- **The problem have multiple inputs like- age, BMI, no. of children they have, gender and habits are the input variables. Based on these parameters, we need to calculate the charges for the particular policy, so charges is the only output**
- **In this case, a defined set of inputs and one -output are given, so we can use supervised learning**
- **Here, the inputs like age, BMI and no. of children are numerical dataset. But the gender and habits are in categorical form (ordinal). So we can use ONE HOT Encoding, to convert those inputs to nominal dataset.**
- **Since we have multiple inputs we cannot use Simple Linear Regressor. We can do the model creation in Multiple Linear Regressor, Support Vector Machine, Decision Tree Regressor and Random Forest Regressor algorithms to find the R^2 value**

Comparison of R^2 value, to find the best model using different Machine Learning Regression algorithms

1. **Multiple Linear Regressor (MLR)**: R^2 value=0.789
2. **Support Vector Machine(SVM)**

s.no	Hyper parameter	Linear	Linear with standardization	RBF	Poly	sigmoid
1.	C=1	-0.011	-0.0101	-0.088	-0.0642	-0.0899
2.	C=10	-0.006	0.4624	-0.0819	-0.0931	-0.0907
3.	C=50	0.398	0.6093	-0.111	-0.1003	-0.0985
4.	C=100	0.5432	0.628	-0.124	-0.0997	-0.1181
5.	C=1000	0.634	0.7649	-0.117	-0.0555	-1.665
6.	C=2000	0.689	0.7440	-0.1077	-0.0027	-5.616

SVM Regressor R^2 value provides better result in linear(c=2000) =0.7440

3. Decision Tree Regressor

s.no	Criterion	Max Features	Splitter	R^2 value
1.	squared_error	sqrt	best	0.692
2.	squared_error	sqrt	random	0.666

3.	squared_error	log2	best	0.610
4.	squared_error	log2	random	0.674
5.	friedman_mse	sqrt	best	0.744
6.	friedman_mse	sqrt	random	0.729
7.	friedman_mse	Log2	best	0.661
8.	friedman_mse	Log2	random	0.572
9.	absolute_error	sqrt	best	0.696
10.	absolute_error	sqrt	random	0.7004
11.	absolute_error	Log2	best	0.657
12.	absolute_error	Log2	random	0.734
13.	poisson	sqrt	best	0.760
14.	poisson	sqrt	random	0.672
15.	poisson	Log2	best	0.7098
16.	poisson	Log2	random	0.682

4. Random forest Regressor

s.no	Criterion	Max Features	n_estimator	R^2 value
1.	squared_error	Sqrt (or)Log2	1	0.735
2.	squared_error	Sqrt (or)Log2	10	0.852
3.	squared_error	Sqrt (or)Log2	100	0.870
4.	squared_error	Sqrt (or)Log2	1000	0.871
5.	friedman_mse	Sqrt (or)Log2	1	0.735
6.	friedman_mse	Sqrt (or)Log2	10	0.853
7.	friedman_mse	Sqrt (or)Log2	100	0.871
8.	friedman_mse	Sqrt (or)Log2	1000	0.871
9.	absolute_error	Sqrt (or)Log2	1	0.748
10.	absolute_error	Sqrt (or)Log2	10	0.855
11.	absolute_error	Sqrt (or)Log2	100	0.871
12.	absolute_error	Sqrt (or)Log2	1000	0.872
13.	poisson	Sqrt (or)Log2	1	0.738
14.	poisson	Sqrt (or)Log2	10	0.854
15.	poisson	Sqrt (or)Log2	100	0.868
16.	poisson	Sqrt (or)Log2	1000	0.871

The best model of Regressor for the given problem is

Random forest Regressor with(criterion=absolute_error; max feature=sqrt or log2; n_estimator=1000**) R^2 value =0.872**