To find the following Machine Learning Regression method using R2 Value

Problem Statement:

- **Stage1 = Machine Learning**
- **Stage2 = Supervised Learning**
- **Stage3 = Regression**

Basic info about the statement:

- **Name of the Dataset: insurance_pre.csv**
- **No of Columns: 6**
- **No of Rows: 1339**

Dataset Pre-processing:

In this Dataset having totally 6 columns. only 2 columns are in ordinal data. so I am using label encoder for pre-processing the data(Converting text format to Numerical form)

All Research Values (r2_score of the models):

1. Multiple Linear Regression: **(R2_Score)= 0.789479034**

2. Support Vector Machine:

S.NO	HYPER PARA METER	LINEAR (R2_Score)	RBF(NON-LIN EAR) (R2_Score)	POLY (R2_Score)	SIGMOID (R2_Score)
1	C=10	0.462468414	-0.03227329	0.038716222	0.039307143
2	C=100	0.628879285	0.320031783	0.617956962	0.527610354
3	C=500	0.763105797	0.664298464	0.826368354	0.444606103
4	C=1000	0.764931173	0.810206485	0.856648767	0.287470694
5	C=2000	0.744041830	0.854776642	0.860557926	-0.59395097
6	C=3000	0.741423659	0.866339395	0.859893008	-2.12441947

The SVM Regression use R2_Score (non-linear (rbf) and hyper parameter (c=3000)) = 0.866339395

3. Decision Tree:

S.NO	CRITERION	MAXFEATURES	SPLITTER	R2_SCORE
1	*squared error*	sqrt	best	0.68004182393
2	*squared error*	Log2	random	0.664667991179
3	*squared error*	Log2	best	0.749032503979
4	*squared error*	Sqrt	random	0.728797145251
5	*Poisson*	sqrt	best	0.641675653346
6	*Poisson*	Log2	random	0.659276406167
7	*Poisson*	Log2	best	0.733649026580
8	*Poisson*	Sqrt	random	0.751833199046
9	Friedman mse	sqrt	best	0.701924642544
10	Friedman mse	Log2	random	0.700978074584
11	Friedman mse	Log2	best	0.674772794272
12	Friedman mse	Sqrt	random	0.699076991246
13	Absolute error	sqrt	best	0.756806406468
14	Absolute error	Log2	random	0.723438637698
15	Absolute error	Log2	best	0.724835228468
16	Absolute error	Sqrt	random	0.680307112952

The Decision Tree Regression use R2_Score is Criterion= 'absolute_error', Maxfeature= 'sqrt' and Splitter= 'best' = 0.756806406468

4. Random Forest:

S.NO	CRITERION	MAX_FEATURES	N_ESTIMATORS	R_VALUE
1	*squared error*	sqrt	10	0.8520006346
2	*squared error*	Log2	100	0.8710271903
3	*squared error*	Log2	10	0.8520006346
4	*squared error*	Sqrt	100	0.8710271905
5	*Poisson*	sqrt	10	0.8544955286
6	*Poisson*	Log2	100	0.8680156984
7	*Poisson*	Log2	10	0.8544955286
8	*Poisson*	Sqrt	100	0.8680156984
9	Friedman mse	sqrt	10	0.8502777994
10	Friedman mse	Log2	100	0.8710544015
11	Friedman mse	Log2	10	0.8502777994
12	Friedman mse	Sqrt	100	0.8710544015
13	Absolute error	sqrt	10	0.8574290080
14	Absolute error	Log2	100	0.8710685856
15	Absolute error	Log2	10	0.8574290080
16	Absolute error	Sqrt	100	0.8710685856

The Decision Tree Regression use R2_Score is Criterion = 'absolute_error', Max_feature = 'sqrt' and n_Estimator =100 = 0.8710685856

Final Model:

In this dataset the final model is Decision Tree Regression use R2_Score is Criteri on = 'absolute_error',Max_feature = 'sqrt' and n_Estimator =100 = 0.8710685856. Because other algorithms r2_score value is comparatively very low so I choose the above model.