

## CSE352 ARTIFICIAL PROJECT SPRING 2022 (40pts)

### TEAM NUMBER:

Team Leader **NAME** and **ID**:

Team members **NAMES** and **IDs**:

- 1.
- 2.
- 3.

### FORMAT OF SUBMISSION

1. The TEAM LEADER submits **ONE PDF** file for the TEAM
2. Name your PDF file as `<TEAMLEADERID>.pdf`

**Example:** 11384578.pdf

**All team members receive the same grade.**

## PROJECT DESCRIPTION

**PROJECT REPORT LENGHT** - the report must be concise and well thought out and can not exceed 15 pages.

### PROJECT GOAL

The main goal of the project is to use the Project Data and Internet based Classification **TOOLS** to conduct 3 Experiments of building a **descriptive** Decision Tree classifier and a **non-descriptive** Neural Network classifier for the same sets of **DATA**.

For each Experiment you have to compare descriptive and non-descriptive obtained results and describe and compare the functionality of the **TOOLS** used.

### PROJECT TOOLS

**TOOL WEKA:** <https://www.cs.waikato.ac.nz/ml/weka/>

**LIST OF TOOLS** <https://www.softwaretestinghelp.com/data-mining-tools/>

### 1. Descriptive Classifier

Use the the TOOL WEKA to build the Decision Tree classifier and present the results in a form of **discriminant rules** describing the content of the data.

### 2. Non-Decriptive Classifier

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

Use TOOL WEKA or a tool from LIST OF TOOLS.

3. Describe specifics of TOOLS used in a way that makes your report comprehensible for others.

**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).

### PROJECT STEPS

Project follows the following steps of **Learning Process** to build the classifiers.

#### **S1: Data Preparation**

Use attributes selection, cleaning the data, filling the missing values, etc... operations to build your

Project DATA - **PD**. Explain shortly your decisions.

#### **S2: Data preprocessing**

1. For the Decision Trees **Descriptive Classifier** you use your chosen method of the **discretization** of Project DATA - **PD** creating a set **PD1** of data with no more than 4 values (bins) for each

attribute. Different attributes do not need to have the same number of values (bins) and you do not need to use the same discretization methods for all of them.

Describe which discretization method you used for each attributes. You must use at least **two methods**.

**2.** For the Neural Network **non -descriptive Classifier** use the Project DATA - **PD** and the TOOL method of your choice of **normalization**. Specify which.

### **S3: 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**, i.e. contrasting class **C1** with a class **notC1** that contains other classes.

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

Write a detailed **Project Report** explaining all methods used, motivations, experiments results and their comparison, and submit it via Blackboard.