

---

# DOSSIER SCIENTIFIQUE

---

**Moncef GAROUANI**

**28 Mars 2023**

## **Plan**

<b>A. Curriculum Vitae.....</b>	<b>1</b>
<b>B. Activités d'enseignement.....</b>	<b>3</b>
<b>C. Activités de recherche.....</b>	<b>6</b>
<b>D. Responsabilités et activités scientifiques .....</b>	<b>10</b>
<b>E. Liste des productions scientifiques.....</b>	<b>11</b>

## A. Curriculum Vitae

**M. Moncef GAROUANI**

Né le 04 Avril 1997 à El Menzel, MAROC

✉ [moncef.garouani@eilco.univ-littoral.fr](mailto:moncef.garouani@eilco.univ-littoral.fr)

🌐 [www.mgarouani.fr](http://www.mgarouani.fr)

☎ [+33 6 44 77 99 07](tel:+33644779907)

Situation professionnelle actuelle : Attaché Temporaire d'Enseignement et de Recherche à l'École des Ingénieurs du Littoral Côte d'Opale (EILCO), Laboratoire d'Informatique Signal et Image de la Côte d'Opale (LISIC).

Domaine d'études : Informatique et applications.

Mots-clés : Intelligence Artificielle, Apprentissage Automatique, Explicabilité des IA.

### 1. Diplômes et titres universitaires

2023	<b>Qualification aux fonctions de maître de conférences</b> <b>Campagne</b> : 2023 <b>N° qualification</b> : 23227378439
Sept.2019 – Sept.2022	<b>Doctorat en Informatique</b> (en cotutelle) Université du Littoral Côte d'Opale (ULCO) - Laboratoire LISIC Université Hassan II (UH2C) - Complex Cyber Physical System Laboratory <b>Titre</b> : Towards Efficient and Explainable Automated Machine Learning Pipelines Design <b>Directeurs</b> : Mourad BOUNEFFA (ULCO), Mohamed HAMLICH (UH2C) <b>Jury</b> : P. Parrend, A. Azmani, S. Ventura, S. Verel, A. Majda, N. Youssfi
2017 – 2019	<b>Master Sciences et Techniques (MST)</b> MST en Systèmes Intelligents et Réseaux à la faculté des science et techniques (FST) - Université Sidi Mohamed Ben Abdellah de Fès (USMBA) <b>Mémoire</b> : Sentiment analysis of Moroccan Tweets using text mining <b>Encadrant</b> : Jamal KHARROUBI <b>Mention</b> : Bien ( <i>Prix USMBA du meilleur mémoire M.S. en informatique</i> )
2014 – 2017	<b>Licence Sciences et Techniques (LST)</b> LST en Génie Informatique à la FST de Fès-USBMA <b>Mention</b> : A. Bien

### 2. Activités de recherche menées avant la thèse et depuis la soutenance

Sept. 2022 – Août 2023	<b>ATER</b> à l'École d'Ingénieurs du Littoral Côte d'Opale (Sec. CNU 27) <b>Référent enseignement</b> : Mohammed Benjelloun (Directeur EILCO) <b>Référent recherche</b> : Mourad Bouneffa (Directeur de l'équipe de recherche Ingénierie des Connaissances du LISIC)
Janv. 2018 – Juin 2018	<b>Stage de recherche M2</b> – FST Fès Laboratoire Systèmes Intelligents et Applications (LSIA) de la faculté des sciences et techniques de Fès au Maroc

### 3. Parcours professionnel

Sept. 2022 – Août 2023	<b>ATER</b> à l'École d'Ingénieurs du Littoral Côte d'Opale (section CNU 27, LISIC), Université du Littoral Côte d'Opale.
Nov. 2021– Juill. 2022	<b>Tuteur pédagogique</b> 27 <sup>ème</sup> section (Informatique) - IUT du Littoral Côte d'Opale - École d'Ingénieurs du Littoral-Côte-d'Opale
Août – Déc. 2019	<b>Data Scientist</b> – The Good Data Factory (Rabat, Maroc)  <b>Missions :</b> Collection et prétraitement de données; Concevoir et développer des modèles et des algorithmes d'apprentissage automatique; Développer des outils pour surveiller et analyser les performances des modèles d'apprentissage automatique
Avril 2016 – Juin 2016	<b>Stage de projet de fin d'études L3</b> – Préfecture de Séfrou  <b>Missions :</b> Conception et réalisation d'un système d'information pour la gestion des stagiaires de la préfecture.

### 4. Formations complémentaires

J'ai suivi de nombreuses formations spécifiques et transversales avec des certifications en Informatique et en Intelligence Artificielle au cours de mon parcours (MOOCS, Formations, certifications). J'ai participé également à des écoles d'été sur l'Intelligence Artificielle et l'Explicabilité de l'IA au niveau national et international. A la date de rédaction de ce document, les formations suivies représentent 336 heures. Le tableau 1 récapitule les principales formations suivies.

