

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from time import time
from IPython.display import display

data = pd.read_csv('ASD.csv')
display(data.head(n=5))
```

	id	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	...	gender	ethnicity	jundi
0	1	1	1	1	1	0	0	1	1	0	...	f	White-European	
1	2	1	1	0	1	0	0	0	1	0	...	m	Latino	
2	3	1	1	0	1	1	0	1	1	1	...	m	Latino	y
3	4	1	1	0	1	0	0	1	1	0	...	f	White-European	
4	5	1	0	0	0	0	0	0	1	0	...	f	?	

5 rows × 22 columns

```
In [2]: # Total number of records
n_records = len(data.index)

print("Total number of records:",n_records)

# TODO: Number of records where individual's with ASD
n_asd_yes = len(data[data['Class/ASD'] == 'YES'])

print("Individuals diagonised with ASD:",n_asd_yes)

# TODO: Number of records where individual's with no ASD
n_asd_no = len(data[data['Class/ASD'] == 'NO'])

print("Individuals not diagonised with ASD:",n_asd_no)

# TODO: Percentage of individuals whose are with ASD
yes_percent = float(n_asd_yes) / n_records *100

print("Percentage of individuals diagonised with ASD:",yes_percent)
```

Total number of records: 704
Individuals diagonised with ASD: 189
Individuals not diagonised with ASD: 515
Percentage of individuals diagonised with ASD: 26.84659090909091

```
In [3]: #preprocessing of data by changing '?' to NaN value
asd_data = pd.read_csv('ASD.csv', na_values=['?'])
asd_data.head(n=5)
```

	id	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	...	gender	ethnicity	jun
0	1	1	1	1	1	0	0	1	1	0	...	f	White-European	
1	2	1	1	0	1	0	0	0	1	0	...	m	Latino	
2	3	1	1	0	1	1	0	1	1	1	...	m	Latino	
3	4	1	1	0	1	0	0	1	1	0	...	f	White-European	
4	5	1	0	0	0	0	0	0	1	0	...	f	NaN	

5 rows × 22 columns

```
In [4]: #checking whether data needs cleaning or not
asd_data.describe()
```

Out[4]:

	id	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	
count	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	7
mean	352.500000	0.721591	0.453125	0.457386	0.495739	0.498580	0.284091	0.417614	0.649148	0.323864	
std	203.371581	0.448535	0.498152	0.498535	0.500337	0.500353	0.451301	0.493516	0.477576	0.468281	
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	176.750000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
50%	352.500000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	
75%	528.250000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	
max	704.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	

So, here we have some missing values so it want to go through a cleaning process.

In [5]:

```
#here we are locating rows which contain null value in any of the column
asd_data.loc[(asd_data['age'].isnull() |(asd_data['gender'].isnull() |(asd_data['ethnicity'].isnull()
|(asd_data['jundice'].isnull()|(asd_data['austim'].isnull() |(asd_data['contry_of_res'].isnull()
|(asd_data['used_app_before'].isnull()|(asd_data['result'].isnull()|(asd_data['age_desc'].isnull(
|(asd_data['relation'].isnull()))]
```

Out[5]:

	id	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	...	gender	ethnicity
4	5	1	0	0	0	0	0	0	1	0	...	f	NaN
12	13	0	1	1	1	1	1	0	0	1	...	f	NaN
13	14	1	0	0	0	0	0	1	1	0	...	m	NaN
14	15	1	0	0	0	0	0	1	1	0	...	f	NaN
19	20	0	0	0	0	0	0	1	1	0	...	m	NaN
20	21	0	1	1	1	0	0	0	0	0	...	m	NaN
24	25	1	1	1	1	0	0	0	1	0	...	m	NaN
25	26	0	1	1	0	0	0	0	1	0	...	f	NaN
62	63	0	0	0	0	0	0	0	0	0	...	m	NaN
79	80	1	1	0	0	0	0	0	0	0	...	f	NaN
80	81	0	0	1	1	0	0	1	1	0	...	m	NaN
81	82	1	0	0	0	0	0	1	0	0	...	m	NaN
91	92	0	1	0	0	1	0	1	0	0	...	f	NaN
216	217	1	0	0	0	0	0	1	1	0	...	f	NaN
221	222	0	0	1	0	0	0	0	0	0	...	f	NaN
238	239	0	1	0	0	0	0	0	1	1	...	m	NaN
257	258	0	1	1	0	1	0	0	1	0	...	m	NaN
270	271	1	0	1	1	1	1	0	0	1	...	f	NaN
276	277	1	0	0	0	1	0	0	0	1	...	m	NaN
277	278	1	1	0	0	0	0	0	0	0	...	f	NaN
285	286	0	0	0	1	1	1	1	0	1	...	f	NaN
306	307	1	0	0	1	1	0	1	1	1	...	f	NaN

315	316	1	1	0	0	0	0	0	1	0	...	m	NaN
324	325	1	0	0	0	0	0	0	0	0	...	f	NaN
337	338	1	1	0	0	0	0	0	0	0	...	m	NaN
338	339	1	0	1	0	0	0	1	1	0	...	f	NaN
339	340	1	0	0	0	0	0	1	0	0	...	f	NaN
340	341	1	0	1	1	0	0	1	1	1	...	f	NaN
341	342	1	1	1	1	1	0	0	1	0	...	m	NaN
342	343	1	0	1	1	0	0	0	0	0	...	f	NaN
...
423	424	0	1	0	1	0	0	0	0	0	...	f	NaN
427	428	1	0	0	1	0	0	1	1	0	...	m	NaN
428	429	0	0	0	0	1	0	1	0	0	...	f	NaN
429	430	1	0	1	0	1	0	1	1	1	...	m	NaN
432	433	1	0	0	1	1	0	0	0	0	...	m	NaN
438	439	1	1	0	0	0	0	0	1	0	...	f	NaN
453	454	0	0	1	1	1	0	0	0	0	...	f	NaN
485	486	0	1	0	0	0	0	0	1	1	...	m	NaN
505	506	1	0	1	0	1	0	1	1	0	...	m	NaN
518	519	0	1	0	0	0	0	0	0	1	...	f	NaN
527	528	0	0	0	0	1	0	0	1	0	...	f	NaN
534	535	0	0	0	1	1	0	1	0	0	...	f	NaN
535	536	0	1	0	0	1	0	0	0	0	...	m	NaN
536	537	0	1	0	0	1	0	0	0	0	...	f	NaN
537	538	0	1	0	0	1	1	0	0	1	...	m	NaN
556	557	1	1	0	0	0	0	0	1	0	...	m	NaN
564	565	0	1	0	0	1	1	0	0	1	...	m	NaN
571	572	1	0	0	0	0	0	0	1	0	...	m	NaN
572	573	1	0	0	1	0	0	0	1	0	...	m	NaN
588	589	1	0	1	1	1	1	1	1	1	...	f	NaN
593	594	1	0	0	0	0	0	0	0	0	...	m	NaN
636	637	0	0	0	0	0	0	0	0	0	...	f	NaN
642	643	0	0	0	0	1	0	0	1	0	...	f	NaN
645	646	1	0	1	1	0	0	1	1	1	...	m	NaN
651	652	1	0	0	0	0	0	1	0	0	...	m	NaN

652	653	0	0	0	0	0	0	0	0	0	...	f	NaN
658	659	0	0	1	1	0	0	1	0	0	...	m	NaN
659	660	1	1	1	1	1	1	0	0	1	...	m	NaN
666	667	0	0	0	0	0	0	0	1	0	...	m	NaN
701	702	1	0	1	1	1	0	1	1	0	...	f	NaN

So here we have missing data rows in a random order so we have to drop them by 'dropna' function and again check the data that is it clean or not

[illegible]

```
In [7]: # Reminder of the features:
print(asd_data.dtypes)

# Total number of records in clean dataset
n_records = len(asd_data.index)

# TODO: Number of records where individual's with ASD in the clean dataset
n_asd_yes = len(asd_data[asd_data['Class/ASD'] == 'YES'])

# TODO: Number of records where individual's with no ASD in the clean dataset
n_asd_no = len(asd_data[asd_data['Class/ASD'] == 'NO'])

# Print the results
print("Total number of records:", n_records)
print("Individuals diagnosed with ASD:", n_asd_yes)
print("Individuals not diagnosed with ASD:", n_asd_no)
```

```

id                int64
A1_Score          int64
A2_Score          int64
A3_Score          int64
A4_Score          int64
A5_Score          int64
A6_Score          int64
A7_Score          int64
A8_Score          int64
A9_Score          int64
A10_Score         int64
age              float64
gender           object
ethnicity        object
jundice          object
austim           object
contry_of_res    object
used_app_before  object
result          int64
age_desc         object
relation         object
Class/ASD        object
dtype: object
Total number of records: 609
Individuals diagonised with ASD: 180
Individuals not diagonised with ASD: 429

```

```

In [8]: # Split the data into features and target label
asd_raw = asd_data['Class/ASD']
features_raw = asd_data[['age', 'gender', 'ethnicity', 'jundice', 'austim', 'contry_of_res', 'result',
                        'relation', 'A1_Score', 'A2_Score', 'A3_Score', 'A4_Score', 'A5_Score', 'A6_Score', 'A7_Score',
                        'A9_Score', 'A10_Score']]

```

When we create a model then we may need the data should be normalized so here we are using MinMaxScaler for preprocessing

```

In [9]: from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
numerical = ['age', 'result']

features_minmax_transform = pd.DataFrame(data = features_raw)
features_minmax_transform[numerical] = scaler.fit_transform(features_raw[numerical])
features_minmax_transform
# Show an example of a record with scaling applied
display(features_minmax_transform.head(n = 5))

```

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\preprocessing\data.py:323: DataConversionWarning: Data with input dtype int64, float64 were all converted to float64 by MinMaxScaler.
return self.partial_fit(X, y)

	age	gender	ethnicity	jundice	austim	contry_of_res	result	relation	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score
0	0.024590	f	White-European	no	no	United States	0.6	Self	1	1	1	1	
1	0.019126	m	Latino	no	yes	Brazil	0.5	Self	1	1	0	1	
2	0.027322	m	Latino	yes	yes	Spain	0.8	Parent	1	1	0	1	
3	0.049180	f	White-European	no	yes	United States	0.6	Self	1	1	0	1	
5	0.051913	m	Others	yes	no	United States	0.9	Self	1	1	1	1	

From above clean data sets we can observe that some have non-numeric values we shouldn't use non-numeric value with a learning algorithm so here we should convert it with some method. Here we will use 'One-Hot-Coding' scheme which will convert non-numeric values to "dummy" variable which can be used with learning algorithm.

In addition, We also need to convert the target to numeric value but we needn't to use 'One-Hot-Coding', we can use simple encoding to binary from as we have only variables as 'Yes' and 'No'.

```

In [10]: #One-hot encode the 'features_minmax_transform' data using pandas.get_dummies()
features_final = pd.get_dummies(features_minmax_transform)
display(features_final.head(5))

# Encode the 'all_classes_raw' data to numerical values
asd_classes = asd_raw.apply(lambda x: 1 if x == 'YES' else 0)

# Print the number of features after one-hot encoding

```

```

encoded = list(features_final.columns)
print("total features after one-hot encoding:", len(encoded))

# Uncomment the following line to see the encoded feature names
print(encoded)

```

	age	result	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	...	contry_of_res_United Arab Emirates
0	0.024590	0.6	1	1	1	1	0	0	1	1	...	0
1	0.019126	0.5	1	1	0	1	0	0	0	1	...	0
2	0.027322	0.8	1	1	0	1	1	0	1	1	...	0
3	0.049180	0.6	1	1	0	1	0	0	1	1	...	0
5	0.051913	0.9	1	1	1	1	1	0	1	1	...	0

5 rows × 94 columns

