# Voice Problem Detection - Detailed Report

## 1. Project Objective

The project aims to develop a system that can detect potential voice-related problems based on speech patterns and spoken content. It uses both linguistic features (pauses, hesitations) and text-based representations (BoW and TF-IDF) to classify or understand speech anomalies.

## 2. Feature Engineering

a. Temporal & Linguistic Features:

- Pauses are extracted based on punctuation (e.g., `. , ? !`) to infer speaking rhythm.

- Hesitation Words like "uh", "um", and "err" are detected using regular expressions, indicating potential speech fluency issues.

b. Text Vectorization Techniques:

i. Bag of Words (BoW):

- Implemented using CountVectorizer from scikit-learn.

- Captures word frequencies across spoken sentences.

ii. TF-IDF:

- Implemented using TfidfVectorizer.

- Focuses on rare terms for distinguishing different speech patterns.

## 3. Machine Learning Models Used

a. KMeans Clustering:

- Used on delay data to distinguish between fluent and disfluent speech sections.

b. Text-based Feature Models:

- BoW and TF-IDF vector outputs are used to feed into classification models.

## 4. Model Pipeline Summary

1. Audio Input -> Speech Recognizer (recognize_google)

2. Transcription -> Cleaned text sentence

3. Pause + Hesitation Detection -> Regex + timing simulation

4. Text Vectorization -> CountVectorizer and TfidfVectorizer

5. ML Clustering/Classification -> KMeans or classifiers

6. Output -> Highlights disfluencies and hesitation patterns

## 5. Potential Next Steps for Clinical Robustness

a. Improve Ground Truth:

- Collect labeled datasets of pathological vs. normal speech.

b. Expand Feature Set:

- Add acoustic signal features.

c. Apply Supervised Learning:

- Use models like logistic regression, SVM, or ensemble classifiers.

d. Deep Learning Integration:

- Use LSTM, BERT, or CNN models.

e. Clinical Trial Integration:

- Test with speech-language pathologists in real-world scenarios.

## 6. Conclusion

| Component | Technique Used |
|---|---|
| Audio Input | Microphone + Google Speech API |
| Pause Detection | Punctuation-based rules |
| Hesitation Words | Regex-based matching |
| Vectorization | BoW (CountVectorizer), TF-IDF |
| Clustering | KMeans for unsupervised pause mapping |
| Output | Fluency markers, potential alerts |