Name: Jaydeep Solanki

Roll: 22BEC059

Practical 5: Splitting the data frame using various methods (BostonHousing.csv)

Code: Using Permutation

```
import math
import time

import numpy as np
import pandas as pd
st = time.process_time()
data = pd.read_csv("BostonHousing.csv", header='infer').values

x = data[:, 0:-1]
y = data[:, -1]
n_rows = data.shape[0]
print("Total Rows:", n_rows)

test_split = float(input("Enter a number between 0 and 1 to specify the test split:"))
n_rows_train = math.floor((1 - test_split) * n_rows)
all_indices = np.random.permutation(n_rows)
x_train = x[all_indices[0:n_rows_train], :]
y_train = y[all_indices[0:n_rows_train]]
x_test = x[all_indices[n_rows_train:], :]
y_test = y[all_indices[n_rows_train:]]

print("Shapes:", x_train.shape, y_train.shape, x_test.shape, y_test.shape)
print("Union:",
len(set(all_indices[0:n_rows_train]).union(all_indices[n_rows_train:])))
print("Intersection:",
len(set(all_indices[0:n_rows_train]).intersection(all_indices[n_rows_train:])))

et = time.process_time()
time_taken_ms = (et - st) * 1000
print("Time Taken", time_taken_ms)
```

Output:

```
PS C:\Users\JaySs\OneDrive\Desktop\Lab Works\AI ML Classs\Lab 5> & "C:/Program Files/Python310/python.exe" "c:/Users/JaySs/OneDrive/Desktop/Lab Works/AI M
L Classs/Lab 5/PermutationMethod.py"
Total Rows: 506
Enter a number between 0 and 1 to specify the test split:0.8
Shapes: (101, 13) (101,) (405, 13) (405,)
PS C:\Users\JaySs\OneDrive\Desktop\Lab Works\AI ML Classs\Lab 5>
```

Code: Using Sklearn library

```python
import time

import pandas as pd
from sklearn.model_selection import train_test_split

st = time.process_time()
data = pd.read_csv("BostonHousing.csv", header='infer').values

x = data[:, 0:-1]
y = data[:, -1]
n_rows = data.shape[0]
print("Total Rows:", n_rows)
test_split = float(input("Enter a value split in terms of % : "))
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=test_split / 100)
print(x_train, "\n", "===" * 30, "\n", y_train)
print("Shapes:", x_train.shape, y_train.shape, x_test.shape, y_test.shape)
et = time.process_time()
print(et - st)
```

Output:

```
PS C:\Users\JaySs\OneDrive\Desktop\Lab Works\AI ML Classs\Lab 5> & "C:/Program Files/Python310/python.exe" "c:/Users/JaySs/OneDrive/Desktop/Lab Works/AI M
L Classs/Lab 5/Sklearn.py"
Total Rows: 506
Enter a value split in terms of % : 80
Shapes: (101, 13) (101,) (405, 13) (405,)
PS C:\Users\JaySs\OneDrive\Desktop\Lab Works\AI ML Classs\Lab 5>
```

Code: Using Random Library

```
import random
import numpy as np
import pandas as pd
df = pd.read_csv("BostonHousing.csv", header="infer").values.tolist()
df2 = []
value_split = int(input("Enter the split value in terms of % :"))
indexes=[]
for i in range(int(len(df) * (value_split / 100))):
    rand_value = random.choice(df)
    indexes.append(df.index(rand_value))
    df.remove(rand_value)
    df2.append(rand_value)
df = np.array(df)
df2 = np.array(df2)
print(df)
print("Shape : ", df.shape)
print("----" * 20)
print(df2)
print("Shape :", df2.shape)
print("----" * 20)
print(indexes)
```