



TapID

PROJET KEYMATCH



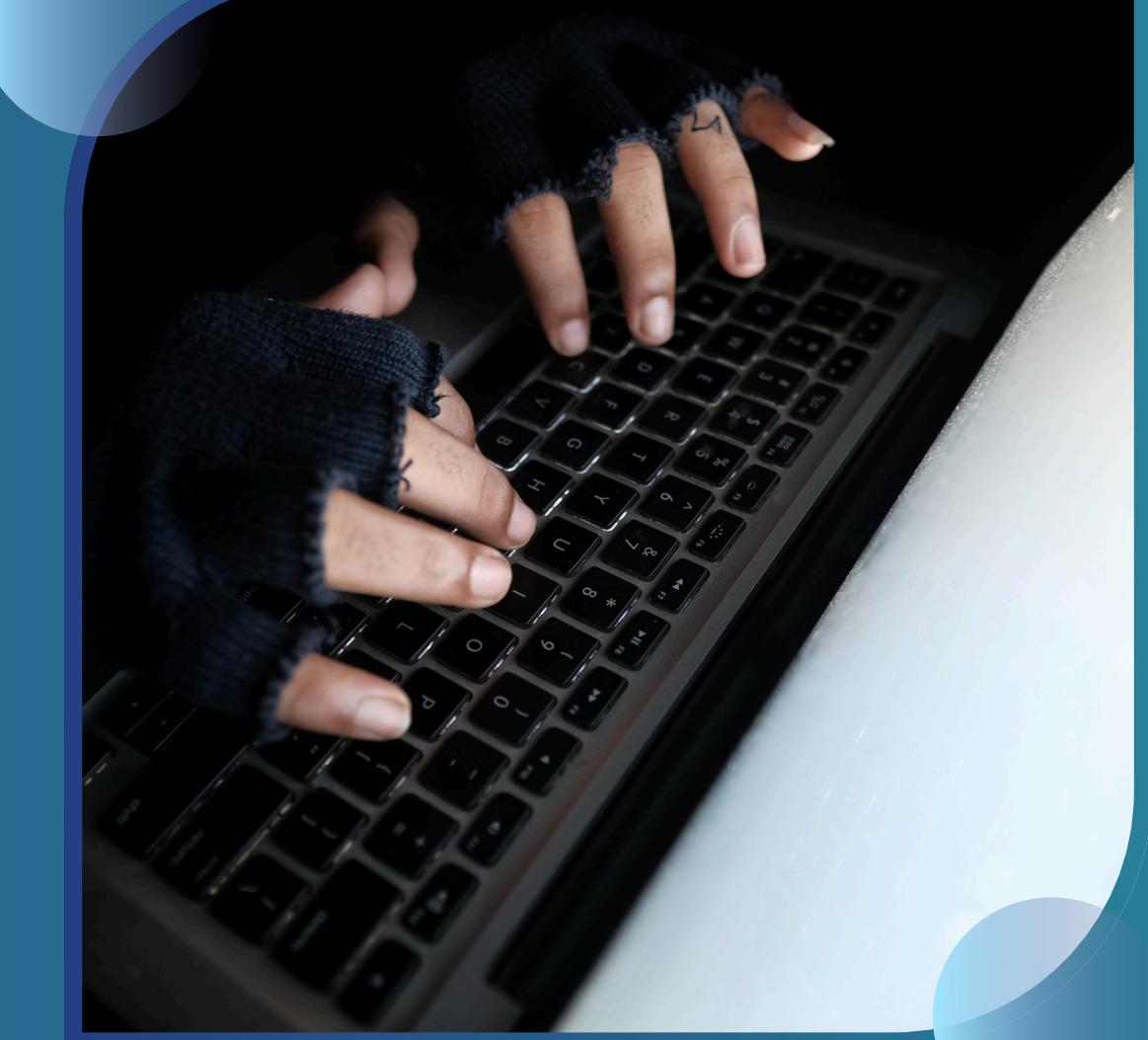


TapID

TAP ID



Fondée en 2021 avec pour mission de redéfinir la sécurité numérique, TapID utilise les dynamiques de frappe au clavier comme une empreinte unique permettant d'authentifier les utilisateurs sans friction. En 2024, TapID est mandatée par un grand groupe alsacien pour développer une API visant à renforcer les mécanismes d'authentification internes.





CYBER IMPORTANCE



La cybersécurité est un enjeu critique dans un monde où la numérisation des services, la croissance des données sensibles, et l'interconnectivité augmentent de façon exponentielle. Cependant, malgré des avancées technologiques, les menaces évoluent également, posant des défis importants aux individus, entreprises et gouvernements.





Augmentation des attaques

Les ransomwares, attaques par phishing, et les vols de données continuent de croître. 81 % des violations de données sont dues à des mots de passe faibles.

Menaces internes

Les employés négligents ou mal intentionnés sont responsables d'environ 30 % des violations, soulignant un manque d'éducation en cybersécurité.



LES MENACES





TapID

LES TENDANCES



Passage à la MFA

Les entreprises adoptent des approches d'authentification multifactorielle.

Biométrie et IA

La biométrie comportementale est de plus en plus reconnue comme une solution robuste.

Zero Trust Security

Le modèle gagne en popularité, exigeant une vérification stricte et continue de chaque utilisateur ou appareil cherchant à accéder aux systèmes.





TapID

IA & BIOMÉTRIE



La biométrie comportementale est une branche émergente des technologies biométriques qui exploite les habitudes et comportements humains pour authentifier des utilisateurs ou détecter des anomalies.

Contrairement à la biométrie traditionnelle (empreintes digitales, reconnaissance faciale), elle analyse des actions spécifiques comme la manière de taper sur un clavier.



LES AVANTAGES



Réduction des failles

Les mots de passe sont vulnérables au piratage. La biométrie comportementale ajoute une couche de sécurité sans dépendre de la mémoire humaine.



Confort utilisateur

Contrairement à la biométrie traditionnelle (empreintes digitales ou reconnaissance faciale), la dynamique de frappe ne nécessite aucun matériel supplémentaire.



Protection

Des outils comme KeyMatch offrent une protection unique en détectant des schémas de frappe incompatibles avec les habitudes enregistrées.



Utilisations multiples

La technologie peut être utilisée dans divers domaines comme les écoles, les banques ou la santé.





LES LIMITES



Variabilité

Le modèle de frappe d'un utilisateur peut varier en fonction de sa fatigue, de son humeur, ou de l'environnement.



Sécurité

Un attaquant très motivé pourrait tenter de reproduire un modèle de frappe en observant ou en enregistrant les frappes d'un utilisateur, compromettant ainsi son profil.



Sensibilité

Les performances peuvent diminuer si l'utilisateur change d'appareil, comme passer d'un clavier mécanique à un clavier tactile ou à un smartphone.



Dépendance

La performance dépend de la qualité et de la diversité des données d'entraînement. L'entraînement et la mise à jour des modèles peuvent devenir coûteux.





TapID

TypingDNA est une entreprise spécialisée dans la biométrie comportementale, utilisant les habitudes de frappe pour authentifier les utilisateurs. Fondée en 2016, son siège est à New York.

L'entreprise propose une technologie d'intelligence artificielle qui analyse les modèles de frappe des utilisateurs sur des claviers, offrant une alternative aux méthodes d'authentification connues.

Leur API repose sur des concepts fondamentaux de machine learning et biométrie comportementale, mais l'implémentation exacte reste confidentielle.

LE CAS TYPING DNA

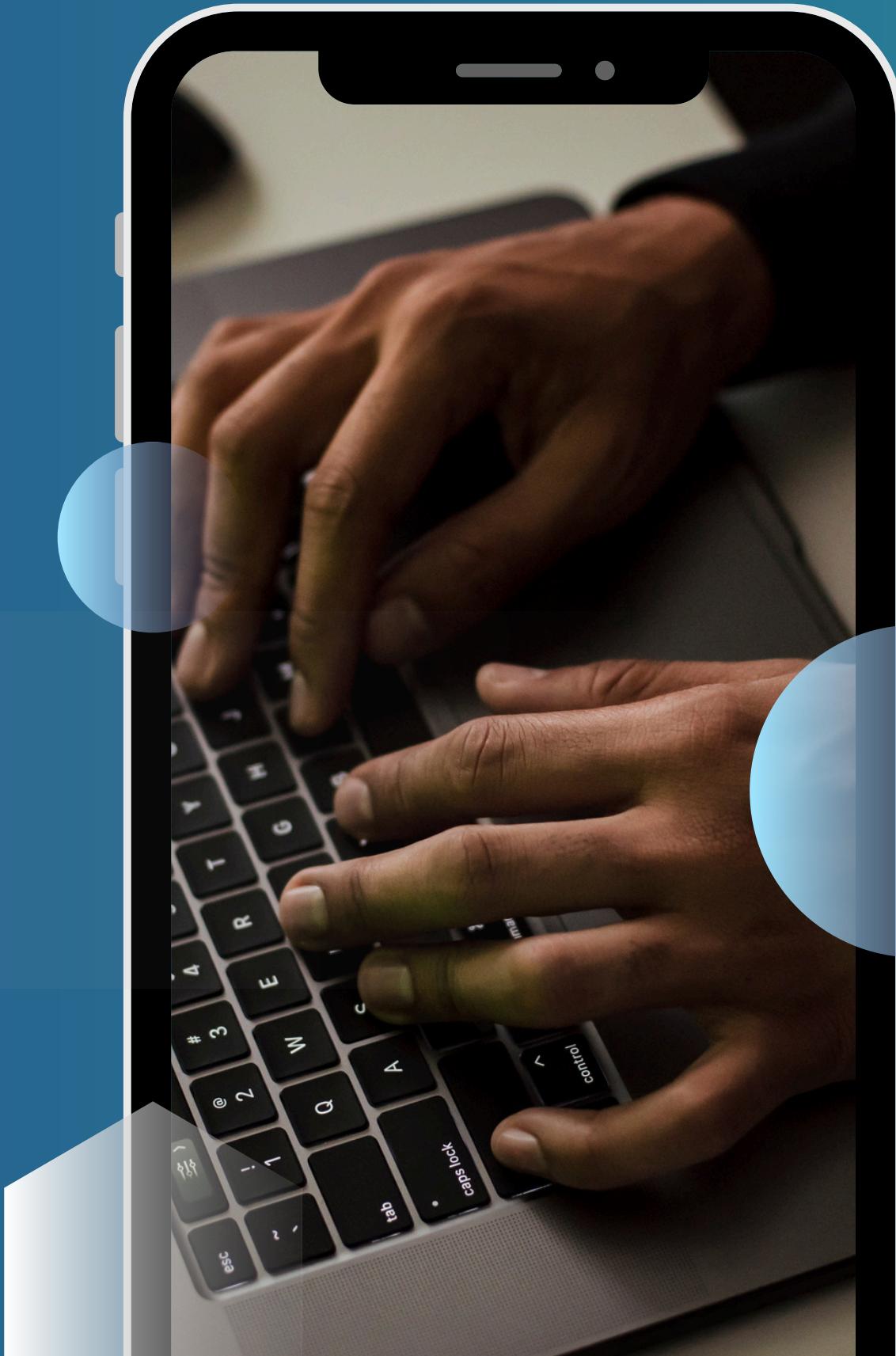




TapID

KEY MATCH

KeyMatch est une API qui analyse les dynamiques de frappe pour authentifier les utilisateurs en temps réel. En exploitant des algorithmes d'intelligence artificielle et de machine learning, KeyMatch détecte des modèles uniques dans la façon dont chaque individu tape sur un clavier.





TapID

11/36

Benchmark Data Set

Ce dataset, publié par le Carnegie Mellon CyLab, est un benchmark dédié à l'étude des dynamiques de frappe (keystroke dynamics).

Description

Les données consistent en des informations sur le timing des frappes provenant de 51 sujets (dactylographes), chacun tapant un mot de passe (.tie5Roanl) 400 fois en 8 sessions.

KEYSTROKE DYNAMICS



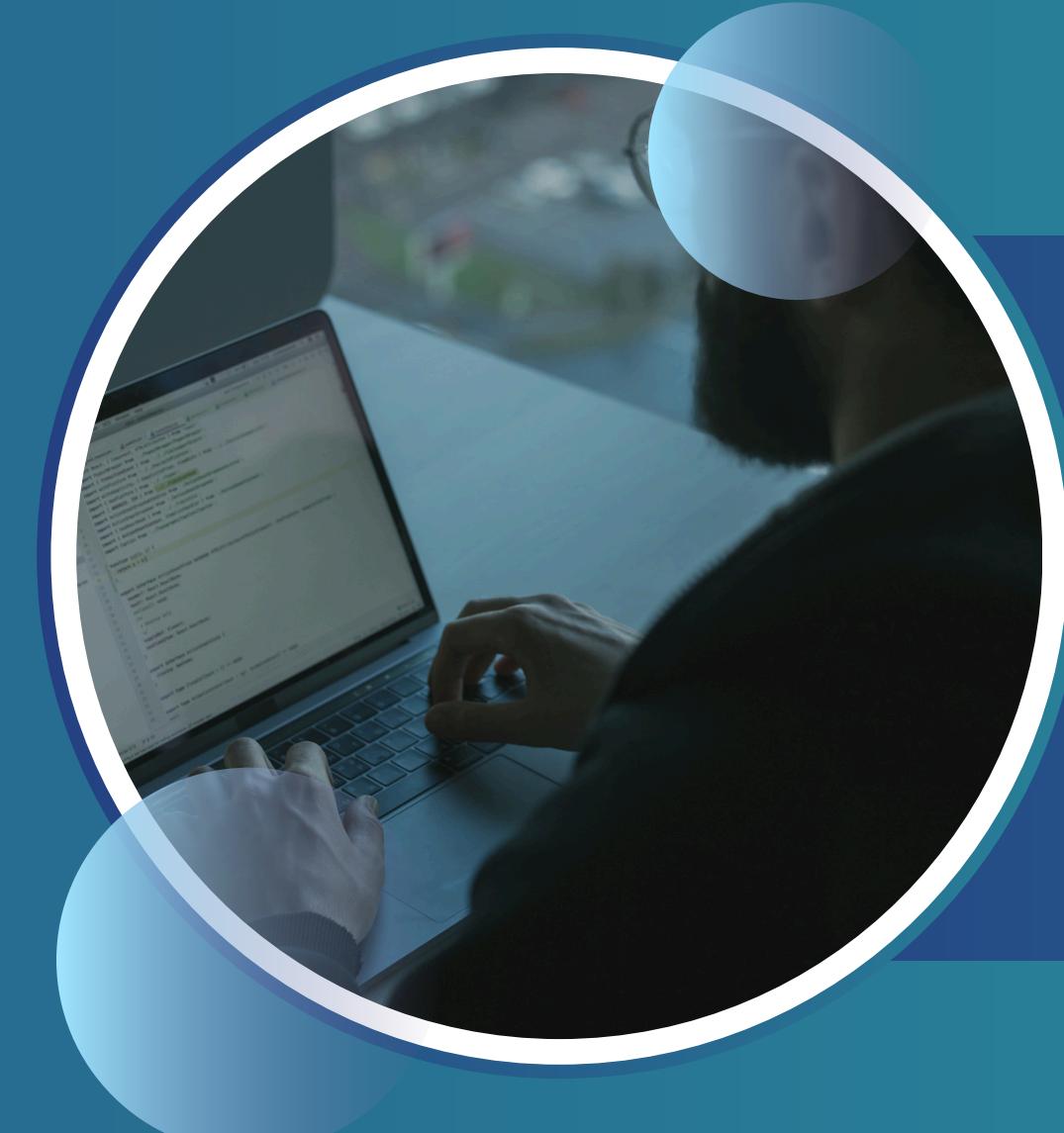


TapID

ETAT DE L'ART

75% de la communauté scientifique estime qu'il est possible de reconnaître un utilisateur grâce à son pattern de frappe au clavier.

Il existe des recherches sur les patterns de frappe sur texte fixe (notre cas) et texte libre. Pour les textes fixes, le meilleur modèle atteint 98.8% de précision.





LES VARIABLES

01

Colonnes H.key

Désignent un temps de maintien pour la touche nommée (c'est-à-dire le temps écoulé entre le moment où la touche a été enfoncée et le moment où elle a été relâchée).

02

Colonnes DD.key1.key2

Désignent un temps d'appui-touche pour le digraphe nommé (c'est-à-dire le temps entre le moment où la touche 1 a été enfoncée et le moment où la touche 2 a été enfoncée).

03

Colonnes UD.key1.key2

Désignent un temps de relâchement de touche pour le digraphe nommé (c'est-à-dire le temps entre le moment où la touche 1 a été relâchée et le moment où la touche 2 a été enfoncée).



