

Projet Introduction aux réseaux de neurones profonds

Le but de ce projet est de vous permettre de mettre en pratique tout ce que vous avez appris sur les réseaux de neurones. Ce projet vous permettra de creuser vos connaissances sur les outils et techniques utilisés pour mettre en place un réseau de neurones artificiel, tant du côté théorique que du côté pratique.

Contexte : Est ce qu'un client va souscrire à une assurance mobile ou pas ?

Les données étudiées sont des données d'une banque aux États-Unis. Outre les services habituels, cette banque propose également des services d'assurance automobile. La banque organise des campagnes régulières pour attirer de nouveaux clients. La banque dispose des données des clients potentiels et les employés de la banque les appellent pour annoncer les options d'assurance automobile disponibles. Nous recevons des informations générales sur les clients (âge, emploi, etc.) ainsi que des informations plus spécifiques sur la campagne de vente d'assurance en cours (communication, dernier jour de contact) et les campagnes précédentes (attributs tels que tentatives précédentes, résultat).

Vous avez des données sur 4000 clients qui ont été contactés lors de la dernière campagne et pour lesquels les résultats de la campagne (le client a-t-il souscrit une assurance ou non) sont connus.

Descriptif des données :

Age

Âge du client (en années). Variable démographique influençant la probabilité de souscrire une assurance.

Job

Profession ou catégorie socio-professionnelle du client. Sert à identifier des profils types.

Marital

Statut marital du client : *Single*, *Married*, *Divorced*. Peut influencer les comportements d'achat.

Education

Niveau d'éducation : primaire, secondaire, universitaire, etc. Souvent corrélé au revenu et au comportement financier.

Default

Indique si le client est en défaut de paiement d'un crédit (*yes/no*). Sert à évaluer le niveau de risque financier.

Balance

Solde moyen du compte bancaire. Représente la capacité financière ou la stabilité économique du client.

HHInsurance

Indique si le client possède une assurance habitation (1 = oui, 0 = non). Utile pour repérer les profils déjà assurés.

CarLoan

Présence d'un prêt automobile (1 = oui, 0 = non). Les clients avec un prêt sont plus enclins à couvrir leur véhicule.

Communication

Type de canal utilisé pour contacter le client : *cellular, telephone*, etc. Certains canaux sont plus efficaces.

LastContactDay

Jour du mois (1–31) du dernier contact effectué lors de la campagne courante.

LastContactMonth

Mois du dernier contact : *jan, feb, mar, ..., dec*. Permet d'analyser les effets de saisonnalité dans la campagne.

NoOfContacts

Nombre total de contacts réalisés durant la campagne actuelle.

DaysPassed

Nombre de jours écoulés depuis le dernier contact issu d'une campagne précédente.

Valeur -1 = aucun contact précédent.

PrevAttempts

Nombre de tentatives effectuées dans les campagnes antérieures. Reflète la difficulté à convertir le client.

Outcome

Résultat de la campagne précédente : *success, failure, other, unknown*.
Indicateur comportemental important.

CallStart

Heure de début de l'appel (format HH:MM:SS). Peut être utilisé pour déterminer le moment de la journée.

CallEnd

Heure de fin de l'appel (format HH:MM:SS). Sert à calculer la durée de l'appel, souvent corrélée à l'intérêt du client.

CarInsurance

Variable cible.

Indique si le client a souscrit une assurance automobile suite à la campagne (1 = oui, 0 = non).

Feature	Description	Example
Id	Unique ID number. Predictions file should contain this feature.	"1" ... "5000"
Age	Age of the client	
Job	Job of the client.	"admin.", "blue-collar", etc.
Marital	Marital status of the client	"divorced", "married", "single"
Education	Education level of the client	"primary", "secondary", etc.
Default	Has credit in default?	"yes" - 1, "no" - 0
Balance	Average yearly balance, in USD	
HHInsurance	Is household insured	"yes" - 1, "no" - 0
CarLoan	Has the client a car loan	"yes" - 1, "no" - 0
Communication	Contact communication type	"cellular", "telephone", "NA"
LastContactMonth	Month of the last contact	"jan", "feb", etc.
LastContactDay	Day of the last contact	
CallStart	Start time of the last call (HH:MM:SS)	12:43:15
CallEnd	End time of the last call (HH:MM:SS)	12:43:15
NoOfContacts	Number of contacts performed during this campaign for this client	
DaysPassed	Number of days that passed by after the client was last contacted from a previous campaign (numeric; -1 means client was not previously contacted)	
PrevAttempts	Number of contacts performed before this campaign and for this client	
Outcome	Outcome of the previous marketing campaign	"failure", "other", "success", "NA"
CarInsurance	Has the client subscribed a CarInsurance?	"yes" - 1, "no" - 0

Ce qui est demandé dans le projet .

Partie 1 :

Dans cette partie, vous ferez une analyse et un traitement de données afin de bien préparer les données qui seront utilisées pour mettre en place le réseau de neurones . Dans cette analyse, les attendues seront :

- Regarder les valeurs manquantes et les remplacées par un moyen simple
- une visualisation de certaines variables par rapport à notre target "CarInsurance"
- Regarder s'il y a corrélation des variables quantitatives avec la matrice de corrélation
- supprimer certaines variables qui ne sont pas pertinentes pour le modèle

Partie 2:

Dans cette partie, vous vous inspirez du premier Tp en mettant en place un réseau de neurones avec l'ensemble de toutes les phases qui ont été vues en TP.

une fois le modèle mis en place, vous choisirez un metric (accuracy, f1-score, précision etc..) selon votre modèle et selon ce que vous voulez prédire.

Vous mettrez aussi en place un moyen permettant d'optimiser les hyperparamètres pour rendre performant le modèle, si vous jugez que les résultats obtenus ne sont pas satisfaisants.

Il est à noter que la partie préparation et traitement des données sera notée sur 70% et la partie mise en place du modèle et optimisation des hyperparamètres sur 30%.

NB : Ce projet est à faire sur jupyter notebook, ce qui facilitera la lecture des résultats sans lancement de votre code.

Vous pouvez me le rendre par mail : **mouhamedbaisidk@groupeisi.com**

Date limite de rendu : 30 décembre 2025

Bonne chance