**PGP in AI/ML**

**Answers (Classification - Assignment 4)**

1. For the dataset -2 (liver disease dataset), implement Decision Tree classifier using Python. **[10 M**]

Please follow the following steps.

1. Implemented in Python code
2. Implemented in Python code.
3. Printed as part of Python code.
4. Minimising the number of leaf nodes and depth of the tree resulted in reducing the complexity of the tree and over-fitting of the model.

max\_leaf\_nodes = 5,

min\_samples\_leaf = 10,

max\_depth= 10

Measures are printed as part of the Python implementation.

1. Summarise effect of pruning on accuracy

In most cases, Pruning helps in increasing the accuracy.

Pruning the tree helps in reducing the complexity of the Tree. And eventually improving generalisation error. Pruning helps in reducing the complexity by,

1. Replacing the node in a tree with a class label of the majority of the samples. This technique is called “Subtree replacement”.

2. Replacing an entire subtree with a branch involving the required attributes. This technique is called “Subtree raising”.

Both these techniques reduces the number of branches, depth and the number of leaf nodes in the Tree. And such tree has less impact due to,

1. Over-fitting caused due to large number of features in the data, with each feature having high variance in the sample.
2. Over-fitting caused due to presence of Noise

Overall, improving the testing error and making the model more general.