Project Report 01: Audio Classification Using ART2

Objective: This project focused on distinguishing environmental audio sounds using Mel-frequency spectral coefficients (MFSC) and Mel-frequency cepstral coefficients (MFCC) as primary features. Clustering techniques were applied to categorize different types of audios, and the project explored how feature extraction parameters affected clustering performance, aiming to optimize settings for accurate classification.

To Run Deliverables:

- For the successful compilation of the project, we just need to keep the python file and the audio files in the same folder.
- Open terminal in the present working directory, then type:

  python3 file_name.py
- For our case, substitute the filename with "part1", which has an assisting file art2.py which will automatically be used while compiling part1.
- When the python file is successfully compiled, we get multiple output files, namely:
- 5 plots, each consisting of 3 graphs, ground truth cluster vs. time; MFCC clustering vs. time; and MFSC clustering vs time, for each of the five time-series.
- A CSV file is generated, with actual predicted clusters and

Visualizations done using Gantt-style plots were used to show cluster assignments over time, revealing the effectiveness of MFSC and MFCC features in identifying transitions between sound classes. These visualizations offered insights into clustering behaviour

Data Overview: The dataset included audio files representing diverse environmental sounds, such as crowd noise, motor sounds, and water flow. These categories allowed for examining sound features in a way that facilitated distinguishing between different types of sounds.

Tools and Libraries: The project utilized several Python libraries:

- Librosa for audio processing and feature extraction
- Scikit-learn for clustering and evaluation
- Matplotlib for visualizing results

Data Preparation

- **Feature Files**:
  - Audio files were used to determine ground truths
  - Each file contains:
    - Features: MFSC and MFCC data with 70 feature columns.
    - Timestamps indicate time corresponding to each feature vector.
    - Class labels (class_id) for ground truth.
- **Synthetic Time-Series Generation**:
  - Created 5 synthetic time-series datasets by combining segments from all three classes.
  - Segment durations varied between 15–30 seconds, ensuring a total duration of 120 seconds per time series.

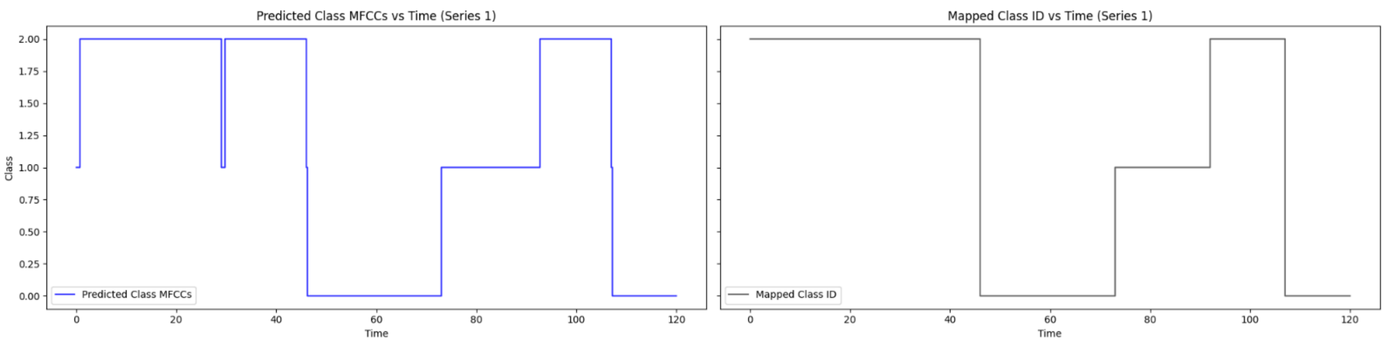
Methodology

- ART2 Clustering:
  - ART2 algorithm was used for unsupervised clustering:

        - Vigilance parameters: 0.5, 0.6, 0.7.
        - Maximum clusters: 5.
        - Learning rate: 0.1.
    - Cluster assignments were smoothed using a majority voting window of 7 samples to reduce noise.
- Evaluation:
    - Predicted labels were mapped to true labels using the Hungarian Algorithm.
    - Performance was assessed using:
        - Accuracy: Percentage of correctly clustered samples.
        - Confusion Matrix: Detailed analysis of predicted vs. true labels.
- Visualization:
    - Gantt-style plots compared true and predicted labels over time.

