# **Introduction**

Using the NJT data, a classification algorithm is trained to detect odometry anomalies.

Data from Minio is stored in the raw/ folder. A subfolder is present to hold nominal data (in other words, data that has not been "labeled" in the NJT Odometry WP). The data are the original zip files produced by DTG SAT.

# **Steps**

1. The notebook 1-data\_extraction creates csv files with a reduced set of variables from the diagnostic files. The resulting files are stored in the folder filtered/
2. For each month of operation, a file named 20YYMM\_window\_events.csv holds the problematic events that occured (with info of Timestamp, VehicleID, type of error, ...). The notebook 2-window\_event\_extraction formats these files and stores all events logged in a single file named window\_events.csv. Note: this step doesn't need to be rerun every time.
3. 3-window\_event\_creation reads the filtered files and extracts the event defined in the window (from window\_events.csv). The data and event information are stored in a pickle file format under events/. Figures showing the speed profiles are plotted.
4. The filtered data from events are in time series format. In order to reduce the data volume and add information, per event features and created and will be used later on. It is done from the pickle file holding speed data and event description.
5. Anomaly detection