



Classification automatique des genres musicaux par apprentissage supervisé

Eliott FLAMENT

TIPE - Session 2025
Candidat n° XXXXX

Sommaire



- 1) Introduction et problématique
- 2) Jeu de données & Extraction
- 3) Modèle k-NN
- 4) Modèle d'arbre de décision
- 5) Modèle Random Forest
- 6) Modèle XGBoost
- 7) Comparaison
- 8) Conclusion & Perspectives

1) Introduction et problématique

Problématique : Peut-on entraîner un modèle capable de reconnaître automatiquement et de manière fiable le genre musical d'un morceau ?



Flaticon

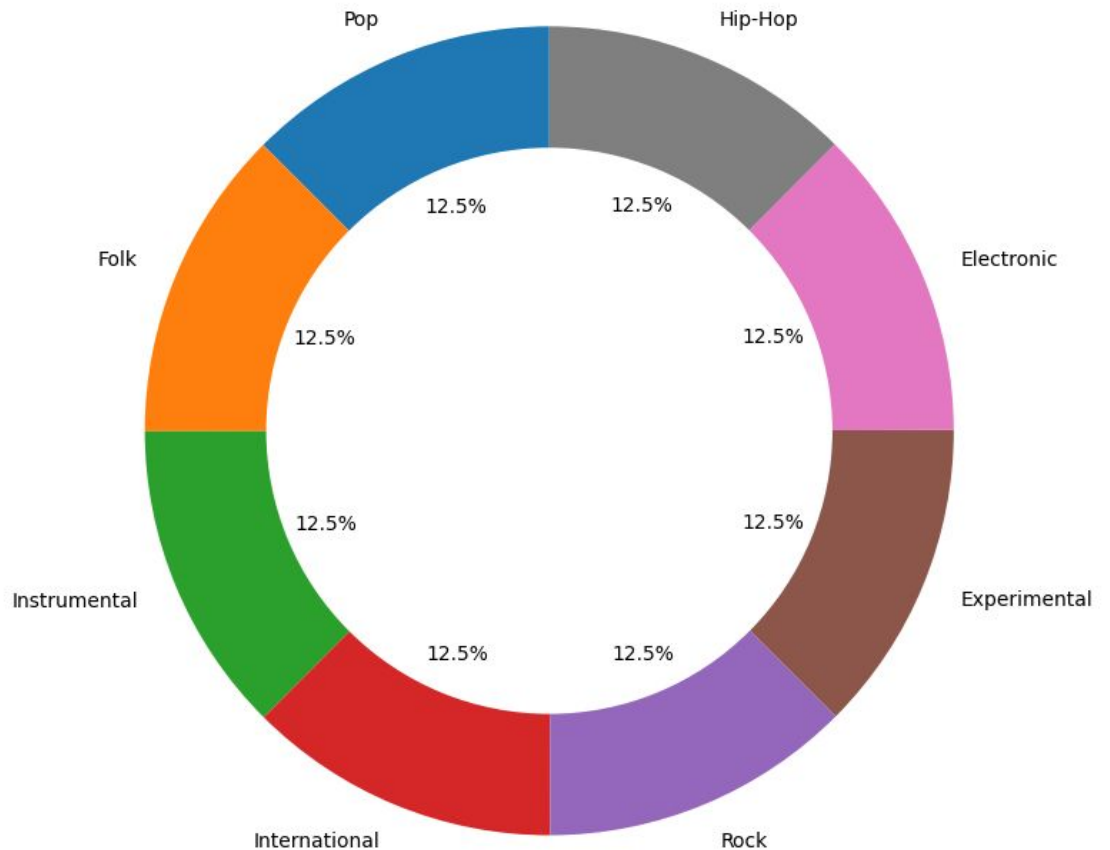
2) Jeu de données & Extraction

 - *Free music archive*

FMA Small

- 8000 morceaux
- 8 genres principaux
- Métadonnées

Répartition des genres musicaux dans le dataset



2) Jeu de données & Extraction

- *Features audio*

33 features audio à extraire pour chaque morceau

- RMS
- ZCR
- Spectral_centroid
- Spectral_rolloff
- Spectral_flatness
- harmonic_RMS
- percussive_RMS
- Tempo
- Mfcc 1 à 13
- Chroma 1 à 12

2) Jeu de données & Extraction


 - *RMS*

Représente le volume perçu d'un signal sonore.

Pour un signal $x[n]$ composé de N échantillons,

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^N x[n]^2}$$

2) Jeu de données & Extraction

 - ZCR

Représente la fréquence à laquelle le signal change de signe.

$$\text{ZCR} = \frac{1}{N-1} \sum_{n=1}^{N-1} \mathbb{1}_{\{x[n] \cdot x[n-1] < 0\}}$$

2) Jeu de données & Extraction

 - *Spectral_centroid*

Fréquence moyenne, pondérée par l'intensité de chaque fréquence

Spectral centroid bas : son grave

Spectral centroid haut : son aigu

$$\text{Spectral Centroid} = \frac{\sum_{k=0}^{K-1} f_k \cdot |X[k]|}{\sum_{k=0}^{K-1} |X[k]|}$$

2) Jeu de données & Extraction

 - *Spectral_rolloff*

Fréquence en dessous de laquelle se trouve 85% de l'énergie spectrale totale.

$$\text{Spectral Rolloff} = 0.85 \cdot \sum_{k=0}^{K-1} |X[k]|$$

2) Jeu de données & Extraction

- *Spectral_flatness*

Mesure à quel point le spectre d'un signal est plat.

