كليـــــة العلـــوم
والتقنيـات – مراكش
FACULTÉ DES SCIENCES ET TECHNIQUE – MARRAKECH

Université Cadi Ayyad
Faculté des Sciences et Techniques
Marrakech
Master SDAD

La reconnaissance des caractères arabe

Préparées par :
ASAKOUR Ihsane
NAIM Kawtar

Encadré par :
Prof. A.OUAARAB

07 Janvier 2023

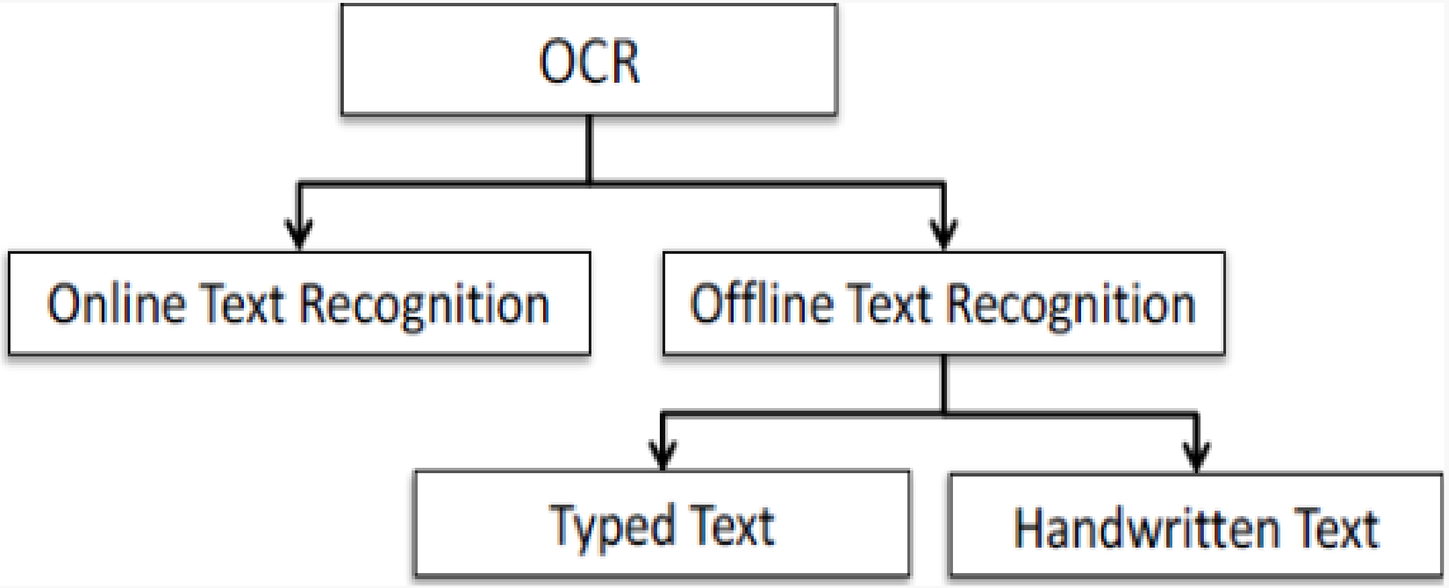
Plan :

Introduction

Introducion

OCR

La reconnaissance optique de caractères (OCR) est un processus qui permet de convertir du texte présent sur une image ou un document scanné en fichier électronique modifiable, comme un document Word ou un fichier de traitement de texte.



Introduction

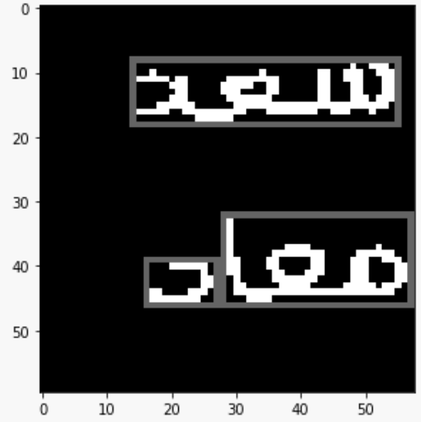
1 OCR

2 Segmentation

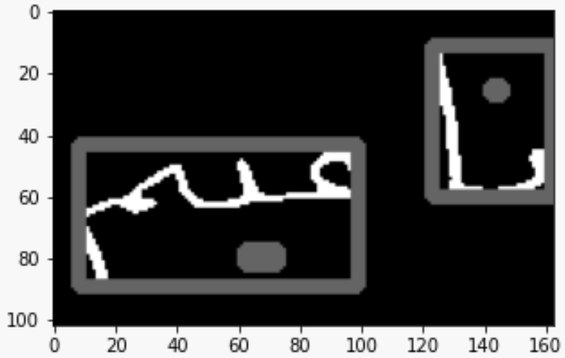
Segmentation

En informatique et en vision par ordinateur, la segmentation d'image est le processus de division d'une image en plusieurs segments ou régions, chacun correspondant à un objet ou à une partie différente de l'image.

- Segmentation des Lignes / Mots : Opencv qui permet de dessiner les boxes entours des mots à chaque fois qu'une espace a été détecté.
  il découpe l'image en utilisant les coordonnées de l'espace comme guide.

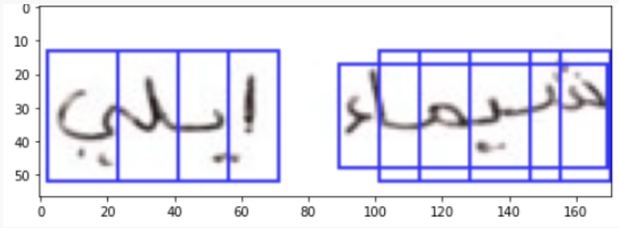


- Segmentation des Caractères
  - Limitation de Opencv



  - Limitation de pytesseract

```
['0' ,'44' ,'23' ,'5' ,'2' ,'ي']
['0' ,'44' ,'41' ,'5' ,'23' ,'ل']
['0' ,'44' ,'56' ,'5' ,'41' ,'ي']
['0' ,'44' ,'71' ,'5' ,'56' ,'إ']
['0' ,'40' ,'169' ,'9' ,'89' ,'ء']
['0' ,'44' ,'113' ,'5' ,'101' ,'ا']
['0' ,'44' ,'128' ,'5' ,'113' ,'م']
['0' ,'44' ,'146' ,'5' ,'128' ,'ي']
['0' ,'44' ,'155' ,'5' ,'146' ,'ش']
```



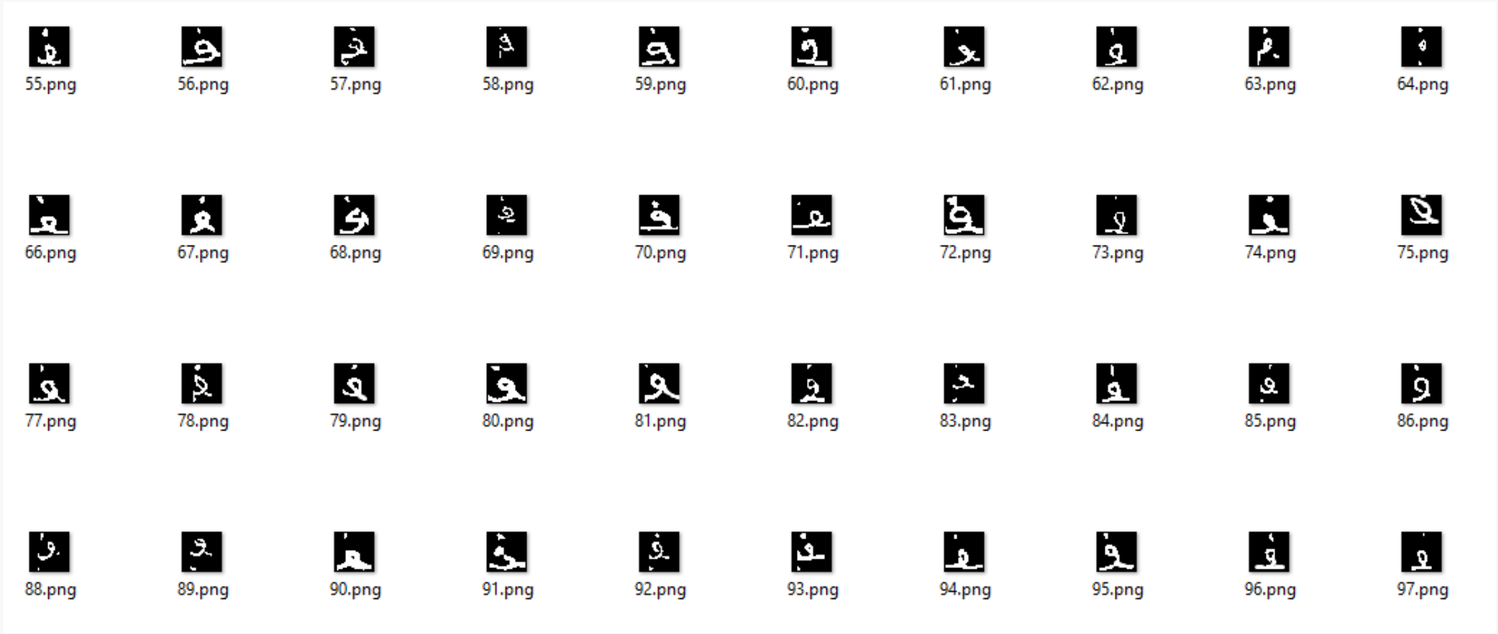
Dataset

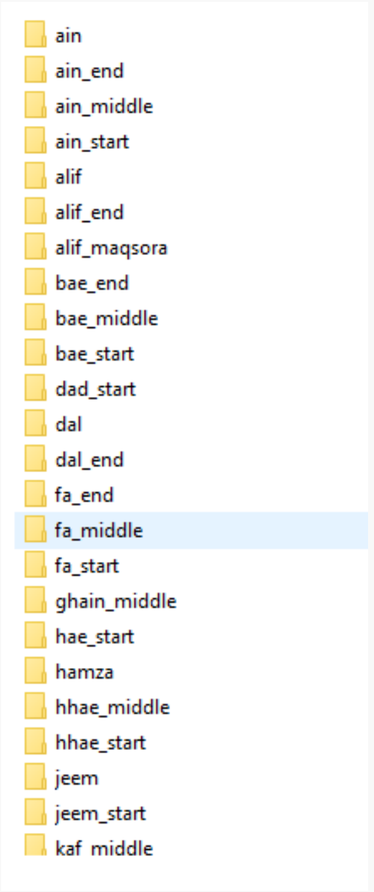
Dataset

Label

Les classes seront définies de manière à ce que chaque lettre dans une position soit comptée comme une classe.

comme illustré ici, chaque dossier présente une classe différente à l'autre, et dans chaque dossier, il y a un ensemble de caractères du même type indiqué dans le nom du dossier.





