Parkinson’s Disease Prediction Project Documentation

1. Project Structure

- main.ipynb: Jupyter Notebook for data exploration, feature engineering, model training, evaluation, and saving models.
- app.py: Streamlit app for loading trained models, taking user input, predicting, and showing results.
- Model & Preprocessing Files (must be present):
- model_rf_all_features.joblib
- best_model_rf_selected_features.joblib
- scaler.joblib
- selected_features.pkl

2. app.py Code Walkthrough

2.1 Importing Libraries

- streamlit: web app UI
- numpy, pandas: data handling
- joblib, pickle: load saved models/features

2.2 App Title

- Displays title in web app.

2.3 Loading Models

- Loads Random Forest models (all features & selected features), scaler, and selected features.
- Uses Streamlit caching for efficiency.
- Stops execution with error if files are missing.

2.4 Feature Setup

- Retrieves original feature names from scaler.

2.5 Model Selection (UI)

- Radio button: choose All Features model or Selected Features model.

2.6 User Inputs

- Dynamically creates numeric input fields for each required feature.

2.7 Preprocessing Input

- Converts user input into DataFrame.
- Scales features.
- If optimized model chosen, keeps only selected features.

2.8 Prediction

- Predicts Parkinson’s likelihood using chosen model.
- Displays predicted class and probabilities.

3. Workflow

- Data preparation in notebook (clean, scale, feature select, train, save models).
- Deployment in Streamlit (load models, take input, predict, display results).

4. Usage Instructions

1. Place all required files in same directory.
2. Run app: streamlit run app.py
3. Open browser, input feature values, click Predict.

5. Future Improvements

- Allow dataset upload for batch predictions.

- Show feature descriptions for users.
- Add feature importance visualization.
- Deploy app on cloud (Streamlit Cloud, HuggingFace Spaces, etc.).