1) **Identify your problem statement**

Client wants to predict the insurance charges based on the given dataset.

1. Machine Learning (here inputs are numbers, hence ML)
2. Supervised Learning (received the dataset for input and output)
3. Regression (Output dataset is Numerical) > Multiple Linear Regression, SVM, DT, RF (ie multiple input)

Machine Learning > Supervised Learning > Regression > Multiple Linear Regression, SVM, DT, RF

2) **Tell basic info about the dataset (Total number of rows, columns)**

The dataset consist of 6 columns and 1338 row.
The input dataset are Age, Sex, BMI, Children, Smoker
The Output dataset is Charges.

3) **Mention the pre-processing method if you're doing any (like converting string to number – nominal data)**

Since the dataset has difference in age , BMI, charges etc. We have to preprocess the inputs using StandardsScaler function in SVM regression.

However, I got the R2 value 0.8663 which is less than Random Forest.

4) **Develop a good model with r2_score. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.**

I saved the Random Forest as the final model which got the accuracy 0.8897 which is higher when compared to other madels.

5) **All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)**

Please refer "Machine Learning method using r2 value" attachment

6) **Mention your final model, justify why u have chosen the same.**

I saved the Random Forest as the final model which got the highest accuracy 0.8897 when compared to other models such as Multiple Linear (accuracy 0.7894), SVM (accuracy 0.8663), Decision tree (accuracy 0.7559).

Tested using various hyper tunning parameters (Criterion, n_estimators**,** *max_depth, random_state) but max_depth and random_state parameters gave the highest accuracy. Hence chosen Random Forest as the good model*

General info

Kindly create Repository in the name Regression Assignment.
Upload all the ipynb and final document in the pdf

Communication is important (How you are representing the document.)