

Sales Prediction

1 Introduction

Sales prediction is a critical task for businesses aiming to optimize revenue and make data-driven decisions. Accurate sales forecasting helps organizations plan inventory, allocate advertising budgets, and design effective marketing strategies. With the increasing availability of historical sales and advertising data, machine learning techniques provide powerful tools to model and predict future sales trends.

This project focuses on predicting product sales based on factors such as advertising expenditure and platform selection. By leveraging machine learning models in Python, the goal is to estimate future sales and assist businesses in maximizing their return on investment through informed decision-making.

2 Dataset Description

The dataset consists of historical sales data along with advertising-related features, including:

- Advertising expenditure across different platforms like TV, Radio and NewsPaper.
- Sales figures (target variable).

The target variable is continuous, making this a regression problem.

3 Methodology

3.1 Data Preprocessing and Exploratory Data Analysis

Exploratory data analysis is performed to understand the distribution of sales and the influence of advertising expenditures:

- Checked for missing or inconsistent values and handled them appropriately
- Visualized the distribution of the **Sales** variable using histograms to analyze spread and frequency.
- Examined the distributions of advertising expenditures across **TV**, **Radio**, and **Newspaper** platforms.
- Analyzed the relationship between each advertising channel and sales using scatter plots.
- Computed a **correlation matrix** to quantify the strength of relationships between sales and advertising features.
- Observed strong positive correlation between TV and Radio advertising with Sales, while Newspaper showed weaker influence.

- Split the dataset into training and testing sets in a 80:20 ratio for model evaluation

3.2 Model Training

A **Linear Regression** model is used to predict sales based on advertising expenditure across different platforms. The model is trained using TV, Radio, and Newspaper advertising costs as input features and Sales as the target variable.

3.3 Evaluation Metrics

Model performance was evaluated using standard regression metrics:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R^2 Score

4 Results

4.1 Model Performance

Table 1: Performance of Linear Regression Model

Metric	Value
Mean Absolute Error (MAE)	1.2748
Mean Squared Error (MSE)	2.9078
Root Mean Squared Error (RMSE)	1.7052
R^2 Score	0.9059

The Linear Regression model achieved a high R^2 score of 0.9059, indicating that approximately 90% of the variance in sales is explained by advertising expenditure. TV and Radio advertising demonstrated a strong positive impact on sales, while Newspaper advertising contributed minimally. The low error values further confirm the effectiveness of the model in predicting sales, making it suitable for supporting business decisions related to advertising budget allocation.

4.2 Regression Coefficients

The coefficient values indicate that TV and Radio advertising have a strong positive influence on sales, with Radio having the highest marginal impact. Newspaper advertising shows a minimal effect, suggesting limited contribution to sales growth compared to other platforms. If no money is spent on TV, Radio, or Newspaper advertising, the model predicts sales of approximately 4.71 units.

Table 2: Linear Regression Coefficients

Feature	Coefficient
Intercept	4.7141
TV	0.0545
Radio	0.1009
Newspaper	0.0043

5 Conclusion

This project successfully applies machine learning techniques to predict sales using advertising and related features. Through data preprocessing, regression modeling, and performance evaluation, a predictive system is developed to support business decision-making.

The results emphasize the impact of advertising investment on sales and demonstrate how data-driven approaches can help optimize marketing strategies. Future work may include incorporating additional features such as seasonal effects, customer demographics, and advanced ensemble models to further enhance prediction accuracy.