

Abstract

The thesis investigates determinants of losing customer (customer churn) in the Czech energy sector. For this purpose, the data from MND Energie, a.s., one of the largest Czech energy suppliers, on average consumption, tariff, and sociodemographic characteristics about 9254 of their customers whose natural gas contracts terminated at the end of 2019 are used. The main goal of this thesis is to build a model capable of predicting probability of non-renewal of the individual customers' contracts. Before the contract termination date, some of the customers randomly selected from the dataset were directly notified of the possibility of a new fixed-price contract. The thesis, in compliance with its main goal, evaluates the influence of this treatment on the churn probability. The experiment has so far only been carried out in 2019. Thus, the thesis deals with supervised machine learning task performed on cross-sectional data. The logit leaf model (LLM) was chosen as the way of obtaining the desired predictions. The LLM algorithm used in this thesis was published in 2018 and it builds on previous research in this area. Its main contribution lies in combining the two generally accepted approaches, decision trees and logistic regression, in order to eliminate their disadvantages. LLM's performance was compared with the performance of its two building blocks used individually. The results were compared with relevant literature.

JEL Classification C53, L97, Q49

Keywords logit leaf algorithm, customer churn prediction, energy industry

Title Application of the logit leaf algorithm for customer churn prediction in the energy distribution industry in the Czech Republic

Author's e-mail 13903645@fsv.cuni.cz

Supervisor's e-mail karel-janda@seznam.cz