

# Predictive Mass Project

Implémentation MLOps dans l'Analytique Prédictive basée sur la Simulation

Aghilas Smail   Idriss Felloussi   Lamara Mouzni

Master 2 Informatique (SID)  
Aix Marseille Université

January 17, 2024

# Table des matières

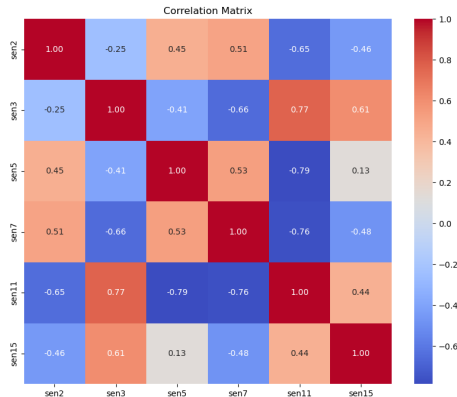
- 1 Introduction
- 2 Exploration des Données
- 3 Développement du Modèle
- 4 Approche MLOps
- 5 Déploiement du Modèle
- 6 outils utilisés

# Introduction

Le projet vise à créer un pipeline MLOps pour prédire la masse dans les simulations. L'objectif principal est d'assurer la précision du modèle, sa scalabilité et sa robustesse. Le processus couvre l'ensemble du cycle, de la collecte des données brutes au déploiement opérationnel du modèle.

# Exploration des Données

Pour amorcer le processus, les données sont chargées et explorées afin de comprendre leur nature. Une analyse des fonctionnalités est effectuée, incluant la création d'une matrice de corrélation pour examiner les relations entre différentes variables. Les statistiques sommaires fournissent un aperçu des tendances centrales et de la dispersion des données. En parallèle, l'identification des valeurs manquantes garantit la transparence dans l'évaluation de l'intégrité des données.

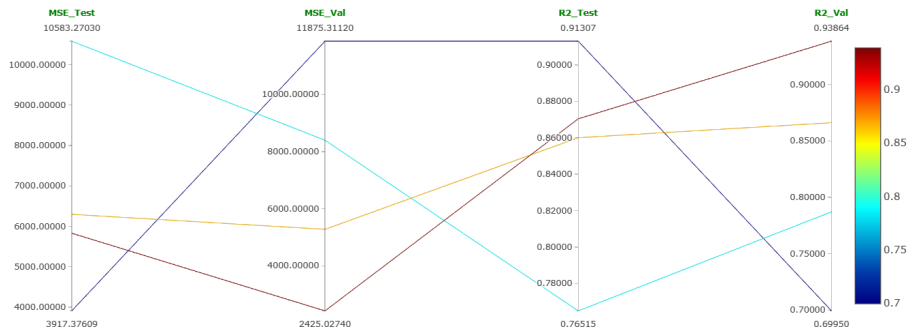


# Développement du Modèle

- Après avoir essayer différent modèles nous avons garder les 4 modèles :
  - **Linear Regressor** : Modèle de base.
  - **Random Forest Regressor** : Capable de modéliser des relations non linéaires, fonctionne bien sur des données avec un grand nombre de caractéristiques.
  - **Gradient Boosting Regressor** : Puissant, peut gérer des relations complexes, robustes aux données manquantes et peuvent gérer des ensembles de données relativement petits.
  - **XGBoost Regressor** : (eXtreme Gradient Boosting) une variante de de Gradient Boosting Regressor.

# Approche MLOps

Nous avons utilisée MIFlow pour suivre les expériences, tracker nos modèles et leurs hyper-paramètres et comparer leurs performances. Voici une de ces comparaisons :



# Déploiement du Modèle

- Application web Flask pour le déploiement des modèles.
- Deux pages HTML distinctes : une pour la saisie et l'autre pour les résultats.

## Flux de l'Application

- ❶ **Page de Saisie (GET)** : L'utilisateur est dirigé vers une page HTML avec un formulaire pour saisir les valeurs des capteurs ou charger un fichier CSV et choisir le modèle.
- ❷ **Prédiction (POST)** : Lorsque l'utilisateur soumet le formulaire, une requête POST est traitée par Flask. Les valeurs sont utilisées pour effectuer la prédiction avec le modèle sélectionné.
- ❸ **Page de Résultats (GET)** : L'utilisateur est ensuite redirigé vers une page HTML des résultats.

# outils utilisés



Trello

mlflow™





# Questions & Discussion

Merci ! Des questions ?