

Ahmed Tidiane BALDE

Ingénieur en Machine Learning

 atidiane |  ahmed-balde |  ahmedt.balde@gmail.com |  +33 6.69.78.36.06

PROFIL

Ingénieur en Machine Learning avec une expérience hybride en recherche et en consulting. En parallèle, mentor passionné par l'enseignement sur des parcours de Data Science et d'IA. Intéressé majoritairement par les domaines de la vision par ordinateur et du traitement naturel du langage en particulier par les grands modèles de langage. Apte à conduire un projet de sa conception au déploiement avec une expérience avérée en optimisation des modèles.

COMPÉTENCES

Analyse de données	Détection et traitement des outliers, imputation des valeurs manquantes, ingénierie des variables, visualisation des données
Machine Learning	Modèles linéaires, arbres de décision, modèles ensemblistes, séries temporelles, réseaux de neurones
Computer Vision	SIFT, CNN, ResNet, Yolo, SORT, DeepSORT, AlphaPose, OpenPose, Segmentation
TAL	Sac de mots, Word2Vec (Skip-gram, CBOW), TF-IDF, RNN, GRU, LSTMs, Transformers

OUTILS

Langages	Python, C/C++, Java, Javascript
Machine Learning	pytorch, tensorflow, OpenCV, scikit-learn, pandas, numpy, scipy, prophet, transformers, peft, langchain, SHAP
Optimisations	deepspeed, TensorRT, onnx
Déploiement	docker, fastapi, mlflow
Tableau de bord	streamlit, gradio, dash
Big Data	Spark, Hadoop
Data Warehousing	SQL, MongoDB
Cloud	AWS, Azure
OS et versionnage	Unix/Linux, Git

FORMATIONS

2017 - 2020	Master DAC (Données, Apprentissage et Connaissance) à Sorbonne Université Sciences (Mention Bien)
2015 - 2017	Licence Mono-Informatique à l' Université de Pierre et Marie Curie (UPMC) (Mention Assez Bien)
2014-2015	Licence 1 Mathématiques-Informatique à l'Université Claude Bernard de Lyon1 (UCBL)
2014	Baccalauréat Scientifique au Lycée Sainte Marie en Guinée

EXPÉRIENCES PROFESSIONNELLES

Senior ML Engineer

Pubstack

Avril 2025 - Présent

Paris

Filtrer les enchères publicitaires

Lors d'enchères publicitaires, l'offre est envoyée à plusieurs acheteurs potentiellement intéressés par l'emplacement publicitaire sur le web. Ayant une place de choix dans l'Adtech, nous développons des modèles de Machine Learning pour filtrer les acheteurs (bidders) et n'appeler que ceux qui ont le plus de chances de manifester un intérêt pour l'espace publicitaire. Tout ceci se passe en quelques microsecondes lorsque vous chargez une page web. Il s'agit donc d'implémenter une architecture robuste et des modèles rapides pour éviter toute latence.

- Implémentation d'algorithmes basés sur les arbres de décision, notamment LightGBM, pour prédire si un acheteur serait intéressé par une enchère donnée;
- Conversion du modèle en ONNX pour assurer l'interopérabilité et accélérer le temps d'inférence;
- Mise en place d'une API robuste et concurrente pour recevoir des centaines de millions de requêtes par jour;
- Mise en place la CI/CD avec Github Actions;
- Gestion de l'infrastructure avec Terraform et Kubernetes pour assurer le réentraînement automatique des modèles via des AWS Step Functions, ainsi que la gestion des ressources (CPU, RAM) des machines déployées;
- Monitoring des performances des machines et des taux d'erreur sur Datadog.

Machine Learning Engineer

Polarys

Feb 2024 - Mars 2025

Colombes

Conception de chatbots

Filiale d'une grande société de transports

Pour remplacer les nombreux FAQs qu'ils avaient à disposition, le client s'est lancé dans la mise en place de chatbots basés sur le RAG (Retrieved Augmented Generation). Ce projet à haute envergure est actuellement déployé en production pour quelques centaines d'utilisateurs pour valider la stabilité du chat. Il sera par la suite adapté à de nombreux cas d'usages pour les différents départements au sein de l'entreprise. L'objectif final étant de le passer à l'échelle de centaines de milliers d'utilisateurs potentiels interne à la compagnie.

- Implémentation de la mémoire de conversation du chatbot avec la base de données Redis;
- Contribution à l'implémentation de techniques pour augmenter les performances du chatbot, entre autres : Prompt Engineering, Sentence Window Retrieval, MMR, reranking, Query Retriever;
- Supervision de stagiaires et d'alternants pour développer des stratégies de chunking avec Azure Document Intelligence et Amazon Textract;
- Amélioration de la CI/CD de l'application avec Jenkins et passage à Gitlab CI;
- Amélioration de l'environnement de travail en local (Docker) pour synchroniser les changements et la communication entre les différents containers;

Chargé de TDs en Probabilités et Statistiques

Université Paris 1 Panthéon-Sorbonne

Fev 2024 - Présent

Paris

Ecole de Management de la Sorbonne

Licence 2

Le programme se concentrera principalement sur les aspects de dénombrement, d'introduction aux probabilités, et sur l'étude des variables aléatoires, tant discrètes que continues. J'aborderai également les concepts de densité de probabilité, les convergences en loi, ainsi que les intervalles de fluctuation et de confiance. Mon rôle consistera à traiter ses concepts avec les étudiants tout en assurant une compréhension profonde et pratique.

