**Question 1:**

**(i): Please provide the intermediate computations.**

See [CSCE\_633\_HW\_2\_Q1.ipynb](http://localhost:8890/notebooks/CSCE_633_HWs/CSCE_633_HW_2_Q1.ipynb) for computations.

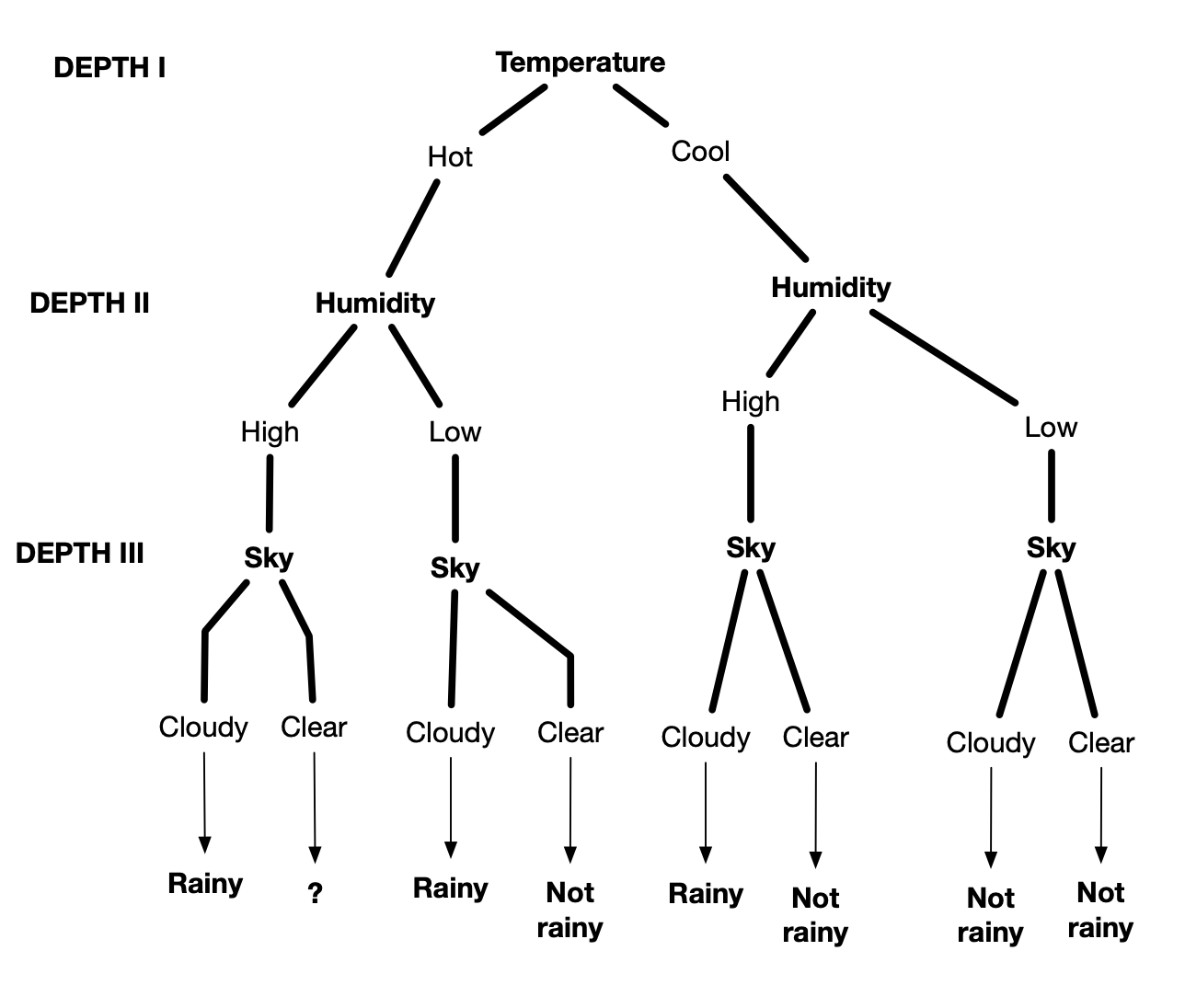
**(ii) the predictor variable/feature that you select for the split in.**

Conditional entropy for attribute temperature: 0.463012016502

Conditional entropy for attribute humidity: 0.463012016502

Conditional entropy for attribute sky: 0.53170034385

**(iii) Draw the resulting decision tree.**



**(b.i) Compute the number of samples belonging to the benign and the number of samples**

**belonging to the malignant case.**

**Count for class benign:** 444

**Count for class malignant:** 239

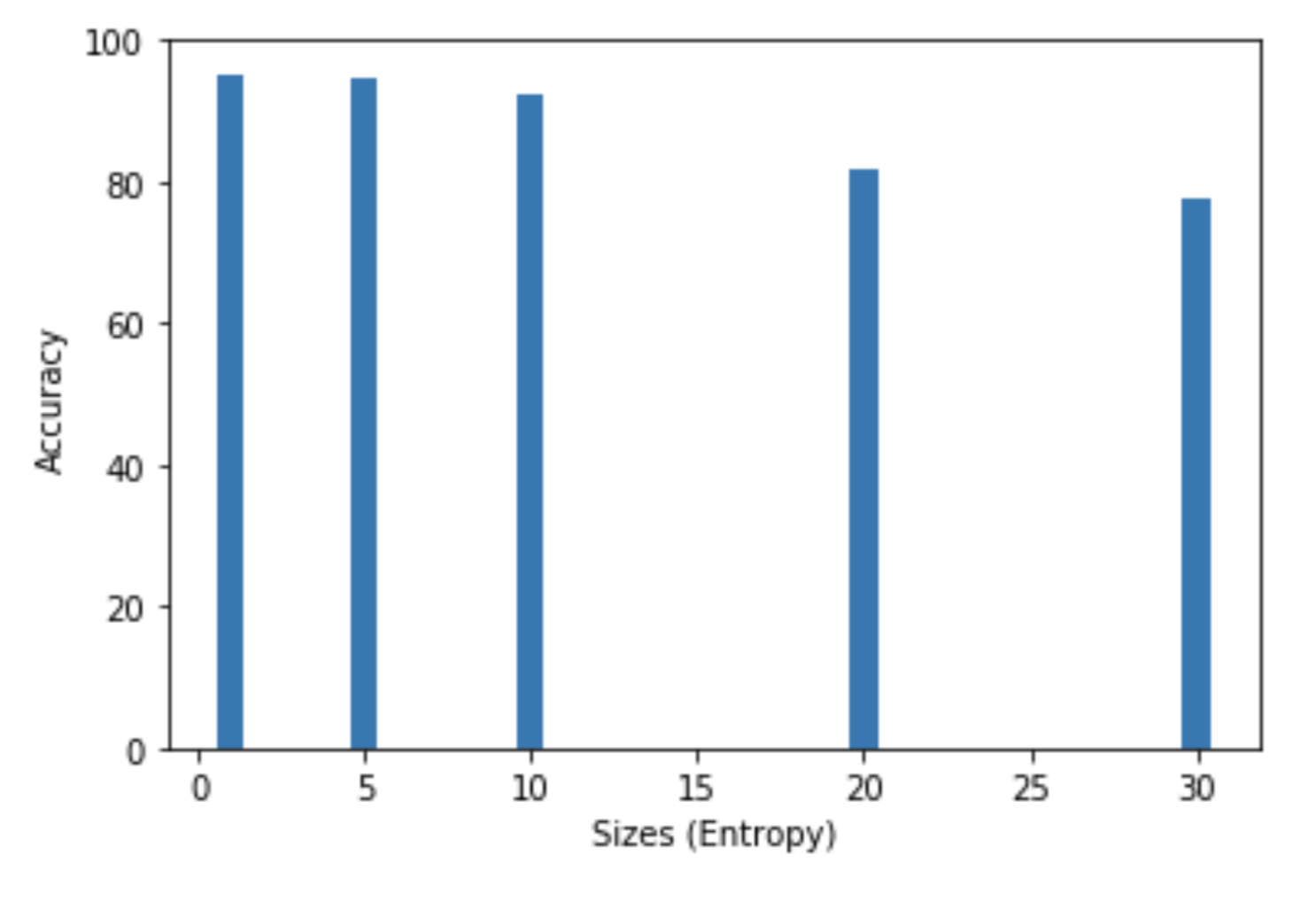
**Proportion of benign in train set:** 0.65