C'est le rapport entre la moyenne géométrique et la moyenne arithmétique des amplitudes spectrales.

$$\text{Spectral Flatness} = \frac{\left(\prod_{k=0}^{K-1} |X[k]| \right)^{1/K}}{\frac{1}{K} \sum_{k=0}^{K-1} |X[k]|}$$

2) Jeu de données & Extraction



- *Harmonic_RMS / Percussive_RMS*

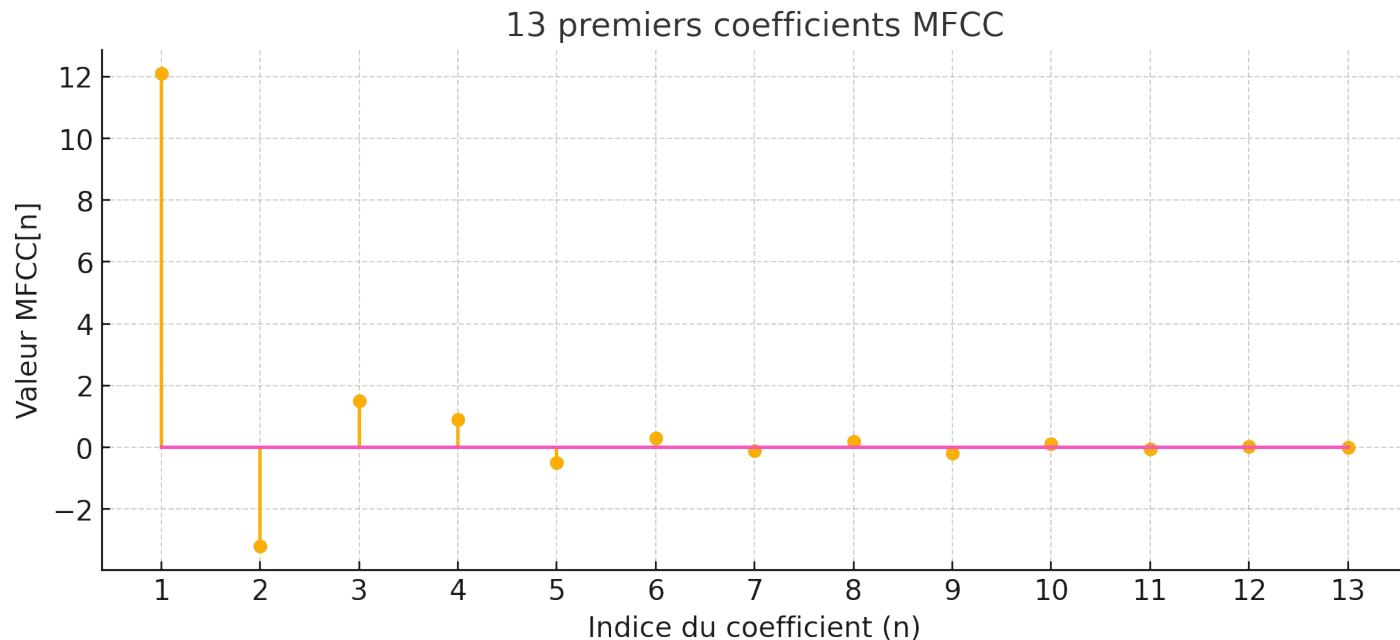
Séparation du signal audio en deux composantes (harmonique et percussif), via l'algorithme HPSS (Harmonic-Percussive Source Separation).

Application de RMS aux deux composantes.

2) Jeu de données & Extraction

 - *Mfcc 1 à 13*

- Transformée de Fourier
- Échelle Mel
- Logarithme
- Transformée en cosinus discrète (DCT)
- 13 premiers coefficients



2) Jeu de données & Extraction

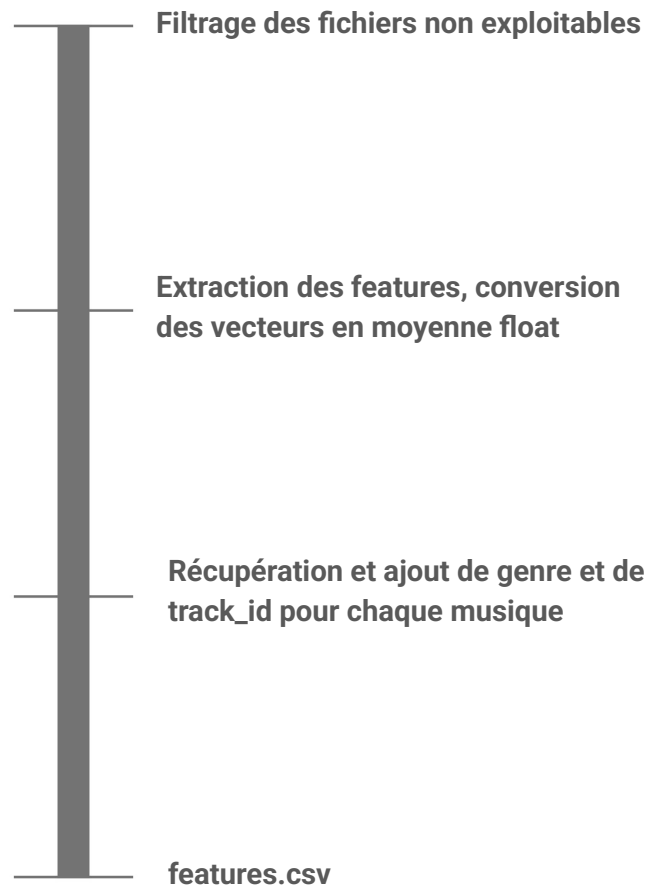
 - *Chroma 1 à 12*

- Distribution de l'énergie
- 12 hauteurs de notes de la gamme musicale occidentale

