Practical No. 02

Aim : To compute similarity and dissimilarity measures for a given dataset using Euclidean distance and Jaccard similarity.

Software Used : IDLE

Theory :

Dataset: File name (customers.csv)

id,age,income,gender

1,25,50000,Male

2,30,60000,Female

3,22,45000,Male

4,35,80000,Female

5,28,52000,Male

6,40,70000,Female

7,50,90000,Male

8,45,85000,Female

9,33,62000,Male

10,27,48000,Female

Code :

```
import pandas as pd
from sklearn.metrics import pairwise_distances
from sklearn.preprocessing import LabelEncoder
import numpy as np
# Load dataset
df = pd.read_csv('customers.csv')
```

```
# Convert categorical data to numerical
le = LabelEncoder()
df['gender'] = le.fit_transform(df['gender'])  # Male: 1, Female: 0

# Compute Euclidean distance
numeric_data = df[['age', 'income', 'gender']].values
euclidean_distances = pairwise_distances(numeric_data, metric='euclidean')

print("Euclidean Distances:")
print(euclidean_distances)

# Compute Jaccard similarity for categorical data (gender)
gender_data = df[['gender']].values
jaccard_similarity = 1 - pairwise_distances(gender_data, metric='jaccard')

print("Jaccard Similarity:")
print(jaccard_similarity)

# Dissimilarity matrix (1 - similarity)
jaccard_dissimilarity = 1 - jaccard_similarity

print("Jaccard Dissimilarity:")
print(jaccard_dissimilarity)
```

Result :


```
Warning (from warnings module):
  File "D:\python 3.11\Lib\site-packages\sklearn\metrics\pairwise.py", line 2361
    warnings.warn(msg, DataConversionWarning)
DataConversionWarning: Data was converted to boolean for metric jaccard
Jaccard Similarity:
[[1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]
 [0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]
 [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]
 [0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]
 [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]
 [0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]
 [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]
 [0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]
 [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]
 [0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]]
Jaccard Dissimilarity:
[[0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]
 [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]
 [0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]
 [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]
 [0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]
 [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]
 [0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]
 [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]
 [0. 1. 0. 1. 0. 1. 0. 1. 0. 1.]
 [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]]
```

Conclusion : In this practical , We have performed to compute similarity and dissimilarity measures for a given dataset using Euclidean distance and Jaccard similarity.