

Prédiction de l'Autisme à l'aide du Machine Learning

Encadré par :
Mr Nafaa Haffar

Réalisées par :
BOUABID Fadwa - KSIKSI Nadia - HMMERCHA Sarra - MOUISSAOUI Leila

PLAN

01

INTRODUCTION

02

PRÉTRAITEMENT DES DONNÉES

03

APPLICATION

04

CONCLUSION

INTRODUCTION

L'autisme est un trouble complexe qui affecte des millions de personnes dans le monde.

La Détection et diagnostic précoce sont essentiels pour un accompagnement adapté.

Ce projet explore comment le machine learning peut contribuer à prédire cette maladie.

PRÉTRAITEMENT DES DONNÉES

1.Compréhension et exploration des données :

- Description des données :

Analyse initiale des variables disponibles : 20 features et une variable cible:CLASS/ASD.

- Problèmes détectés :

Présence de nombreuses valeurs manquantes.

Les features ont des plages différentes

2.Gestion des valeurs manquantes :

- Suppression des enregistrements trop incomplets

3.Normalisation des données:

4.Encodage des variables catégoriques :

- 0 pour "Non", 1 pour "Oui"

5.Division des données :

Séparation du jeu de données en deux parties :

Entraînement (80%) : Pour construire le modèle.

Test (20%) : Pour évaluer la performance du modèle.

APPLICATION

1-Logistic_Regression:

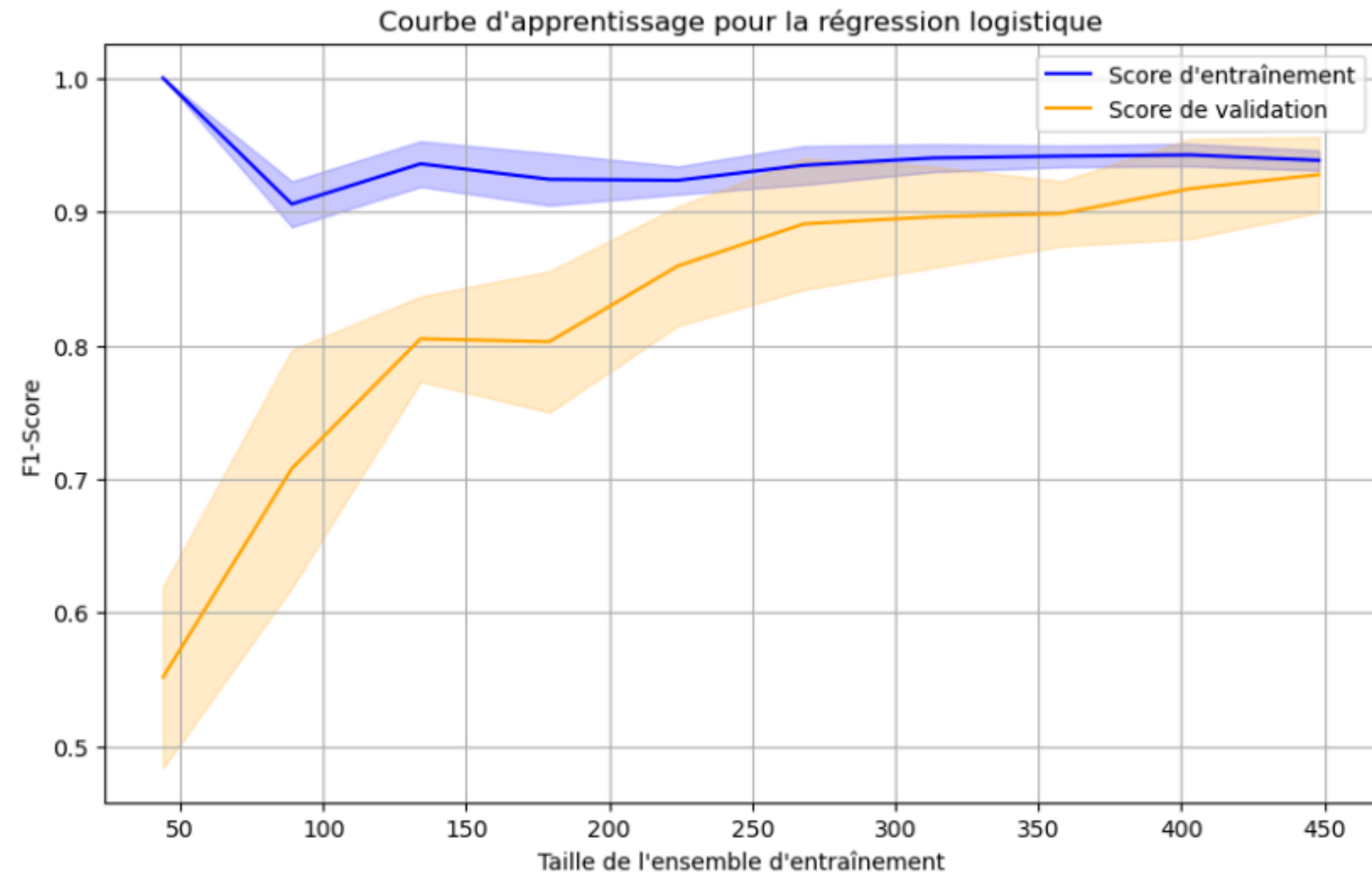
Rapport de classification:

Accuracy : 0.950354609929078

Rapport de classification :

	precision	recall	f1-score	support
0	0.95	0.98	0.96	91
1	0.96	0.90	0.93	50
accuracy			0.95	141
macro avg	0.95	0.94	0.94	141
weighted avg	0.95	0.95	0.95	141

Courbe d'apprentissage:



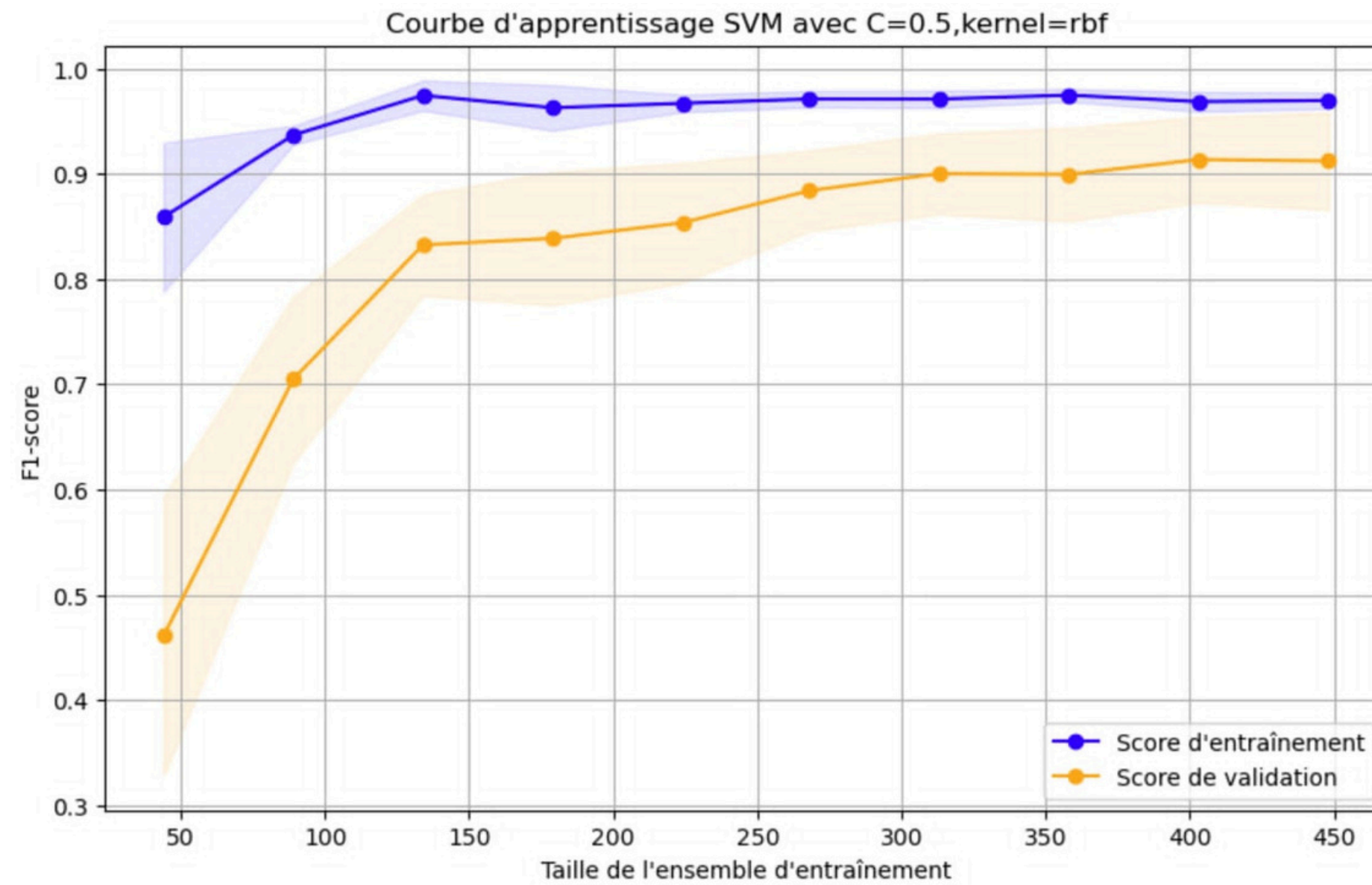
2-SVM(Support Vector Machine ou Machine à vecteurs de support

Classification Report :

	precision	recall	f1-score	support
0	0.92	0.99	0.95	91
1	0.98	0.84	0.90	50
accuracy			0.94	141
macro avg	0.95	0.91	0.93	141
weighted avg	0.94	0.94	0.93	141

Accuracy : 0.9361702127659575

Courbe d'apprentissage:



3-KNN(k-Nearest Neighbors)

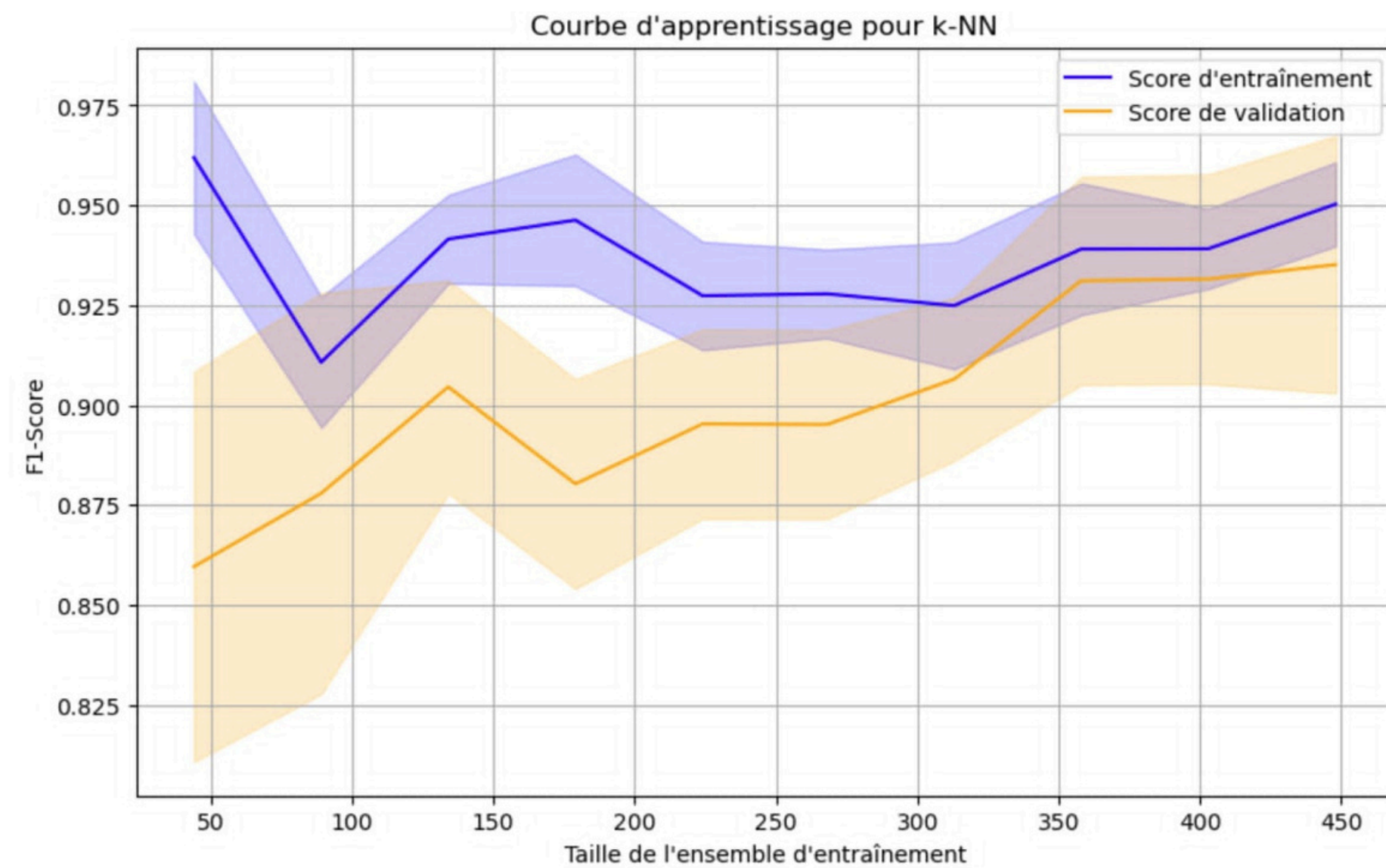
Rapport de classification:

	precision	recall	f1-score	support
0	0.93	0.93	0.93	91
1	0.88	0.88	0.88	50
accuracy			0.91	141
macro avg	0.91	0.91	0.91	141
weighted avg	0.91	0.91	0.91	141

Accuracy : 0.9148936170212766

3-KNN(k-Nearest Neighbors)

Courbe d'apprentissage:



