

# STT3795 - Proposition de projet

Kamen Damov, Simon Langlois, Samir Bio Gbian

February 2024

## 1 Descriptions et objectifs

L'objectif de ce projet est de comparer différentes pipelines d'apprentissage automatique pour classer des langages, des types de musique, et des chiffres prononcés. Notre projet sera une analyse empirique des différentes architectures de classification. Nous allons en développer trois pipelines qui auront des prétraitements de données différents et classificateurs différents:

### 1.1

La première architecture serait d'utiliser une pipeline de prétraitement de données audios soit avec un Fourier Transform à partir duquel on pourra extraire les domaines fréquentiels, et domaines temporels. Suite à cela, nous appliquerons des modèles de classifications traditionnels tel que le SVM, l'arbre de décision, la forêt d'arbres de décisions, et la Naive de Bayes. Finalement, nous allons évaluer la performance des modèles.

### 1.2

La deuxième architecture est d'appliquer la même pipeline de prétraitement qu'au point 1.1 mais auquel on ajoutera un algorithme de d'analyse de données topologiques nommées persistent homology qui nous donnera des caractéristiques additionnelles. Nous nous sommes inspiré de ce papier-là qui présente une méthode d'analyse topologique pour but de classification <https://arxiv.org/abs/1910.08345>. Suite à ce prétraitement, nous allons utiliser les mêmes classificateurs qu'au point 1.1.

### 1.3

La troisième architecture sera d'appliquer la même pipeline que le point 2.1 mais d'appliquer un modèle d'apprentissage profond soit un réseaux de neurones convolutionnel (CNN).

Nous allons utiliser ces trois pipelines sur trois jeux de données audios différents

ce qui nous donnera la possibilité de comparer les types d'architectures entre elles, mais aussi de comparer un type d'architecture sur 3 jeux de données différents.

## 2 Contributions de chaque membre de l'équipe

Chaque membre de l'équipe se chargera de faire les trois pipelines (1.1, 1.2, 1.3) pour un jeux de données spécifiques.

## 3 Source de données

### 3.1

Notre premier jeu de données est un ensemble de prononciations des catégories de MNIST. Voici le lien vers les données: <https://github.com/Jakobovski/free-spoken-digit-dataset/tree/master>

### 3.2

Notre deuxième jeu de données est un ensemble de langues parlé qui est libellé avec la langue qui est parlé dans l'extrait sonore. Voici le lien vers les données: [https://huggingface.co/datasets/common\\_voice](https://huggingface.co/datasets/common_voice)

### 3.3

Notre troisième jeu de données est un ensemble de données musicales qui est libellé avec le genre de musique. Voici le lien vers les données: <https://huggingface.co/datasets/marsyas/gtzan>

## Descriptions et objectifs - Post-commentaires

Notre objectif sera d'explorer et comparer plusieurs méthodes et techniques pour classifier du langage parlé parmi 20 langues.

### Normalisation et prétraitement

Ayant des données audio, nous allons d'abord normaliser les données (borner les spectres entre -1 et 1, réduire le bruit, etc).

### Extraction de représentation

Ensuite nous allons utiliser des algorithmes d'extraction de représentation fréquentielles et spectrographiques telle que Fast Fourier Transform et TDA (persistent homology). Ces techniques augmenteront considérablement la dimensionnalité de nos données.

## Réduction de dimensionnalité

Après avoir extrait ces représentations, nous allons réduire la dimensionnalité en utilisant l'algorithme de PCA.

## Modèles

Ensuite nous allons remodeler les données pour les ajuster aux modèles de classification suivants: SVM, Random Forest, et CNN. Basé sur la performance de ces modèles, nous allons restructurer le prétraitement des données (soit en retirant ou ajoutant des étapes dans le prétraitement) et recomparer la performance des modèles.