UC San Diego Extension
Advanced Web Analytics:
Harnessing the Predictive Power
Winter 2016
Homework#6

Date Given: March 7, 2016   Due Date: March 13, 2016
==

1. Naïve Bayes Method: The following data is captured by Analytics service of a website.

Visitor ID	Previously Visited	Session Length	Page Views	Will Visit Again
1	No	Long	Medium	No
2	Yes	Short	High	No
3	No	Long	Low	No
4	Yes	Long	High	No
5	No	Short	High	Yes
6	Yes	Short	Medium	Yes
7	No	Long	High	Yes
8	No	Long	High	No
9	Yes	Short	High	Yes
10	No	Average	Medium	No
11	Yes	Average	Low	No
12	Yes	Long	Low	Yes
13	No	Average	Low	No
14	Yes	Average	High	No
15	No	Long	Low	Yes
16	Yes	Short	Medium	No
17	No	Average	Low	Yes
18	Yes	Long	High	Yes
19	No	Short	High	No
20	Yes	Short	High	No
21	No	Short	High	No
22	Yes	Long	Low	No
23	Yes	Short	Medium	No
24	No	Long	High	Yes
25	Yes	Long	Medium	Yes

This data is given in "HW06Train.csv" file.

Write R code to predict if the people with the following characteristics will visit again or not. Compute probabilities for each case. If the probability is greater than 0.5, visitor will visit again.

Visitor ID	Previously Visited	Session Length	Page Views
26	No	Short	Low
27	Yes	Short	Medium
28	Yes	Average	Medium
29	Yes	Long	High

This data is given in HW06Test.csv file.

Verify these answers using Excel and KNIME (optional).

===

The answer to this problem is as follows. Make sure your R code generate these answers.

			R		
Previously_Visited	SessionLength	PageViews	P(Will_Visit_Again=No)	P(Will_Visit_Again=Yes)	Prediction
No	Short	Low	0.6239554	0.3760446	No
Yes	Short	Medium	0.7398844	0.2601156	No
Yes	Average	Medium	0.8504983	0.1495017	No
Yes	Long	High	0.4534413	0.5465587	Yes