**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