DEPARTMENT *of* COMPUTING
College of Business & Technology
EAST TENNESSEE STATE UNIVERSITY

CSCI 5260 – Artificial Intelligence
Lab 9 – Naïve Bayesian Inference

Overview

Creating a predictive model is a very common activity in artificial intelligence. In this lab, you apply Naïve Bayes inference to the Kaggle adult Data Set, which asks you to predict whether income exceeds $50,000 per year based on census data.

Step 1 – Download Files

Download the lab9.py file and the adults.csv file from D2L and place them in the same folder.

The lab9.py file uses the following libraries:

- numpy
- pandas
- matplotlib.pyplot as plt # for data visualizations
- sklearn.naive_bayes.GaussianNB
- seaborn
- warnings

Run the lab9.py program to ensure all of the libraries work, and the program loads the CSV file correctly. Others may be added along the way, but this will get you started.

Step 2 – Tutorial

Complete the remainder of the tutorial on the following site, beginning with Step 7 and STOP at the end of Step 15:

- https://www.kaggle.com/prashant111/naive-bayes-classifier-in-python

Step 3 – Questions

Answer the following questions in Lab9.docx.

1. Explain why feature engineering was necessary in this problem.
2. Explain why feature scaling was necessary in this problem.
3. Does this model have a good accuracy? Why or why not?
4. Speculate about ways to improve the accuracy of the prediction.
5. Given these results, and **referring to this problem specifically**, explain why some use the term "idiot Bayes."

SUBMISSION

Create a Word Document named **Lab9.docx** with your responses to the questions. Submit your complete **lab9.py** file.

Submit to the Lab 9 dropbox at or before Monday, April 5, 2021 by 11:59 PM.

GRADING

A letter grade will be assigned for each response. The letter grades are based on both correctness and the adequacy of answers. Points are assigned as follows:

		A	B	C	D	F	Zero
		Excellent	Above Average	Average	Below Average	Poor	No Attempt
		10	8	6	4	2	0
Naïve Bayes Classification	Step 2						
	Step 3-1						
	Step 3-2						
	Step 3-3						
	Step 3-4						
	Step 3-5						