.tie5Roanl



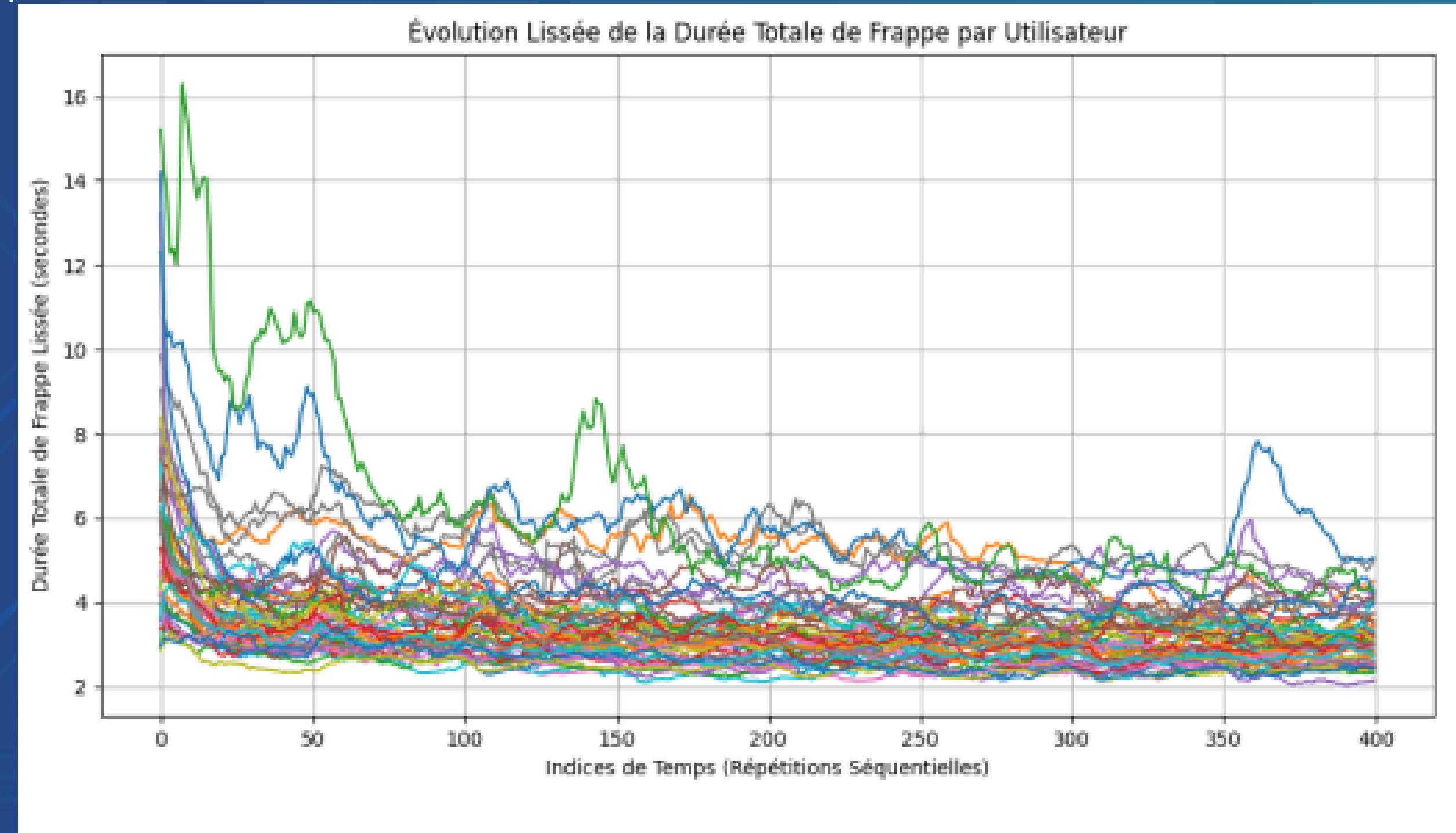
TapID

1	subject	sessionIndex	rep	H.period	DD.period.t	UD.period.t	H.t	DD.t.I	UD.t.I	H.I	DD.i.e	UD.i.e	H.e	DD.e.five	UD.e.five	H.five	DD.five.Shift.r	UD.five.Shift.r	H.Shift.r	DD.Shift.r.o	UD.Shift.r.o	H.o	DD.o.a	UD.o.a	H.a	DD.a.n	UD.a.n	H.n	DD.n.I	UD.n.I	H.I	DD.I.Return	UD.I.Return	H.Return
20400																																		



