**FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE**

Supervisor's statement of a final thesis

Supervisor: Mgr. Martin Jureček, Ph.D.
Student: Mehmet Efe Zorlutuna
Thesis title: Unsupersived Instance Selection for Malware Detection
Branch / specialization: Computer Security 2021
Created on: 5 June 2023

Evaluation criteria

1. Fulfillment of the assignment

▸ [1] **assignment fulfilled**
   [2] assignment fulfilled with minor objections
   [3] assignment fulfilled with major objections
   [4] assignment not fulfilled

I consider the assignment fulfilled.

2. Main written part   50 /100 (E)

The structure of the thesis is mostly correct – some chapters can be merged. The bibliography consists of relevant references; however, it contains several mistakes. Justification for using DBSCAN is insufficient since there are many more options for dealing with unknown clusters. Experimental results were evaluated using three measures, but two required labels that may not be provided. The first elimination technique, NCE1, is incorrect since, as shown in Figure 2.4, the purple cluster was eliminated. Regarding NCE2, I believe the filtering rule should not depend on the number of instances in the cluster. In general, clusters may have various structures, and while some clusters may contain only small representative instances, others may have more complicated structures, so more representative instances are required.

3. Non-written part, attachments   75 /100 (C)

The source code has been written in Python in Jupyter Notebooks, a standard in data science. The source code could be better commented. The dataset and the source code are public so that the experimental result can be verified.

4. Evaluation of results, publication outputs and awards   55 /100 (E)

Unsupervised instance selection is a relatively new research area with few proposed algorithms. If the student improved his proposed algorithm, it could have the potential to be published in a journal and also used in the industry where unlabeled data are used for the decisions.

5. Activity of the student

    [1] excellent activity
    [2] very good activity
▸ **[3] average activity**
    [4] weaker, but still sufficient activity
    [5] insufficient activity

The activity of the student was average. The student was responsive, but several relatively long periods occurred in which the student was not focused on the thesis.

6. Self-reliance of the student

    [1] excellent self-reliance
    [2] very good self-reliance
▸ **[3] average self-reliance**
    [4] weaker, but still sufficient self-reliance
    [5] insufficient self-reliance

I rate the student's self-reliance as average. He solved some technical problems by himself and needed help with other theoretical issues.

The overall evaluation   50 /100 (E)

I did not advise the student to submit the thesis because I believe that the work is not finished and has the potential to be of much higher quality. Even though some parts of the thesis, for example, the description of the NCE 2 technique, can be significantly improved, the student has shown that he can study, implement and even design a new algorithm from the domain of unsupervised instance selection. Most of the student's work was done in the last months before submission, which affected the quality of the work. Nevertheless, I believe that the thesis can be defended. I recommend grade E.

Instructions

Fulfillment of the assignment

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.

Main written part

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. 52/2021, 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.

Non-written part, attachments

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.

Evaluation of results, publication outputs and awards

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.

Activity of the student

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.

Self-reliance of the student

From your experience with the course of the work on the thesis and its outcome, assess the student's ability to develop independent creative work.

The overall evaluation

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.