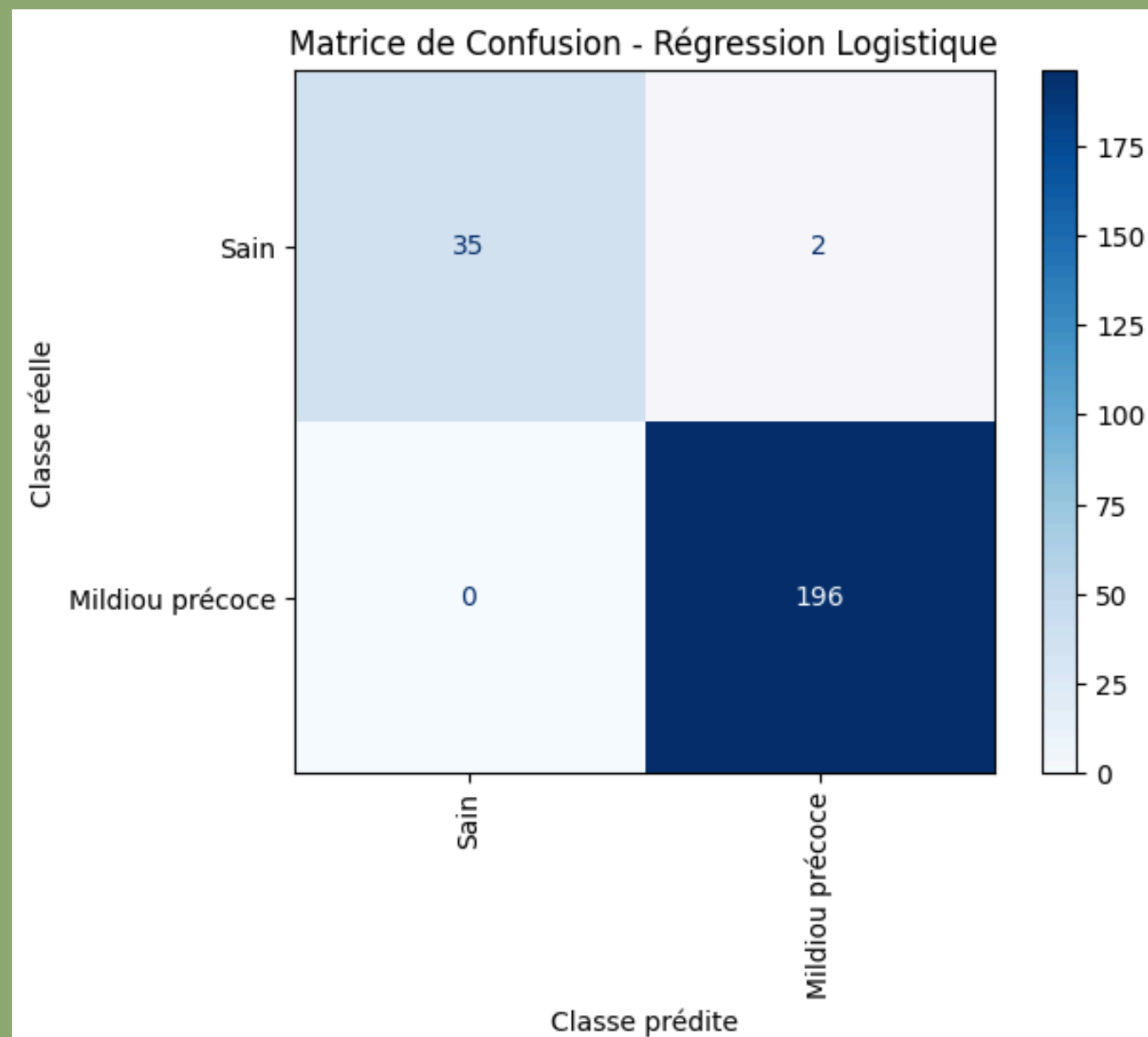
2

**Distinction
entre Mildiou
tardif et sain**

Résultats

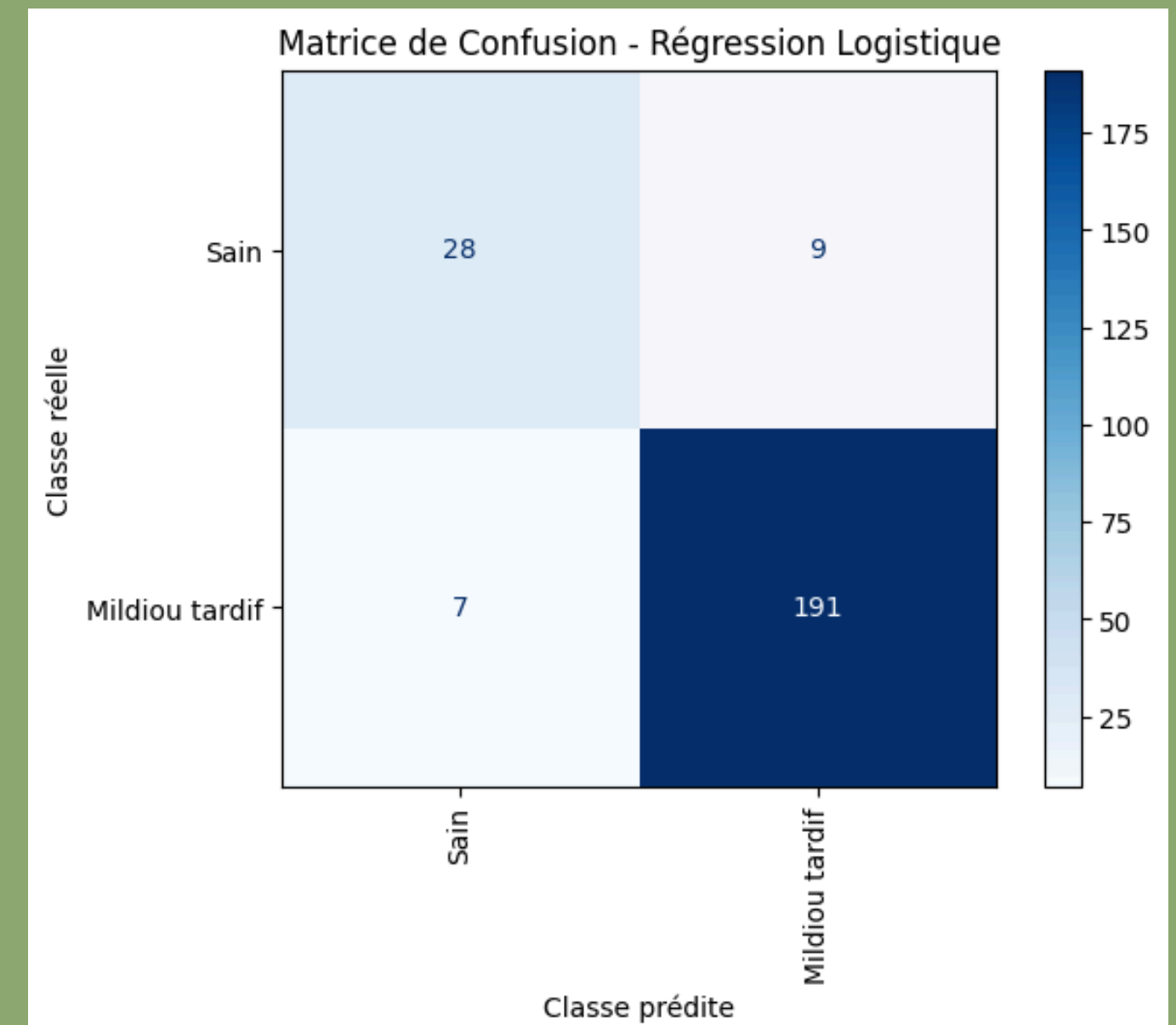
1

Accuracy = 0.99



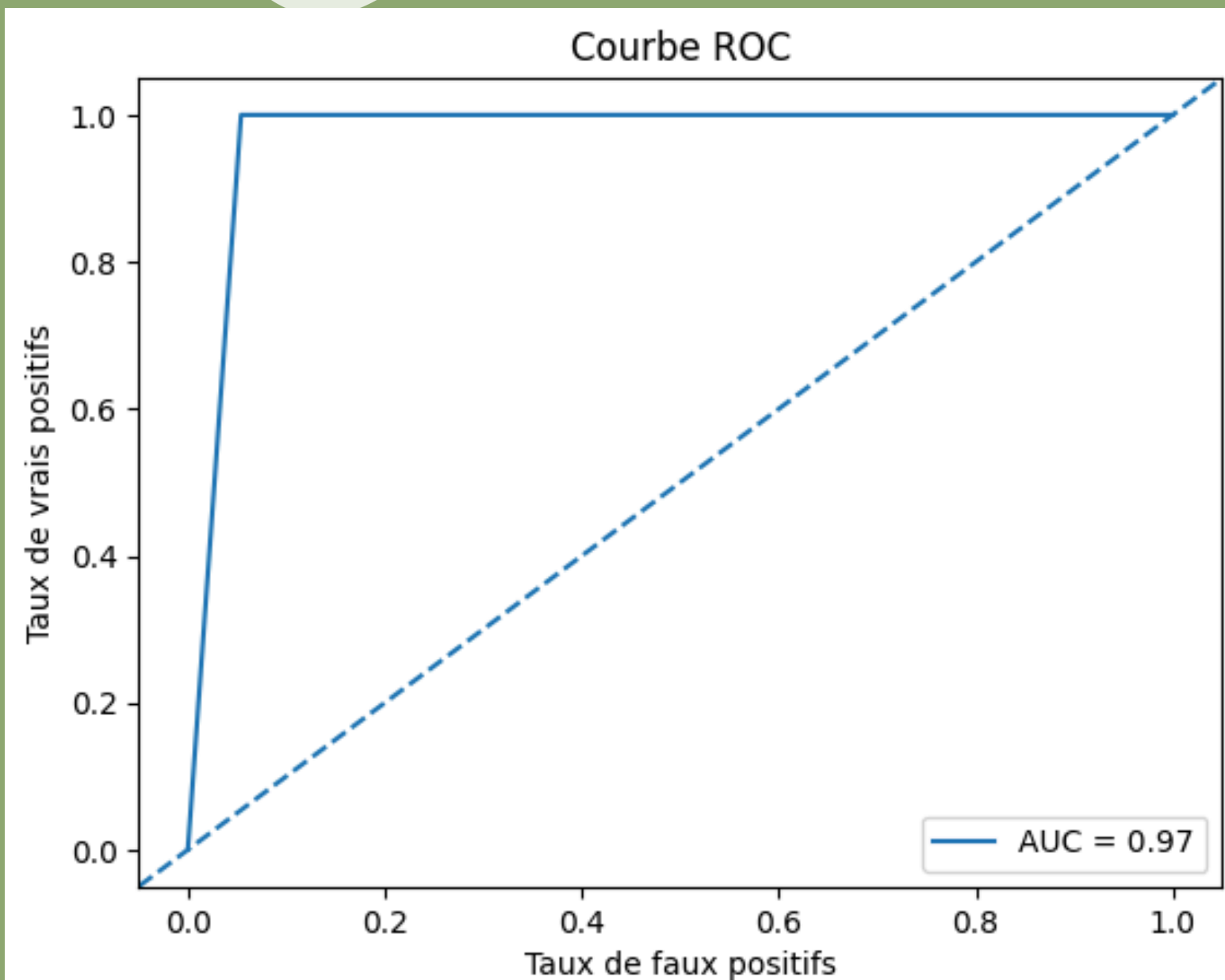
2

Accuracy = 0.93

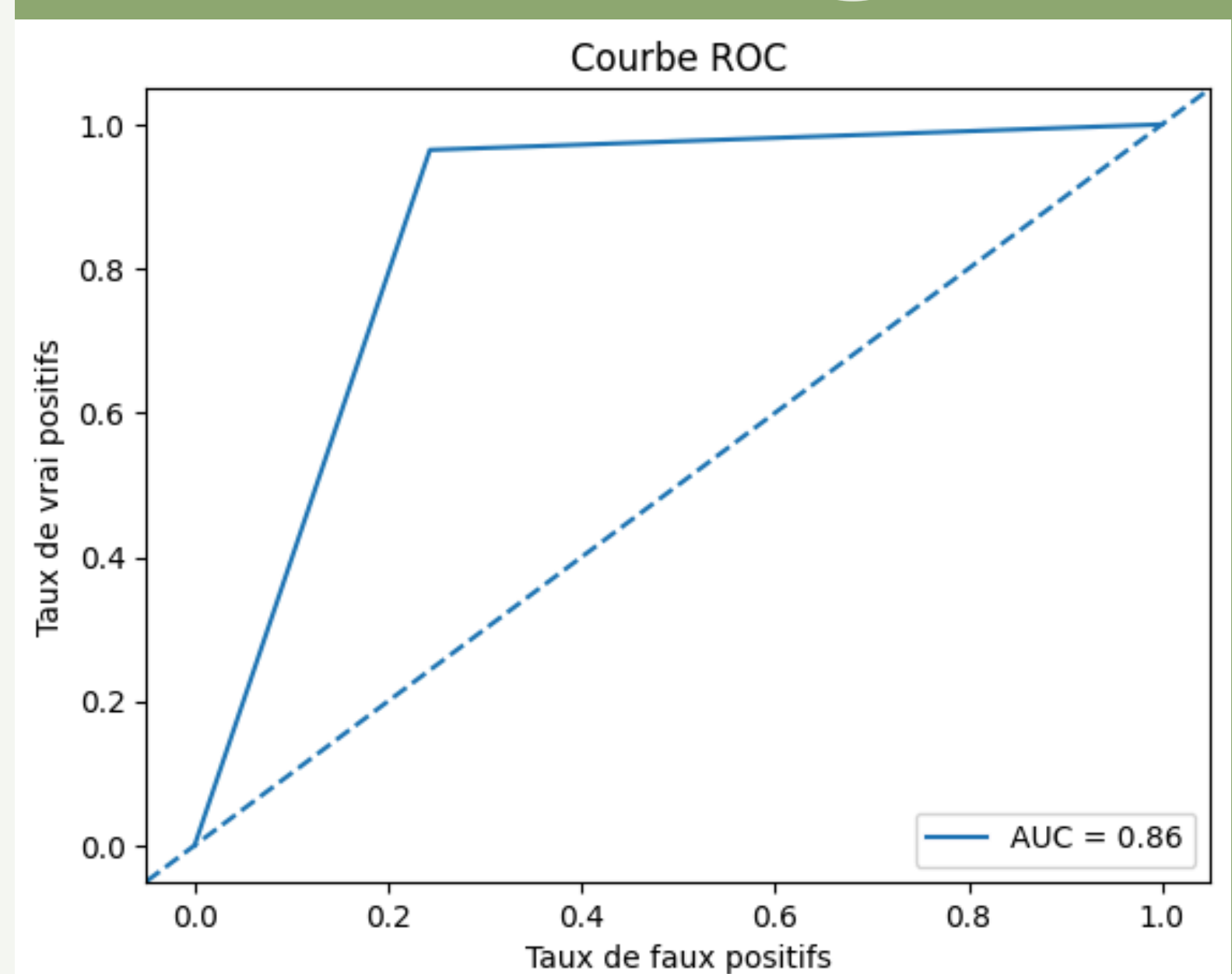


Résultats

1



2





Modèles de Machine Learning

SVM

Objectifs de la recherche

1

**Choix de la
méthode**

2

**Choix des
hyperparamè-
tres**

Choix de la méthode

Ramener un problème de classification ou de discrimination à un hyperplan

1

2

Efficace dans les espaces de haute dimension

Utilise un sous-ensemble de points d'entraînement dans la fonction de décision (appelé vecteurs de support), il est donc également efficace en mémoire.

3

4

Différentes fonctions du noyau peuvent être spécifiées pour la fonction de décision

Choix des hyperparamètres

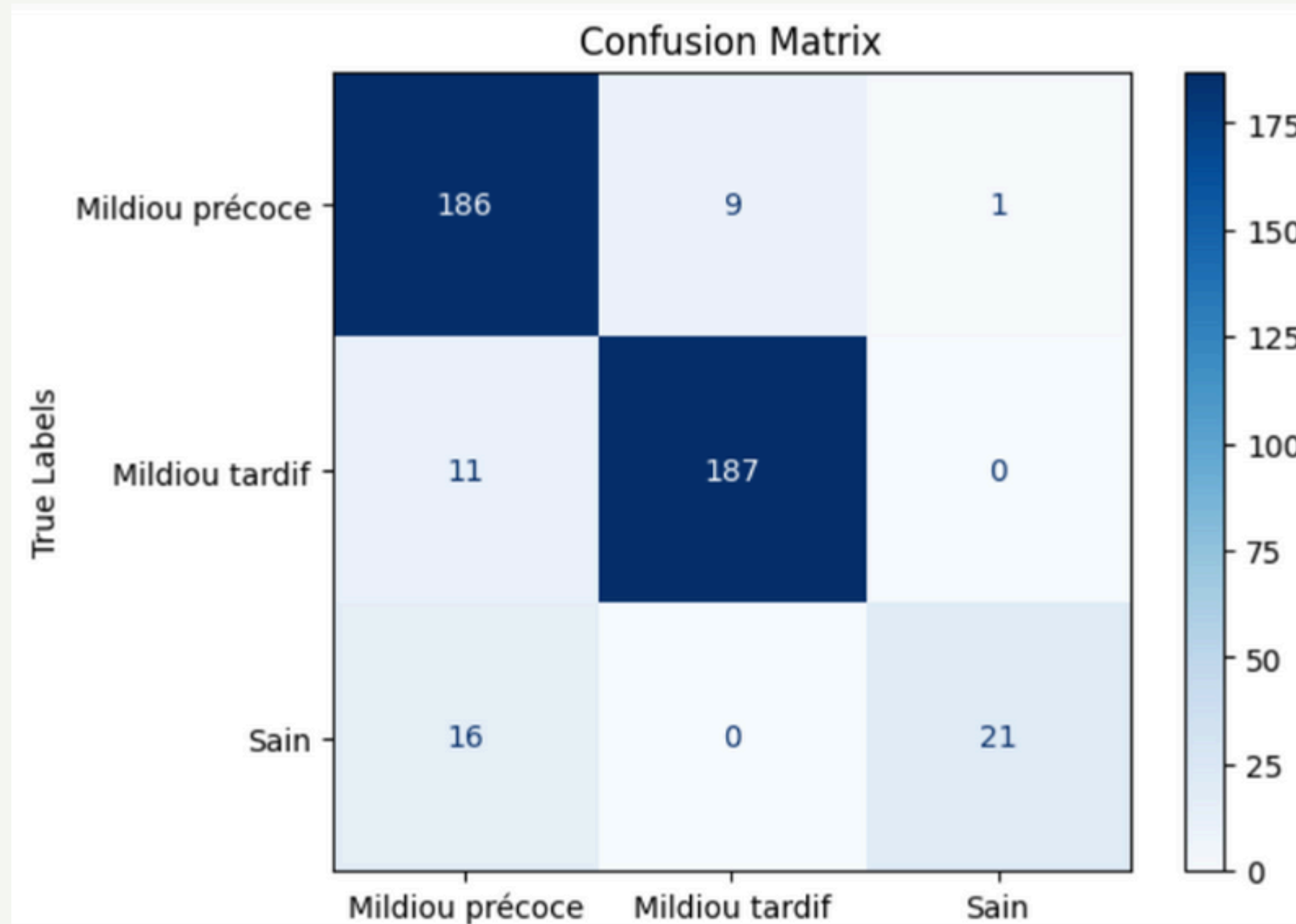
- un contre un pour SVC : des classificateurs sont construits et chacun d'entre eux forme des données à partir de deux classes.
- kernel (noyau) : P
- decision_function_shape : On choisit d'appliquer la méthode de classificateur un contre le reste au lieu d'un contre un

Résultats

Accuracy: 91.42%

	precision	recall	f1-score	support
0	0.87	0.95	0.91	196
1	0.95	0.94	0.95	198
2	0.95	0.57	0.71	37
accuracy			0.91	431
macro avg	0.93	0.82	0.86	431
weighted avg	0.92	0.91	0.91	431

Modèles de Machine Learning





CONCLUSION POUR LES SVM

Le modèle fonctionne très bien pour les classes majoritaires (0 et 1) avec des précisions, rappels et F1-scores élevés

DecisionTreeClassifier

Objectifs de la recherche

1

**Choix de la
méthode**

2

**Choix des
hyperparamètres**

--

CHOIX DE LA MÉTHODE



Choix des hyperparamètres

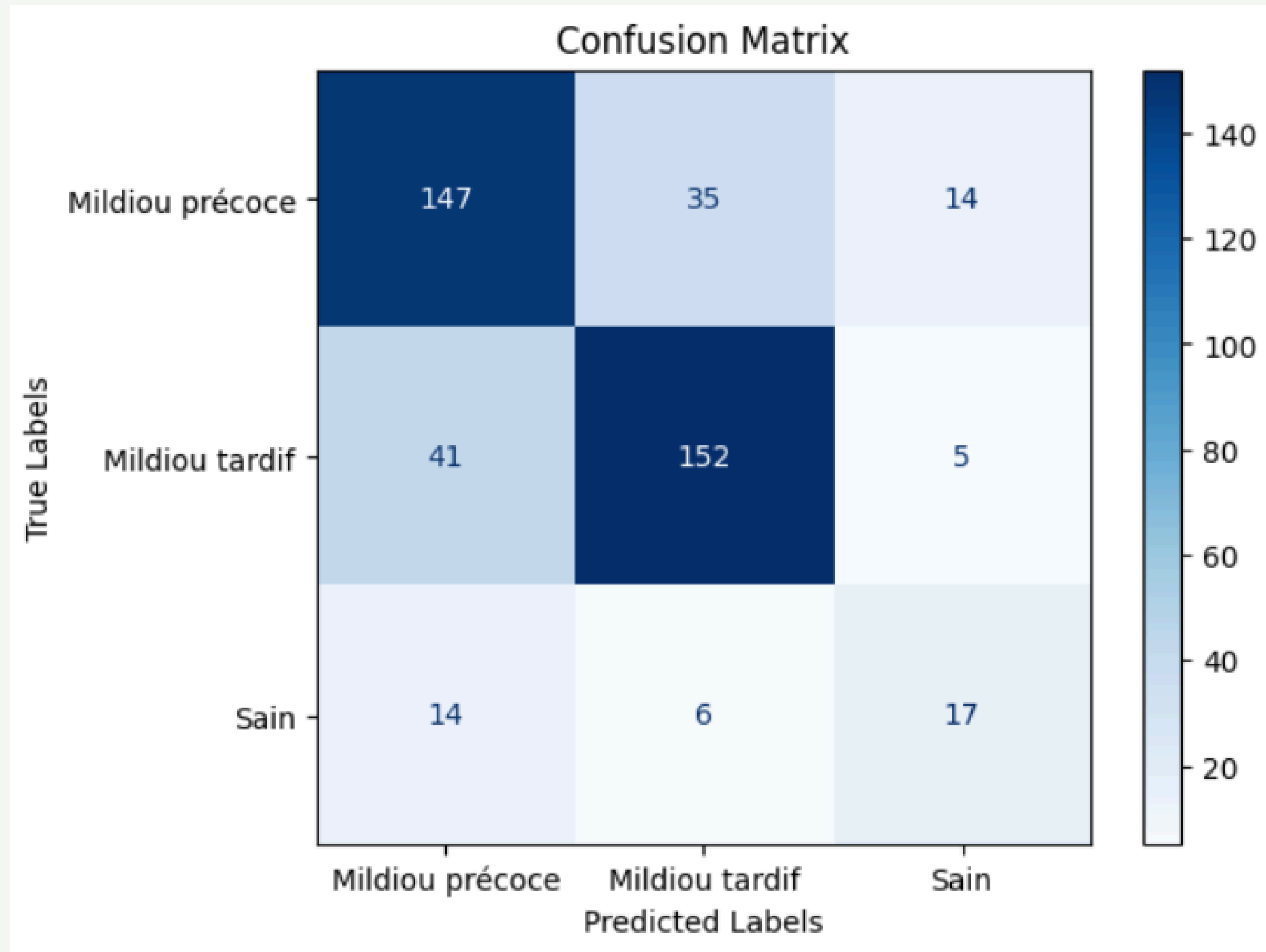
- `max_depth=10` : Limite la profondeur de l'arbre à 10 niveaux. Cela empêche l'arbre de devenir trop complexe et de surapprendre sur les données d'entraînement
- `min_samples_split=2` : C'est le nombre minimum d'échantillons requis pour diviser un nœud. Avec une valeur de 2, on autorise le modèle à diviser un nœud dès qu'il y a au moins 2 échantillons.
- `min_samples_leaf=1` : spécifie le nombre minimum d'échantillons dans une feuille. on autorise des feuilles avec un seul échantillon.
- `criterion='gini'` : l'objectif est de minimiser l'impureté (la diversité des classes) dans les nœuds.

Résultats

Decision Tree Accuracy: 0.73

	precision	recall	f1-score	support
0	0.73	0.75	0.74	196
1	0.79	0.77	0.78	198
2	0.47	0.46	0.47	37
accuracy			0.73	431
macro avg	0.66	0.66	0.66	431
weighted avg	0.73	0.73	0.73	431

Modèles de Machine Learning





CONCLUSION POUR LES DECISION TREE

DecisionTreeClassifier donne des résultats acceptables sur les classes majoritaires, il a des difficultés à traiter la classe minoritaire (classe 2)

Random Forest

Objectifs de la recherche

1

**Choix de la
méthode**

2

**Choix des
hyperparamètres**

Choix de la méthode

Random Forest est également très résistant au bruit dans les données

1

2

Contrairement à un modèle de décision unique (comme le `DecisionTreeClassifier`), le Random Forest utilise plusieurs arbres, ce qui réduit la variance du modèle.

Le modèle Random Forest est connu pour ses performances solides sur les problèmes de classification grâce à la combinaison de plusieurs arbres de décision.

3

4

Mieux équipé pour traiter des classes déséquilibrées que des modèles plus simples, grâce à sa capacité à effectuer un échantillonnage aléatoire lors de la construction de chaque arbre.

Choix des hyperparamètres

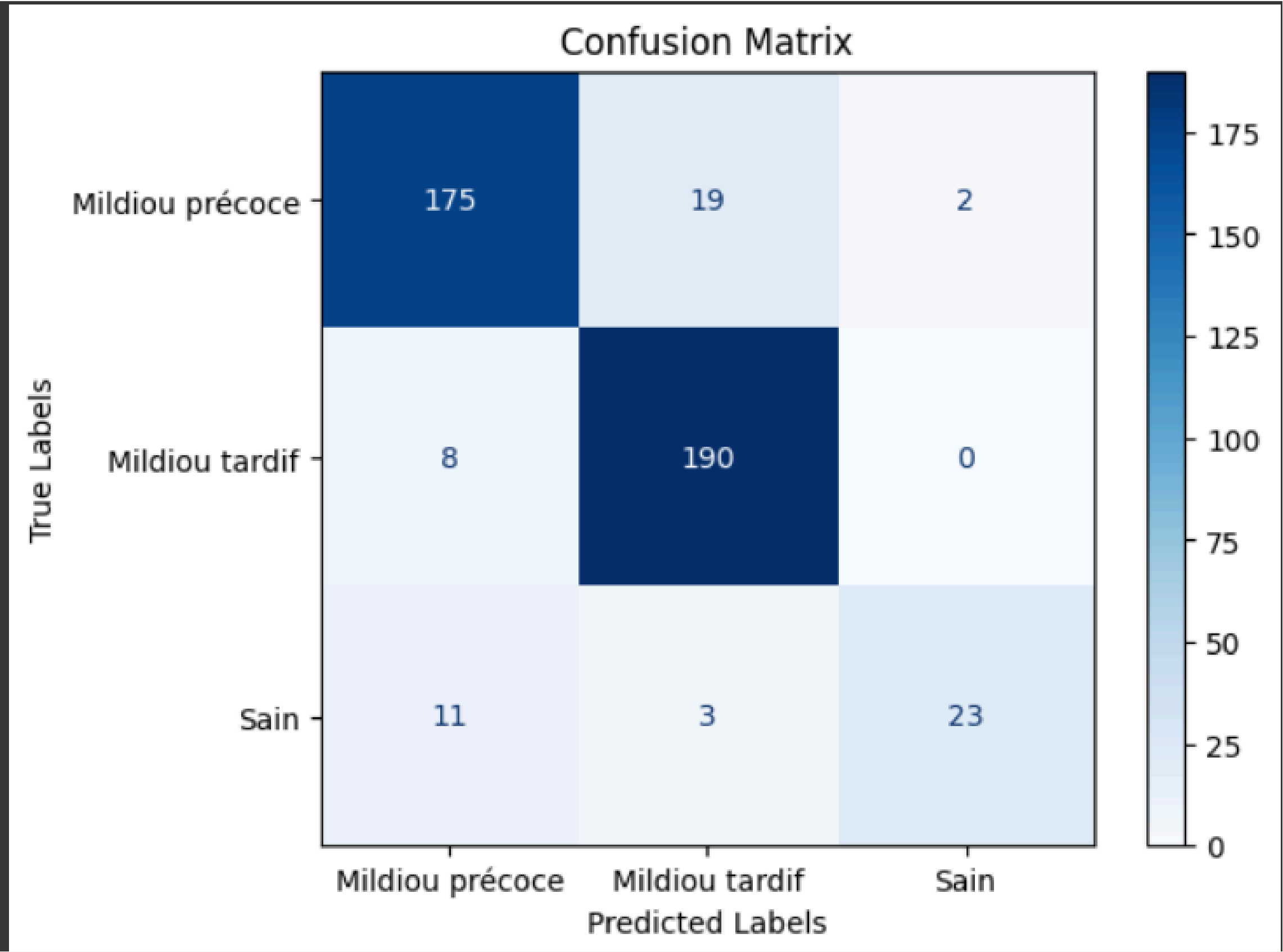
- `n_estimators=100` : 100 arbres de décision se qui améliore la stabilité des prédictions
- `max_features='sqrt'` : Pour chaque division dans un arbre, seulement la racine carrée du nombre total de caractéristiques sera considérée. Cela permet d'accélérer le calcul .

Résultats

Random Forest Accuracy: 0.90

	precision	recall	f1-score	support
0	0.90	0.89	0.90	196
1	0.90	0.96	0.93	198
2	0.92	0.62	0.74	37
accuracy			0.90	431
macro avg	0.91	0.82	0.86	431
weighted avg	0.90	0.90	0.90	431

Modèles de Machine Learning





CONCLUSION POUR LES DECISION TREE

Avec une accuracy de 90% et des précisions et F1-scores élevés pour les classes 0 et 1, le modèle Random Forest se comporte très bien pour les deux premières classes, qui ont plus de données d'entraînement. Le modèle a plus de difficultés à bien identifier la classe 2 (62% de rappel)



Modèle de Deep Learning

Modèle CNN

1

**Choix de
l'architecture**

2

Choix de l'architecture

Modèle CNN

1 couche dense + fonction d'activation Softmax

1 couche dense + fonction d'activation ReLu

1 couche pooling Pooling2D

1 couche convolutive Conv2D

1 couche pooling Pooling2D

CNN



Choix de l'architecture

Modèle CNN

1 couche dense + fonction d'activation Softmax

1 couche dense + fonction d'activation ReLu

1 couche pooling Pooling2D

1 couche convolutive Conv2D

1 couche pooling Pooling2D

CNN







Discussion

Limites de l'étude

Identifiez et discutez des limites et des défis rencontrés lors de votre recherche.

Expliquez les aspects de votre méthodologie, de votre échantillonnage ou d'autres facteurs qui pourraient avoir influencé les résultats obtenus.

L'objectif est de fournir une évaluation critique de votre recherche en reconnaissant ses limites et en suggérant des avenues pour des recherches futures plus approfondies et robustes.



Conclusion

Récapitulez de manière succincte les principaux résultats de votre étude et leurs implications. Soulignez l'importance de votre recherche dans le contexte plus large de votre domaine d'étude et mettez en évidence les contributions spécifiques que vous avez apportées. Terminez en réaffirmant l'objectif principal de votre recherche et en soulignant comment vos résultats répondent à cette question fondamentale.