**Day 17 — 11 July 2025 (Friday)**

**Topic: Improving Spam Model — Feature Tuning & Stopwords**

This session focused on improving model accuracy through feature engineering. We tested different **max\_features** values in TF-IDF, removed noise words, and applied stopword removal. We tested performance by tweaking vocabulary size and comparing NB vs Logistic Regression accuracy. I realized how tuning preprocessing improves prediction quality significantly.  
We also tested random unseen emails to ensure model generalization.

**Experiment Example:**

tfidf = TfidfVectorizer(max\_features=5000, stop\_words='english')

X = tfidf.fit\_transform(df['text']).toarray()

Today strengthened my practical confidence in **fine-tuning ML pipelines**.

.