

Les employers vont-il partir ?

Sommaire :

Remise en contexte

Présentation des datas

Le nettoyage des données

La visualisation

Les modèles utilisés

Quel est le meilleur modèle ?

Conclusion

Remise en contexte :

On a deux semaines pour analyser le dataset et créé un modèle qui soit capable de prédire si des employées vont quitter l'entreprise dans les deux prochaines années

Présentation des datas :

4653 Rows ,
9 Columns ,
une target : "**LeaveOrNot**",
les duplicates ,
les outliers.

Les outliers :

Sur le deuxième box plot on peut apercevoir des outliers .

Cela représente les personnes sous payer dans l'entreprise

```
▶ (df_outliers == 3).sum()
```

```
41]: 3492
```

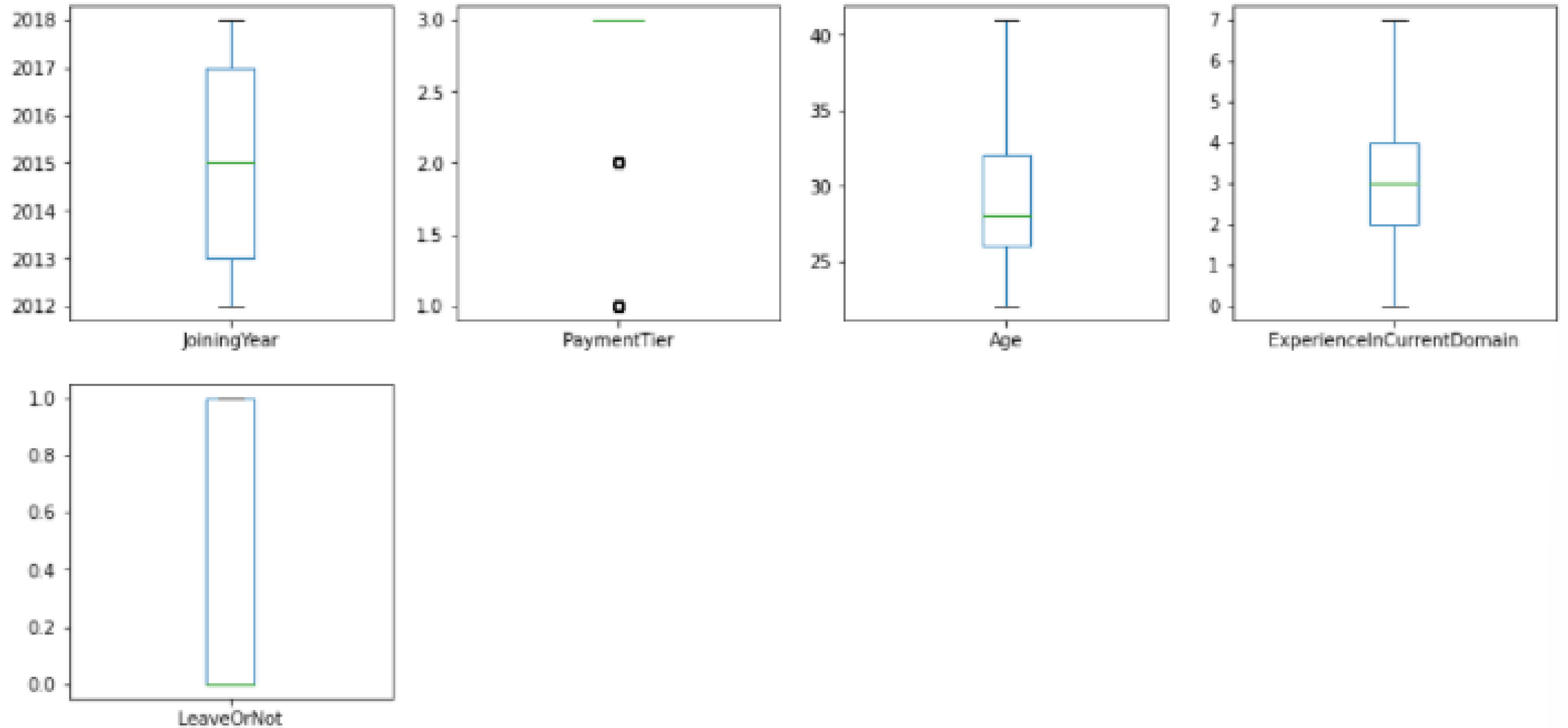
```
▶ (df_outliers == 2).sum()
```

```
42]: 918
```

```
▶ (df_outliers == 1).sum()
```

```
43]: 243
```

Les outliers



Nettoyages des données :

