**Sales Prediction**

**1. Introduction**

**Problem Statement**

The goal of this project is to build machine learning models that predict sales based on advertising expenditures across various media channels such as TV, Radio, and Newspaper.

**2. Data Overview**

**Data Source**

The dataset used in this project is in a CSV file named `Advertising.csv`. It includes data on advertising spends and sales figures.

**Data Description**

The dataset contains the following columns:

* TV: Advertising expenditure on TV (in thousands of dollars).
* Radio: Advertising expenditure on Radio (in thousands of dollars).
* Newspaper: Advertising expenditure on Newspaper (in thousands of dollars).
* Sales: Sales figures (in thousands of units).

**3. Data Preprocessing**

1. Loading the Data: The dataset is loaded into a DataFrame using pandas.
2. Initial Examination: Inspect the first few rows and overall structure of the dataset.
3. Check for Missing Values and Duplicates: Identify and handle any missing values or duplicate entries.
4. Drop Unnecessary Columns: Remove columns that do not contribute to the analysis, such as an index column.

**4. Exploratory Data Analysis (EDA)**

**Visualizations**

1. Scatter Plots: Examine relationships between sales and advertising spends across different media. For instance, plotting TV spend against Radio spend and observing how they correlate with sales.
2. Correlation Heatmap: Visualize the strength of relationships between different features and the target variable (Sales) using a heatmap.
3. Pair Plot: Explore the pairwise relationships between features to understand their interactions and how they relate to sales.

**5. Model Development**

**Feature Selection**

The features used in the model include:

* TV
* Radio
* Newspaper

**The target variable is:**

* Sales

**Data Splitting**

The dataset is divided into training and testing sets to evaluate the performance of the model. A common split ratio is 67% for training and 33% for testing.

**6. Model Training and Evaluation**

**Linear Regression Model**

1. Training: Fit a Linear Regression model to the training data.

2. Evaluation: Assess the model’s performance on the testing data using metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE). These metrics measure the average prediction error and the average squared prediction error, respectively.

**Random Forest Regressor Model**

1. Training: Fit a Random Forest Regressor model to the training data. This model uses an ensemble of decision trees to make predictions.

2. Evaluation: Evaluate the model using the same metrics (MAE and MSE) to compare its performance against the Linear Regression model.

**7. Conclusions**

The project involved building and evaluating two machine learning models—Linear Regression and Random Forest Regressor—to predict sales based on advertising spends. The models were assessed using MAE and MSE, providing insights into their predictive accuracy. Future work could include trying more advanced algorithms or fine-tuning the existing models for better performance.