Hook Document

Every day, millions of recyclable materials are thrown into landfills simply because people do not know how to sort their waste correctly. Cities are losing money, recycling systems are overwhelmed, and contaminated waste streams are accelerating climate impact. Now imagine you've been asked to help a sustainability task force at UVA solve this problem using machine learning.

You've been given access to a real dataset from the UC Irvine Machine Learning Repository called RealWaster, containing thousands of images of everyday waste items—plastic bottles, glass containers, compostable scraps, and more. The task force wants to know: Can a computer vision model reliably distinguish between recyclable, compostable, and non-recyclable trash? And more importantly, can a student just like you build such a system?

In this case study, you are the analyst responsible for designing a simplified but functional image-classification pipeline. You will explore the dataset, learn how to restructure messy real-world labels, and build a basic convolutional neural network (CNN) capable of classifying waste into three broad categories. Along the way, you will confront challenges like class imbalance, noisy categories, and uncertain predictions—exactly the kinds of problems real data scientists face.

Your mission is to produce a deliverable that demonstrates whether machine learning can help everyday people sort their waste more accurately. By the end of this case study, you will not only understand the foundations of image classification, but you will also see how data science can directly contribute to environmental sustainability.

All materials, including data, starter code, and supplemental readings can be found in the GitHub repository linked below: