# Insurance charges prediction – Machine Learning Regression

1.Problem Statement: To predict the insurance charges based on the several parameters

2. Basic info about dataset: There are 1338 rows and 6 columns. It contains several parameters (age, sex, bmi, children, smoker and charges)

3.Pre-processing method: As the sex and smoker columns are string data,

We can use one hot encoding (nominal) pre-processing method

4. Regression methods:

1. Simple Linear Regression: R2 score value - 0.7811302113434095
2. Multiple Linear Regression: R2 score value - 0.78
3. Support Vector Machine: R2 score value - 0.866339395308168

A screenshot of a computer

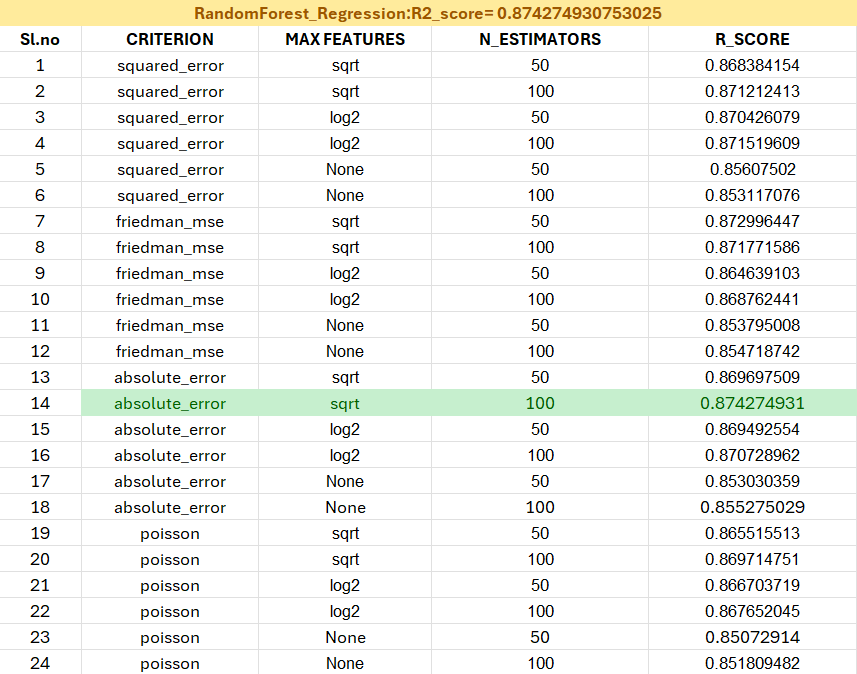
Description automatically generated

1. Decision Tree: R2 score value - 0.771493153484015

A screenshot of a computer

Description automatically generated

1. Random Forest: R2 score value - 0.874274930753025



Since Random Forest R2 score value is 0.874274930753025 and it is near to 1  
Random Forest is the best model for this problem prediction   
I have made the deployment phase for random forest