



Abonia SOJASINGARAYAR

**Data scientist - Machine Learning
Engineer – Natural Language
Processing et Computer Vision
Engineer**

Formation et Qualification

Formation	2021 : M2 - Intelligence Artificielle IA School – Boulogne-Billancourt 2020 : M1 - Chef De Projet Digital Institut F2I – Paris (France) 2015 : Licence en technologie informatique et Ingénierie Université Pondicherry – Karikal (Inde)
Certification et Badge	Google certified professional data engineer IBM Quantum Conversation Badge Machine learning specialization par deeplearning.IA , Andrew NG Deep learning & natural language specialization par deeplearning.IA,Andrew NG Spark par learning Academy Advance Python bootcamp GCP professional data engineer Badges Watson Assistant Hands on Méthodologie Watson Assistant Méthodologie Watson Fondation Watson Assistant Hands on Méthodologie Watson Knowledge Studio Watson Knowledge Studio Hands on
Langues	Français Intermédiaire Anglais Bilingue Tamoul Maternelle Hindi Bilingue
Medium Writer	https://medium.com/@abonia
Linkedin & Publications	Linkedin : https://www.linkedin.com/in/aboniasojasingarayar/ Github : https://github.com/Abonia1 Google Scholar: https://arxiv.org/abs/2212.11804 https://arxiv.org/ftp/arxiv/papers/2006/2006.02767.pdf

Expérience professionnelle

Profil : Je suis passionnée par l'intelligence artificielle, le deep learning, l'apprentissage automatique, le traitement du langage naturel et la vision par ordinateur. J'ai précédemment travaillé dans les domaines des solutions de gestion RH, l'industrie chimique et la cyber sécurité. Ma mission personnelle est de créer des solutions basées sur l'IA qui résolvent des problèmes à fort impact pour les gens du monde entier et simplifient la vie quotidienne.

Compétences fonctionnelles

- Distribution (Ventes- E-Commerce etc...)
- Industrie (Chimique, IT, Solution RH, Cyber sécurité)

Compétences techniques

- Méthodes et Outils → Développent en mode Agile
- Modélisation : Algorithmes de classifications (SVM, SGD, MLP, Arbre de décision, Réseaux de Neurones...), API développement, optimisation de hyperparamètre, régularisation etc
- Recherche et développement d'un outil d'extraction de contenu.
- Implémentation d'une bibliothèque de prétraitement de texte.
- Expert dans l'art de mise en œuvre d'algorithmes de clustering et de classification.
- Expert dans l'art d'implémentation d'algorithmes de requêtes d'expansion.
- Mise en place d'un outil de calcul de similarité lexicale (WordNet et W2V).
- Catégorisation de texte
- Prétraitement des données (nettoyage et formatage des données)
- Implémenter et tester les classifications
- Tests et évaluation des résultats
- Langages et frameworks → ScikitLearn, Tensorflow, Keras, NLTK, spaCy, OpenCV, Numpy, Pandas, Matplotlib, Seaborn, Spark, PyTorch, Flask, Docker, Spyder, Git, TFS, Apache, VScode, Jupyter, Visual Studio, Conda, Wordpress, phpMyAdmin, Python, HTML5, CSS3, PHP, MySQL, JavaScript, JQuery, Bootstrap, PHP, Framework Codeigniter, C#, ASP.NET, VB.Net, Visual Studio., ubuntu, WSL, databricks, MLFlow, GCP, IBM cloud, Cloud Action ...
- Bases de données : MySQL, T-SQL, PostgreSQL, BigQuerySQL

Historique de Carrière

**Juillet 2021 –
En cours**

Projet Intern

- Etude des produits IBM Watson AI : Watson Assistance pour construire un chatbot pour la transformation d'entreprise, Watson Knowledge Studio pour l'annotation de projet NLP.
- Optimisation d'algorithmes de séries temporelles pour l'industrie de la mode - IA4Fashion - IBM Research. Pour prédire la demande future du produit en fonction de la saison - Arima, fbprophet.

- Projet de recherche CodeNet IBM - IBM Research : programme/fonctions de prévision qui effectuent une tâche similaire. Ensemble de données avec 14 millions d'échantillons de code, dont chacun est destiné à résoudre l'un des 4000 problèmes de codage.
- KPI recrutement Dashboard : supprimer les données Trello et visualiser le tableau de bord des données via un superset. . Déclenchez le processus de mise au rebut/l'appel API et ajoutez-le à la base de données postgresql dans le cloud via la fonction IBM Cloud. Visualisation de données superset.
- Analyse du sentiment de satisfaction client avec TextBlob et NLTK Vader.
- Exploration de données
- Prétraitement du texte stemming, lemmatisation etc
- Analyse watson : Améliorez les performances et l'efficacité du chatbot à l'aide de la boîte à outils de l'assistant, des statistiques de métriques.
- Détection humaine par computer vision en temps réel : Utilisation des poids YOLO et réglage fin du modèle pré-entraîné. Modèle YOLOv5 de darknet utilisant un hub de torche.

Outils et techniques : CPFD, Python, NLTK, Seaborn, py-trello, Matplotlib, IBM cloud, IBM object cloud storage, Cloud function, Superset, code engine, BERTopic, scikit learn, PostgreSQL...

Nov – Dec 2022**Mission RATP Data Factory – Data Analyst & Data Engineer**

Construire pipeline ETL

Analyser des données et une étude comparative des cdv et de la localisation du train. Préparation de l'environnement technique.

Outils: python, databricks, MLflow, pyspark, pandas, power BI, ubuntu, wsl,

Oct 2022**Mission RGE – NLP Engineer**

- Préparation d'un conseil pour la création/génération de texte et l'analyseur de courriels
- Documenter la proposition, les idées, la recherche.

Etude accélérée - Transfer learning/modèle pré-entraîné, étude du modèle de résultat de l'état de l'art, données de référence, Huggingface,...

S**ept – Oct 2022****Mission Orange - Assistante de formation technique WDS**

- Préparation de données adapté à la formation
- Watson Discovery Search - Préparation du matériel de formation technique

- Assister à une présentation de formation une fois/semaine

Outils et techniques : Watson discovery search, LDA analysis, Gensim, NLTK, spaCy...

**Juin 2021 –
Juillet 2021**

Mission Allianz France et Whatfix – API integration Engineer

- Assistant Watson et intégration API pour la plateforme Digital Coach - Allianz France.
- Entraînement et intégration du chatbot d'assistance watson
- Spécification de open API : intégration avec IBM app connect
- Intégration de l'API Watson avec l'équipe Whatfix d'Inde

Outils et techniques : Watson Assistant, whatfix platform, WA API...

**Nov 2021 –
Juin 2021**

Mission BNPP – RPA Developer

- RPA et OCR
- Automatisation intelligente des processus bancaires et d'assurance avec Blueprism et Abbyy Flexicapture.
- Détection de date manuscrite à partir d'images de pdf

Outils et techniques : Blueprism, Abbyy Flexicapture

**Juillet 2021 –
Oct 2021**

Mission SNCF – Data Scientist

- Classer les documents pdf scannés avec la technique ML et OCR
- Anonymiser les documents avec la bibliothèque cv2
- Optimisation de l'algorithme de clustering
- Préparation du jeu de données d'images
- Préparation de la documentation technique
- POC - Appliquer SpaCy NER pour détecter les entités nommées à partir du texte rendu du moteur OCR

Outils et techniques : python, openCV, Clustering- DBSCAN, K-Means, pandas, numpy, matplotlib, jupyter, IDE-vscode, SpaCy...

**Mai 2021 –
Juin 2021**

Malt – Freelancer - Deep learning Engineer

- Optimisation des paramètres de AlphaGo model en Tensorflow
- Implémentation de mobinet

Outils et techniques : python, tensorflow, Google colab-GPU...

InFlows AI – Machine Learning Engineer (Alternance)

**Jan 2020 –
Oct 2020**

- Construction d'un pipeline auto-ML
- Utiliser des techniques d'apprentissage automatique et de modélisation statistique pour développer et évaluer des algorithmes afin d'améliorer les performances, la qualité, la gestion des données et la précision.
- Échantillonnage analyse avec des données chimiques
- Expertise en analyse et visualisation de données
- Analyse comparative de ML algorithmes

Outils et techniques : python, pandas, numpy, matplotlib, spark, Scikit learn, XGboost, decision tree, SVM, RF...

**Mars 2018 –
Aout 2019**

Neocase Software – Research Engineer (Alternance)

- POC-Case Classifier est le projet permettant de détecter la catégorie principale des cas(messages) avec ML.NET
- Développer un modèle pour détecter la catégorie et la sous-catégorie de message (Multi Class Classification (SGD, Decision tree), Flask App avec FastCGI Module en Python)
- Développer des API's pour appeler un modèle d'apprentissage machine afin de prédire la catégorie Correspondante du message
- Développer une application Web dans le framework Flask en python pour fournir une interface conviviale permettant de tester le modèle avec une variété de messages.
- Rédaction et gestion de document de recherche.
- Data Processing : Réalisation d'un modèle pour nettoyer les données - Le modèle peut détecter le nom propre, l'adresse électronique, les numéros de téléphone mobile, les mots courants dans l'ensemble de données. Expérimenter avec Stanford POS tagger etc en python.
- New UX- Nouvelle interface de bootstrap dans l'application de service RH optimisée par Neocase en ASP.Net, CSS3, js.
- Translation API - API en C# qui automatise le processus de traduction du back office et du front office de l'application Neocase.
- Resx Management, Développé et expérimenter avec REST API et Swagger Documentation en C#
- Application Windows : Application en C # avec interface utilisateur conviviale pour la traduction automatique en front et back office du

service RH.

- Gestion de base de données : Gestion de la relation entre les tables dans différents clients, Mise à jour d'une opération de données.
- Unit testing API routes.
- Azure Devops-Gérer le déploiement avec devops
- Développement de projet agile.

Outils et techniques : python, MySQL, Azure, pandas, numpy, matplotlib, Scikit-Learn, NLTK, spaCy, tensorflow, tensorboard, C#, REST API, visual studio, swagger, Window form application, .NET, ASP.Net, Bootstrap, javascript, ...

**July 2017 –
Jan 2018**

Enzynov – Full Stack Developer (Stage)

- Recoder le site (enzynov.fr) qui était sous wordpress en framework Bootstrap et Codeigniter (php) pour intégrer la plateforme qui permet aux clients de trouver des produits correspondant à leur secteur et leurs besoins.
- Développer une plateforme en architecture MVC pour gérer les stocks et les produits côté admin.
- Créer le site e-commerce (Enzymarine) et le Panel d'administration et l'outil statistique pour bien gérer le site par l'administration et les employés.
- Développer une application Windows (Enzynov Stock Management System)

Outils et techniques : PHP (Codeigniter), MySQL, phpMyAdmin, apache, jquery, C#, REST API, swagger, Window form application, .NET, ASP.Net, Bootstrap, HTML5, CSS3...

**Jan 2017 –
Apr 2017**

Milk Asus Management – Technicienne d'informatique (Stage)

- Installer Apache 2.4, PHP et MySQL sur Windows Serveur 2012 R2.
- Clonage du Disque Dur.
- Connecter le disque dur dans le PC et configurer les paramètres.
- Développement front-end

Outils et techniques : Clone de disque dur, configuration du système de serveur de windows, HTML, CSS...

**Sep 2016 –
Dec 2016**

Cyber Café – Technicienne d'informatique (Stage)

- Résoudre les problèmes de réseau et la mise en œuvre du réseau.
- Installer, configurer et maintenir des logiciels de sécurité informatique.

- Résolution des problèmes de sécurité et de sauvegarde des données.
- Clonage de disque dur et configuration des paramètres du disque dur

Outils et techniques : Clone de disque dur, maintien du poste PC, réparer l'imprimante...