PROJECT : PROFIT IDENTIFICATION

3 stages of selection

1) Stage1 - Machine Learning Domain
2) Stage2 - Supervised Learning
3) Stage3 - Regression

1. Algorithm : Multiple Linear Regression

R^2 value : 0.9358

2. **Algorithm : Support Vector Machine**

R^2 value : 0.93012

S.No	Parameters		R^2 VALUE	Remarks
	Kernel	Penalty C		
1	Linear	10	-0.03964	
2		100	0.10646	
3		1000	0.78028	
4		5000	0.90037	
5		10000	0.92399	
6		25000	0.93012	Best value
7		50000	0.93012	
8		100000	0.93012	
9	Poly	10	-0.05366	
10		100	-0.01980	
11		1000	0.26616	
12		5000	0.79365	
13		25000	0.63183	
14	rbf	10	-0.05680	
15		100	-0.05072	
16		1000	0.00676	
17		5000	0.21242	
18		25000	0.56372	
19		26000	0.56738	
20	sigmoid	10	-0.05471	
21		100	-0.03045	
22		1000	0.18506	
23		5000	0.73065	
24		10000	0.85353	
25		20000	0.90111	
26		25000	0.88146	
27		26000	0.86835	
28		30000	0.81438	
29		100000	-0.84337	
30	Precomputed			Not applicable for this dataset

3. Algorithm : Decision Tree

R^2 value : 0.94913

S.No	Parameters			R^2 VALUE	Remarks
	Criterion	Splitter	Max features		
1	squared_ error	best	None	0.90348	
2			sqrt	0.53804	
3			log2	0.45088	
4		random	None	0.913387	
5			sqrt	-0.17588	
6			log2	0.88226	
7	friedman_mse	best	None	0.93876	
8			sqrt	0.76841	
9			log2	0.772701	
10		random	None	0.53204	
11			sqrt	0.85095	
12			log2	0.41113	
13	absolute_ error	best	None	0.94913	Best value
14			sqrt	0.71394	
15			log2	0.91887	
16		random	None	0.78195	
17			sqrt	0.72708	
18			log2	0.73241	
19	poisson	best	None	0.68624	
20			sqrt	0.72344	
21			log2	0.41837	
22		random	None	0.67066	
23			sqrt	0.55848	
24			log2	0.75563	

4. Algorithm : Random forest

R^2 value : 0.94590

S.No	Parameters			R^2 VALUE	Remarks
	Criterion	n_estimators	Max_ features		
1	squared_ error	50	None	0.94463	
2			sqrt	0.68300	
3			log2	0.68300	
4		100	None	0.94600	
5			sqrt	0.75915	
6			log2	0.75915	
7	friedman_mse	50	None	0.93889	
8			sqrt	0.68891	
9			log2	0.68891	
10		100	None	0.94127	
11			sqrt	0.76085	
12			log2	0.76085	
13	absolute_ error	50	None	0.94019	
14			sqrt	0.72223	
15			log2	0.72223	
16		100	None	0.94590	Best value
17			sqrt	0.78574	
18			log2	0.78574	
19	poisson	50	None	0.78621	
20			sqrt	0.66418	
21			log2	0.66418	
22		100	None	0.78406	
23			sqrt	0.72725	
24			log2	0.72725	

Conclusion:

Comparison of results

S.No	Algorithm	Parameters	R_Value
1	Multiple Linear Regression		0.9358
2	Support Vector machine	Kernel=Linear, C=25000	0.93012
3	Decision Tree	Criterion=absolute_ error, Splitter=best, Max features =None	0.94913
4	Random Forest	Criterion=absolute_ error, n_estimators=100, Max features =None	0.94590

Result

Finalized Algorithm : Decision tree

Parameters : Criterion=absolute_ error,splitter=best,Max features =None

R_value :0.94913