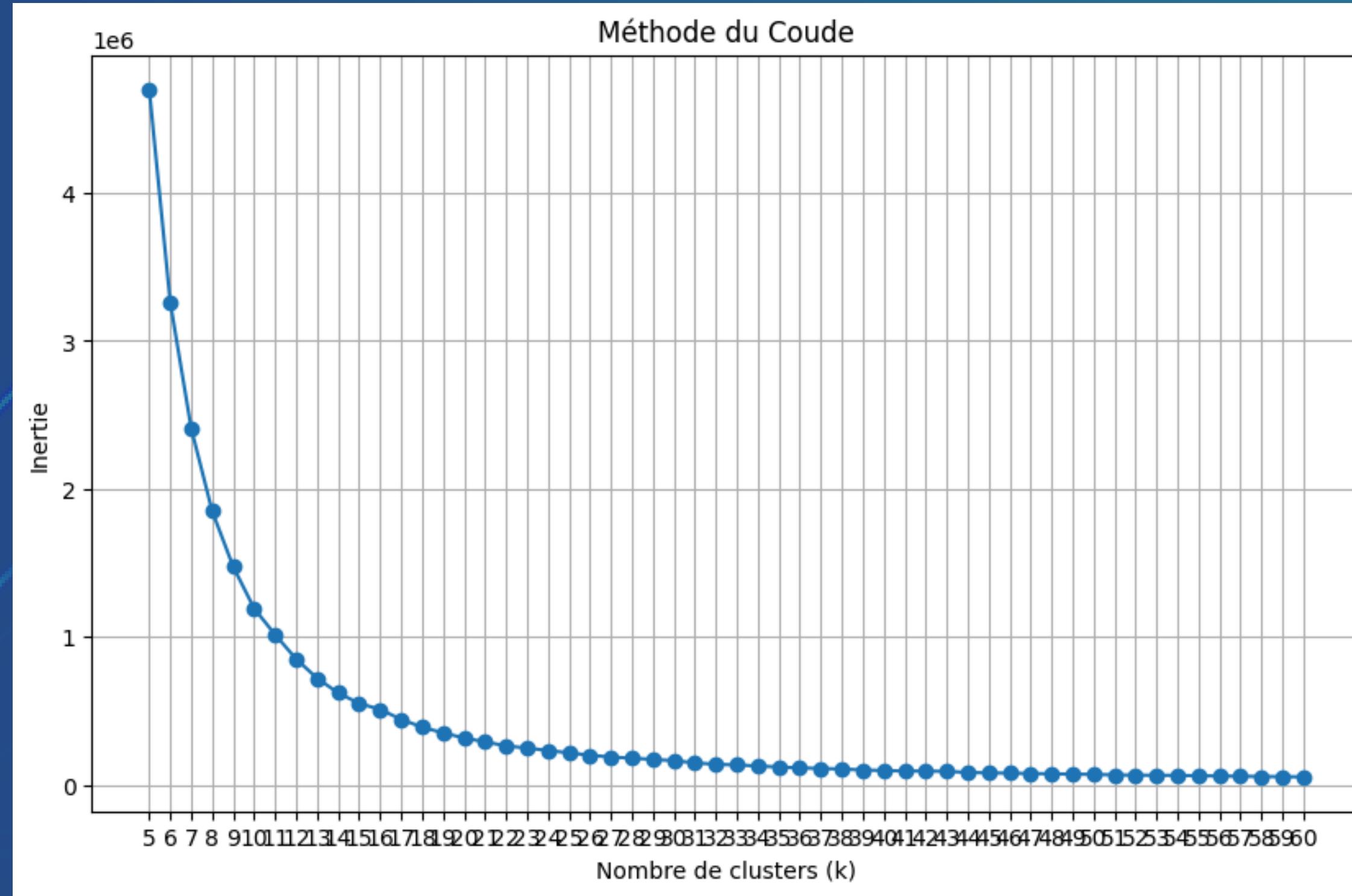


TapID





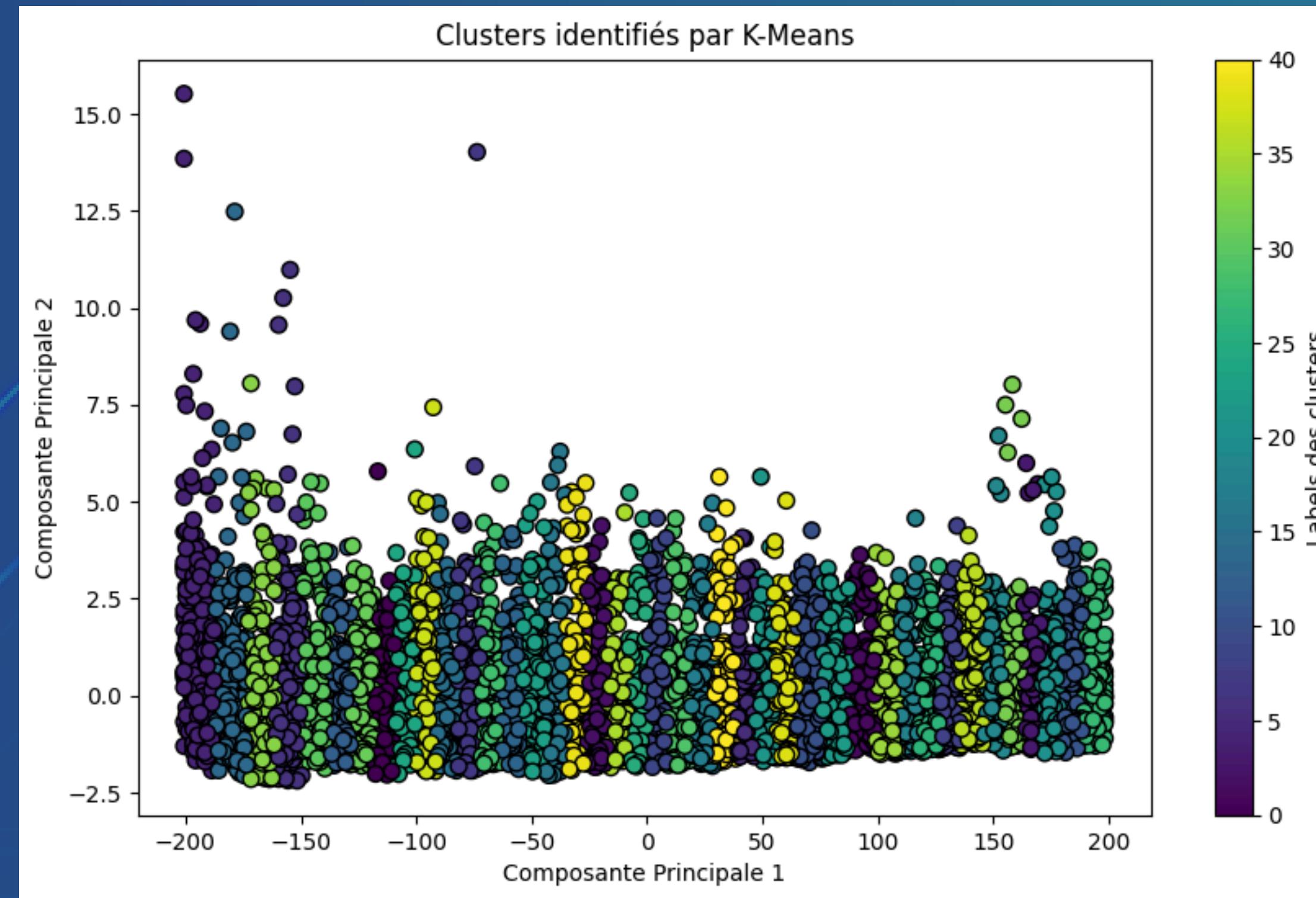
TapID





TapID

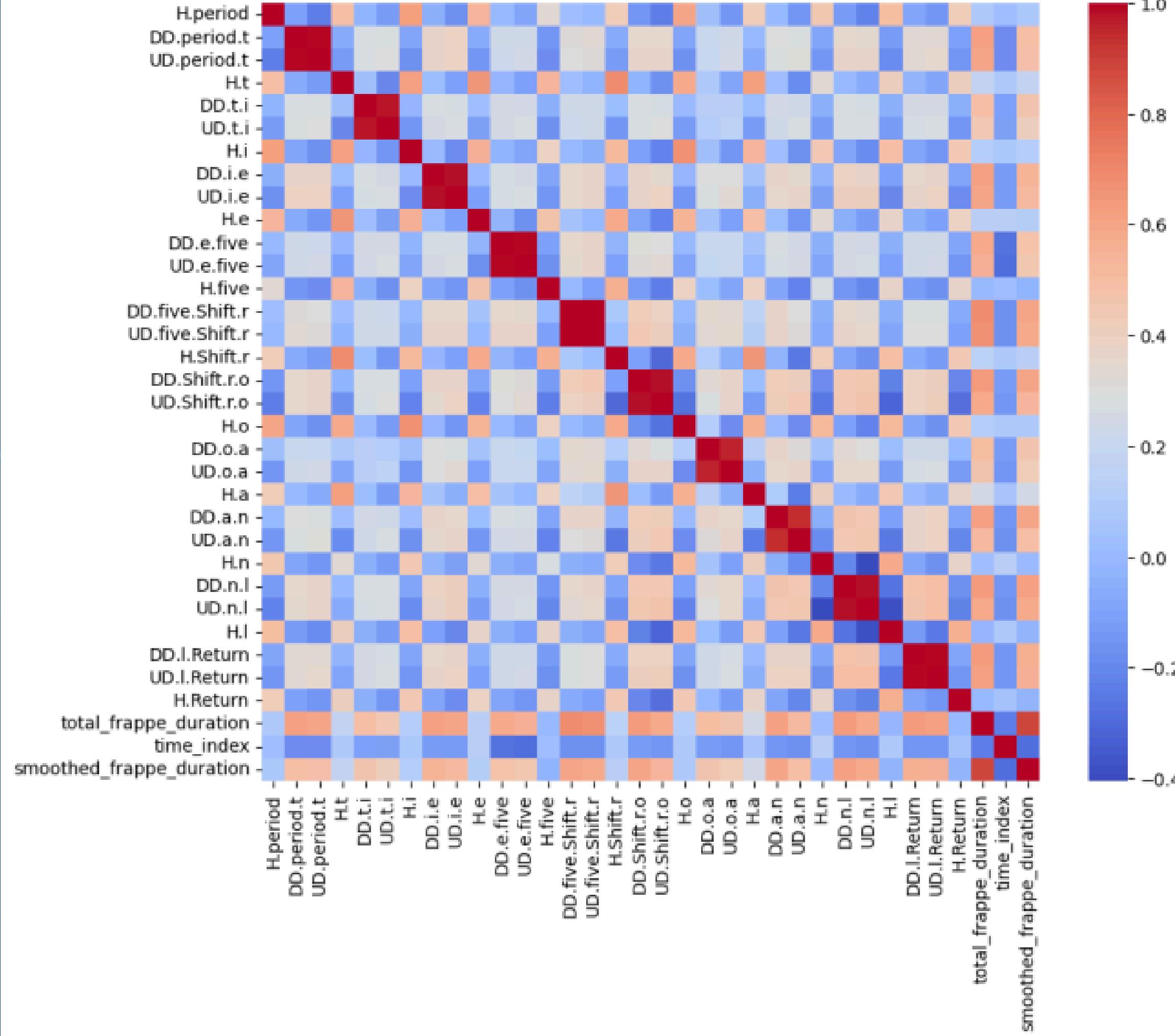
18/36





TapID

Matrice de Corrélation sur les données normalisées



15/36





TapID

19/36

ARCHITECTURE GÉNÉRALE

[Utilisateur]



[Interface Tkinter] --> [API FastAPI] --> [Modèle Machine Learning]



[Base de données ou fichiers]





TapID

COMPOSANTS DE L'API

[API FastAPI]

|--- /verify_capture

|
v

[Modèle Machine Learning]

|--- /record_password

|
v

[Base de données ou fichiers]

|--- /train_model_full_dataset

|
v

[Modèle Machine Learning] <-->
[Base de données ou fichiers]

