

```
In [1]: #steps in machine learning
#1.Import data often comes with csv file

#2.clean the data (remove missing or duplicate values)# EDA

#3.split the dataset into 2 sets training set and testing set

#e.g if we have 1000 pictures of cats and dogs we can reserve 80% for training and 20% for testing

#4.create a model

#This involves selecting an algorithm to analyze the data.so many mahine Learning algorithms out there such
#random forest,neural networks etc.
#each algorithm has pros and cons in terms of acuuracy and performance.
#Libraries out there that provides algorithms.one of most popular library is scikit-learn.
#so build a model using these algorithms

#5.Train the model #method .fit(Training data)
#we feed the training data to the model.model will Learn the patterns in the data .

#6.make predictions  #.predict(Testing data)
#eg.we ask the model is it a cat or dog if we come with a new image from the testing dataset and our model
#predictions.predictions are not always accurate.

#7.Evaluate the predictions
#in this step, evaluate the predictions and measure their accuracy.if accuracy is low,then
#we need to get back to our model and either select a different algorithm that is going to produce a more c
#the kind of problem we are solving or fine tune the parameters of oue model.
#each algorithm has parameters(Hyperparameters) that we can modify to optimize the accuracy.
```

```
In [2]: import pandas as pd
import numpy as np
```

```
In [3]: #step1 =read the data
df=pd.read_csv('Music.csv')
```

```
In [4]: df.head()
#1= Male 0= Female
```

	Age	Gender	Genre
0	20	1	Hiphop
1	23	1	Hiphop
2	25	1	Hiphop
3	26	1	Jazz
4	29	1	Jazz

```
In [5]: #2.clean the data (remove missing or duplicate values)
df.isnull().sum()
```

```
Age      0
Gender    0
Genre     0
dtype: int64
```

```
In [6]: #target column is Genre i.e y
```

```
In [7]: #step 3 split dataset into 2 sets input and output
```

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

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

```
In [10]: X.head()
```

	Age	Gender
0	20	1
1	23	1
2	25	1
3	26	1
4	29	1

```
In [11]: y.head()
```

0	Hiphop
1	Hiphop
2	Hiphop
3	Jazz
4	Jazz

Name: Genre, dtype: object

```
In [12]: X.shape
```

(18, 2)

```
In [13]: y.shape
```

```
(18,)
```

```
In [14]: #step 4
         #create model by using different algorithms
         from sklearn.tree import DecisionTreeClassifier
```

```
In [15]: #sklearn is the package that comes with scikit-learn library
         #In this package we have the module called tree and in this module
         #we have a class called DecisionTreeClassifier.
```

```
In [16]: #create an object of DecisionTreeClassifier
         model=DecisionTreeClassifier()
```

```
In [17]: #step 5 Training the model
         #fit the model
```

```
In [18]: model.fit(X,y)

         DecisionTreeClassifier()
```

```
In [19]: #prediction
         predictions=model.predict([[21,1],[22,0]])
         predictions

         array(['Hiphop', 'Dance'], dtype=object)
```

```
In [20]: from sklearn.metrics import accuracy_score
```

```
In [21]: #Evaluate the Accuracy  
#split the data into training and testing  
#keep 70% for training and 30% for testing
```

```
In [22]: from sklearn.model_selection import train_test_split
```

```
In [23]: X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.30,random_state=42)
```

```
In [24]: X_train.shape
```

```
(12, 2)
```

```
In [25]: y_train.shape
```

```
(12,)
```

```
In [26]: X_train.shape
```

```
(12, 2)
```

```
In [27]: y_test.shape
```

```
(6,)
```

```
In [28]: model=DecisionTreeClassifier()
```

```
In [29]: model.fit(X_train,y_train)
```

```
DecisionTreeClassifier()
```

```
In [30]: predictions=model.predict(X_test)
```

```
predictions
```

```
array(['Hiphop', 'Hiphop', 'Classical', 'Jazz', 'Jazz', 'Acoustaic'],  
      dtype=object)
```

```
In [31]: y_test
```

```
0      Hiphop  
1      Hiphop  
8      Classical  
5         Jazz  
3         Jazz  
13     Acoustaic  
Name: Genre, dtype: object
```

```
In [32]: accuracy_score(y_test,predictions)
```

```
1.0
```

```
In [33]: data={'Actual':y_test,'Predicted':predictions}
```

```
In [34]: pd.DataFrame(data)
```

	<b>Actual</b>	<b>Predicted</b>
0	Hiphop	Hiphop
1	Hiphop	Hiphop
8	Classical	Classical
5	Jazz	Jazz
3	Jazz	Jazz
13	Acoustaic	Acoustaic

```
In [ ]:
```