Parkinson’s disease detection model using XGBoost

XGBoost is a new Machine Learning algorithm designed with speed and performance in mind. XGBoost stands for eXtreme Gradient Boosting and is based on decision trees. In this project, we will import the XGBClassifier from the xgboost library; this is an implementation of the scikit-learn API for XGBoost classification.

Background

Parkinson’s disease is a progressive disorder of the central nervous system affecting movement and inducing tremors and stiffness. It has 5 stages to it and affects more than 1 million individuals every year in India. This is chronic and has no cure yet. It is a neurodegenerative disorder affecting dopamine-producing neurons in the brain.

Problem Statement

To build a model that will accurately detect the presence of Parkinson’s disease in an individual.

Prerequisite: Install the following libraries with pip

In [3]:
```
pip install numpy pandas sklearn xgboost
```

```
Requirement already satisfied: numpy in c:\users\harvey\anaconda3\lib\site-packages (1.23.5)
Requirement already satisfied: pandas in c:\users\harvey\anaconda3\lib\site-packages (1.5.3)
Requirement already satisfied: sklearn in c:\users\harvey\anaconda3\lib\site-packages (0.0.post5)
Requirement already satisfied: xgboost in c:\users\harvey\anaconda3\lib\site-packages (1.7.5)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\harvey\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\harvey\anaconda3\lib\site-packages (from pandas) (2022.7)
Requirement already satisfied: scipy in c:\users\harvey\anaconda3\lib\site-packages (from xgboost) (1.10.0)
Requirement already satisfied: six>=1.5 in c:\users\harvey\anaconda3\lib\site-packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

Prerequisite: Install Jupyter lab with pip

In [4]:
```
pip install jupyterlab
```

