

# Hamza El BACCOURI

## Data Scientist

📍 : Grenoble, France

☎ : 06 15 62 33 24

✉ : baccouri.hamza@gmail.com

Expérience : 6 ans





Secteurs d'activités : Robotique, Télécommunications, Retail, BTP

## COMPÉTENCES TECHNIQUES / EXPERTISES

- **Machine Learning** : Scikit-learn, Keras, TensorFlow, TensorFlow Data Validation, TensorFlow Extended (TFX), PyTorch
- **Outils techniques** : Talend DI, Anaconda (Spyder), PyCharm, Jupyter Notebook, Spark (Pyspark, Sparksql, MLlib), Rstudio, Pandas, NumPy, SciPy, Gradio, LangChain, Hugging Face, IBM Watson, Vertex AI
- **Visualisation de données** : Matplotlib, Seaborn, Dash, Tableau
- **Langages de programmation** : Python, R, C, C++, Java, UML, MySQL, SQL, Access, HTML, CSS, JavaScript
- **Systèmes d'exploitation** : Windows XP, 7, 8, 10, Linux
- **Bases de données relationnelles** : Oracle, MySQL, Postgresql, MongoDB
- **Cloud** : Google Cloud Platform (Cloud storage, cloud SQL, Dataflow, Data Proc, Compute Engine, google Kubernetes engine, Vertex AI), AWS (Amazon SageMaker, AWS data wrangler, AWS glue, AWS Athena), Azure cloud platform (Azure Machine Learning, Azure DevOps, Azure SQL Database, Azure Cosmos DB)
- **GCP ML** : BigQuery, Cloud Composer, AI Platform, Kubeflow
- **Big Data** : Spark, Hadoop, Hive
- **Apache Spark** : Spark SQL, Scala, Pyspark, DataFrame
- **Conteneurisation** : Docker
- **Apprentissage automatique supervisé** : Régression linéaires et logistiques, KNN, arbres de décision, Gradient boosting
- **Apprentissage automatique non supervisé** : Clustering, K-means, Réduction de la dimensionnalité
- **Système de recommandation** : Filtrage basé sur le contenu, Filtrage collaboratif, La factorisation matricielle, Word2Vec, Prod2Vec, NLP, TF-IDF, KNN, FAISS, Implicit, LightFM, RankFM, Surprise
- **Les règles d'association** : Apriori, FP-growth, Eclat, CARMA
- **Deep learning** : Réseaux de neurones (CNN, RNN, DNN), Transformers, GAN (Generative Adversarial Networks), VAE (Variational Autoencoders)
- **Natural language processing (NLP)** : NLTK, Gensim, TF-IDF, Spacy, Hugging Face Transformers
- **Customer segmentation** : RFM (Recency, Frequency, Monetary)
- **IA générative** : LLM (GPT-3, GPT-4, LLaMA-2, T5, DALLÉ, BLIP...), RAG (Retrieval-Augmented Generation),
- **Prompt Engineering** : Zero-shot, Few-shot, Interview Pattern, Chain of Thought (COT), Tree of Thought

- **Chatbots and AI Applications** : Image Captioning, ChatGPT-like Website, Voice Assistant, AI-Powered Meeting Assistant, Private Data Summarization
- **Computer Vision** : Object Detection, KerasCV
- **Time Series Forecasting** : Long Short-Term Memory (LSTM), RNN, Facebook Prophet, SARIMAX, LinkedIn Silverkite
- **Version Control** : Git, GitHub
- Gestion des données
- Capacité d'analyse
- Rédaction des spécifications fonctionnelles et techniques
- Pilotage technique des projets R&D

## PARCOURS PROFESSIONNEL

Date début - fin	Nom de la société	Fonction
02/2023- jusqu'à présent	 Sydev	Data Scientist, Data Engineer
10/2021-01/2023	 Talys Consulting : mission chez Franprix	ETL, Data Engineer
10/2017-12/2020	 LAB-STICC	Data Scientist, Ingénieur de recherche
02/2016-07/2017	 Orange	Ingénieur planification et optimisation du réseau

## FORMATIONS

Année d'obtention	Diplômes / Formation / Ecoles / Habilitation
2020	Doctorat (PHD) en informatique (IA, machine learning) UBO, Brest
2015	Diplôme d'ingénieur en télécommunications, ENIT
2010	Baccalauréat Mathématiques

## CERTIFICATIONS

### Coursera :

- Machine Learning Engineering for Production (MLOps) Spécialisation
- Practical Data Science
- Deep Learning Spécialisation
- Machine Learning Spécialisation
- IBM Data Science certificat professionnel
- Generative AI with Large Language Models
- Microsoft Azure Data Scientist Associate (DP-100)
- Generative AI for Everyone
- Building Generative AI-Powered Applications with Python
- Generative AI : Prompt Engineering Basics

## Datacamp :

- Datacamp Data Scientist avec Python
- Datacamp Data Analyst avec Python
- Datacamp Machine Learning Scientist avec Python

## LANGUES

Langues Maitrisées	Niveau
Anglais	Courant
Français	Courant
Arabe	Langue Maternelle
Espagnol	Notions

## EXPERIENCES PROFESSIONNELLES

SOCIETE - Ville	Sydev - Grenoble
Projet	Système de recommandation de produits
Fonction	Data Scientist
MM/AA - MM/AA	02/23 – jusqu'à présent
Objectif / Mission	Mise en place d'un système de recommandation de produits similaires et complémentaires

### Activités

- Collecte de données à partir de sources diverses (SQL Server, ElasticSearch et PostgreSQL)
- Mise en œuvre d'un système de recommandation de produits similaires :
  - ❖ Extraction des caractéristiques pertinentes des données textuelles via NLP pour calculer la similarité entre les produits
  - ❖ Chargement des images des produits à partir de leurs URL (PIL, ThreadPoolExecutor) et extraction des features à partir des images (MobileNetV2)
  - ❖ Utilisation des techniques d'indexation et de recherche de voisins les plus proches pour accélérer les calculs de similarité (Faiss, Annoy)
  - ❖ Vectorisation des noms de produits, suivies d'un clustering (MiniBatchKMeans) et calcul des mesures de similarité (Cosine) entre les produits d'un même cluster
- Mise en œuvre d'un système de recommandation de produits complémentaires :
  - ❖ Utilisation des données transactionnelles pour identifier des associations entre les produits achetés ensemble.
  - ❖ Application des techniques d'analyse de panier de marché (MBA) et des règles d'association
  - ❖ Mise en œuvre d'un modèle de similarité sémantique entre

	produits basé sur Skip-gram with negative sampling pour générer des recommandations de produits complémentaires. <ul style="list-style-type: none"> <li>• Génération de recommandations précises et non personnalisées en partant d'un grand volume de données historiques (transactionnelles)</li> <li>• Implémentation d'une application Flask et Dash pour l'évaluation des recommandations de produits similaires et complémentaires</li> </ul>
Informatique  Outils	Python, Jupyter Notebook, Dash, Flask, Scikit-learn, Pandas, traitement d'images (PIL, MobileNetV2, VGG16, Resnet), NLP (Gensim, TfidfVectorizer, CountVectorizer, HashingVectorizer, NLTK, Skip-gram, Word2vec, implicit, lightfm), Approximate nearest neighbor (Faiss, Annoy), clustering (KMeans, MiniBatchKMeans), Market Basket Analysis (association Rules, mlxtend, apriori, fpgrowth, pycaret, ECLAT), Elasticsearch (Elasticvue), Azure DevOps, SQL Server, PostgreSQL, ThreadPoolExecutor

<b>SOCIETE - Ville</b>	<b>Talys Consulting - Paris</b>
	<b>Mission chez Franprix Leader Price (Département Informatique)</b>
Projet	Equipe COLOR
Fonction	Data Engineer
MM/AA - MM/AA	04/22 – 01/23
Objectif / Mission	Collecte des données Logistiques et des Référentiels
Activités	<ul style="list-style-type: none"> <li>• Conception des Flux Talend</li> <li>• Développement des Jobs Talend</li> <li>• Collecte, gestion et migration des données (base de données, fichiers, Cloud...)</li> <li>• Alimentation du datawarehouse</li> <li>• Automatisation des traitements des fichiers</li> <li>• Suivi Quotidien de données Référentiels</li> <li>• Faire les demandes d'ordonnancement des flux Talend</li> </ul>
Informatique  Outils	Talend TOS, Talend TAC, MySQL, Oracle, Snowflake, CosmoDb, JSON  Excel, Dbeaver, SQLyog

<b>SOCIETE - Ville</b>	<b>Talys Consulting - Paris</b>
Projet	Viapost
Fonction	Consultant Talend
MM/AA - MM/AA	10/21 – 03/21
Objectif / Mission	Récupération des données (Viapost/Siemens) d'un service web sous forme de fichier Json via TALEND, mapping des fichiers JSON pour alimenter une

## BD MySQL

Activités	<ul style="list-style-type: none"> <li>• Etude et analyse des besoins fonctionnels et techniques à partir d'un cahier de charge</li> <li>• Extraction des fichiers Json (WEB SERVICE GET)</li> <li>• Conception de BD (Target) (BD MYSQL) et création des Tables (Target)</li> <li>• Spécification des différents blocs techniques (Extraction des données, mapping des données,) et la relation côté base de données</li> <li>• Mise à jour des données BD MySQL via Talend</li> <li>• Création de variables de contexte pour exécuter les jobs Talend dans différents environnements (Dev, Test, Prod)</li> </ul>
Informatique Outils	Talend TOS, Talend cloud, MySQL, Web Service, JSON

