NAME:

Medical Insurance Cost Prediction Using Multiple Linear Regression

Dataset Link:

(Local file used: insurance_dataset.csv)

Drive link :
https://drive.google.com/file/d/1RMOM6ZeB3NLgKYR5l5b6NWbqHWfPkoKN/view?usp=sharing

Methodology Name:

Multiple Linear Regression with One-Hot Encoding for Categorical Variables

Inference:

The model attempts to predict medical insurance charges using features such as age, sex, BMI, number of children, smoking status, and region. Exploratory data analysis revealed that BMI and smoking status are significantly correlated with charges. One-hot encoding was applied to handle categorical variables. A multiple linear regression model was trained on 80% of the dataset and tested on the remaining 20%. The predicted charges closely align with actual values for many instances, suggesting the model performs reasonably well. However, further evaluation using metrics like R^2 or MAE would provide a clearer picture of its predictive power.