

APPLIED DATA SCIENCE – Phase 5

Topic:Future Sales Prediction

Team members:

1.K.Jayapriya - (511321104036) - [jayapriyakumaresan02@gmail.com](mailto:jayapriyakumaresan02@gmail.com)

2.S.Anu - (511321104005) [- s.anupriyanka1225@gmail.com](mailto:-%20%20s.anupriyanka1225@gmail.com)

3.M.Charishma - (511321104014) - [Charishmamasarapu@gmail.com](mailto:%20Charishmamasarapu@gmail.com)

4.P.Lahari - (511321104045) [- laharipasala8@gmail.com](mailto:-%20%20%20laharipasala8@gmail.com)

Content:

* Introduction
* Sales prediction
* Problem statement
* Design thinking process
* Model traning process
* Forecasting algorithm
* Evaluation metrics

Required packages and installation:

1.Numpy

2.Pandas

3.Keras

4.Tensor flow

5.CSV

6.Matplotlib.pyplot

INTRODUCION:

Predicting future sales is a critical aspect of business planning and decision-making. In the realm of data science, several techniques and approaches are employed to forecast sales accurately. Here's an overview of how sales prediction is tackled using data science methods:

Data Collection and Cleaning:

Data Sources:

Gathering relevant historical sales data is the first step. This data can include past sales figures, customer demographics, seasonality patterns, marketing expenditure, economic indicators, and more.

Data Cleaning:

Raw data often contains errors and missing values. Data scientists clean and preprocess this data to ensure accuracy in analysis.

2. Exploratory Data Analysis (EDA):

Pattern Identification:

EDA techniques are used to understand the underlying patterns and relationships within the data. Visualization tools help in identifying trends, seasonality, and correlations.

Feature Selection:

Relevant features impacting sales, such as holidays, promotions, and economic factors, are identified through EDA.

3. Feature Engineering:

Creation of New Features:

Data scientists generate new features from existing data to enhance predictive accuracy. For example, deriving metrics like sales velocity or customer lifetime value.

Temporal Features:

Date-related features such as day of the week, month, and holiday indicators are crucial, especially for businesses with seasonal variations.

4. Model Selection:

Regression Models:

Linear regression, polynomial regression, and time-series models like ARIMA (AutoRegressive Integrated Moving Average) are often used for sales prediction tasks.

Machine Learning Algorithms:

Decision trees, random forests, and gradient boosting machines are employed for more complex, non-linear relationships.

Deep Learning:

Networks, particularly recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), are effective for modeling sequential data like time-series sales.

5. Training and Validation:

Splitting Data:

The dataset is divided into training and validation sets. Training data is used to train the model, while the validation set helps tune hyperparameters and assess model performance.

Cross-Validation:

Techniques like k-fold cross-validation ensure the model's reliability by testing it on different subsets of the data.

6. Model Evaluation and Hyperparameter Tuning:

Metrics:

Common metrics for evaluating sales prediction models include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Hyperparameter Tuning:

Models are fine-tuned to optimize their performance. Techniques like grid search or random search are employed to find the best combination of hyperparameters.

7. Deployment and Monitoring:

Deployment:

Once a satisfactory model is developed, it's deployed into the production environment where it can make real-time predictions.

Monitoring:

Continuous monitoring is essential to ensure the model's accuracy over time. If the model's performance degrades, it might need retraining or adjustment of hyperparameters.

8. Advanced Techniques:

Ensemble Learning:

Combining predictions from multiple models often results in a more accurate forecast.

Time Series Forecasting:

Techniques like Prophet (developed by Facebook) are specifically designed for forecasting time-series data, including sales.

9. AI and Predictive Analytics:

AI-Based Predictions:

Artificial intelligence, including natural language processing and sentiment analysis, is integrated to enhance sales predictions, especially in industries heavily reliant on customer feedback and market sentiments. Sales prediction in data science is a dynamic field. The choice of techniques depends on the specific business problem, the nature of the data, and the goals of the prediction.



SALES PREDICTION:

Sales prediction in data science involves using historical sales data and various machine learning algorithms to forecast future sales. There are several steps involved in building a sales prediction model.

1.Data Collection:

Gather historical sales data, which should ideally include information on sales figures, dates, products, prices, promotions, and other relevant factors.

2.Data Preprocessing:

Clean the data by handling missing values and outliers.

Convert categorical variables into numerical representations through techniques like one-hot encoding.

Normalize or scale numerical features to bring them to a similar scale.

3.Feature Selection:

Identify the most relevant features (variables) that affect sales. Feature selection methods like correlation analysis or feature importance from tree-based models can help.

4.Data Splitting:

% for testing Split the dataset into training and testing sets. A common split is 80% for training and 20.

5.Prediction and Visualization:

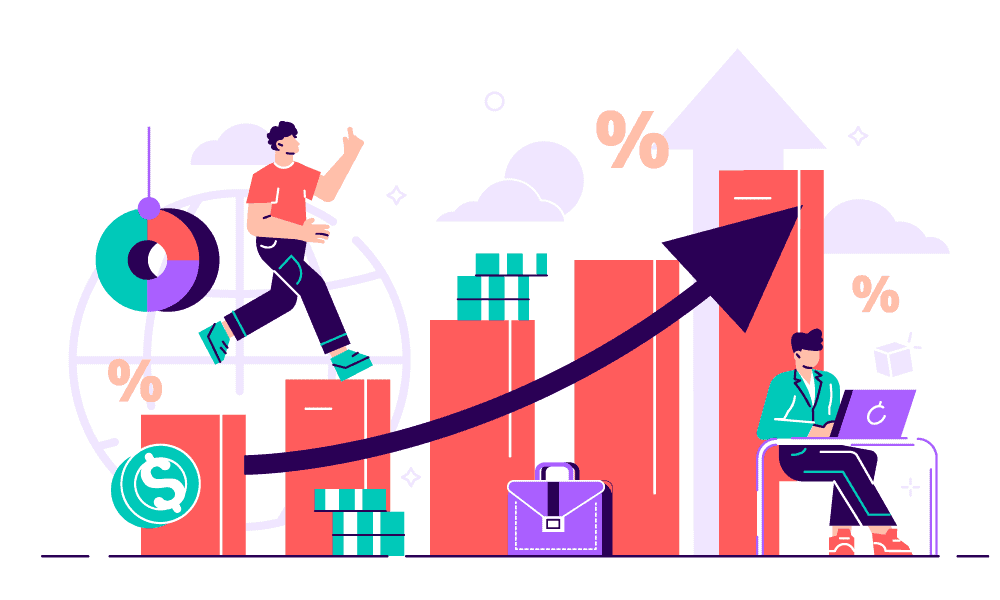
Use the trained model to make predictions on the test data.

Visualize the predicted values against the actual values to assess the model's accuracy visually.

6.Fine-Tuning and Optimization:

Fine-tune the model and retrain it with more data if available.

Perform hyperparameter tuning using techniques like Grid Search or Random Search to optimize the model's performance.



PROBLEM STATEMENT:

 Increase sales by identifying the factors that influence customer behavior.

Objectives:

* Identify the top three factors that influence customer behavior.
* Develop a plan to address the identified factors.

Scope:

* Use sales data, customer data, and marketing data from the past year.
* Focus on a specific product line.

