## 1) Problem Statement Identification

To predict the insurance charges based on several parameters

DOMAIN SELECTION:

1) Machine Learning
2) Supervised Learning
3) Regression

## 2) Data Set

No. of Rows: 1338

No. of Columns: 6

## 3) Pre Processing Method

Preprocessing **Nominal Data** using **One Hot Encoding**

## 4) Machine Learning Algorithms

1. Multiple Linear Regression

2. Support Vector Machine

3. Decision Tree

4. Random forest

## 5) Comparison of Research values of all algorithms

1. **Multiple Linear Regression - 0.7894**

2. **Support vector machine**

| S.NO | R_Score Value | | | | |
|---|---|---|---|---|---|
| | Hyper Parameter | Linear | Rbf (Non-Linear) | Poly | Sigmoid |
| 1 | C=0.01 | -0.0888 | -0.0896 | -0.0895 | -0.089 |
| 2 | C=1.0 | -0.010 | -0.083 | -0.075 | -0.0754 |
| 3 | C=10 | 0.4624 | -0.032 | 0.038 | 0.0393 |
| 4 | C=70 | 0.6195 | 0.222 | 0.5190 | 0.4849 |
| 5 | C=100 | 0.6288 | 0.3200 | 0.6179 | 0.5276 |

| | | | | | |
|---|---|---|---|---|---|
| 6 | C=200 | 0.6356 | 0.4790 | 0.7516 | 0.5455 |
| 7 | C=300 | 0.6871 | 0.5584 | 0.7926 | 0.5000 |

### 3. Decision Tree

| S.NO | Criterion | Max Features | Splitter | R_Score Value |
|---|---|---|---|---|
| 1 | squared_error | None | best | 0.6893 |
| 2 | squared_error | None | random | 0.6951 |
| 3 | squared_error | sqrt | random | 0.6430 |
| 4 | squared_error | sqrt | best | 0.7485 |
| 5 | squared_error | Log2 | best | 0.5753 |
| 6 | squared_error | Log2 | random | 0.6379 |
| 7 | absolute_error | sqrt | random | 0.5887 |
| 8 | absolute_error | sqrt | best | 0.7707 |
| 9 | absolute_error | log2 | best | 0.7237 |
| 10 | absolute_error | log2 | random | 0.7578 |
| 11 | absolute_error | None | random | 0.7319 |
| 12 | absolute_error | None | best | 0.6722 |
| 13 | friedman_mse | None | best | 0.6931 |
| 14 | friedman_mse | None | random | 0.6772 |
| 15 | friedman_mse | sqrt | best | 0.6690 |
| 16 | friedman_mse | sqrt | random | 0.5938 |
| 17 | friedman_mse | Log2 | best | 0.7087 |
| 18 | friedman_mse | Log2 | random | 0.7019 |
| 19 | poisson | None | best | 0.7283 |
| 20 | poisson | None | random | 0.7291 |
| 21 | poisson | sqrt | best | 0.7670 |
| 22 | poisson | sqrt | random | 0.6753 |
| 23 | poisson | Log2 | best | 0.7018 |
| 25 | poisson | Log2 | random | 0.6165 |

**4. Random Forest**

| S. NO | n_estimators | Criterion | Max Features | R_Score Value |
|---|---|---|---|---|
| 1 | 50 | squared_error | None | 0.8529 |
| 2 | 50 | squared_error | sqrt | 0.8684 |
| 3 | 70 | squared_error | Log2 | 0.8705 |
| 4 | 50 | friedman_mse | None | 0.8555 |
| 5 | 50 | friedman_mse | sqrt | 0.8680 |
| 6 | 90 | friedman_mse | Log2 | 0.8721 |
| 7 | 100 | absolute_error | None | 0.8540 |
| 8 | 100 | absolute_error | sqrt | 0.8739 |
| 9 | 200 | absolute_error | Log2 | 0.8720 |
| 10 | 50 | poisson | None | 0.8551 |
| 11 | 50 | poisson | sqrt | 0.8669 |
| 12 | 100 | poisson | Log2 | 0.8676 |

## 6) Best Model

**Random Forest** algorithm gives a best model based on the comparison of research values (r_score value= 0.87) of all algorithms.