

Chapitre 2

# Apprentissage Automatique (Machine Learning)

---

# Plan du chapitre

1. Introduction
2. Domaines d'applications
3. Processus ML



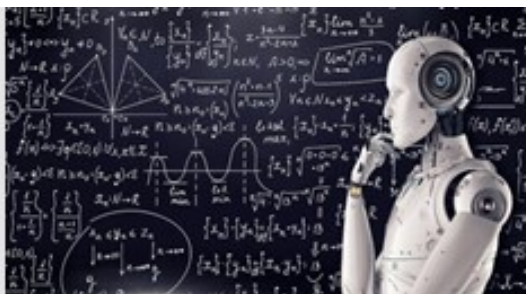
# 01

## Introduction

---



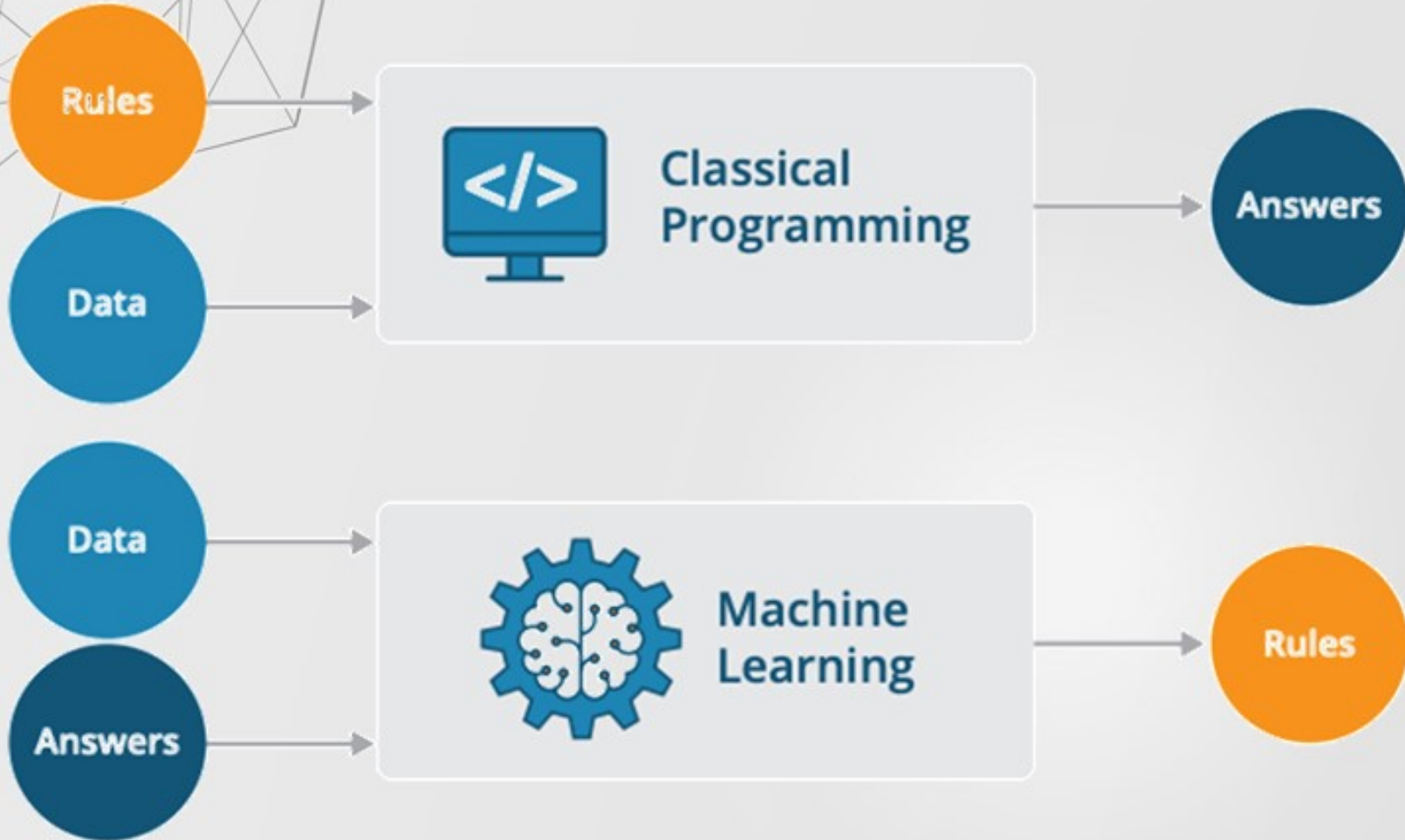
# Contexte



- De manière générale, un programme informatique tente de résoudre un problème pour lequel nous avons la solution.  
Par exemple : calculer la moyenne générale des étudiants, classer les étudiants selon leur moyenne. ...
- Pour certains problèmes, nous ne connaissons pas de solution exacte et donc nous ne pouvons pas écrire de programme informatique.  
Par exemple : reconnaître automatiquement des chiffres écrits à la main à partir d'une image scannée, déterminer automatiquement une typologie des clients d'une banque, jouer automatiquement aux échecs contre un humain ou un autre programme ...

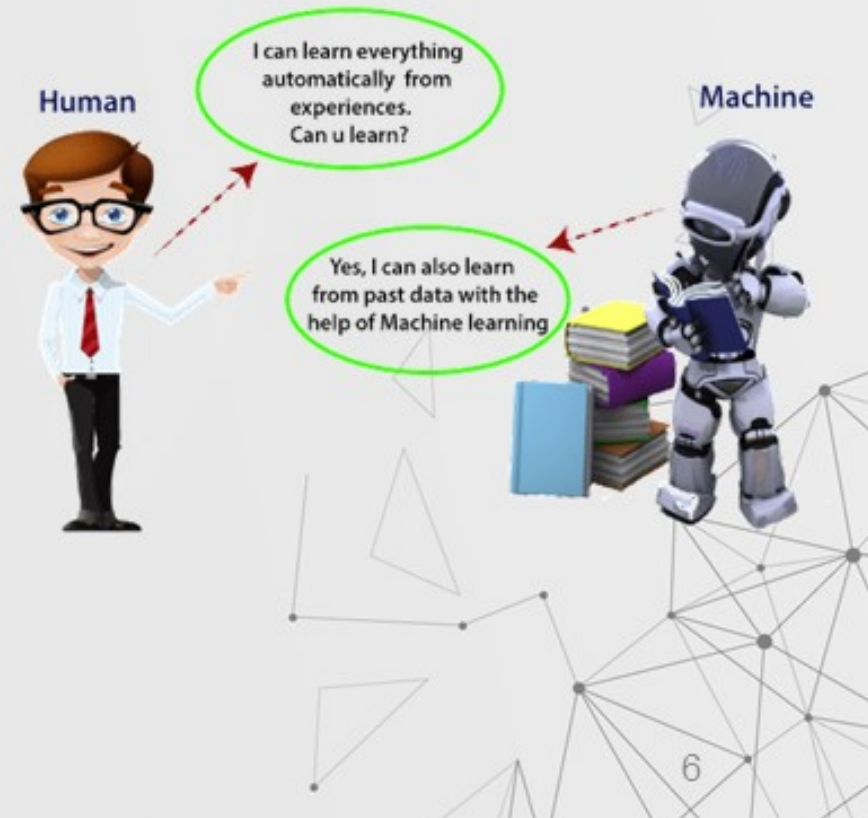
En revanche, pour ces problèmes, il est facile d'avoir une base de données regroupant de nombreuses instances du problème considéré.

# Machine learning vs programmation classique



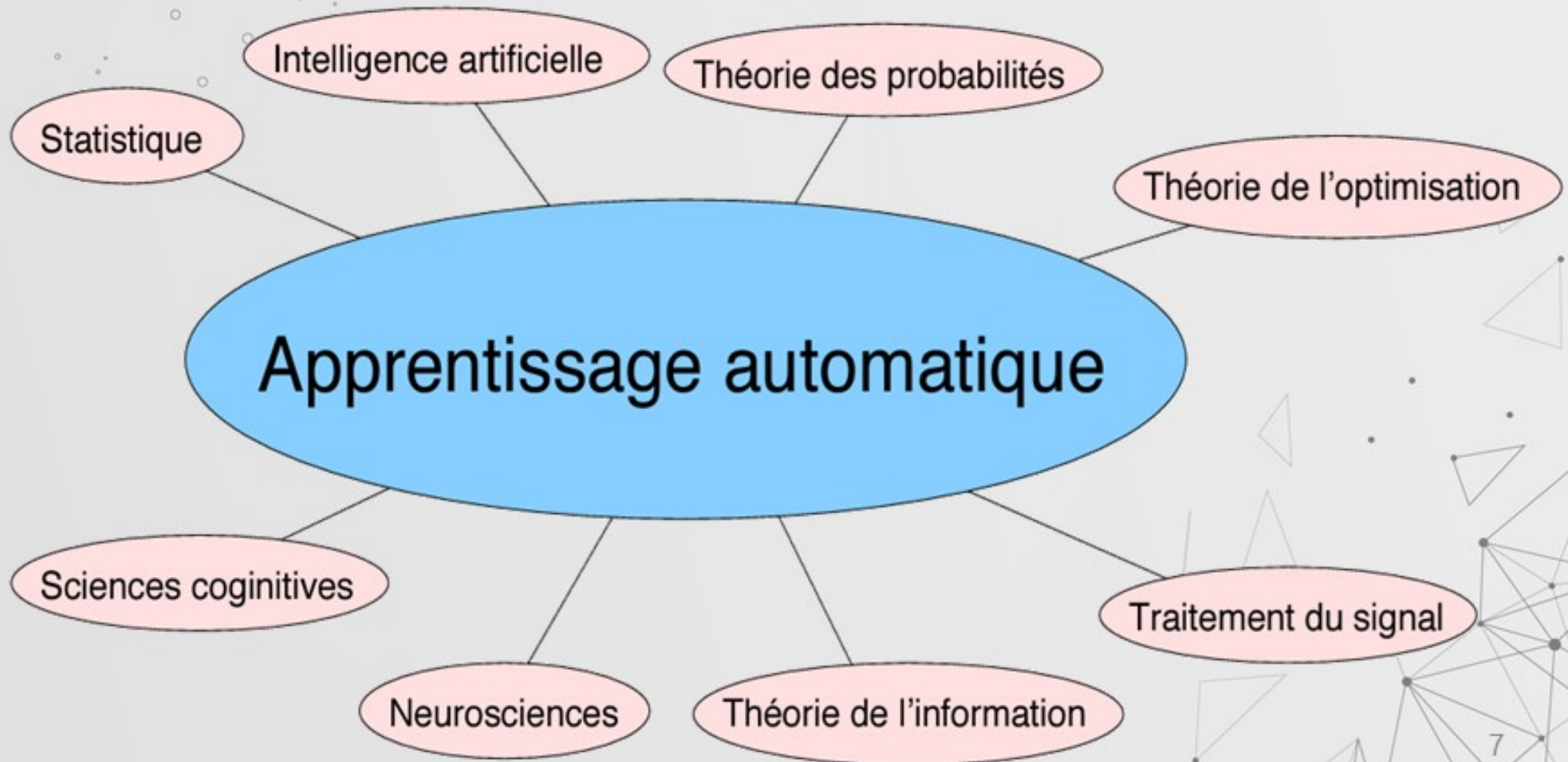
# Définition

- L'**apprentissage automatique** est une des familles essentielles de techniques pour l'Intelligence Artificielle
- Il consiste à amener les ordinateurs à apprendre et à agir comme les êtres humains, et à améliorer leur apprentissage de manière automatique, en se nourrissant de données.
- Cela passe par le développement d'algorithmes permettant de rendre une machine autonome sans avoir été explicitement programmée dans ce but.



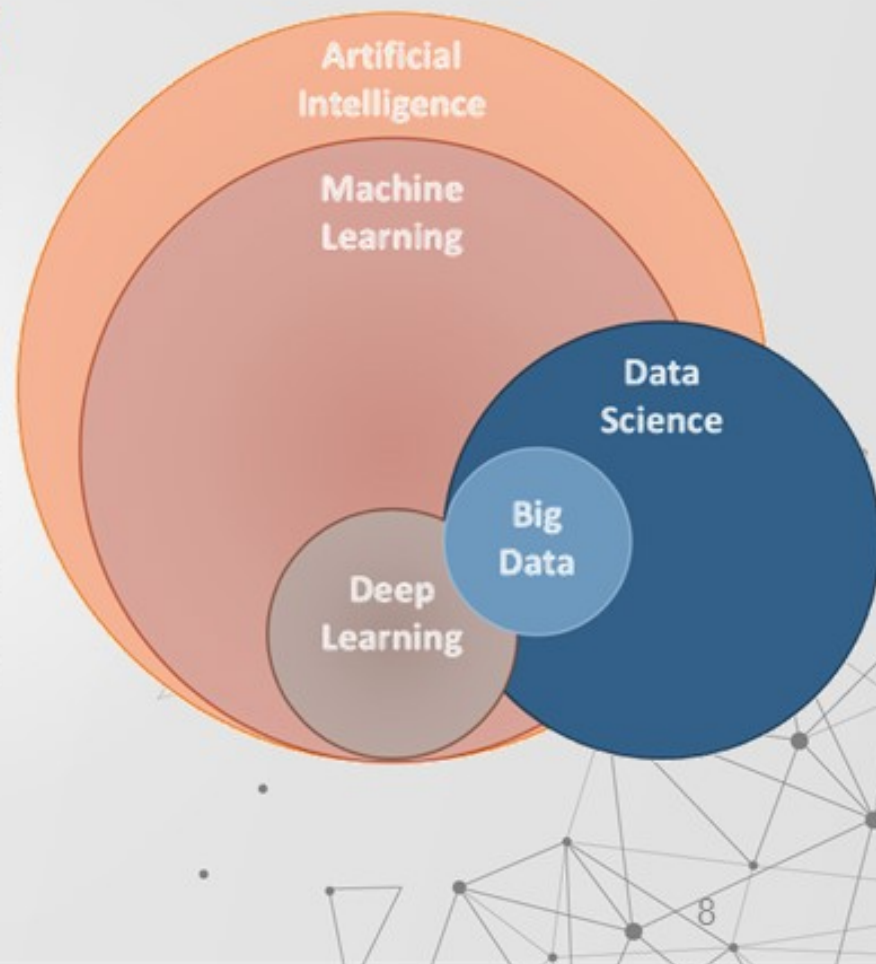


# ML: Un domaine pluridisciplinaire



# ML & matières connexes

- **La science des données (Data Science):** approche pluridisciplinaire pour l'extraction de connaissances à partir de données hétérogènes
- **Les données massives (Big Data):** solution dessinée pour permettre à tout le monde d'accéder en temps réel à des bases de données géantes.







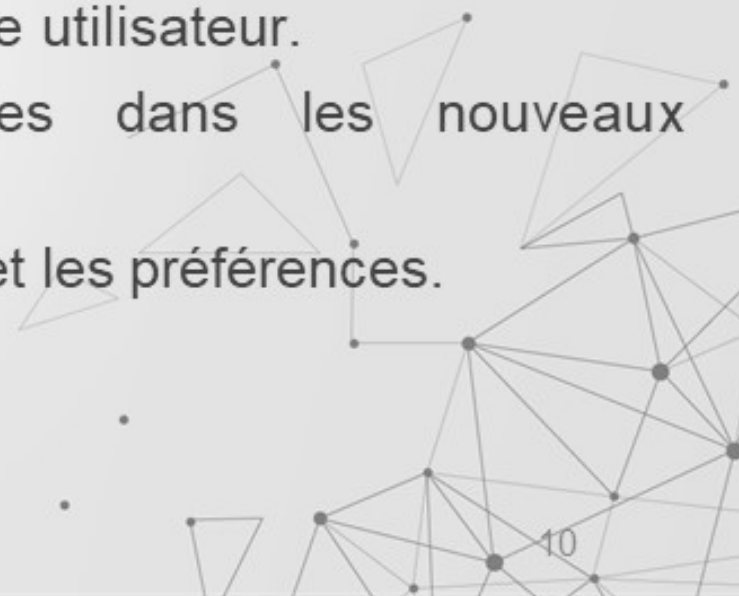
# 02

## Domaines d'applications

---

# Applications du ML

- **Recherche Web** : Classement des informations en fonction de sur quoi vous êtes le plus susceptible de cliquer.
- **Finance** : Evaluation du risque sur les offres de crédit. Comment décider où investir de l'argent.
- **E-commerce** : Prédire le désabonnement des clients. Afficher les produits en fonction des préférences de chaque utilisateur.
- **Robotique** : La gestion des incertitudes dans les nouveaux environnements. Voiture autonome.
- **Réseaux sociaux**: Données sur les relations et les préférences.



# Applications du ML

**Reconnaissance  
Vocale & Image**

**Diagnostic Médical**

**Filtrage des Spams  
sur le Web**

**Recommandation de  
contenu**



**Détection de Fraudes**

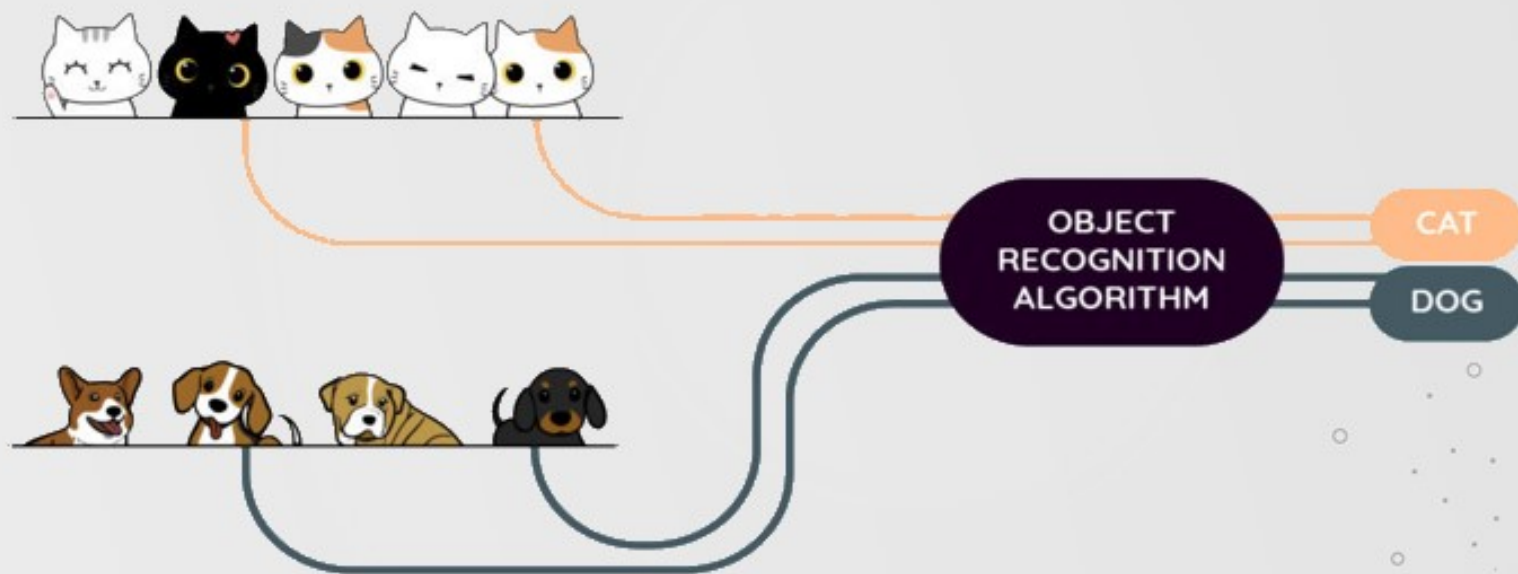
**Traduction**

**Prédiction diverses**

**Assistance Personnel  
vituel**

# Applications : Reconnaissance d'image

la reconnaissance d'images consiste à analyser les pixels et les formes d'une image pour reconnaître l'image comme un objet particulier.



# Applications : Assistant personnel virtuel

Comme leur nom l'indique, les assistants personnels virtuels aident à trouver des informations utiles, lorsqu'on leur demande par texte ou par voix.

Voici quelques-unes des principales applications de l'apprentissage automatique :

- Reconnaissance de la parole
- Conversion parole en texte
- Traitement du langage naturel
- Conversion de texte en parole



SIRI



ALEXA



GOOGLE

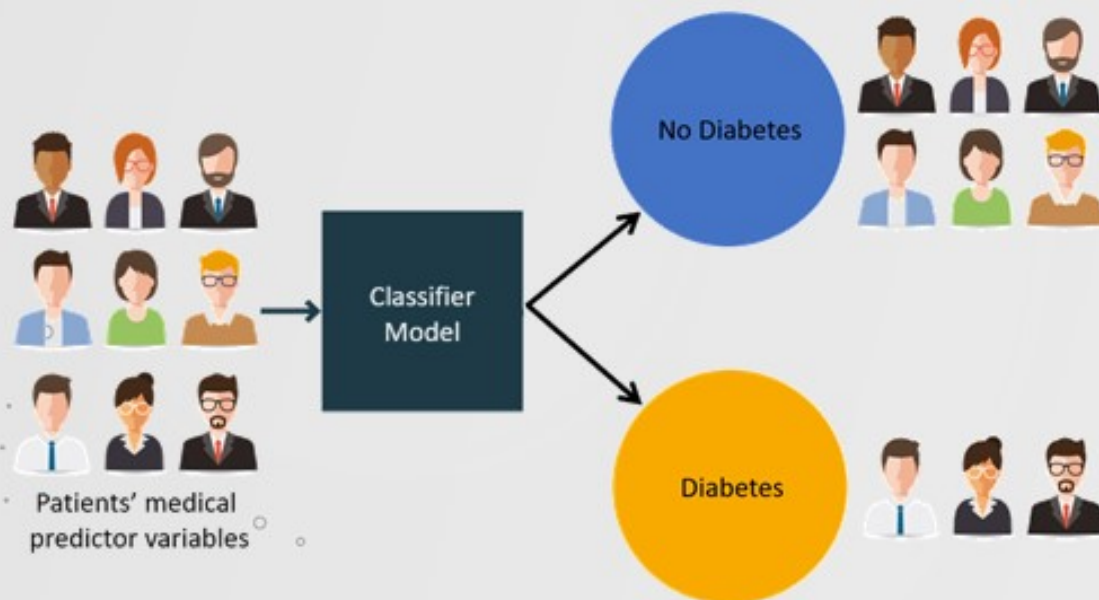


CORTANA



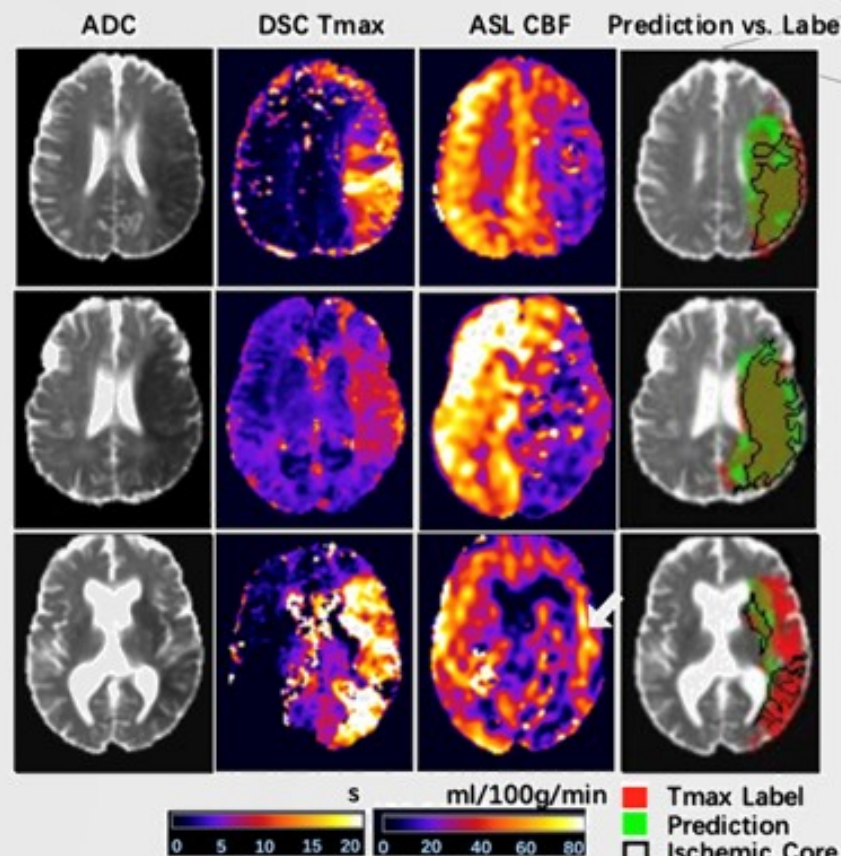
# Applications : Diagnostic médical

Les systèmes de ML sont utilisés dans le domaine de la médecine, pour étudier les résultats et rechercher certaines maladies par exemple, mais également pour établir des diagnostics à partir de rapports de pathologie.

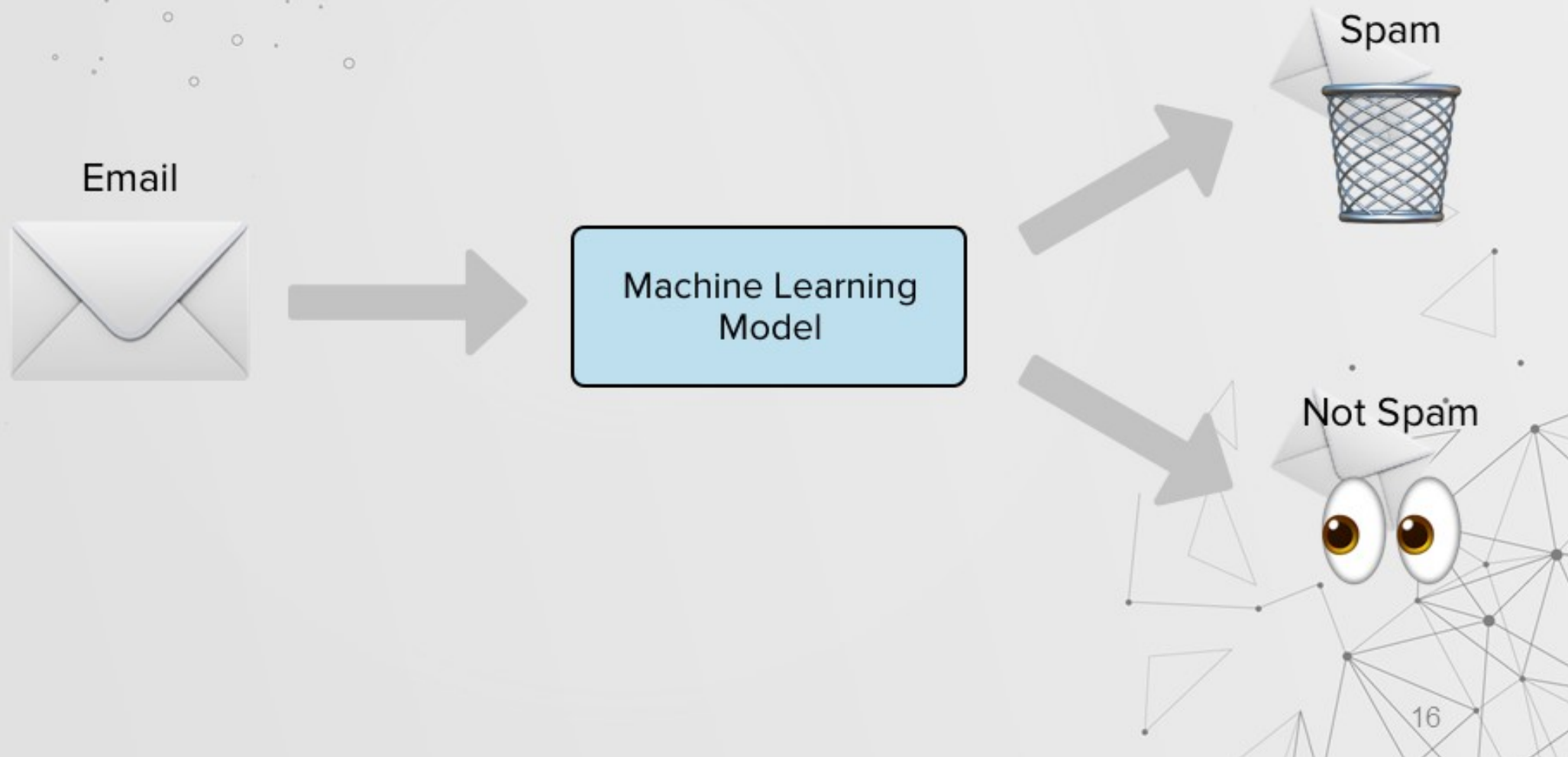


# Applications : Diagnostic médical

- ▷ L'ordinateur apprend à
- ▷ prédire la prédisposition de patients à une maladie sur la base de données génétiques associées à des cas de patients confirmés malades ou sains.



# Applications : filtrage des spams sur le Web



# Applications : Recommandations de produits

35% des revenus d'Amazon sont générés par les recommandations de produits.

⇒ Eh bien, cela se produit parce que Google suit votre historique de recherche et recommande des annonces en fonction de votre historique de recherche.





# Applications : Détection de fraudes



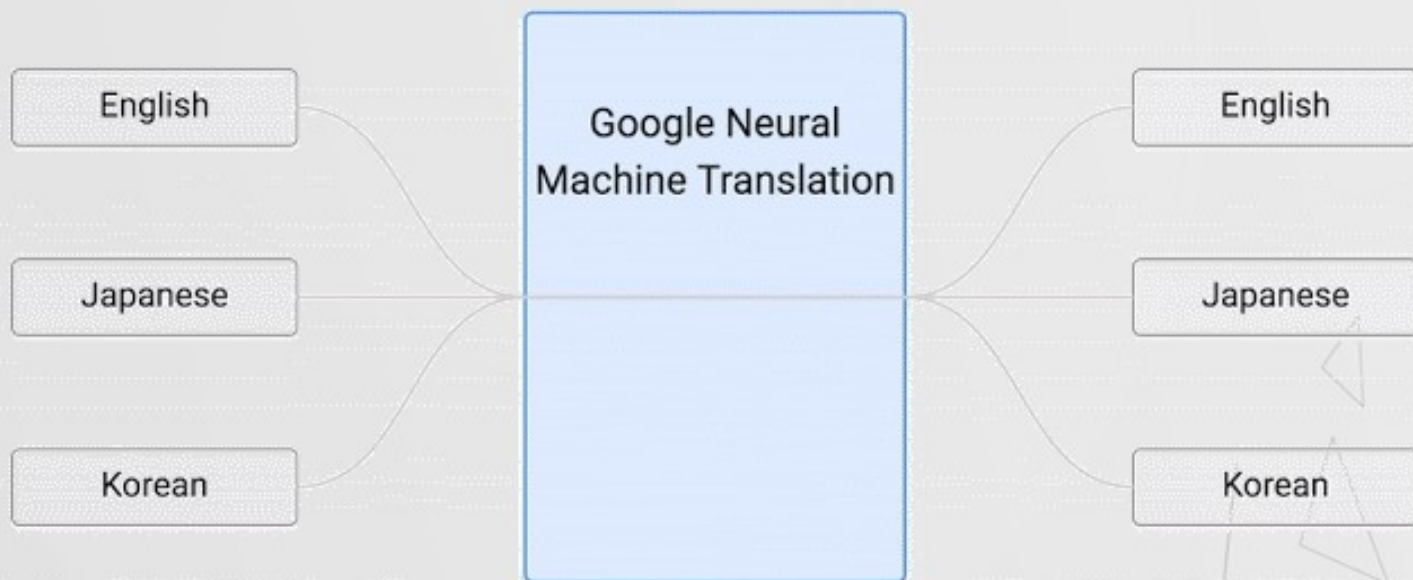
Le nombre de transactions a augmenté. Dans le même temps, de nombreux criminels sont devenus aptes à trouver des failles.

Chaque fois qu'un client effectue une transaction, le modèle d'apprentissage automatique vérifie minutieusement son profil à la recherche de schémas suspects.



# Applications : Traduction

Training



# Applications : Alertes de trafic



Tous ceux qui utilisent des cartes fournissent leur emplacement, leur vitesse moyenne, l'itinéraire sur lequel ils voyagent.

Cela aide Google à collecter des données massives sur le trafic, ce qui leur permet de prédire le trafic à venir et d'ajuster votre itinéraire en fonction de celui-ci.

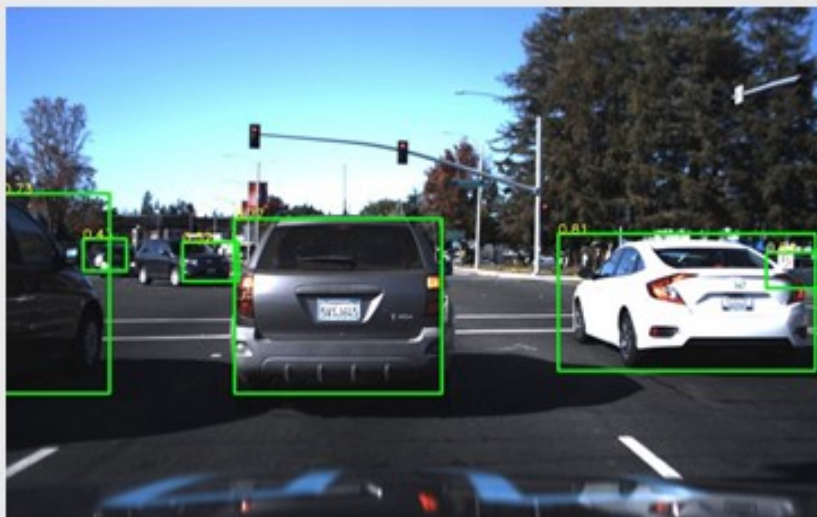
# Applications : Réseaux sociaux

Facebook utilise la détection de visage et la reconnaissance d'image pour trouver automatiquement le visage de la personne qui correspond à sa base de données et nous suggère donc de marquer cette personne en fonction de DeepFace.



Les suggestions automatiques de marquage d'amis sur Facebook ou sur toute autre plate-forme de médias sociaux.

