

# **Rapport TP Machine Learning**

## **Réseau de Neurone**

### **Membre du groupe :**

- Airouche Kafia
- Oulmou Kenza
- Ouamrane Lydia

## Objectif du TP :

L'objectif de ce projet est de développer et d'implémenter un modèle de réseau de neurones artificiels (Artificial Neural Network, ANN) pour prédire la probabilité qu'un client d'une banque décide de quitter (churn) la banque. Ce modèle aidera la banque à identifier les clients à risque de départ, permettant ainsi la mise en œuvre de stratégies proactives pour améliorer la rétention des clients et optimiser les services offerts.

## Contexte :

Dans le secteur bancaire hautement concurrentiel d'aujourd'hui, la fidélisation des clients est cruciale pour la croissance et la stabilité à long terme d'une banque. Un taux de churn élevé peut avoir des implications significatives sur la rentabilité et la réputation d'une institution financière. Par conséquent, comprendre et prédire le comportement de churn des clients devient une priorité stratégique.

## Données :

Le projet s'appuie sur l'ensemble de données Bank\_customers.csv, qui contient des informations détaillées sur les clients d'une banque. Cet ensemble de données inclut plusieurs caractéristiques telles que l'âge du client, le solde du compte, le nombre de produits bancaires utilisés, l'état du crédit, entre autres. L'analyse de ces caractéristiques fournira des insights précieux sur les facteurs influençant la décision d'un client de quitter la banque.

## Méthodologie

Le projet sera développé en suivant une série d'étapes structurées, incluant :

- 1. Exploration et prétraitement des données :** On a analysé l'ensemble des données pour comprendre les caractéristiques disponibles, identifier les valeurs manquantes, et effectuer le nettoyage et la préparation des données nécessaires pour le modèle de prédiction :
  - **Importation des bibliothèques :** On a commencé par importer les bibliothèques nécessaires telles que pandas, keras, sklearn, pour la modélisation et l'évaluation.
  - **Chargement des données :** On a chargé le dataset Bank\_customers.csv dans un DataFrame pandas
  - **Exploration initiale :** On a utilisé les fonctions telles que info(), describe(), et la vérification des valeurs manquantes pour obtenir un aperçu des données, y compris les types de données, les statistiques descriptives, et la propreté des données.
  - **Prétraitement des Données :** Dans le traitement de l'ensemble de données, nous avons éliminé les colonnes non pertinentes pour la prédiction du churn, telles que les identifiants et les noms. Nous avons également converti les variables catégorielles Gender et Geography en formats numériques pour faciliter leur analyse par le modèle de réseau de neurones artificiels, rendant ainsi les données prêtes pour l'entraînement du modèle.
- 2. Sélection des caractéristiques :** Nous avons divisé l'ensemble de données en caractéristiques et cible, puis séparé ces données en ensembles d'entraînement et de test pour préparer le modèle.

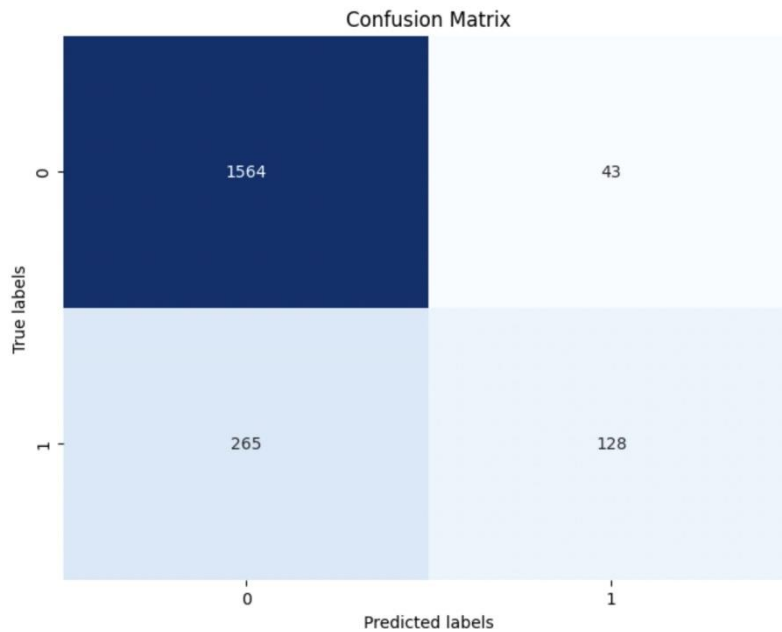
- Division de l'ensemble de données en caractéristiques (X) et cible (y), où la cible est la variableExited.
- Séparation des données en ensembles d'entraînement et de test, avec une répartition de 80% pour l'entraînement et 20% pour le test.
- Standardisation des caractéristiques pour uniformiser leur échelle, facilitant ainsi le processus d'apprentissage du modèle de réseau de neurones.

**3. Construction entraînement et évaluation du modèle ANN :** On a utiliser le modèle de réseau de neurones artificiels pour entraînement et pour prédire des clients de la banque.

- **Construction du modèle :** On a initialisé un modèle séquentiel, suivi de l'ajout de deux couches cachées avec 6 unités chacune, utilisant la fonction d'activation 'relu'. Une couche de sortie avec 1 unité a été ajoutée pour la prédiction finale, utilisant 'sigmoid' comme fonction d'activation pour la classification binaire.
- **Compilation et entraînement :** Le modèle a été compilé avec l'optimiseur 'adam' et la fonction de perte 'binary\_crossentropy', ciblant une métrique d'exactitude ('accuracy'). Il a été entraîné sur l'ensemble d'entraînement avec des batches de taille 10 pendant 100 epochs, optimisant la performance du modèle pour la prédiction du churn.
- **Évaluation et prédiction :** Le modèle a été évalué sur l'ensemble de test pour obtenir la perte de test et l'exactitude, fournissant une mesure de la performance du modèle sur des données non vues. Des prédictions ont été réalisées sur l'ensemble de test, et les classes prédites ont été dérivées en appliquant un seuil de 0.5, classifiant les clients en 'quitter' ou 'rester'.

**4. Optimisation et validation :** Après l'entraînement et l'évaluation du modèle de réseau de neurones pour prédire le churn des clients de la banque, nous avons procédé à l'analyse détaillée des performances du modèle à travers la matrice de confusion et le rapport de classification, suivi de la visualisation de la matrice de confusion

- **Matrice de confusion :** La matrice de confusion a été générée pour visualiser les performances du modèle en termes de prédictions correctes et incorrectes, elle indique que le modèle est relativement bon pour prédire les clients qui ne quitteront pas la banque mais pourrait être amélioré pour identifier de manière plus précise ceux qui sont susceptibles de partir, comme le suggère le nombre plus élevé de faux négatifs.
  - ❖ **Vrais Négatifs (VN) :** 1564 clients ont été correctement prédits comme ne quittant pas la banque.
  - ❖ **Faux Positifs (FP) :** 43 clients ont été incorrectement prédits comme quittant la banque alors qu'ils sont restés.
  - ❖ **Faux Négatifs (FN) :** 265 clients ont été incorrectement prédits comme restant alors qu'ils ont effectivement quitté la banque.
  - ❖ **Vrais Positifs (VP) :** 128 clients ont été correctement prédits comme quittant la banque.



- **Rapport de classification :** Le rapport montre que le modèle est performant pour prédire les clients qui restent mais moins pour identifier ceux qui partent, comme le montre le faible rappel pour la classe 1. Cela pourrait indiquer un besoin d'améliorer la capacité du modèle à détecter les churns sans sacrifier sa capacité à reconnaître les non-churns.
  - **Pour la classe 0 (clients ne quittant pas la banque) :**
    - **Précision :** 0.86, indiquant que 86% des prédictions de non-churn sont correctes.
    - **Rappel :** 0.97, signifiant que le modèle a correctement identifié 97% des clients réels ne quittant pas la banque.
    - **Score F1 :** 0.91, une moyenne harmonique entre précision et rappel, suggérant un bon équilibre pour la classe 0.
    - **Support :** 1607, le nombre de vrais échantillons de la classe 0 dans le jeu de test.
  - **Pour la classe 1 (clients quittant la banque) :**
    - **Précision :** 0.75, ce qui signifie que 75% des prédictions de churn sont correctes.
    - **Rappel :** 0.33, indiquant que seulement 33% des clients réels quittant la banque ont été correctement identifiés par le modèle.
    - **Score F1 :** 0.45, qui est relativement faible, reflétant le déséquilibre entre précision et rappel pour la classe 1.
    - **Support :** 393, le nombre de vrais échantillons de la classe 1 dans le jeu de test.