```

total features after one-hot encoding: 94
['age', 'result', 'A1_Score', 'A2_Score', 'A3_Score', 'A4_Score', 'A5_Score', 'A6_Score', 'A7_Score', 'A8_Score',
 'A9_Score', 'A10_Score', 'gender_f', 'gender_m', 'ethnicity_Asian', 'ethnicity_Black', 'ethnicity_Hispanic',
 'ethnicity_Latino', 'ethnicity_Middle Eastern ', 'ethnicity_Others', 'ethnicity_Pasifika', 'ethnicity_South Asian',
 'ethnicity_Turkish', 'ethnicity_White-European', 'ethnicity_others', 'jundice_no', 'jundice_yes', 'austim_no',
 'austim_yes', 'contry_of_res_Afghanistan', 'contry_of_res_AmericanSamoa', 'contry_of_res_Angola', 'contry_of_r
es_Armenia', 'contry_of_res_Aruba', 'contry_of_res_Australia', 'contry_of_res_Austria', 'contry_of_res_Bahamas',
 'contry_of_res_Bangladesh', 'contry_of_res_Belgium', 'contry_of_res_Bolivia', 'contry_of_res_Brazil', 'contry_of
_res_Burundi', 'contry_of_res_Canada', 'contry_of_res_Chile', 'contry_of_res_China', 'contry_of_res_Costa Rica',
 'contry_of_res_Cyprus', 'contry_of_res_Czech Republic', 'contry_of_res_Ecuador', 'contry_of_res_Egypt', 'contry
_of_res_Ethiopia', 'contry_of_res_Finland', 'contry_of_res_France', 'contry_of_res_Germany', 'contry_of_res_Icela
nd', 'contry_of_res_India', 'contry_of_res_Indonesia', 'contry_of_res_Iran', 'contry_of_res_Ireland', 'contry_of
_res_Italy', 'contry_of_res_Jordan', 'contry_of_res_Malaysia', 'contry_of_res_Mexico', 'contry_of_res_Nepal', 'c
ontry_of_res_Netherlands', 'contry_of_res_New Zealand', 'contry_of_res_Nicaragua', 'contry_of_res_Niger', 'contr
y_of_res_Oman', 'contry_of_res_Pakistan', 'contry_of_res_Philippines', 'contry_of_res_Portugal', 'contry_of_res_
Romania', 'contry_of_res_Russia', 'contry_of_res_Saudi Arabia', 'contry_of_res_Serbia', 'contry_of_res_Sierra Le
one', 'contry_of_res_South Africa', 'contry_of_res_Spain', 'contry_of_res_Sri Lanka', 'contry_of_res_Sweden', 'c
ontry_of_res_Tonga', 'contry_of_res_Turkey', 'contry_of_res_Ukraine', 'contry_of_res_United Arab Emirates', 'con
try_of_res_United Kingdom', 'contry_of_res_United States', 'contry_of_res_Uruguay', 'contry_of_res_Viet Nam', 'r
elation_Health care professional', 'relation_Others', 'relation_Parent', 'relation_Relative', 'relation_Self']

```

In [11]: # histogram of Class/ASD

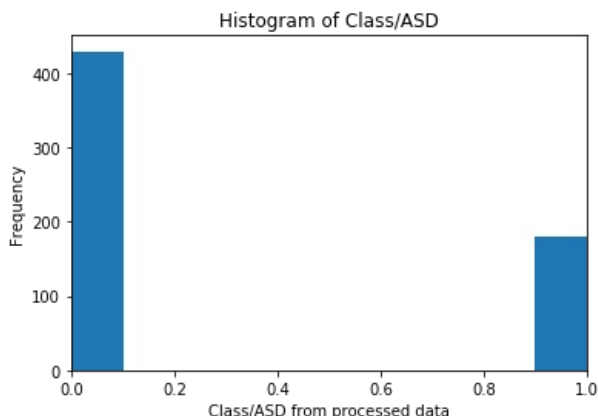
```

# 8 bins
plt.hist(asd_classes, bins=10)

# x-axis limit from 0 to 1
plt.xlim(0,1)
plt.title('Histogram of Class/ASD')
plt.xlabel('Class/ASD from processed data')
plt.ylabel('Frequency')

```

Out[11]: Text(0, 0.5, 'Frequency')



Now we have all non-numerical values converted into numerical value and all numerical values has been normalized form. So we will split the data into training and test data. We will use 80% of data as training data and 20% as test data.

In [12]: from sklearn.model_selection import train_test_split

```

np.random.seed(1234)

X_train, X_test, y_train, y_test = train_test_split(features_final, asd_classes, train_size=0.80, random_state=

# Show the results of the split

```

```
print("Training set has samples:",X_train.shape[0])
print("Testing set has samples:",X_test.shape[0])
#asd_data
```

Training set has samples: 487
Testing set has samples: 122

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\model_selection_split.py:2179: FutureWarning: From version 0.21, test_size will always complement train_size unless both are specified.
FutureWarning)

SVM :

```
In [13]: from sklearn import svm
from sklearn.model_selection import cross_val_score

C = 1.0
svc = svm.SVC(kernel='linear', C=C, gamma=2)
```

```
In [14]: cv_scores = cross_val_score(svc, features_final, asd_classes, cv=10)

cv_scores.mean()
```

Out[14]: 1.0

AUC Score:

```
In [15]: # calculate cross-validated AUC
cross_val_score(svc, features_final, asd_classes, cv=10, scoring='roc_auc').mean()
```

Out[15]: 1.0

F-beta Score:

