Exposee Bayesian Networks - Bank Marketing

By Maarten Berenschot and Daan Brugmans

Problem Domain

Banks always try to acquire more new long-term deposits from their customers. To that purpose, banks try out various methods for reaching out to customers that might be interested in starting a new long-term deposit. One such method is calling potential new depositors via phone. Unfortunately, this method far from guarantees that the potential customers reached will actually subscribe to a term deposit. Preferably, we aim to contact only those potential depositors that have the highest likelihood of subscribing to a new term deposit. In an attempt to model the type of customer that would be most likely to subscribe to a new term deposit, multiple campaigns were undertaken. In these campaigns, current customers of the bank were contacted via phone and were asked about their interest in subscribing. For these customers, we ask ourselves: which characteristics are the most relevant for determining the likelihood of a current bank customer subscribing to a new long-term deposit?

Data

For this project, we will use data collected from the campaigns mentioned in the Problem Domain. The dataset contains data on the calls made to current customers of the bank that may be interested in becoming new long-term depositors. The data is available at the following [URL](URL). The dataset contains 45211 instances of calls to potential new depositors and 16 descriptive variables describing the calls and characteristics of the potential depositor called, and a binary target variable describing if the potential depositor opened a new term deposit. Every data point describes an individual call. For the purposes of this project, only a subset of the descriptive features will be taken into account, alongside the target variable. The descriptive variables used for this project are as follows: age of the callee, job category of the callee, marital status of the callee, highest achieved education level of the callee, the average yearly household balance of the callee, if the callee is currently defaulting on a loan or not, if the callee has a housing loan or not, if the callee has a personal loan or not, duration of the call in seconds, count of contacts performed for the current campaign, count of contacts performed for previous campaigns, and if the previous campaign was considered a success/failure.

Implementation plan

For this project, we will use the programming language [R](R). We will use the dagitty package for analyzing directed acyclic graphs (DAGs) and for testing the dependencies in the graph using d-separation. We will use the BayesianNetwork package to build and analyze Bayesian networks, although we will consider the bnlearn package as an alternative for these purposes. If relevant for our project, we will use the lavaan package for latent variable analysis.

Application

For this project, we will research which variables influence the target variable the strongest. That is to say: which characteristics of the call and the callee influenced whether the callee would open a new term deposit the most? Furthermore, we will look at how two specific descriptive features influence the target feature, namely the average yearly balance of the callee and the duration of the call. As a research question: to what extent did the average yearly household balance of the callee and the duration of the campaign call influence whether the callee would open a new term deposit or not?