**2311CS020228-day23**

**Using the same data set of Civil\_Engineering\_Regression\_Dataset.csv**

**Part 3: Multiple Linear Regression**

1. Fit a multiple linear regression model using **Building Height, Material Quality, Labor Cost, Concrete Strength, and Foundation Depth** as independent variables.
2. What is the equation of the multiple regression model?
3. Which independent variable has the highest impact on Construction Cost based on the regression coefficients?

Code:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

try:

df = pd.read\_csv("Civil\_Engineering\_Regression\_Dataset.csv")

print("CSV file loaded successfully!\n")

except FileNotFoundError:

print("Error: CSV file not found. Check the file path.")

exit()

print("Columns in dataset:", df.columns)

required\_columns = ["Building Height", "Material Quality", "Labor Cost", "Concrete Strength", "Foundation Depth", "Construction Cost"]

for col in required\_columns:

if col not in df.columns:

print(f"Error: Column '{col}' is missing. Check your dataset headers.")

exit()

X = df[["Building Height", "Material Quality", "Labor Cost", "Concrete Strength", "Foundation Depth"]] # Independent variables

y = df["Construction Cost"] # Dependent variable

X = X.apply(pd.to\_numeric, errors="coerce") # Convert to numeric

y = pd.to\_numeric(y, errors="coerce")

df.dropna(inplace=True)

model = LinearRegression()

model.fit(X, y)

coefficients = model.coef\_

intercept = model.intercept\_

equation = f"Construction Cost = {intercept:.2f} "

for var, coef in zip(X.columns, coefficients):

equation += f"+ ({coef:.2f} \* {var}) "

print("\n📌 Regression Equation:")

print(equation)

max\_coef\_index = np.argmax(abs(coefficients))

most\_influential\_var = X.columns[max\_coef\_index]

most\_impactful\_coef = coefficients[max\_coef\_index]

print(f"\n🔍 Most Influential Variable: {most\_influential\_var} (Coefficient: {most\_impactful\_coef:.2f})")

y\_pred = model.predict(X)

r2 = r2\_score(y, y\_pred)

mse = mean\_squared\_error(y, y\_pred)

print(f"\n📊 Model Performance:")

print(f"R-squared: {r2:.4f}")

print(f"Mean Squared Error: {mse:.4f}")

# Plot actual vs predicted values

plt.figure(figsize=(8, 6))

sns.scatterplot(x=y, y=y\_pred, color="blue", alpha=0.6)

plt.plot([y.min(), y.max()], [y.min(), y.max()], color="red", linestyle="--", linewidth=2) # Perfect fit line

plt.xlabel("Actual Construction Cost")

plt.ylabel("Predicted Construction Cost")

plt.title("Actual vs Predicted Construction Cost")

plt.show()