

Objectif

Entraîner notre premier réseau neuronal pour prédire la résiliation de clients

Dans ce notebook, notre objectif principal est d'acquérir une expérience pratique des réseaux neuronaux et de leur application à la résolution de problèmes réels. Nous travaillerons avec un ensemble de données de résiliation de clients, visant à comprendre les fondamentaux de la création et de l'entraînement d'un réseau neuronal pour la modélisation prédictive.

Objectifs clés :

- Prétraiter et explorer l'ensemble de données de résiliation de clients.
- Apprendre les bases de l'architecture des réseaux neuronaux.
- Entraîner un modèle de réseau neuronal pour prédire la résiliation de clients.
- Évaluer l'exactitude et les performances du modèle.
- Prédire si le client suivant va abandonner ou non la banque
- Evaluation Robuste avec Keras Classifier
- Recherche des meilleurs Hyperparametre avec Grid Search

Application 1 :

Préparer le dataset:

- Data Importation: Importer les données du fichier CSV "Churn_Modelling.csv".
- Data Cleaning: Vérifier et traiter les valeurs manquantes dans le dataset.
- Data Visualization: Explorer visuellement les données, notamment en utilisant des graphiques pour les données catégorielles et des histogrammes pour les données numériques.
- Feature Engineering: Créer des variables dummy pour les variables catégorielles comme "Geography" et "Gender".
- Data Preparation and Preprocessing: Diviser le dataset en ensembles d'entraînement et de test, puis mettre à l'échelle les caractéristiques à l'aide d'une transformation StandardScaler.

Data Importation

index	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	IsActiveCard	IsActiveMember	EstimatedSalary	Exited
0	1	15854002	Morgane	619	France	Female	42	2	0	1	1	1	101344.88	1
1	2	15847311	Mel	608	Spain	Female	41	1	81007.86	1	1	1	112542.38	0
2	3	15810954	Olga	562	France	Female	42	8	154600.8	3	1	0	111581.57	1
3	4	15791814	Boni	688	France	Female	19	1	0	0	2	0	93624.63	0
4	5	15791868	Melinda	894	Spain	Female	41	2	125110.62	1	1	1	79854.1	0
5	6	15774913	Chlo	681	Spain	Male	44	8	111376.76	3	1	0	149776.17	1
6	7	15952811	Burkard	832	France	Male	50	7	0	0	2	1	10942.8	1
7	8	15889148	Olivera	776	Germany	Female	29	4	115046.76	4	1	0	119568.86	0
8	9	15762366	Reb	881	France	Male	44	4	142091.07	2	0	1	79469.5	0
9	10	15923389	Hel	684	France	Male	27	2	144031.88	1	1	1	71723.73	0
10	11	15767021	Bernie	568	France	Male	31	0	920716.72	2	0	0	92181.12	0
11	12	15727173	Andreas	497	Spain	Male	24	0	0	2	1	0	76395.01	0
12	13	15822264	Kay	476	France	Female	34	10	0	2	1	0	26203.98	0
13	14	15814493	Chlo	597	France	Female	26	5	0	2	0	0	94002.79	0
14	15	15408862	Scott	635	Spain	Female	15	7	0	0	2	0	65961.65	0
15	16	15847968	Gottfrid	616	Germany	Male	45	3	141129.41	2	0	1	94527.26	0
16	17	15713452	Ramon	493	Germany	Male	50	1	124000.88	1	1	0	2897.57	1
17	18	15786218	Henderson	549	Spain	Female	24	0	0	2	1	1	14404.41	0
18	19	15881507	Melissa	587	Spain	Male	45	0	0	1	0	0	158948.81	0
19	20	15949962	Reb	735	France	Female	24	0	0	2	1	1	14724.89	0
20	21	15778517	McDonald	752	France	Male	41	8	0	2	1	1	170885.17	0

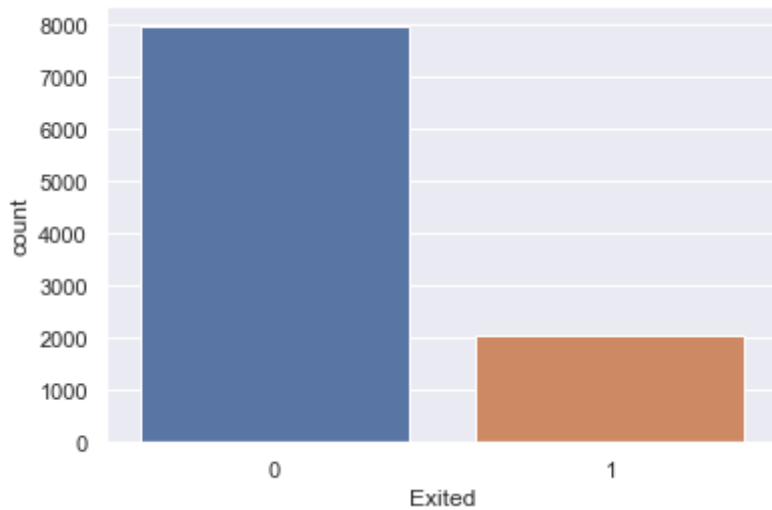
ChurnModelingDataset	
□	RowNumber: int64
□	CustomerId: int64
□	Surname: object
□	CreditScore: int64
□	Geography: object
□	Gender: object
□	Age: int64
□	Tenure: int64
□	Balance: float64
□	NumOfProducts: int64
□	HasCrCard: int64
□	IsActiveMember: int64
□	EstimatedSalary: float64
□	Exited: int64
□	dtype: object

1. **RowNumber:** Numéro de la ligne dans le dataset.
2. **CustomerId:** Identifiant unique du client.
3. **Surname:** Nom de famille du client.
4. **CreditScore:** Score de crédit du client.
5. **Geography:** Pays d'origine du client (France, Germany, Spain).
6. **Gender:** Genre du client (Male, Female).
7. **Age:** Âge du client.
8. **Tenure:** Nombre d'années pendant lesquelles le client a été client de la banque.
9. **Balance:** Solde du compte du client.
10. **NumOfProducts:** Nombre de produits bancaires détenus par le client.
11. **HasCrCard:** Indique si le client possède une carte de crédit (1 pour Oui, 0 pour Non).
12. **IsActiveMember:** Indique si le client est un membre actif (1 pour Oui, 0 pour Non).
13. **EstimatedSalary:** Salaire estimé du client.
14. **Exited:** Variable cible binaire indiquant si le client a résilié (1 pour Oui, 0 pour Non).

