MILESTONE – 4 Final report

Anantha Narayanan:

By understanding the different approaches in designing a data pipeline, I worked jointly on designing the structure of Tech stack 1 and provided my inputs for Tech stack 2, outlining the flow of the components for the garbage classification system. Further, I orchestrated the setup of MinIO as the data lake, for efficient storage of both image and trash can sensor data through object storage. Understanding the intricacies of image data was also a crucial process, where pre-processing steps involving detection of outliers based on image dimensions was carried out, which was challenging. The subsequent step involved visualizing sensor data in Tableau, a crucial aspect for a comprehensive analysis of waste management. Based on the retrieved images, we performed garbage classification using MobileNet V2 deep learning model. Following the ingestion process, I executed the steps to set up and load the pre-processed data from MinIO to the data warehouse used in Tech stack 2. The documentation I compiled offered valuable insights into the comparison of both Tech Stacks implemented and the overall workflow of the system addressing the big data challenges effectively.

Kishan Ramesh:

In the smart waste management project, I collaborated on the design of Tech Stack 1, delineating the component flow in the data pipeline and also contributed my inputs for the design of Tech stack 2. Recognizing the significance of understanding our data before analysis, I contributed to drawing insights into the challenges posed by the sensor data of trash cans and the image data collected for building the classification model. As part of pre-processing of the image data, I implemented data augmentation techniques. This involved applying algorithms for rescaling, resizing, and brightening the images, enhancing the quality of the data for subsequent analysis. A significant portion of my work focused on setting up the hybrid Data Warehouse after completing the data ingestion process. This step was crucial for accommodating versatile data types after they underwent pre-processing. Following this, I actively participated in the analysis of the cleaned data using different tools, to gain meaningful insights on the bin levels, air quality etc. I documented the outcomes derived from the created pipeline, applying Big Data concepts to ensure a comprehensive understanding of the results and implications on waste management.

Anbu Nambi:

My responsibility in the smart waste management project involved designing the Tech stack 2 with additional contributions to the overall pipeline of Tech Stack 2. Gaining insights into both sensor data and image data, a comprehensive analysis of these crucial datasets was performed. This involved implementing standard pre-processing methods such as removing outliers, dropping specific columns, and performing feature extraction to enhance data cleanliness and relevance. After the cleaning process, detailed analytics on the sensor data was conducted using Tableau and plotly for visualization. A pivotal aspect of my contribution was to establishing the MySQL server used in Tech stack 1 to serve as data lake once they are

ingested. The MySQL server served as a centralized repository, for managing and storage of both image and sensor data. I also helped in optimizing the classification model performed for garbage classification that helped achieve an improved prediction accuracy. Overall, the use of Big Data tools to streamline our classification system gave me better understanding, that I meticulously documented throughout the project.

Aparna Shankar:

In the smart waste management project, my role encompassed the designing of pipeline for Tech stack 2, while concurrently providing collaborative inputs for the design of Tech Stack 1. I shared responsibility in understanding sensor data and waste image data collected, before proceeding with analytics and waste classification. The necessary pre-processing steps for the data were implemented to ensure reliability and quality of data. After retrieving the pre-processed data, my major role involved the setup of the data warehouse using PostgreSQL to load the image and sensor data. This step was critical in efficiently loading both image and sensor data, facilitating seamless access to insights derived from the integrated datasets. In the context of waste classification, I focused on enhancing the model performance using MobileNet V2. Visualization of the retrieved data from the warehouse was executed through the creation of a dashboard using Tableau to gain insights. Throughout the project, I diligently documented the outcomes from the pipeline, applying Big Data concepts to provide valuable insights for efficient waste management practices.