A
Major Project

On

Parkinson's Disease Detection using voice measurements

(Submitted in partial fulfilment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

B. Harshith Reddy (167R1A05J5)

M. Sreevathsava Theertha (167R1A05M6)

M. Yogendra (167R1A05M4)

Under the Guidance of
Dr. Raj Kumar Patra
(Associate Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi)
Recognized Under Section 2(f) & 12(B) of the UGC Act.1956, Kandlakoya (V), Medchal Road
Hyderabad - 501401.

2016 - 2020

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled "**Parkinson's Disease Detection using Voice Measurements**" being submitted by **B. Harshith Reddy, M. Sreevathsava Theertha, M. Yogendra** bearing the Roll Numbers **167R1A05J5, 167R1A05M6, 167R1A05M4** respectively in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering of the Jawaharlal Nehru Technological University Hyderabad, during the year 2019-2020. It is certified that they have completed the project satisfactorily.

INTERNAL GUIDE | **DIRECTOR**

Dr. Raj Kumar Patra
(Associate Professor)

Dr. A. Raji Reddy

Head Of Department

Dr. K. Srujan Raju

EXTERNAL EXAMINER

Submitted for viva voce Examination held on _____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. We take this opportunity to express my profound gratitude and deep regard to my guide

Dr. Raj Kumar Patra, Associate Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to Project Review Committee (PRC) Coordinators: **Dr. Raj Kumar Patra, Mr. K. Murali , Dr. B. Krishna** and **Dr. T. S. Mastan Rao** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to the Head of the Department **Dr. K. Srujan Raju** for providing excellent infrastructure and a nice atmosphere for completing this project successfully.

We are obliged to our Director **Dr. A. Raji Reddy** for being cooperative throughout the course of this project. We would like to express our sincere gratitude to our Chairman Sri. **Ch. Gopal Reddy** for his encouragement throughout the course of this project

The guidance and support received from all the members of **CMR TECHNICAL CAMPUS** who contributed and who are contributing to this project, was vital for the success of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement without which this assignment would not be possible. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project.

B. Harshith Reddy (167R1A05J5)

M. Sreevathsava Theertha (167R1A05M6)

M. Yogendra (167R1A05M4)

ABSTRACT

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.

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.

To build a model to accurately detect the presence of Parkinson's disease in an individual.

In this Python machine learning project, using the ***Python libraries*** scikit-learn, numpy, pandas, and xgboost, we will build a model using an XGBClassifier. We'll load the data, get the features and labels, scale the features, then split the dataset, build an XGBClassifier, and then calculate the accuracy of our model.

Table of contents

List of Figures

INTRODUCTION

1.1 INTRODUCTION

Project’s main aim is to detect parkinson’s disease using simple machine learning algorithms to predict more accurate and relevant predictions in less time and cost as there is a lot of open source data of the previous observations of many people having parkinson’s disease.

1.2 PROJECT SCOPE

This has been developed to override the problems prevailing in the practicing manual system. The operations which are performed in the manual system takes more time as they depend on more medical test paperwork required. We need to develop an automated system where we can easily identify the disease and take preventive measures before it increases the severity level. There is no particular way of testing the existence of the disease i.e. Blood test or urine test. The only way to identify is by taking the symptoms into consideration and undergoing the analysis. It’s a big task for a human to compare and analyse the voice. So the vocal recordings are divided into a few important features to identify the level of jitter and shimmering in the voice. these features are scaled up to train the model with supervised learning using ensemble classifier i.e. XGBoost classifier is more efficient and robust way to train a model and this classifier uses decision trees.

1.3 PARKINSONS DISEASE

Parkinson's disease could be a progressive systema nervosum disorder that affects movement. Symptoms start gradually, sometimes starting with a barely noticeable tremor in exactly one hand. Tremors are common, but the disorder also commonly causes stiffness or slowing of movement.

In the early stages of Parkinson's disease, your face may show little or no expression. Your arms might not swing once you walk. Your speech may become soft or slurred. Parkinson's disease symptoms worsen as your condition progresses over time.

Although Parkinson's disease cannot be cured, medications might significantly improve your symptoms. Occasionally, your doctor may suggest surgery to manage certain regions of your brain and improve your symptoms.

Symptoms

Parkinson's disease signs and symptoms can be different for everyone. Early signs may be mild and go unnoticed. Symptoms often begin on one side of your body and usually remain worse on that side, even after symptoms begin to affect both sides.

Tremor. A tremor, or shaking, usually begins in a limb, often your hand or fingers. You may rub your thumb and forefinger back-and-forth, known as a pill-rolling tremor. Your hand may tremor when it's at rest.