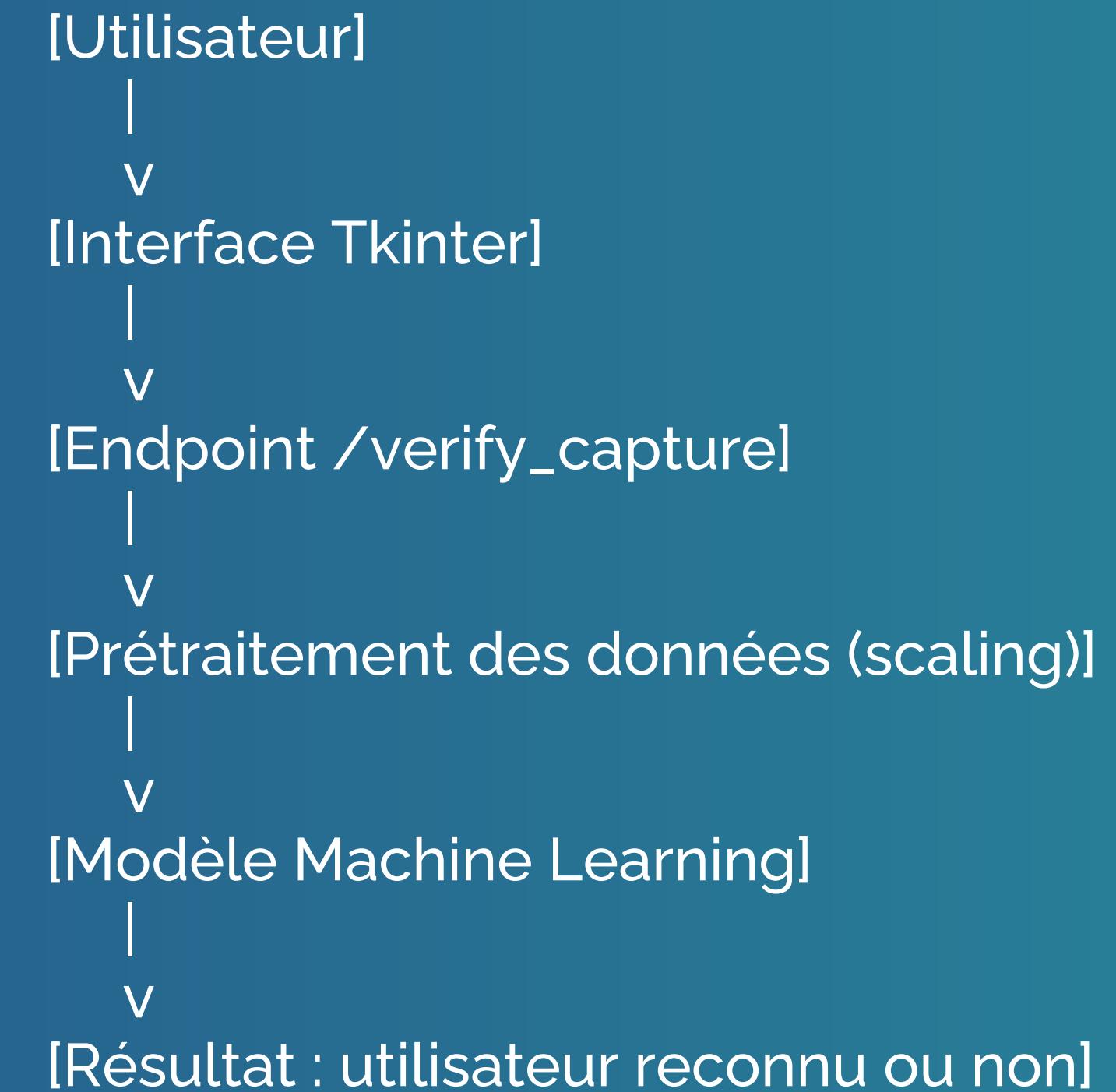




TapID

21/36

FLUX DE DONNÉES WORKFLOW

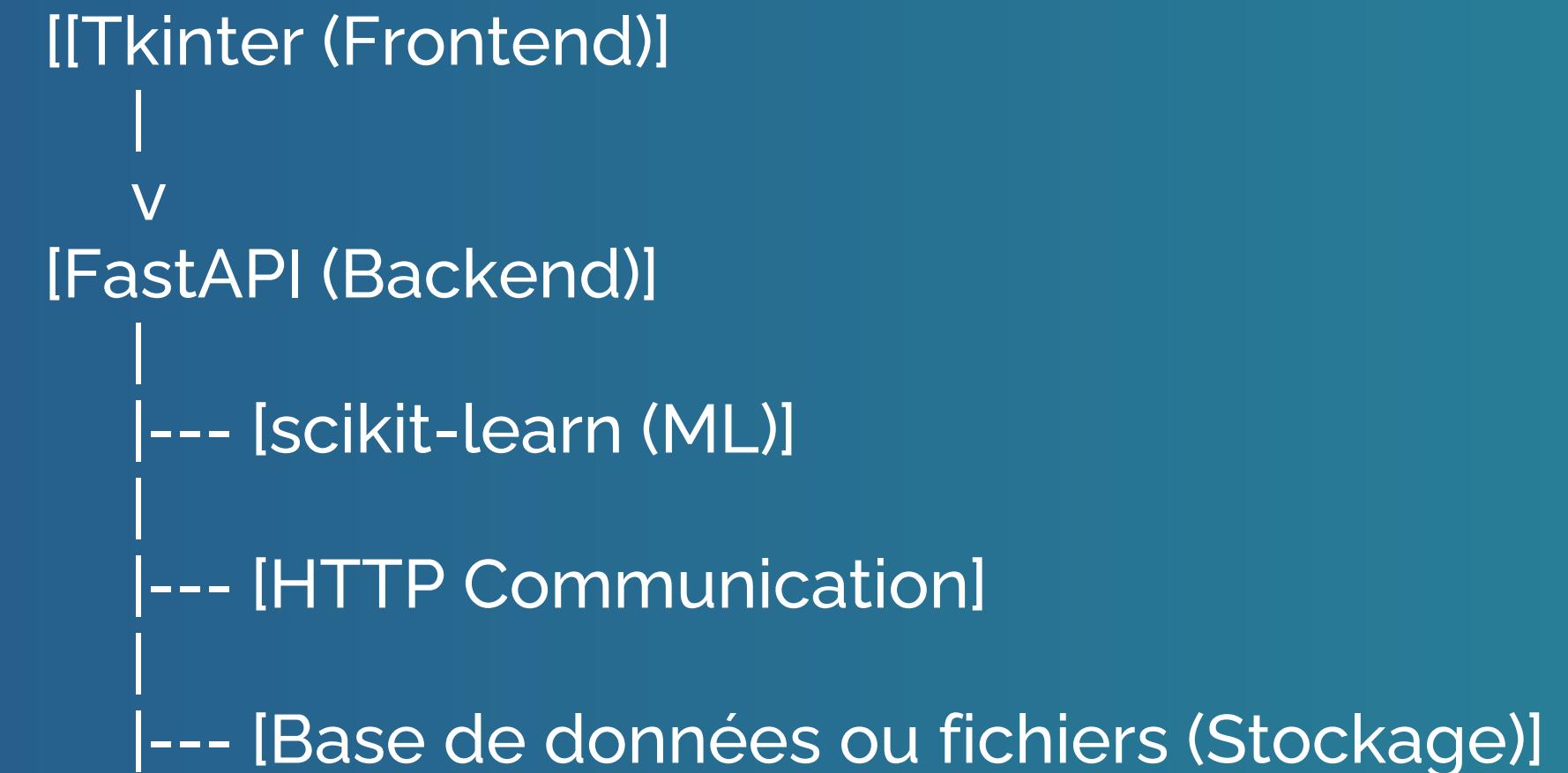




TapID

22/36

ARCHITECTURE TECHNIQUE





TapID

23/36

MODÈLE ML

[Entraînement du modèle]

- |--- Chargement des données
- |--- Prétraitement des captures clavier (scaling)
- |--- Entraînement (SVM ou autre algorithme)

[Prédiction avec le modèle]

- |--- Prétraitement des nouvelles données
- |--- Comparaison avec les classes existantes





CAS D'USAGE

1. **Vérification renforcée d'un mot de passe**
 - Données utilisateur envoyées depuis Tkinter
 - API FastAPI (endpoint /verify_capture)
 - Résultat retourné : utilisateur reconnu ou non
2. **Enregistrement d'un nouvel utilisateur**
 - Données utilisateur envoyées depuis Tkinter
 - API FastAPI (endpoint /record_password)
 - Enregistrement des données dans le stockage
3. **Réentraînement du modèle avec l'ensemble des données**
 - API FastAPI (endpoint /train_model_full_dataset)
 - Chargement des données complètes et réentraînement du modèle
 - Résultat : modèle mis à jour



TapID

25/36



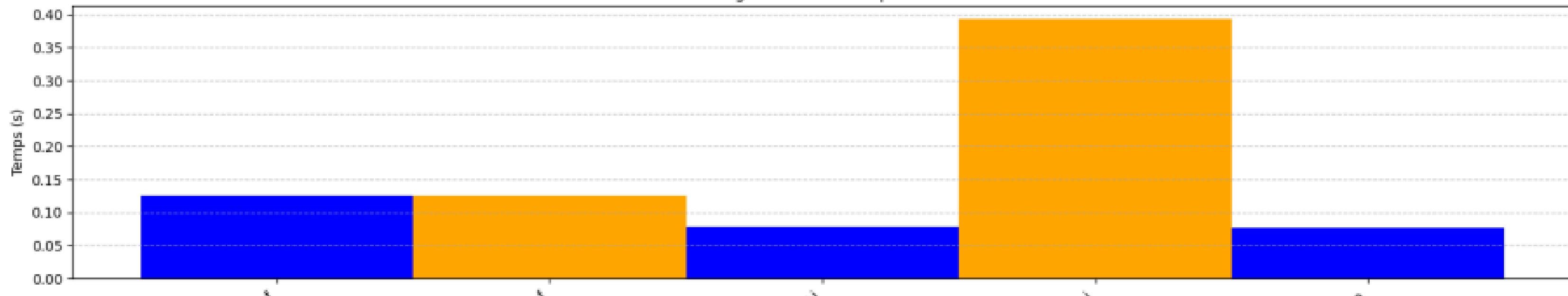
10

10

1

Touche: f - Hold Time H : 0.1263 s + UD : 0.1263, s = DD : 0.2526
Touche: j - Hold Time H : 0.0788 s + UD : 0.3932, s = DD : 0.4720
Touche: o - Hold Time H : 0.0771 s + UD : 0.0000, s = DD : 0.0000

Chronogramme des temps H et U



Chronogramme des temps DD (Cont)

TapID

Reconnnaissance de votre mot de passe par traitement biodynamique.
Veuillez saisir le code suivant puis valider avec la touche RETURN <.*tie5Roenl*>.
Vous avez 10 essais avant réinitialisation.

