

Problem Statement – Linear Regression with OLS Method

Insurance Charges Prediction

Abstract:

HealthyLife Insurance company has gathered the data regarding the customers and wants to analyze it. Different customers have different lifestyles and hence not all can have the same medical expenditures or requirement of health insurance. To better tailor the insurance package to be given to different customers, the company now wants to integrate machine learning into the process and identify the right insurance charges for each customer. You are hired as a ML Engineer to help the company predict the right charges based on the data that they have been maintaining for each customer.

Problem Statement:

Predict how much could be the insurance charges for a beneficiary based on the data provided using Linear Regression.

Dataset Information:

Column	Description
Age	Age of Primary Beneficiary
Sex	Insurance Contractor gender (female, male)
BMI	Body Mass Index, providing an understanding of body, weights that are relatively high or low relative to height, objective index of body weight (kg / m ²) using the ratio of height to weight, ideally 18.5 to 24.9
Children	Number of children covered by health insurance / Number of dependents
Smoker	Whether the beneficiary smokes or not
Region	The beneficiary's residential area in the US, northeast, southeast, southwest, northwest.
charges	Individual medical costs billed by health insurance (target variable)

Scope:

- Exploratory data analysis
- Data Pre-processing
- Training linear regression model with OLS method for prediction
- Tuning the model to improve the performance

Learning Outcome:

The students will get a better understanding of how the variables are linked to each other and how the EDA approach will help them gain more insights and knowledge about the data that we have and train the Linear Regression model.