
OUMAR KANTE

EL AYACHI YASSINE

MASTER 2 MBFA-SIEF

Projet De Data Visualisation (Livrable - Note Manuscrite)

Objectif du Projet

L'objectif de ce projet est d'analyser l'efficacité des campagnes de marketing direct menées par une banque dans le but d'optimiser les souscriptions aux dépôts à terme. En exploitant les données des campagnes passées, nous avons cherché à comprendre quels facteurs influencent la décision des clients de souscrire ou non à ces produits financiers. Ce projet vise également à développer un modèle prédictif permettant de prévoir la probabilité qu'un client souscrive à un dépôt à terme en fonction de ses caractéristiques.

Le but ultime est d'aider la banque à mieux cibler ses campagnes futures en utilisant des modèles basés sur les données pour améliorer les taux de conversion.

Présentation des Données

Les données utilisées dans ce projet proviennent des résultats d'une campagne de marketing direct menée par une banque. Le dataset contient plusieurs milliers de lignes correspondant à des clients, avec un total de 18 variables, comprenant des informations démographiques, des données comportementales ainsi que des indicateurs sur les précédentes interactions avec la banque.

Les données sont divisées en variables explicatives et une variable cible. La variable cible est y , qui indique si un client a souscrit à un dépôt à terme (valeurs : *yes* ou *no*). Les variables explicatives comprennent des informations telles que l'âge, le métier, le statut marital, le solde bancaire moyen, la durée de la dernière communication, le nombre de contacts réalisés durant la campagne, et d'autres.

Variables et Indicateurs Utilisés

Les principales variables utilisées dans ce projet sont :

1. Variables Démographiques :

- *Age* : L'âge du client. Cet indicateur peut influencer la décision de souscription en fonction de l'âge des clients potentiels.
- *Job* : Le type de métier du client. Il peut être un indicateur de stabilité financière et d'influence sur la souscription.
- *Marital* : Le statut marital du client. Ce critère peut affecter les décisions financières.

2. Variables Comportementales :

- *Duration* : La durée de la dernière communication entre la banque et le client. Cette variable est cruciale car des communications plus longues peuvent indiquer un intérêt plus fort du client.
 - *Campaign* : Le nombre de contacts effectués durant la campagne. Cela reflète l'engagement du client avec la campagne marketing.
 - *Previous* : Le nombre de contacts réalisés avant cette campagne. Cela peut aider à mesurer l'historique d'engagement du client avec la banque.
3. **Variables de Souscription** :
- *Housing* et *Loan* : Indiquent respectivement si le client possède un prêt immobilier ou un prêt personnel. Ces informations aident à évaluer la situation financière du client.
4. **Variable Cible** :
- *y* : Indique si le client a souscrit à un dépôt à terme (yes ou no).
-

Difficultés Rencontrées

Lors de la mise en place du projet, plusieurs défis ont été rencontrés :

1. **Données Manquantes et Incomplètes** : Certaines variables avaient des valeurs manquantes, ce qui a nécessité un travail de nettoyage pour supprimer ou imputer les valeurs manquantes. Cela a exigé une réflexion sur les méthodes d'imputation appropriées pour chaque type de variable (numérique vs catégorielle).
 2. **Déséquilibre des Classes** : La variable cible *y* montre un déséquilibre marqué entre les clients ayant souscrit et ceux n'ayant pas souscrit (environ 12% de souscriptions). Cela a posé un défi pour les modèles de classification, car cela pouvait biaiser les résultats.
 3. **Choix des Modèles** : La sélection des modèles de machine learning s'est avérée difficile. Des modèles simples comme la régression logistique ont montré de bons résultats, mais nous avons aussi exploré des modèles plus complexes comme les forêts aléatoires pour améliorer la précision des prédictions. Trouver le modèle le plus performant avec un bon compromis entre complexité et précision a demandé plusieurs essais.
-

Idées pour Aller Plus Loin

Pour aller plus loin et améliorer ce projet, plusieurs axes peuvent être explorés :

1. **Enrichissement des Données** : L'ajout de nouvelles variables, comme les données comportementales en ligne (par exemple, les interactions sur les réseaux sociaux ou le comportement d'achat des clients), pourrait affiner encore davantage les prédictions et la personnalisation des campagnes.
2. **Segmentation des Clients** : Une segmentation plus fine des clients, par exemple par âge, localisation géographique ou segment de revenus, permettrait de personnaliser davantage les campagnes et d'augmenter leur efficacité.
3. **Prédiction en Temps Réel** : Intégrer ce modèle dans un système opérationnel pour effectuer des prédictions en temps réel lors des campagnes pourrait permettre à la banque d'adapter ses actions marketing instantanément et de manière plus ciblée.
4. **Retour sur Investissement (ROI)** : Intégrer une analyse financière pour évaluer le retour sur investissement des campagnes basées sur les prédictions générées par le modèle pourrait être une prochaine étape importante.