Machine Learning Regression Report - 50_Startups Dataset

Multiple Linear Regression

$R^2 = 0.8753$

Support Vector Machine (SVM)

C	Kernel	R^2
2000	linear	0.8703
3000	linear	0.8584
1000	linear	0.7943
3000	poly	0.6603
500	linear	0.5352
3000	sigmoid	0.4987
2000	poly	0.4921
2000	sigmoid	0.3607
1000	poly	0.2113
1000	sigmoid	0.1377
500	poly	0.0718
100	linear	0.0430
3000	rbf	0.0193
500	sigmoid	0.0065
2000	rbf	-0.0306
1000	rbf	-0.0800
100	poly	-0.0842
100	sigmoid	-0.0964

500	rbf	-0.1015
10	linear	-0.1069
100	rbf	-0.1208
10	poly	-0.1216
10	sigmoid	-0.1228
10	rbf	-0.1253

Best SVM Model: C=2000, Kernel=linear, R^2=0.8703

Decision Tree Regressor

Criterion	Max Features	Splitter	R^2
mae	auto	best	0.9070
mse	auto	best	0.8540
friedman_mse	auto	best	0.8399
mse	sqrt	best	0.7282
mse	log2	best	0.7282
friedman_mse	sqrt	best	0.7282
friedman_mse	log2	best	0.7282
mae	auto	random	0.7166
mse	auto	random	0.6256
friedman_mse	auto	random	0.6097
mae	sqrt	best	-0.0575
mae	log2	best	-0.0575
mae	sqrt	random	-0.5610
mae	log2	random	-0.5610
mse	sqrt	random	-0.9510
mse	log2	random	-0.9510

friedman_mse	sqrt	random	-0.9510
friedman_mse	log2	random	-0.9510

Best Decision Tree: {'Criterion': 'mae', 'Max_Features': 'auto', 'Splitter': 'best', 'R^2': 0.9070088484283245}

Random Forest Regressor

Criterion	Max Features	N_Estimators	R^2
mae	auto	10	0.9339
mse	auto	10	0.9289
mae	auto	100	0.9265
mse	auto	100	0.9199
mae	sqrt	100	0.8017
mae	log2	100	0.8017
mse	sqrt	100	0.7781
mse	log2	100	0.7781
mae	sqrt	10	0.7688
mae	log2	10	0.7688
mse	sqrt	10	0.7393
mse	log2	10	0.7393

Best Random Forest: {'Criterion': 'mae', 'Max_Features': 'auto', 'N_Estimators': 10, 'R^2': 0.9338561182244065}

Summary of Best Models

Model	Best R^2
Multiple Linear Regression	0.8753
SVM (Best)	0.8703
Decision Tree (Best)	0.9070
Random Forest (Best)	0.9339