Intitulé de la Formations	Equipe pédagogique	Total heures
Ethique de la Recherche et Intégrité scientifique	École Doctorale	8
Institut d'Automne en Intelligence Artificielle	Sorbonne Université	26
4eu+ Summer School On Artificial Intelligence	Université de Milan	30
Composition efficace du mémoire de thèse (et autres documents) avec latex.	Denis Bitouzé (ULCO)	40
Traitement et Analyse de Données Expérimentales I (TADE I) et Logiciel R	Rebecca Dauwe (ULCO)	27
Mooc Recherche reproductible : principes méthodologiques pour une science transparente	Fun-Mooc.Fr	24
La carrière universitaire du chercheur	Florence Pilard	5
Certification intelligence artificielle	Gilles Kassel	35
MOOC Doctorat et Poursuite de carrière	Phdooc.Moocit.Fr	24
La carrière universitaire ou de chercheur	Ecole Doctorale STS	5
DELL EMC Data science & Big data analytics	FST Fès	3 mois
IBM Data Science Professional Certificate	IBM sur Coursera	5 mois

Tableau 1: Liste des principales formations complémentaires suivies jusqu'au ce jour.

## B. ACTIVITÉS D'ENSEIGNEMENT

Tous les cours que j'ai assurés ont été effectués à l'université du Littoral Côte d'Opale.

- 2022-2023 : **ATER**, 192 heures, École d'Ingénieurs du Littoral Côte d'Opale.
- 2021-2022 : **Tuteur pédagogique**, 169 heures, IUT Calais & EILCO Calais.

À la date de rédaction de ce document, la charge d'enseignement représente **361 h** (Eq. TD) de présence devant les étudiants (Voir détail ci-dessous).

### 1. Activités d'enseignement en qualité d'ATER

Actuellement, je suis recruté à l'université du littoral côte d'opale en qualité d'attaché temporaire d'enseignement et de recherche. Dans ce cadre, j'assure des missions d'enseignement à l'école des ingénieurs du littoral côte d'opale (EILCO). Mes charges d'enseignement sont dispensées principalement dans la spécialité Informatique de l'EILCO (site de Calais) et dans la spécialité Génie Énergétique et Environnement, GEE (site de Dunkerque) pour les étudiants de la première et la dernière année du cycle ingénieur (ING1 et ING3). L'ensemble des enseignements assurés durant cette année est listé dans le tableau 2.

Etablissement	Niveau	Matière	Volume
EILCO Calais	ING1	Bases de données	TD 20h
EILCO Calais	ING1	Encadrement projets d'études en C Algorithmique avancée et programmation C	TD 30h TP 56h
EILCO Calais	ING1	Conception des systèmes d'information	TP 24h
EILCO Dunkerque	ING1	Bases de données	TD 10h TP 32h
EILCO Dunkerque	ING3	Big Data	TP 8h
EILCO Calais	ING3	Business intelligence & initiation au Big Data	TD 16h TP 24h
EILCO Calais	ING3	IA de confiance (Explicabilité de l'IA)	TP 18h

Tableau 2 – Missions d'enseignement en tant que ATER à l'EILCO.

Les responsables de certaines matières m'ont confié la mission de rédaction des sujets de TD/TP, ce qui m'a permis de m'impliquer dans les modules et de réfléchir à une progression pédagogique adaptée aux matières et aux étudiants. En tant que chargé de TD et de TP, j'ai été amené à suggérer aux responsables de modules à la fin de chaque partie, les sujets d'examens adéquats à proposer aux étudiants pour l'évaluation. ➦ [\*Voici l'évaluation de mes enseignements<sup>1</sup>.\*](#)

#### ▪ Algorithmique avancée et programmation C, EILCO Calais

**Responsable :** Rym Guibadj

**Public :** 76 étudiants ING1 Informatique

**Description :** étude des structures de données dynamiques, des algorithmes avancés et la programmation orientée objet en langage C ++ : classe, objet, encapsulation, structures de données.

**Rôle :** chargé de TP et encadrant de 10 groupes en projets d'études en C.

**Volume :** 86 heures en présentiel.

---

<sup>1</sup> [Evaluation des enseignements GAROUANI Moncef](#)

▪ **Bases de données, EILCO Calais & EILCO Dunkerque**

**Responsable :** Mourad Bouneffa

**Public :** 76 étudiants ING1 Informatique, 34 étudiants ING1 GEE

**Description :** conception et manipulation des bases de données relationnelles : algèbre relationnelle de CODD, modèles conceptuels de données, normalisation des bases de données relationnelles, langage SQL.

**Rôle :** chargé de TD et de TP.

▪ **Business intelligence & initiation au Big Data, EILCO Calais & EILCO Dunkerque**

**Responsable :** Mourad Bouneffa

**Public :** 59 étudiants ING3 Informatique, 26 étudiants ING3 GEE

**Description :** introduction à l'informatique décisionnelle, Conception d'un entrepôt de données, Notions d'ETL et intégration de données, les algorithmes de Big Data.

**Rôle :** chargé de TD et de TP.

▪ **IA de confiance (Explicabilité de l'IA), EILCO Calais**

**Responsable :** Mourad Bouneffa

**Public :** 19 étudiants ING3 Informatique- Mineure Intelligence Artificielle

**Description :** acquérir une compréhension des fondements et des enjeux éthiques des algorithmes de l'IA. Développer des compétences pour évaluer la fiabilité et la robustesse des algorithmes d'IA et pour travailler avec ces systèmes de manière éthique et responsable. Explorer les outils et techniques de l'explicabilité de l'IA.

**Rôle :** chargé de TP.

## 2. Activités d'enseignement en qualité de tuteur pédagogique

Pendant ma thèse, j'ai été recruté pour assurer des missions d'enseignement et faire partie de l'équipe pédagogique de l'ULCO en tant que tuteur pédagogique (chargé d'une mission d'approfondissement des cours, animations liées à certains enseignements, sujets d'examens blancs, aide à la recherche documentaire dans les disciplines concernées). Ces missions ont principalement concerné l'approfondissement des cours et travaux pratiques, et dispensées dans la spécialité d'informatique de l'institut universitaire de technologie, et le cycle préparatoire (CP1 et CP2) de l'EILCO(Math/ Info). Le tableau 3 récapitule ces interventions.

Etablissement	Niveau	Matière	Volume
IUT Calais	BUT1	Langage de programmation python	30h
IUT Calais	BUT1	Probabilités et statistiques	15h
IUT Calais	BUT1	Langage de programmation Java	25h
IUT Calais	BUT1	Encadrement des situations d'apprentissage et d'évaluation (SAÉ) en réseau	10h
EILCO	CP1	Architecture des ordinateurs	24h
EILCO	CP1	Algorithmique 2/ le langage python	35h
EILCO	CP1	Algorithmique 2/ le langage C/C++	23h

Tableau 3 – Missions d'enseignement en tant que tuteur pédagogique.

Au final, ces expériences d'enseignement m'ont permis de dispenser un enseignement équilibré entre les différentes disciplines de l'informatique. Elles m'ont également confronté aux différents aspects de l'enseignement (cours, TD, TP, et encadrement de projets) et à leurs exigences (préparation, rédaction et correction des sujets). Parallèlement à mes missions d'enseignement, j'ai eu l'occasion de faire partie des jurys de semestres et à des comités pédagogiques paritaires en lien avec l'enseignement.

### 3. Encadrement

En tant que doctorant puis attaché temporaire d'enseignement et de recherche à l'EILCOLISIC, j'ai encadré et participer à l'encadrement scientifique et technique de plusieurs étudiants. Le tableau ci-dessous récapitule une partie de mes missions de supervision.

Niveau	Année	Sujet
ING3 (Niveau Master 2)	2022-2023	<b>Co-encadrement de 2 Projets de recherche bibliographique</b> <b>Titre 1 :</b> Recherche bibliographique sur « Explainable Artificial Intelligence applied to image classification ». <b>Titre 2 :</b> Etude bibliographique sur « Automated deep learning and neural architectures search ». Durée : 6 mois
ING3 (Niveau Master 2)	2022-2023	<b>Co-encadrement de 2 Projets d'Innovation et Conception (PIC)</b> <b>Titre 1 :</b> Développement d'une application Mobile pour un Système Automated Machine Learning. <b>Titre 2 :</b> Interface Utilisateur Evoluée pour un Système d'Automated Machine Learning. Durée : 6 mois
ING2 (Niveau Master 1)	2022-2023	<b>Encadrement</b> d'un groupe de 6 étudiants dans le cadre d'un Hackathon <b>Titre :</b> Developement d'une application mobile pour la gestion des tournées rendement d'EDF pour surveiller les positions des vannes et détecter les défaillances dans la centrale nucléaire de Graveline. Durée : 6 mois <i>Le groupe a reporté le 1<sup>er</sup> prix de hackathon WIIND</i>
Master 2	2021	<b>Co-encadrement d'un projet master recherche</b> <b>Titre :</b> Développement d'un moteur d'inférence en Tensorflow 2 ontologiquement explicable. Durée : 6 mois
Master 2	2020	<b>Co-encadrement</b> d'un groupe de 2 étudiants, ENSAM-UH2C <b>Titre :</b> Extraction automatique des limites de bâtiments collectées par un drone du port d'Agadir en utilisant les réseaux de neurones profonds. Durée : 6 mois

\* Pour les projets co-encadrés, ma charge était de 50%.

## C. ACTIVITÉS DE RECHERCHE

### 1. Stage de Master

**Titre:** "Sentiment analysis of Moroccan Tweets using text mining"

**Lieux :** Laboratoire des Systèmes Intelligents & Application, FST de Fès, Maroc

**Encadrant :** Jamal Kharroubi

**Publications :**

- 4 papiers dans des conférences internationales [12, 13, 14, 15]
- 1 rapport de recherche (Mémoire de master)

**Thématiques :**

- Natural Language Processing (NLP)
- Extraction de texte
- Détection d'opinion
- Analyse des sentiments

### Résumé de mes travaux de stage de Master

Ma première expérience de recherche fut un stage de recherche M2 au sien du laboratoire des Systèmes Intelligents & Applications (LSIA) de la faculté des sciences et techniques de Fès au Maroc. Ce stage, début d'une longue collaboration entre le LSIA et l'Institut de linguistique computationnelle (ILC-CNR) de Pise en Italie, a pour finalité de mettre en place des solutions intelligentes dans les domaines de la valorisation du patrimoine et des systèmes de transport intelligents. L'objectif de l'axe sur lequel j'ai travaillé, était **l'étude et l'analyse automatique du contenu textuel des réseaux sociaux au Maroc**.

