Eng: Huda Hemdan

Anime Recommendation System Document

1. Project Description

1.1. Introduction: The Anime Recommendation System project aims to provide personalized anime recommendations to users based on their viewing history and ratings. This system utilizes a dataset containing information about user preferences from 73,516 users on 12,294 anime titles.

1.2. Purpose: The main goal is to enhance user experience by offering tailored anime suggestions, helping users discover new shows and movies based on their past interactions and preferences.

1.3. Objectives:

- Clean and prepare the dataset.
- Perform exploratory data analysis (EDA).
- Visualize trends and insights.
- Build a recommendation model.
- Evaluate and interpret the results.

Eng: Huda Hemdan

2. Data Preparation

2.1. Import Required Libraries:
- Begin by importing essential libraries for data manipulation, visualization, and analysis, including pandas, numpy, matplotlib, seaborn, and others.

2.2. Load the Dataset:
- Use a file dialog to select and load the anime.csv and rating.csv files containing anime details and user ratings, respectively.

2.3. Read the Dataset:
- Read the datasets into pandas DataFrames for further processing and analysis.

2.4. Print Head Values and Info:
- Display the first few rows of each dataset and basic information about their structure and data types.

2.5. Print Statistical Data:
- Generate and print summary statistics for both datasets to understand the distribution and characteristics of the data.

2.6. Handle Missing Values and Duplicates:
- Drop any rows with null values and replace -1 ratings (indicating missing ratings) with NaN.
- Remove any duplicate entries to ensure data integrity.

2.7. Clean Anime Names:
- Clean up the anime names by removing unwanted characters and sequences (like " and &) to standardize the text.

2.8. Merge Dataframes:
- Merge the anime and rating DataFrames on the anime_id column, allowing for combined analysis of anime details and user ratings.

3. Exploratory Data Analysis (EDA)

3.1. Top Animes by Members:
- Identify and visualize the top anime titles based on the number of members (users who have interacted with the anime) to understand popular trends.

3.2. Distribution of Anime Types:
- Analyze and visualize the distribution of different types of anime (e.g., TV, Movie, OVA) to see which categories are most common.

3.3. Visualize Average Ratings and User Ratings:
- Create histograms to visualize the distribution of average ratings and user ratings for different types of anime. This helps identify patterns in how different categories are rated.

4. Genre Analysis

4.1. Analyze Genres:
- Break down the genres of anime titles, splitting multiple genres into individual entries. Analyze the frequency of each genre to see which are most popular.

4.2. Generate Word Cloud for Genres:
- Create a word cloud to visually represent the popularity of different genres based on their frequency in the dataset. This provides an intuitive way to grasp genre trends.

5. Interpretation of Results

5.1. Insights:
- TV and Movie categories tend to have higher average and user ratings, indicating their popularity and potentially higher production quality.
- Popular genres include Mystery, Police, Psychological, and Supernatural, which can guide content recommendations.
- Focused recommendations on TV and Movie categories are likely to be more accurate and appreciated by users.

5.2. Conclusion: The Anime Recommendation System effectively identifies trends in anime ratings and genres, offering valuable insights for providing personalized recommendations. Future work could involve building a collaborative filtering model to enhance recommendation accuracy based on user behavior and preferences.