Slowed movement (bradykinesia). Over time, Parkinson's disease may slow your movement, making simple tasks difficult and time-consuming. Your steps may become shorter when you walk. It may be difficult to get out of a chair. You may drag your feet as you try to walk.

Rigid muscles. Muscle stiffness may occur in any part of your body. The stiff muscles can be painful and limit your range of motion.

Impaired posture and balance. Your posture may become stooped, or you may have balance problems as a result of Parkinson's disease.

Loss of automatic movements. You may have a decreased ability to perform unconscious movements, including blinking, smiling or swinging your arms when you walk.

Speech changes. You may speak softly, quickly, slur or hesitate before talking. Your speech may be more of a monotone rather than with the usual inflections.

Causes

In Parkinson's disease, certain nerve cells (neurons) in the brain gradually break down or die. Many of the symptoms are due to a loss of neurons that produce a chemical messenger in your brain called dopamine. When dopamine levels decrease, it causes abnormal brain activity, leading to symptoms of Parkinson's disease.

The cause of Parkinson's disease is unknown, but several factors appear to play a role, including:

- **Your genes.** Researchers have identified specific genetic mutations that can cause Parkinson's disease. But these are uncommon except in rare cases with many family members affected by Parkinson's disease.
  However, certain gene variations appear to increase the risk of Parkinson's disease but with a relatively small risk of Parkinson's disease for each of these genetic markers.
- **Environmental triggers.** Exposure to certain toxins or environmental factors may increase the risk of later Parkinson's disease, but the risk is relatively small.

Researchers have also noted that many changes occur in the brains of people with Parkinson's disease, although it's not clear why these changes occur. These changes include:

- **The presence of Lewy bodies.** Clumps of specific substances within brain cells are microscopic markers of Parkinson's disease. These are called Lewy bodies, and researchers believe these Lewy bodies hold an important clue to the cause of Parkinson's disease.
- **Alpha-synuclein is found within Lewy bodies.** Although many substances are found within Lewy bodies, scientists believe an important one is the natural and widespread protein called alpha-synuclein (a-synuclein). It's found in all Lewy bodies in a clumped form that cells can't break down. This is currently an important focus among Parkinson's disease researchers.

1.4 Risk factors

Risk factors for Parkinson's disease include:

- **Age.** Young adults rarely experience Parkinson's disease. It ordinarily begins in middle or late life, and the risk increases with age. People usually develop the disease around age 60 or older.
- **Heredity.** Having a close relative with Parkinson's disease increases the chances that you'll develop the disease. However, your risks are still small unless you have many relatives in your family with Parkinson's disease.
- **Sex.** Men are more likely to develop Parkinson's disease than are women.

1.5 Complications

Parkinson's disease is often accompanied by these additional problems, which may be treatable:

- **Thinking difficulties.** Patient with parkinson disease may experience cognitive problems (dementia) and thinking difficulties. These usually occur in the later stages of Parkinson's disease. Such cognitive problems aren't very responsive to medications.
- **Depression and emotional changes.** Patient with parkinson disease may experience depression, sometimes in the very early stages. Receiving treatment for depression can make it easier to handle the other challenges of Parkinson's disease.

  Patient with parkinson disease may also experience other emotional changes, such as fear, anxiety or loss of motivation. Doctors may give Patient with parkinson disease medications to treat these symptoms.
- **Swallowing problems.** Patient with parkinson disease may develop difficulties with swallowing as Patient’s condition progresses. Saliva may accumulate in Patient’s mouth due to slowed swallowing, leading to drooling.
- **Chewing and eating problems.** Late-stage Parkinson's disease affects the muscles in Patient’s mouth, making chewing difficult. This can lead to choking and poor nutrition.
- **Sleep problems and sleep disorders.** People with Parkinson's disease often have sleep problems, including waking up frequently throughout the night, waking up early or falling asleep during the day.

  People may also experience rapid eye movement sleep behavior disorder, which involves acting out Patient’s dreams. Medications may help Patient’s sleep problems.
- **Bladder problems.** Parkinson's disease may cause bladder problems, including being unable to control urine or having difficulty urinating.
- **Constipation.** Many people with Parkinson's disease develop constipation, mainly due to a slower digestive tract.

2. Methodologies used in present work

2.1 Introduction

❖ Objectives

● Artificial Intelligence (AI) and understand its relationship with data.

● Machine Learning (ML) and understand its relationship with Artificial Intelligence.

● Machine Learning approach and its relationship with data science.

● Identify the application.

❖ Definition of Artificial Intelligence

Artificial Intelligence refers to intelligence displayed by machines that simulate human intelligence.