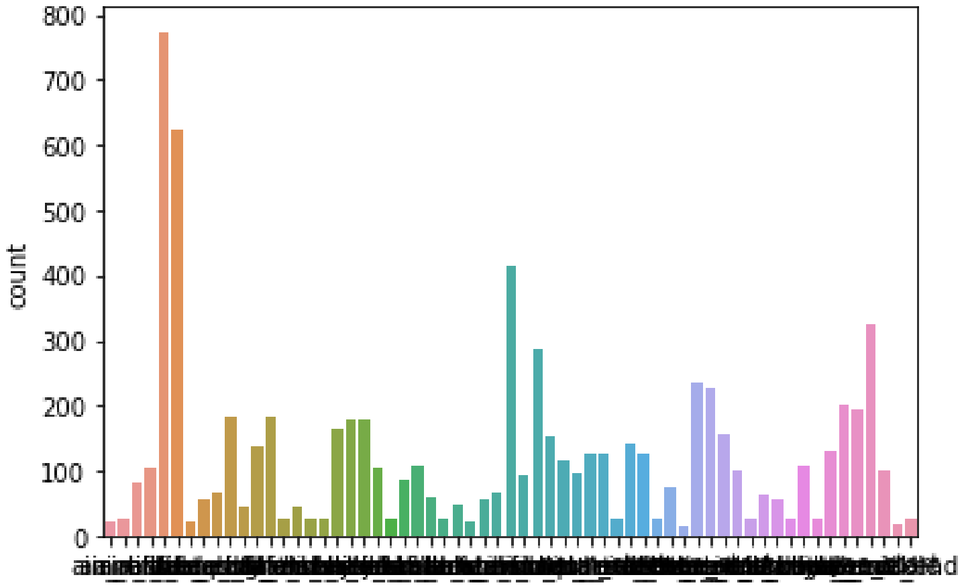
Les Approches Proposées

Les Approches Proposées

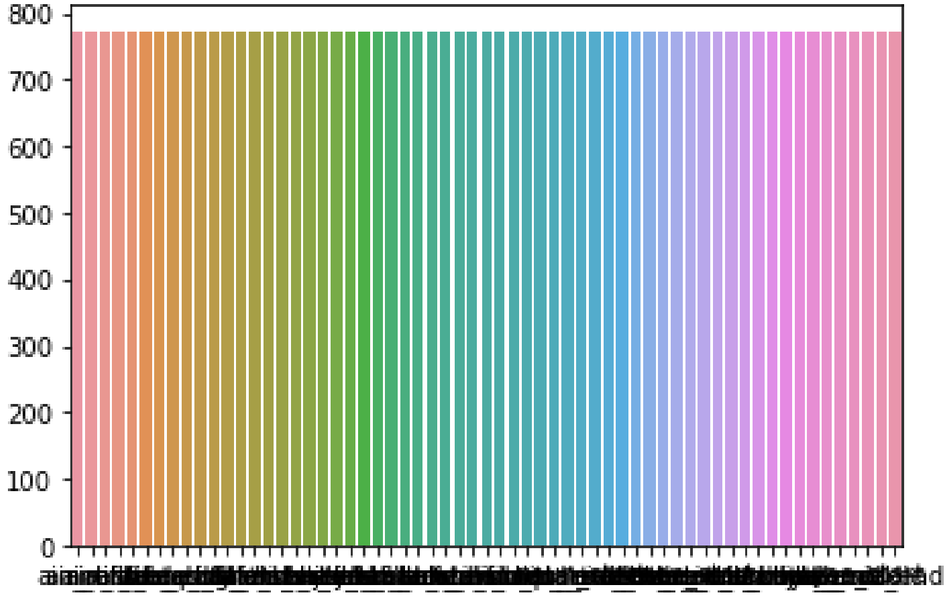
1 Approche 1

2 Approche 2

Equilibrer les classes



Non équilibré



équilibré

Les Approches Proposées

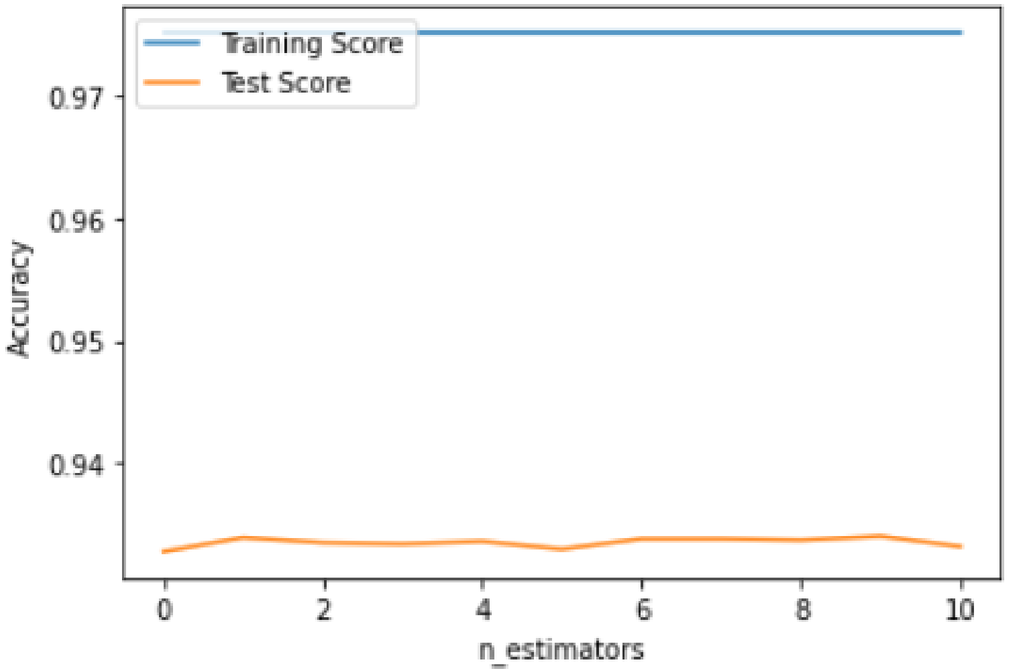
1 Approche 1

2 Approche 2

L’estimation des hyperparamètres

```
ITER  999  TRAINING SCORE  0.9751603838608769  TEST SCORE  0.9327748913158732
ITER  1000  TRAINING SCORE  0.9751603838608769  TEST SCORE  0.9339412575548722
ITER  1001  TRAINING SCORE  0.9751603838608769  TEST SCORE  0.9335171243770544
ITER  1002  TRAINING SCORE  0.9751603838608769  TEST SCORE  0.9334110910825999
ITER  1003  TRAINING SCORE  0.9751603838608769  TEST SCORE  0.9336231576715088
ITER  1004  TRAINING SCORE  0.9751603838608769  TEST SCORE  0.9329869579047821
ITER  1005  TRAINING SCORE  0.9751603838608769  TEST SCORE  0.9338352242604178
ITER  1006  TRAINING SCORE  0.9751603838608769  TEST SCORE  0.9338352242604178
ITER  1007  TRAINING SCORE  0.9751603838608769  TEST SCORE  0.9337291909659633
```

La courbe de résultats



L'application du modèle

Résultats

Performance de méthode machine learning sur la classification multiple

```
rfc = RandomForestClassifier(n_estimators=1000)
rfc.fit(X_train, y_train)
print('train score:', rfc.score(X_train, y_train))
print('test score:', rfc.score(X_test, y_test))
```

```
train score: 0.9744711309050421
test score: 0.9391368889831407
```

