Lab Assignment 4: Classification Decision Tree using the ID3 Algorithm

Duration: 2 hours **Date:** 19-03-2024

Titanic Dataset: Titanic-Dataset from roll no. **msd2024001 to msd2024017**

Objective:

To understand the implementation of the ID3 (Iterative Dichotomiser 3) algorithm for building a classification decision tree and evaluate its performance on a dataset.

Task Instructions:

- **Dataset Selection and Preprocessing**
  - Load the dataset and perform basic data preprocessing, including handling missing values.
  - Split the dataset into training and testing sets (78:22 split).
- **Implementing ID3**
  - Implement the ID3 algorithm from a Python library like scikit-learn to construct a decision tree classifier.
  - Key steps:
    - Calculate entropy for the target class.
    - Calculate information gain for each feature.
      - Select the feature with the highest information gain for splitting.
      - Recursively build the tree by selecting the best features at each node.
- **Training the Decision Tree**
  - Train the decision tree classifier using the training data.
- **Testing and Evaluation**
  - Use the trained decision tree to classify the test data.
  - Evaluate the model performance using metrics such as accuracy, precision, recall, and F1-score.
  - Visualize the decision tree (if possible) using a library matplotlib.
- **Analysis and Conclusion**
  - Discuss the results, including any observations from the decision tree structure.
  - Reflect on the performance of the ID3 algorithm for the chosen dataset.