Experiment - 1 Data Preprocessing

Code

```
import numpy as np
import pandas as pd
from sklearn.impute import SimpleImputer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# Importing the dataset
data_set = pd.read_csv('Dataset.csv', delimiter=',')

# Ensuring all values are numeric and handling incorrect formatting
data_set = data_set.apply(pd.to_numeric, errors='coerce')

# Extracting independent and dependent variables
x = data_set.iloc[:, 1:].values  # Excluding the 'User' column
y = data_set.iloc[:, 0].values  # Keeping 'User' as the dependent variable

# Handling missing data (Replacing missing data with the mean value)
imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
imputer.fit(x)
x = imputer.transform(x)

# Splitting the dataset into the Training set and Test set
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=0)

# Feature Scaling
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)

# Displaying preprocessed data
print("Training data after preprocessing:")
print(x_train[:2])  # Display first 5 rows of processed training data
print("Test data after preprocessing:")
print(x_test[:2])  # Display first 5 rows of processed test data
```

Output

```
PS V:\Deeptanshu Lal\PROJECTS\ML\exp-1> python .\exp-1.py
Training data after preprocessing:
[[-8.26053994e-01 -1.07505153e+00 -7.62096340e-01 -5.24596609e-01
   2.49076318e-01  1.62362277e+00  1.15848016e+00  1.41007834e+00
   1.38276733e+00 -1.02022922e+00 -7.77474412e-01 -4.77520229e-01
   3.23371143e-01  1.75644686e+00  1.62827432e+00 -4.33408312e-01
  -2.65508399e-01 -1.30408487e-01 -1.95430238e-01  5.31193506e-02
```

```
  -2.15603792e-01 -6.29084361e-01 -6.04599981e-01 -6.93245573e-01
  -1.01835375e-11]
 [ 1.31358641e+00  2.28498220e-01 -2.51199243e-01 -2.34719881e-01
  -8.11818626e-01 -1.18899523e+00 -1.43257430e+00 -1.09938813e+00
  -1.32370769e+00 -9.43534790e-01 -6.61477890e-01 -1.02414282e+00
  -9.57272685e-01 -8.98025222e-01  1.62827432e+00 -2.59947081e-01
  -6.92394926e-02  6.95018481e-02 -1.37026766e-01 -1.79607780e-01
   4.38227393e+00  1.93072829e+00  1.57456141e+00  2.91627801e+00
  -1.01835375e-11]]
Test data after preprocessing:
[[-1.73450467e+00 -1.63170250e+00 -8.09992942e-01 -1.04332339e+00
  -1.15548882e+00 -1.14237725e+00 -1.33346293e+00 -9.54958403e-01
  -3.10628240e-01 -7.68877464e-02  5.63735379e-01  6.71179414e-01
   2.02853779e+00  1.75644686e+00  1.62827432e+00 -3.60372004e-01
  -1.72538917e-01 -2.51925213e-02 -3.69065282e-02 -8.36230757e-01
  -1.03975169e+00 -1.09109933e+00 -1.15504559e+00 -1.32025202e+00
  -1.01835375e-11]
 [-8.49494949e-02 -4.83169480e-01 -6.82268668e-01 -8.60243347e-01
  -9.53769358e-01  1.62362277e+00 -6.62532338e-02  5.79607425e-01
   1.38276733e+00  6.82387097e-01  1.77444908e+00 -6.55727705e-02
  -5.61051390e-01 -5.52063220e-01 -5.20259412e-01  1.14363996e-01
  -2.55178456e-01 -2.88232436e-01 -3.62297301e-01 -4.04023228e-01
  -4.75861022e-01 -6.85275370e-01 -4.68873666e-01 -1.67915849e-01
  -1.01835375e-11]]
```