Problem Statement:

Develop a new model of health insurance charges prediction based input parameters of age, sec, bmi, childer and smoker.

Dataset details:

There are six columns and 60 rows in the dataset. The dataset has multiple input and one output. This multiple linear regression.

Independent fields: Age, Sex, BMI, Childern and Is Smoker

Dependent fields: Insurance Charge

Preprocessing Method:

There is a nominal data is available in the dataset, so we need to preprocess using one hot encoding algorithm

Model Evaluation

Support Vector Machine model evaluation:

No	Hyper Parameter	Linear (r value)	RBF (Non Linear)	Poly (r_value)	SigMod (r value)
1	C10	-0.00161763	-0.0819691	-0.09311616	-0.0907832
2	C100	0.54328182	-0.12480368	-0.09976172	-0.11814555
3	C500	0.627046276	-0.12464161	-0.0820288	-0.45629443
4	C1000	0.634036931	-0.11749092	-0.05550594	-1.66590813
5	C2000	0.689326311	-0.10778764	-0.00270245	-5.61643154
6	C3000	0.759089037	-0.09621285	0.048928964	-12.0190481

Decision Tree Model evaluation`

No	Criterion	Splitter	Max Features	R Value
1	squared_error	best	None	0.707627516
2	squared_error	best	sqrt	0.738117225
3	squared_error	best	log2	0.721713998
4	squared_error	random	None	0.723480813
5	squared_error	random	sqrt	0.728163368
6	squared_error	random	log2	0.726140533
7	friedman_mse	best	None	0.686733982
8	friedman_mse	best	sqrt	0.699954467
9	friedman_mse	best	log2	0.728869594
10	friedman_mse	random	None	0.699923934
11	friedman_mse	random	sqrt	0.681997249
12	friedman_mse	random	log2	0.584977089
13	absolute_error	best	None	0.695402527
14	absolute_error	best	sqrt	0.699429673
15	absolute_error	best	log2	0.698224255
16	absolute_error	random	None	0.716456942
17	absolute_error	random	sqrt	0.676764732
18	absolute_error	random	log2	0.751534952
19	Poisson	best	None	0.714667632
20	Poisson	best	sqrt	0.714230843
21	Poisson	best	log2	0.712018404
22	Poisson	random	None	0.685138133
23	Poisson	random	sqrt	0.680125837
24	Poisson	random	log2	0.646979951

Best Model:

Although the Support Vector Machine (SVM) model achieved the highest R^2 score, the best overall performance was observed with the Multiple Linear Regression model. The accuracy of this model is 0.759089037. This model, configured with a linear kernel and a correction value of 3000, provided the most reliable results based on the evaluation metrics.