Mentor Data Scientist
OpenClassrooms

Fev 2023 - Présent
Paris

Accompagnement d'étudiants dans la réalisation des projets

[voir Parcours](#)

En tant que mentor en Data Science chez OpenClassrooms, une plateforme de formations en ligne, je guide les étudiants dans leurs projets pratiques. Ma tâche inclut l'assistance technique, l'explication de concepts théoriques, et la démystification des modèles et algorithmes. Je réalise un suivi hebdomadaire pour fixer des objectifs atteignables, rédige des rapports de progression, et prépare les étudiants à présenter efficacement leurs projets lors des soutenances.

- Présenter les projets et expliquer les attentes et compétences à valider;
- Guider et répondre aux questions techniques et théoriques des étudiants sur les notions de Data Science;
- Définir les objectifs à atteindre pour chaque étudiant pour la session suivante et s'assurer à respecter le délai défini en amont pour la validation de chacun des projets;
- Rédiger un rapport hebdomadaire sur leur progression;
- Valider les livrables des étudiants avant de les présenter à la soutenance.

Machine Learning Engineer
Pivot & Co

Dec 2022 - Fev 2024
Paris

R&D Proof-of-concepts

Internal - [voir HF](#)

Face à la complexité croissante des régulations ESG et à la nécessité pour les entreprises d'y répondre de manière efficace, nous avons identifié le besoin d'un outil interactif pour simplifier l'accès à ces informations. C'est dans cet esprit que le concept d'un chatbot ESG a été développé. Ce chatbot vise à offrir une interface conviviale et intuitive pour les utilisateurs, leur permettant de comprendre et d'interroger facilement les régulations ESG. Il s'agit d'une réponse directe à l'augmentation des exigences en matière de conformité et de transparence, offrant ainsi un moyen rapide et accessible de naviguer dans les complexités des normes ESG.

- Génération de données basées sur les régulations ESG en s'inspirant de ce qui a été fait sur *standford alpaca* sous format (Instructions Input Output);
- FineTuning de Falcon-7B sur les données générées en utilisant du Parameter Efficient Fine Tuning (PEFT) et Low Rank Adaptation (LoRA);
- Déploiement du modèle sur HuggingFace *Inference Endpoints* et développement d'un dashboard avec *streamlit* pour l'interaction;
- Même expérience réalisée avec les API d'OpenAI couplées à *langchain* pour créer un prototype de QA sur des documents sur lesquels nous avons créé des vectors stores avec ChromaDB.

Machine Learning and Software Engineer
Institut National de Recherche en Informatique et en Automatique (INRIA)

Mars 2021 - Nov 2021
Paris

Développement d'un système de protection piétons

[voir IPYA](#)

IPYA (I Predict Your Actions), est un projet innovant financé par l'INRIA, d'une durée de 9 mois, axé sur le développement de modèles IA pour prédire les intentions des piétons et augmenter l'autonomie des véhicules, réduisant ainsi les risques d'accidents. Ma contribution incluait l'ajustement des algorithmes de détection de piétons, et l'exploration de techniques avancées de tracking et de segmentation. J'ai également participé au développement de modèles basés sur des LSTMs et des couches de transformers pour une prédiction précise des actions futures. Le projet s'est conclu par un déploiement réussi sur Nvidia Jetson Nano pour des tests en temps réel.

Classification d'intentions piétons par intelligence artificielle pour le véhicule autonome

- État de l'art, prise en main des données JAAD (Joint Attention for Autonomous Driving) et PIE (Pedestrian Intent Estimation);
- Récupérer les keypoints des différents piétons avec un algorithme de pose estimation notamment AlphaPose;
- Classification de piétons avec des modèles de réseaux de neurones récurrents tels que Long Short Term Memory (LSTMs);
- Exploration des transformateurs dans la continuité du projet;
- Présentation de papiers de recherches (RNNs, GRU) durant les workshops;
- Dépôt de brevet sur les travaux à l'institut national de propriété industrielle (INPI) sous le numéro FR3119480A1 : A method and apparatus for controlling a vehicle based on a prediction of a pedestrian's intention to cross a roadway on which said vehicle is traveling.

Assistant de recherche en Machine Learning
Laboratoire Informatique de Paris 6 (LIP6/CNRS)

Nov 2018 - Fev 2019
Paris

Prédiction de l'affluence dans les stations de RER de l'Île de France

- Prise en main de la base de données MySQL des logs de validation des titres de transports;
- Recherche de données publiques à caractère socio-économiques sur la population de l'Île-de-France;
- Création, pré-traitement des jeux de données et application de la Backward Elimination pour la sélection de features;
- Test de différents algorithmes de régression avancée tels que eXtreme Gradient Boosting (XGBoost), Random Forest;
- Amélioration des résultats par une méthode d'Ensemble Learning, Stacking models.

BREVETS

Fouqueray, Thibault and Ahmed Tidiane Balde (Aug. 2022). "Procédé et dispositif de contrôle d'un véhicule à partir d'une prédiction de l'intention d'un piéton de traverser une chaussée sur laquelle circule ledit véhicule". FR3119480A1. URL: <https://patents.google.com/patent/FR3119480A1/fr> (visited on 08/02/2023).

LANGUES

Pourparl, Français	Bilingue
Anglais	Courant