**EXP-12**

**Apply One Hot Encoding for categorical sequence data.**

One Hot Encoding is a technique used to convert categorical variables into a format that can be provided to ML algorithms to improve prediction accuracy. It converts categorical variables into a binary vector representation.

Here's how you can apply One Hot Encoding to categorical sequence data using Python, specifically utilizing the **OneHotEncoder** from the **sklearn.preprocessing** module:

from sklearn.preprocessing import OneHotEncoder

import numpy as np

# Sample categorical sequence data

categorical\_data = ['cat', 'dog', 'cat', 'bird', 'dog']

# Reshape data to fit OneHotEncoder requirements

categorical\_data = np.array(categorical\_data).reshape(-1, 1)

# Initialize OneHotEncoder

encoder = OneHotEncoder()

# Fit and transform the data

one\_hot\_encoded\_data = encoder.fit\_transform(categorical\_data).toarray()

print("One Hot Encoded Data:")

print(one\_hot\_encoded\_data)

**One Hot Encoded Data:**

[[1. 0. 0.]

[0. 1. 0.]

[1. 0. 0.]

[0. 0. 1.]

[0. 1. 0.]]