CSE532 Artificial Intelligence
PROJECT DESCRIPTION
BAKARY DATA - on the course web page.

This is a classification data with TYPE DE ROCHE (Rock Type) as a CLASS attribute. There are 98 records with 48 attributes and 6 classes.

Classes are:

C1 : R. Carbonatees AND R. Carbonatees impures

C2 : Pyrate

C3 : Charcopyrite

C4 : Galene
C5 : Spahlerite

C6 : Sediments terrigenes

**Most important attributes** (as determined by the expert) are: S, Zn, Pb, Cu, CaO+MgO, CaO, MgO, Fe2O3

This is a real life experimental data and it contains a lot of missing data (no value).

**THE PROJECT GOAL** is to use different Internet based CLASSIFICATION TOOLS (choose one you like) to generate sets of DISCRIMINANT RULES describing the content of the data.

**The project** has to follow all steps of Learning Process:
**Data Preparation** that includes attributes selection, cleaning the data, filling the missing values, etc...

**Data preprocessing**: must use at least 2 methods of data discretization, and compare the final results obtained after each of them.

**Learning Proper**: for each experiment describe below use a classification tool for rules generation applied to the TWO sets of preprocessed data and compare the results.

**Discriminant Rules Generation Experiments**; you have to perform 3 experiments (all on the same preprocessed data)
**Experiment 1**: use all records to find rules for the full classification; i.e. rules describing all classes \( C1 - C6 \) simultaneously.

**Experiment 2**: use all records to find rules contrasting class \( C1 \) with all others.

**Experiment 3**: repeat Experiment 1 for all records with the **most important attributes** only.

**Write a detailed Project Description** with methods, motivations, results and submit it to the Professor in a folder (and CD) on the day of your PROJECT PRESENTATION.

**Project Presentation**: each student, or a group will be given 10-15 minutes to present the project and results.