Problem Definition

Waste management and segregation are major problems hindering India's progress. According to an annual report by CPCB in 2020-21, we generate about 1,60,038 TPD (tonnes per day) of solid waste. 50% of the waste mentioned above is treated by more than 4,500 waste processing plants, half of which are attributed to material recovery facilities. India contributes 12% to global municipal solid waste generation. The above-mentioned statistics suggest that it would be difficult for India to realize the goal of net zero emissions by 2050.

In spite of this, the efforts made by the Indian government to collect and segregate waste and process them is nothing but commendable. With programmes such as 'Swachh Bharat Mission', 'National Water Mission' and 'Waste to Wealth', we are developing solutions to combat waste collection in landfills. The amount of waste sent to landfills has decreased from 54% in 2015-16 to 18.4% in 2020-21, as per CPCB's report. Consequently, India is on the right path to realise net zero emissions.

We would like to help by targeting the waste processing and segregation plants. Our solution identifies the incoming solid waste through images and generates analysis from the given data. We would also perform area-wise analysis that shows which areas are hot spots for a certain type of waste product, following which, better policies can be taken with respect to those areas.

Dataset Description

Our dataset is known as WaRP (Waste Recycling Plant), which is available on Kaggle. This dataset contains three different sub-types of images for different domain applications. These are:

1. WaRP-C for image classification
2. WaRP-D for image detection
3. WaRP-S for image segmentation

We have worked on WaRP-D data that determines the object's presence by drawing a bounding box around it. This dataset contains 2,452 images in the training dataset and 522 images in the validation dataset. The images therein are full-HD with 1920 x 1080 pixels in width and height, respectively. Each image has class-related information present in it along with bounding box information.

Novelty

Our topic here impacts the lives of everyone in India. Our approach to tackling this problem is outlined below :

1. Identification of waste types using image classification techniques
2. Generating a transactional dataset based on the waste identified in the images
3. Infusing the generated data with synthetic area-related information
4. Creating area-wise analysis and relational analysis between different areas

1. Identification of waste types using image classification techniques
We are going to use YOLOv8n deep learning model that can detect objects in images with great efficiency and requires relatively lesser time to train, having about 3 million parameters. YOLO stands for 'You Only Look Once' which stands for its method of only showing the image to the model once. This technique greatly reduces the training time. Another novel feature used by YOLO-based networks is the implementation of a fully-connected convolutional layer. This reduces the multiple passes required for an R-CNN into a single pass, also helping reduce the training time.

2. Generating a transactional dataset based on the waste identified in the images
We assume here that our image set generated from identifying images from data constitutes a single transaction. Under this assumption, we can consider that the waste items in a single image are related to each other.

3. Infusing the generated data with synthetic area-related information
In order to imbue the data with geographical information related to the source of waste, we take 15+ areas in Ahmedabad and randomly allocate transaction data to each area. We then find the frequency count corresponding to each region and generate localized datasets. After performing analysis on these local datasets, we would combine them in a single dataset and perform analysis on the same.

4. Creating area-wise analysis and relational analysis between different areas
On the basis of generated area-wise as well as combined analysis, we then analyze the presence of different waste types in different areas and provide a visualization that can provide insights into relations among different areas and hence generate a better understanding of waste distribution in the city.

This approach of detecting waste images using image classification techniques and providing an area-wise analysis is not common in India and we believe that this approach would provide better insights into the state of waste management and help build solutions for the same.

Algorithms

Algorithm 1 - YOLOv8n algorithm for image classification

Input : A 1920 x 1080 RGB image
Output : Vector that keeps count of items detected in image

```
Configure YOLOv8n model
Train the model on training dataset
Validate on validation data
Predict for test data
Convert output prediction into vector storing count of items found
Generate CSV file for the collected output data
```

Algorithm 2: Apriori Algorithm (Frequent itemset generation)

Input: Dataset
Output: Large itemsets

```
L1 = {large 1-itemsets};
for(k=2; Lk-1≠Φ; k++)
        Ck = apriori_gen(Lk-1);
        forall transaction t ∈ D do
                Ct = subset(Ck, t);
                forall candidate c ∈ Ct do
                c.count++;
        Lk ={c∈Ck|c.count≥min_sup};
        end
end
Answer = Uk Lk;
```

Algorithm 3: K-Means Algorithm (Separating waste for each area)

Input : k, the number of clusters
Transformed data into vectors
Output : Set of desired clusters

```
Specify the number k clusters to assign
Randomly initialize k centroids
repeat
        for each datapoint do
                Assign point to closest centroid
                Recalculate centroid as mean of all points assigned
        end for
```

```
until convergence
```

Algorithm 4 : Fuzzy C-means algorithm

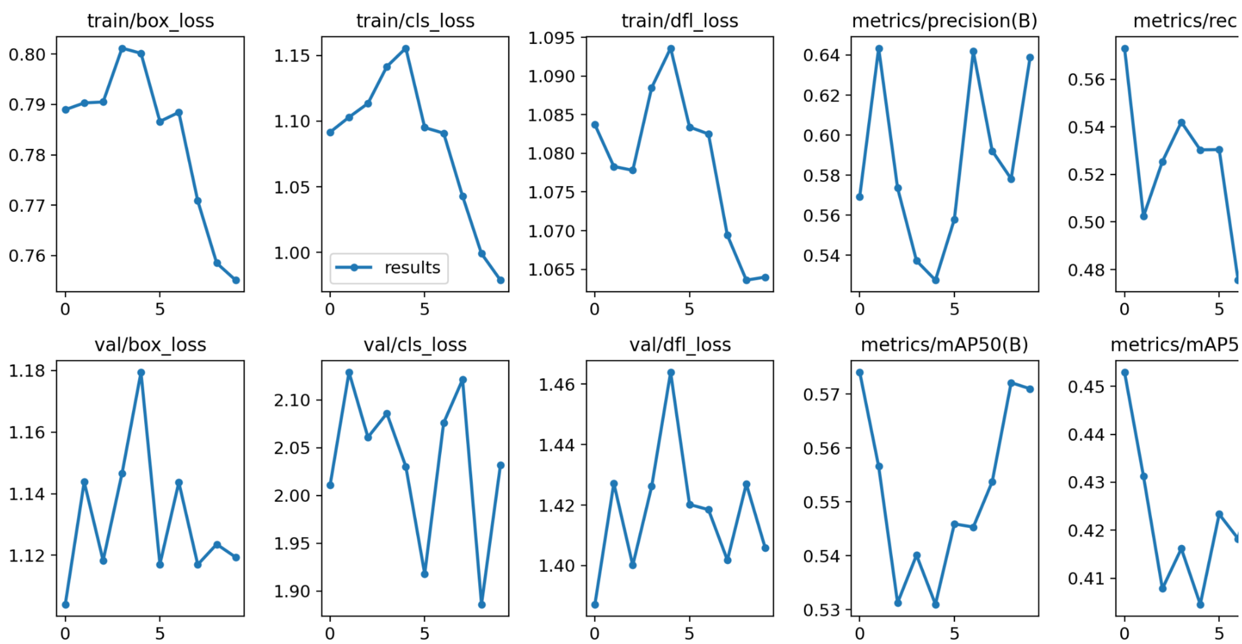
```
Input : k(the number of clusters)
        Transformed data into vectors
Output : Set of desired clusters
Fix k, 2 < k < n;
Fix e (eg : e = 0.001);
Fix MaxIterations;
Choose any inner product norm metric (e.g., Euclidean distance);
Fix m, 1 < m < ∞;
Randomly initialize V0 = v1, v2, ...., vc cluster centers;
for t = 1 to MaxIterations do
        Update the membership matrix U;
        Calculate the new cluster centers Vt;
        Calculate the new objective function Jmt;
        if (abs(Jmt - Jmt-1 ) < e) then
                break
        else
                Jmt-1 = Jmt;
        end if
end for
end
```

Results

We document our results obtained from the techniques and analysis performed above.

