

## CSE352 PROJECT HOMEWORK (10 extra points)

**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 HOMEWORK DESCRIPTION

**WEKA** Machine Learning Project TOOL

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

### DATA PREPARATION

Follow the following **steps** to prepare the BAKARY DATA for your Project.

#### 1. Remove Records

Decide, if you want to remove some records that need to be removed, if any. Describe your reason.

#### 2. Remove attributes

Remove attributes that have a large amount of missing data; for example 10%, or any % you decide appropriate. Describe your decision.

Remove attributes that have out of range values, etc...

#### 3. Fill missing data

In order to fill missing values of the attributes **you decide to keep** you can use a method of your choice, but **must describe your choice**.

You can use the WEKA tool (if you are already ready to do so)

You can compute, for each attribute the average value.

REMEMBER: must do it within the same class.

You can find another method, or write your own program to do so...

### **3. Discretize data**

Use the binning method with 3-4 bins.

You can do it by hand, or write a program, or use the WEKA.

Replace values in the bin by appropriate number, or interval, or a chosen symbol like your name of the bin, etc ...

**4. Submit** prepared data with explanation of all steps via **Blackboard** and KEEP in your records

### **Remember**

THIS IS THE DATA YOU WILL USE FOR YOUR PROJECT.