|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | 23/05/2025 | |
| Documentation Pipeline | | | | |
|  | | | | |
| Cynthia Nseke  Esteban Baigts  Alexandre Delmotte | |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Introduction **Objectif du Projet**  Ce projet met en place un pipeline de machine learning pour prédire  La **longueur des sépales** d’une fleur Iris à partir de leur **largeur**,  Via une architecture containerisée avec **Docker**.  **Contexte**  Ce projet s’inscrit dans un **travail collaboratif en équipe**, réparti sur  **4 jours consécutifs** à partir du **22 avril**.  Chaque groupe est composé de **3 à 4 étudiants**, travaillant ensemble à la  réalisation d’un **pipeline complet de traitement et de modélisation**  **de données**.  Nous avons été missionnés par un **laboratoire de recherche en**  **botanique** souhaitant automatiser l’analyse de spécimens floraux.  Plus précisément, le laboratoire cherche à estimer la **longueur des sépales**  de fleurs **Iris** à partir de leur **largeur**, à l’aide de techniques de machine  learning. **Rôles des membres du projet**  |  |  | | --- | --- | | **Nom** | **Rôles** | | Cynthia | Responsable de la documentation  Développeuse de l’interface Web | | Alexandre | Configuration du Docker,  Configuration du MLflow  Développeur IA | | Esteban | Mise en place du pré-processing et base des données |  Architecture Technique  Outils et Technologies utilisés **Docker** : Utilisé pour containeriser les services et assurer leur isolation et portabilité.      **Docker Compose** : Utilisé pour définir et gérer les services Docker dans un fichier  docker-compose.yml.    **FastAPI** : Framework web utilisé pour créer l'API RESTful qui permet de faire des prédictions.    **PostgreSQL** : Base de données relationnelle utilisée pour stocker les données du jeu de données Iris.    **MLflow** : Plateforme de gestion du cycle de vie des modèles de machine learning,  utilisée pour suivre les expériences, enregistrer les modèles, et gérer les artefacts  .    **Python** : Langage de programmation utilisé pour écrire les scripts de traitement des  dnnées et de modélisation.   * **Pandas** : Bibliothèque Python utilisée pour la manipulation et l'analyse des données.       **HTML-CSS-JS :**  front-end utilisateur  Html/CSS ->Mise en forme    JS--> Envoyer la valeur à l’API FastAPI et afficher le résultant.   Performance du modèle Pour évaluer la performance :  **Lancer le pipeline :** docker compose up – build  **Accéder à l’interface web :** localhost :5000 Interface utilisateur   Connexion en locale avec la requête : local:8000/predict | |
| **Conclusion**Objectifs atteints  * Pipeline fonctionnel et dockerisé * Interface Web fonctionnel  Piste d’améliorations **Modèle IA**   * Ajouter d'autres variables * Tester des modèles avancés (boosting, réseaux de neurones)   **Données**   * Utiliser plus de données d'entraînement * Ajouter des infos environnementales (climat, saisons)   **Front-end**   * Ajouter des infos environnementales (climat, saisons) |  |

|  |  |
| --- | --- |
|  |  |
|  | |