

 <p>Institut Supérieur de Statistique d'Econométrie et de Data Science</p>	Année académique 2024-2025	Chargé de Cours AKPOSSO Didier Martial
	MINI – PROJET DE SYNTHESE MACHINE LEARNING & DEEP LEARNING	Document Etudiant 1/1
Mini – Projet	Durée 1 MOIS	CYCLE Licence – Ingénieur – Master

Ce Mini-projet est relatif à la modélisation prédictive par algorithmes d'Intelligence Artificielle de Machine learning et Deep Learning

NB : Il doit être soigneusement traité, rédigé sous WORD puis converti en PDF et transmis à l'enseignant à l'adresse suivant : **insseds@gmail.com** au plus tard le **05 juillet 2025 à 23H : 59mn.** (Cette date est à respecter scrupuleusement)

À propos de l'ensemble de données tiré de Kaggle

Ensemble de données sur le taux de désabonnement (churn) des clients de la Banque du Botswana

Présentation de l'ensemble de données

Cet ensemble de données synthétiques simule les données clients d'une banque fictive au Botswana. Il est spécialement conçu pour modéliser le comportement de désabonnement. Il comprend un ensemble complet de données démographiques, financières, d'utilisation des produits et d'indicateurs comportementaux susceptibles d'influencer la décision d'un client de quitter la banque. Généré à l'aide de la bibliothèque Python Faker, cet ensemble de données réalistes mais entièrement fictives est utilisé à des fins pédagogiques, de test et de modélisation.

Points forts de l'ensemble de données

Nombre d'enregistrements : 115 640 clients

Taux de désabonnement : déterminé par un score de risque de désabonnement calculé en fonction de plusieurs attributs client

Zone géographique : Botswana

Structure des données : l'ensemble de données est organisé sous forme de tableau, chaque ligne représentant un client unique

Cas d'utilisation

Cet ensemble de données est idéal pour les applications suivantes :

Modélisation de la prévision du taux de désabonnement : création et évaluation de modèles d'apprentissage automatique pour prédire le taux de désabonnement.

Segmentation de la clientèle : analyse des profils clients et segmentation en fonction de diverses données démographiques et financières.

Analyse des produits : identification des produits les plus associés à la fidélisation ou au taux de désabonnement.

Descriptions de fichiers et informations sur les champs de données

Voici la traduction en français des variables du jeu de données

<u>Variable en anglais</u>	<u>Traduction en français</u>
RowNumber	Numéro de ligne
CustomerId	Identifiant client
Surname	Nom de famille
First.Name	Prénom
Date.of.Birth	Date de naissance
Gender	Sexe
Marital.Status	État matrimonial
Number.of.Dependents	Nombre de personnes à charge
Occupation	Profession
Income	Revenu
Education.Level	Niveau d'éducation
Address	Adresse
Contact.Information	Coordonnées (Téléphone ou Email)
Customer.Tenure	Ancienneté du client (en mois ou années selon le contexte)
Customer.Segment	Segment client
Preferred.Communication.Channel	Canal de communication préféré
Credit.Score	Score de crédit
Credit.History.Length	Durée de l'historique de crédit
Outstanding.Loans	Montant des prêts en cours
Churn.Flag	Indicateur de résiliation (0 = actif, 1 = résilié)
Churn.Reason	Motif de résiliation
Churn.Date	Date de résiliation
Balance	Solde
NumOfProducts	Nombre de produits détenus
NumComplaints	Nombre de réclamations

TAF

1. En vous inspirant de la méthodologie du cours de scoring, Faire la modélisation prédictive du CHURN en mettant en compétition les algorithmes de machine learning et de Deep learning
 2. Faire une analyse du profil de risque par règles d'association
 3. Si possible une petite application Shiny ou Streamlit pour l'interactivité de votre analyse
- NB : les analyses se feront dans le logiciel de votre choix (R, PYTHON, EXCEL, POWERBI)