MOVIE RECOMMENDATION ENGINE

MOHAMMAD GULSAR ROSHAN K

AVINESH K K

SURYANANDAN K V

Introduction

-Project Overview: Introduce the concept of a Movie Recommendation Engine that suggests the best movies to users.

Data Preparation

- Library Import: Highlight the use of Python libraries such as Pandas, Numpy, Matplotlib, and Seaborn.

- Dataset Loading: Explain how the dataset is loaded using the Pandas library.

Data Analysis

- Unique Genres: Describe the process of finding unique movie genres within the dataset.

- Unique Ratings: Outline the method to identify all unique ratings of movies.

- Unique Titles: Mention how to find all unique movie titles.

- Visualization: Discuss the creation of a countplot for user ratings and a pie plot for movie ratings distribution.

Data Grouping

- User Behavior: Group data to show each user's viewed movies and their corresponding ratings.

Data Prediction

- Rating Prediction: Predict the percentage for each rating category.

Genre Analysis

- Genre Separation: Use the "set" data structure to eliminate duplication of genres

- Data Structuring: Create a data structure that matches movie titles with their genres.

Statistical Analysis

- Average Counts: Present statistics on movie counts based on average ratings.

Data Presentation

- Title Display: Show how to arrange and display movie titles in order and in reverse order.

Rating Filter

- High Ratings: Filter and display movies with ratings greater than or equal to 4.

Recommendation Engine

- Algorithm Overview: Introduce the Apriori algorithm and other data mining techniques like Decision Tree, Naive Bayes, KNN.

- Data Transformation: Explain the conversion of rows into columns and vice versa for the recommendation process.

- Recommendation Process: Detail how movies are recommended based on 'antecedents' and 'consequents'.

Metrics Explanation

- Confidence: Define confidence and its role in the recommendation strength.

- Lift: Explain the lift metric and its significance in likelihood estimation.

- Support: Clarify the concept of support as a measure of item popularity.

- Conviction: Discuss conviction as a comparison of dependent and actual frequencies of item appearances.

- Threshold: Describe the threshold setting in the Apriori algorithm for identifying significant item sets.