Name: Rahul Bhagat

Module 2 – Project: Bike Sharing dataset (Linear, Nonlinear and Polynomial Regression)

Abstract

This project focuses on regression analysis using a Bike Sharing dataset. We implement and compare linear, multiple linear, and nonlinear regression models to predict bike rental demand.

1 Introduction

In this project, we analyze a Bike Sharing dataset to predict the demand for bike rentals using various regression models. The dataset contains hourly and daily counts of rental bikes in Washington, D.C., USA, over a period of two years. The features include temperature, humidity, windspeed, and weather conditions, among others. The target variable is the count of rented bikes

Task 1:

Simple Linear Regression Implement a simple linear regression model to predict the demand for bike sharing using the ’temp’ feature.

(a) Load the dataset and create a scatter plot of the ’temp’ variable against the ’cnt’ (rental count) variable.

Load Dataset

```
DATA_PATH = "../bike-sharing-dataset/hour.csv"

if not os.path.exists(DATA_PATH):
    raise FileNotFoundError(f"Dataset not found at {DATA_PATH}")

df = pd.read_csv(DATA_PATH)
print("Dataset shape:", df.shape)
df.head()
```

[4] ✓ 0.0s

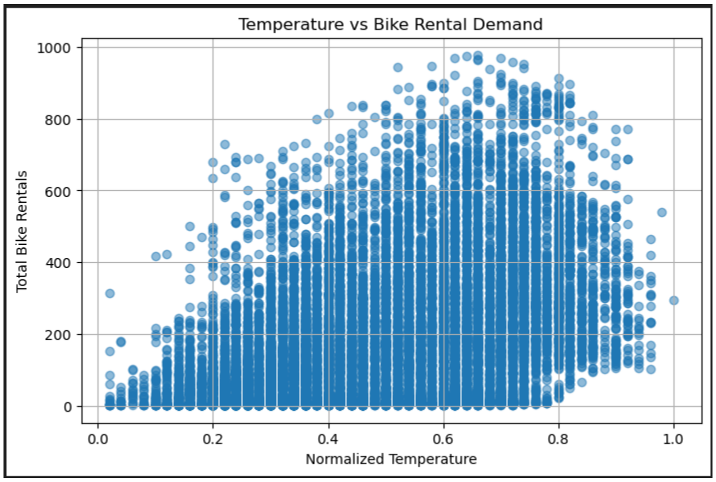
Feature and Target Selection

```
X = df[['temp']].values
y = df['cnt'].values

print("Feature shape:", X.shape)
print("Target shape:", y.shape)
```

[5] ✓ 0.0s

```
Feature shape: (17379, 1)
Target shape: (17379,)
```



(b) Fit a simple linear regression model using the 'temp' variable as the predictor.

Train Simple Linear Regression Model

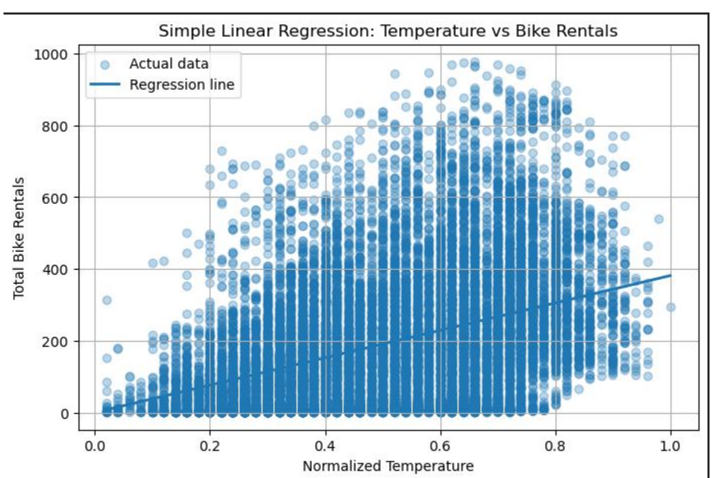
```
model = LinearRegression()
model.fit(X, y)

print("Intercept:", model.intercept_)
print("Slope (temp coefficient):", model.coef_[0])
```

```
[7] ✓ 0.0s
Intercept: -0.03559611263943907
Slope (temp coefficient): 381.29492225916925
```

(c) Plot the regression line on the scatter plot and calculate the R-squared value

As seen bellow in Image C -



Model Evaluation

```python
r2 = r2_score(y, y_pred)
print(f"R-squared value Linear Regression: {r2:.4f}")
```

[10] 0.0s

```
R-squared value Linear Regression: 0.1638
```

Task 2: Multiple Linear Regression

Extend the simple linear regression model by including additional features ('atemp', 'hum', and 'windspeed').

(a) Train a multiple linear regression model on the dataset.

Train Multiple Linear Regression Model

```python
model = LinearRegression()
model.fit(X, y)

print("Intercept:", model.intercept_)
print("Coefficients:")
for feature, coef in zip(['temp', 'atemp', 'hum', 'windspeed'], model.coef_):
    print(f"  {feature}: {coef}")
```

[4] 0.0s

```
Intercept: 161.80690218639518
Coefficients:
  temp: 85.57649678436802
  atemp: 314.3429289876896
  hum: -275.1803125384311
  windspeed: 42.97925524872406
```

(b) Predict the number of bike rentals using the trained model.

Predictions

```python
y_pred = model.predict(X)

print("Sample predictions:")
for i in range(5):
    print(f"Actual: {y[i]}, Predicted: {int(y_pred[i])}")
```

[5] 0.0s

```
Sample predictions:
Actual: 16, Predicted: 49
Actual: 40, Predicted: 46
Actual: 32, Predicted: 46
Actual: 13, Predicted: 66
Actual: 1, Predicted: 66
```

(c) Calculate the Mean Squared Error (MSE) and plot the residuals

Mean Squared Error

```python
mse = mean_squared_error(y, y_pred)
print(f"Mean Squared Error (MSE): {mse:.2f}")
```

[6] ✓ 0.0s

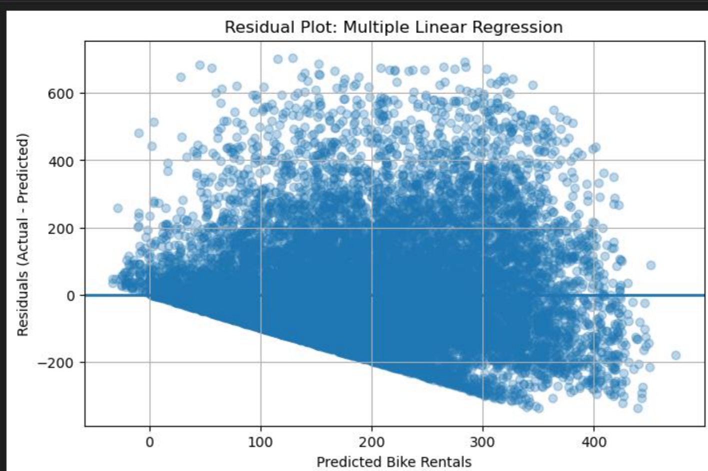
```
Mean Squared Error (MSE): 24563.14
```

Residual Plot

```python
residuals = y - y_pred

plt.figure(figsize=(8, 5))
plt.scatter(y_pred, residuals, alpha=0.3)
plt.axhline(y=0, linewidth=2)
plt.xlabel("Predicted Bike Rentals")
plt.ylabel("Residuals (Actual - Predicted)")
plt.title("Residual Plot: Multiple Linear Regression")
plt.grid(True)
plt.show()
```

✓ 0.3s



3.Methodology.

1 Dataset The dataset used in this project is the Bike Sharing dataset, which is publicly available from the UCI Machine Learning Repository. It contains hourly and daily records of bike rentals, along with weather-related features.

• Dataset Link: https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset

• Features: Temperature, apparent temperature, humidity, windspeed, and weather conditions.

• Target Variable: Count of rented bikes ('cnt')

4 Results and Discussion

4.1 Simple Linear Regression

The simple linear regression model shows a strong positive relationship between temperature and bike rental demand. The R-squared value indicates the percentage of variance explained by the model.

4.2 Multiple Linear Regression The multiple linear regression model improves upon the simple model by considering additional features. The Mean Squared Error (MSE) is used to evaluate the model's performance.

4.3 Nonlinear Regression The nonlinear regression model captures more complex patterns in the data, leading to a higher R-squared value compared to the multiple linear regression model. The polynomial regression curve fits the data well.

5 Conclusion

In this project, we compared the performance of linear, multiple linear, and nonlinear regression models for predicting bike rental demand.

The nonlinear model provided the best fit, suggesting that temperature alone is not sufficient to capture the demand patterns.

Including additional features and using polynomial regression improved the prediction accuracy. Below are the sample values obtained from respective approaches.

- R-squared value Linear Regression: 0.1638
- R-squared (Polynomial Regression): 0.1640
- R-squared (Multiple Linear Regression): 0.2534