

Mini-Projet : Implémentation des Algorithmes K-Means, KNN et Apriori

Pour les étudiants de L3 Data Science et FinTech

January 24, 2025

Introduction

Ce mini-projet s'inscrit dans le cadre de la validation des méthodes déjà vues en cours dans le module de Data Mining. Il a pour objectif de permettre aux étudiants de mettre en pratique leurs connaissances en programmation et en data science en implémentant trois algorithmes fondamentaux : K-Means, K-Nearest Neighbors (KNN) et Apriori. Les données utilisées dans ce projet seront basées sur les travaux dirigés (TD) corrigés à la main. Les étudiants devront comprendre les résultats, reproduire les calculs, et valider les implémentations à l'aide des données corrigées.

Objectifs pédagogiques

- Comprendre les principes fondamentaux des algorithmes K-Means, KNN et Apriori.
- Développer des compétences en programmation Python pour implémenter ces algorithmes.
- Valider les résultats en les comparant avec des solutions corrigées.
- Renforcer la capacité à analyser des résultats et optimiser des algorithmes.

Consignes du projet

1. Préparation des données

Les données des exercices corrigés doivent être enregistrées dans un sous-dossier du projet nommé data sous format CSV ou Excel. Les étudiants devront :

- Importer et analyser les données.
- Vérifier la cohérence des données en comparant avec les corrigés.
- Prétraiter les données si nécessaire (gestion des valeurs manquantes, normalisation, etc.).

2. Implémentation des algorithmes

Les étudiants doivent implémenter les algorithmes suivants :

- a. **K-Means** : Implémenter l'algorithme de regroupement pour partitionner les données en clusters. Visualiser les résultats sur des graphiques 2D ou 3D.
- b. **KNN** : Implémenter un classifieur basé sur les KNN pour prédire les classes des données test.
- c. **Apriori** : Implémenter l'algorithme pour l'extraction des règles d'association à partir des données transactionnelles fournies.

3. Validation et comparaison

- Les étudiants doivent comparer les résultats obtenus avec ceux des corrigés fournis.
- Rédiger un rapport détaillant les étapes suivies, les résultats obtenus et les éventuelles divergences.

Livrables

Les étudiants doivent fournir les éléments suivants :

1. **Code source** : Les implémentations des algorithmes en Python, bien documentées.
2. **Rapport** : Un document PDF contenant :
 - Une introduction expliquant les algorithmes utilisés.
 - Les étapes d'implémentation.
 - Les résultats obtenus et leur validation.
3. **Visualisations** : Graphiques et visualisations des résultats obtenus.

Évaluation

Le projet sera évalué sur les critères suivants :

- Qualité de l'implémentation des algorithmes.
- Cohérence et justesse des résultats obtenus.
- Clarté et structure du rapport.
- Créativité dans la visualisation et l'analyse des données.

Deadline

Le projet doit être soumis avant le **28/01/2025**. Tout retard entraînera une pénalité.

Annexes

Les annexes contiendront les corrigés des TD, les exemples de données et toute autre documentation utile pour le projet.

Bonne chance à tous !