# ­Parkinson’s Disease Detection

## Objectives:

* Select best Machine Learning Model for the data
* Save the model for future use

## Pre-Processing

* Load Dataset and analyze the features
* Total No of features: 452
* No nan values found in the dataset.
* No feature adding and extraction as dataset explanation is NULL
* **Drop** ID as no use of that feature
* Extracting X and y where X is data (Independent Vars) and y is the output (Dependent Var)
* X = All labels except class, as class is the target variable
* y = class
* Label Encoding for class is required as the **class** is a string. (Two possible values = H, P)
* Where H represents Healthy, and P represents Patient
* As we know this will be Binary Classification Either 0 or 1
* The most used Binary classification Methods will be utilized and are the following will be the models that will be used to train the data.

Models need to be compared:

* Logistic Regression
* SVM
* Random Forest Classifier
* XG Boost Classifier
* Light GBM
* Decision Tree

**The following metrics will be used for comparing:**

1. Accuracy Score
2. F1 Score
3. ROC Curve

Note: Scaling Data is Required for the Models Logistic Regression and SVM

Keeping in that mind we need to frame conditions to pass scaled data for both the models.

A screenshot of a computer screen

AI-generated content may be incorrect.

From the above table we can conclude

1. The XG Boost Classifier has the highest Accuracy Score along with F1 Score
2. The Area under Curve (AUC) Computed using ROC Curve Function also states the same that XG Boost Algorithm has highest probability of detecting the disease effectively
3. The image of ROC Curve is given below

A graph with a line

AI-generated content may be incorrect.­­­­­­­­