très peu de données à nettoyer

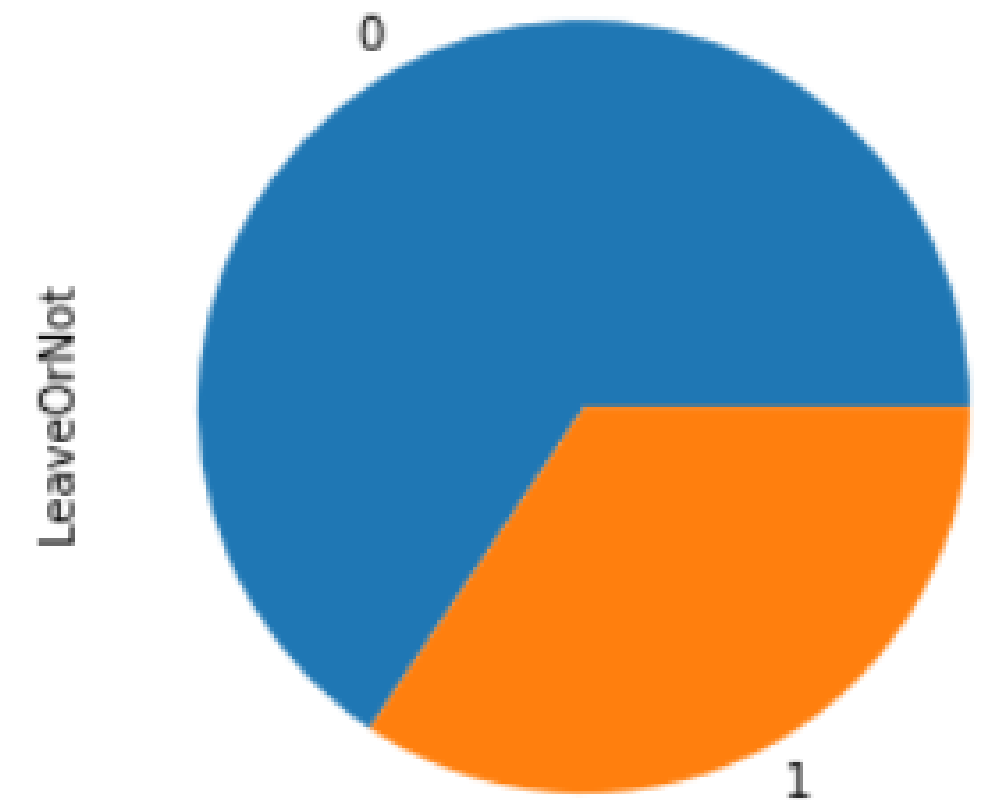
Il y avait 4 colonnes à mettre en 0, 1 grâce au

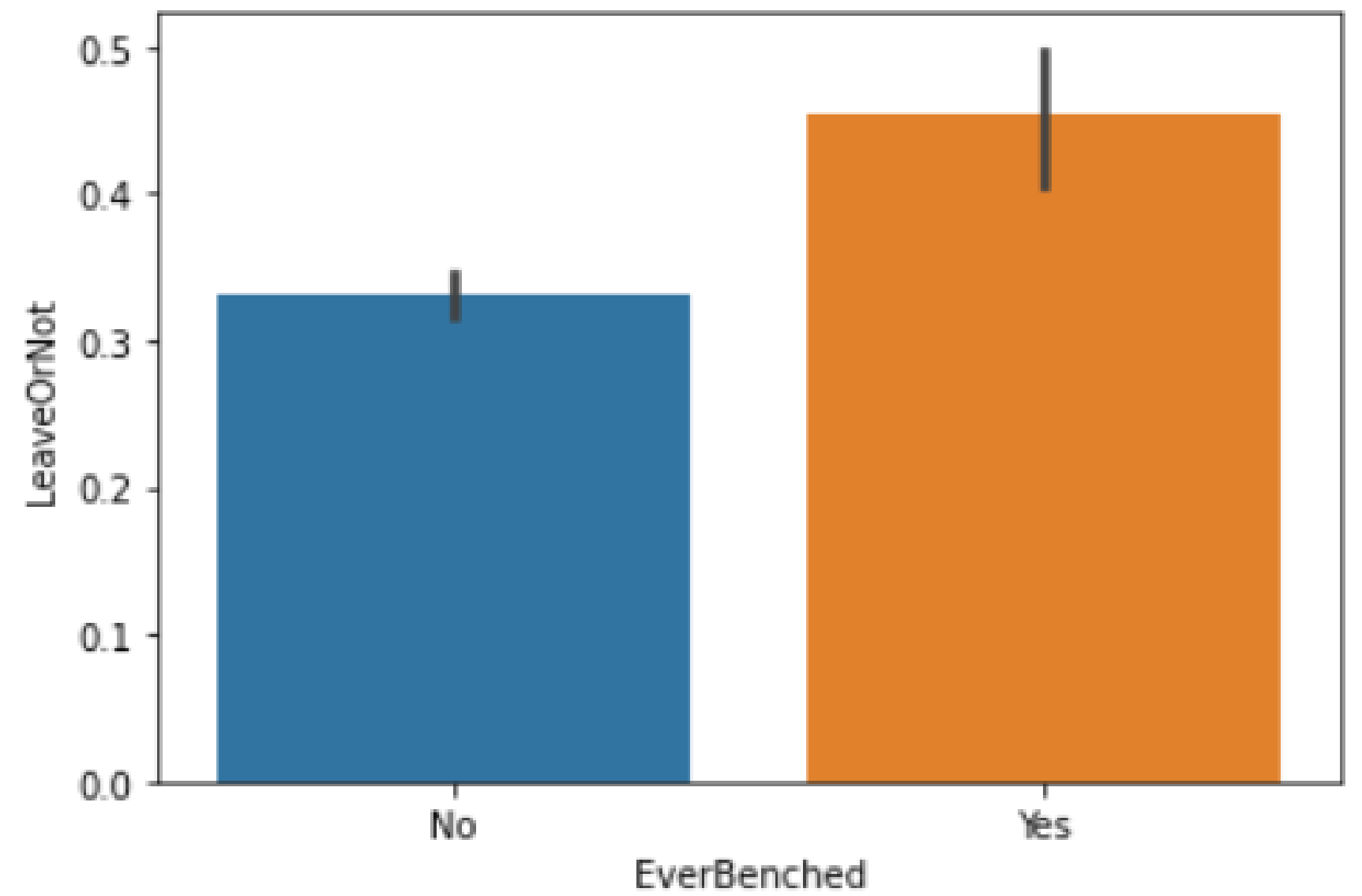
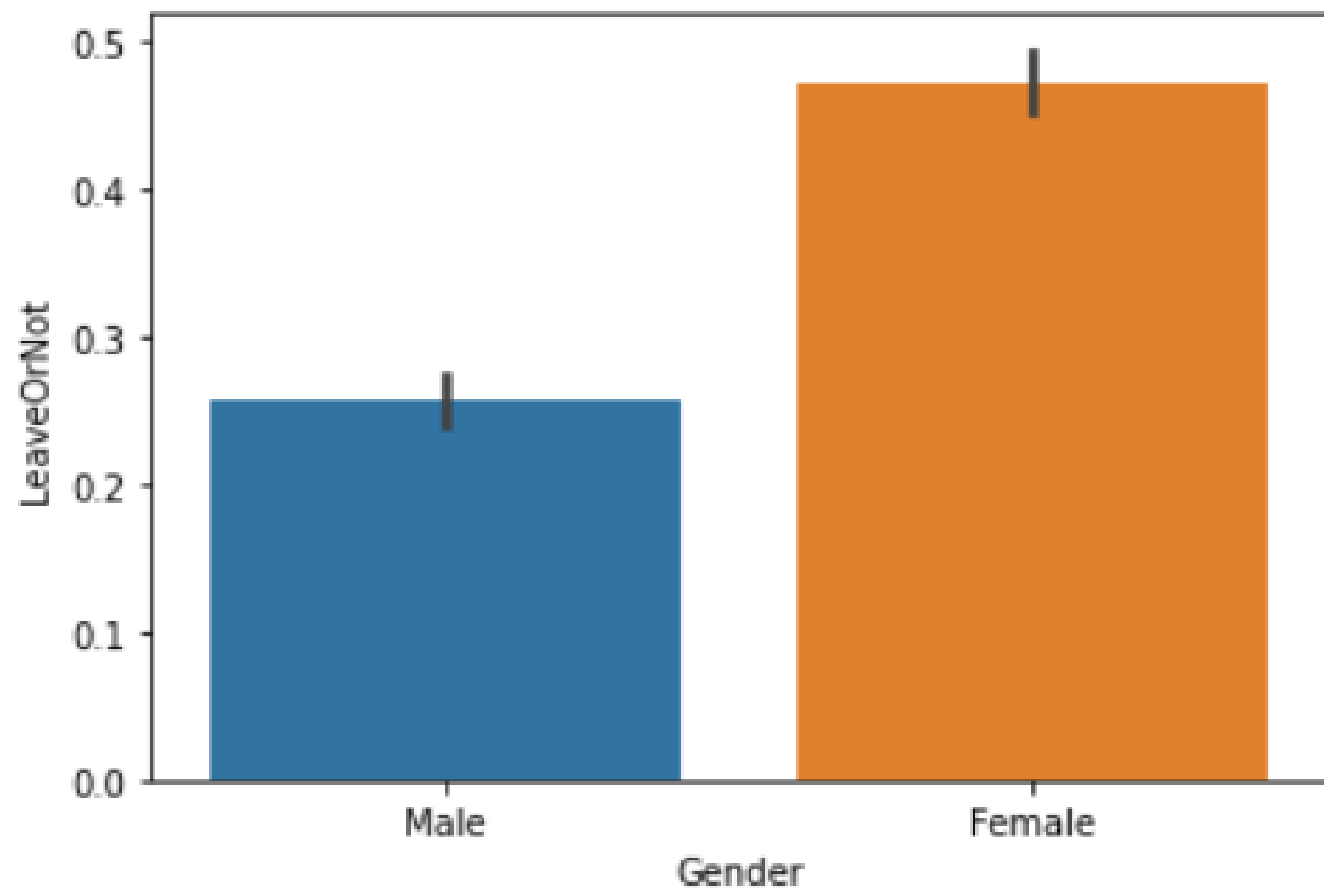
LabelEncoder

