Decision Tree

A decision tree is a type of supervised learning algorithm that is commonly used in machine learning to model and predict outcomes based on input data. It is a tree-like structure where each internal node tests on attribute, each branch corresponds to attribute value and each leaf node represents the final decision or prediction. The decision tree algorithm falls under the category of supervised learning. They can be used to solve both regression and classification problems.Decision tree learning is one of the most widely adopted algorithms for classification.

Decision Tree Terminologies

There are specialized terms associated with decision trees that denote various components and facets of the tree structure and decision-making procedure. :

Root Node:

A decision tree’s root node, which represents the original choice or feature from which the tree branches, is the highest node.

Internal Nodes (Decision Nodes):

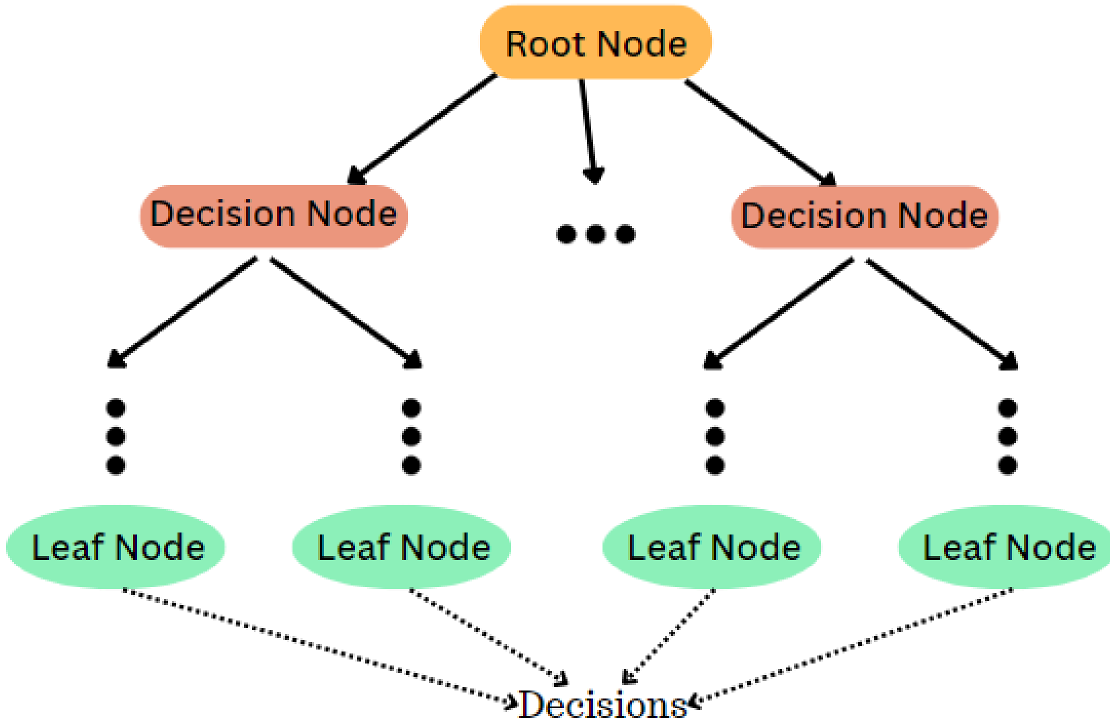
Nodes in the tree whose choices are determined by the values of particular attributes. There are branches on these nodes that go to other nodes.

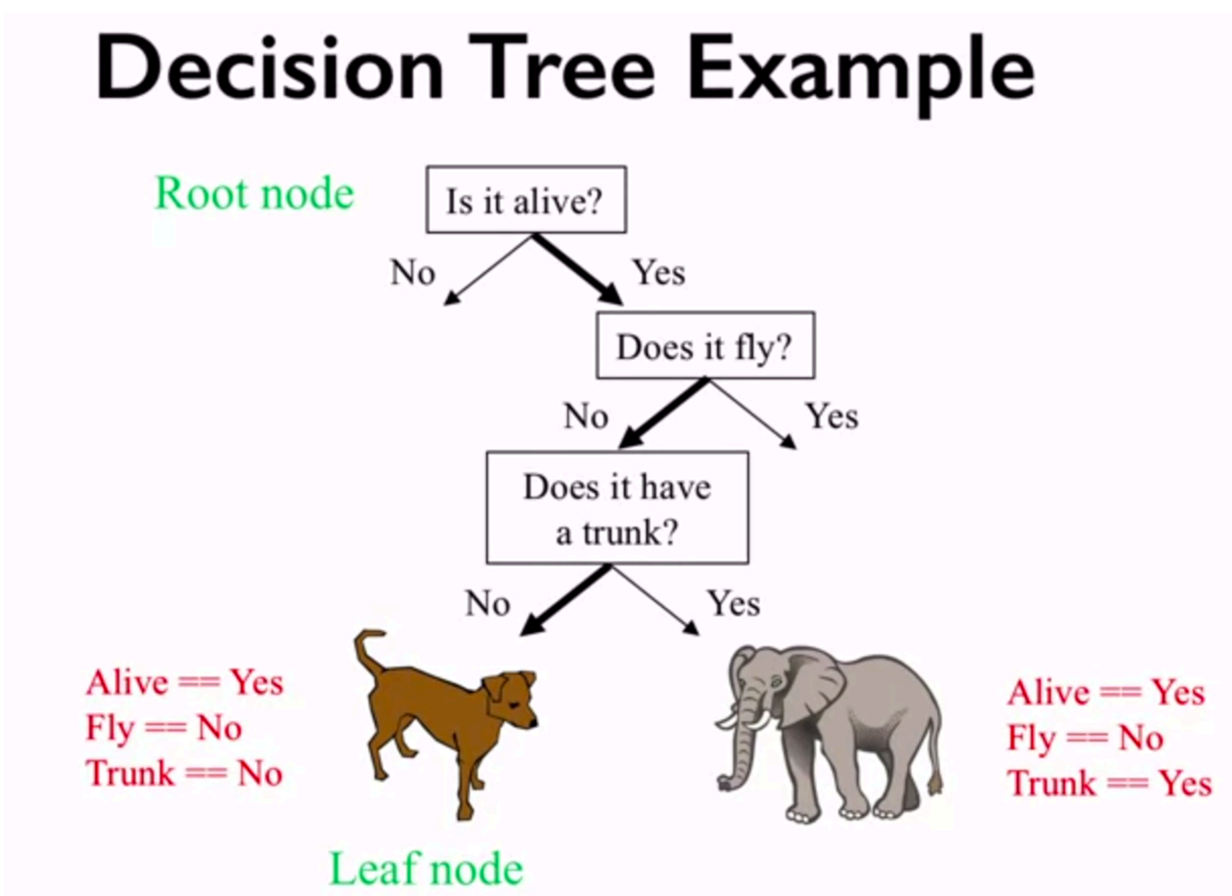
Leaf Nodes (Terminal Nodes):

The branches’ termini, when choices or forecasts are decided upon. There are no more branches on leaf nodes.

Branches (Edges):

Links between nodes that show how decisions are made in response to particular circumstances. Splitting: The process of dividing a node into two or more sub-nodes based on a decision criterion. It involves selecting a feature and a threshold to create subsets of data.





Information Gain:

- Information gain is the measurement of changes in entropy after the segmentation of a dataset based on an attribute.
- It calculates how much information a feature provides us about a class.
- According to the value of information gain, we split the node and build the decision tree.
- A decision tree algorithm always tries to maximize the value of information gain, and a node/attribute having the highest information gain is split first. It can be calculated using the below formula:

```
Information Gain= Entropy(S)- [(Weighted Avg) *Entropy(each feature)
```

Entropy:

Entropy is a metric to measure the impurity in a given attribute. It specifies randomness in data. Entropy can be calculated as:

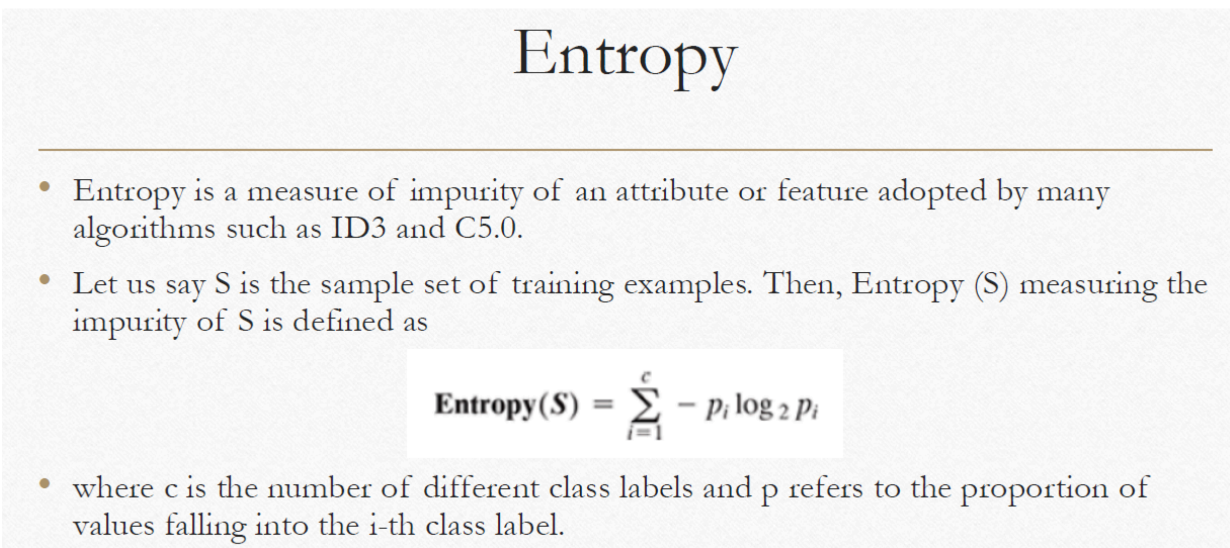
Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)

Where,

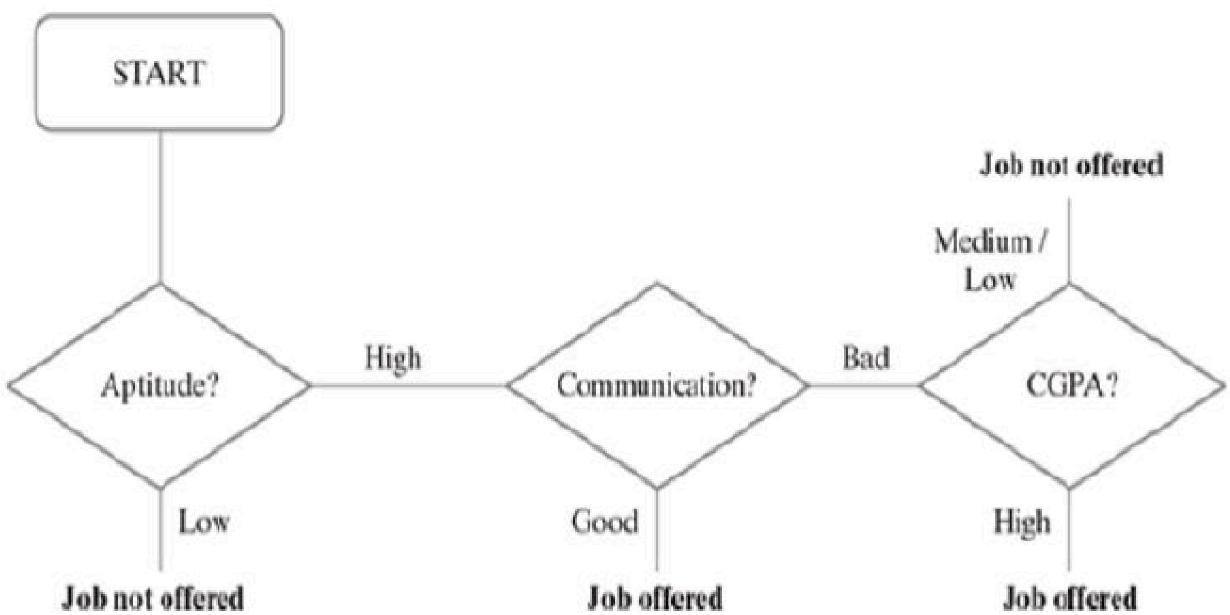
S= Total number of samples

P(yes)= probability of yes

P(no)= probability of no



CGPA	Communication	Aptitude	Programming Skill	Job offered?
High	Good	High	Good	Yes
Medium	Good	High	Good	Yes
Low	Bad	Low	Good	No
Low	Good	Low	Bad	No
High	Good	High	Bad	Yes
High	Good	High	Good	Yes
Medium	Bad	Low	Bad	No
Medium	Bad	Low	Good	No
High	Bad	High	Good	Yes
Medium	Good	High	Good	Yes
Low	Bad	High	Bad	No
Low	Bad	High	Bad	No
Medium	Good	High	Bad	Yes
Low	Good	Low	Good	No
High	Bad	Low	Bad	No
Medium	Bad	High	Good	No
High	Bad	Low	Bad	No
Medium	Good	High	Bad	Yes



There are many implementations of decision tree, the most prominent ones being C5.0, CART (Classification and Regression Tree), CHAID (Chi-square Automatic Interaction Detector) and ID3 (Iterative Dichotomiser3) algorithms.

