

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Andrej Zidek
Advisor:	Karel Janda
Title of the thesis:	Application of the logit leaf algorithm for customer churn prediction in the energy distribution industry in the Czech Republic

OVERALL ASSESSMENT (provided in English, Czech, or Slovak):

Please provide a short summary of the thesis, your assessment of each of the four key categories, and an overall evaluation and suggested questions for the discussion. The minimum length of the report is 300 words.

Short summary This thesis takes an interesting and original data set of 9055 observations of Czech gas consumers whose contract ended in 2019 and estimates the probability of consumer leaving the particular gas distribution company without signing the continuation contract. Since the beginning of this project I was very excited by these data and the serious approach of the student and I really liked this project.

This is a good thesis. Andrej Zidek was working on this thesis diligently, over a long period of time. Visiting me in my office, providing reports on the progression with getting data, since this was very much data driven project. The work with the student was a good one, no significant complaints on my side.

Contribution

Big positive added value of this thesis is that student is not using some general easily publicly accessible data already used in many papers. Student is using the data of a particular Czech energy company (further denoted as Company) and he had to exercise effort to establish relation with this company (he is not working for this Company), to obtain the data and to get some consulting support from the Company. Optimally, the results of this bachelor thesis focused on predicting probability of a customer leaving the Company should be of interest to the Company. It looks like that the student may provide real practical benefit for the Company and society. The results of this project are of interest to the Company, which is still interested in further working with this student on this project.

This thesis project was a very interesting one, rather unusual for a Czech University. Two IES bachelor students, Andrej Zidek and Cyril Skorvaga jointly approached the Company, jointly discussed data access and other issues with Company. I consider this an excellent approach. To be clear: the theses of both students are using different data set, different methodology (with some overlap in the use of decision trees), different research questions. So these two theses have a feel of two different chapters in a one topically consistent PhD thesis.

Methods The modelling approach used in this thesis is relevant to the research question being investigated, and adequate to the author's level of studies. It follows the methodology introduced in a recent paper „A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees“ by A. De Caigny (2018). The thesis topic is comprehensively analyzed. The author takes care to use data properly, to obtain additional data in addition to the main data set on gas consumption obtained from the Company.

Literature The author focuses on the most relevant literature. The thesis demonstrates author's understanding and command of literature. The author quotes relevant literature in a proper way in an overwhelming majority of cases. I was missing enough literature dealing with randomized controlled trial methodology (this may be explained by the different approach being taken by the student).

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Andrej Zidek
Advisor:	Karel Janda
Title of the thesis:	Application of the logit leaf algorithm for customer churn prediction in the energy distribution industry in the Czech Republic

Manuscript form The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The English of the thesis is good. The number of typos is low, significantly below the average IES thesis.

The text effectively refers to graphs and tables and disposes with a complete bibliography.

Overall evaluation and suggested questions for the discussion during the defense

While I consider this bachelor thesis to be good, above the IES average bachelor thesis quality, I have a disagreement with the author with respect to the main methodological focus of the thesis. I consistently and strongly believe that the major potential strong feature of this thesis is a possibility to focus the thesis along the randomized controlled trial (RCT) methodological line. Which was not done by author under (according to my opinion incorrect) opinion that the sample size is too small for such an approach (about 9 thousand data points, 2/3 of them treatment group, 1/3 control group).

While I could ask a number of interesting questions, in this case I focus just on one issue, which I would like Andrej Zidek to prepare for his defense:

How could be the paper extended/improved by using RCT approach. In particular, I would like student to show that he understands RCT methodology and that he is able to conduct such study with the data used in this thesis. I.e. I would like to confirm that a wasted opportunity to focus this thesis on RCT approach was not because of student not knowing appropriate RCT methodology.

In my view, the thesis fulfills the requirements for a bachelor thesis at IES, Faculty of Social Sciences, Charles University, I recommend it for the defense and suggest a grade B. The results of the Turnitin analysis do not indicate significant text similarity with other available sources.

SUMMARY OF POINTS AWARDED (for details, see below):

CATEGORY	POINTS
<i>Contribution (max. 30 points)</i>	26
<i>Methods (max. 30 points)</i>	24
<i>Literature (max. 20 points)</i>	16
<i>Manuscript Form (max. 20 points)</i>	19
TOTAL POINTS (max. 100 points)	85
GRADE (A – B – C – D – E – F)	B

NAME OF THE REFEREE: Karel Janda

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Andrej Zidek
Advisor:	Karel Janda
Title of the thesis:	Application of the logit leaf algorithm for customer churn prediction in the energy distribution industry in the Czech Republic

DATE OF EVALUATION: *August 1, 2023*

Digitally signed August 1, 2023, Karel Janda

Referee Signature

EXPLANATION OF CATEGORIES AND SCALE:

CONTRIBUTION: *The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.*

METHODS: *The tools used are relevant to the research question being investigated, and adequate to the author's level of studies. The thesis topic is comprehensively analyzed.*

LITERATURE REVIEW: *The thesis demonstrates author's full understanding and command of recent literature. The author quotes relevant literature in a proper way.*

MANUSCRIPT FORM: *The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.*

Overall grading:

TOTAL	GRADE
91 – 100	A
81 - 90	B
71 - 80	C
61 – 70	D
51 – 60	E
0 – 50	F