Week 2 – Dataset & Pre-processing

🎯 Objective
Collect a public waste-image dataset and plan preprocessing for the AI-Driven Waste Classification project (Sustainability theme).

🗂️ Dataset Chosen
Name: Garbage Classification (Kaggle)
Classes & counts: cardboard (393), glass (491), metal (400), paper (584), plastic (472), trash (127)
Why this dataset: Clean, popular, 6 practical categories — ideal for transfer learning.

📁 Data Organization (Google Drive)
AI_Waste_Classification/
- dataset_raw/ ← unzipped Kaggle dataset (class folders inside)
- dataset_clean/ ← will store resized/normalized/split data (next week)
- notebooks/ ← Colab notebook will live here

⚙️ Planned Pre-processing (to run next week)
- **Resize** all images to **224×224**
- **Normalize** pixel values to **[0, 1]**
- **Split** into **Train 80% / Validation 20%**
- Target layout:
  - dataset_clean/train/<class_name>/*
  - dataset_clean/val/<class_name>/*

✅ Week-2 Activities Completed
- Selected Kaggle dataset and documented class counts
- Organized Drive project structure (raw/clean/notebooks)
- Wrote clear preprocessing plan and repo documentation

🧰 Tools (for execution in Week 3)
Google Colab, Python, TensorFlow/Keras, NumPy, Matplotlib

📅 Week-3 Plan (Modeling)
- Use transfer learning (MobileNetV2)
- Train on cleaned dataset, evaluate accuracy
- Save example predictions for final PPT