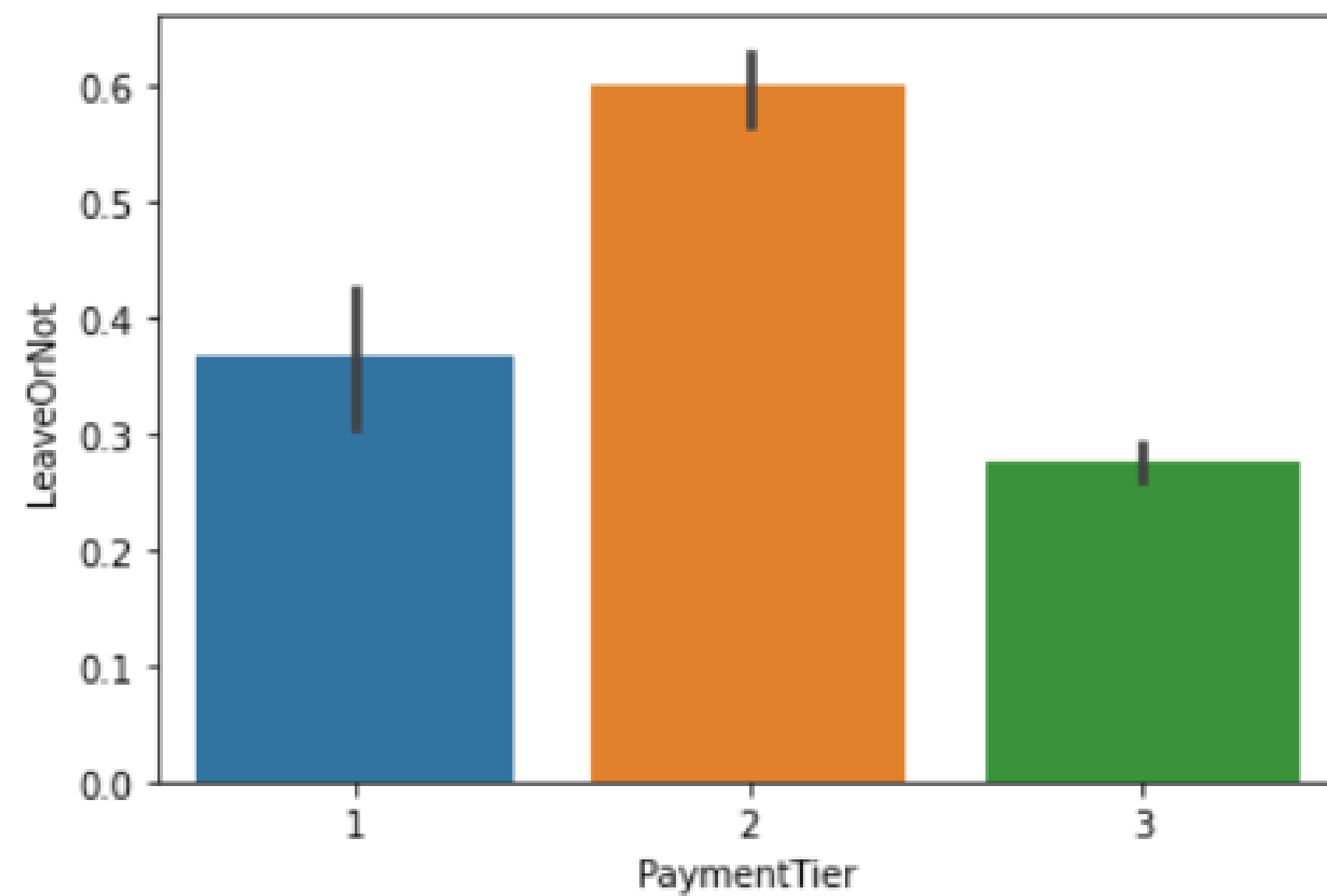
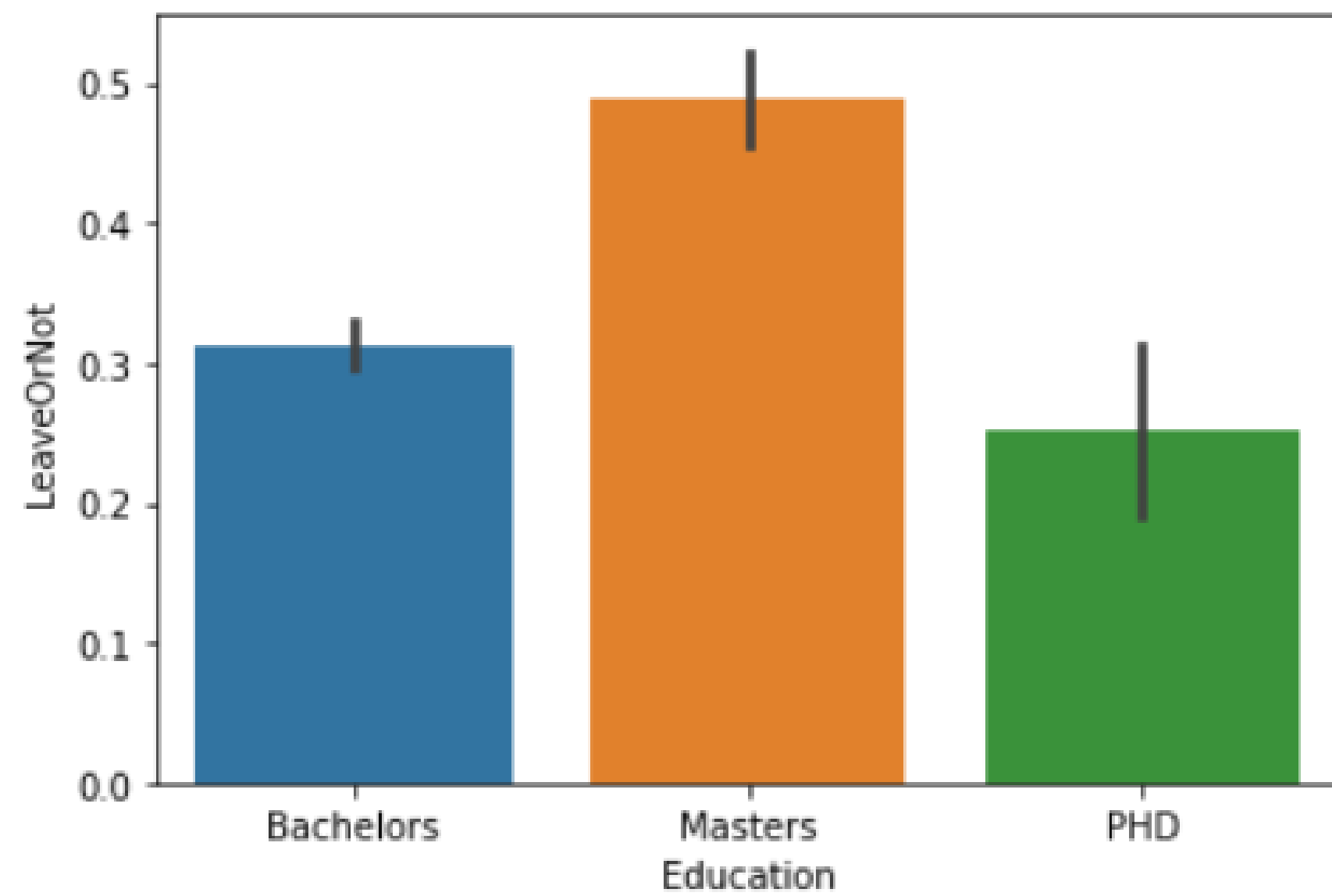
Pas de valeurs nulles

