R.A.I.D.

RAM Artificial Inspection Dump

Sommaire

Contexte

- La RAM présente des données volatiles, fragmentées, et résiduelles

- Les outils forensiques principaux (Rekall, Volatility …) sont basés sur des signatures statiques

- Étendre le travail de Thomas Gougeon sur les cartes à puce vers des dumps avec structures multiples

Problématique

Dans quelle mesure une architecture d'apprentissage profond de type Transformer, exploitant uniquement les relations contextuelles locales et globales d'une séquence binaire brute, permet-elle de segmenter efficacement un dump de RAM hétérogène, et ce, même en l'absence de descripteurs explicites ou des signatures traditionnelles?

Les buts du projet :

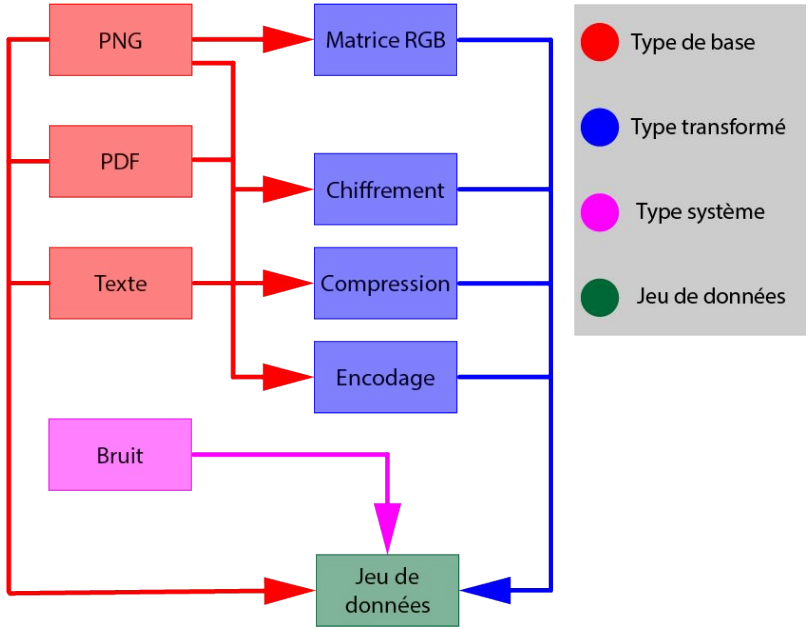
- Mettre en place un transformer entraîné sur des dumps de RAM synthétiques
- Classifier les octets (Chiffrés / Non chiffrés)

Récupération des données

Dataset constitué de données Open-Source:

- Project Gutenberg (Livres)
- ArXiv (PDF)
- Lorem Picsum (Images)

