



OFFRE DE STAGE

Analyse et interprétation des prédictions issues de modèles complexes : Application à la détection de la fraude aux moyens de paiement

LOCALISATION : 50 avenue Pierre Mendès France - 75013 PARIS

DATES : début (souhaité) en Mars-Avril 2018 – 6 mois

Description de l'entreprise

Le Groupe BPCE, deuxième groupe bancaire français issu de la fusion des Caisses d'Épargne et des Banques Populaires, est un groupe diversifié regroupant, outre des activités de banque de détail, des filiales spécialisées, la banque de grande clientèle Natixis, des activités de gestion d'actifs, de banque privée, d'assurance... en France et à l'international.

La Direction des Risques, de la Conformité et du Contrôle Permanent (DRCCP) du Groupe BPCE a pour mission de mettre en œuvre tous les moyens permettant de piloter le groupe en matière de liquidité, de solvabilité, de maîtrise des risques et de contrôle interne.

Au sein de cette direction, le pôle « Modèles Internes & Big Data » est composé de 4 sous pôles :

- Modèles Retail : en charge de développer, modéliser et faire évoluer les modèles internes de crédit (PD, LGD, CCF, ELBE) pour la clientèle des Particuliers et des entreprises de moins de 3 M€ de CA (i.e. les Professionnels), des Associations de petite taille, etc.
- Modèles Hors Retail : en charge de développer, modéliser et faire évoluer les modèles internes (PD, LGD, CCF, ELBE) pour la clientèle Corporate (PME/ETI), Secteur Public, Assurances, Associations (de taille importante).
- Modèles de Portefeuille : en charge de développer les méthodologies de mesure des risques sur base portefeuilles. Les travaux principaux sont les différents Stress tests crédit, les méthodologies de provisionnement statistiques (IAS39, IFRS9), celles autour des mesures internes/pilier 2 sur le crédit et plus globalement les méthodologies projection des mesures de risques de crédit.
- Innovation & Support : en charge d'une part de développer des méthodologies innovantes au sens « Big Data » pour les projets Data identifiés à la DRCCP, en lien avec la direction Digital ; puis d'autre part de développer des méthodologies « transverses » au sens où elles peuvent s'appliquer autant au provisionnement statistique (IFRS9), qu'au périmètre Retail ou Hors Retail, mais également aux différents Stress Tests. Le terme « Support » doit être compris au sens « support méthodologique » qui nécessite une réflexion avancée sur les problématiques à modéliser ainsi qu'une capacité à coordonner et travailler les différents sous pôles.

Poste et mission

Le pôle Innovation et Support recrute un **Data Scientist stagiaire pour travailler sur les sujet Big Data**.

Le pôle Big Data DRCCP pour missions d'identifier et de réaliser les projets Big Data relatifs à la gestion des risques, de la conformité et des contrôles permanents. Son

champ d'action est l'ensemble des métiers de la DRCCP (Risque de crédit, risque opérationnel, risque de marché, risque ALM, conformité, contrôles permanents, sécurité des SI, sécurité des biens et personnes, etc..).

Dans le cadre de ses missions, le pôle Big Data applique des algorithmes de Machine Learning, ce qui aboutit du fait du compromis performance/complexité, à des modèles difficilement appréhendables par des experts métiers. Il est de ce fait nécessaire de pouvoir décrire et expliquer, dans la mesure du possible, le fonctionnement des modèles développés.

Le stagiaire aura pour mission :

- de réaliser une revue de la littérature portant sur l'interprétation des modèles, comprenant notamment les algorithmes permettant d'expliquer ou de synthétiser le lien entre variables explicatives et prédictions (les algorithmes LIME, LOCO par exemple) ;
- de représenter les modèles de manière graphique en utilisant les techniques de *data visualisation* adaptés à ces méthodes ;
- de tester différentes implémentations en python ;
- si besoin, d'implémenter une approche qui soit couverte par la littérature mais non implémentée dans un package existant ;
- de tester les outils sur différents modèles de tests ;
- d'appliquer les techniques sur un modèle boîte noire de détection de fraude élaboré par l'équipe ;
- de présenter les interprétations du modèle aux experts métiers de la lutte contre la fraude.

Apports pédagogiques du stage

Ce stage permettra au candidat d'acquérir des connaissances à la fois techniques (Machine Learning, R, Python, Data visualisation) et fonctionnelles ainsi qu'une bonne vision de la culture bancaire et de la gestion des entreprises.

Contenu statistique/Big Data : algorithmes LIME (Local Interpretable Model-Agnostic Explanations), LOCO (Leave One Covariate Out), Interprétabilité des modèles, Black-box learning, Data visualisation, Détection de fraude, Machine Learning, Data Science.

Profil et compétences requises

Vous êtes étudiant de niveau Bac +5 d'une formation scientifique de type école d'ingénieur (ENSAI, ENSAE, ISUP, X, Mines, Centrale...) ou Ecole Normale Supérieure, avec une spécialisation en statistiques, Data Science ou Big Data.

Vous disposez d'un très bon relationnel et d'un bon niveau de synthèse rédactionnelle.

Les problématiques bancaires vous intéressent. Vous faites preuve de curiosité et d'ouverture, d'une grande rigueur et d'un sens critique développé, et disposez de fortes capacités analytiques.

Informations pratiques

Vous souhaitez postuler ? Envoyez par mail votre candidature (CV + Lettre de motivation) à Ronan GAUTIER (ronan.gautier@bpce.fr) ou à Grégoire Jaffre (gregoire.jaffre@bpce.fr).