THESIS SUPERVISOR'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Pricing and data: long-distance bus routes**
Author's name:	**Mohammad Asad Ali**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Computer Science
Thesis reviewer:	Ing. Ivan Nikolaev
Reviewer's department:	External - Barclays Investment Bank

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment	**challenging**
How demanding was the assigned project?	

The assignment was of ordinary difficulty among the data analysis assignments, with data pre-processing having above average difficulty due to complex nature of the unprocessed datasets.

Fulfilment of assignment	**fulfilled with minor objections**
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 goals of the thesis seem to be achieved. My main objections are to the presentation of the results and to the limited depth of the analysis of the results.

Activity and independence when creating final thesis	**B - very good.**
Assess whether the student had a positive approach, whether the time limits were met, whether the conception was regularly consulted and whether the student was well prepared for the consultations. Assess the student's ability to work independently.	

The student consulted his work regularly, but unfortunately the progress between the consultations has been slow and student created most of the thesis content towards the end of the final semester, leaving little time for incorporating meaningful feedback to the thesis results.

Technical level	**C - good.**
Is the thesis technically sound? How well did the student employ expertise in his/her field of study? Does the student explain clearly what he/she has done?	

The thesis does not clearly explain many important details of the work. For example, it is stated on page 31 that the split between the testing and training data is 25:75, but the split into a validation dataset used in figures on page 35 onward is nowhere to be found. Another example is that the thesis doesn't mention the scraping intervals of the datasets or other details of the scraping process.

Another issue is with the analysis of the results, which doesn't go much further beyond feature importance. While the distinction between the sample, connection and route is described in the thesis, the results are not analyzed in these terms at all.

Formal level and language level, scope of thesis	**C - good.**
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 thesis uses only limited amount of formalism, however, some symbols are confusing and inconsistent, such as the term "adjustedR-squared" on page 13.

The structuring of the thesis into chapters is good. However, overall, the thesis seems incoherent. For example, the choice and number of techniques in section 8 appear to be arbitrary and there is no discussion of the results in this section in aggregate.

The extent of the thesis is sufficient, but some sections are too brief and focus on simple details while omitting the difficult ones. For example, the section on metrics (Section 3.3) describes simple residuals in detail and a figure while the

description of adjusted R^2 is rather brief. Additionally, sections on ensembles and boosting are difficult to follow, with many concepts left unexplained and without reference.

The thesis contains some typos and grammar errors, for example:

p11 – "in" capitalized, p12 – "for therefore", fig. 6.7 and 6.8 – same captions, p40 – missing word "features" in paragraph on feature importance, p46 – y-axis labels missing in figure.

Despite these errors, the text is mostly written in satisfactory and understandable English.

Selection of sources, citation correctness	B - very good.

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?

The number of references on the topic is small but their choice is mostly appropriate. Some unexplained concepts are not referenced in the text.

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

Overall, this thesis achieves its stated goals. It is readable and well structured, but it suffers from the lack of important details on the methods used and shallow analysis of the final results.

The grade that I award for the thesis is **C - good.**

Date: **12.8.2020**   Signature: