Q1) Explain Linear Regression in Detail.

ANS.

- Linear Regression is one of the simplest statistical techniques used in Data Analysis and Machine Learning.
- It is a Method used to find the relationship between a dependent variable and one or more independent variables by fitting a linear Equation to observed data.
- Linear Equation :-

$$y = \beta_0 + \beta_1 x + \epsilon$$

where,

$y \Rightarrow$ Dependent variable
$x \Rightarrow$ Independent variable
$\beta_0 \Rightarrow$ Interrupt
$\beta_1 \Rightarrow$ Coe-fficient
$\epsilon \Rightarrow$ Error Term.

- The Independent variables are the features of the given data using which target variable can be calculated.

STEPS:-

1. Data Collection.
2. Modelling.
3. Training.
4. Prediction
5. Evaluation

- TYPES of Linear Regression:-
  - Simple Linear Regression : Involves one independent variable
  - Multiple Linear Regression : Involves more than one independent variable.

Q2) Compute SST, SSR, SSE, MSE, RMSE, R Square for the given Example.

ANS.

1. SST stands for 'Total Sum of Squares.'

$$SST = \sum (y_i - \bar{y})^2 \quad \text{...(Formula)}$$

SST = 630

2. SSE stands for 'sum of Squares for Error'

$$SSE = \Sigma(\hat{y}_i - \bar{y})^2 \quad \text{...(Formula)}$$

$$= 327.397$$

3. SSE (Error sum of squares for Regression)

$$SSE = \Sigma(y_i - \hat{y}_2)^2 \quad \text{...(Formula)}$$

$$SSE = 302.603$$

4. MSE (Mean Squared Error)

$$\therefore MSE = \frac{SSE}{n} \quad \text{...(Formula)}$$

$$M.S.E. = 65.497$$

5. RMSE (Root Mean squared Error)

$$\therefore RMSE = \sqrt{SSE}$$

$$RMSE = 8.092$$

6. R^2 (Coefficient of Determination)

$$R^2 = 1 - \frac{SSE}{SST} \quad \text{...(Formula)}$$

$$R^2 = 0.480$$

Q.3> Comment whether the Model is Best Fit or not based on the calculated values

Ans. The Linear Regression Model does not provide a very good fit for the Data.

- With an R^2 of only 0.48 and a high RMSE, the Model captures only 48% of the relationships leaving a large portion of variability un-explained.

Q4) Write a python code to calculate the R Square For Boston Dataset.

ANS. CODE:-

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
From sklearn.linear_model import Linear Regression
From sklearn.metrics import r2_score.

df = pd.read_csv('housing.csv') // Reading Boston Dataset.
x = df.drop('MEDV', axis = 'columns')
y = df.MEDV

x_train, x_test, y_train, y_test = train_test_split (x, y, test_
                                    size = 0.3)

model = LinearRegression()
modelFit (x_train, y_train)
y_prod = model.predict (x_test)

r_squared = r2_score( B y_test, y_prod)
print( f " R-Squared value : {r_squared} ")
mse = mse_squared_error( y_test, y_pred)
print ( f " Meanqu squared Error : {mse} ")

# Plotting
plt.figure( figsize(8,6))
plt.scatter( y_test, y_pred)
plt.plot( [min (y_test), max (y_test)], [min(y_test), max(y_test)],
                    color ='red', linestyle = '--')
plt.xlabel ('True Prices')
plt.ylabel ('Predicted Prices.')
```

```
plt.label('True vs Predicted Home Prices')
plt.show()
```