MemoTag Speech Intelligence Module

Objective

To build a basic pipeline that analyzes raw voice recordings to identify potential indicators of cognitive stress or decline using a combination of audio processing, NLP features, and unsupervised machine learning techniques.

Dataset Summary

- Audio Files: 5–10 simulated `.wav` files representing user speech
- Transcripts: Manually associated text transcripts (`transcripts.csv`) for linguistic analysis
- Assumption: Voice samples simulate speech in memory recall or word-association contexts

Feature Engineering

The following features were extracted, combining both acoustic and linguistic cues known to correlate with early cognitive decline.

Feature	What it measures	Why it's important
Pause Score	Amount of silence in the speech, normalized	Excessive pausing may indicate cognitive load or memory retrieval effort
Hesitation Ratio	Ratio of filler/hesitation words (um, uh, etc.) to	Hesitation is a common sign of difficulty in fluency

	total word count	and recall
Pitch Mean & Pitch Standard Deviation	Pitch average and variability using frequency tracking	Reduced pitch variability is common in flat, monotonous speech associated with cognitive decline
Speech Rate	Words per second	Slower speech often indicates cognitive slowing or retrieval issues
Sentence Completion	Binary flag if sentence is complete (ends with punctuation)	Incomplete sentences may suggest disorganized thought processes

Machine Learning Methodology

1. **Isolation Forest (Anomaly Detection)**

Why chosen: Robust to small datasets, unsupervised (no medical labels needed), easily interpretable outlier scores.

How it works: Each sample is analyzed in multidimensional space of the extracted features. The model isolates samples that are statistically "different" from the rest and flags them as potential indicators of cognitive stress.

2. **PCA (Principal Component Analysis)**

Why chosen: Reduces feature space to 2D, allows us to visually inspect clusters and anomalies.

Output: Clear scatterplot showing normal vs. abnormal speech patterns.

Results Summary

Top Features (based on variance and impact in clustering):

- **Pause Score** : Most abnormal samples had high pause ratios
- **Hesitation Ratio** : Top indicator of fluency issues
- **Pitch Variability** : Low variance correlated with flagged samples

Visual Insight:

Clear separation of outlier points in PCA space. Isolation Forest reliably flagged samples with high hesitation, slow speech rate, and monotone pitch.

Next Steps for Clinical Robustness

Step	Description
Use Real Clinical Audio	Collect samples from actual patients under medical supervision
Task-Based Prompts	Add structured tasks (e.g., naming, memory recall) to elicit more cognitive stress markers
Integrate Whisper (or ASR)	Transcribe speech directly from audio to automate the pipeline
Add Time-Domain Features	Use MFCCs, jitter, shimmer for richer voice pattern detection
Build a Risk Score API	Create an API to return a cognitive risk score from a given `.wav` file

Supervised Learning (Future)	When enough labeled patient data is collected, apply classification techniques (e.g., SVM, XGBoost)

Summary

This project provides a foundational, interpretable pipeline that combines voice signal processing and basic NLP with unsupervised machine learning to flag potential early signs of cognitive impairment. It’s lightweight, clinically inspired, and ready to be extended into a real-world tool.