

Le cycle de vie de la science des données comprend plusieurs étapes clés qui permettent de transformer des données brutes en insights exploitables. Voici les principales étapes :

### 1. Définition du problème

Déterminer le problème à résoudre ou la question à poser. Cela implique de comprendre les besoins des parties prenantes.

### 2. Collecte de données

Rassembler les données pertinentes à partir de diverses sources, qu'elles soient internes (bases de données) ou externes (API, web scraping).

### 3. Préparation des données

Nettoyer et transformer les données pour les rendre exploitables. Cela inclut la gestion des valeurs manquantes, la normalisation et la transformation des types de données.

### 4. Exploration des données

Analyser les données pour découvrir des tendances, des patterns ou des anomalies. Cela se fait souvent à l'aide de statistiques descriptives et de visualisations.

### 5. Modélisation

Choisir et appliquer des algorithmes de machine learning ou de statistiques pour créer des modèles prédictifs. Cette étape implique souvent la sélection de caractéristiques pertinentes.

### 6. Évaluation

Tester et évaluer les performances du modèle à l'aide de métriques appropriées (précision, rappel, F1-score, etc.) pour s'assurer qu'il répond aux objectifs fixés.

### 7. Déploiement

Mettre le modèle en production pour qu'il puisse être utilisé dans des applications réelles. Cela inclut la création d'API ou l'intégration dans des systèmes existants.

### 8. Surveillance et maintenance

Suivre les performances du modèle au fil du temps et effectuer des mises à jour ou des retrainings si nécessaire pour s'adapter à de nouvelles données.

### 9. Communication des résultats

Présenter les résultats et les insights de manière claire et compréhensible aux parties prenantes, souvent à l'aide de visualisations et de rapports.

Ce processus est itératif et peut nécessiter plusieurs cycles pour affiner le modèle et améliorer les résultats.