Data:

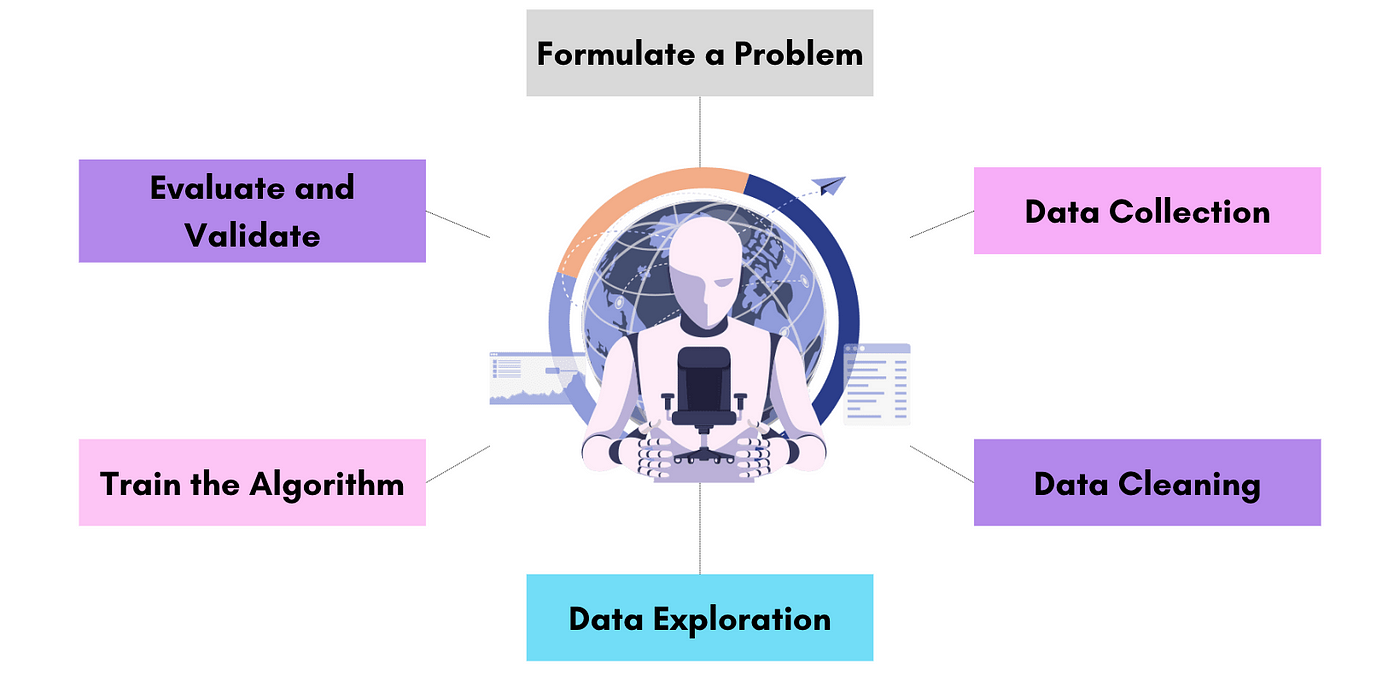
* Sales data
* Customer data
* Marketing data

Methods and Tools:

* Regression analysis
* Decision tree analysis
* Customer segmentation

Success Metrics:

* Increase in revenue after implementing the plan.



DESIGN THINKING PROCESS:

Design thinking can be a powerful approach in shaping the future of sales in the field of data science. By integrating design thinking principles into the sales process, organizations can better understand customer needs, identify innovative solutions, and create products and services that meet those needs effectively.

1. Empathize:

Understanding Customer Needs:

Use data analysis and machine learning algorithms to gain insights into customer behavior and preferences.

User Personas:

Create detailed user personas based on data to understand different customer segments and their unique requirements.

Surveys and Interviews: Conduct surveys and interviews to gather qualitative data, enhancing the quantitative insights obtained from data analysis.

2. Define:

Problem Definition:

Clearly define the problems and challenges faced by customers, backed by data-driven evidence.

Data Analysis:

Utilize data science techniques to analyze patterns and trends, helping in defining specific sales challenges or opportunities.

Define Success Metrics:

Set measurable goals based on data, ensuring that the solutions align with business objectives.

3. Ideate:

Brainstorming Sessions:

Conduct brainstorming sessions involving sales professionals, data scientists, and designers to generate innovative ideas.

Data-Driven Ideation:

Use data analytics to identify potential patterns and correlations, leading to data-driven ideation.