En raison du manque de ressources (bases de données) pour la langue Arabe, et en particulier le dialecte Marocain, la collecte de données a été la première étape de mon travail. Pour mener à bien cette étude, nous avons construit un corpus de 13.550 tweets valides sur la base de 36.114 tweets collectés et qui ont été étiquetés manuellement comme positifs, négatifs, neutres ou mixtes et classés en Dialecte Marocain ou Arabe Standard.

Afin de développer un système d'analyse automatique des sentiments, nous avons implémenté deux approches. Une approche basée sur l'apprentissage automatique (ML) [12], dont le système reçoit des tweets déjà étiquetés. La deuxième approche, nommée approche lexicale [13], nécessite la construction d'un dictionnaire de mots. Pour ce faire, nous avons construit MAC [14], un dictionnaire de 30.000 mots étiquetés comme positif, négatif ou neutre. Plusieurs expériences comparatives ont été réalisées pour évaluer les méthodes les plus utilisées dans le domaine du traitement automatique du langage naturel tel que : N-grammes, Stemming, Vectorisation, Transformation, et Design d'algorithmes [15].

Les résultats de ce travail ont été publiés dans [12, 13, 14, 15]. Ces publications sont basées sur les résultats de mon travail de stage M2, et les résultats de collaboration avec le laboratoire LSIA en parallèle à ma thèse. A titre de distinction, j'ai eu le prix du meilleure mémoire M.S. en informatique de l'université Sidi Mohamed Ben Abdellah de Fès.



## 2. Travaux de Doctorat

**Titre:** “Towards Efficient and Explainable Automated Machine Learning Pipelines Design”

**Lieux:** Université du Littoral Côte d’Opale & Université Hassan II de Casablanca (cotutelle)

**Soutenance:** 27 septembre 2022, École Nationale Supérieure des Arts et des Métiers de Casablanca, Maroc.

**Jury:**

- N. Yousfi, Université Hassan II de Casablanca, Maroc (**Présidente**)
- P. Parrend, EPITA Strasbourg, France (**Rapporteur**)
- A. Azmani, Université Abdelmalek Essaadi, Maroc (**Rapporteur**)
- E. Ben lahmar, Université Hassan II de Casablanca, Maroc (**Rapporteur**)
- A. Majda, Université Moulay Ismail de Meknès, Maroc (**Examineur**)
- S. Verel, Université du Littoral Côte d’Opale, France (**Examineur**)
- S. Ventura, Université de Cordoue, Espagne (**Examineur**)
- M. Bouneffa, Université du Littoral Côte d’Opale, France (**Encadrant**)
- M. Hamlich, Université Hassan II de Casablanca, Maroc (**Encadrant**)

### Publications:

- 5 articles de revues internationales et 1 article soumis
  - 5 articles publiés [1, 2, 3, 4, 6]
  - 1 soumission (ACM Computing Surveys) [7]
- 5 publications dans des conférences internationales [8, 9, 10, 11, 18]
- 1 papier dans un workshop international [19]
- 2 posters dans des écoles d’été internationaux [20, 21]
- 1 logiciel publié dans le dépôt officiel du langage de programmation Python (PyPI)

## Résumé de mes travaux de thèse

### Contexte et objectifs de la thèse

L’industrie du futur introduit de nouveaux concepts, processus et pratiques conduisant à des mutations profondes dans le pilotage des systèmes d’information associés. Une des problématiques cruciales est l’utilisation de quantités importantes de données, notamment celles produites par les différents dispositifs d’acquisition de données, pour en extraire de la connaissance destinée à la maîtrise des processus de l’entreprise à travers un système d’information évolutif, réactif et adapté aux spécificités de l’industrie 4.0.

L’intelligence artificielle et plus particulièrement l’apprentissage automatique fournit des algorithmes, méthodes et outils permettant l’extraction de connaissances et de modèles à partir des données représentant l’activité d’une entreprise et son environnement, et à apporter plus d’automatisation, de commodité et de satisfaction. Cependant, la plupart des entreprises ne disposent pas de moyens humains leur permettant de déployer efficacement des solutions d’apprentissage automatique. Cela s’explique notamment par le fait que la construction de telles solutions est un processus long et difficile qui nécessite une expertise hautement technique et intersectorielle qui est une ressource limitée. Nous nous intéressons donc à ce besoin d’assistance à l’analyse de données, qui commence à recevoir une certaine attention de la communauté scientifique, donnant naissance au domaine de l’apprentissage automatique automatisé.

L’apprentissage automatique automatisé (AutoML) est devenu un domaine en plein essor qui vise à rendre l’application des méthodes d’apprentissage automatique aussi dépourvue d’intervention humaine que possible. Nos travaux dans cette thèse s’inscrivent donc dans le cadre de la mise en œuvre de processus reposant sur les concepts de l’AutoML destinés aux



acteurs de l'industrie 4.0. Le but étant de disposer de moyens permettant une construction de systèmes d'information intelligents dédiés au pilotage des différents processus de l'entreprise.

### Approches proposées

L'apprentissage automatique automatisé a pour objectif de rendre les techniques du ML accessibles aux ingénieurs et chercheurs scientifiques qui souhaitent mener des analyses avancées mais qui n'ont pas l'expertise requise. Cela peut être considéré comme une démocratisation du ML. L'AutoML est généralement traité comme un problème de sélection d'algorithmes ML et de configuration de leurs hyperparamètres. A cet égard, les approches existantes incluent l'optimisation bayésienne, les algorithmes évolutionnaires et l'apprentissage par renforcement. Ces approches focalisent sur l'assistance de l'utilisateur en automatisant une partie ou l'ensemble du processus d'analyse de données, mais sans se soucier de son impact sur l'analyse. L'objectif a généralement été axé sur les facteurs de performance, laissant ainsi de côté d'autres aspects importants, voire cruciaux, tels que la complexité du calcul, la confiance et la transparence.

Cette observation nous a amenés à orienter nos recherches vers le domaine du Meta-Learning (MtL) et à développer des méthodes alternatives qui apportent une aide à la construction des techniques de modélisation appropriées tout en fournissant le rationnel des modèles ML sélectionnés. En particulier, nous considérons cette demande importante d'assistance intelligente comme un processus de méta-analyse, et nous progressons vers la résolution de deux défis de la recherche en AutoML. Tout d'abord, pour adresser le problème de la complexité de calcul, nous avons étudié une formulation de l'AutoML comme un problème de recommandation, puis proposé une nouvelle conceptualisation d'un système expert basé sur le MtL capable de recommander des pipelines ML optimaux pour une tâche donnée. Dans un deuxième temps, nous avons focalisé sur l'explicabilité du processus d'aide à la décision d'AutoML pour adresser le problème de l'acceptation et la confiance dans ces systèmes généralement vus comme des boîtes noires.

### Contributions majeures de la thèse

Afin d'atteindre nos objectifs et de proposer de nouvelles approches performantes de MtL pour l'assistance à l'analyse de données, différents verrous ont dû être levés. On rappellera ci-dessous ce cheminement et le bilan des travaux et contributions qui en ont résulté.

**Assistance à l'analyse de données :** Motivés par la tendance vers des modèles toujours plus efficaces, des espaces de recherche plus grands et les limites des outils existants, le premier verrou à adresser était spécifiquement lié à la problématique d'assistance des utilisateurs dans le processus d'analyse de données. Plus précisément, l'enjeu consistait à définir de nouvelles méthodes efficaces, interopérables et simples à utiliser pour l'assistance des utilisateurs dans l'analyse de données. Par conséquent, notre réponse a été de proposer et de développer un framework avec un tel objectif. À cette fin, nous avons proposé une nouvelle conceptualisation d'un framework qui tire parti des idées du méta-apprentissage reposant sur les dissimilarités entre jeux de données capable de fournir un support dans le but d'améliorer l'analyse et de réduire le temps passé dans la sélection et la paramétrisation des algorithmes. C'est un outil qui, pour la première fois, ne vise pas à fournir un support d'analyse de données uniquement pour la sélection et le paramétrage des algorithmes, mais est plutôt orienté vers une contribution positive à la confiance dans un système d'aide à la décision aussi puissant en fournissant automatiquement un ensemble de niveaux d'explication pour inspecter les résultats fournis sans avoir à dépendre d'un data scientist pour générer et interpréter tous les tracés et résultats extrêmes.

Nous avons implémenté un prototype du framework proposé, AMLBID, sur une architecture client-serveur, où le serveur coordonne en tant que système d'assistance AutoML, qui, étant donné un problème (ensemble de données), une métrique prédictive souhaitée (précision, rappel, score F1) recommande des algorithmes ML avec une configuration d'hyperparamètres associés qui sont classés en fonction de leurs impacts sur le résultat final de l'analyse. Tandis que le côté client est composé d'une interface graphique conviviale qui facilite la manipulation des ensembles de données, prend en charge la simulation visuelle de divers scénarios, et facilite l'interprétation des résultats obtenus. Parallèlement, nous avons implémenté un module basé sur des règles d'inférence qui guide les utilisateurs finaux, en cas de résultats insatisfaisants renvoyés par l'AutoML, et destiné à améliorer les performances prédictives. Par conséquent, cela peut augmenter la transparence, la contrôlabilité et la confiance en AutoML. Une preuve de concept recommandant simplement l'emploi du workflow passé le plus pertinent a été développée, et a permis de valider l'intérêt de l'approche de méta-analyse envisagée. Une importante série d'expériences de méta-apprentissage a été réalisée pour démontrer la praticabilité de ce cadre d'évaluation. Ces contributions ont donné lieu à des publications en conférences et revues internationales [2, 3, 8, 9, 11].