The biggest challenge of a decision tree algorithm is to find out which feature to split upon.

The main driver for identifying the feature is that the data should be split in such a way that the partitions created by the split should contain examples belonging to a single class. If that happens, the partitions are considered to be pure.

Advantages of Decision Tree

- Easy to understand and interpret, making them accessible to non-experts.
- Handle both numerical and categorical data without requiring extensive preprocessing.
- Provides insights into feature importance for decision-making.
- Handle missing values and outliers without significant impact.
- Applicable to both classification and regression tasks.

Disadvantages of Decision Tree

- **Disadvantages** include the potential for overfitting
- Sensitivity to small changes in data, limited generalization if training data is not representative
- Potential bias in the presence of imbalanced data.

Conclusion

Decision trees, a key tool in machine learning, model and predict outcomes based on input data through a tree-like structure. They offer interpretability, versatility, and simple visualization, making them valuable for both categorization and regression tasks. While decision trees have advantages like ease of understanding, they may face challenges such as overfitting. Understanding their terminologies and formation process is essential for effective application in diverse scenarios.

In [18]:

```python
import pandas as pd
```

In [19]:

```python
import pandas as pd
data = pd.read_csv('DecisionTreeDataset -Num.csv')
data
```

Out[19]:

	CGPA	Communication	Apptitude	Programming Skill	Job Offered
0	2	1	1	1	1
1	1	1	1	1	1
2	0	0	0	1	0
3	0	1	0	0	0
4	2	1	1	0	1
5	2	1	1	1	1
6	1	0	0	0	0
7	1	0	0	1	0
8	2	0	1	1	1

	CGPA	Communication	Apptitude	Programming Skill	Job Offered
9	1	1	1	1	1
10	0	0	1	0	0
11	0	0	1	0	0
12	1	1	1	0	1
13	0	1	0	1	0
14	2	0	0	0	0
15	1	0	1	1	0
16	2	0	0	0	0
17	2	1	1	0	1

In [20]:
```
x = data.drop('Job Offered', axis = 1)
y = data['Job Offered']
```

In [21]:
```
x.shape
```

Out[21]: (18, 4)

In [22]:
```
y.shape
```

Out[22]: (18,)

In [23]:
```
from sklearn.tree import DecisionTreeClassifier
dtree_entropy = DecisionTreeClassifier(criterion = 'entropy')
model = dtree_entropy.fit(x,y)
dtree_entropy.get_depth()
```

Out[23]: 3

In [24]:
```
from sklearn import tree
text_representation = tree.export_text(dtree_entropy,feature_names=['CGPA','Communicati
print(text_representation)
```

```
|--- Apptitude <= 0.50
|   |--- class: 0
|--- Apptitude >  0.50
|   |--- Communication <= 0.50
|   |   |--- CGPA <= 1.50
|   |   |   |--- class: 0
|   |   |--- CGPA >  1.50
|   |   |   |--- class: 1
|   |--- Communication >  0.50
|   |   |--- class: 1
```

In [25]:
```
#Predictions
prediction = dtree_entropy.predict(x)
prediction
```

Out[25]: array([1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1], dtype=int64)

In [26]:
```python
diff=pd.DataFrame({'Actual':y,'Predicted':prediction})
diff
```

Out[26]:

	Actual	Predicted
0	1	1
1	1	1
2	0	0
3	0	0
4	1	1
5	1	1
6	0	0
7	0	0
8	1	1
9	1	1
10	0	0
11	0	0
12	1	1
13	0	0
14	0	0
15	0	0
16	0	0
17	1	1

In [27]:
```python
#Metric Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y, prediction)
cm
```

Out[27]:
```
array([[10,  0],
       [ 0,  8]], dtype=int64)
```

		Predicted Class	
		No	Yes
Observed Class	No	TN	FP
	Yes	FN	TP

TN	True Negative
FP	False Positive
FN	False Negative
TP	True Positive

In [28]:
```
TN = cm[0][0]
FP = cm[0][1]
FN = cm[1][0]
TP = cm[1][1]
print(TP, FN, TN, FP)
```

```
8 0 10 0
```

In [29]:
```
accuracy = (TP + TN) / (TP + FP + FN + TN)
accuracy
```

Out[29]:
```
1.0
```

In [30]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y, prediction)
```

Out[30]:
```
1.0
```

In [31]:
```
sensitivity = TP / (TP + FN)
sensitivity
```

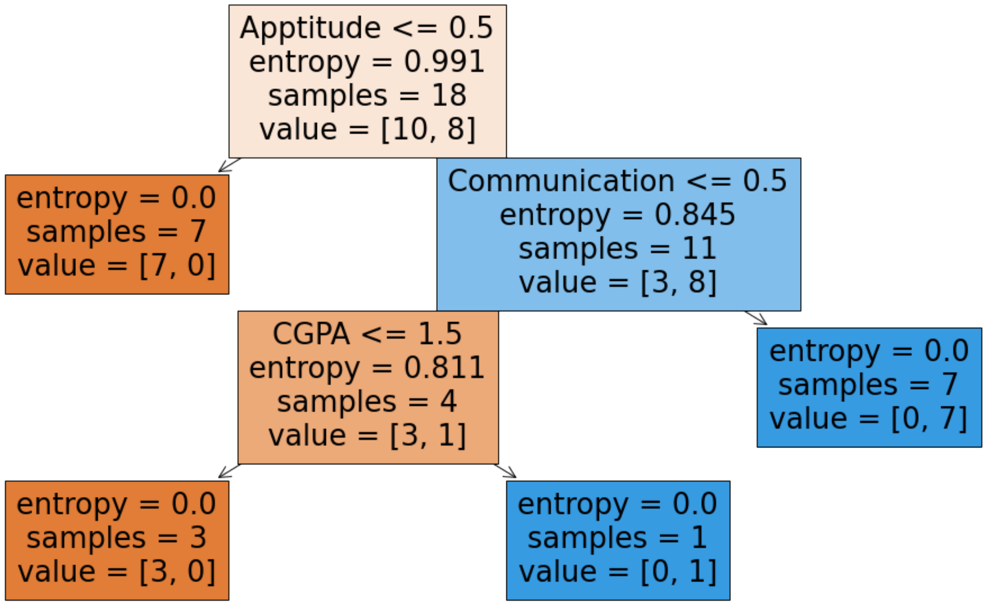
Out[31]:
```
1.0
```

In [32]:
```
data.head(1)
```

Out[32]:

	CGPA	Communication	Apptitude	Programming Skill	Job Offered
0	2	1	1	1	1

In [25]:
```
from sklearn.tree import plot_tree
plt.figure(figsize=(20,10))
plot_tree(dtree_entropy, feature_names=['CGPA','Communication','Apptitude','Programming
plt.show()
```



In [33]:
```python
import pandas as pd
df=pd.read_csv('diabetes.csv')
df.head()
```

Out[33]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Ou
0	6	148	72	35	0	33.6	0.627	50	
1	1	85	66	29	0	26.6	0.351	31	
2	8	183	64	0	0	23.3	0.672	32	
3	1	89	66	23	94	28.1	0.167	21	
4	0	137	40	35	168	43.1	2.288	33	

In [34]:
```python
x = df.drop('Outcome', axis = 1)
y = df['Outcome']
```

In [35]:
```python
x.shape
```

Out[35]: (768, 8)

In [36]:
```python
y.shape
```

Out[36]: (768,)

In [37]:
```python
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=1)
```

In [38]:
```python
print(x_train.shape)
print(x_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(614, 8)
(154, 8)
```

```
(614,)
(154,)
```

In [69]:
```python
from sklearn.tree import DecisionTreeClassifier
dtree_entropy = DecisionTreeClassifier(criterion = 'entropy')
model = dtree_entropy.fit(x_train,y_train)
dtree_entropy.get_depth()
```

Out[69]:
```
14
```

In [68]:
```python
# we can change the depth of tree by max_depth parameter
from sklearn.tree import DecisionTreeClassifier
dtree_entropy = DecisionTreeClassifier(criterion = 'entropy',max_depth=5)
model = dtree_entropy.fit(x_train,y_train)
dtree_entropy.get_depth()
```

Out[68]:
```
5
```

In [70]:
```python
feature=list(x.columns)
feature
```

Out[70]:
```
['Pregnancies',
 'Glucose',
 'BloodPressure',
 'SkinThickness',
 'Insulin',
 'BMI',
 'DiabetesPedigreeFunction',
 'Age']
```

In [71]:
```python
from sklearn import tree
text_representation = tree.export_text(dtree_entropy,feature_names=feature)
print(text_representation)
```

```
|--- Glucose <= 127.50
|   |--- BMI <= 26.45
|   |   |--- BMI <= 9.10
|   |   |   |--- Pregnancies <= 7.50
|   |   |   |   |--- class: 0
|   |   |   |--- Pregnancies >  7.50
|   |   |   |   |--- class: 1
|   |   |--- BMI >  9.10
|   |   |   |--- DiabetesPedigreeFunction <= 0.67
|   |   |   |   |--- class: 0
|   |   |   |--- DiabetesPedigreeFunction >  0.67
|   |   |   |   |--- DiabetesPedigreeFunction <= 0.71
|   |   |   |   |   |--- class: 1
|   |   |   |   |--- DiabetesPedigreeFunction >  0.71
|   |   |   |   |   |--- class: 0
|   |--- BMI >  26.45
|   |   |--- Age <= 28.50
|   |   |   |--- BMI <= 30.95
|   |   |   |   |--- Pregnancies <= 7.00
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- Pregnancies >  7.00
|   |   |   |   |   |--- class: 1
|   |   |   |--- BMI >  30.95
|   |   |   |   |--- Age <= 22.50
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- Age >  22.50
|   |   |   |   |   |--- BMI <= 45.40
|   |   |   |   |   |   |--- BMI <= 38.35
```

```
|   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.50
|   |   |   |   |   |   |   |   |--- BloodPressure <= 53.00
|   |   |   |   |   |   |   |   |   |--- SkinThickness <= 39.50
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- SkinThickness >  39.50
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- BloodPressure >  53.00
|   |   |   |   |   |   |   |   |   |--- BloodPressure <= 73.00
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- BloodPressure >  73.00
|   |   |   |   |   |   |   |   |   |   |--- Insulin <= 36.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- Insulin >  36.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.50
|   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.56
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.56
|   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.65
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.65
|   |   |   |   |   |   |   |   |   |   |--- Insulin <= 63.00
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- Insulin >  63.00
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |--- BMI >  38.35
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- BMI >  45.40
|   |   |   |   |   |   |--- class: 1
|   |   |--- Age >  28.50
|   |   |   |--- Glucose <= 89.50
|   |   |   |   |--- Pregnancies <= 11.50
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- Pregnancies >  11.50
|   |   |   |   |   |--- SkinThickness <= 15.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- SkinThickness >  15.50
|   |   |   |   |   |   |--- class: 1
|   |   |   |--- Glucose >  89.50
|   |   |   |   |--- DiabetesPedigreeFunction <= 0.20
|   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.18
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- DiabetesPedigreeFunction >  0.18
|   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.19
|   |   |   |   |   |   |   |--- Age <= 34.50
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- Age >  34.50
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.19
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- DiabetesPedigreeFunction >  0.20
|   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.61
|   |   |   |   |   |   |--- SkinThickness <= 27.50
|   |   |   |   |   |   |   |--- Age <= 54.50
|   |   |   |   |   |   |   |   |--- BloodPressure <= 83.00
|   |   |   |   |   |   |   |   |   |--- BMI <= 31.15
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- BMI >  31.15
|   |   |   |   |   |   |   |   |   |   |--- Age <= 34.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- Age >  34.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 4
|   |   |   |   |   |   |   |   |--- BloodPressure >  83.00
|   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.40
|   |   |   |   |   |   |   |   |   |   |--- class: 0
```

```
|   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.40
|   |   |   |   |   |   |   |   |   |   |--- BMI <= 31.45
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- BMI >  31.45
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- Age >  54.50
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- SkinThickness >  27.50
|   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.35
|   |   |   |   |   |   |   |   |--- BMI <= 34.85
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- BMI >  34.85
|   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.32
|   |   |   |   |   |   |   |   |   |   |--- BMI <= 41.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- BMI >  41.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.32
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.35
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- DiabetesPedigreeFunction >  0.61
|   |   |   |   |   |   |--- Pregnancies <= 7.50
|   |   |   |   |   |   |   |--- Age <= 30.50
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- Age >  30.50
|   |   |   |   |   |   |   |   |--- BMI <= 28.75
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- BMI >  28.75
|   |   |   |   |   |   |   |   |   |--- Pregnancies <= 3.00
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- Pregnancies >  3.00
|   |   |   |   |   |   |   |   |   |   |--- Age <= 34.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- Age >  34.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |--- Pregnancies >  7.50
|   |   |   |   |   |   |   |--- class: 1
|--- Glucose >  127.50
|   |--- Glucose <= 166.50
|   |   |--- BMI <= 29.95
|   |   |   |--- Pregnancies <= 1.50
|   |   |   |   |--- class: 0
|   |   |   |--- Pregnancies >  1.50
|   |   |   |   |--- BMI <= 23.45
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- BMI >  23.45
|   |   |   |   |   |--- Age <= 61.50
|   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.75
|   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.31
|   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.28
|   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.17
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.17
|   |   |   |   |   |   |   |   |   |   |--- BMI <= 26.70
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- BMI >  26.70
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 4
|   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.28
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.31
|   |   |   |   |   |   |   |   |--- SkinThickness <= 33.50
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- SkinThickness >  33.50
|   |   |   |   |   |   |   |   |   |--- BloodPressure <= 67.00
```

```
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- BloodPressure >  67.00
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.75
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- Age >  61.50
|   |   |   |   |   |   |--- class: 0
|   |   |--- BMI >  29.95
|   |   |   |--- BloodPressure <= 61.00
|   |   |   |   |--- Age <= 40.50
|   |   |   |   |   |--- class: 1
|   |   |   |   |--- Age >  40.50
|   |   |   |   |   |--- Pregnancies <= 7.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- Pregnancies >  7.50
|   |   |   |   |   |   |--- class: 1
|   |   |   |--- BloodPressure >  61.00
|   |   |   |   |--- Age <= 30.50
|   |   |   |   |   |--- Insulin <= 260.00
|   |   |   |   |   |   |--- Glucose <= 156.00
|   |   |   |   |   |   |   |--- BloodPressure <= 85.50
|   |   |   |   |   |   |   |   |--- BloodPressure <= 72.00
|   |   |   |   |   |   |   |   |   |--- Pregnancies <= 1.00
|   |   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.18
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.18
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- Pregnancies >  1.00
|   |   |   |   |   |   |   |   |   |   |--- Age <= 28.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- Age >  28.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- BloodPressure >  72.00
|   |   |   |   |   |   |   |   |   |--- Pregnancies <= 4.50
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- Pregnancies >  4.50
|   |   |   |   |   |   |   |   |   |   |--- BloodPressure <= 79.00
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- BloodPressure >  79.00
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- BloodPressure >  85.50
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- Glucose >  156.00
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- Insulin >  260.00
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- Age >  30.50
|   |   |   |   |   |--- BloodPressure <= 89.00
|   |   |   |   |   |   |--- BMI <= 34.05
|   |   |   |   |   |   |   |--- SkinThickness <= 28.50
|   |   |   |   |   |   |   |   |--- BMI <= 32.20
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- BMI >  32.20
|   |   |   |   |   |   |   |   |   |--- Glucose <= 144.50
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- Glucose >  144.50
|   |   |   |   |   |   |   |   |   |   |--- BMI <= 33.45
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- BMI >  33.45
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- SkinThickness >  28.50
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- BMI >  34.05
|   |   |   |   |   |   |   |--- Age <= 38.50
|   |   |   |   |   |   |   |   |--- class: 1
```

```
|   |   |   |   |   |   |   |--- Age >  38.50
|   |   |   |   |   |   |   |   |--- Age <= 39.50
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- Age >  39.50
|   |   |   |   |   |   |   |   |   |--- Pregnancies <= 7.50
|   |   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction <= 0.70
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- DiabetesPedigreeFunction >  0.70
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- Pregnancies >  7.50
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- BloodPressure >  89.00
|   |   |   |   |   |   |--- class: 1
|   |--- Glucose >  166.50
|   |   |--- Glucose <= 172.50
|   |   |   |--- class: 1
|   |   |--- Glucose >  172.50
|   |   |   |--- SkinThickness <= 32.50
|   |   |   |   |--- Glucose <= 194.50
|   |   |   |   |   |--- SkinThickness <= 30.50
|   |   |   |   |   |   |--- SkinThickness <= 25.50
|   |   |   |   |   |   |   |--- Glucose <= 190.50
|   |   |   |   |   |   |   |   |--- Glucose <= 183.50
|   |   |   |   |   |   |   |   |   |--- Age <= 27.50
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- Age >  27.50
|   |   |   |   |   |   |   |   |   |   |--- Pregnancies <= 5.00
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- Pregnancies >  5.00
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |--- Glucose >  183.50
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- Glucose >  190.50
|   |   |   |   |   |   |   |   |--- BMI <= 24.70
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- BMI >  24.70
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- SkinThickness >  25.50
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- SkinThickness >  30.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- Glucose >  194.50
|   |   |   |   |   |--- class: 1
|   |   |   |--- SkinThickness >  32.50
|   |   |   |   |--- Insulin <= 619.50
|   |   |   |   |   |--- class: 1
|   |   |   |   |--- Insulin >  619.50
|   |   |   |   |   |--- class: 0
```

In [72]:
```python
#Predictions
y_pred = dtree_entropy.predict(x_test)
y_pred
```

Out[72]:
```
array([1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0,
       1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1,
       0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1,
       1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0,
       1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1],
      dtype=int64)
```

In [73]:
```
diff=pd.DataFrame({"Actual":y_test,"Predicted":y_pred})
diff
```

Out[73]:

	Actual	Predicted
285	0	1
101	0	0
581	0	0
352	0	0
726	0	0
...	...	...
563	0	1
318	0	0
154	1	1
684	0	0
643	0	1

154 rows × 2 columns

In [74]:
```
#Metric Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
cm
```

Out[74]:
```
array([[76, 23],
       [24, 31]], dtype=int64)
```

In [75]:
```
TN = cm[0][0]
FP = cm[0][1]
FN = cm[1][0]
TP = cm[1][1]
print(TP, FN, TN, FP)
```

```
31 24 76 23
```

In [76]:
```
accuracy = (TP + TN) / (TP + FP + FN + TN)
accuracy
```

Out[76]:
```
0.6948051948051948
```

In [77]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Out[77]:
```
0.6948051948051948
```

In [78]:
```
sensitivity = TP / (TP + FN)
sensitivity
```

Out[78]:
```
0.5636363636363636
```

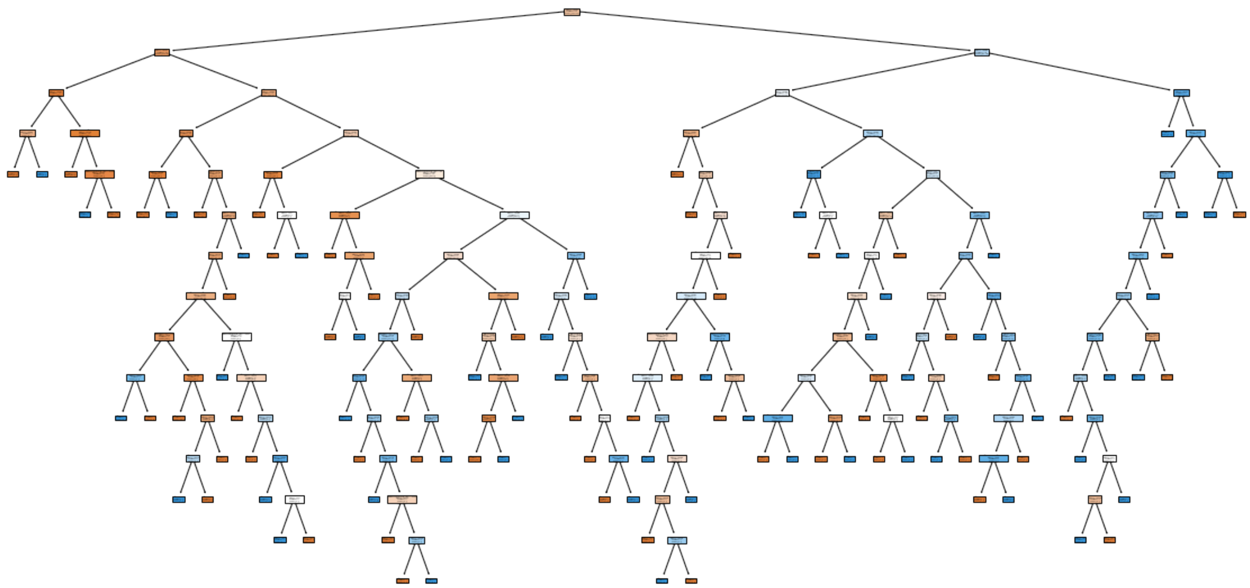
In [79]:

```
specificity=TN/(TN+FP)
specificity
```

Out[79]: 0.7676767676767676

In [80]:

```
import matplotlib.pyplot as plt
from sklearn.tree import plot_tree
plt.figure(figsize=(20,10))
plot_tree(dtree_entropy, feature_names=feature,  filled=True)
plt.show()
```



In [ ]: