

Stage M2 : AutoML et Interprétabilité des modèles

Date d'entrée souhaitée : 06/04/2020

Flexibilité : 2 mois

Durée du stage : 6 mois

Nombre de poste à pourvoir : 1

Entité : Crédit Agricole SA

Direction d'accueil (*pas de sigle*) : Marketing et Communication

Pôle (*sigle*) : DCI

Maître de stage : Walid Erray : walid.erray@credit-agricole-sa.fr

Lieu de travail : Montrouge

Présentation du service :

Au sein du Pôle Développement Clients et Innovation, le DataLab Groupe Crédit Agricole est un centre de compétences dédié aux sciences de la donnée et à leurs applications dans le domaine bancaire. Son rôle est de créer des approches innovantes pour la valorisation de la donnée interne et externe, qu'elle soit structurée ou non structurée. Dans le cadre de ses missions, des thématiques scientifiques à forte valeur ajoutée sont étudiées : Apprentissage Automatique, Auto-ML, Traitement du Langage Naturel, Process Mining, Time Series Mining, Deep Learning, Géomatique, etc. Ces activités sont menées conjointement avec des partenaires internes : les Caisses Régionales, les Entités du Groupe et les Producteurs Informatiques. Le DataLab développe également un réseau de partenaires externes lors de missions industrielles (Editeurs de logiciels, startup, SSII, etc.) ou de collaborations universitaires.

Dans le cadre de ce stage, vous rejoindrez l'équipe Data Science afin de contribuer à la valorisation de la Data et participer à la mise en place d'approches innovantes au service des Clients du Groupe.

Descriptif de la mission :

Contexte et objectifs du stage :

Ce stage de Data Science a pour objectif de réaliser un état de l'art, de concevoir et développer des approches d'interprétabilité de modèles de machine learning complexes et plus particulièrement des approches de recommandation.

Les applications du machine learning dans des secteurs réglementés (banque, assurance, etc.) restent souvent cantonnées aux modèles linéaires ou à arbres compte tenu de leur capacité à produire des décisions relativement faciles à expliquer à un analyste métier. Récemment, des nouvelles techniques permettant d'expliquer les résultats produits par des modèles black-box ont vu le jour et attirent de plus en plus l'intérêt de la communauté data science. En effet, il est désormais possible d'utiliser des modèles très performants comme xGBoost, Deep learning, etc sans sacrifier l'explication du modèle.

Pendant votre stage, vous serez amené(e) à étudier des techniques comme LIME, Influence Functions, Permutation Importance (et autres) afin de les améliorer ou d'en créer d'autres dans le but de ressortir des interprétations compréhensibles par un analyste métier. L'interprétabilité sera appliqué à une chaîne de traitement AutoML (solution Groupe développée au DataLab). Au-delà de la performance, L'interprétabilité des modèles permettra également d'orienter le métier vers des leviers concrets en terme d'accompagnement du client.

Organisation et livrables :

Au cours de son stage, le stagiaire aura pour mission de :

- Etablir un état de l'art scientifique sur les approches d'interprétabilité de modèles (articles, bibliothèques open source)
- Comprendre et documenter les critères et aspects de l'interprétabilité (focus sur les méthodes Deep Learning)
- Etudier les différentes approches (modèle spécifique, agnostique, graphique)
- Améliorer des frameworks existants (LIME, ELI5, etc) -> programmer en Python
- Appliquer les développements proposés dans le cadre d'un projet en cours de Moteur de recommandation
- Réaliser une étude comparative entre les approches de l'état de l'art
- Rédiger un rapport pour décrire la méthodologie proposée et les résultats obtenus

PROFIL RECHERCHE

Niveau d'études préparé : Bac + 5

Formation : Université Ecole d'ingénieur

Ecoles ou formations dans lesquelles vous souhaitez éventuellement diffuser l'annonce :

X, ENS, Supélec, Ecole des Mines, Telecom Paris, Ecole Centrale, ENSAE, ENSAI, INSA, UPMC, Université Paris Sud, Université Paris Dauphine, Université de Lyon

Spécialisation :

Data Science, Machine Learning, Intelligence Artificielle

Compétences techniques ou spécifiques au poste (5 maximum) :

- Automated Machine Learning
- Selection de modèle / hyper-paramètre
- Deep Learning
- Interprétabilité des modèles
- Développement Python
- Algorithmique

Compétences générales et transverses (5 maximum) :

- Autonomie et communication
- Rigueur
- Innovation/R&D
- Force de proposition
- Travail en équipe

Outils informatiques :

- Langage de développement : Python.

Langues : Français

