Team #: *15*
Project Title: *Bin Buddies*
Team Members: *Mia Patrikios Hezekiah McDonald, Yves Song, Isaiah Weaver, Nia Maheshwari*

Project Proposal Q&A

1. **Who is the end user? Who will be using your machine learning model?**

   Kids and younger audiences that interact with products made of plastic, paper, cardboard, glass, or any other recyclable material.

2. **How does your product make the end user's life significantly better? What is the problem to solve?**

   It can inform kids how to recycle products and build this environmentally friendly habit early on, increasing awareness among the next generation.

3. **Is your problem a classification, regression, segmentation, recommendation, etc.? Why is Machine learning suitable?**

   Classification. Machine learning is suitable because we have a dataset of labeled images. We will be able to define features and targets/labels and create a machine learning algorithm based off of it.

4. **What machine learning technique(s) will you use to solve your problem? Why?**

   We'll be using Supervised Learning, since we have a labeled dataset that includes labeled garbage images.

5. **What is the dataset for your project? Do you have enough data for the selected machine learning technique? How much effort will it take to get the data ready for ML training?**

   We will be using the garbage dataset, which should contain enough data for the supervised machine learning technique. Since the data is labeled, does not contain numerical data, there will not be a lot of data cleaning necessary. However, since all of the values are categorical, we will need to one-hot-encode the data.

6. **What are the baseline model(s) and paper(s) you will use for inspiration? Are you replicating results, improving performance metrics, reducing inference time, or enhancing interpretability?**

   Applying machine learning approach in recycling | Journal of Material Cycles and Waste Management Paper related to using machine learning for recycling via image classification.

We will use a baseline CNN model for the image classification. CNNs are good for this project because they're designed for images, automatically learn patterns like shapes and textures, and work well for classifying objects into categories.

7. **How will you measure model application performance?**
   Examples: Classification: Accuracy, F1, ROC-AUC, Precision/Recall; Regression: MSE, RMSE, R^2; Segmentation: IoU, Dice score

   Since we are using a classification model, we will focus on accuracy. This will generally help us improve the model, but precision will also be used since we really want to avoid wrong positive answers. Tracking the F1 score will also be useful to measure performance because it provides a single number that considers both precision and recall.

8. **Do you have enough computational resources to handle the machine learning technique and the data? Can you train the model with the resources you have?**

   Yes, because all the images (13.9k jpg files) require about 8GB and we can run it on Min Kao Building Tesla or Hydra computers, which should be able to support that.

   Tesla Machines in Room 418
   Memory: 32 GB DDR5 RAM
   Processor: Intel Core i7-14700 vPro
   GPU: NVIDIA GeForce RTX 4060 (8 GB VRAM)
   Storage: 1 TB NVMe SSD

9. **What are the key milestones in your roadmap?**

   Add or delete rows as needed. Grid is helpful way to visualize the milestones you envision

	Tasks Description	**Priority**	**Deadline**	**Time**	**Team Member**
1	**Preprocessing Data** - remove redundant information	**Medium**	**9/18/25**	~1 day	Yves Song
2	**One Hot Encoding for Categories** - convert categorical values to numerical	**Medium**	**9/20/25**	~3-4 days	Nia
3	**Remove Collinear Features and Normalize** -	**Low**	**9/27/25**	~2 days	Mia, Isaiah
4	**Prepare Datasets for Training/Validation** - ensure the	**High**	**9/30/25**		Everyone

	proportions of the classes are equal on both sets				
5	**Preliminary test of Model** - Determine what parameters to change to improve model	**High**	**10/10/25**		Hezekiah
6	**Draft Midterm Project Report**	**Low**	**10/18/25**		**TBD**
7	**Midterm Project Report**	**High**	**10/21/25**		**TBD**
8	**Test Dataset Performance** - Feed in test dataset to model and evaluate accuracy of classification	**High**	**10/28/25**		**TBD**
9	**Revise Dataset Training Model**	**High**	**11/03/25**		**TBD**
10	**Draft final project presentation**	**High**	**11/12/25**		**TBD**
11	**Delivery of project presentation**	**High**	**11/20/25**		**TBD**
12	**Delivery of final project report**	**High**	**12/02/25**		**TBD**

10. **What is the expertise of each team member?**

    Isaiah Weaver - coding in python, making API calls, backend programming, simple UI's

    Mia Patrikios - python, displaying data in digestible format, reading academic papers

    Yves Song - python programming, planning, Systems thinking

    Nia Maheshwari - coding in python, jupyter notebooks, divvying up subtasks

    Hezekiah McDonald -

11. **How will the project tasks be distributed within your team?**

    We will use Github issues to assign tasks to individuals in the group and track contributions.

a. Gather dataset
b. Code the model for testing
c. Calibrate the model through training
d. Workout hosting/API for the model
e. Create user interface

12. Do you want to produce a technical publication out of this project?

No.

References: [Add references or URL to papers, websites, blogs, datasets, etc. used in this proposal.]

https://utk.instructure.com/courses/233844/files/27472938?module_item_id=5736757

https://link.springer.com/article/10.1007/s10163-021-01182-y

https://www.kaggle.com/datasets/zlatan599/garbage-dataset-classification/data