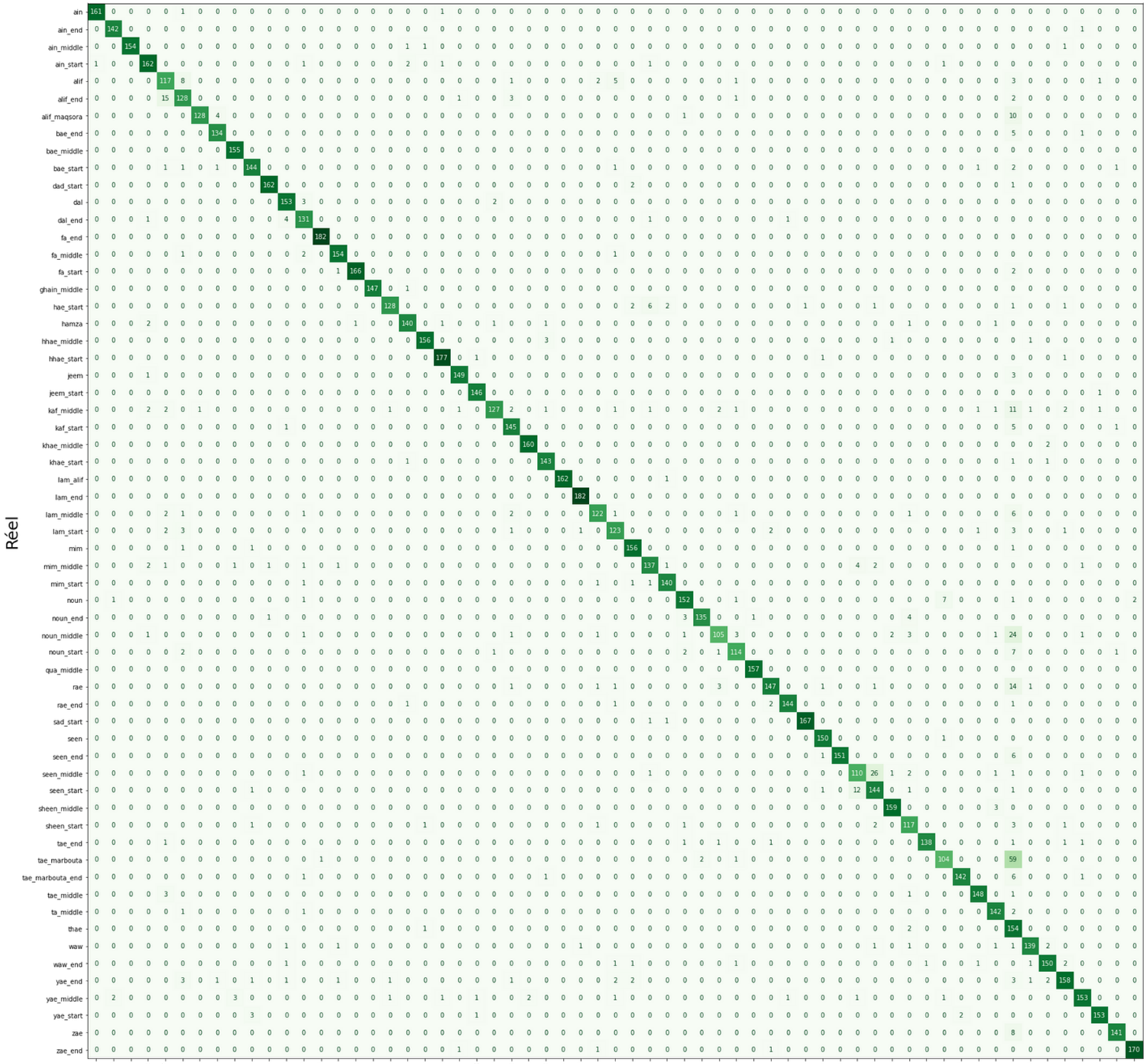
```
y_pred=rfc.predict(X_test)
```

```
precision,recall,fscore,none= precision_recall_fscore_support(y_test, y_pred, average='weighted
print("Taux d'erreur : {}".format(1-accuracy_score(y_test,y_pred)))
print("Taux de réussite : {}".format(accuracy_score(y_test,y_pred)))
print('Précision : '+(str(precision)))
print('Taux de détection : '+(str(recall)))
print('F1-score : '+(str(fscore)))
```

```
Taux d'erreur : 0.06086311101685926
Taux de réussite : 0.9391368889831407
Précision : 0.949831797592206
Taux de détection : 0.9391368889831407
F1-score : 0.941637640480587
```

Résultats

Performance des méthodes machine learning sur la classification multiple



2

Matrice de Confusion

Conclusion et perspective

Conclusion et perspective

Dans ce travail on s'intéresse aux techniques de reconnaissance des caractères arabes, nous avons précisé sur la technique de machine Learning Random Forest.
Cela nous donne une précision de 95%

NEXT ➡

Appliquer des méthodes de deep learning

Rechercher d'autres méthodes plus efficaces sur le jeu de données

Merci !

Pour
Votre
Attention

Asakour & Naim