

Case Study: Predicting Medical Charges Using Linear Regression

Objective:

Develop a predictive model to estimate individual medical charges based on personal attributes using linear regression.

Dataset Overview:

Find dataset [here](#)

The dataset comprises several features:

- age: Age of the individual.
- sex: Gender (male/female).
- bmi: Body Mass Index.
- children: Number of children/dependents.
- smoker: Smoking status (yes/no).
- region: Geographical region.
- charges: Individual medical costs billed by health insurance.

Tasks:

Data Preprocessing:

1. Handle missing or anomalous data.
2. Convert categorical variables (sex, smoker, region) into numerical formats using encoding techniques (e.g., one-hot encoding).
3. Normalize/standardize numerical features if required.

Exploratory Data Analysis (EDA):

1. Analyze the distribution of key variables (e.g., age, bmi, charges).
2. Investigate relationships between features and the target variable (charges).
3. Identify potential outliers or influential points.

Model Development:

1. Split the dataset into training and testing sets.
2. Implement a linear regression model.
3. Evaluate model performance using appropriate metrics (e.g., R-squared, Mean Squared Error).

Model Interpretation:

1. Interpret the coefficients of the linear regression model.
2. Discuss the impact of different features on medical charges.

Advanced Analysis (Optional):

1. Implement regularized linear models (Ridge, Lasso) to see if they yield better results.
2. Conduct a sensitivity analysis to understand the robustness of the model.

Expected Outcomes:**Students should be able to:**

1. Clean and preprocess the dataset for modeling.
2. Conduct a thorough exploratory analysis.
3. Build and validate a linear regression model.
4. Interpret the results and discuss the implications.
5. Address the challenges and limitations of the model.

This case study will enhance students' understanding of linear regression in a practical, real-world context, focusing on healthcare cost prediction