

Université de Bourgogne, UFR Sciences et Techniques, Département I.E.M.

Projet Tutoré - M2 BDIA

Groupe
Elio GENCE
Pierre AUGIER
Constantin PERROT
Aymeric OBLIN

Projet tutoré RDT :
Méthodologie Création d'une IA ML

Contents

1	Sujet	4
2	Arborescence	5
3	Que faire pour tester mon IA	5
4	Bilan et Conclusion	6
5	Sources	6

1 Sujet

Dans le domaine de l'analyse des particules de Résidus De Tir (RDT), une part importante du temps de l'expert est dédiée à la classification des RDT, selon les classes et sous-classes décrites par la norme ASTM1588-20 (i.e. particules caractéristiques, particules compatibles et particules associées). En effet, de nombreuses particules sont faussement classées automatiquement par les techniques actuelles comme étant des RDT, et c'est l'expert humain qui doit passer plusieurs heures par jour afin de vérifier les classifications produites par les logiciels actuels.

Le but de ce projet est donc d'analyser les données mises à disposition puis d'appliquer des algorithmes d'IA afin de classer les données sans intervention humaine.

Environnement technologique – Langage Python, IDE Spyder, Bibliothèque Python (Matplotlib, Scikit-learn, Pandas, Numpy)

2 Arborescence







 datasets	03/12/2024 17:22	Dossier de fichiers
 results	06/12/2024 19:22	Dossier de fichiers
 DataCleaning.py	07/12/2024 20:26	Python File
 FileManipulation.py	08/12/2024 23:10	Python File
 KNNTraining.py	12/12/2024 17:39	Python File
 ModelSetup.py	08/12/2024 22:31	Python File

Figure .1: Exemple de l'arborescence pour KNN

Chaque algorithme a son dossier et dans celui ci nous trouvons donc les fichiers suivants :

Datacleaning.py -> Permet le formater les données brut (incomplet pour le moment)

FileManipulation.py -> Permet la manipulation des documents (load, save, etc)

ModelSetup.py -> Permet la manipulation du model (plot, setup xy, etc)

xxxTraining.py -> Ce qui change entre chaque IA (xxx correspond à l'algo en question), s'adapte à l'algo pour setup les bons paramètres

Nous y trouvons également les documents suivants :

datasets -> Divisé en deux dossiers : predict (pour les données sur lesquelles on veut faire la prédiction) et training (pour les données d'entraînement)

results -> Divisé en deux dossiers également : plot et csv

3 Que faire pour tester mon IA

Maintenant tout vos fichiers sont correctement placés, que faire ???

C'est simple : tester votre algorithme !

Chaque algorithme à plusieurs hyperparamètres, pour les trouver il suffit de consulter la documentation scikit-learn (ou simplement des vidéos yt/chatgpt, personne ne viendras vous juger).

Ensuite il faut réaliser des learnings curves, validation curves ou encore grid search

Ces outils ont pour but de vous simplifier la vie, vous leur donner les hyperparamètres à tester et ces fonctions s'occupent de vous trouver la meilleur valeur, a vous ensuite de sauvegarder le tout sous format png (quel chance, la fonction pour le faire est déjà écrite !)

Je vais pas plus en détail maintenant car c'est long à écrire mais inspirez vous de KNN ou de SGD pour comprendre ce que vous devez faire exactement (au pire posez moi des questions)

4 Bilan et Conclusion

ET VOILA (oui ce doc est mieux foutu que nos rapports hebdos et j'en ai rien à foutre)

5 Sources

Lien vers nos fichiers :

- <https://cloud.u-bourgogne.fr/index.php/s/AWgaY45K4mrEMH3>
- <https://github.com/ElioGence/ResidusDeTir>

Source pour en apprendre plus :

- https://scikit-learn.org/stable/machine_learning_map.html
- https://scikit-learn.org/1.5/modules/cross_validation.html
- <https://www.youtube.com/@MachineLearnia>

Source à vérifier pour la prof :

- https://github.com/Machine-Learning-Tokyo/Interactive_Tools
- <https://github.com/h2oai/h2o-3/blob/master/h2o-docs/src/product/tutorials/dl/dl.md>
- <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html>