# Applications : Voitures autonomes



Tesla est le leader dans ce secteur.

La partie IA est dirigée par le fabricant de matériel NVIDIA.

Le modèle collecte les données de tous ses véhicules et de ses conducteurs.

Il utilise des capteurs internes et externes qui font partie de l'IOT pour détecter des personnes ou des objets



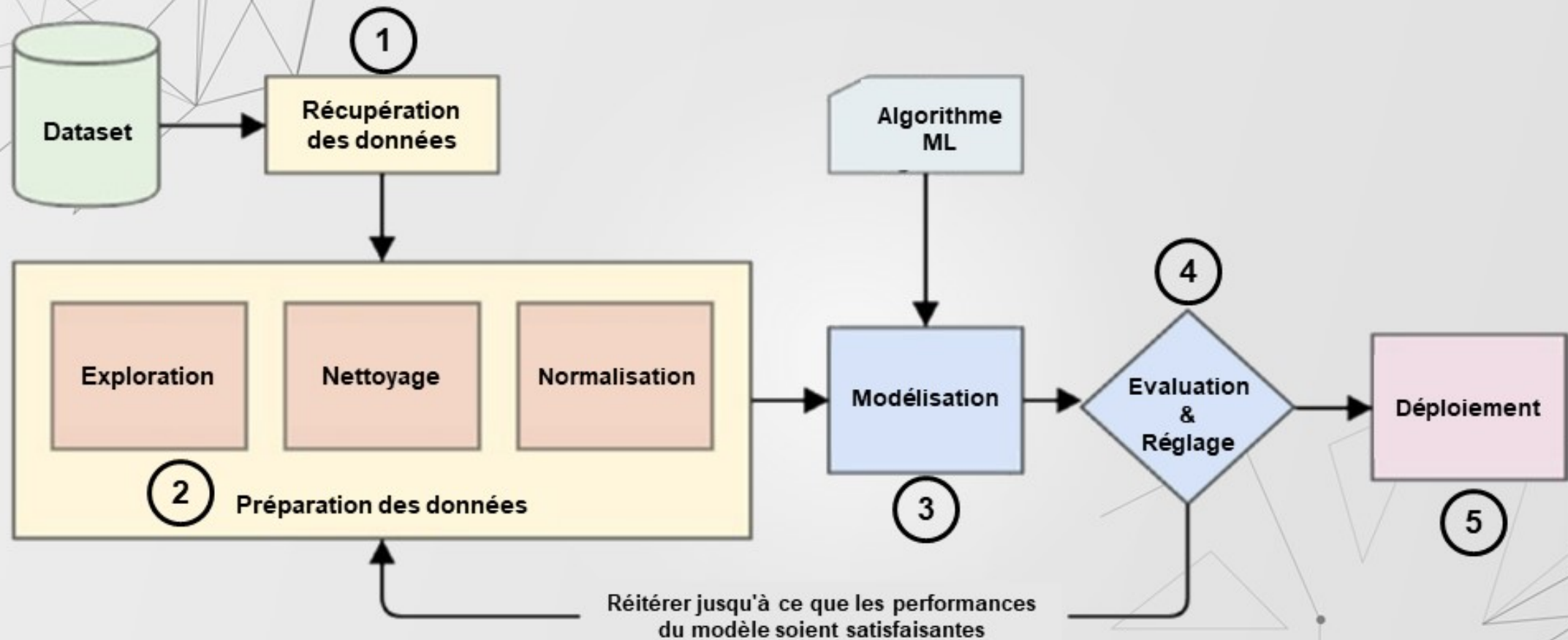
# 03

## Processus ML

---



# Chaîne du processus ML



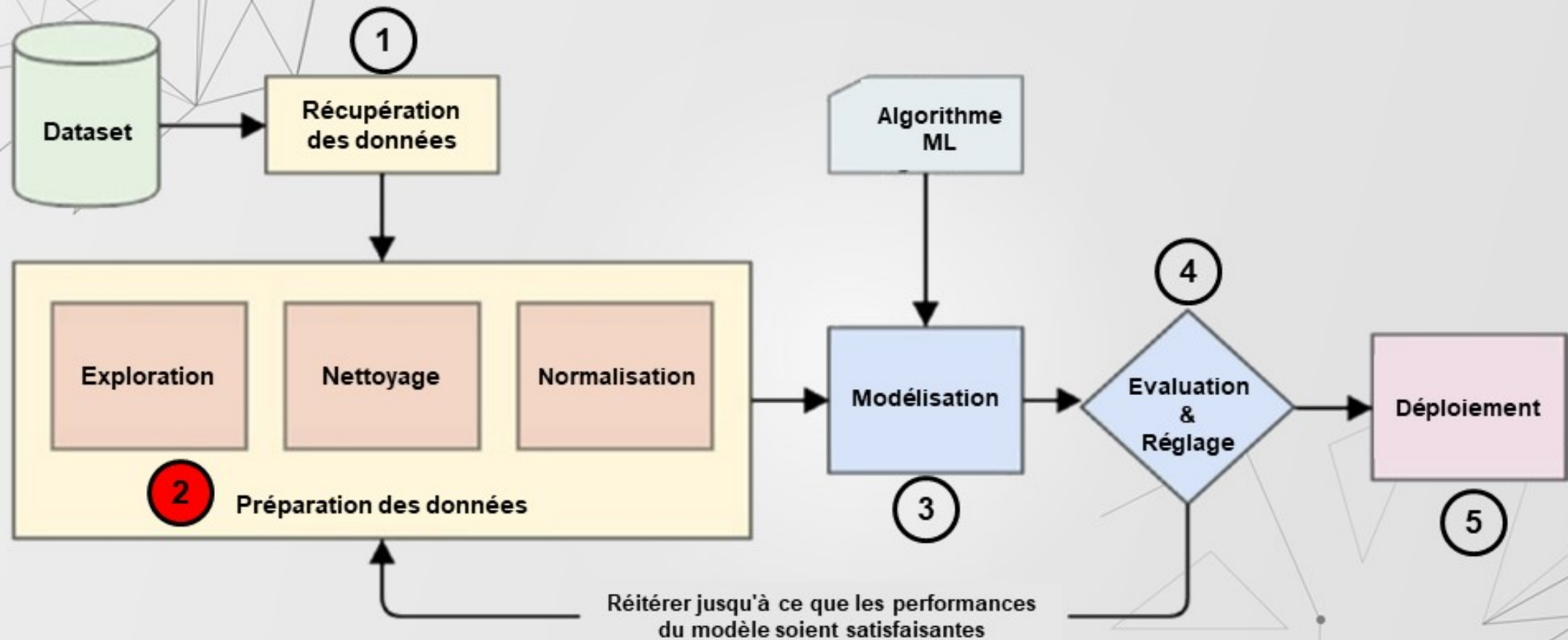
# Chaîne du processus ML

## Etape 1: Récupération des données

Collecter des données qui peuvent être :

- de différentes sources :
  - entrepôts de données
  - Web
  - capteurs,
  - ...
- et de différents types :
  - Numériques
  - Textuelles
  - Images/Graphes
  - Séquences
  - ...

# Chaîne du processus ML



# Chaîne du processus ML

## Etape 2: Préparation des données

### a. Exploration

- Faire un inventaire des données
  - Typologie: Numériques , Temporelles, Textes, Binaires, ...
  - Variables catégorielles, discrètes ou continues
  - Nombre d'observations (nombre de lignes)
  - Nombre de caractéristiques/features/variables (nombre de colonnes)
- Détecter les anomalies:
  - Outliers (valeurs aberrantes)
  - Valeurs manquantes
  - Corrélations

# • Chaîne du processus ML

## • Etape 2: Préparation des données

### b. Nettoyage

Préparer les features/variables afin qu'elles soient utilisables par des algorithmes du ML:

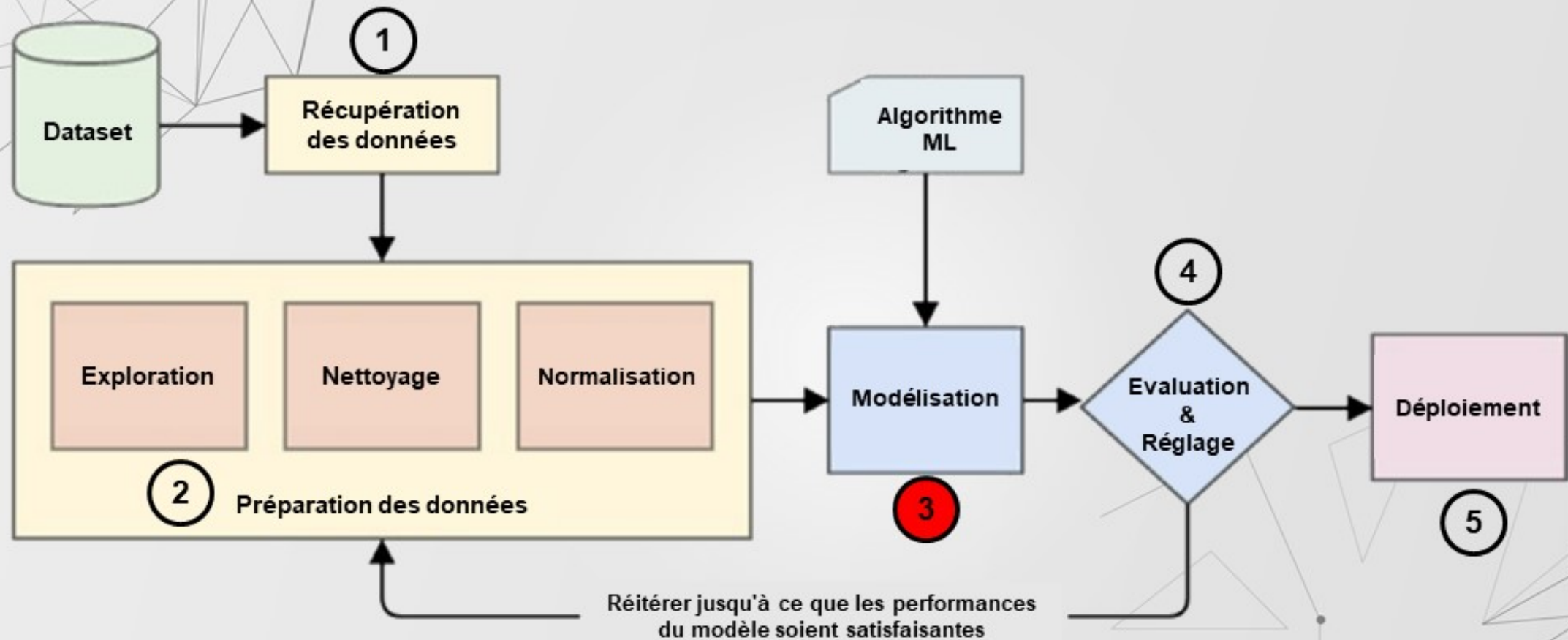
- Remplacer ou supprimer les valeurs manquantes/ aberrantes
- Transformer des données (variables) au format numérique

### c. Normalisation

Il s'agit d'une mise à l'échelle des valeurs numériques pour ne pas avoir un grand écart entre les valeurs..



# Chaîne du processus ML



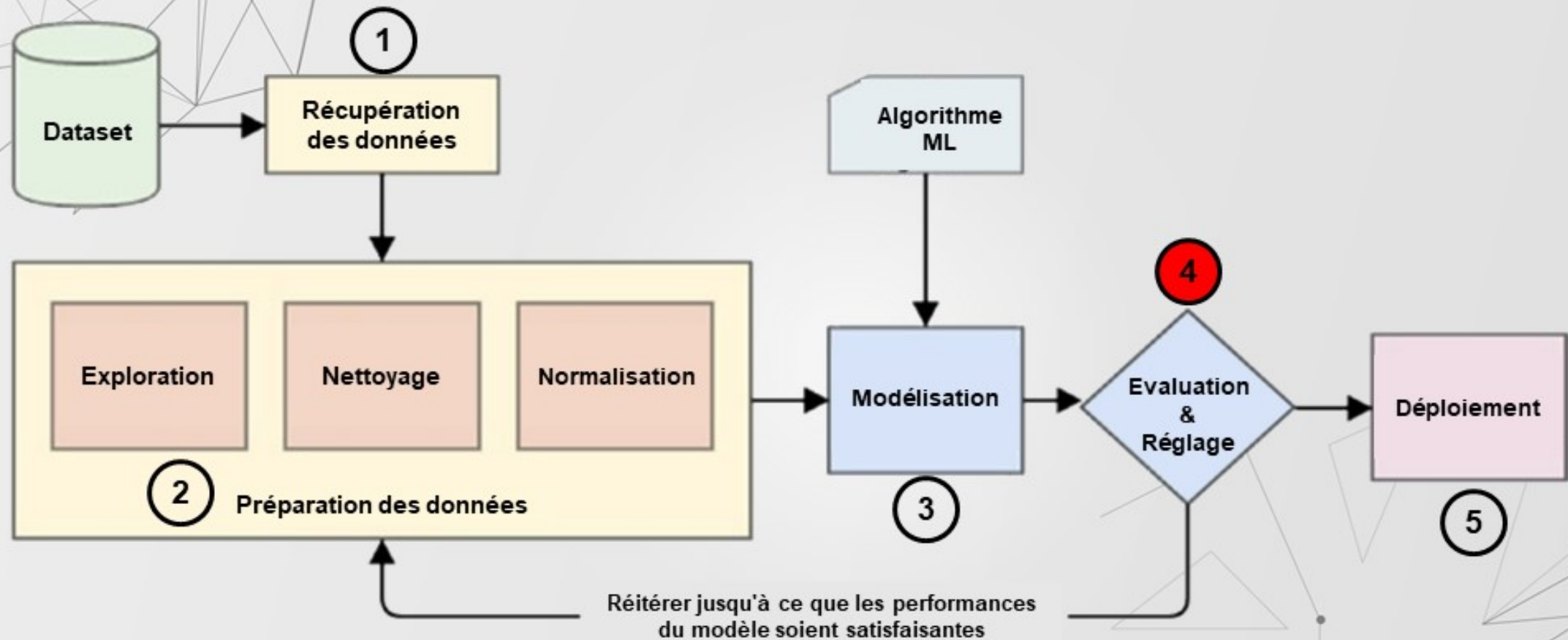
# Chaîne du processus ML

## Etape 3: Modélisation

Transmettre à l'algorithme ML les données divisées en :

- **Données d'entraînement**: sous-ensemble destiné à l'apprentissage d'un modèle.
- **Données de test** : sous-ensemble destiné à l'évaluation du modèle.  
Ce jeu de données ne doit en aucun cas être utilisé dans les données d'entraînement.

# Chaîne du processus ML

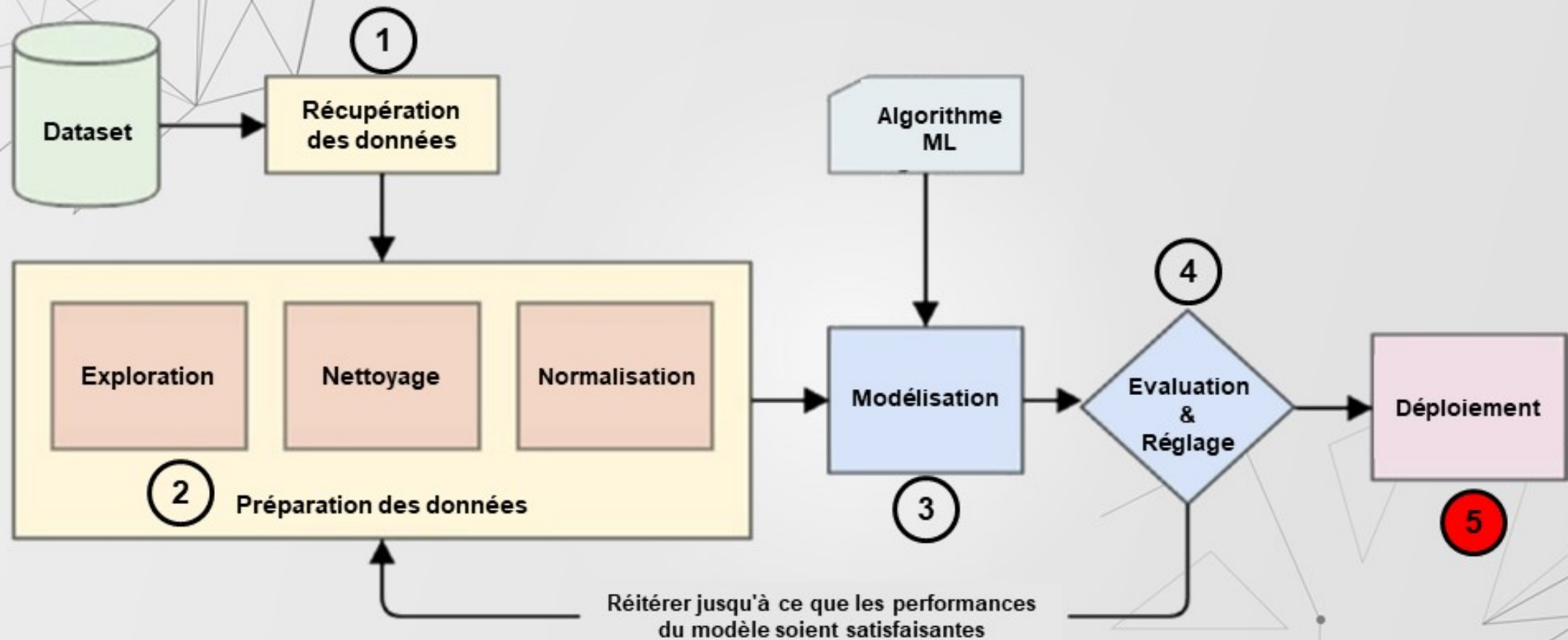


# Chaîne du processus ML

## Etape 4: Evaluation & Réglage

- **Validation du modèle:** Observer les performances du modèle sur de nouvelles données
- La mesure de performances s'effectue à travers un ensemble de **métriques** et selon le type du problème.
- Optimisation des hyperparamètres pour obtenir des modèles avec les meilleurs résultats.

# Chaîne du processus ML





# Chaîne du processus ML

## Etape 5: Déploiement

- **Intégrer la solution:** mettre en place une interface d' exécution
- Prendre en considération l'évolution des données sur lesquelles est basé l'apprentissage
- Surveiller en fonction des prédictions et des résultats
- Créer des stratégies business

# Terminologie

L'apprentissage artificiel apprend à produire à partir de données d'entrée des prédictions utiles sur des données encore jamais vues.

## Dataset

Un ensemble de données qui peuvent être traitées par un ordinateur comme une seule unité à des fins d'analyse et de prédiction. Ce Dataset inclut toujours 2 types de variables :

- Une ou plusieurs variables caractéristiques (**features**)  $x$
- Une variable objectif (**target**)  $y$

## Caractéristiques (features)

Une caractéristique est une variable d'entrée. Par exemple dans un détecteur de spam, les caractéristiques pourraient être les mots du mail, l'adresse de l'expéditeur, ...

# Terminologie

## Étiquettes (labels)

Une étiquette est la “chose” que nous essayons de prédire. Par exemple le prix d'une pizza, la race d'un animal sur une photo, l'auteur d'un texte, ...

## Exemple

Dans un mail, on distingue :

### 1. Exemples étiquetés :

Des mails déjà étiquetés comme spam/non spam.

### 2. Exemples non étiquetés :

Des mails non étiquetés

# Terminologie

## Modèle (Model)

Un modèle est le résultat d'un algorithme ML qui a été entraîné en lui fournissant un sur un ensemble de données, qu'il peut utiliser pour raisonner et apprendre de celles-ci. Un modèle définit la relation entre les caractéristiques et l'étiquette. Il est composé de 2 phases :

1. **L'entraînement (training)** : Le modèle apprend les relations entre les caractéristiques et l'objectif à partir d'exemples.
2. **L'inférence (test)** : consiste à appliquer le modèle entraîné sur des exemples non connus pour faire des prédictions.



# Terminologie

## L'entraînement (training)

L'entraînement consiste à trouver les meilleurs poids et biais possibles à partir des exemples connus. Par exemple, il s'agit de trouver un bon modèle qui permet de prédire le prix d'une pizza en fonction de sa taille, à partir des tailles et prix des pizzas que l'on connaît.

## La perte (loss)

Il existe une infinité de modèles, il faut pouvoir les comparer et mesurer leur efficacité. La perte est la pénalité obtenue pour les mauvaises prédictions. Pour un modèle parfait, la perte sera de 0. Un mauvais modèle aura une grande perte.