Assignment

(Subramani)

To find the best regression model in Machine Learning, we compare different regression methods using their R^2 values on the given dataset (50_Startups.csv)

Sample Data sets (50_Startups.csv)

R&D Spend	Administration	Marketing Spend	State	Profit
165349.2	136897.8	471784.1	New York	192261.8
162597.7	151377.6	443898.5	California	191792.1
153441.5	101145.6	407934.5	Florida	191050.4
144372.4	118671.9	383199.6	New York	182902
142107.3	91391.77	366168.4	Florida	166187.9

1. **Multiple Linear Regression:** Best R^2 value = **0.9358**

2. **Support Vector Machine (SVM):** Best R^2 value = **0.9301**

S. No	Hyper Parameter C	Linear R^2 value	RBF R^2 value	POLY R^2 value	SIGMOID R^2 value
1	C=1	-0.0556	-0.0574	-0.0571	-0.0572
2	C=100	0.1064	-0.0507	-0.0198	-0.0304
3	C=1000	0.7802	0.0067	0.2661	0.1850
4	C=10000	0.9240	0.3718	0.8129	0.8535
5	C=100000	0.9301	0.7085	0.4002	-0.8434
6	C=10000000	0.9301	0.7099	-0.3494	-245.13

3. **Decision Tree:** Best R^2 value = **0.9479**

S. No	Criterion	Splitter	max_depth	R^2 value
1	squared_error	best	None	0.9323
2	squared_error	best	int	0.9238
3	squared_error	random	int	0.9066
4	squared_error	random	None	0.9479
5	friedman_mse	best	None	0.9184
6	friedman_mse	best	int	0.9034
7	friedman_mse	random	int	0.8938
8	friedman_mse	random	None	0.9157
9	absolute_error	best	None	0.9156
10	absolute_error	best	int	0.8934
11	absolute_error	random	int	0.9176
12	absolute_error	random	None	0.9085
13	poisson	best	None	0.8822
14	poisson	best	int	0.9133
15	poisson	random	int	0.9153
16	poisson	random	None	0.9130