```
%cd /content/drive/MyDrive/Colab Notebooks/Diabetes_Data_Set
%ls
```

```
/content/drive/MyDrive/Colab Notebooks/Diabetes_Data_Set
diabetes_data.csv  diabetes_data_upload.csv
```

```
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
%matplotlib inline
mpl.rcParams['figure.dpi'] = 300
plt.rcParams['figure.subplot.hspace'] = 0.4
plt.rcParams['figure.subplot.wspace'] = 0.4

from sklearn.preprocessing import StandardScaler

from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_validation

from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import ExtraTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.ensemble import AdaBoostClassifier
from sklearn.ensemble import GradientBoostingClassifier
```

```
diabetes_data = pd.read_csv('diabetes_data.csv')
diabetes_data.shape
```

```
(520, 17)
```

```
diabetes_data.head()
```

	Age	Gender	Polyuria	Polydipsia	Sudden_weight_loss	Weakness	Polyphagia	Genital_thrush	Visual_blurring	Itching	Irritability	Delayed_healing	Partial_paresis	Muscle_stiffness	Alopecia	Obesity	class
0	40	Male	No	Yes	No	Yes	No	No	No	No	No	No	No	No	Yes	No	Positive
1	58	Male	No	No	No	No	No	No	Yes	No	No	No	Yes	No	No	No	Positive
2	41	Male	Yes	No	No	Yes	Yes	No	No	Yes	No	Yes	No	Yes	No	No	Positive
3	45	Male	No	No	No	No	No	Yes	No	Yes	No	Yes	No	No	No	No	Positive
4	60	Male	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Positive

```
diabetes_data.info()
```

```
<class>
RangeIndex: 520 entries, 0 to 519
Data columns (total 17 columns):
 0   Age                 520 non-null    int64
 1   Gender              520 non-null    object
 2   Polyuria            520 non-null    object
 3   Polydipsia          520 non-null    object
 4   Sudden_weight_loss  520 non-null    object
 5   Weakness            520 non-null    object
 6   Polyphagia          520 non-null    object
 7   Genital_thrush      520 non-null    object
 8   Visual_blurring     520 non-null    object
 9   Itching             520 non-null    object
 10  Irritability        520 non-null    object
 11  Delayed_healing     520 non-null    object
 12  Partial_paresis     520 non-null    object
 13  Muscle_stiffness    520 non-null    object
 14  Alopecia            520 non-null    object
 15  Obesity             520 non-null    object
 16  class               520 non-null    object
dtypes: int64, 1, object: 16
memory usage: 69.1 KB
```

```
diabetes_data.describe
```

```
<bound_method_DataFrame_describe_of_
Age Gender Polyuria Polydipsia Sudden_weight_loss Weakness Polyphagia Genital_thrush Visual_blurring Itching Irritability Delayed_healing Partial_paresis Muscle_stiffness Alopecia Obesity class
0   40   Male       No      Yes                 No      Yes       No            No             No      No          No             No            No             No      Yes     No  Positive
1   58   Male       No       No                 No       No       No            No            Yes      No          No             No           Yes             No       No     No  Positive
2   41   Male      Yes       No                 No      Yes      Yes            No             No     Yes          No            Yes            No            Yes       No     No  Positive
3   45   Male       No       No                 No       No       No           Yes             No     Yes          No            Yes            No             No       No     No  Positive
4   60   Male      Yes      Yes                Yes      Yes      Yes            No            Yes     Yes         Yes             No           Yes            Yes      Yes    Yes  Positive
...
516 48 Female       No       No                 No       No       No            No             No      No          No             No            No             No       No     No  Negative
517 25 Female      Yes      Yes                Yes      Yes      Yes            No             No     Yes         Yes            Yes           Yes            Yes      Yes    Yes  Positive
518 36   Male       No       No                 No       No       No            No             No      No          No             No            No             No       No     No  Negative
519 44   Male       No       No                 No       No       No            No             No      No          No             No            No             No       No     No  Negative
520 62   Male       No       No                 No       No       No            No             No      No          No             No            No             No       No     No  Negative
[520 rows x 17 columns]>
```

