Assignment-Regression Algorithm

- **Problem:**

1.) Identify your problem statement – Insurance Charge Prediction

2.) Tell basic info about the dataset (Total number of rows, columns) – 1338 Rows, 6 Columns

3.) Mention the pre-processing method if you're doing any (like converting string to number –nominal data)- Converting Sex and Smoker categorical data into numerical

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.

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

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

- **Domain:** Machine Learning (Value prediction is 'Number')
- **Learning:** Supervised Learning (Input & Output provided and Requirement is clear)
- **Algorithm:** Regression (Output value is 'Numerical')

To determine the evaluation matrix value for various Regression algorithms using the R2 approach.

Standards in R^2 value: Value between 0-1.

1. **MLR: 0.7894**
2. **Support Vector Machine:**

Sr.No	Hyper Parameter	Linear	Poly	RBF	Sigmoid
1	C=1.0	-0.1116		-0.0894	-0.0897
2	C=10	-0.0016		-0.0874	-0.0897
3	C=100	0.5432		-0.0745	-0.0897
4	C=1000	0.6340		-0.0407	-0.0890

- **Result: The Support Vector Machine algorithm, provides R^2 Value = 0.6340**

  SVR (kernel="linear", C=1000)

3. **Decision Tree:**

Sr.No	Criterion	Max_Features	Splitter	R^2 Value
1	mae(squared_error)	auto	Best	0.6854
2	mae(squared_error)	sqrt	Best	0.7555
3	mae(squared_error)	log2	Best	0.6751
4	mae(squared_error)	auto	Random	0.7173
5	mae(squared_error)	sqrt	Random	0.7760
6	mae(squared_error)	log2	Random	0.6752

7	friedman_mse	auto	Best	0.6809
8	friedman_mse	sqrt	Best	0.7181
9	friedman_mse	log2	Best	0.7564
10	friedman_mse	auto	Random	0.6084
11	friedman_mse	sqrt	Random	0.6271
12	friedman_mse	log2	Random	0.5252
13	Mse	auto	Best	0.6896
14	Mse	sqrt	Best	0.7092
15	Mse	log2	Best	0.6680
16	Mse	auto	Random	0.6427
17	Mse	sqrt	Random	0.7118
18	Mse	log2	Random	0.6950

- **Result: The decision tree algorithm, provides R2 Value = 0.7760**

  DecisionTreeRegressor (criterion='mae', max_features='sqrt', splitter='random')

4. **Random Forest:**

Sr.No	Criterion	Max_Features	n_estimators	R^2 Value
1	mse	auto	100	0.8522
2	mse	sqrt	100	0.8683
3	mse	log2	100	0.8690
4	mse	auto	200	0.8548
5	mse	sqrt	200	0.8715
6	mse	log2	200	0.8730
7	mae	auto	100	0.8546
8	mae	sqrt	100	0.8748
9	mae	log2	100	0.8677
10	mae	auto	200	0.8553
11	mae	sqrt	200	0.8727
12	mae	log2	200	0.8743
13	friedman_mse	auto	100	0.8517
14	friedman_mse	sqrt	100	0.8739
15	friedman_mse	log2	100	0.8703
16	friedman_mse	auto	200	0.8562
17	friedman_mse	sqrt	200	0.8724
18	friedman_mse	log2	200	0.8712

- **Result: The Random Forest algorithm, provides R2 Value = 0.8748**

  RandomForestRegressor (criterion='mae', max_features='sqrt', n_estimators=100)

Best Value selection:

	Multiple Linear Regression	Support Vector Machine	Decision Tree	Random Forest
R^2 Value	0.7894	0.6340	0.7760	0.8748

Ada Boosting – 0.8376

XG Boosting – 0.7884

LG Boosting – 0.8660

Conclusion:

As the above models evaluation matrix value, Random Forest algorithm provides the highest R^2 Value (0.8748), so this model has been considered as good model and saved for further deployment phase.

Title: Insurance Charge Predictor.

Dataset:

insurance_pre.csv