**Day 10 — 2 July 2025 (Wednesday)**

**Topic: Text Preprocessing & Cleaning Techniques**

Today we cleaned email text using Python and NLTK. We applied lowercase conversion, punctuation removal, whitespace trimming, and stemming. The trainer explained that clean text improves ML performance and reduces model confusion. We also discussed why spam emails often use capital words and symbols to bypass filters. This session helped me understand how preprocessing affects model learning, especially in classifying promotional vs legitimate messages.

**Code Practiced:**

import re

from nltk.stem import PorterStemmer

ps = PorterStemmer()

text = "WIN $$$ Cash NOW!!! Limited OFFER!!!"

clean\_text = re.sub('[^a-zA-Z]', ' ', text).lower()

stemmed = " ".join([ps.stem(word) for word in clean\_text.split()])

print(stemmed)