Parkinson's Disease Prediction Using Support Vector Machine (SVM)

Overview

This project focuses on predicting **Parkinson's Disease** using a **Support Vector Machine (SVM)** model. Parkinson's Disease is a neurodegenerative disorder that affects movement, causing symptoms like tremors, stiffness, and difficulty with balance and coordination. Early detection of Parkinson's Disease can significantly improve the quality of life for patients. This project uses a dataset containing various voice measurements to predict whether a person has Parkinson's Disease.

What is Parkinson's Disease?

Parkinson's Disease is a progressive nervous system disorder that affects movement. It occurs when nerve cells in the brain that produce dopamine (a chemical responsible for smooth and coordinated muscle movement) become impaired or die. Common symptoms include:

- Tremors (shaking) in hands, arms, legs, or jaw
- Stiffness of the limbs and trunk
- Slowness of movement (bradykinesia)
- Impaired balance and coordination

Early detection is crucial for managing the disease effectively. This project uses machine learning to predict Parkinson's Disease based on voice measurements, as voice changes are one of the early signs of the disease.

What is Support Vector Machine (SVM)?

Support Vector Machine (SVM) is a powerful supervised machine learning algorithm used for **classification** and **regression** tasks. It works by finding the optimal hyperplane that separates data points of different classes in a high-dimensional space. SVM is particularly effective for:

- **Binary classification problems** (e.g., predicting whether a person has Parkinson's Disease or not).
- Handling **non-linear data** using kernel functions (e.g., linear, polynomial, or radial basis function kernels).

In this project, we use a **linear kernel** SVM to classify whether a person has Parkinson's Disease based on voice features.

Project Workflow

1. Data Collection

- The dataset used in this project is the **Parkinson's Disease Dataset**, which contains **195 rows** and **24 columns**.

- Each row represents a patient, and the columns include various voice measurements such as:
    - **MDVP:Fo(Hz)**: Average vocal fundamental frequency.
    - **MDVP:Jitter(%)**: Variation in fundamental frequency.
    - **MDVP:Shimmer**: Variation in amplitude.
    - **HNR**: Harmonics-to-noise ratio.
    - **RPDE**, **DFA**, **PPE**: Nonlinear measures of voice signals.
- The target variable is **status**, where:
    - **1**: The person has Parkinson's Disease.
    - **0**: The person does not have Parkinson's Disease.

---

2. Data Preprocessing

- **Handling Missing Values**: The dataset is checked for missing values, and no missing values are found.
- **Feature Selection**: The name column is dropped as it is not relevant for prediction. The status column is separated as the target variable.
- **Standardization**: The features are standardized using **StandardScaler** to ensure all features contribute equally to the model. Standardization transforms the data to have a mean of 0 and a standard deviation of 1.

---

3. Splitting the Data

- The dataset is split into **training** and **testing** sets:
    - **80%** of the data is used for training the model.
    - **20%** of the data is used for testing the model.
- The split is stratified to ensure that the proportion of Parkinson's and non-Parkinson's cases is maintained in both sets.

---

4. Model Training

- A **Support Vector Machine (SVM)** with a **linear kernel** is used for training.
- The model is trained on the training data (X_train, Y_train).

---

5. Model Evaluation

- The model's performance is evaluated on both the **training** and **testing** datasets.
- **Accuracy Score** is used to measure the model's performance:

- **Training Accuracy**: ~89.10%
- **Testing Accuracy**: ~89.74%

The high accuracy scores indicate that the model generalizes well to unseen data.

---

6. Building a Predictive System

- A predictive system is built to predict whether a person has Parkinson's Disease based on their voice measurements.
- The system takes input data, standardizes it, and uses the trained SVM model to make predictions.

Key Features of the Project

- **Early Detection**: The model can help in the early detection of Parkinson's Disease, which is crucial for effective treatment.
- **High Accuracy**: The model achieves an accuracy of ~89% on both training and testing data.
- **Scalability**: The model can be easily scaled to handle larger datasets or additional features.

How to Run the Code

1. **Clone the Repository**:
2. **Install Required Libraries**:
3. **Run the Jupyter Notebook**:
    - Open the parkinsons_prediction.ipynb file in Jupyter Notebook.
    - Execute the cells to train the model and make predictions.

Requirements

- Python 3.x
- Libraries:
    - numpy
    - pandas
    - scikit-learn

Conclusion

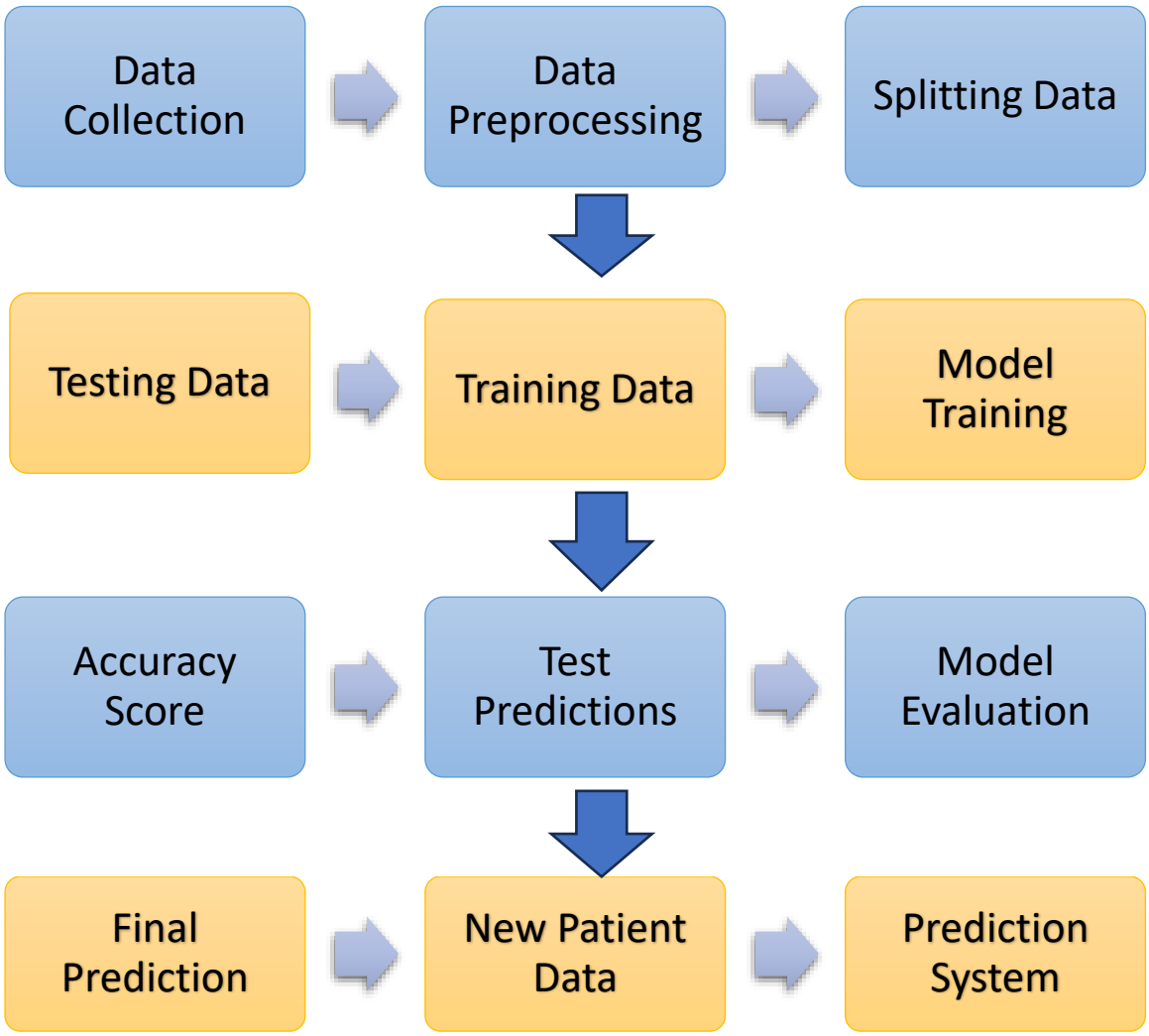
This project demonstrates the use of **Support Vector Machine (SVM)** for predicting Parkinson's Disease based on voice measurements. The model achieves high accuracy and can be a valuable tool

for early detection of the disease. By leveraging machine learning, we can improve the quality of life for individuals affected by Parkinson's Disease.

Acknowledgments

- The dataset is sourced from the **Kaggle.**

Visual Workflow Diagram



This project is a step forward in using machine learning for healthcare applications. By predicting Parkinson's Disease early, we can provide timely interventions and improve patient outcomes.