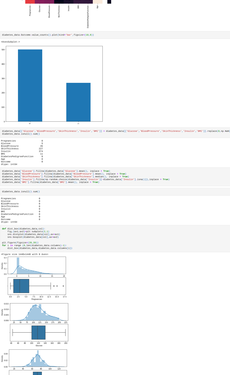
```
diabetes_data.describe()
```

	Age
count	520.000000
mean	48.028846
std	12.151466
min	16.000000
25%	39.000000
50%	47.500000
75%	57.000000
max	90.000000

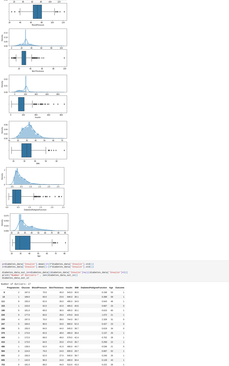
```
diabetes_data.hist(figsize=(20,20))
```

```
sns.pairplot(diabetes_data,hue='class')
```

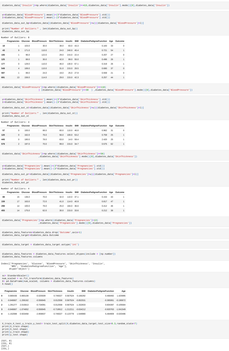
```
sns.pairplot(diabetes_data,hue='Polyuria')
```



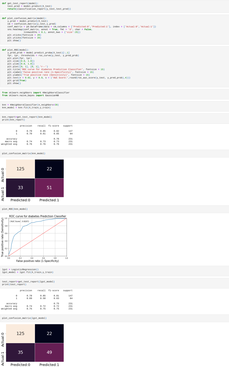

```
plt.figure(figsize=(8,6))
sns.heatmap(diabetes_data.corr(), annot=True)
```




```
diabetes_data.boxplot(column=['Age'],grid=False)
```




```
sns.countplot(diabetes_data['class'])
```




```
diabetes_data['Gender'] = diabetes_data['Gender'].replace('Male',1)
diabetes_data['Gender'] = diabetes_data['Gender'].replace('Female',0)
diabetes_data['Polyuria'] = diabetes_data['Polyuria'].replace('Yes',1)
diabetes_data['Polyuria'] = diabetes_data['Polyuria'].replace('No',0)
diabetes_data['Polydipsia'] = diabetes_data['Polydipsia'].replace('Yes',1)
diabetes_data['Polydipsia'] = diabetes_data['Polydipsia'].replace('No',0)
diabetes_data['Sudden_weight_loss'] = diabetes_data['Sudden_weight_loss'].replace('Yes',1)
diabetes_data['Sudden_weight_loss'] = diabetes_data['Sudden_weight_loss'].replace('No',0)
diabetes_data['Weakness'] = diabetes_data['Weakness'].replace('Yes',1)
diabetes_data['Weakness'] = diabetes_data['Weakness'].replace('No',0)
diabetes_data['Polyphagia'] = diabetes_data['Polyphagia'].replace('Yes',1)
diabetes_data['Polyphagia'] = diabetes_data['Polyphagia'].replace('No',0)
diabetes_data['Genital_thrush'] = diabetes_data['Genital_thrush'].replace('Yes',1)
diabetes_data['Genital_thrush'] = diabetes_data['Genital_thrush'].replace('No',0)
diabetes_data['Visual_blurring'] = diabetes_data['Visual_blurring'].replace('Yes',1)
diabetes_data['Visual_blurring'] = diabetes_data['Visual_blurring'].replace('No',0)
diabetes_data['Itching'] = diabetes_data['Itching'].replace('Yes',1)
diabetes_data['Itching'] = diabetes_data['Itching'].replace('No',0)
diabetes_data['Irritability'] = diabetes_data['Irritability'].replace('Yes',1)
diabetes_data['Irritability'] = diabetes_data['Irritability'].replace('No',0)
diabetes_data['Delayed_healing'] = diabetes_data['Delayed_healing'].replace('Yes',1)
diabetes_data['Delayed_healing'] = diabetes_data['Delayed_healing'].replace('No',0)
diabetes_data['Partial_paresis'] = diabetes_data['Partial_paresis'].replace('Yes',1)
diabetes_data['Partial_paresis'] = diabetes_data['Partial_paresis'].replace('No',0)
diabetes_data['Muscle_stiffness'] = diabetes_data['Muscle_stiffness'].replace('Yes',1)
diabetes_data['Muscle_stiffness'] = diabetes_data['Muscle_stiffness'].replace('No',0)
diabetes_data['Alopecia'] = diabetes_data['Alopecia'].replace('Yes',1)
diabetes_data['Alopecia'] = diabetes_data['Alopecia'].replace('No',0)
diabetes_data['Obesity'] = diabetes_data['Obesity'].replace('Yes',1)
diabetes_data['Obesity'] = diabetes_data['Obesity'].replace('No',0)
diabetes_data['class'] = diabetes_data['class'].replace('Positive',1)
diabetes_data['class'] = diabetes_data['class'].replace('Negative',0)
```

```
diabetes_data.sample(5)
```

	Age	Gender	Polyuria	Polydipsia	Sudden_weight_loss	Weakness	Polyphagia	Genital_thrush	Visual_blurring	Itching	Irritability	Delayed_healing	Partial_paresis	Muscle_stiffness	Alopecia	Obesity	class
250	41	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
267	38	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
229	42	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1
128	40	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
232	38	1	1	1	0	1	1	0	1	1	1	0	1	1	1	1	1

```
diabetes_data.sample(5)
```

	Age	Gender	Polyuria	Polydipsia	Sudden_weight_loss	Weakness	Polyphagia	Genital_thrush	Visual_blurring	Itching	Irritability	Delayed_healing	Partial_paresis	Muscle_stiffness	Alopecia	Obesity	class
283	42	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
302	55	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
198	35	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1
203	40	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1
174	40	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1

```
diabetes_data.sample(5)
```

	Age	Gender	Polyuria	Polydipsia	Sudden_weight_loss	Weakness	Polyphagia	Genital_thrush	Visual_blurring	Itching	Irritability	Delayed_healing	Partial_paresis	Muscle_stiffness	Alopecia	Obesity	class
280	40	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
287	35	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
188	35	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1
189	35	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1
190	35	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1

```
diabetes_data.hist(figsize=(20,20))
```

```
%cd /content/drive/MyDrive/Colab Notebooks/Diabetes_Data_Set
%ls
```

```
/content/drive/MyDrive/Colab Notebooks/Diabetes_Data_Set
diabetes_data.csv  diabetes_data_upload.csv
```

```
sns.boxplot(diabetes_data['Age'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Gender'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Polyuria'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Polydipsia'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Sudden_weight_loss'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Weakness'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Polyphagia'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Genital_thrush'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Visual_blurring'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Itching'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Irritability'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Delayed_healing'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Partial_paresis'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Muscle_stiffness'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Alopecia'],diabetes_data['class'])
```



```
sns.boxplot(diabetes_data['Obesity'],diabetes_data['class'])
```



```
sns.countplot(diabetes_data['Gender'],hue='class')
```



```
sns.countplot(diabetes_data['Polyuria'],hue='class')
```



```
sns.countplot(diabetes_data['Polydipsia'],hue='class')
```



```
sns.countplot(diabetes_data['Sudden_weight_loss'],hue='class')
```



```
sns.countplot(diabetes_data['Weakness'],hue='class')
```



```
sns.countplot(diabetes_data['Polyphagia'],hue='class')
```



```
sns.countplot(diabetes_data['Genital_thrush'],hue='class')
```



```
sns.countplot(diabetes_data['Visual_blurring'],hue='class')
```



```
sns.countplot(diabetes_data['Itching'],hue='class')
```



```
sns.countplot(diabetes_data['Irritability'],hue='class')
```



```
sns.countplot(diabetes_data['Delayed_healing'],hue='class')
```



```
sns.countplot(diabetes_data['Partial_paresis'],hue='class')
```



```
sns.countplot(diabetes_data['Muscle_stiffness'],hue='class')
```



```
sns.countplot(diabetes_data['Alopecia'],hue='class')
```



```
sns.countplot(diabetes_data['Obesity'],hue='class')
```



```
diabetes_data['Age'].sample(5)
```

```
Age
250    41
267    38
229    42
128    40
232    38
dtype: int64
```

```
diabetes_data['Age'].sample(5)
```

```
Age
280    40
287    35
188    35
189    35
190    35
dtype: int64
```

```
diabetes_data['Age'].hist(figsize=(20,20))
```



```
diabetes_data['Gender'].hist(figsize=(20,20))
```



```
diabetes_data['Polyuria'].hist(figsize=(20,20))
```



```
diabetes_data['Polydipsia'].hist(figsize=(20,20))
```



```
diabetes_data['Sudden_weight_loss'].hist(figsize=(20,20))
```



```
diabetes_data['Weakness'].hist(figsize=(20,20))
```



```
diabetes_data['Polyphagia'].hist(figsize=(20,20))
```



```
diabetes_data['Genital_thrush'].hist(figsize=(20,20))
```



```
diabetes_data['Visual_blurring'].hist(figsize=(20,20))
```



```
diabetes_data['Itching'].hist(figsize=(20,20))
```



```
diabetes_data['Irritability'].hist(figsize=(20,20))
```



```
diabetes_data['Delayed_healing'].hist(figsize=(20,20))
```



```
diabetes_data['Partial_paresis'].hist(figsize=(20,20))
```



```
diabetes_data['Muscle_stiffness'].hist(figsize=(20,20))
```



```
diabetes_data['Alopecia'].hist(figsize=(20,20))
```



```
diabetes_data['Obesity'].hist(figsize=(20,20))
```



```
diabetes_data['class'].hist(figsize=(20,20))
```



```
diabetes_data.describe()
```

	Age	Gender	Polyuria	Polydipsia	Sudden_weight_loss	Weakness	Polyphagia	Genital_thrush	Visual_blurring	Itching	Irritability	Delayed_healing	Partial_paresis	Muscle_stiffness	Alopecia	Obesity	class
count	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000	520.000000
mean	48.028846	0.630769	0.503846	0.448077	0.424038	0.423077	0.455769	0.242308	0.446154	0.419231	0.359615	0.417308	0.430769	0.384615	0.269231	0.169231	0.615385
std	12.151466	0.483160	0.500524	0.497776	0.494604	0.494614	0.498503	0.428892	0.497828	0.493889	0.480381	0.493589	0.495561	0.486911	0.444030	0.375083	0.486911
min	16.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	39.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	47.500000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
75%	57.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	1.000000
max	90.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

```
X = data[['Age','Gender','Polyuria','Polydipsia','Sudden_weight_loss','Weakness','Polyphagia','Genital_thrush','Visual_blurring','Itching','Irritability','Delayed_healing','Partial_paresis','Muscle_stiffness','Alopecia','Obesity']]
Y = data['class']
print(X.shape)
print(Y.shape)
```

```
(520, 16)
(520,)
```

```
X = np.array(X)
Y = np.array(Y)
```

```
# Splitting the data into training and testing sets
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.3, random_state=7)
print(X_train.shape, X_test.shape, Y_train.shape, Y_test.shape)
```

```
(364, 16) (156, 16) (364,) (156,)
```

```
# Feature Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
print(X_train[:5])
```

```
[[-0.39888689 -1.30558359 -1.00208468 -0.89442231 -0.84845096 -0.84845096
  -0.91766294 -0.56568542 -0.89442231 -0.84845096 -0.72111025 -0.84845096
  -0.91766294 -0.77459667 -0.60368832 -0.44271887]
 [ 1.26025954  0.76623377  0.99791273  1.11803399  1.17851130  1.17851130
   1.08972056  1.76776695  1.11803399  1.17851130  1.38675049  1.17851130
   1.08972056  1.29099445  1.65647118  2.25943966]
 [-0.57143065 -1.30558359 -1.00208468 -0.89442231 -0.84845096 -0.84845096
  -0.91766294 -0.56568542 -0.89442231 -0.84845096 -0.72111025 -0.84845096
  -0.91766294 -0.77459667 -0.60368832 -0.44271887
```