

Catalogue De Données

MoviesForYou

AMIN ELFAQUIRI

December 13, 2023

1 Introduction

Ce rapport présente un catalogue de données contenant des informations sur les films, les utilisateurs, les évaluations de films, les genres et les professions. Le catalogue de données est essentiel pour comprendre les préférences des utilisateurs, explorer les évaluations de films, créer du contenu basé sur les genres et permettre le traitement des données en temps réel pour des recommandations dynamiques. Ce rapport fournit une description détaillée des différentes tables de données et présente les objectifs et les approches pour chaque cas d'utilisation.

2 Description des données

2.1 Movies (`u_item.txt`)

La table **Movies** contient des informations sur chaque film, notamment son titre, sa date de sortie et ses genres.

Champs :

- Movie ID
- Title
- Release Date
- Genres (Action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western)

2.2 Users (`u_user.txt`)

La table **Users** contient des informations démographiques sur chaque utilisateur, telles que l'âge, le sexe, la profession et le code postal.

Champs :

- User ID
- Age
- Gender
- Occupation
- Zip Code

2.3 Ratings (`u_data.txt`)

La table **Ratings** contient les notes attribuées par les utilisateurs aux films, ainsi que les horodatages.

Champs :

- User ID
- Movie ID
- Rating
- Timestamp

2.4 Genres (`u_genre.txt`)

La table **Genres** contient une liste de genres.

Champs :

- Genre ID
- Genre Name

2.5 Occupations (`u_occupation.txt`)

La table **Occupations** contient une liste de professions.

Champs :

- Occupation ID
- Occupation Name

3 Cas d'utilisation

3.1 Prétraitement des Données avec Apache Spark

Objectif : Charger, nettoyer et prétraiter les données de MoviesForYou pour une utilisation ultérieure.

Étapes de mise en œuvre :

- Charger l'ensemble de données MoviesForYou dans Spark.
- Nettoyer et prétraiter les données, notamment la gestion des valeurs manquantes, la conversion des types de données et la division du champ des genres en une liste pour une analyse facilitée dans Elasticsearch.
- Créer les caractéristiques nécessaires pour le système de recommandation, telles que la mesure de l'activité de l'utilisateur et les évaluations moyennes des films, en fusionnant les notes moyennes avec le DataFrame des films pour inclure les détails du film.

3.2 Construction du Modèle de Recommandation

Objectif : Sélectionner, entraîner et évaluer un modèle de recommandation approprié.

Étapes de mise en œuvre :

- Choisir un modèle approprié pour les recommandations de films, tel que le modèle ALS (Alternating Least Squares).
- Former le modèle en utilisant le jeu de données préparé.
- Évaluer les performances du modèle pour mesurer son efficacité.

3.3 Intégration des Données dans Elasticsearch

Objectif : Préparer et ingérer les données traitées dans Elasticsearch pour une recherche et une analyse efficaces.

Étapes de mise en œuvre :

- Définir et créer des indices Elasticsearch pour les données de films et d'utilisateurs.
- Transformer les données pour les rendre compatibles avec Elasticsearch.
- Ingestion des données traitées dans Elasticsearch pour une utilisation ultérieure.

3.4 Visualisation avec Kibana

Objectif : Créer des visualisations et des tableaux de bord dans Kibana basés sur les données Elasticsearch.

Étapes de mise en œuvre :

- Créer des visualisations et des tableaux de bord dans Kibana pour représenter les évaluations moyennes des films, les niveaux d'activité des utilisateurs, etc.
- Interpréter et expliquer les insights obtenus à partir des visualisations pour mieux comprendre les préférences des utilisateurs.

3.5 Développement de l'API Flask

Objectif : Développer une API qui permet aux utilisateurs de rechercher des recommandations de films en fournissant le titre du film.

Étapes de mise en œuvre :

- Rechercher les utilisateurs ayant interagi avec le film spécifié en utilisant l'API Flask.
- Utiliser les ID utilisateur pour générer des recommandations basées sur le modèle ALS.

4 Conclusion

Ce catalogue de données offre une vue complète des informations sur les films, les utilisateurs, les évaluations de films, les genres et les professions. Il permet d'explorer les préférences des utilisateurs, de créer du contenu basé sur les genres et de traiter les données en temps réel pour des recommandations dynamiques. En utilisant les différentes tables de données et les approches mentionnées, il est possible de mieux comprendre les préférences des utilisateurs, d'adapter les recommandations de contenu et de fournir des recommandations personnalisées basées sur les genres. Ce catalogue de données constitue une ressource précieuse pour les analyses et les prises de décision dans le domaine du divertissement et du marketing.