Prototyping:

Develop prototypes of sales solutions using data simulation and modeling to test and validate ideas before full implementation.

4. Prototype:

Data Prototyping:

Create prototypes of data-driven sales tools and applications, allowing stakeholders to visualize the solutions in action.

A/B Testing:

Implement A/B testing using data analysis to compare the performance of different prototypes and choose the most effective ones.

5. Test:

User Testing:

Conduct user testing with real customers, collecting feedback and analyzing data on user interactions.

Iterative Refinement:

Use data feedback loops to iteratively refine the prototypes, making data-driven decisions on necessary improvements.

6. Implement:

Data-Driven Implementation:

Implement data-driven sales solutions, integrating them with existing systems and processes.

Change Management:

Use data analytics to monitor the impact of changes, ensuring that the implementation aligns with the desired outcomes.

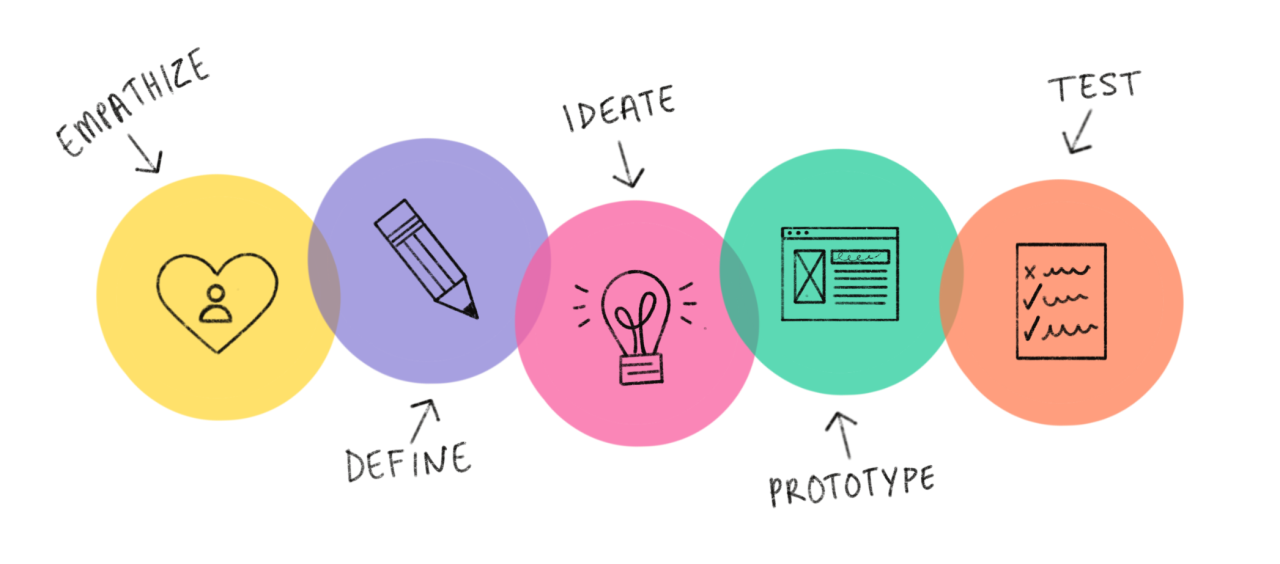
7. Iterate:

Continuous Improvement:

Gather ongoing data from sales activities, customer interactions, and market trends to continuously refine sales strategies.

Predictive Analytics:

Utilize predictive analytics to anticipate future customer needs and market trends, staying ahead of the competition.



Benefits of Applying Design Thinking in Data-Driven Sales:

Customer-Centric Solutions:

Design thinking ensures that sales solutions are tailored to meet the specific needs and preferences of customers, enhancing customer satisfaction and loyalty.

Innovation:

By combining data-driven insights with creative thinking, organizations can innovate new sales approaches, products, and services that are ahead of the curve.

Agility:

Data-driven design thinking allows sales teams to adapt quickly to changing market demands and customer expectations, ensuring agility in the sales process.

