**Problem Definition**: To predict profit by investing in various department across different location of an industry in the US

1. **3 type of stage analysis:** Machine Learning, Supervised Learning, Regression (Categorical

Nominal col expand to more than one col using one hot encoder)

**2. Multiple Linear Regression**: Using MLR the predicted R Value is 0.93 (even after standardised)

**3. Support Vector Machine**: Using SVM the predicted R Value is 0.95 (After standardised with

Kernel =’Linear’, Gama= (either auto or scale), C=100000).

Before standardisation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **Non-Linear** | | |
| **S#** | **Hyper Tuning Param** | **Linear** | **RBF** | **Poly** | **Sigmoid** |
| 1 |  | 0.89 | -0.05 | -0.05 | -0.05 |
| 2 | C=1 | 0.89 | -0.05 | -0.05 | -0.05 |
| 3 | C=10 | -2.43 | -0.05 | 0.02 | -0.05 |
| 4 | C=100 | -357.07 | -0.03 | 0.46 | -0.05 |
| 5 | C=1000 | -36014.02 | 0.16 | 0.64 | -0.07 |
| 6 | C=10000 | Indefinitely running | 0.67 | 0.81 | -0.21 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **Non-Linear** | | |
| **S#** | **Hyper Tuning Param** | **Playable Param** | **Linear** | **RBF** | **Poly** | **Sigmoid** |
| 7 |  | Gama = Auto | -0.05 | -0.05 | -0.05 | -0.05 |
| 8 | C=1 | Gama = Auto | -0.05 | -0.05 | -0.05 | -0.05 |
| 9 | C=10 | Gama = Auto | -0.03 | -0.05 | -0.05 | -0.05 |
| 10 | C=100 | Gama = Auto | 0.10 | -0.05 | -0.01 | -0.03 |
| 11 | C=1000 | Gama = Auto | 0.78 | 0.00 | 0.36 | 0.18 |
| 12 | C=10000 | Gama = Auto | 0.92 | 0.37 | 0.81 | 0.85 |
| 13 | C=100000 | Gama = Auto | 0.93 | 0.70 | 0.40 | -0.84 |
| 14 | C=1 | Gama = Scale | -0.05 | -0.05 | -0.05 | -0.05 |
| 15 | C=10 | Gama = Scale | -0.03 | -0.05 | -0.05 | -0.05 |
| 16 | C=100 | Gama = Scale | 0.10 | -0.05 | -0.01 | -0.03 |
| 17 | C=1000 | Gama = Scale | 0.78 | 0 | 0.26 | 0.18 |
| 18 | C=10000 | Gama = Scale | 0.92 | 0.37 | 0.81 | 0.85 |
| 19 | C=100000 | Gama = Scale | 0.93 | 0.70 | 0.40 | -0.84 |

After standardisation

**4. Decision Tree**: Using DT the predicted R Value is 0.96 (Criterion = ‘absolute\_error’, Splitter = Best).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SI#** | **criterion** | **splitter** | **max\_features** | **R Value** |
| 1 |  |  |  | 0.93 |
| 2 | squared\_error |  |  | 0.91 |
| 3 | friedman\_mse |  |  | 0.90 |
| 4 | absolute\_error |  |  | 0.95 |
| 5 | poisson |  |  | 0.72 |
| 6 | squared\_error | best | auto | 0.91 |
| 7 | absolute\_error | Random |  | 0.93 |
| 8 | absolute\_error | Random | log2 | -1.10 |
| 9 | absolute\_error | Random | auto | 0.75 |
| 10 | absolute\_error | Random | sqrt | 0.63 |
| 11 | squared\_error | best | sqrt | 0.41 |
| 12 | squared\_error | best | log2 | 0.75 |
| 13 | squared\_error | best |  | 0.92 |
| 14 |  | best |  | 0.91 |
| 15 | friedman\_mse | best | auto | 0.90 |
| 16 | friedman\_mse | best | sqrt | 0.88 |
| 17 | friedman\_mse | best | log2 | 0.58 |
| 18 |  |  | log2 | 0.45 |
| 19 |  |  | sqrt | 0.70 |
| 20 |  |  | auto | 0.90 |
| 21 | absolute\_error | best | log2 | 0.78 |
| 22 | absolute\_error | best | auto | 0.93 |
| 23 | absolute\_error | best | sqrt | 0.86 |
| 24 | absolute\_error | best |  | 0.96 |
| 25 | poisson | best | auto | 0.67 |
| 26 | poisson | best | sqrt | 0.54 |
| 27 | poisson | best | log2 | 0.41 |
| 28 | poisson |  |  | 0.75 |

**Conclusion**

**Models R- Score values**

1. MLR = R Value is 0.93
2. SVM = R Value is 0.95
3. DT = R Value is 0.96

Since “Decision tree” gives higher R\_score (0.96) than that of the other models, thus it gives lesser error. Hence, I will save and deploy this model to the production environment so as to allow my business to give input and able to get the output or predicted profit.