Tomas Bata University in Zlín
Faculty of Applied Informatics

OPPONENT´S EVALUATION OF THE DIPLOMA THESIS

Student: **Denis Varaksin**   **Opponent:** **RNDr. Pavel Popela, Ph.D.**

Study program: **Engineering Informatics**
Study discipline: **Information Technologies**
Academic year: **2018/2019**

Thesis topic: **Traffic Data Prediction**

Evaluation of the thesis - outline:

Please enter your evaluation of the submitted work here. The report will focus on:

- The thoroughness of the elaboration, its validity and the difficulties resolved in the thesis
- The method and level of the concepts resolved in the thesis
- The level of the student´s thesis topic and elaboration and its contribution(s)
- The formal aspects of the thesis, errors and mistakes in the technical aspects
- Questions relating to the defence of the thesis
- In conclusion, assess the overall qualities of the submitted thesis and classify it according to the rating scale shown at the end of this form

Evaluation of the thesis:

The proposed theme of the master thesis is actual and interesting and it is based on the real-world data from Zlín company. The thesis goals are clearly specified by supervisor in thesis guidelines. However, the comparison of thesis guidelines and content of thesis shows a lack of detailed analysis of results. For example, the abstract mostly focuses on motivation than on thesis results. In addition, a preface describing the overview of thesis content by chapters is missing. Some ideas are given in short chapter Introduction, however, the unclear use of passive voice decreases a readability of this chapter.

Theoretical chapter 1 on forecasting is quite general without any references to traditional literature and to recent resources about machine learning. Another short chapter 2 on data analysis is based only on one resource reference and again presents a general framework for data treatment. Forward references to the particular use of proposed general ideas is not presented. Chapter 3 deals with classical time series concepts but literature references are again missing and notation is sometimes unexplained (see indices on page 12) or rather intuitive and not unified (cf. x_t and $x(t)$). Therefore, the use of mathematical formulas can significantly mislead the reader who does not know time series literature in detail. Sources of Figures 5-8 are not mentioned, and unfortunately, there are no detail explanations related to these figures and their role in the text. Theory in Chapters 3, 4, and 5 covers many topics typically covered by whole textbooks on time series. Therefore, the presentation of so many concepts is quite short and material is not organized well. Unfortunately, the author also does not show (e.g., by forward references), which concepts and how are further used for real-world data.

Part II Analysis begins with presentation of considered crossroad situation by annotated pictures in Chapter 6. Chapter 7 promises application of Part I concepts. Section 7.1 visualizes the processed data and then Section 7.2 gives the step-by-step description of software implementation by using programming language R tools. RStudio related codes are commented in short within the text and complete codes can by found in enclosed archive file. Section 7.2 is the original and most

interesting part of thesis for the reader. It confirms the understanding of the author to the thesis theme and selected computational tools. However, it is a pity that this section mostly answers how to use the RStudio for real-world data without detail interpretation of results for the analysed real-world situation and without explanatory and motivating details related to the realized computational steps. Even the chapter with conclusions is quite general like in the textbook on time series as the author has not used the opportunity to give the insightful comments although his realized computations and knowledge allowed it. Hence, the thesis text looks as a bit unfinished and it is a pity that the author has not utilized his knowledge and the opportunity given by the challenging theme at a higher level.

At the end, therefore, the thoroughness of the elaboration of thesis does not reach very good level. Some conclusions are not supported either by literature resources or real-world data processing, see e.g. sentence, "The time series of traffic data on the crossroad have strong seasonal factor." in 3.3. without any proof or reference (e.g. to 7.1), or conclusion "By analysing residuals and comparing both: real and predicted data graphs we can say that, beside sharp peak values in real data graph, forecast is accurate" without reference to the method used (page 43), or conclusions about normality of residuals by just graph evaluation without any statistical tests presented. Many symbols used in the formulas are not described, see e.g., page 16. Because of missing details it is hard to evaluate thesis validity and the difficulties resolved in the thesis are not well documented. The method used in thesis is classical and level of the concepts resolved in the thesis is negatively touched by lot of missing details. The level of the student´s thesis topic is high and the topic is challenging, however, the elaboration and its contribution seems to me a bit unconvincing. The formal aspects of the thesis, errors and mistakes in the technical aspects have already been criticized above.

I am suggesting that the author will answer the following questions during the defence:
1. What are the original results of the presented thesis?
2. Can he give the examples of interpretations of computational results for analysed real-world situation?
3. Does he utilized (in Part II) all concepts introduced in Part I?

After all my mostly critical comments, I can still conclude that the author has satisfied most of goals of his thesis at satisfactory level, and so, I recommend his thesis for the thesis defence and classify it by D mark.

Overall evaluation of the thesis:
The Opponent shall grant a mark according to the ECTS classification scale:
A – Excellent, B – Very Good, C – Good, D – Satisfactory, E – Sufficient , F – Insufficient
An "F" grade also means "I do not recommend the thesis for defence."

I recommend this thesis to be defended and suggest the following evaluation:
D - Satisfactory
In the case of an evaluation grade of "F – Insufficient", please supply the main shortages and reasons for this assessment.

Date: 24. 5. 2019   Thesis Opponent's Signature: