

**#Description** 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. The original study published the feature extraction methods for general voice disorders.

1. Matrix column entries (attributes):
2. name - ASCII subject name and recording number
3. MDVP:Fo(Hz) - Average vocal fundamental frequency
4. MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
5. MDVP:Flo(Hz) - Minimum vocal fundamental frequency
6. MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several
7. measures of variation in fundamental frequency
8. MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude
9. NHR,HNR - Two measures of ratio of noise to tonal components in the voice
10. status - Health status of the subject (one) - Parkinson's, (zero) - healthy
11. RPDE,D2 - Two nonlinear dynamical complexity measures
12. DFA - Signal fractal scaling exponent
13. spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation

<https://www.kaggle.com/datasets/thecansin/parkinsons-data-set>

## Step 1 : Importing the Libraries

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.metrics import accuracy_score
```

## Step 2: Loading the dataset

```
# Loading the csv data to a Pandas DataFrame
parkinsons_data = pd.read_csv('/content/parkinsons.csv')
```

## Step 3: Exploratory Data Analysis

Exploratory Data Analysis (EDA), also known as Data Exploration, is a step in the Data Analysis Process, where a number of techniques are used to better understand the dataset being used.

### 3.1) Understanding Your Variables

- 3.1.1) Head of the dataset
- 3.1.2) The shape of the dataset
- 3.1.3) List types of columns