Course ID: CS60050

Course Name: Machine Learning

Medical Insurance Approval using Decision Tree based Learning Model

Submitted by:

Bratin Mondal
Roll Number: 21CS10016

Kapil Bhanwariwal
Roll Number: 23BM6JP20

Group: 23

---

1 Introduction

Medical insurance approval processes are often complex and deterministic rules may not be feasible. This study aims to employ machine learning techniques, specifically decision tree models, to automate the approval process and enhance decision-making transparency.

2 Model Description

The model consists of several components:

- **Decision Tree Nodes:** The decision tree is constructed using nodes, each representing a decision point. Nodes have attributes such as attribute to split on, split value, entropy, and whether they are leaf nodes.
- **Decision Tree Construction:** The Decision_tree function constructs the decision tree recursively. It selects the best attribute to split on based on a criterion (e.g., entropy), splits the samples into left and right branches, and creates child nodes accordingly.
- **Pruning:** The model includes a pruning mechanism to reduce overfitting. Pruning is performed based on a validation set to remove unnecessary nodes from the tree.
- **Prediction:** The model provides functions to predict the class of a single sample or multiple samples using the constructed decision tree.
- **Accuracy Calculation:** Functions are available to calculate the accuracy of predictions and other metrics such as precision, recall, and F1-score.
- **Visualization:** The decision tree can be visualized using the visualize_tree function, which generates a graphical representation of the tree structure.

The machine learning model is trained and used in the following steps:

1. **Import Libraries:** Necessary libraries including the model itself are imported.
2. **Define Constants and Attributes:** Constants such as the maximum height of the tree and a list of attributes are defined.
3. **Main Function:** The main function executes the training and testing process for the model.
   (a) **Train the Model:** The model is trained using the `select_best_tree` function for both entropy and Gini criteria.
   (b) **Training Process:** Within the `select_best_tree` function:
      - The training data is loaded and processed.
      - Trees of different heights are trained using the `Decision_tree` function.
      - Pruning is performed to reduce overfitting based on a validation set.
      - Metrics are calculated and written to files for both unpruned and pruned trees.
      - Graphical representations of the trees are generated and saved.
      - The best unpruned and pruned trees are identified based on test accuracy.

   (c) **Training Complete:** Once training is complete for both criteria, the best trees are displayed along with their heights and accuracies.

3 Results

3.1 Scikit-learn Decision Tree Classifier:

We tested the Scikit-learn package on our dataset using both the entropy and Gini index criteria.

(a) Metrics for Gini Index Criterion

Class	Precision	Recall	F1-score	Support
bad	0.9825	1.0000	0.9912	112
ok	0.9756	0.9091	0.9412	44
good	0.6000	0.8571	0.7059	7
vgood	0.8750	0.7000	0.7778	10
Accuracy			0.9538	

(b) Metrics for Entropy Criterion

Class	Precision	Recall	F1-score	Support
bad	0.9912	1.0000	0.9956	112
ok	0.9500	0.8636	0.9048	44
good	0.4615	0.8571	0.6000	7
vgood	1.0000	0.7000	0.8235	10
Accuracy			0.9422	

Link to Visualization of Decision Tree using Gini Index
Link to Visualization of Decision Tree using Entropy

3.2 Experiment

In our experiment, we train our model using both Entropy Gain and Gini Index criteria. We use the parameter `max_height` to denote the maximum height of the decision tree. We test the trained decision tree and then prune it based on a validation set. Finally, we evaluate the pruned decision tree. We range the `max_height` parameter from 1 to 20.

3.2.1 Unpruned Gini Index Based Decision Tree

Link to visualization of Unpruned Gini Index Based Decision Tree

Table 2: Unpruned Gini Index Based Decision Tree Matrices for class **bad**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.6474	1.0	0.786	112	0.6474
2	2	0.6474	1.0	0.786	112	0.6474
3	3	1.0	0.8393	0.9126	112	0.7977
4	4	0.888	0.9911	0.9367	112	0.815
5	5	1.0	0.9375	0.9677	112	0.8266
6	6	0.9646	0.9732	0.9689	112	0.8844
7	7	0.9244	0.9821	0.9524	112	0.8613
8	8	0.9737	0.9911	0.9823	112	0.9306
9	9	0.9825	1.0	0.9912	112	0.948
10	10	0.9825	1.0	0.9912	112	0.948
11	11	0.9912	1.0	0.9956	112	0.9422
12	12	0.9912	1.0	0.9956	112	0.9595
13	13	0.9912	1.0	0.9956	112	0.9595
14	14	0.9912	1.0	0.9956	112	0.9827
15	13	0.9912	1.0	0.9956	112	0.948
16	13	1.0	1.0	1.0	112	0.9422
17	13	0.9912	1.0	0.9956	112	0.9538
18	13	0.9825	1.0	0.9912	112	0.9364
19	14	0.9912	1.0	0.9956	112	0.9711
20	14	0.9825	1.0	0.9912	112	0.9653

Table 3: Unpruned Gini Index Based Decision Tree Matrices for class **ok**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	44	0.6474
2	2	0.0	0.0	0.0	44	0.6474
3	3	0.557	1.0	0.7154	44	0.7977
4	4	0.625	0.6818	0.6522	44	0.815
5	5	0.775	0.7045	0.7381	44	0.8266
6	6	0.8182	0.8182	0.8182	44	0.8844
7	7	0.8667	0.5909	0.7027	44	0.8613
8	8	0.9722	0.7955	0.875	44	0.9306
9	9	1.0	0.8409	0.9136	44	0.948
10	10	1.0	0.8409	0.9136	44	0.948
11	11	0.9268	0.8636	0.8941	44	0.9422
12	12	0.9535	0.9318	0.9425	44	0.9595
13	13	0.9744	0.8636	0.9157	44	0.9595
14	14	0.9767	0.9545	0.9655	44	0.9827
15	13	0.9268	0.8636	0.8941	44	0.948
16	13	0.95	0.8636	0.9048	44	0.9422
17	13	1.0	0.8864	0.9398	44	0.9538
18	13	0.9737	0.8409	0.9024	44	0.9364
19	14	0.9756	0.9091	0.9412	44	0.9711
20	14	0.9524	0.9091	0.9302	44	0.9653

Table 4: Unpruned Gini Index Based Decision Tree Matrices for class **good**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	7	0.6474
2	2	0.0	0.0	0.0	7	0.6474
3	3	0.0	0.0	0.0	7	0.7977
4	4	0.0	0.0	0.0	7	0.815
5	5	0.25	1.0	0.4	7	0.8266
6	6	0.0	0.0	0.0	7	0.8844
7	7	0.3846	0.7143	0.5	7	0.8613
8	8	0.5	0.7143	0.5882	7	0.9306
9	9	0.5	1.0	0.6667	7	0.948
10	10	0.5	1.0	0.6667	7	0.948
11	11	0.5	0.8571	0.6316	7	0.9422
12	12	0.6	0.8571	0.7059	7	0.9595
13	13	0.6	0.8571	0.7059	7	0.9595
14	14	0.8571	0.8571	0.8571	7	0.9827
15	13	0.5	0.7143	0.5882	7	0.948
16	13	0.4615	0.8571	0.6	7	0.9422
17	13	0.5	0.8571	0.6316	7	0.9538
18	13	0.4615	0.8571	0.6	7	0.9364
19	14	0.7	1.0	0.8235	7	0.9711
20	14	0.75	0.8571	0.8	7	0.9653

