

## CHAPTER 4

# Project -1 Machine Learning

## Prediction Of Different type of Microbes

---

In this project, we will work with Microbe.csv dataset to develop a machine learning algorithm that predicts the microbe. A model like this would be very valuable to predict the microorganism using serial no,raddi,perimeter,.....etc

### 4.1 Problem Statement

Develop a model that has the capacity of predicting the microorganism by making use of the information provided in microbes Dataset

### 4.2 Dataset

The dataset used in this project consists of 25 variables.The main variable we are interested is 'microorganism'. This variable predicts the fracture of the person based on the inputs given in dataset

1. Serial No	Enter the serial no
2. Solidity	Enter the solidity
3. Eccentricity	Enter the eccentricity
4. EquivDiameter	Enter the equivdiameter
5. Extrema	Enter the extrema
6. FilledArea	Enter the filledarea
7. Extent	Enter the extent
.	.
.	.
.	.
25. Raddi	Enter the raddi

The overview of the original dataset is shown in figure with its original features:

## 4.3 Algorithm –KNN Algorithm

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

## 4.4 Programming Steps

- This project requires us to predict the microorganism based on the given input dataset.
- First, we read the given dataset using pandas function.
- Then we print the inputs and output from csv file.
- We initialize the model i.e., K-NN Algorithm.
- We further implement this using Django in order for better representation

### Code:

```
import pandas as pd
path="C:\\Users\\shara\\OneDrive\\Desktop\\Project\\train_dataset.csv"
data=pd.read_csv(path)
print(data)
print(data.info())
print(data.isnull().sum())

inputs=data.drop(['microorganisms'], 'columns')
output=data['microorganisms']
print(inputs)
print(output)

import sklearn
```

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(inputs,output,train_size=0.8)
print(x_train)
print(x_test)
print(y_train)
print(y_test)

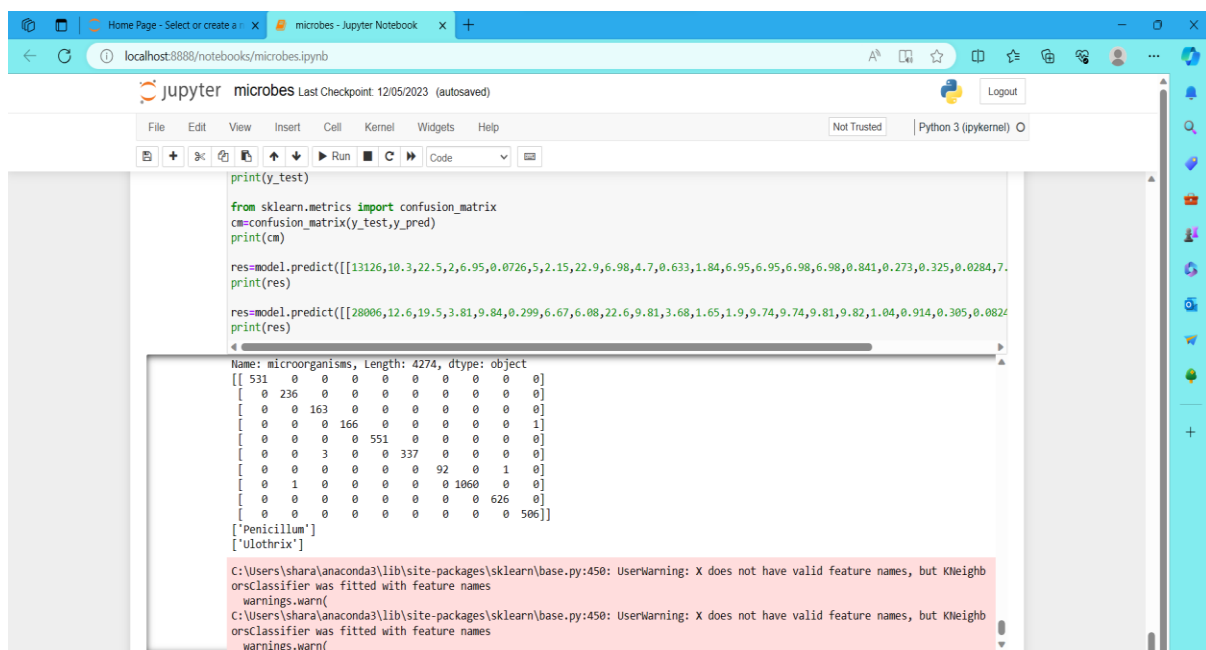
from sklearn.neighbors import KNeighborsClassifier
model=KNeighborsClassifier(n_neighbors=13)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
print(y_pred)
print(y_test)

from sklearn.metrics import confusion_matrix
cm=confusion_matrix(y_test,y_pred)
print(cm)

res=model.predict([[13126,10.3,22.5,2,6.95,0.0726,5,2.15,22.9,6.98,4.7,0.633,1
.84,6.95,6.95,6.98,6.98,0.841,0.273,0.325,0.0284,7.06,5.42,0.184,1.93]])
print(res)

res=model.predict([[28006,12.6,19.5,3.81,9.84,0.299,6.67,6.08,22.6,9.81,3.68,1
.65,1.9,9.74,9.74,9.81,9.82,1.04,0.914,0.305,0.0824,10.5,4.46,0.649,5.35]])
print(res)
```

## OUTPUT:

The screenshot shows a Jupyter Notebook window titled 'microbes - Jupyter Notebook'. The browser address bar shows 'localhost:8888/notebooks/microbes.ipynb'. The notebook interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, cell execution, and output viewing. The code cell contains the same Python code as shown in the previous block. The output cell displays the result of the model predictions, which is a list of predicted class names: ['Penicillium', 'Ulothrix']. Below the output, there are two warning messages from sklearn: 'UserWarning: X does not have valid feature names, but KNeighborsClassifier was fitted with feature names'. The output also shows a confusion matrix for the two classes, with the first row representing 'Penicillium' and the second row representing 'Ulothrix'. The confusion matrix shows that all instances of 'Penicillium' were correctly classified, while all instances of 'Ulothrix' were also correctly classified.

```
print(y_test)

from sklearn.metrics import confusion_matrix
cm=confusion_matrix(y_test,y_pred)
print(cm)

res=model.predict([[13126,10.3,22.5,2,6.95,0.0726,5,2.15,22.9,6.98,4.7,0.633,1.84,6.95,6.95,6.98,6.98,0.841,0.273,0.325,0.0284,7.06,5.42,0.184,1.93]])
print(res)

res=model.predict([[28006,12.6,19.5,3.81,9.84,0.299,6.67,6.08,22.6,9.81,3.68,1.65,1.9,9.74,9.74,9.81,9.82,1.04,0.914,0.305,0.0824,10.5,4.46,0.649,5.35]])
print(res)
```

```
Name: microorganisms, Length: 4274, dtype: object
[[ 531  0  0  0  0  0  0  0  0  0]
 [  236  0  0  0  0  0  0  0  0  0]
 [  163  0  0  0  0  0  0  0  0  0]
 [  166  0  0  0  0  0  0  0  0  1]
 [  551  0  0  0  0  0  0  0  0  0]
 [  337  0  0  0  0  0  0  0  0  0]
 [  92  0  0  0  0  0  0  0  0  0]
 [ 1060  0  0  0  0  0  0  0  0  0]
 [  626  0  0  0  0  0  0  0  0  0]
 [  506  0  0  0  0  0  0  0  0  0]
 ['Penicillium']
 ['Ulothrix']
```

```
C:\Users\shara\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but KNeighborsClassifier was fitted with feature names
warnings.warn(
C:\Users\shara\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but KNeighborsClassifier was fitted with feature names
warnings.warn(
```

Home Page - Select or create a notebook x microbes - Jupyter Notebook x 127.0.0.1:8000/microbes/ x +

127.0.0.1:8000/microbes/

### Prediction of Microbes

Serial No	2658
Solidity	3.44
Eccentricity	6.55
EquivDiameter	2.77
Extrema	5.56
FilledArea	2.4
Extent	5.66
Orientation	8.65
EulerNumber	6.53
BoundingBox1	8.45
BoundingBox2	9.35
BoundingBox3	2.45
BoundingBox4	5.85
ConvexHull1	8.14
ConvexHull2	9.51
ConvexHull3	8.45
ConvexHull4	9.556
MajorAxisLength	7.35
MinorAxisLength	9.36
Perimeter	3.68
ConvexArea	3.98
Centroid1	2.67
Centroid2	5.98
Area	2.66
raddi	7.56

Predict

24°C Haze 19:55 15-12-2023

Home Page - Select or create a notebook x microbes - Jupyter Notebook x 127.0.0.1:8000/microbes/ x +

127.0.0.1:8000/microbes/

### Prediction of Microbes

Serial No	
Solidity	
Eccentricity	
EquivDiameter	
Extrema	
FilledArea	
Extent	
Orientation	
EulerNumber	
BoundingBox1	
BoundingBox2	
BoundingBox3	
BoundingBox4	
ConvexHull1	
ConvexHull2	
ConvexHull3	
ConvexHull4	
MajorAxisLength	
MinorAxisLength	
Perimeter	
ConvexArea	
Centroid1	
Centroid2	
Area	
raddi	

Predict

["Volvox"]

24°C Haze 19:55 15-12-2023