❖ The Emergence of Artificial Intelligence

The data economy with its vast reservoir is enabling unprecedented innovation in data sciences, the field which deals with extracting useful information and insights from the available data.

Data science is going toward a new paradigm where one can teach machines to learn from data and derive a variety of useful insights. This is known as Artificial Intelligence.

❖ Artificial Intelligence in Practice

Given below in this Machine Learning tutorial are a few areas where AI is used widely.

● Self-driving cars

● Applications like Siri that understand and respond to human speech

● Google's AlphaGo AI has defeated many Go champions such as Ke Jie

● Implementing AI in chess

● Amazon ECHO product (home control chatbot device)

● Hilton using Connie – concierge robot from IBM Watson

❖ **Data Facilitates Artificial Intelligence Products**

Amazon pulls in data from its user database to recommend products to users. This functionality helps bring in more users. More users generate even more data that help enhance the recommendations even further.

2.2. Machine Learning:

The capability of Artificial Intelligence systems to learn by extracting patterns from data is known as Machine Learning.

❖ **Machine Learning Benefits**

● Powerful Processing

● Better Decision Making & Prediction

● Quicker Processing

● Accurate

● Affordable Data Management

● Inexpensive

● Analysing Complex Big Data

2.3 Features of Machine Learning

● Machine Learning is computing-intensive and generally requires a large amount of training data.

● It involves repetitive training to improve the learning and decision making of algorithms.

● As more data gets added, Machine Learning training can be automated for learning new data

patterns and adapting its algorithm.

Example: Learning from new spam words or new speech (also called incremental learning)

❖ **Artificial Intelligence and Machine Learning**

Machine Learning is an approach or subset of Artificial Intelligence that is based on the idea that

machines can be given access to data along with the ability to learn from it.

❖ **Traditional Programming vs. Machine Learning Approach**

Traditional programming relies on hard-coded rules.

Machine Learning relies on learning patterns based on sample data.

As you go from rule-based systems to the deep learning ones, more complex features and input-output relationships become learnable.

● Data Science and Machine Learning go hand in hand. Data Science helps evaluate data for

Machine Learning algorithms.

● Data science is the use of statistical methods to find patterns in the data.

● Statistical machine learning uses the same math and techniques as data science.

● These techniques are integrated into algorithms that learn and improve on their own.

● Machine Learning facilitates Artificial Intelligence as it enables machines to learn from the

patterns in data.

❖ **Machine Learning Techniques**

1. Classification

2. Categorization

3. Clustering

4. Trend analysis

5. Anomaly detection

6. Visualization

7. Decision making

❖ **Machine Learning Algorithms**

Let us understand Machine Learning Algorithms in detail.

● Machine Learning can learn from labeled data (known as supervised learning) or unlabelled

data (known as unsupervised learning).

● Machine Learning algorithms involving unlabelled data, or unsupervised learning, are more complicated than those with the labeled data or supervised learning

● Machine Learning algorithms can be used to make decisions in subjective areas as well.

Examples:

● Logistic Regression can be used to predict which party will win at the ballots.

● Naïve Bayes algorithm can separate valid emails from spam.

Applications of Machine Learning

Some of the applications of Machine learning mentioned below.

- Image Processing
- Robotics
- Data Mining
- Video Games
- Text Analysis
- Healthcare

Applications Uses

Image Processing

- Image tagging and recognition
- Self-driving cars
- Optical Character Recognition (OCR)

Robotics

- Human simulation
- Industrial robotics

Data Mining

- Anomaly detection
- Grouping and Predictions
- Association rules

Text Analysis

- Sentiment Analysis
- Spam Filtering
- Information Extraction

2.4 Ensemble learning:

Gradient Boosting Machines fit into a category of ML called Ensemble Learning, which is a branch of ML methods that train and predict with many models at once to produce a single superior output.

Ensemble learning is broken up into three primary subsets:

- Bagging
- Stacking
- Boosting

Ensemble learning is the process by which multiple models, such as classifiers or experts, are strategically generated and combined to solve a particular computational intelligence problem. Ensemble learning is primarily used to improve the (classification, prediction, function approximation, etc.) performance of a model, or reduce the likelihood of an unfortunate selection of a poor one. Other applications of ensemble learning include assigning a confidence to the decision made by the model, selecting optimal (or near optimal) features, data fusion, incremental learning, nonstationary learning and error-correcting. This article focuses on classification related applications of ensemble learning, however, all principle ideas described below can be easily generalized to function approximation or prediction type problems as well.

1. Bagging : Bagging tries to implement similar learners on small sample populations and then takes a mean of all the predictions. In generalized bagging, you can use different learners on different populations. As you can expect this helps us to reduce the variance error.

