Student 1 / Mohamed Gomaa Mostafa

Id/2103129

Student2/ Mazen Ahmed Eldeeb

Id/2103132

Student3/Omar Khaled Ahmed

Id/ 2103113

Student4/Aly Gamal Mohamed

Id/ 2103128

Project Machine Learning

Problem 1:

1-First we face a problem that we found the first column has data type object so we turned it into numbers Using a built in function to

2-After that we split our data into x features and y label

3-After that we split the data into training data with 80 % and testing data take the rest

4-And we used the linear regression to predict some value in our data

5-After that we train our model with the training data finally we can use it to predict values in the data

Problem 1 link:

<https://colab.research.google.com/drive/14BogOfA_LC8E8aZ30lAhmweZF-iAO-F_>

Problem 2:

**First we read our data**

2-we face a problem that we have feature that is not important and it will effect our result so we check the data and found that the id column is not important

3-we have 14 feature and we can handle this in 2 methods

**First method:**

We apply k-means on our data and the number of clusters is (2 to classify that if he has heart disease or not ) then to plot it we used reduction dimension algorithm PCA ,T-SNE

**Second method:**

We apply PCA to reduce 14 features into 2 feature and we take the result and apply it in k-Means and we plot it

**After that:**

we reduce the dimension of feature from 14 features to 2 features by using PCA and plot it

we reduce the dimension of feature from 14 features to 2 features by using T\_SNE and plot it

Problem 2 link:

<https://colab.research.google.com/drive/1pmsT5tAqaljErlLzCQete439X675JdOU>

problem 3 :

1-first we upload the data and we have movie data and rating data

2-we look at the data and we found that it need to be cleaned by removing some columns

3-we will clean the data by using iloc function that function take the name of the columns that we need and we need movie id and title to show only from the movie data

4-we did the same in the rating data and what we need is user id, movie id and rating onlythen we merge 2 data frame (movie,rating) in one data frame and called It data

5- after that we make a pivot table in order to make rows are users and columns are movies. And values are the rating

6- in our model we well find correlation between "Bad Boys (1995)" and other movies then we will recommend that movies to the user

Problem 3 link:

<https://colab.research.google.com/drive/1atLcIsy5LtOu8Gl1ISovbmPWiRVNsf_E>

**references :**

* [**https://pandas.pydata.org/**](https://pandas.pydata.org/)
* [**https://datascience.stackexchange.com/questions/48693/perform-k-means-clustering-over-multiple-columns**](https://datascience.stackexchange.com/questions/48693/perform-k-means-clustering-over-multiple-columns)
* [**https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html**](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html)
* [**https://www.kaggle.com/code/aussie84/clustering-with-kmeans-pca-tsne#Profiling-each-of-the-cluster**](https://www.kaggle.com/code/aussie84/clustering-with-kmeans-pca-tsne#Profiling-each-of-the-cluster)
* [**https://github.com/codebasics/py/blob/master/ML/13\_kmeans/13\_kmeans\_tutorial.ipynb**](https://github.com/codebasics/py/blob/master/ML/13_kmeans/13_kmeans_tutorial.ipynb)
* <https://github.com/SeoK106/knn_and_kmeans_with_heart-disease-dataset/blob/4c844a655136aa719a8ac8ffe798e2baf1f669de/kmeans_with_heart_disease_dataset.py#L34>
* <https://youtu.be/20fbgWm5M2o>
* <https://youtu.be/epI9W3MZ3Ts>