Machine Learning Regression R² Value Report

# #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

# #Domain - Machine Learning. Because it contains numerical inputs and outputs

# #Learning - Supervised Learning. Because client are well defined inputs and outputs.

# #Model - Regression

# 2.) Tell basic info about the dataset (Total number of rows, columns)

# #1338 rows × 6 columns

# 3.) Mention the pre-processing method if you’re doing any (like converting

# string to number – nominal data)

# #To find the which type of categorical data - find nominal data to Convert the numerical the inputs - Data Collection

# dataset=pd.get\_dummies(dataset,drop\_first=True,dtype=int)

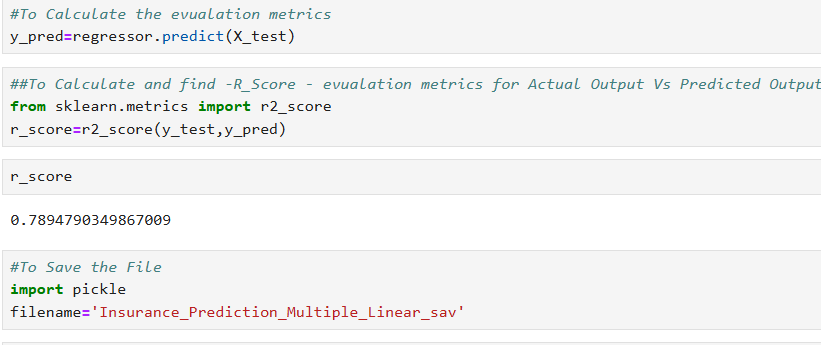
# 4.) Develop a good model with r2\_score. You can use any machine learning

# algorithm; you can create many models. Finally, you have to come up

# with final model.

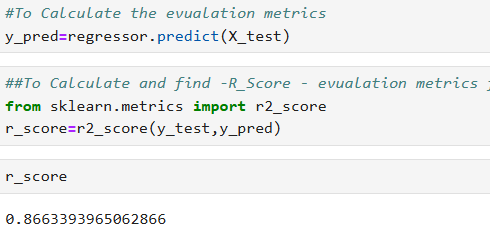
# 1. Multiple Linear Regression

R² value = 0.7894



# 2. Support Vector Machine

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | **Hyper Parameter** | **linear (R²)** | **RBF (Non-Linear R²)** | **Poly (R²)** | **Sigmoid (R²)** |
| **1** | **C10** | **0.46** | **-0.03** | **0.038** | **0.039** |
| **2** | **C100** | **0.62** | **0.32** | **0.61** | **0.52** |
| **3** | **C500** | **0.76** | **0.66** | **0.82** | **0.444** |
| **4** | **C1000** | **0.76** | **0.81** | **0.85** | **0.28** |
| **5** | **C2000** | **0.74** | **0.85** | **0.8605** | **0.054** |
| **6** | **C3000** | **0.74** | **0.866** | **0.85** | **-2.214** |



**Best R²: 0.8669 using rbf kernel with C3000 hyper parameter**

# 3. Decision Tree

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Criterion** | **Split** | **R² Score** |
| **1** | **Friedman mse** | **Best** | **0.71** |
| **2** | **Squared error** | **Best** | **0.78** |
| **3** | **Absolute error** | **Best** | **0.66** |
| **4** | **Poisson** | **Best** | **0.73** |
| **5** | **Friedman mse** | **Random** | **0.65** |
| **6** | **Squared error** | **Random** | **0.73** |
| **7** | **Absolute error** | **Random** | **0.688** |
| **8** | **Poisson** | **Random** | **0.66** |

**Best R²: 0.73 using poission with Best**

**4.Random Forest**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | **Hyper Parameter** | **n\_estimators**  **r2** | **max\_depth**  **r2** | **max\_features** | **random\_state** |
| **1** | **50** | **0.840** | **0.8535** | **0.8535** | **0** |
| **2** | **100** | **0.850** | **0.8535** | **0.8535** | **0** |
| **3** | **500** | **0.851** | **0.8535** | **0.8535** | **0** |
| **4** | **1000** | **0.851** | **0.8535** | **0.8535** | **0** |
| **5** | **2000** | **0.854** | **0.8535** | **0.8535** | **0** |

**Best R²: 0.8535 using max\_depth with tree count = 50**

**5. Result Comparison Table:**

|  |  |
| --- | --- |
| **Model** | **R² Score** |
| **Linear Regression** | **0.78** |
| **#Support Vector Machine** | **0.86** |
| **Decision Tree** | **0.73** |
| **Random Forest** | **0.853** |

**6. Final Recommendation:**

As part of the Insurance Charge Prediction project, we have successfully developed and evaluated multiple regression models to determine the best approach for predicting insurance charges based on customer information such as age, BMI, smoking status, and other features.

### 🔍 Models Evaluated:

* **Multiple Linear Regression**
* **Support Vector Regression (SVR)**
* **Decision Tree Regressor**
* **Random Forest Regressor**

### ✅ Final Recommendation:

After a thorough comparison using the **R² Score** **0.86**, I pleased to recommend the **Support Vector Machine** as the best performing model.

🔹 **Why Support Vector Machine?**

* Achieved the highest R² Score of **0.86**, indicating excellent predictive power.
* More robust and less prone to overfitting compared to other models.
* Performs well even on unseen or new data, ensuring long-term reliability.

I have attached the complete project report, dataset summary, and the trained model for your reference and further use.