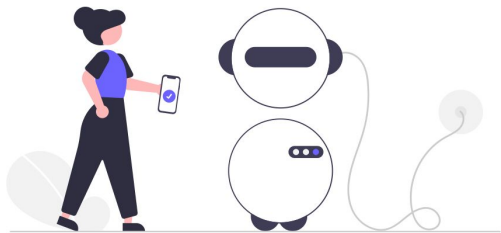
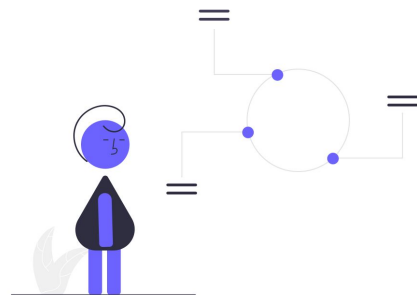


# Le vocabulaire dans le domaine de la Data



By Mahamadou Kaba



# La Data

La data est un terme anglais utilisé dans le secteur des télécommunications pour qualifier les **données** qui peuvent circuler par un réseau téléphonique ou un réseau informatique.

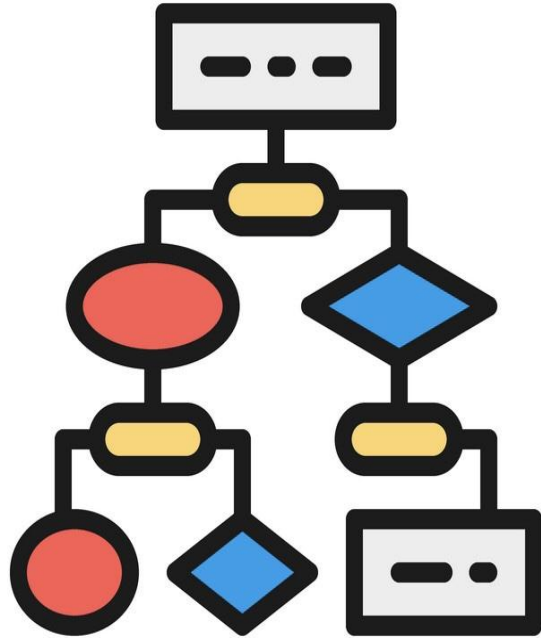


# DATA SCIENCE



La data science englobe la préparation des données pour l'analyse, y compris le nettoyage, l'agrégation et la manipulation des données pour effectuer une analyse avancée des données.



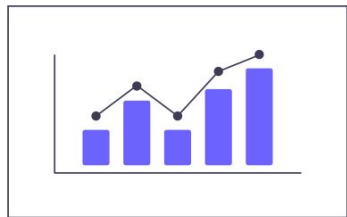


# Algorithme

un ensemble d'instructions que nous  
donnons à un ordinateur pour qu'il puisse  
prendre des valeurs et les manipuler sous  
une forme utilisable

# Big Data

Ils désignent un ensemble très volumineux de données qu'aucun outil classique de gestion de base de données ou de gestion de l'information ne peut vraiment travailler