[Quitter](#)

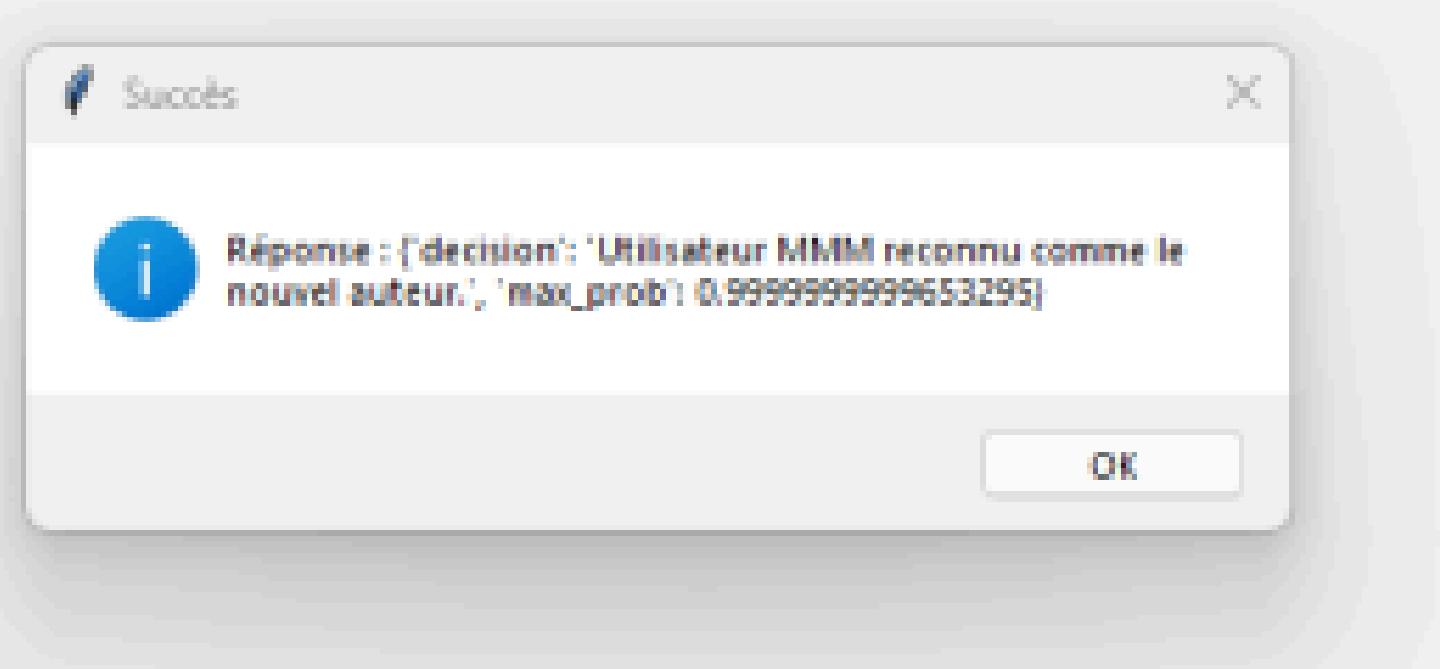
Tentatives restantes : 2

TapID

Reconnnaissance de votre mot de passe par traitement biodynamique.
Veuillez saisir le code suivant puis valider avec la touche RETURN <...Roend>.
Vous avez 10 essais avant réinitialisation.

[Quitter](#)

Tentatives restantes : 2





TapID

29/36

MULTILAYER PERCEPTRON

Modèle de réseau de neurones artificiels capable de capturer des relations complexes dans les données grâce à son architecture multi-couches.



MLPClassifier

Il s'avère être un excellent choix pour notre projet grâce à sa capacité à capturer des relations complexes et sa flexibilité.

Pourquoi ce choix ?

Avec un ajustement optimal des hyperparamètres, il peut offrir une performance robuste et une généralisation satisfaisante.





MLP CLASSIFIER



Capacité

Les données du projet présentent probablement des relations non linéaires. Le MLPC, avec ses couches cachées et ses fonctions d'activation non linéaires, est idéal pour modéliser ces structures.



Adaptabilité

Fonctionne bien avec des données numériques continues, catégoriques encodées ou combinées. Convient pour des projets avec des volumes modérés de données.



Performance

Éprouvé pour résoudre des problèmes de classification multiclass ou binaire avec des performances solides. Permet d'obtenir des résultats compétitifs grâce à des optimisations par rétropropagation.



Flexibilité et Contrôle

Configuration facile des hyperparamètres tels que le nombre de couches cachées ou de neurones par couche. Possibilité de régularisation pour éviter le surapprentissage (*L2 regularization* ou *dropout*).

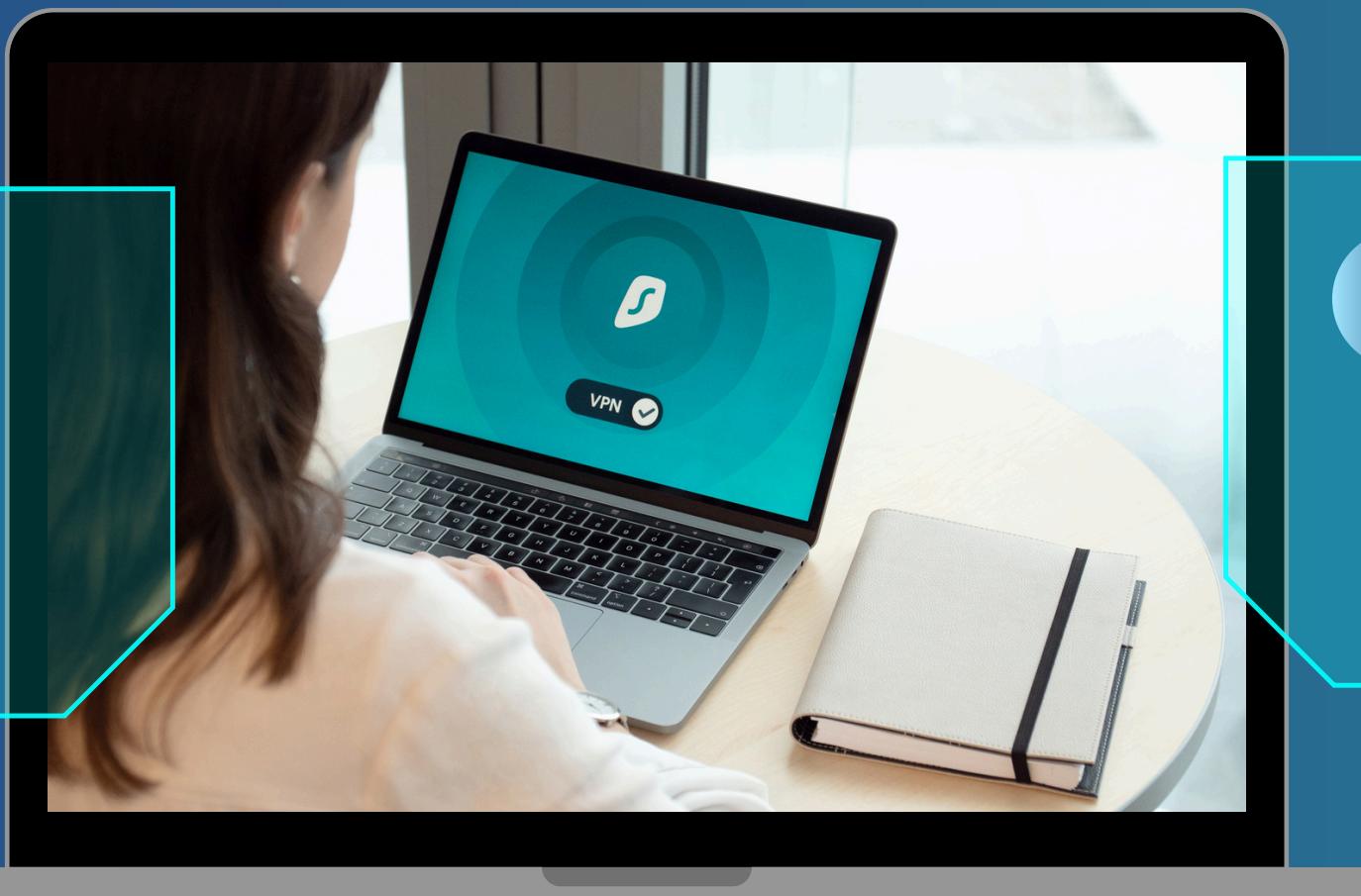




RÉSULTATS PROMETTEURS

Premiers résultats

Les premiers résultats montrent une courbe de perte bien convergente.



Validation croisée

Les scores de validation croisée (moyenne 83.38%) montrent que le modèle généralise bien sur des données non vues.