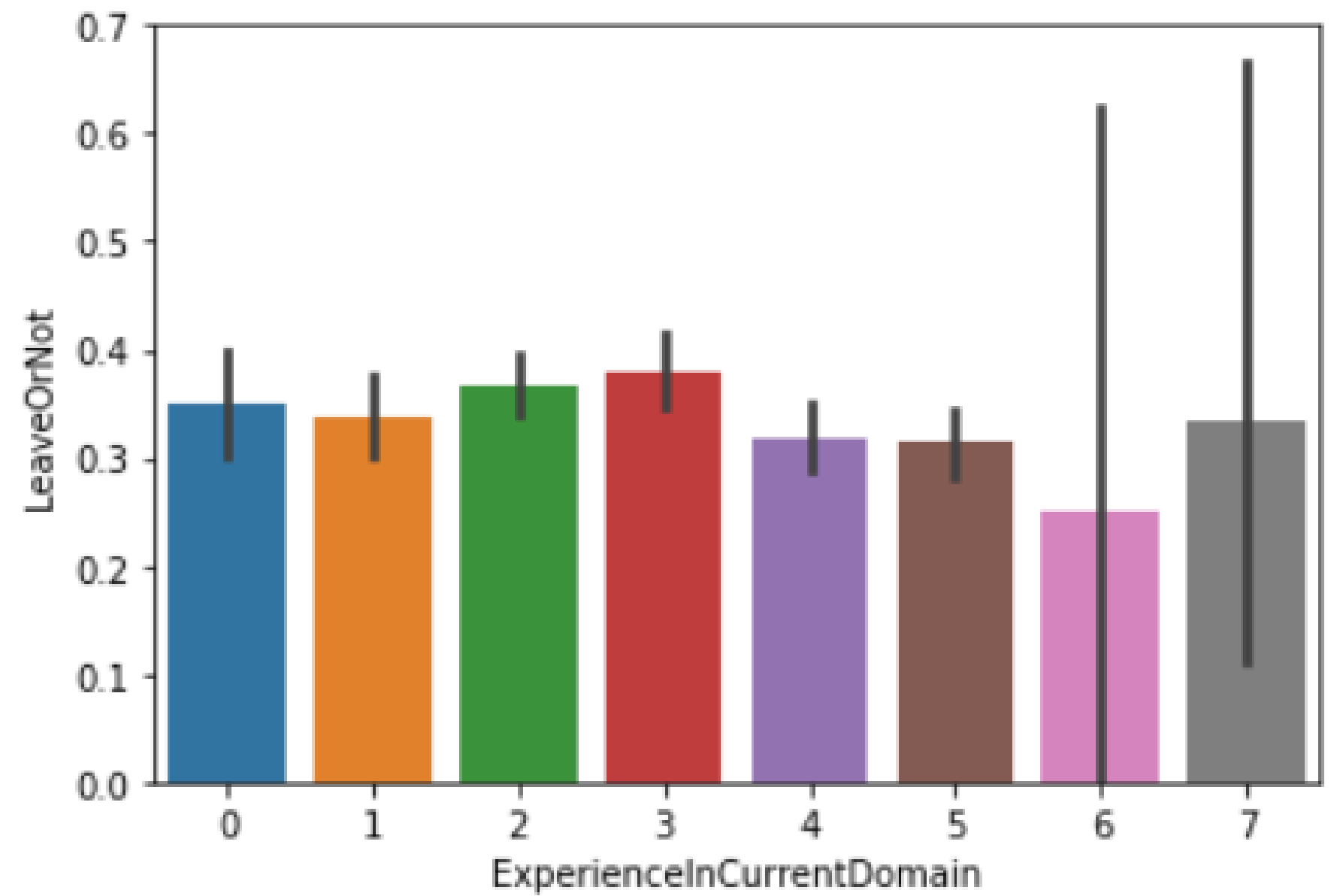
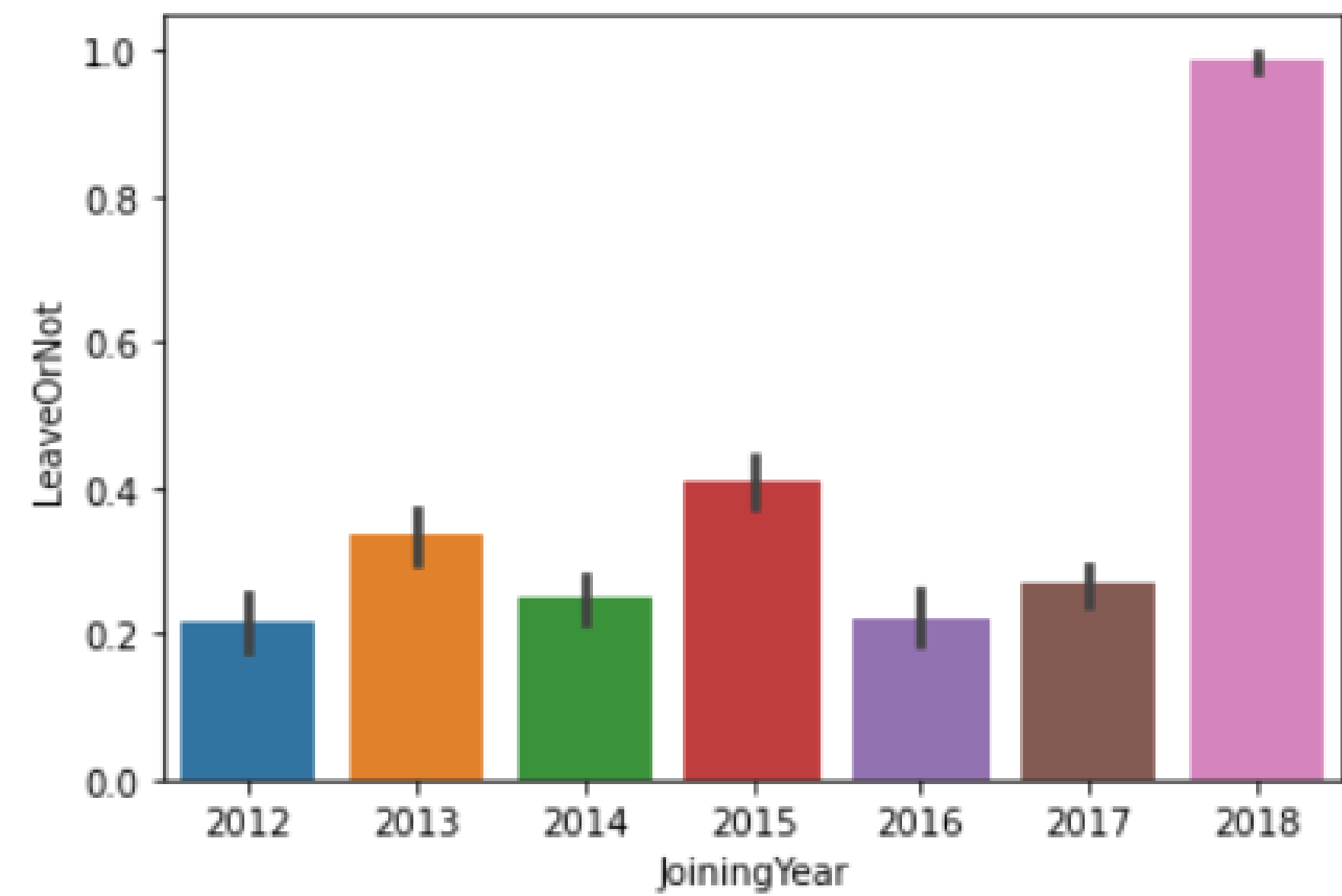
Unbalanced data :

