**Sales profit Prediction**

1. End goal is Sales profit prediction.
2. 49 Rows and 5 Columns.
3. Nominal Data- String convert to Numbers.
4. R2 researched value given below,

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **R2 values:** | | | | | | |
|  |  |  |  |  |  |  |
| **1. Multiple Linear Regression - R2 value=0.93586** | | | | | | |
|  |  |  |  |  |  |  |
| **2. Support Vector Machine-Regression:** | | | | | | |
|  |  |  |  |  |  |  |
| **SL.NO.** | **HYPER PARAMETER** | **LINEAR** | **NON-LINEAR** | **POLY** | **SIGMOID** | **PRECOMPUTED** |
| 1 | C=0.1 | -0.057306 | -0.057479 | -0.057448 | -0.057458 | NONE |
| 2 | C=1000 | 0.780283 | 0.006768 | 0.266123 | 0.185068 | NONE |
| 3 | C=2000 | 0.876772 | 0.067515 | 0.481002 | 0.397065 | NONE |
| 4 | C=3000 | 0.895674 | 0.123227 | 0.637006 | 0.591363 | NONE |
| 5 | C=4000 | 0.89723 | 0.172383 | 0.732637 | 0.628237 | NONE |
| 6 | C=10000 | 0.923998 | 0.371895 | 0.812862 | 0.853531 | NONE |
|  |  |  |  |  |  |  |
| **SVM Regression- R2 value=0.923998(Linear, Hyper parameter=10000)** | | | | |  |  |
|  |  |  |  |  |  |  |
| **3. Decision Tree:** | | | |  |  |  |
|  |  |  |  |  |  |  |
| **SL.NO** | **CRITERION** | **SPLITER** | **R2 VALUES** |  |  |  |
| 1 | ***squared\_error*** | ***best*** | 0.930723 |  |  |  |
| 2 | ***squared\_error*** | ***random*** | 0.603344 |  |  |  |
| 3 | ***friedman\_mse*** | ***best*** | 0.926274 |  |  |  |
| 4 | ***friedman\_mse*** | ***random*** | 0.919507 |  |  |  |
| 5 | ***absolute\_error*** | ***best*** | 0.925167 |  |  |  |
| 6 | ***absolute\_error*** | ***random*** | 0.895339 |  |  |  |
| 7 | ***poisson*** | ***best*** | 0.907179 |  |  |  |
| 8 | ***poisson*** | ***random*** | 0.450146 |  |  |  |
|  |  |  |  |  |  |  |
| **Decision Tree- R2 value=0.930723(Criterion=squared error, splitter-best)** | | | | |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **4. Random Forest:** | | | |  |  |  |
|  |  |  |  |  |  |  |
| **SL.NO** | **CRITERION** | **n\_estimators** | **R2 VALUES** |  |  |  |
| 1 | ***squared\_error*** | ***100*** | 0.946004 |  |  |  |
| 2 | ***squared\_error*** | ***50*** | 0.944633 |  |  |  |
| 3 | ***friedman\_mse*** | ***100*** | 0.94127 |  |  |  |
| 4 | ***friedman\_mse*** | ***50*** | 0.938895 |  |  |  |
| 5 | ***absolute\_error*** | ***100*** | 0.945909 |  |  |  |
| 6 | ***absolute\_error*** | ***50*** | 0.940193 |  |  |  |
| 7 | ***poisson*** | ***100*** | 0.941388 |  |  |  |
| 8 | ***poisson*** | ***50*** | 0.946354 |  |  |  |
|  |  |  |  |  |  |  |
| **RandoForestRegressior- R2 value=0.946354(Criterion=poisson, n\_estimators-50)** | | | | | |  |

Final model is **RandoForestRegressior- R2 value=0.946354 (Criterion=poisson, n\_estimators-50)**

**Final Output:**

**independent=dataset[['R&D Spend', 'Administration', 'Marketing Spend','State\_Florida', 'State\_New York']]**

**I/P Values: [23,43,566,1,0]**

**dependent=dataset[["Profit"]]**

**Output Values:**  **[35031.6728]**

**Note: Accuracy improved from 50% to 90%**