$$\text{Chroma}[n] = \sum_{k, f_k \in \text{classe}_n} |X[k]|$$

2) Jeu de données & Extraction

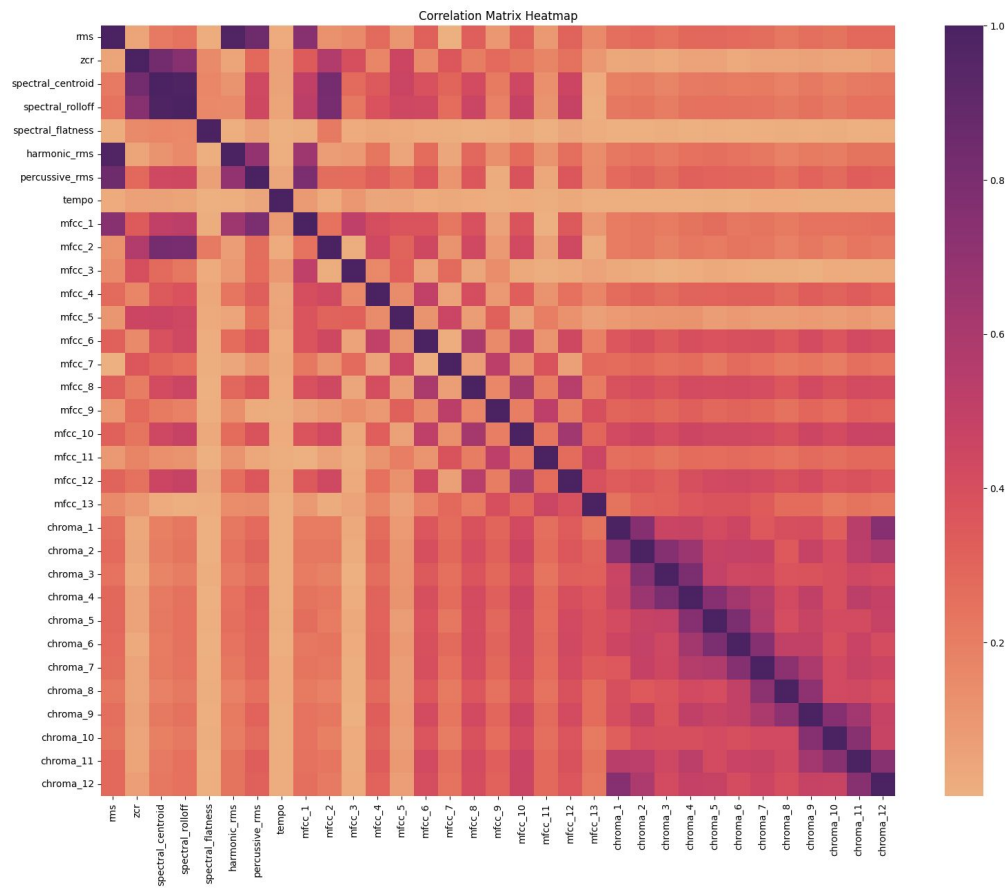
 - *Pipeline*



2) Jeu de données & Extraction

 - *Corrélation*

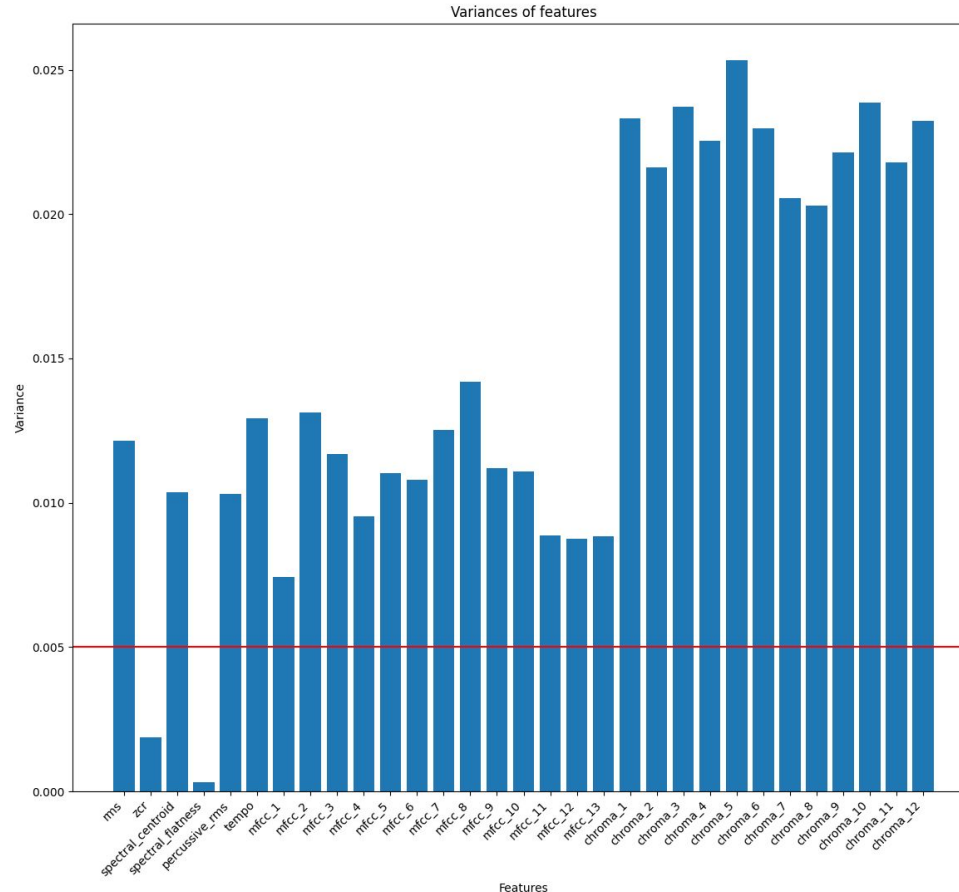
- Seuil = 0.95
- Colonnes supprimées : ['spectral_rolloff', 'harmonic_rms']