**Dissimilarité entre jeux de données:** Il existe en effet une forte dépendance entre l'efficacité du méta-apprentissage et la caractérisation des jeux de données étudiés. Afin de permettre de nouvelles approches de méta-analyse et pouvoir identifier les expériences pertinentes dans le contexte de l'utilisateur, nous nous sommes ensuite intéressés à un verrou majeur du domaine : *la caractérisation des jeux de données*. Suite au constat d'une importante perte d'information dans les méthodes de caractérisation communément employées, nous avons proposé une nouvelle architecture neuronale (ANN) qui permet une meilleure caractérisation des jeux de données et par la suite un calcul de dissimilarité plus significatif entre ces derniers. Nous avons défini un ensemble de propriétés désirables pour proposer des fonctions capables de prendre en compte la totalité de l'information disponible, ce qui passe par la caractérisation des attributs particuliers de ces jeux de données. Nous avons ensuite montré que ces dissimilarités permettent de caractériser l'adéquation d'algorithmes de classification avec des jeux de données plus efficacement que des distances traditionnelles, et qu'elles peuvent être employées avec de bonnes performances dans un contexte de classification au méta-niveau pour la sélection d'algorithmes. Ces fonctions de dissimilarité ont enfin permis le développement d'un prototype d'assistance à l'analyse de données basé sur de nouvelles approches de méta-analyse. Ces contributions ont donné lieu à des publications en conférences et revues internationales [1, 10].

**Reproductibilité des résultats:** À cette fin, nous avons matérialisé nos contributions dans le progiciel AutoML transparent, interprétable et auto-explicable AMLBID, qui, étant donné un problème de classification, recommande des algorithmes d'apprentissage automatique qui impactent positivement l'analyse. AMLBID est implémenté en tant que package Python open source pour reproduire les expériences, les analyses et permettre des études plus approfondies. Nous avons longuement évalué nos recommandations sous trois angles. Dans le premier, nous avons vérifié la précision de nos prédictions. Dans le second, nous avons analysé le gain qu'ils apportaient à l'utilisateur final non expérimenté. Enfin, dans le troisième, nous avons analysé les performances d'AMLBID par rapport aux humains dans un scénario réaliste d'utilité et d'utilisabilité [4, 6].

## D. RESPONSABILITÉS ET ACTIVITÉS SCIENTIFIQUES

### 1. Activités d'administration et d'animation de la recherche

Participation à l'organisation de conférences internationales (PC member):

- [AIGC 2023](#) : the 1<sup>st</sup> International Conference on AI-generated Content.
- [MobiSPC 2022](#) : the 19<sup>th</sup> International Conference on Mobile Systems and Pervasive Computing, 9-11 Août, 2022, Canada, Elsevier (Procedia Computer Science series).
- [ICInPro 2021](#) : the 16<sup>th</sup> International Conference on Information Processing.

### 2. Activités de valorisation et de transfert

Tout au long de ma thèse, je me suis fortement investi dans la médiation scientifique des laboratoires LISIC et LCCPS au travers de différentes activités scientifiques telles que :

- Présentations lors des événements du laboratoire et du pôle Mutations Technologiques et Environnementales (MTE).
- Participation à 2 écoles d'été internationales : 4EU+ Summer school, Italie, 2022, et IA2-Institut d'Automne en IA, Paris 2021.
- Participation à des conférences internationales (ICEIS'23, ICDTA'22, SADASC'22, BDIoT'21, ICABDE '21, ICEIS'21, SCA'21, ICI2C'21, ICDTA'21).
- Exposé grand public (Ma thèse en 180) dans une rencontre autour des collaborations entre la recherche et les entreprises, la cité de l'IA (IAHDF), 2022, Lille.
- Exposé dans 4EU+ summer school on artificial intelligence, Université de Milan, 2022  
Titre : Meta-learning for automated algorithms selection and configuration

### 3. Séjours à l'étranger

2020-2022	3 séjours de 6 mois à L'École Nationale Supérieure d'Arts et Métiers de Casablanca au Maroc dans le cadre de la cotutelle de thèse. Une partie de ces séjours s'est déroulée dans le centre d'études et de recherche d'ingénierie et de gestion du groupe des hautes études des sciences et techniques de l'ingénierie et du management- Hestim.
2022	Séjour d'une semaine dans le cadre d'une invitation de l'université de Milan pour assister à '4EU+ Summer school on artificial intelligence' et participer à '4EU+ International Workshop on Recent Advancements in Artificial Intelligence'.
2023	Mobilités Erasmus+ Enseignement et Formation à l'université Yaşar en Turquie (Séjour d'une semaine prévu fin juin 2023).

### 4. Activités d'évaluation scientifique

Depuis la dernière année de mon doctorat j'ai été sollicité pour des missions de review de publications pour les journaux et conférences suivants :

- **Journaux :**
  - Information Fusion (2023), **Q1**, Impact fact. : 17.564
  - Journal of Computing in Higher Education (2022), **Q1**, Impact fact. : 4.045
  - Intelligent Information Management Journal (2022), Impact fact. : 1.69
- **Conférences :**
  - The 3<sup>rd</sup> International Conference on Cloud and Big Data
  - The 19<sup>th</sup> International Conference on Mobile Systems and Pervasive Computing
  - The 16<sup>th</sup> International Conference on Information Processing

## E. LISTE DES PRODUCTIONS SCIENTIFIQUES

Les publications de cette section peuvent être regroupées selon deux grandes catégories :

- Les travaux de mon doctorat à l'ULCO— Laboratoire d'Informatique Signal et Image de la Côte d'Opale [1-4 ; 6-11 ; 18] ont trait à **l'automatisation de la sélection, la paramétrisation et l'explicabilité automatique des algorithmes de l'IA**. Pour ces travaux j'ai contribué à la majorité des démarches scientifiques, de la proposition et l'implémentation des approches, à l'analyse, la rédaction, et à la présentation des résultats, sous la supervision de mon directeur de thèse Mourad Bouneffa.
- Les publications [12, 13, 14, 15], publiées avant et pendant ma thèse, concernent des travaux réalisés au cours de mon stage de recherche M2 (2018) et portent sur le **traitement automatique du langage naturel** (NLP). Ces publications sont basées sur les résultats de collaboration avec le laboratoire LSIA—USMBA. À ceux-là, pour lesquels j'ai été le principal acteur, s'ajoutent des papiers de collaborations ayant tout particulièrement trait à mes thèmes d'intérêts scientifiques [5, 16, 17].

### Résumé

- 6 publications dans des journaux internationaux et 1 article soumis
  - 6 articles (JBD, SoftwareX, IJAMT, IJAI)
  - 1 soumission (ACM Computing Surveys)
- 11 publications dans des conférences internationales (ICEIS, ICABDE, SADASC, etc.)
- 1 publication dans workshop international & 2 posters dans des instituts d'été internat.
- 1 logiciel publié dans le dépôt officiel du langage de programmation Python (PyPI)

Journal	Base de données	Impact factor	Quartile
Journal of Big Data	Scopus	10.835	Q1
The International Journal of Advanced Manufacturing Technology	Scopus	3.563	Q1
SoftwareX	Scopus	2.868	Q2
ACM Computing Surveys	Scopus	14.324	Q1
IAES International Journal of Artificial Intelligence	Scopus	2.16	Q3
International Conference on Enterprise Information Systems	Scopus	-	B1

Table 4 - Rangs des journaux et conférences<sup>2, 3</sup>

<sup>2</sup> <https://www.scimagojr.com/>

<sup>3</sup> <http://www.conferenceranks.com/>

## Journaux

1. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M. Autoencoder-kNN meta-model based data characterization approach for an automated selection of AI algorithms. *Journal of Big Data* (2023) DOI : [10.1186/s40537-023-00687-7](https://doi.org/10.1186/s40537-023-00687-7)
2. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M., *et al.* Towards big industrial data mining through explainable automated machine learning. *The International Journal of Advanced Manufacturing Technology* (2022). DOI: [10.1007/s00170-022-08761-9](https://doi.org/10.1007/s00170-022-08761-9)
3. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M., *et al.* Using meta-learning for automated algorithm selection and configuration: an experimental framework for big industrial data. *Journal of Big Data* (2022) DOI : [10.1186/s40537-022-00612-4](https://doi.org/10.1186/s40537-022-00612-4)
4. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M., AMLBID: An auto-explained Automated Machine Learning tool for Big Industrial Data. *SoftwareX* (2021) 100919, DOI : [10.1016/j.softx.2021.100919](https://doi.org/10.1016/j.softx.2021.100919)
5. Chaabi M., Hamlich M., and **Garouani M.**, Product defect detection based on convolutional autoencoder and one-class classification. *IAES International Journal of Artificial Intelligence* (2022). DOI: [10.11591/ijai.v12.i2.pp912-920](https://doi.org/10.11591/ijai.v12.i2.pp912-920)
6. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M., AMLBID2.0: An auto-explained Automated Machine Learning tool for Big Industrial Data. *SoftwareX* [In press]
7. **Garouani, M.**, Ahmad, A., Bouneffa, M., A Survey and Perspective View of Meta-Learning for Automated Algorithms Selection and Parametrization. **Submitted** to *ACM Computing Surveys* (2023).

