ERHVERVSAKADEMI AARHUS

Data Processing.

Mandatory Assignment. Autumn 2020.

Complete as many of the following exercises as possible.
Credit will be given for efficient, effective and well functioning solutions, as well as for clear written explanations of the solutions and (important) your own thinking in making the solutions, including the decisions, limitations and interpretations you have made in order to come up with your solution. Illustrations, screendumps in the report are much appreciated, and usually makes the report much easier to read.
It is mandatory to include screen dumps that demonstrates your use of the database, as well as (Flask) data visualizations – So, that it is possible to understand the functionality, even if it is (for some reason) difficult to run the handed in scripts.
Your report must be handed in on WiseFlow on the 27th of November 12.00 at the latest. You can handin the report either on your own, or as a part of a group. Max groupsize is 2. Each group hand in their own reflections and argumentation. Remember to back argumentation with references to relevant sources. If the report consists of multiple files (report, scripts) then it must be zipped into one file. The max size of the report is 8 pages (Not including zipped Sql scripts and Flask solution).

The Case.

The theme in this semesters (Itek, Data Processing) course is self-driving cars.

As you are well aware of by now self-driving cars runs by data, just as much as they run on fuel. And, you will in this exercise deal with a number of issues concerning this data.

E/R Diagrams, Tables & Visualizations.

Autonomous vehicles do, of course, store, and work with, an awful lot of data. Both structured and unstructured data.

We do not have time in this assignment to address all the AV data processing that is going in the real world. So, we will only look at a limited part of the data processing – which nevertheless should give us a gist of the full story.

So, in order to narrow our scope, we will in this, first exercise, only look at structured data fitting in a relational database. And, we won't focus much on all of the sensor data that the car actually needs to drive itself.

Monitoring driver data.

It is expected that coming autonomous vehicles will have various equipment to monitor the drivers performance and alertness (capacity to take control of the vehicle if necessary). Again, here we will probably be dealing with both structured and unstructured data.

For the purposes of this exercise we will assume that the cars programs (that you find interesting) can transform observations into structured data that we can save into a relational database.

E.g. eye tracking could demonstrate both how focused the driver is on the driving task, and the level of general alertness (and you are free to assume that output from such a simple program could be given in a simple format). Geolocation data (and timestamps) along with data about speed would give us one window into potential traffic violations for a given driven.

In your report, you should focus on just one or two kinds of driver monitoring. You decide which, and why you find this kind of information relevant (You describe your solution in more detail in exercise 3). Then, describe in a headline form what kind of sensor output that you would like to see as (output) structured data that we can insert into a relation database.

Create an SQL database to support your solution. Minimum requirements for the database are:

- Relevant and updated E/R diagrams giving an overview of the database, including the relations and cardinalities
- Description of the tables and the database design, including contraints and level of normalization
- The database SQL scripts must be written using PostGreSQL/pgAdmin, and can't be auto-generated by any tools
- Create relevant data, insert and test - in order to verify constraints and database integrity
- Create relevant database queries in order to visualize the data (You can use Python Flask for visualizations. This will also allow you to show route maps with e.g. colors for indications of areas with driver problems. E.g. use Openlayers, introduced in Itek 2 semester).

Big Data & Autonomous Vehicles.

(After the exercise above, then) In this part of the report we will look at some of theoretical aspects of the data tsunami that is going to hit us in the coming years. And you are asked to reflect about what kind of information that we can potentially extract, and where you think the limits are. The purpose of going through these reflections is to prepare for the concept of working with Big Data.

Exercise 1.

Decribe what the 5 V's of Big Data are about. Use autonomous vehicles to make examples to illustrate the 5 V's in a AV context.

Exercise 2.

Part a.

In a headline form, describe in your own words how a cars entertainment center for video and music could (reasonably be assumed, in the future, to) provide content to users based on Big Data.

Part b.

Once in a while the entertain center might pause (videos and music) shortly, in order to show the cars passengers news from their Facebook newsfeed. Facebooks newsfeed algorithm is not publicly available, and deciding what a person should see next in his or her NewsFeed is obviously a hard problem (indeed, personalization is a hard problem), but we can make some educated guesses about how the Facebook newsfeed works, or perhaps rather how we think it should/could work. Describe, shortly and in a headline form, in say about 10 - 20 lines, what you think (guess) might be key ingredients in Facebooks newsfeed algorithm.

Part c.

Again, in a headline form, describe how marketing companies could use Big Data to run campaigns to sell a new autonomous vehicles to potential customers. Try to come up with some headlines and principles on what data could be relevant to target customers, and what you think could be a useful approach. Be realistic about what can be done, but still, importantly, as creative as possible when you describe your solution, without going into a lot of technical details.

Exercise 3.

Give a rough outline of how your car system would realisticly measure driver performance and alertness (I.e. how could you improve on your solution in the previous section). What will be saved as structured and unstructured data? What kinds of alarms will the system give to the driver, and what actions will the car take? Again, be realistic, but still as creative as possible.

How will the system document traffic violations? What kind of violations could realisticly be sent to the authorities for automatic fines (In some "strict" countries …. with instant redrawal on bank accounts?) and what kind of violations could be sent to the authorities for manual processing? Should the car also document violations by other cars or pedestrians?

https://www.independent.co.uk/news/world/asia/china-police-facial-recognition-technology-ai-jaywalkers-fines-text-wechat-weibo-cctv-a8279531.html

https://www.dailymail.co.uk/sciencetech/article-5554579/Facial-recognition-AI-used-China-send-FINES-jaywalkers-text-messages.html

What parts of the data should be sent to insurance companies for assessment of (performance based) insurance prizes? Etc.

Here, you will be given credit for your exploration of the area. But, it is still expected that you hold the documentation of your solution for this exercise on a headline form, and within 1 -2 pages absolute max.

Exercise 4.

Continuing with data ethics. Could the data your system is collecting be used in connection with the cars "moral machine settings" (http://moralmachine.mit.edu/) ?

Your ideas, reflections. Shortly, give your reasoning: Can data that you are collecting (about e.g. driver alertness and performance) give us insights into a drivers preferred morale machine settings? E.g. can we infer that a driver with a history of speed violations (near pedestrian areas) want his or her car to disregard the safety of other people in driving crisis situations?

Should drive-performance data be given to the authorities in case on early onset of mental or physical illnesses that could potentially endanger public safety in traffic conditions? How much data should we store, what kind, and for how long? If you think that we can spot immoral behavior should the authorities then use this red flag to investigate further into the person (What should be investigated) ? What kind of misuse of data could follow from such an approach (If you think that different countries will have different approaches here then elaborate on that)?

Again, keep your reflections on a headline form, and within 1 – 2 pages absolute max.

Exercise 5.

Do you think that the autonomous vehicle data collection system (that you have described in this assignment) is compliant with current GDPR laws? Does the system fulfill reasonable standards of data ethics?

Again, describe your reflections with some details, but still within a reasonable length, see above.