**Détection et Analyse des Biais dans un Moteur de Recommandations : Plan d'Action**

**1. Préparation et Collecte des Données (Jours 1-2)**

**Actions :**

* Téléchargez le dataset MovieLens depuis Grouplens.
* Nettoyez les données avec Pandas :
  + Gérez les valeurs manquantes.
  + Normalisez les colonnes comme les genres, les utilisateurs et les notes.
* Structurez les données dans une base de données relationnelle (SQLite ou PostgreSQL).
* Visualisez un premier aperçu des données (distributions, corrélations) avec Matplotlib/Seaborn.

**Livrables :**

* Fichiers CSV ou base de données SQL structurée.
* Rapport sur les étapes de nettoyage et visualisation des données.

**2. Implémentation du Moteur de Recommandations (Jours 3-7)**

**Actions :**

* **Approche 1 : Filtrage collaboratif**
  + Implémentez l’algorithme k-plus proches voisins pour recommander des films similaires aux préférences utilisateur.
* **Approche 2 : Basé sur le contenu**
  + Utilisez des techniques NLP pour analyser les descriptions de films et recommander ceux avec des caractéristiques similaires.
  + Implémentez cette méthode avec Scikit-learn (TF-IDF ou CountVectorizer).

**Objectifs :**

* Comparez les performances en termes de précision, rappel, et F1-score entre les deux approches.
* Identifiez les forces et faiblesses des méthodes.

**Livrables :**

* Fichiers Python pour chaque approche.
* Résultats de comparaison des performances sous forme de tableau ou graphique.

**3. Analyse des Biais (Jours 8-10)**

**Actions :**

* Identifiez des biais possibles :
  + Films populaires sur-représentés.
  + Recommandations moins adaptées pour certains groupes d'utilisateurs (âge, genre).
* Effectuez une analyse statistique :
  + Testez les hypothèses sur les biais via des tests statistiques (par exemple, test de chi-carré).
* Implémentez des corrections :
  + Ajustez les pondérations dans le modèle.
  + Rebalancez les données (sur-échantillonnage ou sous-échantillonnage).

**Outils :**

* Seaborn, Matplotlib pour la visualisation des biais.
* Scikit-learn ou TensorFlow pour les corrections.

**Livrables :**

* Visualisations des biais.
* Résumé des tests et solutions appliquées.

**4. Visualisation et Interface Utilisateur (Jours 11-13)**

**Actions :**

* Développez une interface utilisateur avec **Streamlit** ou **Dash** pour :
  + Afficher les recommandations.
  + Présenter les analyses des biais sous forme de graphiques interactifs.
* Ajoutez des widgets permettant aux utilisateurs de :
  + Tester les recommandations pour différents profils.
  + Visualiser les biais avant/après correction.

**Livrables :**

* Application interactive hébergée localement ou en ligne (Heroku, Streamlit Sharing).

**5. Documentation et Mise en Valeur (Jour 14)**

**Actions :**

* Rédigez une documentation claire :
  + Expliquez le fonctionnement des algorithmes.
  + Ajoutez des captures d'écran de l'interface et des résultats.
* Rédigez un README structuré sur GitHub :
  + Incluez les étapes pour reproduire le projet.
* Préparez un post LinkedIn et un article pour partager le projet.

**Livrables :**

* Documentation complète.
* Projet publié sur GitHub avec un README professionnel.

**Ressources Recommandées**

1. **Dataset MovieLens :**  
   https://grouplens.org/datasets/movielens/
2. **Tutoriels :**
   * Filtrage collaboratif : [Medium - Collaborative Filtering](https://medium.com/).
   * Streamlit : Documentation officielle.
3. **Visualisation :**
   * Seaborn : Documentation.
   * Dash : Documentation.