Table 5: Unpruned Gini Index Based Decision Tree Matrices for class **vgood**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	10	0.6474
2	2	0.0	0.0	0.0	10	0.6474
3	3	0.0	0.0	0.0	10	0.7977
4	4	0.0	0.0	0.0	10	0.815
5	5	0.0	0.0	0.0	10	0.8266
6	6	0.5	0.8	0.6154	10	0.8844
7	7	0.7273	0.8	0.7619	10	0.8613
8	8	0.7692	1.0	0.8696	10	0.9306
9	9	1.0	0.8	0.8889	10	0.948
10	10	1.0	0.8	0.8889	10	0.948
11	11	1.0	0.7	0.8235	10	0.9422
12	12	1.0	0.7	0.8235	10	0.9595
13	13	0.9091	1.0	0.9524	10	0.9595
14	14	1.0	1.0	1.0	10	0.9827
15	13	1.0	0.9	0.9474	10	0.948
16	13	0.875	0.7	0.7778	10	0.9422
17	13	0.8889	0.8	0.8421	10	0.9538
18	13	0.875	0.7	0.7778	10	0.9364
19	14	1.0	0.9	0.9474	10	0.9711
20	14	1.0	0.9	0.9474	10	0.9653

3.2.2 Pruned Gini Index Based Decision Tree

Link to visualization of Pruned Gini Index Based Decision Tree

Table 6: Pruned Gini Index Based Decision Tree Matrices for class **bad**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.6474	1.0	0.786	112	0.6474
2	1	0.6474	1.0	0.786	112	0.6474
3	3	1.0	0.8393	0.9126	112	0.7977
4	4	0.888	0.9911	0.9367	112	0.815
5	5	1.0	0.9375	0.9677	112	0.8613
6	6	0.9646	0.9732	0.9689	112	0.8844
7	7	0.982	0.9732	0.9776	112	0.896
8	8	0.9652	0.9911	0.978	112	0.9075
9	9	1.0	0.9911	0.9955	112	0.9306
10	10	1.0	0.9911	0.9955	112	0.9306
11	11	0.9737	0.9911	0.9823	112	0.9191
12	10	0.9573	1.0	0.9782	112	0.9364
13	12	0.9912	1.0	0.9956	112	0.9364
14	12	1.0	0.9911	0.9955	112	0.9538
15	12	0.9825	1.0	0.9912	112	0.9538
16	12	1.0	0.9911	0.9955	112	0.9653
17	13	1.0	0.9911	0.9955	112	0.9422
18	12	0.9912	1.0	0.9956	112	0.9653
19	12	0.9737	0.9911	0.9823	112	0.9595
20	12	0.9911	0.9911	0.9911	112	0.9653

Table 7: Pruned Gini Index Based Decision Tree Matrices for class **ok**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	44	0.6474
2	1	0.0	0.0	0.0	44	0.6474
3	3	0.557	1.0	0.7154	44	0.7977
4	4	0.625	0.6818	0.6522	44	0.815
5	5	0.6471	1.0	0.7857	44	0.8613
6	6	0.8182	0.8182	0.8182	44	0.8844
7	7	0.8684	0.75	0.8049	44	0.896
8	8	0.9688	0.7045	0.8158	44	0.9075
9	9	0.9722	0.7955	0.875	44	0.9306
10	10	0.9722	0.7955	0.875	44	0.9306
11	11	0.9706	0.75	0.8462	44	0.9191
12	10	1.0	0.7955	0.8861	44	0.9364
13	12	0.9487	0.8409	0.8916	44	0.9364
14	12	0.9737	0.8409	0.9024	44	0.9538
15	12	1.0	0.8409	0.9136	44	0.9538
16	12	0.9762	0.9318	0.9535	44	0.9653
17	13	0.9737	0.8409	0.9024	44	0.9422
18	12	1.0	0.9091	0.9524	44	0.9653
19	12	0.9744	0.8636	0.9157	44	0.9595
20	12	0.9756	0.9091	0.9412	44	0.9653