Ces variables fou

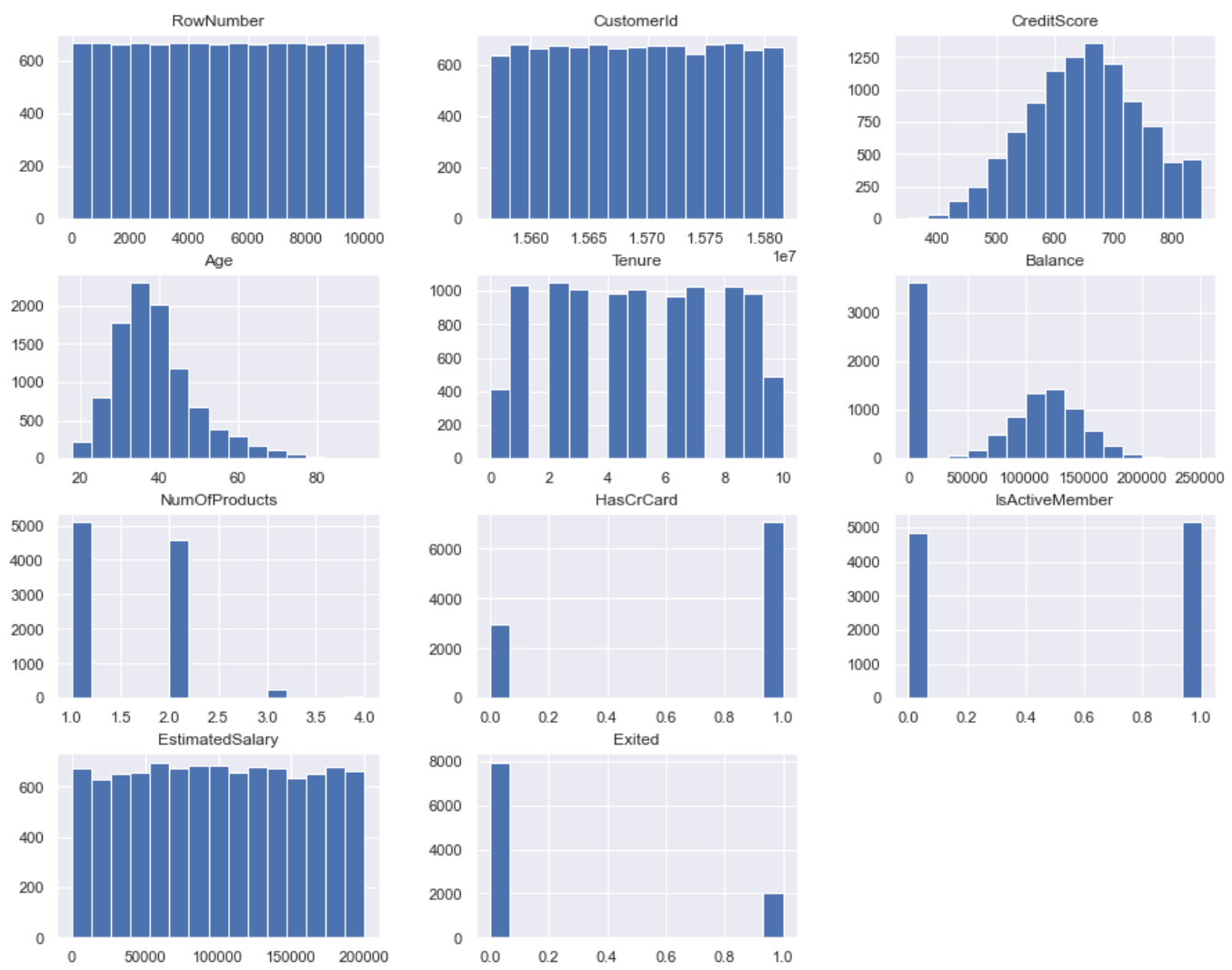
Data Visualisation

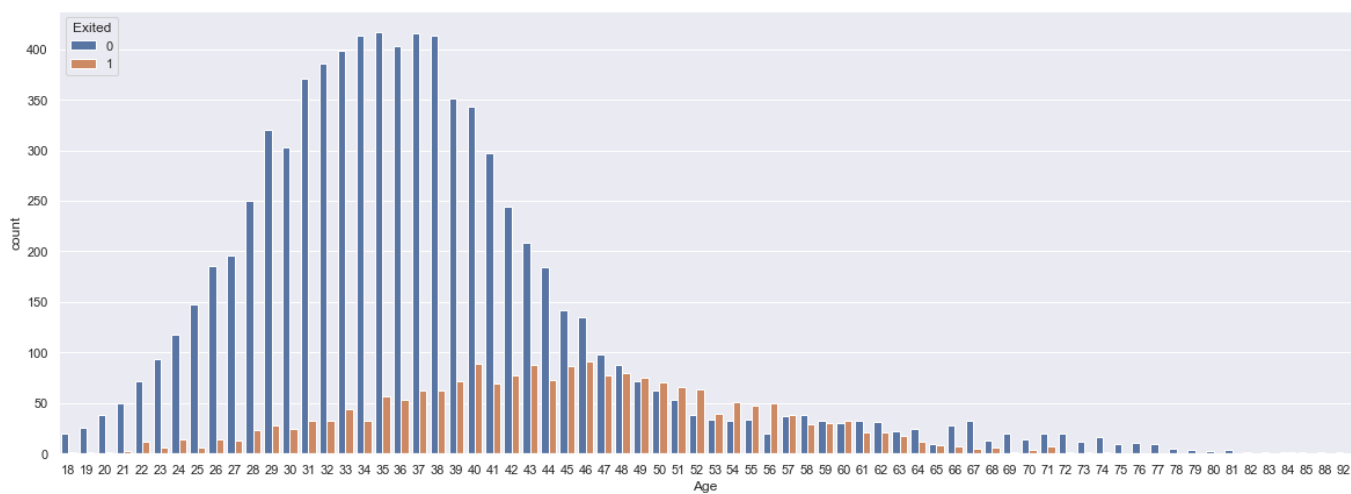
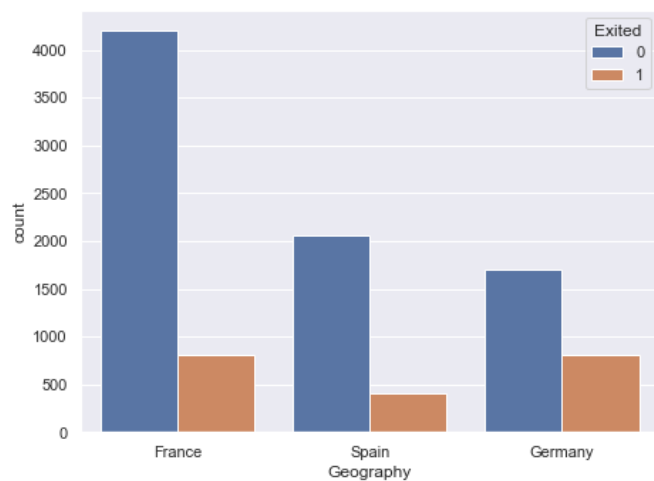
