**Coil Competition Data**

This data set is from the Computational Intelligence and Learning (COIL) competition. The data contains measurements of river chemical concentrations and algae densities.

**Dataset Characteristics**

Multivariate

**Subject Area**

Physical

**Associated Tasks**

-

**Attribute Type**

Categorical, Real

**# Instances**

340

**# Attributes**

17

**Information**

**Additional Information**

This data comes from a water quality study where samples were taken from sites on different European rivers of a period of approximately one year. These samples were analyzed for various chemical substances including: nitrogen in the form of nitrates, nitrites and ammonia, phosphate, pH, oxygen, chloride. In parallel, algae samples were collected to determine the algae population distributions. The competition involved the prediction of algal frequency distributions on the basis of the measured concentrations of the chemical substances and the global information concerning the season when the sample was taken, the river size and its flow velocity. The competition instructions contain additional information on the prediction task: http://kdd.ics.uci.edu/databases/coil/instructions.txt

SHOW MORE

**Has Missing Values**

Symbol: 0

**Attribute Information**

**Additional Information**

There are a total of 340 examples each containing 17 values. The first 11 values of each data set are the season, the river size, the fluid velocity and 8 chemical concentrations which should be relevant for the algae population distribution. The last 8 values of each example are the distribution of different kinds of algae. These 8 kinds are only a very small part of the whole community, but for the competition we limited the number to 7. The value 0.0 means that the frequency is very low. The data set also contains some empty fields which are labeled with the string XXXXX. The training data are saved in the file: analysis.data (ASCII format). Table 1: Structure of the file analysis.data A ... K a ... g CC1,1 ... CC1,11 AG1,1 ... AG1,7 ... CC200,1 ... CC200,11 AG200,1 ... AG200,7 Explanation: CCi,j: Chemical concentration or river characteristic AGi,j: Algal frequency The chemical parameters are labeled as A, ..., K. The columns of the algaes are labeled as a, ..,g.