```
Requirement already satisfied: jupyterlab in c:\users\harvey\anaconda3\lib\site-packages (3.5.3)
Requirement already satisfied: packaging in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (22.0)
Requirement already satisfied: notebook<7 in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (6.5.2)
Requirement already satisfied: jupyterlab-server~=2.10 in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (2.19.0)
Requirement already satisfied: ipython in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (8.10.0)
Requirement already satisfied: jinja2>=2.1 in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (3.1.2)
Requirement already satisfied: jupyter-server<3,>=1.16.0 in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (1.23.4)
Requirement already satisfied: nbclassic in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (0.5.2)
Requirement already satisfied: tornado>=6.1.0 in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (6.1)
Requirement already satisfied: tomli in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (2.0.1)
Requirement already satisfied: jupyter-core in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab) (5.2.0)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\harvey\anaconda3\lib\site-packages (from jinja2>=2.1->jupyterlab) (2.1.1)
Requirement already satisfied: nbconvert>=6.4.4 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (6.5.4)
Requirement already satisfied: nbformat>=5.2.0 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (5.7.0)
Requirement already satisfied: prometheus-client in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (0.14.
1)
Requirement already satisfied: argon2-cffi in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (21.3.0)
Requirement already satisfied: pywinpty in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (2.0.10)
Requirement already satisfied: anyio<4,>=3.1.0 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (3.5.0)
Requirement already satisfied: jupyter-client>=6.1.12 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab)
(7.3.4)
Requirement already satisfied: terminado>=0.8.3 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (0.17.
1)
Requirement already satisfied: pyzmq>=17 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (23.2.0)
Requirement already satisfied: Send2Trash in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (1.8.0)
Requirement already satisfied: traitlets>=5.1 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (5.7.1)
Requirement already satisfied: websocket-client in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-server<3,>=1.16.0->jupyterlab) (0.58.
0)
Requirement already satisfied: pywin32>=1.0 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-core->jupyterlab) (305.1)
Requirement already satisfied: platformdirs>=2.5 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-core->jupyterlab) (2.5.2)
Requirement already satisfied: json5>=0.9.0 in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab-server~=2.10->jupyterlab) (0.9.6)
Requirement already satisfied: babel>=2.10 in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab-server~=2.10->jupyterlab) (2.11.0)
Requirement already satisfied: requests>=2.28 in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab-server~=2.10->jupyterlab) (2.28.1)
Requirement already satisfied: jsonschema>=4.17.3 in c:\users\harvey\anaconda3\lib\site-packages (from jupyterlab-server~=2.10->jupyterlab) (4.17.
3)
Requirement already satisfied: ipython-genutils in c:\users\harvey\anaconda3\lib\site-packages (from notebook<7->jupyterlab) (0.2.0)
Requirement already satisfied: nest-asyncio>=1.5 in c:\users\harvey\anaconda3\lib\site-packages (from notebook<7->jupyterlab) (1.5.6)
Requirement already satisfied: ipykernel in c:\users\harvey\anaconda3\lib\site-packages (from notebook<7->jupyterlab) (6.19.2)
Requirement already satisfied: notebook-shim>=0.1.0 in c:\users\harvey\anaconda3\lib\site-packages (from nbclassic->jupyterlab) (0.2.2)
Requirement already satisfied: pygments>=2.4.0 in c:\users\harvey\anaconda3\lib\site-packages (from ipython->jupyterlab) (2.11.2)
Requirement already satisfied: colorama in c:\users\harvey\anaconda3\lib\site-packages (from ipython->jupyterlab) (0.4.6)
Requirement already satisfied: jedi>=0.16 in c:\users\harvey\anaconda3\lib\site-packages (from ipython->jupyterlab) (0.18.1)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.30 in c:\users\harvey\anaconda3\lib\site-packages (from ipython->jupyterlab) (3.0.36)
Requirement already satisfied: pickleshare in c:\users\harvey\anaconda3\lib\site-packages (from ipython->jupyterlab) (0.7.5)
Requirement already satisfied: matplotlib-inline in c:\users\harvey\anaconda3\lib\site-packages (from ipython->jupyterlab) (0.1.6)
Requirement already satisfied: decorator in c:\users\harvey\anaconda3\lib\site-packages (from ipython->jupyterlab) (5.1.1)
Requirement already satisfied: backcall in c:\users\harvey\anaconda3\lib\site-packages (from ipython->jupyterlab) (0.2.0)
Requirement already satisfied: stack-data in c:\users\harvey\anaconda3\lib\site-packages (from ipython->jupyterlab) (0.2.0)
Requirement already satisfied: idna>=2.8 in c:\users\harvey\anaconda3\lib\site-packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.16.0->jupyterla
b) (3.4)
Requirement already satisfied: sniffio>=1.1 in c:\users\harvey\anaconda3\lib\site-packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.16.0->jupyte
rlab) (1.2.0)
Requirement already satisfied: pytz>=2015.7 in c:\users\harvey\anaconda3\lib\site-packages (from babel>=2.10->jupyterlab-server~=2.10->jupyterlab)
(2022.7)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in c:\users\harvey\anaconda3\lib\site-packages (from jedi>=0.16->ipython->jupyterlab) (0.8.3)
Requirement already satisfied: attrs>=17.4.0 in c:\users\harvey\anaconda3\lib\site-packages (from jsonschema>=4.17.3->jupyterlab-server~=2.10->jupy
terlab) (22.1.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in c:\users\harvey\anaconda3\lib\site-packages (from jsonschema>=4.17.
3->jupyterlab-server~=2.10->jupyterlab) (0.18.0)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-client>=6.1.12->jupyter-server<
3,>=1.16.0->jupyterlab) (2.8.2)
Requirement already satisfied: entrypoints in c:\users\harvey\anaconda3\lib\site-packages (from jupyter-client>=6.1.12->jupyter-server<3,>=1.16.0->
jupyterlab) (0.4)
Requirement already satisfied: beautifulsoup4 in c:\users\harvey\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=1.16.0->jup
yterlab) (4.11.1)
Requirement already satisfied: lxml in c:\users\harvey\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=1.16.0->jupyterlab)
(4.9.1)
Requirement already satisfied: jupyterlab-pygments in c:\users\harvey\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=1.16.0
->jupyterlab) (0.1.2)
Requirement already satisfied: nbclient>=0.5.0 in c:\users\harvey\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=1.16.0->ju
pyterlab) (0.5.13)
Requirement already satisfied: bleach in c:\users\harvey\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=1.16.0->jupyterlab)
(4.1.0)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\harvey\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=1.16.
0->jupyterlab) (1.5.0)
Requirement already satisfied: tinycss2 in c:\users\harvey\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=1.16.0->jupyterla
b) (1.2.1)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\users\harvey\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=1.16.0->
jupyterlab) (0.8.4)
Requirement already satisfied: defusedxml in c:\users\harvey\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=1.16.0->jupyter
lab) (0.7.1)
Requirement already satisfied: fastjsonschema in c:\users\harvey\anaconda3\lib\site-packages (from nbformat>=5.2.0->jupyter-server<3,>=1.16.0->jupy
terlab) (2.16.2)
Requirement already satisfied: wcwidth in c:\users\harvey\anaconda3\lib\site-packages (from prompt-toolkit<3.1.0,>=3.0.30->ipython->jupyterlab) (0.
2.5)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\harvey\anaconda3\lib\site-packages (from requests>=2.28->jupyterlab-server~=2.10->jup
yterlab) (2022.12.7)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\harvey\anaconda3\lib\site-packages (from requests>=2.28->jupyterlab-server~=2.1
0->jupyterlab) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\harvey\anaconda3\lib\site-packages (from requests>=2.28->jupyterlab-server~=2.10->
jupyterlab) (1.26.14)
Requirement already satisfied: argon2-cffi-bindings in c:\users\harvey\anaconda3\lib\site-packages (from argon2-cffi->jupyter-server<3,>=1.16.0->ju
pyterlab) (21.2.0)
Requirement already satisfied: psutil in c:\users\harvey\anaconda3\lib\site-packages (from ipykernel->notebook<7->jupyterlab) (5.9.0)
Requirement already satisfied: debugpy>=1.0 in c:\users\harvey\anaconda3\lib\site-packages (from ipykernel->notebook<7->jupyterlab) (1.5.1)
Requirement already satisfied: comm>=0.1.1 in c:\users\harvey\anaconda3\lib\site-packages (from ipykernel->notebook<7->jupyterlab) (0.1.2)
Requirement already satisfied: pure-eval in c:\users\harvey\anaconda3\lib\site-packages (from stack-data->ipython->jupyterlab) (0.2.2)
Requirement already satisfied: asttokens in c:\users\harvey\anaconda3\lib\site-packages (from stack-data->ipython->jupyterlab) (2.0.5)
Requirement already satisfied: executing in c:\users\harvey\anaconda3\lib\site-packages (from stack-data->ipython->jupyterlab) (0.8.3)
Requirement already satisfied: six in c:\users\harvey\anaconda3\lib\site-packages (from websocket-client->jupyter-server<3,>=1.16.0->jupyterlab)
(1.16.0)
Requirement already satisfied: cffi>=1.0.1 in c:\users\harvey\anaconda3\lib\site-packages (from argon2-cffi-bindings->argon2-cffi->jupyter-server<
3,>=1.16.0->jupyterlab) (1.15.1)
Requirement already satisfied: soupsieve>1.2 in c:\users\harvey\anaconda3\lib\site-packages (from beautifulsoup4->nbconvert>=6.4.4->jupyter-server<
3,>=1.16.0->jupyterlab) (2.3.2.post1)
Requirement already satisfied: webencodings in c:\users\harvey\anaconda3\lib\site-packages (from bleach->nbconvert>=6.4.4->jupyter-server<3,>=1.16.
0->jupyterlab) (0.5.1)
Requirement already satisfied: pycparser in c:\users\harvey\anaconda3\lib\site-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->jupyt
er-server<3,>=1.16.0->jupyterlab) (2.21)
Note: you may need to restart the kernel to use updated packages.
```

Data Source

The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals. Parkison's Disease data is downloaded from https://archive.ics.uci.edu/ml/machine-learning-databases/parkinsons/ which contains 24 columns and 195 records.

In [5]:
```python
# Import libraries

import numpy as np
import pandas as pd
import os, sys
from sklearn.preprocessing import MinMaxScaler
from xgboost import XGBClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
```

In [6]:
```python
# Load dataset

data = pd.read_csv("parkinsons.data.csv")
data.head()
```

Out[6]:

	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	...	Shimmer:DDA	NHR	HNR
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109	0.04374	...	0.06545	0.02211	21.033
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	...	0.09403	0.01929	19.085
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633	0.05233	...	0.08270	0.01309	20.651
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	...	0.08771	0.01353	20.644
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	...	0.10470	0.01767	19.649

5 rows × 24 columns

Data Preparation

In [7]:
```python
# Get the features (columns except ‘status’) and labels (those in the ‘status’ column) from the dataset
## The features are all the columns except ‘status’, and the labels are those in the ‘status’ column

features = data.loc[:,data.columns!='status'].values[:,1:]
labels = data.loc[:,'status'].values
```

In [8]:
```python
# The ‘status’ column has values 0 and 1 as labels; get the counts of these labels for both, 0 and 1

print(labels[labels==1].shape[0], labels[labels==0].shape[0])
```

```
147 48
```

There are 147 ones and 48 zeros in the status column in the dataset

In [9]:
```python
# Initialize a MinMaxScaler and scale the features to between -1 and 1 to normalize them
# Scale the features to between -1 and 1

scaler = MinMaxScaler((-1,1)) # transforms the features by scaling them to a given range (between -1 and 1) to normalize
x = scaler.fit_transform(features) # fit_transform() fits to the data and then transforms it (we don’t need to scale the labels)
y = labels

# It is a good habit to scale the data so that the algorithm will better fit the data
# It is a rear case to get a higher accuracy without scaling
```

Model Selection and Training

Split data to training and testing

In [10]:
```python
# Split the dataset into 80% training and 20% testing

x_train, x_test, y_train, y_test = train_test_split (x, y, test_size = 0.2, random_state = 7)
```

Initialize an XGBClassifier and train the model

In [11]:
```python
# Initialize an XGBClassifier and train the model
# This classifies using eXtreme Gradient Boosting - using gradient boosting algorithms for modern data science problems
# It's under the category of Ensemble Learning in ML, where we train & predict using many models to produce 1 superior output

# Train the model
model = XGBClassifier()
model.fit(x_train,y_train)
```

Out[11]:

▾ XGBClassifier

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, gpu_id=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, ...)
```

Result

Calculate the accuracy of the model

In [12]:
```python
# Generate y_pred (predicted values for x_test) and calculate the accuracy of the model then print it

y_pred = model.predict(x_test)
print (accuracy_score(y_test, y_pred)*100)
```

```
94.87179487179486
```

Summary

Using the train_test_split() function, we split the inputs and the output into 2 parts containing 80% (named x_train and y_train, to train the model) and 20% (to test and find the accuracy) data. The resulting model yielded an accuracy of 94.87% in detecting the presence of Parkinson’s disease in an individual, which is great considering the number of lines of code.

End