Experiment-6

6. AIM: Write a program to implement Categorical Encoding, One-hot Encoding.

Program for demonstration of one hot encoding:

```
import numpy as np

import pandas as pd

# Create a sample dataset

data = {

'Employee_ID': [1, 2, 3, 4, 5],

'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eva'],

'Gender': ['Female', 'Male', 'Male', 'Male', 'Female'],

'Remarks': ['Excellent', 'Good', 'Average', 'Good', 'Excellent']

}

df = pd.DataFrame(data)

print("Original Dataset:")

print(df)

# Perform one-hot encoding

df_encoded = pd.get_dummies(df, columns=['Gender', 'Remarks'],
prefix=['Gender', 'Remarks'])

print("\nDataset after One-Hot Encoding:")

print(df_encoded)
```

Output:

```
Original Dataset:
   Employee_ID     Name  Gender    Remarks
0            1    Alice  Female  Excellent
1            2      Bob    Male       Good
```

```
2            3  Charlie    Male    Average
3            4    David    Male       Good
4            5      Eva  Female  Excellent

Dataset after One-Hot Encoding:
   Employee_ID     Name  Gender_Female  Gender_Male  Remarks_Average  \
0            1    Alice              1            0                0
1            2      Bob              0            1                0
2            3  Charlie              0            1                1
3            4    David              0            1                0
4            5      Eva              1            0                0

   Remarks_Excellent  Remarks_Good
0                  1             0
1                  0             1
2                  0             0
3                  0             1
4                  1             0
```

Program for demonstration of Categorical encoding:

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import OneHotEncoder

# Sample dataset
data = {
        'Employee_ID': [1, 2, 3, 4, 5],
        'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eva'],
        'Gender': ['Female', 'Male', 'Male', 'Male', 'Female'],
        'Remarks': ['Excellent', 'Good', 'Average', 'Good', 'Excellent']
        }
df = pd.DataFrame(data)

# Converting type of columns to category
df['Gender'] = df['Gender'].astype('category')
df['Remarks'] = df['Remarks'].astype('category')

# Assigning numerical values and storing it in another columns
df['Gen_new'] = df['Gender'].cat.codes
df['Rem_new'] = df['Remarks'].cat.codes

# Create an instance of One-hot-encoder
enc = OneHotEncoder()

# Passing encoded columns
enc_data = pd.DataFrame(enc.fit_transform(df[['Gen_new', 'Rem_new']]).toarray())

# Merge with the main dataframe
new_df = pd.concat([df, enc_data], axis=1)
print("Original Dataset:")
```

```
print(df)
print("\nDataset after One-Hot Encoding:")
print(new_df)
```

Output:

```
Original Dataset:
   Employee_ID     Name  Gender    Remarks  Gen_new  Rem_new
0            1    Alice  Female  Excellent        0        1
1            2      Bob    Male       Good        1        2
2            3  Charlie    Male    Average        1        0
3            4    David    Male       Good        1        2
4            5      Eva  Female  Excellent        0        1

Dataset after One-Hot Encoding:
   Employee_ID     Name  Gender    Remarks  Gen_new  Rem_new    0    1    2
0            1    Alice  Female  Excellent        0        1  1.0  0.0  0.0
1            2      Bob    Male       Good        1        2  0.0  1.0  0.0
2            3  Charlie    Male    Average        1        0  0.0  1.0  1.0
3            4    David    Male       Good        1        2  0.0  1.0  0.0
4            5      Eva  Female  Excellent        0        1  1.0  0.0  0.0

     3    4
0  1.0  0.0
1  0.0  1.0
2  0.0  0.0
3  0.0  1.0
4  1.0  0.0
```