Florent Lefort

Data Analyst

Expérience

10

ans

Motivations

Passioné par la Data, j'implémente des solutions techniques pour répondre à des besoins métier et j'explique simplement ces choix aux profils non techniques.

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 :

Analyse des données : Collecte, transformation, création d'indicateurs
Restitution des résultats : Création de dashboards et démonstrateurs
Communication : Vulgarisation, création et animation de conférences et formations
Référent technique : Garant des bonnes pratiques, rédaction des tests unitaires et d'intégration
Assistance : Accompagnement et aide à la montée en compétences
Machine learning : Apprentissage supervisé
Approche : Agile, Jira, Scrum

Expériences professionnelles

France Travail **– Data Analyst (2 ans)**
Projet 1 : Construction d'un Hub d'offres et restitution d'indicateurs dans un dashboard
Projet 2 : Création d'un dashboard pour suivre la consommation des modèles LLM

Carrefour **– Data Analyst (1 an)**
Elaboration d'un modèle explicatif des campagnes publicitaires et promotions sur les chiffres de vente et restitution des résultats dans un dashboard

Havas **– Data Analyst (1 an et 5 mois)**
Développement d'un outil d'aide à la décision pour optimiser le budget publicitaire entre les médias

Valeuriad **– Data Scientist (11 mois)**
Elaboration d'un modèle pour identifier les meilleurs profils à un appel d'offre

Verlingue **– Data Analyst (4 mois)**
Elaboration et déploiement d'un modèle pour extraire de la valeur dans des bulletins de salaire

BPCE **– Data Analyst / Scientist (2 ans et 2 mois)**
Projet 1 : élaboration d'un modèle pour prédire les rejets/validations des opérations bancaires
Projet 2 : élaboration d'un modèle pour prédire le risque d'attrition des clients

Servier **– Data Scientist (2 ans et 7 mois)**
Projet 1 : élaboration d'un modèle pour prédire la progression de l'arthrose
Projet 2 : réalisation de clustering de nuits de sommeil à partir de l'IoT

Dell **– Formateur Data (2 ans et 5 mois)**
Formation sur l'utilisation d'un logiciel de statistiques et vulgarisation de ces méthodes pour des clients

Formation

Master 2 Ingénierie Mathématique
Nantes

Certifications

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

Langues

Français
Anglais

France Travail

Data Analyst

2 ans

Projet 1 :

Créer un Hub d'offres et un dashboard de restitution d'indicateurs

Contexte :

Le département Offre et Marque Employeur de la DSI de France Travail a besoin d'exploiter et de calculer des indicateurs sur les offres plus facilement. Le département a aussi besoin de suivre certains indicateurs comme le nombre d'offres diffusées et le nombre d'offres nouvelles chaque jour.

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 du Hub aux PO/PM pour leur faciliter le requêtage des offres
- Mise à disposition du dashboard aux PO/PM pour répondre à des questions fonctionnelles

Environnement technique :

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

Projet 2 :

Créer un dashboard pour suivre l'utilisation des modèles LLM

Contexte :

Le département Agence Data Services de la DSI de France Travail développe et met à disposition des autres départements des modèles de Machine Learning, de Deep Learning et d'IA Générative comme les LLM. Afin de suivre l'utilisation des LLM à travers les différents cas d'usage, le département a besoin d'avoir un dashboard de suivi.

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 :

- Utilisation du dashboard par les Product Managers (PM) et utilisation des distributions pour définir des quotas par minute pour chaque modèle LLM afin de contrôler leur consommation

Environnement technique :

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

Carrefour

Data Analyst

1 an

Projet :
Élaborer un modèle explicatif des campagnes publicitaires et des promotions sur les ventes

Contexte :
Le département Data de Carrefour veut connaître l'impact des campagnes publicitaires (radio, TV, online) et des promotions sur les chiffres de ventes afin de quantifier leurs contributions. Le département veut aussi répartir le budget des futures campagnes entre les différents médias afin de maximiser les ventes.

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 :
- Contribution des campagnes publicitaires à hauteur de 5% des ventes totales
- Utilisation du dashboard par l'équipe métier pour simuler les budgets des prochaines campagnes à répartir entre les différents médias

Environnement technique :
SQL, R (stats, caret, nloptr, clustofvar, shiny, plotly), CSS, HTML, Javascript, Oracle

Verlingue

Data Analyst

4 mois

Projet :
Élaborer un modèle pour extraire de la valeur dans des bulletins de salaire

Contexte :
Le département Digital Factory de Verlingue souhaite extraire le salaire net, le salaire brut et les primes dans les bulletins de salaire déposés par ses clients pour faire gagner du temps au gestionnaire lors de la saisie d'un arrêt de travail.

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 :
- Gain de temps et réduction du risque d'erreur lors de la ressaisie manuelle en proposant directement les valeurs extraites depuis le bulletin de salaire au gestionnaire lors de la saisie de l'arrêt de travail

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

BPCE

Data Analyst / Data Scientist

2 ans et 2 mois

Projet 1 :
Élaborer un modèle pour prédire les rejets/validations des opérations bancaires

Contexte :
Le département Data Science - Usages Business Avancés de BPCE travaille sur différents cas d'usages notamment la prédiction des opérations bancaires afin d'aider les conseillers en agences à les traiter plus rapidement.

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 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 et aide à la décision quotidienne pour les conseillers avec une suggestion d'action pour chaque opération bancaire : rejet ou validation
- Restitution sur l'écran du conseiller des 3 variables les plus contributives aux prédictions du modèle de chaque opération bancaire
- Suivi du modèle dans le temps avec alertes dans un dashboard si les seuils sont dépassés

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

Projet 2 :
Élaborer un modèle prédictif pour détecter les clients sur le point de résilier un produit de BPCE

Contexte :
Le département Data Science - Usages Business Avancés de BPCE travaille sur différents cas d'usages notamment la détection du risque d'attrition, ie la probabilité qu'a un client de résilier un produit du groupe (produits bancaires et assurantiels).

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 :

- Utilisation du modèle lors de campagnes de prises de contact par les conseillers bancaires avec les clients "à risque" afin de leur proposer un produit plus adapté

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

Havas

Data Analyst

1 an et 5 mois

Projet :
Développer un outil d'aide à la décision média

Contexte :
Le département CSA Data Consulting de Havas veut répartir le budget publicitaire entre les différents médias pour maximiser l'exposition d'une cible spécifique à la publicité.

Missions :
- Développement d'un outil d'optimisation de l'allocation budgétaire entre 2 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 les 2 médias
    - 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 :
- Utilisation de l'outil par l'équipe métier pour déterminer la meilleure répartition du budget entre 2 médias de façon à toucher le plus de personnes d'une cible donnée

Environnement technique :
SAS, SQL, Excel VBA

Valeuriad

Data Scientist

11 mois

Projet :
Élaborer un modèle pour identifier les meilleurs profils correspondant aux appels d'offres

Contexte :
L'ESN Valeuriad souhaite aider les commerciaux à mieux répondre aux appels d'offres en utilisant les mots contenus dans les offres et dans les dossiers de compétences des collaborateurs. Par ailleurs, le second objectif est de trouver des liens de similarités entre les profils des collaborateurs pour effectuer de meilleur remplacement de mission.

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 pour les commerciaux pour trouver le top 5 des profils qui correspondent le mieux à un appel d'offre
- Meilleure compréhension des profils des collaborateurs et de leurs compétences par les commerciaux

Environnement technique :
Python (pandas, sentence-transformers, scikit-learn, flask), Docker, Git/GitLab

Servier

Data Scientist

2 ans et 7 mois

Projet 1 :

Élaborer un modèle pour prédire la progression de l'arthrose sur des IRM de genoux

Contexte :

Le département R&D Data Science de Servier souhaite améliorer la sélection des patients à inclure dans la phase 3 de l'étude clinique d'un médicament contre l'arthrose.

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 :

- Meilleure sélection des patients à inclure dans les essais cliniques du médicament à tester

Environnement technique :

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

Projet 2 :

Réaliser le clustering des nuits de sommeil de collaborateurs de Servier

Contexte :

Le département R&D Data Science de Servier souhaite analyser les séquences de nuits de sommeil de collaborateurs de Servier volontaires dans le cadre d'un projet IoT. Ce projet de 4 mois a pour but de tester l'utilisation d'objets connectés afin de récolter des données à analyser.

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 les réponses issues d'un questionnaire soumis aux collaborateurs du projet

Résultats :

- Meilleure compréhension de l'IoT
- Découverte de résultats logiques parmi certains groupes (exemple : le groupe des collaborateurs levés tôt est caractérisé par un temps de transport en moyenne plus élevé que dans les autres groupes)

Environnement technique :

R (traminer), Git/GitLab