

Assignment 2 – Naïve Bayesian Classification

In this assignment, you are to write Python code (on your own, no use of built-in libraries, aside from numpy) to perform Naïve Bayes classification on the following testing data where the attribute “Private/Public” is the target variable:

University Name	University State	Private/Public	SAT verbal	SAT math	Academics (1 – 5)	Social (1 – 5)	Quality of Life (1 – 5)
University at Albany	New York	public	525	575	4	3	3
Syracuse	New York	private	600	600	4	4	4
Tufts	Massachusetts	private	580	620	4	4	4
Golden Gate College	California	private	500	500	2	1	1
San Jose State	California	public	425	565	4	2	3

Using the following data as training data:

University Name	University State	Private/Public	SAT verbal	SAT math	Academics (1 – 5)	Social (1 – 5)	Quality of Life (1 – 5)
Adelphi	New York	private	500	475	2	2	2
Boston College	Massachusetts	private	500	550	4	5	3
Boston University	Massachusetts	private	550	575	4	4	3
Cal Tech	California	private	650	780	5	1	3
City College NY	New York	public	480	525	3	2	2
Columbia	New York	private	625	650	5	3	3

Harvard	Massachusetts	private	700	675	5	3	4
Hofstra	New York	public	500	525	2	2	2
MIT	Massachusetts	private	650	750	5	3	3
Stanford	California	private	625	675	5	4	5
Berkely	California	public	530	600	5	3	3
Davis	California	public	550	600	4	4	4
UCLA	California	public	500	550	4	3	3
San Diego	California	public	550	600	4	4	4
USC	California	private	475	525	4	4	3
Worcester	Massachusetts	private	550	560	4	3	4
Queens	New York	public	450	450	4	3	3
University of Lowell	Massachusetts	public	472	535	3	3	3
University of Mass.	Massachusetts	public	480	510	3	3	3
University at Buffalo	New York	public	450	525	4	2	2

You should treat all of the ordinal data as nominal for the purposes of training and testing. For this data, do NOT scale or otherwise adjust any of the data when training or testing. Your code must simply output the confusion matrix for the testing data.