**CME4432 ASSIGNMENT – I**

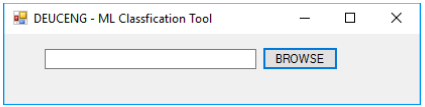
**(Due Date: 18.11.2019 11:59 PM)**

Dummy attribute technique has a problem like causing the big size of the dataset. Therefore, regression trees can use to convert the nominal attributes to numeric without any size problem. Twoing and Gini algorithms as regression trees find a numeric gain value for each nominal cell value. In this assignment, you will convert all attributes to numeric values using Twoing and Gini algorithms separately.

* The file format of the dataset must be ARFF (of WEKA tool). While reading the dataset from the code, you must use the following expressions in the WEKA.dll for C# implementations;

weka.core.Instances insts = new weka.core.Instances(new java.io.FileReader("iris.arff"));

* iris.arff is an example file. You must browse the file from the window of your application like the following figure;



* Some attributes can be already numeric. You won’t change any value for these attributes, but for using Twoing and Gini, you must discretize the dataset firstly like the following code;

weka.filters.unsupervised.attribute.Discretize myDiscretize = new weka.filters.unsupervised.attribute.Discretize();

myDiscretize.setInputFormat(insts);

myDiscretize.setFindNumBins(true);

insts = weka.filters.Filter.useFilter(insts, myDiscretize);

* You can catch the attribute you control; it is either numeric or not, like the following code;

insts.attribute(i).isNumeric()

* You can find the sub categories for each attribute like the following code;

for (int i=0;i<insts.numAttributes();i++)

{

for (int j = 0; j < insts.attribute(i).numValues(); j++)

string sub\_category = insts.attribute(i).value(j);

}

* Lastly, you can put all values of the dataset in a list data structure like the following code;

List<List<string>> lst = new List<List<string>>();

for(int i=0;i<insts.numInstances();i++)

{

lst.Add(new List<string>());

for(int j=0;j<insts.instance(i).numValues();j++)

{

lst[lst.Count - 1].Add(insts.instance(i).toString(j));

}

}