futureskills prime
A MeitY - NASSCOM Digital Skilling Initiative

इलेक्ट्रॉनिकी एवं
सूचना प्रौद्योगिकी मंत्रालय
MINISTRY OF
ELECTRONICS AND
INFORMATION TECHNOLOGY
सत्यमेव जयते

CDAC

Boot Camp on Artificial Intelligence

Practical Assignment - 5

Date of Submission: **Maximum Marks:**

Lab Assignment: Text Preprocessing with Python

Objective:

To clean and preprocess raw text data, preparing it for further analysis or input into machine learning models. Students will use Python libraries such as **NLTK**, **re**, and **Pandas** to complete this task.

Tasks:

You will be working with a dataset containing product reviews. Your task is to:

1. Load the dataset.

2. Perform the following text preprocessing tasks:

   o Convert all text to lowercase.

   o Remove punctuation, numbers, and special characters.

   o Tokenize the text into individual words.

   o Remove stopwords.

   o Perform stemming and lemmatization.

   o Calculate the **Term Frequency (TF)** of words in each review

Dataset:

Use the CSV file named product_reviews.csv