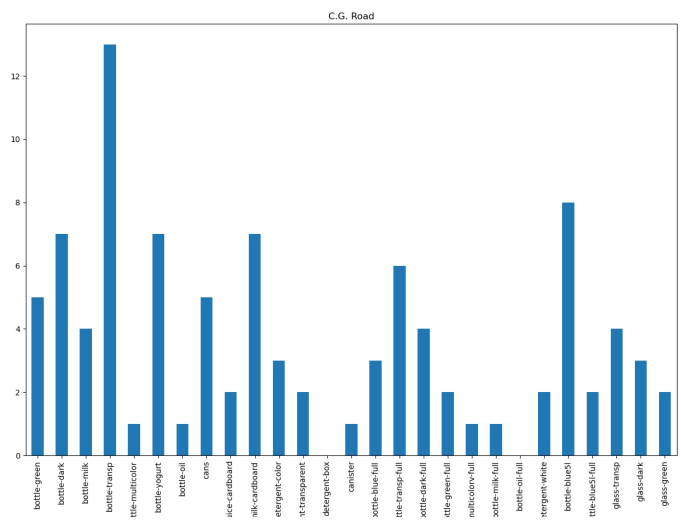
1. Classification model accuracy measures

   Our model YOLOv8n was trained on 2,452 training images. Each epoch was monitored using WANDB.ai, a tool that monitors performance measures as well as model training parameters. We documented various parameters like training loss, box loss and class loss and metrics such as precision, recall and mean average precision over a duration of five epochs. The weights and biases used here have been pre-trained over 40 epochs preceding this training duration. Here are the accuracy measures for our model.

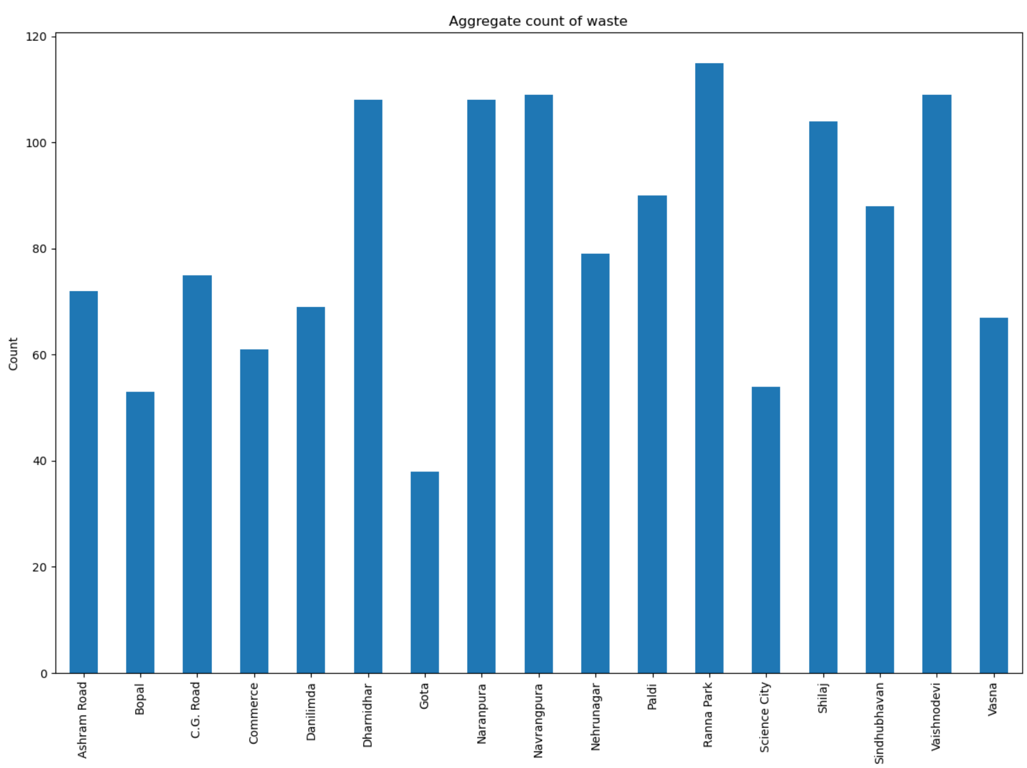


2. Area-wise distribution statistics

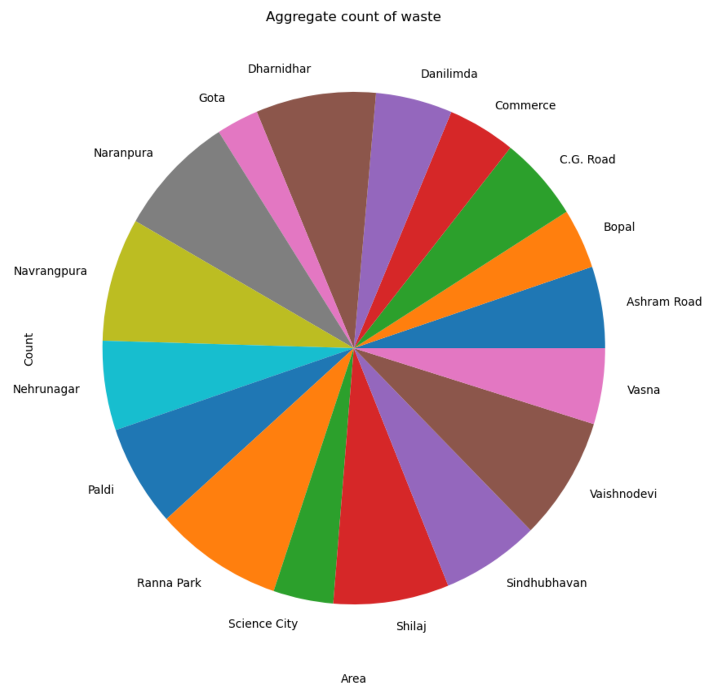
Using randomized area allocation, we have generated the dataset wherein we have distributed waste transactions to different areas. A bar graph representing such distribution for C.G. Road is as follows :



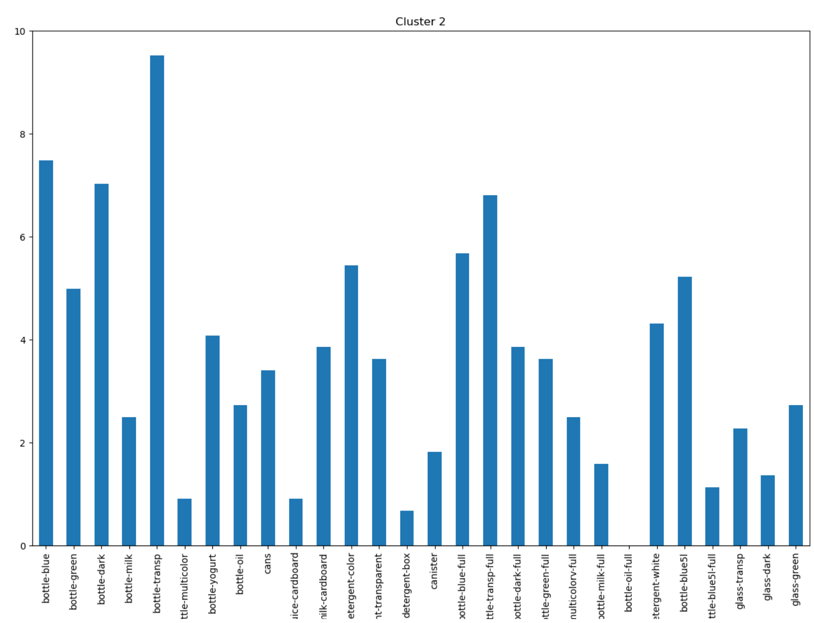
Area wise waste of C.G Road



Area-wise count distribution (bar)



Area-wise count distribution (bar)



Cluster-wise Data of C-2

Result of model's prediction on Validation Batch