David Rosenfeld
UCF ID: 1550340

Assignment 1

1) Decision Tree Basics
   a) Given A = 5 and D = 3, we calculate permutations P(5, 3), meaning that 3 are chosen from a set of 5 where order matters:

$$P(5,3) = \frac{5!}{(5-3)!} = \frac{120}{2} = 60$$

For the general case A >> D, the formula is:

$$P(A, D) = \frac{A!}{(A-D)!}$$

   b) Here is the output for the tree of depth 1:

```
{'Finished HMK':
  {
    1: {'Entropy': 0.863120568566631, 'Positives': 5, 'Negatives': 2},
    0: {'Entropy': 0.9852281360342516, 'Positives': 3, 'Negatives': 4}
  }
}
```

Here is the information gain for all attributes at depth 1:

```
{
  'Early': 0.020244207153756l9,
  'Finished HMK': 0.06105378373381032,
  'Senior': 0.011265848648557286,
  'Likes Coffee': 0.03914867190307081,
  'Liked The Last Jedi': 0.0013397424044413464
}
```

Here is the output for the tree of depth 2:

```
{'Finished HMK':
  {
  1: {
    'Entropy': 0.863120568566631,
    'Positives': 5,
    'Negatives': 2,
    'Next Split': 'Early'},
  0: {
    'Entropy': 0.9852281360342516,
    'Positives': 3,
    'Negatives': 4,
    'Next Split': 'Likes Coffee'}
  }
}
```

(See next page for information gain)

Information gain at depth 2 for Finished HMK = 1:

```
{
  'Early': 0.41379956460568024,
  'Senior': 0.2916919971380598,
  'Likes Coffee': 0.18385092540042136,
  'Liked The Last Jedi': 0.2916919971380598
}
```

Information gain at depth 2 for Finished HMK = 0:

```
{
  'Early': 0.02024420715375619,
  'Senior': 0.12808527889139454,
  'Likes Coffee': 0.46956521111470706,
  'Liked The Last Jedi': 0.02024420715375619
}
```

The code for this is in the appendix at the end of the file.

c) Here is the output when the tree is extended to a depth of 3:

```
{'Finished HMK':
  {
  1: {
    'Entropy': 0.863120568566631,
    'Positives': 5,
    'Negatives': 2,
    'Next Split': 'Early',
    'Depth 3 Splits': {
      'Early': {
        1: None,
        0: 'Senior'}
     }
  },
  0: {
    'Entropy': 0.9852281360342516,
    'Positives': 3,
    'Negatives': 4,
    'Next Split': 'Likes Coffee',
    'Depth 3 Splits': {
      'Likes Coffee': {
         0: 'Senior',
         1: None}
     }
     }
  }
}
```

Here is the information gain at depth 3:

Information gain at depth 3 for Finished HMK = 1, Early = 0:

```
{
  'Senior': 0.29650626049338447,
  'Likes Coffee': -0.014771863965748366,
  'Liked The Last Jedi': 0.29650626049338447
}
```

Information gain at depth 3 for Finished HMK = 0, Likes Coffee = 0:

```
{
  'Early': 0.43425063560155797,
  'Senior': 0.5852281360342516,
  'Liked The Last Jedi': 0.43425063560155797
}
```

The code for this is in the appendix at the end of the file.

Overall, I would choose the depth=2 tree. A depth of 1 likely does not capture enough of the complexity in the dataset, given the number of features, and a depth of 3 may cause overfitting.

d) A decision tree is realizable if there is some tree that perfectly classifies the given data set.
A tree of "no fixed depth" means that there is no set limit to the depth of the tree, so the tree can grow to arbitrary depth.
A dataset where two instances (table rows) have the same attribute (input) values but a different label (output) value will not be realizable, because no decision tree can be constructed that will label these two instances correctly. This can be proved by contradiction:
Assume two instances, x_1 and x_2, that have the same attributes, but x_1 has label *y*, and x_2 has label *z*, where $y \neq z$.
Because x_1 and x_2 have the same attributes, their feature vectors are identical. A deterministic decision tree must assign the same label to identical feature vectors. Since we know that $y \neq z$, we know that the tree does not classify one of x_1 or x_2 correctly. And since a realizable decision tree must classify all instances correctly, no such tree can exist.

2) Application of Decision Tree on Real-World Data Set

a) Here is the training accuracy for cut-off depths between 2 and 10:

```
'Depth': 2, 'Training Accuracy': 0.9444575312119404
'Depth': 3, 'Training Accuracy': 0.9448083679575788}
'Depth': 4, 'Training Accuracy': 0.945615292472547
'Depth': 5, 'Training Accuracy': 0.9492238989991129
'Depth': 6, 'Training Accuracy': 0.9497551660710796
'Depth': 7, 'Training Accuracy': 0.9520305929642197
'Depth': 8, 'Training Accuracy': 0.9526671110598778}
'Depth': 9, 'Training Accuracy': 0.9536695017617016
'Depth': 10, 'Training Accuracy': 0.9550177172556548}]
```

Based on these results, a depth of k = 10 gives the best training accuracy.

b) Here is the test accuracy for a depth of 10:

```
Test accuracy for depth=10: 0.9506525530763217
```

This equates top 95.07% accuracy in classifying the test data.

c) Based on the results from the test data, I do not see any overfitting issues. The accuracy of 95.07% with the test data was very close to the 95.50% accuracy with the training data.

The code for this is in the appendix at the end of this file.

3) Independent Events and Bayes Theorem

a) Bayes theorem states:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Event *B* can occur in two ways: with event *A* occurring, and without event *A* occurring. This means that *P(B)* is equivalent to the following:

$$P(B) = P(B|A)P(A) + P(B|\neg A)P(\neg A)$$

This gives the expression shown in the denominator of the question. So substituting this expanded expression for *P(B)* in the definition of Bayes theorem gives the proof.

b)

i) Is *X* independent of *Y*? Why or why not?

X and *Y* are independent if *P(X, Y) = P(X)P(Y)*

First, compute *P(X)*:
P(X = 0) = P(0,0,0) + P(0,1,0) + P(0,0,1) + P(0,1,1)
P(X = 0) = 0.1 + 0.2 + 0.1 + 0.175 = 0.575

P(X = 1) = P(1,0,0) + P(1,1,0) + P(1,0,1) + P(1,1,1)
P(X = 1) = 0.05 + 0.1 + 0.1 + 0.175 = 0.425

Next, compute *P(Y)*:
P(Y = 0) = P(0,0,0) + P(1,0,0) + P(0,0,1) + P(1,0,1)
P(Y=0) = 0.1 + 0.05 + 0.1 + 0.1 = 0.35

P(Y = 1) = P(0,1,0) + P(1,1,0) + P(0,1,1) + P(1,1,1)
P(Y = 1) = 0.2 + 0.1 + 0.175 + 0.175 = 0.65

Next, compute *P(X,Y)*:
P(X = 0, Y = 0) = P(0,0,0) + P(0,0,1) = 0.1 + 0.1 = 0.2
P(X = 0, Y = 1) = P(0,1,0) + P(0,1,1) = 0.2 + 0.175 = 0.375
P(X = 1, Y = 0) = P(1,0,0) + P(1,0,1) = 0.05 + 0.1 = 0.15
P(X = 1, Y = 1) = P(1,1,0) + P(1,1,1) = 0.1 + 0.175 = 0.275

(continued on next page)

Next, compare P(X)P(Y) with P(X,Y):

X	***Y***	***P(X, Y)***	***P(X)P(Y)***
0	0	0.2	0.20125
0	1	0.375	0.37375
1	0	0.15	0.14875
1	1	0.275	0.27625

Since *P(X,Y) ≠ P(X)P(Y)*, *X* and *Y* are not independent.

ii) Is *X* conditionally independent of *Y*, given Z? Why or why not?
X and *Y* are conditionally independent, given Z, if:

$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

This must be true for all values of X, Y, and Z.

Step 1: Calculate *P(Z)*:

P(Z = 0) = P(0,0,0) + P(0,1,0) + P(1,0,0) + P(1,1,0)
= 0.1 + 0.2 + 0.05 + 0.1 = 0.45
P(Z=1) = P(0,0,1) + P(0,1,1) + P(1,0,1) + P(1,1,1)
= 0.1 + 0.175 + 0.1 + 0.175 = 0.55

Step 2: Calculate *P(X|Z)*:

For X = 0:
P(X=0 | Z=0) = (P(0,0,0) + P(0,1,0)) / P(Z=0) = (0.1 + 0.2) / 0.45 = 0.6667
P(X=0 | Z=1) = (P(0,0,1) + P(0,1,1)) / P(Z=1) = (0.1 + 0.175) / 0.55 = 0.5000
For X = 1:
P(X=1 | Z=0) = (P(1,0,0) + P(1,1,0)) / P(Z=0) = (0.05 + 0.1) / 0.45 = 0.3333
P(X=1 | Z=1) = (P(1,0,1) + P(1,1,1)) / P(Z=1) = (0.1 + 0.175) / 0.55 = 0.5000

Step 3: Calculate *P(Y|Z)*

For Y = 0:
P(Y=0 | Z=0) = (P(0,0,0) + P(1,0,0)) / P(Z=0) = (0.1 + 0.05) / 0.45 = 0.3333
P(Y=0 | Z=1) = (P(0,0,1) + P(1,0,1)) / P(Z=1) = (0.1 + 0.1) / 0.55 = 0.3636
For Y = 1:
P(Y=1 | Z=0) = (P(0,1,0) + P(1,1,0)) / P(Z=0) = (0.2 + 0.1) / 0.45 = 0.6667
P(Y=1 | Z=1) = (P(0,1,1) + P(1,1,1)) / P(Z=1) = (0.175 + 0.175) / 0.55 = 0.6364

Step 4: Calculate *P(X,Y | Z)*:

Using the formula:
P(X, Y | Z) = P(X, Y, Z) / P(Z)

For Z = 0:
P(0,0 | 0) = P(0,0,0) / P(Z=0) = 0.1 / 0.45 = 0.2222
P(0,1 | 0) = P(0,1,0) / P(Z=0) = 0.2 / 0.45 = 0.4444
P(1,0 | 0) = P(1,0,0) / P(Z=0) = 0.05 / 0.45 = 0.1111
P(1,1 | 0) = P(1,1,0) / P(Z=0) = 0.1 / 0.45 = 0.2222
For Z = 1:
P(0,0 | 1) = P(0,0,1) / P(Z=1) = 0.1 / 0.55 = 0.1818
P(0,1 | 1) = P(0,1,1) / P(Z=1) = 0.175 / 0.55 = 0.3182
P(1,0 | 1) = P(1,0,1) / P(Z=1) = 0.1 / 0.55 = 0.1818
P(1,1 | 1) = P(1,1,1) / P(Z=1) = 0.175 / 0.55 = 0.3182

Step 5: Compute P(X | Z)P(Y | Z):
P(0 | 0) P(0 | 0) = (0.6667)(0.3333) = 0.2222
P(0 | 0) P(1 | 0) = (0.6667)(0.6667) = 0.4444
P(1 | 0) P(0 | 0) = (0.3333)(0.3333) = 0.1111
P(1 | 0) P(1 | 0) = (0.3333)(0.6667) = 0.2222
P(0 | 1) P(0 | 1) = (0.5000)(0.3636) = 0.1818
P(0 | 1) P(1 | 1) = (0.5000)(0.6364) = 0.3182
P(1 | 1) P(0 | 1) = (0.5000)(0.3636) = 0.1818
P(1 | 1) P(1 | 1) = (0.5000)(0.6364) = 0.3182

Given that the values obtained in Step 4 and Step 5 all match, we have proven that

$$P(X,Y|Z) = P(X|Z)P(Y|Z)$$

Therefore, *X* and *Y* are conditionally independent, given *Z*.

iii) Calculate *P(X ≠ Y | Z = 0)*

$$P(X \neq Y|Z=0) = \frac{P(X \neq Y, Z=0)}{P(Z=0)}$$

$$P(X \neq Y, Z=0) = P(0,1,0) + P(1,0,0) = 0.2 + 0.05 = 0.25$$

$$P(Z=0) = P(0,0,0) + P(1,0,0) + P(0,1,0) + P(1,1,0)$$
$$P(Z=0) = 0.1 + 0.05 + 0.2 + 0.1 = 0.45$$

$$P(X \neq Y|Z) = \frac{0.25}{0.45}$$

$$P(X \neq Y|Z) = 0.556$$

David Rosenfeld
UCF ID: 1550340

Implementing Naïve Bayes

The following results were obtained from the Naïve Bayes classifier:

```
Training Accuracy: 0.9693
Testing Accuracy: 0.9823
Training Time: 0.1093 seconds
```

The code for this implementation is in the appendix.

Appendices

The following pages contain the code for sections 1, 2, and 4.