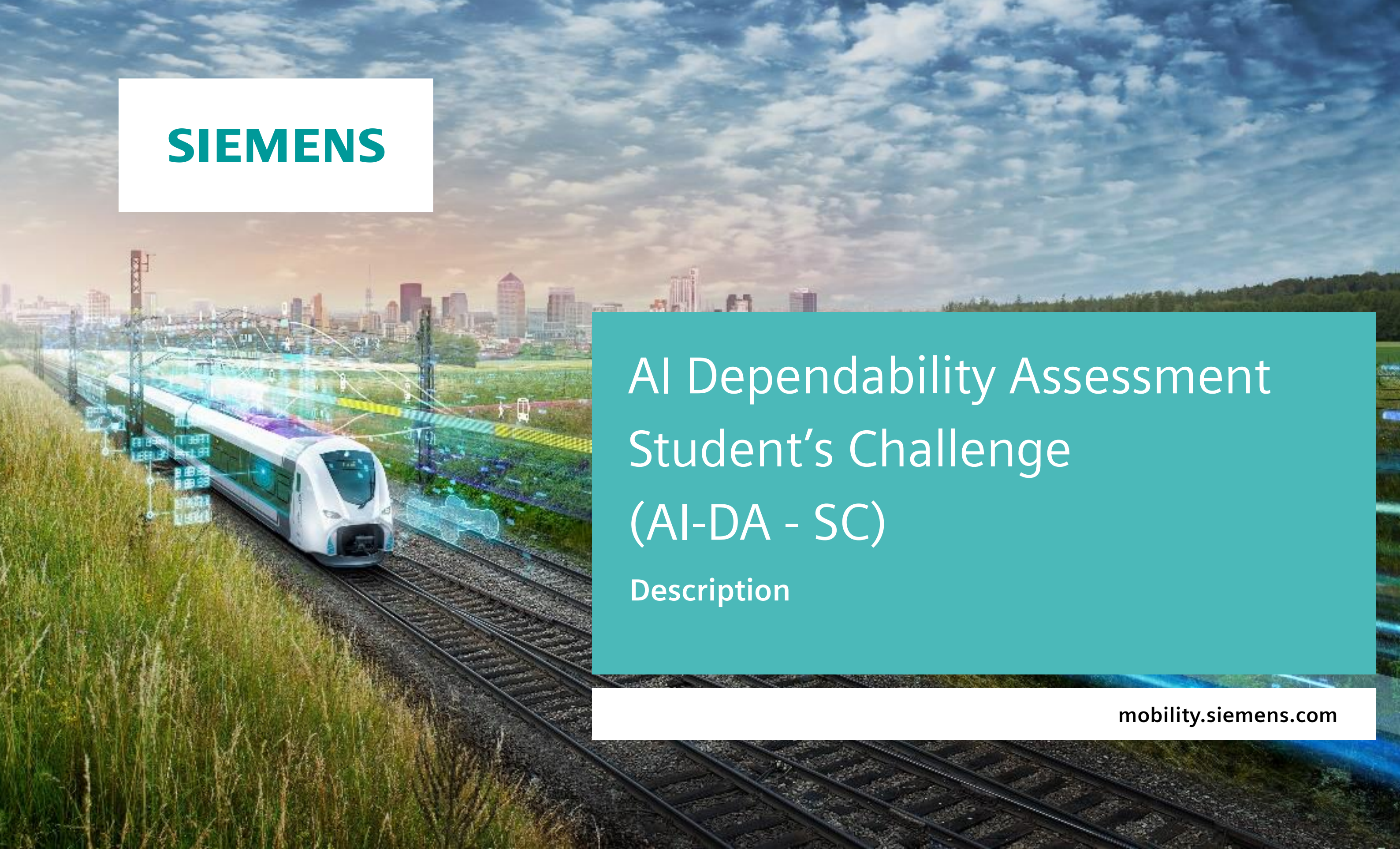


SIEMENS

AI Dependability Assessment Student’s Challenge (AI-DA - SC)

Description

mobility.siemens.com

Motivation

The application of Machine Learning (ML) techniques for safety decisions, e. g. in autonomous driving, is heavily disputed. E. g., new standards or certification rules are proposed, while others think that better explainability and transparency of the algorithms are needed. However, the problems, that are discussed, are quite complex, e. g., including multi-sensor fusion, and usually high-dimensional data. The idea pursued here is to reduce the use case complexity dramatically in order to focus on the generic challenges. This may somewhat oversimplify the problem, but the plain truth is that if we can’t solve this simpler problem, how should we solve more complex problems? Or the other way around: if we solve the simple problem, we may be able to subsequently lift it to higher dimensions and complexity.

The problem presented here is the basic introductory 2D classification problem which can be found in almost any textbook, because it is intuitive and can easily be visualized and understood. The goal is to base a safety-related binary decision only on the ML algorithm. Such algorithms may be considered to be the basic blocks in more complex systems, e. g., the decision to change lanes safety on the motorway; or consider the decision about a simple signal aspect, e. g. distinguishing a red from a green light. Assume that the sensor data have been preprocessed and compressed to only two relevant inputs x and y (in real problems the dimensions are higher) and that the ML algorithms computes a binary function f(x,y). We may assume that x and y are normalized to the unit interval and that there is a sufficient number n of labeled samples on which to train the ML algorithm.

