

# Offre de stage - M2 Informatique

## Clustering dynamique avec DyClee et sa visualisation graphique

(English version below)

### Description et contexte

Dyclee met en œuvre un algorithme de regroupement dynamique qui traite efficacement les flux de données et atteint plusieurs propriétés importantes qui ne sont généralement pas réunies dans le même algorithme. L'algorithme de regroupement dynamique fonctionne en ligne en deux étapes temporelles différentes, une étape rapide basée sur la distance qui génère des micro-clusters et une étape basée sur la densité qui regroupe les micro-clusters en fonction de leur densité et génère les clusters finaux. L'algorithme permet de détecter les nouveautés et la dérive des concepts grâce à une fonction d'oubli qui permet aux micro-clusters et aux clusters finaux d'apparaître, de dériver, de fusionner, de se diviser ou de disparaître. L'algorithme sur lequel repose Dyclee a été conçu pour pouvoir détecter des modèles complexes, même dans des distributions à densité multiple et sans présupposer la convexité des clusters.

**Mots clés :** Machine Learning, Dynamic clustering, Concept drift, data visualization

### Environnement de travail

Le stage s'effectuera au sein de la startup *PostLab* (<https://www.postlab.fr/>) à Marseille, en collaboration avec le *Laboratoire d'analyse et d'architecture des systèmes* (<https://www.laas.fr/public/>). PostLab possède une forte expertise en ingénierie scientifique, autour de projets de recherche en IA. Elle est spécialisée en développement de démonstrateurs et d'applications web issus de travaux de recherche académique. Le stage sera encadré par Driss Sadoun (PostLab) et Louise Travé-Massuyès (LAAS-CNRS).

Durée du stage : 4 à 6 mois

### Objectifs

1. Faire évoluer le code de DyClee (Python)
2. Améliorer l'interface web graphique de DyClee (Symfony, PHP, JS)

### Profil

Diplôme de fin d'études d'ingénieur ou Master2

- Bonnes compétences en programmation (notamment en Python, PHP et JS)
- Bonne connaissance de l'IA, de l'apprentissage automatique et de la science des données
- Connaissance de Git / Gitlab
- Compétences en lecture et rédaction de documentation en français et/ou en anglais
- Une connaissance du framework Symfony et de Docker serait un plus.

### Comment postuler ?

Veuillez envoyer dès que possible votre CV et votre lettre de motivation à [driss.sadoun@postlab.fr](mailto:driss.sadoun@postlab.fr) et [louise.trave-massuyes@laas.fr](mailto:louise.trave-massuyes@laas.fr)

# Internship offer - M2 computer science

## Dynamic clustering with Dyclee and its graphical visualization

### Description

Dyclee implements a dynamic clustering algorithm that efficiently deals with data streams and achieves several important properties which are not generally found together in the same algorithm. The dynamic clustering algorithm operates online in two different time-scale stages, a fast distance-based stage that generates micro-clusters and a density-based stage that groups the micro-clusters according to their density and generates the final clusters. The algorithm achieves novelty detection and concept drift thanks to a forgetting function that allows micro-clusters and final clusters to appear, drift, merge, split or disappear. The algorithm supporting Dyclee has been designed to be able to detect complex patterns even in multi-density distributions and making no assumption of cluster convexity.

**Keywords :** Machine Learning, Dynamic clustering, Concept drift, data visualization

### Work environment

The internship will be conducted within the startup PostLab (<https://www.postlab.fr/>) at Marseille, in collaboration with the *Laboratoire d'analyse et d'architecture des systèmes* (<https://www.laas.fr/public/>). PostLab has a strong expertise in scientific engineering in the scope of AI research project. It specializes in the development of web demonstrators and applications based on academic research. The internship will be supervised by Driss Sadoun (PostLab) and Louise Travé-Massuyès (LAAS-CNRS).

Duration : 4 to 6 month

### Goals

1. Improve the code of the algorithm (Python)
2. Improve the graphic web interface of DyClee (Symfony, PHP, JS)

### Profile

End of engineering degree, M1/M2

- Good skills in programming (especially using Python, PHP and JS)
- Background in AI, Machine Learning and Data Science
- Knowledge in Git / Gitlab
- Skills for documentation reading and writing in French and English
- Some knowledge of the framework Symfony and Docker would be a bonus

### How to apply ?

Please send as soon as possible your CV and motivation letter to [driss.sadoun@postlab.fr](mailto:driss.sadoun@postlab.fr) and [louise.trave-massuyes@laas.fr](mailto:louise.trave-massuyes@laas.fr)

## References

- N. BARBOSA ROA, L. Travé-Massuyès, V. HUGO GRISALES, "DyClee: Dynamic clustering for tracking evolving environments." *Pattern Recognition* 94 (2019): 162-186.
- N. BARBOSA ROA, L. Travé-Massuyès, V. HUGO GRISALES, Diagnosability improvement of dynamic clustering through automatic learning of discrete event models, 20th World Congress of the International Federation of Automatic Control (WC IFAC'17), Toulouse, France, 9-14 July 2017.
- N. BARBOSA ROA, L. Travé-Massuyès, V. HUGO GRISALES, A novel algorithm for dynamic clustering: properties and performance, 15th IEEE International Conference on Machine Learning and Applications, Anaheim, California, USA, 2016.