## Multiple Linear Regression

- R2 Score = 0.7894

## Support Vector Machine

| S.No | Kernel | C | R2 Score |
|---|---|---|---|
| 1. | linear | 1000 | 0.6340 |
| 2. | rbf | 2000 | -0.1077 |
| 3. | poly | 3000 | 0.0489 |
| 4. | sigmoid | 4000 | -20.720 |
| 5. | linear | 2000 | 0.6893 |
| 6. | linear | 3000 | 0.7651 |
| 7. | linear | 4000 noSC | 0.7648 |
| 8. | linear | 4000 withSC | 0.7413 |

## Decision Tree

| S.No | citerion | splitter | Max_features | R2 Score |
|---|---|---|---|---|
| 1. | Squared_error | best | - | 0.7205 |
| 2. | Squared_error | best | auto | 0.6855 |
| 3. | Squared_error | best | sqrt | 0.6102 |
| 4. | Squared_error | best | Log2 | 0.7157 |
| 5. | Squared_error | random | auto | 0.7128 |
| 6. | Squared_error | random | sqrt | 0.5629 |
| 7. | Squared_error | random | Log2 | 0.6324 |
| 8. | Friedman_mse | best | auto | 0.7056 |
| 9. | Friedman_mse | best | sqrt | 0.7244 |
| 10. | Friedman_mse | best | Log2 | 0.6499 |
| 11 | Friedman_mse | random | auto | 0.7167 |
| 12 | Friedman_mse | random | sqrt | 0.6325 |

| | | | | |
|---|---|---|---|---|
| 13 | Friedman_mse | random | Log2 | 0.6875 |
| 14 | Absolute_error | best | auto | 0.6802 |
| 15 | Absolute_error | best | sqrt | 0.7506 |
| 16 | Absolute_error | best | Log2 | 0.7026 |
| 17 | Absolute_error | random | auto | 0.6715 |
| 18 | Absolute_error | random | sqrt | 0.6542 |
| 19 | Absolute_error | random | Log2 | 0.6516 |
| 20 | poisson | best | auto | 0.7073 |
| 21 | poisson | best | sqrt | 0.7004 |
| 22 | poisson | best | Log2 | 0.6873 |
| 23 | poisson | random | auto | 0.6745 |
| 24 | poisson | random | sqrt | 0.6688 |
| 25 | poisson | random | Log2 | 0.6890 |

## Random Forest

| S.No | N_estimators | criterion | Max_features | R2 Score |
|---|---|---|---|---|
| 1. | 100 | Squared_error | 1.0 | 0.8532 |
| 2. | 25 | Squared_error | 1.0 | 0.8514 |
| 3. | 25 | Absolute_error | auto | 0.8407 |
| 4. | 25 | Absolute_error | sqrt | 0.8611 |
| 5. | 25 | Friedman_mse | auto | 0.8508 |
| 6. | 25 | Friedman_mse | Log2 | 0.8644 |
| 7 | 25 | poisson | sqrt | 0.8611 |
| 8 | 25 | poisson | Log2 | 0.8644 |

- Problem Statement Identification

  Predict the insurance charges based on the several parameters
- Data preprocessing - is nominal data so the technique used is One Hot Encoding
- Many models have created but random forest algorithm gives higher accuracy with corresponding parameters. So the higher accuracy model is saved, loaded, and moved to deployment phase.