**Day 12 — 4 July 2025 (Friday)**

**Topic: Train-Test Split & Introduction to Naïve Bayes Algorithm**

Today we learned about splitting data into training and testing sets to evaluate ML performance. The instructor introduced the **Naïve Bayes algorithm**, widely used for spam classification due to its probabilistic nature and efficiency in text tasks. We discussed prior probability, likelihood, and Bayes theorem. Understanding why NB works well for small text datasets made me confident in selecting it for our spam classifier.

**Code Practiced:**

from sklearn.model\_selection import train\_test\_split

from sklearn.naive\_bayes import MultinomialNB

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, df['label'], test\_size=0.2)

model = MultinomialNB()

model.fit(X\_train, y\_train)