THE COLLEGE OF SAINT ROSE
CSC 535: INTRODUCTION TO ARTIFICIAL INTELLIGENCE
PROJECT FOUR

REQUIREMENTS:

There are three (3) phases to this project:

Phase One:

Do some additional reading and research into Naïve Bayes classifiers. There will be additional reading material posted with the assignment. These are very simple classification mechanisms, so I do not anticipate that you will have trouble understanding how to implement one.

Phase Two:

Using a programming language of your choice, implement a naïve Bayes classifier to classify pages of text, based on which words appear on the page. To do this, you will first need to train the classifier with pre-classified examples of pages. You should choose two (2) classifications: interesting and not interesting, and try to make the interesting category fairly narrow: for example, it might be "pages about the Yankees." Very broad categories would make this assignment much more difficult (for example, "pages about sports" would present a much more challenging assignment).

Phase Three:

Provide a brief write-up of your project. Describe the steps you went through in the design and implementation of your naïve Bayes classifier. Detail what you used as your criteria for determining whether or not a page was "interesting". Summarize the results.

NOTES:

As for input data, feel free to create a bunch of text files, where you just copy plain text from different sources (for example, online books, text-based web pages, and so on). You will will have to do this "data collection" phase yourself so you have enough to train your Bayesian network. About 15 – 20 pages would be plenty! Also collect a handful of pages that you can use to test your classifier.

ADDITIONAL REQUIREMENTS:

Your program must be robust and relatively easy to use. There should be some elegance to the code you write and must contain some documentation. Your may lose a point if your code looks "hacked".

PROGRAMMING LANGUAGES ALLOWED:

You may use any standard programming language to complete this assignment (Java, C, C++, C#, Python). You are responsible for ensuring that your program works given the installed compiler versions.

SUBMISSION:

Please submit all of the source code for your assignment, along with your code walk-though video. You must also include a readme.txt file that explain how to get your program running, just in case there are any issues. Also explain any non-standard user-interface issues.