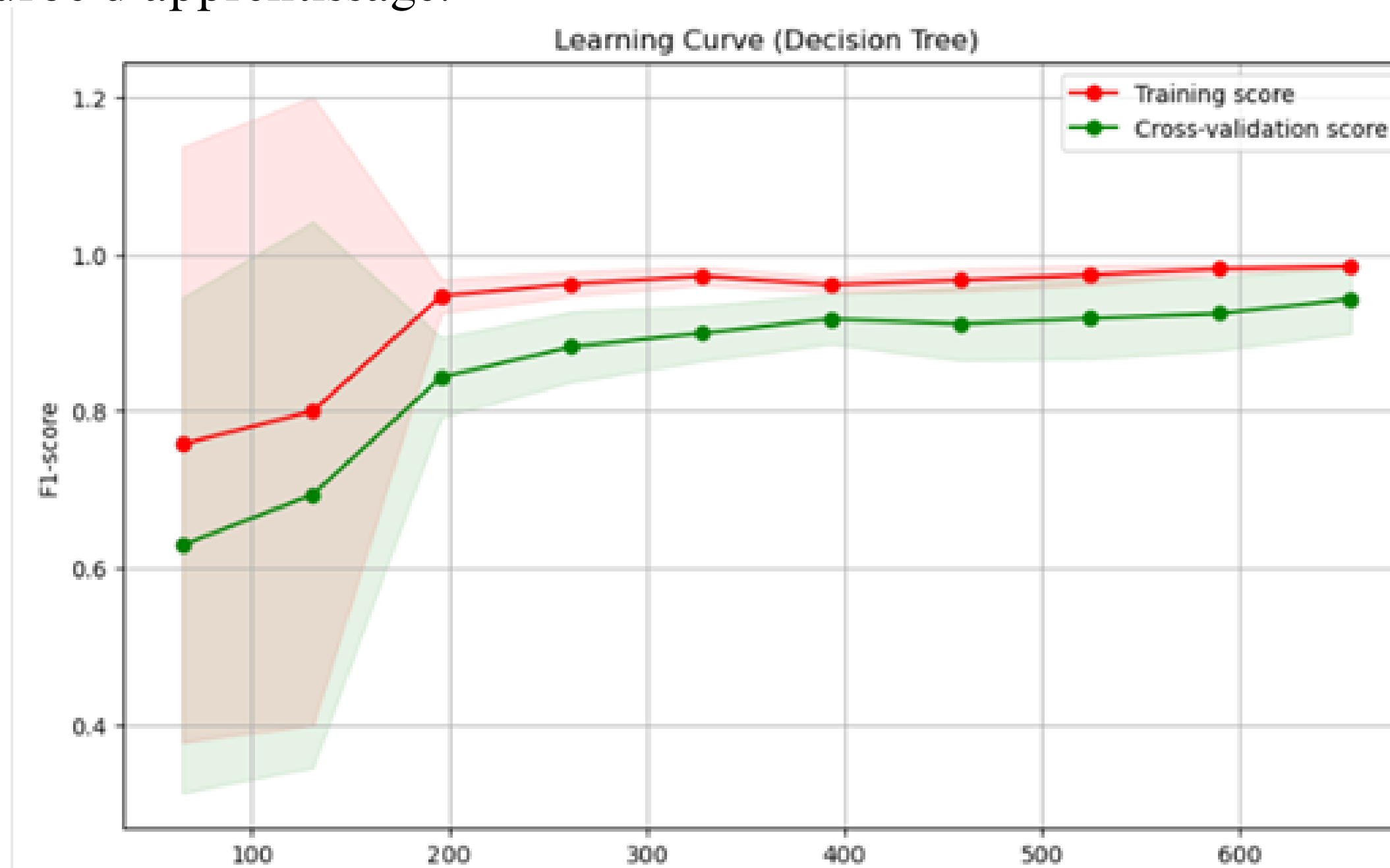
4-Arbre de Décision

Rapport de classification:

```
Accuracy: 0.82  
Precision: 0.70  
Recall: 0.61  
F1-score: 0.65
```

4-Arbre de Décision

Courbe d'apprentissage:



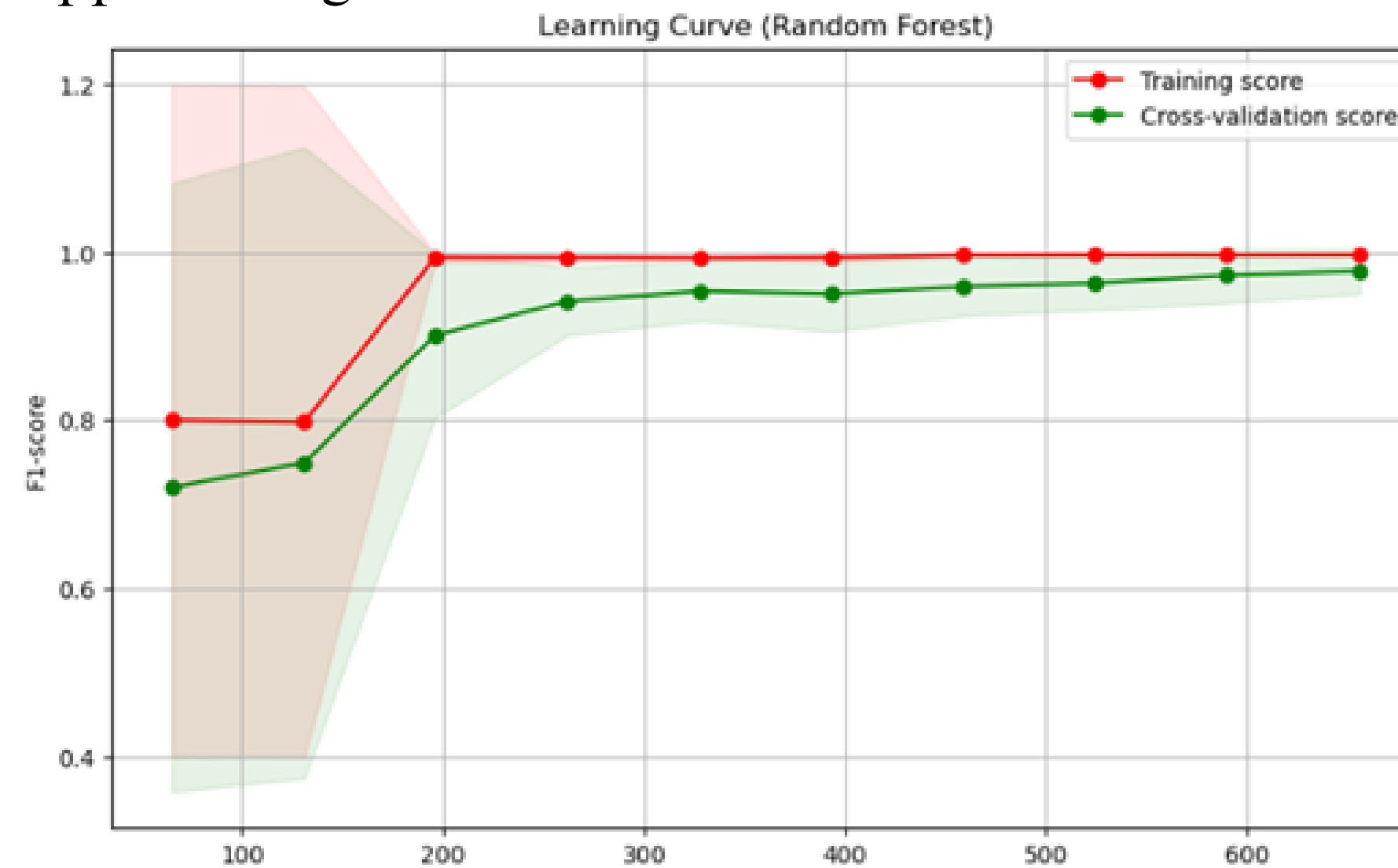
5-Random Forest

Rapport de classification:

```
Accuracy: 0.92  
Precision: 0.91  
Recall: 0.79  
F1-score: 0.85
```

5-Random Forest

Courbe d'apprentissage:

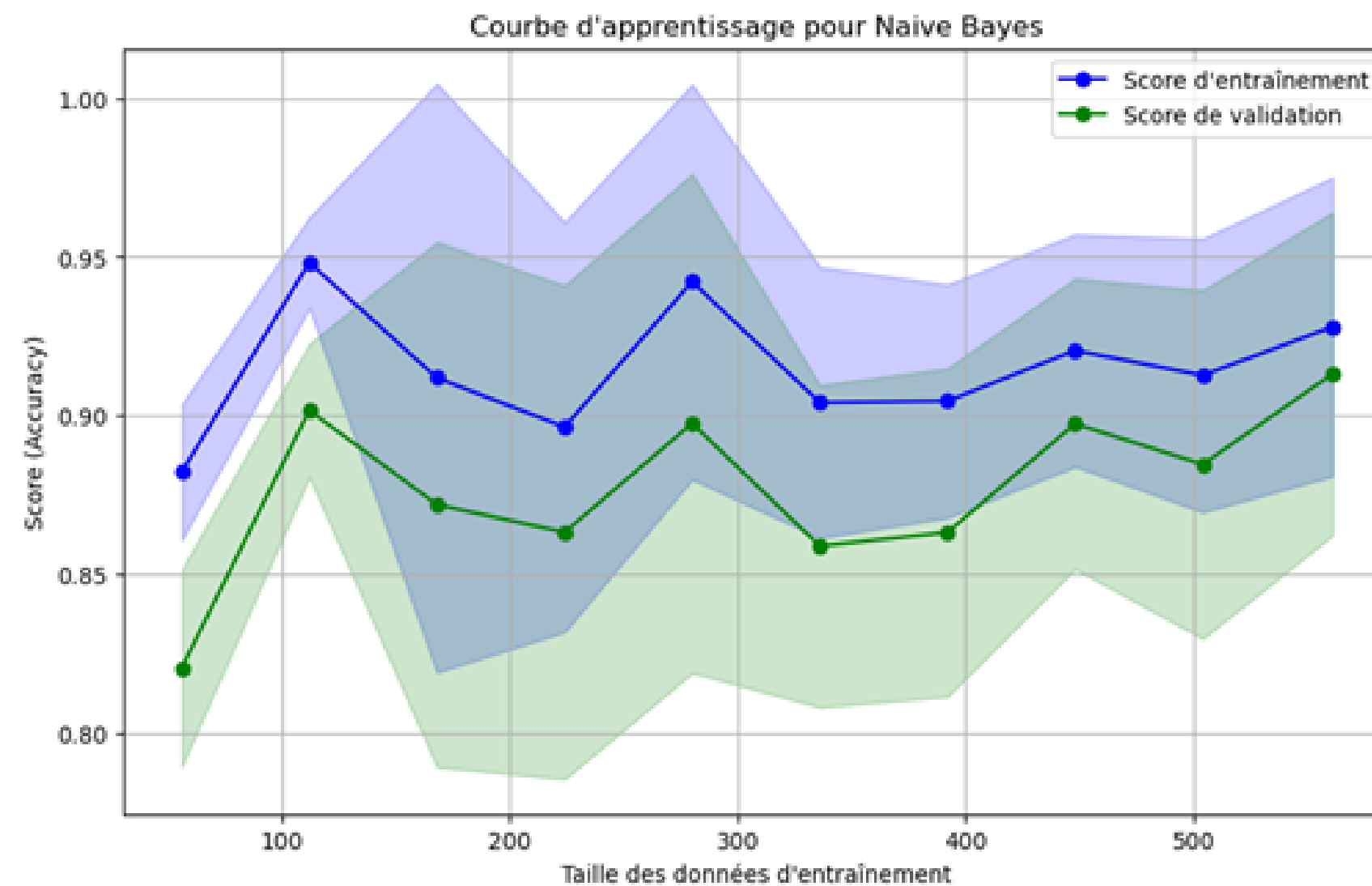


6-Naive Bayes :

```
Accuracy: 0.9645390070921985
Confusion Matrix:
[[88  3]
 [ 2 48]]
F1-score: 0.964616598018095
Recall: 0.9645390070921985
Précision sur l'ensemble de test : 0.9645390070921985
```

5-Random Forest

Courbe d'apprentissage:



	Logistic	KNN	SVM	ARBRE	RANDOMFOREST	NAIVE BAYES
Accuracy	0.95	0.91	0.93	0.82	0.92	0.964
F1-Score	0.96	0.93	0.95	0.65	0.85	0.96
precision	0.95	0.93	0.92	0.70	0.91	0.96
Risqué Underfitting	✖	✖	✔	✖	✖	✖
Risqué Overfitting	✖	✖	✖	✔	✔	✖

CONCLUSION

En conclusion, étant donné la simplicité du dataset, les modèles simples comme Naive Bayes et la régression logistique offrent de meilleurs résultats que les modèles complexes comme l'arbre de décision, qui montrent des limites en termes de performance et de généralisation.



MERCI !