



Zoidberg 2.0

Alcaro Matthieu,
Lambert Emilien,
Marcel Léo,
Qedira Fares,
Vallera Antonio



Sommaire

- Objectifs
- Ensembles de données
- Algorithmes
- Transformations
- Statistiques
- Résultats
- Conclusion



Objectifs

L'objectif du projet est d'apprendre et de comprendre les tenants et aboutissants des points nécessaires à la réalisation d'un programme de reconnaissance d'images

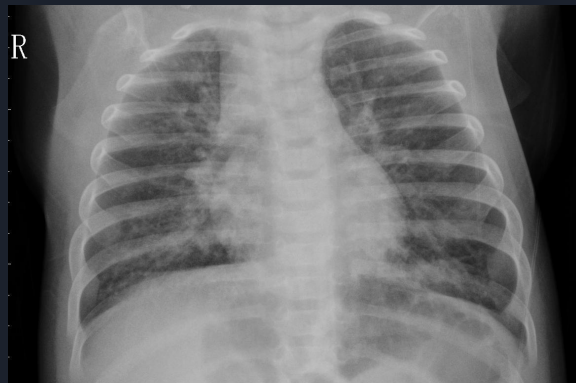
- Les ensembles de données
- Les familles d'algorithmes
- Les transformations
- Les graphiques et statistiques

Présentation du projet

Le thème du projet est de concevoir un algorithme de reconnaissance d'image d'examen radiologique des poumons.

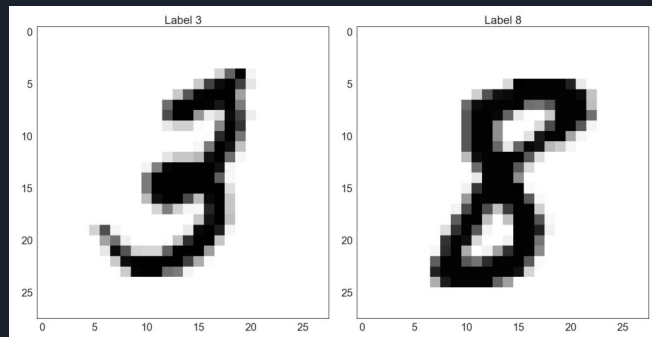
Pour différencier :

- Les patients sains
- Les patients malades par infection
- les patients malades par bactérie



Ensemble de données

- Mnist :
nombres de 0 à 9 sous forme de fichiers binaires
- Zoidberg 2.0 :
photo d'examen radiologique sur des patients
- Particularités physiques :
situs inversus "cœur à droite"
sutures métalliques sur le sternum
- Particularités informatiques :
certaines images en RGB noir et blanc
et non en véritable grayscale.





Algorithmes

- SVC
- SVC Linear
- MLP Classifier
- Prédiction KNN
- Naive Bayes
- Decision Tree Classifier
- Random Tree forest
- Extremely Randomized Trees



Transformations

L'objectif d'une Transformation est de diminuer la quantité d'information à analyser par l'algorithme en conservant uniquement les données nécessaires.

- Nystroem
- PCA

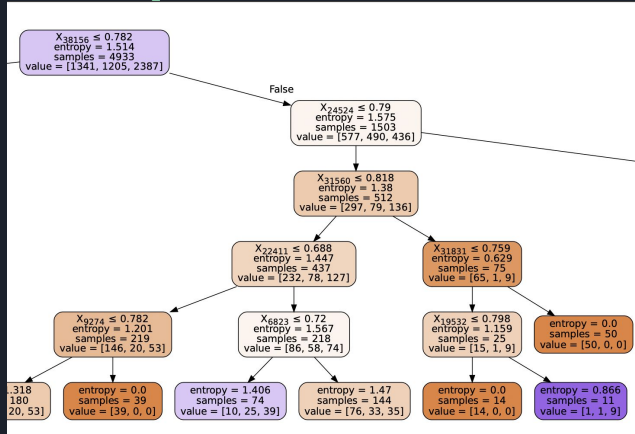


Statistiques

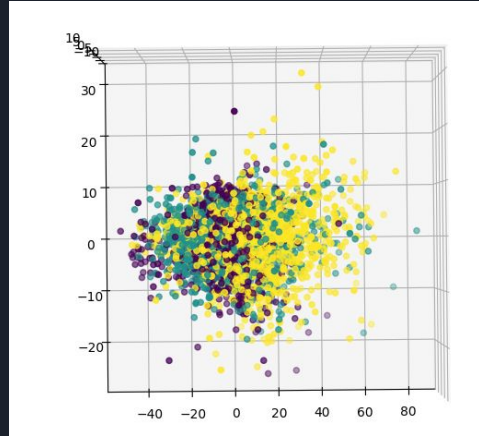
Pour avoir une bonne vision des performances de chaque algorithme nous avons mis en place plusieurs indicateurs.

- Des timers pour calculer le temps d'exécution de chaque étape.
 - Chargement des données
 - Transformation des données
 - Apprentissage etc...
- Des représentations graphiques 2D et 3D de plusieurs indicateur.
 - Matrices de confusion
 - PCA 3D
 - Tree Graph

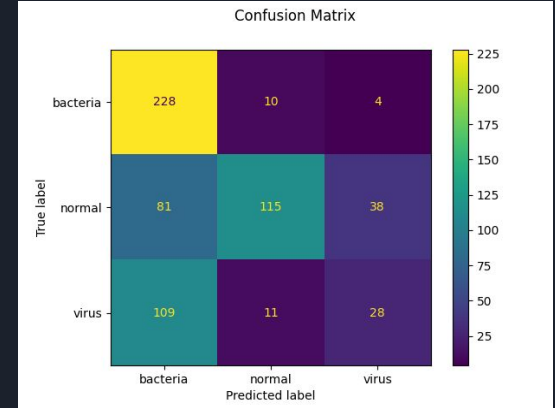
Statistiques



Résultat de l'algorithme
Decision Tree Classifier



Training dataset
réduit en 3
dimensions grâce au
PCA



Matrice de confusion obtenue
après l'exécution de
l'algorithme SVC.

Résultats

La qualité de la prédiction et le temps d'exécution, grandement impactés par les paramètres utilisés.

Scores :

Algorithme	Scores (précision)	Paramètres algorithme	Temps d'exécution (seconde)
SVC	bacteria: 0.55 (55%) normal: 0.85 (85%) virus: 0.40 (40%)	/	934.9
SVC (avec PCA)	bacteria: 0.52 (52%) normal: 0.71 (71%) virus: 0.35 (35%)	/	26.5
SVC LINEAR	0.74741536590310 (74%)	/	1355.9
SVC LINEAR (avec PCA)	0.63085343604297 (63%)	/	18.0
MLP CLASSIFIER	0.38782051282051 (38%)	solver= lbfgs alpha= 1e-5 hidden_layer_sizes= (784, 3) random_state= 1	118.6
KNN	0.45993589743589 (46%)	k= 70 (calculer par la fonction)	3.5
NAIVE BAYES	0.46794871794871 (47%)	/	10.8
DECISION TREE CLASSIFIER	0.52083333333333 (52%)	criterion= entropy max_depth= 5 min_sample_split= 2 random_state= 0	44.4
RANDOM TREE FOREST	0.55288461538461 (55%)	n_estimators= 100 max_depth= 5 min_sample_split= 2 random_state= 0	16.2
EXTREMELY RANDOMIZED TREES	0.60096153846153 (60%)	n_estimators= 100 max_depth= 10 min_sample_split= 2 random_state= 0	13.8

Conclusion

De nombreux algorithmes sont performants et réalisent l'objectif avec succès. Mais leur technologie est relativement ancienne et demande des jeux de données très spécifiques avec beaucoup de traitement.

En revanche :

Le MLP Classifier qui est un réseau de neurones lui peut travailler avec des données plus vastes et moins ordonnées.

Et avec les bons paramètres nous obtenons les meilleurs résultats et les meilleurs temps de traitement.

