SUPERVISOR'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	**Stroke Mortality Prediction**
Author's name:	**Regina Mavrina**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Computer Science
Thesis supervisor:	Ing. Matěj Klíma
Supervisor's department:	Department of Computer Science

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment   **ordinarily challenging**

Evaluation of thesis difficulty of assignment.

Compared to other assignments of master thesis in the software engineering field, I find this one ordinarily challenging. The development part of the work is easy and doesn't require advanced programming skills. On the other hand, the analytical part of the assignment I find challenging, and the student had to prove her knowledge of statistics and manipulation with data.

Satisfaction of assignment   **fulfilled with minor objections**

Assess that handed thesis meets assignment. Present points of assignment that fell short or were extended. Try to assess importance, impact or cause of each shortcoming.

The purpose of the assignment was to implement a stroke mortality forecast among the population, using various statistical methods. Also, a comparison of results obtained by those methods should be discussed in the thesis.
The assignment was fulfilled. A reader can find an introduction to the problem, a description of the data set, and a description of statistical methods that can be used for a stroke mortality prediction. The final comparison of the results can be found in chapter 4.6. Moreover, the author showed a quick reaction to the actual situation when she added a part dedicated to the actual COVID-19 pandemic and its relationship to stroke, although it wasn't part of the assignment. However, the part of the thesis dedicated to the description of the software project development doesn't contain some important parts, such as requirements on the software and the description of design and final implementation.

Technical level   **C - good.**

Assess level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.

A combination of knowledge from statistics and programming was necessary to solve the assignment of this thesis. The author proved that she has both of them. She comprehensively described the discussed statistical methods and illustrated her statements and data with many charts. Most of the expert terms and symbols are introduced to the reader and backed by a reference.
However, sometimes it is hard to follow the text, mainly because the author dives into a big detail very quickly and sometimes forgets to give a proper introduction to the discussed problem. Also, a proper analysis of results calculated by some statistical methods is missing (e.g., for the Bootstrap method or for the Jackknife method).

Formal and language level, scope of thesis   **C - good.**

Assess correctness of usage of formal notation. Assess typographical and language arrangement of thesis.

The thesis contains some formal errors. It would be better if the results gathered from the execution of the R code were transformed into some tables. The presentation the author has chosen (screenshots of the result) seems confusing. The Contents look differently from the rest of the text as if it was a screenshot or a picture. The spaces between numbers and the text are heterogeneous. The code segments in some places, e. g., on page 32, have a weird grey background and are shown with inconsistent font sizes (e.g., on page 31 the Figure 2.4 versus Figure 2.5).
The figures 3.9, 3.10, and 3.11 contain information connected too much to the internal implementation in R (name of functions or variables). Some of the figures (e.g., Figure 1.5 and Figure 1.7) are too small to be readable.
On the other hand, the text contains only very few typing and grammar errors.

Selection of sources, citation correctness	**B - very good.**

Present your opinion to student's activity when obtaining and using study materials for thesis creation. Characterize selection of sources. Assess that student used all relevant sources. Verify that all used elements are correctly distinguished from own results and thoughts. Assess that citation ethics has not been breached and that all bibliographic citations are complete and in accordance with citation convention and standards.

The student used in her work 38 different sources, which I find satisfactory. Statements that are not her own she marked with a reference, although sometimes the citation style varies (sometimes the number of the reference is before the dot, sometimes it is after it).
The student correctly referred to the articles, from which she used some data or statements. Most of the charts and figures she made on her own, the others are correctly marked with the reference.

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Summarize thesis aspects that swayed your final evaluation.

The thesis describes an interesting research topic of stroke prediction using various statistical methods, their description, and evaluation. The author has found a data set for the research, described it, and implemented a project in R language that calculates the probability of stroke based on the selected data using the selected statistical methods. All of the methods are described in the thesis, although sometimes a little incomprehensible.

The scope of the thesis is alright. The author uses the sources thoroughly and shows in the thesis an understanding of various statistical methods.

However, some formal errors can be found in the thesis, and the thoughts and statements of the author could be described in the thesis more coherently. Moreover, the description of the development of the software project is incomplete.

Questions for defense
1) Would you like to continue in the work you wrote about in this thesis?
2) What is an outlier, and did you remove them in the data sample?

I evaluate handed thesis with classification grade **C - good.**

Date: **26.8.2020**   Signature: