

Team Data Science Process

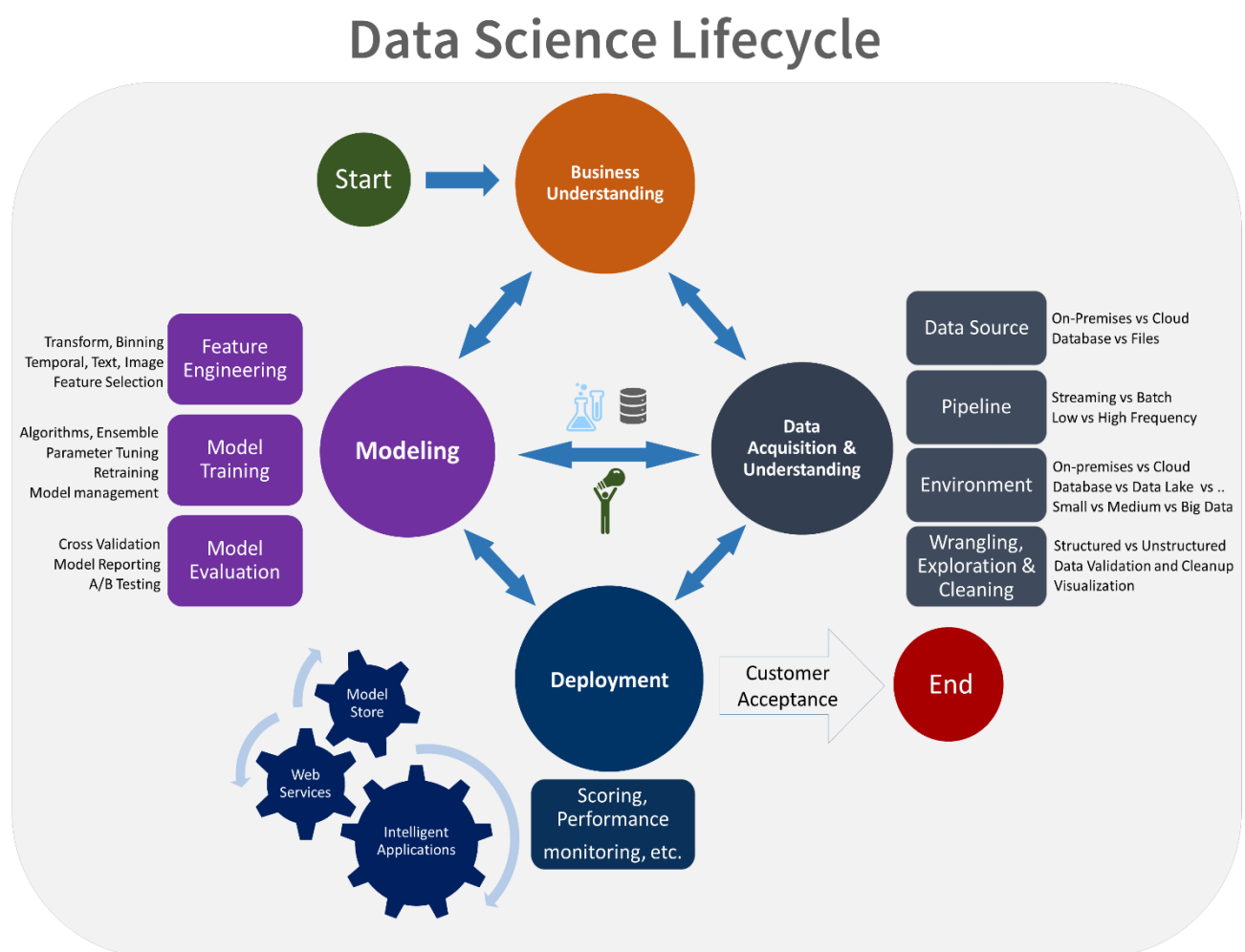
Aperçu

Le Team Data Science Process (TDSP) est une méthodologie de data science agile et itérative pour délivrer une solution d'analyse prédictive et des applications intelligentes efficacement. Le TDSP vise à améliorer la collaboration en équipe, il inclut les meilleurs pratiques ainsi que des structures issues de Microsoft pour contribuer à l'implémentation de projets de Data Science

Les composants clé du TDPS

- ❖ Un cycle de vie Data science
- ❖ Une structure projet standardisé
- ❖ Des infrastructures et ressources recommandé pour des projets data science
- ❖ Des Outils recommandé pour l'exécution du projet

Cycle de vie



Compréhension du métier :

1. L'objectif de cette étape est d'identifier les variables métier clés que l'analyse doit prédire.
2. Définir les objectifs du projet en posant des questions précises, spécifiques et non ambiguës.
3. Définir l'équipe du projet en spécifiant les rôles et responsabilités de chaque membre
4. Définir des métriques de succès tel qu'un pourcentage de précision de l'analyse .

Acquisition et compréhension des données

1. Produire un data set propre, de qualité qui est en relations avec les variables métier clés
2. Développer une architecture d'un pipeline de données pour rafraichir les données régulièrement
3. Exploration des données pour acquérir une meilleure compréhension des données, obtenir des données de meilleure qualité et définir si elles sont suffisantes pour l'analyse

Modélisation

1. Extraire des caractéristiques à utiliser dans l'analyse à partir des données brutes
2. Entraîner un model suivant ces étapes
 - a. Diviser les données aléatoirement en données d'entraînement ainsi que des données de test
 - b. Construire des modèles à travers les données d'entraînements
 - c. Evaluer les données d'entraînement et de test en utilisant des algorithmes Machine Learning différents
 - d. Déterminer la meilleure solution en comparant les métriques de succès définies auparavant des différentes méthodes.

Déploiement

Déployer les modèles performant à travers plusieurs applications selon les besoins tel que : des Dashboard, Site web en ligne ou des application back end.

Acceptation par le client

Confirmer que le modèle déployé satisfait le besoin du client puis livrer le projet à l'entité qui sera amenée à exécuter le système en production

Rôles et tâches

Group Manager

Gère l'entière équipe data science au sein de l'entreprise

Team Lead

Gère une équipe au sein de l'unité data science de l'entreprise ce rôle peut être confondu avec le group manager si l'unité de data science est réduite

Project Lead

Gère les activités journalières des data Scientists assignés a un projet spécifique

Project Individual Contributors

Comprend les Data Scientists, Business Analysts, Data Engineer, Architectes et autres qui contributeurs au projet data science

