MLOps Assignment – 2 Report

AIM: The objective is to predict the number of bike rentals (target variable: cnt) using the Bike Sharing dataset but with added complexity and optimization as follows:

1. Creating two new features
   a. temp_hum = temp*hum

```
[34] #Adding two new features
     df['wind_temp'] = df['windspeed'] * df['temp']
     df['hum_temp'] = df['hum'] * df['temp']
```

2. Using Target Encoder for Categorical Features

Results:

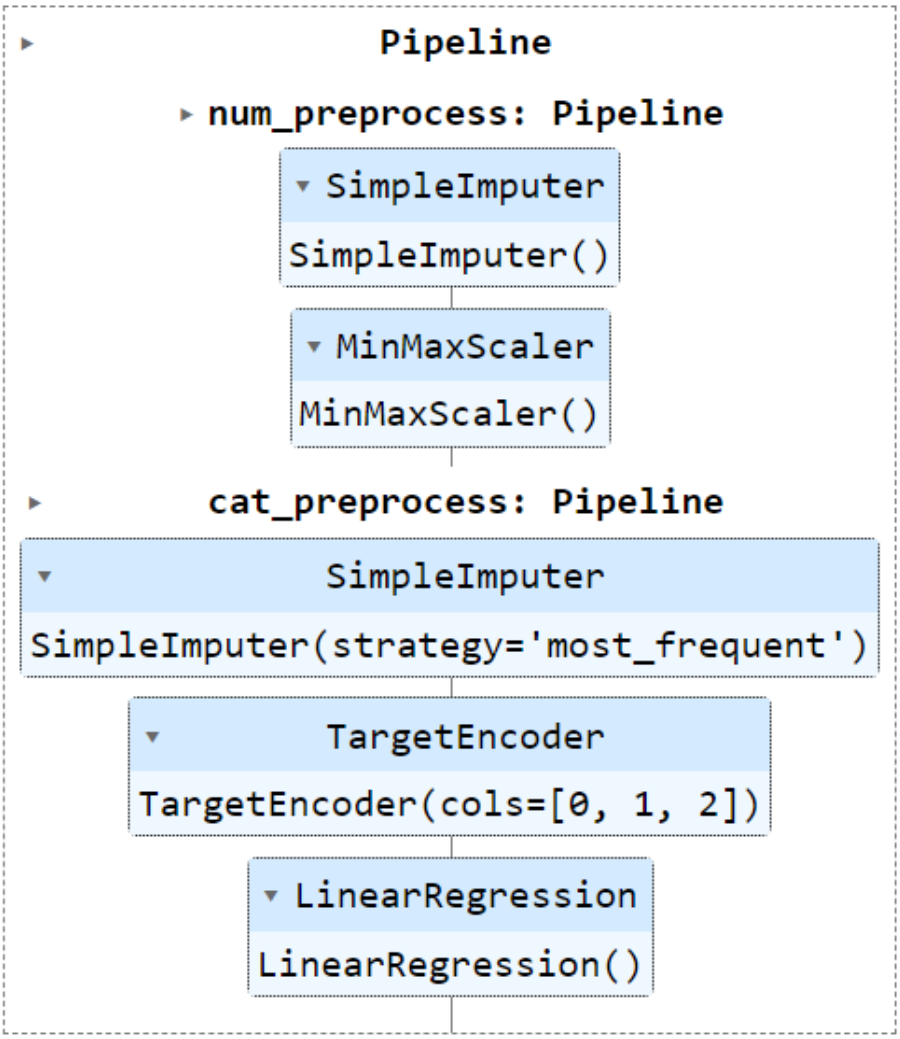
1. Using Linear Regression Package:
   a. Mean Square Error: 1.3539963244568922e-21
   b. R-squared: 1

```
Mean Squared Error: 1.359273477023023e-21
R-squared: 1.0
```

2. Using Linear Regression from Scratch:
   a. Mean Square Error: 3.1348747725858174e-20
   b. R-squared: 1

```
Mean Squared Error: 3.1348747725858174e-20
R-squared: 1.0
```

Pipeline:



Observation:

1. Mean Square Error has been reduced to a large extent nearly zero which was around 1800 in random forest regressor using one hot encoder which suggests that Linear Regression model may overfit the training data.