

Practical No-4

Date of Conduction :

Date of Checking:

Data Analytics I

Create a Linear Regression Model using Python/R to predict home prices using Boston Housing Dataset (<https://www.kaggle.com/c/boston-housing>).

The Boston Housing dataset contains information about various houses in Boston through different parameters. There are 506 samples and 14 feature variables in this dataset.

The objective is to predict the value of prices of the house using the given features.

Python Code

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt

# Load the Boston Housing Dataset
boston_data = pd.read_csv('train.csv')

# Explore the dataset
print(boston_data.head())

# Separate features (X) and target variable (y)
X = boston_data.drop(['ID', 'indus'], axis=1)
y = boston_data['indus']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Create a Linear Regression model
model = LinearRegression()

# Train the model
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f'Mean Squared Error: {mse}')
print(f'R-squared: {r2}')

# Plotting predicted vs actual prices
plt.scatter(y_test, y_pred)
plt.xlabel('Actual Prices')
plt.ylabel('Predicted Prices')
plt.title('Actual Prices vs Predicted Prices')
plt.show()
```

OUTPUT

"C:\Users\Ram Kumar Solanki\PycharmProjects\pythonProject\venv\Scripts\python.exe"

"C:\Users\Ram Kumar Solanki\PycharmProjects\MBA_BFS\main.py"

ID	crim	zn	indus	chas	...	tax	ptratio	black	lstat	medv
0	1	0.00632	18.0	2.31	0 ...	296	15.3	396.90	4.98	24.0
1	2	0.02731	0.0	7.07	0 ...	242	17.8	396.90	9.14	21.6
2	4	0.03237	0.0	2.18	0 ...	222	18.7	394.63	2.94	33.4
3	5	0.06905	0.0	2.18	0 ...	222	18.7	396.90	5.33	36.2
4	7	0.08829	12.5	7.87	0 ...	311	15.2	395.60	12.43	22.9

[5 rows x 15 columns]

Mean Squared Error: 15.3919055697973

R-squared: 0.6933129963581055

Process finished with exit code 0

