



## Motivations

Traducteur entre l'univers technique et les enjeux métier, je conçois des solutions Data qui optimisent les budgets, anticipent les risques et automatisent les processus en rendant chaque choix technique accessible aux profils business.

## Compétences

### Compétences techniques :

**Manipulation des données** : SQL, HiveQL, BigQuery, Pyspark, Pandas, Numpy  
**Langages** : Python, Bash, SAS, R  
**Dashboard** : Qlik, Power BI, Streamlit, Shiny  
**Plateformes** : Dataiku, Docker  
**SGBD** : Teradata, Oracle, MySQL  
**Cloud** : GCP  
**API** : Flask  
**Gestion du code** : Git, GitLab, GitHub, Bitbucket  
**Big Data** : Hadoop, HDFS  
**Machine Learning** : Scikit-learn, TensorFlow, PyTorch, MLflow, DVC, Optuna, Shap

### Compétences fonctionnelles :

**Pilotage par la donnée** : Construction de hubs de données unifiés et de dashboards de pilotage donnant aux équipes métier une vision 360° en temps réel  
**Gains de productivité** : Automatisation de tâches manuelles chronophages (extraction de données, scoring, reporting) réduisant les erreurs et libérant les équipes pour des activités à forte valeur ajoutée  
**Maximisation du ROI** : Modélisation de l'impact business des investissements (marketing, publicitaires, opérationnels) et optimisation de l'allocation budgétaire  
**Réduction des risques** : Anticipation des comportements clients (attrition, fraude) pour des actions préventives ciblées

## Expériences professionnelles

### France Travail – Data Analyst (2 ans)

Projet 1 : Création d'un Hub d'offres centralisé et d'un dashboard de suivi de l'activité de diffusion des offres  
Projet 2 : Création d'un dashboard de suivi en temps réel de l'utilisation et de la consommation des modèles LLM

### Carrefour – Data Analyst (1 an)

Modélisation de l'impact des campagnes publicitaires sur le chiffre d'affaires pour optimiser l'allocation budgétaire

### Havas – Data Analyst (1 an et 5 mois)

Création d'un outil d'optimisation de l'allocation budgétaire entre médias pour maximiser l'exposition publicitaire

### Valeuriad – Data Scientist (11 mois)

Développement d'un modèle de matching pour identifier les meilleurs profils de consultants face aux appels d'offres

### Verlingue – Data Analyst (4 mois)

Développement d'un modèle d'extraction automatique des données salariales depuis les bulletins de salaire

### BPCE – Data Analyst / Scientist (2 ans et 2 mois)

Projet 1 : Création d'un modèle de prédiction et d'explicabilité pour automatiser la validation ou le rejet des opérations bancaires  
Projet 2 : Développement d'un modèle prédictif pour identifier les clients à risque de résiliation de produits assurantiels

### Servier – Data Scientist (2 ans et 7 mois)

Projet 1 : Développement d'un modèle sur IRM pour prédire la progression de l'arthrose chez les patients  
Projet 2 : Analyse de données IoT issues de bracelets connectés pour identifier des typologies de sommeil

### Dell – Formateur Data (2 ans et 5 mois)

Formations techniques sur un logiciel de statistiques et vulgarisation des méthodes d'analyse pour les clients

## Formation

**Master 2 Ingénierie Mathématique**  
Nantes

## Certifications

**Machine Learning** - Coursera, Andrew Ng  
**Machine Learning en production** - Coursera

## Langues

**Français**  
**Anglais**  
**Allemand**

**Projet 1 :**

Création d'un Hub d'offres centralisé et d'un dashboard de pilotage pour suivre l'activité de diffusion des offres

**Contexte :**

Le département Offre et Marque Employeur doit disposer d'une vision unifiée, fiable et exploitable des offres d'emploi. L'objectif est de réduire le temps passé à croiser manuellement les données, d'accélérer l'accès aux indicateurs clés et de permettre aux équipes produit de piloter l'activité de diffusion des offres au quotidien.

**Missions :**

- Construction d'un Hub d'offres par agrégation de plusieurs tables du Data Lake : offres, alertes, candidatures, déclarations d'embauche
- Calcul d'indicateurs à partir des variables du Hub : nombre d'offres diffusées, nombre d'offres nouvelles, nombre de candidatures
- Développement d'une application avec filtre interactif pour restituer ces indicateurs
- Accompagnement des profils fonctionnels sur la prise en main du Hub
- Participation à des réunions transverses intra-départements pour coordonner la création de plusieurs Hubs et les connecter entre eux

**Résultats :**

- Mise à disposition d'un Hub d'offres centralisé, utilisé comme référentiel commun par les Product Managers, éliminant les requêtes SQL complexes et réduisant le temps consacré à l'extraction de données
- Pilotage quotidien de l'activité rendu possible grâce au dashboard automatisé permettant de détecter en temps réel les anomalies de diffusion et répondre aux questions métier récurrentes (volumétrie, attractivité des offres)

**Environnement technique :**

HiveQL, Python (pandas, streamlit), Dataiku, Bash, Git/GitLab

**Projet 2 :**

Création d'un dashboard de suivi en temps réel de l'utilisation et de la consommation des modèles LLM

**Contexte :**

Avec la multiplication des cas d'usage IA générative au sein de France Travail, le département Agence Data Services doit maîtriser les coûts, anticiper les dérives de consommation et sécuriser le passage à l'échelle des modèles LLM. L'objectif est de détecter les usages anormaux et définir des garde-fous techniques pour prévenir les dépassements.

**Missions :**

- Récupération des données depuis le Data Lake et agrégation pour avoir une table qui contient les différents indicateurs à représenter : nombre de requêtes par minute, nombre de tokens ingérés, nombre de tokens générés, coûts associés
- Création d'un dashboard avec différents graphiques : suivi du nombre de requêtes dans le temps, suivi de la consommation de tokens dans le temps, distributions et boîtes à moustache du nombre de requêtes et du nombre de tokens totaux par minutes
- Mise en place de pipelines de déploiement pour mettre à jour les données du dashboard quotidiennement

**Résultats :**

- Établissement de quotas par minute basés sur l'analyse statistique des distributions de consommation évitant les surconsommations et permettant une allocation budgétaire prévisionnelle par trimestre
- Dashboard de pilotage utilisé quotidiennement par les Product Managers pour suivre l'adoption des modèles LLM

**Environnement technique :**

HiveQL, Qlik, Kubernetes (CronJob), Bash, Git/GitLab



## Carrefour

### Data Analyst

1 an

#### Projet :

Modélisation de l'impact des campagnes publicitaires sur le chiffre d'affaires pour optimiser l'allocation budgétaire

#### Contexte :

Carrefour investit plusieurs millions d'euros annuellement dans des campagnes publicitaires sans pouvoir mesurer précisément leur impact sur les ventes. L'objectif est double : quantifier l'incrémental de ventes généré par la publicité et fournir aux équipes métier un outil leur permettant de simuler et optimiser les futurs arbitrages budgétaires.

#### Missions :

- Construction d'un Datamart à partir du SGBD : agrégation des ventes par rayon et famille de produits
- Construction d'indicateurs pour prendre en compte l'effet des promotions
- Feature Engineering sur les investissements médias pour créer des indicateurs dérivés
- Modélisation linéaire explicative des indicateurs pour expliquer les ventes
- Construction d'un dashboard avec restitution des résultats de modélisation et optimisation de la répartition du budget publicitaire entre les médias

#### Résultats :

- Quantification de l'impact publicitaire à hauteur de 5% des ventes totales, soit plusieurs millions d'euros de ventes directement attribuables aux campagnes médias
- Outil de simulation adopté par les équipes marketing pour planifier l'allocation budgétaire entre canaux (TV, radio) avec des scénarios d'optimisation permettant de maximiser le ROI des futures campagnes

#### Environnement technique :

SQL, R (data.table, stats, caret, doparallel, nloptr, clustofvar, shiny, plotly), CSS, HTML, Javascript, Oracle



## Verlingue

### Data Analyst

4 mois

#### Projet :

Développement d'un modèle d'extraction automatique des données salariales depuis les bulletins de salaire

#### Contexte :

Les gestionnaires du courtier en assurance Verlingue perdent un temps précieux à ressaisir manuellement les données de bulletins de salaire lors du traitement des arrêts de travail. L'objectif est d'automatiser la saisie tout en s'intégrant aux processus existants sans rupture pour les équipes.

#### Missions :

- Echanges avec les équipes RH afin de bien comprendre la structure des bulletins de salaire
- Mise à disposition de 10 000 bulletins de salaire ainsi que des valeurs saisies à la main par les gestionnaires
- Création d'expressions régulières (regex) pour récupérer les informations souhaitées dans les documents
- Déploiement des modèles basés sur ces regex dans des API et conteneurisation des API

#### Résultats :

- Réduction drastique du temps de traitement des arrêts de travail grâce à la pré-saisie automatique des valeurs extraites libérant les gestionnaires pour des activités à plus forte valeur ajoutée comme le conseil client
- Élimination des erreurs de frappe dans la saisie des montants salariaux garantissant la fiabilité des déclarations et réduisant les risques de litiges avec les clients et organismes

#### Environnement technique :

Python (pandas, pypdf2, pdfplumber, re, flask), Docker, Git/GitLab

**Projet 1 :**

Création d'un modèle de prédiction et d'explicabilité pour automatiser la validation ou le rejet des opérations bancaires

**Contexte :**

Les conseillers de BPCE doivent traiter quotidiennement des centaines d'opérations bancaires (virements importants, dépôt de chèques) en décidant manuellement de leur validation ou rejet. L'objectif est d'aider à prioriser et fiabiliser la prise de décision tout en gardant une explicabilité forte.

**Missions :**

- Construction de la table d'apprentissage depuis le SGBD avec les données répertoriées par le métier
- Construction d'indicateurs pertinents pouvant avoir un impact sur la prédiction des opérations bancaires
- Modélisation prédictive des opérations bancaires à traiter : rejet ou validation
- Calcul de seuils pour chaque variable du modèle
- Calcul des plus grosses contributions pour chaque prédiction effectuée par le modèle
- Création d'un dashboard de suivi des alertes dans le temps

**Résultats :**

- Déploiement en production d'un système d'aide à la décision utilisé quotidiennement par les conseillers leur fournissant une recommandation instantanée (validation/rejet) pour chaque opération bancaire
- Transparence totale des décisions grâce à l'affichage des 3 facteurs les plus contributifs à la recommandation renforçant la confiance des conseillers dans le système et facilitant leur validation finale
- Monitoring continu du modèle via un dashboard avec système d'alertes automatiques permettant de détecter toute dérive de performance et garantissant la fiabilité dans le temps

**Environnement technique :**

SQL, Python (pandas, scikit-learn, xgboost, mlflow, optuna, shap, dvc), GCP, BigQuery, Power BI, Teradata, Bash, Git/Bitbucket

**Projet 2 :**

Développement d'un modèle prédictif pour identifier les clients à risque de résiliation de produits assurantiels

**Contexte :**

La résiliation de produits assurantiels représente une perte directe de chiffre d'affaires pour BPCE. Les actions de rétention sont souvent déclenchées trop tard ou de manière trop large. L'objectif est de passer d'une posture réactive à une stratégie de prévention en identifiant les signaux faibles de départ des clients sur les produits d'assurance.

**Missions :**

- Construction de la table d'apprentissage depuis le SGBD avec les données répertoriées par le métier
- Analyse des données et construction d'indicateurs pertinents pouvant avoir un impact sur l'attrition
- Feature Engineering pour créer des indicateurs dérivés
- Modélisation prédictive du risque que les clients résilient leur assurance habitation dans les 3 prochains mois

**Résultats :**

- Modèle prédictif déployé en production permettant d'identifier 3 mois à l'avance les clients à risque élevé de résiliation et donnant aux conseillers une fenêtre d'action suffisante pour intervenir
- Campagnes de rétention ciblées orchestrées sur la base des prédictions permettant aux conseillers de proposer proactivement des produits plus adaptés aux clients à risque et d'améliorer le taux de fidélisation

**Environnement technique :**

SQL, Python (pyspark, scikit-learn, xgboost, mlflow, optuna, dvc), Teradata, Hadoop/HDFS, Bash, Git/Bitbucket

**Projet :**

Création d'un outil d'optimisation de l'allocation budgétaire entre médias pour maximiser l'exposition publicitaire

**Contexte :**

Les clients de Havas investissent des budgets publicitaires importants sur plusieurs médias (TV, digital, radio, presse) sans outil pour optimiser la répartition entre ces canaux. L'objectif est de rendre l'optimisation média exploitable directement par les équipes métier via un outil simple de simulation pour maximiser l'exposition publicitaire d'une cible.

**Missions :**

- Développement d'un outil d'optimisation de l'allocation budgétaire entre les médias dans l'objectif de maximiser l'exposition médiatique à une publicité
- Création d'une interface graphique pour : sélectionner la cible publicitaire, définir les options pour chaque média, modéliser les performances en fonction du budget pour chaque média, simuler différents scénarios et sélectionner le meilleur en terme d'exposition globale
- Restitution des résultats dans différentes feuilles

**Résultats :**

- Interface de simulation utilisée par les équipes métier pour tester différents scénarios d'allocation budgétaire et sélectionner la stratégie optimale en fonction de la cible démographique et des objectifs d'exposition
- Capacité de tester instantanément plusieurs scénarios d'allocation réduisant le temps de conception des stratégies média

**Environnement technique :**

SAS, SQL, Excel VBA

**Projet :**

Développement d'un modèle de matching pour identifier les meilleurs profils de consultants face aux appels d'offres

**Contexte :**

Les commerciaux de l'ESN Valeuriad passent un temps important à rechercher manuellement les profils adaptés aux appels d'offres avec un risque de passer à côté de compétences pertinentes. L'objectif est d'exploiter la richesse des dossiers de compétences pour accélérer le staffing avant-vente et identifier, lors des remplacements de mission, un collaborateur aux compétences similaires.

**Missions :**

- Récupération des dossiers de compétences des collaborateurs et pre-processing du texte
- Utilisation de méthodes NLP d'Embedding pour transformer les mots en vecteurs de nombres
- Sélection du meilleur dossier pour une offre selon une métrique de similarité
- Clustering des profils des collaborateurs et visualisation dans un espace en 2D pour observer les profils proches
- Déploiement et mise à jour automatique du graphique et des dossiers vectorisés en fonction des arrivées et des départs au sein de Valeuriad

**Résultats :**

- Gain de temps majeur pour les commerciaux capables d'identifier le "Top 5" des experts pertinents en quelques secondes au lieu de plusieurs heures de recherche manuelle
- Cartographie visuelle 2D des compétences des +150 collaborateurs offrant aux commerciaux une vue instantanée des profils disponibles et facilitant l'identification de candidats de remplacement

**Environnement technique :**

Python (pandas, spacy, sentence-transformers, scikit-learn, flask), Docker, Git/GitLab

### Projet 1 :

Développement d'un modèle sur IRM pour prédire la progression de l'arthrose chez les patients

#### Contexte :

Le laboratoire Servier développe un médicament contre l'arthrose dont le coût s'élève à plusieurs millions d'euros et doit sélectionner les patients pour la phase 3 des essais cliniques. Inclure des patients dont l'arthrose ne progresse pas suffisamment réduit les chances de prouver l'efficacité du médicament. L'objectif est d'améliorer la sélection des patients en incluant ceux présentant un potentiel de progression de la pathologie.

#### Missions :

- Récupération d'un jeu de 10 000 IRM de genoux avec les stades d'arthrose associés ainsi que des données démographiques et cliniques liées aux patients
- Modélisation avec des méthodes de Deep Learning du niveau d'arthrose à 1 an en fonction des IRM à baseline : le but est de prédire s'il y a une progression ou pas dans le temps
- Modélisation avec des méthodes de Machine Learning du niveau d'arthrose à 1 an en fonction des données démographiques et cliniques

#### Résultats :

- Aide à la sélection de profils patients plus pertinents pour les phases avancées d'essais cliniques du médicament augmentant significativement les chances de succès de la phase 3 et optimisant l'investissement R&D
- Valorisation des données IRM comme levier décisionnel dans la stratégie d'essais

#### Environnement technique :

Python (pandas, scikit-learn, tensorflow, pytorch, keras, mlflow), Git/GitLab

### Projet 2 :

Analyse de données IoT issues de bracelets connectés pour identifier des typologies de sommeil

#### Contexte :

Dans le cadre d'un programme d'innovation en santé connectée, Servier souhaite explorer le potentiel des objets connectés pour la collecte de données patients en conditions réelles. L'objectif de cette initiative pilote de 4 mois est de transformer des signaux bruts issus de bracelets connectés en profils de sommeil interprétables et enrichir les études cliniques futures.

#### Missions :

- Equipement de 1500 collaborateurs avec des bracelets connectés capables d'identifier les différentes phases de sommeil (profond et léger) et la phase d'éveil
- Récupération des données et clustering des nuits en utilisant des méthodes d'analyse de séquences
- Caractérisation des groupes obtenus avec des variables comportementales issues d'un questionnaire soumis aux collaborateurs du projet

#### Résultats :

- Déploiement réussi de bracelets connectés auprès de 1 500 collaborateurs volontaires démontrant la capacité à orchestrer un projet IoT à grande échelle avec collecte automatisée de données physiologiques
- Identification d'insights actionnables comme la corrélation entre horaires de lever précoces et temps de transport élevés validant l'intérêt de l'IoT pour comprendre les comportements en vie réelle
- Constitution d'une expertise technique et méthodologique en analyse de données IoT positionnant l'équipe R&D Data Science pour intégrer ces technologies dans de futurs protocoles cliniques digitaux

#### Environnement technique :

R (data.table, traminer), Git/GitLab