

## Week 4: Model Improvement & Optimization

### Task 1: Feature Engineering & Optimization (Tai)

#### Steps:

- Experiment with different feature combinations (MFCCs, Pitch, Formants).
- Try dimensionality reduction techniques (PCA, t-SNE) to reduce feature size while maintaining information.
- Analyze the impact of different feature sets on the model's performance.
- Save the optimized feature set as **data/optimized\_features.csv**.
- Upload final feature dataset to GitHub.

#### Files Will Work With:

- data/training\_data.csv
- data/testing\_data.csv
- data/optimized\_features.csv (**new file**)

### Task 2: Train an Improved Model (James)

#### Step:

- Load data/optimized\_features.csv.
- **Train different machine learning models and compare results:** Logistic Regression, Support Vector Machine (SVM), Random Forest, Neural Networks (optional)
- Compare accuracy of different models.
- Save the best model as **models/improved\_model.pkl**.
- Upload final model to GitHub.

#### Files Will Work With:

- data/optimized\_features.csv
- models/improved\_model.pkl (**new file**)

### Task 3: Compare & Choose Best Model (Kyle & Wissam)

#### Steps:

- Load **models/improved\_model.pkl** and **testing\_data.csv**.
- **Evaluate models using:** Accuracy, Precision, Recall, and F1-score
- Confusion Matrix (**analyze errors**).
- Compare results to Week 3 baseline model (**James' 70% model**).
- Write a final model report: Explain which model worked best and why.
- Save results as **results/final\_model\_evaluation.txt**.
- Upload evaluation report to GitHub.

#### Files Will Work With:

- models/improved\_model.pkl
- data/testing\_data.csv
- results/final\_model\_evaluation.txt (**new file**)