**Problem Statement or Requirement:**

A client’s requirement is, he wants to predict the insurance charges based on

the several parameters. The Client has provided the dataset of the same.

As a data scientist, you must develop a model which will predict the insurance

charges.

1. **Identify your problem statement:**

Insurance charge prediction – need to predict the insurance charges based in various parameters

1. **Tell basic info about the dataset (Total number of rows, columns):**

Dataset contains various columns such as age, sex, bmi, children, smoker and charges.

Number of columns – 6

Number of rows – 1338

1. **Mention the pre-processing method if you’re doing any (like converting string to number – nominal data):**

Pre-processing method has been done

1. For all model: Convert nominal (categorial) col to numeric using one hot encoding
2. For creating SVM model – Standardize the inputs
3. **Develop a good model with r2\_score. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model:**

Random Forest model has been deployed as it shows best result and it is named as “finalized\_model\_Random\_Forest.sav”

1. **All the research values (r2\_score of the models) should be documented. (You can make tabulation or screenshot of the results.):**

Regression Algorithms

**Multiple Regression:**

1. Slope/weight: array([[ 257.8006705 , 321.06004271, 469.58113407, -41.74825718, 23418.6671912 ]])
2. Bias/intercept: array([-12057.244846])
3. R2\_score = 0.7894

**Support Vector Machine (SVM):**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Kernel** | **Gamma** | **R Value** |
| 1 | rbf | scale (default) | 0.8696 |
| 2 | rbf | auto | 0.8696 |
| 3 | poly | scale (default) | 0.8579 |
| 4 | poly | auto | 0.8579 |
| 5 | sigmoid | scale (default) | -368141.4907 |
| 6 | sigmoid | auto | -368141.4907 |
| 7 | precomputed | scale (default) | Required square matrix |
| 8 | precomputed | auto | Required square matrix |
| 9 | linear | scale | 0.7414 |
| 10 | linear | auto | 0.7414 |

**Decision Tree:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.No** | **Criterion** | **Max Feature** | **Splitter** | **R Value** |
|  | squared\_error (default) | auto | best (default) | (keyword error) |
| 1 | mse | auto | best (default) | 0.7094 |
| 2 | mse | auto | random | 0.6821 |
| 3 | mse | sqrt | best | 0.7305 |
| 4 | mse | sqrt | random | 0.6985 |
| 5 | mse | log2 | best | 0.6944 |
| 6 | mse | log2 | random | 0.7057 |
| 7 | friedman\_mse | auto | best | 0.6870 |
| 8 | friedman\_mse | auto | random | 0.7106 |
| 9 | friedman\_mse | sqrt | best | 0.5956 |
| 10 | friedman\_mse | sqrt | random | 0.6404 |
| 11 | friedman\_mse | log2 | best | 0.7901 |
| 12 | friedman\_mse | log2 | random | 0.6769 |
|  | absolute\_error | auto | best | (keyword error) |
| 13 | mae | auto | best | 0.6766 |
| 14 | mae | auto | random | 0.7458 |
| 15 | mae | sqrt | best | 0.7338 |
| 16 | mae | sqrt | random | 0.7340 |
| 17 | mae | log2 | best | 0.7357 |
| 18 | mae | log2 | random | 0.7044 |
| 19 | poisson | auto | best | 0.6666 |
| 20 | poisson | auto | random | 0.6521 |
| 21 | poisson | sqrt | best | 0.7134 |
| 22 | poisson | sqrt | random | 0.7299 |
| 23 | poisson | log2 | best | 0.5791 |
| 24 | poisson | log2 | random | 0.7171 |

**Random Forest:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Criterion** | **Max Feature** | **R Value** |
|  | squared\_error (default) | auto | (keyword error) |
| 1 | mse | auto | 0.8498 |
| 2 | mse | sqrt | 0.8695 |
| 3 | mse | log2 | 0.8695 |
| 4 | friedman\_mse | auto | 0.8500 |
| 5 | friedman\_mse | sqrt | 0.8702 |
| 6 | friedman\_mse | log2 | 0.8702 |
|  | absolute\_error | auto | (keyword error) |
| 7 | mae | auto | 0.8526 |
| 8 | mae | sqrt | 0.8708 |
| 9 | mae | log2 | 0.8708 |
| 10 | poisson | auto | 0.8275 |
| 11 | poisson | sqrt | 0.8287 |
| 12 | poisson | log2 | 0.8287 |

1. **Mention your final model, justify why u have chosen the same:**

Random Forest model has shown better result (r2\_score) value when compare to other regression models – Multiple linear regression, Support Vector Machine (SVM) and Decision Tree.