Here it is assumed that the inputs for the algorithm, i.e., the training data, are correct and that classes do not overlap. It is obvious that the problem can probably not satisfactorily be solved without other assumptions (remember ”no free lunch” or similar theorems). Also, a solution should not be limited to 2D, but it should scale to higher dimensions.

So the task is: Under which assumptions and by which arguments can we derive safety assertions for the decision derived from the trained classifier, most prominently: can we give a meaningful (very low) upper bound for the probability, that a new sample (x,y) is misclassified? Or can we even prove that a misclassification cannot occur?

Note that the task is not to recognize the pattern in the data. This is often obvious for 2D data. The real challenge is to bound the misclassification probability in a trustworthy manner, so that even life and limb of people might depend on the result. For such safety applications the guaranteed misclassification probabilities must be very low. Depending on the applications, a target of 10-3 or lower could be sufficient for safety-related applications, while safety-critical applications must achieve much lower numbers. So, part of the competition is to achieve very small numbers with a high level of trust. You may compare this to a confidence interval in statistics: we want to derive an upper bound on the misclassification probability with some measure of trust.

Concerning the assumptions, it is obvious that they must be validated in practice. So, for any assumption there must be an argument for its validity. Such arguments may differ in rigor, e. g. they may just be based on plausibility, physical properties, or statistical tests.

The challenge

The following game is proposed:

Player A (the Assessor) chooses a subset S of the unit square $I=[0,1]^2$ and a probability distribution P on I.

Then A chooses a sample size n and generates n random variates x_i from P. Additionally labels l_i are assigned: $l_i=1$ (red), if x_i is an element of S, and 0 (green), else – thus, the Assessor uses an indicator function I_S.

For an illustrative example see Figure 1.

Player E (the Safety Expert) has now the task to provide a machine learning algorithm and safety arguments for this classification problem, including all assumptions and, possibly, safety-related application rules. In particular, she must provide a (non-trivial) upper bound for the probability, that the next random variate x_{n+1} is misclassified.

If a red data point is labeled green, then this decision is safety-critical, and an accident may happen. If a green data point is labeled red, then this may cause or some cost, but not directly a safety problem. Take traffic lights as an example...

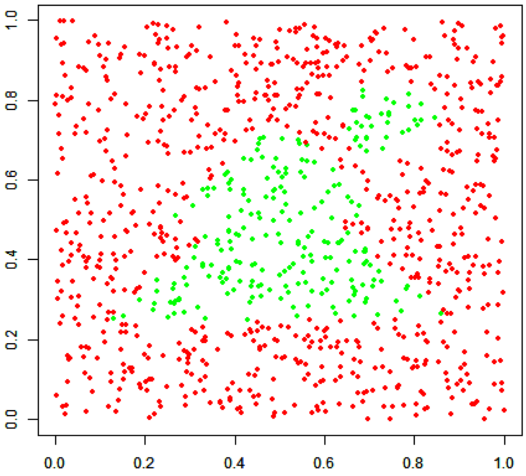


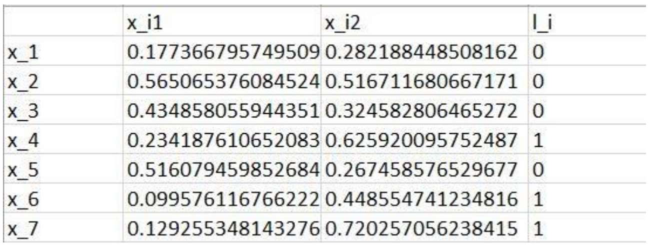
Figure 1: Example data set

The name of the game

The competition is open for student teams from universities. A team may consist of up to three members. All members must be enrolled as full-time students and they must not yet have completed a PhD.

Each team receives three data sets A, B and C as csv files with labeled data. The data sets vary in the complexity of the Set S, the distribution P and the samples size n.

An example for a data set one can see in Figure 2.



	x_i1	x_i2	l_i
x_1	0.177366795749509	0.282188448508162	0
x_2	0.565065376084524	0.516711680667171	0
x_3	0.434858055944351	0.324582806465272	0
x_4	0.234187610652083	0.625920095752487	1
x_5	0.516079459852684	0.267458576529677	0
x_6	0.099576116766222	0.448554741234816	1
x_7	0.129255348143276	0.720257056238415	1

Figure 2: Example .csv file

As an entry to the competition they should submit until 02.05.2021 an expose of their solution proposal (not more than 10 A4 pages excluding appendices, font size not less than 10 pt), including

- A description of the employed ML approach
- All assumptions made, including an exhaustive proposal/justification for their validation
- For each dataset, an upper bound for the misclassification error, including its justification
- An outline how the approach could be scaled to higher dimensions
- In an appendix, the documented code should be supplied
- References to results that are used

The proposal shall be accompanied by short CVs of the team members and a certificate of their enrollment as full-time students.

The proposals will be evaluated by an expert jury, composed of representatives from Siemens and research institutions. The evaluation will be based on the criteria outline above.

In a first phase the proposals will be checked for their correctness and comprehensibility.

Those teams qualifying for the second phase will receive from 18.05.2021 on additional unlabeled data sets A', B' and C', which they shall label. These validation results will be taken into account in the final assessment of the proposals.

Finally, the jury may decide to grant up to three awards with a joint reward of 10,000 €. It is up to the discretion of the jury to decide on the number of prices and the sharing of the reward.

Siemens Mobility GmbH

If you have any questions about the challenge, please feel free to contact us:

research-and-education.ct@siemens.com

Further information:
https://ecosystem.siemens.com/ai-da-sc