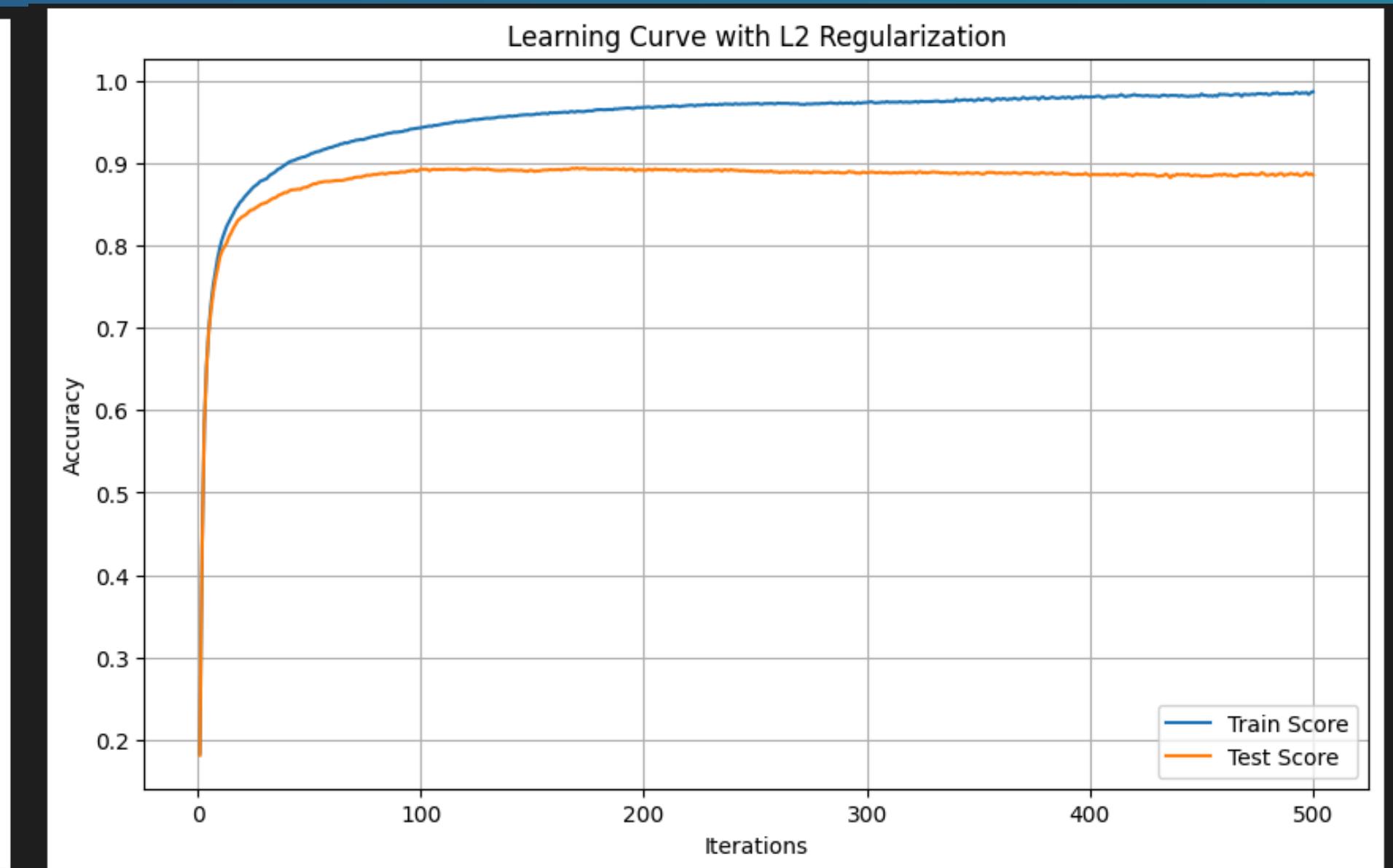
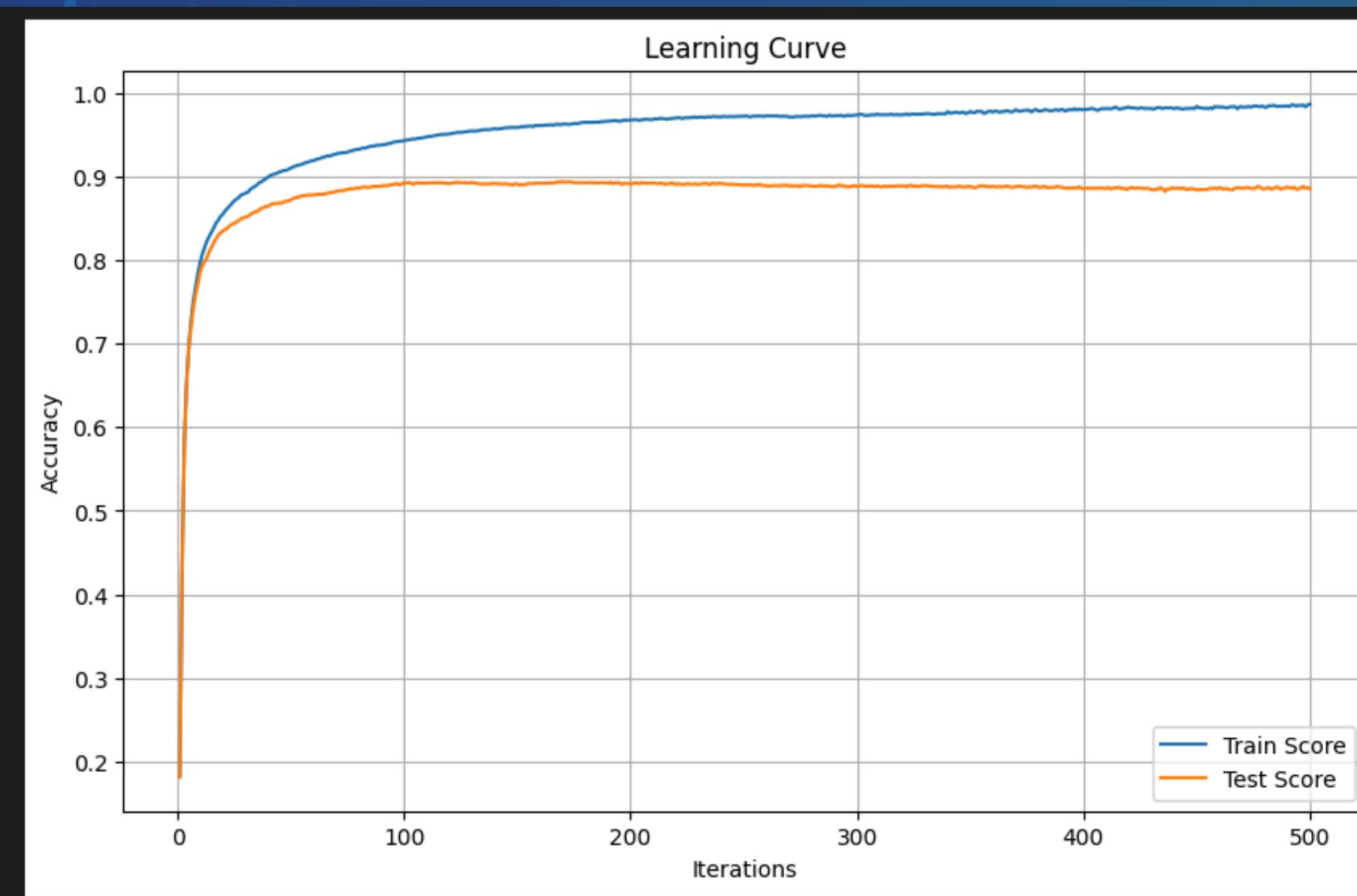




Comparaison avec d'autres modèles

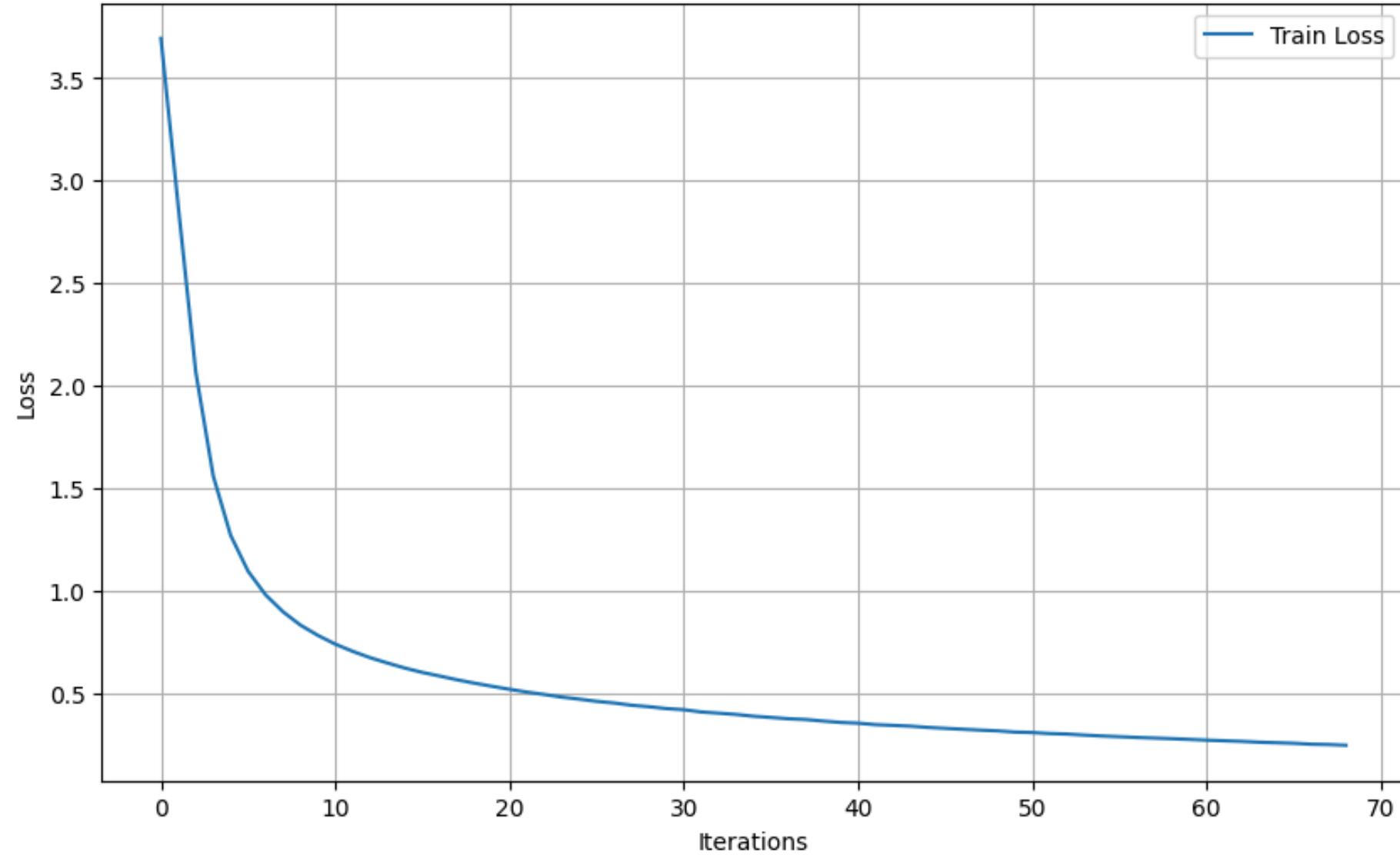
Modèle	Avantages	Inconvénients
MLPClassifier	Non-linéarité, flexibilité, performant	Demande d'ajustement des hyperparamètres
Régression Logistique	Simplicité, rapide à entraîner	Limité aux relations linéaires
Random Forest	Résistant aux outliers, peu sensible à l'échelle	Moins performant sur des relations complexes
SVM	Bonne séparation des classes	Complexité computationnelle sur grands ensembles



