Report

Decision Tree Classification for Restaurant wait Example.

Procedure to Compile and Run

This solution can be used for any type of variables, not only for categorical but also for numerical and dichotomous variables.

The solution was programmed using Python programming language. The program includes two python files as follows.

1. **DecisionTree.py** – include classes Node and Decision tree with all the necessary functionalities.
2. **main.ipynb** – include the main program which preprocess data, train a model, test the model and do predictions.

To compile and run the program the main program called "main.ipynb" should be executed using any IDE that supports interactive python notebooks (Google-colab, Jupyter Lab). The program will provide a non-GUI representation of the decision tree.

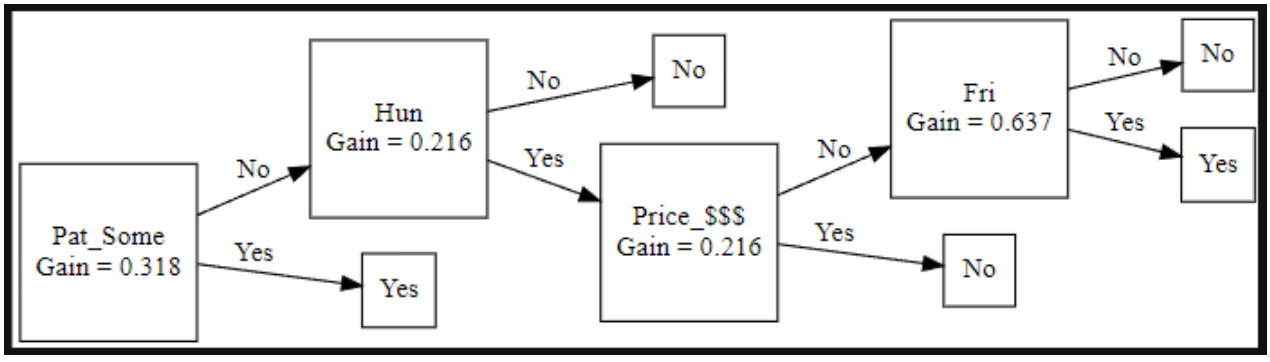
When successfully executed, the program will present the following results,

- Graphical representation of the Decision tree.
- Information gain results for each split
- Test results and accuracy
- Prediction of unseen data

Results

Graphical Representation of the Decision Tree

The solution is greedy where in case the best/highest information gain value is same in several features when deciding the best split, the **first feature** will be selected for the split. As I have implemented random selection when calculating the information gain, the tree will look different in different attempts, but the results will always be same.



Information Gain calculation results

Depth 0 – First split

```
Depth: 0 - Split: 1
        Feature: Type_Thai
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Est_0-10
                Threshold: 0.0 Gain: 0.057
                Threshold: 1.0 Gain: 0
        Feature: Type_Burger
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Bar
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Pat_Full
                Threshold: 0.0 Gain: 0.057
                Threshold: 1.0 Gain: 0
        Feature: Est_>60
                Threshold: 0.0 Gain: 0.132
                Threshold: 1.0 Gain: 0
        Feature: Est_10-30
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Price_$
                Threshold: 0.0 Gain: 0.014
                Threshold: 1.0 Gain: 0
        Feature: Alt
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Price_$$$
                Threshold: 0.0 Gain: 0.019
                Threshold: 1.0 Gain: 0
        Feature: Fri
                Threshold: 0.0 Gain: 0.014
                Threshold: 1.0 Gain: 0
        Feature: Hun
                Threshold: 0.0 Gain: 0.136
                Threshold: 1.0 Gain: 0
        Feature: Type_Italian
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Pat_Some
                Threshold: 0.0 Gain: 0.318
                Threshold: 1.0 Gain: 0
```

```
            Feature: Price_$$
                    Threshold: 0.0 Gain: 0.132
                    Threshold: 1.0 Gain: 0
            Feature: Est_30-60
                    Threshold: 0.0 Gain: 0.0
                    Threshold: 1.0 Gain: 0
            Feature: Res
                    Threshold: 0.0 Gain: 0.014
                    Threshold: 1.0 Gain: 0
            Feature: Type_French
                    Threshold: 0.0 Gain: 0.0
                    Threshold: 1.0 Gain: 0
            Feature: Pat_None
                    Threshold: 0.0 Gain: 0.132
                    Threshold: 1.0 Gain: 0
            Feature: Rain
                    Threshold: 0.0 Gain: 0.0
                    Threshold: 1.0 Gain: 0
```

Depth 1 – Second split

```
Depth: 1 - Split: 2
        Feature: Price_$$
                Threshold: 0.0 Gain: 0
        Feature: Hun
                Threshold: 0.0 Gain: 0.216
                Threshold: 1.0 Gain: 0
        Feature: Price_$$$
                Threshold: 0.0 Gain: 0.085
                Threshold: 1.0 Gain: 0
        Feature: Est_30-60
                Threshold: 0.0 Gain: 0.051
                Threshold: 1.0 Gain: 0
        Feature: Fri
                Threshold: 0.0 Gain: 0.142
                Threshold: 1.0 Gain: 0
        Feature: Pat_Full
                Threshold: 0.0 Gain: 0.085
                Threshold: 1.0 Gain: 0
        Feature: Est_10-30
                Threshold: 0.0 Gain: 0.051
                Threshold: 1.0 Gain: 0
        Feature: Type_Italian
                Threshold: 0.0 Gain: 0.039
                Threshold: 1.0 Gain: 0
        Feature: Res
                Threshold: 0.0 Gain: 0.085
                Threshold: 1.0 Gain: 0
        Feature: Alt
                Threshold: 0.0 Gain: 0.142
                Threshold: 1.0 Gain: 0
        Feature: Type_Thai
                Threshold: 0.0 Gain: 0.011
                Threshold: 1.0 Gain: 0
        Feature: Est_0-10
                Threshold: 0.0 Gain: 0.085
                Threshold: 1.0 Gain: 0
        Feature: Pat_Some
                Threshold: 0.0 Gain: 0
        Feature: Pat_None
                Threshold: 0.0 Gain: 0.085
                Threshold: 1.0 Gain: 0
        Feature: Bar
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
```

```
          Feature: Est_>60
                  Threshold: 0.0 Gain: 0.085
                  Threshold: 1.0 Gain: 0
          Feature: Rain
                  Threshold: 0.0 Gain: 0.085
                  Threshold: 1.0 Gain: 0
          Feature: Type_French
                  Threshold: 0.0 Gain: 0.039
                  Threshold: 1.0 Gain: 0
          Feature: Price_$
                  Threshold: 0.0 Gain: 0.085
                  Threshold: 1.0 Gain: 0
          Feature: Type_Burger
                  Threshold: 0.0 Gain: 0.011
                  Threshold: 1.0 Gain: 0
```

Depth 2 – Third split

```
Depth: 2 - Split: 3
        Feature: Bar
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Pat_None
                Threshold: 0.0 Gain: 0
        Feature: Pat_Some
                Threshold: 0.0 Gain: 0
        Feature: Price_$$$
                Threshold: 0.0 Gain: 0.216
                Threshold: 1.0 Gain: 0
        Feature: Est_10-30
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Type_Italian
                Threshold: 0.0 Gain: 0.216
                Threshold: 1.0 Gain: 0
        Feature: Pat_Full
                Threshold: 1.0 Gain: 0
        Feature: Rain
                Threshold: 0.0 Gain: 0
        Feature: Est_30-60
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Type_Burger
                Threshold: 0.0 Gain: 0.216
                Threshold: 1.0 Gain: 0
        Feature: Hun
                Threshold: 1.0 Gain: 0
        Feature: Price_$$
                Threshold: 0.0 Gain: 0
        Feature: Type_French
                Threshold: 0.0 Gain: 0
        Feature: Est_0-10
                Threshold: 0.0 Gain: 0
        Feature: Res
                Threshold: 0.0 Gain: 0.216
                Threshold: 1.0 Gain: 0
        Feature: Price_$
                Threshold: 0.0 Gain: 0.216
                Threshold: 1.0 Gain: 0
        Feature: Alt
                Threshold: 1.0 Gain: 0
        Feature: Type_Thai
                Threshold: 0.0 Gain: 0.0
                Threshold: 1.0 Gain: 0
        Feature: Fri
                Threshold: 0.0 Gain: 0.216
                Threshold: 1.0 Gain: 0
        Feature: Est_>60
                Threshold: 0.0 Gain: 0
```

Depth 3 – Fourth split

```
Depth: 3 - Split: 4
        Feature: Price_$$
                Threshold: 0.0 Gain: 0
        Feature: Res
                Threshold: 0.0 Gain: 0
        Feature: Hun
                Threshold: 1.0 Gain: 0
        Feature: Pat_None
                Threshold: 0.0 Gain: 0
        Feature: Fri
                Threshold: 0.0 Gain: 0.637
                Threshold: 1.0 Gain: 0
        Feature: Type_Italian
                Threshold: 0.0 Gain: 0
        Feature: Type_Burger
                Threshold: 0.0 Gain: 0.174
                Threshold: 1.0 Gain: 0
        Feature: Est_30-60
                Threshold: 0.0 Gain: 0.174
                Threshold: 1.0 Gain: 0
        Feature: Est_0-10
                Threshold: 0.0 Gain: 0
        Feature: Type_French
                Threshold: 0.0 Gain: 0
        Feature: Bar
                Threshold: 0.0 Gain: 0.174
                Threshold: 1.0 Gain: 0
        Feature: Est_10-30
                Threshold: 0.0 Gain: 0.174
                Threshold: 1.0 Gain: 0
        Feature: Pat_Some
                Threshold: 0.0 Gain: 0
        Feature: Alt
                Threshold: 1.0 Gain: 0
        Feature: Price_$
                Threshold: 1.0 Gain: 0
        Feature: Rain
                Threshold: 0.0 Gain: 0
        Feature: Type_Thai
                Threshold: 0.0 Gain: 0.174
                Threshold: 1.0 Gain: 0
        Feature: Price_$$$
                Threshold: 0.0 Gain: 0
        Feature: Pat_Full
                Threshold: 1.0 Gain: 0
        Feature: Est_>60
                Threshold: 0.0 Gain: 0
```

Test results

Accuracy = 100%

	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	Output
0	No	No	No	No	None	$$$	No	No	Thai	0-10	No
1	No	No	Yes	Yes	Full	$	Yes	No	Burger	10-30	Yes
2	Yes	No	No	No	Full	$$	Yes	No	French	0-10	No

```
Test Results:

Example 0 - > No
Example 1 - > Yes
Example 2 - > No
```

```
Accuracy of testing:
1.0
```

Prediction results

	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est
0	No	No	No	No	None	$$$	No	No	Thai	0-10
1	No	No	No	No	None	$$$	Yes	No	Burger	10-30
2	No	No	No	No	None	$	Yes	No	French	0-10

```
Predictions:

Example 0 - > No
Example 1 - > No
Example 2 - > No
```

References

1. https://www.youtube.com/watch?v=LDRbO9a6XPU
2. https://mljar.com/blog/visualize-decision-tree/