Assignment 3 Solution Manual

Task#1

Steps to implement ID3 on Mushroom Data:

1. First of all create a Tree class.
2. Then Read data from file.
3. Separate the 70% data for training and 30% for testing.
4. Now start training on training data.
5. Count the number of positive("edible") and negative("poisonous") sample.
6. Calculate the "ENTROPY" using these counts.
7. Now calculate the counts of all existed values in each attribute.
8. Calculate the "AVERAGE ENTROPY" using all counts.
9. Now calculate the "GAIN" of each attribute by subtracting average entropy from entropy.
10. Which attribute has highest gain will become root of our tree.
11. Now divide the data into count of variety of attribute the root contain.
12. Now repeat process from (5) for the further children this will results a Decision Tree.
13. Now test the testing data on the Decision Tree to get the Accuracy.

Decision Tree:

odor

almond: edible

anise: edible

creosote: poisonous

fishy: poisonous

foul: poisonous

musty: poisonous

none: spore-print-color

black: edible

brown: edible

buff: edible

chocolate: edible

green: poisonous

orange: edible

white: habitat

grasses: edible

leaves: population

clustered: poisonous

several: edible

paths: edible

waste: edible

woods: ring-type

evanescent: poisonous

pendant: edible

yellow: edible

pungent: poisonous

spicy: poisonous

Accuracy:

Accuracy of Classifier is 100%.

Output Screen Shot:

```
C:\Users\Adeel\Desktop\AI>python task1.py
odor
    almond: edible
    anise: edible
    creosote: poisonous
    fishy: poisonous
    foul: poisonous
    musty: poisonous
    none: spore-print-color
                          black: edible
                          brown: edible
                          buff: edible
                          chocolate: edible
                          green: poisonous
                          orange: edible
                          white: habitat
                                       grasses: edible
                                       leaves: population
                                                        clustered: poisonous
                                                        several: edible
                                       paths: edible
                                       waste: edible
                                       woods: ring-type
                                                     evanescent: poisonous
                                                     pendant: edible
                          yellow: edible
    pungent: poisonous
    spicy: poisonous
Accuracy of Classifier is 100.00%
```

Task#2

Steps to implement Naïve Bayes on Mushroom Data:

1. Read data from file.
2. Separate the 70% data for training and 30% for testing.
3. Now start training on training data.
4. Now further split the training data into positive(“edible”) and negative(“poisonous”) data.
5. Now find the probabilities of both positive and negative data by dividing count of both of the one by count of total training data.
6. Now calculate the counts of all existed values in each attribute.
7. And then calculate their probabilities for positive and negative data separately and store these probabilities.
8. Now test the testing data on the Decision Tree to get the Accuracy.

Resultant data for all attributes():

cap-shape:

['b', 269, 33, 'bell', 0.09171496760995568, 0.011982570806100218]

['c', 0, 4, 'conical', 0.0, 0.0014524328249818446]

['x', 1347, 1185, 'convex', 0.4592567337197409, 0.43028322440087147]

['f', 1128, 1116, 'flat', 0.3845891578588476, 0.40522875816993464]

['k', 162, 416, 'knobbed', 0.055233549266962154, 0.15105301379811184]

['s', 27, 0, 'sunken', 0.009205591544493692, 0.0]

cap-surface:

['f', 1110, 523, 'fibrous', 0.3784520968291851, 0.18990559186637618]

['g', 0, 4, 'grooves', 0.0, 0.0014524328249818446]

['y', 1037, 1228, 'scaly', 0.35356290487555403, 0.4458968772694263]

['s', 786, 999, 'smooth', 0.26798499829526085, 0.3627450980392157]

cap-color:

['n', 889, 718, 'brown', 0.3031026252983294, 0.2607116920842411]

['b', 35, 88, 'buff', 0.011933174224343675, 0.03195352214960058]

['c', 23, 8, 'cinnamon', 0.007841800204568702, 0.002904865649963689]

['g', 721, 600, 'gray', 0.2458233890214797, 0.2178649237472767]

['r', 11, 0, 'green', 0.0037504261847937266, 0.0]

['p', 40, 62, 'pink', 0.013637913399249914, 0.022512708787218 59]

['u', 10, 0, 'purple', 0.0034094783498124785, 0.0]

['e', 425, 593, 'red', 0.14490282986703035, 0.21532316630355847]

['w', 498, 235, 'white', 0.16979202182066144, 0.08533042846768336]

['y', 281, 450, 'yellow', 0.09580634162973065, 0.16339869281045752]

bruises?:

['t', 1946, 454, 'bruises', 0.6634844868735084, 0.16485112563543936]

['f', 987, 2300, 'no', 0.33651551312649164, 0.8351488743645606]

odor:

['a', 289, 0, 'almond', 0.09853392430958063, 0.0]

['l', 287, 0, 'anise', 0.09785202863961814, 0.0]

['c', 0, 135, 'creosote', 0.0, 0.049019607843137254]

['y', 0, 398, 'fishy', 0.0, 0.14451706608569354]

['f', 0, 1535, 'foul', 0.0, 0.5573710965867829]

['m', 0, 24, 'musty', 0.0, 0.008714596949891068]

['n', 2357, 80, 'none', 0.8036140470508012, 0.029048656499636893]

['p', 0, 187, 'pungent', 0.0, 0.06790123456790123]

['s', 0, 395, 'spicy', 0.0, 0.14342774146695716]

gill-attachment:

['a', 123, 14, 'attached', 0.041936583702693486, 0.005083514887436456]

['d', 0, 0, 'descending', 0.0, 0.0]

['f', 2810, 2740, 'free', 0.9580634162973065, 0.9949164851125636]

['n', 0, 0, 'notched', 0.0, 0.0]

gill-spacing:

['c', 2109, 2679, 'close', 0.7190589839754518, 0.9727668845315904]

['w', 824, 75, 'crowded', 0.28094101602454824, 0.027233115468409588]

['d', 0, 0, 'distant', 0.0, 0.0]

gill-size:

['b', 2728, 1203, 'broad', 0.9301056938288442, 0.43681917211328974]

['n', 205, 1551, 'narrow', 0.06989430617115582, 0.5631808278867102]

gill-color:

['k', 240, 42, 'black', 0.08182748039549949, 0.015250544662309368]

['n', 660, 80, 'brown', 0.22502557108762358, 0.029048656499636893]

['b', 0, 1197, 'buff', 0.0, 0.434640522875817]

['h', 146, 389, 'chocolate', 0.04977838390726219, 0.1412490922294844]

['g', 172, 359, 'gray', 0.05864302761677463, 0.13035584604212055]

['r', 0, 15, 'green', 0.0, 0.005446623093681917 5]

['o', 32, 0, 'orange', 0.010910330719399931, 0.0]

['p', 581, 455, 'pink', 0.19809069212410502, 0.16521423384168482]

['u', 313, 35, 'purple', 0.10671667234913058, 0.012708787218591141]

['e', 63, 0, 'red', 0.021479713603818614, 0.0]

['w', 680, 167, 'white', 0.23184452778724854, 0.060639070442992014]

['y', 46, 15, 'yellow', 0.015683600409137403, 0.005446623093681917 5]

stalk-shape:

['e', 1117, 1344, 'enlarging', 0.38083873167405385, 0.4880174291938998]

['t', 1816, 1410, 'tapering', 0.6191612683259461, 0.5119825708061002]

stalk-root:

['b', 1354, 1320, 'bulbous', 0.46164336856460964, 0.4793028322440087]

['c', 358, 29, 'club', 0.12205932492328674, 0.010530137981118372]

['u', 0, 0, 'cup', 0.0, 0.0]

['e', 601, 187, 'equal', 0.20490964882372997, 0.06790123456790123]

['z', 0, 0, 'rhizomorphs', 0.0, 0.0]

['r', 144, 0, 'rooted', 0.049096488237299694, 0.0]

['?', 476, 1218, 'missing', 0.162291169451074, 0.4422657952069717]

stalk-surface-above-ring:

['f', 270, 109, 'fibrous', 0.09205591544493692, 0.03957879448075526]

['y', 12, 5, 'scaly', 0.004091374019774974, 0.0018155410312273058]

['k', 107, 1563, 'silky', 0.036481418342993524, 0.5675381263616558]

['s', 2544, 1077, 'smooth', 0.8673712921922946, 0.39106753812636164]

stalk-surface-below-ring:

['f', 308, 106, 'fibrous', 0.10501193317422435, 0.03848946986201888]

['y', 156, 50, 'scaly', 0.05318786225707467, 0.01815541031227306]

['k', 98, 1516, 'silky', 0.03341288782816229, 0.5504720406681191]

['s', 2371, 1082, 'smooth', 0.8083873167405387, 0.39288307915758897]

stalk-color-above-ring:

['n', 12, 303, 'brown', 0.004091374019774974, 0.11002178649237472]

['', 0, 0, 'buff', 0.0, 0.0]

['c', 0, 24, 'cinnamon', 0.0, 0.008714596949891068]

['g', 413, 0, 'gray', 0.14081145584725538, 0.0]

['o', 123, 0, 'orange', 0.041936583702693486, 0.0]

['p', 398, 913, 'pink', 0.13569723832253666, 0.33151779230210604]

['e', 68, 0, 'red', 0.023184452778724857, 0.0]

['w', 1919, 1198, 'white', 0.6542788953290146, 0.43500363108206247]

['y', 0, 5, 'yellow', 0.0, 0.0018155410312273058]

stalk-color-below-ring:

['n', 41, 314, 'brown', 0.013978861234231163, 0.1140159767610748]

['', 0, 0, 'buff', 0.0, 0.0]

['c', 0, 24, 'cinnamon', 0.0, 0.008714596949891068]

['g', 414, 0, 'gray', 0.1411524036822366, 0.0]

['o', 123, 0, 'orange', 0.041936583702693486, 0.0]

['p', 409, 911, 'pink', 0.13944766450733037, 0.3307915758896151]

['e', 63, 0, 'red', 0.021479713603818614, 0.0]

['w', 1883, 1189, 'white', 0.6420047732696897, 0.43173565722585333]

['y', 0, 15, 'yellow', 0.0, 0.0054466230936819175]

veil-type:

['p', 2933, 2754, 'partial', 1.0, 1.0]

['u', 0, 0, 'universal', 0.0, 0.0]

veil-color:

['n', 58, 0, 'brown', 0.019774974428912375, 0.0]

['o', 65, 0, 'orange', 0.02216160927378111, 0.0]

['w', 2810, 2749, 'white', 0.9580634162973065, 0.9981844589687727]

['y', 0, 5, 'yellow', 0.0, 0.0018155410312273058]

ring-number:

['n', 0, 24, 'none', 0.0, 0.008714596949891068]

['o', 2572, 2682, 'one', 0.8769178315717695, 0.9738562091503268]

['t', 361, 48, 'two', 0.12308216842823048, 0.017429193899782137]

ring-type:

['c', 0, 0, 'cobwebby', 0.0, 0.0]

['e', 686, 1223, 'evanescent', 0.23389021479713604, 0.44408133623819 9]

['f', 27, 0, 'flaring', 0.009205591544493692, 0.0]

['l', 0, 918, 'large', 0.0, 0.3333333333333333]

['n', 0, 24, 'none', 0.0, 0.008714596949891068]

['p', 2220, 589, 'pendant', 0.7569041936583703, 0.2138707334785766]

['s', 0, 0, 'sheathing', 0.0, 0.0]

['z', 0, 0, 'zone', 0.0, 0.0]

spore-print-color:

['k', 1169, 164, 'black', 0.39856801909307876, 0.05954974582425563]

['n', 1217, 158, 'brown', 0.41493351517217864, 0.05737109658678286]

['b', 31, 0, 'buff', 0.010569382884418685, 0.0]

['h', 27, 1131, 'chocolate', 0.009205591544493692, 0.4106753812636166]

['r', 0, 48, 'green', 0.0, 0.017429193899782137]

['o', 29, 0, 'orange', 0.009887487214456188, 0.0]

['u', 38, 0, 'purple', 0.012956017729287419, 0.0]

['w', 390, 1253, 'white', 0.13296965564268667, 0.4549745824255628]

['y', 32, 0, 'yellow', 0.010910330719399931, 0.0]

population:

['a', 269, 0, 'abundant', 0.09171496760995568, 0.0]

['c', 190, 35, 'clustered', 0.0647800886464371, 0.012708787218591141]

['n', 265, 0, 'numerous', 0.09035117627003068, 0.0]

['s', 620, 268, 'scattered', 0.21138765768837367, 0.09731299927378359]

['v', 846, 1995, 'several', 0.2884418683941357, 0.724400871459695]

['y', 743, 456, 'solitary', 0.2533242413910672, 0.1655773420479303]

habitat:

['g', 975, 521, 'grasses', 0.3324241391067167, 0.18917937545388525]

['l', 152, 420, 'leaves', 0.051824070917149676, 0.15250544662309368]

['m', 175, 25, 'meadows', 0.059665871121718374, 0.00907770515613653]

['p', 103, 690, 'paths', 0.03511762700306853, 0.25054466230936817]

['u', 75, 200, 'urban', 0.02557108762359359, 0.07262164124909223]

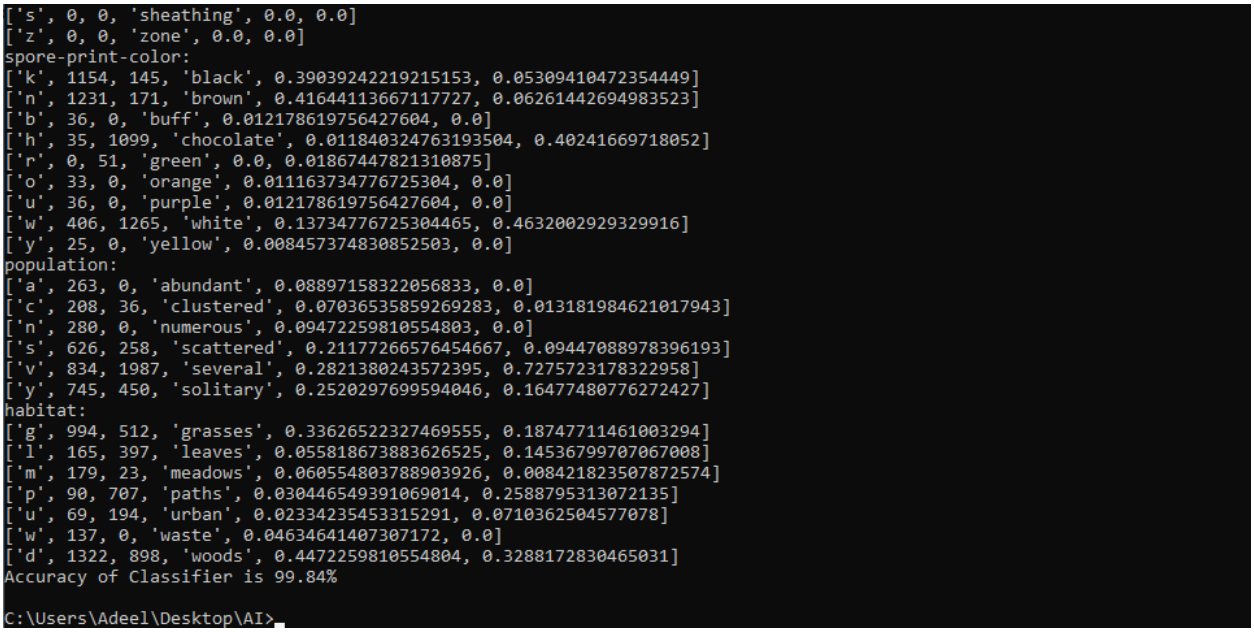
['w', 131, 0, 'waste', 0.044664166382543474, 0.0]

['d', 1322, 898, 'woods', 0.4507330378452097, 0.3260711692084241]

Accuracy:

Accuracy of Classifier is 99.6-99.9% (Variable due to random splitting).

Output Screen Shot:



Comparison Between Both of them

1. Naïve Bayes implementation is easier than ID3 implementation.
2. Accuracy of ID3 Classifier is more efficient than Naïve Bayes Classifier according to the result of Mushroom Data Testing.