

## Homework 10

Project name: ADL - Cleaning the input data for training self driving cars

Team members: Henri Laats, Kristjan Laid, Helena Sokk

Link to repo: [https://github.com/henrilaats/A5-ADL1-Autonomous\\_Driving](https://github.com/henrilaats/A5-ADL1-Autonomous_Driving)

### Task 2: Business understanding (The report of task 2 should be 400-800 words.)

#### Identifying business goals

- Background -

Tartu University's Autonomous Driving Lab has managed to get the driving recordings obtained with the self-driving Lexus on Rally Estonia tracks. They want to use the gathered data to help them develop a self-driving vehicle themselves. The problem they are facing is that currently the dataset contains information that they don't need or that would influence the outcome. Therefore the dataset is unusable at the moment. That's the part where our team comes in to help them clean the dataset into an usable one.

- Business goals -

The main goal for this project is to help Autonomous Driving Lab, UT with creating data that they can use for developing self-driving cars. In other words we need to produce a cleaned dataset with just proper driving as input for machine learning models.

- Business success criteria -

Since there are three teams doing this project then the ultimate success would be that our team has produced the best cleaned dataset. And it would be actually used by the ADL for machine learning models.

#### Assessing your situation

- Inventory of resources

- Team members: Henri, Helena and Kristjan
- Time: 19 days(At the time of writing). Poster session date: 16th of December at 2pm
- Data: Given by the instructor in the form of a ROS bag
- Hardware: Personal computers
- Software: Jupyter notebook, ROS, Google drive, some poster creating tool and Filmora for editing the presentation video

- Requirements, assumptions, and constraints

We assume that the dataset is quite a mess as the whole aim of this project is to clean it.

A constraint might be the lack of knowledge when it comes to working with ROS and huge data like that.

- Risks and contingencies

The main two risks that we currently see is time management and accessing data as none of us have worked with ROS bags.

- Terminology

We feel that most of the terminology should be self explanatory. Otherwise this part will be updated if a new terminology should occur. Expecting that there might be few car/driving related terminology that we haven't seen/used before.

- Costs and benefits

The cost of electricity and internet connection is probably the biggest cost moneywise, but that isn't something out of reach for any of the members. So the project won't be constrained due to money issues. But probably the biggest cost can be measured in time and effort which availability depends largely on other subjects and obligations.

The benefit of this project is knowledge gain and possibly helping the world of developing self-driving cars.

### Defining your data-mining goals

- Data-mining goals - The main goal is to have the driving dataset cleaned and organized using different data mining methods. The team needs to present the project code of the data preparation and cleaning via github. Meanwhile the team also needs to create a good presentation for the poster session and prepare a video for it.
- Data-mining success criteria - The biggest success would be that we have cleaned the data and separated the data from straight sections of the road and curved regions of the road. Also hopefully we manage to separate parts of the data where the road has intersections. In terms of grading aspect there are 20 points available. Out of the 20 there are 10 points for the project itself and 10 points for presentation. For the team the success criteria is to achieve at least 15 points while the course's minimum criteria is 10 points for the whole project.

### Task 3: Data understanding (The report of task 3 should be 400-800 words.)

- Gathering data requirements

- The necessary data will be all the data given excluding camera images and Lidar data. It will be given in the form of a ROS bag that we have to manage to use in a suitable way.

- Verify data availability

- The data is not publicly available, but we will get access to the virtual machine with the data.

- Define selection criteria - All driving recordings obtained with the self-driving Lexus on Rally Estonia tracks, removing parts of the recording where the car does not move, the turn signal is on and the car is reversing so that as a result we will have a dataset with only proper driving.

- Describing data - Couldn't figure out how to use .bag files, therefore didn't do these tasks. But we contacted our instructor so maybe he could explain to us a little bit what to do or where to start from.
- Exploring data
- Verifying data quality

## **Task 4: Planning your project (The report of task 4 should be 100-300 words.)**

### **Plan for the project with tasks:**

- Installing ROS and getting access to virtual machine with the data. - For each member about 2 hours
- Extract the data you need from the ROS bag - 2 hours
- Start cleaning data and removing things we don't need(car not moving, reversing, turn signal on) - 10 hours total
- Separate the data from straight sections of the road and curved regions of the road - must be differentiated whether the driver was actually turning or avoiding something on the road(but the road was straight) - 10 hours total
- Produce cleaned dataset which contains only proper driving and that can be used for machine learning models. 10 hours
- Creating a poster and preparing to present it. Also a 3 minute video to present the project(5 hours)

### **List of the methods and tools**

- ROS in conda environment
- Github - [https://github.com/henrilaats/A5-ADL1-Autonomous\\_Driving](https://github.com/henrilaats/A5-ADL1-Autonomous_Driving) for uploading the work we've done