Insurance Charge Prediction

Aim: To find the best model using machine learning algorithms.

1.Identify your problem statement

Domain Selection: Machine Learning

Learning Selection: Supervised Learning

Output-Numerical value: Regression

2.Dataset Information

Number of inputs: 5 (age, bmi, children, sex, smoker)

Output: charges

Total number of rows: 1338

Total number of columns: 6

3. Pre-processing method

Input 'sex and smoker' were provided under categorical values as nominal data which has been converted to numerical values using 'one-hot-encoding'.

Standardised the values in order to create a best model.

4. Develop a good model with r2_score

Performed the below mentioned machine learning algorithms to find the best model using Hyper Tunning Parameters.

SIMPLE LINEAR REGRESSION

R2 value=0.732

MULTIPLE LINEAR REGRESSION

R2 value=0.764

SUPPORT VECTOR MACHINE

Kernel	***c***	***R2 Value***
rbf	1000	0.810
Poly	1000	0.856
Linear	1000	0.764
sigmoid	1000	0.287

DECISION TREE

Criterion	*Splitter*	*R2 Value*
squared_error	best	0.686
friedman_mse	best	0.699
poisson	best	0.747
absolute_error	best	0.668
squared_error	random	0.723
friedman_mse	random	0.633
poisson	random	0.732
absolute_error	random	0.722

RANDOM FOREST

n_estimators	*Random_state*	*R2 Value*
10	0	0.833
	1	0.839
20	0	0.846
	1	0.848
30	0	0.851
	1	0.851
40	0	0.854
	1	0.852
50	0	0.850
	1	0.855
60	0	0.851
	1	0.854
70	0	0.853
	1	0.855
80	0	0.854
	1	0.853
90	0	0.854
	1	0.854
100	0	0.854
	1	0.855

Result: It is concluded that the Support Vector Machine with Kernel=ploy and c=1000 provides the best model.