**ASSIGNMENT IN MACHINE LEARNING REGRESSION**

1. **Problem Statement:**

Based on the customer category the Insurance charges should be predict using Machine Learning Algorithm.

1. **Dataset Details:**

Number of rows and columns = 1338 rows × 6 columns

Input Columns: Age, BMI, Children, sex, Smoker

Output Column: Charges

1. **Pre-processing:**

In the dataset the sex column and Smoker column having nominal datas that is converted into numerical data.

1. **Support Vector Machine**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C | Kernel -  linear -R\_score | Kernel-Non linear | | |
| rbf | sigmoid | poly |
| C=10 | 0.462468414233968 | -0.03227329390671052 | 0.03930714378274347 | 0.28747069486976207 |
| C=100 | 0.6288792857320358 | 0.3200317832050831 | 0.5276103546510404 | 0.6179569624059795 |
| C=1000 | 0.7649311738596382 | 0.8102064851758545 | 0.28747069486976207 | 0.8566487675946569 |
| C=1100 | 0.7649488704789191 | 0.8102064851758545 |  | 0.858368618430418 |

1. **Decision Tree**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hyper tuning parameters** | | | | |
| **Criterion** | **Splitter** | | **Max features** | **R score** |
| Squared error | best | | Auto | 0.6796953481207556 |
| Sqrt | 0.6843187719214159 |
| Log2 | 0.5625874065207388 |
| random | | Auto | 0.7423791701709614 |
| Sqrt | 0.6956732924974154 |
| Log2 | 0.6799728128152169 |
| Friedman \_mse | best | Auto | | 0.6984524549626255 |
| Sqrt | | 0.7457462062762186 |
| Log2 | | 0.720400545454866 |
| random | Auto | | 0.6712190765194331 |
| Sqrt | | 0.6867626374627028 |
| Log2 | | 0.609235638058919 |
| MSE | best | Auto | | 0.7082794116169788 |
| Sqrt | | 0.7499038708267396 |
| Log2 | | 0.698868852005945 |
| random | Auto | | 0.7440644282342701 |
| Sqrt | | 0.6860969841372521 |
| Log2 | | 0.6680315030471853 |
| MAE | best | Auto | | 0.6801994245732044 |
| Sqrt | | 0.7100344297274037 |
| Log2 | | 0.7031649937442285 |
| random | Auto | | 0.711448424004641 |
| Sqrt | | 0.7236865973050574 |
| Log2 | | 0.7343056762775368 |

1. **Random Forest**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hyper tuning parameters** | | | | |
| **Criterion** | **n\_estimator** | | **Max \_features** | **R score** |
| Squared error (default) | 50 | | 1.0 (default) | 0.8498823801108987 |
| Sqrt | 0.8694967410829306 |
| Log2 | 0.8694967410829306 |
| 100(default) | | 1.0 (default) | 0.8539235792996565 |
| Sqrt | 0.8709953890446224 |
| Log2 | 0.8709953890446224 |
| Friedman \_mse | 50 | 1.0 (default) | | 0.8499979973593528 |
| Sqrt | | 0.8700444213599546 |
| Log2 | | 0.8700444213599546 |
| 100(default) | 1.0 (default) | | 0.8540051307771396 |
| Sqrt | | 0.8709457661833552 |
| Log2 | | 0.8709457661833552 |

1. **Comparative Result Analysis for Insurance Charges Prediction using Machine Learning Algorithms**

|  |  |  |
| --- | --- | --- |
| SI. No | Algorithm | R score |
| 1. | Multiple Linear Regression | 0.7894790349867009 |
| 2. | Support Vector Machine  kernel-poly | 0.858368618430418 |
| 3. | Decision Tree  criterion-Friedman\_mse, splitter=best, max features=sqrt | 0.7457462062762186 |
| 4. | Random Forest  Criterion=Friedman\_mse, n\_estimators=100,max features=sqrt | 0.8709953890446224 |

1. **Best Model**

Random Forest Regression- r\_score = 0.8709953890446224