Table 8: Pruned Gini Index Based Decision Tree Matrices for class **good**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	7	0.6474
2	1	0.0	0.0	0.0	7	0.6474
3	3	0.0	0.0	0.0	7	0.7977
4	4	0.0	0.0	0.0	7	0.815
5	5	0.0	0.0	0.0	7	0.8613
6	6	0.0	0.0	0.0	7	0.8844
7	7	0.3846	0.7143	0.5	7	0.896
8	8	0.3846	0.7143	0.5	7	0.9075
9	9	0.3846	0.7143	0.5	7	0.9306
10	10	0.3846	0.7143	0.5	7	0.9306
11	11	0.4118	1.0	0.5833	7	0.9191
12	10	0.625	0.7143	0.6667	7	0.9364
13	12	0.4667	1.0	0.6364	7	0.9364
14	12	0.5	1.0	0.6667	7	0.9538
15	12	0.5455	0.8571	0.6667	7	0.9538
16	12	0.5833	1.0	0.7368	7	0.9653
17	13	0.4545	0.7143	0.5556	7	0.9422
18	12	0.7143	0.7143	0.7143	7	0.9653
19	12	0.7	1.0	0.8235	7	0.9595
20	12	0.6667	0.8571	0.75	7	0.9653

Table 9: Pruned Gini Index Based Decision Tree Matrices for class **vgood**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	10	0.6474
2	1	0.0	0.0	0.0	10	0.6474
3	3	0.0	0.0	0.0	10	0.7977
4	4	0.0	0.0	0.0	10	0.815
5	5	0.0	0.0	0.0	10	0.8613
6	6	0.5	0.8	0.6154	10	0.8844
7	7	0.7273	0.8	0.7619	10	0.896
8	8	0.7692	1.0	0.8696	10	0.9075
9	9	0.7692	1.0	0.8696	10	0.9306
10	10	0.7692	1.0	0.8696	10	0.9306
11	11	1.0	0.8	0.8889	10	0.9191
12	10	0.7692	1.0	0.8696	10	0.9364
13	12	1.0	0.6	0.75	10	0.9364
14	12	1.0	1.0	1.0	10	0.9538
15	12	0.9091	1.0	0.9524	10	0.9538
16	12	1.0	0.8	0.8889	10	0.9653
17	13	0.7692	1.0	0.8696	10	0.9422
18	12	0.7692	1.0	0.8696	10	0.9653
19	12	1.0	1.0	1.0	10	0.9595
20	12	0.9091	1.0	0.9524	10	0.9653

3.2.3 Unpruned Information Gain Based Decision Tree

Link to visualization of Unpruned Information Gain Based Decision Tree

Table 10: Unpruned Information Gain Based Decision Tree Matrices for class **bad**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.6474	1.0	0.786	112	0.6474
2	2	0.6474	1.0	0.786	112	0.6474
3	3	1.0	0.8393	0.9126	112	0.7977
4	4	1.0	0.8393	0.9126	112	0.7977
5	5	1.0	0.9375	0.9677	112	0.8266
6	6	0.973	0.9643	0.9686	112	0.8844
7	7	0.925	0.9911	0.9569	112	0.8786
8	8	0.9652	0.9911	0.978	112	0.9364
9	9	0.9739	1.0	0.9868	112	0.9306
10	10	0.9739	1.0	0.9868	112	0.948
11	11	1.0	1.0	1.0	112	0.9422
12	12	0.9825	1.0	0.9912	112	0.948
13	13	0.9912	1.0	0.9956	112	0.9422
14	12	0.9912	1.0	0.9956	112	0.9422
15	14	0.9825	1.0	0.9912	112	0.9364
16	13	1.0	1.0	1.0	112	0.9595
17	13	1.0	1.0	1.0	112	0.9538
18	14	0.9912	1.0	0.9956	112	0.9364
19	13	0.9912	1.0	0.9956	112	0.9422
20	14	0.9912	1.0	0.9956	112	0.9422

Table 11: Unpruned Information Gain Based Decision Tree Matrices for class **ok**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	44	0.6474
2	2	0.0	0.0	0.0	44	0.6474
3	3	0.557	1.0	0.7154	44	0.7977
4	4	0.557	1.0	0.7154	44	0.7977
5	5	0.775	0.7045	0.7381	44	0.8266
6	6	0.8043	0.8409	0.8222	44	0.8844
7	7	0.963	0.5909	0.7324	44	0.8786
8	8	0.973	0.8182	0.8889	44	0.9364
9	9	0.9722	0.7955	0.875	44	0.9306
10	10	0.9744	0.8636	0.9157	44	0.948
11	11	0.9744	0.8636	0.9157	44	0.9422
12	12	0.975	0.8864	0.9286	44	0.948
13	13	0.9737	0.8409	0.9024	44	0.9422
14	12	0.95	0.8636	0.9048	44	0.9422
15	14	0.9487	0.8409	0.8916	44	0.9364
16	13	0.9762	0.9318	0.9535	44	0.9595
17	13	0.975	0.8864	0.9286	44	0.9538
18	14	0.9487	0.8409	0.8916	44	0.9364
19	13	0.9474	0.8182	0.878	44	0.9422
20	14	0.95	0.8636	0.9048	44	0.9422

Table 12: Unpruned Information Gain Based Decision Tree Matrices for class **good**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	7	0.6474
2	2	0.0	0.0	0.0	7	0.6474
3	3	0.0	0.0	0.0	7	0.7977
4	4	0.0	0.0	0.0	7	0.7977
5	5	0.25	1.0	0.4	7	0.8266
6	6	0.0	0.0	0.0	7	0.8844
7	7	0.3846	0.7143	0.5	7	0.8786
8	8	0.625	0.7143	0.6667	7	0.9364
9	9	0.4667	1.0	0.6364	7	0.9306
10	10	0.5833	1.0	0.7368	7	0.948
11	11	0.4615	0.8571	0.6	7	0.9422
12	12	0.6	0.8571	0.7059	7	0.948
13	13	0.4286	0.8571	0.5714	7	0.9422
14	12	0.4615	0.8571	0.6	7	0.9422
15	14	0.5	0.8571	0.6316	7	0.9364
16	13	0.5455	0.8571	0.6667	7	0.9595
17	13	0.5	1.0	0.6667	7	0.9538
18	14	0.4286	0.8571	0.5714	7	0.9364
19	13	0.4615	0.8571	0.6	7	0.9422
20	14	0.4615	0.8571	0.6	7	0.9422

Table 13: Unpruned Information Gain Based Decision Tree Matrices for class **vgood**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	10	0.6474
2	2	0.0	0.0	0.0	10	0.6474
3	3	0.0	0.0	0.0	10	0.7977
4	4	0.0	0.0	0.0	10	0.7977
5	5	0.0	0.0	0.0	10	0.8266
6	6	0.5	0.8	0.6154	10	0.8844
7	7	0.7692	1.0	0.8696	10	0.8786
8	8	0.7692	1.0	0.8696	10	0.9364
9	9	1.0	0.7	0.8235	10	0.9306
10	10	1.0	0.7	0.8235	10	0.948
11	11	0.7778	0.7	0.7368	10	0.9422
12	12	0.7778	0.7	0.7368	10	0.948
13	13	1.0	0.8	0.8889	10	0.9422
14	12	1.0	0.7	0.8235	10	0.9422
15	14	0.875	0.7	0.7778	10	0.9364
16	13	0.875	0.7	0.7778	10	0.9595
17	13	1.0	0.7	0.8235	10	0.9538
18	14	1.0	0.7	0.8235	10	0.9364
19	13	1.0	0.9	0.9474	10	0.9422
20	14	1.0	0.7	0.8235	10	0.9422

3.2.4 Pruned Information Gain Based Decision Tree

Link to visualization of Pruned Information Gain Based Decision Tree

Table 14: Pruned Information Gain Based Decision Tree Matrices for class **bad**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.6474	1.0	0.786	112	0.6474
2	1	0.6474	1.0	0.786	112	0.6474
3	3	1.0	0.8393	0.9126	112	0.7977
4	1	0.6474	1.0	0.786	112	0.6474
5	5	1.0	0.9375	0.9677	112	0.8613
6	6	1.0	0.9375	0.9677	112	0.8844
7	7	0.9391	0.9643	0.9515	112	0.8728
8	8	0.9821	0.9821	0.9821	112	0.9249
9	9	0.9823	0.9911	0.9867	112	0.9191
10	10	0.9825	1.0	0.9912	112	0.9711
11	11	0.9912	1.0	0.9956	112	0.948
12	12	0.9912	1.0	0.9956	112	0.948
13	12	0.9912	1.0	0.9956	112	0.9306
14	12	0.9912	1.0	0.9956	112	0.9422
15	12	0.9912	1.0	0.9956	112	0.9595
16	12	0.9912	1.0	0.9956	112	0.9595
17	12	0.9739	1.0	0.9868	112	0.948
18	12	0.9912	1.0	0.9956	112	0.9595
19	13	0.9739	1.0	0.9868	112	0.9364
20	12	0.9912	1.0	0.9956	112	0.9595

Table 15: Pruned Information Gain Based Decision Tree Matrices for class **ok**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	44	0.6474
2	1	0.0	0.0	0.0	44	0.6474
3	3	0.557	1.0	0.7154	44	0.7977
4	1	0.0	0.0	0.0	44	0.6474
5	5	0.6471	1.0	0.7857	44	0.8613
6	6	0.7692	0.9091	0.8333	44	0.8844
7	7	0.875	0.6364	0.7368	44	0.8728
8	8	0.9459	0.7955	0.8642	44	0.9249
9	9	0.9706	0.75	0.8462	44	0.9191
10	10	1.0	0.9318	0.9647	44	0.9711
11	11	1.0	0.8409	0.9136	44	0.948
12	12	1.0	0.8409	0.9136	44	0.948
13	12	1.0	0.7727	0.8718	44	0.9306
14	12	1.0	0.8182	0.9	44	0.9422
15	12	1.0	0.8864	0.9398	44	0.9595
16	12	1.0	0.8864	0.9398	44	0.9595
17	12	1.0	0.8409	0.9136	44	0.948
18	12	1.0	0.8864	0.9398	44	0.9595
19	13	0.9487	0.8409	0.8916	44	0.9364
20	12	1.0	0.8864	0.9398	44	0.9595

Table 16: Pruned Information Gain Based Decision Tree Matrices for class **good**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	7	0.6474
2	1	0.0	0.0	0.0	7	0.6474
3	3	0.0	0.0	0.0	7	0.7977
4	1	0.0	0.0	0.0	7	0.6474
5	5	0.0	0.0	0.0	7	0.8613
6	6	0.0	0.0	0.0	7	0.8844
7	7	0.3846	0.7143	0.5	7	0.8728
8	8	0.4545	0.7143	0.5556	7	0.9249
9	9	0.3889	1.0	0.56	7	0.9191
10	10	1.0	0.7143	0.8333	7	0.9711
11	11	0.5	0.7143	0.5882	7	0.948
12	12	0.5	0.7143	0.5882	7	0.948
13	12	0.3846	0.7143	0.5	7	0.9306
14	12	0.4545	0.7143	0.5556	7	0.9422
15	12	0.5385	1.0	0.7	7	0.9595
16	12	0.625	0.7143	0.6667	7	0.9595
17	12	0.625	0.7143	0.6667	7	0.948
18	12	0.5385	1.0	0.7	7	0.9595
19	13	0.5385	1.0	0.7	7	0.9364
20	12	0.5385	1.0	0.7	7	0.9595

Table 17: Pruned Information Gain Based Decision Tree Matrices for class **vgood**

