**Name**: Shubham Dahane  
**Class**: CSE(AI) B div  
**Roll No.**: 282015  
**PRN No.**: 22310312  
**Subject**: Machine Learning

**Assignment 6 – Regression Technique on TV Marketing Dataset**

**Objective**

The objective of this assignment is to build a Linear Regression model to analyze the relationship between TV advertising spending and product sales. This helps in predicting sales figures based on investment in television advertisements, thereby supporting strategic marketing decisions.

**Dataset Used**

**TV Marketing Dataset (tvmarketing.csv)**  
This dataset contains marketing spend and sales data with the following attributes:

* **TV**: Amount spent on TV advertisements
* **Sales**: Sales in units (target variable)

**a) Data Preprocessing and Exploration**

* The dataset was loaded and inspected using .head(), .tail(), .info(), .shape, and .describe() to understand the structure and distribution.
* Checked for null or missing values using .isnull().sum(). No missing values were found.
* Initial visualization was done using scatter plots to explore the linear relationship between **TV advertisement spending** and **Sales**.

**Insights:**

* A positive correlation is visually evident between TV spending and sales.
* Data is clean and ready for modeling.

**b) Feature and Label Selection**

* **Feature (X)**: TV advertisement expenditure
* **Target (y)**: Sales  
  Both columns were extracted as Pandas Series/DataFrames for model training.

**c) Train-Test Split**

The dataset was split into:

* **80% Training set**
* **20% Testing set**  
  Using train\_test\_split() from sklearn.model\_selection with random\_state=42 for reproducibility.

**d) Model Training and Prediction**

* A **Linear Regression** model was created using LinearRegression() from sklearn.linear\_model.
* The model was trained on the training set and used to predict on the test set.
* Predictions were stored in y\_pred.

**e) Performance Evaluation Metrics**

| **Metric** | **Value** |
| --- | --- |
| Mean Absolute Error (MAE) | 2.444 |
| Mean Squared Error (MSE) | 10.205 |
| Root Mean Squared Error | 3.194 |
| R² Score | 0.677 |

**Interpretation:**

* The R² score indicates that around **67.7% of the variation** in Sales can be explained by TV advertisement spending.
* MAE and RMSE values are moderate, suggesting a decent fit for a simple linear model.

A graph with blue dots

AI-generated content may be incorrect.

**f) Visualization of Regression Results**

1. **Test Set Prediction Plot:**
   * X-axis: TV advertisement spend
   * Y-axis: Sales
   * Blue points: Actual values
   * Red line: Model predictions
2. **Complete Dataset Visualization:**
   * Overlaid regression line on full dataset to show overall model behavior.

Both plots confirm a **linear upward trend** between TV spend and sales.

A graph with blue dots and a red line

AI-generated content may be incorrect.

A graph with blue dots and a red line

AI-generated content may be incorrect.

**g) Additional Insights and Future Scope**

* The model provides a **simple and interpretable baseline** for predicting sales.
* Future improvements can include:
  + **Adding more advertising channels** (e.g., Radio, Newspaper)
  + **Multiple Linear Regression**
  + **Polynomial Regression** for non-linearity
  + **Advanced Models**: Random Forest, XGBoost, or even Neural Networks for larger datasets

**Conclusion**

Linear Regression was successfully implemented to estimate the effect of TV advertising on sales. Despite its simplicity, the model captured key trends in the data. With additional features and refined techniques, predictive performance could be further enhanced for real-world business applications.