**CSCI 6951 Data Science and Machine Learning**

**Lab5 Rule-Based and Naïve Bayes Classifiers**

1. (5 points) Apply Naive Bayes as probabilistic learning algorithm on the dataset below and create a table with counts and probabilities following the model

|  |  |  |
| --- | --- | --- |
| **snow** | **weather** | **go skiing** |
| fresh | foggy | **no** |
| sticky | windy | **no** |
| sticky | sunny | **yes** |
| fresh | windy | **yes** |
| fresh | foggy | **yes** |
| frosted | sunny | **no** |

|  |  |  |
| --- | --- | --- |
| **snow** | **weather** | **go skiing** |
| |  |  |  | | --- | --- | --- | |  | yes | no | | fresh | 2 | 1 | | sticky | 1 | 1 | | frosted | 0 | 1 | | |  |  |  | | --- | --- | --- | |  | yes | no | | sunny | 1 | 1 | | foggy | 1 | 1 | | windy | 1 | 1 | | |  |  | | --- | --- | | yes | no | | 3 | 3 | |
| |  |  |  | | --- | --- | --- | |  | yes | no | | fresh | 2/3 | 1/3 | | sticky | 1/3 | 1/3 | | frosted | 0/3 | 1/3 | | |  |  |  | | --- | --- | --- | |  | yes | no | | sunny |  |  | | foggy | 1/3 | 1/3 | | windy | 1/3 | 1/3 | | |  |  | | --- | --- | | yes | no | | 3/6 | 3/6 | |

Classify as **yes** or **no** the following new instance "**snow=fresh and weather=sunny**"

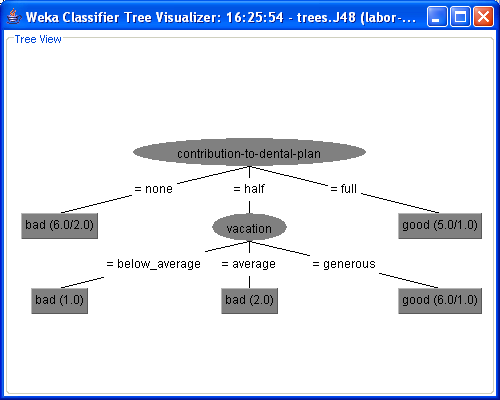
Likelihood of Yes:L=\*\*==0.11

Likelihood of No: L=\*\*==0.05

P(YES)==0.67

P(NO)==0.03

2. (10 points) Consider the following decision tree and a testing set in the table below.





1. Derive a rule set from the above decision tree.

If Contribution-to-dental-plan=”none” then class is “bad”

If Contribution-to-dental-plan=”half” then class is “bad”

If Contribution-to-dental-plan=”half” and vacation=”below” then class is “bad”

If Contribution-to-dental-plan=”half” and vacation=”average” then class is “bad”

If Contribution-to-dental-plan=”full” and vacation=”generous” then class is “good”

Contribution-to-denial-plan=”full” then class is “good”

1. Fill in the following Confusion Matrix.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Predicted Class | |
| Class = ‘good’ | Class =’bad’ |
| Actual Class | Class=’good’ | 11 | 2 |
| Class=’bad’ | 2 | 9 |

1. Compute the overall Classification Accuracy.

Acc=(7+9)/(7+9+2+2)=16/20=8/10=4/5=0.80=80%

Err=(2+2)/(7+2+2+9)=0.20=20%

1. Classify the following test instances:
   1. Education = yes, Vacation=Average, Contribution to dental plan=half

-> Class =? Bad

* 1. Contribution to dental Plan = full -> Class =?Good

3. (5 points)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Weather Condition | Driver's Condition | Traffic Violation | Seat Belt | Crash Severity |
| Good | Alcohol-impaired | Exceed speed limit | No | Major |
| Bad | Sober | None | Yes | Minor |
| Good | Sober | Disobey stop sign | Yes | Minor |
| Good | Sober | Exceed speed limit | Yes | Major |
| Bad | Sober | Disobey traffic signal | No | Major |
| Good | Alcohol-impaired | Disobey stop sign | Yes | Minor |
| Bad | Alcohol-impaired | None | Yes | Major |
| Good | Sober | Disobey traffic signal | Yes | Major |
| Good | Alcohol-impaired | None | No | Major |
| Bad | Sober | Disobey traffic signal | No | Major |
| Good | Alcohol-impaired | Exceed speed limit | Yes | Major |
| Bad | Sober | Disobey stop sign | Yes | Minor |

Using the 1R algorithm complete the table below and find out which is the best attribute to use for rule generation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Attribute** | **Rules** | **Errors** | **Total Errors** |
| **1** | Weather Condition | Good -> | 2/7 | 4/12 |
|  |  | Bad -> | 2/5 |  |
| **2** | Driver’s Condition | Alcohol-impaired -> | 1/5 | 4/12 |
|  |  | Sober -> Major | 2/5 |  |
| **3** | Traffic Violation | Exceed Speed Limit -> Major | 0/3 | 1/12 |
|  |  | None -> Major | 1/3 |  |
|  |  | Disobey stop sign -> Minor | 0/3 |  |
|  |  | Disobey traffic signal ->Major | 0/3 |  |
| **4** | Seat Belt | No -> Major | 0/4 | 4/12 |
|  |  | Yes -> Minor | 4/8 |  |

**If Traffic\_violation=”Exceed speed limit” then class is “Major”**

**If Traffic\_violation=”None” then class is “Major”**

**If Traffic\_violation=”Disobey Stop Sign” then class is “Minor”**

**If Traffic\_violation=”Disobey Traffic Sign” then class is “Major”**

**Acc=11/12**

**Err=1/12**

**Confusion Matrix**