Assignment 01: Project Proposal

Group Number: 07

ENEL645: Data Mining and Machine Learning

Ali Mohammadi Ruzbahani , Amirreza Hosseini

Instructor:

Dr. Roberto Medeiros de Souza

Profile of group members

Name	UCID
Ali Mohammadi Ruzbahani	**30261140**
Amirreza Hosseini	**30263633**

Introduction:

Proper sorting of waste is significant in urban centres especially now that recycling as well as proper disposal of waste is of upmost importance. The primary challenge lies in correctly identifying whether the item belongs in one of the following categories:

Table 1: Overview of Experimental Setup

Category	Bin color	Description
General Waste	Black Bin	Non-recyclable and non-compostable materials like, plastic bags, bottle caps, pop cans, candy wrappers, straws, cigarette butts, tissues, used napkins, old flyers, and newspapers.
Organic Waste	Green Bin	This was further categorised into organic which encompass wastes such as food wastes, yard wastes among others.
Recycling	Blue Bin	Others are Recycling particularly plastic, paper and cardboard and glass recyclable items.
Other	N/A	Products that cannot be described as food, animals, medicine or live animals (i.e., dangerous goods, perishable items).

This system will assist every resident understand the proper methods for disposing of waste and separating waste streams in a clean and efficient manner for the city as a whole, increasing the amount of recyclable materials and compostable materials that is collected as waste.

Data Collection:

The quality and the variety of the input data have to be chosen to construct a stable and precise garbage classification system. We will be primarily interested in images along with textual descriptions of each of the trash items we identified under the "Blue," "Green," "Black," and "Other" categories.

1) **Image Data Collection:** The images will be collected using cellphone cameras under the following guidelines:
   a) **Object Centralization:** Every picture will represent an item that will be placed in the centre foreground of the scene. This makes the objects in each image more uniform concerning the rest of the dataset, and thus, the main body can more easily be distinguished by the model.
   b) **Background:** Whenever possible, a homogenised or plain background and technical lighting shall be employed in order to reduce pin and object salience. But some variations will be allowed just to cheque the models' performance under other circumstances.
   c) **Image Quality:** Pictures will be taken using different models of cellphones to ensure high quality and different picture resolution. This in turn will help the model to handle the situation where it is to be implemented across different devices.
   d) **Photo Composition:** This arrangement will be useful to the system due to reduced complexity that will arise from having many items within a single photograph.

2) **Textual Data Collection:** With each image, there will be a short, natural language description of the object in form of a simple three to seven words sentence. The guidelines for collecting textual data include:
   a) **Conciseness:** Labels will be brief but descriptive conveying general characteristics of the object, such as the type (e.g., "greasy pizza box," or "empty plastic bottle").

b) **Standardization:** Descriptions will only be textual and will worn a format that is proven to be quite effective in the collection of such data. For instance, expressing the object's condition by using adjectives such as "used", "dirty", "empty" and then the name of the object.

c) **Natural Language Variability:** Besides following the general format, there is going to be variation in the language so as to come up with actual inputs which should be expected when handling the system.

Experimental Design:

1) **Data Pre-processing:** In order to avoid the problem of having uniform data for training the garbage classification model, the following data pre-processing techniques will be used: Images will be preprocessed for a start by reducing their size to a common width and height so that all the images are square. The pixel values will be preprocessed by normalizing and the like, and the images will be augmented through rotation, cropping and brightness modification. Furthermore, image quality will be examined, and where necessary, poor quality images or images with broken links will be removed.In the case of text cleaning, all text descriptions will be converted to lowercase and all stop words will be eradicated from the text to capture only the important data. Any words will also be stemmed (e.g., changing from 'bottles' to 'bottle') in order to eliminate divergence between the word arrangements in text data. Post pre-processing that will follow is the division of data into training, validation and test set in order to ensure model performance and its ability to generalize on new data.

2) **Model Selection:** However, we suggest a multimodal deep learning framework where both images and texts can be utilized. For the image classification activity, we will use a Convolutional Neural Network (CNN), such as ResNet or MobileNet by applying a transfer learning method. This will help in boosting the accuracy of the proposed model by passing knowledge from the overall image tasks to the specific garbage categorization task.