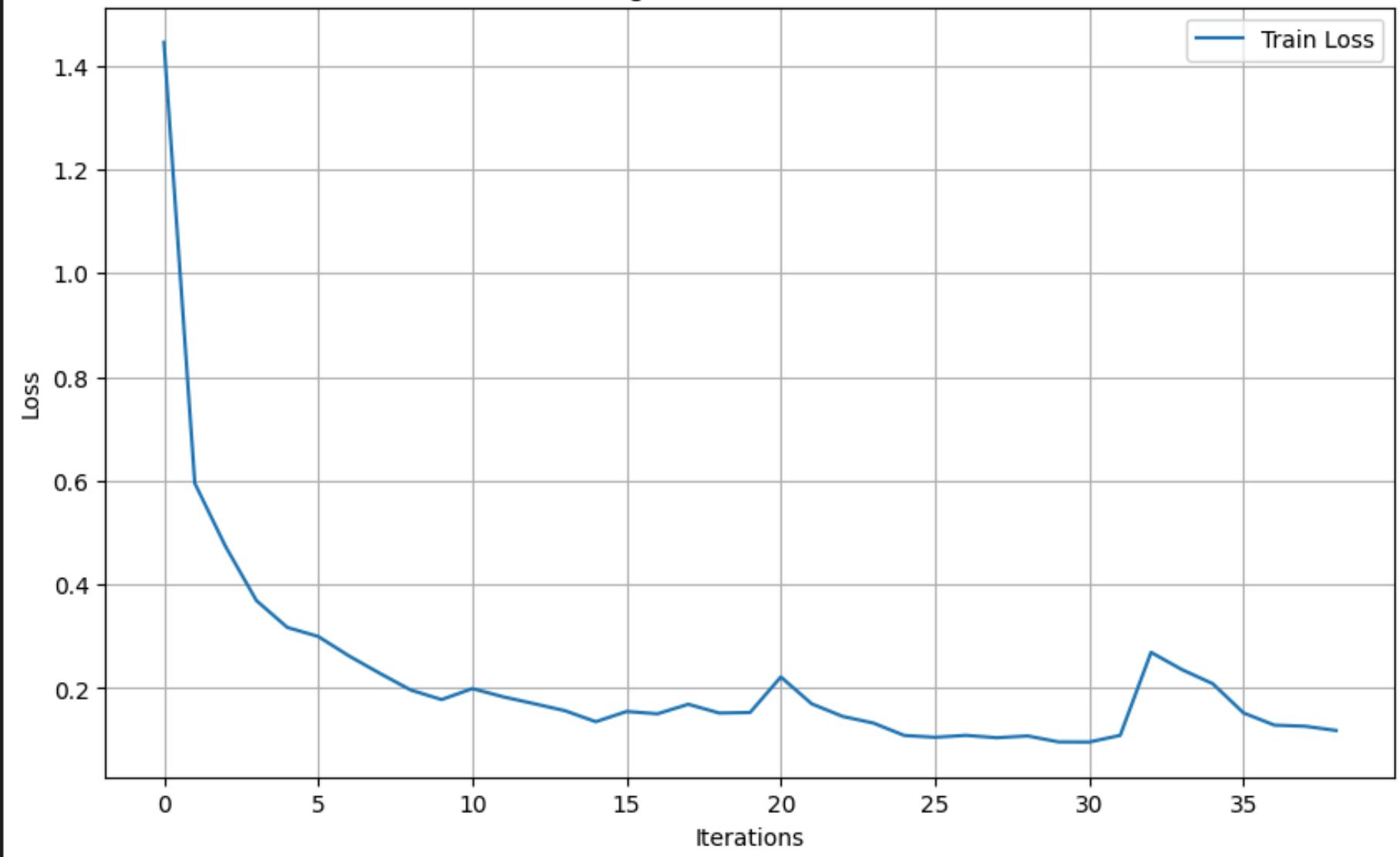




Learning Curve with Early Stopping



Learning Curve with Grid Search

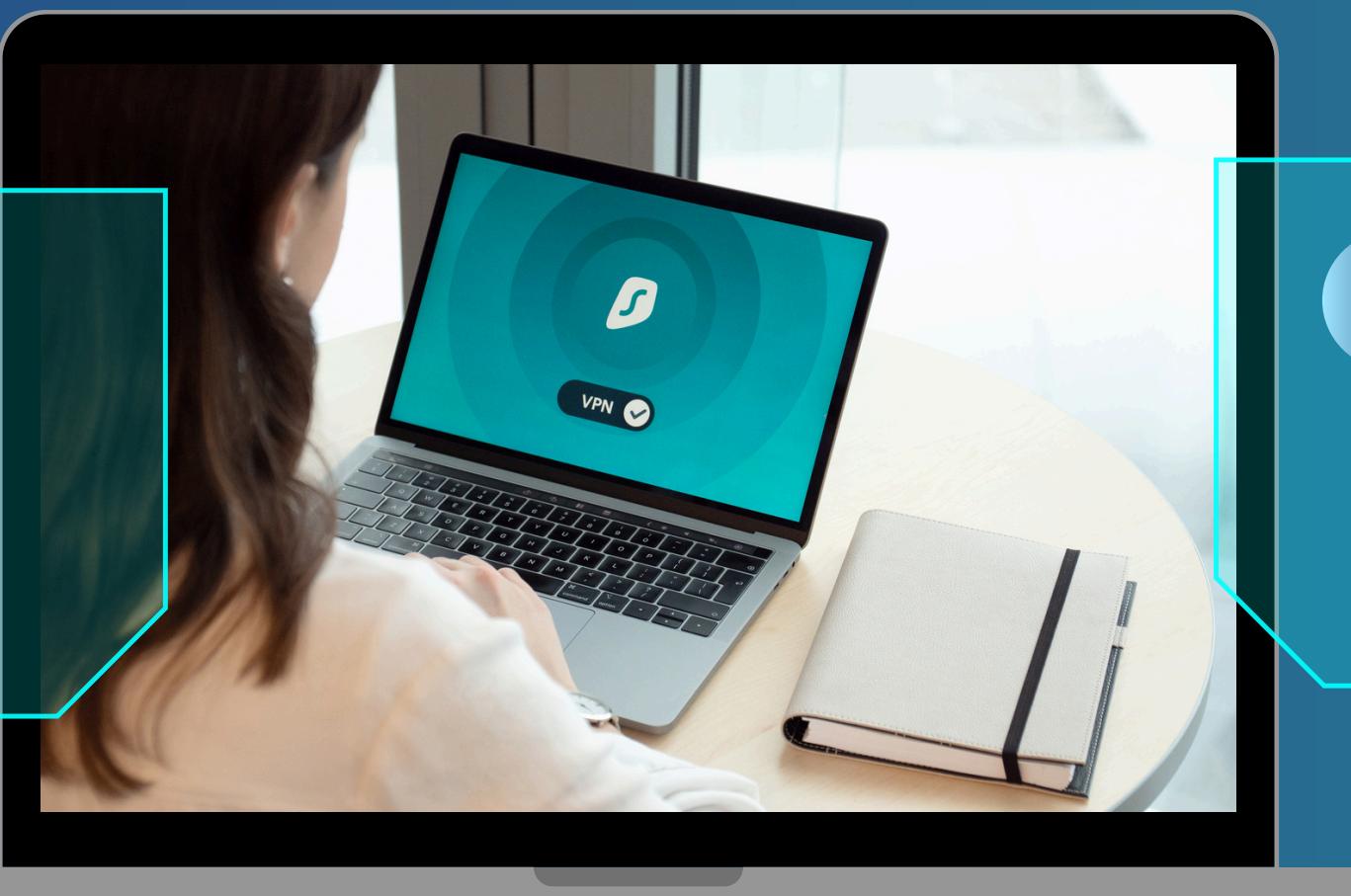




CONCLUSION

Performance

Le MPLClassifier présente une performance équilibrée entre biais et variance.



Généralisation

Les techniques appliquées (régularisation, early stopping, validation croisée) assurent une bonne généralisation sur des données non vues.





Studio Shodwe

36/36

MERCI !

