

Boot Camp on Artificial Intelligence

Practical Assignment – 5

Group Members:

Mihir Patil
Siddhesh Patil
Mihir Pande
Rohan Mengal
Shrawani Pande
Yash Mulay
Mubin Tamboli
Shailesh Galande
Dhanashree Patil
Shivani Pandey

PRN:

122B1B213
122B1B218
122B1B200
122B1B175
122B1B201
122B1B178
22TCOMP2071
22TCOMP2024
22TCOMP2093
22TCOMP2109

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:
 - Convert all text to lowercase.
 - Remove punctuation, numbers, and special characters.
 - Tokenize the text into individual words.
 - Remove stopwords.
 - Perform stemming and lemmatization.
 - Calculate the **Term Frequency (TF)** of words in each review

Dataset:

Use the CSV file named product_reviews.csv