

# project-2

March 7, 2024

```
[25]: import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error
from math import sqrt

# Load the dataset
data = pd.read_csv("D:/ANACONDA/Advertising.csv")
data.head()
```

```
[25]:      Unnamed: 0      TV  Radio  Newspaper  Sales
0           1  230.1   37.8         69.2    22.1
1           2   44.5   39.3         45.1    10.4
2           3   17.2   45.9         69.3     9.3
3           4  151.5   41.3         58.5    18.5
4           5  180.8   10.8         58.4    12.9
```

```
[26]: data.shape
```

```
[26]: (200, 5)
```

```
[27]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0  200 non-null    int64
1   TV          200 non-null    float64
2   Radio       200 non-null    float64
3   Newspaper   200 non-null    float64
4   Sales       200 non-null    float64
dtypes: float64(4), int64(1)
memory usage: 7.9 KB
```

```
[28]: data.describe()
```

```
[28]:
```

	Unnamed: 0	TV	Radio	Newspaper	Sales
count	200.000000	200.000000	200.000000	200.000000	200.000000
mean	100.500000	147.042500	23.264000	30.554000	14.022500
std	57.879185	85.854236	14.846809	21.778621	5.217457
min	1.000000	0.700000	0.000000	0.300000	1.600000
25%	50.750000	74.375000	9.975000	12.750000	10.375000
50%	100.500000	149.750000	22.900000	25.750000	12.900000
75%	150.250000	218.825000	36.525000	45.100000	17.400000
max	200.000000	296.400000	49.600000	114.000000	27.000000

```
[29]: data.isnull()
```

```
[29]:
```

	Unnamed: 0	TV	Radio	Newspaper	Sales
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
..	...	...	...	...	...
195	False	False	False	False	False
196	False	False	False	False	False
197	False	False	False	False	False
198	False	False	False	False	False
199	False	False	False	False	False

```
[200 rows x 5 columns]
```

```
[30]: data.dtypes
```

```
[30]:
```

Unnamed: 0	int64
TV	float64
Radio	float64
Newspaper	float64
Sales	float64

```
dtype: object
```

```
[31]: X = data[['TV', 'Radio', 'Newspaper']]
      y = data['Sales']

      # Training and testing sets
      X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
      ↪random_state=42)
```

```
[32]: # Initialize and fit the linear regression model
      model = LinearRegression()
      model.fit(X_train, y_train)
```

```
[32]: LinearRegression()
```

```
[33]: # Predicting on the testing set
y_pred = model.predict(X_test)

# Calculating R squared
r_squared = r2_score(y_test, y_pred)
```

```
[34]: # Calculating Mean Absolute Error
mae = mean_absolute_error(y_test, y_pred)

# Calculating Mean Squared Error
mse = mean_squared_error(y_test, y_pred)
```

```
[35]: # Calculating Root Mean Squared Error
rmse = sqrt(mse)

print("R squared:", r_squared)
print("Mean Absolute Error:", mae)
print("Mean Squared Error:", mse)
print("Root Mean Squared Error:", rmse)
```

```
R squared: 0.899438024100912
Mean Absolute Error: 1.4607567168117606
Mean Squared Error: 3.1740973539761046
Root Mean Squared Error: 1.7815996615334502
```