## Regression Assignment :

**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.**

1)      **Stage - 1 :- Machine Language**

            **Stage - 2 :- Supervised Learning**

            **Stage - 3 :- Regression (Numeric)**

2)      **dataset 1338 rows × 6 columns**

3)      **In preprocessing for converting string to number get_dummies function used.**

4)      **Final model is SVM regression is giving the best result.**

5)      **Models are below :**

            **MultipleLinear_Regression**        **0.7894**

            **Random Forest - Regression**        **0.871 (n_estimator = 100)**

            **Support Vector Machine Regression**    **0.8779**

| DECISION TREE REGRESSION (11-04-2024) | | | | |
|---|---|---|---|---|
| S.no | criterion | max features | splitter | r2 value |
| 1 | squared_error | None | best | 0.6997 |
| 2 | squared_error | None | random | 0.7266 |
| 3 | squared_error | sqrt | best | 0.6924 |
| 4 | squared_error | sqrt | random | 0.6518 |
| 5 | squared_error | log2 | best | 0.7021 |
| 6 | squared_error | log2 | random | 0.6009 |
| 7 | friedman_mse | None | best | 0.6784 |
| 8 | friedman_mse | None | random | 0.6446 |
| 9 | friedman_mse | sqrt | best | 0.6961 |
| 10 | friedman_mse | sqrt | random | 0.6928 |
| 11 | friedman_mse | log2 | best | 0.6767 |
| 12 | friedman_mse | log2 | random | 0.7053 |
| 13 | absolute_error | None | best | 0.6717 |

| | | | | |
|---|---|---|---|---|
| 14 | absolute_error | None | random | 0.7359 |
| 15 | absolute_error | sqrt | best | 0.687 |
| 16 | absolute_error | sqrt | random | 0.724 |
| 17 | absolute_error | log2 | best | 0.7068 |
| 18 | absolute_error | log2 | random | 0.727 |
| 19 | poisson | None | best | 0.7262 |
| 20 | poisson | None | random | 0.6747 |
| 21 | poisson | sqrt | best | 0.6977 |
| 22 | poisson | sqrt | random | 0.6026 |
| 23 | poisson | log2 | best | 0.7357 |
| 24 | poisson | log2 | random | 0.695 |

Decision Tree  Regression S.no :- 14 & 23 are the best value r2 = 0.7359 and 0.7357

(criterion='absolute_error', splitter='random',max_features=None)

(criterion='paisson', splitter='best',max_features=log2)

| RANDOM FOREST REGRESSION (14-04-2024) | | | | |
|---|---|---|---|---|
| S.no | n_estimators | criterion | max features | r2 value |
| 1 | 100 | squared_error | None | 0.8538 |
| 2 | 100 | squared_error | sqrt | 0.871 |
| 3 | 100 | squared_error | log2 | 0.871 |
| 4 | 100 | friedman_mse | None | 0.854 |
| 5 | 100 | friedman_mse | sqrt | 0.871 |
| 6 | 100 | friedman_mse | log2 | 0.871 |
| 7 | 100 | absolute_error | None | 0.852 |
| 8 | 100 | absolute_error | sqrt | 0.871 |
| 9 | 100 | absolute_error | log2 | 0.871 |
| 10 | 100 | poisson | None | 0.8526 |
| 11 | 100 | poisson | sqrt | 0.868 |
| 12 | 100 | poisson | log2 | 0.868 |

As per Random Forest Regression r2_value 0.871 is repeating.

6)   **SVM Regressor** is the final Model.
   *In this model I got best result compare with other models.*