

1. Introduction

As part of the Coursera Capstone project, I have taken a Classification problem. The data has been obtained from UCI Machine Learning Library.

2. Business Problem

To predict the presence of amphibian species near the water reservoirs based on the features obtained from GIS and satellite information

3. Data

The data used was derived from GIS and satellite information, as well as from information gathered from the natural inventories that were prepared for the environmental impact assessment (EIA) reports for two planned road projects (Road A and Road B) in Poland. These reports were mostly used to gather information on the size of the amphibian population in each of the 189 occurrence sites.

Road A project concerned part of the planned A1 motorway section in Pyrzowice; the section is located along the northern border of the Silesian Voivodship and is about 75 km long. The field research involved a strip of land with a width of 500 m on both sides of the proposed project area. The field inventory was carried out in 2010 and 2011. The results of these inventories were complemented by Marek Sołtysiak observations, which were conducted between 2014 and 2016. Finally, the first project included 80 amphibian breeding sites.

Road B inventory was prepared in the vicinity of two variants of the planned Beskid Integration Way on the Bielsko Biala-Wadowice-Głogoczów section of the S52 motorway. The length of this section of road is approximately 60 km. During the inventory, which was taken in 2010, 125 real and potential amphibian occurrence sites were described in. The methodology of the herpetological inventory included map analysis, literature and archive data analysis, and, then, field observations. As in the first case, the inventory was made in the spring time and consisted of the observation of the occurrence of amphibians in water reservoirs. The research area included a 500 m wide belt for each of the considered variants of the planned road. In order to conduct the final experiments, 109 amphibian occurrence sites were taken into account.

Attribute information and description have been included in the Appendix.

Appendix

Attribute Information

List of attributes and types:

ID -> Integer

MV -> Categorical

SR -> Numerical

NR -> Numerical

TR -> Categorical

VR -> Categorical

SUR1 -> Categorical

SUR2 -> Categorical

SUR3 -> Categorical

UR -> Categorical

FR -> Categorical

OR -> Numerical

RR -> Ordinal;

BR -> Ordinal;

MR -> Categorical

CR -> Categorical

Green frogs -> Categorical; Label 1

Brown frogs -> Categorical; Label 2

Common toad -> Categorical; Label 3

Fire-bellied toad -> Categorical; Label 4

Tree frog -> Categorical; Label 5

Common newt -> Categorical; Label 6

Great crested newt -> Categorical; Label 7

Name and symbol type description:

1) ID - vector ID (not used in the calculations)

2) MV - motorway (not used in the calculations)

3) SR -> Surface of water reservoir numeric [m2]

4) NR -> Number of water reservoirs in habitat - Comment: The larger the number of reservoirs, the more likely it is that some of them will be suitable for amphibian breeding.

5) TR -> Type of water reservoirs:

a. reservoirs with natural features that are natural or anthropogenic water reservoirs (e.g., subsidence post-exploited water reservoirs), not subjected to naturalization

b. recently formed reservoirs, not subjected to naturalization

c. settling ponds

d. water reservoirs located near houses

e. technological water reservoirs

f. water reservoirs in allotment gardens

g. trenches

h. wet meadows, flood plains, marshes

i. river valleys

j. streams and very small watercourses

6) VR - Presence of vegetation within the reservoirs:

a. no vegetation

b. narrow patches at the edges

c. areas heavily overgrown

d. lush vegetation within the reservoir with some part devoid of vegetation

e. reservoirs completely overgrown with a disappearing water table

Comment: The vegetation in the reservoir favors amphibians, facilitates breeding, and allows the larvae to feed and give shelter. However, excess vegetation can lead to the overgrowth of the pond and water shortages.

7) SUR1 - Surroundings 1 - the dominant types of land cover surrounding the water reservoir

8) SUR2 - Surroundings 2- the second most dominant types of land cover surrounding the water reservoir

9) SUR3 - Surroundings 3 - the third most dominant types of land cover surrounding the water reservoir

Comment: The surroundings feature was designated in three stages. First, the dominant surroundings were selected. Then, two secondary types were chosen.

a. forest areas (with meadows) and densely wooded areas

b. areas of wasteland and meadows

c. allotment gardens

d. parks and green areas

e. dense building development, industrial areas

f. dispersed habitation, orchards, gardens

g. river valleys

h. roads, streets

i. agricultural land

The most valuable surroundings of water reservoirs for amphibians are areas with the least anthropopressure and proper moisture.

10) UR - Use of water reservoirs:

a. unused by man (very attractive for amphibians)

b. recreational and scenic (care work is performed)

c. used economically (often fish farming)

d. technological

11) FR - The presence of fishing:

a. lack of or occasional fishing

b. intense fishing

c. breeding reservoirs

Comment: The presence of a large amount of fishing, in particular predatory and intense fishing, is not conducive to the presence of amphibians.

12) OR - Percentage access from the edges of the reservoir to undeveloped areas (the proposed percentage ranges are a numerical reflection of the phrases: lack of access, low access, medium access, large access to free space):

a. 0-25%lack of access or poor access

b. 25%-50%- low access

c. 50%-75%-medium access,

d. 75% - 100%-large access to terrestrial habitats of the shoreline is in contact with the terrestrial habitat of amphibians.

13) RR Minimum distance from the water reservoir to roads:

a. <50 m

b. 50-100 m

c. 100-200 m

d. 200-500 m

e. 500-1000 m

f. >1000 m

Comment: The greater the distance between the reservoir and the road, the more safety for amphibians.

14) BR - Building development - Minimum distance to buildings:

a. <50 m

b. 50-100 m

c. 100-200 m

d. 200-500 m

e. 500-1000 m

f. >1000 m

Comment: The more distant the buildings, the more favorable the conditions for the occurrence of amphibians.

15) MR - Maintenance status of the reservoir:

a. Clean

b. slightly littered

c. reservoirs heavily or very heavily littered

Comment: Trash causes devastation of the reservoir ecosystem. Backfilling and leveling of water reservoirs with ground and debris should also be considered.

16) CR - Type of shore

a. Natural

b. Concrete

Comment: A concrete shore of a reservoir is not attractive for amphibians. A vertical concrete shore is usually a barrier for amphibians when they try to leave the water.

17) Label 1 -the presence of Green frogs

18) Label 2 - the presence of Brown frogs

19) Label 3 - the presence of Common toad

20) Label 4 - the presence of Fire-bellied toad

21) Label 5 - the presence of Tree frog

22) Label 6 - the presence of Common newt

23) Label 7 - the presence of Great crested newt