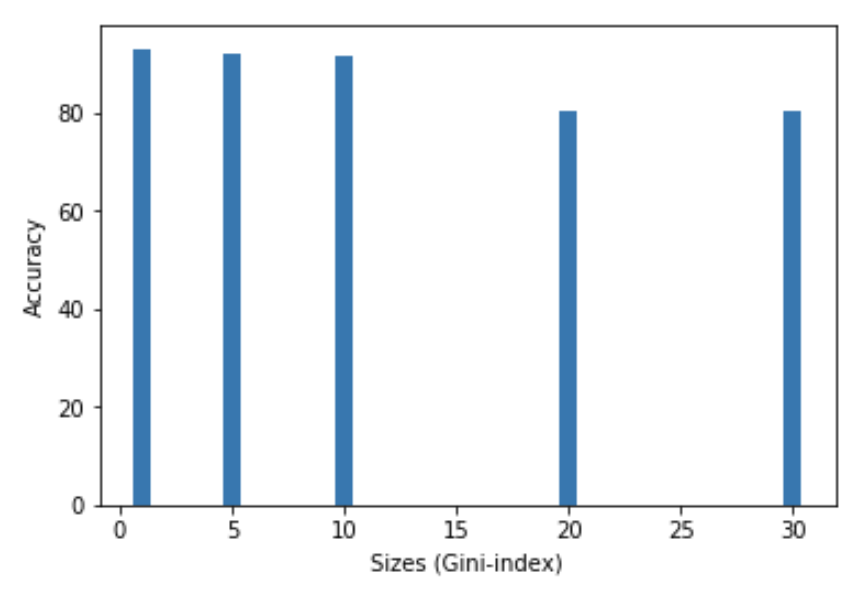
**Proportion of malignant in train set:** 0.349

**Proportion of benign in train set:** 0.649

**Proportion of malignant in train set**: 0.35

**(b.ii)** See CSCE\_633\_HW\_2\_DEC\_TREE.ipynb

****

****

|  |  |  |
| --- | --- | --- |
| **Node Size:** | **Entropy** | **Gini** |
| 1 | 95.18% | 92.98% |
| 5 | 94.74% | 92.11% |
| 10 | 92.11% | 91.67% |
| 20 | 81.58% | 80.26% |
| 30 | 77.63% | 80.26% |

**Entropy as the splitting criterion appears to perform better with respect to decision accuracy against using Gini-index.**

**Question 2:**

1. **Data pre-processing:** See implementation in[CSCE\_633\_HW\_2\_SVM.ipynb](http://localhost:8889/notebooks/CSCE_633_HW_2_SVM.ipynb)
2. **Use linear SVM in LIBSVM:** 
   1. **Best C value for accuracy:** 1

**Using best C value on test set:**

**Train time for C value 1** : 0.000626087188721 s

**Accuracy for C value 1 :** 66.66666666666666 %

1. **Use kernel SVM in LIBSVM:**
   1. **RBF:**
      1. **Best degree value for accuracy:** 5

**Using best degree value on test set:**

**Train time for degree value 5 :** 0.000859975814819 s

**Accuracy for degree value 5 :** 100.0 %

* 1. **Polynomial:**
     1. **Best Gamma value for accuracy:** 1
     2. **Using best Gamma value on test set:**

**Train time for gamma value 1 :** 0.000965118408203 s

**Accuracy for gamma value 1 :** 66.66%