Insurance Cost Analysis Using Ridge Regression

Project_Description:

This project aims to analyze insurance costs using Ridge Regression technique and apply it to a dataset related to health insurance. The primary objective of the project is to understand the factors that influence insurance costs and develop a predictive model that can be used to estimate insurance costs for individuals.

Data Used: A dataset containing the following information was used:

Age	Sex	Bmi	Children	Smoker	Region	Annual Insurance Charges
19	Female	27.9	0	Yes	Southwest	16884.92
18	Male	33.77	1	No	Southeast	1725.552
28	Male	33	3	No	Southeast	4449.462
33	Male	22.705	0	No	Northwest	21984.47
32	Male	28.88	0	No	Northwest	3866.855

Steps Accomplished:

1. **Exploratory Data Analysis (EDA):** Data was analyzed, and exploratory analysis was conducted to examine the relationships between different variables using visualizations.
2. **Data Cleaning:** Missing values in the data were handled and corrected.
3. **Data Encoding:** Categorical variables were encoded using Label Encoding.
4. **Data Preparation and Splitting:** The data was split into a training set and a test set.
5. **Data Standardization:** Data scaling was performed for consistency.
6. **Classification using Ridge Regression:** A Ridge Regression model was applied to classify records and analyze the effects

Model Performance Evaluated

```
r2 = r2_score(y_test, y_pred)
print("R2 equal :", r2)
```

```
R2 equal : 0.7447324977388576
```

Visualization