   As for the text input, we will rather use a deep natural language processing (NLP) model, for example, BERT or a Transformer-based model, for feature extraction from the short sentence given under each photo. The features extracted from the text and image will be merged together, and then fed into a final classifier.

   The architecture will consist of training and optimizing the CNN and the NLP on our dataset and then followed by a classification layer. Our next step will be to divide the data in proportion 4:1:1 into training, validation and testing sets, respectively.

Table 2: Overview of Experimental Setup

Component	Model/Method	Description
Image Classification	ResNet / MobileNet	Pre-trained CNNs for image classification tasks, fine-tuned on our dataset.
Text Classification	BERT / Transformer	Pre-trained models for NLP tasks, fine-tuned to classify garbage-related sentences.
Final Classification	Fully Connected Layer	Combines features from both image and text models for final prediction.

3) **Evaluation Metrics:** Preliminary evaluation metrics include accuracy, precision, recall and F1 score to assess both the image as well as text classification performance from the system. In formulating

the model, cross-validation is going to be used so as to enhance its capability to generalize on other datasets.

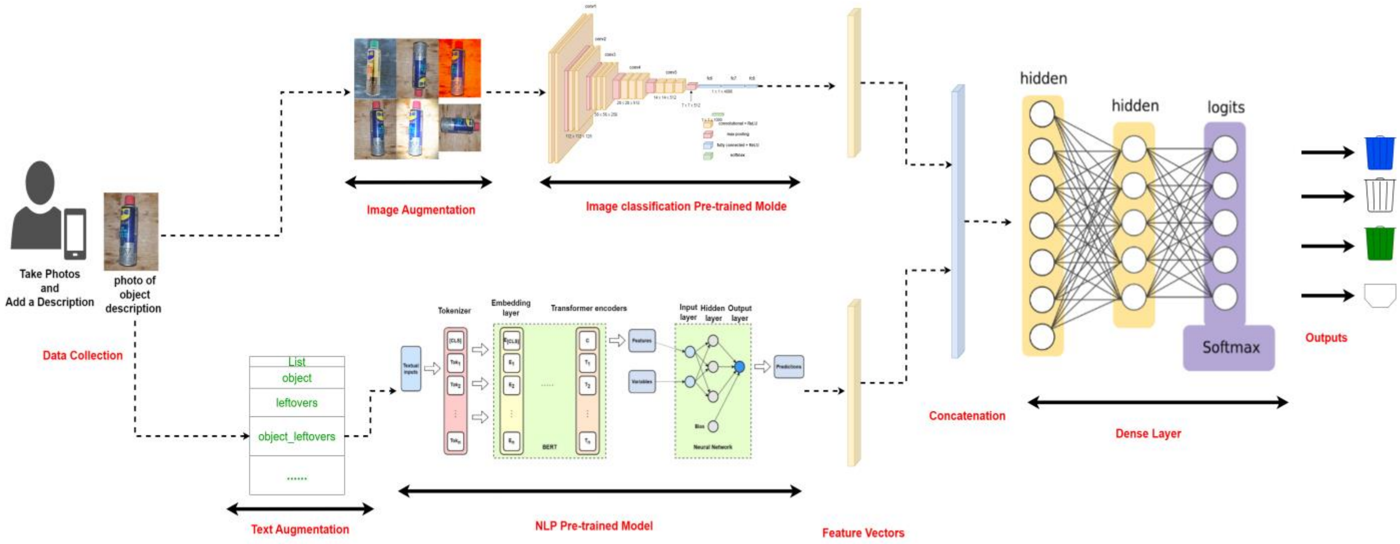

Figure 1. Model workflow involves several steps. Initially, input data undergoes data augmentation techniques to improve the model's learning process. Subsequently, the data undergoes preprocessing before being fed into a pre-trained model to extract feature vectors for both the image and NLP components of the data. Finally, these two parts are combined and analyzed through a dense layer to classify the input into respective garbage classes.

References:

1. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.
2. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning (ICML).
3. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171-4186.
4. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.
5. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.
6. Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization.
7. Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition.
8. Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). mixup: Beyond Empirical Risk Minimization.