<b>SOCIETE - Ville</b>	<b>LAB-STICC, Brest</b>
Projet	Automatisation de la mise au point de lois de contrôle de systèmes cyber-physiques (drones, engins flottants...) évoluant en environnement incertain et perturbé,
Fonction	Data Scientist
MM/AA - MM/AA	10/2017 – 12/2020
Objectif / Mission	<p>Proposer une approche pour automatiser le réglage et la mise au point de lois de contrôle de systèmes cyber-physiques</p> <ul style="list-style-type: none"> <li>• Hors-ligne : via simulation, l'utilisation d'algorithme de réduction de dimensionnalité et de clustering</li> <li>• En ligne : selon l'analyse des objectifs visés, utilisation des réseaux de neurones pour la régression</li> </ul>

Activités	<ul style="list-style-type: none"> <li>• Collecter et préparer les données en vue de l'analyse, les normaliser tout en conservant leur intégrité</li> <li>• Effectuer une analyse statistique univariée et multivariée</li> <li>• Interpréter des données : identifier des tendances et évaluer la pertinence des données (Feature Selection)</li> <li>• Manipuler des données avec des bibliothèques Python spécialisées</li> <li>• Mettre en œuvre des techniques de réduction de dimension</li> <li>• Développer des algorithmes de Machine Learning en python (Régression, clustering, classification)</li> <li>• Analyse, conception et réalisation de pipelines complètes de Machine Learning</li> <li>• Automatisation du contrôle de systèmes cyberphysiques (drones, engins flottants...) en environnement incertain et perturbé, réglage et mise au point grâce à l'intelligence artificielle</li> <li>• Régression, analyse statistique univariée et multivariée, clustering, classification, réseaux de neurones, réduction de dimensionnalité</li> <li>• Évaluation des performances du paramétrage trouvé avec un code ROS (Robot Operating System)</li> <li>• Création d'un Dashboard sur un navigateur web et y afficher les graphiques et indicateurs pertinents</li> <li>• Assurer avec d'autres intervenants le portage des modèles sur les applications cibles</li> <li>• Reporting et présentation de résultats et être capable de vulgariser des techniques proches de l'état de l'art</li> <li>• Faire des benchmarks des performances de différents algorithmes</li> </ul>
-----------	--

- Rédiger de la documentation technique

Informatique Outils	Python (Scikit-learn, Pandas, Numpy, Matplotlib, Seaborn, Dash, Spark), SQL, Keras, TensorFlow, Docker
<b>SOCIETE - Ville (dept)</b>	<b>Orange</b>
Projet	Réseau 2G/3G/4G
Fonction	Ingénieur planification et optimisation du réseau
MM/AA - MM/AA (durée)	02/2016 – 07/2017
Objectif / Mission	
Activités	<ul style="list-style-type: none"> <li>• Faire le pré-design du réseau d'accès et les études techniques spécifiques pour améliorer la qualité de service voix et data</li> <li>• Dimensionnement et positionnement des sites, prise en compte des contraintes environnementales, validation des spécifications du site : Coordonnées, le type d'antenne, la hauteur de l'antenne, les azimuts de l'antenne</li> <li>• Analyse des performances des nouveaux sites après intégration (drive test and/or OSS KPIs)</li> <li>• Suivi de la qualité du réseau d'accès radio (GSM, UMTS et LTE) pour assurer la couverture des zones ainsi que la qualité de service conformément aux objectifs du marketing.</li> <li>• Suivre, analyser les KPIs, lancer les actions d'optimisation et traitement des plaintes des clients.</li> <li>• Adapter, tester et évaluer les nouvelles méthodes et technologies afin d'optimiser le réseau radio (Features, Trials &amp; Tuning)</li> <li>• Suivi régulier de l'état d'avancement des SSV 4G (s'assurer de la bonne implémentation des paramètres Radio)</li> </ul>
Informatique Outils	Environnement technique : PRS, MAPINFO, U2000, Astellia, CME, ASSET, Microsoft Office, LMT, OSIRIS
<b>SOCIETE - Ville (dept)</b>	<b>MITSUBICHI ELECTRIC, RENNES</b>
Projet	Communication par Satellite
Fonction	Ingénieur de recherche
MM/AA - MM/AA (durée)	04/2015 – 10/2015
Objectif / Mission	Projet de fin d'études
Activités	<ul style="list-style-type: none"> <li>• Développement et mise en place des algorithmes de Pre-distortion afin de combattre les non-linéarités introduites par les amplificateurs de puissance dans le cadre de la modulation SC-FDMA</li> </ul>
Informatique Outils	Outils : Matlab, Simulink