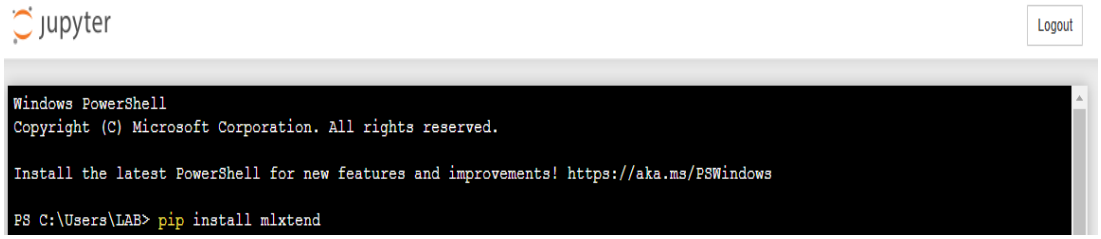
Experiment-5

AIM: Develop a program for Bias, Variance, remove duplicates, Cross Validation:

Install module from terminal:

>pip install mlxtend



import csv file housing.csv from path or user and enter the path where it exists:

for example: 'D:/housing.csv'

program:

```
from pandas import read_csv
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from mlxtend.evaluate import bias_variance_decomp
import numpy as np

# Load dataset
url = 'D:/housing.csv'

dataframe = read_csv(url, header=None)
# Separate into inputs and outputs
data = dataframe.values
X, y = data[:, :-1], data[:, -1]
```

```
# Ensure that X and y are numpy arrays

X = np.array(X)

y = np.array(y)


# Split the data

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)


# Define the model

model = LinearRegression()


# Estimate bias and variance

mse, bias, var = bias_variance_decomp(model, X_train, y_train, X_test, y_test, loss='mse',

                        num_rounds=200, random_seed=1)


# Summarize results

print(f'MSE: {mse:.3f}')

print(f'Bias: {bias:.3f}')

print(f'Variance: {var:.3f}')
```

Output:

```
MSE: 135.372
Bias: 113.311
Variance: 22.061
```