

Projet de Data Mining/Machine Learning

Un responsable d'une banque souhaite réduire le nombre de clients qui quittent leurs services de carte de crédit. Il aimerait pouvoir anticiper le départ des clients afin de leur fournir de meilleurs services et ainsi les retenir.

Votre travail sera de mettre en place un modèle de Machine Learning capable de prédire les départs des clients. Pour cela, la banque met à votre disposition une base de données de 10127 clients. La variable cible est Attrition_Flag.

Les éléments à retourner pour ce projet sont :

- Un rapport au format PDF ou PowerPoint
- Un fichier de python contenant le script (commenté) de votre travail

La date limite est fixée au Jeudi 25 février 2021 à 23 H 59 MN. Chaque jour de retard équivaut à un malus de 1 point.

Les recommandations qui suivent sont relatives au rapport de l'étude que vous allez effectuer.

1. Consignes relatives à la forme du rapport écrit

La forme du rapport écrit doit être soignée :

- La page de couverture doit mentionner : le titre du rapport, l'école ENSAE, la formation, le nom des étudiants et celui de l'enseignant.
- Un sommaire est indispensable ainsi qu'un résumé d'une demi-page maximum.
- Il n'y a pas de contraintes sur le nombre de pages, mais il faut numéroter les pages.
- Soignez la rédaction de votre rapport. Évitez les copier / coller des sources web.

2. Sur le fond, le rapport doit être concis et pédagogique

Le rapport d'étude doit présenter de façon claire, dense et agréable à lire :

- le contexte de l'étude

- les données à étudier (taille du jeu de données, nombre de variables, statistiques descriptives de chaque variable, liaison en les variables deux à deux, ...)
- les principales difficultés à surmonter
- les techniques testées, les moyens mis en œuvre
- les résultats obtenus et l'analyse de ceux-ci
- la technique retenue et les raisons de votre choix
- une conclusion sous forme de bilan et de perspectives de l'étude ainsi qu'un bilan pour l'étudiant indiquant notamment les domaines où il a progressé et les compétences acquises

Les passages trop techniques doivent être renvoyés en annexe et le rapport d'étude doit constituer un document de travail utile à toute personne non initiée à la science des données.

Description du jeu de donnée :

Variable	Description
CLIENTNUM	Numéro client. Identifiant unique du client titulaire du compte
Attrition_Flag	Variable cible (activité client) - si le compte est fermé, 1 sinon 0
Customer_Age	Âge du client en années
Gender	Sexe du client (M = Homme, F = Femme)
Dependent_count	Nombre de personnes à charge
Education_Level	Niveau d'éducation
Marital_Status	situation maritale
Income_Category	Catégorie de revenu annuel
Card_Category	Type de carte
Months_on_book	Période de relation avec la banque
Total_Relationship_Count	Nombre total de produits détenus par le client
Months_Inactive_12_mon	Nombre de mois d'inactivité au cours des 12 derniers mois
Contacts_Count_12_mon	Nombre de contacts au cours des 12 derniers mois
Credit_Limit	Limite de crédit sur la carte de crédit
Total_Revolving_Bal	Solde renouvelable total sur la carte de crédit

Avg_Open_To_Buy	Ligne de crédit ouverte à l'achat (moyenne des 12 derniers mois)
Total_Amt_Chng_Q4_Q1	Changement du montant de la transaction (T4 par rapport au T1)
Total_Trans_Amt	Montant total de la transaction (12 derniers mois)
Total_Trans_Ct	Nombre total de transactions (12 derniers mois)
Total_Ct_Chng_Q4_Q1	Changement du nombre de transactions (T4 par rapport au T1)
Avg_Utilization_Ratio	Taux d'utilisation moyen de la carte