****

**Online Learning Completion**

**Submitted by:**

**Name: Khushi kathak**

**Roll No: 202401100300136**

**Course: Artificial Intelligence – Mid-Semester Exam**

**Submitted to:**

**Mr. Bikki Gupta**

**Department of Computer Science**

**Kiet Group Of Institution**

**Introduction:**

This project aims to analyze movie preferences based on user data to segment users into groups with similar tastes. By understanding these clusters, we can provide personalized movie recommendations and enhance user experience.

**Methodology:**

**Data Acquisition:** We utilized a dataset containing user data, including their watch time for movies, preferred genres, and average ratings given to movies.

1. **Data Preprocessing:** Genre preferences, initially in text form, were converted into numerical representations using Label Encoding. This step is crucial for applying machine learning algorithms.
2. **Clustering:** The K-Means clustering algorithm was employed to group users into clusters based on their watch time, encoded genre preferences, and average ratings. The number of clusters was set to 3, but this can be adjusted based on further analysis and desired granularity.
3. **Analysis:** The resulting clusters were examined to understand the distinct characteristics of each group, such as their preferred genres, watch time patterns, and rating tendencies.

CODE:

import pandas as pd

from sklearn.cluster import KMeans

from sklearn.preprocessing import LabelEncoder

# 1. Data Preparation

df = pd.read\_csv("/content/movie\_watch.csv")  # Replace with your actual file path

le = LabelEncoder()#using labelencoder() as le

df['genre\_encoded'] = le.fit\_transform(df['genre\_preference'])

features = df[['watch\_time\_hour', 'genre\_encoded', 'avg\_rating\_given']]

# 2. Clustering

kmeans = KMeans(n\_clusters=3)  # You can change the number of clusters here

kmeans.fit(features)

labels = kmeans.labels\_

# Assign cluster labels to the DataFrame

df['cluster'] = labels

# Print the DataFrame with cluster assignments

print(df[['watch\_time\_hour', 'genre\_encoded', 'avg\_rating\_given', 'cluster']])

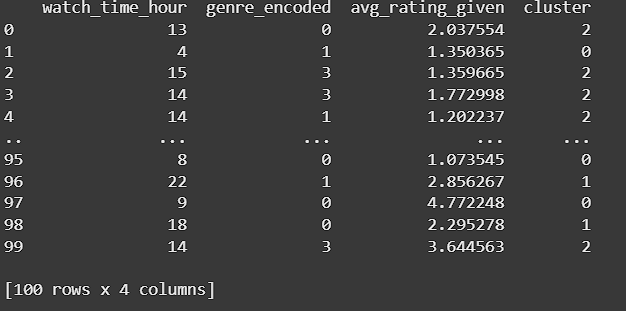
**Results and Output:**

The clustering process resulted in three distinct user groups with varying movie preferences. These clusters could represent, for instance, users who prefer action movies with high watch times and ratings, users who favor comedies with moderate watch times, or users who primarily watch documentaries with lower ratings. Further analysis of these clusters can reveal detailed insights into each group's preferences.

**Applications:**

The insights gained from this project can be applied to various aspects of movie recommendation systems, including:

* **Personalized Recommendations:** Tailoring movie suggestions for individual users based on their cluster membership.
* **Targeted Marketing:** Promoting specific movies or genres to relevant user segments.
* **Content Creation:** Identifying audience preferences to guide the development of new movies or content.



Reference:

# Datasets provided.