## Conférences Internationales

8. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M. (2022). Scalable Meta-Bayesian Optimization for Machine Learning Hyperparameters optimization. In: *Smart Applications and Data Analysis*. SADASC 2022. Communications in Computer and Information Science, vol 1677. Springer, Cham., DOI: [10.1007/978-3-031-20490-6\\_14](https://doi.org/10.1007/978-3-031-20490-6_14)
9. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M., *et al.* (2021). Towards the Automation of Industrial Data Science: A Meta-learning based Approach. In *Proceedings of the 23rd International Conference on Enterprise Information Systems - Volume 1: ICEIS*, pages 709-716. DOI : [10.5220/0010457107090716](https://doi.org/10.5220/0010457107090716)
10. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M., *et al.* (2021). Towards meta-learning based data analytics to better assist the domain experts in industry 4.0. In: *The International Conference on Artificial Intelligence and Big Data in Digital Era (ICABDE)*, Lecture Notes on Data Engineering and Communications Technologies. Springer, Cham. DOI : [10.1007/978-3-030-97610-1\\_22](https://doi.org/10.1007/978-3-030-97610-1_22)
11. **Garouani, M.**, Hamlich, M., Ahmad, A., Bouneffa, M., *et al.* (2021). Towards an automatic assistance framework for the selection and configuration of machine-learning-based data analytics solutions in industry 4.0. In *the Proceedings of the 5th International Conference on Big Data and Internet of Things*. BDIoT 2021. Lecture Notes in Networks and Systems, vol 489. Springer, Cham. DOI : [10.1007/978-3-031-07969-6\\_1](https://doi.org/10.1007/978-3-031-07969-6_1)



12. **Garouani M.**, Chrita H., Kharroubi J. (2021). Sentiment Analysis of Moroccan Tweets Using Text Mining. In: *Digital Technologies and Applications. International Conference on Digital Technologies and Applications (ICDTA21)*. Lecture Notes in Networks and Systems, vol 211. Springer, Cham. DOI: [10.1007/978-3-030-73882-2\\_54](https://doi.org/10.1007/978-3-030-73882-2_54)
13. **Garouani M.**, Kharroubi J. (2022). Towards a New Lexicon-Based Features Vector for Sentiment Analysis: Application to Moroccan Arabic Tweets. In: *Advances in Information, Communication and Cyber security (ICI2C)*. Lecture Notes in Networks and Systems, vol 357. Springer, Cham. DOI : [10.1007/978-3-030-91738-8\\_7](https://doi.org/10.1007/978-3-030-91738-8_7)
14. **Garouani M.**, Kharroubi J. (2022). MAC: An open and free Moroccan Arabic corpus for sentiment analysis. In: *Innovations in Smart Cities Applications Volume 5. SCA'2021*. Lecture Notes in Networks and Systems. vol 393. DOI: [10.1007/978-3-030-94191-8\\_68](https://doi.org/10.1007/978-3-030-94191-8_68)
15. **Garouani M.**, Zaysa, K. (2022). Leveraging the Automated Machine Learning for Arabic Opinion mining: A Preliminary Study on AutoML tools and comparison to human performance. In: *Digital Technologies and Applications (ICDTA22)*. Lecture Notes in Networks and Systems. vol 455. Springer, Cham. DOI: [10.1007/978-3-031-02447-4\\_17](https://doi.org/10.1007/978-3-031-02447-4_17)
16. Chaabi, M., Hamlich, M., **Garouani, M.**, (2022). Evaluation of AutoML tools for manufacturing applications. In: *Proceedings of the 12th International Conference on Integrated Design and Production*. Lecture Notes in Mechanical Engineering. [in Press]
17. M. Choai, **M.Garouani**, et al. Automated Decision Support Framework for IoT : Towards a Cyber Physical Recommendation System". In *Proceedings of the 25th International Conference on Enterprise Information Systems (ICEIS 2023)*. [in Press]
18. **Garouani, M.**, Bouneffa, M., Ahmad, A., Explaining Meta-features Importance in Meta-learning through Shapley Values". In *Proceedings of the 25th International Conference on Enterprise Information Systems (ICEIS 2023)*. [in Press]

## Workshops & Posters

19. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M., "Meta-learning for automating industrial data science". In *the 4EU+ International Workshop on Recent Advancements in Artificial Intelligence (2022)*, Gargnano del Garda, Italy.
20. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M., et al. "Towards industrial data science through explainable automated machine learning". POSTER In: *MTE Pole's Doctoral Day (2021)*, ULCO University, Calais, France.
21. **Garouani, M.**, Ahmad, A., Bouneffa, M., Hamlich, M., "Explainable Automated Machine Learning". POSTER In : *IA<sup>2</sup> – Institut d'Automne en Intelligence Artificielle (2021)*, Sorbonne Université, Paris, France. DOI: [10.54985/peeref.2208p4898652](https://doi.org/10.54985/peeref.2208p4898652)

## Logiciels

22. **Garouani, M.**, Bouneffa, M., Ahmad., AMLBID: An auto-explained Automated Machine Learning tool for Big Industrial Data. Available at: [Python Packages Index](#)

## Thèses

**Garouani, M.** (2019-2022) Towards efficient and explainable Automated machine learning pipelines design: Application to industry 4.0 data. **Thèse de doctorat** en cotutelle entre l'université du littoral côte d'Opale et l'université Hassan II. Sous la direction de Mourad Bouneffa et Mohamed Hamlich. [theses.fr/2022DUNK0620](https://theses.fr/2022DUNK0620)