

PRÉDICTION DU DÉCROCHAGE SCOLAIRE AVEC UN MLP

Projet de Deep Learning

Présenté par :
Romuald Courtois, Fall Pape Ahmed

Encadrant : Tristan Vanrullen

Sommaire

- 1 INTRODUCTION
- 2 THÉORIE DU MLP
- 3 ANALYSE EXPLORATOIRE
- 4 PRÉPARATION ET FEATURE ENGINEERING
- 5 IMPLÉMENTATION D'UN MLP EN NUMPY
- 6 OPTIMISATION DES HYPERPARAMÈTRES
- 7 MODÉLISATION KERAS
- 8 COMPARAISON KERAS VS NUMPY
- 9 CONCLUSION ET PERSPECTIVES



Introduction

CONTEXTE

Appliquer MLP sur le jeu de données
“Predict Students’ Dropout and Academic Success” .

OBJECTIF

Prédire l'issue scolaire des etudiants

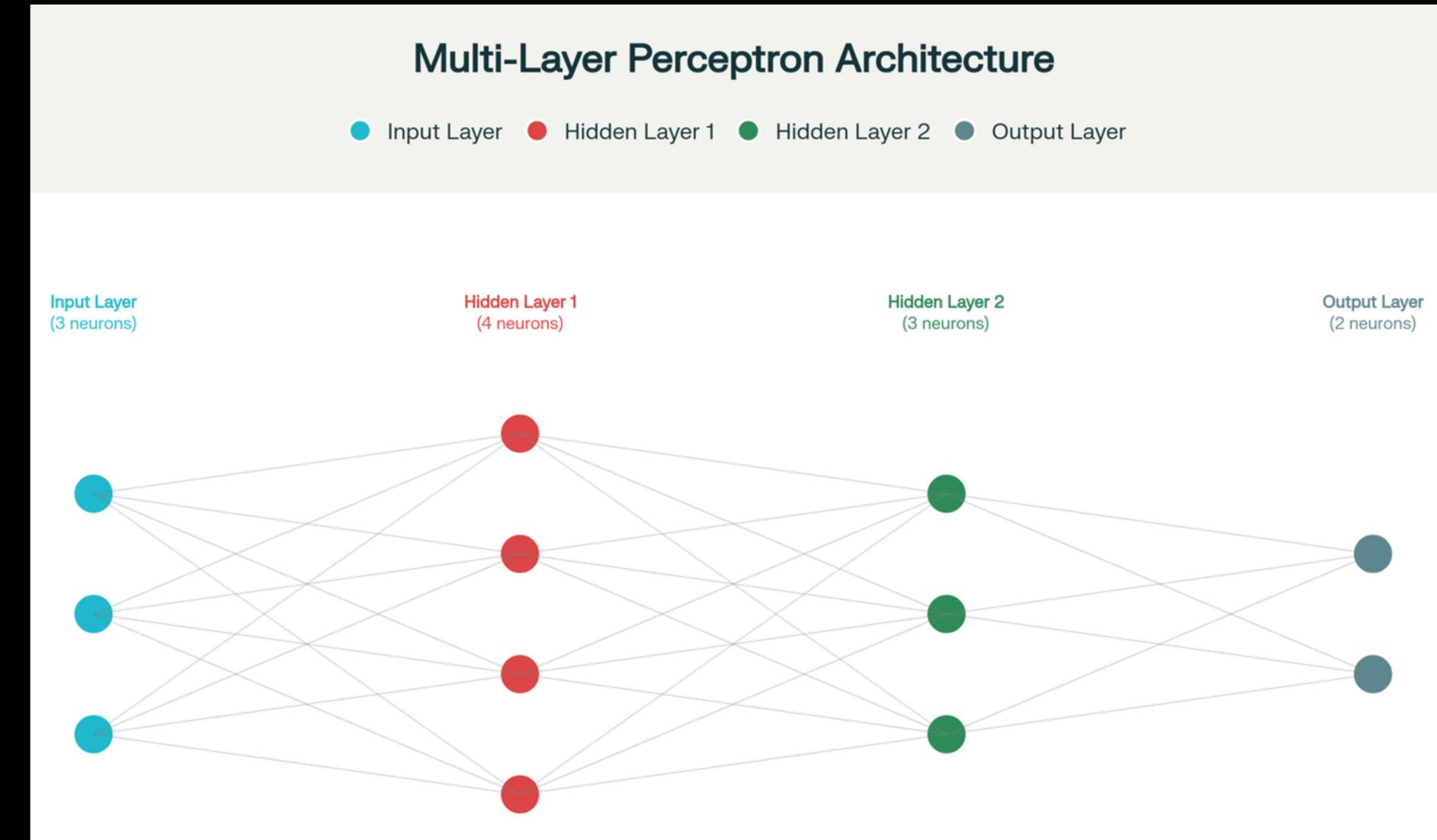
Théorie du MLP

Un MLP contient :

- Input Layer : variables explicatives
- Hidden Layers : neurones + activation
- Output Layer : probabilité binaire (sigmoïde)

Vocabulaire :

- Propagation avant
- Fonction d'activation : relu, tanh, sigmoid
- Loss function : cross-entropy
- Descente de gradient & backpropagation
- Vanishing Gradients : saturation tanh/sigmoid

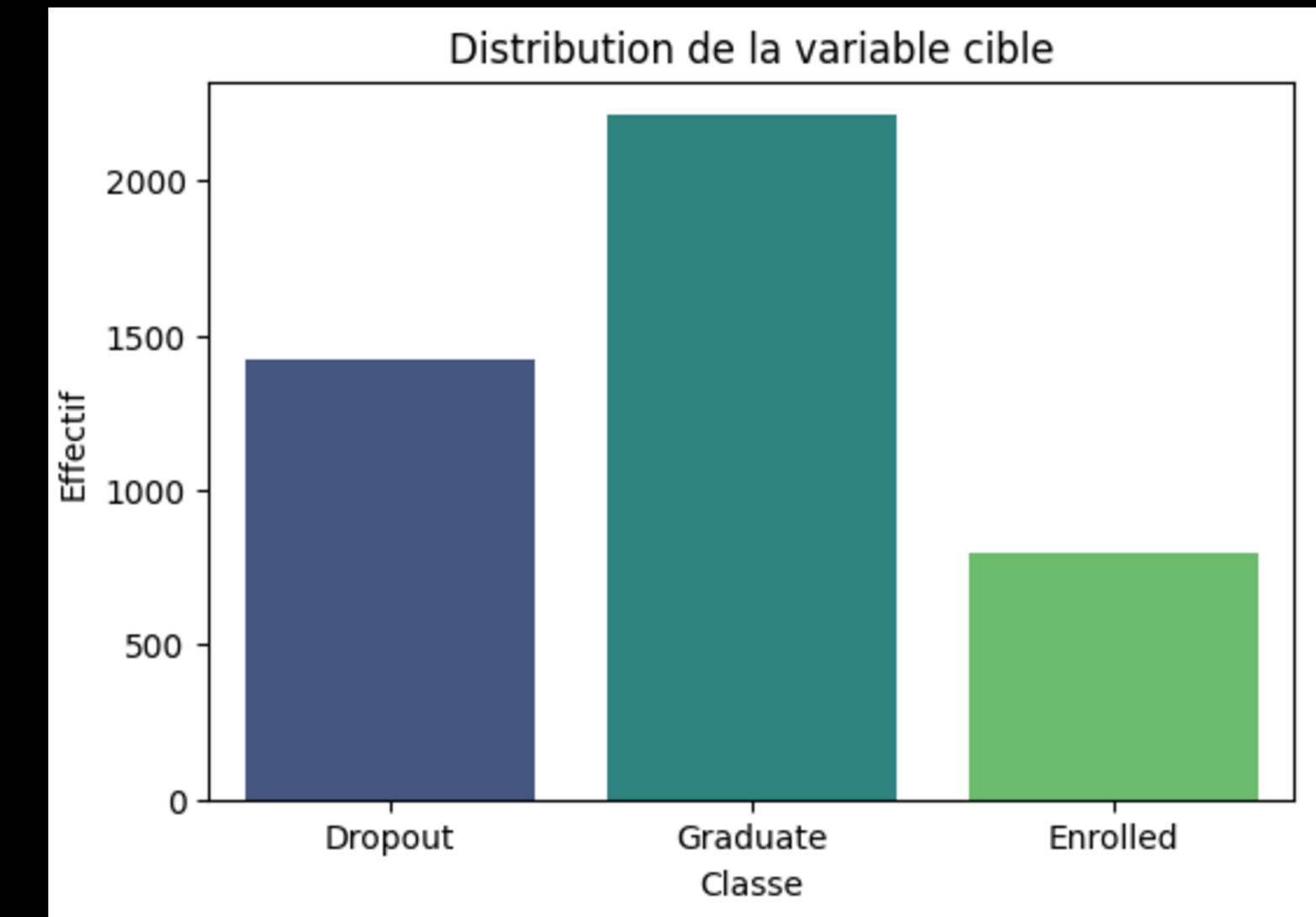


EDA

“Predict Students’ Dropout and Academic Success”

Target :

- Graduate (49.9 %)
- Dropout (32.1 %)
- Enrolled (17.9 %)



4424

Observations

36

Features

1

Target

Préparation et feature engineering

- Pas de doublons ni de valeurs manquantes
- Encodage de “Target”
 - Dropout = 0, Graduate = 1

Séparation du Dataset

- Graduate + Dropout (61% + 39%)

→ SMOTE rééquilibrer artificiellement (train)

3436
Observations

36
Features

1
Target

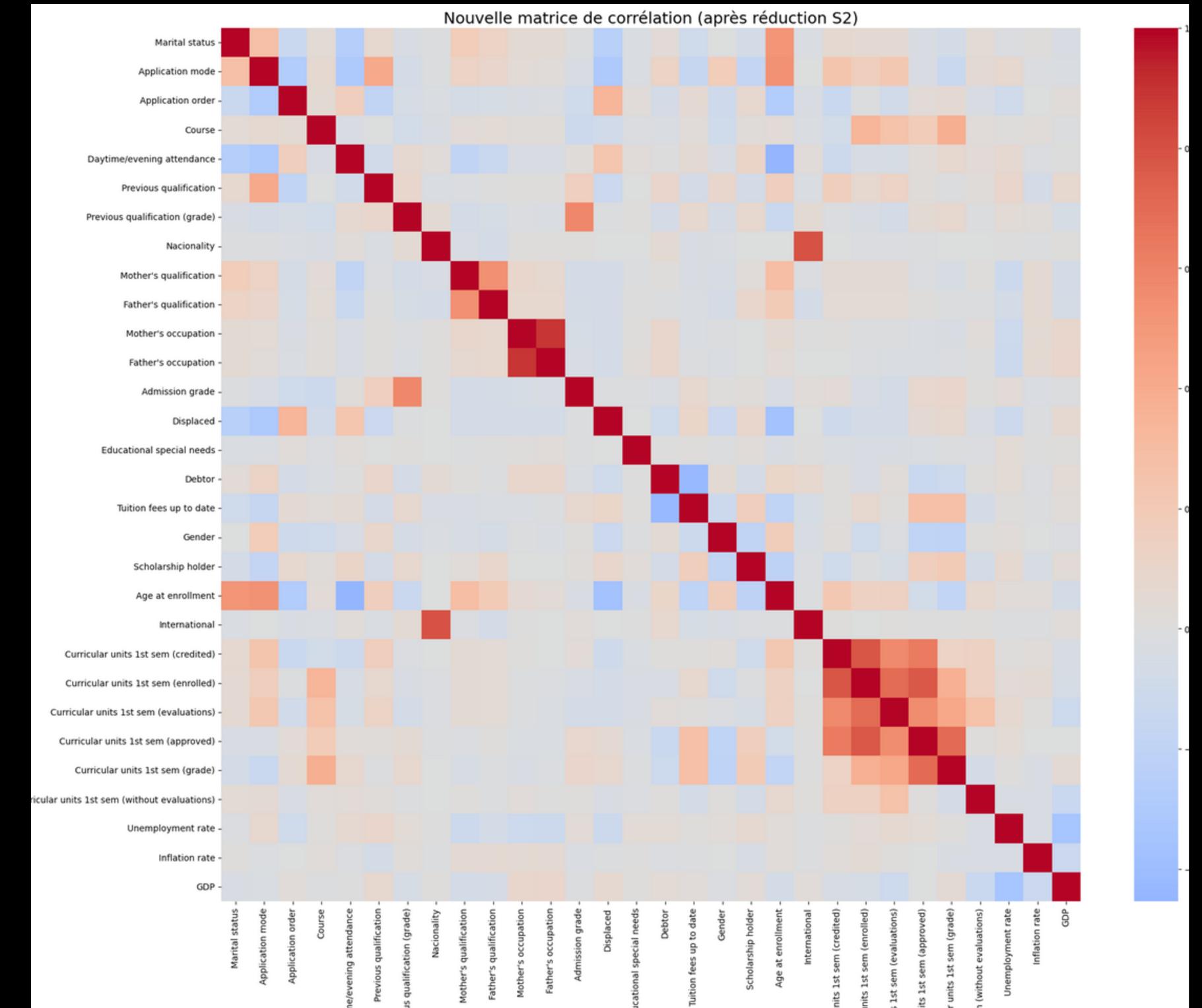


Matrice de corrélation

→ Plusieurs caractéristiques fortement corrélées entre elles

→ Multicolinéarité

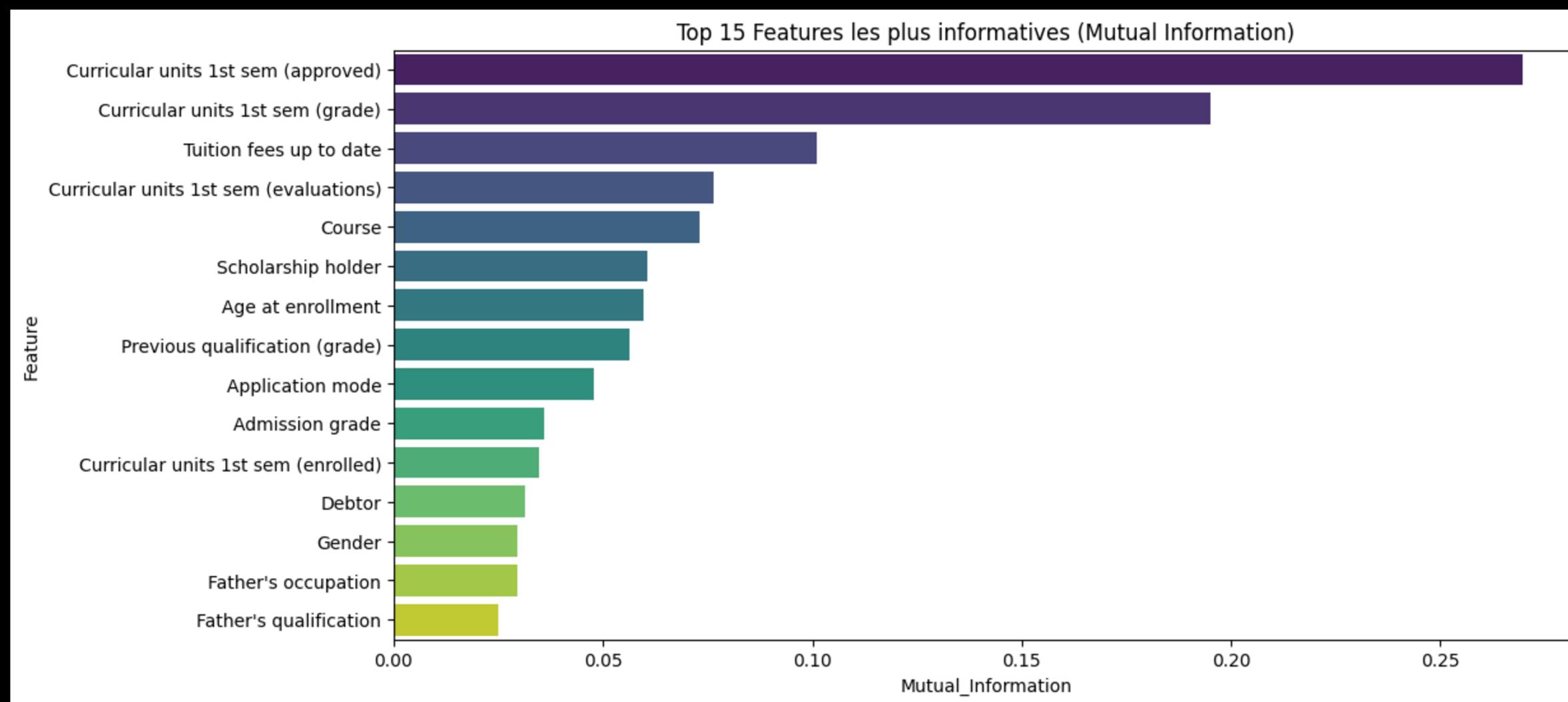
→ Suppression variables (corrélattion > 0.85)



Sélection des features

Basée sur l'importance des features

Après suppression des fortements corrélées

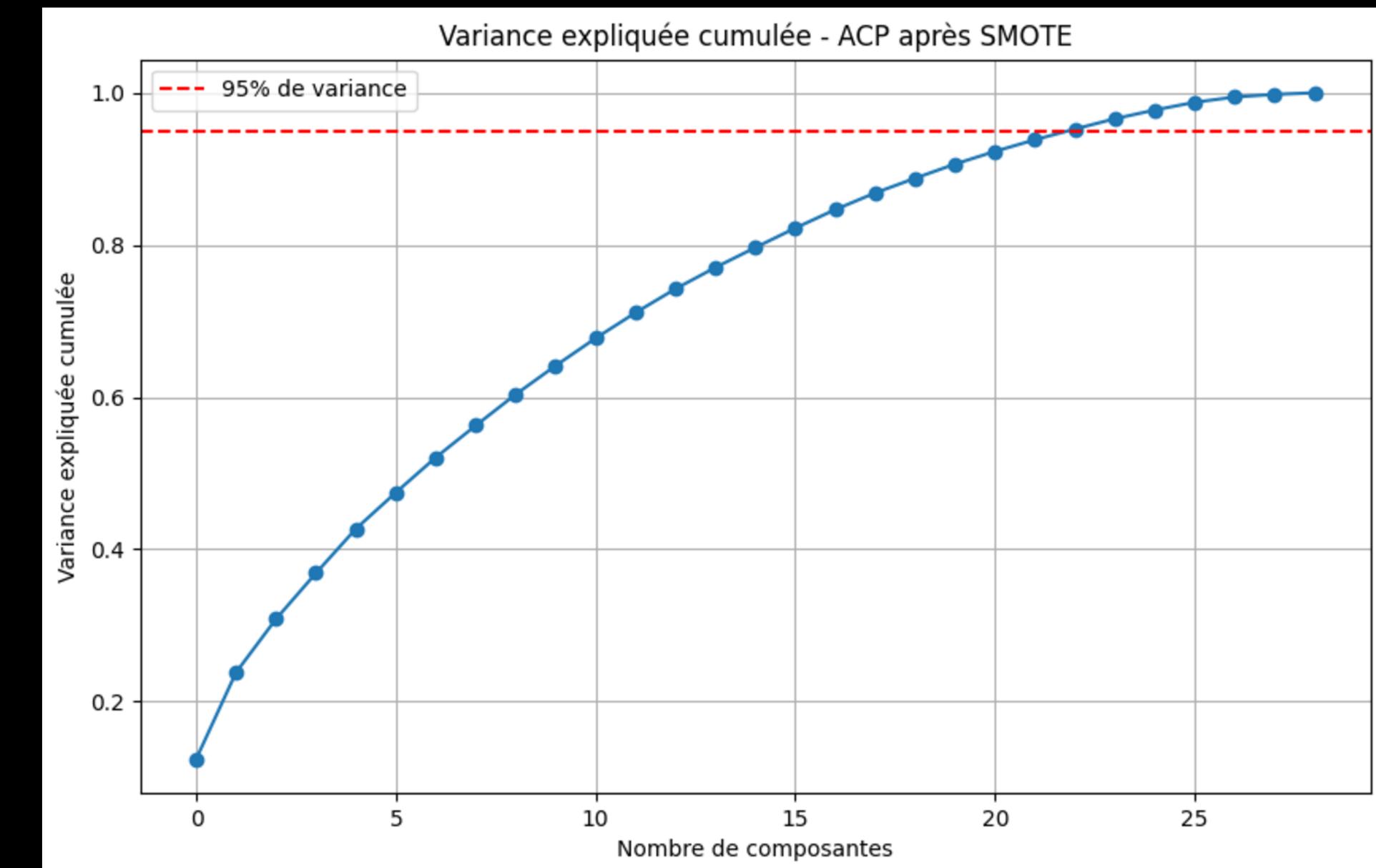


Réduction de dimension

Analyse en Composantes Principales
(ACP)

Sélection des composantes qui expliquent 95% de la variance

23 composante suffisent pour conserver 95% de l'information



Modélisation Keras

**Deux couches cachées
(64 et 32)**

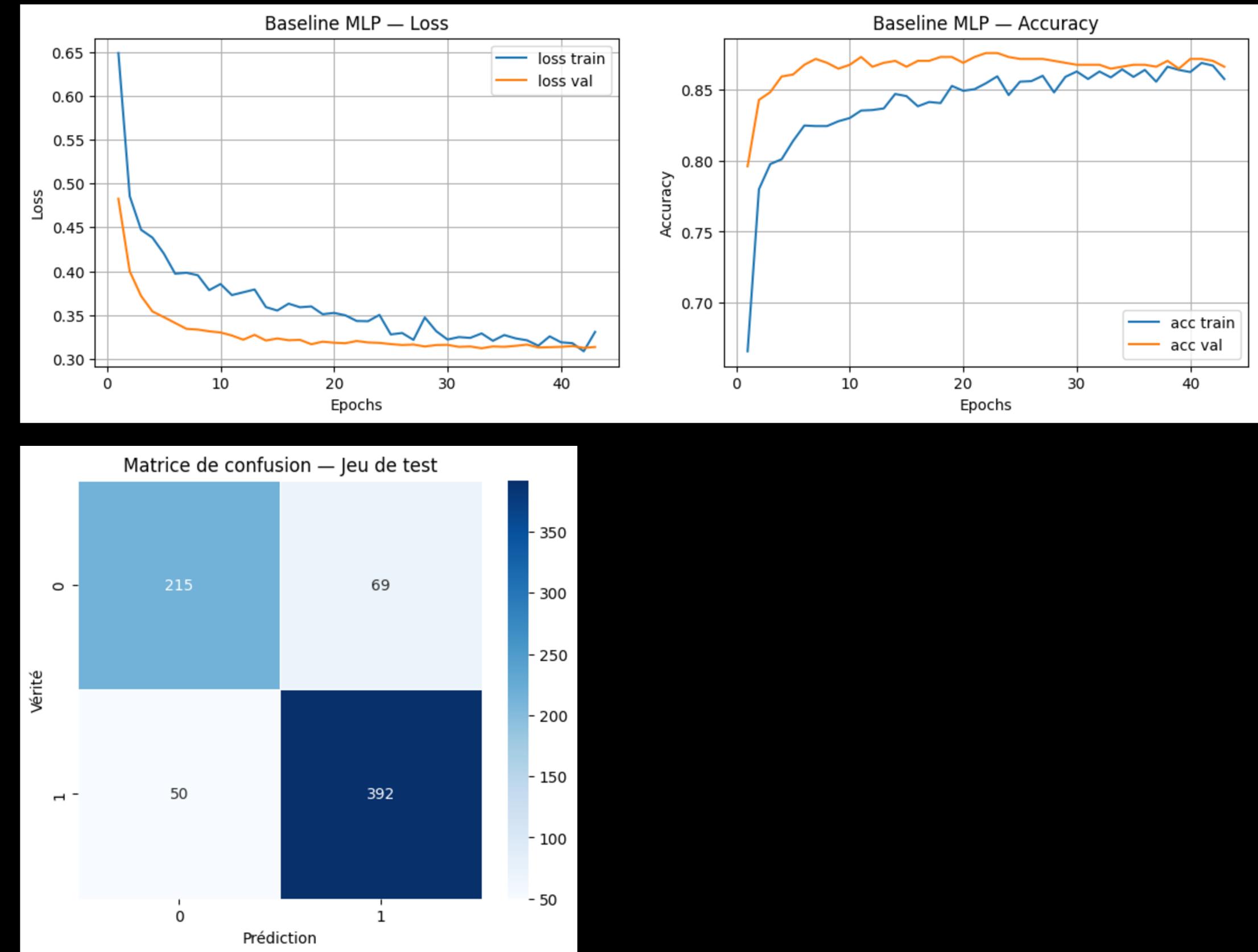
**BatchNorm + Dropout
(Stabilisation et régularisation)**

**Activation : Relu
(éviter vanishing gradient)**
**Optimiseur : Adam
(lr : 0.001)**

**Sortie sigmoid
(Binaire)**

Performances avec Keras

- Accuracy Test : 0.8457
- F1-score Graduate : 0.88
- F1-score Dropout : 0.80



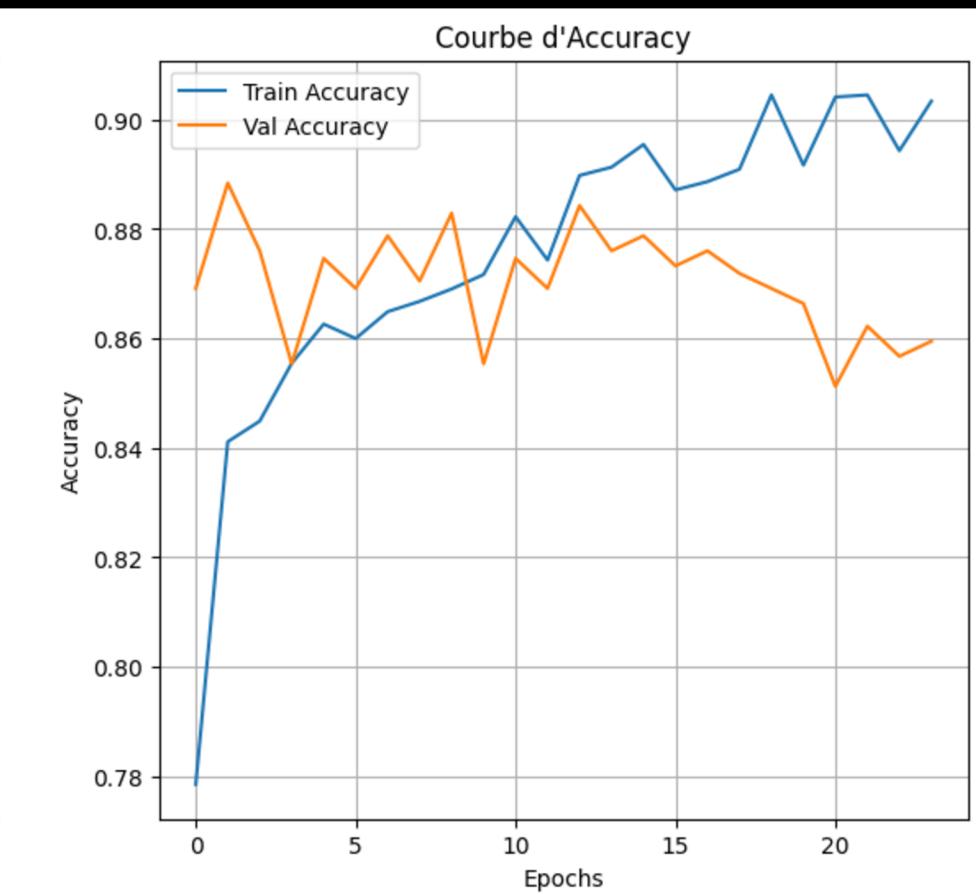
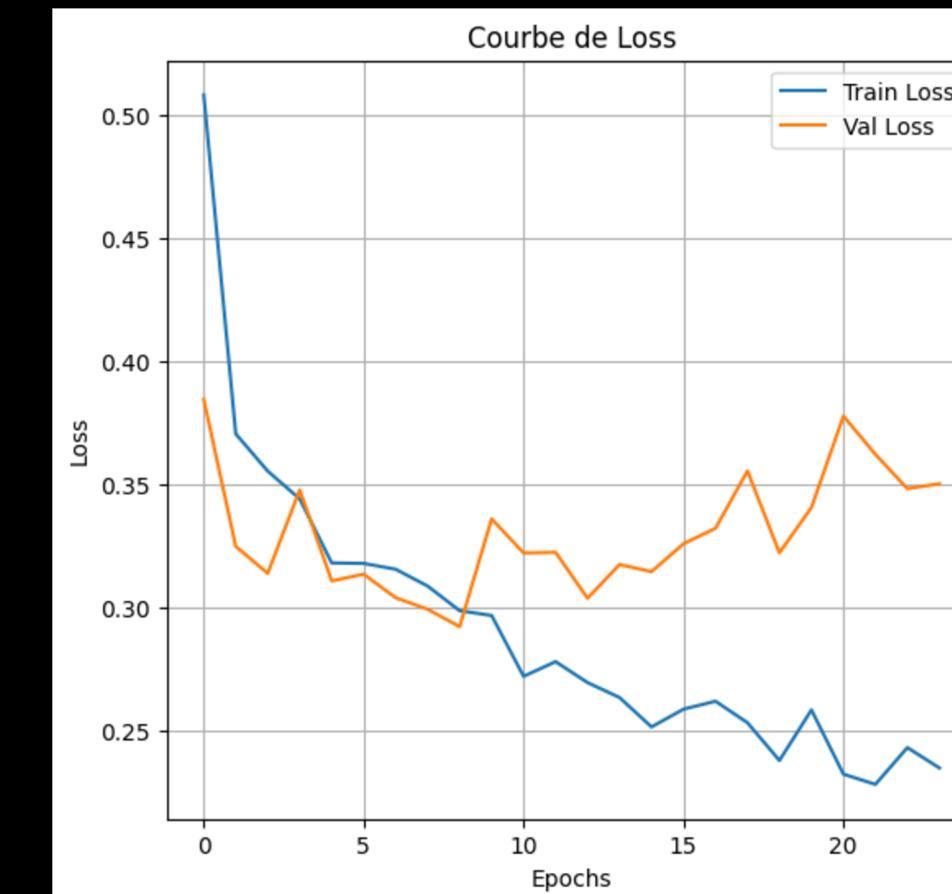
Hyperparamètres

- Recherche :
 - 1-3 couches
 - 32-256 neurones
 - Activation : ReLU vs Tanh
 - Dropout 0-0.5
 - Learning rate log-sampling
 - Early stopping : ON / OFF
- Meilleur modèle :
 - 96 neurones + 160 neurones
 - ReLU
 - Dropout(0.1 / 0.4)
 - LR \simeq 0.005
 - Early Stopping : ON

Performances du modèle

Résultat modèle optimisé

- Accuracy test : 83.7%
(similaire à avant)
- Gain en stabilité
- Validation plus stable
(Le tuning affine le modèle
mais ne transforme pas
radicalement la performance.)



Implémentation MLP

Initialisation Xavier
(Donner un meilleur départ au réseau)

Forward pass avec activations relu + sigmoid
(prédiction sans Keras)

Backpropagation manuelle
(Apprendre en corigeant ses poids)

Mesurer l'erreu : cross-entropy
(adaptée à la classification binaire)

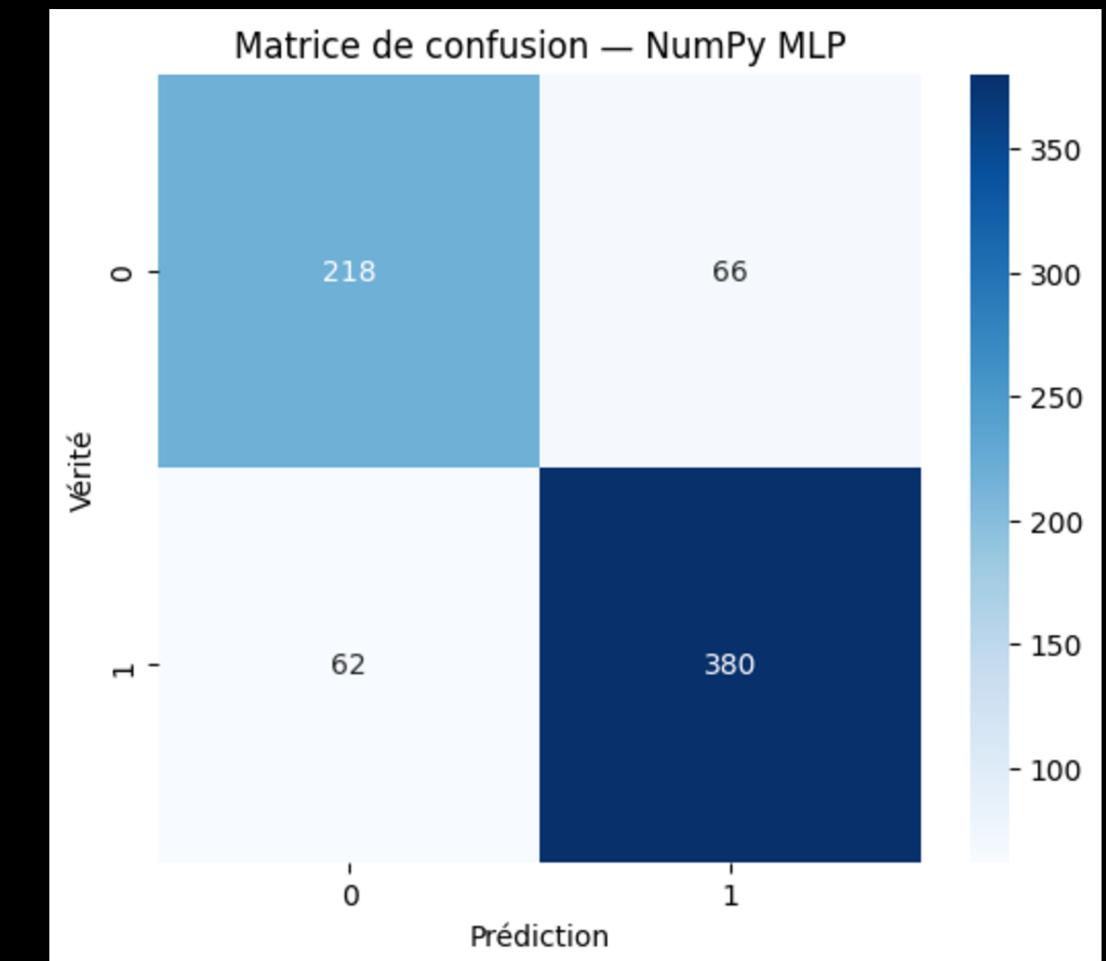
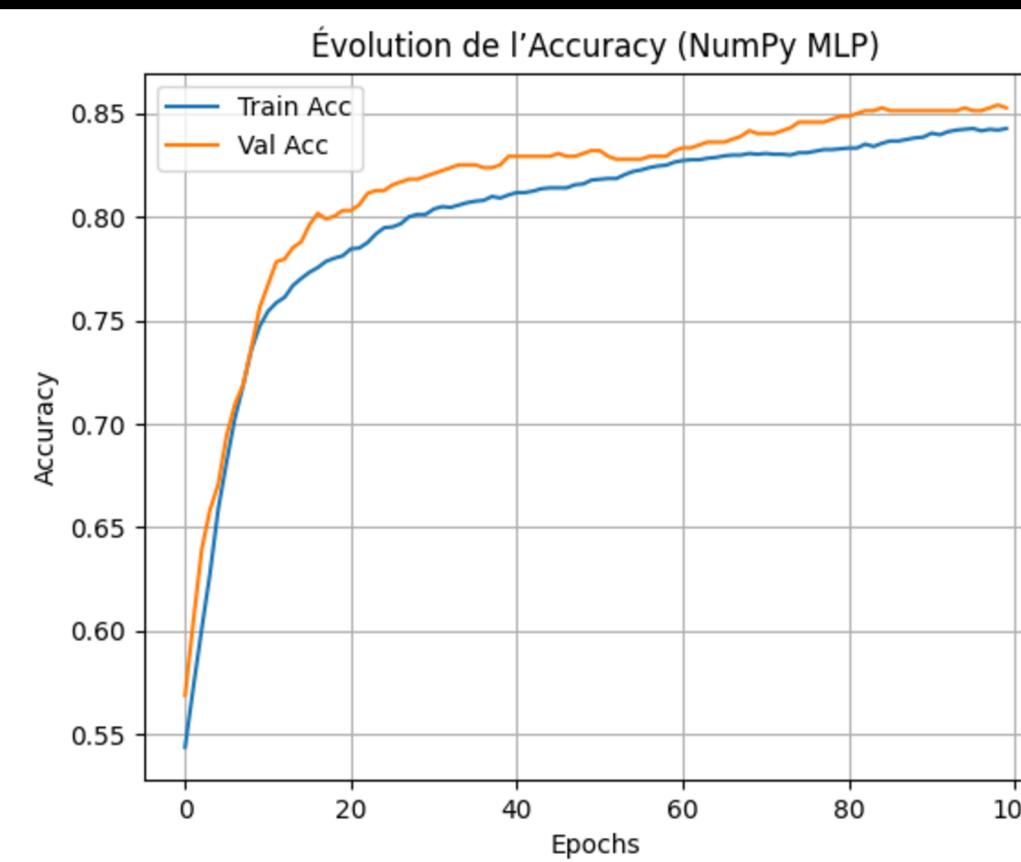
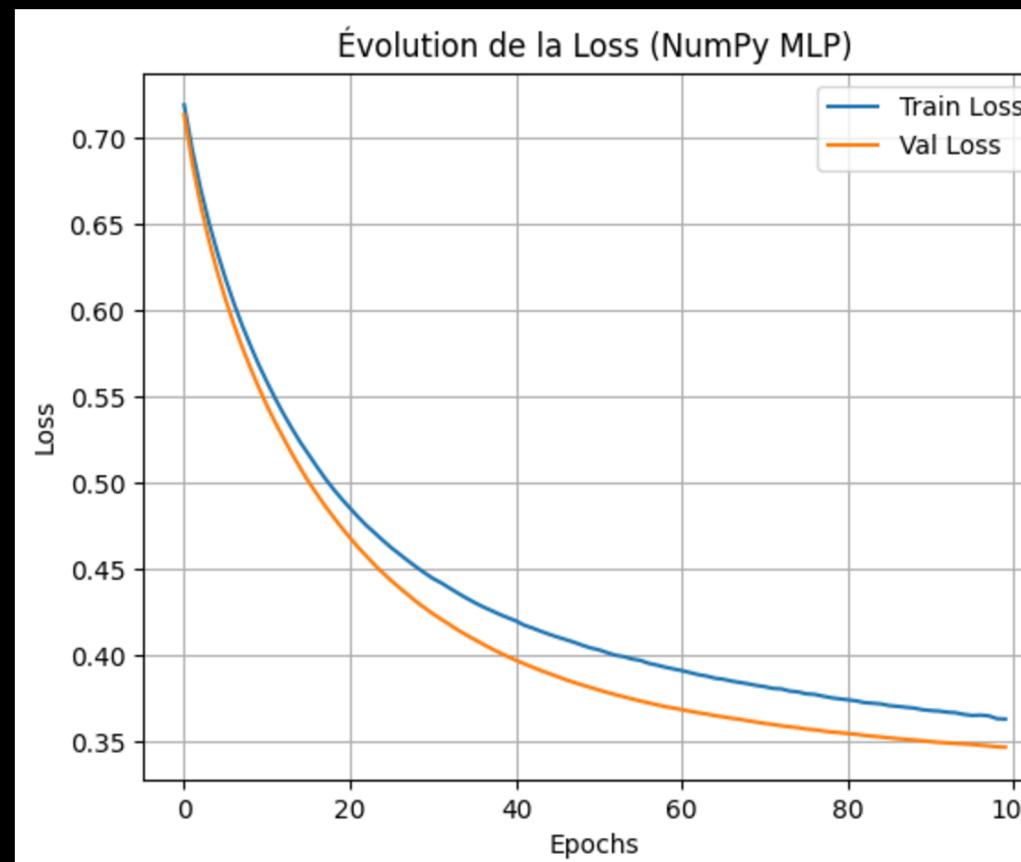
Descente de gradient mini-batch
(entraînement plus stable et plus rapide)

Courbes loss/accuracy

Performances Numpy

- Accuracy test : 82.3%
- Courbes propres
Convergence progressive
Pas d'overfitting sévère

Keras garde un léger avantage



Comparaison

Critère	Keras	NumPy
Performance	★★★★★	★★★
Stabilité	★★★★★	★★★
Complexité	★★	★★★★★
Optimisation du gradient	Avancée	Manuelle
Courbes	Stables	Correctes

Conclusion et perspectives

- Résultats acceptables de notre modèle MLP
- Keras performant, NumPy validé
- Dropout + BatchNorm : impact positif
- Utiliser une recherche hyperparamétrique plus large
- Tester un réseau plus profond
- Ajouter d'explication (SHAP)
- Prédire les résultats des étudiants enroled

Merci pour votre
attention

SUPPRESSION DES DONNEES

Suppression de :

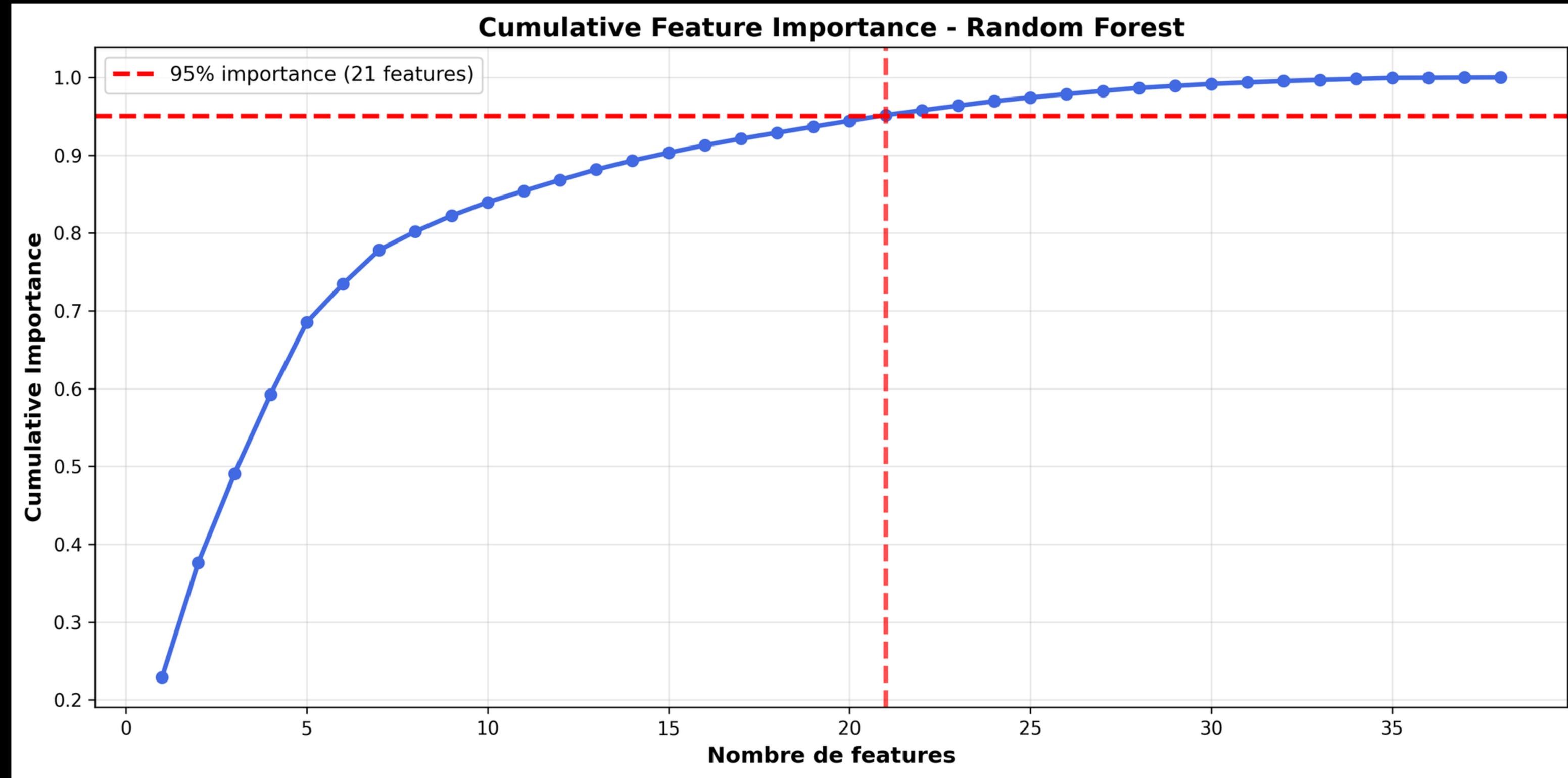
- Previous qualification
- Nacionality
- Mother's qualification
- Father's qualification
- Mother's occupation
- Father's occupation
- Course (171)
 - aucune évaluation réalisée pour cette filière (4.89 %)

AJOUT DES FEATURES

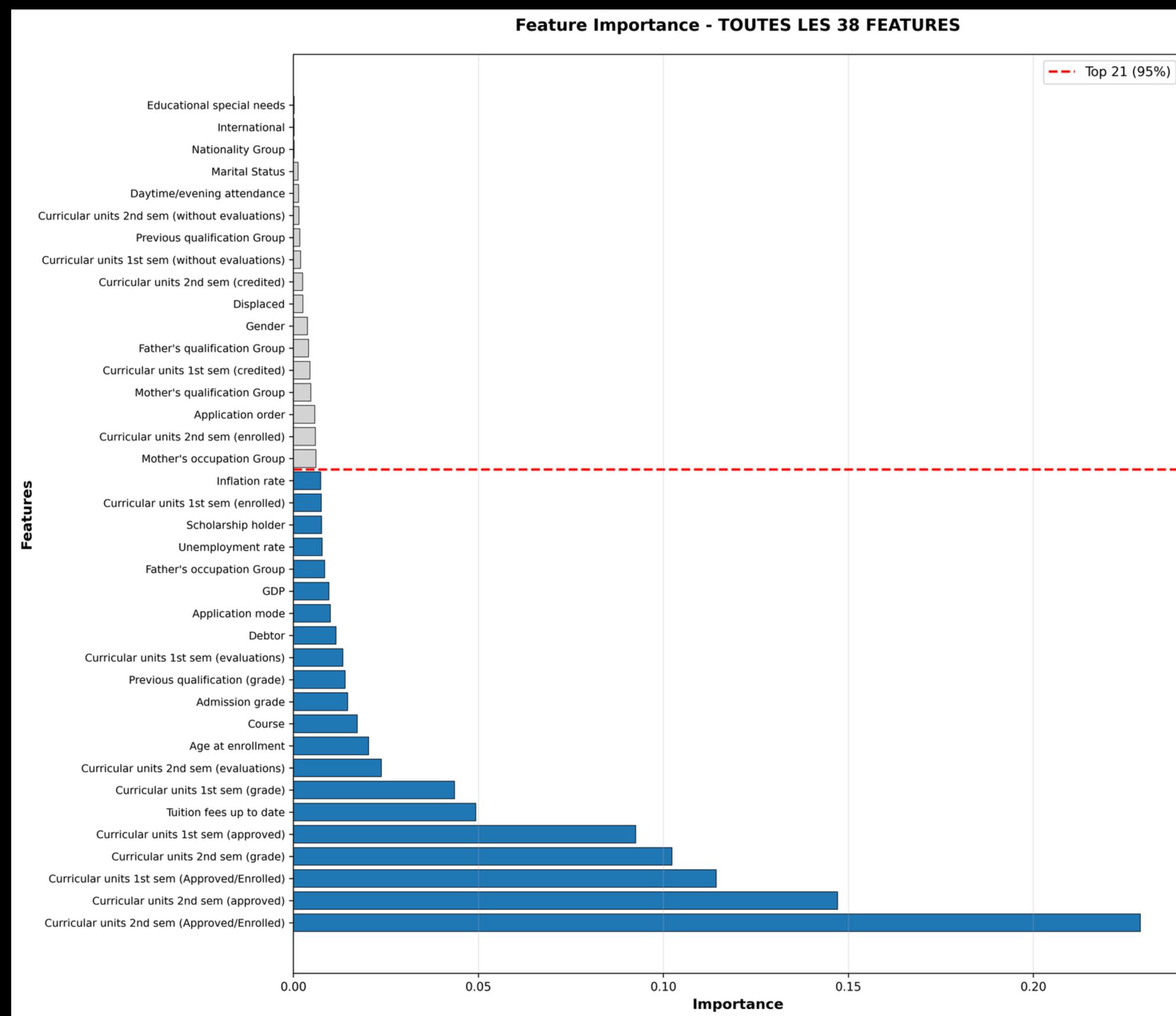
Ajout de :

- Nationality Group
- Mother's qualification Group / Father's qualification Group
- Mother's occupation Group / Father's occupation Group
- Curricular units 1st sem (Approved/Enrolled) et Curricular units 2nd sem (Approved/Enrolled)
-