2) Jeu de données & Extraction

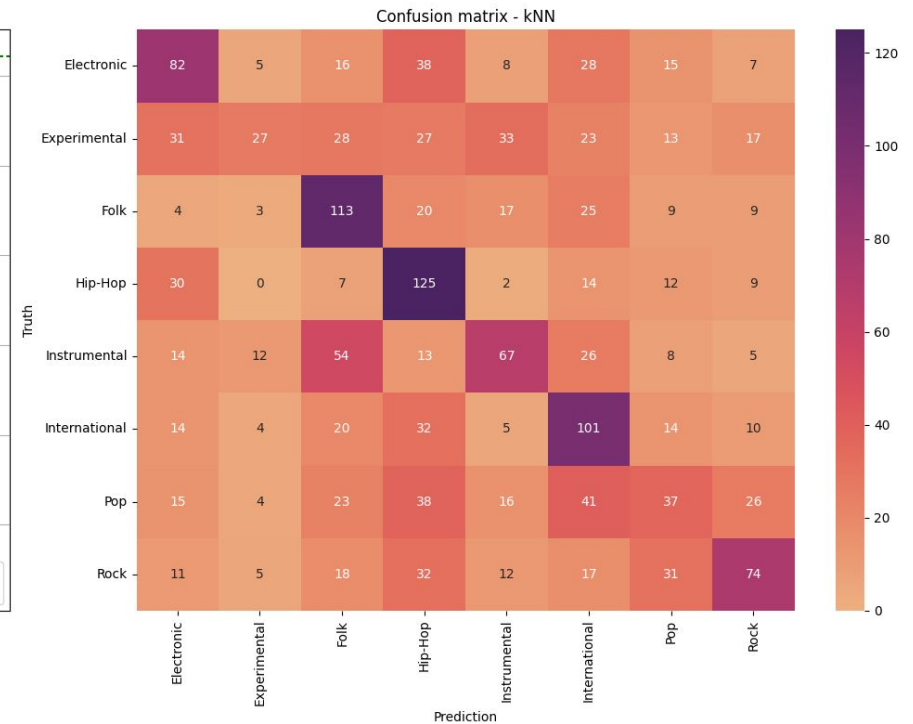
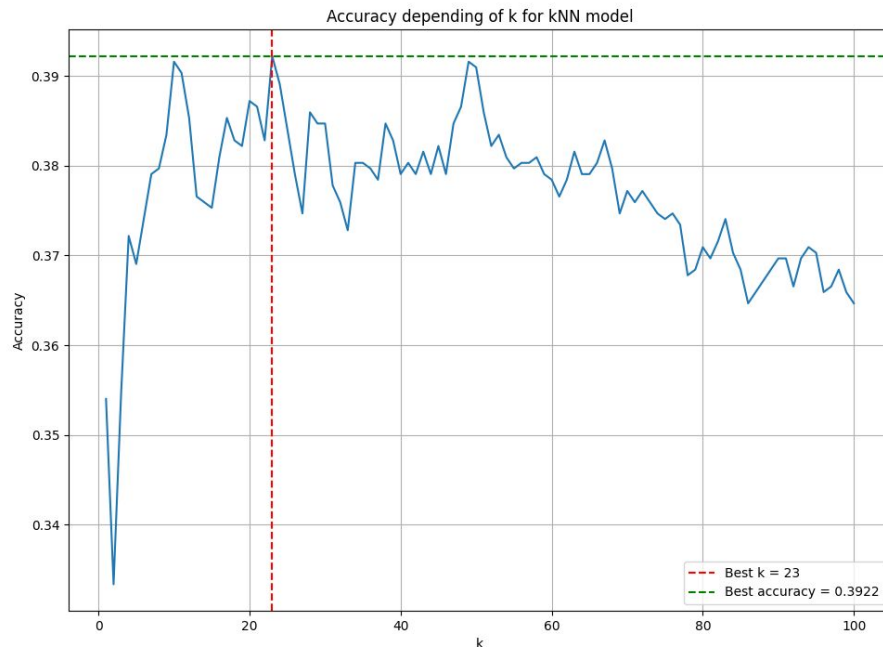
 - *Variance*

- Seuil = 0.005
- Colonnes à supprimer : ['zcr', 'spectral_flatness']



3) Modèle k-NN

— Valeurs de k ?



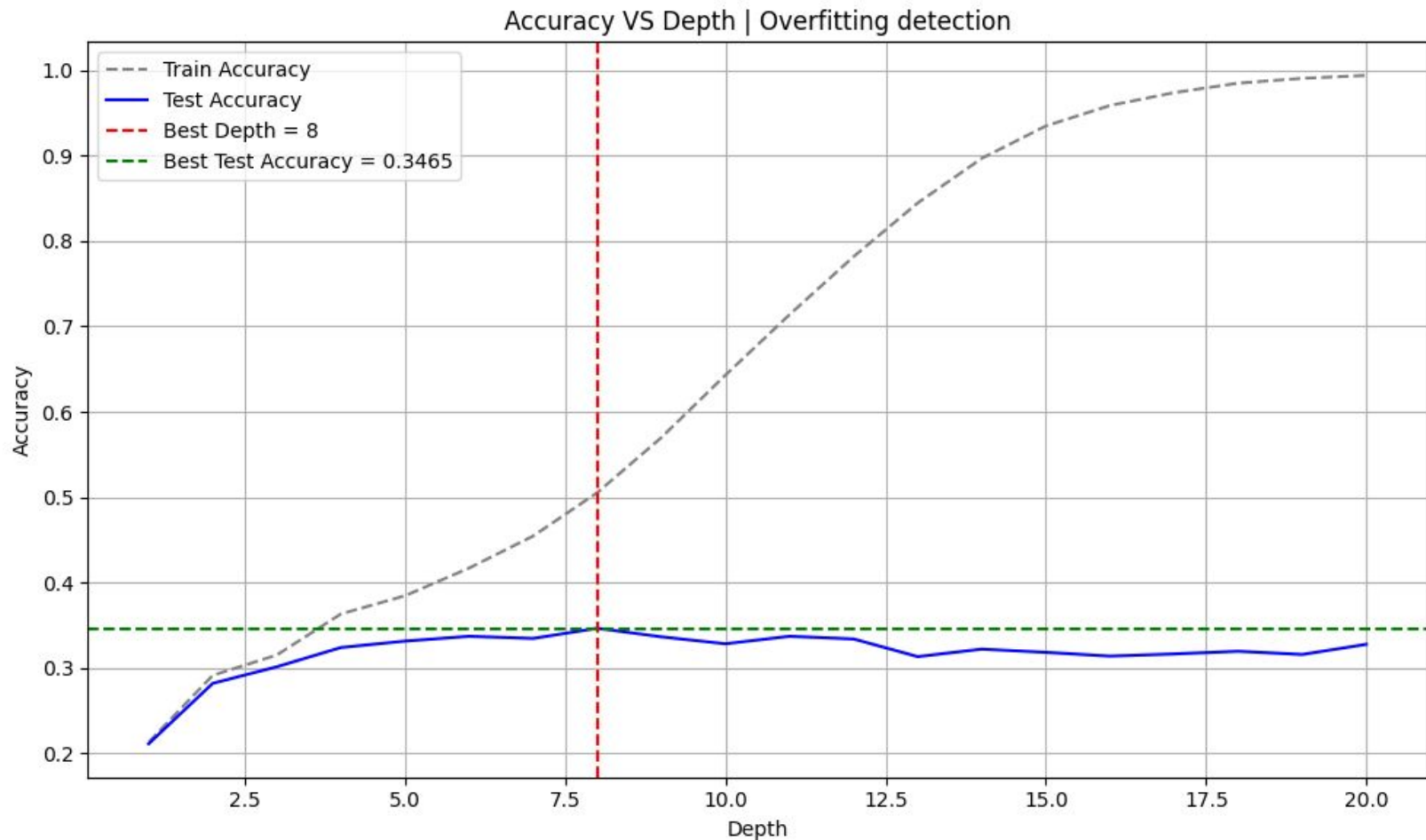
Les + : Folk, Hip-Hop

Accuracy : 0.39

Les - : Experimental, Pop

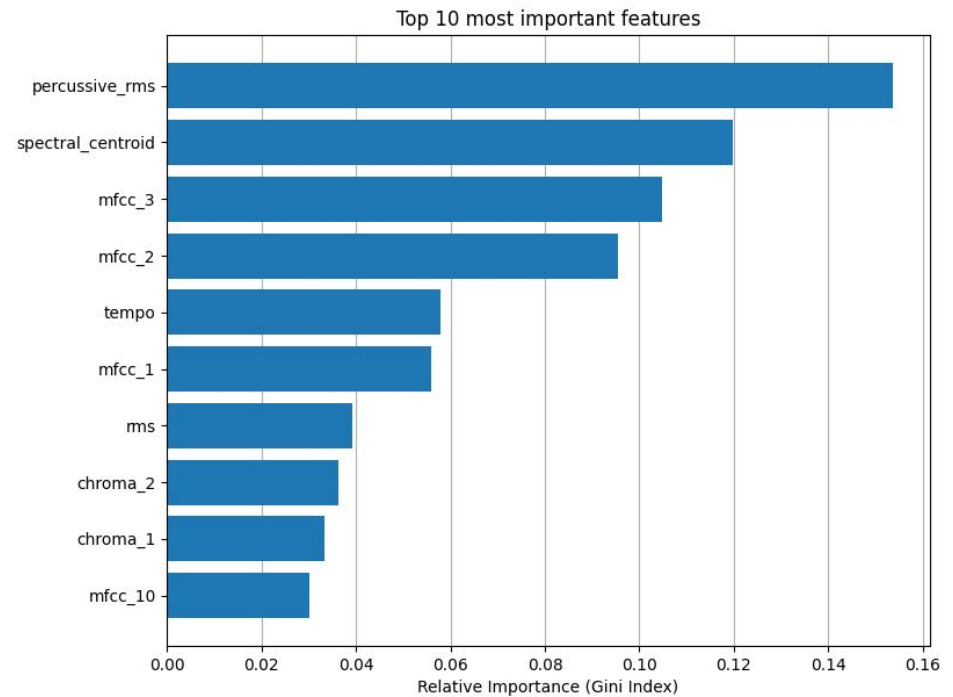
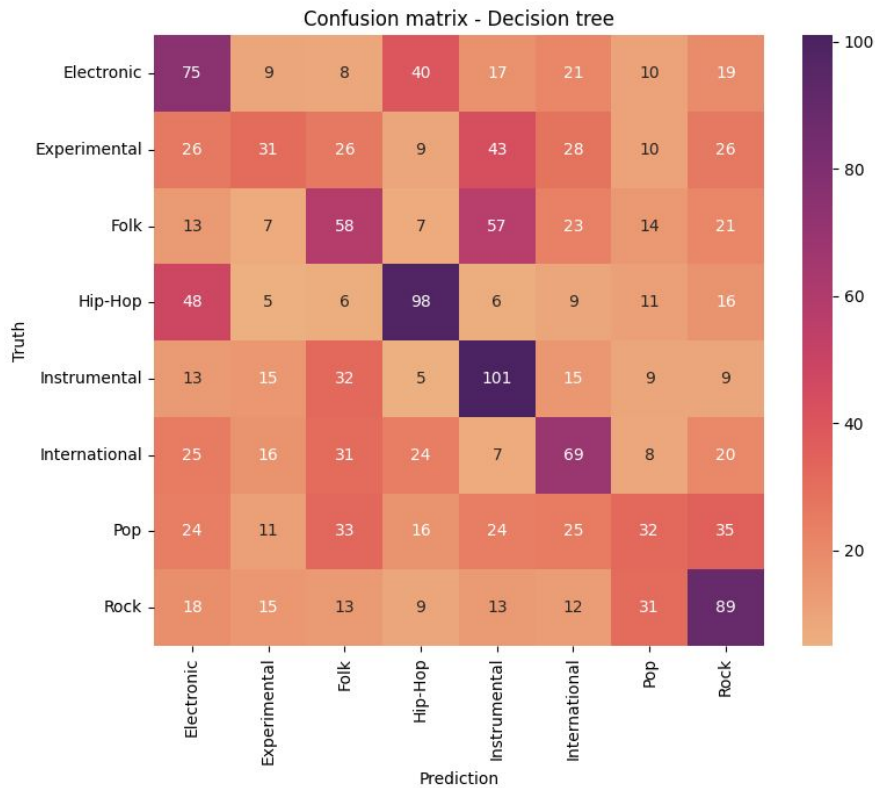
4) Modèle d'arbre de décision

— Profondeur ?



4) Modèle d'arbre de décision

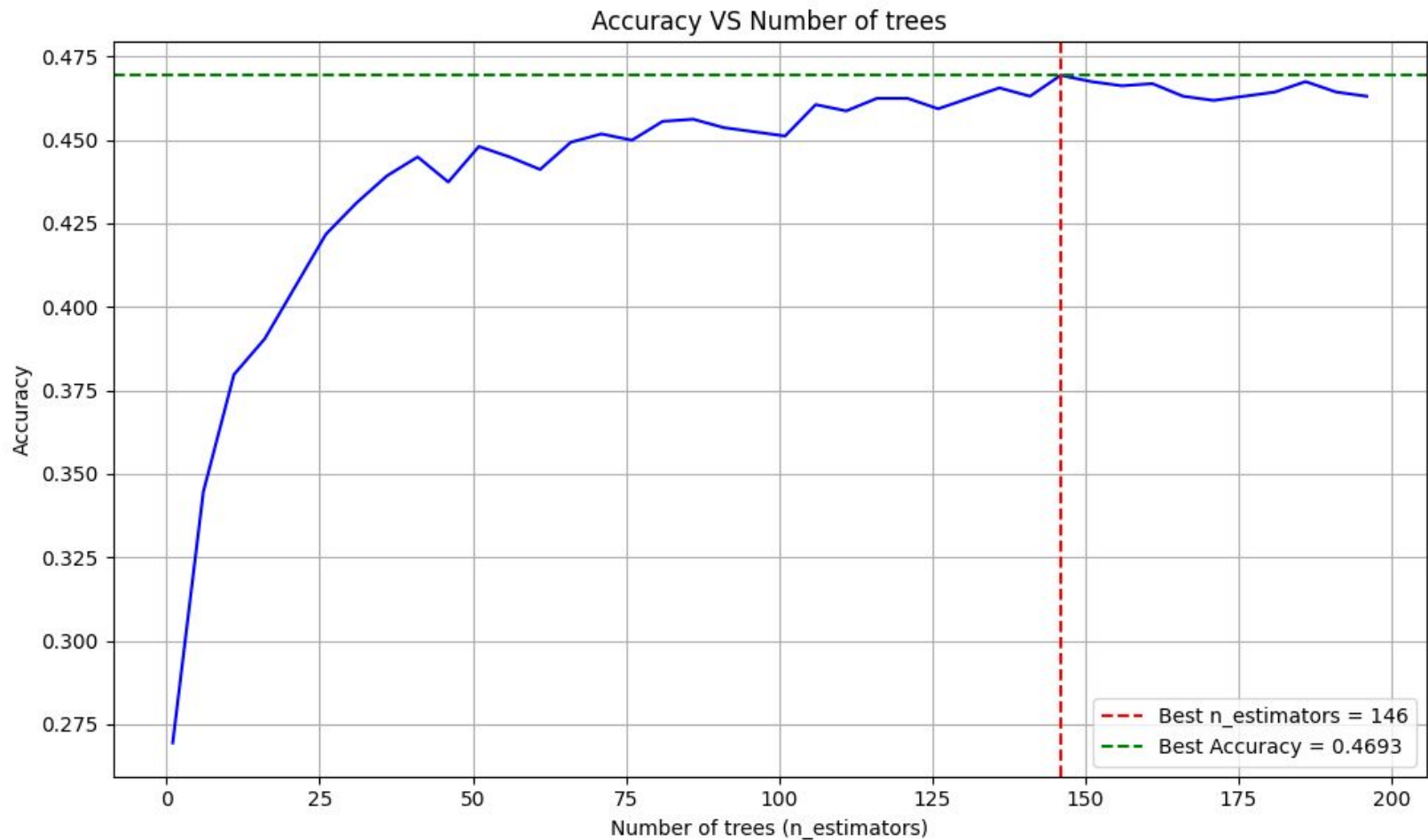
 - Bilan



Accuracy : 0.35

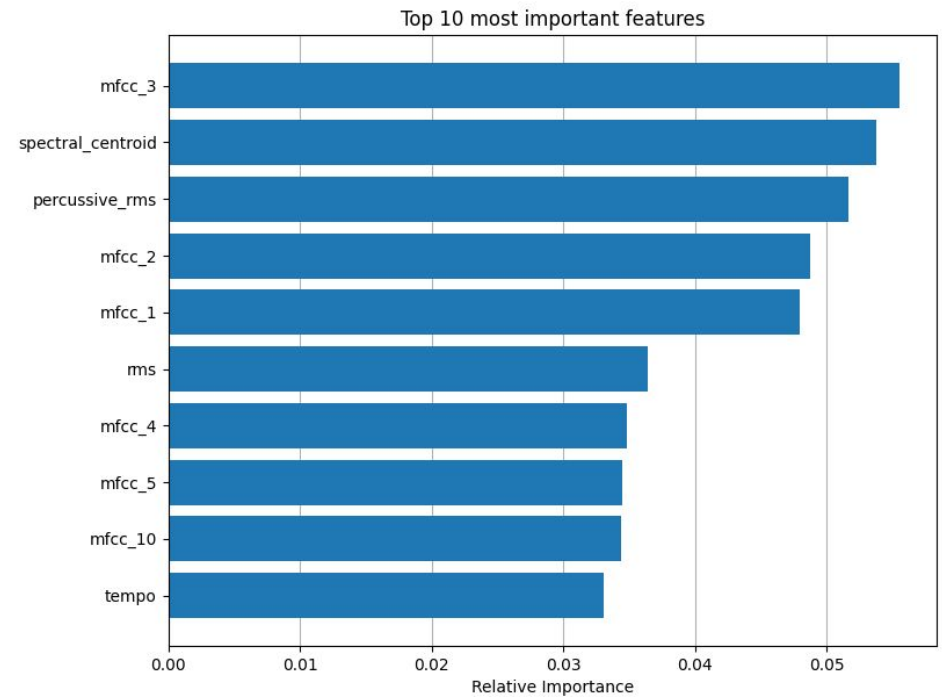
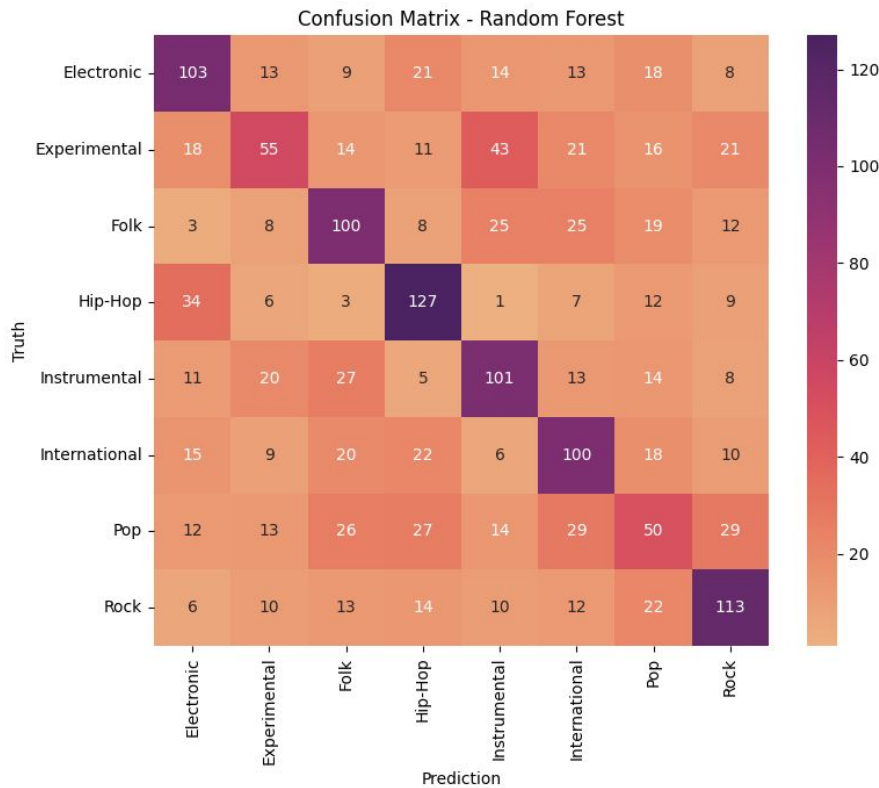
5) Modèle Random Forest

— - Nombre d'arbres ?



5) Modèle Random Forest

 - Bilan



Accuracy : 0.47

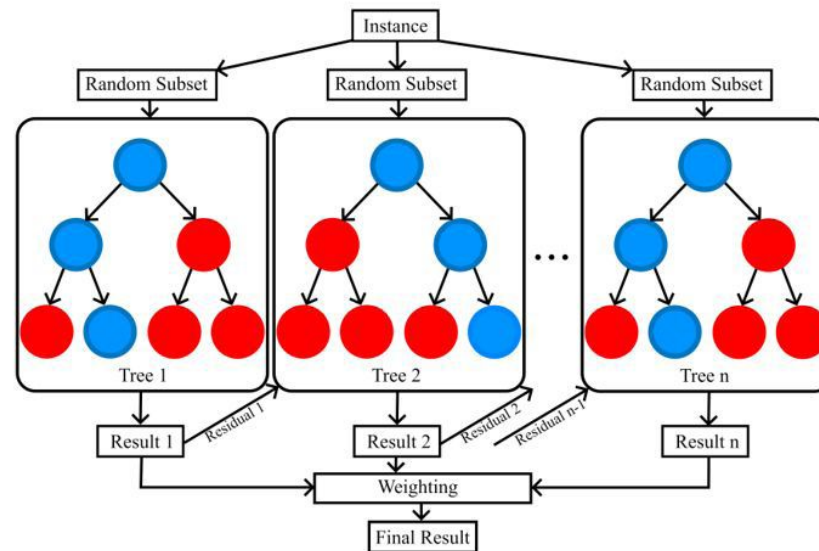
6) Modèle XGBoost



Taux d'apprentissage : 0.1

Arrêt Anticipé : 20

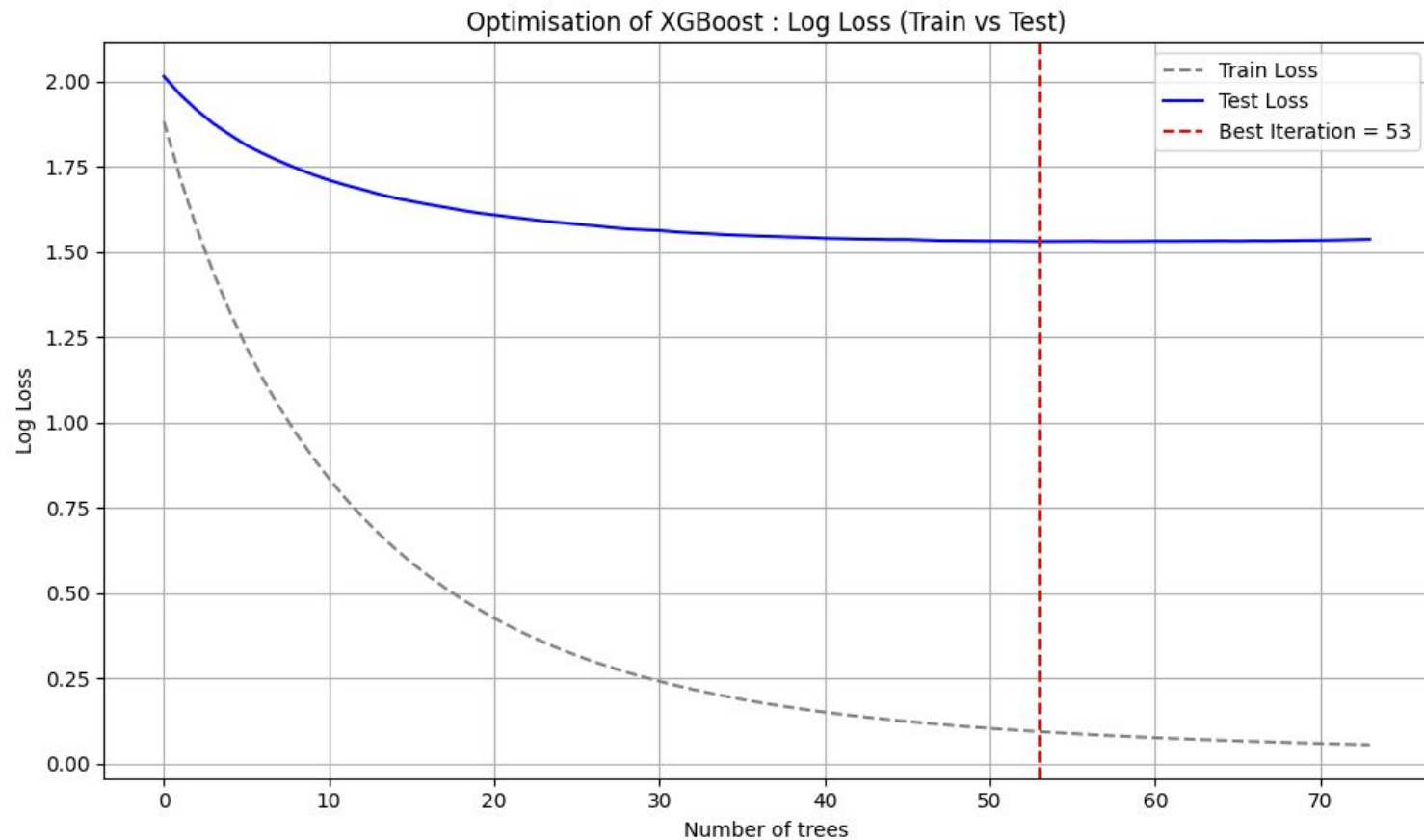
Fonction de coût : log loss



ResearchGate

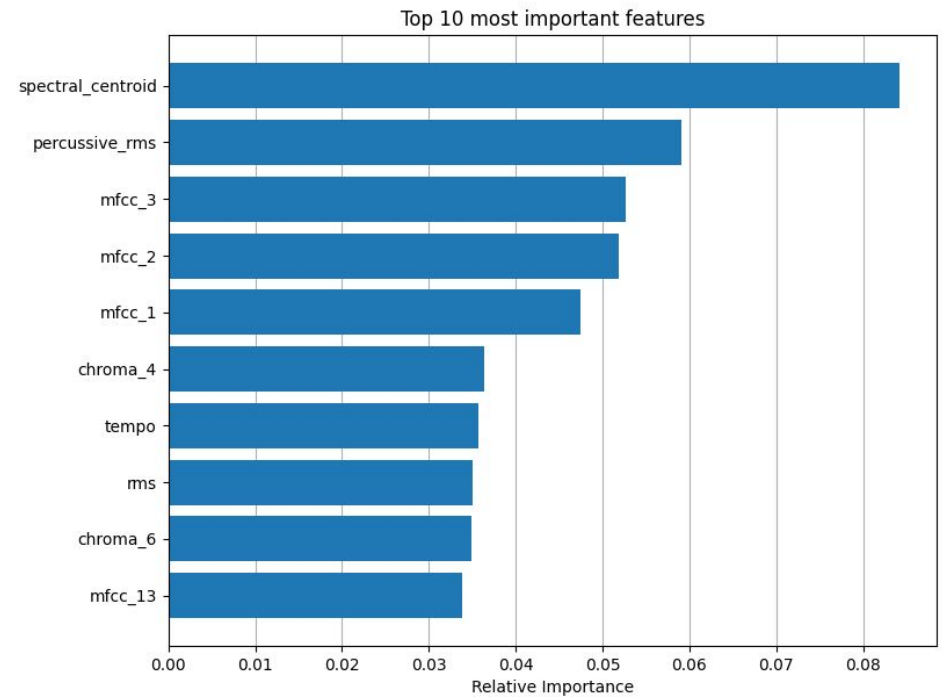
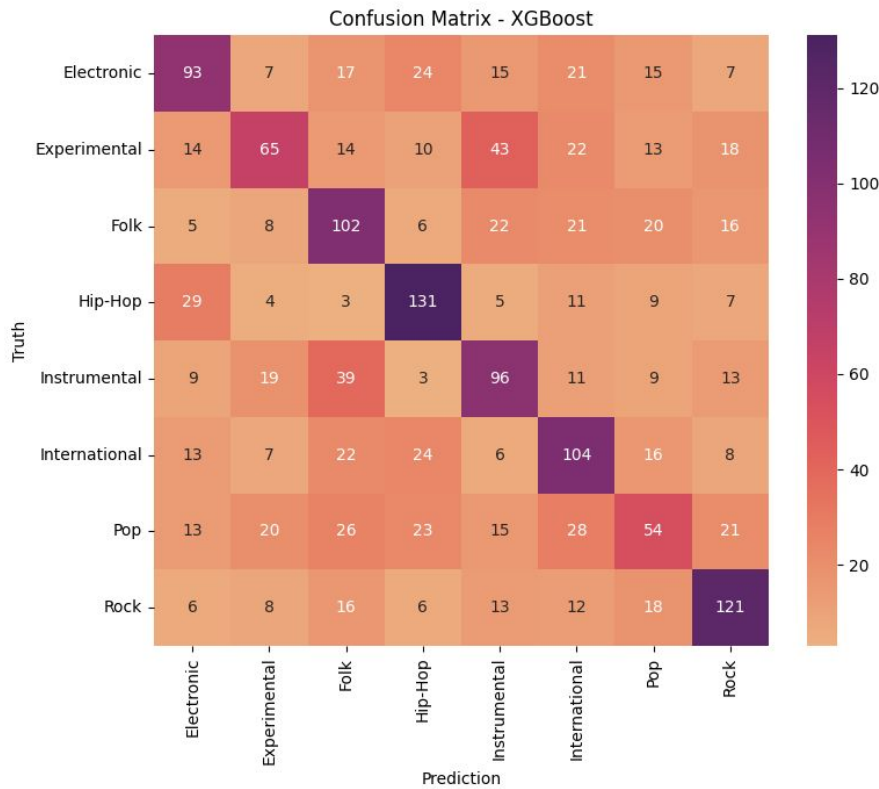
6) Modèle XGBoost

 - Surapprentissage



6) Modèle XGBoost

 - Bilan



Accuracy : 0.48

7) Comparaison



Accuracy k-NN : 0.39

Accuracy arbre de décision : 0.35

Accuracy random forest : 0.47

Accuracy XGBoost : 0.48

Tous les modèles ont du mal à deviner qu'un morceau est de l'expérimental ou de la pop.

Dans les 3 modèles utilisant des arbres, les features spectral_centroid, percussive_rms et les mfcc de 1 à 3 sont à chaque fois décisives.

8) Conclusion & Perspectives



Convergence vers une accuracy de 0.50

Meilleur modèle existant sur FMA : accuracy de 0.88

Plus grande base de données

Le genre d'un morceau reste subjectif

Tester avec un réseau de neurones