CSE 4/574 Assignment2

Due date: 4th March 11:59PM (Submit on Brightspace)

Implement a basic decision tree algorithm to classify instances of the Iris dataset into their respective classes. (50 points)

Data Loading: Load the Iris dataset using following python code:

```
from sklearn.datasets import load_iris
iris = load_iris()
```

Data Preprocessing: Perform any necessary preprocessing steps, such as handling missing values or encoding categorical variables. Ensure that the data is ready for training. (10 points)

Decision Tree Implementation: Write functions to (20 points)

- Calculate entropy or Gini impurity.
- Find the best split point for a given feature.
- Recursively build the decision tree.
- Make predictions using the trained tree.

Tree Training: (5 points)

- Split the dataset into training and testing sets.
- Train your decision tree on the training set using the implemented algorithm.

Model Evaluation: (10 points)

- Evaluate the accuracy of your decision tree model on the testing set.

Documentation and Analysis: (5 points)

- Provide clear and concise documentation for your code, explaining each function and its purpose.
- Discuss any design choices you made during the implementation.
- Reflect on the performance of your decision tree model. What are its strengths and limitations?

Submission Guidelines:

1. Submit a Python script or a Jupyter Notebook containing your code.

2. Include comments in your code to explain key steps and decisions.

The goal is not necessarily to create a highly optimized or feature-rich decision tree but to demonstrate a fundamental understanding of how the algorithm works.