

Projet du Cours MGL7811

Chapitre 3 – Validation et gestion des données

Utilisez les questions ci-dessous pour guider le chapitre 2 de votre rapport de projet.

Caractéristique du jeu de données

Explorez votre jeu de données et écrivez vos observations.

- Consultez le notebook (Kagle) et la présentation de la semaine 2 pour les exemples de différents types d'analyse.

Décrivez chaque attribut du jeu de données:

- Type de valeur (catégorie, numérique)
- Distribution des valeurs
- Problèmes potentiels (données manquantes, les biais)

Traitement de données

Décrive le traitement des données/l'ingénierie des fonctionnalités

- Par attribut:
 - Normalisation, Scaling, transformation, etc.
 - Consultez l'article "The Data Linter" pour obtenir plusieurs exemples de traitement et validation de données
- Sélection des fonctionnalités
 - Consultez la présentation de la semaine 5

Validation des données

Créer un schéma pour toutes les fonctionnalités, notamment:

- Fréquence attendue dans l'ensemble de données
- Valeurs de plage attendues
- Consultez l'article » Data Validation for Machine Learning «

Planifier un composant de validation qui:

- Lit le schéma défini
- Lit votre jeu de données
- Valide le jeu de données