# **Phase 2: Detect Trending Topics Per Category**

## Steps to Complete Phase 2 – Trend Detection

### **Preprocessing**

Created a Preprocessing Script File: **preprocess\_articles.py** inside the preprocessing/ folder.

Implemented a **clean\_text()** function that lowercases the text, removes punctuation, tokenizes words, removes stopwords, and lemmatizes each word using NLTK. This function is applied to each article by combining the title and description fields into one.

The script processes all CSV files in data/raw/, applies the cleaning function, adds a new cleaned\_text column, and saves the cleaned files to data/processed/ with the same filenames.

### **Tokenization & Vectorization**

### **Topic Extraction / Trend Detection**

### **Ranking & Selection**

### **Output & Presentation**