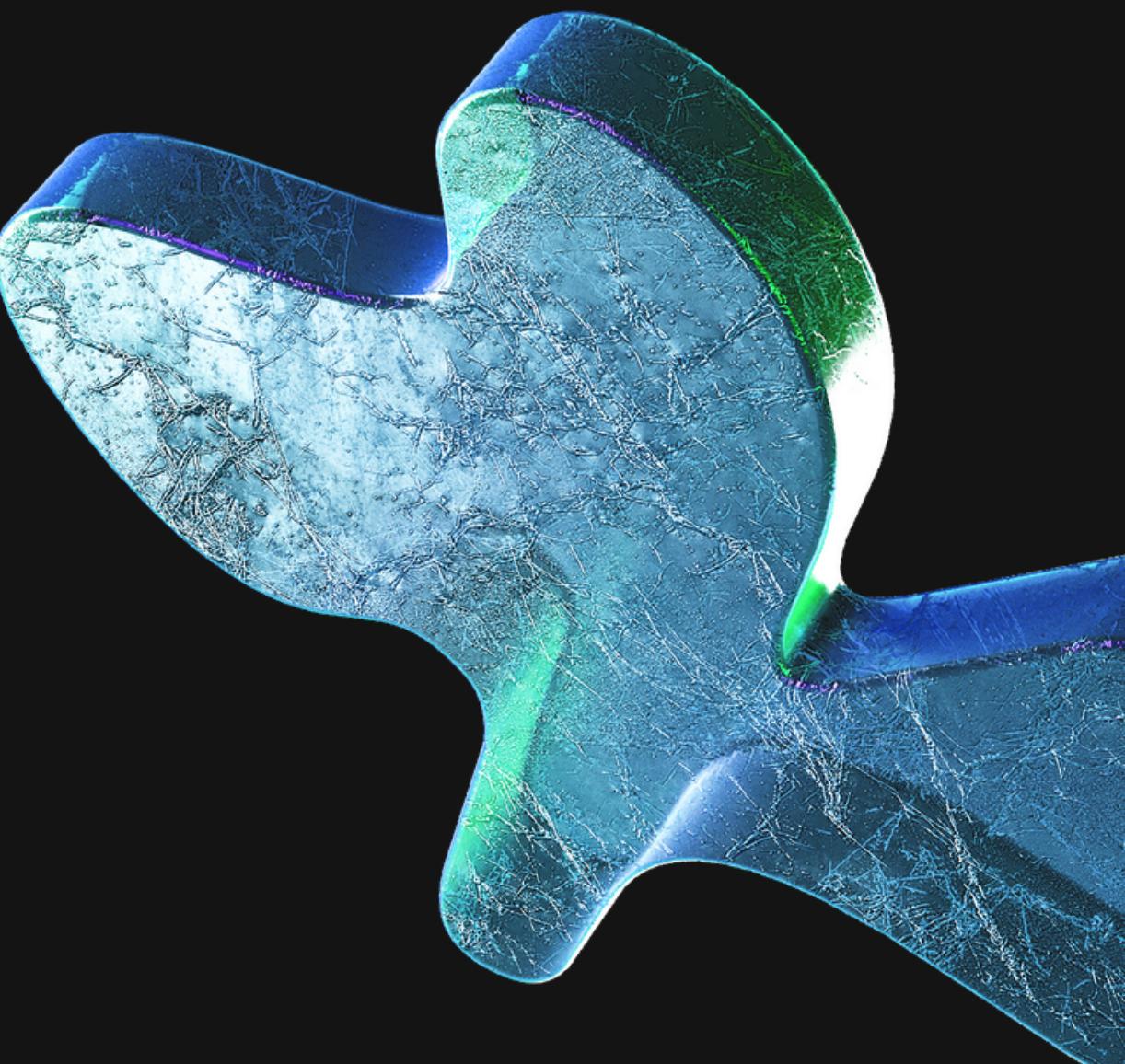


Projet file rouge

Certification Simplon.co : E2 , E1



Sommaire

Prédiction de prêt bancaire

E1/E2

E2 Optimisation du modèle

Présentation du projet

Présentation des données

Transformation des données

Modèle

Résultat

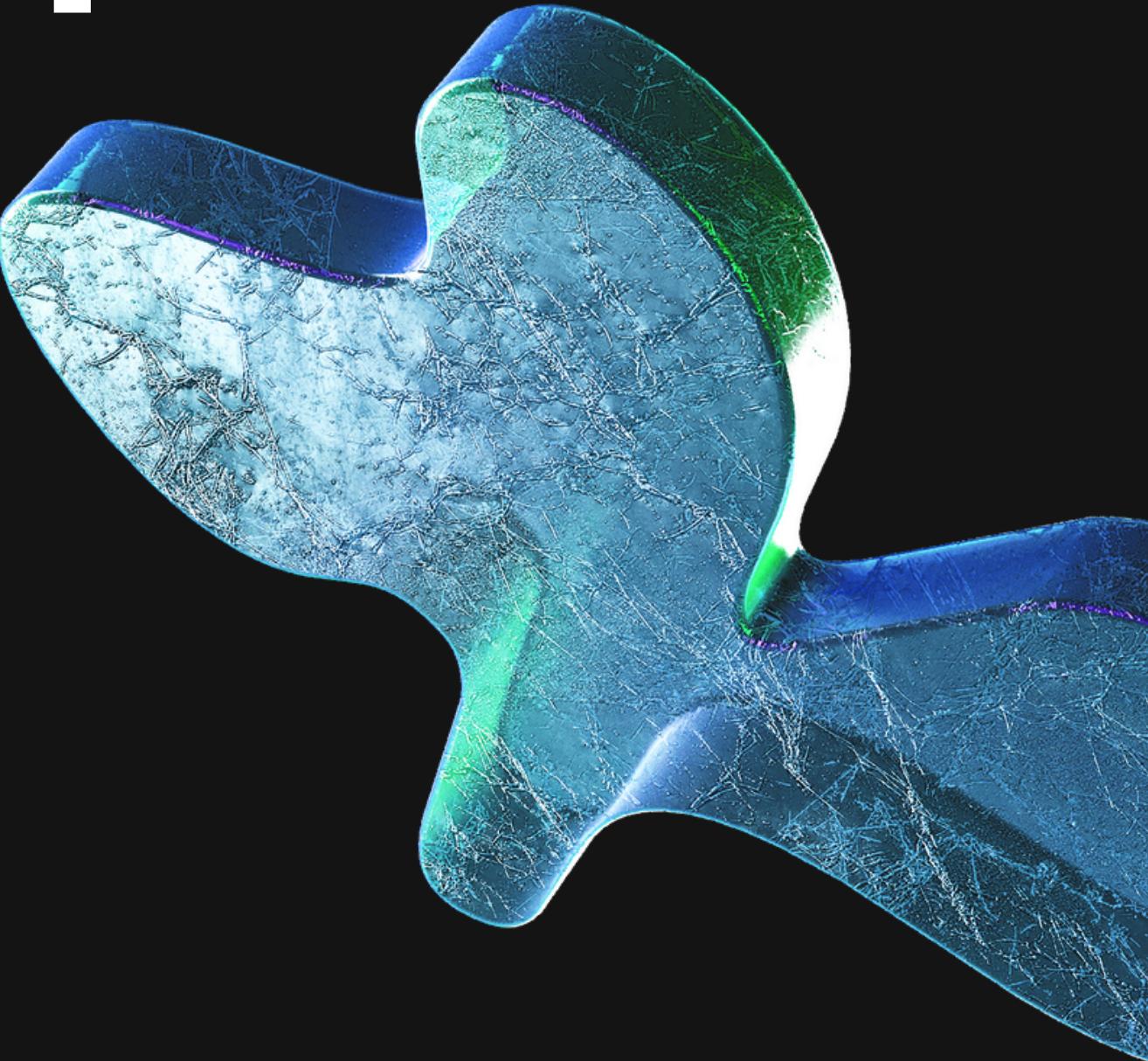
Déploiement

Organisation

Conclusion

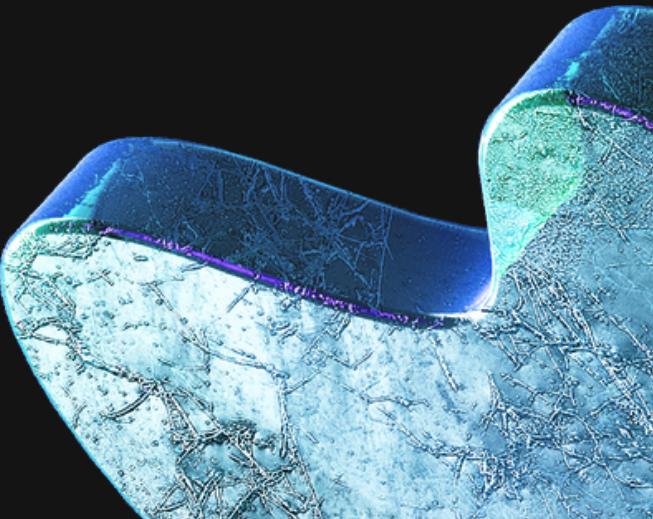
Optimisation d'un model

Optimisation d'un model de prédiction de crise cardiaque



Mise en situation

Sanofi l'entreprise de mon alternance ma demandé d'optimisé un model d'intelligence artificiel qu'il on déjà crée mai qui n'est pas optimisé .
Je devrais ensuite présenté mes amélioration lors d'un entretien .



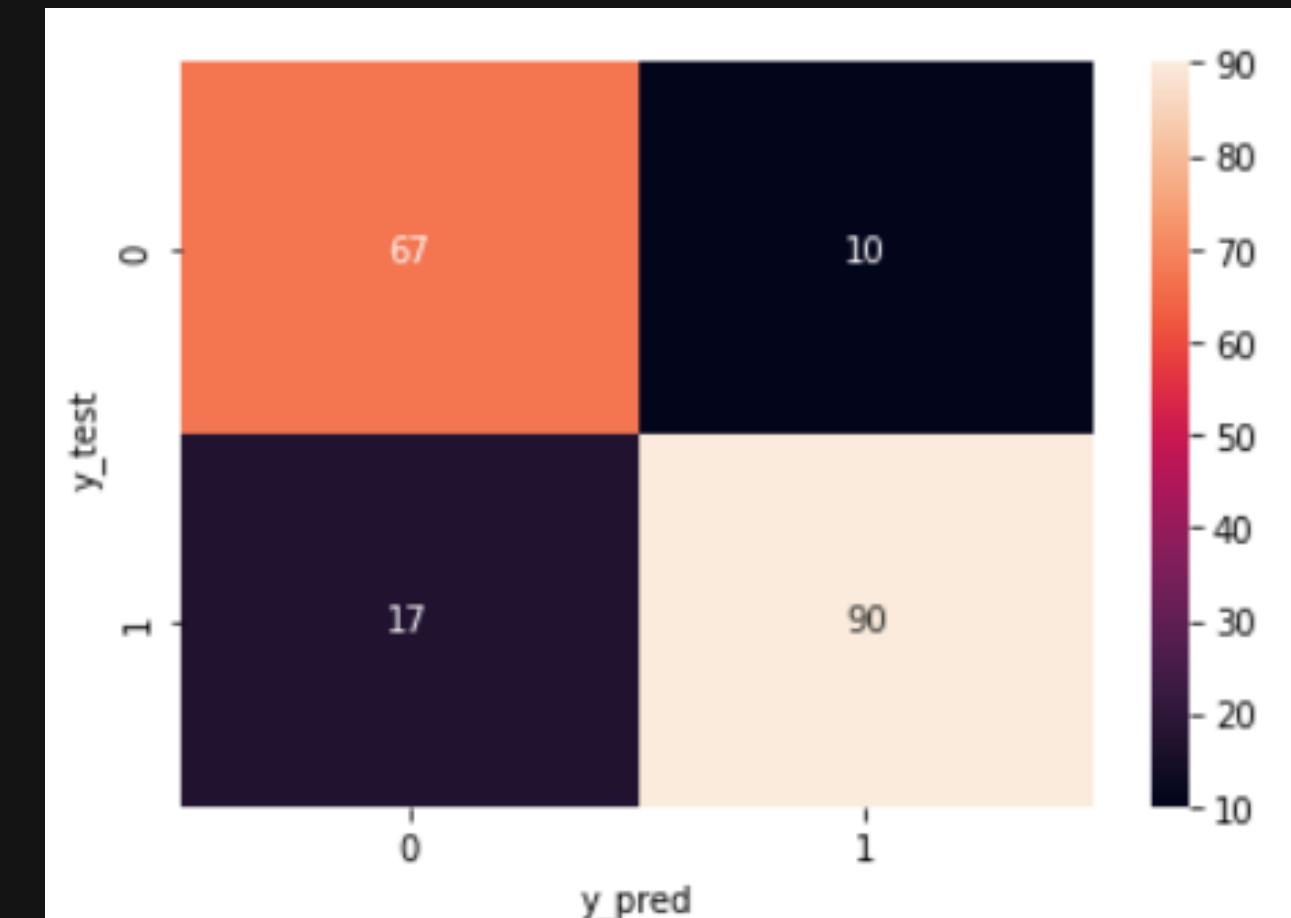
Modèle de référence

RandomForestClassifier

Classification Report

accuracy train : 1.000				
accuracy test : 0.891				
	precision	recall	f1-score	support
0	0.85	0.90	0.87	77
1	0.92	0.89	0.90	107
accuracy			0.89	184
macro avg	0.89	0.89	0.89	184
weighted avg	0.89	0.89	0.89	184

Confusion Matrix



Optimisation du modèle

SMOTE

Smote est utilisée pour balancée les données .

GridSearchCV

Algorithme qui permet de trouvée les meilleurs valeurs à mettre pour nos Hyper Paramètre du modèle

HyperParamètre

Les hyper paramètre utilisée sont :
max_features
n_estimators
max_depth
criterion

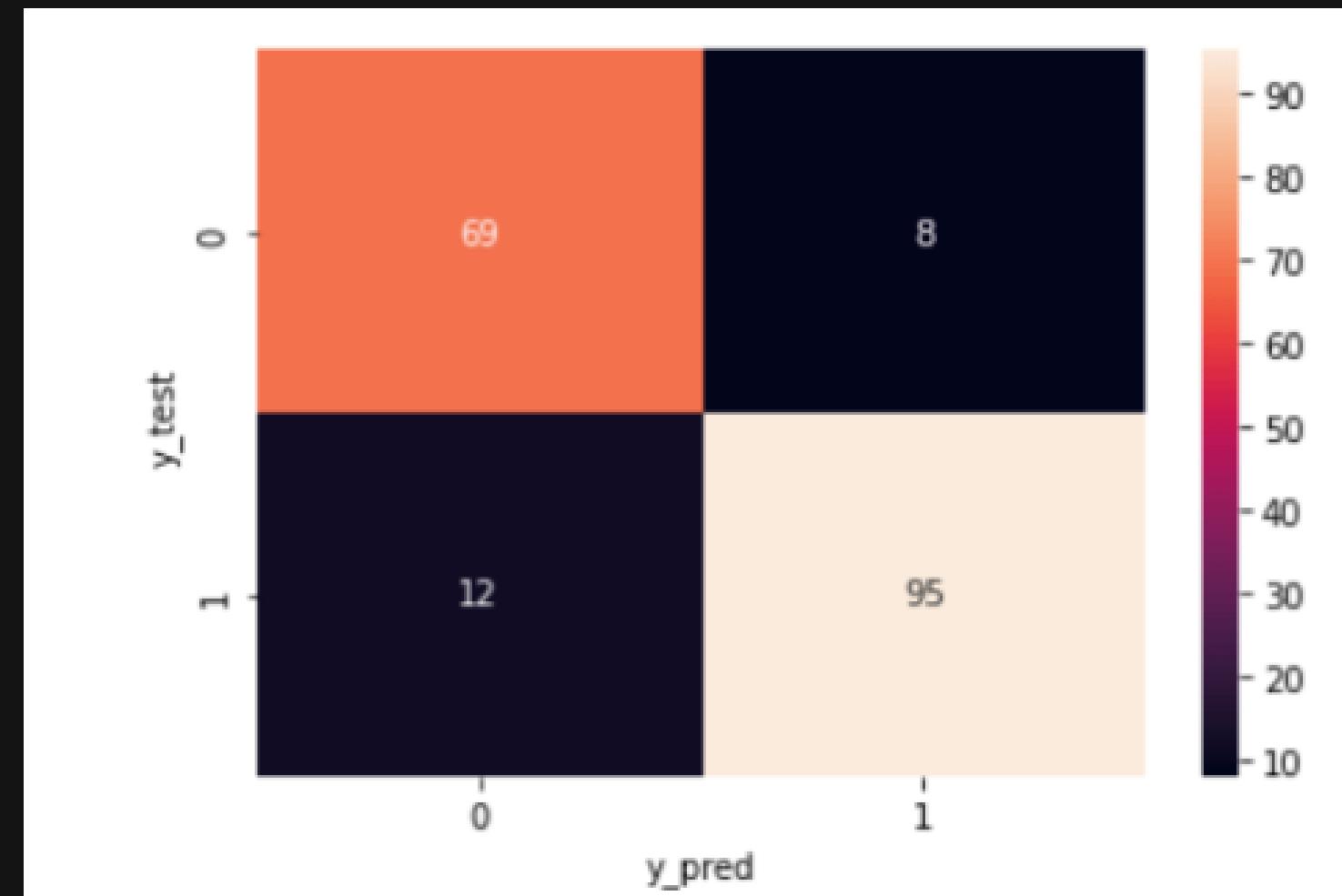
Modèle de optimisé

RandomForestClassifier

Classification Report

	precision	recall	f1-score	support
0	0.76	0.88	0.82	77
1	0.91	0.80	0.85	107
accuracy			0.84	184
macro avg	0.83	0.84	0.84	184
weighted avg	0.85	0.84	0.84	184

Confusion Matrix



Nouvelle fonctionnalité

The screenshot displays a user interface for a machine learning application, likely a credit risk or medical diagnosis model. It consists of three main sections: User Input Features, Prediction Probability, and Prediction.

User Input Features: This section contains various input fields for the user to enter their characteristics. The fields include:

- Enter your age: A numeric input field with a value of 0,00 and increment/decrement buttons.
- Sex: A dropdown menu with the value 0.
- Chest pain type: A dropdown menu with the value 0.
- Resting blood pressure: A numeric input field with a value of 0,00 and increment/decrement buttons.
- Serum cholestral in mg/dl: A numeric input field with a value of 0,00 and increment/decrement buttons.
- Fasting blood sugar: A dropdown menu with the value 0.
- Resting electrocardiographic results: A numeric input field with a value of 0,00 and increment/decrement buttons.
- Maximum heart rate achieved: A numeric input field with a value of 0,00 and increment/decrement buttons.

Prediction Probability: This section shows a table of prediction probabilities for different values of the input features. The table has two columns: the first column represents the input value (e.g., 0, 1, 2, ..., 9) and the second column represents the probability (e.g., 1.0000, 0.6700, 0.9800, ...).

	0	1
0	1.0000	0.0000
1	0.3300	0.6700
2	0.9800	0.0200
3	0.1000	0.9000
4	1.0000	0.0000
5	0.9800	0.0200
6	1.0000	0.0000
7	1.0000	0.0000
8	0.0900	0.9100
9	1.0000	0.0000

Prediction: This section displays the predicted output for the entered features. It includes a table of predicted values and a "Prediction" button.

	0
0	0
1	1
2	0
3	1
4	0
5	0
6	0
7	0
8	1
9	0

Prediction: A red-bordered button labeled "Prediction".

True: A table showing the true values for each feature.

	0	1
0	1.0000	0.0000
1	0.3300	0.6700
2	0.9800	0.0200
3	0.1000	0.9000
4	1.0000	0.0000
5	0.9800	0.0200
6	1.0000	0.0000
7	1.0000	0.0000
8	0.0900	0.9100
9	1.0000	0.0000

Ajout d'un bouton

Le bouton permet de montré les prédition une fois le bouton cliqué est non avant comme précédemment

Test unitaire

Différent test unitaire sur
l'application :

- Vérification que le dataset soit de la bonne taille .
- Vérification que les prédictions de l'application reste entre 0 et 1 si mon modèle aller au dessus ou en dessous de ces nombres cela voudrais dire que mon modèle est défectueux .
- Vérification que le nombre de prédition soit égale au nombre d'échantillons donner

Planning E2

E2 file rouge Visible par l'espace de travail | 000 Tableau Power-ups | Automatisation | Filtre | EL Partager |

a faire Ajoutez une autre liste

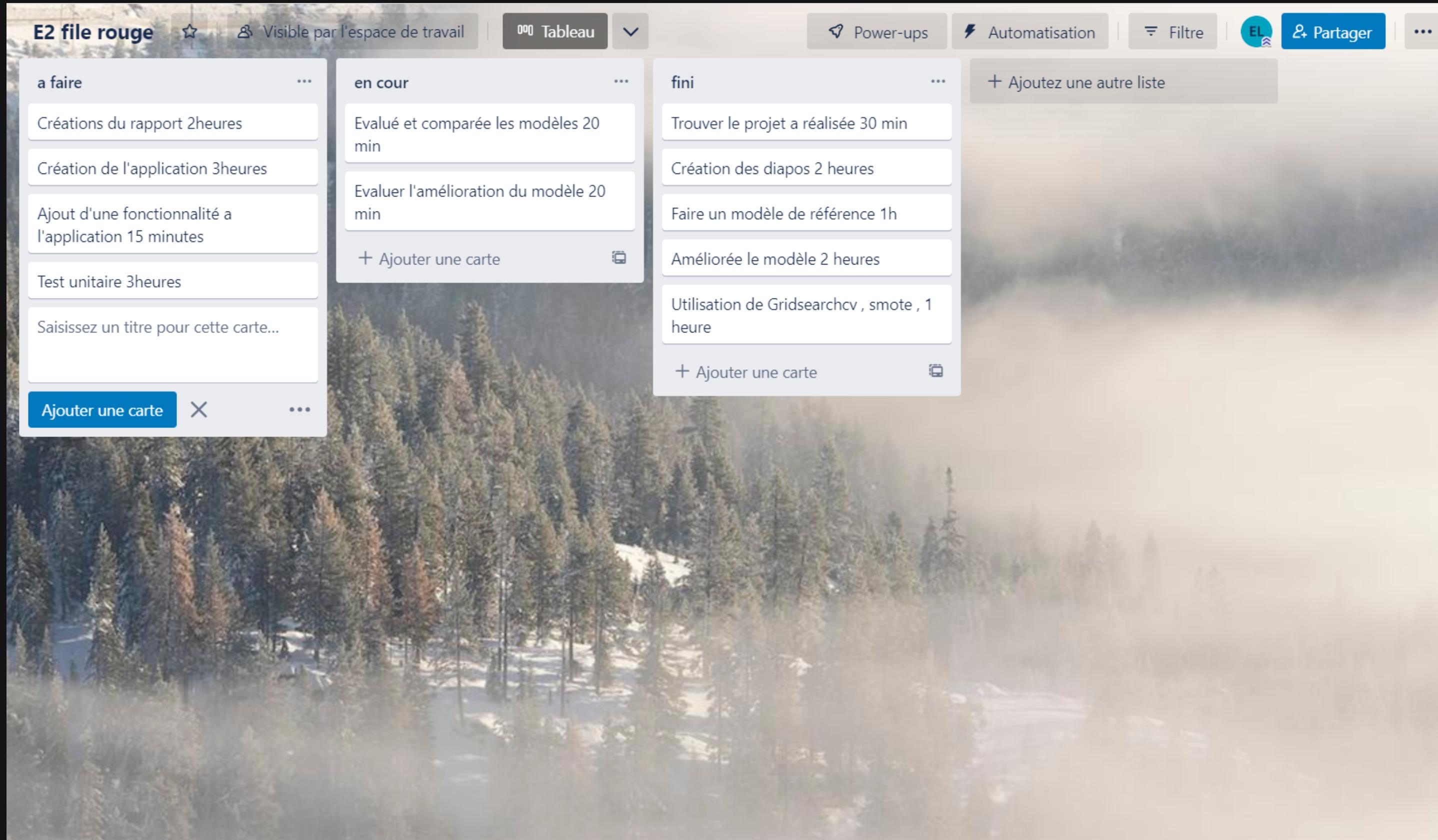
- Créations du rapport 2heures
- Création de l'application 3heures
- Ajout d'une fonctionnalité a l'application 15 minutes
- Test unitaire 3heures
- Saisissez un titre pour cette carte...

en cours Ajouter une carte

- Evalué et comparée les modèles 20 min
- Evaluer l'amélioration du modèle 20 min

fini Ajouter une carte

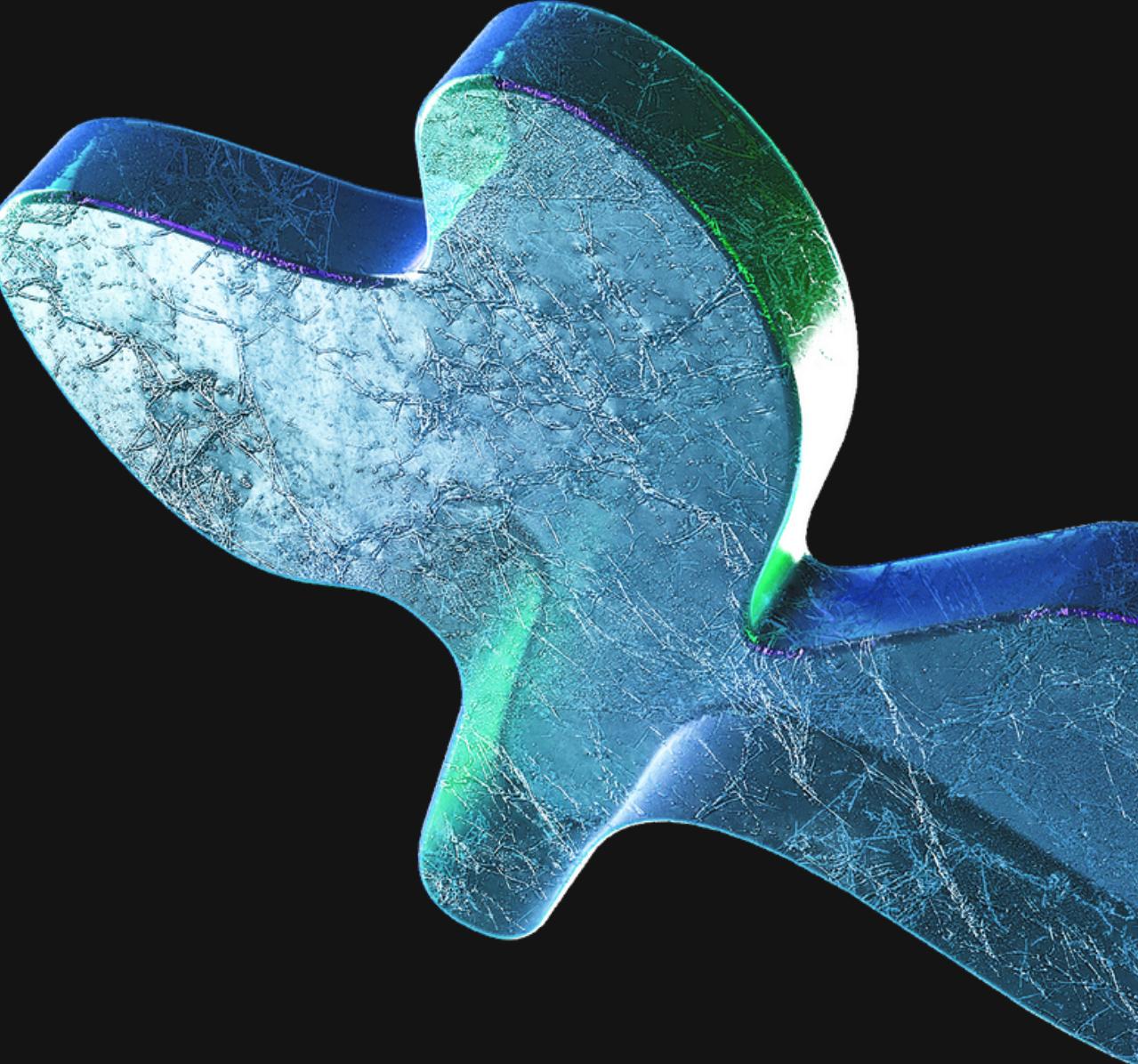
- Trouver le projet a réalisé 30 min
- Création des diapos 2 heures
- Faire un modèle de référence 1h
- Améliorée le modèle 2 heures
- Utilisation de Gridsearchcv , smote , 1 heure



E2

Prédiction de prêt

Allez vous avoir le prêt ?



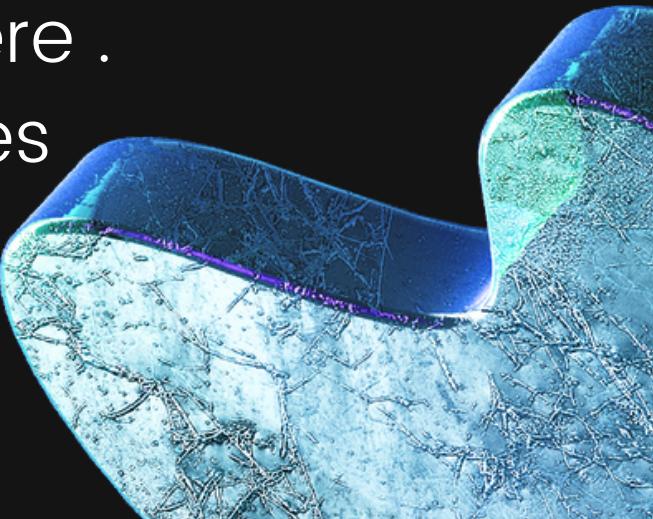
Mise en situation

Je suis data scientist mon client est banquier et m'a demandé de réaliser un projet pour l'aider lors de la décision d'une demande de prêts bancaires il veut être sûr de choisir les meilleures personnes possibles pour leur accorder un prêt et ne pas se tromper .

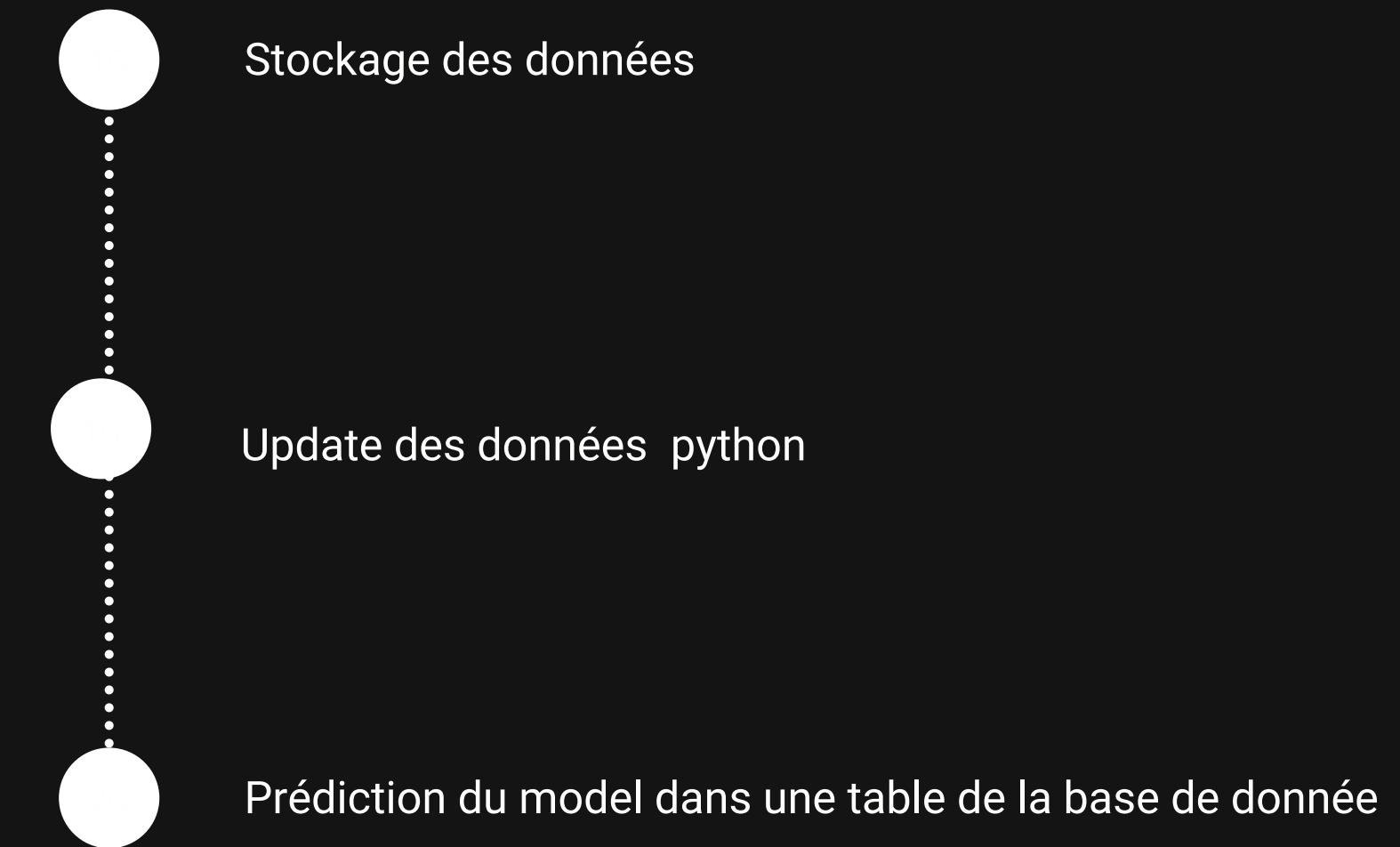
Le projet est donc de réaliser un modèle de machine learning qui nous donneront l'indication de si oui ou non cette personne peut prétendre à un prêt il nous fournira des données qu'il a lui-même collectée pour réaliser notre modèle .

Pourquoi certains prêts bancaires sont refusés ?

La France est placé 23ème sur 42 pays à l'échelle mondiale sur l'inclusion financière . 45% des prêts sont refusés voyons donc comment on peut optimiser ces chances avec mon intelligence artificielle .



11 MySQL Workbench



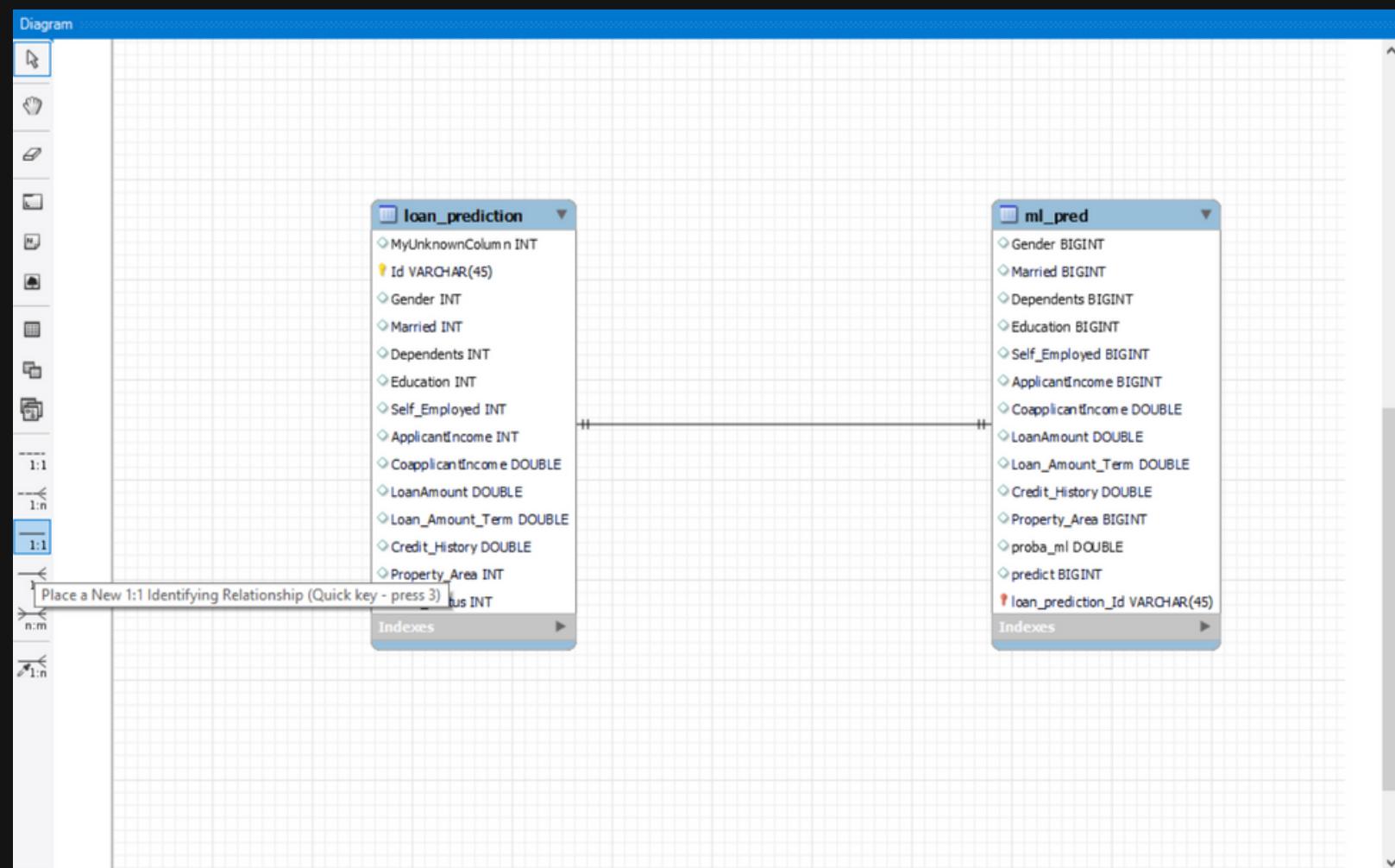
MySQL Workbench

	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	proba_ml	predict
▶	1	1	2	0	0	6250	5654	188	180	1	1	0.8125	1
	1	0	0	1	0	6783	0	130	360	1	1	0.8125	1
	0	0	0	0	0	3159	0	100	360	1	1	0.8125	1
	1	1	0	1	0	3000	1666	100	480	0	2	0.8125	0
	1	1	2	0	0	3283	2035	148	360	1	2	0.8125	1
	1	1	0	0	0	3727	1775	131	360	1	1	0.8125	1
	1	1	2	0	0	3510	4416	243	360	1	0	0.8125	1
	1	1	0	0	0	2083	3150	128	360	1	1	0.8125	1
	0	1	0	0	0	2484	2302	137	360	1	1	0.8125	1
	1	0	0	0	0	3229	2739	110	360	1	2	0.8125	1
	1	1	1	0	0	12841	10968	349	360	1	1	0.8125	1
	1	1	0	0	0	2499	2458	160	360	1	1	0.8125	1

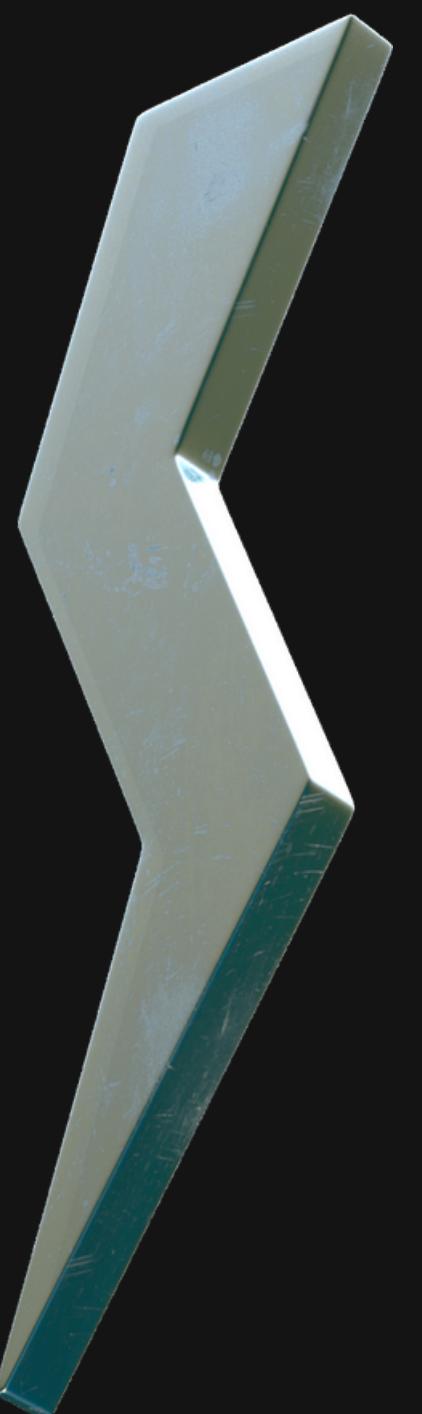
	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
▶	1	1	0	1	0	2000	1900	700	360	0	0	1
	1	1	1	0	0	4583	1508	128	360	1	0	0
	1	1	0	0	1	3000	0	66	360	1	2	1
	1	1	0	1	0	2583	2358	120	360	1	2	1
	1	0	0	0	0	6000	0	141	360	1	2	1
	1	1	2	0	1	5417	4196	267	360	1	2	1
	1	1	0	1	0	2333	1516	95	360	1	2	1
	1	1	3	0	0	3036	2504	158	360	0	1	0
	1	1	2	0	0	4006	1526	168	360	1	2	1
	1	1	1	0	0	12841	10968	349	360	1	1	0
	1	1	2	0	0	3200	700	70	360	1	2	1

MySQL Workbench

Identifying relationship car la seconde table n'a pas de clé primaire



Présentation des données



Présentation des données

Les valeurs nulles



	Nan	%nan
Credit_History	50	8.14
Self_Employed	32	5.21
LoanAmount	22	3.58
Dependents	15	2.44
Loan_Amount_Term	14	2.28
Gender	13	2.12
Married	3	0.49
Loan_ID	0	0.00
Education	0	0.00
ApplicantIncome	0	0.00
CoapplicantIncome	0	0.00
Property_Area	0	0.00
Loan_Status	0	0.00

Présentation des données



Types de colonne :
object , float64 , int64

0	Loan_ID	614	non-null	object
1	Gender	601	non-null	object
2	Married	611	non-null	object
3	Dependents	599	non-null	object
4	Education	614	non-null	object
5	Self_Employed	582	non-null	object
6	ApplicantIncome	614	non-null	int64
7	CoapplicantIncome	614	non-null	float64
8	LoanAmount	592	non-null	float64
9	Loan_Amount_Term	600	non-null	float64
10	Credit_History	564	non-null	float64
11	Property_Area	614	non-null	object
12	Loan_Status	614	non-null	object

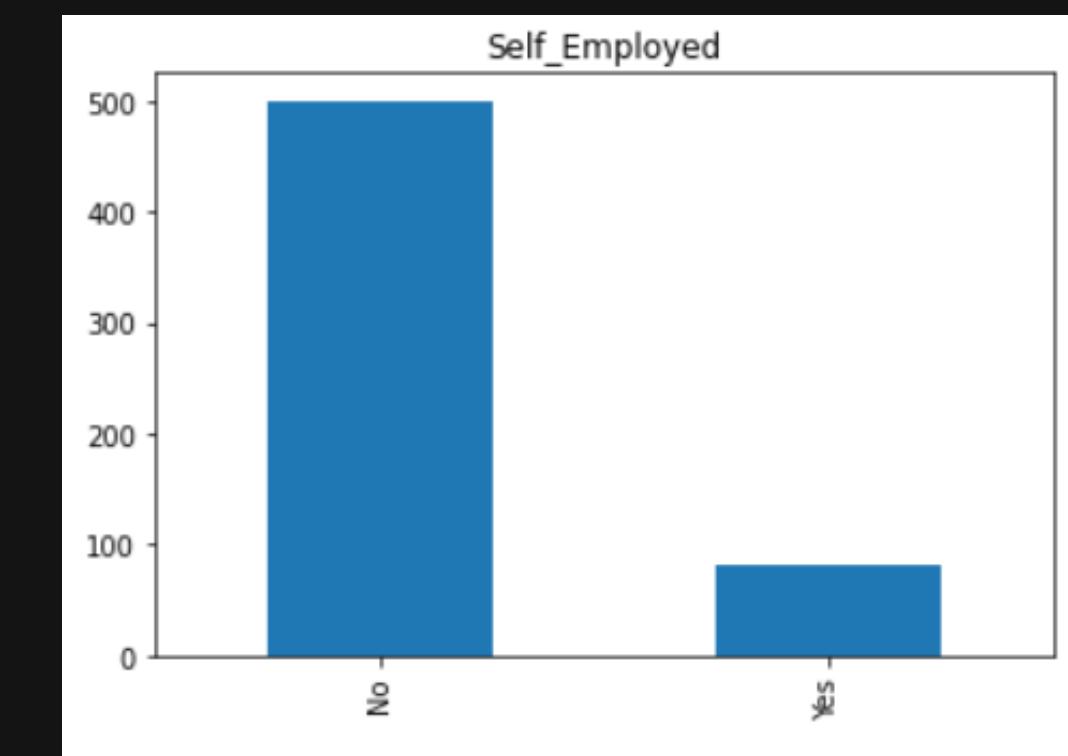
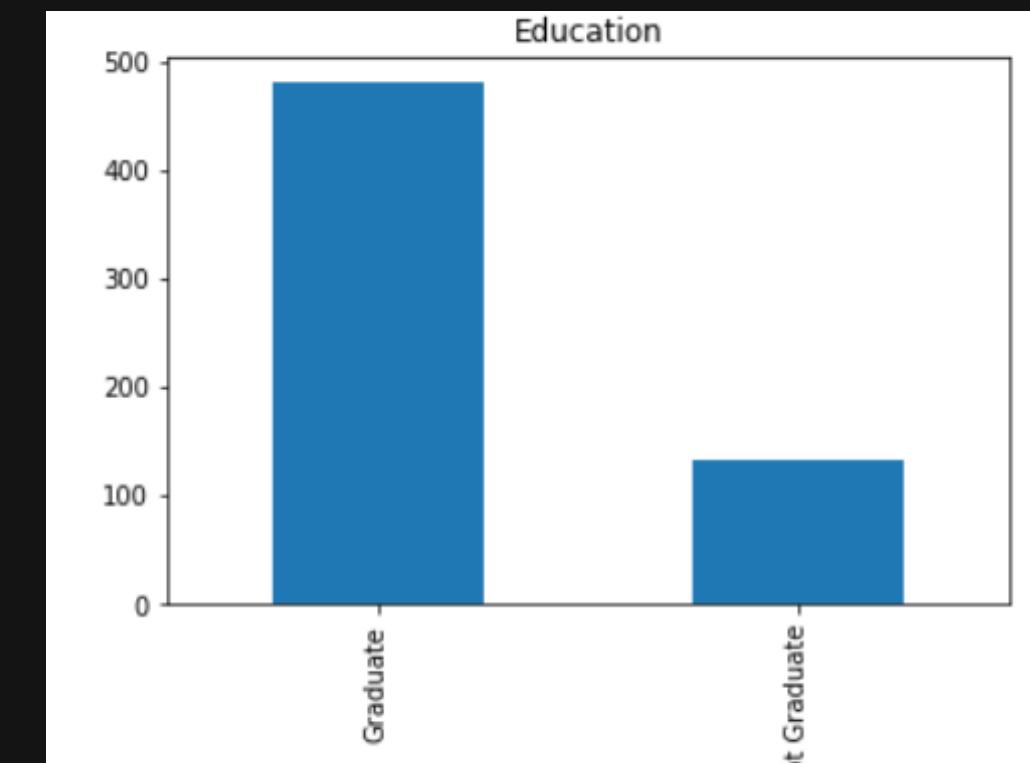
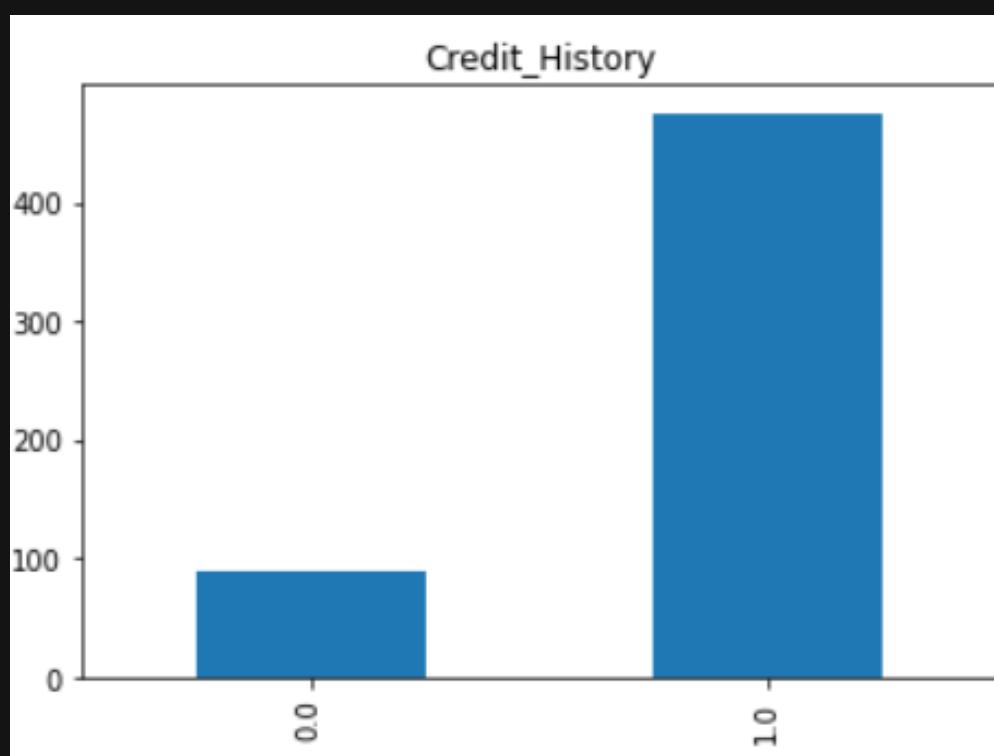
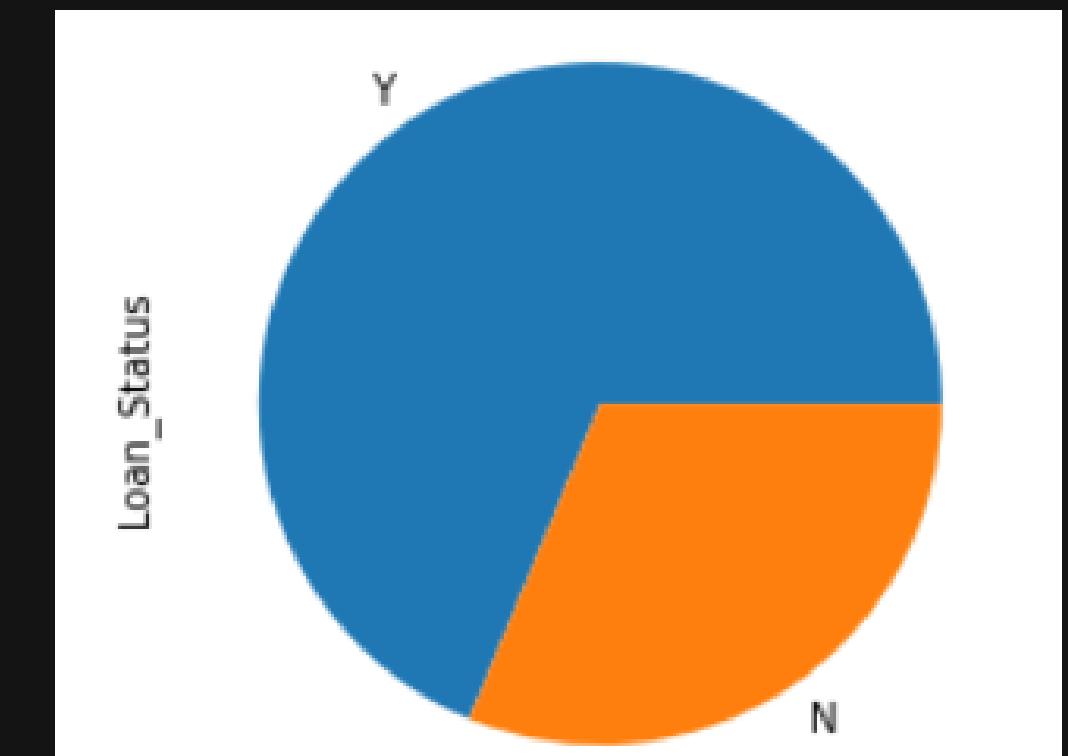
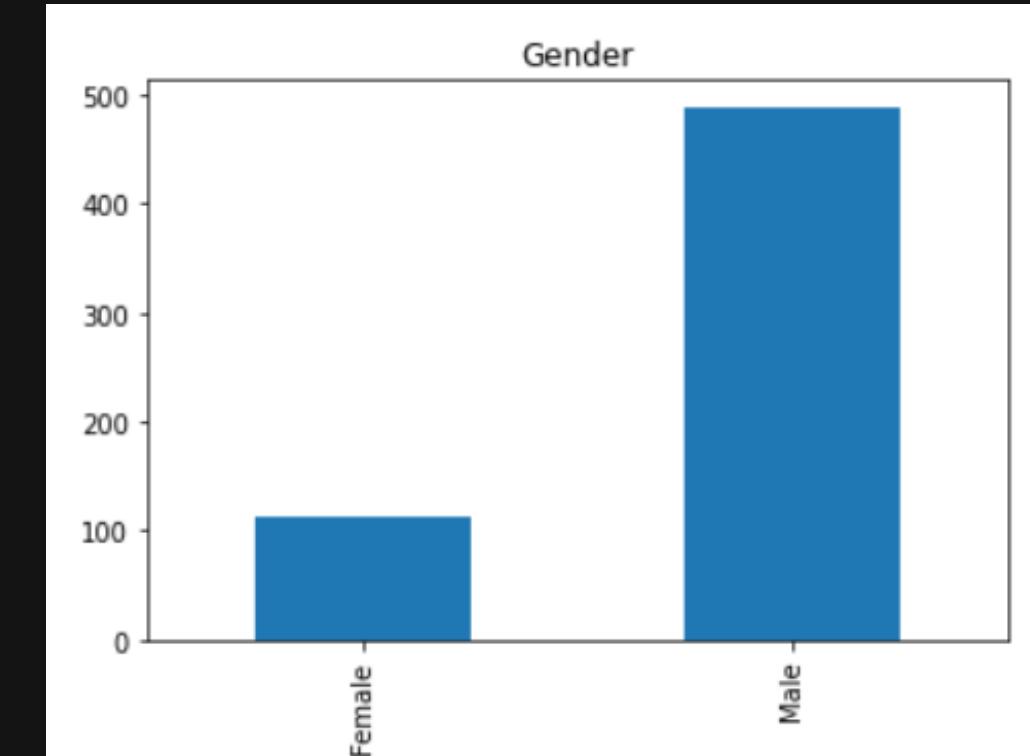
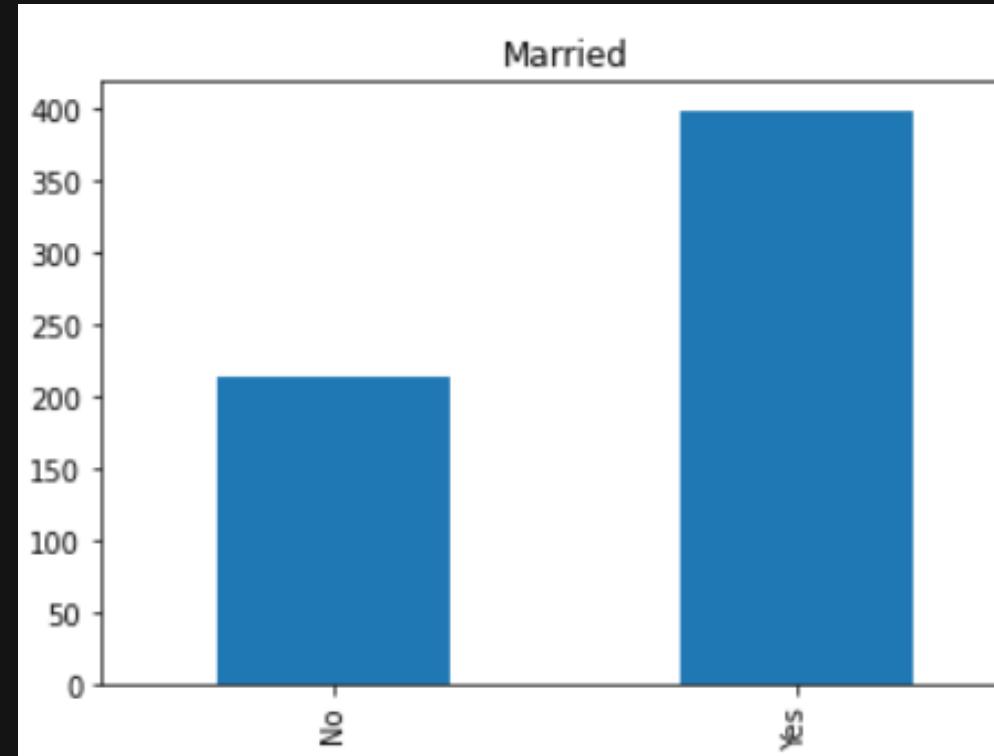
Descriptive of the dataset :				
	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term
count	614.000000	614.000000	592.000000	600.00000
mean	5403.459283	1621.245798	146.412162	342.00000
std	6109.041673	2926.248369	85.587325	65.12041
min	150.000000	0.000000	9.000000	12.00000
25%	2877.500000	0.000000	100.000000	360.00000
50%	3812.500000	1188.500000	128.000000	360.00000
75%	5795.000000	2297.250000	168.000000	360.00000
max	81000.000000	41667.000000	700.000000	480.00000
 Credit_History				
count	564.000000			
mean	0.842199			
std	0.364878			
min	0.000000			
25%	1.000000			
50%	1.000000			
75%	1.000000			
max	1.000000			

Data visualisation

Power Bi présentation



Data visualisation



Power Bi présentation

Genre

- prêt reçue
 - Femme
 - Homme
- prêt refusé
 - Femme
 - Homme

prêt reçue

importance de l'éducation

Importance de l'éducation	Prise en compte
Diplômé	Non
Non diplômé	Oui

Marriage

Statut matrimonial	Prise en compte
Non Marié	Non
Marié	Oui

Travailleur indépendant ou non

Type de travailleur	Prise en compte
Travailleur non indépendant	Non
Travailleur indépendant	Oui

Somme de Credit_History **Somme de Loan_Amount_Term** **ApplicantIncome** **Somme de CoapplicantIncome**

	1,00	360,00	1025	2 773,00
1,00	1,00	120,00	1299	1 086,00
1,00	1,00	360,00	1538	1 425,00
1,00	1,00	360,00	1625	1 803,00
1,00	1,00	360,00	1759	3 541,00
1,00	1,00	360,00	1782	2 232,00
1,00	1,00	360,00	1800	1 213,00
1,00	1,00	360,00	1809	1 868,00
1,00	1,00	360,00	1820	1 719,00
1,00	1,00	360,00	1863	1 041,00
1,00	1,00	360,00	1875	1 875,00
272,00	94 584,00		458 203,92	

Endroit d'habitations

Endroit d'habitation	Nombre de prêts
Urbain	100
Semi Urbain	80
Rural	60

Type de dépendence

Type de dépendance	Nombre de prêts
Indépendants	100
Dépendants... (1)	80
Dépendants (1)	60
Dépendants... (1)	40

Power Bi présentation

Genre

- prêt reçue
 - Femme
 - Homme
- prêt refusé
 - Femme
 - Homme

prêt reçue

Importance de l'éducation

Category	Value
Diplômé	~25%
Non diplômé	~75%

Marriage

Category	Value
Marié	~25%
Non Marié	~75%

Travailleur indépendant ou non

Category	Value
Travailleur non indé...	~25%
Travailleur indép...	~75%

Somme de Credit_History **Somme de Loan_Amount_Term** **ApplicantIncome** **Somme de CoapplicantIncome**

	1,00	480,00	645	3 683,00
1,00	480,00	1811	1 666,00	
1,00	360,00	1928	1 644,00	
1,00	360,00	1963	0,00	
1,00	480,00	2031	1 632,00	
0,00	360,00	2137	8 980,00	
1,00	360,00	2165	0,00	
1,00	360,00	2213	0,00	
1,00	360,00	2330	4 486,00	
1,00	360,00	2423	505,00	
1,00	360,00	2484	2 302,00	
53,00	19 224,00	40 205,00		

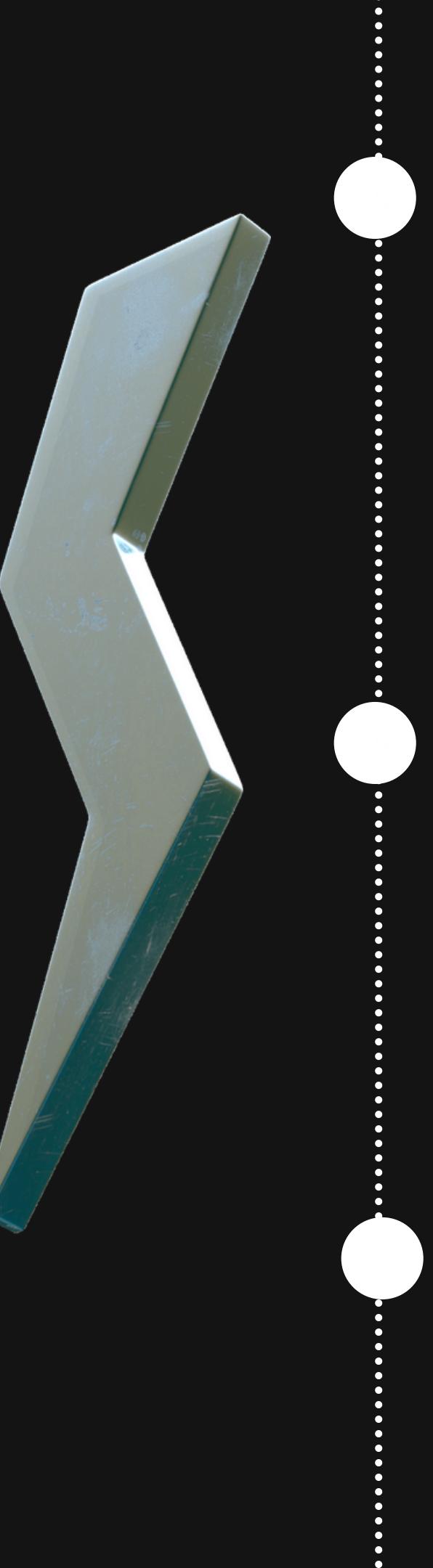
Endroit d'habitations

Endroit d'habitations	Count
Urbain	~30
Semi Urbain	~12
Rural	~10

Type de dépendence

Type de dépendence	Count
Indépendants	~45
Dépendants	~5
Dépendants...	~1
Dépendants...	~1

Transformation des données



LabelEncoder

- Les valeurs seront transformée en donnée numérique

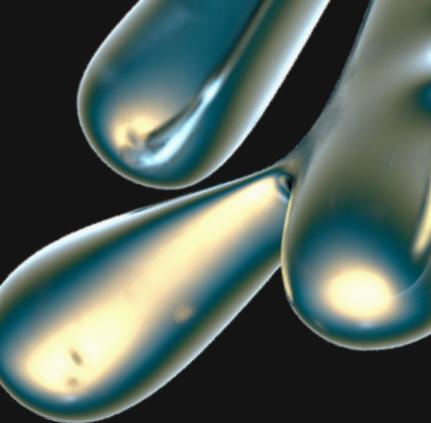
Standard Scaler

- Le standard scaler permet de changer l'intervalle des valeurs

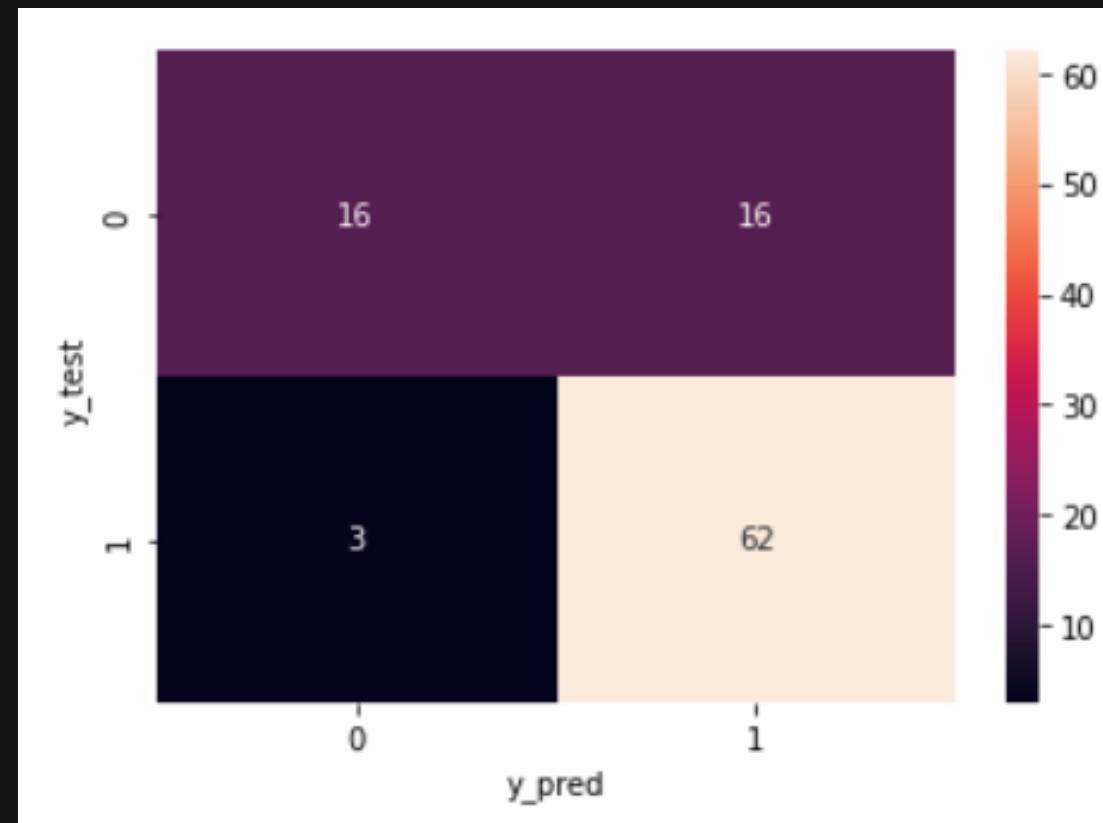
Smote

- Permet d'équilibré les données de la target

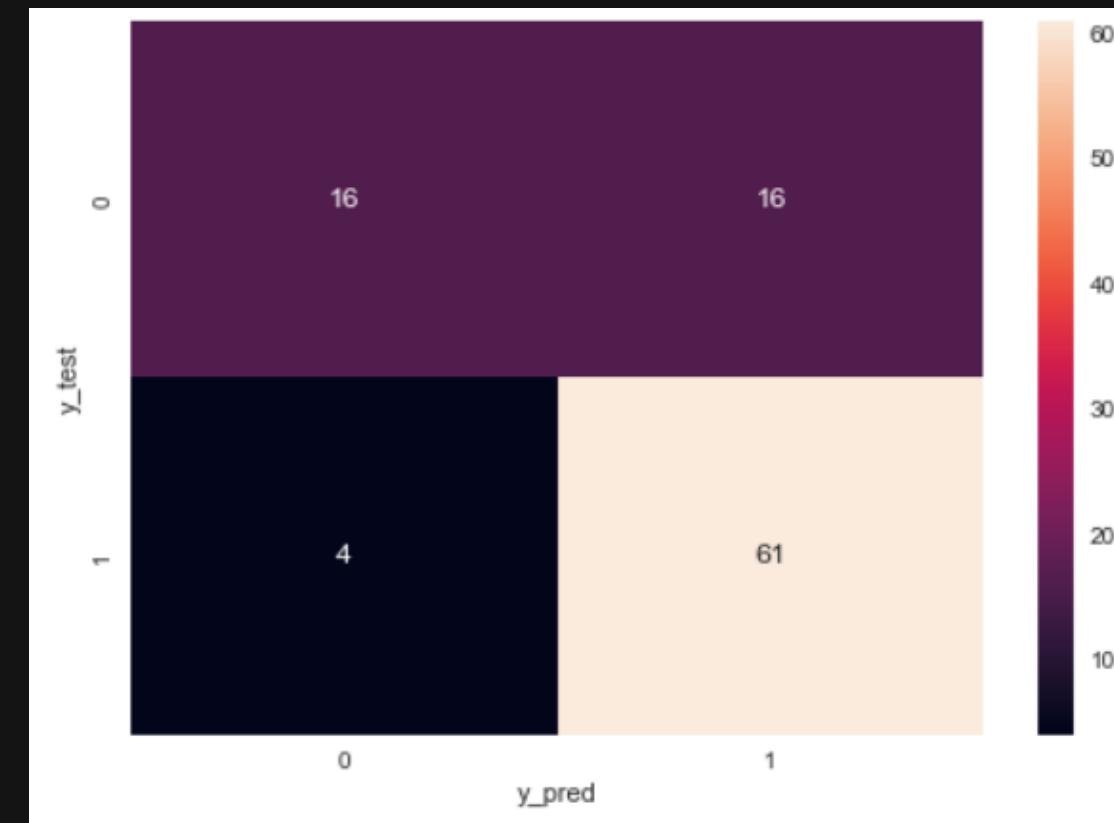
Les différent modèles



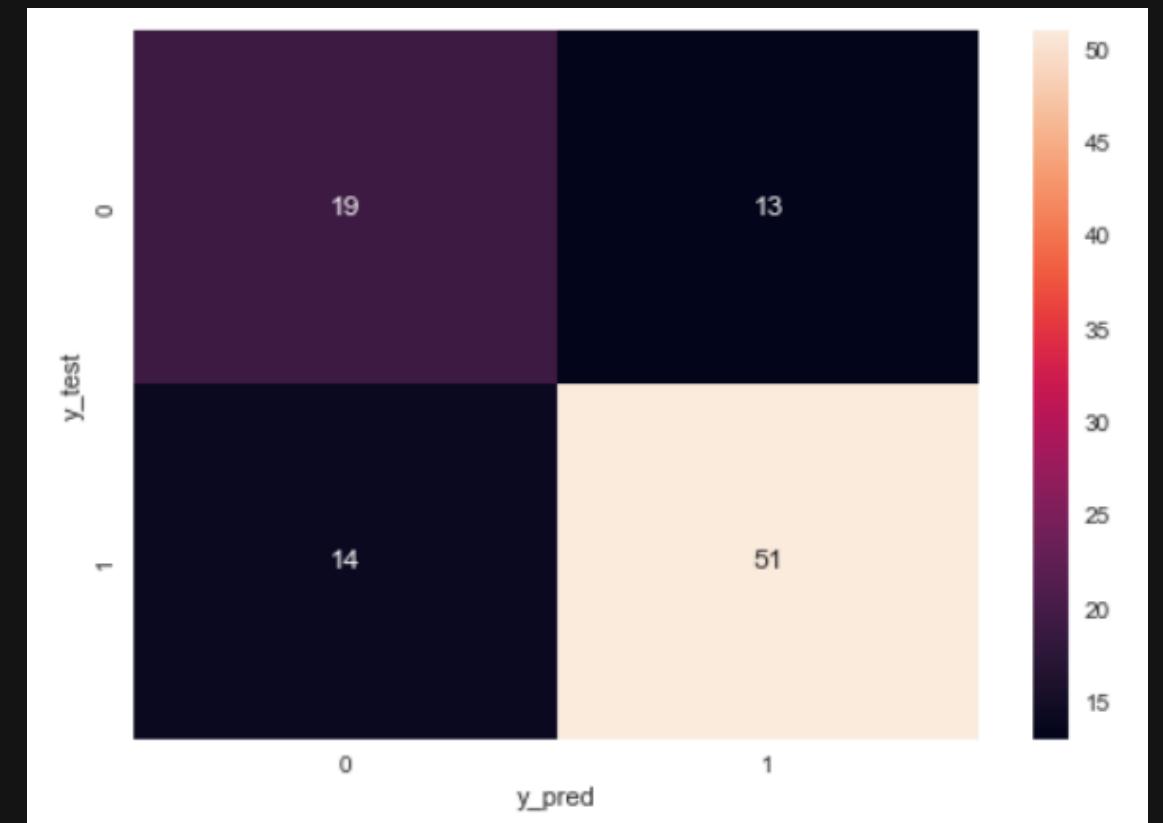
LOGISTIC REGRESSION



K NEAREST NEIGHBORS

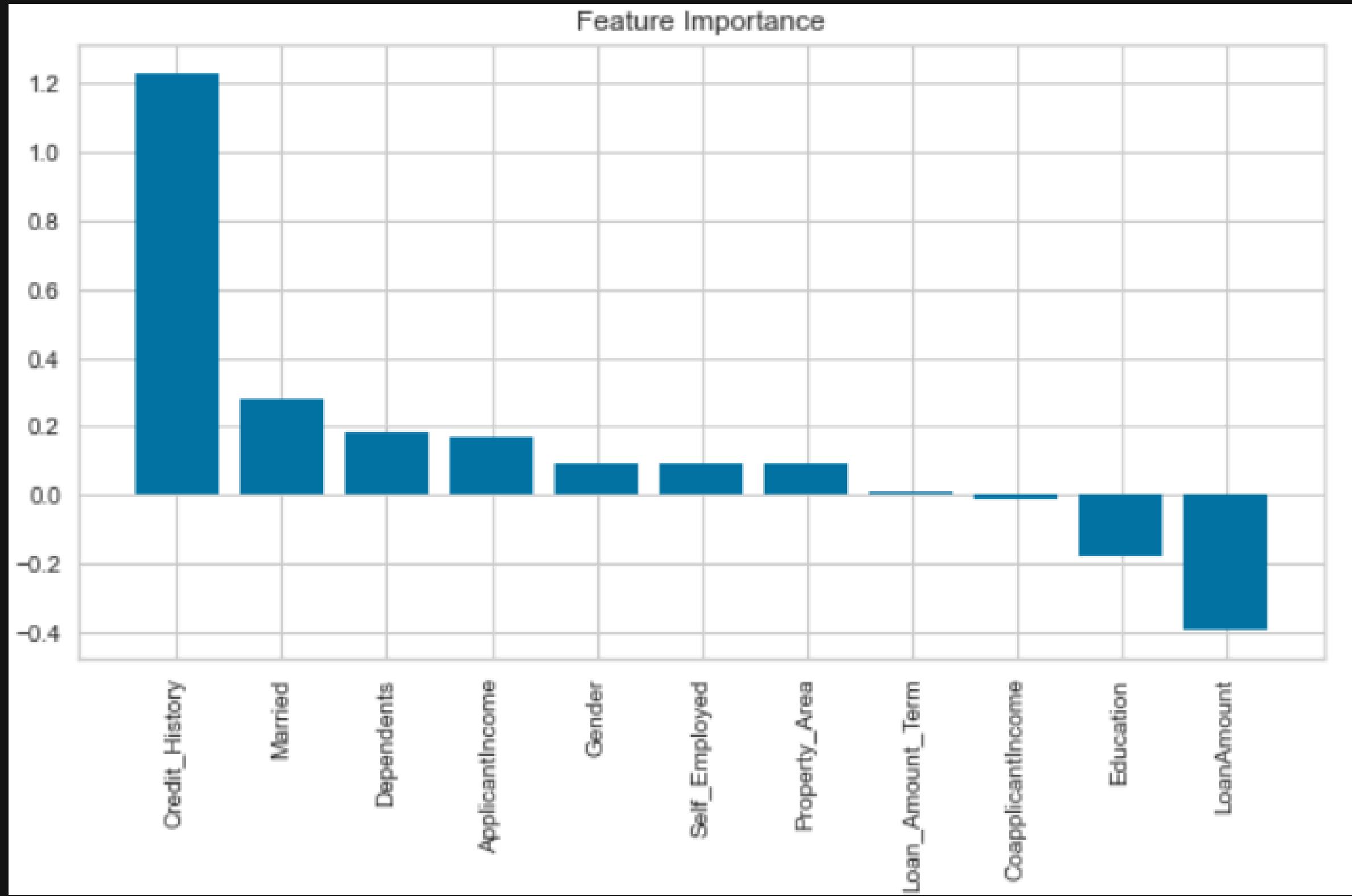
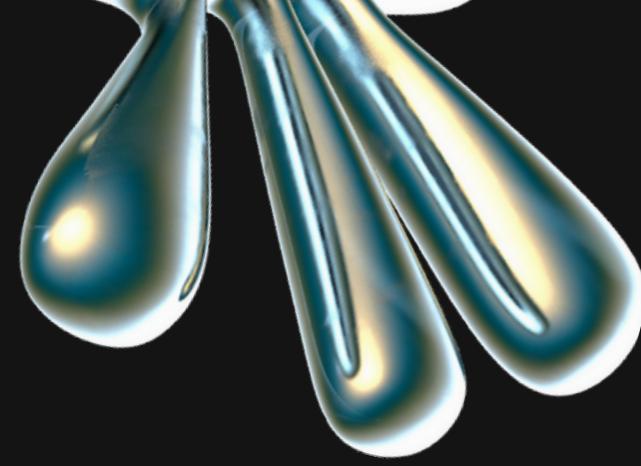


DECISION TREE
CLASSIFIER

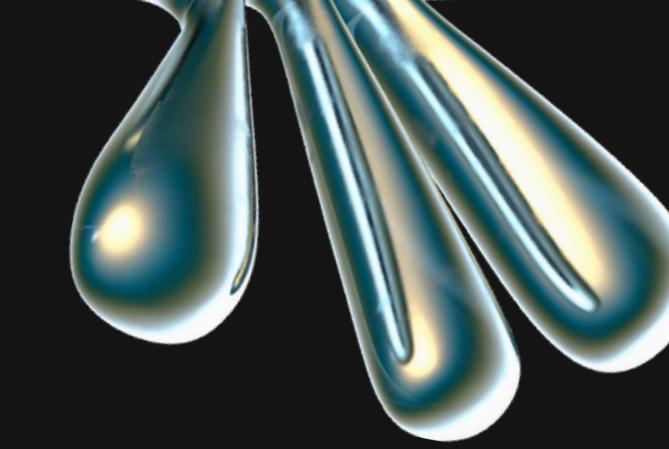


Feature Importance

Logistic Regression



Metrics



Accuracy , recall , f1 score

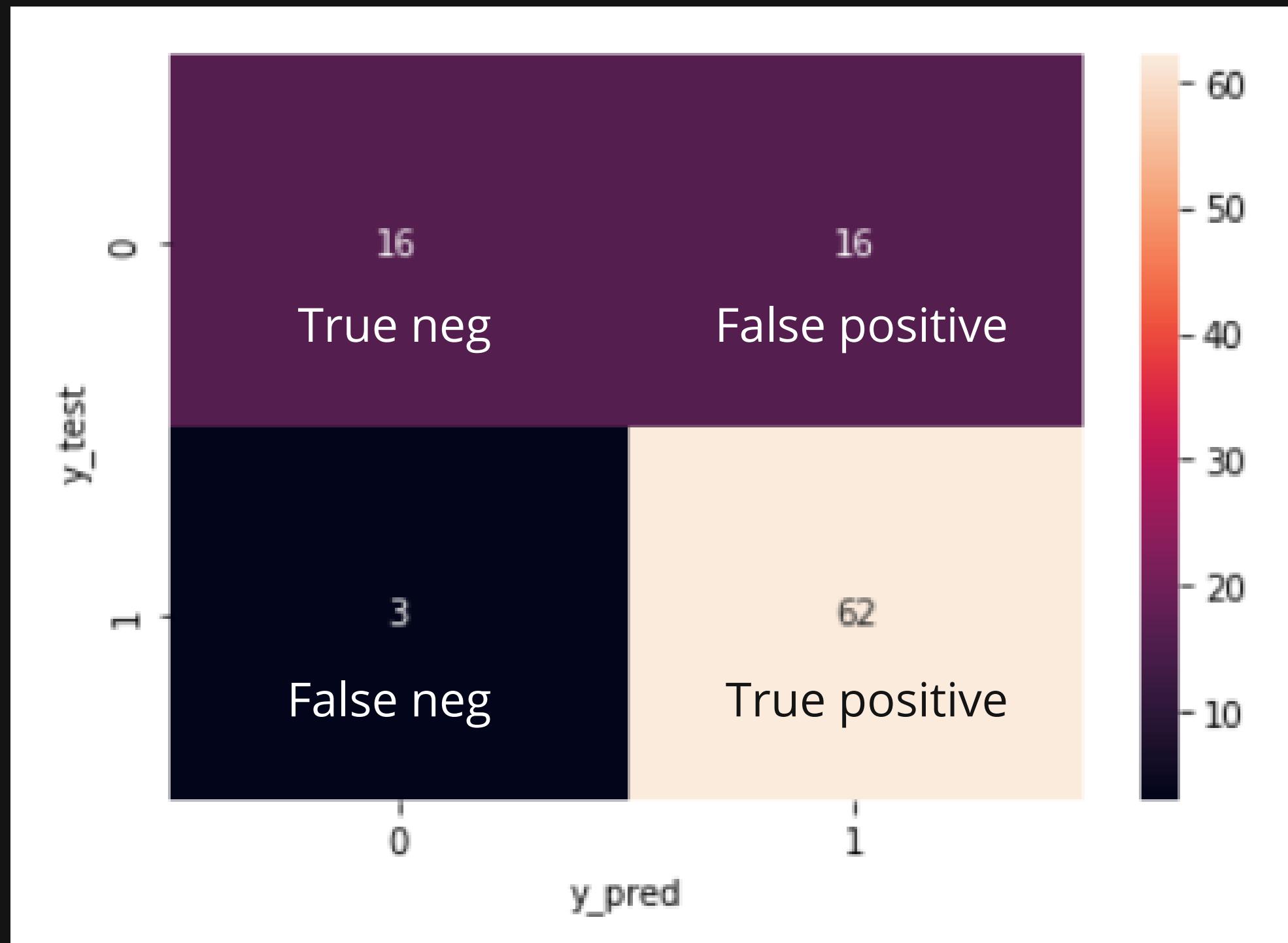
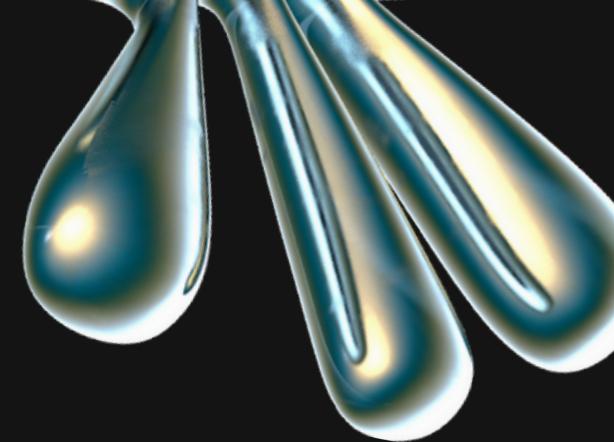
accuracy train : 0.745

accuracy test : 0.732

	precision	recall	f1-score	support
0	0.60	0.56	0.58	32
1	0.79	0.82	0.80	65
accuracy			0.73	97
macro avg	0.70	0.69	0.69	97
weighted avg	0.73	0.73	0.73	97

Confusion Matrix

Logistic Regression



Deployment

Heroku

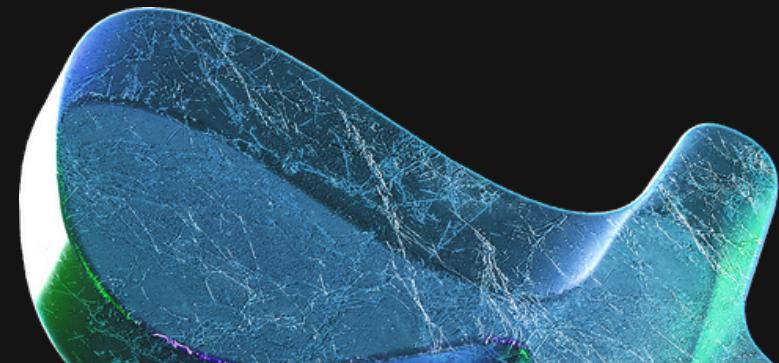
Utilisation de Heroku permettant de déployer a partir de github et également en local

Streamlit

Utilisation de la librairie python du nom de streamlit pour crée l'application tester en local puis en ligne

Github

Github utilisée comme backup et pour le déploiement de mon application .

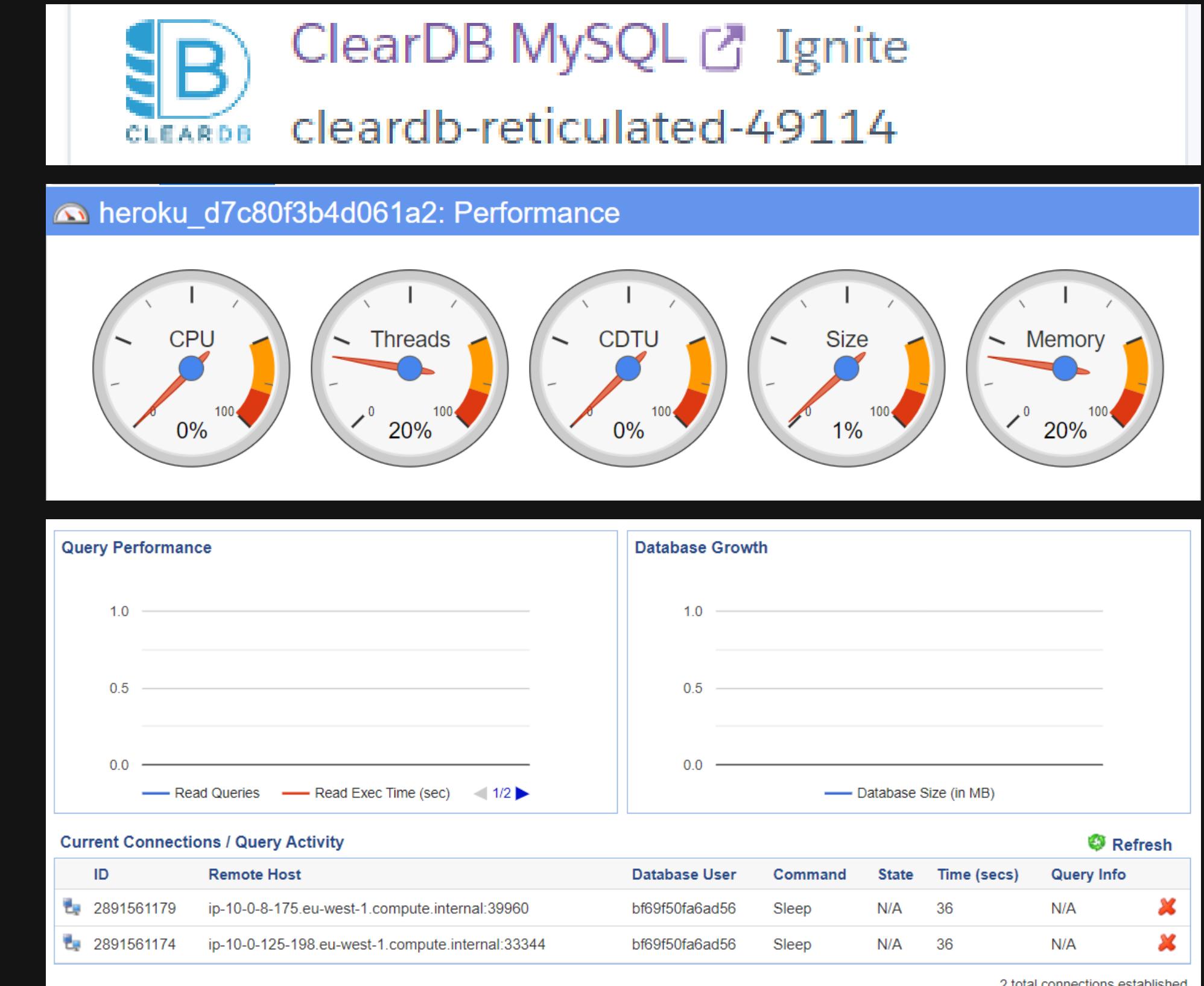
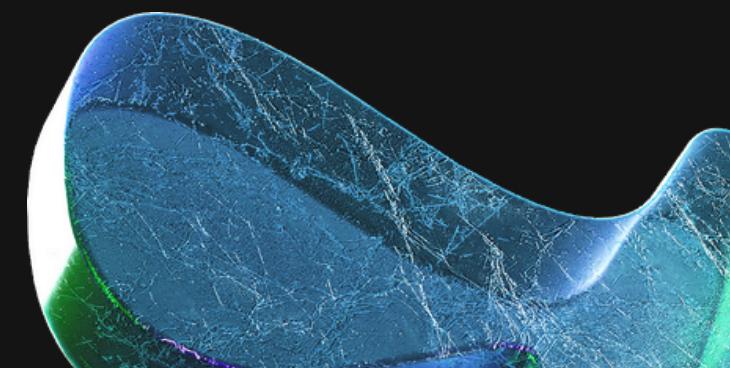


Deployment

Heroku add-on

Utilisation de heroku add-on se qui me permet d'ajouter la possibilité d'utilisée ClearDB pour connecter ma base de donnée à l'application heroku

C12/C7/C12/C13



Deployment

EdenLecarpentier / loan_prediction_file_rouge Public

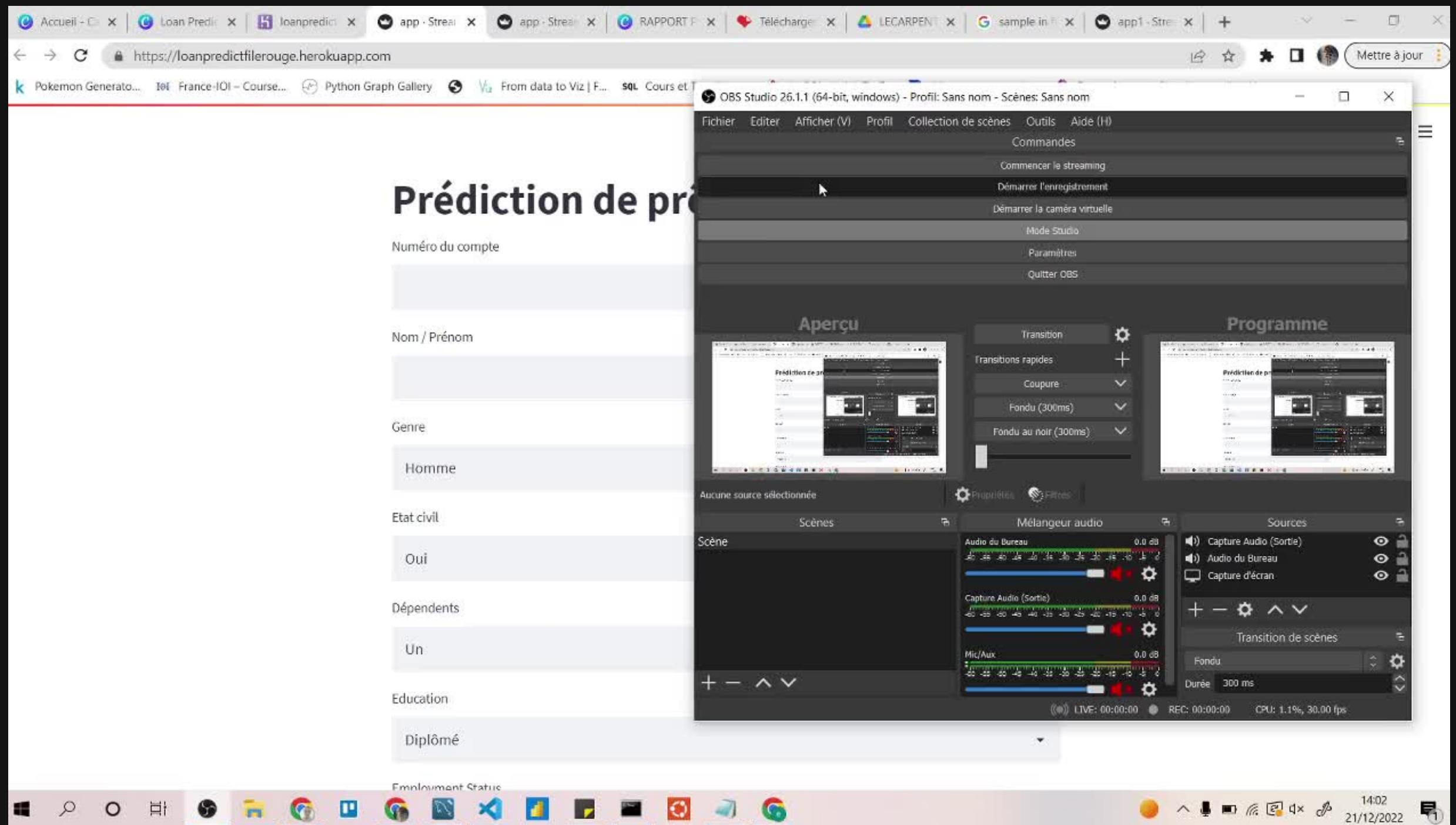
Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main ▾ 1 branch 0 tags Go to file Add file ▾ < Code ▾

EdenLecarpentier e2	91b9f38 3 days ago	95 commits
__pycache__	rapport	29 days ago
mlflow	e2	3 days ago
notebook	e2	3 days ago
pickle_model	deploying	last month
power_bi_visualisation	e2	3 days ago
rapport	powerbipush	15 days ago
Procfile	deploying	last month
app.py	deployment	29 days ago
model_pk1.pickle	e2	3 days ago
requirements.txt	deploying	last month
runtime.txt	r	last month
setup.sh	main	last month

c12/c7/c12/c13

Deployment



c12/c7/c12/c13

MLFlow Tracking

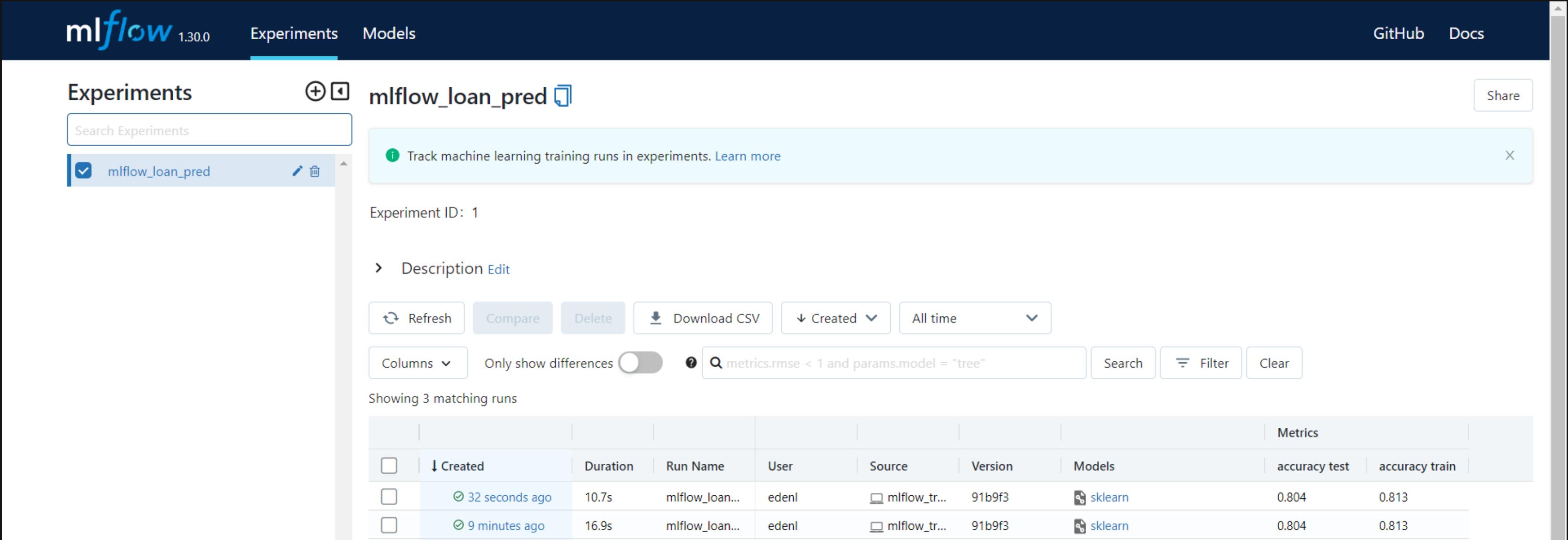
Tracking

- Utilisation de MLFlow tracking pour vérifier si le modèle continue de toujours bien performé
- Le modèle sera arrêté si il régresse de trop
- Utilisation en local et non sur l'application

Heroku

MLFlow tracking a ensuite était connecté à l'application Heroku pour tracker si le modèle de l'application était toujours aussi performant

MLFlow Tracking



The screenshot shows the MLFlow Tracking interface. At the top, there is a navigation bar with the MLFlow logo (1.30.0), Experiments, Models, GitHub, and Docs links. The main area is titled "Experiments" and shows a list of experiments. One experiment, "mlflow_loan_pred", is selected and highlighted in blue. A tooltip for this experiment states: "Track machine learning training runs in experiments. Learn more". Below the experiment name, there is a search bar labeled "Search Experiments" and a list of experiments. The selected experiment has a "Description Edit" button, a "Refresh" button, a "Compare" button, a "Delete" button, a "Download CSV" button, and dropdown menus for "Created" and "All time". There are also "Columns", "Only show differences", and search/filter/clear buttons. The table below shows "Showing 3 matching runs" with columns: Created, Duration, Run Name, User, Source, Version, Models, accuracy test, and accuracy train. The first run was created 32 seconds ago, took 10.7s, and has metrics of 0.804 and 0.813. The second run was created 9 minutes ago, took 16.9s, and has metrics of 0.804 and 0.813.

	Created	Duration	Run Name	User	Source	Version	Models	accuracy test	accuracy train
<input type="checkbox"/>	32 seconds ago	10.7s	mlflow_loan...	edenl	mlflow_tr...	91b9f3	sklearn	0.804	0.813
<input type="checkbox"/>	9 minutes ago	16.9s	mlflow_loan...	edenl	mlflow_tr...	91b9f3	sklearn	0.804	0.813

Organisation

The screenshot shows a digital workspace interface with a rocky landscape background. The top navigation bar includes icons for 'Visible par l'espace de travail' (Visible by workspace), 'Tableau' (Tableau), 'Power-ups', 'Automatisation', 'Filtre', 'Partager', and 'Ajouter une autre liste' (Add another list). The main area is organized into a timeline from July to December, with each month containing a list of tasks:

- Juillet**:
 - importer les données dans python 5 min
 - Analyser et nettoyer les données en python 1heure
 - visulisation des données en python 1 heure
- Aout**:
 - préparation des données pour l'entraînement du model : séparation des données x et y , preprocessing , Standardizing
 - création de la pipeline et choix des models a utilisée 20 minutes
 - Evaluation des models 30 minutes
- Septembre**:
 - choisir le meilleur model qui sera utilisé pour l'Api / Application 20 minutes
 - début du rapport 2 heures
 - exporter le model comme pickle pour l'Api/Application 5 minutes
 - Conception de la base de donnée sql avec les données du model 2 heures
 - MDP , relations entre les tables 10 minutes
 - backup des données de sql bank loan 15 minutes
- octobre**:
 - amélioré le rapport corriger les fautes 2 heures
 - coder l'application : import des données , dropdown list avec les différents features , import du model 5heures
 - Veille de 5 pages 2heures
- Novembre**:
 - MLFlow 1 heure
- Décembre**:
 - Rapport 3 heures
 - Power BI 3 heures
 - Diapo 3 heures

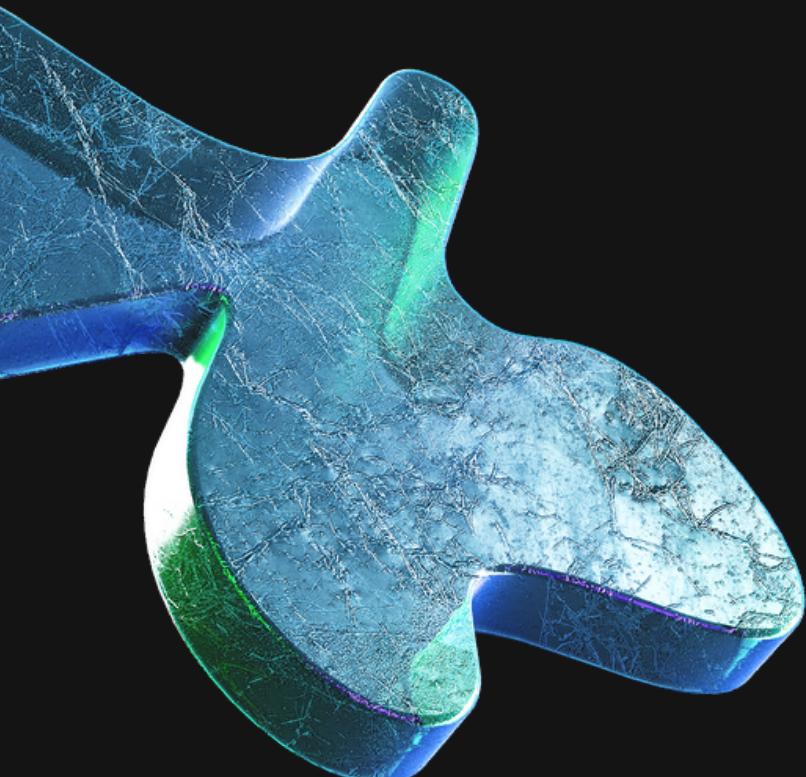
Each task card includes a 'Ajouter une carte' (Add a card) button.

Conclusion

- Grand manque de donné se qui par défaut entraîneras des résultats peu exploitable
- Donnée non variée
- Donnée non balancé target possédant plus de prêt homme que femme
- Résultat moyen suite au différentes constatation

Slide annexe sommaire

- Oversampling / Undersampling
- Underfitting / Overfitting
- Les différents modèles

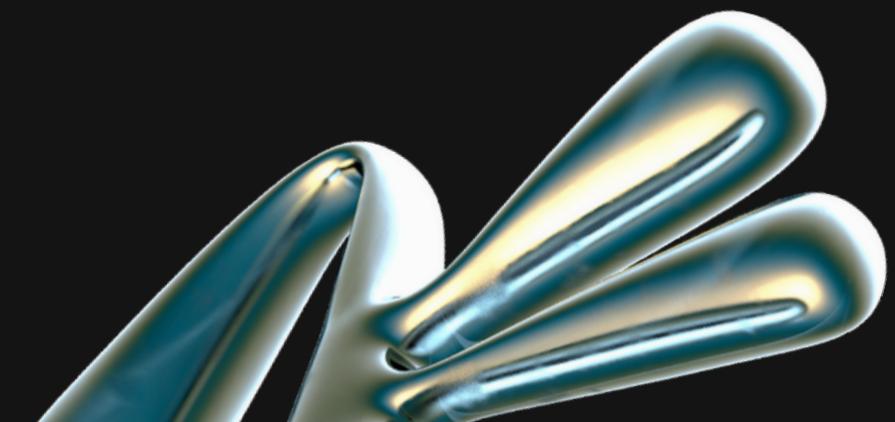


Oversampling

- Provoquer par des données non balancée
- Les différentes méthodes vont crée de nouvelle donnée dans la class minoritaire dans notre cas 'Les fraudes bancaires'
- Efficace individuellement mieux si on utilise l'Undersampling avec
- 1 méthode principal : Random Oversampling

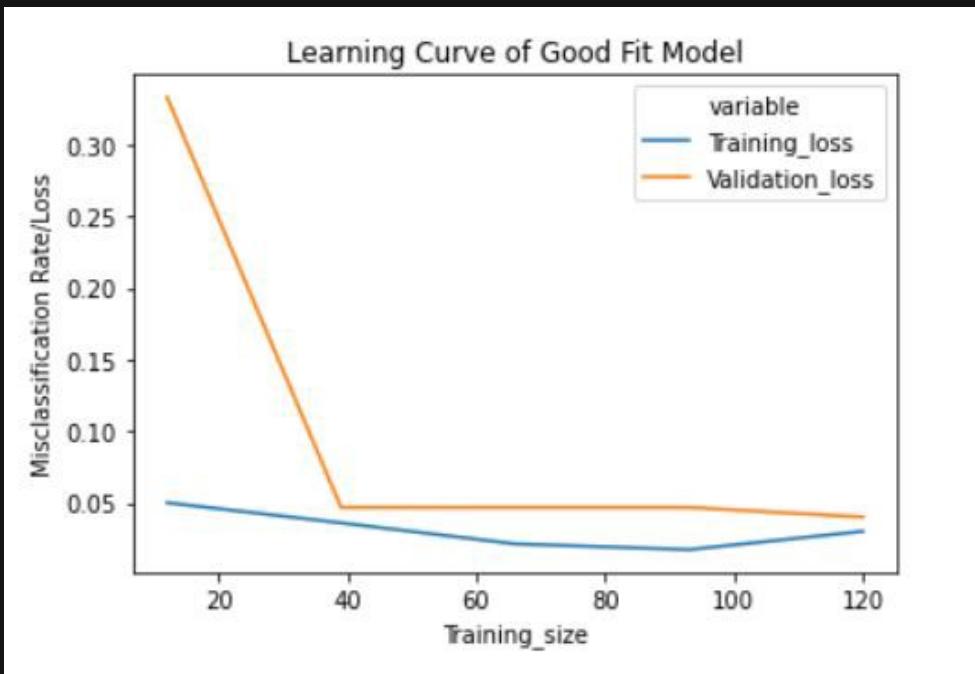
Undersampling

- Provoquer par des données non balancée
- Les différentes méthode vont supprimée ou fusionner des exemples dans la class majoritaire dans notre cas 'Les transaction bancaires'
- Efficace individuellement mieux si on utilise l'Oversampling avec
- 1 méthode principal : Random Oversampling



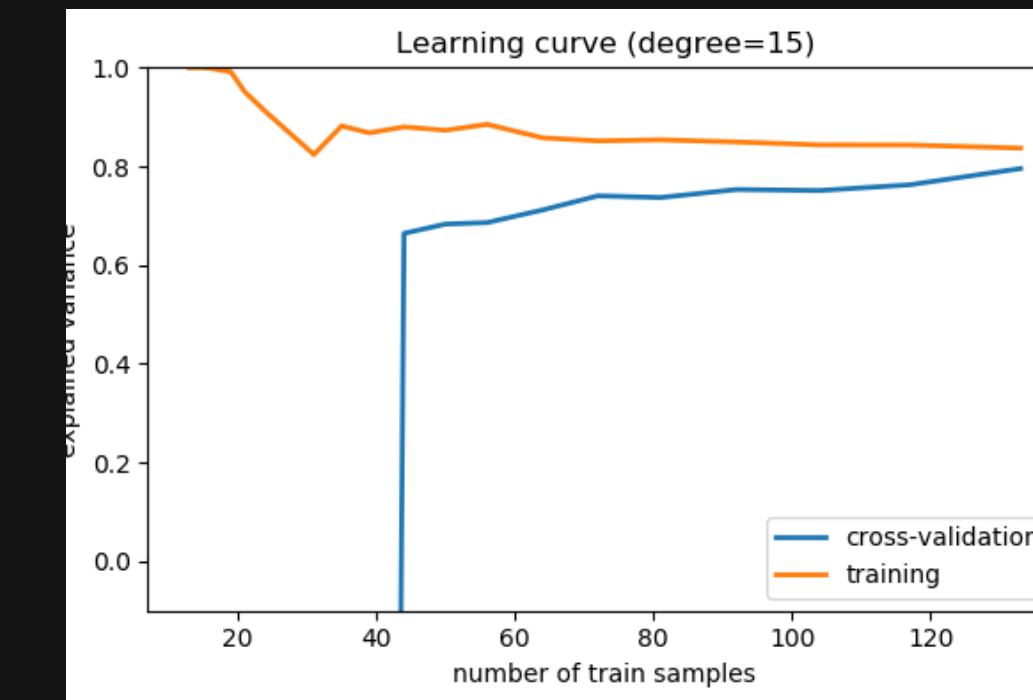
Underfitting

- Utilisation de la Learning Curve pour l'identifier
- Les données se sont trop peu entraîner la barre de train fera donc une chute
- Les données ne se sont pas assez entraîner ou très mal entraîner



Overfitting

- Utilisation de la Learning Curve pour l'identifier
- Les données se sont trop bien entraîner et la barre du train sur la Learning Curve est anormalement haute
- Les données sont trop entraîner pour autant ne donne pas des modèles vraiment performant



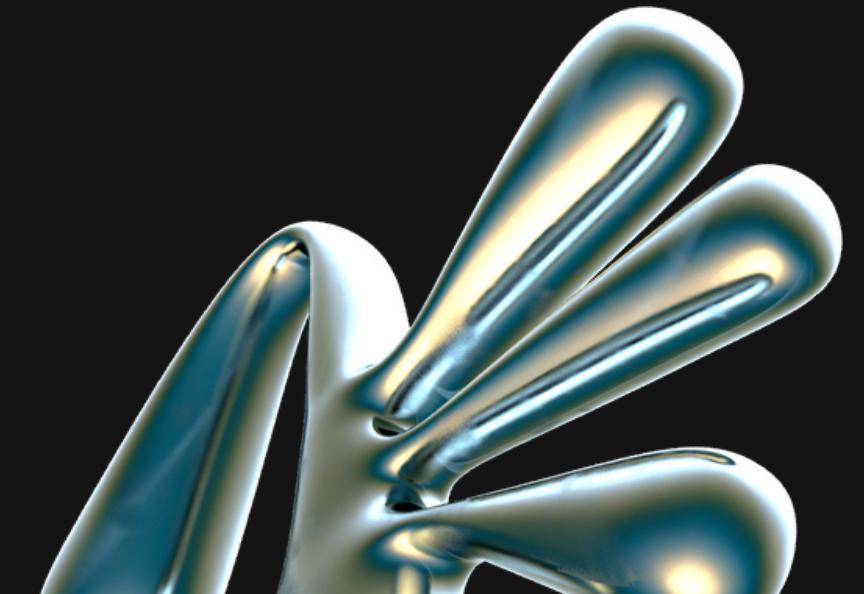
Déploiement

Procfile

- Fichier qui permet de spécifier les commandes exécutées
- Permet plus de possibilité de configuration
- Exécute le fichier app.py

Setup.sh

- Permet la configuration pour que tout soit compris lors du déploiement par Heroku



Hyperparamètre de modèles

Logistic Regression

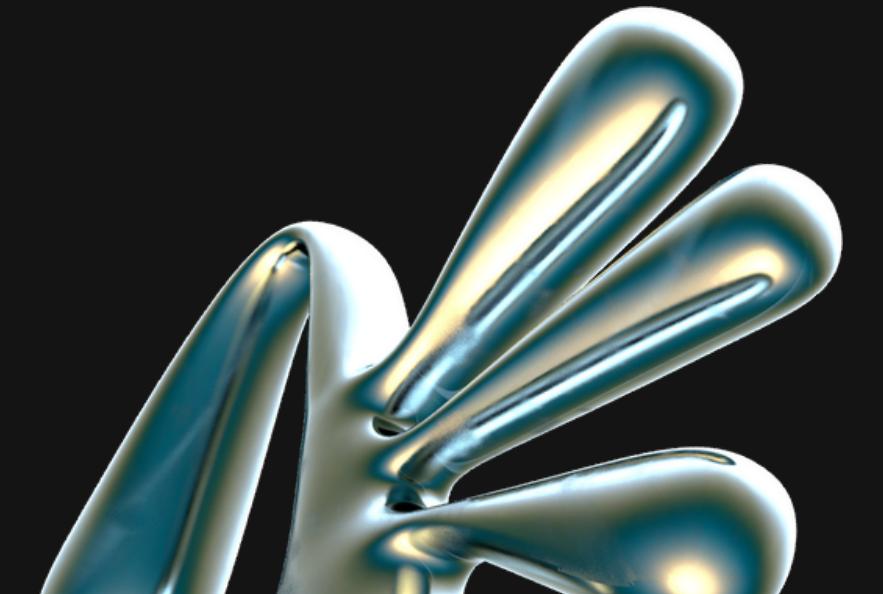
- penalty
- c
- solver

K Nearest Neighbors

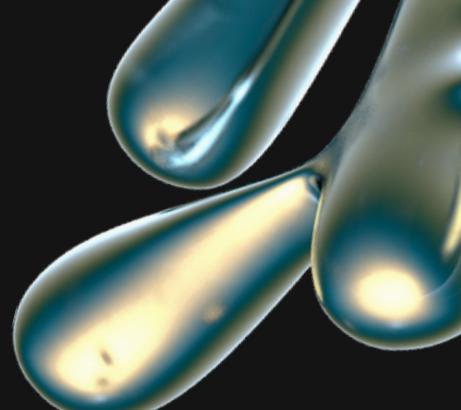
- n_neighbors
- weights
- metric

Decision Tree Classifier

- max_depth
- criterion
- min_samples_leaf

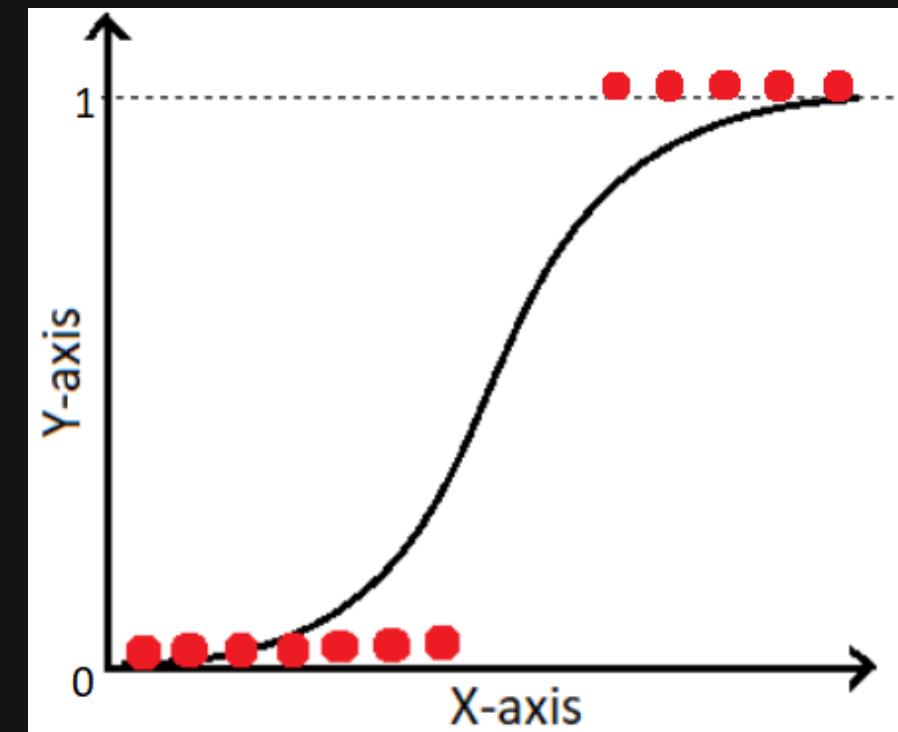


Explication des différents modèles



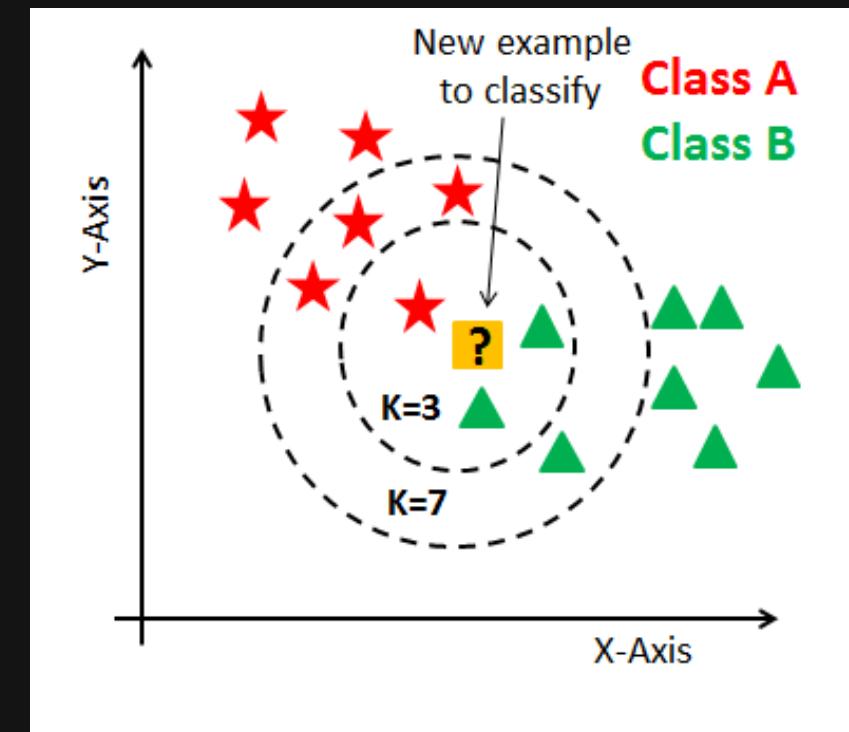
LOGISTIC REGRESSION

- Modèle de classification
- Analyse de toutes les différentes features
- Vérification des corrélations entre les données



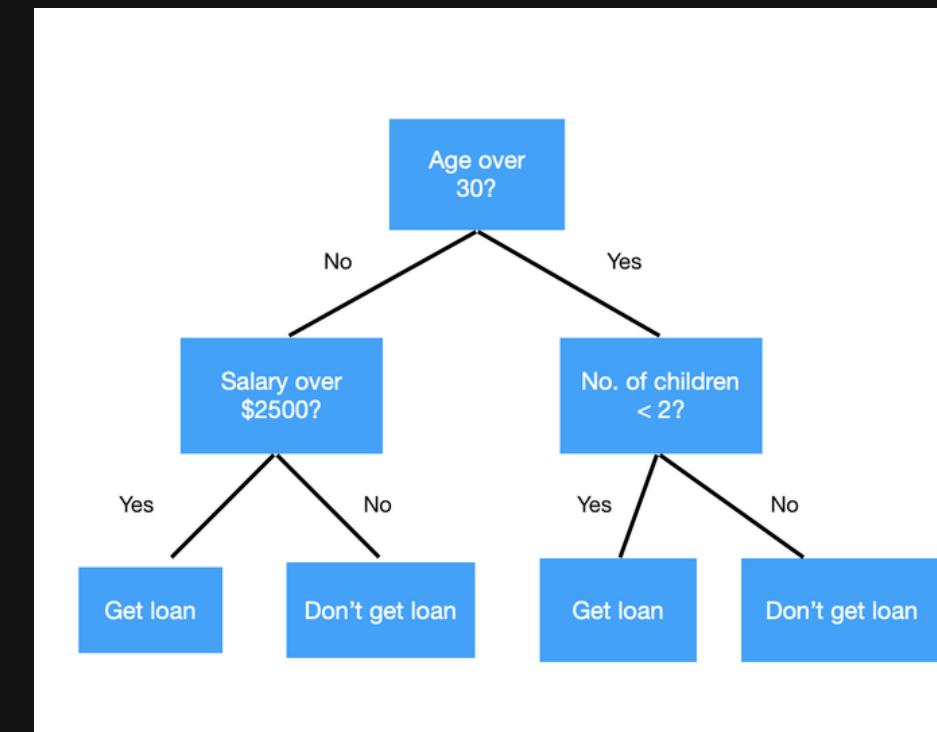
K NEAREST NEIGHBORS

- Modèle de classification
- Trouve les groupes de données
- Analyse les points de données autour des groupes
- Les points de données sont ajouté au groupe



DECISION TREE CLASSIFIER

- Modèle de classification
- Fonctionne comme un arbre
- Toutes les branches possibles sont analysées pour trouver toutes les corrélations entre les données



Classification report

F1 SCORE

- C'est la moyenne de la précision et du recall
- Utilisation peut recommandé sur un dataset non balancée
- Utilisée comme comparaison .

PRECISION

- La précision doit se rapprocher le plus possible de 1 sans l'atteindre
- L'accuracy permet de déterminer que les cas positif .
- Indique le pourcentages de prédiction correct

SUPPORT

- C'est le nombre d'occurrence qu'il y a eux dans chaque class

MACRO AVG

- Moyenne de la précision de toutes les classes

RECALL

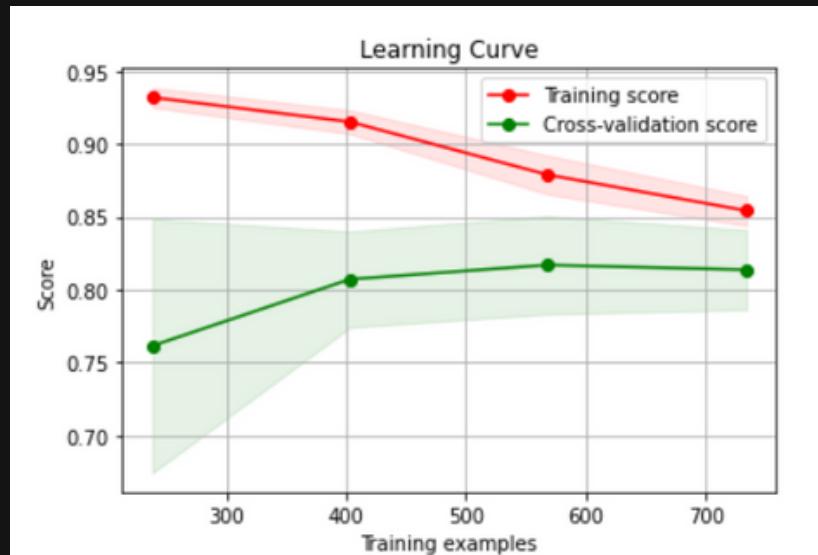
- le 0 devras être le plus proche possible de 0
- le 1 devras être le plus proche possible du 1 sans devenir 1
- Le recall permet d'identifier comment un modèle a identifié tout les cas positif dans les données (True positive)

WEIGHTED AVG

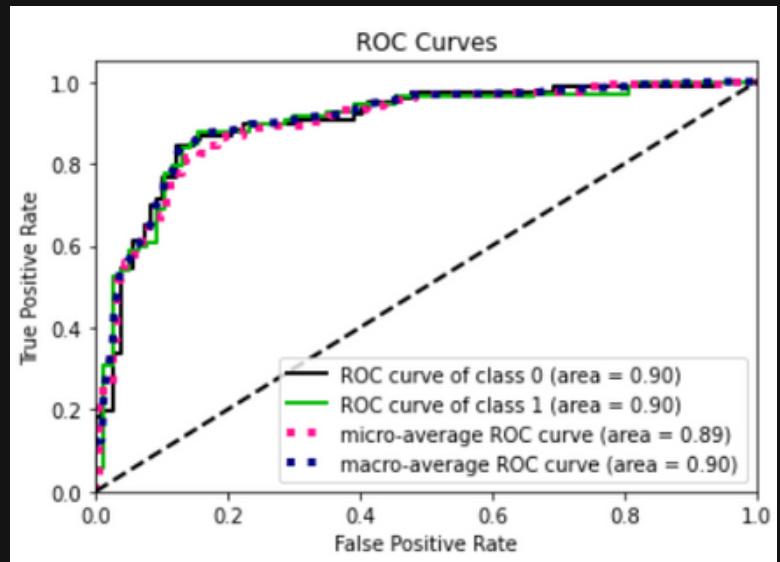
- Toutes les classes qui contribue à la moyenne sont pessée par leur taille

Explication des différents graphiques

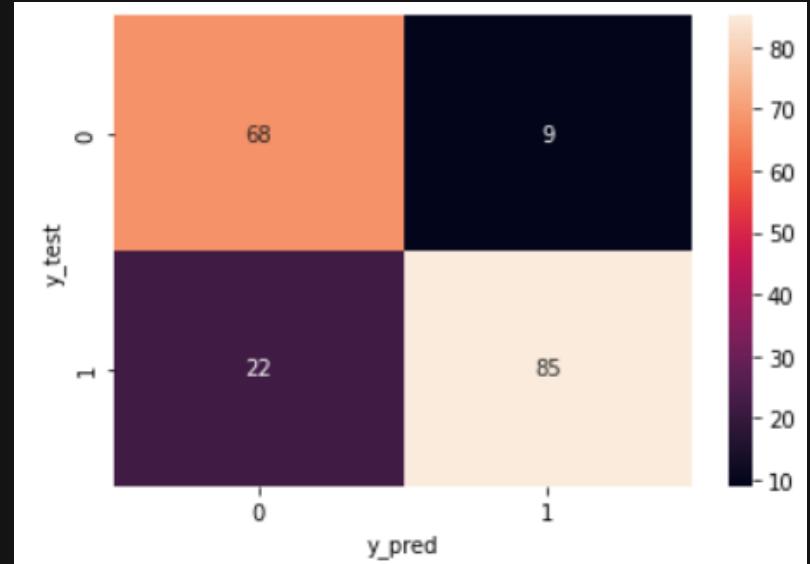
LearningCurve



ROC Curve



Confusion Matrix



- Montre les relations qu'il peut y avoir entre le résultat du train et la cross validation test score
- Permet de démontré : si il y a suffisament , trop ou pas suffisament de données .
- Le modèle était sensible au erreur en comparant le taux d'erreur possible
- l'axe y montre le score de l'entraînement
- l'axe x montre le temp

- La ROC Curve montre les true positive sur l'axe y et sur l'axe x nous montre les False Positive .
- Les lignes plus elle se dirigeront en haut a gauche mieux sa seras .
 - False positive axis X
 - True positive axis Y

- Démontre les True Positive , True negative , False Positive , False negative

Equilibrée le dataframe

SMOTE

- Performe de la data augmentation en crée des point de donnée synthétic basée sur les données originel
- Pour cela on auras besoin d'indentifié Le k Nearest Neighbors
- On prendras ensuite un de ces neighborset il seras décalée très légèrement de la donnée originelle

Pipeline

Sklearn pipeline

- Permet un code plus simple et facile à comprendre
- Possibilité d'effectuer plusieurs transformations en un seul bloc de code
- Moins de code pour un résultat
- Débugger plus facilement

Relation MySQLWorkbench



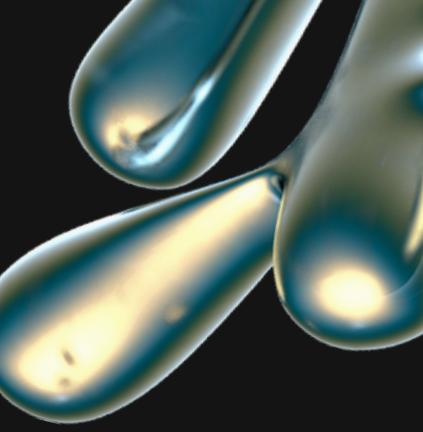
NON IDENTIFYING KEY

- Une relation ou les 2 tables non pas de relation parents enfants pour être identifier

IDENTYFING KEY

- Une ralton ou les tables on besoin d'une relation parent enfant pour être identifier

Encoder



LABEL ENCODER

Les valeurs seront directement
changée entre 0 et 1+
se qui peut créer une hiérarchie

ONE HOT ENCODER

Chaque valeur 0 , 1 , 2 seront séparé
dans une nouvelle colonne

Scaler

MINMAXSCALER

Les valeurs seront entre 1 et 0 ou -1,1 si il y a des valeurs négative dans le dataset

STANDARD SCALER

Les valeurs seront entre un min et un max

GridSearchCV - HyperParamètre



HYPER PARAMÉTRE

GRIDSEARCHCV

Algortyhme qui permet de trouvée les hyper paramètre idéal pour le modèle par rapport au donnée

- Valeurs utilisée pour controlée et améliorée l'entraînement
- n_estimators : Indication sur le nombre d'arbres dans la forets j'en utilise 200
- max_features : Démontre le nombre maximal de features possible pour chaque arbres
- max_depth : Le nombre de split possible pour chaque arbre permet de lutté contre l'overfit , underfit
- criterion : Permet de vérifier la qualité d'un split en utilisant Gini Impurity qui permet de savoir a quel fréquences des éléments du set sont indiquer incorrectement .

Regression / Classification

REGRESSION

- C'est des prédition qui sont utilisée pour des résultats continu .
- Le modèle sert également à comprendre les relations entre les features et les variable
- par exemple : Prédiction de prix ou prédition du temps

CLASSIFICATION

- C'est le prédition de plusieurs class
- Cela permet de classifier une prédition comme vue précédemment "Obtention d'un prêt"