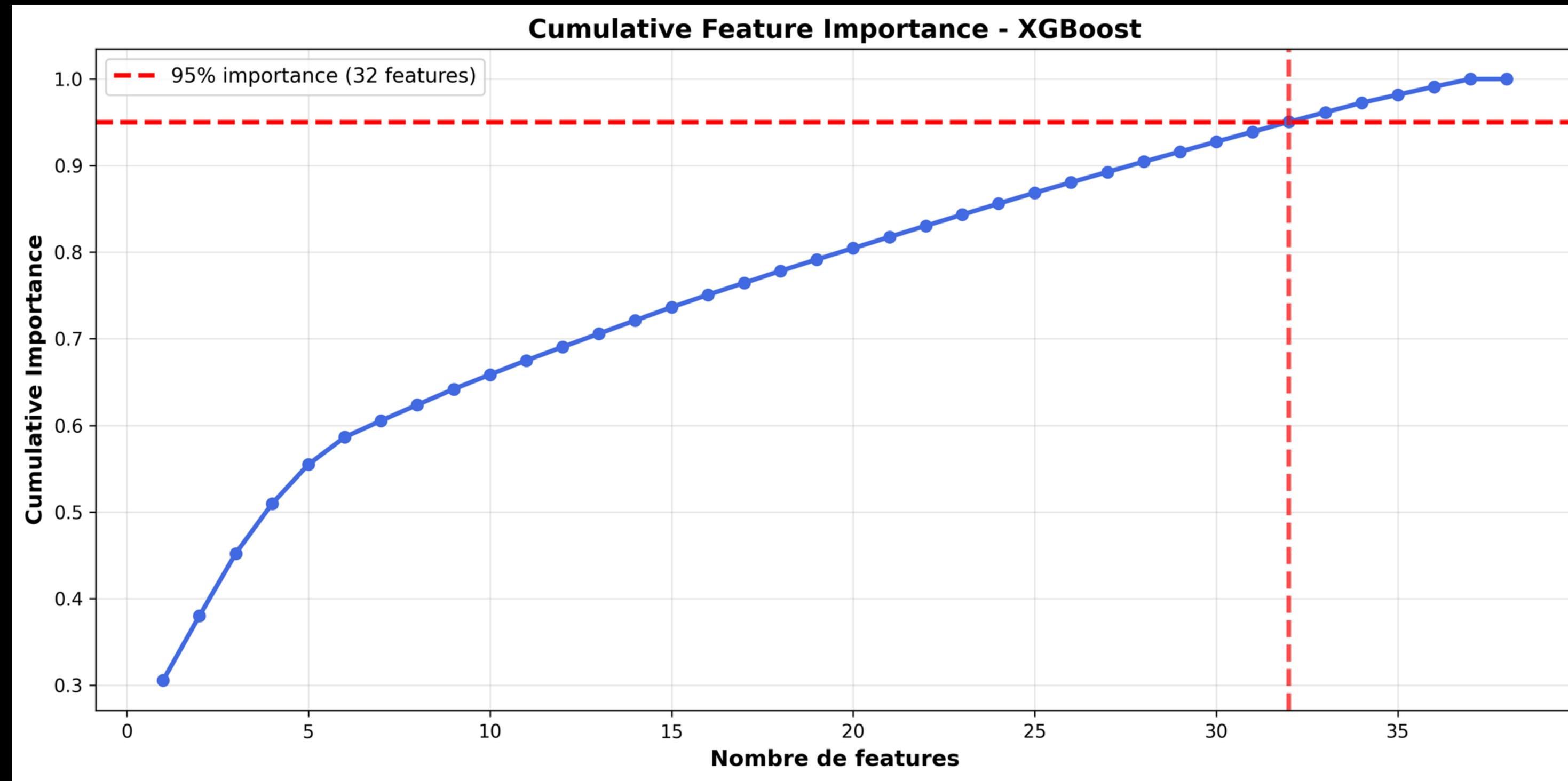
RANDOM FOREST



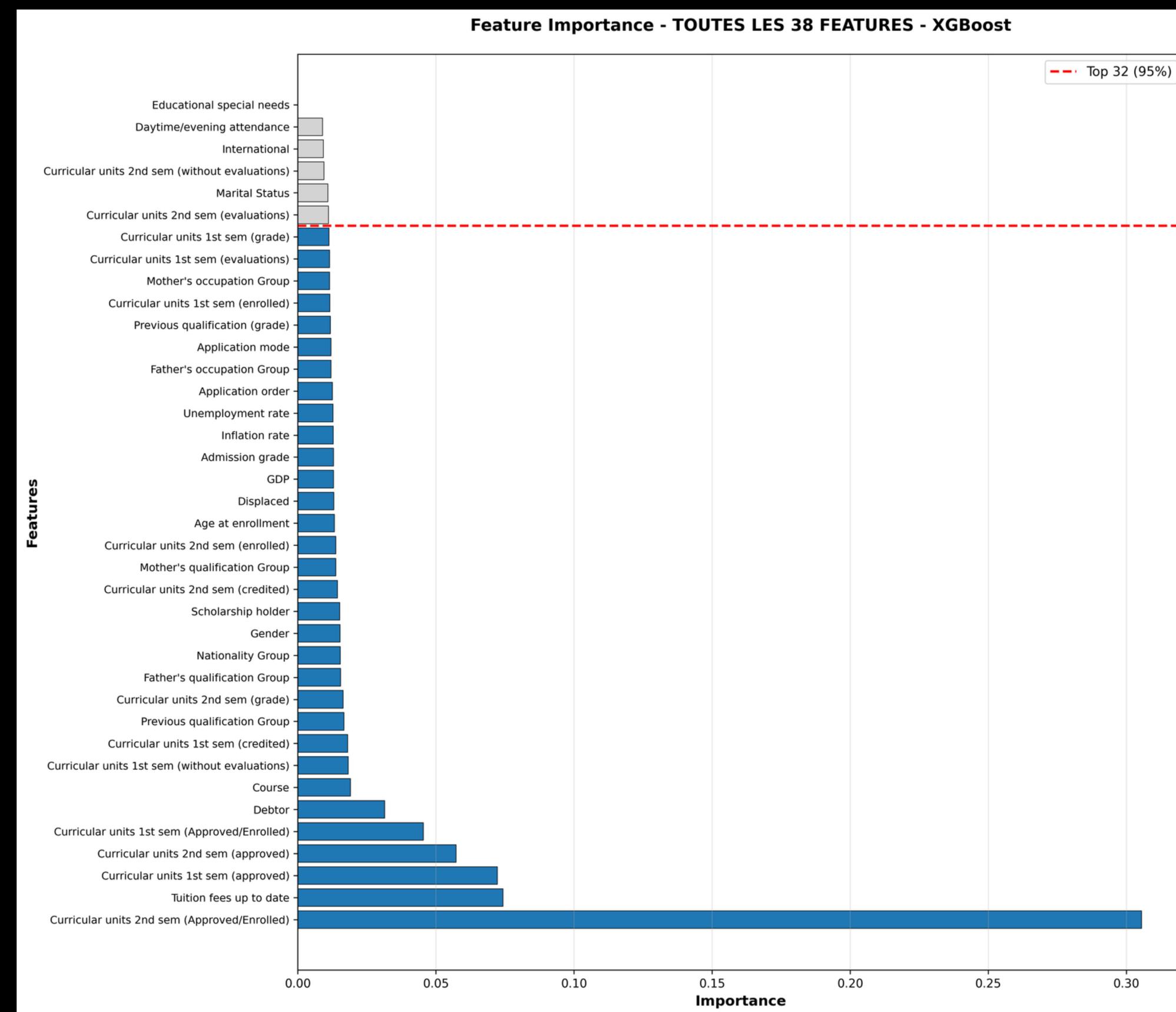
RANDOM FOREST



XGBOOST



XGBOOST



FAMD

