FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

# Supervisor's statement of a final thesis

| | |
|---|---|
| **Student:** | Ard Kelmendi |
| **Supervisor:** | Ing. Veronika Maurerová |
| **Thesis title:** | Feature Importance for Black-box Models |
| **Branch of the study:** | Computer Science |

**Date:** 25. 1. 2021

| *Evaluation criterion:* | *The evaluation scale: 1 to 4.* |
|---|---|
| **1. Fulfilment of the assignment** | *1 = assignment fulfilled,*<br>***2 = assignment fulfilled with minor objections,***<br>*3 = assignment fulfilled with major objections,*<br>*4 = assignment not fulfilled* |

*Criteria description:*
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.

*Comments:*

The student should fulfill four main goals in the thesis. He described the Feature importance calculation for Black-Box models very well, however with a few inaccuracies (listed in the next part). He also implemented the algorithm to the H2O-3 library; he had to orient in the huge open-source project and use three programming languages. The implementation looks functional as he evaluated it in the tests and compared it with a different library. However, it is still not production-ready and needs some work to finish, so the implementation goal is not fully achieved.

| *Evaluation criterion:* | *The evaluation scale: 0 to 100 points (grade A to F).* |
|---|---|
| **2. Main written part** | *65 (D)* |

*Criteria description:*
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.

*Comments:*
In the main written part, the student used the correct structure. What is wrong or should be better:

Sometimes the sentences are too rich and less understandable (for example this sentence in the "Train and Learn data" section: "After the model outputs a prediction a second dataset is needed to verify the performance of the model, the testing dataset." Also the phrase "Learn data" is not understandable.).

In the Introduction chapter goals of the thesis are very poorly described. In the Machine Learning chapter, the students described basic algorithms very well, except for the Generalised Linear model section where he probably Generalised Linear Models definition confused with the term General Linear Model and logistic regression (page 14).

The Implementation chapter could be more comprehensible to the reader. The student started with a list of parts instead of introducing what the chapter is about. Then the essential part of the Map-Reduce principle is 'hidden' after algorithm description. Also, data description could be in its subsection and the student could describe more in detail what the features are about. There are also missing names of axes in the graphs, the figures with the graphs are in really bad resolution to see names of the features and the numbers of their importance. The student compared two feature importances (figure 4.2 and 4.3), however, there is just a bar graph to compare, where it is very hard to understand if the numbers are similar or not. There is no comment about similarity and no conclusion if the results make sense. Also why the GBM algorithm is selected is not specified.

The implementation of the OAT technique is not compared with any other implementation and there is no information about if some implementation exists or not. The student did not cite figure 4.4.

The subsection about big data is very poor. The student did not describe why it is important to test his implementation also with big data. There are no results on how fast/slow his implementation is and what is expected complexity.

Section 4.3 Advantages and disadvantages in the Implementation chapter should be in the theoretical part of the thesis in my opinion.

Section 4.4 could be better written out. Figure 4.11 has the label "First rows of cars.csv dataset", however, the image shows rows 0-4 and 401-405. The student compared his implementation with Sklearn implementation, but the results could be displayed more readable (for example in the table) instead of description in two sentences. More datasets to compare results could be used. The comment about results is missing. There are also missing model parameter settings to compare results. Also, the speed of these two algorithms could be nice to compare. Figure 4.12 does not make sense to me.

The conclusion chapter could be better structured. There is information about the complexity of the algorithm which should be also mentioned in the previous chapter. The students did not mention if he achieved the goals of the thesis or not. He did not mention in which state his implementation currently is and what are his plans for the future.

| *Evaluation criterion:* | *The evaluation scale: 0 to 100 points (grade A to F).* |
|---|---|
| **3. Non-written part, attachments** | *50 (E)* |

*Criteria description:*
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.

*Comments:*
The student uploaded the implementation via GIT to the H2O-3 public repository at Github. It was a new technology for him. He has to understand the huge project structure of H2O-3, which is not easy.

The implemented code is working in Java and Python, however not fully for R (incorrect calling OAT method via API). The quality of code is ok, just a few namings could be better. The student implemented tests to verify his implementation is working. However, a code of test for comparing his implementation with Sklearn is missing, as well as the test complexity of the algorithm.

| *Evaluation criterion:* | *The evaluation scale: 0 to 100 points (grade A to F).* |
|---|---|
| **4. Evaluation of results, publication outputs and awards** | *60 (D)* |

*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:*
The student wrote his thesis in cooperation with the H2O.AI company. The primary goal for the company was the implementation of the algorithm. The goal was almost achieved, but the implementation still needs some work to be production-ready.

| *Evaluation criterion:* | *The evaluation scale: 1 to 5.* |
|---|---|

| | |
|---|---|
| **5. Activity and self-reliance of the student** | *5a:*<br>*1 = excellent activity,*<br>***2 = very good activity,***<br>*3 = average activity,*<br>*4 = weaker, but still sufficient activity,*<br>*5 = insufficient activity*<br>*5b:*<br>*1 = excellent self-reliance,*<br>*2 = very good self-reliance,*<br>*3 = average self-reliance,*<br>***4 = weaker, but still sufficient self-reliance,***<br>*5 = insufficient self-reliance.* |

*Criteria description:*
From your experience with the course of the work on the thesis and its outcome, review the student's activity while working on the thesis, his/her punctuality when meeting the deadlines and whether he/she consulted you as he/she went along and also, whether he/she was well prepared for these consultations (5a). Assess the student's ability to develop independent creative work (5b).

*Comments:*

The student was active and well prepared for meetings. The situation was hard for consulting results due to COVID restrictions. The student organized several calls to discuss the result of his thesis continuously. However, he could be more independent.

| *Evaluation criterion:* | *The evaluation scale: 0 to 100 points (grade A to F).* |
|---|---|
| **6. The overall evaluation** | *65 (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:*

I suggested grade D. In my opinion, the theoretical part is well written up to a few details. What affected my grade the most is the implementation part of the thesis. The code of the algorithm is ok. However, the results of testing and comparing with other implementations were very poorly described and not fully implemented.

I would appreciate it that the student understood a huge amount of code in the H2O-3 library and implemented an almost production-ready algorithm using three languages. He also had to understand the not trivial part of the Machine Learning field to describe and implement the algorithm.

Signature of the supervisor: