**Tag based Collaborating Filtering Movie Recommendation System**

* + - * Need to show Tag based collaborative filtering is better than Method 1
* PPT must be share of working done & related document of theory used. Coding explaination need to required in document.

**Description:**

**Method1:**

1. **Input:** The dataset, **Movie lens Dataset** is implemented as input. The input dataset is taken from dataset repository. The dataset is in the format ‘.csv’ or ‘.xlsx’ (<https://www.kaggle.com/datasets/grouplens/movielens-20m-dataset?select=rating.csv>)
2. **Pre-processing:** The collected input data’s are subjected to pre-processing. In the Pre-processing step,

* Can handle the missing data
* Can perform label encoding.

1. **Data Splitting**:In this step, the pre-processed data’s are split into Train set (70%) and test set (30%) for decision

* **Train data** is used for evaluate the model.
* **Test data** is used for predict the model.

1. **Recommendation**: In this step, we can recommend the movie to users based on their **user input ratings and ratings present in movielens dataset**.

* Collaborative filtering algorithm----calculate the recommendation score, Movie name ,Genre (User-User Collaborative Filtering)
* **Pearson correlation coefficient -** finding similarity between users.

1. **Classification**: In this step, we can implement the different machine learning algorithms.

* Hybrid RF + DT

1. **Performance metrics**: In this step, we can analyse some performance metrics such as

* Accuracy
* Precision
* Recall
* F1-score.
* MAP
* RMSE

1. **Numerical & graph result**: graph of accuracy precision recall F1, MAP, RMSE need to plot. And merge graph of two method in terms of accuracy precision recall F1, MAP, RMSE. Accuracy table of two method need to display.

**Method2:**

**1.Input:** The dataset, **Movie lens Dataset** is implemented as input. The input dataset is taken from dataset repository. The dataset is in the format ‘.csv’ or ‘.xlsx’

**2.Pre-processing:** The collected input data’s are subjected to pre-processing. In the Pre-processing step,

* Can handle the missing data
* Can perform label encoding.

1. **Data Splitting**:In this step, the pre-processed data’s are split into Train set (70%) and test set (30%) for decision

* **Train data** is used for evaluate the model.
* **Test data** is used for predict the model.

1. **Recommendation**: In this step, we can recommend the movie to users based on their **user input tags and tags present in movielens dataset**.

* Using movie lens data set collaborative filtering need to done by user input tag and already tag present in movie lens dataset for that movie this will be done by wordnet library distance value need to calculate after that collaborative filtering.
* Distance value need to display
* Recommended movie name & genre ,recommendation score .

1. **Classification**: In this step, we can implement the different machine learning algorithms.

* Hybrid RF + DT

1. **Performance metrics**: In this step, we can analyse some performance metrics such as

* Accuracy
* Precision
* Recall
* F1-score.
* Pearson correlation coefficient
* MAP
* RMSE

1. **Numerical & graph result:** Graph of accuracy precision recall F1, MAP, RMSE need to plot. And merge graph of two method in terms of accuracy precision recall F1, MAP, RMSE. Accuracy table of two method need to display.

**Note:**

This is the document for confirmation of the working procedure.

The process will be implemented as per the description given above.

The Dataset will be provide by us.

Not a real time project.

No GUI is provided.

**Language**: Python

**Front** **End**: Anaconda Navigator – **Spyder IDE**