Milestone 3: Design Document Report

Abstract:

This project proposal aims to establish a comprehensive Smart Waste Management and Classification System to enhance waste management practices. By using sensor data, the system will keep track of the fill levels and weights of trash bins, which will aid in predicting future fill rates and determining the frequency of pick-ups. The project aims to streamline collection routes and reduce operational costs by concentrating on bins that are over 50% full, thereby promoting cleanliness in the community.

The system will also generate waste collection notifications based on bin levels, effectively reducing waste overflow. An integral part of the project is an automated waste classification system that uses image recognition technology to classify waste as recyclable or non-recyclable, with data stored in the cloud. This feature not only improves segregation and disposal but also prevents unhygienic accumulation of waste around bins.

In addition to these, the project will gather a comprehensive set of real-world images to improve model generalization. It will also focus on data visualization for detailed analysis of air quality, bin fill levels, and waste weights. The goal is to improve waste collection efficiency, reduce waste overflow, and encourage recycling in the community, thereby making waste management more efficient and cost-effective.

Project Design:

There are 2 phases in this project flow as can be seen in the activity diagrams below. In the initial phase, sensors are deployed to gather data from waste bins, to perform data analysis on the bin fill levels, weight and the surrounding air quality. When the sensors detect that a bin is completely filled, waste collection is initiated. The subsequent phase involves the image classification of the garbage into different categories.

The data collected from the sensors in the bins is stored securely in cloud storage using platforms like Firestore and Amazon S3. This approach not only accommodates the ever-growing data (scalability) but also ensures data durability, meaning the data remains intact. Subsequently, the data preprocessing stage is initiated, wherein data ingestion (Airbyte) is performed to guarantee that the data is systematically collected and prepared for analysis. As part of data preprocessing, standard methods of data imputation are also applied. Data imputation is a technique used to handle missing or incomplete data points, ensuring the dataset is complete and suitable for analysis. For effective data visualization and monitoring, Tableau is utilized as a data visualization tool, facilitating the creation of interactive dashboards and reports. These visualizations assist in keeping track of waste bin fill levels and air quality data, thereby aiding in informed decision-making based on the collected data.

Once the authorities receive alerts triggered by sensor detection, the objective of waste classification is to automatically identify and separate different types of waste such as

recyclable (Paper, metal, glass, plastics etc) and non-recyclable materials (organic, bio wastes etc). The captured images of waste are stored in the Amazon S3 bucket. S3 is commonly used for data backup, archiving purposes and disaster recovery. It can be replicated to multiple S3 buckets across different AWS regions for high availability and data redundancy. To prepare the stored images for analysis, preprocessing methods such as data augmentation, noise reduction, colour correction etc will be applied. Data augmentation includes image transformation, colour manipulation, noise addition, normalization methods that help improve the model's performance by exposing it to a more diverse range of data, reducing overfitting, and enhancing its ability to generalize to unseen examples of garbage images. Further, the Deep Learning models (CNNs) like VGG-16 and ResNet-50 will be used to perform classification of wastes. Transfer learning with these models can save significant training time and resources while providing accurate results in classifying different types of waste.

The VGG-16 excels in capturing fine-grained features in images because of its multiple convolutional layers. Similarly, the ResNet model has the ability to train deep networks effectively, which can capture both low and high-level features in images. Employing these CNN models, we aim to assess and compare their classification performance in the final stage.

Design Diagrams:

The activity diagram in *Figure 1* outlines the process of collecting and analyzing data from garbage bins.

1. Data Collection: The process commences with sensors located inside the garbage bins collecting various data such as fill levels, weight of the trash, and air quality in the vicinity.
2. Garbage Level Check: The sensors issue an alert if the level of garbage in the bin surpasses 50%. If it does not, no alert is sent, thereby reducing the frequency of unnecessary pickups.
3. Data Storage: Post garbage collection, the trash is classified. For this purpose, an image of the trash is stored in an Amazon S3 Bucket. The data collected from the sensor is also stored.
4. Image Data Collection: The system collects image data that is stored in Amazon S3.
5. Image Pre-processing and Analysis: The image data undergoes pre-processing which includes increasing brightness, resizing, and scaling. Subsequently, the pre-processed image is classified using deep learning models.
6. Waste Classification: In the final step, deep learning algorithms such as VGG16 and ResNet50 are employed to classify the waste into different categories like plastic, paper, glass, and trash.

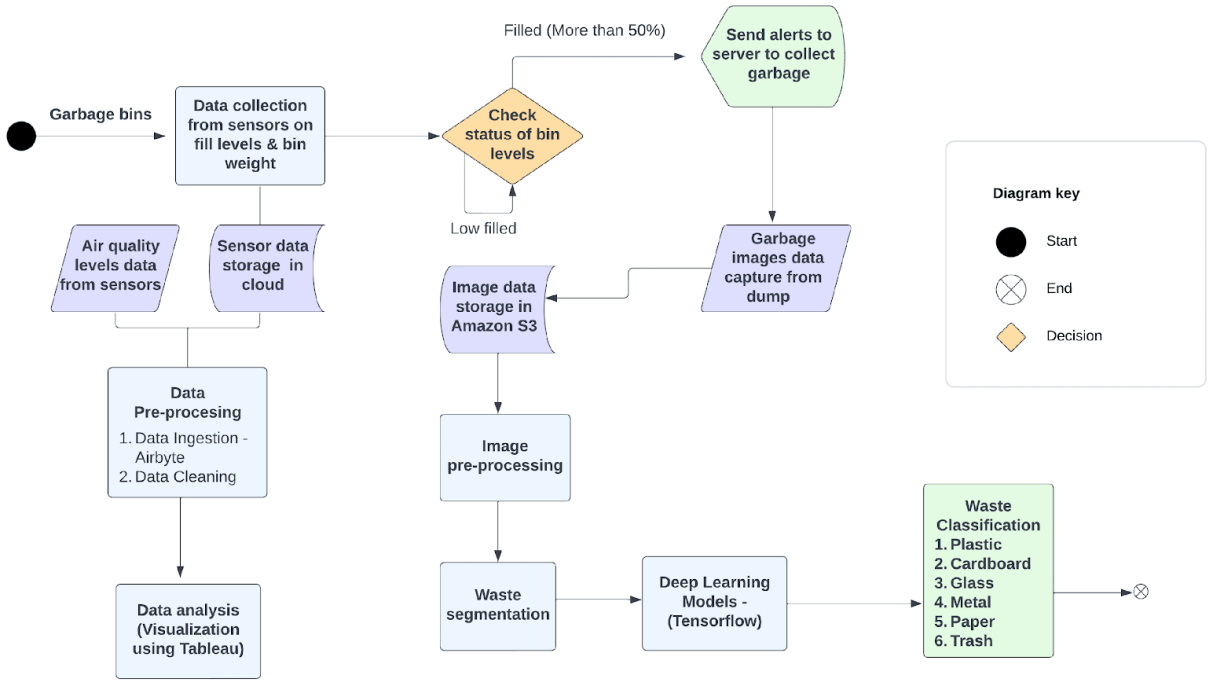

Figure 1

The activity diagram in *Figure 2* outlines the process of collecting and analyzing data from garbage bins.

1. Data Collection: The process commences with sensors located inside the garbage bins collecting various data such as fill levels, weight of the trash, and air quality in the vicinity.
2. Garbage Level Check: The sensors issue an alert if the level of garbage in the bin surpasses 50%. If it does not, no alert is sent, thereby reducing the frequency of unnecessary pickups.
3. Data Collection: The data is collected from the sensors located on bins. This data is then sent to Amazon S3 for storage.
4. Data Pre-Processing: The stored sensor data undergoes pre-processing using tools like Airbyte and Snowflake.
5. Image Pre-processing and Analysis: The image data undergoes pre-processing which includes increasing brightness, resizing, and scaling. Subsequently, the pre-processed image is classified using deep learning models.
6. Waste Classification: In the final step, deep learning algorithms such as VGG16 and ResNet50 are employed to classify the waste into different categories like plastic, paper, glass, and trash.

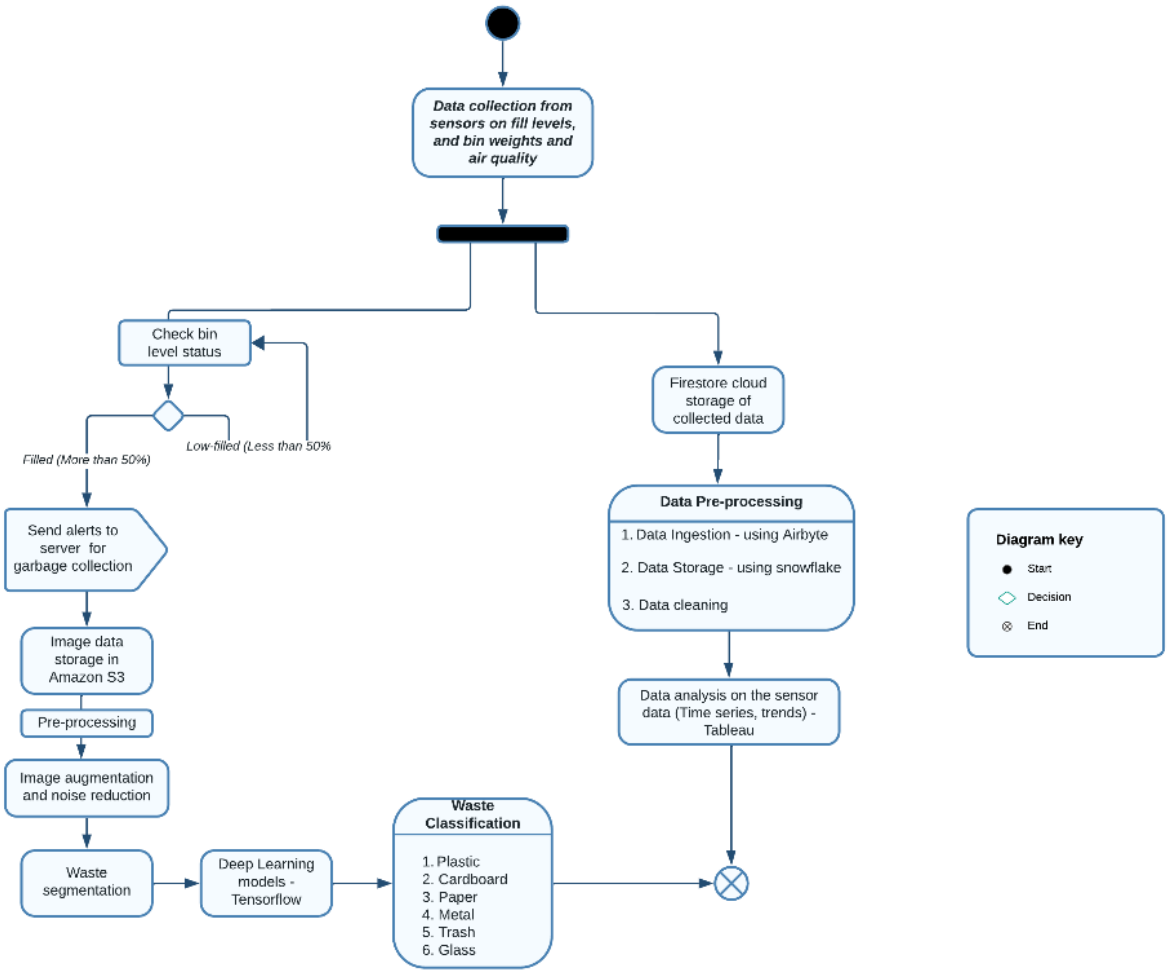

Figure 2

The use case diagram in *Figure 3* outlines the process of garbage collection and disposal.

