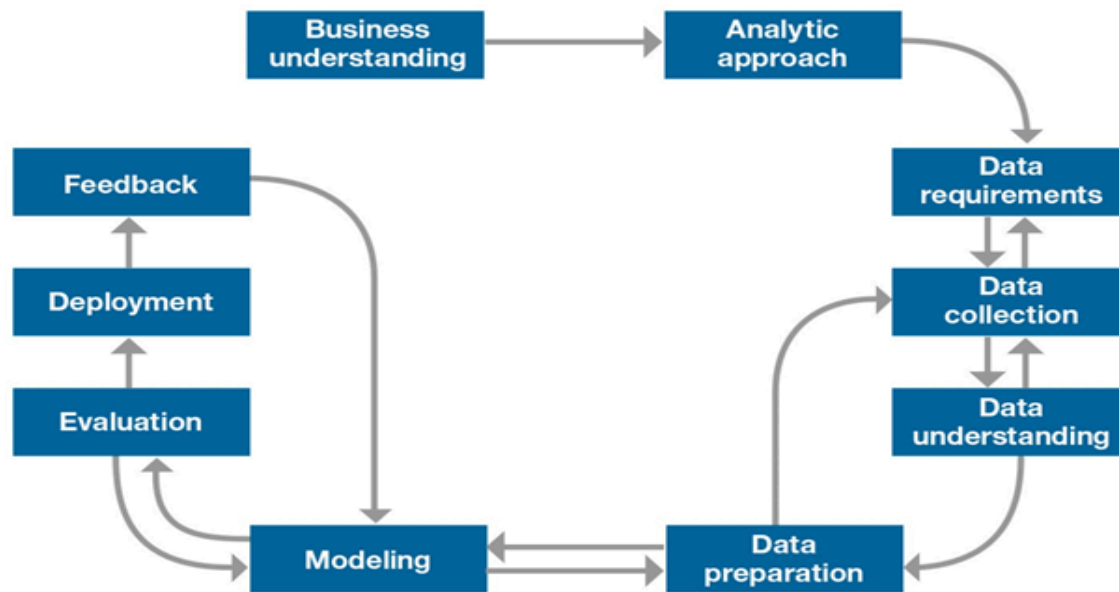




1-Business understanding (compréhension Métier)

Le rôle essentiel de cette étape Est de définir le problème et d'identifier les objectifs métiers et les exigences de la solution du point de vue commercial.

2-Analytic Approach (Approche analytique)

Le rôle de cette étape Définir les objectifs dataScience et d'identifier les techniques ML permettant d'atteindre le résultat souhaité.

3-Data Requirement (Exigences niveaux données)

Le rôle de cette étape est de déterminé les exigences en matière de données; elles nécessitent un contenu, des formats et des représentations de données particuliers, guidés par les connaissances du domaine.

4-Data Collection (Collecte de données)

Le rôle de Cette étape consiste à identifier et à rassembler des ressources de données- structurées, non_structurées et semi-structurées qui sont pertinentes pour le problème.

NB : si il rencontre des lacunes il faudra retourne al l'étape précédente qui est l'exigences en matière de données et de collecter davantage de données.

5-Data Understanding (Compréhension des données)

Le rôle de cette étape et de comprendre le contenu des données, à évaluer leur qualité et à découvrir les informations initiales relatives à ces données.

La collecte de données pourrait être nécessaire pour combler les lacunes de la compréhension.

6-Data preparation (préparation des données)

Généralement c'est l'étape la plus longue donc il est recommandé d'automatiser certaines étapes de préparation des données.

Les activités de préparation des données comprennent le nettoyage des données : gestion des valeurs manquantes ou non valides (remplacer), élimination des doublons,...

Et la combinaison des données de plusieurs sources (fichiers, tableaux, plates-formes)

Et la transformation des données en variables utiles plus on prépare correctement les données, plus on facilite la tâche aux modèles utilisés (modélisation)

Process :

-Features engineering : il vérifie les valeurs manquantes, les critères ou la manière de remplacer des valeurs manquantes et remplace les valeurs manquantes et réduit l'échelle des données numériques

-Features sélection : il élimine les informations inutiles.

-dimensions réduction : réduction des dimensions.

7-Modeling (modélisation)

C'est un processus itératif permet de développer des modèles.

Il y a l'apprentissage non supervisé ou analyse descriptive qui Désigne la situation d'apprentissage automatique où les données ne sont pas étiquetées.

Et il y a l'apprentissage supervisé ou analyse prédictive qui consiste à apprendre une fonction de prédiction à partir d'exemples annotés.

8-EVALUATION(évaluation des modèles)

Avant le déploiement, il faut évaluer le modèle de l'étape précédente, comprendre sa qualité et s'assurer qu'il répond correctement et entièrement au problème de l'entreprise

Pour un modèle prédictive on utilise un lot de données test (test dataset), construire des tableaux et des graphiques représentatifs.

On peut utiliser également des tests de signification statistique afin de s'assurer de sa qualité.

9-Deployment(déploiement)

Une fois qu'un modèle satisfaisant a été développé et approuvé par les sponsors commerciaux, il est déployé dans l'environnement de production ou dans un environnement de test comparable.

Le déploiement d'un modèle dans un processus métier opérationnel implique généralement plusieurs groupes, compétences et technologies.

10-feedback

En collectant les résultats du modèle mis en œuvre, l'organisation reçoit des informations en retour sur les performances du modèle et observe son impact sur son environnement de déploiement.

L'analyse de ces commentaires permet au 'spécialiste des données' d'affiner le modèle, en augmentant sa précision et donc son utilité.