

# Automatisation avec les pipelines

Atelier #12





# Qu'est-ce qu'un pipeline?

**Définition** : Un pipeline est une succession d'étapes prédéfinies et ordonnées qui peut être utilisée pour automatiser un flux de travail.

## Pourquoi utiliser les pipelines?

1. Utiliser les pipelines pour minimiser la fuite d'information provenant des données.
2. Construire les pipelines pour préparer les données et modéliser.
3. Construire les pipelines pour extraire/sélectionner les variables et modéliser.

**Les options et étapes pouvant être incluses dans les pipelines sont quasi infinies, ce qui les rends très puissants et versatiles.**



# Étapes possibles à inclure dans un pipeline

Nettoyage  
des données

Transformation  
des données  
textuelles en  
numérique

Apprentissage  
d'un modèle

Traitement des  
valeurs nulles

Préparation  
des données

Inclure  
une  
validation  
croisée

Sélection  
des  
variables

Application de  
méthodes  
d'ensemble

**Aperçu d'un projet réel  
utilisant les pipelines.**

