PROJECT DESCRIPTION

THE PROJECT GOAL is to use Internet based Classification Tools to **build wo classifiers:** one descriptive and one statistical and **discuss** the results and **compare** these to approaches on the basis of obtained results.

1. Descriptive Classifier

Use a **Decision Tree** tool to generate sets of **discriminant rules** describing the content of the data.

You can choose one you like, or use WEKA:
http://www.cs.waikato.ac.nz/~ ml/weka/index.html)

2. Non-Decsriptive Classifier

Use **Neural Networks** tool to build your Classifier

Choose your own tool: here are some suggestions:

http://www.mathworks.com/products/neural-network/?requestedDomain=www.mathworks.com

http://www.simbrain.net/

PROJECT DATA is provided on the course web page.

This is a real life 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 has to follow all steps of **Learning Process**

Data Preparation that includes attributes selection, cleaning the data, filling the missing values, etc...

Data preprocessing

For the **Descriptive Classifier** you must use at least 2 methods of data discretization, and compare the final results obtained after each of them.

Learning Proper

For each of the **experiments 1- 3** describe below use a classification tool for rules generation applied to your 2 sets of preprocessed data and compare the results.

For the **Non -descriptive Classifier** use your chosen method of preprocessing and perform the **Learning Proper** for each the **experiments 1- 3** describe below.

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 Experiments 1, 2 for all records with the **most important attributes** as defined thy the expert only.

Write a detailed Project Description with methods, motivations, results and submit via e-mail to TA and Professor.