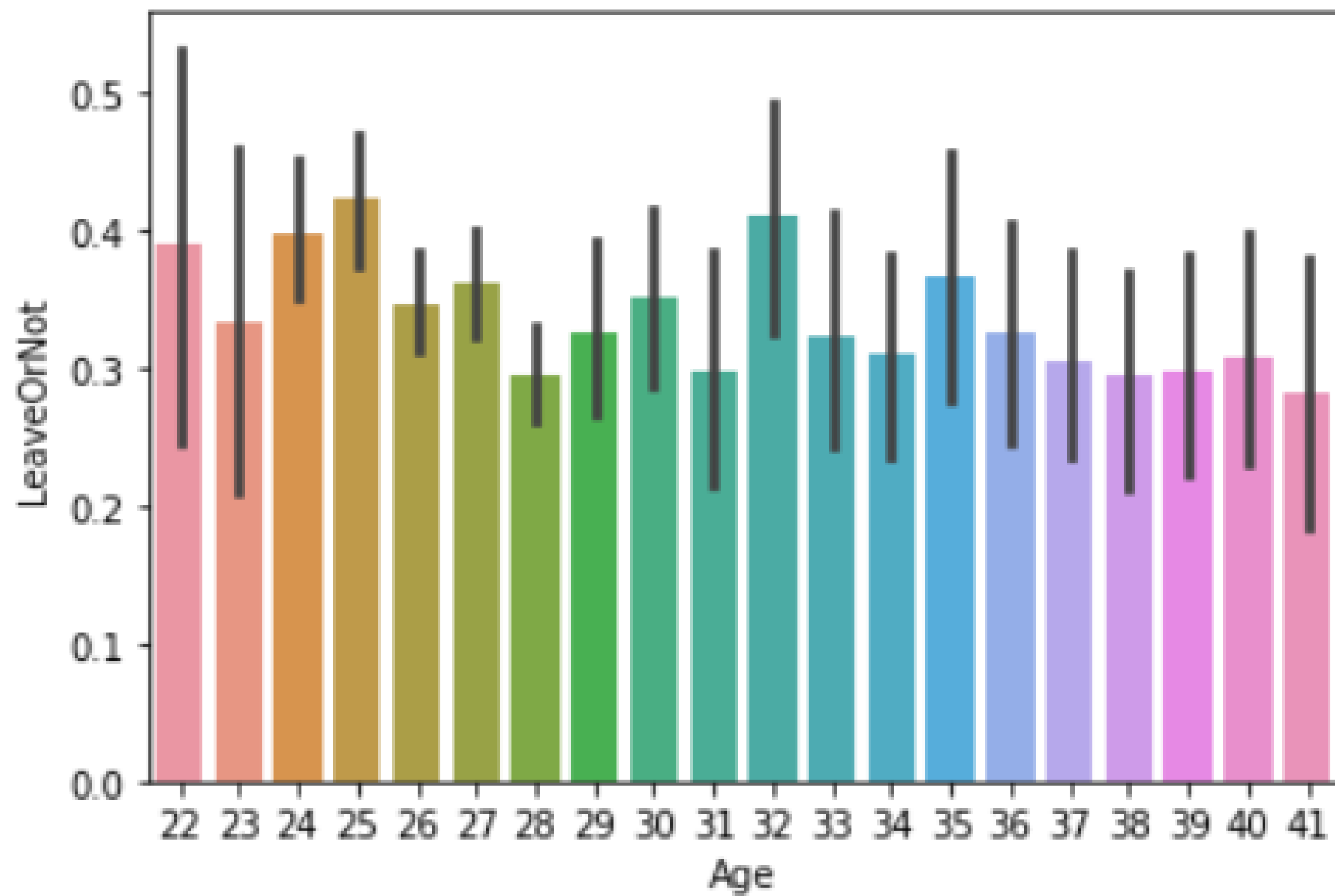
La colonne de la target n'avait pas
le même ratio de 1 et de 0











