CST8390 - Lab 2
Explore CSV file and transform it into ARFF file

Due Date: Sept 20, 24, 25 in corresponding lab sessions

Introduction

The goal of this lab is to explore a CSV file and transform it to ARFF format. Then, load the file in Weka and find statistics.

Steps:

1. Download EmployeesSalary.csv file from Brightspace
2. Open EmployeesSalary.csv in excel and explore it
3. Read https://www.cs.waikato.ac.nz/ml/weka/arff.html to find the expectations of an ARFF file
4. Identify the attributes of the data. Record the attributes and the type of attribute for the data.
5. Closely analyse data. In excel, do the required modifications to match with the requirements for an ARFF file. (Hint: Check the requirements if the data has spaces in it.)
6. Open the file in notepad++. Add the required section headers and corresponding information in the file. Save the file as EmployeesSalary.arff. This involves creating the @relation line, one @attribute line per attribute, and @data to signify the start of data. It is also considered good practice to add comments at the top of the file describing where you obtained this data set, what its summary characteristics are, etc. A comment in the ARFF format is started with the percent character % and continues until the end of the line.
7. Open the ARFF file as you did in lab 1 (by selecting ‘Open file’ in the ‘Preprocess tab’.). You may run into errors as you load your ARFF file. If so, check the requirements to troubleshoot your problem.
8. Which are the four important attributes that are relevant for data analysis?
    a.
    b.
    c.
    d.

9. For the nominal attributes from Question 8, fill in the following table:

   Attribute: _____

Label	Count

   Attribute: _____

Label	Count

   Attribute: _____

Label	Count

In order to get the credit for this lab:

1. Show the loaded file in Weka (5 marks)
2. Fill in the tables for questions 8 & 9 (5 marks)