Project: Parkinson Disease Prediction using Machine Learning

- **Objective: Develop a machine learning model to predict whether a person has Parkinson's disease based on health metrics and voice features.**
- **Dataset: Parkinson's Disease Dataset**

Project Goals:

1. Importing Libraries and Dataset:

- **Load required Python libraries:**
  - **Pandas for handling datasets.**
  - **NumPy for numerical operations.**
  - **Matplotlib/Seaborn for data visualization.**
  - **Sklearn for data preprocessing, model training, and evaluation.**
  - **XGBoost for high-performance machine learning classification.**
  - **Imblearn to handle imbalanced data.**
- **Load the dataset using Pandas and explore its structure.**

2. Data Preprocessing:

- **Handle missing values if any.**
- **Normalize/scale numerical features for better model performance.**
- **Encode categorical variables if required.**
- **Check for class imbalance and apply SMOTE (Synthetic Minority Over-sampling Technique) if necessary.**
- **Split dataset into training and testing sets (80%-20%).**

3. Exploratory Data Analysis (EDA):

- **Visualize feature distributions using histograms and boxplots.**
- **Use correlation heatmaps to identify the most important features for prediction.**
- **Identify trends in voice features that contribute to Parkinson's disease.**

4. Model Training and Selection:

- **Train different machine learning models:**
    - **Logistic Regression**
    - **Random Forest**
    - **Support Vector Machine (SVM)**
    - **XGBoost (Extreme Gradient Boosting)**
- **Compare model performance using accuracy, precision, recall, and F1-score.**

5. Model Evaluation and Prediction:

- **Evaluate the best model based on classification metrics.**
- **Test the model with new patient data to predict Parkinson's disease.**
- **Use ROC-AUC curves to assess model performance.**

Conclusion:

- **This model helps in early Parkinson's disease detection using voice and movement-related data.**
- **Future improvements can involve Deep Learning models for better accuracy.**