

# Project Statement: Parkinson's Disease Detection Using Voice Measurements

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects motor and non-motor functions. This project aims to develop a machine-learning model to detect Parkinson's Disease using voice measurements from patients. The dataset contains various voice features extracted from recordings of individuals, both with and without Parkinson's Disease. The goal is to classify whether a patient has Parkinson's Disease based on these voice features.

## Dataset Description

The dataset contains 195 rows (samples) and 24 columns (features). Each row represents a voice recording from an individual, and the columns represent various voice measurements. The dataset includes:

- 23 features: These are voice measurements such as frequency, jitter, shimmer, and other acoustic properties.
- 1 target variable: `status` (binary classification: 1 = Parkinson's Disease, 0 = Healthy).
- Handle missing values (if any) and normalize/standardize the data for better model performance.

## Model Development and Model Evaluation( use appropriate metrics):

- Build a multidimensional Logistic regression model **(KEC 3301-3310)**
- Perform feature selection or dimensionality reduction to identify the most relevant features for classification using principal component analysis. **(KEC 3311-3320)**
- Build a DNN model for the binary classification to predict whether a patient has Parkinson's Disease based on voice measurements. **(KEC 3321- rest)**

## Possible Model Evaluation Metrics:

- Precision, Recall, F1-Score
- MSE
- ROC-AUC (Receiver Operating Characteristic - Area Under Curve)

## Data File:

The file provided ( `**parkinsons.csv**` ) contains the necessary data for this project.

## **Project Report format:**

1. Introduction,
2. Data description,
3. Theory,
4. Python code with description (mandatory)
5. Result description\*, and
6. Conclusion

\*For Logistic regression model Results should includes model description, metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, and related figures

\*For PCA analysis should includes, the amount of variance and number of principal components, factor analysis, the features and principal components contributes in classification

\*Description of DNN model with the description of different layers, model compilation factors such optimizer, activation function, and loss function, accuray, loss with epoch, AUC, ROC etc

## **Submission Guideline:**

Please make a pdf with ID (i.e. project\_ML4\_your\_ID.pdf). Then email to [edgeexam@kuet.ac.bd](mailto:edgeexam@kuet.ac.bd) with a subject Capston Project report ML4 and ID (ex KEC 33..) within Feb 6, 2025 1:00 PM (Thursday).