1. The user interacts with the system, which could be a smart garbage bin or a waste management application.
2. The system plays a crucial role in coordinating the entire process. It communicates with the user, the garbage truck, and the sensors.
3. Sensors, positioned within the garbage bins, gather data pertaining to fill level, weight, and surrounding air quality. Upon reaching a certain threshold, they alert the authorities to dispatch a garbage truck for bin clearance.
4. An authorized individual, possibly a waste management official, supervises the process by ensuring timely dispatch of the garbage truck.
5. The garbage truck is responsible for the collection and transportation of garbage.

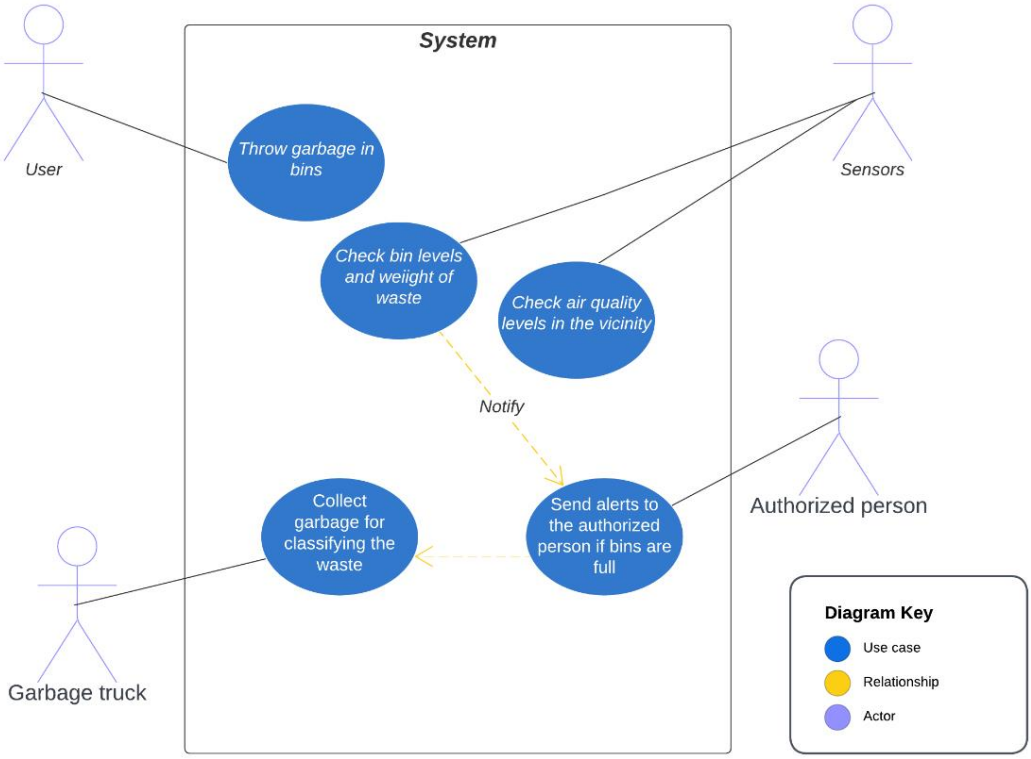

Figure 3

Project Schedule

The goal is to optimize waste management, monitor bin fill levels and air quality, and classify waste materials. The project will be completed in stages, with each prototype adding functionality and complexity.

Prototype 1: Sensor Data Collection and Pre-processing

- Establish a data pipeline to transmit sensor data to a cloud storage system (e.g., Amazon S3).
- Develop data preprocessing algorithms to handle missing values and outliers in sensor data.
- Implement data ingestion using a tool like Airbyte.

Prototype 2: Data Visualization

- Integrate Tableau for data visualization.
- Create interactive dashboards to monitor waste bin fill levels and air quality.

Prototype 3: Waste Classification Model

- Begin collecting a diverse dataset of waste images.
- Develop preprocessing techniques for image data, including data augmentation, noise reduction etc.

Prototype 4 (Week 9-10): Waste classification using CNN

- Integrate pre-trained CNN models (e.g., VGG-16 and ResNet-50) for waste classification.
- Fine-tune the models on the waste image dataset and assess the performance of the classified wastes.