

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: df = pd.read_csv('PimaIndiansDiabetes.csv')
df
```

```
Out[2]:
```

	TimesPregnant	GlucoseConcentration	BloodPrs	SkinThickness	Serum	BMI	DiabetesFunct	Age	Class
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
...
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

768 rows × 9 columns

```
In [3]: df.head()
```

```
Out[3]:
```

	TimesPregnant	GlucoseConcentration	BloodPrs	SkinThickness	Serum	BMI	DiabetesFunct	Age	Class
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

```
In [4]: df.dtypes
```

```
Out[4]: TimesPregnant      int64
GlucoseConcentration      int64
BloodPrs                  int64
SkinThickness             int64
Serum                     int64
BMI                       float64
DiabetesFunct             float64
Age                       int64
Class                     int64
dtype: object
```

```
In [5]: df.Class.unique()
```

```
Out[5]: array([1, 0], dtype=int64)
```

```
In [6]: df.Class.value_counts()
```

```
Out[6]: 0      500
1       268
Name: Class, dtype: int64
```

```
In [7]: df.shape
```

```
Out[7]: (768, 9)
```

```
In [8]: X = df.iloc[:,0:8]
X.head()
```

```
Out[8]:
```

	TimesPregnant	GlucoseConcentration	BloodPrs	SkinThickness	Serum	BMI	DiabetesFunct	Age
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 [9]: y = df.iloc[:,8]
y.head()
```

```
Out[9]: 0    1
1    0
2    1
3    0
4    1
Name: Class, dtype: int64
```

```
In [10]: from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.20)
```

```
In [11]: from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier(criterion='entropy')
model.fit(X_train,y_train)
```

```
Out[11]: DecisionTreeClassifier(criterion='entropy')
```

```
In [12]: y_pred = model.predict(X_test)
```

```
In [13]: y_pred
```

```
Out[13]: array([0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0,
0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0,
0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1,
0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1,
0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1,
0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0,
0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1],
dtype=int64)
```

```
In [14]: model.score(X_train,y_train)*100
```

```
Out[14]: 100.0
```

```
In [15]: model.score(X_test,y_test)*100
```

```
Out[15]: 70.77922077922078
```

```
In [16]: from sklearn.metrics import confusion_matrix
```

```
In [17]: confusion_matrix(y_test,y_pred)
```

```
Out[17]: array([[82, 27],
               [18, 27]], dtype=int64)
```

```
In [18]: from sklearn.metrics import accuracy_score
```

```
In [19]: accuracy_score(y_test,y_pred)*100
```

```
Out[19]: 70.77922077922078
```

```
In [20]: from sklearn import tree
text_representation = tree.export_text(model)
print(text_representation)
```

```

|--- feature_1 <= 127.50
|   |--- feature_7 <= 28.50
|   |   |--- feature_5 <= 30.95
|   |   |   |--- feature_0 <= 7.50
|   |   |   |   |--- feature_6 <= 0.67
|   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- feature_6 > 0.67
|   |   |   |   |   |   |--- feature_6 <= 0.69
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- feature_6 > 0.69
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- feature_0 > 7.50
|   |   |   |   |   |   |--- class: 1
|   |   |--- feature_5 > 30.95
|   |   |   |--- feature_5 <= 45.40
|   |   |   |   |--- feature_6 <= 0.50
|   |   |   |   |   |--- feature_2 <= 53.00
|   |   |   |   |   |   |--- feature_5 <= 32.45
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- feature_5 > 32.45
|   |   |   |   |   |   |   |   |--- feature_3 <= 10.50
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- feature_3 > 10.50
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- feature_2 > 53.00
|   |   |   |   |   |   |--- feature_5 <= 33.55
|   |   |   |   |   |   |   |--- feature_2 <= 72.00
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- feature_2 > 72.00
|   |   |   |   |   |   |   |   |   |--- feature_1 <= 100.50
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- feature_1 > 100.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- feature_5 > 33.55
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |--- feature_6 > 0.50
|   |   |   |--- feature_5 <= 32.20
|   |   |   |   |--- class: 0
|   |   |   |--- feature_5 > 32.20
|   |   |   |   |--- feature_5 <= 38.80
|   |   |   |   |   |--- feature_5 <= 37.60
|   |   |   |   |   |   |--- feature_5 <= 34.30
|   |   |   |   |   |   |   |--- feature_3 <= 30.50
|   |   |   |   |   |   |   |   |--- feature_5 <= 32.35
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- feature_5 > 32.35
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- feature_3 > 30.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- feature_5 > 34.30
|   |   |   |   |   |   |   |   |--- feature_0 <= 3.50
|   |   |   |   |   |   |   |   |   |--- feature_6 <= 0.74
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- feature_6 > 0.74
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- feature_0 > 3.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- feature_5 > 37.60
|   |   |   |   |   |   |--- class: 1
```

```

| | | | | --- feature_5 > 38.80
| | | | | |--- feature_2 <= 79.00
| | | | | | |--- class: 0
| | | | | |--- feature_2 > 79.00
| | | | | | |--- feature_2 <= 84.00
| | | | | | | |--- class: 1
| | | | | | |--- feature_2 > 84.00
| | | | | | | |--- class: 0
| | | | | --- feature_5 > 45.40
| | | | | |--- class: 1
--- feature_7 > 28.50
|--- feature_5 <= 26.35
| |--- feature_5 <= 9.80
| | |--- class: 1
| |--- feature_5 > 9.80
| | |--- class: 0
|--- feature_5 > 26.35
| |--- feature_2 <= 85.50
| | |--- feature_6 <= 0.22
| | | |--- feature_5 <= 32.40
| | | | |--- feature_0 <= 7.00
| | | | | |--- feature_6 <= 0.16
| | | | | | |--- class: 0
| | | | | |--- feature_6 > 0.16
| | | | | | |--- feature_5 <= 29.90
| | | | | | | |--- class: 0
| | | | | | |--- feature_5 > 29.90
| | | | | | | |--- class: 1
| | | | | |--- feature_0 > 7.00
| | | | | | |--- class: 1
| | | | |--- feature_5 > 32.40
| | | | | |--- class: 0
| | |--- feature_6 > 0.22
| | | |--- feature_1 <= 99.50
| | | | |--- feature_5 <= 34.65
| | | | | |--- feature_0 <= 11.00
| | | | | | |--- class: 0
| | | | | |--- feature_0 > 11.00
| | | | | | |--- class: 1
| | | | |--- feature_5 > 34.65
| | | | | |--- feature_3 <= 12.50
| | | | | | |--- class: 0
| | | | | |--- feature_3 > 12.50
| | | | | | |--- feature_3 <= 35.50
| | | | | | | |--- class: 1
| | | | | | |--- feature_3 > 35.50
| | | | | | | |--- feature_1 <= 40.50
| | | | | | | | |--- class: 1
| | | | | | | |--- feature_1 > 40.50
| | | | | | | | |--- class: 0
| | | | |--- feature_1 > 99.50
| | | | | |--- feature_1 <= 105.50
| | | | | | |--- class: 1
| | | | | |--- feature_1 > 105.50
| | | | | | |--- feature_7 <= 55.00
| | | | | | | |--- feature_3 <= 25.50
| | | | | | | | |--- feature_5 <= 30.85
| | | | | | | | | |--- feature_0 <= 0.50
| | | | | | | | | | |--- truncated branch of depth 2
| | | | | | | | | |--- feature_0 > 0.50
| | | | | | | | | | | |--- class: 1
| | | | | | | | | |--- feature_5 > 30.85
| | | | | | | | | |--- feature_1 <= 107.00
| | | | | | | | | | |--- class: 0
| | | | | | | | | |--- feature_1 > 107.00
| | | | | | | | | | |--- truncated branch of depth 5
| | | | | | |--- feature_3 > 25.50
| | | | | | | |--- feature_2 <= 79.00
| | | | | | | | |--- feature_3 <= 39.50
| | | | | | | | | |--- truncated branch of depth 4
| | | | | | | | |--- feature_3 > 39.50
| | | | | | | | | |--- class: 0
| | | | | | | | |--- feature_2 > 79.00
| | | | | | | | |--- class: 1
| | | | | |--- feature_7 > 55.00
| | | | | | |--- class: 0
|--- feature_2 > 85.50
| |--- feature_3 <= 26.50
| | |--- class: 0
| |--- feature_3 > 26.50
| | |--- feature_1 <= 104.00
| | | |--- class: 0
| | |--- feature_1 > 104.00

```

```

|--- feature_5 <= 41.20
|--- feature_4 <= 52.50
|--- class: 1
|--- feature_4 > 52.50
|--- class: 0
|--- feature_5 > 41.20
|--- class: 1
|--- feature_1 > 127.50
|--- feature_5 <= 29.95
|--- feature_7 <= 22.50
|--- class: 0
|--- feature_7 > 22.50
|--- feature_7 <= 60.50
|--- feature_1 <= 160.00
|--- feature_0 <= 10.50
|--- feature_4 <= 127.50
|--- feature_5 <= 29.60
|--- feature_6 <= 0.31
|--- feature_1 <= 133.50
|--- class: 0
|--- feature_1 > 133.50
|--- feature_3 <= 13.50
|--- truncated branch of depth 4
|--- feature_3 > 13.50
|--- class: 0
|--- feature_6 > 0.31
|--- feature_6 <= 0.99
|--- class: 1
|--- feature_6 > 0.99
|--- class: 0
|--- feature_5 > 29.60
|--- class: 0
|--- feature_4 > 127.50
|--- class: 0
|--- feature_0 > 10.50
|--- class: 1
|--- feature_1 > 160.00
|--- feature_2 <= 81.00
|--- feature_4 <= 275.00
|--- class: 1
|--- feature_4 > 275.00
|--- class: 0
|--- feature_2 > 81.00
|--- class: 0
|--- feature_7 > 60.50
|--- class: 0
|--- feature_5 > 29.95
|--- feature_1 <= 158.50
|--- feature_7 <= 42.50
|--- feature_2 <= 61.00
|--- feature_7 <= 40.50
|--- class: 1
|--- feature_7 > 40.50
|--- class: 0
|--- feature_2 > 61.00
|--- feature_4 <= 260.00
|--- feature_6 <= 0.72
|--- feature_5 <= 34.10
|--- feature_3 <= 32.00
|--- class: 0
|--- feature_3 > 32.00
|--- feature_1 <= 138.00
|--- class: 0
|--- feature_1 > 138.00
|--- class: 1
|--- feature_5 > 34.10
|--- feature_6 <= 0.67
|--- feature_2 <= 75.50
|--- class: 1
|--- feature_2 > 75.50
|--- feature_2 <= 85.50
|--- truncated branch of depth 4
|--- feature_2 > 85.50
|--- class: 1
|--- feature_6 > 0.67
|--- class: 0
|--- feature_6 > 0.72
|--- feature_0 <= 11.00
|--- class: 1
|--- feature_0 > 11.00
|--- class: 0
|--- feature_4 > 260.00
|--- class: 0

```

```

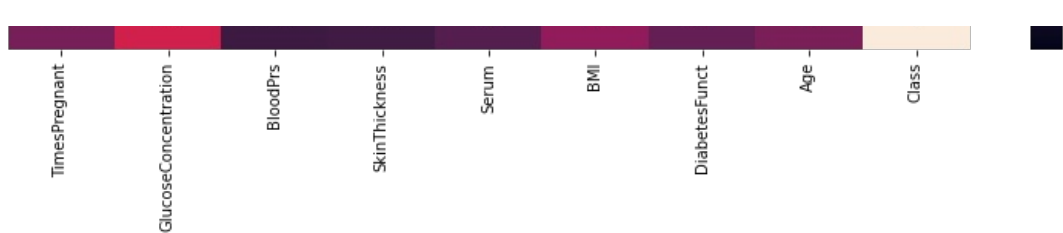
|--- feature_7 > 42.50
|   |--- feature_0 <= 3.50
|   |   |--- feature_4 <= 70.00
|   |   |   |--- feature_1 <= 137.00
|   |   |   |   |--- class: 1
|   |   |   |   |--- feature_1 > 137.00
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- feature_4 > 70.00
|   |   |   |   |--- class: 1
|   |   |--- feature_0 > 3.50
|   |   |--- class: 1
|--- feature_1 > 158.50
|   |--- feature_7 <= 44.00
|   |   |--- feature_6 <= 0.43
|   |   |   |--- feature_6 <= 0.42
|   |   |   |   |--- feature_5 <= 35.30
|   |   |   |   |   |--- feature_1 <= 177.00
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- feature_1 > 177.00
|   |   |   |   |   |   |--- feature_6 <= 0.31
|   |   |   |   |   |   |   |--- feature_1 <= 192.00
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- feature_1 > 192.00
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- feature_6 > 0.31
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- feature_5 > 35.30
|   |   |   |   |--- class: 1
|   |   |   |--- feature_6 > 0.42
|   |   |   |--- class: 0
|   |   |--- feature_6 > 0.43
|   |   |--- class: 1
|   |--- feature_7 > 44.00
|   |   |--- feature_0 <= 7.50
|   |   |   |--- feature_6 <= 1.16
|   |   |   |   |--- feature_5 <= 34.60
|   |   |   |   |   |--- feature_5 <= 31.65
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- feature_5 > 31.65
|   |   |   |   |   |   |--- feature_6 <= 0.89
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- feature_6 > 0.89
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- feature_5 > 34.60
|   |   |   |   |   |--- class: 1
|   |   |   |--- feature_6 > 1.16
|   |   |   |--- class: 0
|   |   |--- feature_0 > 7.50
|   |   |--- class: 1

```

```
In [21]: plt.figure(figsize=(14,8))
sns.heatmap(df.corr(),annot=True)
```

Out[21]: <AxesSubplot:>





correlation between Blood pressure, Glucose and outcome

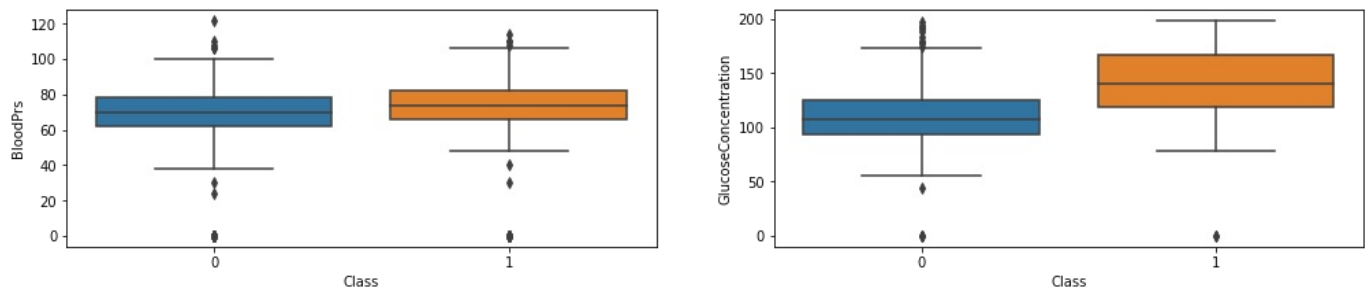
```
In [22]: plt.figure(figsize=(17,7))
plt.subplot(2,2,1)
sns.boxplot(df['Class'],df['BloodPrs'])
plt.subplot(2,2,2)
sns.boxplot(df['Class'],df['GlucoseConcentration'])
plt.show()
```

C:\Users\Nilesh koli\anaconda3\anaconda\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(

C:\Users\Nilesh koli\anaconda3\anaconda\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(

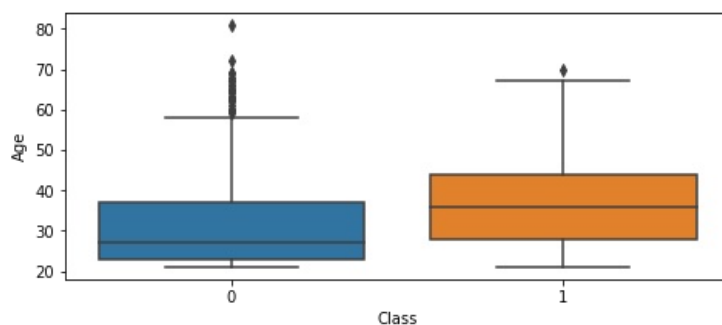


```
In [23]: plt.figure(figsize=(17,7))
plt.subplot(2,2,1)
sns.boxplot(df['Class'],df['Age'])
```

C:\Users\Nilesh koli\anaconda3\anaconda\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(

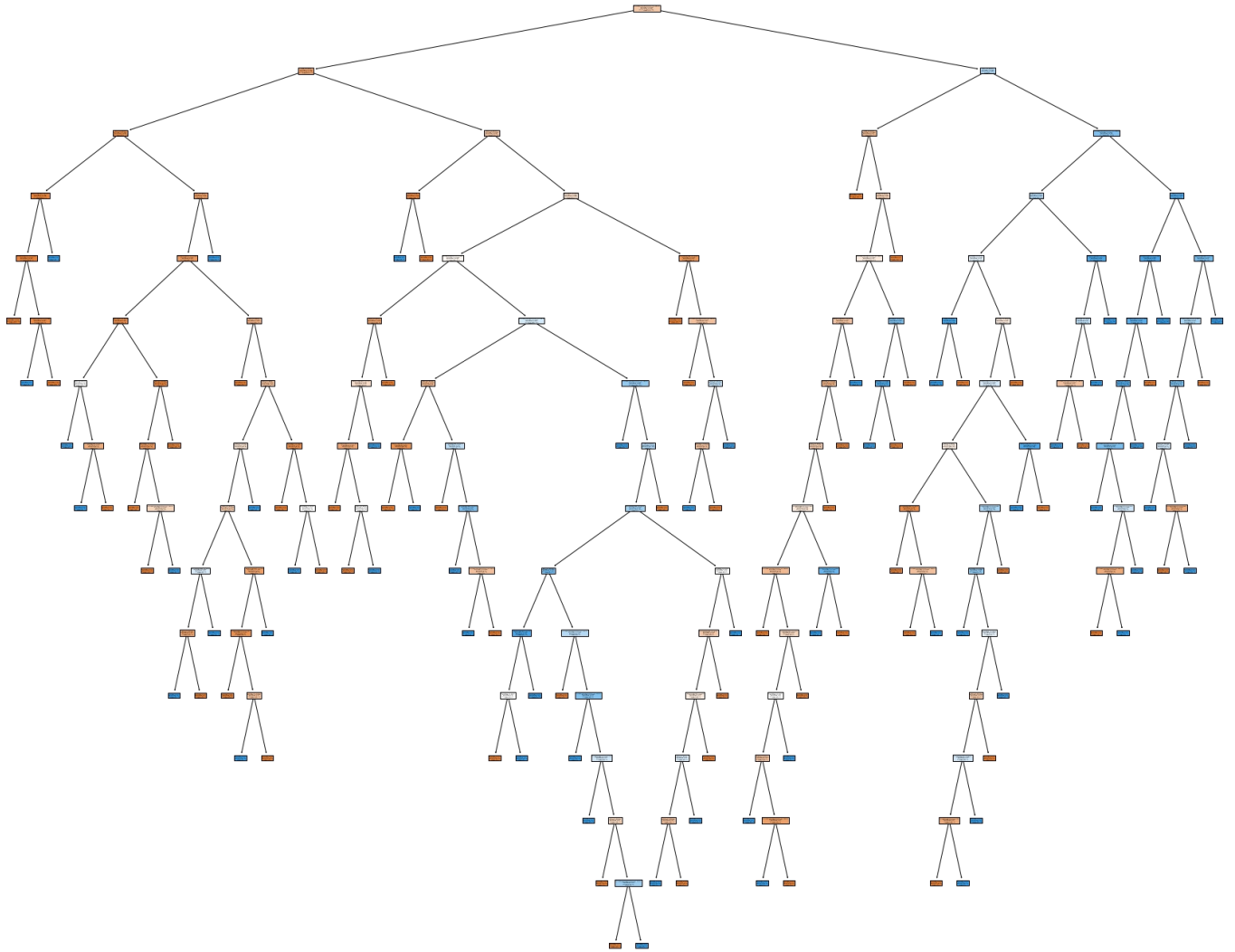
Out[23]: <AxesSubplot:xlabel='Class', ylabel='Age'>



```
In [24]: with open("decistion_tree.log", "w") as fout:
fout.write(text_representation)
```

```
In [25]: import matplotlib.pyplot as plt
fig = plt.figure(figsize=(30,25))
_ = tree.plot_tree(model,
feature_names=df.columns,
```

```
class_names='Class',  
filled=True)
```



In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js