CS60050 MACHINE LEARNING
ASSIGNMENT 01 - DECISION TREE LEARNING

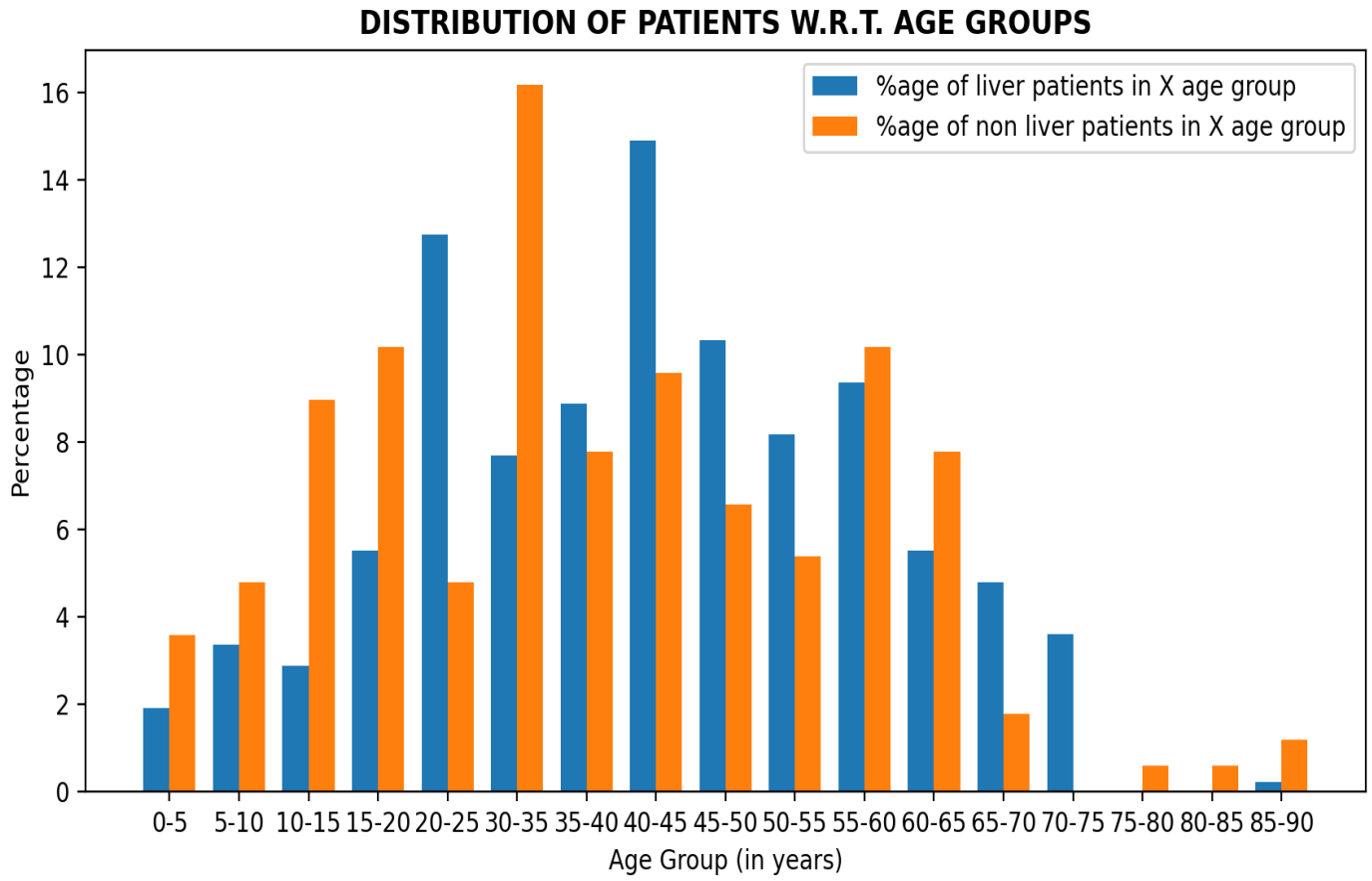
Group 23
Amrta Chaurasia (19EE10004)
Nakul Aggarwal (19CS10044)

PREFACE

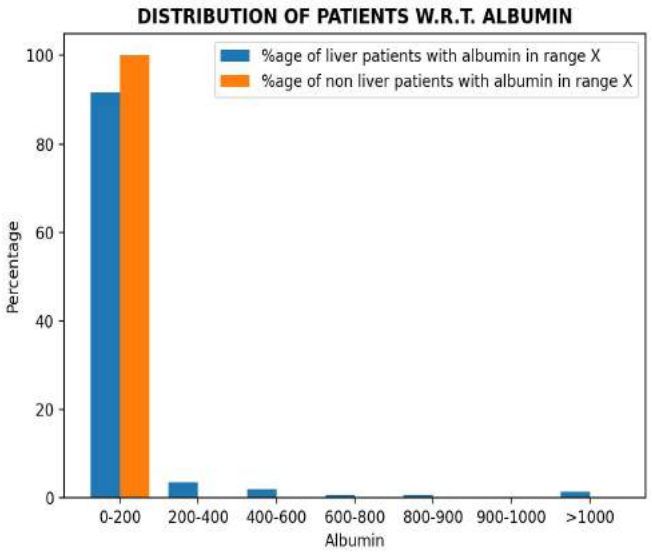
➢ **Dataset used: Indian Liver Patient Dataset (ILPD) - Data about Liver Patients in India**

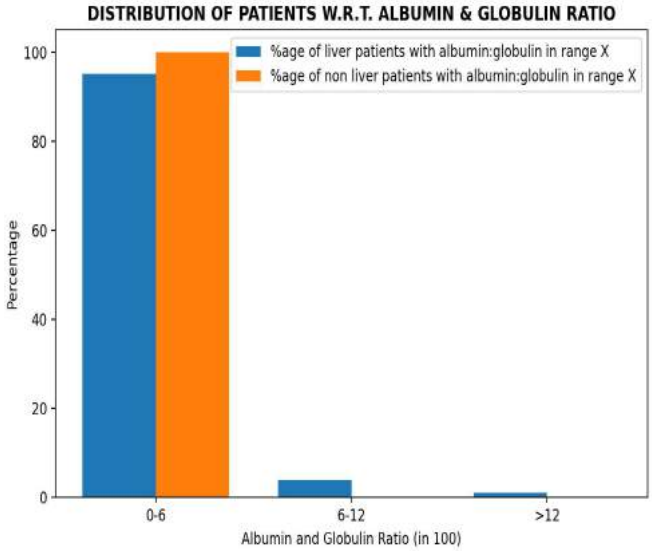
UNDERSTANDING ILPD DATASET

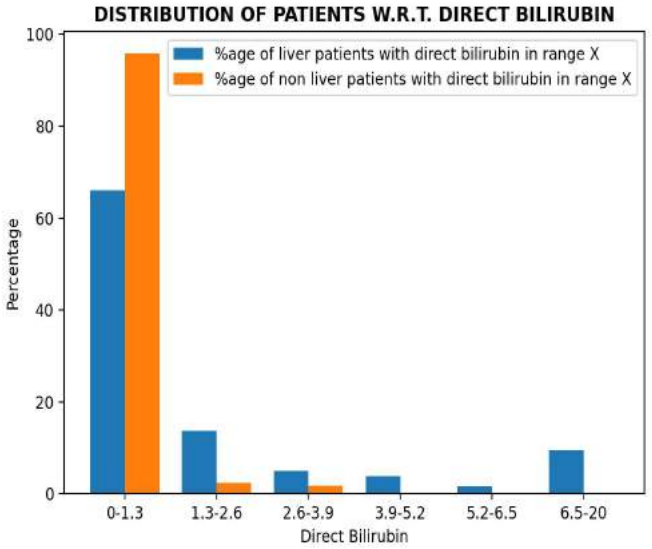
In data science and machine learning, visualization of datasets becomes of paramount importance because it gives us a clear idea of what the information means by giving it a visual context through maps or graphs. This makes data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns and outliers within large datasets. Accordingly, it helps the engineers to design their project pipeline and choose re-sampling techniques, if any, to rework their dataset.

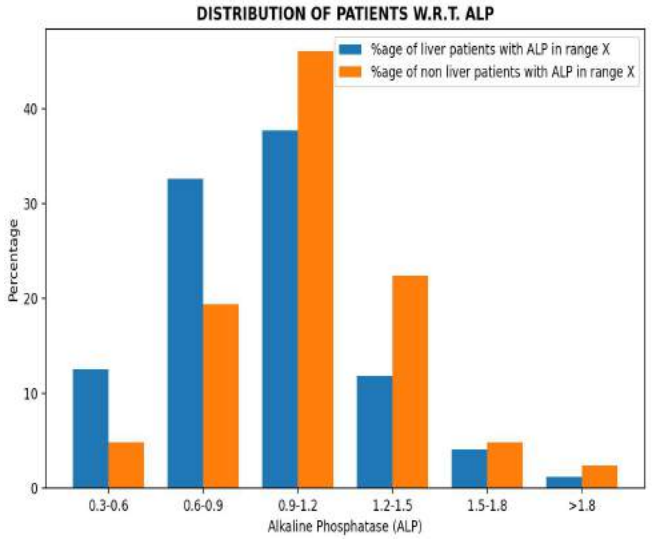


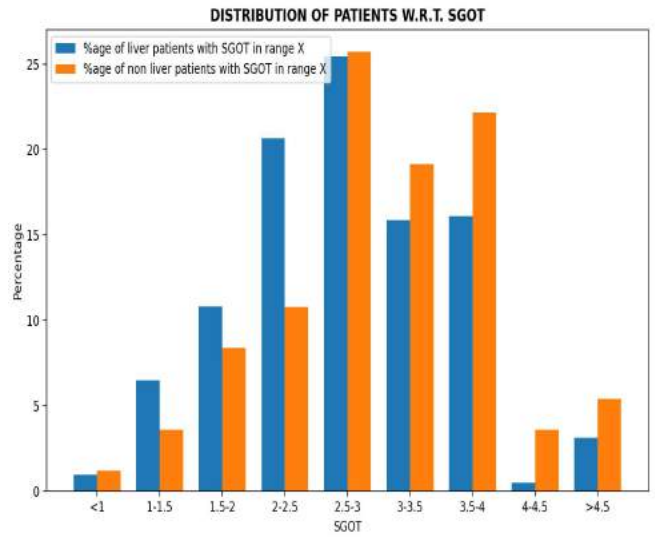
• Distribution of patients and non-patients seems to be relatively more sensitive to their age, amount of *ALP*, *SGOT* and *SGPT* than to other attributes.
• Majority of the liver patients are in the age group of *20-25* and *40-45*. Majority of the liver patients have *ALP*, *SGOT* and *SGPT* in the range of *0.6-1.2*, *2-3* and *5.4-7.2* respectively.

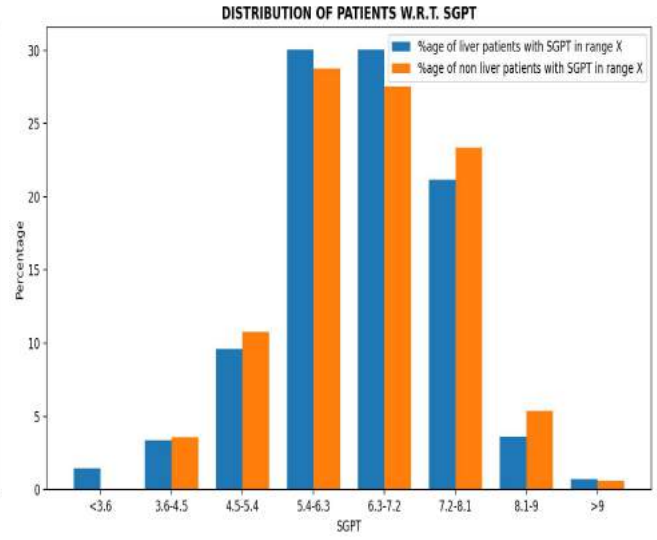




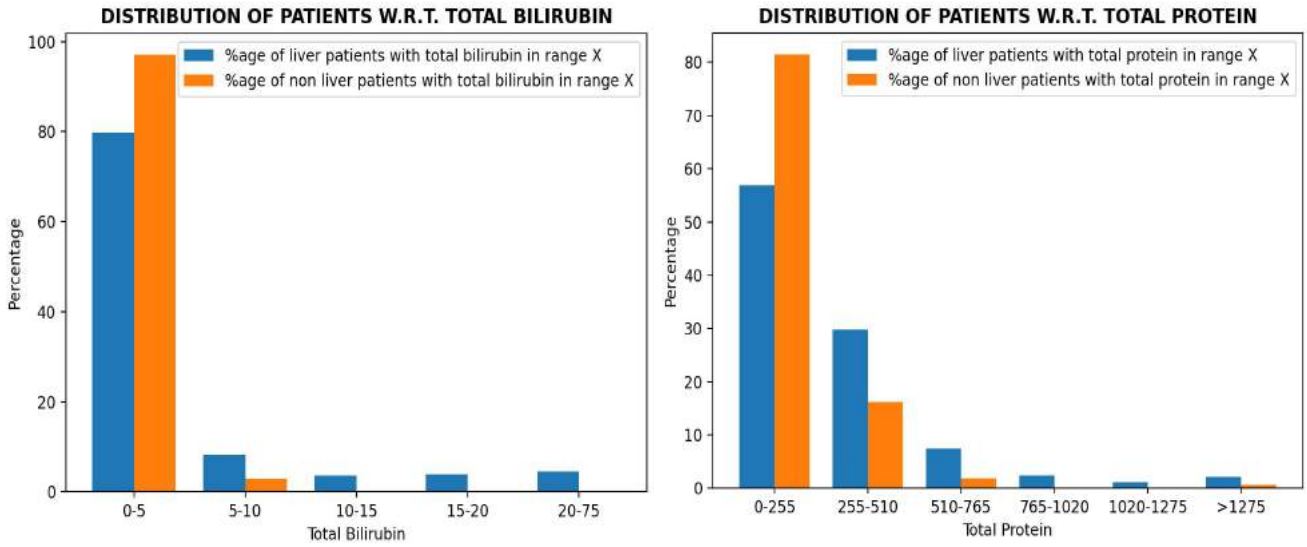


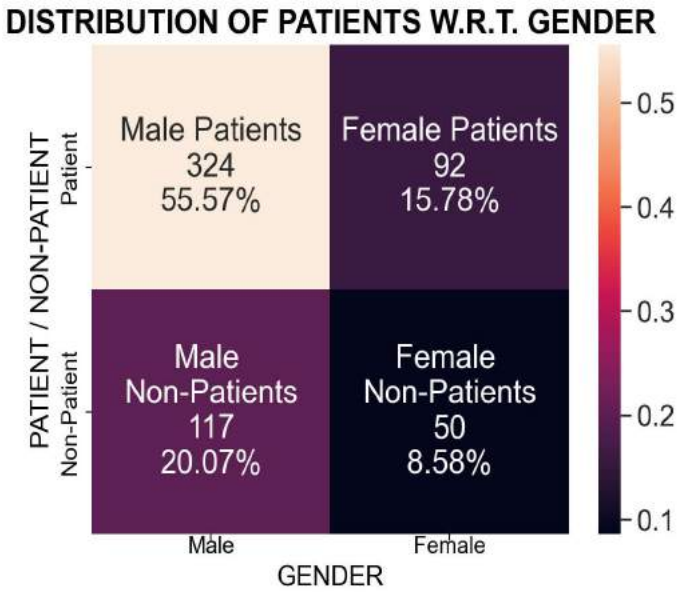






• Liver patients are much more probable to have high albumin, albumin-globulin ratio, direct bilirubin, total bilirubin and total protein than the non-liver patients. Very rarely does any non-liver patient have these values high. Despite that, the majority of the liver patients have these parameters in the low range like the non-liver patients.





Similarly, numerous other important inferences can be deduced from the visualized data very easily. This also helps in getting an intuition about the diversity of the dataset, both with respect to the target values and the attributes values. For example, the heatmap shows that the data collected for females is far less than for males. Also, the data has far more samples for patients than for non-patients.

REWORKING ILPD DATASET

The ILPD dataset has two classes - patient (positive) and non-patient (negative). Though in an ideal scenario it is always better to have a balanced distribution of examples among different classes, here more than *71%* of the examples are positive! This is a huge imbalance considering that the size of the dataset is also not that large (only *583*). So the dataset has *416* positive examples and *167* negative examples. Certainly, we cannot use the ILPD dataset in its original form to train our decision trees. The consequences of an extremely imbalanced dataset like this one include --

• Insufficient evidence of the minority class to learn about their favorable criteria.

• Considerably high performances of naive untrained classifiers, comparable to the trained classifiers. For example, if we build a decision tree with a single node (zero depth) using the ILPD dataset, it will assign an average test accuracy of approximately *71%* to randomly selected test sets by always predicting a positive class label!

Therefore, at least for this dataset, it is imperative that we *rework* the dataset before using it for training our decision trees. In this project, we use ***Synthetic Minority Oversampling Technique (SMOTE)*** to generate new synthetic points from the minority class to increase its cardinality. In the implementation, we have used ***imblearn.over_sampling.SMOTE*** class to perform over sampling using SMOTE. We resample ILPD dataset examples to a synthetic balanced dataset. Then we choose about *125* synthetic negative examples from this resampled dataset and add them to the original ILPD dataset examples. Now in the *reworked-ILPD dataset*, we have *416* positive (all original) and *292* negative (original + synthetic) that are better balanced.

	POSITIVE		NEGATIVE	
	COUNT	PROPORTION	COUNT	PROPORTION
BEFORE	416	71.72 %	167	28.28 %
AFTER	416	58.76 %	292	41.24 %

PART 01

TASK SUMMARY

Decision tree classifiers are built on a randomly *80/20* split dataset, using gini index and information gain as the impurity measures. The impact of using the individual impurity measures is analyzed by their prediction accuracies and some structural properties of the resultant decision trees.

PROCEDURE

The re-sampled dataset is randomly split into *80%* training set and *20%* testing set. Two decision trees are trained, both on the same training data. One of them uses information gain as the impurity measure to choose the best splitting attribute at every internal node and the other uses gain in the gini index. The accuracies of the trees are computed on the same test data. Besides accuracy, the depth, size and training time for the two decision trees are also compared to deduce some informative results.

RESULTS

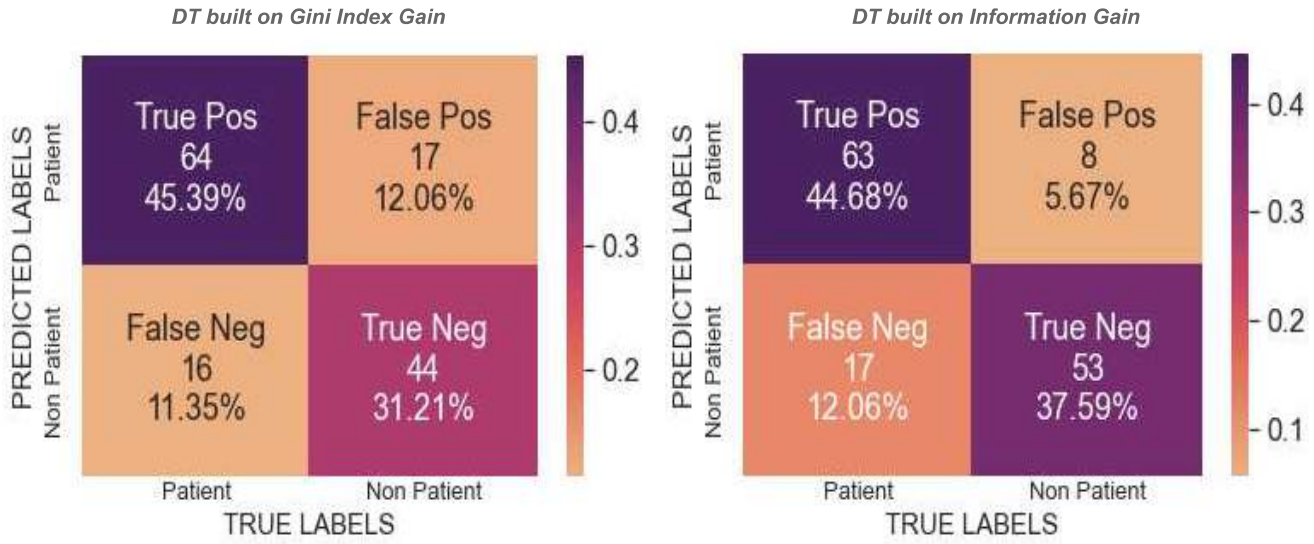
• **COMPLEXITY COMPARISON**

IMPURITY MEASURE	DEPTH	SIZE	TRAINING TIME
Gini Index	21	195	0.567 sec
Information Gain	24	203	0.710 sec

• **ACCURACY COMPARISON**

IMPURITY MEASURE	TEST ACCURACY	TRAIN ACCURACY
Gini Index	76.596 %	100.000 %
Information Gain	82.269 %	100.000 %

• **CONFUSION MATRICES**



CONCLUSION

COMPARING STRUCTURAL PROPERTIES AND TRAINING TIMES

• The train accuracies of both the trees are *100%* because of over-fitting and no pruning.

• Both the size (number of nodes) and height (depth) of the decision tree built with information gain as the impurity measure are more than in the case of gini index as the impurity measure.

• Decision tree took more time to be built when information gain was chosen as the impurity measure than when the gini index was chosen.

COMPARING PREDICTION CAPABILITIES

• The test accuracy of the tree built with information gain as the impurity measure is *5.67%* higher than the one built with gini index as the impurity measure on the test set.

• According to the confusion matrices, the tree built with information gain classifies *8* more examples correctly from the test set than the one built with the gini index.

• According to the confusion matrices, though both the trees have approximately the same TPR and FNR, the tree built with information gain impurity measure has considerably lower FPR, and hence significantly higher TNR.

In conclusion, overall the prediction accuracy is higher in this case when information gain is used, and hence it constructs a tree that is able to generalize better. In particular, the significant improvement in using information gain over gini index was observed for negative examples. Gini index is almost equally efficient in predicting classes of positive examples correctly but information gain is quite better in predicting classes of negative examples.

COMMENTS

A better estimate of the relative performances and relative structural complexities of information gain and gini index can be obtained if we compare their average behaviours. This is exactly what we have done in the next part.

PART 02

TASK SUMMARY

Accuracies are provided for decision trees built on information gain and gini index impurity measures by averaging them over 10 random *80/20* splits. The best tree (the one with the

highest test accuracy) is selected for the following parts.

PROCEDURE

Do 20 random *80/20* splits on the dataset. On the first ten splits, train ten distinct decision trees with gini index as the impurity measure and evaluate their performances on their respective test sets. For each of the trees, also compute their depth, size and training times. Finally compute the average test accuracy, average depth, average size and average training time over ten random splits. Also keep track of the highest test accuracy observed, the tree for which the highest test accuracy was observed, the least depth, the least size and the least training time observed.

Repeat these steps for the next ten splits with information gain as the impurity measure. Compare the best trees (ones with highest test accuracies) obtained in the two cases and select the better one.

RESULTS

• AVERAGE COMPLEXITY & ACCURACY COMPARISON

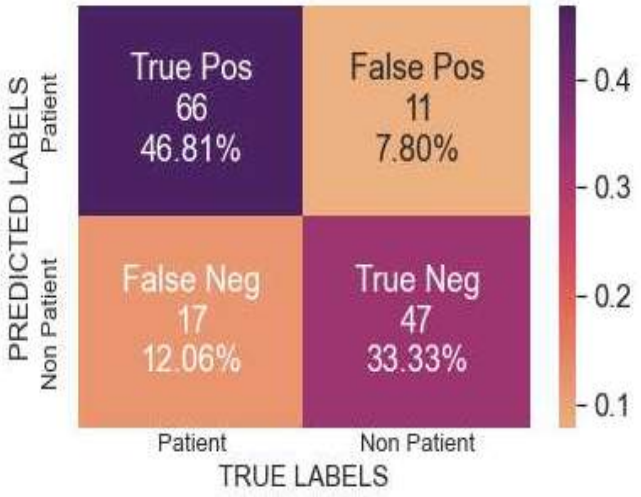
IMPURITY MEASURE	AVERAGE TEST ACCURACY	AVERAGE DEPTH	AVERAGE SIZE	AVERAGE TRAINING TIME
Gini Index	73.759 %	18.9	211.6	0.586 sec
Information Gain	73.830 %	20.3	190.8	0.618 sec

• BEST COMPLEXITY & ACCURACY COMPARISON

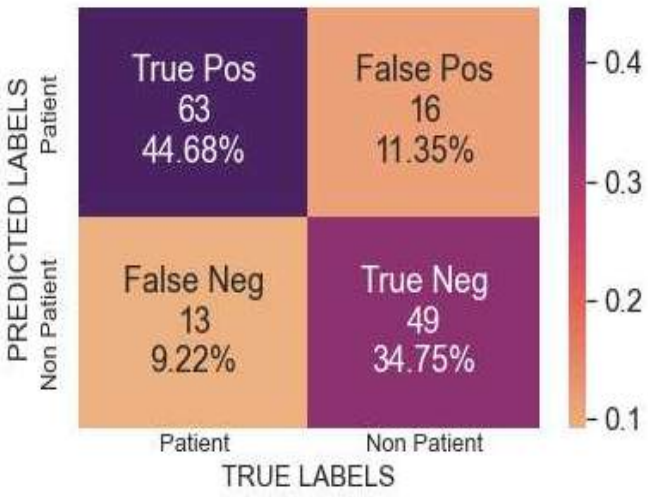
IMPURITY MEASURE	BEST TEST ACCURACY	BEST DEPTH	BEST SIZE	BEST TRAINING TIME
Gini Index	80.142 %	15	199	0.519 sec
Information Gain	79.433 %	15	179	0.516 sec

• CONFUSION MATRICES OF THE BEST DECISION TREES

Best DT built on Gini Index Gain



Best DT built on Information Gain



CONCLUSION

SELECTING THE BEST DECISION TREE

The highest test accuracy observed over the ten decision trees constructed on gini index impurity measure was *80.14%* and the highest test accuracy observed over the ten decision trees constructed on information gain impurity measure was *79.43%*. So we select the decision tree with the test accuracy of *80.14%* constructed with gini index impurity measure.

COMMENTS

This part highlights the average performances and structural complexities of the trees built on the two impurity measures. Though the average size of the decision tree is slightly higher for the gini index impurity measure, the average test accuracy and average depth for them are very close. This shows that both the impurity measures are capable of giving good results. It depends on the training data that which impurity measure might perform better; like in the first part, information gain measure performs significantly better.

PART 03

TASK SUMMARY

The best depth limit for the dataset is computed and its value justified with a plot. A plot of the test accuracy vs the total number of nodes in the tree is also provided.

PROCEDURE

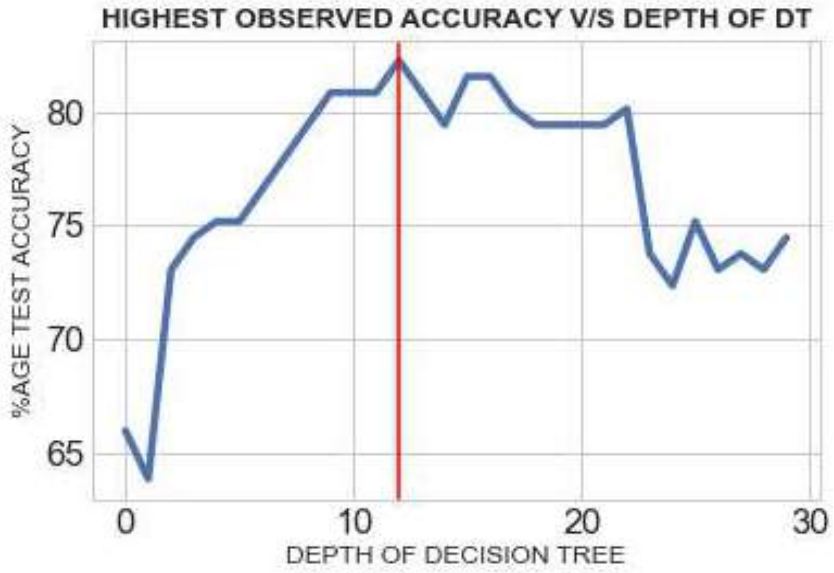
Build 20 decision trees on 20 random *80/20* splits of dataset, with half constructed on information gain and other half on gini index as the impurity measure (may use the 20 trees constructed in the second part). To find the best depth in terms of the test accuracy, explore the training histories of each of the 20 trees (training history of a tree keeps track of the test and train accuracy of partially constructed trees vs the intermediate depth and size of the tree while training) and find for every possible depth, the highest observed test accuracy across all the trees of that depth and consider only that. Tabulate the highest observed test accuracy vs depth data points and plot them on a graph. Locate the global maxima on the plot. Repeat the steps to find for every possible tree size, the highest observed test accuracy across all the trees. Plot their data points on a separate graph.

RESULTS

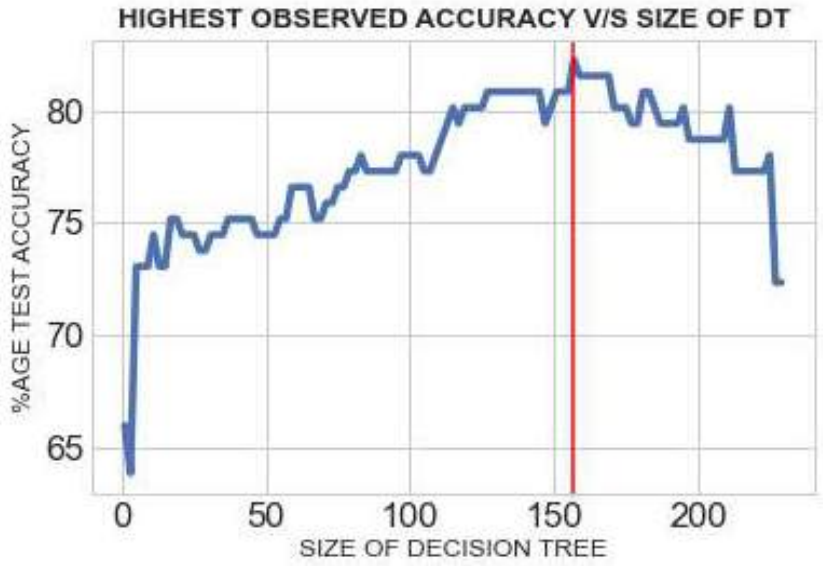
• HIGHEST OBSERVED TEST ACCURACY V/S DEPTH OF DECISION TREE

DEPTH	HIGHEST %AGE TEST ACCURACY	DEPTH	HIGHEST %AGE TEST ACCURACY	DEPTH	HIGHEST %AGE TEST ACCURACY
0	65.957	10	80.851	20	79.433
1	63.829	11	80.851	21	79.433
2	73.050	**12**	**82.270**	22	80.142
3	74.468	13	80.851	23	73.759
4	75.177	14	79.433	24	72.340
5	75.177	15	81.560	25	75.177
6	76.596	16	81.560	26	73.050
7	78.014	17	80.142	27	73.759

8	79.433	18	79.433	28	73.050
9	80.851	19	79.433	29	74.468



• **HIGHEST OBSERVED TEST ACCURACY V/S SIZE OF DECISION TREE**



CONCLUSION

BEST DEPTH OF DECISION TREE

• The test accuracy of the decision tree is observed to gradually increase with its depth till it reaches the global maxima. Beyond that it decreases for the most part, with some irregularities in between.

• According to the observations, the most optimal depth limit is ***12*** at which the highest test accuracy of ***82.27%*** is observed. It shows that among all the 20 trees and all of their partially constructed prototypes seen while training, the highest observed accuracy was *82.27%* for a tree of depth *12*.

BEST SIZE OF DECISION TREE

• The test accuracy of the decision tree is observed to gradually (approximately) increase with its size (number of nodes) till it reaches the global maxima. Beyond that it decreases for the most part, with some irregularities in between.

• According to the observations, the most optimal size is ***157*** at which the highest test accuracy of ***82.27%*** is observed. It shows that among all the 20 trees and all of their partially constructed prototypes seen while training, the highest observed accuracy was *82.27%* for a tree of size *157*.

The gradual decrease in test accuracy beyond the global maxima in both the plots indicates the overfitting of the decision tree on the training data that reduces its capability to generalize over unseen examples.

COMMENTS

This part was first conducted without oversampling the ILPD dataset. The graphs obtained in that case were monotonically non-increasing. That means, the highest test accuracy was observed for trees of depth *0* and size *1*. This is a clear consequence of a highly imbalanced dataset in which the tree with the most optimal depth performs no better than a naive classifier that always predicts the majority class as the target value.

PART 04

TASK SUMMARY

The tree with the highest test accuracy obtained in the second part was pruned using a valid statistical test for comparison.

BACKGROUND

STATISTICAL TESTING

The problem is that we split whenever the impurity measure instructs us to but we never check if the change in entropy (or gini index) is statistically significant. Are the number of examples of a particular class in the child nodes significantly different from the expected number? If not, it means that the split is statistically insignificant, i.e, the resulting children are not significantly different from what we would get by splitting the parent node based on a random distribution.

CHI SQUARE TESTING

χ^2 criterion is the criterion to detect useless splits. χ^2 statistic K measures how much the split deviates from what we would get if the data were split randomly. Small K values indicate low statistical significance. Mathematically, for a node P the χ^2 statistic K is calculated as $K = \sum_{(all\ classes\ i)} \sum_{(all\ children\ j\ of\ P)} \frac{(N_{ij} - N'_{ij})^2}{N'_{ij}}$ where N_{ij} means number of examples from class i in child j following the splitting condition and N'_{ij} means number of examples from class i in child j assuming a random selection.

THRESHOLD

A threshold t is chosen on the values of K to make the decision of whether to prune or not a particular node. Note that leaf nodes cannot be pruned (have no children). If the K value of an internal node is less than the threshold, it must be pruned otherwise not. Lower threshold causes bigger trees (more overfitting). Larger threshold causes smaller trees (less overfitting).

PROCEDURE

Write a recursive member function for the *Decision_Tree* class that conducts χ^2 statistical tests on the internal nodes of the tree in a post-order fashion, i.e, the child nodes must be checked for pruning before the parent nodes.

Find the optimal value of the threshold (within some predefined bounds) by pruning copies of the original tree (the best tree obtained in the second part) with variable thresholds, while incrementing the threshold by a small value at every iteration. Choose the threshold that produces the tree with the maximum test accuracy.

Select the tree produced by pruning the original tree with the most optimal threshold value. Compare the structural attributes and the accuracies of the trees before and after pruning.

RESULTS

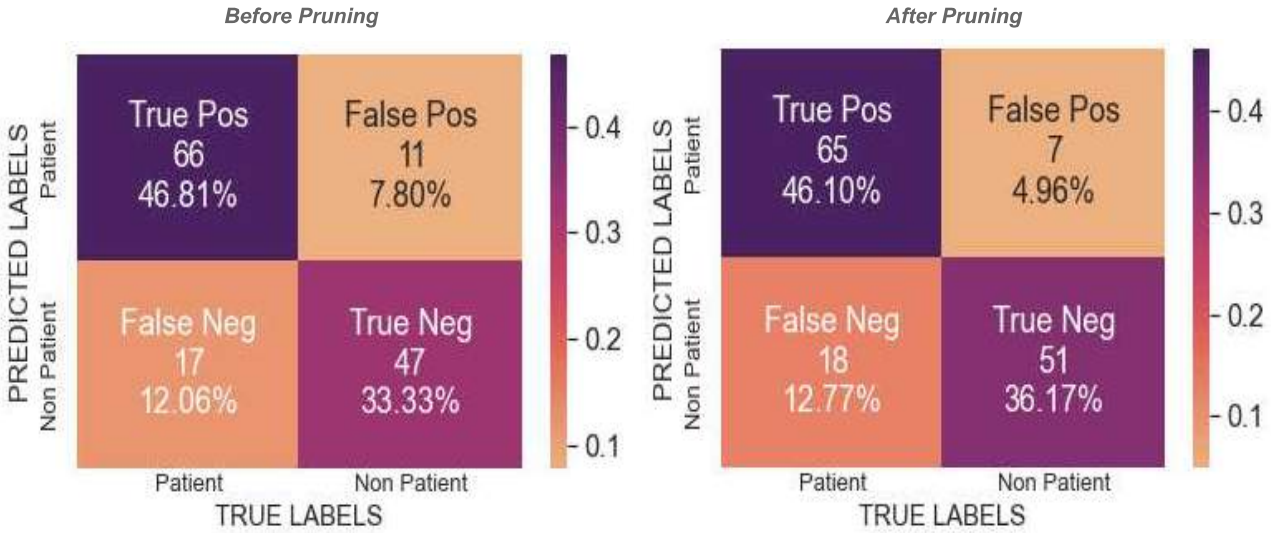
• BEFORE & AFTER PRUNING COMPLEXITY COMPARISON

	DEPTH OF DECISION TREE	SIZE OF DECISION TREE
BEFORE	22	211
AFTER	18	157

• BEFORE & AFTER PRUNING ACCURACY COMPARISON

	ACCURACY ON TEST SET	ACCURACY ON TRAINING SET
BEFORE	80.142 %	100.000 %
AFTER	82.270 %	95.936 %

• CONFUSION MATRICES



CONCLUSION

• The depth of the decision tree is reduced by 4 levels after pruning.

• The size of the decision tree is significantly reduced by 54 nodes.

• The changes in the structural attributes caused by pruning indicate that the complexity of the tree is significantly reduced due to pruning.

• The accuracy of the pruned decision tree on the training dataset has reduced relative to the unpruned decision tree's (reduction in overfitting).

• The accuracy of the pruned decision tree on the test dataset has increased relative to the unpruned decision tree's (better generalization abilities).

PART 05

TASK SUMMARY

The final trees obtained in the second (best tree without pruning), third (tree with the most optimal depth) and fourth (best tree after pruning) parts are printed adhering to the protocols stated in the procedure below.

PROCEDURE

Write a recursive member function for the *Decision_Tree* class that for each and every node prints the following in a well formatted manner such that its distance from the left margin is proportional to its hierarchical level in the tree.

- The number of positive and negative training examples associated with this node.
- A code (*{R, T, F}*) that reflects the condition to arrive at this node. *R* indicates a root node, *T* indicates that this node will be reached if the splitting condition in its parent node is true and similarly *F* indicates it will be reached if the condition is false.
- For a leaf node, the class label must be written. For an internal node, the splitting condition must be written.

RESULTS

• **BEST DECISION TREE BEFORE PRUNING (OBTAINED IN PART 2)**

Please find attached on page 12.

• **DECISION TREE WITH OPTIMAL DEPTH (OBTAINED IN PART 3)**

Please find attached on page 13.

• **BEST DECISION TREE AFTER PRUNING (OBTAINED IN PART 4)**

Please find attached on page 14.

COMMENTS

Due to a lot of queries and confusion among the students surrounding the tree that should be printed in this part, we decided to print the final trees obtained in all the second, third and fourth parts of the assignment.

```
->R(330+ 336-){DIRECT_BILIRUBIN < 1.0}
--|->T(202+ 312-){ALBUMIN < 61.53}
--|--|->T(159+ 303-){TOT_PROTEINS < 217.94}
--|--|--|->T(106+ 259-){AGE < 71.62}
--|--|--|--|->T(94+ 257-){SGPT < 5.11}
--|--|--|--|--|->T(14+ 14-){TOT_BILIRUBIN < 1.05}
--|--|--|--|--|--|->T(14+ 7-){GENDER = MALE}
--|--|--|--|--|--|--|->T(13+ 3-){AG_RATIO < 103.5}
--|--|--|--|--|--|--|--|->T(12+ 1-){TOT_PROTEINS < 208.0}
--|--|--|--|--|--|--|--|--|->T(11+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(1+ 1-){AGE < 52.5}
--|--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(1+ 2-){AGE < 60.0}
--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(1+ 4-){AGE < 49.0}
--|--|--|--|--|--|--|--|->T(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(0+ 7-)[NON LIVER PATIENT]
--|--|--|--|--|->F(80+ 243-){TOT_BILIRUBIN < 1.79}
--|--|--|--|--|--|->T(72+ 239-){TOT_BILIRUBIN < 0.6}
--|--|--|--|--|--|--|->T(10+ 10-){TOT_PROTEINS < 176.5}
--|--|--|--|--|--|--|--|->T(7+ 2-){SGPT < 6.95}
--|--|--|--|--|--|--|--|--|->T(7+ 1-){AG_RATIO < 33.0}
--|--|--|--|--|--|--|--|--|--|->T(6+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(1+ 1-){AGE < 34.5}
--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(3+ 8-){AG_RATIO < 29.5}
--|--|--|--|--|--|--|--|--|->T(1+ 8-){AGE < 24.5}
--|--|--|--|--|--|--|--|--|--|->T(1+ 3-){AGE < 20.5}
--|--|--|--|--|--|--|--|--|--|--|->T(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 5-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(62+ 229-){SGPT < 7.9}
--|--|--|--|--|--|--|--|->T(50+ 212-){AGE < 36.0}
--|--|--|--|--|--|--|--|--|->T(9+ 76-){TOT_PROTEINS < 211.5}
--|--|--|--|--|--|--|--|--|--|->T(7+ 76-){AG_RATIO < 16.75}
--|--|--|--|--|--|--|--|--|--|--|->T(3+ 4-){AGE < 25.0}
--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(4+ 72-){DIRECT_BILIRUBIN < 0.69}
--|--|--|--|--|--|--|--|--|--|--|--|->T(3+ 71-){AG_RATIO < 85.48}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(2+ 70-){TOT_PROTEINS < 197.93}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 63-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(2+ 7-){ALBUMIN < 39.5}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 7-){SGPT < 6.9}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(2+ 7-){ALBUMIN < 39.5}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 7-){SGPT < 6.9}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 7-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 1-){AGE < 28.6}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 1-){AGE < 34.26}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(41+ 136-){SGPT < 6.9}
--|--|--|--|--|--|--|--|--|--|->T(20+ 99-){TOT_PROTEINS < 195.2}
--|--|--|--|--|--|--|--|--|--|->T(20+ 67-){DIRECT_BILIRUBIN < 0.41}
--|--|--|--|--|--|--|--|--|--|->T(20+ 50-){DIRECT_BILIRUBIN < 0.3}
--|--|--|--|--|--|--|--|--|--|->T(15+ 48-){TOT_PROTEINS < 132.97}
--|--|--|--|--|--|--|--|--|--|->T(3+ 1-){GENDER = MALE}
--|--|--|--|--|--|--|--|--|--|->T(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(12+ 47-){TOT_PROTEINS < 167.2}
--|--|--|--|--|--|--|--|--|--|->T(2+ 26-){AG_RATIO < 62.72}
--|--|--|--|--|--|--|--|--|--|->T(1+ 25-){TOT_PROTEINS < 143.5}
--|--|--|--|--|--|--|--|--|--|->T(1+ 3-){AGE < 47.54}
--|--|--|--|--|--|--|--|--|--|->T(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 22-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(1+ 1-){AGE < 51.54}
--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(10+ 21-){AG_RATIO < 18.77}
--|--|--|--|--|--|--|--|--|--|->T(5+ 2-){GENDER = MALE}
--|--|--|--|--|--|--|--|--|--|->T(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(5+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(5+ 19-){SGPT < 6.22}
--|--|--|--|--|--|--|--|--|--|->T(5+ 7-){TOT_PROTEINS < 183.24}
--|--|--|--|--|--|--|--|--|--|->T(5+ 3-){AG_RATIO < 41.5}
--|--|--|--|--|--|--|--|--|--|->T(5+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 12-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(5+ 2-){AG_RATIO < 30.5}
--|--|--|--|--|--|--|--|--|--|->T(1+ 2-){AGE < 58.0}
--|--|--|--|--|--|--|--|--|--|->T(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(4+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 17-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 32-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(21+ 37-){SGPT < 7.21}
--|--|--|--|--|--|--|--|->T(15+ 12-){TOT_PROTEINS < 196.41}
--|--|--|--|--|--|--|--|->T(10+ 12-){GENDER = MALE}
--|--|--|--|--|--|--|--|->T(6+ 11-){TOT_PROTEINS < 172.5}
--|--|--|--|--|--|--|--|->T(5+ 3-){ALBUMIN < 25.5}
--|--|--|--|--|--|--|--|->T(1+ 3-){AGE < 63.69}
--|--|--|--|--|--|--|--|->T(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(4+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(1+ 8-){TOT_BILIRUBIN < 0.73}
--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 8-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(4+ 1-){AGE < 60.88}
--|--|--|--|--|--|--|--|->T(4+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(5+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(6+ 25-){AG_RATIO < 22.54}
--|--|--|--|--|--|--|--|->T(0+ 13-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(6+ 12-){TOT_PROTEINS < 150.64}
--|--|--|--|--|--|--|--|->T(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(4+ 12-){TOT_BILIRUBIN < 1.12}
--|--|--|--|--|--|--|--|--|->T(3+ 12-){DIRECT_BILIRUBIN < 0.15}
--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(2+ 12-){AGE < 48.28}
--|--|--|--|--|--|--|--|->T(2+ 4-){AGE < 44.35}
--|--|--|--|--|--|--|--|--|->T(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 8-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(12+ 17-){AGE < 36.0}
--|--|--|--|--|--|--|--|->T(10+ 4-){DIRECT_BILIRUBIN < 0.5}
--|--|--|--|--|--|--|--|->T(10+ 2-){AGE < 29.5}
--|--|--|--|--|--|--|--|->T(7+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(3+ 2-){AG_RATIO < 19.0}
--|--|--|--|--|--|--|--|->T(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(2+ 13-){ALBUMIN < 45.0}
--|--|--|--|--|--|--|--|->T(1+ 13-){AG_RATIO < 19.0}
--|--|--|--|--|--|--|--|->T(1+ 2-){AGE < 47.5}
--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(0+ 11-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(8+ 4-){SGPT < 5.89}
--|--|--|--|--|--|--|->T(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(8+ 2-){DIRECT_BILIRUBIN < 0.94}
--|--|--|--|--|--|--|--|->T(8+ 1-){ALBUMIN < 17.22}
--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(8+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|->F(12+ 2-){AGE < 81.0}
--|--|--|--|--|->T(12+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|->F(53+ 44-){AGE < 12.67}
--|--|--|--|->T(1+ 8-){GENDER = MALE}
--|--|--|--|--|->T(0+ 8-)[NON LIVER PATIENT]
--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|->F(52+ 36-){SGPT < 6.06}
--|--|--|--|--|->T(11+ 17-){ALBUMIN < 36.89}
--|--|--|--|--|--|->T(5+ 16-){TOT_PROTEINS < 271.31}
--|--|--|--|--|--|--|->T(1+ 12-){TOT_BILIRUBIN < 0.65}
--|--|--|--|--|--|--|--|->T(1+ 1-){AGE < 50.0}
--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 11-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(4+ 4-){AGE < 62.5}
--|--|--|--|--|--|--|--|->T(1+ 4-){TOT_PROTEINS < 303.0}
--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(6+ 1-){ALBUMIN < 55.0}
--|--|--|--|--|--|--|--|->T(6+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(41+ 19-){AG_RATIO < 34.28}
--|--|--|--|--|--|--|->T(22+ 2-){ALBUMIN < 11.0}
--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(22+ 1-){SGOT < 4.1}
--|--|--|--|--|--|--|--|--|->T(21+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(1+ 1-){AGE < 22.0}
--|--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(19+ 17-){TOT_PROTEINS < 246.5}
--|--|--|--|--|--|--|--|->T(7+ 1-){ALBUMIN < 56.04}
--|--|--|--|--|--|--|--|--|->T(7+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(12+ 16-){SGPT < 7.95}
--|--|--|--|--|--|--|--|--|->T(9+ 16-){AGE < 48.43}
--|--|--|--|--|--|--|--|--|--|->T(3+ 12-){ALKPHOS < 1.39}
--|--|--|--|--|--|--|--|--|--|--|->T(1+ 12-){ALBUMIN < 17.51}
--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 12-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(6+ 4-){AGE < 59.15}
--|--|--|--|--|--|--|--|--|--|--|->T(5+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(1+ 4-){AG_RATIO < 70.5}
--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(3+ 0-)[LIVER PATIENT]
--|--|->F(43+ 9-){AG_RATIO < 53.54}
--|--|--|->T(3+ 6-){TOT_BILIRUBIN < 0.8}
--|--|--|--|->T(0+ 6-)[NON LIVER PATIENT]
--|--|--|--|->F(3+ 0-)[LIVER PATIENT]
--|--|--|->F(40+ 3-){SGPT < 7.29}
--|--|--|--|->T(32+ 0-)[LIVER PATIENT]
--|--|--|--|->F(8+ 3-){AGE < 40.5}
--|--|--|--|--|->T(3+ 3-){AGE < 27.27}
--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(3+ 2-){TOT_BILIRUBIN < 1.1}
--|--|--|--|--|--|--|->T(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(1+ 2-){TOT_BILIRUBIN < 1.7}
--|--|--|--|--|--|--|--|->T(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(5+ 0-)[LIVER PATIENT]
--|->F(128+ 24-){TOT_PROTEINS < 147.48}
--|--|->T(4+ 9-){ALBUMIN < 23.5}
--|--|--|->T(3+ 0-)[LIVER PATIENT]
--|--|--|->F(1+ 9-){AGE < 30.56}
--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|->F(0+ 9-)[NON LIVER PATIENT]
--|--|->F(124+ 15-){DIRECT_BILIRUBIN < 1.28}
--|--|--|->T(20+ 10-){AGE < 39.0}
--|--|--|--|->T(3+ 8-){SGOT < 2.65}
--|--|--|--|--|->T(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(1+ 8-){ALBUMIN < 25.31}
--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(0+ 8-)[NON LIVER PATIENT]
--|--|--|--|->F(17+ 2-){TOT_BILIRUBIN < 2.78}
--|--|--|--|--|->T(16+ 1-){DIRECT_BILIRUBIN < 1.05}
--|--|--|--|--|--|->T(4+ 1-){TOT_BILIRUBIN < 2.15}
--|--|--|--|--|--|--|->T(4+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(12+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(1+ 1-){AGE < 45.71}
--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|->F(104+ 5-){ALBUMIN < 12.0}
--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|->F(104+ 4-){TOT_PROTEINS < 547.98}
--|--|--|--|--|->T(85+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(19+ 4-){ALBUMIN < 28.04}
--|--|--|--|--|--|->T(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(19+ 2-){SGPT < 5.95}
--|--|--|--|--|--|--|->T(3+ 2-){AGE < 52.5}
--|--|--|--|--|--|--|--|->T(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(16+ 0-)[LIVER PATIENT]
```

```
->R(332+ 234-){DIRECT_BILIRUBIN < 1.2}
--|->T(215+ 223-){TOT_PROTEINS < 208.0}
--|--|->T(109+ 172-){ALBUMIN < 93.5}
--|--|--|->T(103+ 172-){AGE < 71.0}
--|--|--|--|->T(94+ 170-){TOT_PROTEINS < 132.5}
--|--|--|--|--|->T(13+ 9-){TOT_BILIRUBIN < 0.85}
--|--|--|--|--|--|->T(12+ 2-){AGE < 62.0}
--|--|--|--|--|--|--|->T(12+ 1-){ALKPHOS < 0.8}
--|--|--|--|--|--|--|--|->T(1+ 1-){AGE < 40.0}
--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(11+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(1+ 7-){ALBUMIN < 60.5}
--|--|--|--|--|--|--|->T(0+ 7-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(81+ 161-){AGE < 25.92}
--|--|--|--|--|--|->T(6+ 28-){SGPT < 8.0}
--|--|--|--|--|--|--|->T(3+ 27-){TOT_BILIRUBIN < 1.6}
--|--|--|--|--|--|--|--|->T(2+ 26-){TOT_PROTEINS < 184.59}
--|--|--|--|--|--|--|--|--|->T(0+ 16-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(2+ 10-){TOT_PROTEINS < 187.0}
--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(1+ 10-){ALKPHOS < 1.25}
--|--|--|--|--|--|--|--|--|--|--|->T(0+ 9-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(1+ 1-){AGE < 19.5}
--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(1+ 1-){AGE < 20.0}
--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(3+ 1-){TOT_BILIRUBIN < 1.2}
--|--|--|--|--|--|--|--|--|->T(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(75+ 133-){TOT_BILIRUBIN < 1.95}
--|--|--|--|--|--|--|--|->T(71+ 132-){ALBUMIN < 39.5}
--|--|--|--|--|--|--|--|--|->T(63+ 100-){AG_RATIO < 31.5}
--|--|--|--|--|--|--|--|--|--|->T(38+ 79-){GENDER = MALE}
--|--|--|--|--|--|--|--|--|--|--|->T(16+ 52-){SGOT < 2.75}
--|--|--|--|--|--|--|--|--|--|--|--|->T(5+ 2-){ALKPHOS < 0.88}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(4+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(11+ 50-){SGOT < 4.0}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(5+ 40-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(6+ 10-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(22+ 27-){AGE < 48.5}
--|--|--|--|--|--|--|--|--|--|--|--|->T(17+ 10-){SGPT < 5.95}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(17+ 6-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(5+ 17-){ALKPHOS < 1.35}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(3+ 17-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(25+ 21-){AG_RATIO < 34.0}
--|--|--|--|--|--|--|--|--|--|->T(6+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(19+ 21-){AG_RATIO < 39.5}
--|--|--|--|--|--|--|--|--|--|--|->T(2+ 9-){AG_RATIO < 34.15}
--|--|--|--|--|--|--|--|--|--|--|--|->T(2+ 1-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 8-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(17+ 12-){TOT_PROTEINS < 163.0}
--|--|--|--|--|--|--|--|--|--|--|--|->T(7+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(10+ 12-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(8+ 32-){TOT_PROTEINS < 145.5}
--|--|--|--|--|--|--|--|--|--|->T(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(6+ 32-){AG_RATIO < 38.5}
--|--|--|--|--|--|--|--|--|--|--|->T(4+ 6-){AGE < 37.5}
--|--|--|--|--|--|--|--|--|--|--|--|->T(3+ 1-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 5-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(2+ 26-){TOT_BILIRUBIN < 1.8}
--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 26-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(4+ 1-){TOT_BILIRUBIN < 2.63}
--|--|--|--|--|--|--|--|--|->T(4+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|->F(9+ 2-){ALBUMIN < 18.0}
--|--|--|--|--|->T(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|->F(9+ 0-)[LIVER PATIENT]
--|--|--|->F(6+ 0-)[LIVER PATIENT]
--|--|->F(106+ 51-){AGE < 18.0}
--|--|--|->T(5+ 13-){TOT_BILIRUBIN < 0.95}
--|--|--|--|->T(2+ 13-){GENDER = MALE}
--|--|--|--|--|->T(1+ 13-){ALKPHOS < 1.45}
--|--|--|--|--|--|->T(0+ 13-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|->F(3+ 0-)[LIVER PATIENT]
--|--|--|->F(101+ 38-){TOT_PROTEINS < 275.5}
--|--|--|--|->T(44+ 28-){TOT_BILIRUBIN < 0.65}
--|--|--|--|--|->T(8+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(36+ 28-){SGPT < 6.36}
--|--|--|--|--|--|->T(12+ 18-){ALBUMIN < 49.0}
--|--|--|--|--|--|--|->T(5+ 18-){AG_RATIO < 53.5}
--|--|--|--|--|--|--|--|->T(2+ 16-){SGOT < 3.5}
--|--|--|--|--|--|--|--|--|->T(1+ 16-){AG_RATIO < 21.22}
--|--|--|--|--|--|--|--|--|--|->T(1+ 1-){AGE < 52.5}
--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 15-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(3+ 2-){TOT_BILIRUBIN < 1.95}
--|--|--|--|--|--|--|--|--|->T(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(7+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(24+ 10-){AG_RATIO < 66.0}
--|--|--|--|--|--|--|->T(22+ 5-){TOT_PROTEINS < 212.5}
--|--|--|--|--|--|--|--|->T(3+ 3-){AGE < 49.5}
--|--|--|--|--|--|--|--|--|->T(3+ 1-){AGE < 29.5}
--|--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(19+ 2-){ALKPHOS < 1.45}
--|--|--|--|--|--|--|--|->T(19+ 1-){TOT_PROTEINS < 257.0}
--|--|--|--|--|--|--|--|--|->T(17+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(2+ 1-){AGE < 62.5}
--|--|--|--|--|--|--|--|--|--|->T(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(2+ 5-){SGPT < 7.15}
--|--|--|--|--|--|--|--|--|->T(2+ 1-){AGE < 36.5}
--|--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|->F(57+ 10-){AGE < 25.5}
--|--|--|--|--|->T(4+ 5-){AG_RATIO < 46.0}
--|--|--|--|--|--|->T(1+ 5-){TOT_BILIRUBIN < 2.35}
--|--|--|--|--|--|--|->T(0+ 5-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(53+ 5-){SGPT < 7.9}
--|--|--|--|--|--|->T(47+ 1-){AGE < 62.0}
--|--|--|--|--|--|--|->T(38+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(9+ 1-){AGE < 64.5}
--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(9+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(6+ 4-){TOT_BILIRUBIN < 1.15}
--|--|--|--|--|--|--|->T(5+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(1+ 4-){AGE < 48.0}
--|--|--|--|--|--|--|--|->T(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|->F(117+ 11-){TOT_PROTEINS < 147.0}
--|--|->T(7+ 4-){TOT_PROTEINS < 144.5}
--|--|--|->T(7+ 1-){TOT_BILIRUBIN < 3.85}
--|--|--|--|->T(1+ 1-){AGE < 43.0}
--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|->F(6+ 0-)[LIVER PATIENT]
--|--|--|->F(0+ 3-)[NON LIVER PATIENT]
--|--|->F(110+ 7-){ALBUMIN < 12.0}
--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|->F(110+ 6-){TOT_PROTEINS < 400.0}
--|--|--|--|->T(82+ 0-)[LIVER PATIENT]
--|--|--|--|->F(28+ 6-){TOT_PROTEINS < 412.5}
--|--|--|--|--|->T(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|->F(28+ 3-){ALBUMIN < 27.5}
--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(28+ 2-){TOT_PROTEINS < 977.71}
--|--|--|--|--|--|--|->T(23+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(5+ 2-){TOT_BILIRUBIN < 7.75}
--|--|--|--|--|--|--|--|->T(1+ 2-){AGE < 41.5}
--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(4+ 0-)[LIVER PATIENT]
```

```
->R(330+ 336-){DIRECT_BILIRUBIN < 1.0}
--|->T(202+ 312-){ALBUMIN < 61.5}
--|--|->T(162+ 305-){TOT_PROTEINS < 217.94}
--|--|--|->T(108+ 261-){AGE < 71.62}
--|--|--|--|->T(96+ 259-){TOT_BILIRUBIN < 1.79}
--|--|--|--|--|->T(87+ 253-){SGPT < 5.11}
--|--|--|--|--|--|->T(14+ 12-){TOT_BILIRUBIN < 1.05}
--|--|--|--|--|--|--|->T(14+ 7-){GENDER = MALE}
--|--|--|--|--|--|--|--|->T(13+ 3-){AG_RATIO < 103.5}
--|--|--|--|--|--|--|--|--|->T(12+ 1-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(1+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(1+ 4-){AGE < 49.0}
--|--|--|--|--|--|--|--|--|->T(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(0+ 5-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(73+ 241-){TOT_BILIRUBIN < 0.6}
--|--|--|--|--|--|--|->T(10+ 10-){TOT_PROTEINS < 176.5}
--|--|--|--|--|--|--|--|->T(7+ 2-){SGPT < 6.95}
--|--|--|--|--|--|--|--|--|->T(7+ 1-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(3+ 8-){AG_RATIO < 29.5}
--|--|--|--|--|--|--|--|--|->T(1+ 8-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(63+ 231-){SGPT < 7.9}
--|--|--|--|--|--|--|--|->T(51+ 214-){SGPT < 7.61}
--|--|--|--|--|--|--|--|--|->T(51+ 197-){AGE < 22.0}
--|--|--|--|--|--|--|--|--|--|->T(0+ 15-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(51+ 182-){SGPT < 6.9}
--|--|--|--|--|--|--|--|--|--|--|->T(26+ 129-){ALKPHOS < 1.1}
--|--|--|--|--|--|--|--|--|--|--|--|->T(23+ 88-){AG_RATIO < 43.5}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(17+ 81-){TOT_PROTEINS < 195.22}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(16+ 55-){TOT_PROTEINS < 167.75}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(4+ 40-){TOT_PROTEINS < 143.5}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(3+ 7-){AG_RATIO < 22.88}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 7-){TOT_PROTEINS < 110.3}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 7-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 33-){AG_RATIO < 16.11}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 3-){AGE < 40.0}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 30-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(12+ 15-){TOT_PROTEINS < 182.86}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(9+ 4-){AGE < 41.5}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(8+ 1-){AGE < 64.5}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(8+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(3+ 11-){TOT_BILIRUBIN < 0.88}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 10-){AGE < 63.93}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 10-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(2+ 1-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(1+ 26-){AGE < 35.05}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(6+ 7-){SGPT < 6.33}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(6+ 3-){ALBUMIN < 27.95}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 3-){AGE < 37.25}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(5+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(3+ 41-){SGOT < 3.8}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 28-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(3+ 13-){TOT_PROTEINS < 185.22}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(0+ 10-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(3+ 3-){ALBUMIN < 45.0}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(25+ 53-){AGE < 42.0}
--|--|--|--|--|--|--|--|--|--|--|--|->T(7+ 32-){TOT_PROTEINS < 202.5}
--|--|--|--|--|--|--|--|--|--|--|--|->T(4+ 31-){AG_RATIO < 15.2}
--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(3+ 31-){ALBUMIN < 42.84}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 29-){SGOT < 3.05}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 29-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(2+ 2-){TOT_BILIRUBIN < 0.87}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|->F(3+ 1-){TOT_PROTEINS < 215.5}
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->T(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|--|--|->F(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|--|->F(18+ 21-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(0+ 17-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(12+ 17-){AGE < 36.0}
--|--|--|--|--|--|--|--|--|--|->T(10+ 4-){DIRECT_BILIRUBIN < 0.5}
--|--|--|--|--|--|--|--|--|--|--|->T(10+ 2-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|->F(2+ 13-){ALBUMIN < 45.0}
--|--|--|--|--|--|--|--|--|--|--|->T(1+ 13-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|--|--|->F(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(9+ 6-){SGPT < 5.44}
--|--|--|--|--|--|--|->T(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(9+ 3-){AG_RATIO < 35.75}
--|--|--|--|--|--|--|--|->T(3+ 3-){TOT_BILIRUBIN < 1.8}
--|--|--|--|--|--|--|--|--|->T(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(6+ 0-)[LIVER PATIENT]
--|--|--|--|->F(12+ 2-){AGE < 81.0}
--|--|--|--|--|->T(12+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(0+ 2-)[NON LIVER PATIENT]
--|--|--|->F(54+ 44-){TOT_PROTEINS < 523.5}
--|--|--|--|->T(46+ 44-){AGE < 29.06}
--|--|--|--|--|->T(8+ 19-){TOT_PROTEINS < 235.3}
--|--|--|--|--|--|->T(3+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(5+ 19-){ALKPHOS < 1.35}
--|--|--|--|--|--|--|->T(3+ 19-){GENDER = MALE}
--|--|--|--|--|--|--|--|->T(1+ 18-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(2+ 1-)[LIVER PATIENT]
--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|->F(38+ 25-){SGPT < 6.29}
--|--|--|--|--|->T(13+ 18-){AG_RATIO < 22.5}
--|--|--|--|--|--|->T(4+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(9+ 18-){AG_RATIO < 52.14}
--|--|--|--|--|--|--|->T(2+ 15-){ALKPHOS < 0.83}
--|--|--|--|--|--|--|--|->T(0+ 12-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(2+ 3-){TOT_BILIRUBIN < 0.73}
--|--|--|--|--|--|--|--|--|->T(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(7+ 3-){AGE < 47.43}
--|--|--|--|--|--|--|--|--|->T(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(7+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(25+ 7-){ALBUMIN < 53.62}
--|--|--|--|--|--|--|->T(25+ 3-){AGE < 64.5}
--|--|--|--|--|--|--|--|->T(23+ 1-){ALBUMIN < 11.0}
--|--|--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(23+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(2+ 2-){AGE < 65.5}
--|--|--|--|--|--|--|--|--|->T(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|--|--|->F(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|->F(8+ 0-)[LIVER PATIENT]
--|--|->F(40+ 7-){AGE < 39.5}
--|--|--|->T(18+ 7-){TOT_PROTEINS < 184.09}
--|--|--|--|->T(0+ 4-)[NON LIVER PATIENT]
--|--|--|--|->F(18+ 3-)[LIVER PATIENT]
--|--|--|->F(22+ 0-)[LIVER PATIENT]
--|->F(128+ 24-){AGE < 39.51}
--|--|->T(28+ 18-){AGE < 34.07}
--|--|--|->T(25+ 4-){ALBUMIN < 31.04}
--|--|--|--|->T(1+ 3-){AGE < 22.27}
--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(0+ 3-)[NON LIVER PATIENT]
--|--|--|--|->F(24+ 1-)[LIVER PATIENT]
--|--|--|->F(3+ 14-){ALBUMIN < 74.77}
--|--|--|--|->T(1+ 14-){SGOT < 2.55}
--|--|--|--|--|->T(1+ 0-)[LIVER PATIENT]
--|--|--|--|--|->F(0+ 14-)[NON LIVER PATIENT]
--|--|--|--|->F(2+ 0-)[LIVER PATIENT]
--|--|->F(100+ 6-){SGPT < 6.25}
--|--|--|->T(40+ 6-){SGOT < 2.3}
--|--|--|--|->T(25+ 0-)[LIVER PATIENT]
--|--|--|--|->F(15+ 6-){AGE < 43.5}
--|--|--|--|--|->T(0+ 2-)[NON LIVER PATIENT]
--|--|--|--|--|->F(15+ 4-){TOT_PROTEINS < 586.5}
--|--|--|--|--|--|->T(14+ 2-){DIRECT_BILIRUBIN < 1.05}
--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|->F(14+ 1-){TOT_PROTEINS < 149.05}
--|--|--|--|--|--|--|->T(0+ 1-)[NON LIVER PATIENT]
--|--|--|--|--|--|--|->F(14+ 0-)[LIVER PATIENT]
--|--|--|--|--|--|->F(1+ 2-)[NON LIVER PATIENT]
--|--|--|->F(60+ 0-)[LIVER PATIENT]
```

APPENDIX

ABBREVIATIONS

%AGE	Percentage
ACC	Accuracy
ALP	Alkaline Phosphatase
AVG	Average
DT	Decision Tree
FNR	False Negative Rate
FPR	False Positive Rate
GG	Gini Gain
IG	Information Gain
ILPD	Indian Liver Patient Dataset
SGOT	Aspartate Aminotransferase
SGPT	Alanine Aminotransferase
SMOTE	Synthetic Minority Oversampling Technique
TNR	True Negative Rate
TPR	True Positive Rate
WRT	With Respect To

CODE EXECUTION

- **requirements.txt**

Specifies the python packages and modules required to run the code.
Ensure all the necessary dependencies of required versions and latest version of Python3 are available with the following command

```
>>> pip3 install -r requirements.txt
```

- **ILPD.csv**

Contains the data used for training and testing of decision trees.

- **GRP23_A01_IMPLEMENTATION.ipynb**

Jupyter Notebook written in Python containing all the code relevant to the assignment.

Launch jupyter notebook through terminal using the command: `jupyter notebook`
Then open this notebook to view the code written and run to see the output results.

Note: Please make sure before running the notebook, the *Indian Liver Patients Dataset* is present by the name of *ILPD.csv* and *GRP23_A01_IMPLEMENTATION.ipynb are* in the same directory.

(other details are present in README.txt)

REFERENCES

- Chi-Squared Statistical Test Wikipedia
- Colorama Python Documentation
- Decision Trees by Carnegie Mellon's School of CS
- Heatmaps on Seaborn Python
- Imbalanced-Learn SMOTE Python Documentation
- Indian Liver Patient Dataset on Kaggle
- Matplotlib Python Documentation
- Synthetic Minority Oversampling Technique