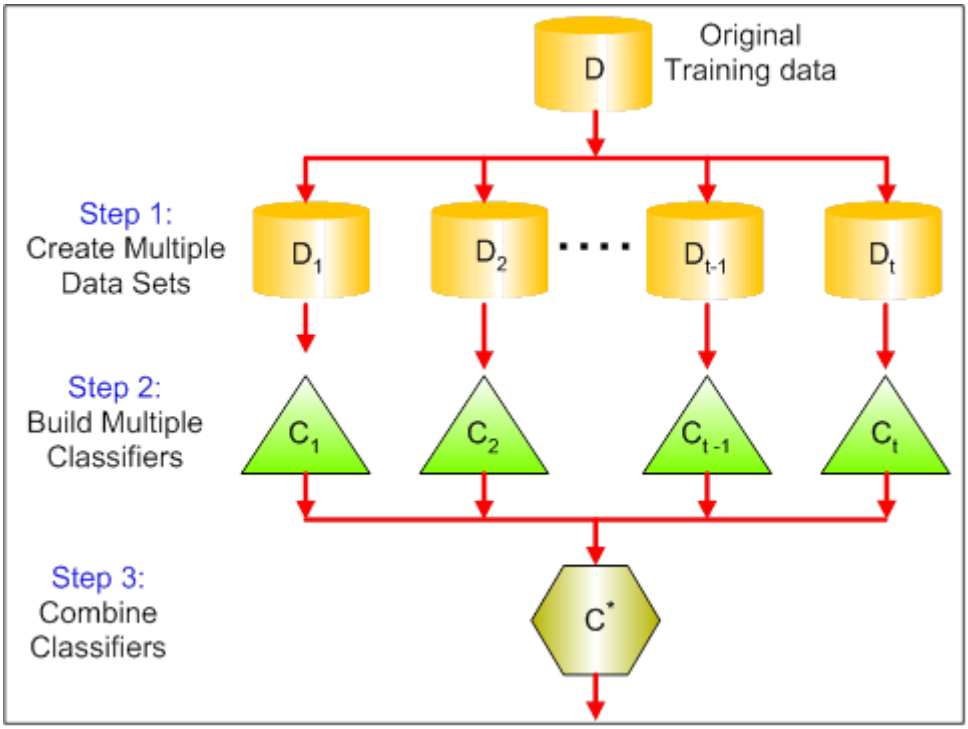


figure 1.1

2. Boosting : Boosting is an iterative technique which adjust the weight of an observation based on the last classification. If an observation was classified incorrectly, it tries to increase the weight of this observation and vice versa. Boosting in general decreases the bias error and builds strong predictive models. However, they may sometimes over fit on the training data.

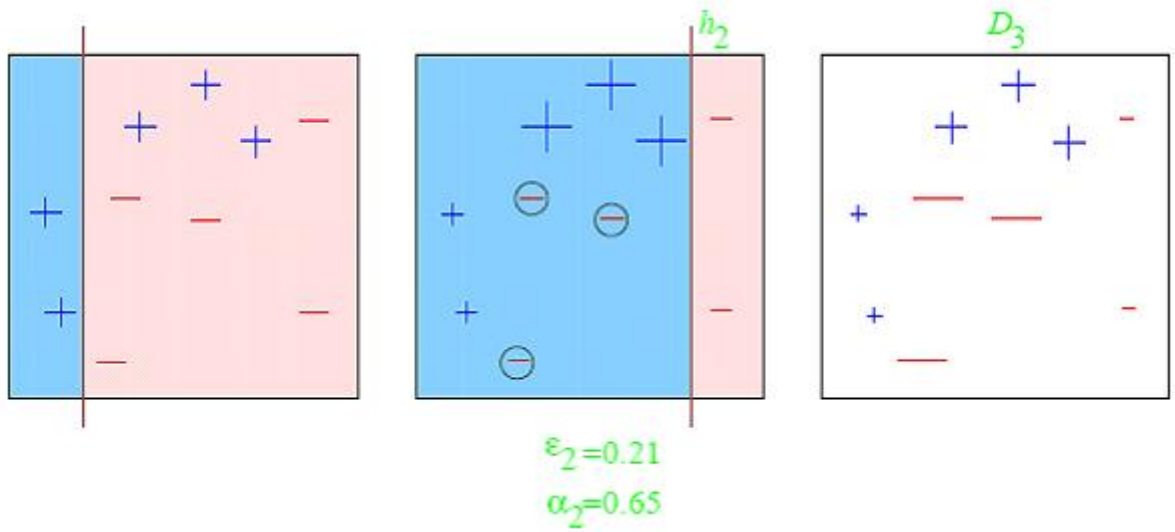


figure 1.2

3. Stacking : This is a very interesting way of combining models. Here we use a learner to combine output from different learners. This can lead to decrease in either bias or variance error depending on the combining learner we use.

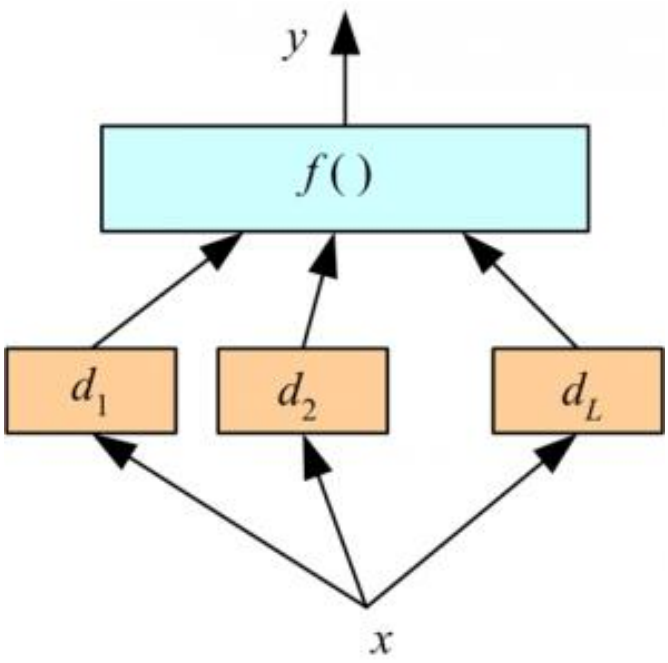


figure 1.3

2.5 XGBoost Classifier: XGBoost is a popular technique and a neat alternative to traditional regression/neural nets. It stands for EXtreme Gradient Boosting, and basically builds something of a decision tree to compute gradients. Here's a popular graphic from the XGBoost website as an example:

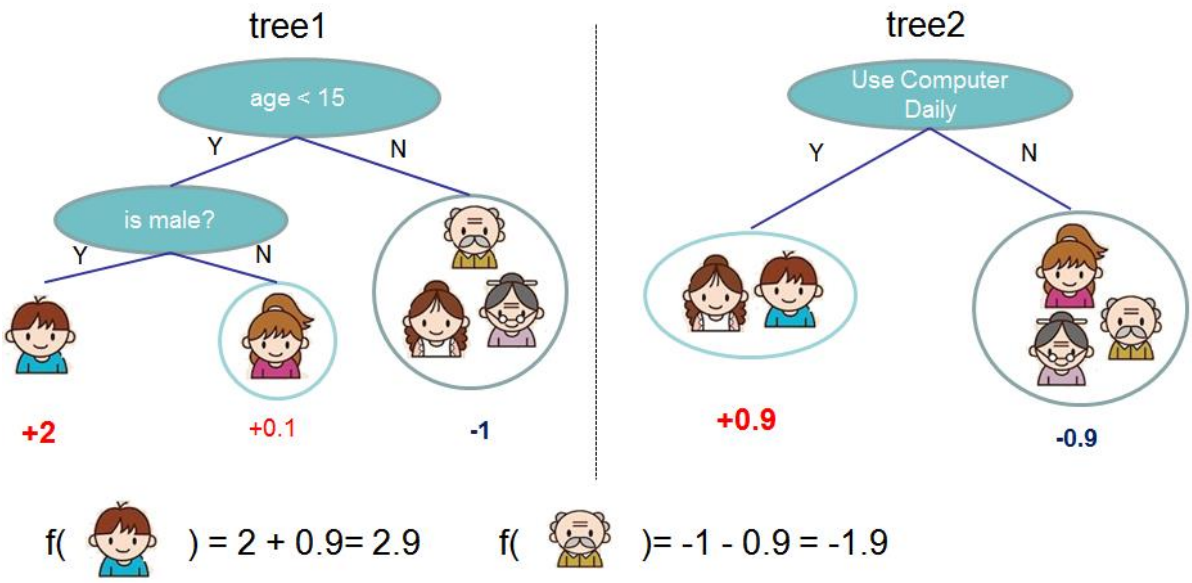


figure 1.4

This sounds simple in practice but can be extremely powerful. Take, for example, Parkinson's detection: we have several metrics that we can analyze and ultimately we need to diagnose Parkinson's (classification!). This is a perfect problem for XGBoost (especially since there is a single output so we don't need to use the MultiOutput wrapper — more on that later).

XGBoosting is extremely powerful and can definitely be a useful tool for Patient's next project! This goes much deeper though — for multiple outputs you'd need a MultiOutput model (SciKit Learn has a great wrapper for this), and for getting more accurate you'll need to fine tune Patient's XGBoost model. In the Jupyter notebook before (and in the repo linked below) I explore using Keras for the continuous data, but the MultiOutput wrapper from sklearn can serve almost as a drop-in replacement for the Keras model.

Advantages:

1. Regularization: XGBoost has in-built L1 (Lasso Regression) and L2 (Ridge Regression) regularization which prevents the model from overfitting. That is why, XGBoost is also called regularized form of GBM (Gradient Boosting Machine).

While using Scikit Learn libarary, we pass two hyper-parameters (alpha and lambda) to XGBoost related to regularization. alpha is used for L1 regularization and lambda is used for L2 regularization.

2. Parallel Processing: XGBoost utilizes the power of parallel processing and that is why it is much faster than GBM. It uses multiple CPU cores to execute the model.

While using Scikit Learn libarary, nthread hyper-parameter is used for parallel processing. nthread represents number of CPU cores to be used. If you want to use all the available cores, don't mention any value for nthread and the algorithm will detect automatically.

3. Handling Missing Values: XGBoost has an in-built capability to handle missing values. When XGBoost encounters a missing value at a node, it tries both the left and right hand split and learns the way leading to higher loss for each node. It then does the same when working on the testing data.

4. **Cross Validation**: XGBoost allows user to run a cross-validation at each iteration of the boosting process and thus it is easy to get the exact optimum number of boosting iterations in a single run. This is unlike GBM where we have to run a grid-search and only a limited values can be tested.

5. Effective Tree Pruning: A GBM would stop splitting a node when it encounters a negative loss in the split. Thus it is more of a greedy algorithm. XGBoost on the other hand make splits upto the max_depth specified and then start pruning the tree backwards and remove splits beyond which there is no positive gain.

2.6 Data Drives Machine Learning.

As more data is available, we have better information to provide patients. Predictive algorithms and machine learning can give us a better predictive model of mortality that doctors can use to educate patients.

But machine learning needs a certain amount of data to generate an effective algorithm. Much of machine learning will initially come from organizations with big datasets. Health Catalyst

is developing Collective Analytics for Excellence (CAFÉ™), an application built on a national de-identified repository of healthcare data from enterprise data warehouses (EDWs) and third-party data sources. It is enabling comparative effectiveness, research, and producing unique, powerful machine learning algorithms. CAFÉ provides a collaboration among our healthcare system partners, big and small.

2.7 Prerequisites

You'll need to install the following libraries with pip:

pip install numpy pandas

pip install numpy sklearn

pip install numpy xgboost

You'll also need to install Jupyter Lab, and then use the command prompt to run it:

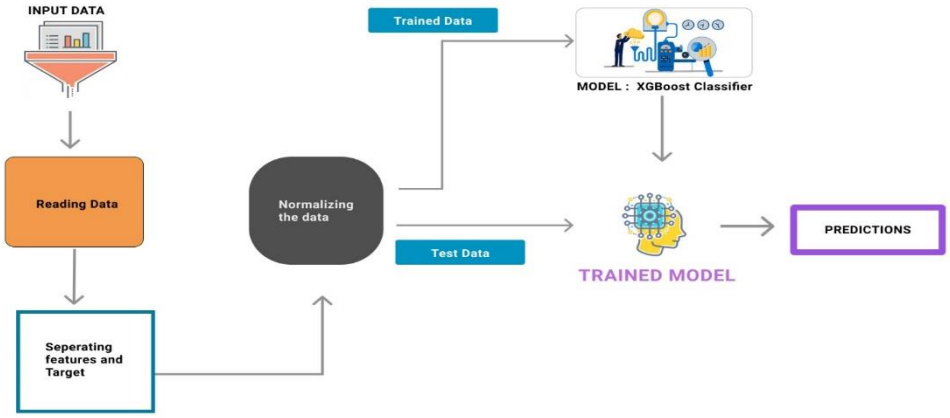
C:\Users\Desktop>jupyter lab

This will open a new JupyterLab window in your browser. Here, you will create a new console and type in your code, then press **Shift+Enter** to execute one or more lines at a time.

3. Architecture

3.1 Introduction

The below architecture shows the procedure followed for disease detection using machine learning, starting from input to final prediction.



3.2 Description

Input Data: Input data is generally in .csv format or .data format where the data is fetched and mapped in the data framed from the source columns.

Reading Data: Pandas library is used to read the data into the data frame

Separating features and target: In this following step we are going to separate the features which we take to training the model by giving the target value i.e. 1/0 for the particular combination of features.

Normalization: Normalization is very important step while we are drealing with the large values in the features as the higher bit integers will cost high computational power and time. To achieve the efficiency in computation we are going to normalize the data values.

Training and test data: Training data is passed to the XGBoost classifier to train the model. Test data is used to test the trained model whether it is making correct predictions or not.

XGBoost Classifier: the purpose of choosing the xgboost classifier for this project is the efficiency and accuracy that we have observed when compared to other classifiers.

Finally, we pass the test data to the trained model instance to make predictions.

4. System Analysis

4.1 System Analysis:

System Analysis is the important phase in the system development process. The System is studied to the minute details and analysed. The system analyst plays an important role of an interrogator and dwells deep into the working of the present system. In analysis, a detailed study of these operations performed by the system and their relationships within and outside the system is done.

A key question considered here is, “what must be done to solve the problem?

4.2 Existing system

As the most common neurodegenerative movement disorder, Parkinson’s Disease (PD) has enormous impact on health care systems around the world. While traditionally much emphasis has been placed on the motor aspects of the disease, it is increasingly recognized that PD is a multisystem disorder, and many of the non-motor features play at least as important a role in impacting overall quality of life.

There are a number of new technologies that may greatly assist in early diagnosis, monitoring and treatment of this disease. Modern neuroimaging technologies, including MRI, EEG, MEG, PET, and CT are able to noninvasively examine the diseased brain and investigate the underlying neural systems in PD, resulting in powerful approaches for disease detection and monitoring. New data fusion methods can combine information from complementary technologies allowing for comprehensive assessment. New lightweight and wireless sensors can monitor movement, electrodermal responses, temperature and heart rate. Non-invasive electrical stimulation can modulate brain activity, providing new unexplored avenues of treatment.

In order to have an overview of new technologies for PD research and create a platform to discuss how these can be used to diagnosis, assess and treat PD, for this Research Topic, we are interested in high quality original research and review articles. Potential subtopics include

but are not limited to the following:

- Biomarkers of PD based on structural or functional minitrial stimulation.
- Data fusion methods pertinent to disease-related biodata.
- Motion Biomarkers Showing Maximum Contrast Between Healthy Subjects and Parkinson's Disease Patients Treated With Deep Brain Stimulation of the Subthalamic Nucleus. A Pilot Study
- Alterations of Regional Homogeneity in Parkinson's Disease Patients With Freezing of Gait: A Resting-State fMRI Study
- Altered Global Synchronizations in Patients With Parkinson's Disease: A Resting-State fMRI Study

4.3 Proposed System

Using machine learning algorithms, we can achieve Parkinson's disease detection with much appropriate results in very less time and cost using powerful, robust and efficient classifiers available using decision trees.

Details of the Dataset:

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

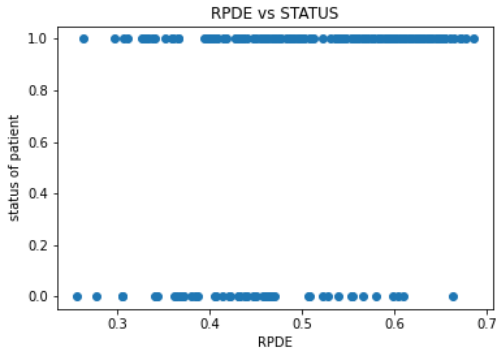
Data Set Information:

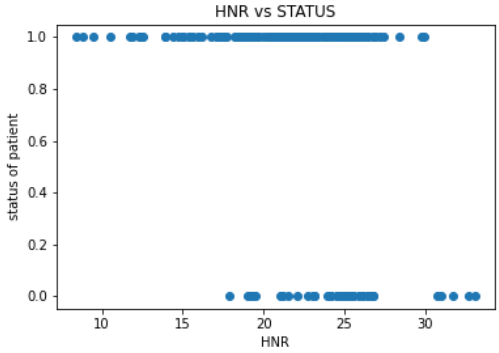
This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD.

Attributes of the Data Used:

Name - ASCII subject name	MDVP: Shimmer
MDVP: Fo(Hz) - Average vocal	MDVP: Shimmer(dB)
MDVP: Fhi(Hz) - Maximum vocal	Shimmer:APQ3
MDVP: Flo(Hz) - Minimum vocal	Shimmer:APQ5
MDVP: Jitter(%)	Shimmer: DDA - Several measures of variation in amplitude
MDVP: Jitter(Abs)	NHR,HNR - Two measures of ratio of noise to tonal components in the voice
MDVP:RAP	RPDE,D2 - Two nonlinear dynamical complexity measures
MDVP:PPQ	DFA - Signal fractal scaling exponent spread1,spread2,PPE
MDVP:APQ	status - Health status of the subject (one) - Parkinson's, (zero) - healthy
Jitter: DDP - Several measures of variation in fundamental frequency	

Relation between prominent features and status variable:





- RPDE is a nonlinear dynamical complexity measure VS The status of the patient (0 – healthy and 1 – Diseased)
- HNR is a measure of ratio of noise to tonal components in the voice VS The status of the patient (0 – healthy and 1 – Diseased)

5. Feasibility Study

The feasibility of the project is analysed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company.

