

STAGIAIRE DATA SCIENCESIT / MACHINE LEARNING

InvestorSight est une AI fintech startup américaine spécialisée dans l'analyse de données financières, l'évaluation des risques de gouvernance et la détection de signaux d'activisme actionnarial.

Notre activité repose sur l'exploitation de grandes quantités de données structurées et non structurées (marchés, fondamentaux, actualités, documents réglementaires) et sur le développement de modèles analytiques avancés permettant d'identifier des comportements atypiques au sein des entreprises cotées en bourse.

Nous concevons des outils internes reposant sur :

- des modèles de machine learning supervisés et non supervisés
- des techniques de traitement statistique avancées
- des pipelines de données fiables et scalables
- des méthodologies d'analyse appliquées au domaine financier et extra-financier

L'objectif du stage est de contribuer à ces travaux en renforçant la qualité de nos modèles et en participant à la structuration de nos processus data et analytiques.

Le/la stagiaire participera à l'ensemble du cycle de développement de nos modèles d'analyse, notamment :

- Préparation, structuration et nettoyage de jeux de données financiers complexes (séries temporelles, ratios, métriques de marché, historiques d'événements)
- Construction de variables explicatives pertinentes via un travail approfondi de feature engineering
- Développement, évaluation et ajustement de modèles de machine learning (régression, classification, modèles d'ensemble, méthodes bayésiennes si applicable)
- Analyse critique des performances, robustesse des modèles, et documentation des choix méthodologiques
- Collaboration à la mise en place ou à l'amélioration de pipelines de données automatisés
- Production de rapports techniques synthétiques destinés à alimenter les décisions pour les clients

Compétences indispensables (Must-Have)

- Excellentes bases en statistiques appliquées, probabilités, modélisation et analyse quantitative
- Maîtrise de **Python** et des bibliothèques scientifiques fondamentales (*pandas*, *numpy*, *scikit-learn*)
- Maîtrise de **SQL** et compréhension des architectures ETL/ELT

- Expérience dans la préparation, la structuration et le nettoyage de jeux de données complexes
- Bonne compréhension des algorithmes de Machine Learning (supervisé / non supervisé), des méthodes d'évaluation et de validation
- Capacité à analyser et interpréter les résultats modèles, identifier les limites et formuler des pistes d'amélioration
- Rigueur dans l'écriture du code, sens du détail et capacité à documenter les travaux de façon claire
- Intérêt marqué pour les problématiques financières, la gouvernance d'entreprise ou l'activisme actionnarial

Compétences appréciées (Nice-to-Have)

- Connaissance ou première expérience avec des frameworks de deep learning : **PyTorch**, **TensorFlow**
- Pratique ou familiarité avec des environnements **cloud** (idéalement **Google Cloud Platform**), data warehouses ou outils BigQuery
- Connaissance des principes d'**MLOps** : automatisation de pipelines, CI/CD, monitoring, versioning de modèles/données
- Expérience en **conception ou optimisation de bases de données** (relationnelles ou columnaires).
- Notions en entraînement ou utilisation de **modèles LLM**, techniques de **NLP**, ou analyse de textes (plus particulièrement appliquées au domaine financier)
- Compréhension générale des outils et approches d'IA les plus récents, ainsi que de la manière de **les exploiter efficacement dans un contexte produit**
- **Anglais courant** (la documentation, une partie importante des échanges techniques et le contexte du stage étant anglophones)
- Participation à des projets personnels, compétitions (Kaggle...), ou travaux académiques avancés en data science, ML ou finance

Niveau d'étude requis : Master, MSc ou Programme Grande Ecole

Date de début : Février/Mars 2026

Durée : 5 à 6 mois

Salaire : 1500 à 1800EUR/mois