```
In [16]: svc.fit(X_train, y_train)
from sklearn.metrics import fbeta_score
predictions_test = svc.predict(X_test)
fbeta_score(y_test, predictions_test, average='binary', beta=0.5)
```

Out[16]: 1.0

K-Nearest-Neighbors (KNN) :

```
In [17]: from sklearn import neighbors

knn = neighbors.KNeighborsClassifier(n_neighbors=10)
cv_scores = cross_val_score(knn, features_final, asd_classes, cv=10)

cv_scores.mean()
```

Out[17]: 0.9474590163934427

AUC Score:

```
In [18]: # calculate cross-validated AUC
cross_val_score(knn, features_final, asd_classes, cv=10, scoring='roc_auc').mean()
```

Out[18]: 0.9930078749846192

F-beta Score:

```
In [19]: knn.fit(X_train, y_train)
from sklearn.metrics import fbeta_score
predictions_test = knn.predict(X_test)
fbeta_score(y_test, predictions_test, average='binary', beta=0.5)
```

Out[19]: 0.9192825112107623

Choosing K is tricky, so I can't discard KNN until we've tried different values of K. Hence we write a for loop to run KNN with K values ranging from 10 to 50 and see if K makes a substantial difference.

```
In [20]: for n in range(10, 50):
knn = neighbors.KNeighborsClassifier(n_neighbors=n)
cv_scores = cross_val_score(knn, features_final, asd_classes, cv=10)
print (n, cv_scores.mean())
```

```

10 0.9474590163934427
11 0.9507377049180328
12 0.9507377049180328
13 0.9540437158469945
14 0.9507650273224044
15 0.944207650273224
16 0.9507650273224044
17 0.9523770491803278
18 0.9523770491803278
19 0.9540163934426229
20 0.9523770491803278
21 0.9523770491803278
22 0.9474590163934424
23 0.9490983606557375
24 0.9507377049180326
25 0.9507377049180328
26 0.9523770491803278
27 0.9507377049180328
28 0.9507377049180326
29 0.9507377049180328
30 0.9523770491803278
31 0.9474863387978143
32 0.9491256830601094
33 0.9474863387978143
34 0.9507650273224044
35 0.9491256830601094
36 0.9491256830601091
37 0.9507650273224044
38 0.9540710382513661
39 0.9524316939890708
40 0.9524316939890708
41 0.9524316939890708
42 0.9507923497267757
43 0.9507923497267757
44 0.9507923497267757
45 0.9507923497267757
46 0.9524316939890708
47 0.9524316939890708
48 0.9557103825136611
49 0.9524316939890708

```

Logistic Regression :

```
In [21]: from sklearn.linear_model import LogisticRegression
```

```

logreg = LogisticRegression()
cv_scores = cross_val_score(logreg, features_final, asd_classes, cv=10)
cv_scores.mean()

```

```

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver w
ill be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)

```

```
Out[21]: 0.9704371584699454
```


AUC Score:

```
In [22]: # calculate cross-validated AUC
cv_scores_roc = cross_val_score(logreg, features_final, asd_classes, cv=10, scoring='roc_auc').mean()
cv_scores_roc.mean()
```

[illegible]

```
Out[22]: 0.9974098683401008
```

F-beta Score:

```
In [23]: logreg.fit(X_train, y_train)
from sklearn.metrics import fbeta_score
predictions_test = logreg.predict(X_test)
fbeta_score(y_test, predictions_test, average='binary', beta=0.5)
```

```
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
```

```
Out[23]: 0.9307359307359306
```

Decision Trees :

```
In [24]: from sklearn import tree
          from sklearn.tree import DecisionTreeClassifier

          dectree = DecisionTreeClassifier(random_state=1)

          # Train the classifier on the training set
          dectree.fit(X_train, y_train)
```

```
Out[24]: DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
                                max_features=None, max_leaf_nodes=None,
                                min_impurity_decrease=0.0, min_impurity_split=None,
                                min_samples_leaf=1, min_samples_split=2,
                                min_weight_fraction_leaf=0.0, presort=False, random_state=1,
                                splitter='best')
```

```
In [25]: # make class predictions for the testing set
y_pred class = dectree.predict(X_test)
```