Three key considerations involved in the feasibility analysis are

- Economic Feasibility
- Technical Feasibility
- Social Feasibility

5.1 Economic Feasibility

This study is carried out to check the economic impact that the system will have on the hostels. The amount of funds that the hostels can pour into the research and development of the system is limited. The expenditures must be justified.

5.2 Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

5.3 Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of responding the user to his request efficiently. The user must not feel hesitated, instead must accept it as a necessity.

6. System Requirement

6.1 Hardware Requirement

Processor	:	core i3 or higher
Hard Disk	:	500GB or higher
RAM	:	4GB or higher

6.2 Software Requirement

Operating System	:	Windows / Linux
IDE	:	Jupyter Notebook
Programming Language	:	Python 3
Python Packages	:	Numpy, Pandas, Sklearn, XGBoost

7. Project Implementation

Importing Required 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
```

Reading the DataSet

```
df=pd.read_csv('parkinsons.data')
df.head()
```

	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	M
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	

5 rows × 24 columns

figure 1.6

Separating Features and Target variable from the dataset

```
features=df.loc[:,df.columns!='status'].values[:,1:]
labels=df.loc[:,'status'].values
print(labels[labels==1].shape[0], labels[labels==0].shape[0])
```

Normalising the data

```
scaler=MinMaxScaler((-1,1))

x=scaler.fit_transform(features)

y=labels
```

Splitting the DataSet into Train and Test Data

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

Model Definition

```
model=XGBClassifier()

model.fit(x_train,y_train)

XGBClassifier(base_score=0.5, booster=None, colsample_bylevel=1,

              colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,

              importance_type='gain', interaction_constraints=None,

              learning_rate=0.300000012, max_delta_step=0, max_depth=6,

              min_child_weight=1, missing=nan, monotone_constraints=None,

              n_estimators=100, n_jobs=0, num_parallel_tree=1,

              objective='binary:logistic', random_state=0, reg_alpha=0,

              reg_lambda=1, scale_pos_weight=1, subsample=1, tree_method=None,

              validate_parameters=False, verbosity=None)
```

Making predictions and Calculating the accuracy

```
y_pred=model.predict(x_test)

print(accuracy_score(y_test, y_pred)*100)
```

94.87179487179486

8.TESTING

9.1. INTRODUCTION TO TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

9.2. TYPES OF TESTING

● UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

● INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic 31outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the

combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

- **FUNCTIONAL TESTING**

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

- Valid Input : identified classes of valid input must be accepted.
- Invalid Input : identified classes of invalid input must be rejected.
- Functions : identified functions must be exercised.
- Output : identified classes of application outputs must be exercised.
- Systems/Procedures: interfacing systems or procedures must be invoked.

9. Conclusion and Future Enhancements

9.1 Conclusion

In this Python machine learning project, we learned to detect the presence of Parkinson's Disease in individuals using various factors. We used an XGBClassifier for this and made use of the sklearn library to prepare the dataset. This gives us an accuracy of 94.87%, which is great considering the number of lines of code in this python project.

Using machine learning algorithms we can achieve Parkinson's disease detection with much appropriate results in very less time and cost using powerful, robust and efficient classifiers available using decision trees.

We'll load the data, get the features and labels, scale the features, then split the dataset, build an XGBClassifier, and then calculate the accuracy of our model.

9.2 Future Enhancements:

Current there is no more efficient data available as open source as medical Industry's current research work is being done to cure the disease. To get more accurate preditions loat of data is required with extra features i.e currently we are identifying this using jitter and shimmer in the voice in the future we can also consider the standing position of the patient using image processing and considering many more symptoms detections becomes more efficient .we can make this an open source by providing a web interface where the patient's data can be given as an input and results will be fetched through the interface.

10. BIBLIOGRAPHY

[1] DATA : https://archive.ics.uci.edu/ml/machine-learning-databases/parkinsons/

[2] INSPIRATION : https://data-flair.training/

[3] DISEASE www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/

[4] XGBOOST https://towardsdatascience.com/exploring-xgboost-4baf9ace0cf6

[5] ENSEMBLE https://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf