**Parkinson’s Disease Classification using Machine Learning**

**Abstract**

Parkinson’s Disease (PD) is a progressive neurological disorder that affects millions worldwide. Early detection is crucial for effective treatment and improved patient outcomes. This project employs machine learning techniques to classify PD based on voice measurements. Using a dataset of vocal attributes, various algorithms are evaluated to identify the most accurate model. The system is further deployed as a web application to provide a user-friendly interface for disease prediction.

## 1.Introduction

Parkinson’s Disease primarily impacts motor functions due to the degeneration of dopamine-producing neurons in the brain. Traditional diagnostic methods rely on clinical evaluations, which can be subjective and time-consuming. Machine learning provides a data-driven, objective approach to early PD detection, potentially improving diagnosis accuracy and efficiency.

## 2. System Analysis

### **2.1 Existing System**

Currently, Parkinson’s Disease diagnosis depends on physical examinations, neurological assessments, and medical imaging. These methods are expensive, require expert analysis, and may lead to misdiagnosis due to human subjectivity.

### **2.2 Proposed System**

The proposed system utilizes machine learning models trained on voice data to classify individuals based on Parkinson’s Disease presence. It automates the prediction process, providing a faster, more objective assessment method.

## 3. System Specification

### **Hardware Requirements**

* Processor: Intel Core i5 or higher
* RAM: 8GB or more
* Storage: 256GB SSD or higher
* GPU: Optional (for deep learning enhancement

### **Software Requirements**

* Operating System: Windows
* Programming Language: Python
* Libraries: Pandas, Scikit-learn, Flask, Joblib
* Web Framework: Flask for deployment

## 4. Dataset Description

The dataset used in this project is sourced from the UCI Machine Learning Repository and contains multiple voice recordings from individuals, both with and without Parkinson's Disease. The dataset includes 24 biomedical voice measurements that capture variations in fundamental frequency, jitter, shimmer, and harmonics-to-noise ratio, which are critical indicators of Parkinson’s Disease.

### **Key Features in the Dataset**

* **MDVP:Fo(Hz)** - Fundamental frequency of the voice
* **MDVP:Fhi(Hz)** - Highest frequency
* **MDVP:Flo(Hz)** - Lowest frequency
* **Jitter (%)** - Variability in frequency
* **Shimmer (dB)** - Variability in amplitude
* **HNR (dB)** - Harmonic-to-noise ratio
* **Spread1, Spread2, D2** - Nonlinear dynamical measures
* **Status** - Classification label (1 = Parkinson’s, 0 = Healthy)

The dataset is preprocessed by handling missing values, normalizing the features, and selecting the most relevant attributes for classification.

## 5. Methodology

* **Data Collection**: Using a dataset containing voice measurements.
* **Preprocessing**: Handling missing values, normalization, and feature selection.
* **Model Training**: Evaluating multiple machine learning algorithms.
* **Deployment**: Developing a Flask web app for real-time prediction.

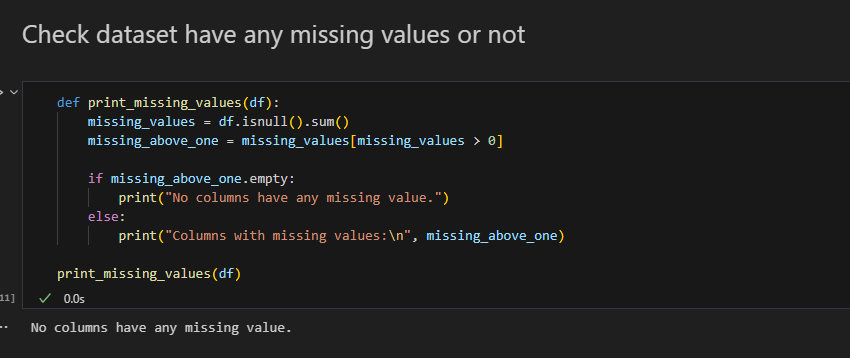
## 6. Module Description

### **Module 1: Data Collection and Preprocessing**

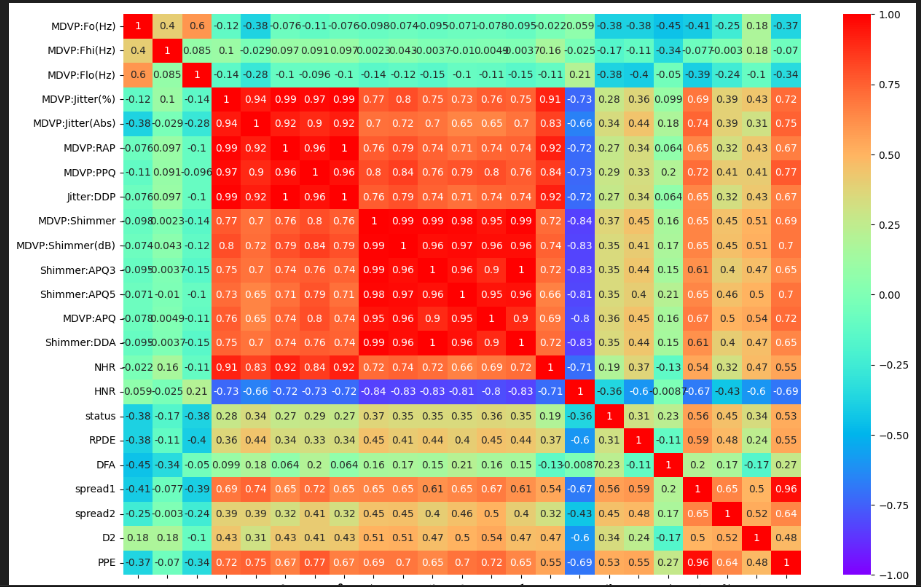
* Collecting voice-based dataset for Parkinson’s detection.
* Handling missing values, removing duplicates, and normalizing features.

### **Module 2: Exploratory Data Analysis**

* Understanding data distribution and correlation between features.
* Identifying feature importance through visualization techniques.



### “ Here I check Any missing values but no missing values occure in the dataset”



### 

### “ Print most important features”

### **Module 3: Model Training and Evaluation**

* Splitting dataset into training and testing sets.
* Training models like Random Forest, SVM, and Logistic Regression.
* Evaluating model performance using accuracy, precision, recall, and F1-score.

### **Module 4: Web Deployment**

* Developing a user-friendly web interface using Flask.
* Integrating the trained model into the web application for real-time predictions.

## 7. Testing

### **Unit Testing**

Each module, including data preprocessing, model training, and deployment, is tested independently to ensure functionality.

### **Integration Testing**

The system is tested as a whole to ensure seamless interaction between modules.

### **Performance Testing**

Evaluating model performance based on:

* Accuracy
* Precision
* Recall
* Confusion Matrix



## 

## 8. Conclusion and Future Enhancements

This project successfully demonstrates a machine learning-based classification system for Parkinson’s Disease. Future enhancements may include:

* Utilizing deep learning models for improved performance.
* Integrating additional biomarkers and patient history for enhanced accuracy.
* Expanding the system for real-time voice-based PD detection.

By incorporating these improvements, the system can provide even more reliable and accessible Parkinson’s Disease classification.

**Github Repository:**[**Parkinson-s-Disease-Detection**](https://github.com/anuselva1905/Parkinson-s-Disease-Detection)