**1. Student bi-weekly performance summary**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Adm. No. | Name | No. of hours present | Progress1 | Remarks |
| 1. 2007476 | Johnnie | 8 | A | On schedule |
| 2. 2112790 | Jayden | 8 | B | Distracted by other modules |
| 3. 2112802 | Wee Loon | 8 | A | On schedule |

1 State whether: A=On Schedule B=Ahead Schedule for no. of days C=Behind Schedule for no. of days

**2.** **Weekly Scrum 1**

|  |  |
| --- | --- |
| Week No: 12- 13 Date: 02/01/2023 | |
| Member Name 1: | **Johnnie (Data Engineer)** |
| This week deliverables | * Implemented DASK framework when importing data and used appropriate datatypes to save RAM. * Outlier detection: Apply domain knowledge to univariate outlier detection using interquartile ranges (IQR). * Outlier detection: Apply isolation forest algorithm to remove outliers, some domain knowledge still needed to remove values that do not make sense (e.g., negative speed) * Feature Selection: Used correlation matrix to remove features that I highly correlation (after feature engineering), to prevent multi-collinear issues within data. |
| Obstacles | * Good domain knowledge needed for univariate and even multivariate outlier detection. |
| Member Name 2: | **Jayden (Backlog Owner)** |
| This week deliverables | * Product Backlog: Allocated number of man hours our team will spend for each product (task). * Product Backlog: Created a list of products that we need to complete for project organisation. * Product Backlog: Created Gantt chart for visualisation of project timeline * Data Pre-processing: Imputed missing values in data using Iterative Imputer |
| Obstacles | * Slight difficulty to allocate hours before knowing what issues might come up later and to decide priorities between tasks. * Choosing method of imputation |
| Member Name 3: | **Wee Loon (Data Engineer)** |
| This week deliverables | * Feature Engineering: Aggregated sensor data to summarize relevant information such as maximum, minimum, mean, standard deviation, kurtosis, and skewness. * Feature Engineering: Generated new columns such as age from driver date of birth. * Feature Engineering: Generated net acceleration from acceleration x, y and z. |
| Obstacles | * Difficult to come up with features that are relevant and will be helpful for our machine learning model. * Issues with multicollinearity may arise if too many features are made. E.g., aggregating the sensor data. |