

Date d'entrée souhaitée : 04/03/2019

Flexibilité :  1 mois  2 mois  > 3 mois

Durée du stage : 6 mois

## INFORMATIONS GENERALES

Entité : Crédit Agricole SA

Maître de stage : Walid Erray : [walid.erray@credit-agricole-sa.fr](mailto:walid.erray@credit-agricole-sa.fr)

Type de métier : *Marketing et Communication*

Lieu de travail : Montrouge

## DESCRIPTION DE LA MISSION

Intitulé du poste : Stage Data Scientist

### 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 concevoir et développer de nouvelles approches de fouille de séries temporelles et de les appliquer dans le cadre de projets de détection d'événements, de détection d'anomalies, d'IOT, d'Open Data, etc.

Les données de types séries temporelles constituent un élément majeur du patrimoine des entreprises. Elles apportent une information sur la dynamique des comportements sur des sujets tels que : la relation client, la fraude, le risque, la finance, etc. L'extraction des connaissances à partir de ces données nécessite la mise en place d'approches spécifiques issues du domaine du Time Series Mining.

L'objectif du stage consiste à enrichir une librairie développée en interne, TSBOX, qui reprend l'état de l'art. Un focus particulier sera porté sur la conception d'approches de fouille de séries temporelles supervisées et l'optimisation de la complexité pour l'analyse de grandes volumétries de données.

#### Organisation et livrables :

Au cours de son stage, le Data Scientist aura pour mission de :

- Etablir un état de l'art scientifique sur les approches Time Series Mining (articles, librairies open source)

- Prendre en main la librairie existante TSBOX
- Proposer une structure de données adaptée pour gérer des données structurées et non-structurées volumineuses
- Concevoir et développer : une méthode de représentation symbolique supervisée, une approche d'extraction d'évènements et de détection d'anomalies, une méthode de clustering de séries à tailles variables
- Appliquer les développements proposés dans le cadre des projets en cours : détection de changement comportement utilisateurs ; détection de fraude aux moyens de paiement ; Analyse de données IOT...
- Réaliser une étude comparative entre l'approche proposée et 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 + 2     Bac + 3     Bac +4     Bac +5

Formation :

- Université     Ecole de commerce     Ecole d'ingénieur     Autre précisez :

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

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

Spécialisation :

Data Science, Machine Learning, Intelligence Artificielle

Compétences techniques ou spécifiques au poste (5 maximum) :

- Machine Learning
- Time series mining
- Détection d'anomalie
- Représentation symbolique
- Deep Learning
- 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.
- Des connaissances en Cython, Numba (voire en C/C++) sont un plus.N/A

Langues 1 : Français

(Si aucune langue souhaitée indiquer N/A)

Niveau requis : Opérationnel (usage régulier)

Langues 2 : N/A

Niveau requis : Choisissez un élément.

(Si aucune langue souhaitée indiquer N/A)

## ASPECTS BUDGETAIRES

### **PARTIE A COMPLETER OBLIGATOIREMENT**

**CGB de facturation :**

**Unité d'affectation opérationnelle (UO) :**

**Votre correspondant budgétaire**

**Avez-vous reçu l'accord de votre correspondant budgétaire ?**  OUI  NON

Signature et nom du Manager :

Signature et nom du correspondant budgétaire :