**IDS ASSIGNMENT - 3**

**GROUP - 157 - Pothula Maruthi Chowdary , Ajit Mohan Pattanayak , Jayadhar Alla**

**Objective :**

Predicting Top 5 Movies for a User based on his Movie Watching History

**Dataset given from here > Data** : https://www.kaggle.com/netflix-inc/netflix-prize-data

Our Group has 2 files for the above requirement .

As per the question we have calculated Performance with the 2 Classifiers :   
  
( Logistic Regression and Decision Tree ) - file **: IDS\_Assignment3\_G157.pynb**

**Note : I kept all the required input files and the pynb files in the same directory so you will see direct file path rather than absolute path in our Program.**

**combined\_data\_1.txt** and **movie\_titles.csv** are considered as inputs .

Input files has different type of data , Ordinal, Categorical etc..

Feature Engineering Techniques used : **Handling Outliers, Binning , Scaling**

> As part of Preprocessing and Data Preparation to handle the Empty Rows and missing values properly .

> Joint point for number of Ratings and histogram considering the Ratings feature are plotted and shown .

Based on the Requirement few of the features were removed and also few were modified .

> Rating is modified as Rating Category and used as the target Variable .

> Proper segregation is done for the Test and train splits .

LOGISTIC REGRESSION is used and the Predict probabilities are calculated for the provided Data model .

DECISOIN TREE : This classifier is used for the same data , tree model fit is performed on it and predicted for the Test set.

**Note : To display the Decision tree in the pynb file, We have installed graphviz, condo-forge and pydotplus by following commands from terminal :  
  
conda install graphviz , conda install -c condo-forge pydotplus**

After Predicting and calculating the tree\_model.score , by using the modules mentioned in the above note , we have shown the DECISION TREE .

After that Classification report is also generated to know the Precision Recall F1-score and Support and shown as output .

######################################################

2. File name : **IDS\_G157\_assignment.ipynb**

This is python file which is for recommending the Movies using the Pearson Correlation .

Similar to the first file , We have followed all the required flow to prepare the data and performed required changes to get the recommendation top 5 movies base on the input .

All the methods are executed and outputs are shown.

Thanks

Best Regards

GROUP 157