Baseline model :

J'ai utilisé un DummyClassifier
comme modèle ainsi que
StandardScaler et
MinMaxScaler

On utilise le résultat du recall

	precision	recall	f1-score	support
0	0.65	0.50	0.57	610
1	0.34	0.49	0.40	321
accuracy			0.50	931
macro avg	0.50	0.50	0.49	931
weighted avg	0.55	0.50	0.51	931

Les autres modèle utiliser :

Smote + DummyClassifier
StandardScaler + MinMaxScaler +
KNeighborsClassifier
Smote + LogisticRegression
StandardScaler + MinMaxScaler +
LogisticRegression

KNN :

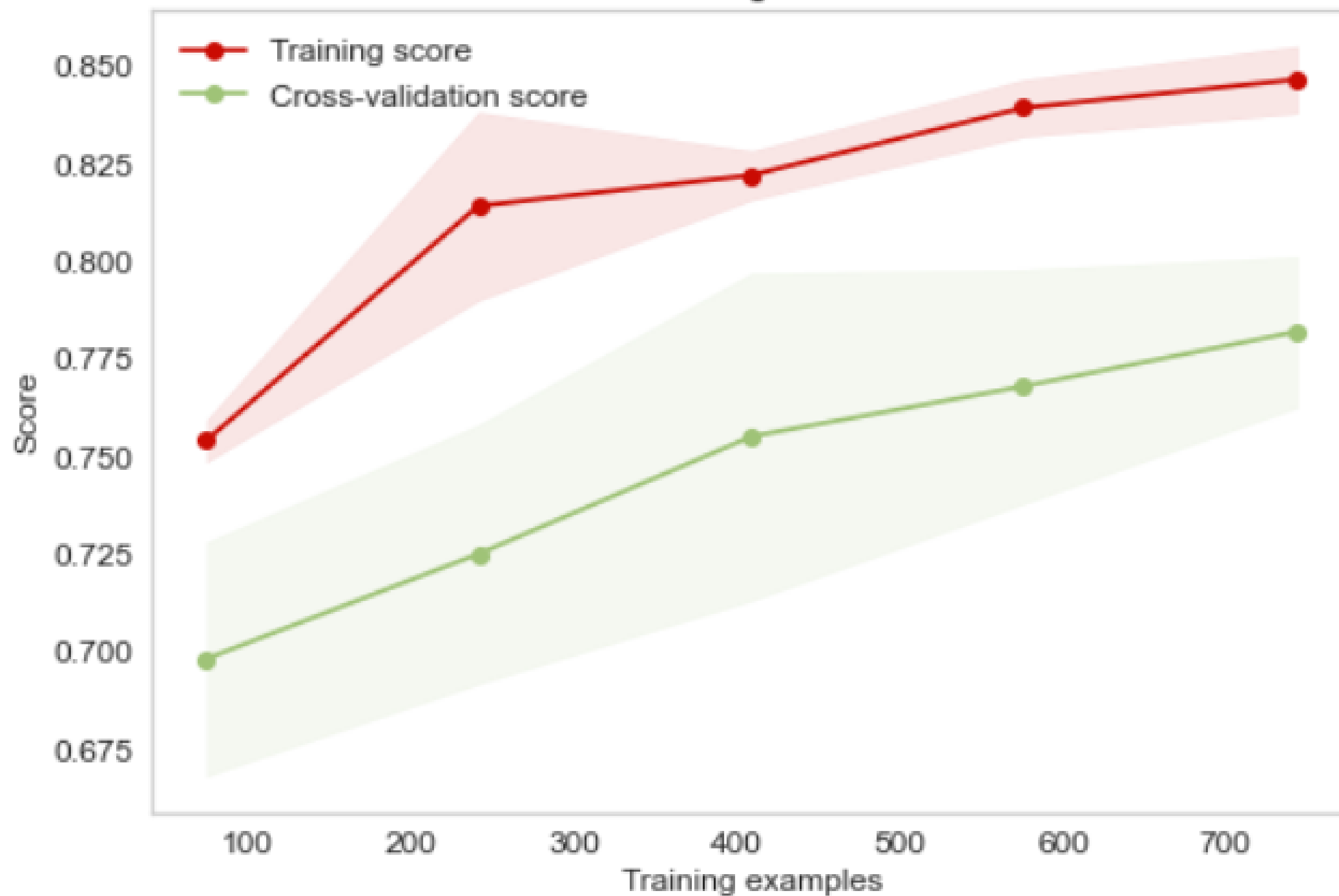
```
accuracy train : 0.858  
accuracy test  : 0.824
```

