# Assignment-Regression Algorithm

**1. Since there are multiple inputs and output are known. It is a supervised and the output in numerical so Supervised Linear Regression.**

**2. There are totally 5 columns as a Input (Age,Sex,BMI,Children,Smoker) and only one output column (Charges)**

**3. In dataset there are sex and smoker columns comes under classification data so here we are converting classification data to nominal data using get dummies. Then splitting up inputs and output(Charges) values separately**

**4. As per given assignment. I have done with multiple algorithm using grid searchCV. In that specifically generated output for the inputs Age:32, BMI: 43, Children:2, Sex Male 0 Or 1:0, Smoker 0 or 1:1 with and without testset .**

**5. Testset with $R^2$ value, Best parameters, Future predictions**

| S.No | Age | BMI | Children | Sex Male | Smoker | Algorithm | $R^2$ value | Best parameters | Future predictions |
|------|-----|-----|----------|----------|--------|-----------|-------------|-----------------|--------------------|
| 1 | 32 | 43 | 2 | 0 | 1 | **SVR-GridSearchCV** | 0.85 98 | C': 3000, 'gamma':'scale','kernel':'poly' | 2291197.32 |
| 2 | 32 | 43 | 2 | 0 | 1 | **SVR GridSearchCV ( without input and output set)** | - | 'criterion':'friedman_mse', 'max_features':'auto','splitter':'best' | 40932.42 |
| 3 | 32 | 43 | 2 | 0 | 1 | **DecisionTree-GridSearchCV** | 0.69 36 | 'criterion':'friedman_mse', 'max_features':'sqrt','splitter':'best' | 63770.42 |
| 4 | 32 | 43 | 2 | 0 | 1 | **DecisionTree-GridSearchCV ( without input and output set)** | - | 'criterion':'friedman_mse', 'max_features':'auto','splitter':'best' | 58571.07 |

| 5 | 32 | 43 | 2 | 0 | 1 | **RandomForest - GridSearchCV** | 0.8719 | **'criterion':'poisson','max_features':'sqrt','n_estimators':100** | 48322.27 |
|---|----|----|---|---|---|---------------------------------|--------|-------------------------------------------------------------------|----------|
| 6 | 32 | 43 | 2 | 0 | 1 | **RandomForest - GridSearchCV ( without input and output set)** | – | **'criterion':'friedman_mse','max_features':'log2','n_estimators':100** | 44859.83 |

6. Among checking all the values for each and every Algorithm. I have found ed that the best $R^2$ **value** is for the Algorithm (RandomForest- GridSearchCV). While comparing to other algorithms **RandomForest- GridSearchCV** $R^2$ **value** valu es is nearly to **1**. Since the value of **RandomForest- GridSearchCV = 0.8719** n early to **1**. So considering the $R^2$ **value** of **RandomForest- GridSearchCV** is the best and final model.