WASP Software Engineering and Cloud Computing
Assignment 1

Konstantin Malysh

konstantin.malysh@cs.lth.se

What did you learn, in your individual session, about static analysis for ML and the pynblint tool?

During the preparation for the session I encountered the Pynblint tool for the first time ever. to me Pynblint seemed like a reasonable extension to any pep8-enforcing tool (usually built into an IDE). During the session we discussed what is the purpose of the tool. We did not talk a lot about the concept of the static analysis itself or ML, just got straight to the point of looking at the results of the tool processing my notebooks and figuring out what the tool is looking for or not.

Will pynblint be useful to you in your WASP PhD project? Why or why not?

Due to my PhD research not being in the areas of ML or software development (only adjacent, but I will not need to write any Jupyter notebooks for that), I do not think I would ever use Pynblint. I could potentially use it as a teacher assistant for the Machine Learning courses I help with, but still the idea of those courses is to just provide solutions with students themselves fixing their errors, so I also doubt I could use it there.

Ideas for how the tool could be improved?

Some of the warnings seemed too empirical: the amount of markdown cells or their lack in the certain areas of a notebook, lengths of cells (those can be long for a reason), presence of non-executed cells (I could see those to sometimes be meaningful). However I have to admit that usually a Jupyter notebook plays a role of a sandbox for me, so, for example, some non-linear execution is usually fine for me as soon as there is numeration of cells (Jupyter enumerates the cells as soon as they are executed), but I can see that being a problem for a serious presentation. Also, as of now, the tool itself seemed too basic, as in being in the early stage of development due to amount of warnings, or maybe my notebooks were tidy enough to not trigger many warnings, so more rules could be carefully introduced in the future.

What do you see as the limits for static analysis tools in ML? For code, models, and for data?

The fact that the analysis is static is in my opinion the main limitation, because ML is all about the data and environment, and static analysis does not take that into account.