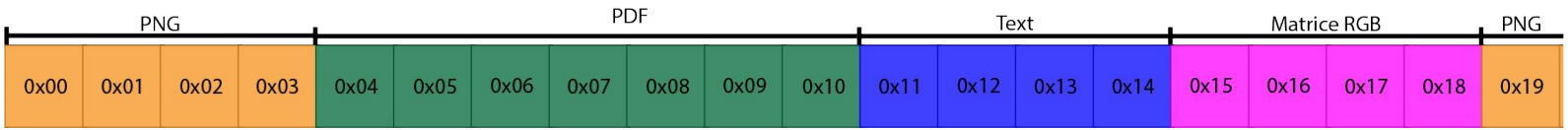
Diversification des données



Représentation visuelle des transformations

Création du segment RAM synthétique

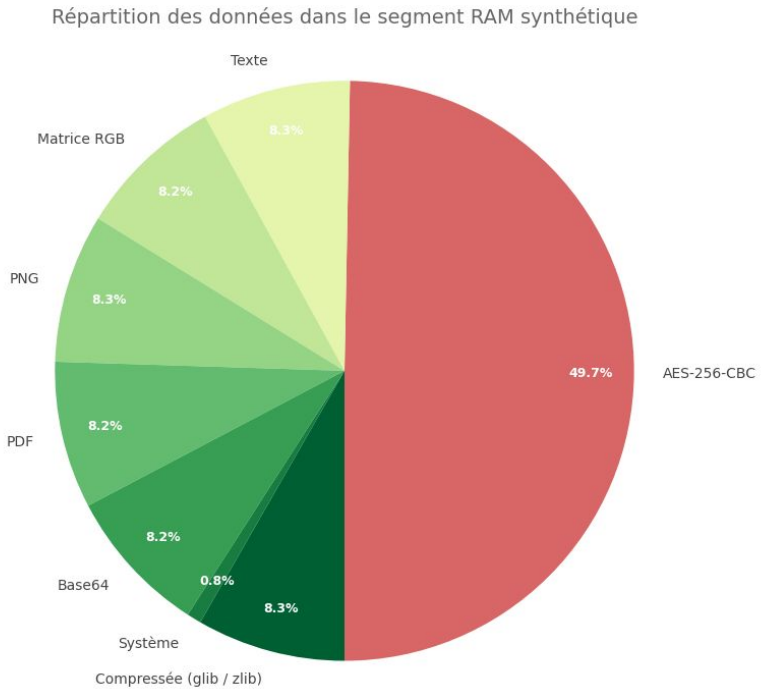
- Concaténation des données
- Fragmentation
- Utilisation de graines pour la reproductibilité avec le même jeu de données



Exemple de RAM synthétique

Répartition dans les datasets

Nous utilisons un dump de RAM synthétique équilibré avec environ 50% de classe chiffrée.



Architecture du Classifieur : Le Transformer (encoder-only)

- Segmentation : Analyse par blocs (chunks) de 1024 octets, dans un batch de 128
- Analyse granulaire : Évaluation individuelle de chaque octet.
- Sortie : Génération d'une distribution de probabilités (logits) par segment.

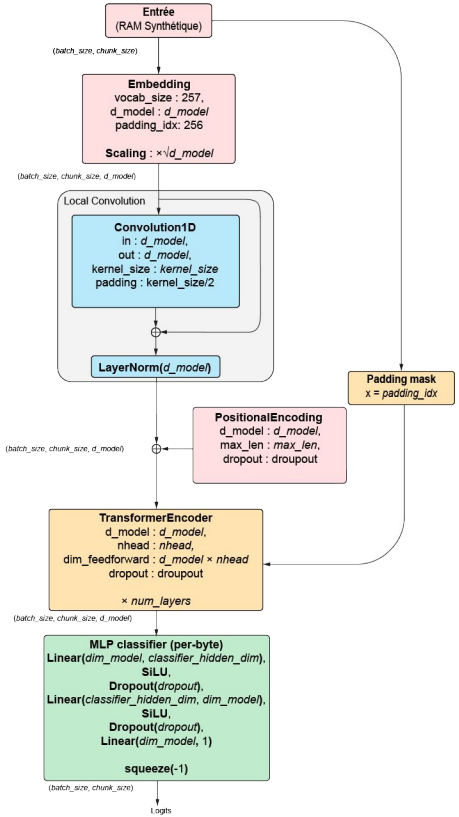


Schéma du Transformer

Entrainement

- 20 époques
- 0.2 de dropout
- LR initial à $1e^{-4}$

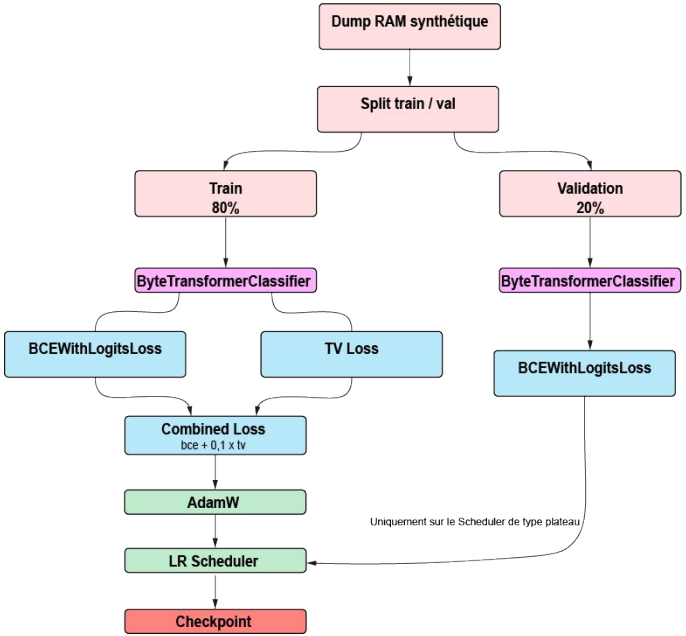
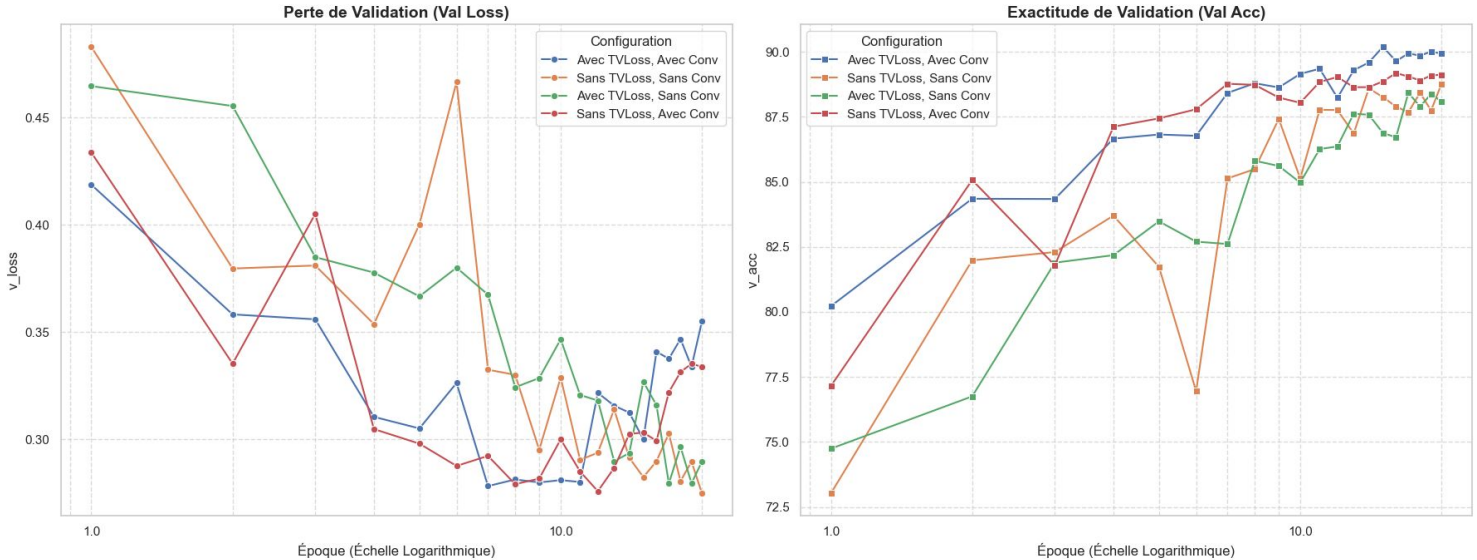
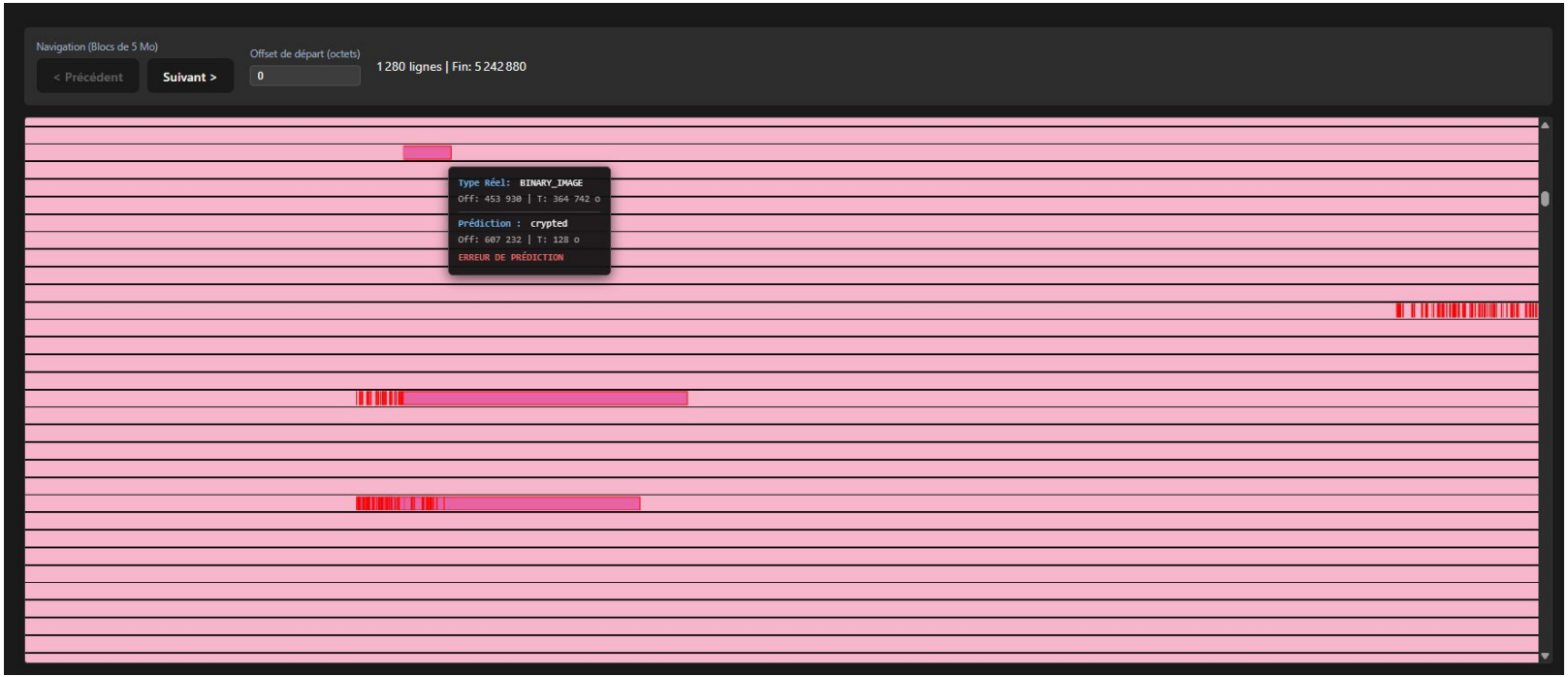


Schéma du processus de l'entraînement

Justification de l’architecture



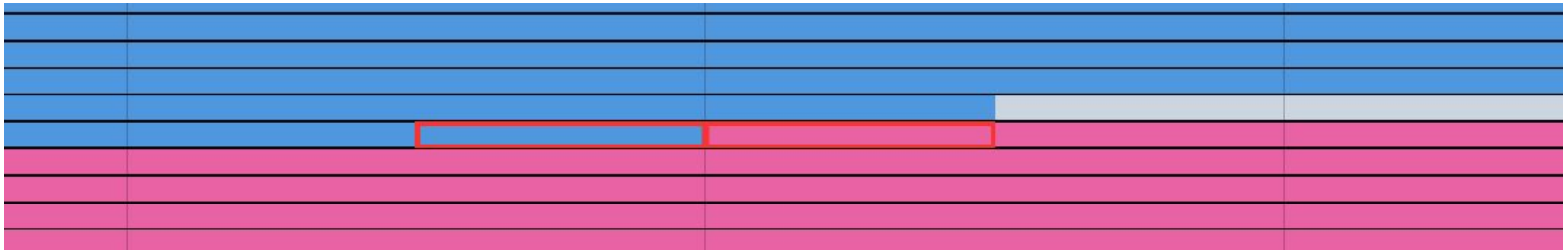
Visualiseur



Vue tronquée d’une partie du visualiseur avec des octets mal classés (en rouge)

Visualisation des erreurs de chevauchement

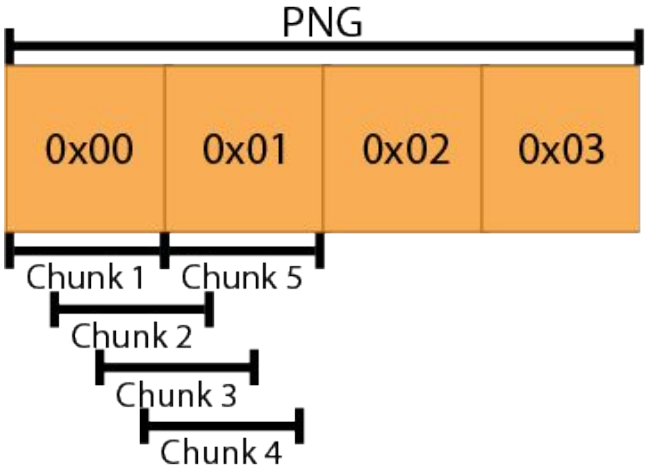
Environ 30% des erreurs provenaient de zones de “chevauchement”



Vue tronquée d’une partie du visualiseur

Mise en place du chevauchement

Le dataset est découpé en chunks de 1024 octets avec un offset de 256 octets.



Exemple d'une découpage en chunks

Post-traitement des logits

Pour convertir les logits en probabilité, on utilise une activation sigmoïde.

$$\text{Classe} = \begin{cases} 1 \text{ (Chiffré)} & \text{si } \sigma(\text{logit}) > 0.5 \\ 0 \text{ (Non Chiffré)} & \text{si } \sigma(\text{logit}) \leq 0.5 \end{cases}$$

Avec une fenêtre de 1024 octets et un pas de 256 octets, chaque octet (hors extrémités) est couvert par 4 fenêtres chevauchantes, produisant 4 logits indépendants par octet.

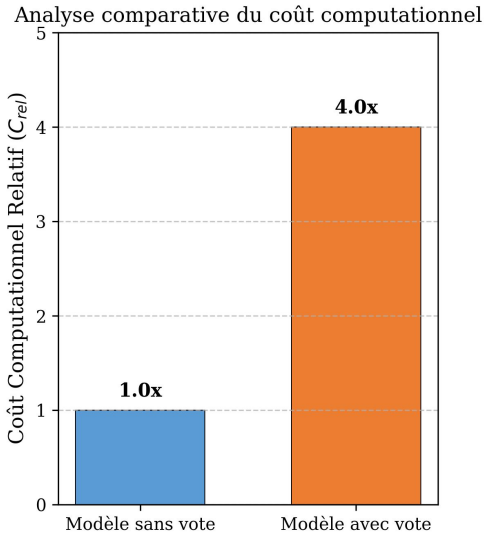
L'agrégation se fait par moyenne pondérée par la confiance :

$$\text{Confiance} = 2 \times |p - 0.5|$$

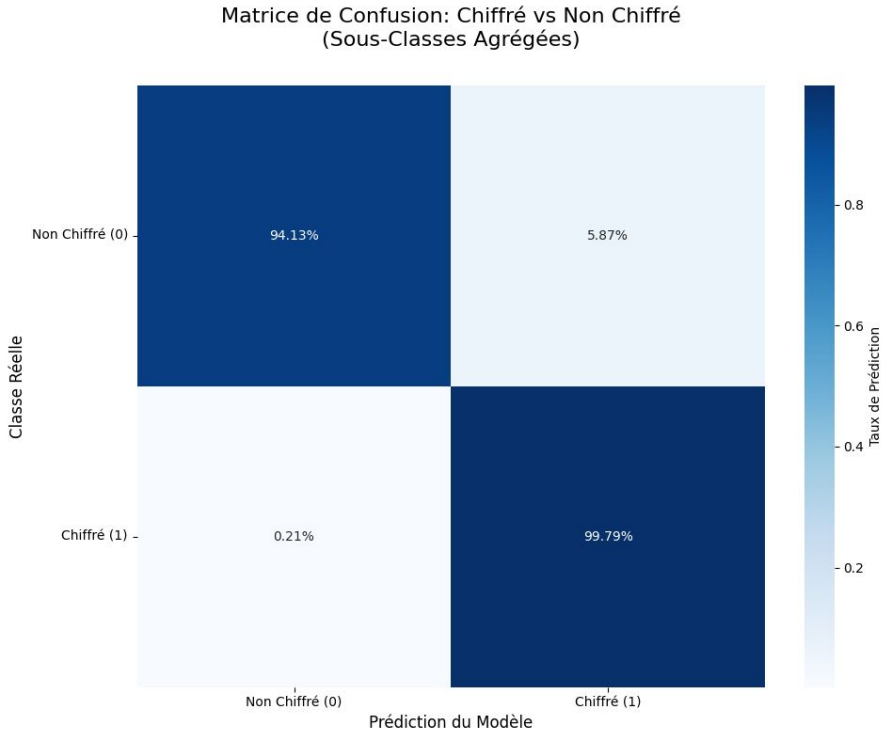
Comparaison des bénéfices du vote

Accuracy		
Sans vote pondéré	Avec vote pondéré	Diminution de l'erreur
92.77%	94.79%	27.94%

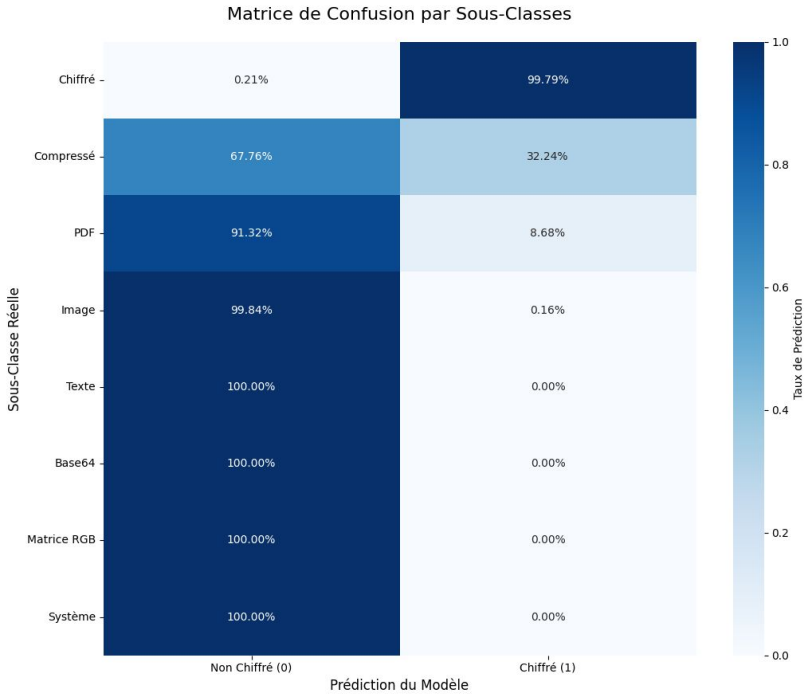
Performances du vote pondéré



Résultats

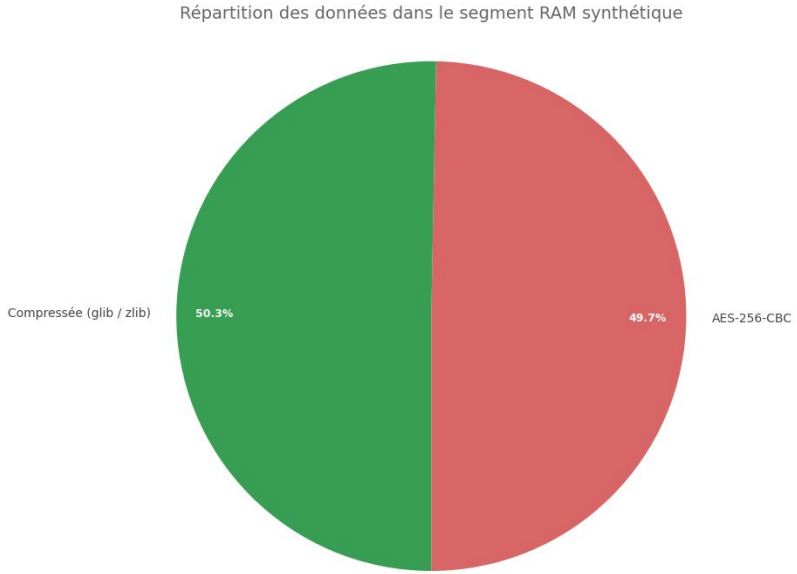

Matrice de Confusion: Chiffré vs Non Chiffré
(Sous-Classes Agrégées)
Non Chiffré (0)
Chiffré (1)
94.13%
5.87%
0.21%
99.79%
Classe Réelle
Non Chiffré (0)
Chiffré (1)
Prédiction du Modèle
0.8
0.6
0.4
0.2
Taux de Prédiction

Résultats

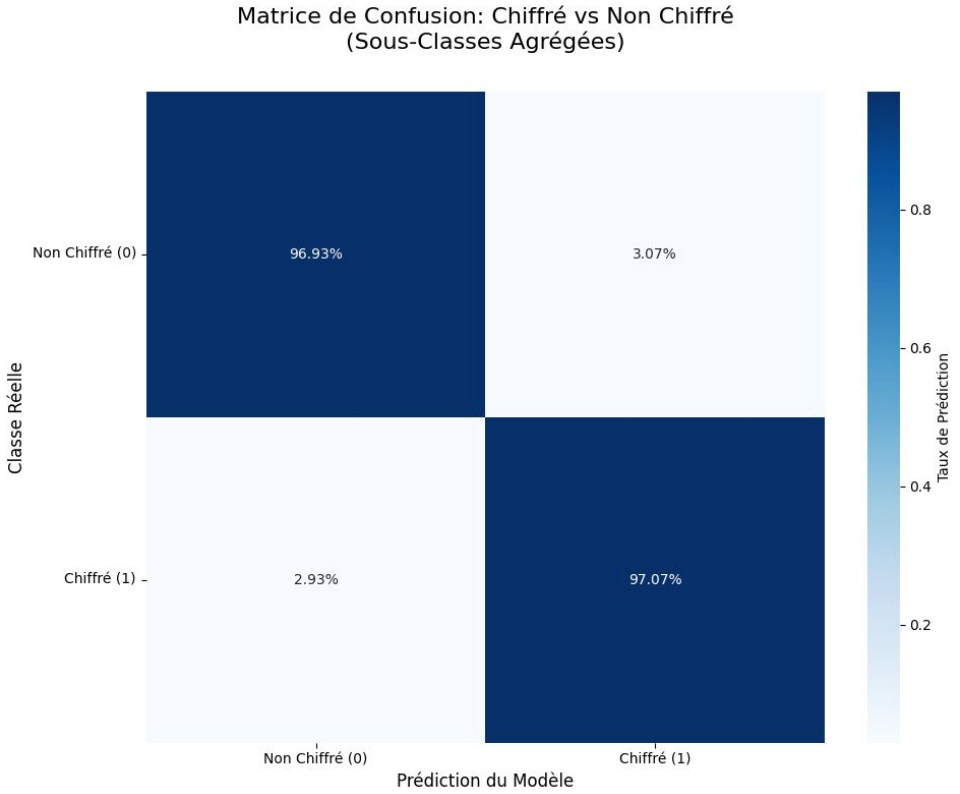

Matrice de Confusion par Sous-Classes
Chiffré
Compressé
PDF
Image
Texte
Base64
Matrice RGB
Système
0.21%
99.79%
67.76%
32.24%
91.32%
8.68%
99.84%
0.16%
100.00%
0.00%
100.00%
0.00%
100.00%
0.00%
100.00%
0.00%
Sous-Classe Réelle
Non Chiffré (0)
Chiffré (1)
Prédiction du Modèle
Taux de Prédiction
1.0
0.8
0.6
0.4
0.2
0.0

Fine-tuning

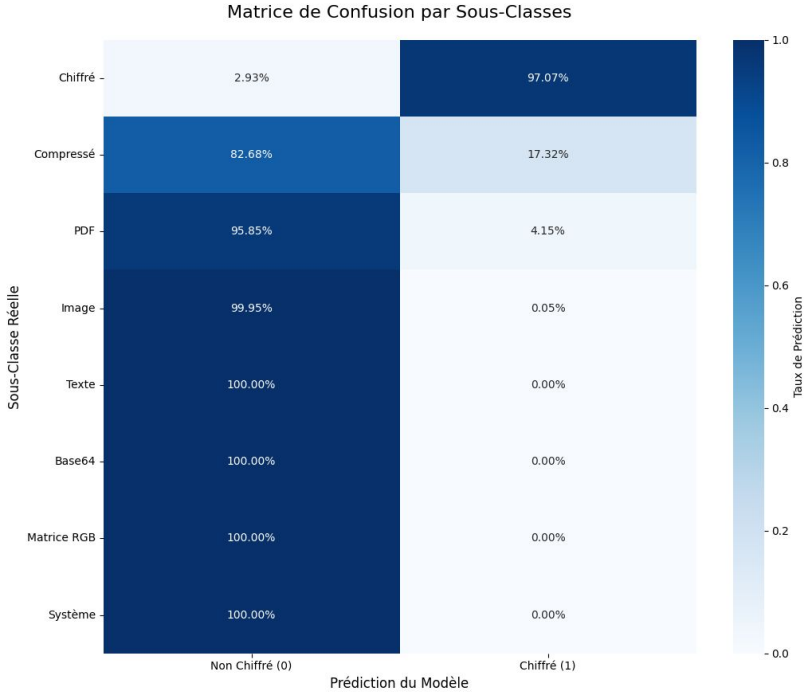
3 époques de fine-tuning



Résultats fine-tuning



Résultats fine-tuning



Résultats Globaux

Scores finaux du modèle

	Score	
Métrique finale	sur le chiffré (0)	sur les non chiffrés (1)
Precision	97.06%	96.94%
Recall	96.93%	97.07%
F1-Score	97.00%	97.00%

Les limites

- L’entraînement à été réalisé sur des “petits” dumps en raison d’une puissance de calcul limitée
- Les dumps étant synthétiques une transposition à la réalité n’est pas garantie

Perspectives

- Une extension du volume de donnée sur l'entraînement
- Un passage à l’apprentissage non supervisé
- RAID comme base pour un outil de file carving “intelligent”