## Applied Data Science:

**Project Name: Future Sales Prediction**

**Phase 4: Development Part 2**

**Description:**

In this part you will continue building your project.

Continue building the "Future Sales prediction" model by:

* Feature engineering
* Model training
* Evaluation.

Input:

import pandas as pd

import numpy as no

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.express as px

from datetime import datetime,timedelta

data = pd.read\_csv("C://Users/Admin/Documents/Phase4/Sales\_data.csv",encoding = "ISO-8859-1")

data\_date = data.copy()

data.head(5)

Output:

0 TV Radio Newspaper Sales

1 230.1 37.8 69.2 22.1

2 44.5 39.3 45.1 10.4

3 17.2 45.9 69.3 12

4 151.5 41.3 58.5 16.5

**Feature Engineering:**

Feature engineering is the process of creating new features from existing ones, or transforming existing features in a way that makes them more informative for the machine learning model.

We'll create a new feature called "Total Advertising" by summing the expenses from TV, Newspaper, Radio and Sales.

Input:

import pandas as pd

data = {pd.read\_csv("C://Users/Admin/Documents/Phase4/Sales\_data.csv",encoding = "ISO-8859-1")

}

data\_date = data.copy()

df = pd.DataFrame(data)

df['Total Advertising'] = df['TV'] + df['Newspaper'] + df['Radio']

print(df)

Output:

TV Radio Newspaper Sales Total Advertising

0 230.1 37.8 69.2 22.1 337.1

1 44.5 39.3 45.1 10.4 128.9

2 17.2 45.9 69.3 12.0 132.4

3 151.5 41.3 58.5 16.5 251.3

4 180.8 10.8 58.4 17.9 250.0

…. …. …. ….

**Model Training:**

Model training in machine learning refers to the process of using a dataset, typically with input features (independent variables) and corresponding target values (dependent variable), to teach a machine learning model to make predictions or decisions.

Input:

import pandas as pd

from sklearn.linear\_model import LinearRegression

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

data = { pd.read\_csv("C://Users/Admin/Documents/Phase4/Sales\_data.csv",encoding = "ISO-8859-1")

}

data\_date = data.copy()

df = pd.DataFrame(data)

X = df[['TV', 'Newspaper', 'Radio']]

y = df['Sales']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

model = LinearRegression()

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

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

print("Coefficients:", model.coef\_)

print("Intercept:", model.intercept\_)

print("Predicted Sales:")

print(y\_pred)

Output:

Coefficients: [0.04555832 0.18878124 0.1897103]

Intercept: 2.979067908422406

Predicted Sales:

[12.5181044 13.10443243]

**Evaluation:**

In machine learning, evaluation refers to the process of assessing the performance and accuracy of a trained model on a dataset. It helps determine how well the model generalizes to new, unseen data. Common evaluation metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2).

Input:

import pandas as pd

from sklearn.linear\_model import LinearRegression

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

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

data = {pd.read\_csv("C://Users/Admin/Documents/Phase4/Sales\_data.csv",encoding = "ISO-8859-1")

}

df = pd.DataFrame(data)

X = df[['TV', 'Newspaper', 'Radio']]

y = df['Sales']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

model = LinearRegression()

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

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

mse = mean\_squared\_error(y\_test, y\_pred)

rmse = mean\_squared\_error(y\_test, y\_pred, squared=False)

mae = mean\_absolute\_error(y\_test, y\_pred)

r2 = r2\_score(y\_test, y\_pred)

print("Mean Squared Error (MSE):", mse)

print("Root Mean Squared Error (RMSE):", rmse)

print("Mean Absolute Error (MAE):", mae)

print("R-squared (R2):", r2)

Output:

Mean Squared Error (MSE): 4.233428596014785

Root Mean Squared Error (RMSE): 2.0573745712091754

Mean Absolute Error (MAE): 1.2725084639076798

R-squared (R2): 0.8776191512296577

**Conclusion:**

We had performed feature engineering, model training, evaluation for our future sales prediction dataset.