Programming Assignment 1: Decision Trees

Part 1: Implementation [7 points]

Your job in this exercise is to predict whether you will have a good night-out in Jerusalem for the coming New Year's Eve. Assume that you have kept a record of your previous night-outs with the following attributes.

- The **size** of the place {Large, Medium, Small}
- How densely the place is usually **occupied** {High, Moderate, Low}
- How the **prices** are {Expensive, Normal, Cheap}
- Volume of the **music** {Loud, Quiet}
- The **location** {Talpiot, City-Center, Mahane-Yehuda, Ein-Karem, German-Colony}
- Whether you are a frequent customer (**VIP**) {Yes, No}
- Whether this place has your **favorite beer** {Yes, No}
- Whether you **enjoyed** {Yes, No}

We have provided a data file (dt-data.txt) that contains the relevant records.

(a) Write a program to construct decision trees based on the idea of splitting by Information Gain.
(b) Run your program on the data file.
(c) Make a prediction for (size = Large; occupied = Moderate; price = Cheap; music = Loud; location = City-Center; VIP = No; favorite beer = No).

You can write your program in any programming language. However, you will have to implement the decision tree algorithm yourself instead of using library functions. Provide a description of the data structures you use, any code-level optimizations you perform, any challenges you face, and of course, the requested prediction.

Your code should print the decision tree that it produces. The format is

attribute on the 1st level
1st attribute on the 2nd level, 2nd attribute on the 2nd level, 3rd attribute on the 2nd level ...
...

For example, if we use **size** as the root, **occupied**, **prices** and **music** as the attributes on the second level corresponding to size={Large, Medium, Small} respectively, then it should look like the following:

size
occupied, prices, music
… More levels and attributes …
(The final level)Yes, No, Yes, Yes, …

Part 2: Software Familiarization [2 points]

Do your own research and find out about a library function that offers a good implementation of the

decision tree algorithm. Learn how to use it. Compare it against your implementation and suggest some ideas for how you can improve your code. Describe all this in your report.

Part 3: Applications [1 point]

Do your own research and describe some interesting applications of decision trees.

In your report, please include the names of all group members and mention their individual contributions. Submissions should include the code as well as the report and are due before 09/17, 11:59pm. Please submit by email to **both** hongx@usc.edu and nakarese@usc.edu (do not email to the instructor).