Regression Assignment

Problem Statement or Requirement

- Predict the insurance charges based on the several parameters

Dataset - Has 1338 rows × 6 columns

Preprocessing - To convert nominal to numbers one hot encoding method is followed

Multilinear Algorithm

R_score = `0.7894790349867009`

Support Vector Machine - SVM

S.no	Kernel	C	R_score
1	linear	-	-0.111661
2	linear	10	-0.0016
3	linear	50	0.398
4	linear	1000	0.63403
5	linear	3000	0.759
6	rbf	-	-0.06429
7	rbf	10	-0.0819
8	rbf	50	-0.1113
9	rbf	1000	-0.117
10	poly	-	-0.0642
11	poly	10	-0.09311

12	poly	50	-0.10033
13	poly	1000	-0.055
14	*sigmoid*	-	-0.0899
15	*sigmoid*	10	-0.090
16	*sigmoid*	50	-0.0985
17	*sigmoid*	1000	-1.6659

Decision Tree

S.no	Criterion	Spliter	R_score
1	-	-	0.69778
2	squared_error	-	0.69119
3	squared_error	best	0.6900
4	squared_error	random	0.7137
5	friedman_mse		0.7023
6	friedman_mse	best	0.69
7	friedman_mse	random	0.7086
8	*absolute_error*		0.6926
9	absolute_error	best	0.67968
10	absolute_error	random	0.71226
11	poisson	-	0.7216

12	poisson	best	0.71869
13	poisson	random	0.7376

Random Forest

S.no	n_estimators	Criterion	random_state	R_score
1	100	-	0	0.85366
2	50	-	0	0.8578
3	100	squared_error	0	0.85383
4	50	squared_error	0	0.8498
5	100	friedman_mse	0	0.8540
6	50	friedman_mse	0	0.8500
7	100	*absolute_error*	0	0.852009
8	*50*	*absolute_error*	0	0.852665
11	100	poisson	0	0.8526
12	50	poisson	0	0.84910

regressor = RandomForestRegressor(n_estimators=600,max_depth=10,min_samples_split=10,min_samples_leaf=5,max_features='log2')
R_score = 0.8809

Final Model Selection and Justification

Final Model Chosen:
RandomForestRegressor with tuned hyperparameters

Final Performance:

- R^2 Score = 0.88

Conclusion:-

The final RandomForestRegressor model with an optimized **`max_features`** parameter achieved an R^2 score of 0.88, indicating strong predictive power. This configuration was chosen because it:

- Improves generalization
- Reduces overfitting
- Delivers stable and reliable performance

Thus, the model is well-suited for the regression task and meets the performance objectives effectively.