

## Create materials to teach data science via self-driving (BSc or MSc)

Developing a course (8 academic hours one-day seminar or a 3 ECTS course) teaching key aspects of data science in a practical way with self-driving toy cars. The particular emphasis is on aspects that become relevant when **deploying a solution to the real world**. The toy cars are equipped with a camera and a Raspberry Pi. The used **machine learning** solution imitates human driving, learning the function between the camera image and the command a human would give when seeing this image. As the task is fun and understandable, the failures are evident (car crashes or behaves weirdly), it is a very good medium to teach data science and machine learning.

The course should cover key concepts that get mentioned in every data science, but rarely practically experienced by students, such as:

- **Garbage in, garbage out.** If your training data is low quality, you get a bad model. Can be exemplified by using too low camera resolution for the task (garbage input) or bad driving examples (garbage training labels).
- **Overfitting** and finding **non-causal relationships** in the data, which later lead to bad generalization ability. Exemplified by visualizing the image areas that influence the decision the strongest (saliency maps). If top half of the image is not cut away, the model will learn to rely on lamp locations in the ceiling, not track walls for deciding when to turn.
- Models **struggle to extrapolate, generalize**. Exemplified by models failing to drive if light conditions change.
- Test set metrics are not the final product, the model will be **deployed in the real world** and will find examples it can not deal with. Or will experience a shift in input distribution. Monitoring performance, characterizing failures and reacting to them is important.
- **Dataset management, iteratively building a better dataset** (to fix the failures observed), and data cleaning are important. Exemplified by results before and after.
- **Computational efficiency** can be important in certain tasks. If the vehicle takes too few decisions per second it will crash. In other domains - if every query costs too much (on compute bill) or takes too long to compute, the product is not viable.
- etc.

The thesis should make use of known theories of developing educational materials, e.g. define learning objectives (what should the student know) and work backwards from there. In particular for a MSc thesis, the educational theory background is needed for a good mark. MSc thesis would also benefit from running the developed course once with test students. BSc thesis can get away without actual experimentation due to time constraints.