

RÉPUBLIQUE DE CÔTE D'IVOIRE



UNION - DISCIPLINE - TRAVAIL

MINISTÈRE DU PLAN ET DU
DÉVELOPPEMENT



ÉCOLE NATIONALE SUPÉRIEURE DE
STATISTIQUE ET D'ÉCONOMIE
APPLIQUÉE

PROJET SECURITE DES DONNEES

BIBLIOTHEQUE PERMETTANT L'ANONYMISATION DES DONNEES

Présenté par :



TRAORÉ Malick



RANDRIAMIALISOA Herizo



ZONGO Kouamé William

JUIN 2022

Sommaire

Sommaire	2
I. Introduction.....	3
II. Description des librairies	3
III. Conclusion.....	5

I. Introduction

La protection des données personnelles est devenue une préoccupation majeure dans notre société actuelle, où les informations sensibles sont régulièrement partagées et stockées numériquement. Face à l'augmentation des risques de violation de données et de l'importance croissante de la confidentialité, l'anonymisation des données est devenue une méthode essentielle pour assurer la sécurité des informations tout en permettant leur utilisation. Dans le cadre de ce projet, nous avons développé une bibliothèque Python dédiée à l'anonymisation des données. Cette bibliothèque vise à fournir des outils efficaces et simples à utiliser pour anonymiser des fichiers .txt, des paragraphes, des phrases, ou même des mots, en appliquant diverses méthodes d'anonymisation étudiées en classe ainsi que des techniques supplémentaires.

II. Description des librairies

Notre projet est structuré de manière modulaire pour faciliter l'extension et la maintenance. Le dossier principal, intitulé `anonymisation`, contient plusieurs fichiers Python, chacun dédié à une méthode spécifique d'anonymisation¹. Voici une description des principales librairies et de leurs fonctionnalités :

1. *cryptography.py* :

- ✚ Fonctionnalité : Contient les fonctions nécessaires pour chiffrer les données sensibles à l'aide de méthodes cryptographiques robustes.
- ✚ Usage : Assure une protection avancée des données, nécessitant une clé de déchiffrement pour accéder aux informations originales.

2. *deletion.py* :

- ✚ Fonctionnalité : Réalise la suppression des données, en supprimant totalement les informations sensibles des fichiers ou des textes.

¹ https://github.com/malicktraore/Projet_securite_donnees_ISE

- ✚ Usage : Méthode radicale mais efficace pour éliminer toute trace d'information sensible.

3. *generalization.py* :

- ✚ Fonctionnalité : Applique la généralisation, qui remplace des valeurs spécifiques par des valeurs plus générales ou des catégories, réduisant ainsi la précision de l'information.
- ✚ Usage : Utilisée principalement pour anonymiser des données numériques ou catégorielles, telles que des âges ou des codes postaux.

4. *hashing.py* :

- ✚ Fonctionnalité : Implémente le hachage, qui transforme les informations sensibles en une chaîne de caractères fixe, généralement irréversible.
- ✚ Usage : Utile pour protéger les mots de passe ou d'autres données qui n'ont pas besoin d'être reconverties en leur forme originale.

5. *masking.py* :

- ✚ Fonctionnalité : Implémente la méthode de masquage, qui consiste à substituer les informations sensibles par des caractères ou des symboles sans signification.
- ✚ Usage : Cette méthode est utile pour anonymiser des chaînes de caractères tout en conservant leur longueur et leur format.

6. *pseudonymisation.py* :

- ✚ Fonctionnalité : Met en œuvre la pseudonymisation, remplaçant les informations identifiables par des pseudonymes ou des identifiants non liés directement à l'individu.
- ✚ Usage : Conserve la possibilité de réidentifier les données sous certaines conditions, offrant un équilibre entre anonymisation et utilité.

7. *randomization.py* :

- ✚ Fonctionnalité : Ajoute du bruit ou des valeurs aléatoires aux données sensibles, en modifiant légèrement les informations pour masquer leur valeur réelle.

- ✚ Usage : Particulièrement utile pour les données numériques où une légère altération n'affecte pas significativement l'analyse globale.

8. *substitution.py* :

- ✚ Fonctionnalité : Réalise la substitution, en remplaçant les informations sensibles par d'autres valeurs valides mais fausses.
- ✚ Usage : Méthode efficace pour anonymiser des données textuelles tout en conservant un aspect réaliste.

III. Conclusion

La bibliothèque d'anonymisation que nous avons développée offre une gamme complète de méthodes pour protéger les données personnelles tout en permettant leur utilisation sécurisée. En organisant notre code de manière modulaire, nous avons créé un outil flexible et extensible qui peut être facilement intégré dans divers projets nécessitant la gestion de données sensibles. Ce projet répond non seulement aux exigences académiques mais fournit également une ressource précieuse pour les professionnels soucieux de la confidentialité des données. La prochaine étape consistera à tester et à améliorer continuellement cette bibliothèque en fonction des retours des utilisateurs et des évolutions des meilleures pratiques en matière de sécurité des données.