# Base de données

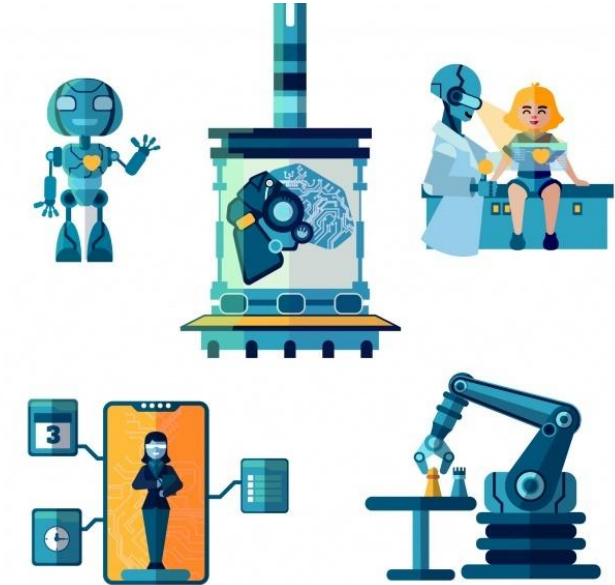
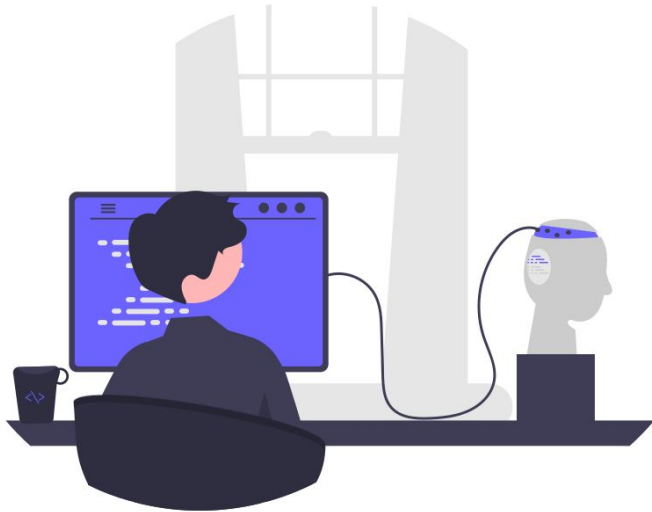


il s'agit d'un espace de stockage pour les données.

Nous utilisons principalement des bases de données avec un système de gestion de base de données (SGBD), comme PostgreSQL ou MySQL.

# Intelligence artificielle

Une discipline qui implique la recherche et le développement de machines conscientes de leur environnement

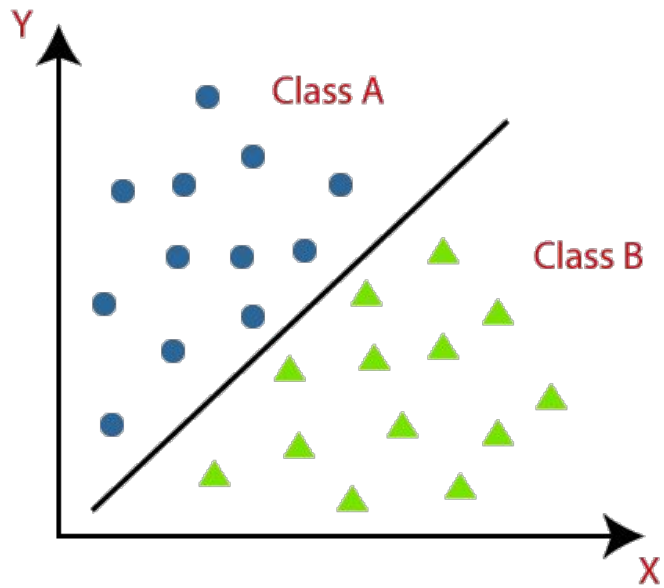


# Machine Learning

Processus par lequel un ordinateur utilise un algorithme pour comprendre un ensemble de données, puis effectue des prédictions en fonction de sa compréhension.

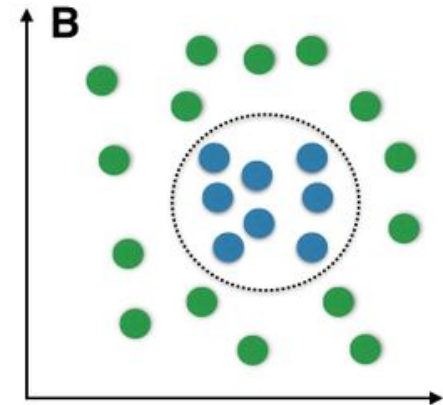
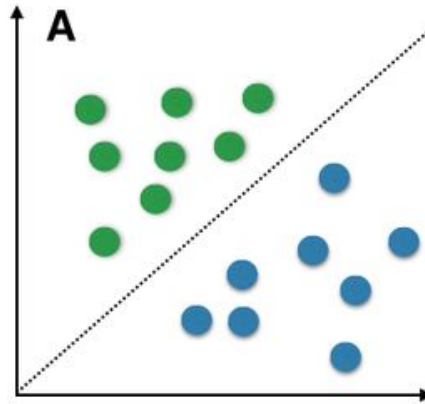






La classification est un problème d'apprentissage automatique supervisé. Il traite de la catégorisation d'un point de données en fonction de sa similarité avec d'autres points de données.

Classification

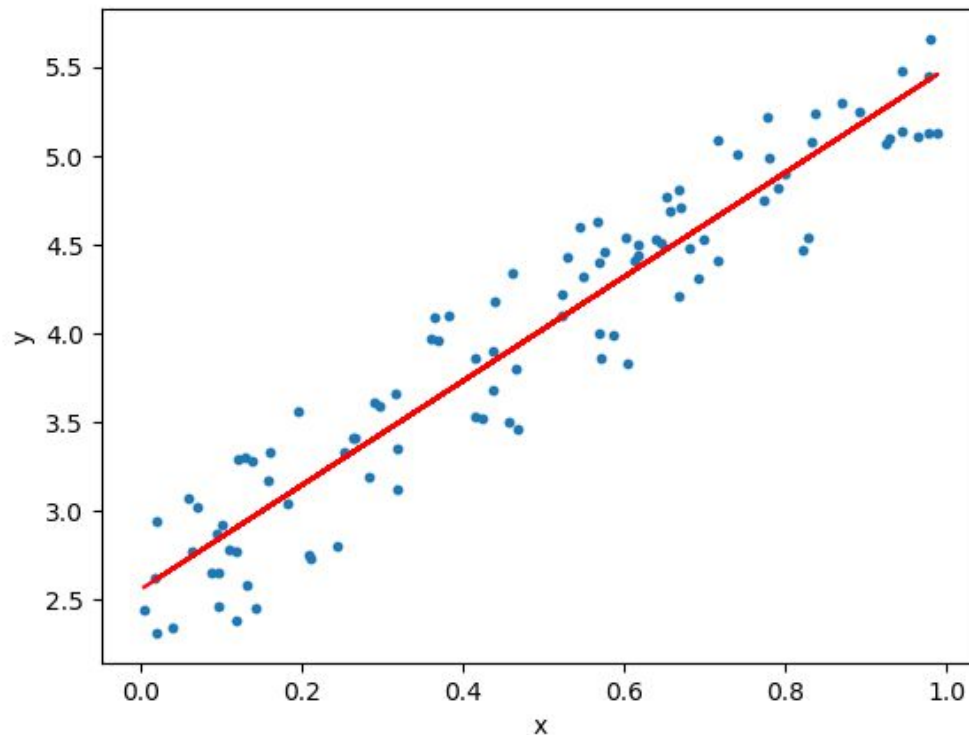


La régression est un autre problème d'apprentissage automatique supervisé. Il se concentre sur la façon dont une valeur cible change lorsque d'autres valeurs d'un jeu de données changent.

$$Y_i = \beta_0 + \beta_1 X_i$$

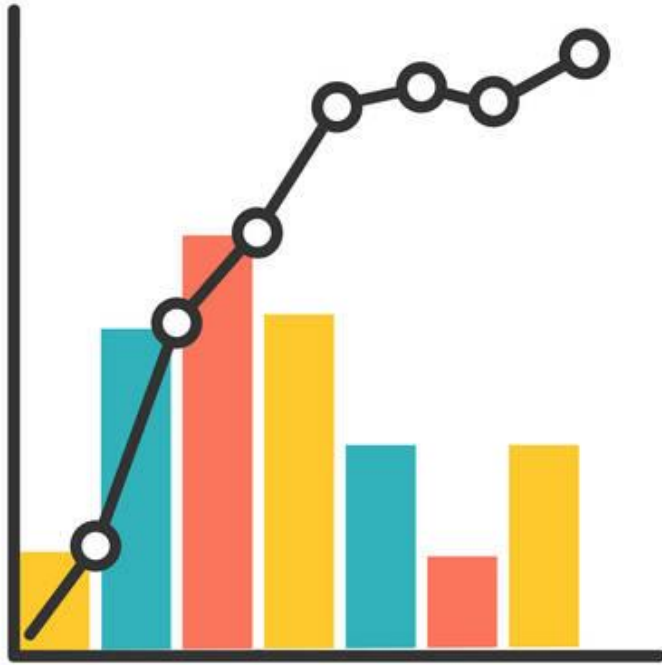
Diagram illustrating the components of the linear regression equation:

- $Y_i$  is the **Dependent Variable**.
- $\beta_0$  is the **Constant/Intercept**.
- $\beta_1$  is the **Slope/Coefficient**.
- $X_i$  is the **Independent Variable**.



Régression

# Corrélation

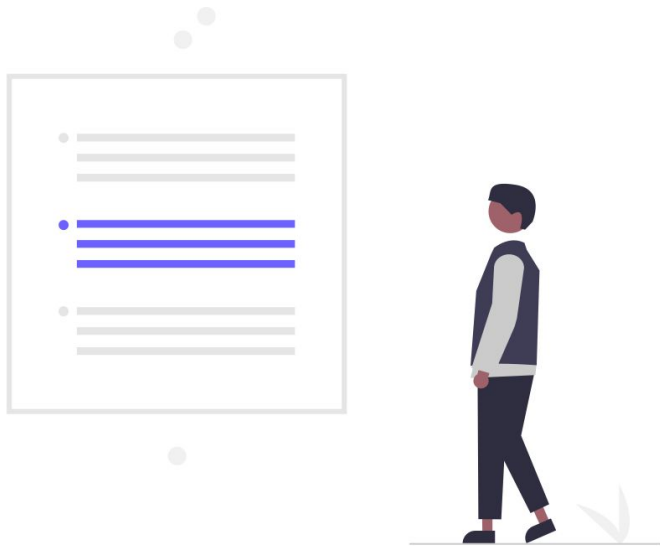


la mesure dans laquelle un ensemble de valeurs dépend d'une autre. Si les valeurs augmentent ensemble, elles sont positivement corrélées. Si l'une des valeurs d'un ensemble augmente à mesure que l'autre diminue, elles sont négativement corrélées.

$$\text{corr}(X, Y)$$
$$P_{X, Y}$$

# Moyenne

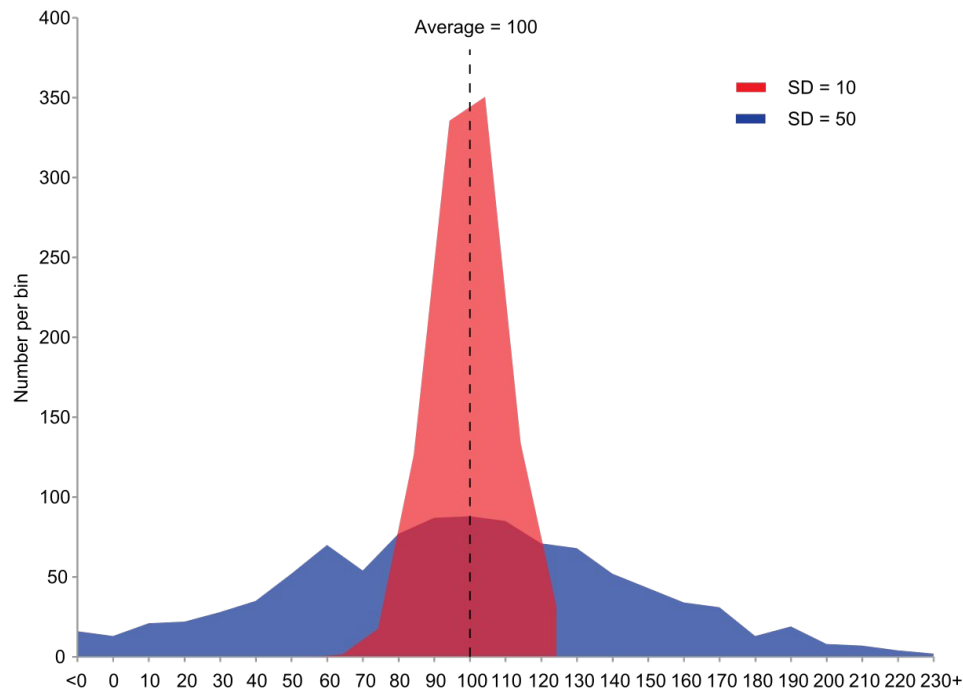
La moyenne est la somme d'une liste de valeurs divisée par le nombre de valeurs de cette liste.

 $\bar{x}$ 

# Variance

La variance d'un ensemble de valeurs mesure l'étendue de ces valeurs.

Mathématiquement, il s'agit de la différence moyenne entre les valeurs individuelles et la moyenne de l'ensemble des valeurs.



# Data Mining



- Processus consistant à extraire des informations exploitables d'un ensemble de données et à les utiliser à bon escient.
- nettoyage et l'organisation des données;
- l'analyse pour trouver des modèles et des relations significatives;
- communiquer ces connexions de manière à aider les décideurs à améliorer leur produit ou leur organisation.

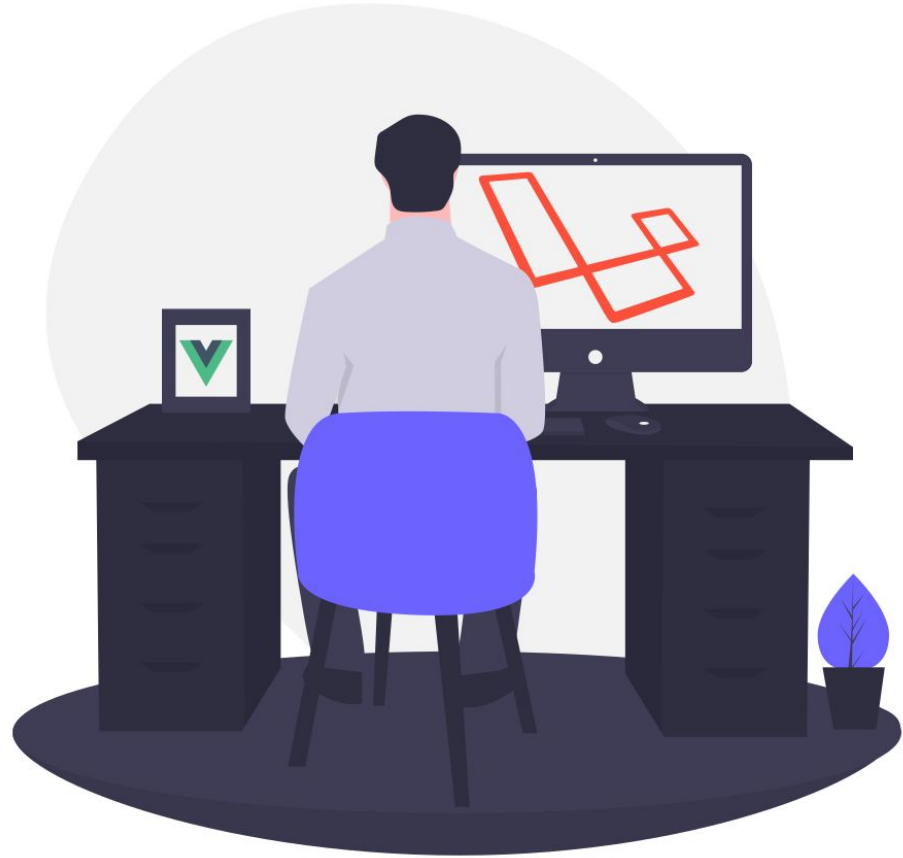
# Web Scraping

Le Web Scraping consiste à extraire des données du code source d'un site Web. Il s'agit généralement d'écrire un script qui identifiera les informations souhaitées par un utilisateur et les extraira dans un nouveau fichier pour une analyse ultérieure.



# Front End

Le Front End comprend tout ce qu'un client ou un utilisateur peut voir et avec lequel il interagit directement. Cela inclut les tableaux de bord de données, les pages Web et les formulaires.





# Back End



Le back-end est constitué de tout le code et de la technologie qui fonctionne en coulisse pour alimenter le frontal avec des informations utiles. Cela inclut les bases de données, les serveurs, les procédures d'authentification et bien plus encore.