
Traffic Prediction in Telecommunications

20596 – MACHINE LEARNING. Prof. Daniele Durante

March 5, 2019

1 PROBLEM DESCRIPTION

The dataset consists of information from **10179 customers of a European telephone company** who were active in the 10 consecutive months preceding the study and made at least one outgoing call in month 10. For **6000** of these customers you know the **logarithm of the total duration of outgoing calls, in seconds, for month 10 [denoted as y]**, and you possess additional **input variables describing several features of each customer's behavior**. For the other **4179** customers, you have only information on the inputs and not on their outgoing traffic in month 10.

Your goal is to predict y for the held-out **4179** customers.

There are **98 input variables**, which are described below.

- **Characteristic variables of customer and of company contract.**
 - **tariff.plan**: customer tariff plan [5 possible plans]
 - **payment.method**: method of payment [3 possible methods]
 - **gender**: gender [M-male, F-female, B-company]
 - **age**: customer's age in years
 - **activ.area**: geographical activation zone [4 possible zones]
 - **activ.chan**: channel of activation [8 possible channels]
 - **vas1**: presence of a first value-added service [N-no, Y-yes]
 - **vas2**: presence of a second value-added service [N-no, Y-yes]
- **Variables for the customer's traffic in the 9 months preceding the 10th. For each month $m = 1, \dots, 9$, the following variables are available**
 - **q0m.out.ch.peak**: total number of outgoing calls at peak tariff times in month m
 - **q0m.out.dur.peak**: duration of total outgoing calls at peak tariff times in month m
 - **q0m.out.val.peak**: total outgoing call value at peak tariff times in month m
 - **q0m.out.ch.offpeak**: total number of outgoing calls at off-peak tariff times in month m

- `q0 m .out.dur.offpeak`: duration of total outgoing calls at off-peak tariff times in month m
- `q0 m .out.val.offpeak`: total outgoing call value at off-peak tariff times in month m
- `q0 m .in.ch.tot`: total number of incoming calls in month m
- `q0 m .in.dur.tot`: duration of total incoming calls in month m
- `q0 m .ch.sms`: total number of SMS sent in month m
- `q0 m .ch.cc`: number of calls to Customer Services in month m

As already mentioned the above variable are available for each month $m = 1, \dots, 9$.