Tutorial Business Analytics

Tutorial 8a: DMC Support

Decision Sciences & Systems (DSS)
Department of Informatics
TU München

Tutorial Business Analytics

Outline

Today's topics:

- DMC Task
- Rules
- Support

Tutorial Business Analytics

DMC Task

Predict if an electric vehicle charging station is in use or available.

The car manufacturer BMW would like to know which of its charging stations for electric vehicles are in use or not at specific timestamps and locations.

18 features (incl. ID columns) and one binary target value (1 „in use“, 0 „available“)

Features give basic information about the time and place of each charging station. In addition, it is known when the station was last used („modified“) and how. The target value is the information whether or not the station is in use at the given timestamp.

Tutorial Business Analytics

DMC Task

Column name	Description	Range of values	Missing values
Id	Ping id	Natural number	No
portNumber	Port at charging station	Natural number	No
TimeStamp	Time of ping	Date	No
EI65_GEO_ID	Station id	Natural number	No
HOUSEHOLD_COUNT	Number of households in neighborhood	Natural number	Yes
TYP1_COUNT	Socket type	Natural number	No
FREECHARGE	No cost for charging	String	Yes
ADDR_LATITUDE	Location	Number	No
ADDR_LONGITUDE	Location	Number	No
ADDR_MUNICIPALITY	City	String	Yes
ADDR_POSTALCODE	Postal code	Number	No
ADDR_REGION	State	String	No
ADDR_STREET	Street	String	No
LAST_MODIFIED	Update of station details	Date	No
PREFERRED_PARTNER	Partner of charge now	String	No
VALIDATION_LAST_MODIFIED	Last time a BMW customer charged there	Date	No
IS_LSC_VALIDATED	Validation technique	String	No
Status	Availability of station	{0,1}, 0 for available	No

Data

- History of availability of electric vehicle charging stations
- Train: Use this set to train your model
- Test: Use this set to predict the classes with your trained model
- Data Preparation for both (!) training and test data set

Task

- Predict if a charging station is “in use” (1) or available (0)

Tutorial Business Analytics

Rules

Teams

- Team size: 1 – 3 members
- **Teams must be built before the first submission** (teams will be fixed after first submission!).
- Each student can only be member of one team within one Data Mining Cup.

Submissions

- Maximum number of valid submissions for each DMC: 10.
- Best ranked submission, only, will be taken into account for the ranking.
- For reasons of traceability you must use a fixed seed of 42 (**set.seed(42)**).
- So far, **only 4 teams** have submitted a solution.

Disqualification reasons:

- **Non-reproducible** submissions (submitted predictions must be reproducible using the submitted R script)
- **Hard-coded** classifications (even if the best ranked submission is not hard-coded!)
- **Copies** from other groups (disqualification of both teams)

Tutorial Business Analytics

Rules

Reproducability

- You must use a fixed seed of 42.

```
set.seed(42)
```

- Clear the whole environment once and then run the whole script from the beginning. Otherwise e.g. the repeated call of functions could lead to different results.

```
rm(list = ls())
```

Tutorial Business Analytics

Support

Let’s start!

Please ask if you have any questions.