



Stage : Gestion des Valeurs Manquantes en Machine Learning

Le **Comité de Recherche et Projets Collaboratifs** de l'association "**Ensemble pour Chehim**" propose ce stage en collaboration avec des **enseignants-chercheurs de l'ESEO**, Grande Ecole d'Ingénieurs en **France**.

1 Contexte et Problématique

Dans les systèmes d'apprentissage automatique, les jeux de données réels sont souvent incomplets, contenant des valeurs manquantes dues à des erreurs de collecte, des contraintes de stockage ou des limites des capteurs. Ces données manquantes peuvent nuire à la performance des modèles, altérer leur interprétabilité et compliquer les analyses.

La gestion des valeurs manquantes est une étape clé du pipeline de machine learning. Les méthodes traditionnelles comme la suppression des observations ou l'imputation simple (moyenne, médiane) peuvent introduire des biais ou entraîner une perte d'information. De nouvelles approches exploitant des techniques avancées (imputation multivariée, réseaux neuronaux, modèles bayésiens) permettent de mieux prendre en compte ces valeurs manquantes, mais leurs limites et performances dépendent fortement du contexte des données.

2 Objectif du stage

Ce projet vise à explorer et développer des techniques innovantes pour gérer les valeurs manquantes en machine learning.

3 Missions

1. État de l'art : Réaliser une revue critique des méthodes existantes pour traiter les données manquantes.
2. Méthodologies : Proposer de nouvelles approches intégrant des modèles génératifs (par exemple, Variational Autoencoders ou GANs) ou des techniques d'apprentissage actif pour imputer ou exploiter les données manquantes de manière plus robuste.
3. Expérimentation : Comparer les performances des approches développées sur des jeux de données en santé par exemple.

4 Compétences souhaitées

1. Maîtrise de la programmation en Python.



Modalité de candidature

Les candidats intéressés sont priés d'envoyer leur candidature (CV, lettre de motivation et relevé de notes de tous les semestres précédents) **par mail** à : togetherforchehim@gmail.com



Internship : Handling Missing Values in Machine Learning

The **Collaborative Research and Projects Committee** of the "Together for Chehim" association offers this internship in collaboration with **researchers from ESEO**, an engineering school in **France**.

1 Context and Problem Statement

In machine learning systems, real-world datasets are often incomplete, containing missing values due to data collection errors, storage constraints, or sensor limitations. These missing data can negatively impact model performance, impair interpretability, and complicate analyses.

Handling missing values is a key step in the machine learning pipeline. Traditional methods, such as removing observations or simple imputation (mean, median), can introduce bias or lead to loss of information. New approaches using advanced techniques (multivariate imputation, neural networks, Bayesian models) offer better handling of missing values, but their limitations and performance heavily depend on the data context.

2 Internship Objective

This project aims to explore and develop innovative techniques for handling missing values in machine learning.

3 Tasks

1. State of the Art : Conduct a critical review of existing methods for handling missing data.
2. Methodologies : Propose new approaches integrating generative models (e.g., Variational Autoencoders or GANs) or active learning techniques to impute or utilize missing data in a more robust manner.
3. Experimentation : Compare the performance of the developed approaches on datasets, such as those in healthcare.

4 Preferred Skills

1. Proficiency in Python programming.



Application procedure

Interested candidates are requested to send their application (CV, motivation letter, and grade transcript for all the previous semesters) **by e-mail** to : togetherforchehim@gmail.com