MDPI

The Analysis of Expenditures for Sewage Management and Water Protection in Poland [†]

Katarzyna Oszczapińska * and Iwona Skoczko

Faculty of Civil and Environmental Engineering, Białystok University of Technology, Wiejska 45A, 15-351 Białystok, Poland; i.skoczko@pb.edu.pl

* Correspondence: k.oszczapinska@doktoranci.pb.edu.pl; Tel.: +48-782-178-194

† Presented at Innovations-Sustainability-Modernity-Openness Conference (ISMO'19), Bialystok, Poland, 22–23 May 2019.

Abstract: The main aim of the research was to analyze expenditures for sewage management and water protection and to identify Polish commune types that are characterized by the highest expenses. The analyses were made using classification trees in Statistica software (Dell Statistica (data analysis software system), version 13, Dell Inc., Round Rock, TX, USA). For this purpose, a database consisting of seven independent variables was prepared.

Keywords: classification trees; sewage management; water protection

1. Introduction

The expenditures on fixed assets aimed at preserving the environment in the year 2017 reached PLN 6.8 billion. Despite increase in expenditures on environment protection, its relation to the Poland's Gross Domestic Product (GDP) is still at the low level. Apart from air and climate protection, expenditures on sewage management and water protection have the largest share in expenditures on environment protection, which in Poland in 2016, represented 1% of total expenditures, while average expenditures for EU was 1.6% [1,2].

Increasing support for water protection and sewage management can significantly improve the quality of environment. Water quality is affected by a wide range of natural and human influences, which are generally larger when available water quantities are low and maximum use must be made of the limited resource. Human activities may cause various changes on water quality, depending on pollution source. Inadequate amount of community facilities for water disposal or their inappropriate use contributes to fecal pollution [3].

2. Materials and Methods

Classification and Regression Trees Algorithm (C&RT) is the most advanced method of regression and classification trees. They are made by repeatedly data splitting. At each split, the data are partitioned into two diverse, maximum homogenous groups. The splitting procedure is then provided to each group separately. Splitting is continued until an overlarge tree is grown, which is then pruned back to the desired size. The size of a tree equals the number of final groups. Each group is typically characterized by either the distribution (categorical response) or mean value (numeric response) of the response variable, group size, and the values of the explanatory variables that define it.

Due to the response variable type, a classification tree based on the C&RT algorithm was used. In this research, our response variable are expenditures on sewage management and water protection, divided into seven classes (Table 1). The predictor variables were chosen to be:

Voivodeship, commune type and area, population, income, and percentage of citizens connected with waterworks and sewage system.

The impact of the mentioned predictors on expenditures was analyzed using STATISTICA software (Dell Statistica (data analysis software system), version 13, Dell Inc., Round Rock, TX, USA). Firstly, a database related to the year 2017 was created. Information about Polish communes was collected from Local Data Bank within Central Statistics Poland [4].

Data included 2478 communes. To avoid over-fitting of model, pruning and 10-fold cross validation were used.

3. Results and Discussion

After analyzing seven predictors, the most important was percentage of citizens connected with the sewers, and second was voivodeship with a validity value of 0.91 (Table 1). Assessed a priori probability was chosen.

Table 1. Importance of the predictors in C&RT model (own study).

Predictors	Range	Validity
% of citizens connected with the sewage system	100	1.00000
Voivodeship	92	0.91532
Commune's income	84	0.83742
Population	73	0.73085
Area	46	0.46407
% of citizens connected with the water system	45	0.45156
Commune's type	16	0.16403

As a result, a classification tree with 11 final nodes was obtained. Risk evaluation in the learning set reached 0.53 and in 10-fold cross validation, reached 0.54.

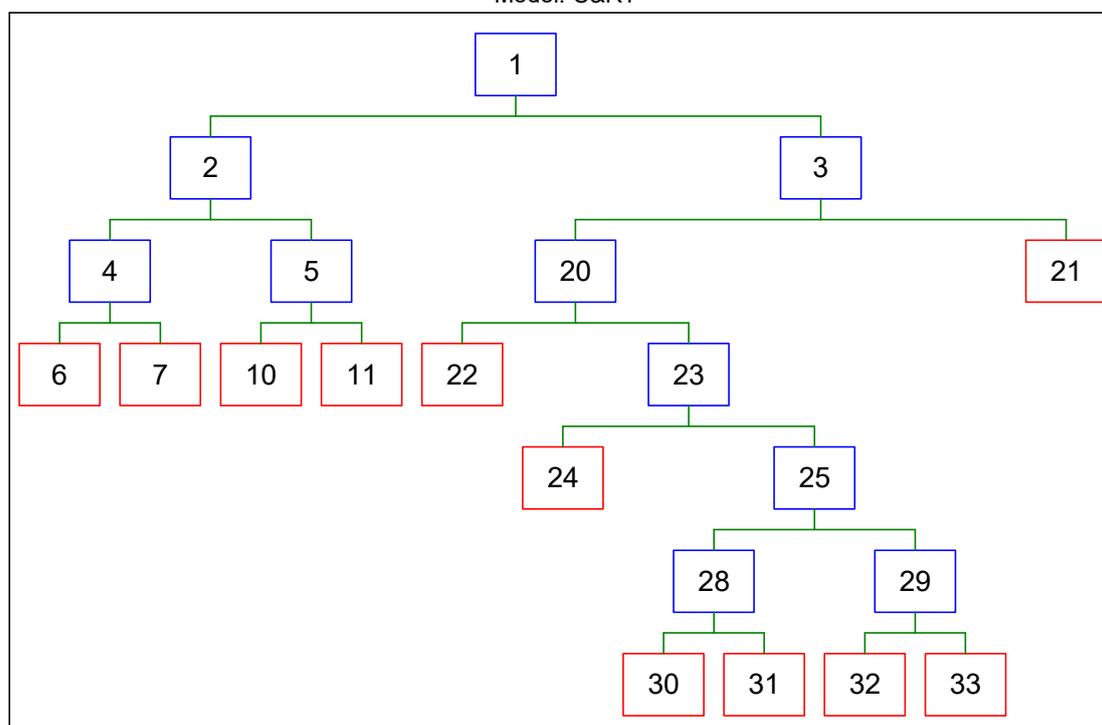Structure of the chosen tree is presented in the Figure 1.

Figure 1. Tree structure for expenditure for sewage management and water protection (own study).

Optimal tree structure was chosen based on v-fold cross validation and substitution cost, taking into account the lowest possible final node number (Table 2).

Table 2. Trees sequence and chosen tree (own study).

Tree No.	Final Nodes	Substitution Cost	Substitution Error	Node Complexity
Tree 1	17	0.520985	0.010035	0.000000
Tree 2	14	0.522195	0.010034	0.000404
Tree 3 *	11	0.525827	0.010031	0.001211
Tree 4	7	0.531073	0.010025	0.001312
Tree 5	4	0.535916	0.010018	0.001614
Tree 6	3	0.550444	0.009993	0.014528
Tree 7	1	0.632768	0.009684	0.041162

* chosen tree number.

Generally, it can be said that over half of the rural communes from Opole and Lower Silesia were characterized by the highest expenditures with respect to small area and population. But the most common were communes under PLN 1000 of expenditures, from Lublin, Masovian, Podlaskie, Greater Poland, West Pomeranian, Silesian, and Pomeranian voivodeship, with income over 42 million, population above 5940, area higher than 95.5 km^2, and percentage of citizens connected to water and sewage systems adequately over 5% and 35.5% (final node number 21). In turn, over half of the communes from Subcarpathian and Lesser Poland voivodeships, with comparable area and population as in Opole and Lower Silesia node, had the lowest expenditures, which was possibly correlated with poorly developed water and sewage systems in rural areas.

It has to be mentioned that about 40% of rural communes had expenditures under PLN 1000. More importantly, only 9% of all rural communes were characterized by higher expenditures on sewage management and water protection than 1 million.

Author Contributions: K.O. and I.S. conceived and designed the experiments, performed the experiments, analyzed the data, and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chojnacki, J.J. Nakłady na środki trwałe służące ochronie środowiska i gospodarce wodnej w Polsce a nakłady inwestycyjne w gospodarce narodowej - badania GUS. *Przegląd Nauk Ekonomicznych* **2017**, *25*, 163–173.
2. Eurostat. Government expenditure on environmental protection, 2018. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php/Government_expenditure_on_environmental_ protection (accessed on 9 January 2019).
3. World Health Organization & United Nations Environment Programme. *Water Quality Monitoring—A Practical Guide to the Design and Implementation of Fresh Water Quality Studies and Monitoring Programmes*; Bartram, J., Ballance, R., Eds.; E & FN Spon: London, UK, 1996.
4. Central Statistics Poland: Local Data Bank. Available online: https://bdl.stat.gov.pl/BDL/start (accessed on 7 January 2019).