

## PROJECT DESCRIPTION

**THE PROJECT GOAL** is to use Internet based Classification Tools to build two type classifiers: **descriptive** and **non-descriptive**. Discuss the results. **Compare** these two 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.

Use WEKA:

<http://www.cs.waikato.ac.nz/~ml/weka/index.html>)

### 2. Non-Descriptive Classifier

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

Use WEKA or a tool of your choice. Describe specifics of your tool in a way that makes your report comprehensible for others.

Here are some tools 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 the following steps of **DM Process** to build the classifiers.

**S1: Data Preparation** that includes attributes selection, cleaning the data, filling the missing values, etc... to build Project DATA - **PD**.

**S2: Data preprocessing**

1. For the Decision Trees **Descriptive Classifier** you use 2 methods of data discretization to the Project Data **PD** creating two data sets: **PD1** and **PD2**. Describe which methods you used.
2. For the Neural Network **Non -descriptive Classifier** use the Project DATA - **PD** and your tool method of normalization of your choice. Specify which.

### Building Classifiers

For each sets of data **PD1**, **PD2** ( for Decision Trees), and **PD** (for Neural Networks) perform the following **Experiments 1- 3**.

For each Experiment **compare** the resulting **Descriptive Classifiers** with each other and compare each **Descriptive Classifier** with the resulting **Non-Descriptive Classifier**.

### Experiments 1- 3

**Experiment 1** : use all records to perform the **full classification** (learning), i.e. build a classifier for all classes **C1- C6** simultaneously.

**Experiment 2** : use all records to perform the **contrast classification** (contrast learning), i.e. contrasting class **C1** with a class **notC1** that contains other classes.

**Experiment 3** : repeat Experiments 1, 2 for all records with the **most important attributes** as defined by the expert only.

**Write a detailed Project Description** with methods, motivations, results and submit via Blackboard and e-mail to TA Sayontan Ghosh [sayontan.ghosh@stonybrook.edu](mailto:sayontan.ghosh@stonybrook.edu)