



Motivations

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

Compétences

Compétences techniques :

Langages : Python, Bash, R, SAS, Dataiku
Manipulation des données : SQL, HiveQL, Pandas, Numpy, Pyspark, BigQuery
Visualisation des données : Matplotlib, Plotly, Seaborn
Dashboard : Qlik, Power BI, Streamlit, Shiny
API : Flask
SGBD : Teradata, Oracle, MySQL
Cloud : GCP
Gestion du code : Git, Gitlab, Github, Bitbucket
Machine learning : Scikit-learn, Tensorflow, Pytorch, Keras, Mlflow, Optuna, Shap
Gestion des données : Hadoop, HDFS, DVC
Conteneurisation : Docker

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 (1 an et 3 mois)

Construction d'un Hub d'offres et restitution d'indicateurs comme le nombre quotidien d'offres diffusées dans un dashboard

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

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

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

1 an et 3 mois

Projet :

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

Contexte :

Le département Offre 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 Datalake : 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 Hub 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 (streamlit), Bash, Git/Gitlab, Dataiku



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 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 une équipe 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

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 une équipe 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

Projet :

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

Contexte :

Le département Data de Verlingue souhaite extraire le salaire net, le salaire brut et les primes dans les bulletins de salaire déposés par les clients de Verlingue 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



Projet 1 :

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

Contexte :

Le département Data Science - Usages Avancés 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
- Participation à la 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, sklearn, xgboost, mlflow, optuna, shap), GCP, BigQuery, Power BI, Git/Bitbucket, Bash, Teradata

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 Avancés travaille sur différents cas d'usages notamment la détection du risque d'attrition, ie la probabilité qu'a un client de résilier un produit de BPCE (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, sklearn, xgboost, mlflow, optuna), Git/Bitbucket, Bash, Teradata, Hadoop/HDFS