Optimized Performance:

Continuous iteration and improvement based on data feedback lead to optimized sales performance, maximizing revenue and efficiency.

MODEL TRANING PROCES:

The process of training a data science model for predicting future sales involves several steps.

1. Model Selection:

Choose appropriate algorithms for time series forecasting or regression tasks. Common choices include Linear Regression, Decision Trees, Random Forest, Gradient Boosting, and Neural Networks (LSTM, GRU for sequential data).

2. Training the Model:

Feed the training data into the selected model.

3. Evaluation:

Use the testing data to evaluate the model's performance. Common metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared.

Iteratively refine the model by adjusting hyperparameters or trying different algorithms based on evaluation results.

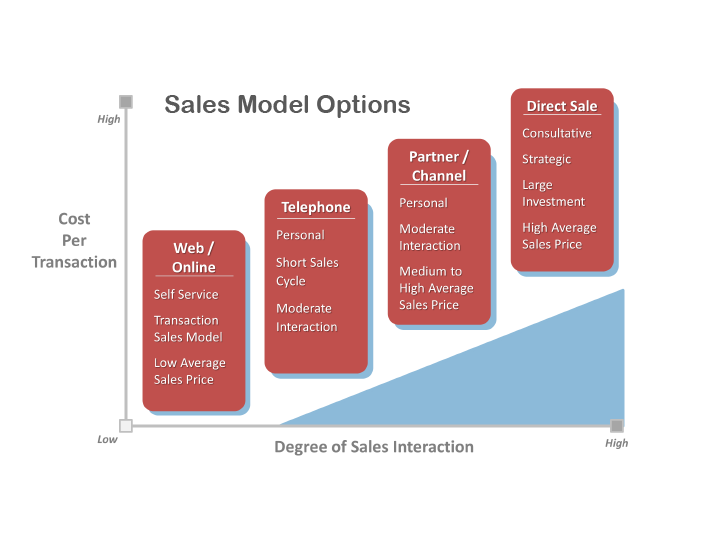
4. Future Enhancements:

Advanced Algorithms:

As technology evolves, consider integrating more complex algorithms like deep learning models for better accuracy, especially when dealing with large and complex datasets.

Automated Machine Learning (AutoML):

Utilize AutoML tools to automate the process of algorithm selection, hyperparameter tuning, and feature engineering.



FORECASTING ALGORITHM:

Linear Regression:

Linear regression models can be used for sales forecasting if there is a linear relationship between the sales and the predictors (features).

Random Forest:

Random forest models can capture complex patterns in the data and handle non-linearity. They are robust and perform well with large and complex datasets.

Gradient Boosting Machines (GBM):

GBM algorithms like XGBoost and LightGBM are powerful for regression tasks, including sales forecasting. They build multiple decision trees to make predictions.

Neural Networks:

Deep learning techniques, such as recurrent neural networks (RNNs) or long short-term memory networks (LSTMs), can capture intricate patterns in sales data. However, they require large amounts of data and computational resources.

SOURCE CODE:

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

import os

import statsmodels.formula.api as sm

from sklearn.linear\_model import LinearRegression, Ridge, Lasso, ElasticNet

from sklearn.metrics import mean\_squared\_error, r2\_score

from sklearn.model\_selection import train\_test\_split

from sklearn.model\_selection import cross\_val\_score

from sklearn.model\_selection import GridSearchCV

import warnings

warnings.simplefilter(action='ignore', category=FutureWarning)

os.getcwd()

df = pd.read\_csv("/kaggle/input/advertisingcsv/Advertising.csv")

df.head()

df.columns

df.rename(columns={'Unnamed: 0': 'Index'}, inplace=True)

