**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:])))

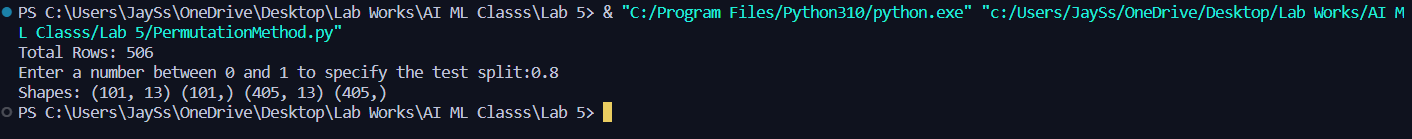
print("Intersection:", len(set(all\_indices[0:n\_rows\_train]).intersection(all\_indices[n\_rows\_train:])))

et = time.process\_time()

time\_taken\_ms = (et - st) \* 1000

print("Time Taken", time\_taken\_ms)

Output:



Code: Using Sklearn library

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)

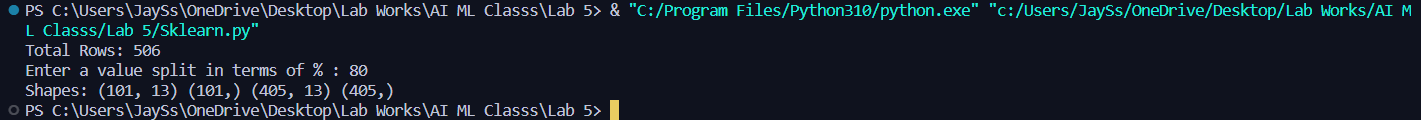
print(x\_train, "\n", "===" \* 30, "\n", y\_train)

print("Shapes:", x\_train.shape, y\_train.shape, x\_test.shape, y\_test.shape)

et = time.process\_time()

print(et - st)

Output:



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)