

STAGE 6M : DATA SCIENTIST JUNIOR



Référence de l'offre : dxehn7wwph
Intitulé : STAGE 6M : DATA SCIENTIST JUNIOR
Date de publication : 23/10/2026
Filières : Numérique
Type de contrat : Stage
Catégorie de contrat : Non cadre
Localisation de l'offre : Paris (75013)
Éligibilité télétravail : Non
Régime de travail : Badgeage
Prime variable sur objectif : Aucune
Encadrant : Non

DESCRIPTION DE LA CAISSE DES DÉPÔTS

Avec ses filiales et ses partenaires stratégiques, la Caisse des Dépôts constitue le premier pôle financier public. Investisseur de long terme au service de l'intérêt général et du développement économique du pays, la Caisse des Dépôts se charge de plusieurs missions, dont la protection de l'épargne populaire, le financement et soutien du logement social, la gestion des mandats publics, l'accompagnement des Français dans leurs parcours de vie et la contribution au développement économique.

La Caisse des Dépôts Etablissement Public s'organise en quatre métiers : la Banque des Territoires, la Gestion d'actifs, la Gestion des participations stratégiques, et les Politiques sociales.

La Banque des Territoires (BDT) rassemble dans une même structure les expertises internes à destination des territoires. Porte d'entrée client unique, elle propose des solutions sur mesure de conseil et de financement en prêts et en investissement pour répondre aux besoins des collectivités locales, des organismes de logement sociale, entreprises publiques locales et des professions juridiques.

La Direction des prêts conçoit et apporte des financements de long terme, financements qui sont adaptés aux besoins d'investissement des clients historiques, comme le secteur du logement sociale (organismes de logement social, secteur associatif et médico-social), ou en lien avec sa mission d'intérêt général pour les acteurs du secteur public local : collectivités locales et leurs satellites, sociétés d'économie mixte, entreprises privées partenaires et co-investisseurs.

DESCRIPTION DU POSTE

La **Direction des Prêts** de la Banque des Territoires s'attache plus particulièrement à financer les projets d'intérêt général qui sont liés au développement du logement social et du secteur public local.

Au sein de la direction des prêts, le service Data Etudes et Pilotage regroupe plusieurs axes complémentaires autour de la donnée du métier prêteur.

- Le pôle Etudes réalise des publications sur les secteurs que finance la Direction des Prêts de la Banque des Territoires (logement social et secteur public local), à destination des clients (collectivités locales, bailleurs sociaux...) et des partenaires (ministères, organismes publics, etc.) afin d'éclairer les décisions en matière de politique publique du logement et de l'aménagement du territoire ;
- Le pôle Analyses Avancées réalise des analyses sur les prêts et met à disposition des collaborateurs différentes productions expertes (Tableaux de Bords et data visualisations, extraction d'information pour le reporting taxonomie) ;
- Le pôle Data Science et Intelligence Artificielle produit des outils et services à destination des clients et des collaborateurs (ChatBot interne spécialiste du domaine prêteur, textmining de documents pour le pré-remplissage de la demande de prêt et l'aide au contrôle de premier niveau des pièces justificatives, notamment).

Contexte :

La Banque des Territoires accorde chaque année près de 12 milliards d'euros de prêts aux bailleurs sociaux, pour la construction d'environ 90 000 logements sociaux et la rénovation de 90 000 autres. Chaque demande de prêt s'accompagne de nombreuses pièces justificatives, qui doivent être contrôlées et validées par les instructeurs. Pour les aider à gagner en temps et en fiabilité, nous

développons des outils d'aide automatique au contrôle de premier niveau, construits sur plusieurs briques de traitement automatique du langage (NLP) :

- Extraction du texte de PDF scannés : modèle OCR entraîné au sein de l'équipe et modèles vision-langage (VLM) ;
- Extraction d'informations propres à chaque type de document : modèles de reconnaissance d'entités nommées (NER) entraînés sur des tâches spécifiques, et LLM ;
- Interprétation de clauses juridiques et rapprochement avec les formulations attendues : algorithmes de similarité sémantique et LLM.

Vous interviendrez sur deux missions principales :

1. Classification automatique de documents. Vous concevrez et développerez un modèle de classification permettant d'identifier le type de chaque pièce justificative reçue, brique d'entrée indispensable à l'orientation des documents dans les bons pipelines de traitement. De la constitution du jeu de données à l'évaluation des performances, vous serez responsable de la démarche de bout en bout.

2. Optimisation et fiabilisation des pipelines de traitement. Il s'agit d'un travail à dominante recherche et développement : explorer des alternatives plus légères et moins coûteuses aux grands modèles que nous utilisons aujourd'hui (spécialisation de modèles de plus petite taille sur nos tâches, approches de compression et d'allègement d'architectures récentes), mesurer les compromis performance / coût / latence, et proposer des pistes d'industrialisation. Vous contribuerez également à la fiabilisation des sorties de modèles (contrôle de structure, de types et de valeurs).

Selon votre appétence et l'avancement de vos travaux, vous pourrez aussi contribuer au croisement automatique entre les données extraites des documents et celles présentes dans notre système d'information.

PROFIL RECHERCHÉ

La CDC fonde le recrutement sur les compétences en excluant tout critère de discrimination. Tous nos postes sont ouverts aux personnes en situation de handicap.

La CDC fonde le recrutement sur les compétences en excluant tout critère de discrimination. Tous nos postes sont ouverts aux personnes en situation de handicap.

De formation scientifique, vous êtes en dernière année d'école d'ingénieur ou en Master 2.

Vous avez une forte appétence pour la data science et pour les évolutions récentes de l'IA et du deep learning, en particulier dans le traitement automatique du langage. Vous appréciez les sujets ouverts, où la solution reste à construire.

Compétences indispensables

- programmation en Python et maîtrise des principales bibliothèques de data science ;
- pratique du machine learning, mise en œuvre dans le cadre de projets académiques ou professionnels ;
- utilisation de Git.

Compétences appréciées

- un projet ou une expérience autour du traitement automatique du langage ;
- une familiarité avec les frameworks de deep learning (PyTorch / TensorFlow, Hugging Face) ;
- des notions de mise en production de modèles (conteneurisation, suivi d'expérimentations).

Qualités personnelles

Notre organisation est attachée à promouvoir au quotidien un mode de travail collaboratif. Vous avez un sens poussé de l'analyse, vous êtes rigoureux et curieux. Autonome et force de proposition, vous savez mener une démarche exploratoire jusqu'à un résultat exploitable, et restituer clairement vos travaux à des interlocuteurs techniques comme métier.