

Industrialiser les projets de Data science

L'objectif de cette formation est de savoir industrialiser un projet de Data Science en respectant les bonnes pratiques de développement. À l'issue de cette formation, vous serez capables de développer et maintenir un modèle de Machine Learning, de passer d'un modèle à sa production, d'identifier les acteurs et le coût de l'industrialisation d'un projet Data Science.

Pré-requis : Notions en Machine Learning, analyse de données, langage Python.



- Modalité :**
- Distanciel en classe virtuelle
 - E-learning : à venir
 - Présentiel

Communauté: community.reconvert.net

Durée totale : 14 H (2 jours)

Acteurs et rôles dans un projet Data Science

Équipes et compétences nécessaires
Data Scientists, Data Engineers, MLOps
Collaboration avec les parties prenantes
TP: atelier définir les rôles dans un projet

Coûts d'industrialisation d'un projet

Coût des ressources humaines et technologiques
Estimation du temps et des efforts
Analyse de retour sur investissement (ROI)
TP: Estimer les coûts d'un projet fictif
Conclusion et perspectives
Évolutions récentes en Data Science et MLOps
Innovations et tendances à surveiller
Questions et échanges avec les participants
TP: perspectives en Data Science (discussion)

PLAN DETAILLE

Introduction

Importance de l'industrialisation
Projet pilote vs projet industrialisé
Exemples de succès et d'échecs
TP: étude de cas : projet réussi vs projet échoué

Bonnes pratiques de développement en Data Science

Gestion de version
Documenter le projet, Sphinx
Packaging: setuptools, dépendances
Tests et validation des modèles
Rendre le code déployable
Le plan de déploiement
Production des rapports
Reproductibilité des expériences
TP: gestion de version sur un projet

Architecture d'un projet Data Science

Composants d'une architecture de projet
Intégration des outils et technologies
Flux de données et pipelines de traitement
TP: architecturer un projet fictif

Transition d'un modèle à la production

Stratégies de déploiement
Monitoring et maintenance des modèles
Gérer les drifts, mise à jour des modèles
TP: déployer un modèle en environnement de test

Industrialiser avec Kubeflow

Déploiement dans le cloud
Les plateformes d'IA
Qu'est-ce que Kubeflow, composants
Notions de Kubernetes et Docker
Kubeflow sur GCP : configurer
Kubeflow : notebook, modéliser, training
Créer un Pipeline d'apprentissage
Déploiement avec Kubeflow