```
In [26]: dectree.score(X test, y test)
```

```
Out[26]: 1.0
```

```
In [27]: from sklearn.model_selection import cross_val_score

         dectree = DecisionTreeClassifier(random_state=1)

         cv_scores = cross_val_score(dectree, features_final, asd_classes, cv=10)
```

```
cv_scores.mean()
```

Out[27]: 1.0

AUC Score:

```
In [28]: # calculate cross-validated AUC
cross_val_score(dectree, features_final, asd_classes, cv=10, scoring='roc_auc').mean()
```

Out[28]: 1.0

F-beta Score:

```
In [29]: dectree.fit(X_train, y_train)
from sklearn.metrics import fbeta_score
predictions_test = dectree.predict(X_test)
fbeta_score(y_test, predictions_test, average='binary', beta=0.5)
```

Out[29]: 1.0

```
In [30]: # TODO: Import 'GridSearchCV', 'make_scorer', and any other necessary libraries
from sklearn.metrics import fbeta_score
from sklearn.metrics import accuracy_score

from sklearn.metrics import make_scorer
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV

def f_beta_score(y_true, y_predict):
    return fbeta_score(y_true, y_predict, beta = 0.5)

# TODO: Initialize the classifier
clf = SVC(random_state = 1)

# TODO: Create the parameters list you wish to tune, using a dictionary if needed.
# HINT: parameters = {'parameter_1': [value1, value2], 'parameter_2': [value1, value2]}
parameters = {'C':range(1,6), 'kernel':['linear', 'poly', 'rbf', 'sigmoid'], 'degree':range(1,6)}

# TODO: Make an fbeta_score scoring object using make_scorer()
scorer = make_scorer(f_beta_score)

# TODO: Perform grid search on the classifier using 'scorer' as the scoring method using GridSearchCV()
grid_obj = GridSearchCV(estimator = clf, param_grid = parameters, scoring = scorer)

# TODO: Fit the grid search object to the training data and find the optimal parameters using fit()
grid_fit = grid_obj.fit(X_train, y_train)

# Get the estimator
best_clf = grid_fit.best_estimator_

# Make predictions using the unoptimized and model
predictions = (clf.fit(X_train, y_train)).predict(X_test)
best_predictions = best_clf.predict(X_test)
```

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\model_selection_split.py:2053: FutureWarning: You should specify a value for 'cv' instead of relying on the default value. The default value will change from 3 to 5 in version 0.22.

warnings.warn(CV_WARNING, FutureWarning)

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.

"avoid this warning.", FutureWarning)

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.

"avoid this warning.", FutureWarning)

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.

"avoid this warning.", FutureWarning)

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.

"avoid this warning.", FutureWarning)

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.

"avoid this warning.", FutureWarning)

C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
re is ill-defined and being set to 0.0 due to no predicted samples.  
    'precision', 'predicted', average, warn_for)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1143: UndefinedMetricWarning: F-score  
is ill-defined and being set to 0.0 due to no predicted samples.  
    'precision', 'predicted', average, warn_for)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1143: UndefinedMetricWarning: F-sco  
re is ill-defined and being set to 0.0 due to no predicted samples.  
    'precision', 'predicted', average, warn_for)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1143: UndefinedMetricWarning: F-sco  
re is ill-defined and being set to 0.0 due to no predicted samples.  
    'precision', 'predicted', average, warn_for)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)  
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma wi  
ll change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly t  
o 'auto' or 'scale' to avoid this warning.  
    "avoid this warning.", FutureWarning)
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.
  "avoid this warning.", FutureWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\model_selection\_search.py:841: DeprecationWarning: The default of the `iid` parameter will change from True to False in version 0.22 and will be removed in 0.24. This will change numeric results when test-set sizes are unequal.
  DeprecationWarning)
C:\Users\ganpa\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.
  "avoid this warning.", FutureWarning)
```

```
In [31]: # Report the before-and-afterscores
print("Unoptimized model\n-----")
print("Accuracy score on testing data:", accuracy_score(y_test, predictions))
print("F-score on testing data:", fbeta_score(y_test, predictions, beta = 0.5))
print("\nOptimized Model\n-----")
print("Final accuracy score on the testing data:", accuracy_score(y_test, best_predictions))
print("Final F-score on the testing data:", fbeta_score(y_test, best_predictions, beta = 0.5))
```

Unoptimized model

Accuracy score on testing data: 0.9508196721311475

F-score on testing data: 0.930232558139535

Optimized Model

Final accuracy score on the testing data: 1.0

Final F-score on the testing data: 1.0