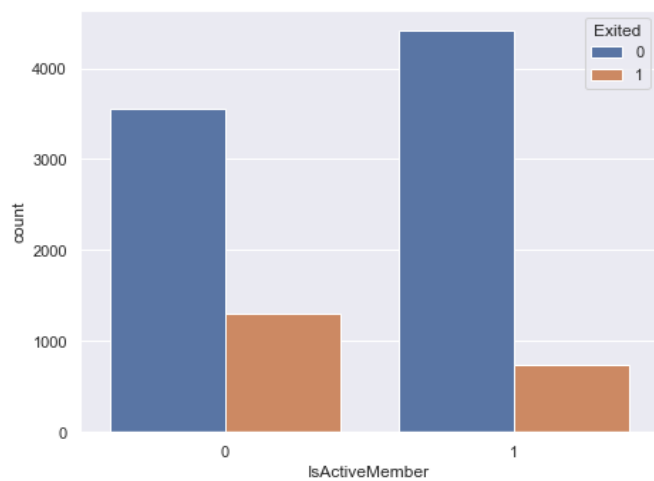
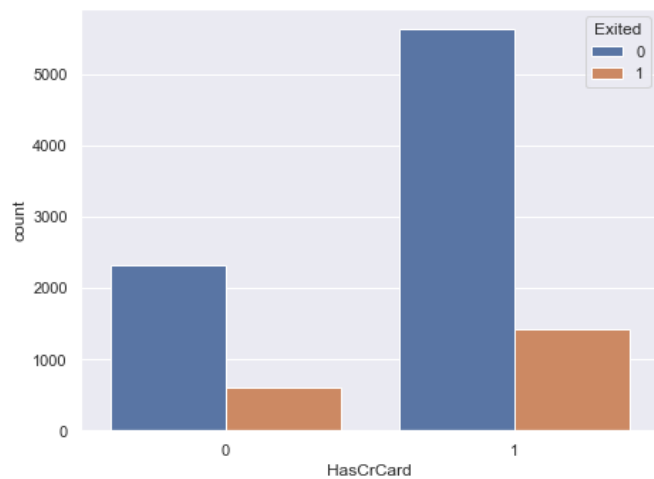
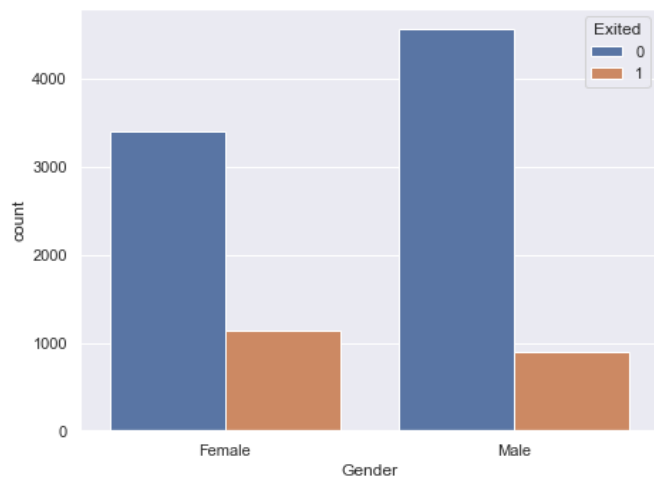
Récupération du nombre de personnes sorties



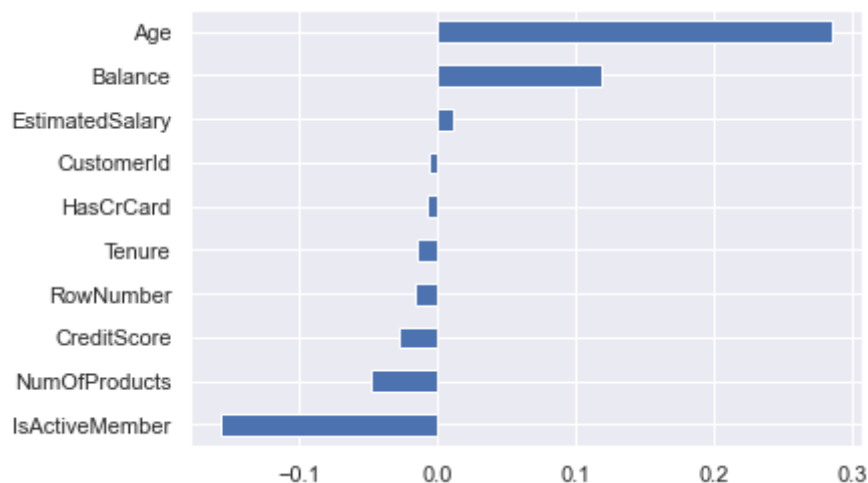
Explorons visuellement les données, en utilisant des graphiques pour les données catégorielles et des histogrammes pour les données

numériques.





Exploration de la corrélation avec la variable cible "Exited"



Feature Engineering

Remplacer des variable categorielle (Geography et Gender)qui alimentront le modèle selon un encodage qui vont jouer un role crucial pour influcrer notre modèle et prédire avec de nouvelle donnée

Data Preprocessing

Standariser les donnèer avec Feature Scaling et Découper les données pour l'entrainement et le test

2. Créer le réseau de neurone adéquat

- Le réseau de neurones créé a une couche d'entrée avec 11 neurones, deux couches cachées avec 20 et 15 neurones respectivement, des fonctions d'activation 'relu' pour les couches cachées, une couche de sortie avec 1 neurone et une fonction d'activation 'sigmoid' pour la classification binaire.
- Les couches de Dropout sont utilisées pour la régularisation, ce qui peut aider à prévenir le surapprentissage du modèle.

3. Compiler le réseau de neurones:

- j'ai choisi l'algorithme d'optimisation
- Adam pour la mise à jour des poids.
- Sélectionner la fonction de perte `binary_crossentropy`, adaptée aux problèmes de classification binaire.
- Utiliser la métrique 'accuracy' pour évaluer les performances du modèle

4. Entraîner le modèle:

On utilisant le jeu de données d'entraînement pour ajuster les poids du réseau. 100 epochs est effectué pour l'entraînement du modèle.

5. Calculer la matrice de confusion:

- Prédire les classes sur le jeu de données de test et les transformer en donnée boolean
 - Calculer la matrice de confusion en comparant les prédictions aux valeurs réelles.
- Matrice de Confusion :

Expectation \ Prediction	True	False
Actual True	1568	27
Actual False	238	167

- Rapport de Classification

	Precision	Recall	F1-Score	Support
Class 0	0.87	0.98	0.92	1595
Class 1	0.86	0.41	0.55	405
Accuracy			0.87	2000

	Precision	Recall	F1-Score	Support
Macro Avg	0.87	0.69	0.73	2000
Weighted Avg	0.87	0.87	0.85	2000

6. Mesurer l'accuracy du modèle:

Le Resultat de l'accuracy du modèle est

0.867

7. Prédiction d'un nouveau client:

- On utilisant le modèle entraîné nous concluons que le nouveau client va abandonner la banque



Application 2 :

8. Création du Modèle:

- Modèle séquentiel avec deux couches cachées et une couche de sortie.
- Couches cachées : ReLU, He_normal, dropout pour la régularisation.
- Couche de sortie : sigmoïde pour la classification binaire.
- Compilation : Optimiseur Adam, perte binaire 'binary_crossentropy', métrique 'accuracy'.

Évaluation avec Cross Validation:

- Validation croisée StratifiedKFold (5 plis).
- Prétraitement des données : Variables catégorielles transformées en variables dummy.
- Mise à l'échelle des caractéristiques avec StandardScaler.
- Cross_val_score : Accuracy moyenne de 79.71%, écart-type de 0.86%.

»

Application 3 :

Grid Search pour les Hyperparamètres (100-200 epochs):

- Grid search sur batch_size et epochs (100-200).
- Meilleurs paramètres : {'batch_size': 25, 'epochs': 100}, score : 0.857