CS 584 – Data Mining
CA2 – Processing ARFF file

The Problem

In this assignment, you are to write a C++ program (using Visual C++ 2017 or above) to perform a relatively simple task: read training data from arff file and store them in data structures of your choice. ARFF (Attribute Relation File Format) is used in variety of application including weka machine learning workbench. For this assignment, you (actually, your program) read the data, store them, and print the data. Your program will print the list of attributes followed by their values and then print the table data.

There are many ways to store the data efficiently for easy update and retrieval. For example, you could store attributes and their value lists in a vector<attributeType> where attributeType could be a struct containing attribute name, and a vector<string> to store values. You could store table either in a simple 2D array of int or a vector<vector<int>> where the int contains the index in the value vector for the particular attribute whose index is the same as the column value of the table.

Sample Input

weather.nominal.arff
contact-lenses.arff
restaurant.arff

Sample Output

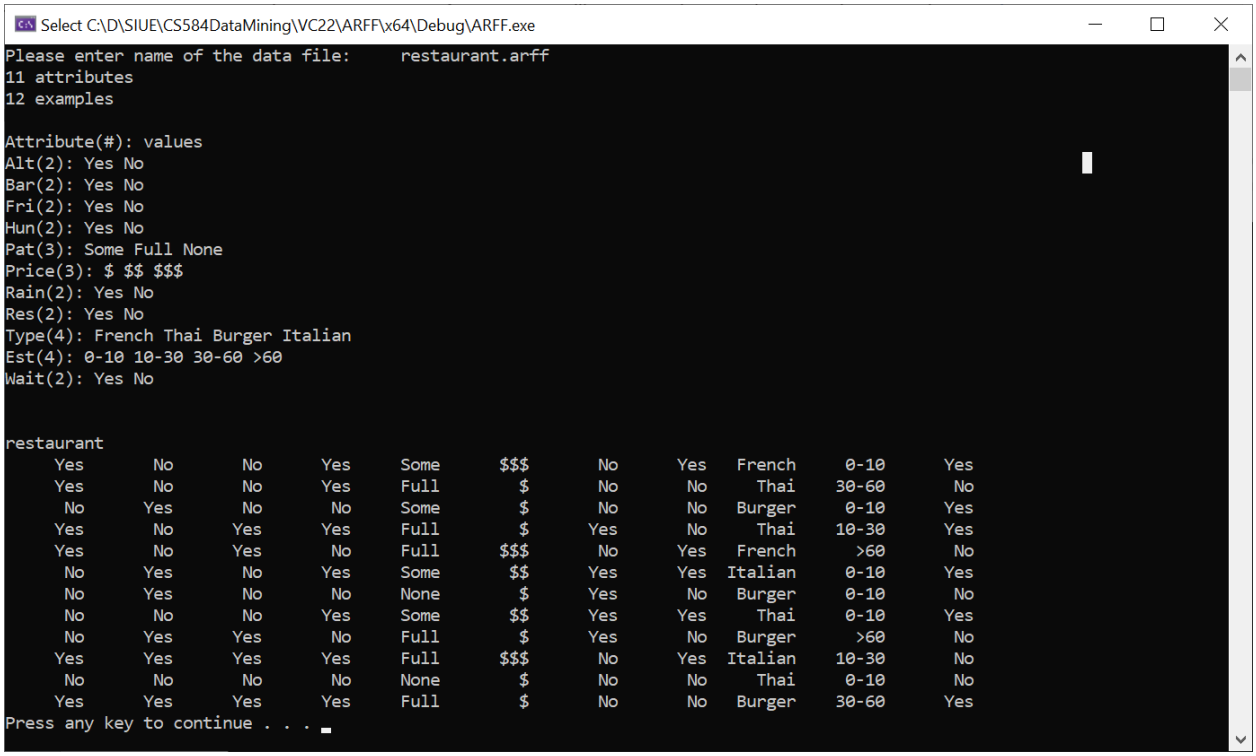
Run arff.exe on the sample input file. Sample output screenshots are also included at the end of this document.

Required Comments

For all programming assignments, you must include your name(s), due date, and problem description at the beginning of your program. Additional comments are necessary for blocks of code such as classes, functions, etc. For this assignment, you must also include a paragraph briefly discussing your data structure and the advantages/disadvantages. If you used ChatGPT or any AI tool, describe what you used and learned. These comments are 20% of the grade.

Due Time

This assignment is *due Tuesday (2/18) at 10 PM.* Submit a zip file which contains your *entire project folder* (or simply your .cpp file(s)) on Moodle by the due time. As noted previously, you may work in **groups of two** for all CA assignments, in which case only one submission is necessary. Late submission will be accepted until Thursday (2/20) at 10 pm. Standard late penalty applies.


Select C:\D\SIUE\CS584DataMining\VC22\ARFF\x64\Debug\ARFF.exe
Please enter name of the data file: restaurant.arff
11 attributes
12 examples
Attribute(#): values
Alt(2): Yes No
Bar(2): Yes No
Fri(2): Yes No
Hun(2): Yes No
Pat(3): Some Full None
Price(3): $ $$ $$$
Rain(2): Yes No
Res(2): Yes No
Type(4): French Thai Burger Italian
Est(4): 0-10 10-30 30-60 >60
Wait(2): Yes No
restaurant
Yes No No Yes Some $$$ No Yes French 0-10 Yes
Yes No No Yes Full $ No No Thai 30-60 No
No Yes No No Some $ No No Burger 0-10 Yes
Yes No Yes Yes Full $ Yes No Thai 10-30 Yes
Yes No Yes No Full $$$ No Yes French >60 No
No Yes No Yes Some $$ Yes Yes Italian 0-10 Yes
No Yes No No None $ Yes No Burger 0-10 No
No No No Yes Some $$ Yes Yes Thai 0-10 Yes
No Yes Yes No Full $ Yes No Burger >60 No
Yes Yes Yes Yes Full $$$ No Yes Italian 10-30 No
No No No No None $ No No Thai 0-10 No
Yes Yes Yes Yes Full $ No No Burger 30-60 Yes
Press any key to continue . . .