

ESSOMBA Chris

Candidature au poste d'ingénieur Data

✉ cessomba.ir2025@esaip.org ☎ 0779455983 [in](#) /in/chris-essomba-360780242/ [github.com/ChrisEssomba](#)
[https://chrisessomba.github.io/](#)

PROFIL

Je suis étudiant apprenti en dernière année d'école d'ingénieur en informatique, avec une spécialité en Big Data et en cloud computing. Je suis aussi trilingue et passionné par la Data Science. Mon objectif est de mettre à profit mes compétences techniques et transversales pour apporter une réelle valeur ajoutée aux professionnels métiers grâce à la donnée.

PROJETS ACADÉMIQUES

Modélisation Prédictive des Décisions de Crédit

02/2023 – 05/2023

Objectif : Réaliser une analyse exploratoire approfondie des données de demandes de crédit afin d'identifier les facteurs clés influençant l'octroi ou le refus d'un prêt. Développer un modèle prédictif performant permettant d'anticiper la décision d'octroi sur la base d'informations client et financières.

Insights:

- **Le taux d'endettement** (debt-to-income ratio) supérieur à 40 % est associé à un taux élevé de refus de crédit.
- **Le modèle prédictif final atteint une précision de 87 %**, avec une AUC de 0.91, assurant une prédiction fiable pour l'octroi de crédit.

Conception d'une IA de résumé de textes

03/2025 – 06/2025

Objectifs: Développer une solution complète de traitement automatique du langage (NLP) permettant de générer des résumés de textes, et la déployer sur le cloud via une interface web intuitive.

Taches réalisées:

- Mise en place d'un pipeline d'ingestion automatisé des données d'entraînement à partir du hub Hugging Face.
- Entraînement d'un modèle de résumé de texte sur un corpus de plus de 10 000 dialogues.
- Évaluation des performances du modèle à l'aide de métriques standards telles que ROUGE-L (0,80) et ROUGE-Lsum (0,88).

Technologies : Python, Transformers, Huggingface, Streamlite, Tensorflow, scikit-learn

EXPÉRIENCES PROFESSIONNELLES

Data Scientist Alternant (01 an),

Institut national de l'environnement industriel et des risques (INERIS)

10/2024 – aujourd'hui
Verneuil-en-Halatte, France

- Développement d'un workflow basé sur des grands modèles de langage (LLMs) pour extraire des données à partir de fichiers PDF, réduisant le temps d'extraction manuelle de 30 %.
- Conception d'une application utilisant un modèle de machine learning supervisé pour classifier les composés chimiques selon leur structure, facilitant la détection de similarités structurales dans de larges jeux de données (>1000 molécules).
- Déploiement d'un algorithme de clustering non supervisé pour regrouper les substances selon leur profil toxicologique, permettant une estimation automatique de la toxicité des substances inconnues avec une précision de 80%.

Technologies : Python, LLMs, NLP, Azure (Data Lake, Synapse, Kubernetes), Docker, GitHub Actions (CI/CD)

COMPÉTENCES TECHNIQUES

Langages de Programmation :

- Python, Java, SQL, R, NoSQL, PySpark, Oracle(PL/SQL).

Machine Learning & Deep Learning :

- TensorFlow, PyTorch, Transformers, NLP, LLM, RAG.

Visualisation des Données :

- Matplotlib, Seaborn, Power BI, Tableau, Qlikview

Cloud Computing

- Azure cloud (DataFactory, Databricks, Kubernetes).

CERTIFICATIONS

Azure AI Fundamentals (AI-900) | Azure Data Engineer Associate (DP-203) | IBM Data Science

FORMATION ACADÉMIQUE

Master 2 -Mathématiques appliquées à la science des données,
Université Paris Cité

09/2025 – 09/2026 | Paris, France

Cycle ingénieur en informatique et réseaux - Spé Big Data & IA,
ESAIP, école d'ingénieurs

09/2022 – 09/2025 | Angers, France