



Review report of a final thesis

Student: Ard Kelmendi
Reviewer: Ing. Karel Klouda, Ph.D.
Thesis title: Feature Importance for Black-box Models
Branch of the study: Computer Science

Date: 25. 1. 2021

<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 4.</i>
1. Fulfilment of the assignment	<i>1 = assignment fulfilled, 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled</i>
<i>Criteria description:</i> Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.	
<i>Comments:</i> The author fulfilled all the assignment points, but some of them in a very sloppy manner. For instance, the testing is indeed insufficient, and comparison with other implementations even more.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
2. Main written part	<i>55 (E)</i>
<i>Criteria description:</i> Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies? Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 26/2017, Art. 3. Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.	
<i>Comments:</i> The text is full of various typing errors, strange sentences, and other things that make reading and understanding uneasy. In particular, whenever a formula appears, it is not easy to decode what every letter stands for. See, e. g., Section 1.4.1, where the linear model is described: What is the subscript l in the text, what is n , the subscripts are written inconsistently, what is θ squared, and so on. I also do not understand the content layout and order. What is the purpose of the first chapter? For instance, why the author decided to describe the general linear model and AUC on page 13? The description is so brief that it is useless for anyone who has not been familiar with the matter before. Moreover, AUC does not appear in the text ever again. Some parts are mentioned and explained repetitively (e. g., the interpretability). Some important parts are hidden in sections where the reader would not expect them. Comparison of the existing implementations of PFI is (very briefly) mentioned at the end of Chapter 4, called Implementation. The computational complexity of the implemented algorithm is given in the conclusion. The conclusion also contains some essential implementation details regarding the H2O-3 framework.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
3. Non-written part, attachments	<i>70 (C)</i>
<i>Criteria description:</i> Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.	
<i>Comments:</i> The selected method for measuring the feature importance, namely PFI, is not very complicated and hard to understand. However, its incorporation into the H2O framework is not a trivial task. The resulting code is now in the pull request in the H2O repository. There is also a Jupyter notebook with a showcase usage. What I miss in the text is some structured description of how the code is organized and so.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>

4. Evaluation of results, publication outputs and awards

70 (C)

Criteria description:

Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.

Comments:

Once the pull request is merged, the users of the H2O framework will have new useful functionality. It would be better if the merge has already happened.

Evaluation criterion:

No evaluation scale.

5. Questions for the defence

Criteria description:

Formulate questions that the student should answer during the Presentation and defence of the FT in front of the SFE Committee (use a bullet list).

Questions:

- 1) Why has the pull request not been merged yet? What is missing?
- 2) What are the weaknesses of the PFI method? Does it always identify the most important features?

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

6. The overall evaluation

60 (D)

Criteria description:

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.

Comments:

With all that has been written above, I recommend the thesis for defense with grade D.

Signature of the reviewer: