Assignment 5

Rishab Katta
Introduction to Big Data

April 27, 2019

Results

```
    Initial dictionary made by grouping values for all non target attr
    for a given target attr:
{'alt': {'F': {'T': [2, 5, 7], 'F': [6, 8, 10]},'T': {'F': [1, 4, 9], 'T': [0, 3, 11]}},

'bar': {'T': {'F': [6, 8, 9], 'T': [2, 5, 11]},'F': {'T': [0, 3, 7], 'F': [1, 4, 10]}},

'fri': {'T': {'F': [4, 8, 9], 'T': [3, 11]},'F': {'T': [0, 2, 5, 7], 'F': [1, 6, 10]}},

'hun': {'F': {'F': [4, 6, 8, 10], 'T': [2]},'T': {'T': [0, 3, 5, 7, 11], 'F': [1, 9]}},

'pat': {'Full': {'F': [1, 4, 8, 9], 'T': [3, 11]},'Some': {'T': [0, 2, 5, 7]},
'None': {'F': [6, 10]}},

'price': {'$$$': {'F': [4, 9], 'T': [0]}, '$$': {'T': [5, 7]},
'$': {'F': [1, 6, 8, 10], 'T': [2, 3, 11]}},

'rain': {'F': {'F': [1, 4, 9, 10], 'T': [0, 2, 3, 11]},'T': {'F': [6, 8], 'T': [5, 7]}},

'res': {'T': {'T': [0, 5, 7], 'F': [4, 9]},'F': {'T': [2, 3, 11], 'F': [1, 6, 8, 10]}},

'type': {'Thai': {'T': [3, 7], 'F': [1, 10]}, 'Burger': {'F': [6, 8], 'T': [2, 11]},
'French': {'F': [4], 'T': [0]}, 'Italian': {'T': [5], 'F': [9]}},

'est': {'>60': {'F': [4, 8]}, '0-10': {'F': [6, 10], 'T': [0, 2, 5, 7]},
'10-30': {'T': [3], 'F': [9]}, '30-60': {'T': [11], 'F': [1]}}}
```

```
Initial Entropy calculated for all non target attr:
{'alt': 0.6020599913279623,
'bar': 0.6020599913279623,
'fri': 0.5888684747570518,
'hun': 0.47714719228675473,
'pat': 0.27643459094367495,
'price': 0.5730178124620979,
'rain': 0.6020599913279623,
'res': 0.5888684747570518,
'type': 1.2041199826559248,
'est': 0.8784945822716372  }

Attributes Decision Tree split on, in order: ['pat', 'alt', 'price', 'fri']


--- 0.5107934474945068 seconds for program to run ---
```

Final Decision Tree on next page

Final Decision Tree