CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Porovnání reprezentací souborů logů pro metody detekce anomálií**
Author's name:	**Martin Hubal**
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	CS
Thesis reviewer:	Ing. Matej Uhrín
Reviewer's department:	CS

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment — **ordinarily challenging**

How demanding was the assigned project?

The thesis aims to compare the robustness of two approaches: Vanilla autoencoder, best performing BoW model and AECNN1D, best performing FastText model taken from a previous thesis [1]. The hypothesis was that the AECNN1D will be better able to deal with unseen log types thanks to the use of FastText. I find the assignment ordinarily challenging as a significant amount of work had already been done by the author of the previous thesis.

[1] Anomaly Detection Methods for Log Files, https://dspace.cvut.cz/handle/10467/95290

Fulfilment of assignment — **fulfilled**

How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.

The author fulfilled all of the requirements.

Methodology — **correct**

Comment on the correctness of the approach and/or the solution methods.

The experiment part of the thesis is solid. The author uses correct methodology for evaluating and comparing proposed approaches under various scenarios.

Technical level — **A - excellent**

Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?

The thesis is technically sound. The student employed expertise beyond his field of study.

Formal and language level, scope of thesis — **A - excellent**

Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?

The language of the thesis is understandable and the thesis is shorter but sufficiently extensive. I only have few minor objections:
- Sec. 2.3.4 is not very well placed and perhaps would be better omitted or included in the previous sections: 2.2.1, 2.2.2
- Sec. 4.2.2 would be easier to read if a diagram was included.

Selection of sources, citation correctness **D - satisfactory**

Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?

Bibliographic citations meet the standards and the selection of sources is adequate. However,

Chapter 2 reads almost as 1:1 translation of Chapter 3 from the previous thesis with the same exact figures, (or very similar). The structure of the text and naming of subsections is also 1:1, but translated.
- Figure 2.1 is almost the same as 3.1. in [1]
- Figure 2.5 is exactly the same figure as 3.4 in [1] - no citation to the original author.
…
- even when examples are listed in the text, they are exactly the same e.g.:
Sec. 2.3.4:
Supervizovanými modely jsou pak například logistická regrese, rozhodovací strom a metoda podpůrných vektorů, které byly porovnány v[9] s dalšími supervizovanými modely.

Sec. 3.1.2 in [1]
Support vector machines (SVM) [31], a logistic regression [32], and decision trees [32, 33] are representative examples of supervised anomaly detection methods.

I understand that this is an introductory chapter and perhaps different explanations, examples, figures are difficult to come up with, but in my opinion the similarity is so strong that it requires a reference of the previous thesis in almost every single subsection. Further, if this is an extension of the previous thesis, I would expect the introductory chapters to be somewhat "better", to extend the previous thesis with better explanations and new ways of understanding. This is not the case here, the introductory chapters in this text seem to be shortened copied sections of the previous thesis.

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.

The thesis is solid and there is a good amount of work behind the experiments. However, I can't help but compare it to the previous thesis that flows better and honestly provides the reader with a much better understanding of the overall problem. I also believe that quite a few sections in this work should be treated as citations of the previous thesis. Given that this is a bachelor's thesis and the student applied expertise far beyond his field of study,

the grade that I award for the thesis is **B - very good.**

Date: **2.6. 2022** Signature: