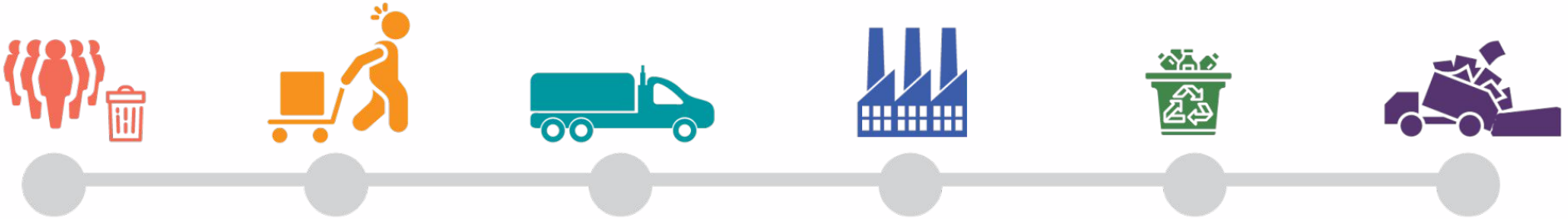
RealWaste Classification

Team 10:
Audrey Sellers, Jaishankar Govindaraj, Pengru Lin

Table of contents

Problem Definition



Waste generation
Residential and commercial waste produced and lying at waste producer's site (before pickup)

Primary collection
Waste and tariff collection from primary producers (and related activities)

Secondary collection
Municipal collection from dumpsters and depots to transfer station

Transfer station management
Monitoring, operations, and evaluation of transfer station activities

Recycling and treatment
Segregation, recycling, and treatment of waste processed at the transfer station

Disposal
Dumping the waste into the landfill

Image source: https://ieg.worldbankgroup.org/evaluations/transitioning-circular-economy/chapter-1-background-and-context

Problem Definition

“The recycling rate has increased from less than 7% in 1960 to the current rate of 32%.”

- Environmental Protection Agency, 2023

“roughly one out of four items (or 25%) are incorrectly placed in the recycling bin.”

- California Management Review, 2023

RealWaste

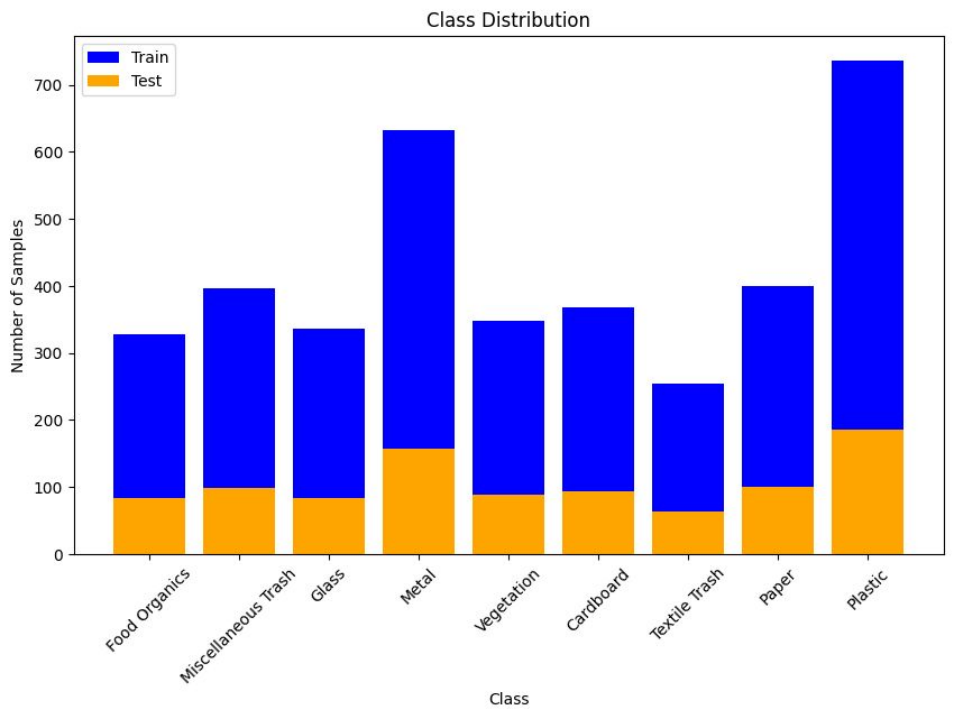
Donated on 10/17/2023

An image classification dataset of waste items across 9 major material types, collected within an authentic landfill environment.

Dataset Characteristics	Subject Area	Associated Tasks
Image	Computer Science	Classification
Feature Type	**# Instances**	**# Features**
-	4752	-

- **Food Organics: 411 images**
- **Miscellaneous Trash: 495 images**
- **Glass: 420 images**
- **Metal: 790 images**
- **Vegetation: 436 images**
- **Cardboard: 461 images**
- **Textile Trash: 318 images**
- **Paper: 500 images**
- **Plastic: 921 images**

Preprocessing



- **Food Organics_weight** = 11.56
- **Miscellaneous_Trash_weight** = 9.60
- **Glass_weight** = 11.31
- **Metal_weight** = 6.02
- **Vegetation_weight** = 10.90
- **Cardboard_weight** = 10.31
- **Textile_Trash_weight** = 14.94
- **Paper_weight** = 9.50
- **Plastic_weight** = 5.16

MODELS

Benchmark - KNN

- Initially performed a Logistic Regression but the scores were really low less than 10% so we disregarded it
- Built **K-Nearest Neighbors (KNN)** model with **3 neighbors** for the classification of the waste dataset into 9 classes
- Reshaped to a consistent format of 524x524 pixels with 3 color channels to standardize input data
- The KNN model showed a **low accuracy of 22.2%**, showing challenges in handling the complex variations in the dataset effectively

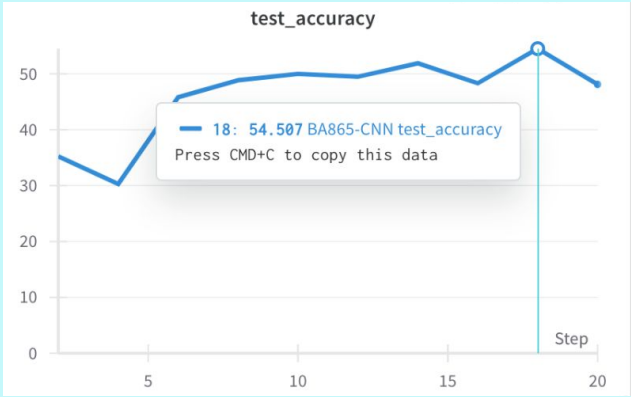
Benchmark - KNN

- Most classes showed low scores across precision, recall, and F1-score
- Class 2 showed a high recall of 83%, precision at low at 16%

	precision	recall	f1-score	support
0	0.09	0.09	0.09	93
1	0.23	0.14	0.18	83
2	0.16	0.83	0.27	84
3	0.28	0.22	0.25	158
4	0.26	0.12	0.16	99
5	0.54	0.13	0.21	100
6	0.32	0.32	0.32	185
7	0.43	0.05	0.08	64
8	0.00	0.00	0.00	88
accuracy			0.22	954
macro avg	0.26	0.21	0.17	954
weighted avg	0.26	0.22	0.20	954

CNN

- **Convolutional Layers** (Conv2D)
- **Max Pooling** (MaxPool2D)
  - Reduce spatial dimensions, but identify important features
  - Helped with reducing computation time
- **ReLU** Activation
- **Loss and Optimization**
  - **Cross Entropy Loss**:
    - Accounted for class imbalance using calculated weights from preprocessing
  - **SGD**:
    - Provides computational efficiency for complex data

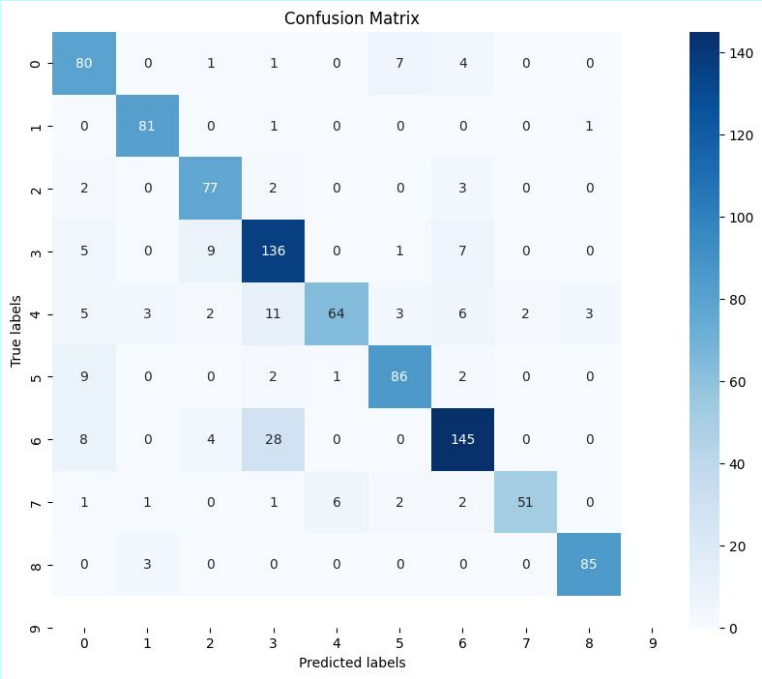


Results

- **54.5%** test accuracy
  - Model classifies better than random guessing

ResNet

- Utilized deep learning for feature extraction and a simple model for classification
- Leveraged a pretrained ResNet model to output high-level features
- **Integration with KNN:** The extracted features were used as inputs to the KNN classifier
- The features from ResNet50 greatly enhanced the accuracy of **85.32%**
- ResNet extracted high-level features from images, which were then passed by KNN for classification.



Real World Applications

- **Waste Management Efficiency**:
  Automated sorting systems in recycling facilities can use this technology to accurately separate different types of waste, increasing recycling rates and reducing the need for manual sorting.

- **Integration with IoT Devices**:
  Embedding this technology in IoT devices that are in waste bins for real-time waste sorting and management, enhancing the capabilities of smart bins and recycling units.

Challenges

- Small dataset for training
- Handling imbalanced classes
- Computational power on both Colab and SCC
    - SCC access at random times
    - Unable to process large batch sizes
    - Too complex of models led to crashing sessions

THANK YOU!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon** and infographics & images by **Freepik**