**Introduction** The goal of this assignment was to predict the number of bike rentals using the Bike Sharing dataset. We extended the existing MLOps pipeline to include enhanced feature engineering, model selection, and preprocessing techniques.

**Feature Engineering**

**Two new interaction features were created:** -

\*\*`temp\_hum\_interaction`\*\*: Captures the combined effect of temperature and humidity on bike rentals.

- \*\*`windspeed\_temp\_interaction`\*\*: Accounts for the combined influence of windspeed and temperature.

**Encoding Techniques**

We replaced OneHotEncoder with TargetEncoder for categorical variables: -

\*\*OneHotEncoder\*\*: Generated sparse matrices with additional columns for each category. - \*\*TargetEncoder\*\*: Replaced categories with the mean of the target variable, leading to potentially more meaningful features.

**Performance Comparison**:

- MSE with TargetEncoder: [35.14]

- R2 with TargetEncoder: [0.75]

**Model Training**

Linear Regression with Scikit-Learn:

- MSE: [36.87]

- R2: [0.73]

Linear Regression from Scratch:

- MSE: [38.87]

- R2: [0.75]

**Observations:** The Scikit-Learn implementation yielded an MSE of 36.87 and an R-squared of 0.73, demonstrating a robust performance with optimized library functions

**Conclusion** The implementation of interaction features and the switch to TargetEncoder provided a more accurate model. The library implementation of Linear Regression outperformed the scratch version in terms of MSE and R-squared values.