

Résumé des cours de Data Mining

Introduction

Le Data Mining, ou exploration de données, est une discipline qui permet d'extraire des informations utiles et des modèles cachés dans des ensembles de données complexes. Ces cours introduisent les concepts clés, les outils, et les méthodes pratiques nécessaires pour travailler avec des données, tout en explorant des techniques comme les règles d'association utilisées dans de nombreux domaines, tels que le marketing, la finance ou les systèmes de recommandation.

Cours 1 : Introduction et Préparation des Données

1. Qu'est-ce que le Data Mining ?

- Processus d'analyse de grandes quantités de données pour identifier des relations et des tendances importantes.
- Permet de résoudre des problèmes comme la prédiction des ventes ou l'analyse des comportements des clients.

2. Importance de Python :

- Python est un outil essentiel en science des données grâce à :
 - Sa facilité d'utilisation et sa syntaxe intuitive.
 - Ses bibliothèques puissantes (Pandas, NumPy, Matplotlib) pour analyser et visualiser les données.
 - Son adoption massive dans les entreprises pour le traitement de données.

3. Collecte des Données :

- **Sources possibles :**
 - Fichiers : CSV, Excel, JSON, XML.
 - Bases de données SQL ou NoSQL.
 - Sites web et APIs.
 - **Outils principaux :**
 - **Pandas** : Pour lire, manipuler et analyser des fichiers.
 - **BeautifulSoup** : Pour extraire des données à partir de pages web.
 - **Requests** : Pour collecter des données via des APIs.
-

Cours 2 : Règles d'Association

1. Qu'est-ce qu'une règle d'association ?

- C'est une méthode pour découvrir des relations entre des éléments dans des transactions.
- Exemple : Si un client achète {Couches}, il est probable qu'il achète aussi {Bière}.

2. Concepts importants :

- **Itemset** : Ensemble d'items (par exemple {Pain, Lait}).
 - **Support** : Fréquence d'apparition d'un itemset dans les transactions.
 - Exemple : {Pain, Lait} apparaît dans 40% des transactions.
 - **Confiance** : Probabilité qu'un item soit acheté lorsqu'un autre l'est.
 - Exemple : Si {Pain} est acheté, il y a 80% de chances que {Lait} soit aussi acheté.
-

Conclusion

Ces cours permettent de comprendre les bases du Data Mining, notamment la collecte et la préparation des données, ainsi que les règles d'association, une technique clé pour analyser des transactions. Grâce à Python et ses bibliothèques, ces méthodes deviennent accessibles et applicables à des domaines variés comme le marketing, les prévisions ou les systèmes de recommandation. Ce savoir est essentiel dans un monde où les données jouent un rôle central dans la prise de décision.