

The Insurance Company Benchmark (COIL 2000)

Task Type

regression, description

Problem Description

For the COIL Competition there were two tasks:

- Predict which customers are potentially interested in a caravan insurance policy.
- Describe the actual or potential customers; and possibly explain why these customers buy a caravan policy.

Prediction

We want you to predict whether a customer is interested in a caravan insurance policy from other data about the customer. Information about customers consists of 86 variables and includes product usage data and socio-demographic data derived from zip area codes. The data was supplied by the Dutch data mining company Sentient Machine Research and is based on a real world business problem. The training set contains over 5000 descriptions of customers, including the information of whether or not they have a caravan insurance policy. A test set contains 4000 customers.

For the prediction task, the underlying problem is to find the subset of customers with a probability of having a caravan insurance policy above some boundary probability. The known policyholders can then be removed and the rest receives a mailing. The boundary depends on the costs and benefits such as of the costs of mailing and benefit of selling insurance policies. To approximate this problem, we want you to find the set of 800 customers in the test set that contains the most caravan policy owners.

Description

The purpose of the description task is to give a clear insight to why customers have a caravan insurance policy and how these customers are different from other customers.

Descriptions can be based on regression equations, decision trees, neural network weights, linguistic descriptions, evolutionary programs, graphical representations or any other form. of solutions (e.g. minimize a loss function, maximize comprehensibility, minimize response time, etc.)?

The descriptions and accompanying interpretation must be comprehensible, useful and actionable for a marketing professional with no prior knowledge of computational learning technology. The value of a description is inherently subjective.

Results

Visit the [TIC Benchmark website](#) for background information and results on this data. For example it contains an online report on the CoIL Challenge 2000, in which this dataset was used, consisting of 29 papers on the dataset and an overview.