recall

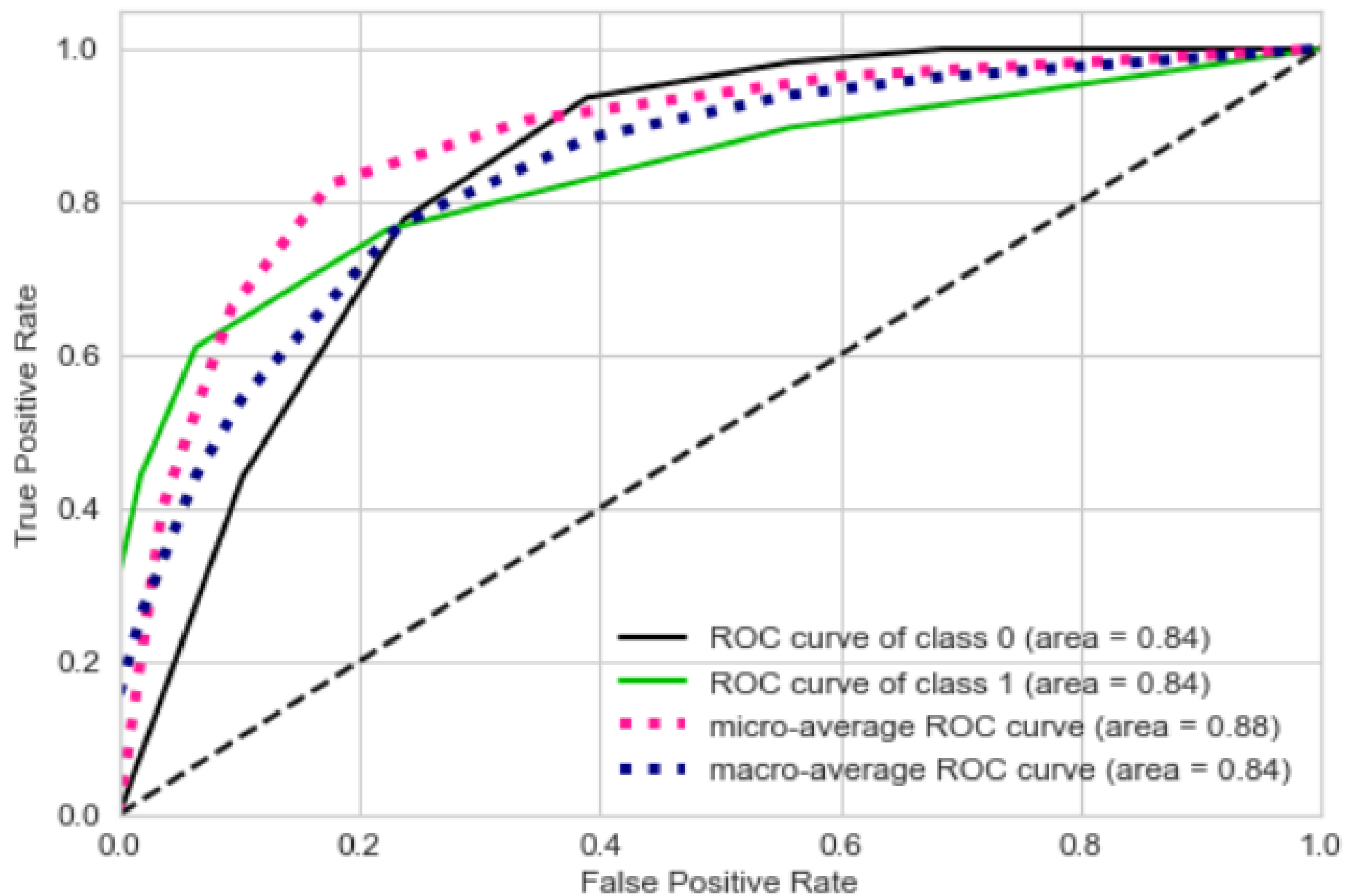
0.94

0.61

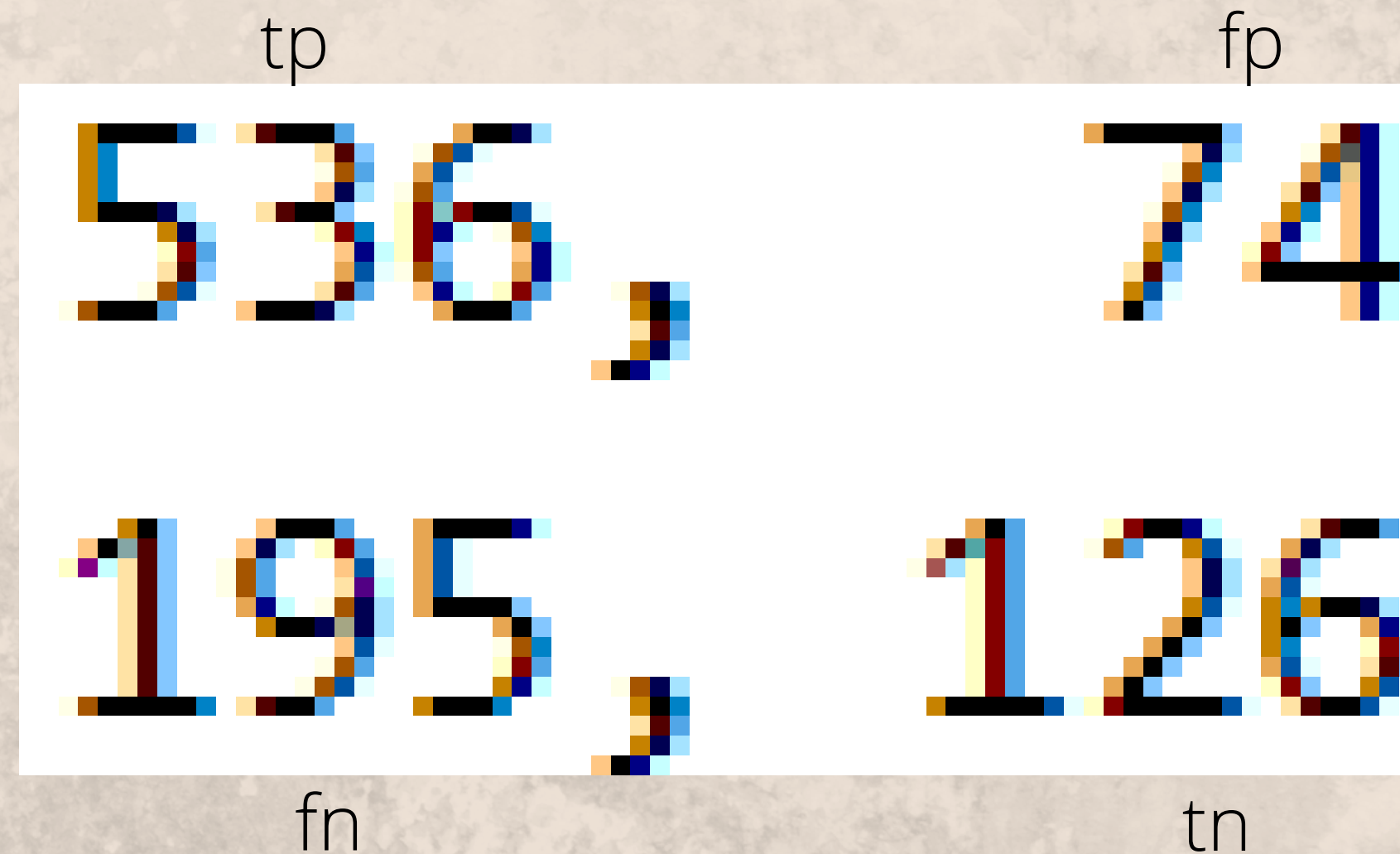
Learning Curve



ROC Curves



Confusion matrix



Ensemble methods :

StandardScaler + MinMaxScaler + Bagging + KNN

Recall 0.21

StandardScaler + MinMaxScaler + GradientBoosting

Recall 0.61

StandardScaler + MinMaxScaler + sampling + KNN

Recall 0.68

Quel sont les types d'employer qui risque de partir ?

Une femme, qui a déjà été temporairement expulsé, qui possède un master ou bachelier, payée entre 1 et 2 , âgée de 22 à 35 ans, arrivé entre 2015 et 2018 .