Project Report: Parkinson's Disease Detection

Project Description: This project aims to detect Parkinson's disease based on a dataset of voice recordings and various acoustic features. The goal is to build a machine learning model that can accurately classify individuals into two groups: those with Parkinson's disease (status=1) and those without (status=0).

Table of Contents

1. Introduction

Parkinson's disease is a neurodegenerative disorder that affects millions of people worldwide. Early detection of this disease can significantly improve the quality of life

for patients. In this project, we employ machine learning techniques to develop a model that can classify individuals as either having or not having Parkinson's disease based on acoustic features extracted from voice recordings.

2. Data Collection and Preprocessing

2.1 Dataset Description

The dataset used for this project is the "P_disease.csv" file, containing 195 rows and 24 columns. Each row represents a voice recording of a patient, and the columns contain various acoustic features. The "status" column is our target variable, with values 1 (indicating Parkinson's disease) and 0 (indicating no Parkinson's disease).

2.2 Data Exploration

Before developing the model, we conducted some data exploration to gain insights into the dataset. Key observations include:

- The dataset contains 195 samples with 24 features.
- The "status" column shows that 147 individuals have Parkinson's disease (status=1), while 48 do not (status=0).

2.3 Data Preprocessing

Data preprocessing was conducted to prepare the dataset for modeling. Key steps include:

- Splitting the data into features (X) and the target variable (Y).
- Splitting the data into training and testing sets (80% train, 20% test).
- Scaling the features using the StandardScaler to standardize the data.

3. Model Development

3.1 Feature Selection

In this project, we used all 22 acoustic features as input variables, excluding the "name" and "status" columns.

3.2 Data Splitting

The dataset was split into training (80%) and testing (20%) sets. This ensures that we have a separate dataset for training and evaluating the model.

3.3 Model Training

We employed a Support Vector Machine (SVM) classifier to build our predictive model. The model was trained on the training data after scaling.

4. Model Evaluation

4.1 Performance Metrics

We evaluated the model's performance using common classification metrics, including accuracy.

4.2 Results

The results of the model evaluation are as follows:

- Accuracy: 88%

5. Conclusion

In conclusion, we successfully developed a machine learning model for Parkinson's disease detection. The model demonstrated promising results in classifying individuals with and without the disease. Early detection of Parkinson's disease can have a significant impact on patient care and quality of life.