PRCP-1021-InsCostPred

Problem Statement

Task 1:- Prepare a complete data analysis report on the given data.

Task 2:-

- Prepare the data, identifying and extracting key features (both input and output parameters) relevant to the problem you will solve.
- Build and train a machine learning model. Here you can evaluate different algorithms, settings and see which model is best for your scenario.

Task 3:- Create a machine learning model to predict the insurance price charged to the customer. The charge depends on various features such as age, type of coverage, amount of coverage needed, gender, body mass index (BMI), region, and other special factors like smoking to determine what health risks the person possesses.

Dataset Description and Link:

In order to make their profit, insurance companies must collect a higher premium than the amount which may become due to the insured person. To achieve this, insurance companies invest a lot of time, effort, and money in creating models that accurately predict healthcare expenses.

Note: The insurance.csv dataset contains 1338 observations (rows) and 7 features (columns). The dataset contains 4 numerical features (age, bmi, children and expenses) and 3 nominal features (sex, smoker and region) that were converted into factors with numerical value designated for each level.

Domain: Finance

Link : https://d3ilbtxij3aepc.cloudfront.net/projects/CDS-Capstone-Projects/PRCP-1021-InsCostPred.zip

Attribute Information:

- age: age of primary beneficiary

- sex: beneficiary’s gender - female or 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: Smoking or non-smoking
- region: the beneficiary's residential area in the US; northeast, southeast, southwest, northwest.
- charges: Individual medical costs billed by health insurance (to be predicted)
- Id: id of beneficiary

Model Comparison Report

Create a report stating the performance of multiple models on this data and suggest the best model for production.

Report on Challenges faced

Create a report which should include challenges faced on the data and what techniques were used with proper reasoning.

Note:- All the above tasks must be created on a single jupyter notebook and the same should be shared as part of final submission of the project.