df

df.shape

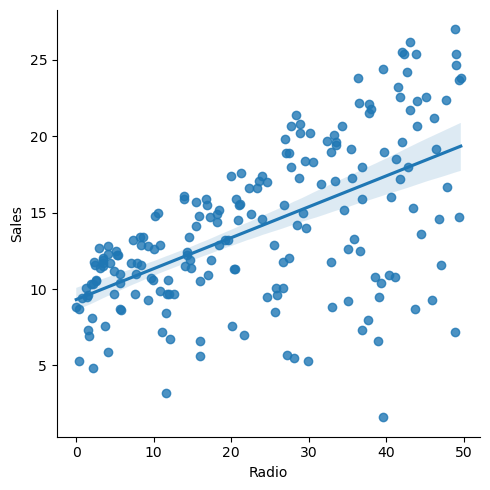
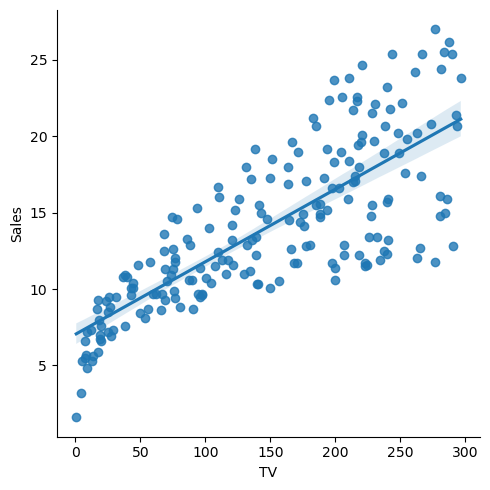
df.info()

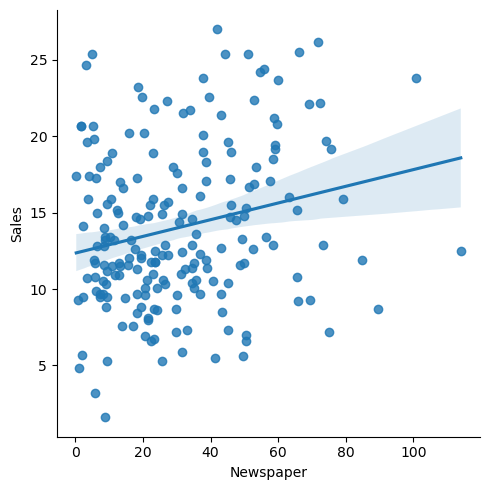
df.describe().T

df.isnull().values.any()

df.isnull().sum()

sns.pairplot(df, x\_vars=["TV", "Radio", "Newspaper"], y\_vars="Sales", kind="reg")



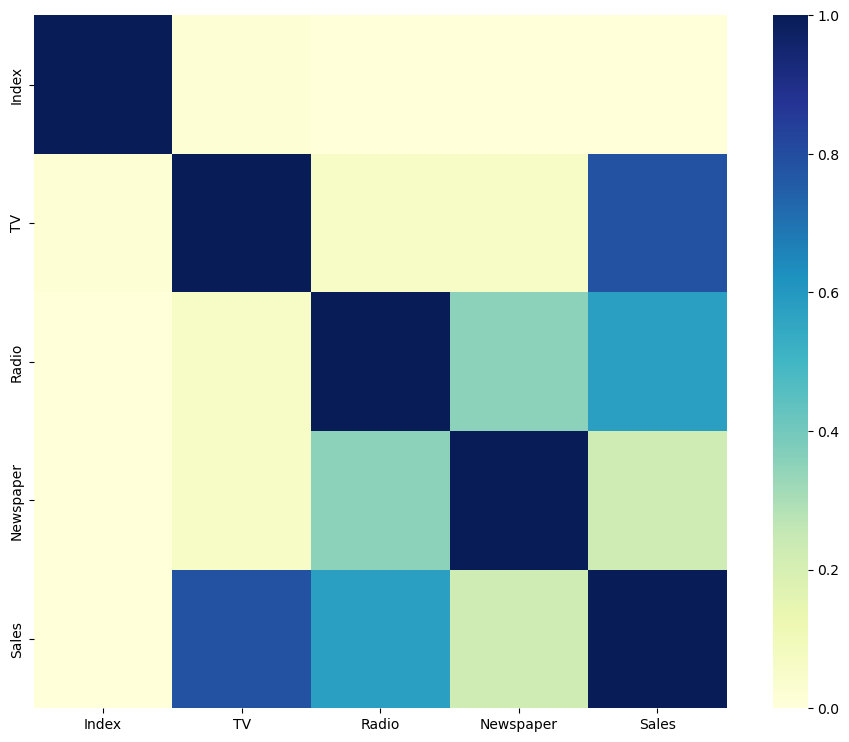


corrmat = df.corr()

f, ax = plt.subplots(figsize=(12, 9))

sns.heatmap(corrmat, vmin=0, vmax=1, square=True, cmap="YlGnBu", ax=ax)

plt.show()



X = df.drop('Sales', axis=1)

y = df[["Sales"]]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.20, random\_state=46)

lin\_model = sm.ols(formula="Sales ~ TV + Radio + Newspaper", data=df).fit()

print(lin\_model.params, "**\n**")

results = []

names = []

For name, model **in** models:

model.fit(X\_train, y\_train)

y\_pred = model.predict(X\_test)

result = np.sqrt(mean\_squared\_error(y\_test, y\_pred))

results.append(result)

names.append(name)

msg = "**%s**: **%f**" % (name, result)

print(msg)

new\_data = pd.DataFrame({'TV': [100], 'Radio': [50], 'Newspaper': [25]})

predicted\_sales = lin\_model.predict(new\_data)

print("Predicted Sales:", predicted\_sales)

new\_data = pd.DataFrame({'TV': [25], 'Radio': [63], 'Newspaper': [80]})

predicted\_sales = lin\_model.predict(new\_data)

print("Predicted Sales:", predicted\_sales)

output:

EVALUATION METRICES: When evaluating future sales, businesses often rely on various metrics and key performance indicators (KPIs) to assess their performance and make informed decisions.

Sales Revenue:

This is the total amount of money generated from sales of products or services. Tracking revenue helps in understanding the overall financial health of the business.

Sales Growth:

Sales growth indicates the increase in revenue over a specific period of time. It is often expressed as a percentage and can be calculated monthly, quarterly, or annually predictable.

Customer Acquisition Cost (CAC):

CAC measures the cost of acquiring a new customer. By understanding how much it costs to acquire a customer, businesses can assess the efficiency of their sales and marketing strategies.

Customer Lifetime Value (CLV):

CLV represents the total revenue a business can expect to earn from a customer throughout their entire relationship. It helps in understanding the long-term value of customers and guides sales and marketing efforts.

Conversion Rate:

Conversion rate measures the percentage of potential customers who take a desired action, such as making a purchase. It is crucial for evaluating the effectiveness of sales and marketing campaigns.

Churn Rate:

Churn rate calculates the percentage of customers who stop using a product or service over a given period. For subscription-based businesses, this metric is vital in predicting future revenue.

Average Purchase Value:

This metric shows the average amount a customer spends on each transaction. Increasing this value can boost overall sales revenue.

Inventory Turnover:

For businesses dealing with physical products, inventory turnover measures how quickly products are sold and replaced within a specific time frame. A high turnover rate indicates efficient sales and stock management.

Lead-to-Customer Conversion Rate:

This metric tracks the percentage of leads (potential customers) that are converted into actual paying customers. Improving this rate can boost future sales by optimizing the sales funnel.

Market Basket Analysis:

This technique analyzes the products that are frequently purchased together. Understanding these patterns can help in cross-selling and upselling, leading to increased sales.

Customer Satisfaction and Net Promoter Score (NPS):

Satisfied customers are more likely to make repeat purchases and recommend the business to others. Monitoring customer satisfaction and NPS scores can provide insights into future sales opportunities.

Sales Pipeline and Conversion Ratios:

Tracking the sales pipeline stages and conversion ratios at each stage helps in identifying bottlenecks and optimizing the sales process for future sales growth.

