

Azaria TANDJA

Paris | azaria.tandja@yahoo.com | +33674135272

Data Scientist | Spécialisé en Analyse du Risque

Data scientist spécialisé en risques financiers, j'interviens sur la modélisation des risques (crédit, ALM), la détection de fraude et l'optimisation des décisions financières à partir de données complexes. À l'aise avec les environnements techniques (Python, SAS, SQL, R), les techniques de machine learning et les cadres réglementaires (Bâle, IFRS 9), je cherche à intégrer une équipe où la donnée sert directement la stratégie de gestion des risques.

Education

Université d'Orléans, Master Econométrie et Statistique Appliquée

2022 – 2024

Master 1 | Master 2

Experience

Data Scientist risque, Crédit Agricole

Stage de 6 mois de Avril à Septembre 2024 | Gestion Financière CARCENTRE

Mission: Refonte des hypothèses de remboursements anticipés (RA) des Dépôts à Terme (DAT) utilisées dans le pilotage ALM, la couverture du risque de taux de 4 Caisses Régionales (Centre Loire, Centre Ouest, Touraine et Poitou, Val de France) représentant plusieurs milliards d'euros d'encours.

- Construction d'une base de données consolidée à partir de +250 fichiers bruts issus des 4 caisses régionales, couvrant 64 mois d'historique sur les remboursements anticipés des DAT.
- Détection et correction de +120 anomalies critiques (doublons, valeurs aberrantes, incohérences intersources), garantissant la cohérence des séries utilisées en modélisation.
- Automatisation de l'extraction des informations via l'outil Monarch, réduisant le temps de préparation des données de plus de 50 %.
- Estimation de modèles via un code Python automatisée (sélection du modèle par minimisation de l'AIC).
- Réduction de l'erreur moyenne absolue (MAE) de -62%, avec des gains majeurs sur les segments les plus volatils.
- Production des matrices trimestrielles de taux RA désormais utilisées par l'équipe RTIG dans les simulations ALM et la stratégie de couverture (swaps).

Environnement technique: SQL, Python (pandas, statsmodels, scikit Learn), SAS 9.4, Excel, Monarch.

Professeur de mathématiques

De 2022 à 2025 | Complétude

- Cours personnalisés à domicile de mathématique à des lycéens et collégiens.

Analyste de données

De 2020 à 2022 | IREEP (Institut de Recherche Empirique en Économie Politique)

- Participation à la collecte et à l'exploitation de 15000+ enquêtes terrain sur les impacts économiques et sociaux du COVID-19 dans 3 pays (Bénin, Burkina Faso, Afrique du Sud).
- Suivi d'indicateurs tels que la perte de revenu, le chômage, les aides publiques reçues.

Environnement technique: STATA, EvIEWS, KoboCollect, R, Excel.

Connaissances réglementaires

- Normes de Bâle : bonne compréhension des principes de gestion du risque de crédit (PD, LGD, CCF, EAD).
- Normes IFRS 9 : notions sur la comptabilisation des pertes de crédit attendues (ECL).
- ALM (Asset and Liability Management) : familiarité avec la gestion du risque de taux et de liquidité.

Projets appliqués en science des données

Évaluation du risque de crédit | partenariat Mobilize Financial Services

Objectif: Construire une grille de score basée sur un modèle prédictif de la probabilité de défaut (PD) à 12 mois sur un portefeuille de crédits automobiles.

- Traitement d'un dataset réel de 460 000 contrats de financement automobiles, avec plus de 20 variables explicatives (durée du crédit, apport, âge client, score d'octroi, etc.).
- Nettoyage et préparation des données : identification et traitement de valeurs aberrantes, gestion des valeurs manquantes, transformation de variables catégorielles.

- Réduction de dimension : suppression des variables fortement corrélées, choix des variables via critères statistiques.
- Comparaison de plusieurs modèles supervisés: régression logistique, arbre de décision, Random Forest, XGBoost.
- Évaluation via AUC, indice de Gini, courbe lift (10/X): la régression logistique finale atteint un AUC de 0.77 et une Gini de 0.54, tout en conservant une bonne interprétabilité.
- Construction d'une grille de score client, permettant de définir des seuils d'acceptation différenciés selon les profils, avec impact direct sur le taux d'acceptation et le risque embarqué.

Apport métier: aide à la décision pour les gestionnaires de crédit, contribution à l'amélioration des stratégies d'octroi dans le respect du cadre réglementaire (Bâle III, IFRS 9).

Modélisation de la LGD | Partenariat LCL Banque et Assurance

Objectif: Estimer la perte en cas de défaut (LGD) sur des crédits sous caution, à partir d'une base réelle de clients LCL, afin d'améliorer la quantification du risque de crédit.

- Nettoyage avancé d'une base de 11 779 contrats contenant 30 variables (caractéristiques clients, prêts, performances observées).
- Refonte des variables (valeurs hors bornes, âges impossibles...).
- Modélisation supervisée de la LGD via 4 modèles comparés : arbre de régression, Random Forest, Bagging, Boosting (500 arbres chacun, validation croisée, sélection des paramètres par minimisation de la MSE).
- Réalisation de tests statistiques sur l'impact des étiquettes DPE/GES sur la LGD.

Apport métier: segmentation efficace de la LGD sur des prêts sous garantie, applicable dans le cadre d'IFRS 9 ou des modèles IRB avancés.

Détection de fraude bancaire sur données déséquilibrées

Objectif: Identifier les transactions potentiellement frauduleuses dans un jeu de données hautement déséquilibré, à l'aide de techniques de machine learning supervisées et non supervisées.

- Analyse d'un jeu de données contenant plus de 280 000 transactions, avec un taux de fraude réel < 0,2% ; traitement des déséquilibres par undersampling.
- Construction de plusieurs modèles supervisés et non supervisés :régression logistique, Random Forest, XGBoost, Isolation Forest, clustering k-means.
- Évaluation multi-métriques : précision, rappel, F1-score, ROC-AUC, avec focalisation sur le rappel des fraudes (sensibilité).
- **+40 % d'amélioration** du rappel par rapport au modèle naïf, à taux de faux positifs constant.

Environnement technique de travail

- Langages et outils : Python, SAS, SQL, R, Excel
- Visualisation et reporting : Power BI, R Markdown.

Certifications

- [SAS Certified Specialist : Base Programming Using SAS 9.4](#)
- [TOEIC® Listening and Reading test - Niveau B2](#)
- [Dataiku core designer](#)
- [Dataiku ML Practitioner](#)
- [World Bank Group : Impact Evaluation Surveys course](#)

Soft Skills

Curieux, adaptable, secret professionnel, bon esprit d'équipe.

Langues

- Français : Langue maternelle
- Anglais : Courant

Centres d'intérêts

Musique, Football loisir, Lecture.