Tree Max Height	Tree Actual Height	Precision	Recall	F1-Score	Support	Accuracy
1	1	0.0	0.0	0.0	10	0.6474
2	1	0.0	0.0	0.0	10	0.6474
3	3	0.0	0.0	0.0	10	0.7977
4	1	0.0	0.0	0.0	10	0.6474
5	5	0.0	0.0	0.0	10	0.8613
6	6	0.5	0.8	0.6154	10	0.8844
7	7	0.7692	1.0	0.8696	10	0.8728
8	8	0.7692	1.0	0.8696	10	0.9249
9	9	1.0	0.8	0.8889	10	0.9191
10	10	0.7692	1.0	0.8696	10	0.9711
11	11	0.7692	1.0	0.8696	10	0.948
12	12	0.7692	1.0	0.8696	10	0.948
13	12	0.7692	1.0	0.8696	10	0.9306
14	12	0.7692	1.0	0.8696	10	0.9422
15	12	1.0	0.8	0.8889	10	0.9595
16	12	0.7692	1.0	0.8696	10	0.9595
17	12	0.7692	1.0	0.8696	10	0.948
18	12	1.0	0.8	0.8889	10	0.9595
19	13	1.0	0.6	0.75	10	0.9364
20	12	1.0	0.8	0.8889	10	0.9595

4 Conclusion

The pruned decision tree, constructed with an information gain approach and a height of 10, demonstrated superior performance on the test dataset, achieving an accuracy of 97.11%. It's noteworthy that the model was built without utilizing any machine learning libraries.

Pruning played a crucial role in reducing overfitting in the decision tree model. By systematically removing unnecessary branches, pruning helped simplify the model, making it more generalizable to unseen data. This reduction in complexity prevented the model from capturing noise or irrelevant patterns in the training data, thereby improving its performance on the test dataset.

The choice of criterion, whether Gini index or information gain, significantly influenced the construction and performance of the decision tree. Information gain measures the reduction in entropy achieved by splitting the data on a particular attribute, while Gini index quantifies the impurity of the dataset. Both criteria aim to find the attribute that maximizes the homogeneity of the resulting subsets, leading to better decision boundaries and ultimately improving the model's predictive accuracy.

In conclusion, pruning effectively mitigated overfitting, and the selection of an appropriate criterion played a crucial role in optimizing the decision tree model for medical insurance approval.

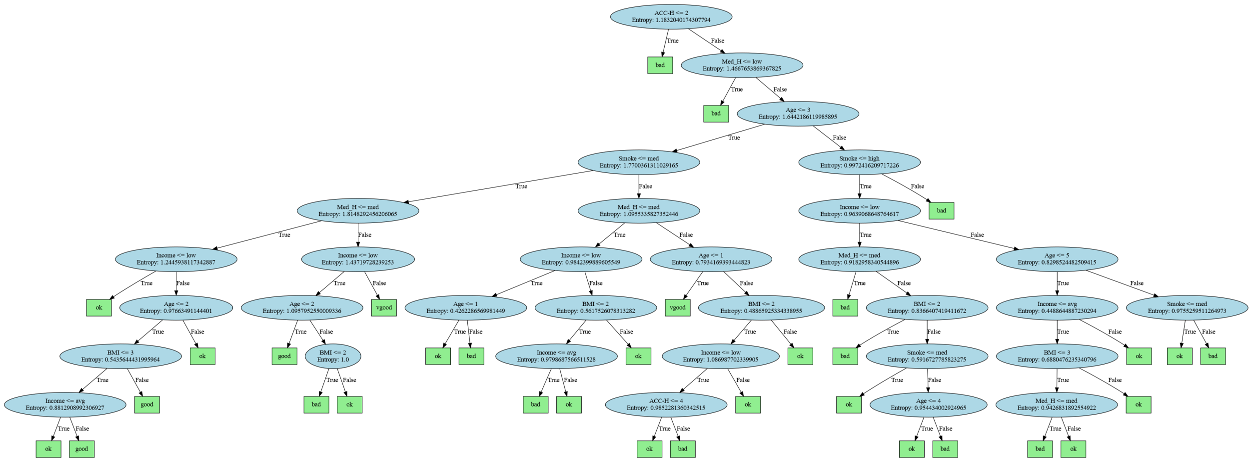

Figure 1: Pruned Decision Tree with Information Gain, Height 10