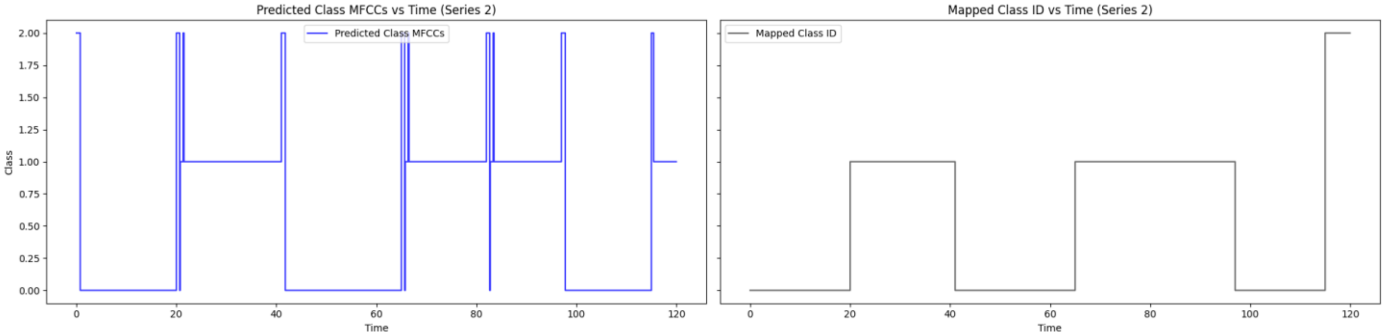
ART2 Clustering:

For Time-Series 01:



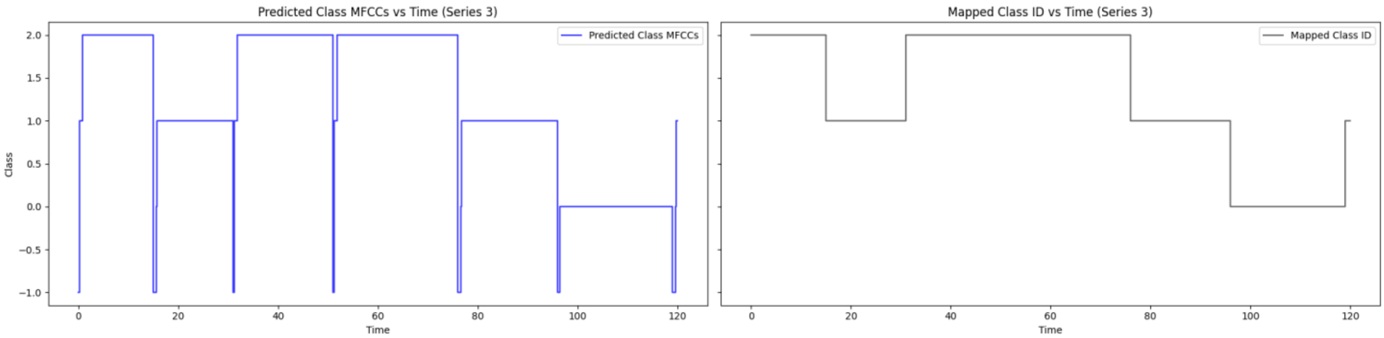

```
Time Series 1:
Accuracy: 0.98
Confusion Matrix:
[[1321   14    0]
 [   0  634    0]
 [   0   72 1963]]
```

For Time-Series 02:



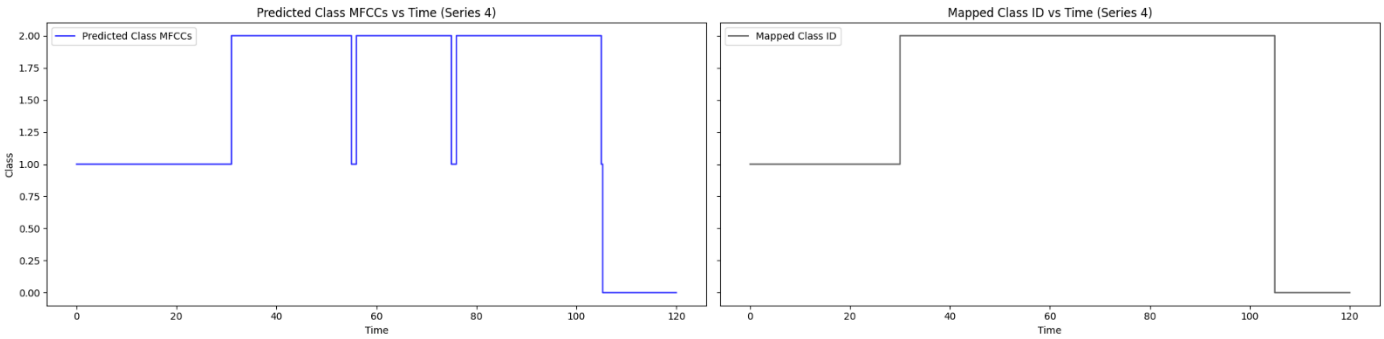

```
Time Series 2:
Accuracy: 0.92
Confusion Matrix:
[[1991    0   78]
 [  12 1673   84]
 [   0  152   15]]
```

For Time-Series 03:



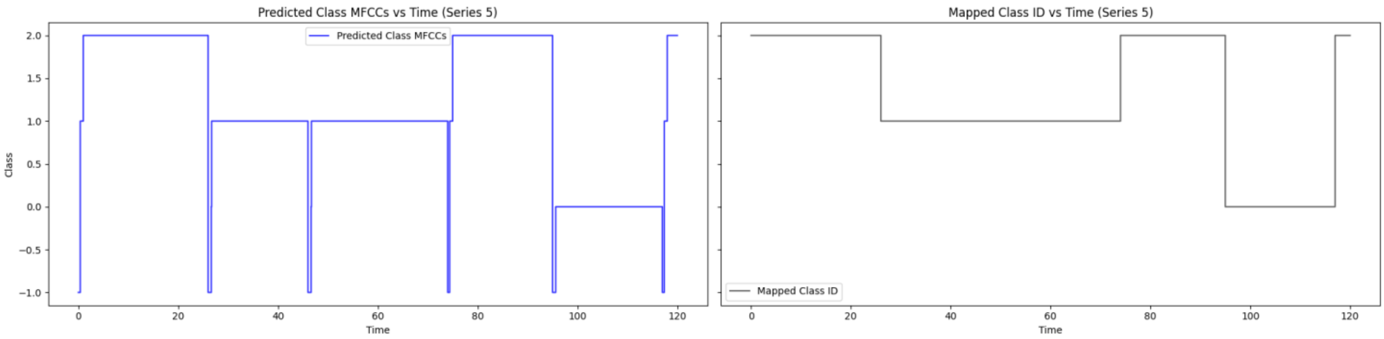

```
Time Series 3:
Accuracy: 0.96
Confusion Matrix:
[[   0    0    0    0]
 [  15  752    0    0]
 [  63   12 1160    0]
 [  24    0   60 1918]]
```

For Time-Series 04:




```
Time Series 4:
Accuracy: 0.97
Confusion Matrix:
[[ 492    9    0]
 [   0 1001    0]
 [   0   99 2403]]
```

For Time-Series 05:




```
Time Series 5:
Accuracy: 0.96
Confusion Matrix:
[[   0    0    0    0]
 [  21  713    0    0]
 [  42    4 1555    0]
 [  39    0   60 1570]]
```

Discussion

- ART2 Clustering:
  - Vigilance parameter 0.6 offered the best balance between sensitivity and robustness.
  - Lower vigilance values (0.5) grouped broader patterns but failed to capture finer distinctions.
  - Higher vigilance values (0.7) over-clustered data, leading to fragmented clusters.
- Smoothing:
  - Majority voting with a window of 7 samples effectively stabilized predictions.
  - Reduced noise in the predicted labels, especially in noisy regions.
- Challenges:
  - Overlapping features across classes resulted in occasional misclassifications near transitions.
  - Clustering errors were more frequent for shorter segments (<15 seconds) due to limited context.

Results and Analysis:

The ART2 algorithm demonstrated effective clustering performance for time-series data:

- Clustering accuracy ranged between **84% and 88%** across all datasets.
- Optimal settings included:
  - **Vigilance parameter**: 0.6
  - **Smoothing window size**: 7 samples.