

Basé sur le code fourni, voici un rapport descriptif de la solution adoptée pour la détection des débuts de crises épileptiques en utilisant les pratiques MLOps :

## Contexte et Objectif

- **Nom de l'équipe** : DSTune +
- **Emails** : malek.Sahlia@ensi-uma.tn
- **Défi** : AI Night Challenge 24 - MLOps pour la détection des débuts de crises épileptiques.

La solution vise à mettre en place un pipeline automatisé pour la détection des débuts de crises épileptiques en s'appuyant sur des pratiques MLOps. Le code inclut l'importation des données, leur prétraitement, la division en ensembles d'entraînement, de validation et de test, la normalisation, l'entraînement d'un modèle XGBoost, et l'évaluation de la performance du modèle. Un aspect important du pipeline est également l'explication du modèle via SHAP (SHapley Additive exPlanations).

## Pipeline Détaillé

1. **Initialisation et Importation des Données** :
  - Clonage du dépôt Git contenant les données et les instructions nécessaires.
  - Lecture et affichage du contenu du fichier README.md pour des instructions détaillées.
2. **Préparation des Données** :
  - Assemblage des fichiers CSV individuels en un DataFrame pandas unique.
  - Séparation et équilibrage des données en ensembles d'entraînement, de validation et de test.
3. **Prétraitement des Données** :
  - Normalisation des caractéristiques (features) à l'aide de **StandardScaler**.
  - Préparation des matrices de caractéristiques et des vecteurs cibles pour l'entraînement et la validation.
4. **Entraînement et Évaluation du Modèle** :
  - Utilisation de **XGBClassifier** pour l'entraînement du modèle sur les données équilibrées.
  - Évaluation des performances du modèle sur les ensembles d'entraînement et de validation à l'aide de diverses métriques, y compris l'AUC, la précision, le rappel et la spécificité.
5. **Explication du Modèle** :
  - Application de SHAP pour expliquer les prédictions du modèle, offrant ainsi une visibilité sur l'importance des différentes caractéristiques pour les prédictions du modèle.

## Automatisation et Pratiques MLOps

- **ZenML** : Utilisation de ZenML pour automatiser et orchestrer le pipeline, de l'importation des données à l'explication du modèle.

- **Dashboard ZenML** : Mise en place d'un tableau de bord pour le suivi et la gestion du pipeline.
- **Reproductibilité et Scalabilité** : La structure du pipeline favorise la reproductibilité et la scalabilité des expérimentations ML.

## Conclusion

Cette solution adopte une approche complète et automatisée pour la détection des crises épileptiques en utilisant des modèles de machine learning avancés. L'intégration de pratiques MLOps telles que l'orchestration du pipeline avec ZenML et l'explication des modèles avec SHAP souligne l'importance de l'automatisation, de la transparence et de la reproductibilité dans les projets d'apprentissage automatique.

---