

## DESCRIPTION DU JEU DE DONNEES

Les données sont liées aux campagnes de marketing direct d'une institution bancaire portugaise. Ces campagnes de marketing étaient basées sur des appels téléphoniques. Souvent, plusieurs contacts avec le même client étaient nécessaires pour savoir si le produit (dépôt bancaire à terme) serait ('oui') ou non ('non') souscrit.

L'objectif de la classification est de prédire si le client va souscrire (oui/non) un dépôt à terme (variable y).

Variables d'entrée :

### # données du client de la banque :

- 1 - âge (numérique)
- 2 - emploi : type d'emploi (catégories : 'admin.', 'col bleu', 'entrepreneur', 'femme de ménage', 'direction', 'retraité', 'indépendant', 'services', 'étudiant', 'technicien', 'chômeur', 'inconnu')
- 3 - état civil : état civil (catégorique : 'divorcé', 'marié', 'célibataire', 'inconnu' ; remarque : 'divorcé' signifie divorcé ou veuf)
- 4 - education (catégories : 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
- 5 - défaut de paiement : a un crédit en défaut de paiement ? (catégorique : 'non', 'oui', 'inconnu')

6 - housing : a un prêt au logement ? (catégories : 'non', 'oui', 'inconnu')

7 - prêt : a un prêt personnel ? (catégories : 'non', 'oui', 'inconnu')

**# lié au dernier contact de la campagne en cours :**

8 - contact : type de communication du contact (catégorique : 'cellulaire', 'téléphone')

9 - month : mois de l'année du dernier contact (catégorique : 'jan', 'feb', 'mar', ..., 'nov', 'dec')

10 - day\_of\_week : jour de la semaine du dernier contact (catégorique : 'lun', 'tue', 'wed', 'thu', 'fri')

11 - duration : durée du dernier contact, en secondes (numérique). Note importante : cet attribut affecte fortement la cible de sortie (par exemple, si duration=0 alors y='no'). Pourtant, la durée n'est pas connue avant qu'un appel ne soit effectué. De même, après la fin de l'appel, y est évidemment connu. Ainsi, cette entrée ne devrait être incluse qu'à des fins de référence et devrait être écartée si l'intention est d'avoir un modèle prédictif réaliste.

**# Autres attributs :**

12 - campaign : nombre de contacts effectués pendant cette campagne et pour ce client (numérique, inclut le dernier contact).

13 - pdays : nombre de jours qui se sont écoulés depuis le dernier contact du client lors d'une campagne précédente (numérique ; 999 signifie que le client n'a pas été contacté auparavant).

14 - previous : nombre de contacts effectués avant cette campagne et pour ce client (numérique)

15 - poutcome : résultat de la campagne marketing précédente (catégorique : 'échec', 'inexistant', 'succès')

**Variable de sortie (objectif souhaité) :**

21 - y - le client a-t-il souscrit un dépôt à terme ? (binaire : 'oui', 'non')