**Problem 1:**

You are stranded on a deserted island. Mushrooms of various types grow widely all over the island, but no other food is anywhere to be found. Some of the mushrooms have been determined as poisonous and others as not (determined by your former companions’ trial and error). You are the only one remaining on the island. You have the following data to consider:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Example** | **NotHeavy** | **Smelly** | **Spotted** | **Smooth** | **Edible** |
| A | 1 | 0 | 0 | 0 | 1 |
| B | 1 | 0 | 1 | 0 | 1 |
| C | 0 | 1 | 0 | 1 | 1 |
| D | 0 | 0 | 0 | 1 | 0 |
| E | 1 | 1 | 1 | 0 | 0 |
| F | 1 | 0 | 1 | 1 | 0 |
| G | 1 | 0 | 0 | 1 | 0 |
| H | 0 | 1 | 0 | 0 | 0 |
| U | 0 | 1 | 1 | 1 | ? |
| V | 1 | 1 | 0 | 1 | ? |
| W | 1 | 1 | 0 | 0 | ? |

You know whether or not mushrooms A through H are poisonous, but you do not know about U through W.

1. Build a ID3 decision tree to classify mushrooms as poisonous or not.

34

A diagram of a network

Description automatically generated

0.003229

A diagram of a diagram

Description automatically generated

A diagram of a network

Description automatically generated

A diagram of smooth and smooth

Description automatically generated

1. Classify mushrooms U, V and W using the decision tree as poisonous or not poisonous.
2. If the mushrooms A through H that you know are not poisonous suddenly became scarce, should you consider trying U, V and W? Which one(s) and why? Or if none of them, then why not?

**Problem 2**

As of September 2012, 800 extrasolar planets have been identified in our galaxy. Supersecret surveying spaceships sent to all these planets have established whether they are habitable for humans or not, but sending a spaceship to each planet is expensive. In this problem, you will come up with decision trees to predict if a planet is habitable based only on features observable using telescopes.

1. In below table you are given the data from all 800 planets surveyed so far. The features observed by telescope are Size (“Big” or “Small”), and Orbit (“Near” or

“Far”). Each row indicates the values of the features and habitability, and how many times that set of values was observed. So, for example, there were 20 “Big” planets “Near” their star that were habitable

|  |  |  |  |
| --- | --- | --- | --- |
| **Size** | **Orbit** | **Habitable** | **Count** |
| Big | Near | Yes | 20 |
| Big | Far | Yes | 170 |
| Small | Near | Yes | 139 |
| Small | Far | Yes | 45 |
| Big | Near | No | 130 |
| Big | Far | No | 30 |
| Small | Near | No | 11 |
| Small | Far | No | 255 |

Derive and draw the decision tree learned by ID3 on this data. Make sure to clearly mark at each node what attribute you are splitting on, and which value corresponds to which branch. By each leaf node of the tree, write in the number of habitable and inhabitable planets in the training data that belong to that node

A diagram of size and size

Description automatically generated

A diagram of a space structure

Description automatically generated

1. For just 9 of the planets, a third feature, Temperature (in Kelvin degrees), has been measured, as shown in the nearby table. Redo all the steps from *part 1* on this data using all three features. For the Temperature feature, in each iteration you must maximize over all possible binary thresholding splits (such as T ≤ 250 vs. T > 250, for example).

|  |  |  |  |
| --- | --- | --- | --- |
| **Size** | **Orbit** | **Temperature** | **Habitable** |
| Big | Far | 205 | No |
| Big | Near | 205 | No |
| Big | Near | 260 | Yes |
| Big | Near | 380 | Yes |
| Small | Far | 205 | No |
| Small | Far | 260 | Yes |
| Small | Near | 260 | Yes |
| Small | Near | 380 | No |
| Small | Near | 380 | No |

According to your decision tree, would a planet with the features (Big, Near, 280) be predicted to be habitable or not habitable?

A diagram of a size

Description automatically generated

A diagram of a diagram

Description automatically generated

0.018311

A diagram of a temperature

Description automatically generated

Threshold split: T<=260 and T>260

**Problem 3:**

Given dataset about animal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name** | **Give Birth** | **Can**  **Fly** | **Live in**  **Water** | **Have Legs** | **Class** |
| human | yes | no | no | yes | mammals |
| python | no | no | no | no | non-mammals |
| salmon | no | no | yes | no | non-mammals |
| whale | yes | no | yes | no | mammals |
| frog | no | no | sometimes | yes | non-mammals |
| komodo | no | no | no | yes | non-mammals |
| bat | yes | yes | no | yes | mammals |
| pigeon | no | yes | no | yes | non-mammals |
| cat | yes | no | no | yes | mammals |
| Leopard shark | yes | no | yes | no | non-mammals |
| turtle | no | no | sometimes | yes | non-mammals |
| penguin | no | no | sometimes | yes | non-mammals |
| porcupine | yes | no | no | yes | mammals |
| eel | no | no | yes | no | non-mammals |
| salamander | no | no | sometimes | yes | non-mammals |
| Gila monster | no | no | no | yes | non-mammals |
| platypus | no | no | no | yes | mammals |
| owl | no | yes | no | yes | non-mammals |
| dolphin | yes | no | yes | no | mammals |
| eagle | no | yes | no | yes | non-mammals |

Using **Naïve Bayes Classifier** to find whether this anime is mammal or not (Using Laplacian smoothing if need)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Give Birth** | **Can**  **Fly** | **Live in**  **Water** | **Have Legs** | **Class** |
| 1 | Yes | No | Yes | No | ? |