

Medical Insurance Cost Prediction Model Implementation Overview

A comprehensive guide on the model's implementation process and methodology

Lakshya Thakur and Simranpreet Singh

Roll nos: 102203051

102203049

Introduction to Model Implementation

Detailed Model Implementation Overview

| Attribute | Description |
|---------------|--|
| Age | The age of the individual. |
| Sex | The gender of the individual. |
| BMI | Body Mass Index of the individual. |
| Children | Number of children covered by the insurance. |
| Smoker Status | Whether the individual is a smoker. |
| Region | The region of residence. |
| Expenses | Medical expenses incurred. |

Data Preprocessing Steps

A comprehensive overview of dataset preparation for modeling

Utilizing pandas to load the dataset for further analysis.

Loading the dataset

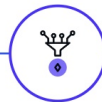


Checking for null values

Ensuring data completeness by identifying and addressing any null entries.

Mapping categorical variables like 'sex' and 'smoker' to binary values for numerical processing.

Converting categorical data





Linear Regression

A simple and interpretable algorithm, ideal for understanding relationships between variables.



Random Forest Regressor

An ensemble method that enhances prediction accuracy by combining multiple decision trees.

Model Selection and Training

Detailed Model Implementation Overview

Model Evaluation Metrics

An overview of model performance metrics used in evaluating machine learning models

| Model | R-squared (R^2) |
|-------------------|---------------------|
| Random Forest | 0.862 |
| Linear Regression | lower score |

Predicting New Data

Detailed Model Implementation Overview



Input Attributes

The model takes several key parameters: age, sex, BMI, number of children, smoking status, and region.



BMI Factor

A BMI of 40.3 indicates obesity, which is associated with higher medical expenses.



Region Influence

Region is coded as 2, reflecting potential geographical variations in healthcare costs.



Age Parameter

Age is set at 40 years, influencing the medical expense prediction significantly.



Children Count

The model accounts for 4 children, which can influence the overall family medical expenses.



Prediction Description

The model effectively predicts medical expenses based on the provided attributes, showcasing its utility.



Sex Parameter

Sex is coded as 1 (male), which plays a role in determining healthcare costs.



Smoking Status

The individual is a smoker (1), typically leading to increased healthcare costs.

Model Persistence and Deployment

Detailed Model Implementation Overview



Model Saving with Joblib

The Random Forest model was saved using Joblib, which allows the model to be easily loaded later for making predictions.

Interactive GUI Development

A graphical user interface (GUI) was developed using Tkinter, providing an interactive platform for users to input data and receive predictions.



1

Successful Prediction of Medical Expenses

The implementation successfully predicts medical expenses using the Random Forest algorithm, demonstrating its effectiveness in this domain.