Unit 3.1 Assignment Label Encoder

Peer Members:

• Syed Muhammad Raqim Ali Shah (2303.KHI.DEG.008)

• Maaz Javaid Siddique (2303.KHI.DEG.004)

Question:

Implement a label encoder for categorical data
using pure Python, Pandas and NumPy.

Solution:

```python
Encoder.py > ...
1   import pandas as pd
2   data=pd.read_csv('Final.csv')
3   def encode_data(data,column):
4       label={}
5       for col in column:
6           uni=data[col].unique()
7           label[col]={uni[x]:x for x in range(len(uni))}
8       new_data=data.replace(label)
9       return new_data
10  column=data.select_dtypes(['object']).columns.tolist()
11  print(encode_data(data,column))
```

First imported the necessary libraries, "pandas" for loading the dataset.

The "encode_data" function takes two arguments, "data" and "column". "data" is a pandas DataFrame object that contains the data to be encoded. "column" is a list of column names that contain object data types.

And as "Final.csv" file is used as dataset, so list "column" will contain "Entity" and "Code" as both are object data types.

The function first creates an empty dictionary called "label" to store the label encodings for each column. Then loops through each column in the "column" list and gets the unique values of that column using the "unique()" function. Then creates a dictionary of label encodings for that column, where each unique value is mapped to its corresponding label encoding. The dictionary is stored in the "label" dictionary using the column name as the key.

The function then uses the "replace()" function to replace the values in the input data with their corresponding label encodings. The "replace()" function takes the label dictionary as input and replaces each unique value in the data with its corresponding label encoding.

Finally, the function returns the encoded data.

Output:

```
muhammadraqim@all-MS-7D35:~/Documents/python$ cd /home/muhammadraqim/Documents/python ; /usr/bin/env /bin/python3 /home/muhamm
023.6.1/pythonFiles/lib/python/debugpy/adapter/../../debugpy/launcher 37909 -- /home/muhammadraqim/Documents/python/Encoder.py
      Unnamed: 0  Entity  Code  Year  Cellular Subscription  Internet Users(%)  No. of Internet Users  Broadband Subscription
0            0       0     0  1980               0.000000           0.000000                      0                0.000000
1            1       0     0  1981               0.000000           0.000000                      0                0.000000
2            2       0     0  1982               0.000000           0.000000                      0                0.000000
3            3       0     0  1983               0.000000           0.000000                      0                0.000000
4            4       0     0  1984               0.000000           0.000000                      0                0.000000
...        ...     ...   ...   ...                    ...                ...                    ...                     ...
8862      8862     228   215  2016              91.793457          23.119989                3341464                1.217633
8863      8863     228   215  2017              98.985077          24.400000                3599269                1.315694
8864      8864     228   215  2018              89.404869          25.000000                3763048                1.406322
8865      8865     228   215  2019              90.102287          25.100000                3854006                1.395818
8866      8866     228   215  2020              88.755806          29.299999                4591211                1.368916

[8867 rows x 8 columns]
```