jupyter dt (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

Run Code

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

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

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

Out[3]:

	Kyphosis	Age	Number	Start
0	absent	71	3	5
1	absent	158	3	14
2	present	128	4	5
3	absent	2	5	1
4	absent	1	4	15

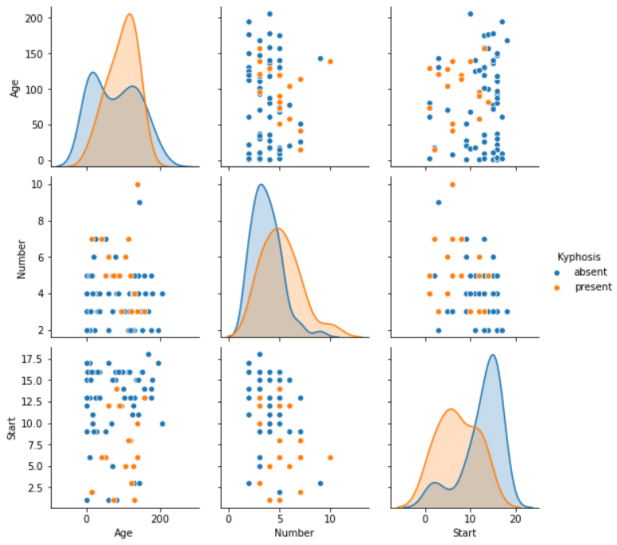
In [4]:
```
#Age is in months
#Number is number of vertabrae involved in operation
#Start is the number of the first (top most) vertabrae operated on
```

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 81 entries, 0 to 80
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Kyphosis  81 non-null     object
 1   Age       81 non-null     int64
 2   Number    81 non-null     int64
 3   Start     81 non-null     int64
dtypes: int64(3), object(1)
memory usage: 2.7+ KB
```

In [6]:
```
sns.pairplot(df,hue='Kyphosis')
```

Out[6]:
```
<seaborn.axisgrid.PairGrid at 0x207fb418ac0>
```



In [7]:
```
from sklearn.model_selection import train_test_split
```

In [8]:
```
X = df.drop('Kyphosis',axis=1)
```

In [9]:
```
y = df['Kyphosis']
```

In [10]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)
```

In [11]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [12]:
```
rfc = RandomForestClassifier(n_estimators=200)
```

In [13]:
```
rfc.fit(X_train,y_train)
```

Out[13]:
```
RandomForestClassifier(n_estimators=200)
```

In [14]:
```
rfc_pred = rfc.predict(X_test)
```

In [15]:
```
df['Kyphosis'].value_counts()
```

Out[15]:
```
absent     64
present    17
Name: Kyphosis, dtype: int64
```

In [ ]:
```
```